Characterisation of a Stereo Matching and Object Location System
Report for BAe Systems

G. A. Buonaccorsi, A. J. Lacey, N. A. Thacker
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Stopford Building, Oxford Road
Manchester, M13 9PT

12 August 2003

Abstract

In this report we demonstrate and characterise a 3D wireframe model matcher, applied to a stereo object location problem typical of many industrial machine vision applications, and relate its design to the methodology we proposed as an earlier component of this project [42]. We begin by outlining the stereo requirements of a 3D model matcher, and in this context discuss the general characteristics of two approaches to solving the stereo correspondence problem: Feature Based Stereo (FBS) and Area Based Stereo (ABS). We show that FBS has significant advantages in terms of the accuracy of the recovered stereo data and of reduced susceptibility to image variations, but that ABS generally has a simpler flow of control and is therefore suitable for parallel processor or hardware implementations. We go on to describe the FBS and ABS algorithms implemented in our TINA vision system: PMF, a classical feature-based solution, and the Stretch Correlator (SC), a re-formulation of the algorithmic constraints of PMF into a feature driven, but fundamentally area-based, solution.

We present a theoretical analysis of the errors involved in stereo calculations, and use it to show how the results from both stereo algorithms can be characterised in terms of correspondence mismatch and 3D reconstruction error. This leads to the conclusion that geometric fitting to 3D data should be done in disparity space, to exploit the property of isotropic errors. We also summarise an analysis of the mismatch probability of a stereo correspondence algorithm, and show that mismatches are inevitable and must be accounted for in later algorithm stages.

We discuss the consequences of our analysis for the design of practical, modular 3D vision systems and describe the components of TINA relevant to this project. We then provide the results of evaluations of the performance of the TINA stereo algorithms, based on our analyses. In particular, we have found that SC has a mismatch probability of typically 1%-2% of the stereo data and in the worst case up to 8%.

Finally, we give a practical demonstration of TINA's 3D Wireframe Model Matcher, using both PMF and SC to reconstruct the 3D structure from stereo views of a brake shoe assembly and then locating a wireframe model of the assembly in the recovered 3D scenes. This is done with the object placed in a variety of orientations in a general scene with no special lighting. For similar object views, the number of features recovered by PMF and SC is shown to be identical, within statistical limits. For a set of 33 views of the object in different orientations, PMF provides matches in 18 cases and SC in 17 (rising

to 21 and 20 matches respectively when the Canny double thresholds are reduced). A breakdown into sub-groups with similar object orientations shows that the numbers of matches remain identical within statistical limits in each sub-group.

On the basis of these results, we conclude that our methodology can successfully be used to design and incrementally refine a modular vision system capable of stereo matching and 3D object location for “difficult” stereo reconstruction tasks. Moreover, the performance of our PMF and SC algorithms is to all intents and purposes identical, which given the computational simplicity of the latter may have positive implications for developing parallel or hardware implementations suitable for temporal stereo tasks.

1 Introduction

Computer vision may be defined as the process of extracting useful information from images with the goal of performing a specific task. This practical emphasis is an important part of the definition, and means that any evaluation of a computer vision algorithm should be framed in terms of the task for which the algorithm is required.

Part I of this project describes a methodology [42] which directly addresses this issue using quantitative statistical methods. This report now presents a practical example of the application of this methodology to computer vision problems In particular, we explain how the data generated by a stereo algorithm can be characterised, to aid in the design of further algorithms that make practical use of these data. This is a requirement of vision systems employing a modular philosophy, such as our TINA vision system.

TINA has been conceived and written as a development environment for research into machine vision algorithms. It comprises a set of algorithms gathered into self-consistent modules and presented using a graphical user interface that normally runs under UNIX with X-windows (X11), although it has been ported to other platforms.

With TINA’s philosophy of integration, algorithms are built up in stages, often from existing algorithm blocks. This allows researchers to build on earlier work in a structured manner. The results generated by each stage must have structural and statistical characteristics that match the assumptions used in the design of subsequent stages. The result is a comprehensive vision system that is currently being applied to real problems in industrial and medical computer vision settings.

We start this report by describing the general stereo requirements for 3D wireframe model matching, focusing the discussion on constraints, camera calibration and the characteristics of Feature Based (FBS) and Area Based (ABS) stereo reconstruction methods. Later, we present the TINA implementations of these two approaches — PMF and the Stretch Correlator (SC). We present a mathematical model of the errors that arise in stereo matching calculations, and with reference to a probablistic analysis of a corner matching algorithm we show that correspondence mismatches inevitably give rise to outlier errors in the recovered geometry. Based on these analyses, we propose that correctly recovered matches can be characterised (using errors estimated from feature location and stereo calibration) in a manner suitable for use in the construction of further processing modules. We conclude with a demonstration of 3D model matching and object location, using both of TINA’s stereo reconstruction algorithms to locate an object in a typical industrial-type matching problem.

2 Stereo reconstruction of 3D data

A number of textbooks give adequate descriptions of the theory and practice of stereo reconstruction of 3D data and of object location [17, 39, 44], therefore it is not our intention to reproduce such material in this report. However, the integrated philosophy of TINA has in some cases necessitated novel or non-standard approaches to certain elements of these problems and we will discuss these issues where it is helpful to do so.

In this section we address (from the perspective appropriate to TINA) stereo constraints, the general background to the different approaches for 3D reconstruction from stereo images, and stereo camera calibration.

2.1 Stereo correspondence and reconstruction

Any stereo system must solve two basic problems: which feature in the left image corresponds to which feature in the right? (correspondence) and what is the real-world location of the world feature that gave rise to the corresponding image feature? (reconstruction) We briefly consider this point.

Stereo applications take stereo pairs of images as input and in the process of reconstructing 3D data try to match primitives (points, blocks or features) in one image to corresponding primitives in the other[1]. Solving this correspondence problem requires a searching and matching technique, the robustness of which determines in turn the robustness of the final 3D result. This stereo matching process is a key component in any 3D Model Matching (3DMM) system, therefore the performance of the stereo reconstruction algorithm affects the performance of the whole system. Consequently, it is useful to discuss the general characteristics of these algorithms (Section 2.3.3), and to present details of their TINA implementations (Section 4).

Most stereo reconstruction methods use models based on the pinhole camera, and a parallel camera geometry[2] [9, 17, 39, 44]. With these models, given two points (one from the left image and one from the right image) identified by the stereo matching process as corresponding the depth can be estimated from the disparity, or the separation of the two points in image co-ordinates [17, 39]. Given image co-ordinates (x_L, y_L) in the left camera and (x_R, y_R) in the right, the 3D world location co-ordinates (X, Y, Z) can be written as:

$$X = \frac{x_L b}{d} \quad ; \quad Y = \frac{y_L b}{d} \quad ; \quad Z = \frac{fb}{d} \tag{1}$$

where $d = (x_L - x_R)$ is the disparity, b is the inter-ocular separation (camera baseline), and f is the focal length of the camera lenses (assumed to be the same for left and right cameras).

The accuracy of the reconstruction depends on the accuracy of the disparities and therefore on the accuracy with which the image plane co-ordinates can be determined. Section 3.1 discusses the impact of measurement errors on the stereo reconstruction process.

2.2 Stereo constraints

Solving the stereo correspondence problem is the major difficulty for most stereo applications. To enable stereo algorithms to address the correspondence problem in as robust a manner

[1] In this context, "corresponding" means originating from the same point in the 3D world space.

[2] Deviations from these models are unavoidable in real-world systems, but can be accounted for via camera calibration (which we discuss in Section 2.4).

as possible, several stereo constraints have been derived from the logical properties of the surfaces in the world view. While these are fairly standardised, it is worth summarising the TINA approach to their use.

2.2.1 Epipolar constraint

For any point in the world that is visible to both cameras in a stereo system, the epipolar plane passes through that point and the epipole (the line connecting the optical centres of the two cameras). It can be shown by elementary geometry [2] that the perspective projection of the point in each camera lies on the corresponding epipolar line (defined by the intersection of the epipolar plane with the image plane of that camera). This simplifies the correspondence problem because, for all practical stereo camera configurations, epipolar geometry can be exploited to reduce a 2D search of order N^2 (for an image of $N \times N$ pixels) to a 1D search of order N for each corresponding pair of points.

For parallel cameras the epipolar lines lie parallel to the image rasters, but parallel cameras are not used in practical systems. From Equation (1) the estimated depth, Z, is inversely proportional to the camera baseline separation, b, therefore the errors in the depth estimates are similarly inversely proportional to b. As larger baselines give less scene overlap between the cameras less depth can be computed if parallel cameras are used. The cameras in our demonstrator therefore have a convergent configuration (Section 6.1), which misaligns the epipolars and the image rasters.

Therefore, to use the epipolar constraint we must rectify either the image data or the extracted features from their original camera co-ordinates into a parallel camera co-ordinate system. In general this is always possible as the epipolar geometry is entirely defined by the relative positions of the two stereo cameras and can be recovered using a calibration process even without knowledge of the 3D camera geometry. We prefer to exploit the epipolar constraint via an explicit calibration stage, as we discuss in Section 2.4.

2.2.2 Disparity gradient limit

The disparity gradient (DG) between a pair of stereo points is the ratio of the difference in the two disparities to the separation of the points in the cyclopean camera space [9, 39]. The DG limit originated from the psychophysical observation on human stereo vision that objects are less likely to be successfully stereo fused when there are nearby objects at different depths from the observer - Burt [5] demonstrated that the DG for the human vision system has a critical value of 1, above which perceptual stereo fusion fails. Additional research [43] has shown that constraining the DG to be less than 2 ensures that no matches are made between points that would have to be located on a surface visible to only one eye (assuming surfaces to be opaque).

When using the DG constraint, the probability of a neighbouring match falling within the limit by chance increases almost linearly with its distance away from the match under consideration [32]. Also, large changes in disparity along the epipolar lines are more likely to result from objects at different depths. Therefore, the DG is typically calculated only for (neighbouring) features within a local pixel window, with the gradient scores being weighted by the reciprocal of their distance from the match being scored. Thus more remote pairs, which are more likely to satisfy the limit by chance, count for less.

2.2.3 Figural continuity

The figural continuity constraint uses continuous edge information to resolve ambiguous matches by preventing sections from the same edge in one image from matching to different, non-continuous edge sections in the opposite image [12, 25, 26]. Figural continuity is closely related to surface smoothness (as is the disparity gradient constraint). Surface contours are assumed to be spatially continuous, at the scale at which the match points are described [32]. Scenes may contain many discontinuities, providing the continuity of points comprising individual objects is maintained.

2.2.4 Ordering

Stereo projection almost always preserves the order of points from two images along matching epipolar lines [1, 32]. Applications using the disparity gradient limit DG ≤ 2 (such as TINA) gain the ordering constraint implicitly because if two elements on the same epipolar line become coincident in one of the stereo images, the disparity gradient becomes equal to 2 and for points that are reversed in order, the DG value exceeds 2 [8].

2.2.5 Uniqueness

The uniqueness constraint reflects the assumption that any point on a physical surface has a unique position in space at any one time. When uniqueness is enforced, matchable elements in each stereo image can be assigned at most one disparity value, corresponding to that unique physical position in the world [24, 32].

2.3 Feature-based vs area-based stereo

Algorithms designed to solve the stereo correspondence problem can basically be divided into two groups: area-based stereo (ABS) and feature-based stereo (FBS). We will briefly describe the approach of each group, and in keeping with our principle that any evaluation of a computer vision algorithm should be framed in terms of the task for which the algorithm is required (Section 1), we will discuss the general characteristics of each approach in the context of 3D object location in general and the TINA vision system in particular.

2.3.1 Feature based stereo (FBS)

Many areas of computer vision exploit a feature-based approach. The definition of a feature is arbitrary but in general a feature is a high level data structure that captures some of the information content in an image, indicating for example the position and orientation of part of an object. Since stereo vision involves the extraction of 3D data from stereo pairs of images of real-world scenes, the most useful features for stereo applications are those that describe the underlying 3D structure of the objects in the scene. In the main, and particularly in man-made environments, this structure is well-described using features such as edges and corners. For this reason, the extraction of edges and corners from raw image grey-level data is the focus of much FBS work. In some cases the extracted edges are linked into high-level data structures, often called edge-strings, and further processing may be done at the level of these higher data structures.

It is important to appreciate the fact that image features do not always correspond to world features in the way we might expect from the feature name. For example, "edges" or "corners" detected in an image *may* correspond exactly to edges or corners of an object as we would recognise them in the "real" world, but may also correspond to particular patterns of shading or to other artefactual sources. When using FBS, therefore, the precise definition of an image feature must be kept in mind when interpreting it as representing its real world counterpart.

FBS algorithms use feature extraction as a preprocessing stage because focusing the stereo matching process on distinctive image features has some significant advantages (discussed in Section 2.3.3). The most basic feature matching algorithms initially establish correspondences between features in the two stereo images using a local similarity metric based on feature properties (e.g. edge contrast and orientation) and the epipolar constraint. Then additional constraints such as disparity gradient, figural continuity and uniqueness (Section 2.2) can be used to resolve ambiguous matches. Algorithms using these approaches include [3] and the original PMF algorithm presented in [31, 32].

The result of applying an FBS stereo algorithm is a sparse map of depth data and a list of matching features. The advantages of this sparse depth map include reducing the correspondence search space[3], as discussed in Section 2.3.3, but it is worth noting that FBS can be particularly useful in man-made scenes because of the lack of image texture.

2.3.2 Area based stereo (ABS)

In ABS the primitives for the matching process are blocks of image data. Matching is based on the expectation that, given two views of the same scene observed from similar viewpoints, the image regions surrounding corresponding points will "look similar" and so will be identifiable using some block similarity metric. Image blocks may be rectangular or shaped and may have a fixed or variable size. The most basic ABS approaches directly match fixed $N \times M$ blocks of pixels between images by finding the block pair that maximises a chosen metric in the 2D search space.

Similarity metrics can be selected from a large number of formulae that can be used to calculate correlation scores. Metrics derived from probability density functions theoretically offer the best solution, but in practice approximations such as the Euclidean distance and dot-product metrics have been used for computational simplicity - these have proved adequate for most purposes.

Rectangular window ABS schemes require the epipolar lines to be coincident with the image rasters. In practice, this means that image rectification must be performed at an early stage in the stereo algorithm [21]. Fixed window techniques effectively assume front-o-parallel surfaces between the world scene and the stereo cameras, and so neglect the distortion effects caused by object rotation between views. Window-shaping ABS applies (usually affine) image warps to the matched image blocks to find the correct disparity estimates, and in TINA's implementation (Section 4.4) this allows us to align the orientations of edges within the blocks. Window-sizing algorithms vary the overall size of the correlation window [23, 27, 28], and must be used with care - if the size is too small, noise will dominate

[3]In computational terms, while FBS *does* reduce the correspondence search space it may incur a significant preprocessing overhead in extracting the features in the first place. However, features can be well-located (to sub-pixel accuracy) and their extraction is often worth this additional computational effort (Section 2.3.3).

the match metric [23]; if the size is too large this will unnecessarily constrain the surface structure of scene being recovered.

ABS algorithms that use correlation-based metrics perform matching on the basis of similar grey levels and are only applicable for stereo pairs of images in which equivalent pixel regions have approximately the same grey levels. Achieving such a situation requires distant light sources (i.e. at a distance indistinguishable from infinity) and minimal figural dissimilarity or distortion between the two views. This limits the most basic ABS algorithms [23, 27] to photogrammetry applications such as terrain mapping. To reduce this dependence on consistency of illumination, techniques such as image pre-processing using rank order filters [29, 19] have been used to focus the stereo algorithm on surface texture alone.

The usual result of applying an ABS algorithm is a dense map of depth data, approximating a continuous surface for a scene [13, 14, 16, 23, 27, 29, 30]. As we have noted, this is particularly useful for photogrammetry applications such as terrain mapping using images from satellites/aircraft [7], but it is less appropriate for many man-made scenes. This contributed to our decision to use a hybrid ABS/FBS approach in TINA's Stretch Correlator algorithm (Section 4.4).

2.3.3 FBS vs ABS

We proposed in Section 1 that any evaluation of a computer vision algorithm should be based on its suitability for the task in hand. Therefore, it is not possible in general to say whether ABS is "better" than FBS or vice versa. It is nevertheless instructive to make some general comparisons that are relevant to the TINA implementations of FBS and ABS algorithms presented in Section 4.

FBS focuses the stereo matching process on distinctive image features and so has the following advantages.

Data reduction. Edge and corner features normally cover just a few percent of the total image area (1–5% for edgels and $< 1\%$ for corners). Moreover, a man-made scene containing a few thousand edgels will typically give rise to a few tens or hundreds of individual line sections (edge strings). This significantly reduces the quantity of data to be analysed, and therefore reduces the search space for stereo matching.

Improved information. The skeletal 3D results of FBS provide a great deal of structural information for object identification, including information about the properties of individual features and about their spatial relationships with neighbouring features. Edges for example have properties such as contrast (the grey level gradient across the edge) and orientation, both of which can be used as parameters in a stereo match similarity metric. Edges exhibiting similar contrast and orientation properties can also be linked together to form edge strings, and higher level image primitives can then be formed by fitting geometric shapes such as straight lines and conic arcs to the edge string data [46]. Such higher level features are information-rich, but may require very specific analysis techniques.

Reduced susceptibility to illumination variations. Absolute grey-level similarity measures (typical of ABS) are unsuitable when the illumination of the objects differs appreciably between the two stereo views. FBS algorithms address this problem by using primitives that are more directly related to the underlying 3D structure of the scene than to the illumination. As a result, similarity metrics relying on feature properties or on higher level geometric primitives are less affected than ABS metrics by luminance changes and large perspective distortions between stereo images.

Improved accuracy and statistics. FBS representations exploit the flexibility and the good statistical properties inherent in edge and corner data. For example the location accuracy of 3D results is higher with FBS than with ABS, as feature detection algorithms such as the Canny edge detector [6] can locate features to sub-pixel accuracies. As a consequence, the reliability of FBS matches is nearly always better than that of ABS matches.

To achieve sub-pixel accuracy with ABS requires careful interpolation — ABS therefore has the additional problem of selecting the correct local interpolant function. In theory this requires us to know enough about the scene to select the correct function for each part of the image (i.e. we need to know what objects we are looking at). One implication of this fact is that dense stereo can only be reliably extracted from arbitrary scenes in the context of a larger vision system with sophisticated scene interpretation capabilities.

These advantages appear to weigh heavily in favour of FBS solutions, but ABS has its own advantages. Local consistency constraints are implicitly applied in ABS via the block correlation metric, and do not have to be explicitly included as with FBS. Moreover, the convolution and correlation operations required for ABS matching result in much simpler matching algorithms than FBS, where the match metrics depend on the precise feature representation and are often feature-type specific and rather complex. Their simplicity means that ABS algorithms are more accessible to parallel processor or hardware implementations, and this may result in speeds fast enough for multi-frame processing. Interest in ABS algorithms has therefore been sustained in the research community, and in developing the TINA Stretch Correlator we have tried to exploit the inherent advantages of both ABS and FBS approaches (Section 4.4).

Despite their apparent strengths and weaknessess, on a fundamental level we can summarise the quality of the data from FBS and ABS algorithms in the same way. Both are capable of returning some proportion of the useful information in a stereo pair of images, and both will make correspondence errors at some level. We will show in Section 3.1 that for FBS techniques we can compute the expected errors on recovered geometry using the technique of error propagation, and we will discuss the problems of quantifying the errors experienced using ABS methods.

2.4 Camera calibration and the epipolar constraint

The epipolar constraint reduces the stereo correspondence problem to a 1D search (Section 2.2.1) and is therefore one of the most fundamentally useful pieces of information that can be exploited during stereo matching. To make use of it, we need to know the epipolar geometry of our specific camera configuration and once we do so we can apply either full image rectification or direct rectification of extracted features (TINA uses both approaches, the former for SC and the latter for PMF, see Section 4).

Finding the epipolar geometry of a particular camera system requires some form of calibration. This can be done either with an explicit calibration stage or by extracting the calibration information from the scene images [11]. The rationale for the latter process (misleadingly termed “uncalibrated”stereo) is that it is possible to calibrate cameras from the data available during use rather than having to rely on special purpose calibration data — it is worth noting that an effective calibration is still being done.

In an explicit calibration process, there is immense freedom for the specification of the camera model, for the definition of the cost function and for the numerical implementation. There are a large number of published methods of which the most commonly referenced is probably [45]. However, there is a small number of basic principles that lead to an optimal calibration process (e.g. if the data for calibration is to be obtained from any automatic matching process then the cost function must be in the form of a robust statistic because outliers are guaranteed to be present [41]). Explicit calibration is robust and allows the intrinsic parameters of the individual cameras, such as lens distortions or an off-centre principal point, to be determined at the same time. Moreover, because the physical dimensions of the calibration object are known, absolute values can be derived for the cameras' geometric parameters and the 3D measurements produced by the system can be given absolute physical dimensions.

With "uncalibrated"stereo it is not possible to correct the intrinsic camera parameters to account for lens distortions or an off-centre principal point. Results cannot reliably be given absolute physical dimensions because, while dimensions can be based on some fixed, known dimension in the scene, their accuracy depends on a small number of 3D measurements and there is insufficient ground truth data to reliably determine the scale factor[4].

Moreover, uncalibrated methods are generally designed to work with static scenes [47]. While this report addresses only static scenes, we have designed the TINA system to work with unconstrained motion, for robotic tasks. Although it is true that the visual feed-back loop can be closed using projective geometry [4], we believe that much more can be achieved with good Euclidean geometry.

The advantages discussed above have convinced us that if an explicit camera calibration is possible there is so much to be gained that it ought to be done. For this reason, the stereo algorithms in TINA assume that explicit calibration has been performed, and indeed rely on it both to determine the epipolar geometry for stereo matching and to calculate 3D position from disparity.

3 Performance Characterisation

3.1 Errors in stereo calculations

The result of any solution to the stereo correspondence problem can be reduced to a set of paired points, where each pair comprises a point in one image and the corresponding point in the other. These paired points can be used to reconstruct depth information. This process is subject to measurement and calculation errors, which should therefore be modelled and accounted for in implementing any stereo algorithm - an issue discussed in our methodology document [42]. In this section we present the error modelling appropriate to the work of this report, and discuss the consequences of this modelling for the algorithms in TINA.

[4] In contrast, calibrated stereo uses many uncorrelated 3D measurements, combined with the corresponding ground truth data points, to arrive at a consensus calibration using error minimisation [40].

3.1.1 Errors in depth estimation

We saw in Section 2.1 that with the pinhole camera model in a parallel camera geometry, the 3D world co-ordinates (X, Y, Z) of a point in a scene are given by Equations 1. Following camera calibration, and working with rectified images, these equations can be held to be valid and so the depth, Z, of the scene point that corresponds to the image co-ordinates (x_L, y_L) in the left camera image plane and (x_R, y_R) in the right camera image plane can be written:

$$Z = \frac{fb}{x_L - x_R} = \frac{fb}{d}$$

where f is the common focal length of the left and right camera lenses, b is the inter-ocular separation (camera baseline), x_L and x_R are the positions of the features on the epipolars in each image, and d is the disparity.

Neglecting errors in f and b and assuming that x_L and x_R are independent of each other, we can determine the sensitivity of errors in Z to errors in disparity d:

$$\Delta Z^2 = \left(\frac{\partial Z}{\partial d}\Delta d\right)^2$$

and solving for ΔZ yields:

$$\Delta Z = \frac{Z^2 \Delta d}{fb} \tag{2}$$

As Δd arises from errors in x_L and x_R this result shows that the accuracy of depth estimation (and therefore of 3D reconstruction) depends on the accuracy of location of image features. The error in Z also has a strong (Z^2) dependence on depth, therefore depth estimates are much less reliable for distant objects.

3.1.2 Feature location errors in world co-ordinates

We have seen that FBS algorithms often use edge data to extract depth information. TINA's ABS algorithm (SC) also makes use of edge information in its matching process, and is therefore subject to the same types of error. For an individual edge, errors arise when calculating the locations of both the edge and the corresponding epipolar line(s) - we may therefore write:

$$\Delta x = \sqrt{\Delta x_f{}^2 + \Delta x_e{}^2}$$

where Δx_f is the error in edge location (typically 0.1 pixels) and Δx_e is the error due to the epipolar placement.

The first term is a statistical error while the second is a systematic error, arising from the calibration process (Section 3.1.4). The latter may be written [15]:

$$\Delta x_e = \frac{\Delta e}{\tan\theta} \tag{3}$$

where: Δe is the error in epipolar placement and θ is the internal angle between the tangent to the edge and the epipolar line.

Note that as $\theta \to 0$, $\Delta x_e \to \infty$ (and so, therefore, does Δx). This is a restatement of the well-known principle that depth cannot reliably be determined from near-horizontal (i.e. epipolar-aligned) lines (Section 3.1.4).

The propagation of disparity errors was previously described by ourselves in [15] and similarly only in Z by Krotkov [18]. The disparity errors for edge based stereo are based on the above observations and have been shown [15] to be:

$$\Delta d = \sqrt{2\Delta x_f{}^2 + \frac{2\Delta e^2}{\tan^2\theta}} \tag{4}$$

where all terms are as defined above.

Equation 4 may be substituted in Equation 2 to obtain ΔZ. As we saw in Equations 1, the world co-ordinates X and Y can be found using the equations:

$$X = \frac{bx}{d} \quad \text{and} \quad Y = \frac{by}{d}$$

where x and y are image co-ordinates. This allows us to determine the errors in X and Y as follows [15].

As x can be uniquely determined from x_L and x_R, we need only consider errors in these values. Assuming that $\Delta x_L = \Delta x_R = \Delta x$, the sensitivity of X with errors in x_L (left image x location) and x_R (right image x location) may be written:

$$\Delta X = \frac{Z\Delta x}{\sqrt{2}f}\left(\frac{X^2}{h^2} + 1\right)^{\frac{1}{2}} \tag{5}$$

A similar analysis for Y gives the result:

$$\Delta Y = \frac{Z}{\sqrt{2}f}\left(\frac{Y^2}{h^2}\Delta x^2 + \Delta y^2\right)^{\frac{1}{2}} \tag{6}$$

We can see from Equations 2, 5 and 6 that the errors in the world co-ordinate estimates are not uniform in the directions X, Y and Z. This is a problem for geometric fitting in the world co-ordinate space, and we address this problem using a transformation into disparity space.

3.1.3 Transformation into disparity space

To make valid statistical measurements using statistical methods which assume equal errors (such as simple least-squares), we need to have isotropic errors in our co-ordinate space. This can be achieved by transforming the world space (X, Y, Z) into disparity space (X', Y', Z').

$$X' = \frac{x}{\Delta x} \quad ; \quad Y' = \frac{y}{\Delta x} \quad ; \quad Z' = \frac{x_1 - x_2}{\sqrt{2}\Delta x} \tag{7}$$

By performing error sensitivity calculations in this new space, we see that the relative error is now unity in each direction [15]:

$$\Delta X' = \left(\frac{1}{\Delta x}\right).\Delta x = 1$$

$$\Delta Y' = \left(\frac{1}{\Delta y}\right).\Delta y = 1$$

$$\Delta Z' = \left[\left(\frac{\Delta x}{\sqrt{2}\Delta x} \right)^2 + \left(\frac{-\Delta x}{\sqrt{2}\Delta x} \right)^2 \right]^{\frac{1}{2}} = 1$$

Finally, we can describe disparity space in terms of the world co-ordinate system by substitution of the image co-ordinates:

$$X' = \frac{fX}{\Delta x Z} \quad ; \quad Y' = \frac{fY}{\Delta x Z} \quad ; \quad Z' = \frac{fI}{\sqrt{2}\Delta x Z} \tag{8}$$

In TINA, the fitting of geometric features to 3D scene data is done in disparity space, to exploit this property of isotropic errors.

3.1.4 Consequences of stereo error modelling

One well known limitation of stereo algorithms is the problem of matching straight edge features that are aligned with the epipolar planes as defined by the stereo camera geometry. For these epipolar-aligned edge features it can be difficult, if not impossible, to calculate accurate disparities, as any point on a line could match any point along its corresponding line in the other image. Most FBS algorithms, including those in TINA, choose to ignore edge features that are closely aligned to the epipolar plane (e.g. within ± 5 degrees) because disparity estimates become less accurate as lines approach epipolar alignment.

This behaviour is a direct consequence of the calibration error Δe (Equation 3). Although in our analysis we combined Δe directly with the statistical error on feature location, we may expect Δe to be systematic within any local region of the data, for example along a single linear feature. Appropriate use of the data therefore requires that we combine information from individual features on the basis of their statistical errors first, before estimating the systematic error on the position of the entire geometrical primitive. When considered along with the property of isotropic errors, this explains why transforming the data into disparity space (Equations 7) is appropriate for geometry fitting (Section 3.1.3).

The presence of outliers[5] and the difficulty of selecting an appropriate interpolant function mean that error modelling for ABS with a fixed regional constraint function cannot be done in the same way as for FBS. It is also true that the results of empirical evaluations of ABS will vary depending upon the test data set. ABS is fundamentally an interpolation process and we cannot know if we have the correct interpolation model for any given part of the scene; we can be sure, however, that a single interpolation process must be wrong for some surface structures. This is an important point if error estimates are needed for reliable use of the data in subsequent stages of processing within a vision system [41], and it provides further justification for the hybrid ABS/FBS approach we have adopted in TINA.

3.2 Modelling correspondence errors

The analysis of Section 3.1 illustrates the benefits of error modelling as proposed in our methodology document [42]. Not only did the stereo error model give a quantitative estimate of the accuracy of the derived world co-ordinates (and therefore the accuracy of stereo reconstruction), it also led us to conclude that geometric fitting should be done in disparity space to exploit the error isotropy. However, this implements only part of the methodology, which we address further in this section.

[5] Due to incorrect matches, specularities, illumination artefacts or smooth occluding boundaries.

We may consider the stereo correspondence problem as identifying from the right image a set of candidate matches for features (or regions) in the left image and selecting the best of these on the basis of some similarity measure, while enforcing consistency with the set of matches available for the right image. Viewed this way, the robustness of a stereo algorithm can be evaluated in terms of the mismatch probability, which must be minimised. This issue was also addressed in our methodology document [42], where we proposed that it is possible to predict the expected success of matching based upon the shapes of the probability distributions of the similarity measures for correct and incorrect matches.

To implement this part of the methodology, we have analysed the performance of a corner matching algorithm by building a simple statistical model of the matching process, using cross-correlation as our match strength measure and using theoretical approximations to the relevant probability distributions [41]. This analysis showed that the probability of an incorrect stereo match is expected to vary proportionally with the area searched during matching. Moreover, the proportion of mismatches to good matches was found to reduce when applying a strong uniqueness criterion and increasing the minimum threshold for cross-correlation. However, it was also shown that when correlation match scores were used to generate "signal" (correct match) and "noise" (mismatch) distributions, it is never possible to guarantee non-overlapping match distributions and so it is not possible to define a combination of threshold values that will exclude the possibility of mismatches [21].

A major conclusion of this work is that all practical stereo algorithms will inevitably produce correspondence mismatches. In a modular vision system, such as TINA, subsequent algorithms must therefore deal with the possibility of mismatches.

4 The TINA vision system

TINA is our modular computer vision system. Its design and implementation has taken into account the principles expressed in our methodology document [42], and the results of our statistical modelling (Section 3). TINA therefore serves as a practical demonstration of our methodology, as well as a working computer vision system. It incorporates both ABS and FBS approaches to stereo reconstruction (Section 2.3) and may use either for 3D object location.

In this section we provide a brief description of the TINA components that are relevant to this report.

4.1 General description of the TINA 3D model matcher

The TINA 3D model matching (3DMM) system, as used for the work of this report, is illustrated in Figure 1 and summarised in Table 1. The letters in the following descriptions (and in Table 1) identify the relevant components of Figure 1.

The original version of the TINA 3DMM system was presented in [35] and [37]. It used a sparse edge based (FBS) depth map (A, B) extracted from stereo pairs of images, together with the corresponding camera calibration information (E). A geometric interpretation of the scene (C) was constructed by fitting lines and arcs to the depth map data. Statistical matching (D) of 3D scene descriptions to a stored wireframe model enabled the model to be located within the scene.

Since the original papers were published, the system has been substantially extended and improved. Of particular interest in the context of this report is the addition of the Stretch Cor-

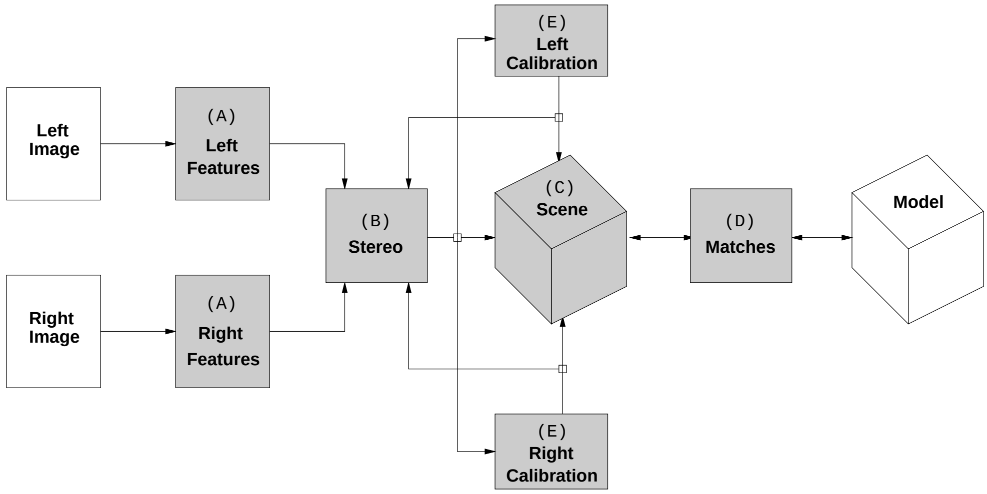


Figure 1: Block diagram of the basic 3D object location system in TINA

relator (SC), an ABS algorithm for stereo reconstruction, but other additions to the original system include the full calibration method discussed in Section 2.4 and improved geometric reconstruction. Table 1 lists the current implementations of the relevant system components.

4.2 Calibration

TINA's calibration method defines its cost function in the image plane as the errors between back projected positions for points[6]. The camera parameters are initially estimated in a first pass based on Tsai [45], using a grid-based calibration tile. This can be done once prior to use of the system. We then apply an iterative robust optimisation (with optimal combination of covariances via the use of a regularisation term) to minimise epipolar errors during use. TINA is able to track an existing calibration, integrating statistical information from known epipolar errors, robot motion, and known object targets [40]. We can therefore avoid the need to perform "calibration object" specific recalibration during use of the stereo vision system.

4.3 PMF

PMF [31] was the original depth from stereo algorithm developed within TINA and has been cited many times as a successful feature based algorithm. The current TINA implementation is a later version of PMF [34] which employs the constraints presented in Section 2.2 and includes matching support at the edge string level.

PMF uses Canny edge information [6] extracted from both images with sub-pixel accuracy (0.1 pixels by experiment). The extracted edge maps are rectified (using the camera calibration information) to allow the epipolar constraint to be applied. The edge data is processed to produce a list of connected edge elements (edge strings), which are defined between epipolar tangencies. Figural continuity is applied by accumulating matches along the length of a string to give a string match score. A high match score is generated when edges

[6]This is the only domain in which measurement errors can be expected to be uniform so that inappropriate statistical error estimates are not used during the calibration process.

Key	Purpose	Algorithms	Assumptions
A	Edge & Corner Detection 2D Geometry Fitting	Canny [6] Bias Corrected Kalman Filter [36] Pollard [33]	Edges present in expected locations Curves and lines can be correctly linked and fitted
B	Stereo Matching 3D Geometry Fitting	PMF [32, 34] Stretch Correlation [20] GDB	Accurate epi-polar geometry and match metrics Improved performance 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	Thacker [40]	Updatable on-line calibration

Table 1: Algorithms and assumptions used in the TINA 3DMM with key for shaded components in Figure 1

from one edge string match edges from the corresponding edge string in the opposite image. Disambiguation is further enhanced using edge contrast and orientation consistency, and the disparity gradient constraint is used to reject edges failing the disparity gradient limit within a 7 pixel radius. As discussed in Section 2.2, the application of the disparity gradient constraint ensures that the ordering constraint is also fulfilled.

4.4 Stretch correlation

In Section 2.3.3 we saw that FBS algorithms (e.g. PMF) are more accurate and robust for "difficult" stereo problems, but that ABS generally has a simpler flow of control and can be more efficiently implemented for real time tasks such as robotic control. While in TINA we wanted to exploit the computational efficiency of ABS, the advantages of FBS are not easily thrown away so we have developed the technique of Stretch Correlation (SC) [20, 21, 22], a hybrid approach that can improve computation efficiency while retaining the accuracy and robustness of PMF.

The SC algorithm [20] has the basic characteristics of ABS, using a dot product correlation to generate match hypotheses for discrete blocks in the left and right images. However, the matching process is focused on blocks containing non-horizontal edges, which reduces the photometric dependence of conventional correlation matching algorithms and improves accuracy and robustness. Stretching or shearing the right image blocks counters the assumption of front-o-parallel surfaces and allows SC to model the surface rotations due to the different views in the left and right cameras.

Canny edge detection is performed after correspondence matching in the rectified stereo images. The resulting edge locations are matched using the SC correspondence results so that the data is available as connected edge strings, in the same way as when using PMF. This allows us to interchange the use of these algorithms within TINA.

4.5 Geometry reconstruction

The list of left image edge strings is processed to establish a feature based (lines and ellipses) geometric representation of the scene [36]. The matched right image edge strings are then projected onto the left image geometric features and fitting performed in disparity space (see Section 3.1.4) to produce a 3D geometric scene representation. This geometric interpretation of the data is passed to the Scene Model Matcher (SMM), which makes a sub-optimal assumption regarding the errors on the geometry to generate a rapid closed form solution to the object location problem. The result of this process is a transformation of the object model into the world co-ordinates, and this is the endpoint of the object location experiments presented in this report (Section 6.3).

TINA, however, does allow further refinements to be made. The initial object location estimate is passed to a closed loop validation routine that refines the derived location on the basis of a robust likelihood measure. This process is described in a further report for this project, discussing in particular the relationship of TINA's methodology to approaches such as that used in BAe Systems' TALEOS.

5 Testing vision algorithms

As with any development activity, testing is essential when developing computer vision algorithms. In this field, however, testing is made particularly difficult by a number of factors:

- large numbers of test images are required to obtain statistically meaningful performance measures[7];
- it is very difficult to select test images to ensure that they are representative;
- assumptions may be valid in one application domain but not in another[8].

As we describe in our methodology document [42], we believe in a test methodology based on a theoretical framework in which the statistical data distributions that affect algorithm performance are identified and evaluation is performed by modelling the algorithm. While this is a significant undertaking, after such a method has been developed for a given algorithm, the algorithm can be rapidly re-evaluated for any new image data set. This re-usability makes it easier to evaluate incremental changes to algorithms and the combinatorial advantage obtained by working directly with data distributions may considerably reduce the required number of test images.

We have implemented such test methods within the TINA framework. This work has been published elsewhere (e.g. [8, 20, 41]) so in this section we will provide only a short overview and the conclusions relevant to this report.

5.1 Test data sets

The most robust test for any vision system is to directly compare its output to ground truth data. However, this approach may be problematic, as ground truth data is extremely difficult to generate accurately. The are several approaches to the solution of this problem. One

[7] Partly because a large number of failure cases is needed to accurately estimate reliability.

[8] This makes it very difficult to relate the results of one evaluation to other users' requirements.

particularly elegant technique involves aquiring not two but three "stereo" images at a time and confirming the reconstructed geometry from two pairings of images in the common image coordinate system [10]. Such data can be synthesised from temporal sequences of static scenes. Alternatively, we can attempt to reconstruct a well measured object which can be aligned with the image data via a projection optimisation strategy. As part of our strategy for evaluating our vision algortihms for accuracy and mismatch probability, we have developed two test image sets with verifiable ground truth data [8]. The first is a set of real test images of a planar calibration tile, the second is a set of synthetic images generated using 3D ray tracing software, which supports testing with more complex scenes. The 3D data produced by the stereo algorithms is compared to the ground truth data to estimate accuracy by counting mismatched features (for the reasons discussed in Section 3.2). This process involves the use of stereo error estimates, and therefore also provides a means for confirming the error model presented earlier. In this section we describe the test methods used with these data sets and in the next we briefly summarise the results of testing our SC algorithm.

5.1.1 Planar grid testing

Ground truth data is relatively easy to generate for simple scenes - we have done so for stereo images of a planar grid of black squares on a white background [8]. With this “planar” stereo data, mismatches result in outliers that lie off-plane in both 3D and disparity space and are therefore easily identified.

In the test process, we first use the stereo algorithm to generate a set of matched point features. Next, we calculate the surface normal angles of a set of planes defined by pseudo-randomly combining groups of three points. The plane orientation of the test grid is estimated by interpolating the maximum from a histogram of these surface normal angles. Because this is a critical part of the process, the stereo correspondence is manually guided to avoid forming a new plane at a different depth to the correct one[9].

Having recovered the plane orientation, mismatched points are identified by pseudo-randomly combining each point with several other points from the 3D result. Those points with surface normal angles consistently more than 4 standard deviations from the recovered plane orientation[10] are labelled as outliers.

5.1.2 Testing using artifical scene data

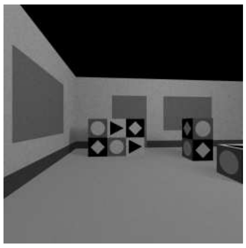
To model more complex scenes, we used the POV-Ray [38] ray-tracing package to define test image sets with precise control over object size, shape, texture, image noise and lighting [8]. This process also gave us precise control over the simulated stereo camera parameters. The 3D scene description files provided ground truth data. Example scenes are shown in Figures 2(a) and 2(b).

In this case, each point in the stereo result is evaluated by calculating its theoretical stereo accuracy, σ, based on its source feature properties and the current stereo camera configuration [8]. The residuals on resulting depth when plotted as $\Delta Z/Z^2$ (Figure 2(d)) were found to have a Normal distribution as expected [11], with a width given by the scaling terms in the

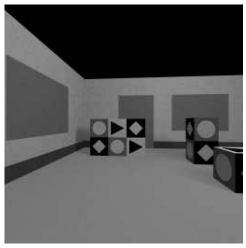
[9] This is possible because of the “wallpaper effect” [8].

[10] The standard deviation is also recovered from the histogram of surface normal angles.

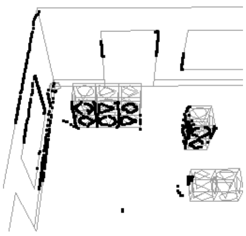
[11] With the exception of minor errors introduced in edge location due to interpolation errors and interations bewteen closely placed edge features.



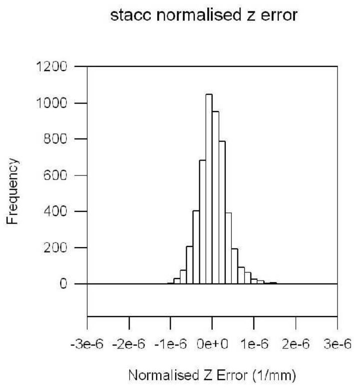
(a) left image



(b) right image



(c) projected 3D result



(d) Residuals scaled by $1/Z^2$

Figure 2: Artificially rendered cube-room scene images with 3D result

theory and an effective edge location error of 0.034 pixels. mismatches were found to lie in the extreme tails of this distribution [8]. If the point in 3D space is consistent with the geometric scene description to within a distance of 4σ, then it may be accepted as a good stereo match, otherwise it must be considered as an outlier.

5.2 Testing TINA's Stretch Correlator

We have applied our test methodology [42] to TINA's SC algorithm, and referred to these tests in various parts of this report. It is appropriate to gather together the relevant conclusions.

In Section 3.1.4 we concluded that our (feature-based) error model of Section 3.1 is appropriate to SC because it is not a pure ABS algorithm but takes a hybrid approach, and so is susceptible to the same types of stereo reconstruction errors.

In Section 3.2 we discussed our theoretical probabilistic analysis of a corner-matching algorithm [41], concluding that mismatch probability is proportional to the search area, and that it may be reduced by applying a strong uniqueness criterion and increasing the minimum threshold for cross-correlation. Experimental studies using SC [20, 21] confirmed that the mismatch probability is proportional to the search area by generating "signal" (correct match) and "noise" (mismatch) distributions for stereo image pairs accompanied by ground truth data, for several disparity search ranges. We also used these studies to find the most robust warp correlation technique by optimising the algoritm parameters.

We have used the test image sets of Section 5.1 to evaluate the performance of SC with regard to feature mismatches [8], using the techniques described to identify the outliers in the 3D data, by comparison to the ground truth data for these image sets. We found that, typically, SC could be expected to mismatch between 1%-2% of the data and in the worst case up to 8%. Figure 2(c) shows a 3-D result from this study.

6 An experiment in 3D object location

Thus far, we have shown how part of our TINA vision system has been designed, built and tested based on our proposed methodology [42]. We now present a practical demonstration of the results achievable when using this system to perform a "difficult" stereo matching and 3D object location task in a setting typical of many industrial computer vision applications.

For this purpose, we assembled a stereo camera system to an appropriate specification [Appendix B] and used it to collect a set of stereo pairs of images of a car brake shoe assembly in a representative range of orientations with no special lighting. We used both PMF and SC to recover the 3D structure from the images, applied TINA's geometric fitting and then used TINA's 3DMM to locate the object in the reconstructed 3D scenes, based on a wireframe model of the brake shoe assembly.

This experiment shows both the capabilities and the limitations of TINA's 3DMM for this task. We conclude by proposing developments that would improve the system performance by addressing these limitations.

6.1 Camera system

As part of this project, we specified and assembled a camera system appropriate for the object location task. The system comprises the hardware (cameras and frame grabber) and controller software required to collect stereo pairs of images and transfer them to a computer system.

We use a *Leutron PicPort Stereo* frame grabber coupled to two *Hitachi KFP1E* cameras. The cameras have CCD arrays with square pixels, they run in non-interlaced mode and they have full frame shutters[12]. It is worth noting that the latter two characteristics are essential for any system intended to analyse moving objects, which is one of our goals for TINA (see Section 6.4). The *PicPort Stereo* can synchronise capture from both cameras simultaneously (the frame grabber has two A/D converters and so can digitise the two data channels concurrently at a frame rate of 25 fps), and uses DMA transfer to move the data to the computer system.

This is an adequate system for accurate stereo image analysis. However, with an estimated 1-2 bits of noise, the 8-bit sensitivity is a minimum specification. Moreover, we cannot use pixel clocking, which would stabilise jitter noise to less than ± 1pixel. Nevertheless the performance is sufficient for most applications, and more than adequte for the static object location task presented in this demonstration.

The camera system was calibrated at the start of the experiment and again at the end (to ensure there had been no calibration drift), using the method described in Section 2.4.

[12]The term "progressive scan" is sometimes used to imply full frame shuttering but this does not necessarily mean odd/even fields are captured simultaneously.

6.2 Wireframe model

One of the inputs to the TINA 3DMM is a wireframe model representation of the real-world object (labelled "Model" in Figure 1). The model is built up using simple geometric feature elements (straight lines and ellipses) and it comprises two ASCII text files:

- a "model file" (`*.poly`) listing the feature elements with their type, dimensions, locations and orientations;
- a "clique file"(`*.ffg, i.e. "focus feature group"`) grouping the feature elements that are visible from each of a set of viewpoints and orientations.

These files are generated by hand for each object to be modelled. We provide further details of file structure and summarize how they are generated in Appendix A. The model may be optimised by iterative refinement, but within the scope of this study it was not necessary to do so, as our main aim was to demonstrate the practical application of our methodology [42]. It may therefore be possible to obtain better object location results with a fully optimised model.

6.3 Results of the 3D model-matching experiments

This section presents the results of the 3DMM and object location experiments. First, we show examples of the inputs to the 3DMM system and representative outputs, then we give a summary of the matching and object location results and discuss the outcomes of the experiments.

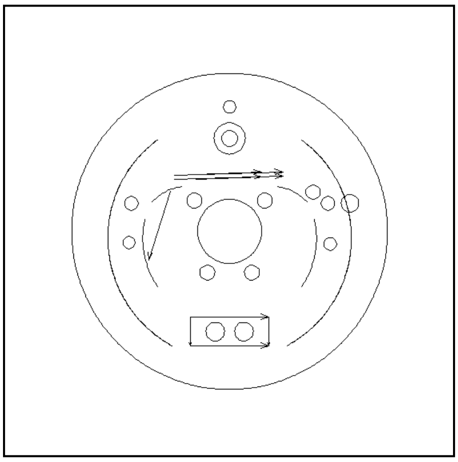
6.3.1 3DMM inputs: images and the wireframe model

The inputs to the TINA 3DMM system are shown as unshaded boxes in Figure 1. For a given run of the object location experiment, these are a stereo pair of images and the wireframe model. Figure 3 shows examples of the inputs used for this experiment - we show left camera images of the brake shoe assembly and views of the wireframe model in three orientations, to show the 3D structure.

The front views (Figures 3(a) and 3(b)) show that most of the visually obvious features of the brake shoe in this orientation have been captured in the model. The side views (Figures 3(c) and 3(d)) show one of the problems with TINA's wireframe models - the external black shell of the brake shoe assembly has curved occluding surfaces that are obvious to the eye but cannot be included in the model. The rear view of the object (Figure 3(e)) has little detail, and also has curved occluding surfaces. As the model is a wireframe, its rear view is simply a mirror image of the front, so we have shown an angled view (Figure 3(f)) to give a clearer impression of the 3D structure.

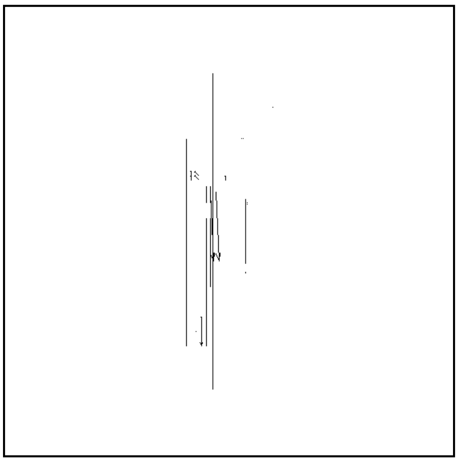
Note that certain features (e.g., the three circles near the top of the model in Figure 3(b)) are not visible from the front of the object, and most are not visible from the rear. As we discuss in Section 6.2 and Appendix A, we address this problem using the clique files that represent different object orientations and viewpoints.

(a) Front view of brake shoe object



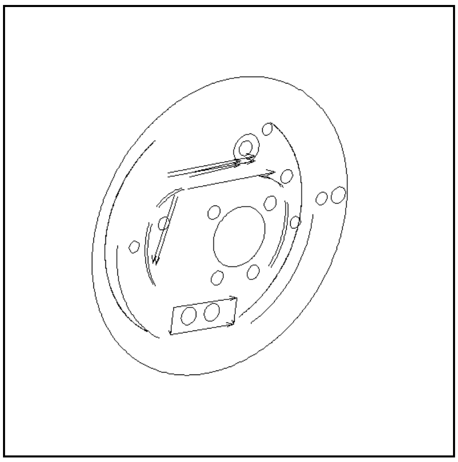
(b) Front view of brake shoe model

(c) Side view of brake shoe object



(d) Side view of brake shoe model

(e) Rear view of brake shoe object



(f) Angled view of brake shoe model

Figure 3: Views of the inputs to the 3DMM - (a), (c) and (e) show left camera images of the brake shoe object while (b), (d) and (f) show views of the wireframe model

	Number of features present						
		PMF			SC		
View	Model	(Good)	(Bad)	(Eye)	(Good)	(Bad)	(Eye)
front	24	14	4	20	16	1	20
rear	12	7	2	9	8	3	10

Table 2: Numbers of features identified for the example stereo outputs of Figure 3. (*Good* $\Rightarrow$ stereo match, *Bad* $\Rightarrow$ stereo mismatch - both as identified by eye; *Eye* $\Rightarrow$ visual match)

6.3.2 Stereo outputs: 3D scene geometry

Figure 4 shows examples of the stereo data outputs from both PMF and SC[13]. The front (Figures 4(a) and 4(b)) and rear (Figures 4(e) and 4(f)) views correspond to the front and rear views of Figure 3; the side (Figures 4(c) and 4(d)) views, however, do not correspond to the side views of Figure 3 - these show the stereo data reconstructed from the front views, to illustrate the 3D nature of the reconstruction process.

For the front and rear views of the input images and the model, all of the important model features are clearly visible by eye in the input images (Figure 3). It is unsurprising, therefore, that from these favourable viewing angles most of the features are also present in the reconstructed stereo data for both algorithms (Figure 4). We can make the qualitative assessment that the stereo data is of comparable quality from both algorithms, and this is borne out by the numbers of recovered features (Table 2), which are very similar for these views.

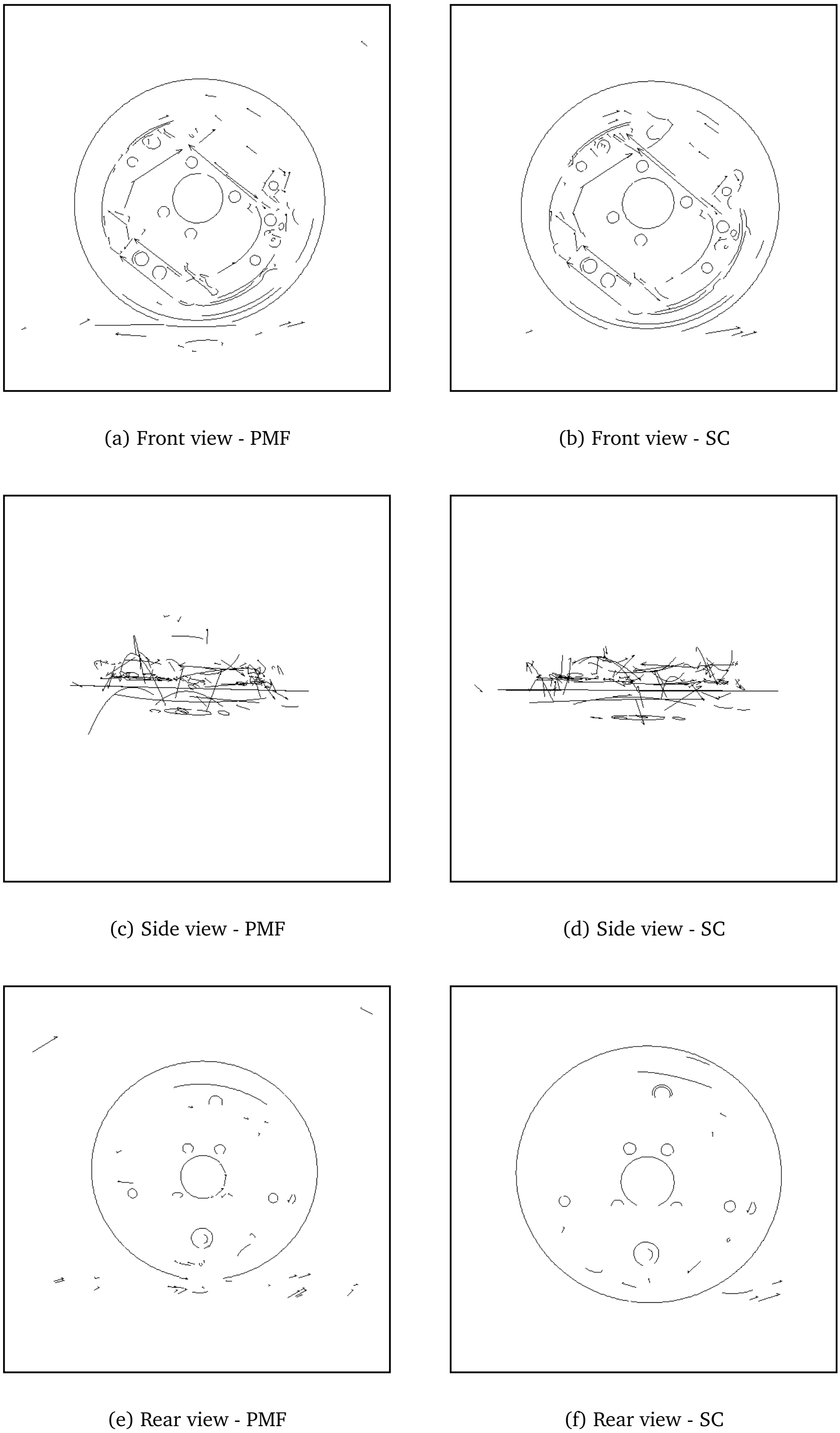
In Table 2, the column headed *Model* gives the number of model features that should be visible in each orientation. The columns headed *Good* give the numbers of matches returned by the stereo algorithms that were judged by eye to be correct (i.e. the matched feature was judged by eye to correspond to the actual object feature), and those headed *Bad* give the numbers of matches returned by the stereo algorithms that by eye were judged to be incorrect. The columns headed *Eye* give the numbers of features that were detectable by eye in the images. The numbers of identified features match closely for each algorithm in each column. While there may appear to be a slight tendency for SC to give better results, this would not be borne out by statistical analysis - there is no statistically significant difference in these results.

Although it is not particularly evident from these viewpoints, the 3D structure recovered using SC is generally cleaner (i.e. has fewer artefactual features) than that recovered using PMF. In Figure 4 the side views give some support to this notion, with more out-of-plane features in the PMF data than the SC data. However, when the object is more clearly tilted or rotated PMF often gives rise to out-of-plane ellipses (or arcs) and this is rarely true of SC (data not shown).

6.3.3 3DMM outputs: numbers and quality of matches

Table 3 summarises (in terms of successful matches) the results of applying TINA's 3DMM to the stereo data recovered from stereo images of the brake shoe assembly in a variety of

[13]The data correspond to rows 2 and 10 in Table 4.



(a) Front view - PMF

(b) Front view - SC

(c) Side view - PMF

(d) Side view - SC

(e) Rear view - PMF

(f) Rear view - SC

Figure 4: Views of the stereo outputs of PMF (a, c and e) and SC (b, d and f) for the object and model of Figure 3. Note that the side views were reconstructed from the front view images

Stereo Algorithm	Default thresholds		Reduced thresholds	
	Matches	Transforms	Matches	Transforms
PMF	18	17	21	19
SC	17	16	20	18

Table 3: Overall numbers of 3DMM matches for the brake shoe assembly, for a total of 33 object orientations (summarised from Table 4)

orientations, with no special lighting. Using the default algorithm parameters, good matches were obtained in 18/33 object orientations (55%) with PMF and 17/33 (52%) with FSC - these results are identical within statistical limits. The number of good model transforms (i.e. accurate object locations) is also identical within statistical limits. This is as expected, given the comparable quality of the stereo data (Section 6.3.2).

It is interesting to note that reducing the values used for double thresholding in TINA's Canny routine increases the number of successful matches to 21/33 (64%) with PMF and 20/33 (61%) with FSC (Table 3). This implies that the fixed Canny thresholds normally used in TINA are perhaps inappropriate. We will return to this point in Section 6.4.

Table 4 presents a more detailed breakdown of the 3DMM results, taking into consideration the different object orientations used in the experiment. The orientations Upright, SideOn and Reverse correspond to the front, side and rear views of Figure 3. Tilted and TiltRvs correspond to an approximately 45 degree tilt of the object plane from the vertical, viewed from the front and rear respectively, while in Flat and FlatRvs the brake shoe is horizontal, with respectively the front or rear surface visible (upwards). The column Twist refers to a rotation about the vertical axis, e.g. Figure 3(c) $\Rightarrow$ row 13 (SideOn) $\Rightarrow$ Twist is 90 degrees. The column Rotation refers to a rotation about the centre of the brake shoe, e.g. Figure 3(a) $\Rightarrow$ row 2 (Upright) $\Rightarrow$ Rotation is 45 degrees.

In Table 4, the numbers of candidate matches are shown for each object orientation, along with the position in the candidate match list of the first good model transform. For the upright front and rear views (i.e. those illustrated in Figure 3), large numbers of candidate matches were generated[14]. The much smaller number of candidate matches for the rear view reflects the smaller number of permutations available for the smaller number of matched features.

The responses within each orientation sub-group are comparable for PMF and SC. In favourable orientations (rows 1-12, positions Upright and Reverse, object plane near-parallel to the image plane) both algorithms give 10/12 matches for the default parameters. When the Canny thresholds are reduced, this improves to 12/12.

Even with a significant tilt of the object plane away from the parallel (rows 1-12, positions Tilted and TiltRvs), SC gets 6/9 and PMF 7/9 matches. Only when the object is in the least favourable orientations (rows 13-14, Side On and rows 24-33, FlatUp and FlatDn) does the matching process consistently fail, when each algorithm has only 1/12 matches (rising to 2/12 with reduced Canny thresholds).

These results reflect the fact that, as we move away from the favourable object orientations many features are no longer visible in the images. Nevertheless, these results show that in most cases matching is still possible, and only when the object is side on or flat and face

[14]The exact number of matches depends on the details of the model, and will alter if features or cliques are added or deleted (Appendix A).

ID No.	Orientation	Twist	Rotation	SC No. Matches	SC 1st Transform	PMF No. Matches	PMF 1st Transform
1	Upright	none	none	155	4	136	1
2	Upright	none	+45	1183	1	1604	1
3	Upright	none	+90	1445	2	1430	5
4	Upright	+45	none	x (34)	x (1)	x (13)	x (2)
5	Upright	+45	+45	575	1	605	1
6	Upright	+45	+90	294	8	592	1
7	Upright	-45	none	832	1	634	1
8	Upright	-45	+45	19	1	33	4
9	Upright	-45	+90	x (48)	x (6)	827	1
10	Reverse	180	none	105	1	57	2
11	Reverse	225	none	48	2	x (36)	x (x)
12	Reverse	135	none	85	1	60	1
13	SideOn	+90	none	x	x	x	x
14	SideOn	-90	none	x	x	x	x
15	Tilted	none	none	826	1	556	1
16	Tilted	none	+90	888	1	1732	2
17	Tilted	+45	none	x	x	x	x
18	Tilted	+45	+90	x	x	58	42
19	Tilted	-45	none	402	1	239	1
20	Tilted	-45	+90	698	3	1316	1
21	TiltRvs	180	none	353	14	254	11
22	TiltRvs	225	none	64	4	40	4
23	TiltRvs	135	none	x	x	x	x
24	Flat	none	none	x	x	x	x
25	Flat	none	+45	x (16)	x (x)	192	x
26	Flat	none	+90	x	x	x	x
27	Flat	none	+135	48	x	x (4)	x (2)
28	Flat	none	+180	x	x	x	x
29	FlatRvs	none	none	x	x	x	x
30	FlatRvs	none	+45	x	x	x	x
31	FlatRvs	none	+90	x	x	x	x
32	FlatRvs	none	+135	x	x	x	x
33	FlatRvs	none	+180	x	x	x	x

Table 4: Number of candidate matches and place in match list of first good transform for 33 orientations of the brake shoe object, using both PMF and SC (x $\Rightarrow$ no match or poor transform)

down (when very few features are visible) does matching prove impossible. These results may have been improved had the model been optimised, but it is still likely that in certain orientations no match would have been possible.

6.4 Discussion of results

These results show that TINA's stereo and 3DMM systems can be used along with our demonstrator hardware (Section 6.1) to identify and locate an object in a "difficult" 3D setting. In favourable orientations, we achieved matching and model location in better than 80% of cases and with only minor adjustments this was improved to 100% of the cases tested. While in the most unfavourable object orientations matching proved very difficult, we should keep in mind that, within the scope of this study, we did not optimise the wireframe model, nor did we optimise the parameter settings for the stereo algorithms. It is reasonable to assume that had we done so, our results would have improved even for these unfavourable object orientations. Nevertheless, it is likely that matching could not be achieved in all cases even with optimisation, and there is clear scope to improve the system.

For example, the improvement in matching results observed when making a small reduction in the Canny double thresholds indicates a potential system improvement. The fixed double threshold values used with TINA's Canny implementation were chosen to work with the worst expected signal to noise ratio. However, this ratio will vary with the illumination conditions, and for many cases the fixed threshold will be too high. It is possible within TINA to make a global estimate of the noise present in each image - such an estimate could be used to set the Canny double thresholds at a level appropriate to each image, which would improve matching in the same way as we achieved by reducing the thresholds manually in this experiment.

We must note, however, that it is important to be conservative in interpreting the numerical data of Table 4. These data do not constitute a solid, statistically valid, quantitative evaluation of the two stereo algorithms, but rather a broad qualitative comparison. To be more specific or quantitative it would be necessary to invest a significant research effort into the optimisation of the model and of each stereo method and into devising a full probablistic analysis (along the lines described in Section 3.2 and in [41]) for this particular experimental configuration. As discussed in Section 6.2, this was outwith the scope of this project.

Overall, we may conclude from this experiment that there is no significant difference between the stereo and 3DMM results obtained using PMF or SC. This is a very positive result, as part of the motivation for developing SC was to facilitate the implementation of a simpler stereo correspondence algorithm on parallel processors or in hardware. The fact that the much simpler computational model does not reduce matching or object location performance is very encouraging, and suggests that SC has a viable future as a replacement for pure FBS reconstruction.

7 Conclusions

This report summarises part of a large body of work carried out to design, build and incrementally refine a modular stereo vision system for a robotic vision application. Through our experience of carrying out this work, we have arrived at the conclusions that knowledge of the errors present in derived results is essential for the correct subsequent use of these results in successive components of a modular vision system, and that probabilistic models of algorithmic performance are needed to fully and properly evaluate the methods employed. These issues were discussed in general terms in the methodology document submitted as Part I of this project [42].

In this report, we first considered the relative merits of feature-based (FBS) and area-based (ABS) algorithms for the stereo correspondence problem (Section 2.3). We saw that FBS has significant advantages in terms of the accuracy of the recovered stereo data and its reduced susceptibility to image variations, but that ABS generally has a simpler flow of control and is therefore suitable for parallel processor or hardware implementations. We described how this motivated our development of Stretch Correlation (SC), our hybrid ABS/FBS stereo algorithm, which we described more fully in Section 4.4.

We presented an analysis of the errors involved in stereo calculations, and showed that we need to consider both statistical errors and systematic errors that arise from the calibration process (Sections 3.1.2 and 3.1.4). Our error model also led us to the conclusion that errors are not isotropic in the recovered world space, and for geometric fitting to 3D data it is best to transform to disparity space, where the errors *are* isotropic, using Equations 7.

The model also showed that once a FBS approach has been adopted, the spatial accuracy of the algorithm is largely fixed and can be treated independently from the stereo correspondence algorithm. The largest remaining practical problem is therefore the reduction of the numbers of outliers resulting from poor correspondence matching. We summarised the results of both theoretical and experimental studies (based on our quantitative probabilistic methodology [42]) carried out in an attempt to quantify the statistical limits of correspondence algorithms in terms of the mismatch probability. We were able to conclude that mismatches are inevitable in recovered stereo data. The quantity of these mismatches will vary according to scene contents, any subsequent algorithm using that data must therefore account for the presence of variable quantities of mismatches.

As we have always maintained the goal of using the data generated from our algorithms for the practical purpose of 3D object location for robot manipulation, in this project we have assembled a demonstrator to illustrate the capability of our TINA vision system for this task (Section 6.1). Using this demonstrator, we have performed an experimental study of stereo matching and 3D object location for a set of image pairs of a car brake shoe assembly in a variety of orientations with no special lighting. The results of these experiments showed that our FSC stereo algorithm is as effective as an established FBS algorithm, PMF, and that the TINA 3DMM system overall is effective for object location. The study also helped to highlight some areas in which performance could be improved, which we address in the next section.

7.1 Future work

This project allowed us to identify areas in which improvements to our vision system could be made, in particular in model and algorithm optimisation for specific tasks, based on our general statistical approach to the evaluation of stereo reconstruction.

As mentioned in Section 4.5, TINA allows the initial object location estimate to be passed to a closed loop validation routine that refines the derived location on the basis of a robust likelihood measure. This has not been fully integrated with the SC algorithm at present, but is under investigation in a separate thread of this project, which considers the relationship of TINA's methodology to approaches such as that used in BAe Systems' TALEOS. The integration of the full system would allow further refinement not only of the object location result, but also of the wireframe model, as it allows the model to be translated into the original image co-ordinates and so allows direct evaluation of the feature location accuracy. This would facilitate a fully quantitative evaluation of the results of our 3DMM and object location experiments.

We also have a temporal stereo capability, with a multi-scale bootstrap technique, but once again this is not fully integrated into the TINA system as described in this report. In particular, the temporal bootstrap helps to properly set up the disparity space co-ordinates and avoids the sub-optimal error assumptions of the wireframe model matcher (Section 4.5).

Future projects should address the issues of fully integrating these elements of the TINA system.

The work presented here also has general characteristics which can be compared to other vision tasks, such as object localisation as attepted in the TALEOS system. The wireframe model matcher presented here is not based upon a robust optimisation but uses a naive least squares apprach, which is not expected to give good results in the presence of outliers. It derives it's robustness from the way that multiple hypotheses are generated from a subset of data (typically of the order of ten) for comparison to the image data. A similar approach may also lead to a succesful extension of the current TALEOS system. Multilple hypotheses for the location of the model in the image data could be generated from subsets of the data, for confirmation using the original image. This is essentially the same approach as taken in the RANSAC algorithm (RANdom SAmpling with Consensus). In addition, the use of a fast but statistically sub-optimal transformation estimation necessitates the return to the use of a more justifiable maximum likelihood estimate as a final position refinement step. This has been described in Part II of this work.

A Wireframe model: details of construction

We stated in Section 6.2 that the TINA wireframe model representation comprises two ASCII text files:

- the "model file" (`*.poly`);
- the "clique file"(`*.ffg`).

The model file contains the geometric representation of the object and the clique file seeds the searches in the 3DMM matching algorithm and limits its search space. We now provide further details of the structure of these files and explain how they are generated by hand.

A.1 The `*.poly` file

The `*.poly` file contains the set of geometric feature elements (one line in the file represents one element of the model) that together describe the real-world object in a manner suitable for the TINA 3DMM. In the current implementation, only straight lines and ellipses can be included — point features and occluding boundaries (caused by gradually curving surfaces) cannot be represented.

For these experiments, the wireframe model was built by first identifying by eye the real-world object features thought most likely to give rise to detectable image features. A line was added to the `*.poly` file for each feature, specifying an integer identifier, the feature type (line or ellipse) and its measured dimensions, location and orientation relative to a fixed datum. Images of the object were then taken under a range of illumination conditions and these were examined before and after running TINA's Canny edge detector. This allowed the model to be adjusted by adding feature elements that had not been apparent to the naked eye and removing elements that in practice proved difficult to detect.

A.2 The `*.ffg` file

The geometric wireframe model itself is not sufficient to represent the object. As opaque surfaces are not explicitly represented, all feature elements are visible from any viewpoint in any orientation. This can lead to impossible views being interpreted by the 3DMM as valid object orientations. To address this problem, TINA uses the concept of a "clique" or "focus feature group". Each clique represents a valid viewpoint and object orientation, by grouping the features that are visible from the corresponding viewpoint in the corresponding orientation.

The `*.ffg` file contains a line for each clique, listing the integer identifiers of the visible feature elements from the `*.poly` file. One feature from each clique is chosen as the "focus feature" — this is expected `always` to be visible for the viewpoint and orientation corresponding to that clique. The `*.ffg` file is used by the 3DMM to seed the matching algorithm and to restrict its search space.

For these experiments, the clique file was built by selecting a set of viewpoints and object orientations then, for each of these, compiling a list of the identifiers of the feature elements from the `*.poly` file expected to be clearly visible. One or more focus features was chosen for each clique, and a line was added to the `*.ffg` file for each combination of clique (feature group) and focus feature.

The set of cliques was iteratively refined by trial and error — in cases where the 3DMM had no match but visual inspection of the stereo data suggested that a match should be possible, a new clique was added and the analysis repeated. If the match was still unsuccessful the new clique was deleted.

B Nominal Specification for the Stereo Camera System

This section summarises the considerations we made when constructing the stereo camera system. Two possible imaging systems were proposed. The first is a lower cost solution using interlaced cameras. The second is a progressive scan (full frame shutter) camera based system.

B.1 Frame-grabber

The suggested frame-grabber is the Leutron PicPort Stereo. There are several variants on this device and the exact m odel would depend on the choice of camera system (either the basic model for the interlaced solution or the H4S or H4 D for the full frame shutter solution). This card has been selected for several reasons:

- The PicPort is a PCI based card supporting CCIR camera signals. The analogue model captures 8-bit data and can transfer this to the host at 132Mb/s (peak) as PCI bus-master which ensure almost total bus control and transfer of data with the minimum of CPU intervention. The data rate necessary in order to capture stereo sequences at frame rates is calculated as

$$datarate = width \times height \times bits \times 2 \times framerate$$

which for typical stereo frames of 512×512 at $8bit$ at a rate of $25\ frames/s$ is

$$512 \times 512 \times 8 \times 2 \times 25 = 100Mb/s$$

- The card is a *true* dual camera digitiser able to capture frames from both cameras simultaneously. It is able to do this because it has dual A/D converters and dual data channels (multiplexing occurs after signal capture a t the board PCI output which is buffered).
- Camera synchronisation signals can be generated on the card and served to multiple cameras. These are the horizontal and vertical sync pulses and pixel clock signals. These signal outputs provide the greatest level synchronisa tion over current analogue CCD cameras resulting in $< 1ns$ pixel jitter with the LV-8500 (PSS) cameras.
- Linux interface libraries are supplied as standard which enable complete control over all of the cards functionality. MS Windows libraries and application software are also provided.

B.2 Cameras

Ideally the cameras should capture full frame images in order to overcome the problems of interlacing which severely corrupts the geometric integrity of the signal (albeit in a known manner). This is because, in an interlaced image consecutive lines in the have a larger temporal separation than the separation between every third line. In fact the camera shutter captures all the odd lines first (odd field) and then all the even lines (even field) and it is the frame capture hardware which must re-interlace the result to give a pseudo-full frame image. Often, particularly for static scene, the benefits of the increased spatial information provided by interlacing greatly out weigh the inter-field time effects. However, in temporal imaging the classic 'striping' effect visible in moving interlaced images often has catastrophic effects on many algorithms.

Most standard low-cost CCIR cameras are interlaced although many provide a non-interlaced output mode. However, the data output in this mode is not a true full frame, rather it is derived from the interlaced image usually by constant output of either the odd or even fields (often with no capture of the alternate field and thus allowing double frame -rates). This results in a characteristic squashed (half-height) frame. The problem now is the loss of correct spatial information in the scene due to the large inter-line gap (where the alternate field data would have been) which h as a detrimental effect on spatial feature extraction such as edge detection. More sophisticated cameras will produce full sized images by interpolating values to artificially create an alternate line (which could anyway be done in software or even encoded into the algorithms themselves). Some systems even subsample the images to create a quarter size frame. However, all these techniques are, for machine vision purposes, pointless. Once the information is lost, it is lost forever.

The solution is acquire proper non-interlaced images. However, because manufactures have used the term non-interlace to mean all odd or even fields or some derivative of, a new term is being used; progressive scan, i.e. the data is scanned from the CCD array in line order 1,2,3,4... However, this still does not ensure that all lines were captured simultaneously. There are devices which capture the odd field and then the even field, leaving the odd field still on the array or in a backing store, before outputting all lines in numeric order. With such devices there is still a temporal lag between alternate fields and thus the interlacing effect is still present (although the capture hardwar e no longer has to reconstruct the image from the fields). The only way to ensure full frame images is to use a *full frame shutter* camera. Such cameras shutter all lines simultaneously (?) and then employs a progressive scan to export the data.

Field shuttered If the budget can only stretch to field frame cameras then the LV-75CE (or compatible Sony XC-75CE) or the Pulnix TM- 300 are recommended.

Full frame shuttered Several full frame shuttered cameras are available. At the lower end of the market are many the VGA compatible cameras. These have imaging sensors with dimensions consistent with the VGA standard (approximately 640×480). Such dimensions are slightly below the 512×512 minimum image usually acquired. Such devices include the Pulnix TM-6702. At the other end of the scale are large size imaged sensor devices such as the Pulnix TM-1300 which is a megapixel device with digital output, but will likely have a cost to match. The recommended device is the LV-8500 (or compatible Sony XC-8500) family. Briefly these devices have the following features/benefits:

- 782×582 full frame shutter, progressive scan image sensor with $> 55dB$ S/N ratio ($55dB$ means $\approx$ 9.13bits of data).
- Square pixels.
- H & V synchronisation (and pixel clock on PSS model) inputs and outputs.
- Shutter speeds up to 1/10,000 sec with full frame shuttering.
- Good integration with the PicPort frame-grabber.

B.3 Lenses

The story with lenses is you get what you pay for really. We want lenses with minimal radial distortion (although we can calibrate for some level of distortion).

Focal length In order to calculate the required focal length of the lenses we need to estimate the dimensions of the scenes that w ill be typically imaged. We will assume therefore, that this system will predominantly be used in a typical office/laboratory environment. Thus we shall assume that typical working distances will be $0 - 4M$ and typical object heights will be $0 - 0.5M$. We can then use the similar triangle formula below to calculate the focal length

$$f = \frac{object\ distance \times CCD\ V\ Spec}{object\ height}$$

$$\frac{4,000\ mm \times 6.3mm}{500} = 50.4mm$$

References

[1] H H Baker and T O Binford, Depth from edge and intensity based stereo. *Proc. VII Int. Joint Conf. Artifical Intelligence*, pp. 631–636, August 1981.

[2] D H Ballard and C M Brown, *Computer vision*. Prentice Hall, 1982.

[3] S T Barnard and W B Thompson, Disparity analysis of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2(4):333–340, 1980.

[4] M Brown, T Drummond and R Cipolla, 3D model acquisition by tracking 2D wireframes. *Proc. BMVC*, Sep. 2001.

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

[6] J F Canny, A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[7] M Chevrel, M Courtis and G Weill, The SPOT satellite remote sensing mission. *Photogrammetric Eng. Remote Sensing*, 47(8):1163–1171, 1981.

[8] S Crossley, *Robust temporal stereo computer vision*. PhD Thesis, University of Sheffield, 2000.

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

[10] Y G Leclerc, Q T Luong and P Fua, Measuring the Self-Consistency of Stereo Algorithms. *Proc. Computer Vision - ECCV 2000*, 1:282–298, 2000.

[11] O D Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig? *Proc. ECCV II*, pp. 563–578, 1992.

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

[13] A W Gruen, Adaptive least squares correlation - a powerful image matching technique. *S. African J. Photogrammetry, Remote Sensing and Cartography*, 14(3), 1985.

[14] M J Hannah, A system for digital stereo image matching. *Photogrammetric Eng. Remote Sensing*, 55:1765–1770, 1989.

[15] A J Harris, N A Thacker and A J Lacey, Modelling feature based stereo vision for range sensor simulation. *Proc. European Simulation Multiconference*, pp. 417–421, June 1998.

[16] Inria Research Laboratories, France, *A parallel stereo algorithm that produces dense depth maps and preserves image features*. Research Report, 1991, No. 1369.

[17] R Jain, R Kasturi and B G Schunk, *Machine vision*. McGraw-Hill, 1995.

[18] E P Krotkov, *Active computer vision by cooperative focus and stereo*. Springer series in Perception Engineering, Springer-Verlag, 1989.

[19] A J Lacey, N A Thacker, S Crossley and R B Yates, A multi-stage approach to the dense estimation of disparity from stereo SEM images. *Image and Vision Computing*, 16:373–383, 1998.

[20] 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.

[21] R A Lane, *Edge based stereo vision with a VLSI implementation*. PhD Thesis, University of Sheffield, 1995.

[22] R A LANE, N A THACKER, N L SEED AND P A IVEY, A generalised computer vision chip. *Real Time Imaging*, 2:203–213, 1996.

[23] M D LEVINE, D A O'HANDLEY AND G M YAGI, Computer determination of depth maps. *Computer Graphics and Image Processing*, 2(2):131–150, 1973.

[24] D MARR AND T POGGIO, Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

[25] J E W MAYHEW AND J P FRISBY, Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, 1981.

[26] J E W MAYHEW AND J P FRISBY, *3D model recognition from stereoscopic cues*. MIT Press, 1991.

[27] H P MORAVEC, Towards automatic visual obstacle avoidance. *Proc. V Int. Joint Conf. Artifical Intelligence*, 1977.

[28] M OKUTOMI AND T KANADE, A locally adaptive window for signal matching. *Int. J. Computer Vision*, 7(2):143–162, 1992.

[29] M O'NEILL AND M DENOS, Automated system for coarse to fine pyramidal area correlation stereo matching. *Image and Vision Computing*, 14:225–236, 1996.

[30] G P OTTO AND T K W CHAU, Region growing algorithm for the matching of terrain images. *Image and Vision Computing*, 7(2):83–93, 1989.

[31] S B POLLARD, *Indentifying correspondences in binocular stereo*. PhD Thesis, University of Sheffield, 1985.

[32] 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.

[33] S B POLLARD, J PORRILL AND J E W MAYHEW, Recovering partial wireframe descriptions from stereo data. *Proc. BMVC*, pp. 31–36, 1990.

[34] S B POLLARD, J E W MAYHEW AND J P FRISBY, Implementation details of the PMF stereo algorithm. In *3D model recognition from stereoscopic cues*, Eds. J E W MAYHEW AND J P FRISBY, *3D model recognition from stereoscopic cues*. MIT Press, 1991.

[35] 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. *Proc. Alvey Vision Conference*, 1987.

[36] J PORRILL, Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter. *Proc. Alvey Vision Conference*, pp. 175–185, 1989.

[37] 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.

[38] *The Persistence of Vision Ray-Tracer (POV-Ray)*. http:\\www.povray.org, www reference valid 31 July 2003.

[39] M SONKA, V HLAVAC AND R BOYLE, *Image processing, analysis and machine vision*. PWS, 1998.

[40] 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.

[41] N A THACKER AND P COURTNEY, Statistical analysis of a stereo matching algorithm. *Proc. BMVC*, pp. 316–326, 1992.

[42] N A THACKER, A J LACEY, P COURTNEY AND G S REES, *An empirical design methodology for the construction of machine vision systems*. Tina memo 2002-005, http:\\www.tina-vision.net\docs\memos.php, www reference valid 31 July 2003.

[43] H P TRIVEDI AND S A LLOYD, The role of disparity gradient in stereo vision. *Perception*, 14:685–690, 1985.

[44] E TRUCCO AND A VERRI, *Introductory techniques for 3D computer vision*. Prentice-Hall, 1998.

[45] R Y TSAI, An efficient and accurate camera calibration technique for 3D machine vision. *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 364–374, 1986.

[46] V VENKATESWAR AND R CHELLAPPA, Extraction of straight lines in aerial images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:1111–1114, 1992.

[47] A ZISSERMAN, P BEARDSLEY AND I REID, Metric calibration of a stereo rig. *IEEE Workshop on Representation of Visual Scenes*, pp. 93–100, Boston, 1995.