Epipolar Geometry for Trinocular Active Range-sensors

A. Blake, D. McCowen, H.R. Lo, D. Konash

Department of Engineering Science
Oxford University
Oxford OX1 3PJ, U K

We investigate recent ideas in the design of trinocular active range-sensors. Such devices have the advantage of freedom from mechanical scanning, and rapid image capture. The main technical problem is overcoming the correspondence problem. This requires careful geometric design to take account of epipolar geometry and thorough modelling of image-measurement error. We present a novel design that, so far, seems to work well. Curiously it involves setting up the projector–camera geometry to be degenerate — so that depth computation is ill-conditioned — and then backing off a little.

Work on active rangefinders for 3–D inspection and robot vision has been in progress for about two decades. Overviews of the most widely–investigated approaches can be found in (Bastuscheck 1989) and (Jarvis 1983a). Some of these are Moiré fringing, ratio image techniques (Bastuscheck & Schwarz 1984), or time–of–flight range systems (Jarvis 1983b). We are concerned with structured–light systems (Shirai 1972, Altschuler et al. 1981, Altschuler et al. 1987). Typical of systems that are quite well–developed for industrial applications in this area (e.g. profiling turbine blades) is the system of Mundy and Porter (1987). This device uses special hardware to give 60,000 range readings/sec, albeit over a small depth range. Like the less depth–limited but slower system of Case, Jalkio and Kim (1987), it involves mechanical scanning of a light pattern across a scene.

There are still other systems which, like ours, attempt to get away from the expense and complexity of a mechanical scanner and move closer to simultaneous acquisition of all range data, essential for a moving scene. Such systems include that of Godin and Levine(1989) and that of Hu and Stockman(1989), the latter of which uses a grid of light to illuminate the scene, providing easily–matched artificial "surface features". However there is a correspondence problem. Regarding the camera and projector as a stereo pair (figure 1), and given an image point, it cannot be uniquely matched to a node point on the grid mask. This is the *node labelling problem.* Some solutions which have been proposed include colour (Boyer and Kak 1987) or thickness (Le Moigne and Waxman) coding, or space coding (Posdamer and Altschuler 1982, Altschuler et al. 1987). There are objections to all of these for certain types of scene, so we hope to be able to label lines and nodes successfully without any type of coding.

The solution of Stockman and Hu to the node labelling problem can be explained as follows. Given a projected grid-crossing on the image plane, it is required to determine which node in the projected pattern corresponds to that crossing. The position of the grid-crossing in the image plane is associated with an epipolar line in the projector plane (on the left of figure 1). Possible solutions are those nodes that lie on (or sufficiently near) the epipolar line. This generates a *set* of possible solutions for each grid crossing which can then be reduced somewhat by constraint propagation, between crossings, along grid-lines in the image plane.

The prototype being developed at Oxford is founded on a somewhat different philosophy, although the optical hardware is similar. Our system is *trinocular*, having two projectors and one camera — the advantage of this will be explained later. We aim to avoid the computational expense of constraint propagation and the problem of remaining ambiguity in node labels. That can be achieved by ensuring that only *one* node lies near the epipolar line. This is done by

- Orientating the projected pattern, with respect to the camera, to minimise the incidence of false node solutions.
- Improving the basic image measurement accuracy of the system. The trinocular configuration is important here.
- Detailed error analysis to predict image measurement tolerances — how close must a node be to the epipolar line to be a feasible solution?
- Restricting the working volume of the system.

1 CURRENT STATE OF DEVELOPMENT

1.1 Introduction

This section introduces the notation which is necessary both to describe the current system and to illustrate the design principles in Section 2. It then goes on to describe the prototype rangefinder system which is now in operation, showing results for a typical scene.

BMVC 1990 doi:10.5244/C.4.6

1.2 Notation

World and image coordinates
Figure 2 shows the general configuration of the current rangefinder. The following section defines the notation which has been developed to describe the geometry illustrated by figure 2; all the terms defined here will be used to illustrate the reasoning behind the design.

The notation to be used in this paper employs homogeneous coordinates throughout. The imaging process is expressed as a perspective transformation $\mathbf{X} = C\mathbf{x}$ from world position $\mathbf{x}$ to image position vector $\mathbf{X}$, via the camera matrix C.

Light planes and their intersections:
The two projectors each produce a set of light planes

$$A = \{\mathbf{a}_1, ..., \mathbf{a}_N\} \text{ and } B = \{\mathbf{b}_1, ..., \mathbf{b}_N\}$$

where $\mathbf{a}_i$ is a dual vector defining a plane

$$\mathbf{a}_i.\mathbf{x} = 0.$$

and similarly for $\mathbf{b}_i$. Light planes $\mathbf{a}_i$ and $\mathbf{b}_j$ intersect along a line L_{ij} in space (see figure 2), which is projected onto the image plane as a line defined by the dual vector $\boldsymbol{\lambda}_{ij}$. The line is the intersection of the plane

$$\lambda_{\mathbf{ij}}.\mathbf{X} = \mathbf{0} \tag{1}$$

with the image plane.

Camera rays and epipolar planes:
The camera ray C_x is a line in space joining the centre $\mathbf{c}$ of the camera to a scene point $\boldsymbol{x}$, passing through the image point $\mathbf{X}$ (figure 2). Epipolar planes $\mathbf{e}$ are defined as those containing both a given camera ray C_x and the centre $\boldsymbol{p}_A$ or $\boldsymbol{p}_B$ of a projector. Stripe edges, the intersections of projector planes $\boldsymbol{q}_A$ or $\boldsymbol{q}_B$ with $\mathbf{a}_i$ and $\mathbf{b}_i$ are denoted S_{Ai} and S_{Bj}. The intersection of $\boldsymbol{q}_A$ or $\boldsymbol{q}_B$ with $\mathbf{e}$ is an epipolar line E. In a single projector system, the two projection planes $\mathbf{q}_A$ and $\mathbf{q}_B$ are deemed to coincide.

1.3 The Current System

Two projectors are used in our current system because the precision of our edge detection software (Canny 1986) exceeds considerably the precision of available corner detectors. We therefore acquire two precisely registered images, one with each projector, and combine them in memory, as in figure 3. First the images of the scene illuminated with A and then B stripes are shown (a,c). (b) and (d) show the result of the Canny edge finder on (a) and (c) respectively. (f) is an edge map ((b) and (d) combined) showing the grid at whose (scene) nodes we make 3–D measurements. Figure 4 shows the results of our current algorithm for a part of this scene. It is an orthographic projection of computed 3–D stripe edge positions and clearly shows the convex surface of one of the forks.

Given an image point $\mathbf{X}$, there are 2N candidate 3D positions $\boldsymbol{x}_{A1}, \boldsymbol{x}_{A2}, \ldots, \boldsymbol{x}_{AN}$ and $\boldsymbol{x}_{B1}, \boldsymbol{x}_{B2}, \ldots, \boldsymbol{x}_{BN}$. These correspond to the intersections of the camera ray C_x, passing through $\boldsymbol{X}$, with light planes $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_N$. Clearly one of $\boldsymbol{x}_{Ai}$ and one of $\boldsymbol{x}_{Bj}$ each refer to the point $\mathbf{x}$ we are looking for. The next step is therefore to match values of $\boldsymbol{x}_{Ai}$ to $\boldsymbol{x}_{Bj}$. In practice we find which of the z coordinates of $\boldsymbol{x}_{Ai}$ and $\boldsymbol{x}_{Bj}$ fall within a given tolerance z_T of one another. We can see, therefore, that it is vital to minimise the number of candidates chosen by minimising the system error which imposes an upper limit on z_T. Thus accuracy requirements for the system are high even if the final range map does not require extreme precision. This precision is used, in effect, to buy a reduction of ambiguity.

2 DESIGNING THE NEXT STAGE

The phase I prototype used no purpose–built hardware. It has given us enough insight into the operation of the rangefinder to design an improved system, to be built at NEL East Kilbride. We start with several degrees of design freedom.

- whether to obtain both sets of planes A & B from one projector, or to continue with two.
- what patterns to use for the masks on planes $\boldsymbol{q}_A$ and $\boldsymbol{q}_B$ (e.g. parallel stripes or not).
- number of stripes per unit distance on these masks.
- relative positions and orientations of the camera and projector(s).

2.1 Analysis on Projection Plane—Minimising Ambiguity

Figure 5 shows pattern A, on the projection plane of a single projector system, with an epipolar line E. The point is that the 3–D situation can be represented in 2 dimensions. Each crossing of E with a pattern line S_{An} corresponds to an intersection of the camera ray C_x with the light plane $\boldsymbol{a}_n$. The aim is to minimise the number of solutions for the stripe pair (i, j). Minimising the number of intersections help achieve this. Let the total number of such intersections (for reasons explained later, we cannot generally reduce N_I to 1) be N_I. To obtain at most N_I intersections we require

$$\theta \leq \tan^{-1} \frac{N_I \delta}{L}, \tag{2}$$

where θ is the angle between the epipolar line and pattern lines (figure 5).

It can be seen from figure 7 that there is one additional constraint on the node label (i, j). We can tell whether (i, j) are odd or even by determining the contrast sign, i.e. whether $\boldsymbol{a}_i$ is the "left–hand" or "right–hand" edge of a light stripe; similarly whether $\boldsymbol{b}_j$ is the "top" or "bottom" edge. There is no ambiguity about edge contrast sign because we know the angles at which light planes fall on the baseplane ($z = 0$) from calibration, and it can be shown that they cannot deviate from this angle by more than 90^o for any scene. The intersections $\nu_{i,j}$ of lines S_{Ai} and S_{Bj} shown are denoted *mask nodes.*

2.2 Use of a Single Centre of Projection

The smaller the value of θ, the more ill–conditioned will be surface measurements $\mathbf{x}$ from intersecting C_x with planes from projector A. This is because, as $\theta \rightarrow 0$, we are effectively trying to triangulate from a zero baseline, equivalent to moving stereo cameras too close together. Therefore we still need the second projected plane set B, with planes set at a much larger angle θ to the epipolar lines. One way of generating the second plane set would be to use just one projector and interchange A and B masks between image acquisitions. This has the drawback of requiring some pattern switching device, an additional complexity and possibly prejudicial to calibration accuracy. Alternatively, it turns out there is a special arrangement of projectors (see figure 6) that provides us with a single *virtual* centre of projection. All light planes can be considered as projected from this virtual centre through stripe-edges S_{Ai} and S_{Bj} on a virtual projector plane $\boldsymbol{q}_{\text{virtual}}$.

In an ideal system, free of measurement error, an epipolar line would generally intersect only one node (i, j) on the virtual projector plane. In the presence of error (see next section) all mask nodes must be accepted within a band of width 2τ around E. A lower limit on θ is therefore imposed by the need (see figure 7) to make sure the tolerance band around E does not overlap other mask nodes with the same contrast sign. Hence and from equation (2) we can define limits for θ (assuming same spatial period δ on plane $\boldsymbol{q}_{\text{virtual}}$ for both A and B):

$$\sin^{-1} \frac{2\tau}{\delta} \leq \theta \leq \tan^{-1} \frac{N_I \delta}{L} \qquad (3)$$

Figure 7 also shows a second epipolar line E. In general epipolar lines on the projection plane are not parallel. This presents the problem that θ cannot be maintained constant over the entire image plane. Figure 8 shows, however, that the epipolar lines can be made parallel on $\boldsymbol{q}_{\text{virtual}}$. This relies on the result from 3–D geometry that, if several planes ($\boldsymbol{a}_i$ here), intersect in one line joining $\boldsymbol{p}_{\text{virtual}}$ and $\mathbf{c}$, their lines of intersection with another plane $\boldsymbol{q}_{\text{virtual}}$ are parallel if and only if the line from $\boldsymbol{p}_{\text{virtual}}$ to $\mathbf{c}$ is parallel to projection plane $\boldsymbol{q}_{\text{virtual}}$. (This could also apply, of course, to $\boldsymbol{q}_A$ in place of $\mathbf{q}_{\text{virtual}}$.) Figure 8 also shows that if the vector $\boldsymbol{p}_{\text{virtual}} - \boldsymbol{c}$ coincides with the line of intersection of all the planes $\boldsymbol{a}_i$, then epipolar planes and light planes coincide. It then only remains to rotate projector A by a small angle θ, as determined by equation (3).

2.3 Testing Intersections in the Image Plane

The virtual projector is useful for formulating the problem in two dimensions, and it has given us some useful insights, such as the need to make epipolar lines parallel on $\boldsymbol{q}_{\text{virtual}}$. However, it is not, after all, obligatory to have a virtual projector centre, since there is an alternative way of representing the geometry of camera rays and their intersections with light planes. This uses the images $\boldsymbol{\lambda}_{ij}$ of the light plane intersections $\boldsymbol{L}_{ij}$ illustrated in figure 2. The amount of ambiguity in line labelling can be analysed using an image–plane representation, as follows. For an error–free system we could simply use equation (1) to determine whether or not a given image point $\mathbf{X}$ lies on the line $\boldsymbol{\lambda}_{ij}$, and thus whether its camera ray intersects $\boldsymbol{L}_{ij}$ (see figure 2). However, the calibration and measurement process introduces errors (see section 3), so we need to use a tolerance band around $\boldsymbol{\lambda}_{ij}$ in the image plane, like that shown in figure 7 for the mask plane. There are several advantages to using the image plane for labelling.

- The lines $\boldsymbol{\lambda}_{ij}$ exist whether or not there is a special camera–projector geometry.
- Calculation of 3D positions using coordinate transforms is slow; using the image plane we do not need to find $\mathbf{x}$ to label the planes $\boldsymbol{a}_i$, $\boldsymbol{b}_j$ which have produced it.
- The dominant source of error (see next section) occurs in localisation of edges in the image plane. The associated measurement tolerance can be used directly to determine when an image point $\mathbf{X}$ lies sufficiently close to a line $\boldsymbol{\lambda}_{ij}$ to be regarded as lying on it.

3 ERROR SOURCES IN RANGEFINDER

We have done an exhaustive analysis of error sources in the rangefinder system, with a view to discovering which sources and types of error predominate. Figure 9 illustrates how different sources of error appear and propagate through the various stages of operation. These stages are: firstly, the use of an image of several points $\mathbf{x}$ whose world coordinates are accurately known, to calibrate the camera using least–squares fitting. Clearly the measurement of points in the image introduces some error; it comes about mainly as a result of optical distortion, random noise in the image, and deviations from an ideal step-edge model. This will produce inaccuracies in the camera calibration and hence in the transformation matrix C.

The second stage uses the matrix C to determine each of the light planes $\boldsymbol{a}_i$, and similarly $\boldsymbol{b}_j$. Here, errors come both from C and again from point estimation in the image. Least–squares fitting is used again, for several points along each light stripe edge. Figure 10 shows how each of the above types of error vary with number of $\mathbf{X}$ points used for obtaining the information. The main conclusion to be drawn from this graph is that the most significant source of error is the on–line estimation of $\mathbf{x}$, provided that enough points are used in the least–squares estimation of $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$, and in camera calibration. Off-line errors, both random and systematic, can be made negligible.

SUMMARY

Our system attempts to avoid the complexity of hardware and computational expense which structured–light

rangefinders often require. It uses no coding of light patterns and requires only two frames to solve a 3–D scene. No connectivity or smoothness of surface patches need be assumed, yet the current system appears to be capable of unambiguous nodel-labelling. Considerations of epipolar geometry suggest a surprising design. Rather than placing both stripe sets, symmetrically, running at about $\pm 45^o$ to the epipolar lines, a highly unsymmetrical design is indicated. One set is nearly degenerate, running almost *parallel* to the epipolar lines. That avoids ambiguity but at the price of ill-conditioned depth computation. The other set is complementary. On its own, it would be highly ambiguous, but computed depths would be relatively free of noise-generated error. The combination of the two sets retains the virtues of each and the vices of neither.

ACKNOWLEDGEMENTS

We are most grateful to the ACME directorate of the SERC for support, to Computer Recognition Systems (CRS) Ltd, and to the University of Oxford. We acknowledge the support of members of the Robotics Research Group at the University of Oxford, particularly Michael Brady, Roberto Cipolla and Andrew Zisserman. We are grateful for stimulating discussions with Bill Thomas of CRS Ltd and Robert Rixon of NEL.

REFERENCES

1. Altschuler,M.D., Altschuler,B.R., and Taboceda,J. Laser electro-optic system for rapid three-dimensional topographic mapping of surfaces. *Optical Engineering* 20: 953—961, 1981.

2. Altschuler,M.D., Bae,K., Altschuler,B.R., Dijak,J.T., Tamburino,L.A., Woolford,B. Robot vision by encoded light beams. In *Three-dimensional machine vision*, Kanade,T. (ed). pp 97—150, Kluwer, 1987.

3. Bastuscheck,C.M. and Schwartz,J.T. Preliminary implementation of a ratio image distance sensor. *Robotics Report #28, Robotics Activity,* New York Univ, 1984.

4. Bastuscheck,C.M., Techniques for Real-time generation of range images. *IEEE Proc. CVPR,* pp 262—268, 1989

5. Canny,J.F. A computational approach to edge detection. *IEEE Trans. PAMI,* vol.8 no.6, pp 679—698, 1986.

6. Case,S.K., Jalkio,J.A., and Kim,R.C. 3–D Vision system analysis and design. In *Three-dimensional machine vision*, Kanade,T. (ed). pp 63—96, Kluwer, 1987.

7. Hu,G. and Stockman,G. 3–D surface solution using structured light and constraint propagation. *IEEE Trans. PAMI*, no. 4, pp 390—402, 1989

8. Godin,G.D. and Levine,M.D. Structured edge map of curved objects in a range image. *IEEE Proc. CVPR*, pp 276—281, 1989

9. Jarvis,R.A. A perspective on range–finding techniques for computer vision. *IEEE Trans. PAMI*, no. 5, pp 122—139, 1983a.

10. Jarvis,R.A. A laser time–of–flight sensor for robotic vision. *IEEE Trans. PAMI*, no. 5, pp 505—512, 1983b.

11. LeMoigne,J. and Waxman,A.M. Multi–resolution grid patterns for building range maps. *University of Maryland Technical Report.*

12. Mundy,J.L. and Porter,G.B. A three–dimensional sensor based on structured light. In *Three–dimensional machine vision*, Kanade,T. (ed). pp 3—62, Kluwer, 1987.

13. Posdamer,J.L. and Altschuler,M.D. Surface measurement by space–encoded projected beam systems. *Computer Graphics and Image Processing*, vol. 18, pp 1–17, 1982.

14. Shirai,Y. Recognition of polyhedrons with a rangefinder. *Journal of AI*, vol. 4, no. 3, pp 243—250, 1972.

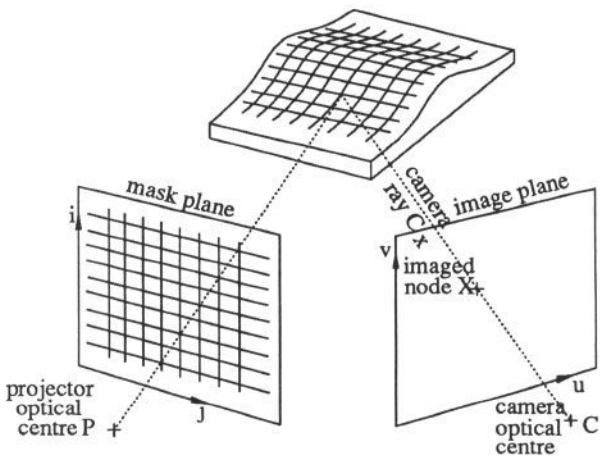

Figure 1: Projector and camera as "stereo pair" showing node/line labelling problem

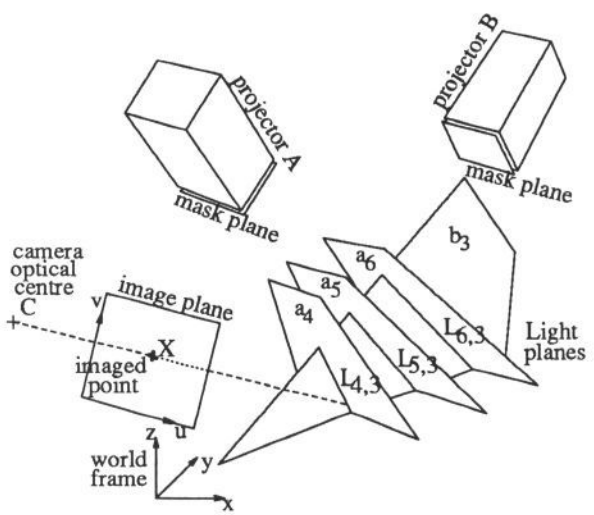

Figure 2: Coordinate systems used in the rangefinder

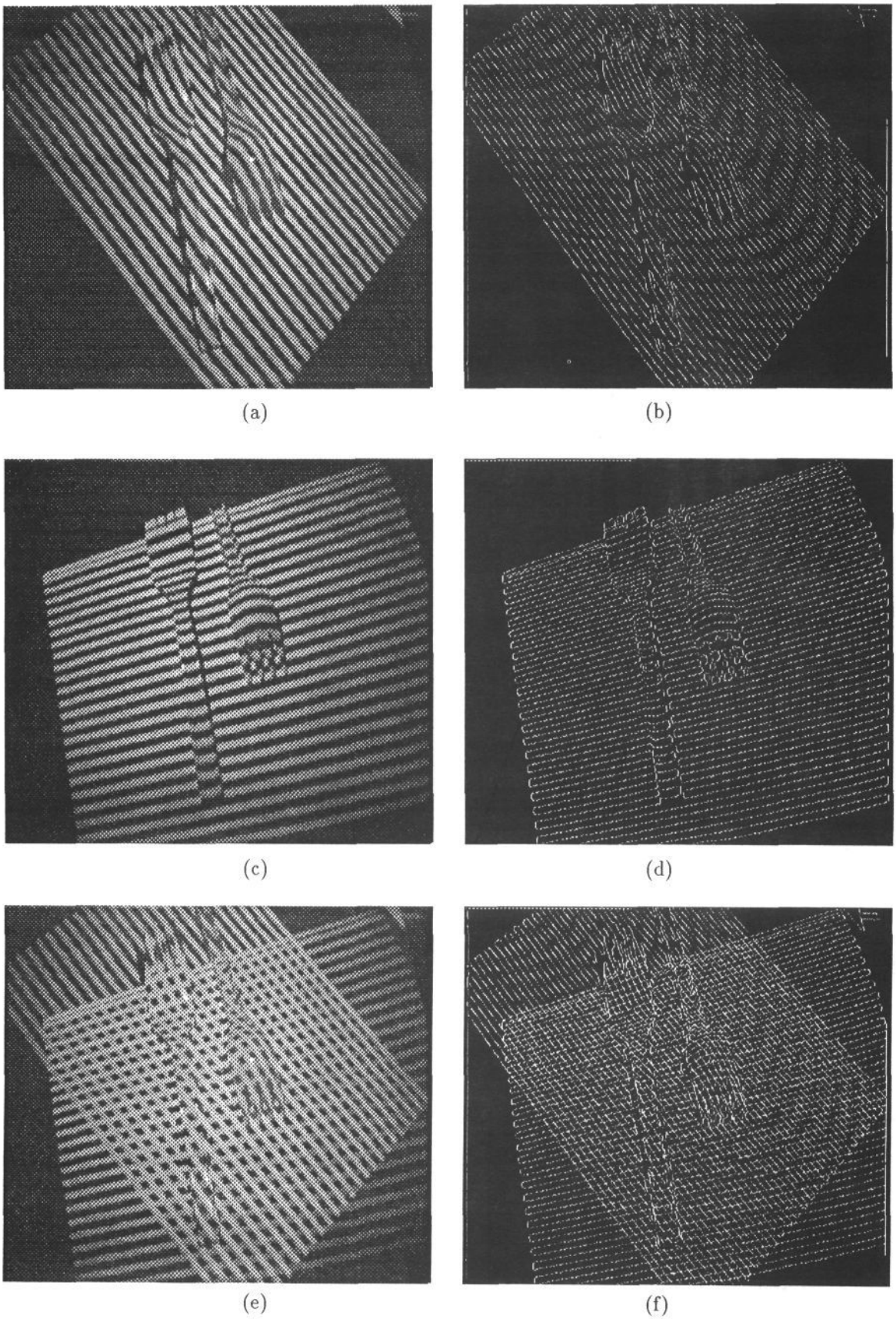

Figure 3: Initial structured-light images and edge maps used in rangefinder

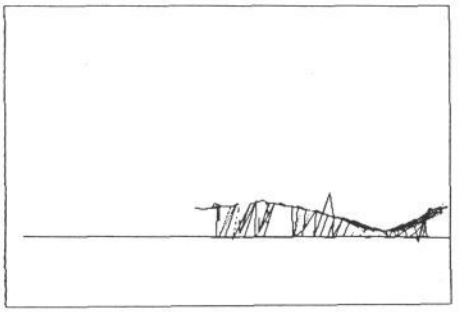

Figure 4: Rangefinder output for part of scene in the past figure

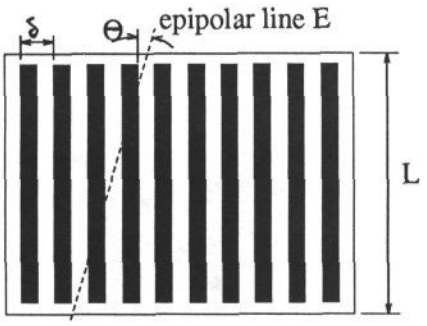

Figure 5: Mask plane q_A showing epipolar line E_{ij}

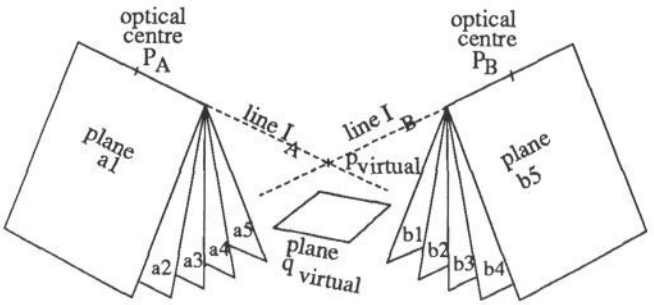

Figure 6: The virtual center of projection; a special relative positioning of the projectors allows the lines I_A and I_B to intersect, making light planes a_i and b_j effectively emanate from the same center of projection

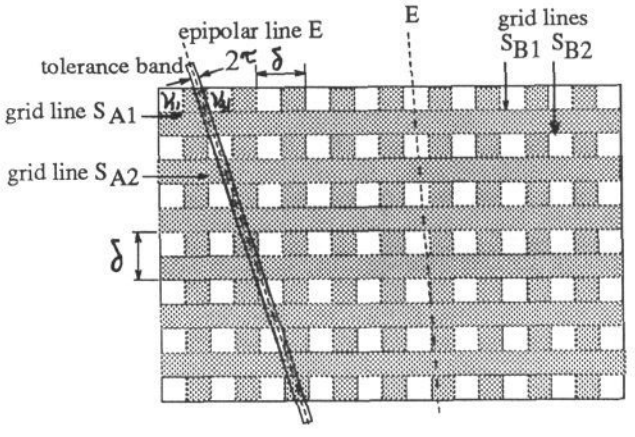

Figure 7: View of plane $q_{virtual}$ showing epipolar lines

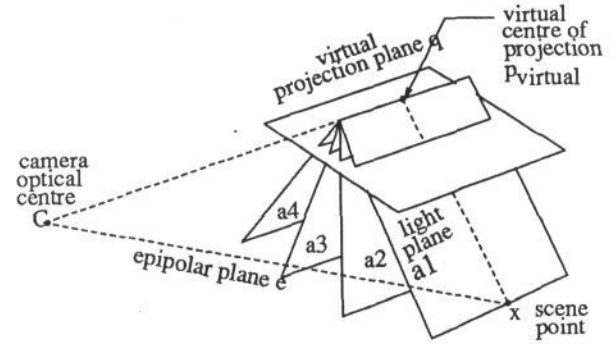

Figure 8: Condition for epipolar lines parallel on mask plane

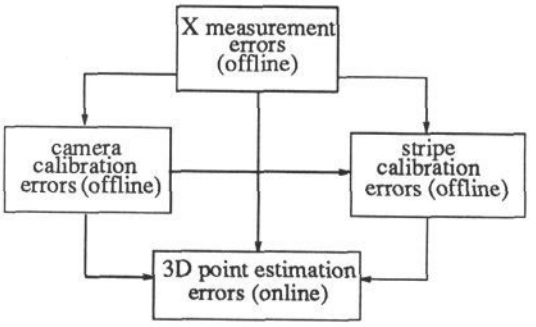

Figure 9: Sources and propagation of errors

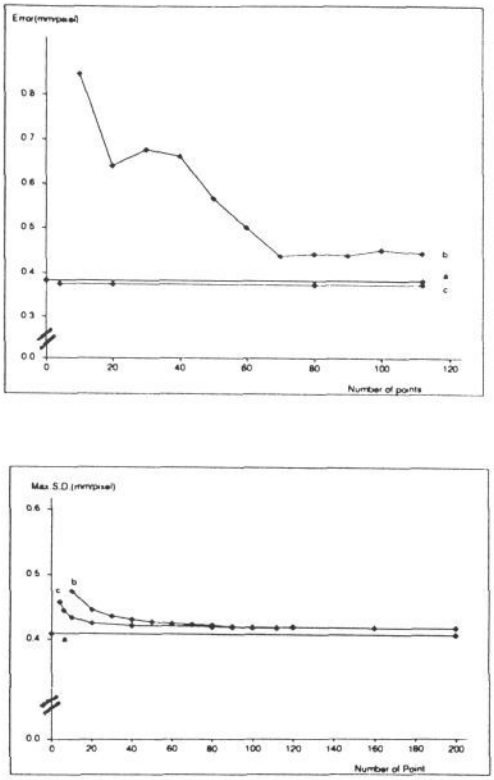

Figure 10: Variation of a 3D estimation error with number of X points used in calibration. (A) Systematic error; (B) Random error. For both: (a) is due to on-line localisation error only; (b) includes camera calibration error; (c) includes stripe calibration error with camera calibration error minimised.