*Sādhanā*, Vol. 18, Part 2, June 1993, pp. 239–249. © Printed in India.

# Improved depth map through optimal axis stereo

S SENGUPTA and S C SAHASRABUDHE*

Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur 721 302, India
*Department of Electrical Engineering, Indian Institute of Technology, Bombay 400 076, India

**Abstract.** Feature-based stereo correspondence techniques suffer from the major limitation that it is difficult to match along epipolar lines and this often results in a sparse set of depth points. Past researchers attempted to solve this problem through trinocular stereo. In this paper, a new method has been proposed for reducing the sparsity of depth points by orienting the epipolar line of the cameras in a direction that maximizes the number of feature points. The corresponding epipolar axis has been termed as the optimal axis. Our analytical as well as simulation results have established that for a limited edge scenario, the proposed approach can lead to considerable improvement in the number of feature points that can be matched. We have introduced a figure-of-merit for the optimal axis and discussed how it is qualitatively related to the variance of the probability density function (*pdf*). We have also presented the results of our simulation experiment, termed as the random stick experiment. Finally, we have also shown the results of improved reconstructed surface of a synthetic image using optimal axis alignment.

**Keywords.** Machine vision; 3-D depth perception; stereo correspondence; epipolar axis.

## 1. Introduction

Depth perception through stereo has emerged as one of the popular fields in recent years. In stereo-based depth perception techniques, two views of the same scene are taken from two different positions and matching is performed. The relative positional difference (also called "disparity") between the corresponding points in the two views gives a measure of the depth of the points from the camera position. The process of matching, also known as "stereo correspondence" is not a trivial task and has attracted the attention of researchers over the last few years (Marr & Poggio 1977; Baker & Binford 1981; Ayache & Faverjon 1985; Medioni & Nevatia 1985; Ohta & Kanade 1985; Lloyd *et al* 1987; Ji & Huang 1988; McIntosh & Mutch 1988; Horaud & Skordas 1989). In order to reduce the computational complexity of area correlation-based matching for every pixel in the image, most of the recent stereo correspondence techniques are based on matching specific features like zero-crossings of Laplacian-

of-Gaussian filtered images (Marr & Poggio 1977; Grimson 1981, 1983, 1985), edge intervals (Ohta & Kanade 1985; Lloyd *et al* 1987), edge segments (Ayache & Faverjon 1985; Medioni & Nevatia 1985; Ji & Huang 1988; McIntosh & Mutch 1988), structures (Horaud & Skordas 1989) etc. In these techniques, we select candidate features from one of the views and search for correspondence in the other view. To reduce the search space to one dimension, matching is performed along epipolar lines. The camera geometry can be constrained so that the epipolar lines lie in a direction parallel to the base line joining the cameras. However, one of the major limitations in zero-crossing or edge-pixel based matching is that by scanning along the epipolar line, features which are oriented along the epipolar lines or close to it are missed. One can match the vector features (Marr & Poggio 1977) between the two views but a pixel by pixel match is difficult when there is foreshortening or occlusion.

The matching problem along epipolar lines is of concern to researchers in trinocular vision (Ito & Ishii 1986; Ohta *et al* 1986; Stewart & Dyer 1988). By adding a third view with the cameras forming a triangle (Ito & Ishii 1986) or placed at the vertices of a right-angled triangle (Ohta *et al* 1986; Steward & Dyer 1988), it is possible to improve the density of the depth map. The features which are along the epipolar lines of the first and the second view and hence, missed, may be picked up while matching the first and the third or the second and the third view. The solution based on trinocular stereo requires alignments of three camera positions. We argue that it may not be necessary to introduce a third view, if instead, we are given the flexibility to orient the epipolar axis of the cameras. In this paper, a new method has been proposed for reducing the sparsity of the depth points by orienting the epipolar line of the cameras in a direction that maximizes the number of feature points. We have termed the corresponding epipolar axis of the cameras as the *optimal axis*. Our analytical as well as simulation results have established that for a limited edge scenario, the proposed approach can lead to considerable improvement in the number of feature points that can be matched. The effectiveness of this approach is scene dependent and the improvements one can get depend upon the statistics of lengths and orientations of the linear edge segments in a scene. We have shown that the variance of the probability density function (*pdf*) of the sum of projections of linear segments along the axis indicates the expected improvements through a best axis choice.

The problem of depth map sparsity is discussed in § 2. In § 3, we have derived the expression for the total number of matchable points in a scene. We have also presented the expressions for *pdf* of the sum of the projections of the linear edge segments, considering our forward random walk model. Our proposed approach on stereo matching has been discussed in § 4. The results are presented in § 5. Section 6 concludes this paper by summarizing the results and outlining the scope for further work.

## 2. Depth map sparsity in stereo correspondence

In a typical scene, edge orientations and edge lengths are random in nature. If the camera axes are fixed and there is no control over them, then it may so happen that in some scenes most of the edge pixels are aligned along the epipolar line, resulting in very poor density of the depth map. This may be illustrated by a simple example, as shown in figure 1. In this, we consider the rectangular projection ABCD of an object on the $x-y$ plane and let $O_{x-}OO_{x+}$ be the direction of the epipolar axis,

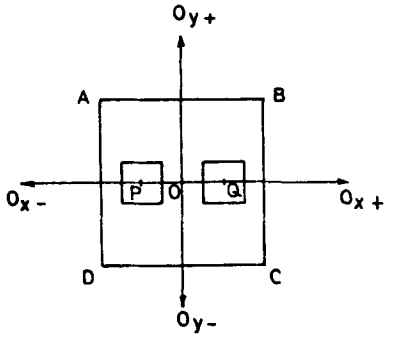


**Figure 1.** A rectangular object viewed from two positions.

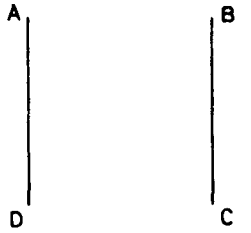


**Figure 2.** Sparse set of depth points.

which is perpendicular to the sides AD and BC. Let us suppose that the object is being viewed from two camera positions having centres of optical axes at $P$ and $Q$, situated at some distance $d$ from the viewed surface ABCD in the $z$ direction. In the process of stereo correspondence, the disparity and hence the depth values can not be computed for the points lying along the segments AB and CD. Depth values will be defined only along AD and BC, as shown in figure 2. Without the existence of lines AB and CD, it is impossible for any surface interpolation technique to interpolate correct depth values in between and the reconstructed surface will not resemble a rectangular object.

We argue that if we are given the flexibility to orient the epipolar axis of the cameras, it is possible to get some position, where the number of matchable feature points will be maximum. To illustrate this, let us consider the same example of the rectangular projection of an object.

If we now rotate the cameras w.r.t. $O$, in the epipolar plane, by some angle $\theta$ w.r.t. $OO_{x+}$, the epipolar axis will get aligned to its new position $O_{t-}OO_{t+}$, as shown in figure 3. Let ($\theta = \theta_{max}$) be the orientation at which the number of matchable feature points of ABCD can be maximized. At this position, it is possible to compute the depth values at all the points along the contour and the resulting set of depth points will appear as shown in figure 4. Here, the object will appear inclined by an angle $\theta_{max}$ w.r.t. its original position shown in figure 1, but has an improved set of depth

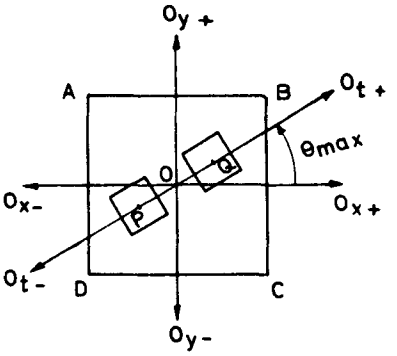


**Figure 3.** Rectangular object viewed after camera epipolar axis rotation.

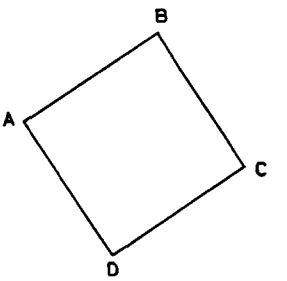


**Figure 4.** Dense set of depth points at rotated epipolar axis.

points, as compared to figure 2. Real world scenes may not be as simple as our example, but for limited edge scenarios, where sparse sets of data points make surface representation difficult, optimal axis alignment should be meaningful. This was our motivation in establishing the optimal axis concept.

It may be noted that our scheme will not be practical for obtaining stereo pairs of satellite images, as there will not be any flexibility of camera rotations. As a matter of fact, for satellite image pairs, the scheme may not even be required, since for satellite stereo image pairs, it is possible to get dense depth maps. Our scheme will be helpful for limited edge s narios and controlled camera environments like robotics and machine vision applications. The scheme should be easier to implement as compared to trinocular vision, since the cumbersome process of aligning the cameras can be avoided.

## 3. Matchable feature points in a scene

To compute the number of matchable feature points, let us consider a single, linear edge segment having length $a$ and orientation $\theta$ with respect to the epipolar axis, as shown in figure 5. We consider this edge segment in the digital pixel space after sampling and let $R$ be the sampling resolution, expressed in units of *number of pixels per units of length*. If we wish to match this segment with its counterpart in the other view, the matching is to be performed for every scan line. The scanning is done in the direction of the epipolar axis, which is the $x$-direction in this case.

The number of scanlines occupied by the edge segment is given by

$$n_s = (a \cos \theta)/R. \tag{1}$$

In the process of digital sampling of the edge segment, there may be more than one pixel existing in the scanline, as shown in figure 6. In this case, only the circled pixels will be picked up as the zero-crossing candidate for matching, i.e., there can be only one matchable feature point per scanline per edge segment.

Thus, the number of matchable feature points is given by

$$M = (a \cos \theta)/R. \tag{2}$$

If we consider $N$ number of edge segments in a scene, with segment length $a_k$ and

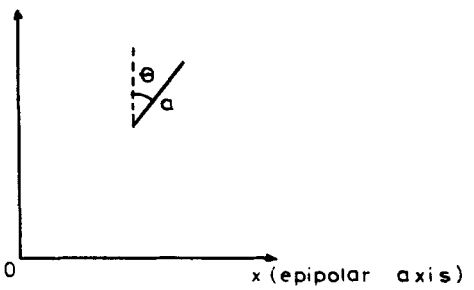


**Figure 5.** Single linear edge segment.

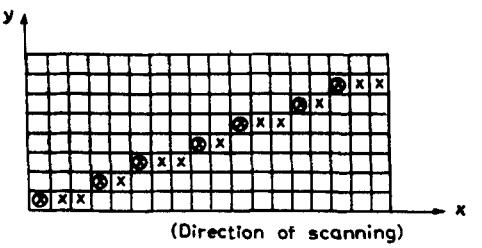


**Figure 6.** Sampling of single linear edge in digital pixel space.

orientation $\theta_k$, the total number of matchable feature points is given by

$$M = \sum_{k=1}^{N} a_k \cos\theta_k. \tag{3}$$

For the sake of convenience, we have chosen $R$ to be unity. The right hand side of (3) indicates the *sum of projections of the linear segments in the direction perpendicular to the epipolar axis.*

To improve the density of the depth map, we are aiming for the maximization of the number of feature points as given in (3), by proper choice of epipolar axis. Considering a general class of scenes, both $a_k$ and $\theta_k$ and even $N$ are random variables and hence a prediction of optimal axis is not possible without performing explicit computation of (3) for all possible positions of the epipolar axis. However, before doing this rigorous exercise, we must know whether the improvement is going to be significant and it is worth trying for a best axis. For a very marginal and insignificant improvement, the search exercise could be avoided.

We attempted to solve the above decision making problem through a probabilistic approach. Given an epipolar axis, it is worth finding, for randomly distributed edge lengths and orientations, what the probability is that the sum of projections will be between $y$ and $y+dy$. This leads to the computation of probability density function (*pdf*) for the sum of projections. If the *pdf* has a wide variance, then wide variations are also expected in the sum of projections by rotating the epipolar axis. For such cases, improvements achievable through *best axis* (as well as deteriorations one may get through *worst axis*) would be significant. From the *pdf*, it is therefore possible to qualitatively predict the improvements.

### 3.1 *Probability density function (pdf) for the sum of projections*

The *pdf* computation problem can be modelled as the popular *random walk* problem of statistics (Papoulis 1984). We have modified the classical random walk model to suit our problem and developed a forward random walk model as described below. In a classical two-dimensional random walk model, one can imagine that a person starts from the origin on the $x$–$y$ plane and moves along a straight line in any arbitrary direction for an arbitrary number of steps. The person then changes direction and walks along another straight line for another arbitrary number of steps and so on. After the person traverses $N$ such straight-line paths, the problem is to determine the probability that the projection of the straight line joining the starting and the end positions on any arbitrary axis lies within some interval. The classical random walk model does not impose any restriction on the direction of movement, i.e., the person is allowed to walk in any arbitrary direction between $-\pi$ to $\pi$. Hence, the projection can be positive or negative. In our problem, the number of matchable feature points, which is proportional to the sum of projections, is always positive. We therefore need to impose some restrictions on the traditional random walk to appropriately model our problem. To make the projections of each straight line segment positive, we restrict the movement direction between $-\pi/2$ to $\pi/2$. Hence, the person always moves forward, as shown in figure 7. Using our forward random walk model, we have derived the combined characteristic function for the sum of projections for $N$ linear segments as

$$\phi_y(\omega) = \frac{1}{2^N} \prod_{k=1}^{N} [J_0(\omega a_k) + jH_0(\omega a_k)], \tag{4}$$

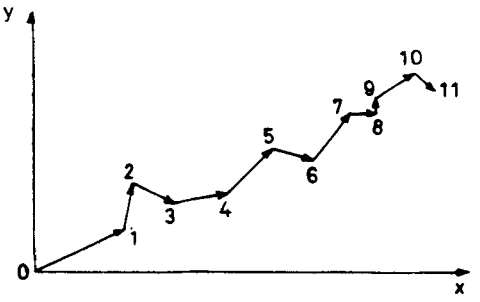


**Figure 7.** Example of our forward random walk.

where, $J_0$ and $H_0$ are the Bessel and Struve functions (Watson 1958) of order zero respectively. The *pdf* computation is performed by expressing it as the Fourier series summation of the samples of the complex characteristic function. Since, in this case, the *pdf* is a real and causal function, it is possible to express the *pdf* as a cosine series summation of the samples of characteristic function. The *pdf* is given by the following equation

$$f_y(mB_1) = \frac{2}{B}\sum_{n=0}^{P-1} \phi_y\left(-\frac{2\pi n}{B}\right)\cos\left(\frac{2\pi mn}{P}\right). \tag{5}$$

3.1a *Mean and variance of the probability density function*: For the simplest case of $N = 1$, the *pdf* is expressed by

$$\begin{aligned} f_{y1}(y1) &= 1/[\pi(a_0^2 - y_1^2)^{1/2}], \quad \text{for } 0 \leqslant y_1 \leqslant a_0 \\ &= 0, \text{ elsewhere,} \end{aligned} \tag{6}$$

the mean of $y$ is given by

$$\begin{aligned} y_{\text{mean}} &= \int_0^{a_1} y f_y(y)\,\mathrm{d}y, \\ &= 2a_1/\pi \end{aligned} \tag{7}$$

and the variance is given by

$$\begin{aligned} \sigma_y^2 &= \int_0^{a_1} y(f_y(y) - f_y(y)_{\text{mean}})^2\,\mathrm{d}y \\ &= [(1/2) - (4/\pi^2)]a_1^2, \end{aligned} \tag{8}$$

where $a_1$ is the length of the single edge element. If we extend these results for the sum of $N$ random variables, the mean and the variances are given by

$$y_{\text{mean}} = [2\Sigma_{k=1}^{N} a_k]/\pi \tag{9}$$

$$\sigma_y^2 = (1/N)[(1/2) - (4/\pi^2)]\sum_{k=1}^{N} a_k^2, \tag{10}$$

where $a_1, a_2, a_3, \ldots, a_N$ are the lengths of the individual edge elements.

### 3.2 *Figure of merit for best axis*

Just like there is a best axis for the sum of projections, there is a worst axis too, where the sum of projections will be least, leading to a more sparse set of feature points. A figure of merit, termed as *improvement factor* may be defined as follows:-

$$I_{\text{factor}} = [(y_{\text{best}} - y_{\text{worst}})/y_{\text{best}}] \times 100\% \tag{11}$$

where, $y_{\text{best}}$ and $y_{\text{worst}}$ are the sum of projections in the case of the best and the worst axes respectively.

## 4. Approach on stereo matching

For stereo matching with best axis alignment, we propose an approach as follows. As we do not have an *a priori* idea of the statistics of edge lengths and orientations, it is possible to start with any arbitrary axis. For a small angle stereo, it is reasonable to make an assumption that the edge statistics will be almost the same for the left and the right images. Thus, we can consider either of the two views at the starting position of the axis. For feature extraction, we can apply the Laplacian-of-Gaussian operator at the finest $\sigma$ and find the total number of zero-crossings. This number will indicate the sum of projections of zero-crossings in a direction perpendicular to the starting epipolar axis. The epipolar axis may now be rotated and for each incremental position of the axis, the total number of zero-crossing points are to be found out. Within the total rotational span between 0 and $\pi$, the position $\theta_{\max}$ which gives the maximum number of zero-crossing points is the best axis. Now, the cameras are to be re-aligned for the best axes and the stereo matching may be performed.

## 5. Results and discussions

It may be noted that the best axis concept will be more effective for scenes having smaller numbers of linear edge segments. If its use is restricted to a very limited environment as in robotics and machine vision applications, where the number of objects in a scene is limited, edges and depth cues are not in abundance and consequently, the set of feature points is sparse, alignment of cameras to the optimal axis will improve the reconstructed 3-D surface by adding more feature points.

For a limited number of $N$, we tried to establish the validity of best axis by conducting a computer simulated probabilistic experiment, which we prefer to call the *random stick experiment*. The experiment is described below.

We pick up a set of sticks of varying lengths and throw them at random. We then compute their sum of projections on any reference axis. The axis is rotated between $-\pi/2$ to $\pi/2$ and the sum of projections $y$ is computed for every position of the axis. We record $y_{\text{best}}$ and $y_{\text{worst}}$ and compute the improvement factor as defined in (11). The experiment is repeated a large number (1000) of times and histograms of improvement factor plotted. By repeating the experiment a large number of times, the histogram would indicate the factor of improvement one can expect. The results are shown in the subsequent graphs.

Figure 8 shows a typical distribution of sticks in one of the occurrences of the experiment. In figure 9, we show the improvement factor for a number of sticks, equal

**Figure 8.** Sticks distributed at random.

to 80, of constant length. The average improvement factor is 16%. If we now consider sticks having a uniform distribution of lengths, we can expect an increase in the improvement factor. This can be confirmed from our results shown in figure 10. The average improvement factor is around 20% and hence, it is better than that for sticks of constant lengths. In our next experiment, the results of which are shown in figure 11, we increased the total number of sticks to 150. The average impovement factor dropped to 14%. Average improvement factor was computed for $N = 80$, 120, 150, 200 and 250 with sticks having uniform distribution of length. The results are shown in figure 12. It may be noted that even for $N = 250$, the improvement factor is around 11%. A scene having number of linear segments equal to 250 can not be termed as a sparse depth map scene, but even for such a scene, optimal axis is worth trying.

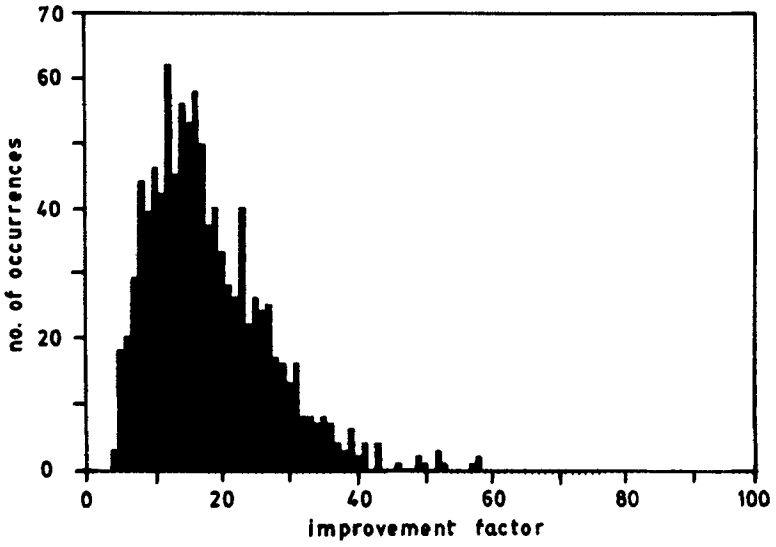


**Figure 9.** Results of random sticks experiment. $N = 80$ (constant length).

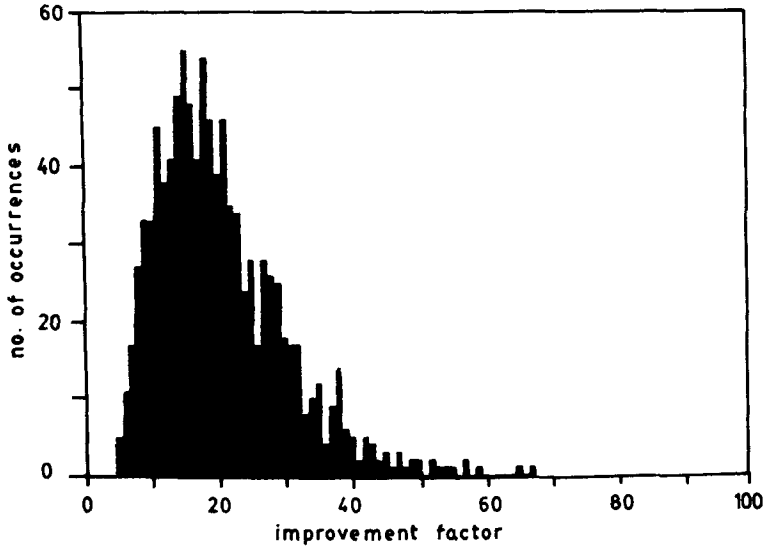


**Figure 10.** Results of random sticks experiment. $N = 80$ (uniformly distributed length).

As an illustrative example of depth map improvements, we have considered a synthetic image of a rectangular object. If the depth map is only defined along two of its sides (as is the case if the other two sides are oriented along the epipolar axis), we are unable to perceive the surface, as shown in figure 13, whereas, by orienting the cameras in the optimal axis, it is possible to get the depth map along all its four sides and the interpolated surface becomes more realistic, as shown in figure 14.

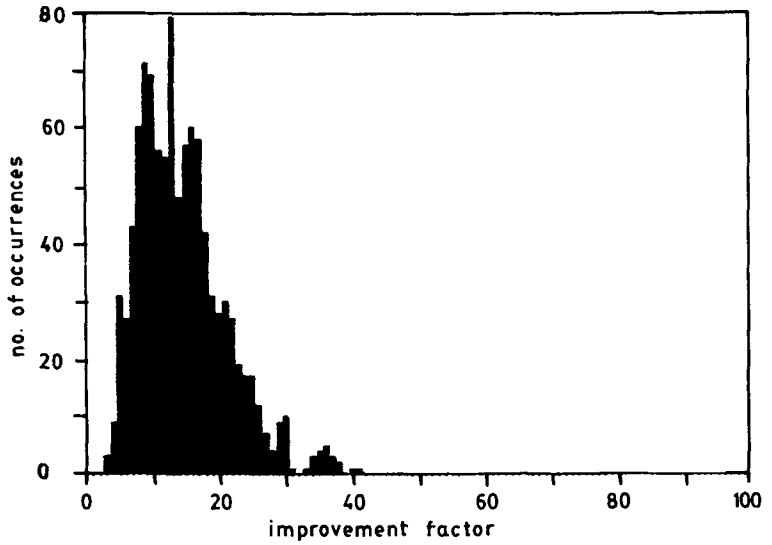


**Figure 11.** Results of random sticks experiment. $N = 150$ (uniformly distributed length).

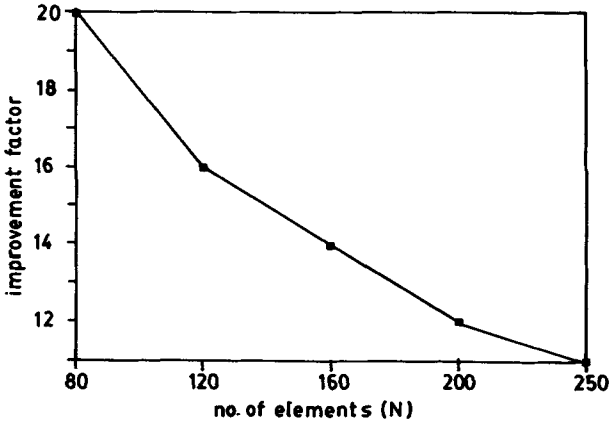


**Figure 12.** Average improvement factors $N$.

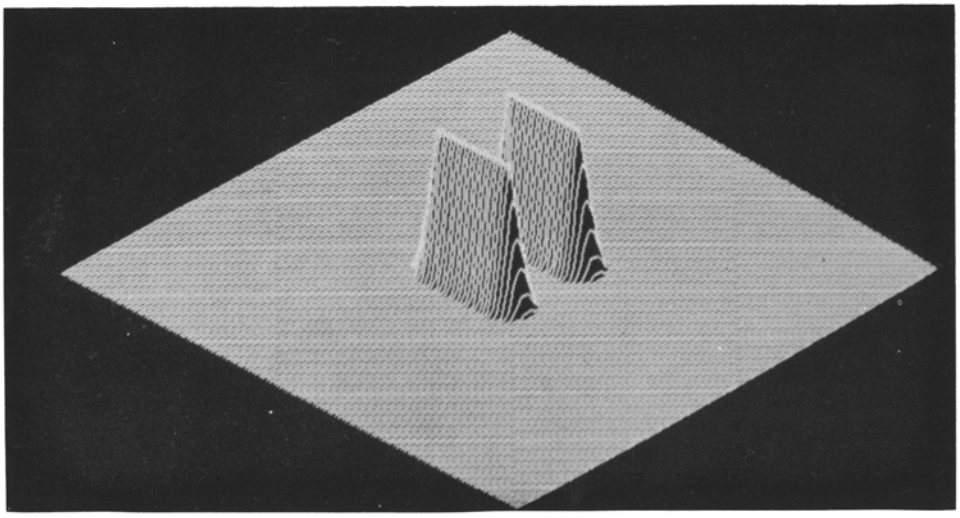

**Figure 13.** Example of the interpolated surface of a rectangular object (depth map defined along two sides only).

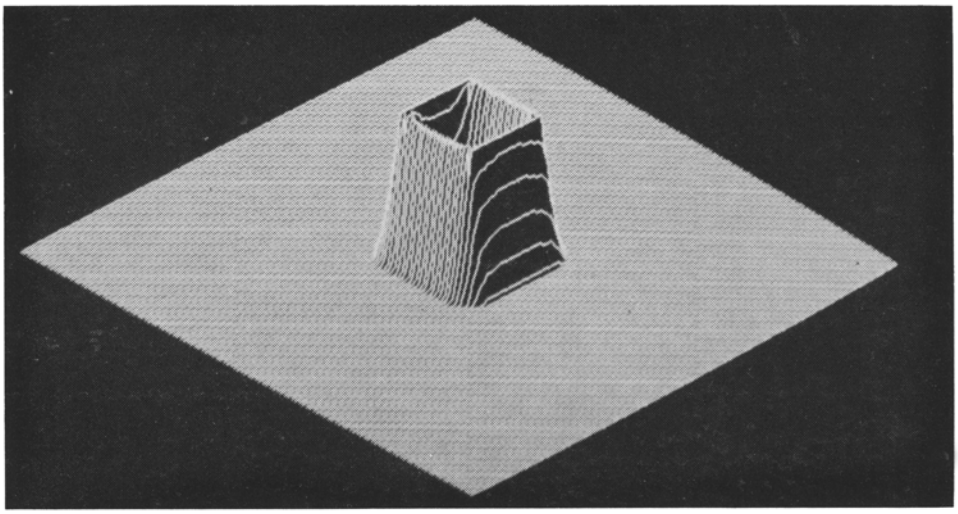

**Figure 14.** Example of the interpolated surface of a rectangular object (depth map defined along all the four sides).

## 6. Conclusions

We have proposed a new concept of optimal axis for stereo cameras in order to reduce the sparsity of depth map. The approach appears to be a good alternative to trinocular stereo. We have confirmed the existence of optimal axis through probabilistic analysis and statistical experiments. The proposed approach can be used in machine vision applications where flexibility to rotate the axes of the cameras can be made feasible. In this approach, we are actually taking multiple views of the same object, although we have not exploited the advantages one may obtain from it. One is the increased field of view and the other is reduction of occlusion. This may be the direction for further research.

## References

Ayache N, Faverjon B 1985 Fast stereomatching of edge segments using prediction and verification of hypothesis. *Proc. Comput. Vision Pattern Recogn.* pp. 662–664

Baker H H, Binford T O 1981 Depth from edge and intensity based stereo. *Proc. 7th Int. Joint Conf. on Artif. Intell.* pp. 631–636

Grimson W E L 1981 A computer implementation of a theory of human stereo vision. *Philos. Trans. R. Soc. (London)* B292: 217–253

Grimson W E L 1983 An implementation of a computational theory of visual surface interpolation. *Comput. Vision Graph. Image Process.* 22: 39–69

Grimson W E L 1985 Computational experiments with a feature based stereo algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-7: 17–34

Horaud R, Skordas T 1989 Stereo correspondence through feature grouping and maximal cliques. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-11: 1168–1180

Ito M, Ishii A 1986 Three view stereo analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-8: 524–532

Ji C X, Huang Z P 1988 Stereo match based on linear feature. *Proc. Int. Conf. Pattern Recogn. (ICPR)*, pp. 875–878

Lloyd S A, Haddow E R, Boyce J F 1987 A parallel binocular stereo algorithm utilizing dynamic programming and relaxation labelling. *Comput. Vision Graphics Image Process.* 39: 202–225

Marr D, Poggio T 1977 A theory of human stereo vision, MIT Artificial Intelligence Memo No. 451, November

McIntosh J H, Mutch K M 1988 Matching straight lines. *Comput. Vision Graph. Image Process.* 43: 386–408

Medioni G, Nevatia R 1985 Segment based stereo matching. *Comput. Vision Graph. Image Process.* 31: 2–18

Ohta Y, Kanade T 1985 Stereo by intra and inter-scanline search using dynamic programming. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-7: 139–154

Ohta Y, Watanabe M, Ikeda K 1986 Improving depth map by right angled trinocular stereo. *Proc. Int. Conf. Pattern Recogn.'86*, pp. 519–521

Papoulis A 1984 *Probability, random variables and stochastic processes*, 2nd edn. (McGraw Hill)

Stewart C V, Dyer C R 1988 The trinocular general support algorithm: A three camera stereo algorithm for overcoming binocular matching errors. *Proc. Int. Conf. Comput. Vision'88*, pp. 134–138

Watson G N 1958 *A treatise on the theory of Bessel functions*, 2nd edn. (Cambridge: University Press)