TRANSLATIONAL MOTION ALGORITHM WITH GLOBAL FEATURE CONSTRAINTS

Igor Pavlin
Edward Riseman
Allen Hanson

COINS Technical Report 86-58

December 1986

This work was supported by the Defense Advanced Research Agency, DARPA, under grant N00014-82-K-0464.

TRANSLATIONAL MOTION ALGORITHM WITH GLOBAL FEATURE CONSTRAINTS

Igor Pavlin, Edward Riseman and Allen Hanson

This work was supported by the Defense Advanced Research Projects Agency, DARPA, under grant N00014-82-K-0464.

ABSTRACT

This paper presents an evaluation and improvement of an algorithm [LAW84] for the recovery of translational motion parameters of a sensor moving through a static environment. The algorithm computes the direction of translational motion using a search through a two-dimensional parameter space for values that are optimal in terms of their consistency with the image dynamics of features over a pair of frames. The direction of translation is computed from the minimum of an error surface constructed in the two-dimensional parameter space. We evaluated the effectiveness of the algorithm using synthetic images on the basis of the number of feature points matched between frames, the relative angle between the camera orientation and the direction of translation, the influence of noise, and computational cost.

The algorithm is robust across a very wide range of camera translations, using as few as eight feature points. Accuracy generally improves when the number of feature points is increased to 16. The algorithm experiences difficulties when the angle between the direction of translation and the direction of view approaches ninety degrees. In such cases the error surface in the vicinity of the minimum is flat causing ambiguity in the location of the minimum. In addition, fluctuations in the error surface due to different sources of noise become more important if the error surface around the minimum is shallow.

A significant improvement in the robustness and computational speed of

the algorithm is achieved by incorporating a new method for the search for the minimum of the error surface. The approach is based on an assumption of global smoothness of the error surface, which allows the use of a regularization constraint. The approach uses a multi-step hierarchical search beginning with a coarse sampling of the error surface (the first step). In the second step, using a smaller neighborhood and a finer resolution, the error surface is sampled around the minimum found in the previous step. In this step a more faithful representation of the error surface near the exact minimum is obtained. In the third step the error surface obtained in the second step is smoothed using a regularization technique. An error surface with much finer resolution and lower noise is found, whose minimum is determined by the *shape* of the error surface found in the second step. The second and third step can be repeated using the previous minimum until sufficient accuracy is achieved. Lawton [LAW84] used a global search technique, followed by a local hill-climbing technique, in the search for the error minimum. The local-hill climbing technique failed in cases of non-convex, i.e., noisy, error surfaces. Our hierarchical approach with smoothing results in more robust and faster algorithm. We believe that these translational motion results will be qualitatively similar to other forms of camera motion, such as pure rotation and motion constrained to a plane.

1. INTRODUCTION

1.1 Background

This paper focuses on analysis of the motion of a camera through a static environment. In particular, we examine the vector field describing the changes in the positions of the images of environmental points over time. Gibson [GIB50] first introduced the idea of "optic flow", which is essentially a vector field representation of the image displacement of environmental points induced by motion. Gibson observed that the patterns and extent of displacements of image points should provide some information about the sensor motion and environment. We will refer to the vector field as the displacement field. For more details about the computation of displacement fields, see [BAL82, ANA85, ANA86].

The first part of this paper describes the results of an extensive testing of the Lawton algorithm [LAW84] for the recovery of translational motion parameters, based on the idea of consistency of image feature dynamics. It should be noted that a motion search algorithm in this same class of algorithms was developed earlier by Williams [WIL80]. The later uses consistency of image dynamics, as well as surface depth computation, in order to recover both the translational motion parameters of the camera and the depth of image regions under the assumption that each region

represents a planar surface oriented at one of directions orthogonal or parallel to the image plane of the camera. Both of these algorithms avoid computation of the entire image displacement field prior to the recovery of translational motion parameters. They compute motion parameters and displacements simultaneously. In that respect, they are different from the algorithms that use the entire displacement field to recover parameters of motion such as [ADI84], or generalized Hough transform approaches in [AGG81], [BAL81], [ORO81], [DAV83].

The axis of translation of a sensor may be uniquely specified by the point where it intersects the image plane (the axis of translation is assumed to start from the focal point). The intersection of the axis of translation and the image plane is called the focus of expansion/contraction (abbreviated to FOE/C), see [BAL82]. The search for the translational axis (FOE/C) is done by hypothesizing many possible positions of the FOE/C on the image plane. The consistency of the location of the hypothesized FOE/C with the displacement of several interesting points is checked and expressed as an error value. Error values found at these hypothesized positions of the FOE/C on the image plane determine an error surface. The search for the minimum error is then equivalent to the search for the most probable position of the FOE/C. Equivalently, the error surface minimum determines the camera translation which is most consistent with the displacement field. The robustness and accuracy of the method are due to its unique combination of local and global characteristics.

The local nature of the technique and its accuracy are the consequence of correlation measurements of intensity values between a window in one image and a window in a subsequent image. These windows are centered around corresponding "interesting" features in two images. The correlation between these windows is used to find the displacement of a feature from one image to another (from one instance in time to another), along the line between the hypothesized FOE/C and the feature point, and to associate with this displacement a measure of confidence, or, equivalently, a measure of error. The smaller the error, the greater the confidence about the orientation and magnitude of the feature displacements.

The global nature of the technique arises from the contribution of many points which ideally are spread across the image at different relative orientations to the FOE/C; the points selected should be distinguishable or trackable, and if few are used should be selected via an interest operator [LAW84, MOR81]. The solutions are stable despite the presence of noise, false feature matches, occlusion of features, and other causes of unreliable information. The reason is that the search for the position of the FOE/C depends on the set of points whose motion is jointly constrained. Displacements not in agreement with the majority of other displacements will, in effect, be disregarded during the search. The direction of displacement of a local feature as suggested by the optimal position of the FOE/C should be but is not necessarily consistent with the displacement obtained by local computations. The determination of the optimal position of the FOE/C, which is consistent with the image dynamics of interesting points over a sequence of images,

is in a sense an averaging process of the contribution of displacement vectors.

This paper considers only translational motion. However, there are two closely related cases of camera motion in a static environment which should be amenable to a similar approach [LAW84] and which should show equivalent results. One of them occurs when the camera motion is constrained to a known plane, where the camera rotation is about the axis perpendicular to that plane. The other is pure rotation about a fixed axis, described by only three parameters. While the algorithm is efficient in low-dimensional parameter spaces (specifically the 2D parameter spaces of pure translational motion or motion constrained to a plane, or the 3D parameter space of pure rotations), it has yet to be proven better in higher-dimensional parameter spaces, i.e., for motions with arbitrary sensor rotation and translation.

1.2 Summary of Paper

In the first part of this paper (Chapter 2) Lawton's algorithm for constrained search for the translational motion parameters is described and tested on both synthetic images and outdoor motion sequences. The influence of the following parameters is examined:

1. The number of feature points in the image plane which are matched between two frames;
2. The accuracy with which the direction of translation can be recovered as a function of the relative orientation of the camera line of sight to the direction of motion;
3. The resolution of the sampling of the FOE/C during the global search;

4. The size of the window used for feature correlation between frames;
5. The step size of feature displacements during feature matching;
6. The efficiency of computation;
7. The robustness of the method with respect to uncorrelated noise.

The experiments show that the search for the minimum of the error surface, used by Lawton, failed in some cases [PAV85]. The algorithm started with a sparse sampling of the error surface followed by a finer-resolution local hill-climbing search for the global minimum of the error surface. There are problems and inefficiencies in this search algorithm. Namely, if the global search in the two-dimensional parameter space is too sparse, there is a possibility of missing the correct minimum. This happens when the error surface around the minimum is relatively flat and exhibits small fluctuations. The local hill-climbing technique fails because the assumption of monotone convexity of the error surface is violated in such cases. The local search gets "trapped" in one of the locally convex regions, not necessarily at the real minimum. These cases of ambiguity cannot be resolved by any local search method [PAV85].

On the other hand, if the global search is performed with much finer resolution, there is unnecessary computation involved in computing many error values in which we are not interested. Indeed, we are interested only in the position of the minimum of the error surface. This aspect becomes more and more important when the number of interesting points followed from one frame to another increases and when the speed of the computation is critical. Thus, it is not feasible to solve

the search failure problem with higher FOE/C sampling rate.

In the second part of the paper we present an improvement of the algorithm by imposing a smoothness constraint on the error surface. The suggested approach is based upon a strong conjecture, which is supported by experimental evidence, that the error surface should be smooth if more than a few points are tracked between a pair of frames. The idea of lower spatial resolution, which can be used in conjunction with the smoothing of the error surface, is consistent with the general perception of how the motion is detected in biological systems [ULL81]. Smoothing of the error surface corresponds to a spatial smoothing of the image intensity values, and thus to a decrease of the acuity of perception derived from temporal analysis of multiple frames. Smoothing of the error surface then reflects our belief that in "the guidance mode" a high acuity of perception is not essential for the evaluation of motion parameters. A high acuity of perception is undoubtedly important for the estimation of the environmental structure.

The assumption of smoothness of the error surface is viewed as a kind of regularization constraint [POG84]. To simulate a smooth error surface which still passes close to the error values computed by displacement vectors of interesting features, one can use a variety of regularization functionals. We have chosen one which is similar to the regularization functional used in the work on extraction of depth maps from stereo [GRI81]. The methodology can be visualized as a problem of finding the equilibrium position of a thin metal plate forced to pass close to a

given sparse set of points in the space. The best plate (surface) is the one which follows the given data within some tolerance and at the same time has the smallest possible potential energy of deformation.

The constraint of smoothness that has been introduced leads to a variational problem which was solved numerically by a method of conjugate gradients [HES80]. The regularization method and the conjugate gradient technique used in this paper are described in Chapter 3.

Not only should the robustness of the algorithm be improved by the smoothing modification, but the computational speed should also be improved since the expensive computation of the error surface via a high sampling rate can also be significantly reduced. Only a few sparse error values near the minimum are necessary to determine the overall shape of the error surface. The shape of the error surface determines the position of the minimum and that happens with reasonably satisfactory accuracy. Interleaving the computation of the local error surface with its smoothing results in a robust end efficient algorithm. The experimental results and their discussion are presented in Chapter 4.

A similar idea appeared recently in the work by Medioni [MED85], but in the context of Adiv's work [ADI84]. The authors have chosen a regularization functional which reflects constraints applied directly to the displacement field; however, the chosen constraints are only partially justified. We do not introduce constraints that early, but only when the error surface is already created, in the last step of the

evaluation of motion parameters.

2. EARLY EXPERIMENTS

ROBUST RECOVERY OF TRANSLATIONAL MOTION PARAMETERS USING GLOBAL FEATURE CONSTRAINTS

In this chapter we introduce more formally the method of analyzing a displacement field for the recovery of the translational motion parameters of a camera moving in a static environment [LAW84]. Then the results of experiments which were designed to test the robustness and limits of the algorithm are described. The analysis of this set of experiments leads to some refinements for which the theoretical foundation is given in Chapter 3 and for which the results can be found in Chapter 4.

2.1 Translational Motion and Displacement Fields

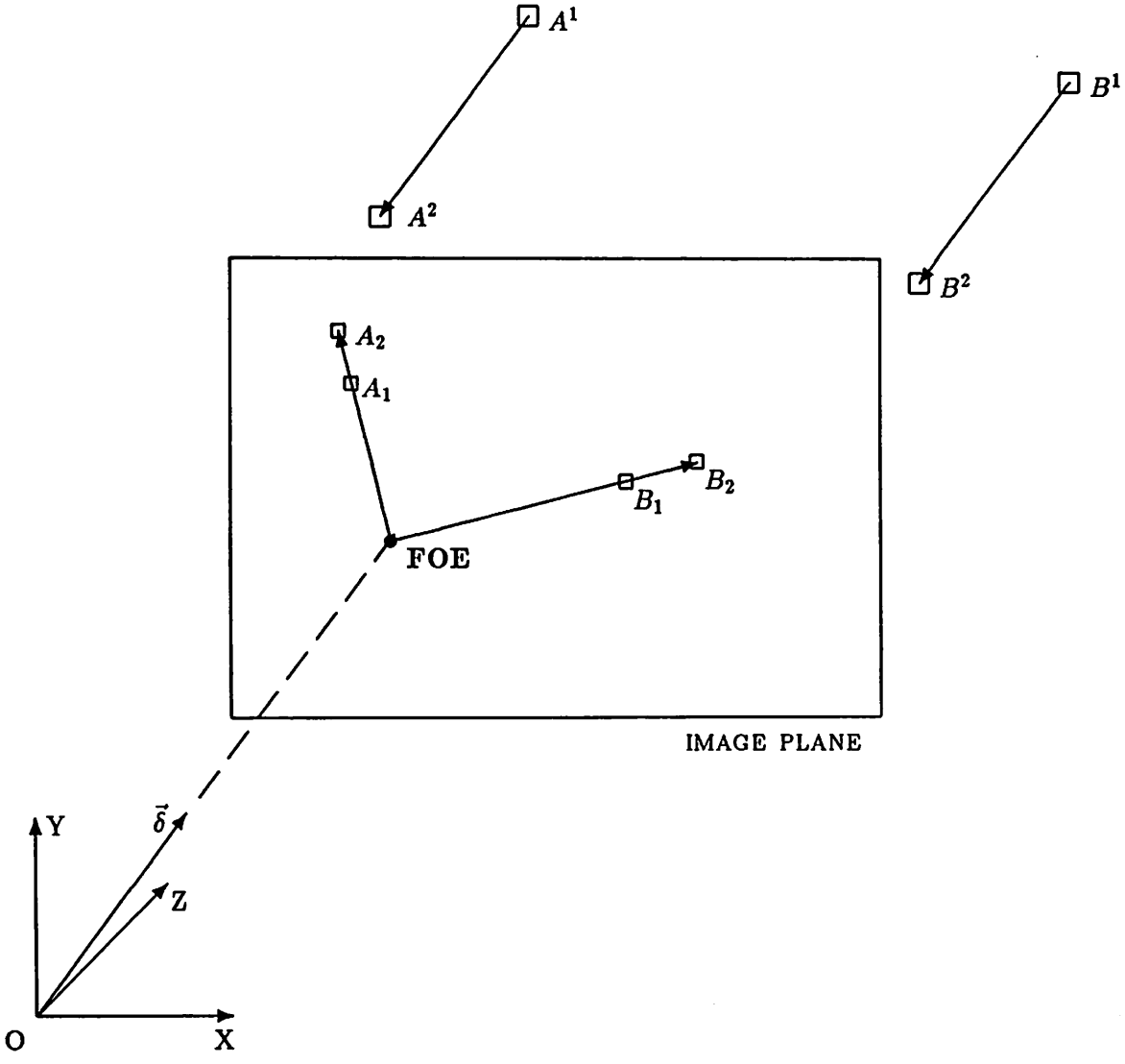
In this section we present a brief review of the FOE/C algorithm and the general ideas that are necessary to follow the discussion of the experiments in the following section; the reader is encouraged to consult [LAW83, LAW84] for more details.

In the case of pure translational motion of the camera in a static environ-

ment, the intersection of the image plane and the vector describing the translational motion of the camera focal point is called the FOE/C. The displacement vector of any image feature lies on the radial line connecting the FOE/C and the feature (see Figure 1).

As few as two image points would be sufficient to identify the correct translational motion, providing that the non-collinear displacements of the two points can be accurately found. In practice, however, the proper correspondence between the projection of environmental points in two or more consecutive frames is difficult to find. Typically, the correspondence between similar parts in two images is established using techniques such as scalar valued correlation functions [MOR81, AGG81, LAW83, ANA86] or symbolic token matching of key image events [FAN82, DAV83, SMI86]. There are a variety of problems that make such mechanisms unreliable: the size of the local area to be searched, the presence of noise in the image, occlusion of surfaces, insufficient resolution of the image, etc.

Redundancy and global constraints on the feature dynamics across frames can be used to overcome these difficulties [BHA85]. Since the displacement vectors of all features in a static environment (i.e., no independently moving objects) are constrained to emanate from the FOE/C, the use of additional features should increase the confidence about the correct position of the FOE/C. Use of redundant features would compensate for features that provide weak or incorrect information (e.g., features in low contrast or homogeneous regions, features at occlusion bound-



The camera model and the displacement field produced by two environmental points A and B during the motion of the camera specified by the translational vector $\vec{\delta}$. The Z axis is along the line of sight and passes through the center of the image plane. The origin of the camera centered coordinate system is at the camera's focal point and images of environmental points are obtained by the central projection.

Figure 1: Camera model and FOE/C.

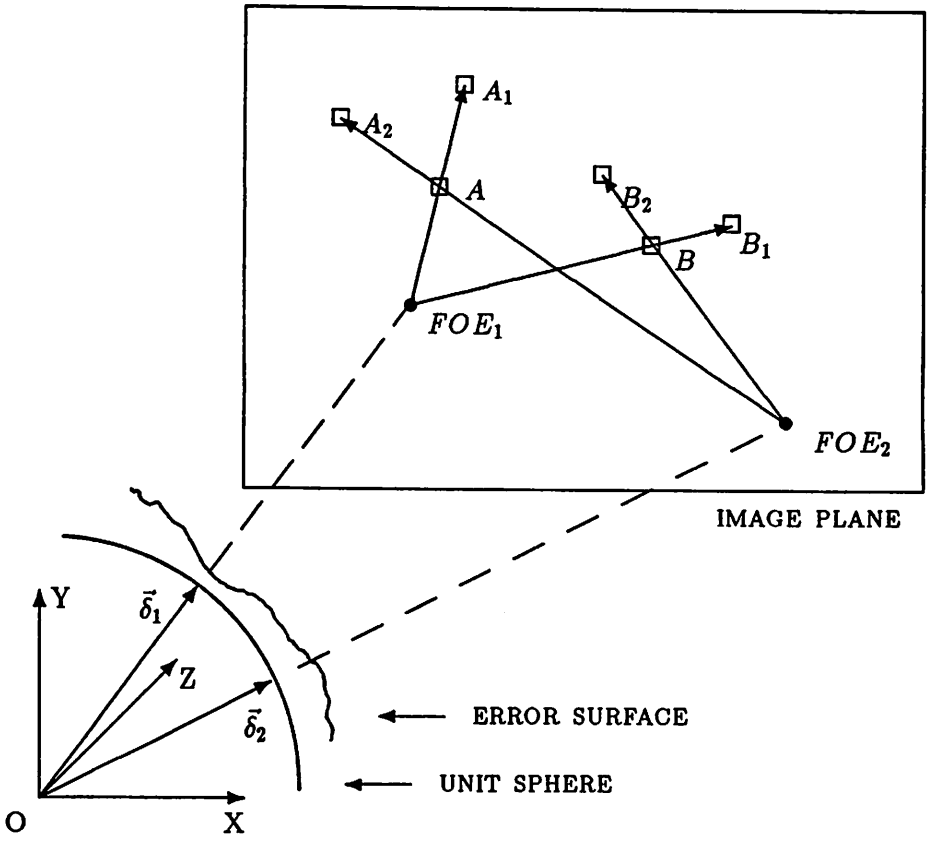
aries). However, the use of too many features is computationally expensive.

2.2 Error function

The position of the FOE/C can be obtained in two basic steps. The first step is the extraction of features and the second step is the search for the position of the FOE/C which is most consistent with their displacements. The whole search procedure is symbolized in Figure 2.

The feature extraction process is responsible for the extraction of distinguishing points which can be tracked from one frame to another. Contour points of high curvature are a good choice because they are less likely to produce ambiguous matches in succeeding frames. The contours can be extracted using a variety of techniques: thresholding, zero-crossing, boundary curvature measures, local contrast measurements, etc., [LAW83, LAW84].

The direction of translation is then found by a search across hypothesized FOE/C positions in the image plane. A hypothesized position of the FOE/C and a feature in the first frame determine a line. The feature is expected to move in the second frame along this radial line not more than certain number of pixels. To determine the extent of the feature displacement, a correlation measure between a window centered at the position of the feature in the first frame and a window moving along the radial line is computed. If a feature, say A, has moved in the subsequent image to the position A_1 some distance d along the radial line, then a window around A is expected to correlate nearly perfectly d units along the radial line in the next frame. Equivalent information is given in the error measure between



The position of the FOE/C in the image plane is hypothesized. For the position FOE_1 the best correlation matches for features A and B are found along radial paths from FOE_1 through A and B at, let us assume, positions A_1 and B_1, respectivelly. FOE_1 itself is at the intersection of the image plane and the translation vector $\vec{\delta}_1$. The position of the FOE_1 also uniquely determines a point on the unit sphere, centered at the camera focus, where the translation vector $\vec{\delta}_1$ pierces the unit sphere. At that point, the error associated with the best match for features A and B is recorded. For a different hypothesized position of the FOE/C, for example, FOE_2, the best match for features A and B might be found along a radial path at the positions A_2 and B_2. The error is expected to be large if the assumed position of the FOE/C is incorrect. This process is repeated for many hypothesized positions of the FOE/C resulting in an error surface over the unit sphere. The position of the minimum of the error surface determines the correct axis of translation (or, equivalently, the correct FOE/C).

Figure 2: Search for the FOE/C.

features A and A_1:

$$error(feature_A, feature_{A_1}) = 1 - corr(feature_A, feature_{A_1}). \qquad (2.1)$$

Several measures were examined by Lawton for feature matching: the normalized correlation, the Moravec correlation (which was used in this paper) [MOR77], and the normalized absolute value difference. These correlation functions differ in speed and precision. The size of the $n \times n$ correlation window can affect accuracy. In addition, alternative forms of interpolation should be considered, because windows from one frame can be displaced to positions with non-integer coordinates in the second frame (i.e., the windows in the second frame are not aligned with pixels). Thus, approximate matching at integral pixel positions can be used, but when higher precision is required, some form of interpolation should be used. In this work a bilinear interpolation [ABR68] is used to compute the intensity values at non-integer pixel positions.

As an alternative to correlation matching, we are currently examining the use of symbolic features, or "tokens" (interesting points with a set of attribute-value pairs [SMI86]). The replacement of correlation matching with symbolic matching has the potential of significantly increasing the speed of computation[1], but it was not used here.

If the position of the hypothesized FOE/C is very close to the correct FOE/C, then most of the features will return good matches and small errors, resulting in a

[1] Lawton, private communication

very small total error for that particular position of the FOE/C. If, on the other hand, the position of the hypothesized FOE/C is far from the correct position, the computed total error will be significant, since many features will return a poor match. By hypothesizing FOE/C positions across the image plane, an error surface can be constructed, with the minimal error expected at the position of the correct FOE/C. By representing the orientation of the axis of translation and associated error values on the unit sphere centered at the camera focus, rather than in the image plane, a more uniform sampling of the hypothesized position of the FOE/C can be achieved.

In the experiments that follow, the hypothesized position of the FOE/C on the unit sphere is specified by angles ϑ and φ, where ϑ is the angle between the translational axis and the Z-axis (the Z-axis is chosen along the line of sight), and φ is the angle between the projection of the translational axis on the X-Y plane and the x-axis (left-handed polar coordinate system). We stress here that the angle ϑ is the key parameter, since it represents the deviation of motion from the line of sight. The variation of the results as a function of φ is expected to be less significant, because it is the distance of the FOE/C from the center of the image that matters and not its angular position.

The error surface on the unit sphere can be constructed starting with a coarse sampling over polar angles. The resolution of the sampling is one factor determining the precision of the method and speed of the search. The grid spacing in the (ϑ, φ)

coordinate system in the set of experiments described in Section 2.3 was roughly 45, 22.5, and 11.25 degrees.

Once the coarsely sampled error surface is found, the search for the minimum of the error surface is continued locally around the minimum found by the coarse sampling. (The smallest error value of the coarse global search is taken as the initial guess for a finer resolution local search). Lawton used a very simple hill-climbing technique for the local search under the assumption that the error surface is a strictly convex function in the vicinity of the global minimum. Error values at eight new neighborhood points around the current minimum were computed. The new neighborhood points were chosen at half the distance from the previous neighbors. The point with the smallest error among these nine points was found and became the new starting point for the next iteration of the local search. The new neighborhood was searched, again with a radius equal to one half of the previous spacing between the neighbors. The procedure was repeated until the radius of the neighborhood became smaller than a limiting value δ_∞. If the error surface near the minimum were convex, this would guarantee that, when the local search was completed, the position of the minimum of the error surface would be found with an accuracy of $\pm\delta_\infty$. The purpose of the global search was to localize the search for the minimum of the error surface in such a convex neighborhood.

2.3 Description of Experiments

We describe now an initial set of experiments that we performed with the motion algorithm just discussed. This is followed by a discussion of the results (Section 2.4) which sets the stage for the improvements described in Chapter 3 and Chapter 4. The experiments were performed using the VISIONS image operating system [HAN84]. The goals of these experiments were:

- to determine the minimum number of features required for robust recovery of the motion parameters
- to determine the susceptibility of the method to image noise
- to determine the precision of the recovery of translational parameters
- to determine the applicability of the method over the range of all possible translational motions.

Initially, we analyzed the algorithm on the same natural outdoor image sequence (the road-sign images) used by Williams [WIL80] and Lawton [LAW83]. These images have a spatial resolution of 128×128 pixels. We repeated these experiments with only a slight discrepancy from the values determined by Lawton, probably due to different features. The results were satisfactory even with only four feature points, and the use of eight feature points was as effective as 64 feature points. The global search was performed with a resolution of 22.5° and the local search parameter was $\delta_\infty = 0.005$ radians ($\approx 0.29°$). The total number of the

hypothesized FOE/C positions was around 160 in each of the experiments. The results of the experiments are shown in the following table:

Results for the Road-Sign Sequence

Exper. No.	No. of features	Translational axis coordinates		
		x coord.	y coord.	z coord.
1	4	-0.8451	-0.4353	0.3101
2	8	-0.8338	-0.4150	0.3638
3	16	-0.8226	-0.4285	0.3736
4	32	-0.8439	-0.4218	0.3313
5	64	-0.8439	-0.4218	0.3313
6*	140	-0.8373	-0.4204	0.3493
Difference Exps. 2 and 6		-1.19%	-2.38%	2.86%
Difference Exps. 5 and 6		-0.79%	-0.33%	5.15%

*[LAW84]

The exact translational axis for the road-sign images is not known. Therefore, the absolute precision of the search is unknown, and it is not clear whether there is any systematic deviation from the correct translational axis, nor whether the algorithm would perform as well for other viewing directions relative to the direction of motion.

We constructed synthetic images to assure controllable experiments. One of the images is shown in Figure 3. It is a part of a motion sequence created with a computer graphics system which incorporates ray-tracing techniques [WHI80]. The sequence was a reasonable substitute for real-world images since light was handled somewhat realistically. For example, shadows, specular reflection, decrease of the light intensity with distance, and the physical laws of reflection and refraction were

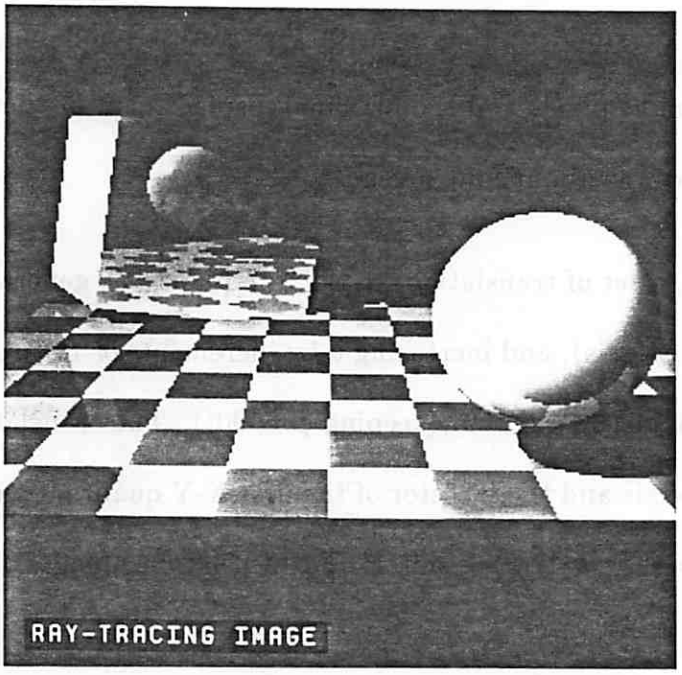


Figure 3: Synthetic image used in experiments.

incorporated. Synthetic images permit accurate control of camera motion in a static environment. Evaluation of the actual algorithm performance on real world images should be performed as data bases of those motion sequences become available.

The images have a spatial resolution of 256×256 pixels. In these experiments the camera position and motion were known exactly. Feature points on this image appeared at some of the corners of checkerboard squares, on the prism, and on the sphere boundaries. The camera field of view was $43.6° \approx 45°$ as opposed to $90°$ in Lawton's experiments. The line of sight was inclined $15°$ towards the horizontal

checkerboard. In all experiments the displacement of the camera was calculated so that the maximal displacement of the nearest point on the checkerboard did not exceed eight pixels and the maximal displacement of the furthest point on the checkerboard was about four pixels.

First, a set of translations in the Y-Z plane was generated, starting from the line of sight (Z-axis), and increasing ϑ by increments of 15° from 0° to 90°, and then by increments of 30° to 180° (keeping $\varphi = 90°$). The translations in the directions along the X-axis and the bisector of the first X-Y quadrant were also analyzed (i.e., for $(\vartheta, \varphi) = (90°, 45°)$), as well as the translation along an "arbitrary" direction specified by the angles $(\vartheta, \varphi) = (62.4°, 64.8°)$.

Parameters were varied from a default set of parameters. The default parameters used were: correlation window displacements of 1 pixel (with a maximum allowed displacement of 10 pixels), correlation window size of 7×7, sampling in polar coordinates of roughly every 45°, and a local search precision of $\delta_\infty = 0.005$ radians. Tables 1 through 6, given in Appendix A, present the results for images without noise. Table 7 presents results when uncorrelated noise of varying strength is added to the first image.

The number of hypothesized FOE/C positions during the entire search (combined global and local) was about 60, 160, and 375 (Tables 1, 2, and 3, respectively). The construction of the error surface was the time-consuming part of the search and it is very important from the point of view of efficiency to minimize the number

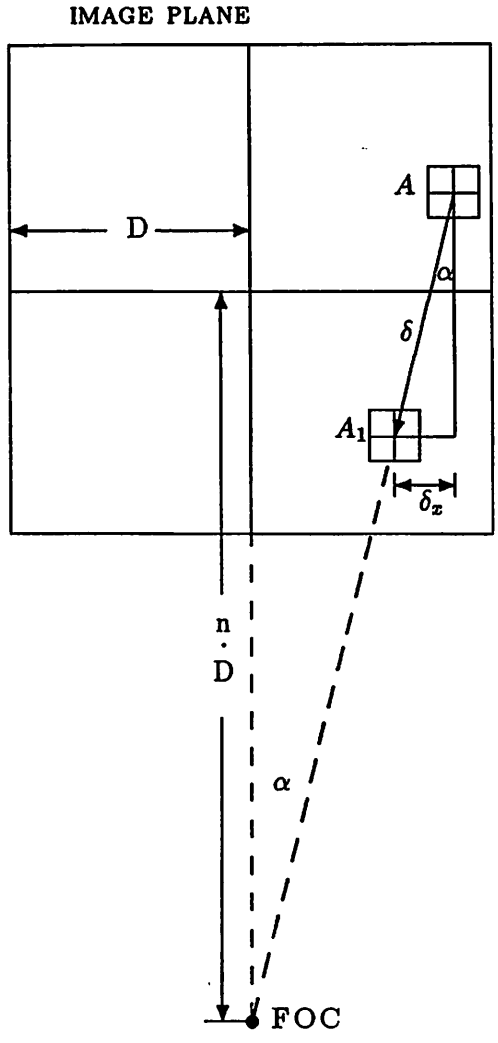
of hypothesized FOE/C positions. We would like to stress here, however, that the algorithm and the environment in which it was performed were designed for programming flexibility in an experimental development process, and not optimized for speed. Thus, the computational speed can be greatly improved.

2.4 Discussion of Early Experimental Results

Before we consider the results of individual experiments, several general remarks about the shape of the error surface for a camera moving in a static environment are in order:

1. The greater the number of features, the "smoother" the error surface is expected to be. With fewer features, the contribution of one feature to the total error value is greater. If some features are "weak" or if their number is small, the error surface is rougher.
2. The use of more features proportionally increases the computation time.
3. The closer a feature is to the FOE/C, the less reliable its contribution.
4. Features further from the FOE/C usually have larger displacements (although the displacement is also a function of the depth of the environmental point) and therefore predict more accurately the orientation of the radial line on which the FOE/C should lie.
5. Since the position of the FOE/C is initially unknown, the features should be spaced more or less uniformly throughout the image plane.

Let us consider the case of camera translational motion parallel (or almost parallel) to the image plane. This corresponds to the situation when $\vartheta \approx 90°$ to the direction of the line of sight. In this case the FOE/C is far away from the image center (Figure 4). We will show in the following brief analysis that the error surface in these situations is flat and that the FOE/C cannot be accurately recovered.



When the FOC is **n** times the image size away from the image plane, a displacement of a feature d units towards the FOC causes the lateral displacement of the feature of $d_x \approx d/n$.

Figure 4: Search for a distant FOC.

Let us assume that a window A in Figure 4 centered at (x, y) moved a distance d towards the FOC (analysis is similar for the FOE). Let us further assume that the FOC is on the Y-axis at the distance $n \cdot D$, $n \gg 1$ from the center of the image, where $2D \times 2D$ is the size of the image plane. Then, according to Figure 4, we have:

$$d_x/d = x/(n \cdot D + y) \approx x/n \cdot D \tag{2.2}$$

or for $x \approx D$,

$$d_x \approx d/n. \tag{2.3}$$

For example, for translations 75° and 85° away from the line of sight, the values for n are $n \approx 4$ and $n \approx 10$, respectively. From these equations we can see that the correlation function must be able to detect a change of feature position *in the x-direction* of the size d_x in order to accurately predict the position of the FOE/C on the Y-axis. However, since the size of the displacement d is limited, at best, by the image size, d_x is expected to be very small when the position of the FOE/C is far from the image center. In general, the displacement d is usually assumed to be small to avoid ambiguities during the matching process. If the correlation function is not able to detect lateral displacement of a feature of the size d_x, then the position of the FOE/C is unknown with an angular uncertainty of d_x/d. Thus, in the cases of translational motions more perpendicular to the line of sight there is a need for very accurate correlation measurements which often cannot be met. That is the reason why the error surface for these directions becomes flat around the minimum.

The sensitivity of the correlation function also depends on the window size. The averaging nature of a correlation function (being a sum of products) works against its sensitivity. Windows that are larger do not necessarily imply better results.

The tables presented in Appendix A are considered next. In these tables, the second column represents the directions (in polar coordinates) in which the camera is translated and hence are the values which the algorithm is supposed to recover. We refer to these directions as "correct" values. The other columns specify deviations (in degrees) from the correct values for experimental runs for 4, 8 and 16 feature points, respectively. The quantity $\Delta\Omega = \sqrt{\Delta\vartheta^2 + \Delta\varphi^2}$ is a measure of total deviation of experimentally obtained direction of motion from its correct value. It should be emphasized again, however, that the value of ϑ is the important motion parameter because it measures deviation of the direction of translational motion from camera orientation.

In some cases the local search fails and it is not possible to recover the correct axis of translation. These cases are marked with an A (for ambiguous) in the tables. Tables 1 through 3 show that the ambiguous results appear mostly for motions tending towards being perpendicular to the line of sight. In all these cases, the value returned by the algorithm was the coarse grid sampling point with the smallest error. Very often this was the direction $(\vartheta, \varphi) = (90°, 90°)$. Because of the noisy nature of the error surface, the assumption of the convexity of the error

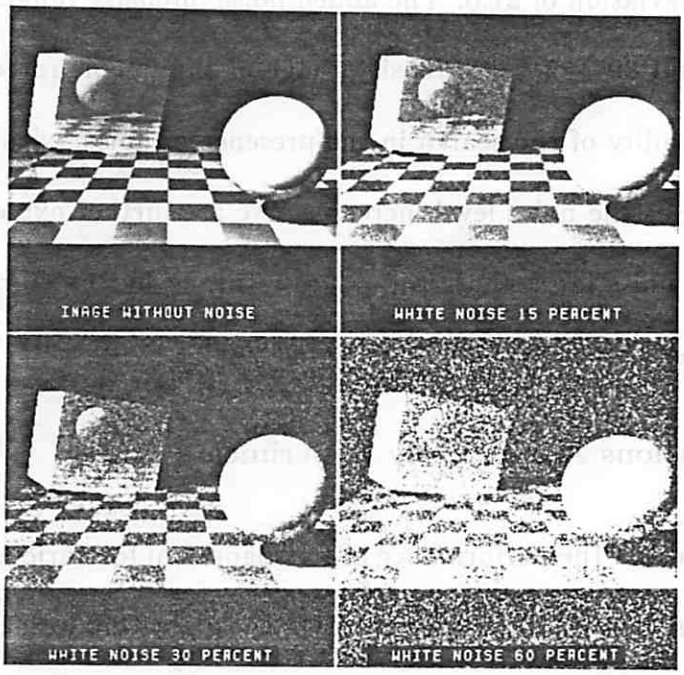
surface near the minimum, a necessary condition for the local hill-climbing search, is false. Rather, the local search returns a value close to the value returned by the global search. Note that decreasing sparseness of the global search results in a smaller number of ambiguous values (compare Tables 1, 2, 3, Appendix A). Thus, failure to recover the correct axis is indeed due to the local part of the search, and not a failure of the FOE/C method. (That is why we label these cases as ambiguous and not as errors.)

Most experiments suggest the improvement of accuracy in determination of camera motion parameters as the number of features is increased. It seems that 16 and often 8 feature points give adequately reliable results (see Tables 1 to 3), with errors of only a few degrees from the correct axis. For four feature points the results are not accurate, though they are relatively close to correct results when the camera translation is along the line of sight. In real-world images, this would be not a sufficient number of features. It is plausible that 32 or 64 feature points would improve the precision of the search for the FOE/C, but in most cases where the camera is oriented approximately towards the direction of motion this will not be necessary.

A finer resolution of the window displacements (sampling every third of a pixel along the radial line, see Table 4) does not produce much improvement with respect to the default set of parameters, demonstrating that the correlation measure between the two windows has a relatively broad peak at the maximum.

Surprisingly, a window of smaller size (3×3) often gave quite satisfactory results and in some cases recovered the axis of translation when larger windows failed to do so (compare Tables 1 and 5). One conjecture is that the error function, Eq. (2.1), is more peaked because there is less averaging effect in the summation of the window intensity values. Since the computation time with smaller windows is shorter, this window size might be used in the initial stages of computation if the noise level in the images is low. A window of size 11×11 did not give results any better than the default window of size 7×7.

Supplying an initial guess close to the correct translational axis (Table 6) gave very good results in many cases. The search is done locally in a neighborhood δ_0 around the initially supplied axis. The significance of this result is in showing that the correlation measure is *able* to detect the changes of axial positions to within a few degrees. However, one should not be misled by the quality of the results. Since we have limited our search space to a very small δ_0 neighborhood around one of the points on the error surface, any returned value for the minimum will be not more than $\pm 2 \cdot \delta_0$ away from the initial guess. We find this upper bound by the following simple theoretical analysis. The value $\pm 2 \cdot \delta_0$ of the bound is a consequence of the fact that each time a new local minimum is found, the new area in which the search is continued has half the diameter of the old one. Thus, in the best case we can search in an area with boundary not further than $\delta_0 + \delta_0/2 + \delta_0/4 + \ldots = 2 \cdot \delta_0$. Thus, the primary problem is to determine the correct neighborhood for local search and therefore the global search must have sufficient resolution.



White noise added in various amounts to the image shown in Figure 3.

Figure 5: Synthetic images with added white noise.

Table 7 represents results of the search for the correct translational axis when one of the images was corrupted with uniform white noise. Some corrupted images are shown in Figure 5. The program parameters were the same as those of Table 3, which was judged to represent the best results. The number of features was 16 and the last column of the Table 3 should be compared with the results in Table 7. Noise was uniformly distributed, with an intensity range of 0 to 100, a mean of 50, and a standard deviation of 30. The gray level values of the first frame

ranged from a minimum of -88 to a maximum of 86, with an average of -54.0 and a standard deviation of 27.0. The added noise intensity ranges were approximately 12%, 30% and 60% of the intensity range of the uncorrupted image. The results show the stability of the search in the presence of noise, with a degradation of the performance as the noise level increases. We are further exploring the relationship between the noise level, the error values returned by the correlation function, and the overall performance of the algorithm.

2.5 Conclusions About Early Experiments

We tested the performance of the algorithm for various translations on synthetic images. The overall performance of the algorithm for cases where the camera motion is within $\approx 45°$ off the direction of the line of sight was very good: robust with respect to noise and accurate. The algorithm can return the translational axis to within a few degrees of the correct axis even if the image is subjected to significant disturbance by noise.

We found that the algorithm is not adequately sensitive to translational motion almost normal to the line of sight. This is due to the insensitivity of the correlation function and the failure of the local search in the case when the error surface around the minimum has a broad valley-like shape and exhibits small fluctuations.

We suggest a new approach which would impose a constraint of smoothness of the error surface and avoid the local hill-climbing technique. In the next chapter

we will give reasons for such a constraint and in Chapter 4 we will demonstrate that this approach increases both speed and reliability of the search for the correct translational axis.

3. SMOOTHNESS CONSTRAINT OF THE ERROR SURFACE

In this chapter we address theoretical aspects of the proposed improvement of Lawton's [LAW84] algorithm and discuss the search for the FOE/C from the point of view of the regularization analysis [POG84]. We will introduce the assumption of smoothness of the error surface in order to gain in robustness and speed of the algorithm. The smoothness assumption is viewed as yet another type of the regularization constraint, among many types used so far in the vision research. In the rest of the chapter we outline the mathematical method of the conjugate gradient [HES80] used to solve a variational problem introduced as a result of the regularization analysis. The details of the conjugate gradient method used in this paper are given in the Appendix C. The regularization method, as well as the conjugate gradient method, are described in greater detail elsewhere, [GRI81, TER82, POG84], but we present it for the sake of completeness.

It has been shown that the regularization method is implicitly found in many low-level vision algorithms [POG84]. In our opinion it reflects a global constraint introduced into low-level computations in an attempt to incorporate high-level knowledge. Research in low-level vision is faced with the problem of reconstructing environmental information from its projection on different media. As in any projection, a lot of information is lost and the inverse problem of reconstructing the environment is difficult. Typically, a global constraint on the solution is assumed and then ap-

plied to the local reconstructive computations; examples of these constraints involve the smoothness of the displacement field, the assumption of the least surface variation, the smoothness of the velocity field along a contour, etc. These constraints are definitely a step forward in the process of the recovery of the environmental data, but they might be good for one set of images and properties, while they fail for some other set of images. Also, one type of constraint may not be enough to guarantee the uniqueness of the solution for the inverse problem.

For the interested reader some references which are relevant to the application of regularization analysis in low-level vision are given below. Tikhonov [TIK63] or Tikhonov and Arsenin [TIK77] are sources for a rigorous formulation of regularization analysis. Horn and Schunck [HOR80] and Hildreth [HIL84] used the constraint of the smoothest velocity field consistent with the data to compute flow fields. The smoothness constraint used later in this chapter is similar to those developed by Grimson [GRI81, GRI82] and Terzopoulos [TER82, TER84], although Grimson and Terzopoulos were more interested in the problem of visual surface interpolation. Ikeuchi and Horn [IKE81] derived shape from shading also using a variational principle. Ullman's work on structure from motion [ULL79] belongs to this class of problems as well.

We now proceed with the statement of the regularization method and in Section 3.2 we give a short account of the numerical procedure we used to derive the results presented in Chapter 4.

3.1 A Short Overview of Regularization Analysis

In the introduction to this chapter we have mentioned that many low-level vision algorithms are of an inverse or ill-posed kind, and as such usually violate one (or more) of the conditions required for well-posed problems:

1. existence of the solution
2. uniqueness of the solution
3. a continual dependence of the solution on the initial data

Roughly speaking, ill-posed problems do not provide a unique solution and the space of acceptable solutions has to be restricted by some criterion. The process of restricting the space of solutions is called the regularization method.

Mathematically one can formulate an inverse problem as one in which the role of the solution and data is exchanged. For example, a linear set of equations

$$y = A \cdot z \tag{3.1}$$

asks for solution y, given an operator A and a set of data z. The inverse problem would be

$$z = A^{-1} \cdot y \tag{3.2}$$

i.e., given A and y, find z.

If the solution is unique both problems can be solved by requiring that a norm (symbolized here as $\| \|$) of the difference of the vectors $A \cdot z$ and y

$$\|A \cdot z - y\| \tag{3.3}$$

is minimal (ideally 0). If, for some z, given A and y, this requirement is achieved, z would be considered the solution. When the solution is not unique, it is necessary to impose a constraint on solutions z, to eliminate implausible solutions. For example, a constraint which requires that z minimizes the functional P

$$\|Pz\|^2 = \int \sum_{r=0}^{p} p_r(\xi) \left(\frac{d^r z}{d\xi^r} \right)^2 d\xi \tag{3.4}$$

is found to be very effective in many cases. These functionals are also called the Tikhonov stabilizers, and have the property that they provide the "smoothest" solutions z. In addition, they have the important property that if the generalized derivatives of up to the p-th order (integrand in Eq. (3.4)) are square-integrable functions, the solution can be shown to be unique (up to the null space of the functional P) if A is linear and continuous (which is often the case). Most stabilizing functionals used so far in early vision are Tikhonov stabilizers (see [TER84]).

One of the key tools in the regularization method is therefore the choice of the functional P. The choice is not unique and usually depends on the nature of the problem, as well as the effectiveness of the functional in restricting the space of possible solutions so that the problem becomes well-posed.

There are three types of approaches to ill-posed problems. The acceptable solutions are searched among those that:

1. Minimize
$$\|A \cdot z - y\|$$
subject to a constraint $\|Pz\| < Const.$ In other words, the solution z should best fit the data, while keeping the generalized variation of its derivatives

bounded.

2. Minimize

$$\|Pz\|$$

subject to the requirement $\|A \cdot z - y\| < Const.$ In other words, the solution z should have the smallest possible functional (Eq. (3.4)) while still being sufficiently close to data. The larger the noise in the image, the larger the constant *Const* should be. Examples are the smoothest surface or the smoothest velocity variations which follow data within some limits.

3. Minimize

$$\|A \cdot z - y\|^2 + \lambda \|Pz\|^2.$$

Lambda is called the regularization parameter. Note that the first and the second case are limiting cases of this case. For large λ the second term is dominant and the problem reduces to the second case, i.e., the search for the most regular solution. When λ is very small, the emphasis is on the solution that best fits the data. An optimal λ is needed which will provide the solution that follows the data as close as possible while still satisfying the regularization constraint. Regularization theory provides techniques for determining the best λ [TIK77], but ad hoc values of λ are used in this paper.

As an example, let us examine the case of fitting a smooth surface through a sparse set of unreliable data. This problem can be viewed as a problem of representing a very rough surface with a similar but a smoother one. Grimson [GRI82] approached this type of problem by searching for solutions using the third functional discussed earlier. In the analysis of a problem from stereo vision, Grimson concluded that it is necessary to fit a surface through values specified on a zero-crossing contour which would have the smoothest variation (the smallest curvature) and would still be consistent with the supplied data. After expressing the problem

as a variational problem, the next step was the minimization of the expression

$$\sum_{(x,y)} \|S(x,y) - C(x,y)\|^2 + \lambda \int \left(S_{xx}^2 + S_{yy}^2 + 2S_{xy}^2\right) dx\, dy. \tag{3.5}$$

The integral in this equation is over the region where the surface is defined. The integrand is a function of the second order partial derivatives of the surface being fitted, for example, $S_{xy} \equiv \partial^2 S(x,y)/\partial x \partial y$. This integral corresponds to the $\|Pz\|^2$ term and expresses the smoothness constraint. The sum is over a set of indices (x,y) for which the sparse data $C(x,y)$ are known. The sum is of the $\|A \cdot z - y\|$ type. λ is the regularization parameter. The greater the λ the more relative importance is given to the second term in the minimization process and the smoother the function will be. At the same time the solution is more likely to depart from the initial sparse set of data $C(x,y)$.

Assumption of a Smooth Error Surface. Since so many problems in low-level vision are ill-posed, is the determination of the motion from the displacement field using the FOE/C method an ill-posed problem? If so, in what sense is it ill-posed and what is a good choice for the stabilization functional? We argue that this problem is ill-posed mainly because the solution is so susceptible to small fluctuations in the error surface. If there is a noise in the error surface in order to gain robustness and a meaningful solution we impose a smoothness constraint on the error surface, similar to the way Grimson imposes a smoothness constraint on a surface to fit stereo depth data.

From the experiments reported in Chapter 2 we concluded that the error surface exhibits smooth behavior on a large scale, say over several pixels, but is not smooth over distances on the order of pixel or less. Globally, the error function is smooth because the correlation function between the two features is a relatively smooth function of the feature displacement provided that the displacement is small, that there is no occlusion, and that no dramatic local changes take effect.

Global smoothness of the error function is due ultimately to its construction from many smooth functions. The correlation function is computed along the radial line connecting the hypothesized position of the FOE/C and a feature. This has the advantage of being a one-dimensional search for the best correlation, as opposed to a two-dimensional search in all directions around the feature. One can think of this one-dimensional function as a cross-section along the radial line of the two-dimensional correlation function. The two-dimensional correlation function in question should have its peak exactly on the radial line connecting the correct position of the FOE/C and the feature. The maxima of all correlation functions give rise to the global minimum of the error surface. If the interesting points and their two-dimensional correlation functions are fixed and the FOE/C moved slightly off its best position, a new set of cross sections determined by new radial lines and the two-dimensional correlation surfaces will be found. The maxima of the new set of one-dimensional correlation functions are just slightly off in value and position from the old maxima, since the two-dimensional correlation functions are relatively smooth functions. Therefore, the total error will change slowly in the vicinity of

the minimum. Consequently, the total error function, which is determined by many smooth functions, should be also a smooth function. In addition, if there is a good spatial distribution of the points around the FOE/C, as the FOE/C moves significantly from its correct position, it is very likely that many points will not have a good match anywhere on their new radial lines; consequently, the error function should increase in all directions as the FOE/C is moved.

We stated in Chapter 2 that in some cases the local search for the minimum resulted in unreliable data. The local search sometimes failed to return the correct location of the FOE/C because it found a false minimum due to the presence of noise in the error surface. The noise can be a consequence of digitization effects, external factors, imperfect sensors, not enough features, etc. For example, digitization noise causes the correlation functions along the radial lines to have a non-smooth behavior if the characteristic length is about a pixel or less.

Our hypothesis is that the problems that led to the failure of the recovery of the correct translational axis are due to non-smoothness of the error surface due to small local fluctuations, and that a regularization constraint is needed in these cases to guide the search for the minimum of the error surface. The information about the position of the minimum is supported by the overall shape of the error surface in this area, and therefore it reflects the contribution of many points on the error surface.

The other advantage of this approach is that the fitting procedure can speed

up the search process. Because of the assumption of global smoothness of the error surface, the search for the minimum can start with a rather sparse set of error values. A smooth surface is fitted through this set of points. Then, the search continues around the minimum of the smoothed surface with a finer resolution and in a smaller region. The selection of the initial number of points in the sparse set of data requires a compromise between the validity of the surface shape and the efficiency of the method.

The smoothing procedure critically depends on the choice of the regularization parameter λ (see Eq. (3.5)). Too much smoothing results in less accurate data, while not enough smoothing causes results to be less reliable. It is hoped that the choice of the parameter will be supplied by a high-level vision module, or that it can be suggested from estimation of the noise in the surface. The global shape and the noisiness of the surface determine the size of the regularization parameter. The greater the overall curvature of the surface, the harder it is to fit the surface (to follow the given data) and the smaller the parameter λ has to be (less weight is given to the smoothing functional). If the surface near the minimum has low principal curvatures and there is more noise in an image, the larger the regularization parameter should be (the stiffer the interpolating surface). These last two requirements require therefore that the regularization parameter should decrease as we approach the minimum.

The regularization theory for linear A and P is equivalent, according to

[POG85], to the method of generalized splines, whose order depends on the order of the stabilizer P, see Eq. (3.4). Consequently, our technique can be also viewed as one form of this particular interpolation technique.

3.2 Conjugate Gradient Method

In this section we state the variational problem which results from the smoothness constraint on the error surface. The solution of the variational problem is achieved by the conjugate gradient method. The flavor of the conjugate gradient method, without proofs of equations, but with enough rigor to help the reader understand the basic concepts and steps, is given in Appendix C. For more detail about the conjugate gradient method reader is referred to Hestenes [HES80].

The problem with which we are concerned here is the minimization of the positive-definite quadratic functional

$$f(x) = \frac{1}{2} x^* A\, x - h^* x + C. \tag{3.6}$$

x is an n-dimensional vector of unknown values, A is an $n \times n$ positive-definite matrix, and h is a vector of given data. The symbol $*$ denotes the transposition of matrices and vectors; C is an arbitrary scalar. The solution of the problem x_o satisfies the equation

$$A\, x_o - h = 0. \tag{3.7}$$

(Compare with Eq. (3.3) in the first section of this chapter.)

To find the minimum of $f(x)$ in Eq. (3.6), the set of linear equations rep-

resented in Eq. (3.7) must be solved. There are many methods for solving this equation, such as the well-known Jacobi and Gauss-Seidel procedures (see, for example, [STU80]), but the conjugate gradient methods are the best for cases where A is a sparse matrix. We will see that the matrix A in our problem is a sparse matrix because it has only a few off-diagonal elements (it is quasi-diagonal), and therefore it is appropriate to take advantage of the conjugate gradient method. The only disadvantage of the method is that it does not converge monotonically towards the solution, though the convergence is guaranteed in not more than n iterative steps.

In Appendix C we derive the conjugate gradient (CG) algorithm for the solution of Eqs. (3.6) and (3.7). Here we concentrate only on the derivation of matrices, A and h, which appear in our formulation of the regularization problem.

Thin metal plate problem. We approximate an error surface represented by a sparse set of points with a surface that a thin metal plate would form, if forced to pass close to the given set of sparse data. A plate is [COU53] an elastic two-dimensional body, in the shape of a plane when in equilibrium, whose potential energy under deformation is given by the integral of a quadratic form in principle curvatures ρ_1 and ρ_2:

$$C_1\left(\frac{1}{\rho_1^2}+\frac{1}{\rho_2^2}\right)+\frac{C_2}{\rho_1\rho_2}, \tag{3.8}$$

where C_1 and C_2 are some constants. For small deflections it follows that

$$\frac{2}{\rho_1}+\frac{2}{\rho_2}=\Delta S\equiv S_{xx}+S_{yy}, \qquad \frac{1}{\rho_1\rho_2}=S_{xx}S_{yy}-S_{xy}^2, \tag{3.9}$$

where S_{xy} is the second partial derivative of the surface with respect to x and y coordinates. By using the assumption of the "thinness" of the metal plate we find that the potential energy of the plate deformation is proportional to

$$\iint_{\Omega} \left(S_{xx}^2 + S_{yy}^2 + 2S_{xy}^2\right) dx\, dy,$$

where Ω is the domain in the X-Y plane over which the surface (thin plate) has to be found. The plate should pass close to values

$$C(i,j) \qquad (i,j \in \Sigma),$$

where Σ defines a sparse set of data ($\Sigma \subset \Omega$). After incorporating the last set of constraints $C(i,j)$ by the method of Lagrange multipliers, the problem becomes one of the minimizing of the expression

$$\mathcal{E}(S) = \iint_{\Omega} \left(S_{xx}^2 + S_{yy}^2 + 2S_{xy}^2\right) dx\, dy + \beta \sum_{i,j \in \Sigma} |S(i,j) - C(i,j)|^2. \tag{3.10}$$

Here β is the *inverse* of the regularization parameter λ from the first section of this chapter. We assume for simplicity that Ω is a quadratic grid and Σ is a subgrid of lower resolution.

By introducing numerical approximations for the second-order partial derivatives [ABR68], in the case of equally spaced coordinates ($x_{i+1} - x_i = h = y_{i+1} - y_i$) Eq. (3.10) becomes:

$$\mathcal{E}_h(S) = \frac{1}{h^2}\{\sum_{i=1}^{n-2}\sum_{j=0}^{n-1} \left(S_{i-1,j} - 2S_{i,j} + S_{i+1,j}\right)^2 + \sum_{i=0}^{n-1}\sum_{j=1}^{n-2} \left(S_{i,j-1} - 2S_{i,j} + S_{i,j+1}\right)^2$$

$$+ 2\sum_{i=0}^{n-2}\sum_{j=0}^{n-2}(S_{i,j} - S_{i+1,j} - S_{i,j+1} + S_{i+,j+1})^2\} + \beta \sum_{i,j\in\Sigma} |S_{i,j} - C_{i,j}|^2 \quad (3.11)$$

The problem can be rewritten as

$$\mathcal{E}_h(S) = \tfrac{1}{2}S^*QS - \beta C^*S,$$

where

$$Q = A + \beta\Delta_\Sigma \quad (3.12)$$

and

$$\Delta_\Sigma(i,j) = \begin{cases} 1, & \text{if } C_{i,j} \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The matrix A with elements $A_{ij,kl}$ for an $n \times n$ dimensional surface is an $n^2 \times n^2$ matrix. It will be convenient for n to be an odd integer. Moreover, we choose $n = 2^m + 1$ (m is an integer) in order to generate the subgrids of sparse data whose dimensionality is $2^{m-i} + 1$, where $i = 1, \ldots, (m-1)$.

Considering pairs (i,j) as row and column indices, matrix A can be visualized as an $n \times n$ matrix with six different types of elements. These elements can be viewed as masks used to compute the new interpolated surface values from the old surface values. They will be also called *templates* (after [GRI81]). The values in the templates are found by expanding Eqs. (3.11) and (3.12) and comparing elements in these equations with the same pairs of indices (i,j). Special care should be taken with the boundary values of i and j. The boundary values cannot be neglected because the dimensions of the matrices are usually low, and because the boundary values propagate inwards during the interpolation procedure. If we label templates

a, b, c, d, e, f, we represent A symbolically as

$$A = \begin{pmatrix} a & b & c & c & \dots & c & c & b & a \\ b & d & e & e & \dots & e & e & d & b \\ c & e & f & f & \dots & f & f & e & c \\ c & e & f & f & \dots & f & f & e & c \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ c & e & f & f & \dots & f & f & e & c \\ c & e & f & f & \dots & f & f & e & c \\ b & d & e & e & \dots & e & e & d & b \\ a & b & c & c & \dots & c & c & b & a \end{pmatrix},$$

where the templates are:

$$a = \begin{pmatrix} \mathbf{4} & -4 & 1 \\ -4 & 2 & \\ 1 & & \end{pmatrix} \qquad b = \begin{pmatrix} -4 & \mathbf{10} & -6 & 1 \\ 2 & -6 & 2 & \\ & 1 & & \end{pmatrix}$$

$$c = \begin{pmatrix} 1 & -6 & \mathbf{11} & -6 & 1 \\ & 2 & -6 & 2 & \\ & & 1 & & \end{pmatrix} \qquad d = \begin{pmatrix} 2 & -6 & 2 & \\ -6 & \mathbf{18} & -8 & 1 \\ 2 & -8 & 2 & \\ & 1 & & \end{pmatrix}$$

$$e = \begin{pmatrix} & 2 & -6 & 2 & \\ 1 & -8 & \mathbf{19} & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{pmatrix} \qquad f = \begin{pmatrix} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & \mathbf{20} & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{pmatrix}$$

The templates in A are masks of the given shapes, with matrix elements not shown equal to zero. To compute a new value of the surface in row r and column c the convolution of the template with the surface is performed so that the position of the bold-faced integer is at the pixel (r, c). For example, the new value of the surface

$S_{0,1}^{new}$ is (we use the template b because it is in row 0, column 1):

$$S_{0,1}^{new} = -4S_{0,0} + 10S_{0,1} - 6S_{0,2} + S_{0,3} + 2S_{1,0} - 6S_{1,1} + 2S_{1,2} + S_{2,1}.$$

The only other consideration is that *the masks have to be mirror reflected and rotated so that the complete matrix $A_{ij,kl}$ is symmetric for reflections across diagonals, rows, and columns.* Thus the letters (e.g. $a, b, c, \ldots$) in the matrix A only denote the same kind of template, but they do not show its appropriate reflection and/or rotation. For example, the template b in the next to last row and the first column of matrix A, is template b above rotated 90° counterclockwise, i.e., the new surface value is

$$S_{n-2,0}^{new} = -4S_{n-1,0} + 10S_{n-2,0} - 6S_{n-3,0} + S_{n-4,0} + 2S_{n-1,1} - 6S_{n-2,1} + 2S_{n-3,1} + S_{n-2,2}.$$

The template b in the last row and the second column is the mirror-image along rows of the shown template b, etc. For more details consult Appendix C and [GRI81].

This completes our exposition of the theoretical tools used in this paper. The result of applying these tools is to improve the detection of motion; this is the subject of the next chapter.

4. IMPROVED SEARCH FOR THE ERROR MINIMUM

Problems with noisy error surfaces can be partially overcome by smoothing. We repeated some of the experiments reported in Chapter 2 to examine the influence of the smoothing procedure on the accuracy of the recovery of the correct translation.

The improved version of the search for the error minimum consists of the following steps: In the first step the error surface is very coarsely sampled over the grid in the two-dimensional space of polar angles determining the direction of camera translation. (By error sampling we mean computation of error values using correlation among the features.) The coarsely sampled error surface exhibits a minimum which is the starting point for the next, finer, search. The eight adjacent neighbors on the square grid of the minimum just found determine a neighborhood where a finer search should continue.

The second step is a finer sampling of the error surface in the neighborhood found in the previous step. This error surface is rough and it would be a mistake, as we have seen in Chapter 2, to immediately continue with finer error sampling at some new minimum in a smaller neighborhood.

Instead, in the third step, the local error surface found in the previous step is smoothed. The smoothing eliminates fluctuations in the error surface and de-

termines the correct position of the minimum based on the global support of all points in the local error surface. Besides eliminating fluctuations, the smoothing procedure can increase the resolution of sampling without deterioration of quality. Since the smoothing operation takes a fraction of the computational effort involved in sampling the error surface, the speed of search for the same resolution is also significantly improved.

Finally, the second and the third step are repeated, as a group, as many times as necessary, to get the desired angular accuracy of the direction of motion. In practice, this repetition is often unnecessary; the precision of computations with angular resolutions smaller than 0.5° is limited by other factors such as image digitization, imprecision of the correlation matching, etc. The results after the third step are already very accurate.

As will be seen in the next section, the problems connected with the local search reported in Chapter 2 were overcome. For the cases labeled as ambiguous in earlier experiments, the newly found axes were clearly closer to the correct axes. Recall that most of the cases labeled as ambiguous had incorrectly returned the translational direction to be the (90°, 90°) axis. The reason was that, due to the shallowness of the error surface, the global search returned this axis as the best estimate, and the local search was not able to overcome fluctuations in the error surface and converge closer to the actual axis. The precision of the recovery of the exact axis continues to deteriorate as the translations become more orthogonal to

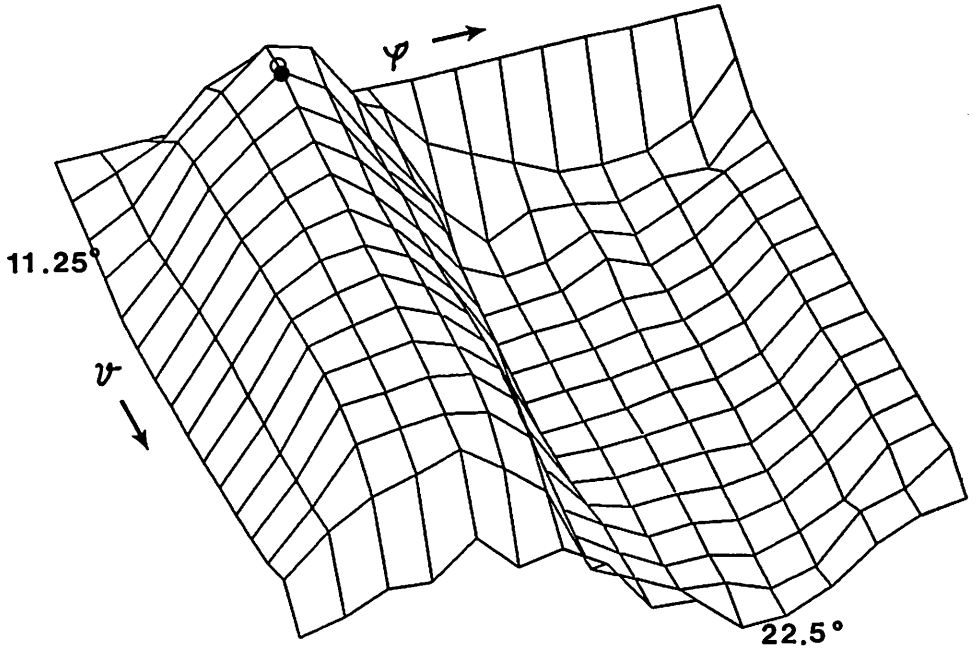
the line of sight; this remains a problem with the correlation matching.

The other advantage of the new approach is in increased computational speed. Since the only interesting point on the error surface is its minimum, we would like to avoid computing unnecessary error values where they are not needed. We increased the sparsity of the error surface to test the limits of the algorithm. Since the error surface is rather flat, the results with very coarse sampling of the error surface were very good, although there was a limit to the sparseness of the sampling. These experiments indicate that a coarse error surface around the minimum provides enough support for the actual position of the minimum. All these conclusions follow from the experiments described in the next sections.

4.1 Description of the Experimental Procedure

In these experiments we used the same images and the same camera, environmental, and algorithm parameters as described in Section 2.2. A subset of the runs presented in Appendix A was repeated to examine the aspects of the algorithm relevant to the error surface smoothing. The numerical data for the new set of experimental runs with smoothed error surfaces are presented in Appendix B.

4.1.1 Coarse sampling of the error surface. The experiments started, as we have mentioned, with a construction of a sparse error surface defined on the entire parameter domain, ($\vartheta \in [0,180)$, $\varphi \in [0,360)$). The step size in ϑ, $\delta\vartheta$, was typically 11.25°, and the step size in φ, $\delta\varphi$, was 22.5°. This choice resulted in a 16×16 dimensional error surface. Error surfaces sampled so coarsely returned in all

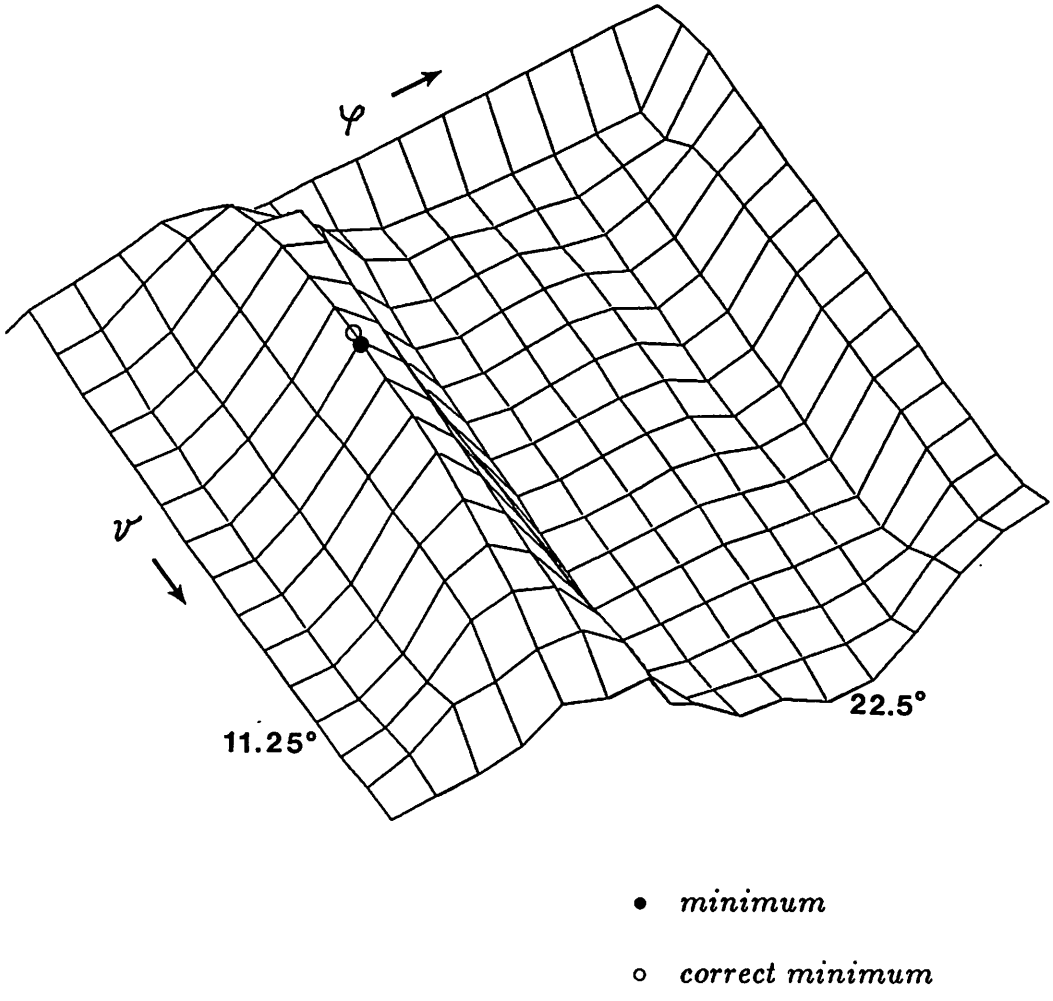


Inverted error surface for translational motion 30° off the line of sight. ϑ changes from 0° – 180° in increments of 11.25°. φ changes from 0° – 360° in increments of 22.5°. The minimum (•) at (33.75°, 90°) corresponds to the error value **0.2734** (Compare Table 1, Appendix B). The correct (○) direction of translation is (30°, 90°). 16 feature points are used. The principal curvatures near the maximum are small, and the valley is quite pronounced.

Figure 7: Coarsely sampled error surface for 30° translational motion off the line of sight.

the cases the position of the minimum which was the closest to the known correct value. The inaccuracy was determined by the resolution in (ϑ, φ) parameter space.

The coarsely sampled error surfaces for translational motions 30° and 75° off the line of sight are shown in Figures 7 and 8. The error surfaces are shown for the entire range of parameter values of ϑ and φ. In all the figures, the error surfaces are shown inverted for a better presentation. Note that the curvature in the ϑ–direction (see Figure 8) is smaller when the translational axis is more off the line of sight. For example, the ratio of the curvatures in ϑ–direction for the error surfaces in Figures 8 and 7 is equal to 0.47. A larger neighborhood around the minimum has to be chosen for surfaces with smaller curvature, such as the one presented in Figure 8, to guarantee that a finer sampling will not miss the error surface minimum. The criterion was simply that all the error values on the boundary determined by $\vartheta_L \leq \vartheta \leq \vartheta_H$, $\varphi_L \leq \varphi \leq \varphi_H$ should be significantly larger than the minimum found. For motions that are more along the line of sight, the boundaries were determined as the first neighbors of the minimum (on the coarse grid). In other words, if $\delta\vartheta$ and $\delta\varphi$ are current sampling intervals in (ϑ, φ) parameter space, then the boundaries are: $\vartheta_L = \vartheta_{min} - \delta\vartheta$, $\vartheta_U = \vartheta_{min} + \delta\vartheta$ and $\varphi_L = \varphi_{min} - \delta\varphi$, $\varphi_U = \varphi_{min} + \delta\varphi$. For motions along axes more perpendicular to the line of sight, the error surface is flatter and the boundaries are determined as the second-closest neighbors. In other words, the boundaries are: $\vartheta_L = \vartheta_{min} - 2\delta\vartheta$, $\vartheta_U = \vartheta_{min} + 2\delta\vartheta$ and $\varphi_L = \varphi_{min} - 2\delta\varphi$, $\varphi_U = \varphi_{min} + 2\delta\varphi$.



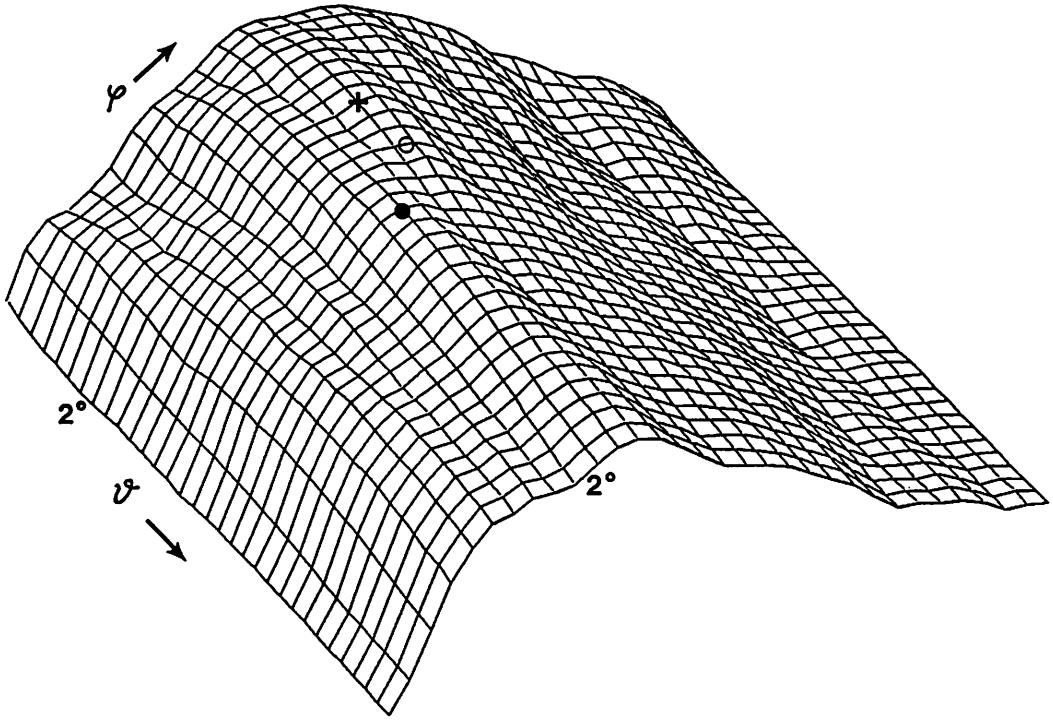
Inverted error surface for translational motion 75° off the line of sight. ϑ changes from 0° – 180° in increments of 11.25°. φ changes from 0° – 360° in increments of 22.5°. The minimum (•) at (78.75°, 90°) corresponds to the error value **0.3297** (Compare Table 1, Appendix B). The correct (o) direction of translation is (75°, 90°). 16 feature points are used. The curvature in the ϑ direction is much smaller than that in Figure 7, indicating a shallow valley and increased difficulty of the search for the minimum.

Figure 8: Coarsely sampled error surface for 75° translational motion off the line of sight.

4.1.2 Possibility of several local minima. Figure 9 shows a densely sampled error surface illustrating local fluctuations. It is possible that several local minima exist for the error surface, which led to the failure of the recovery of the correct minimum in the error surface when the local hill-climbing technique was used. The error surface is for eight feature points in the image and it would be smoother for more features; however the presence of noise would increase the roughness of the surface and a similar situation could occur. In other experiments the error surface was not sampled so densely before smoothing because this would be computationally too expensive.

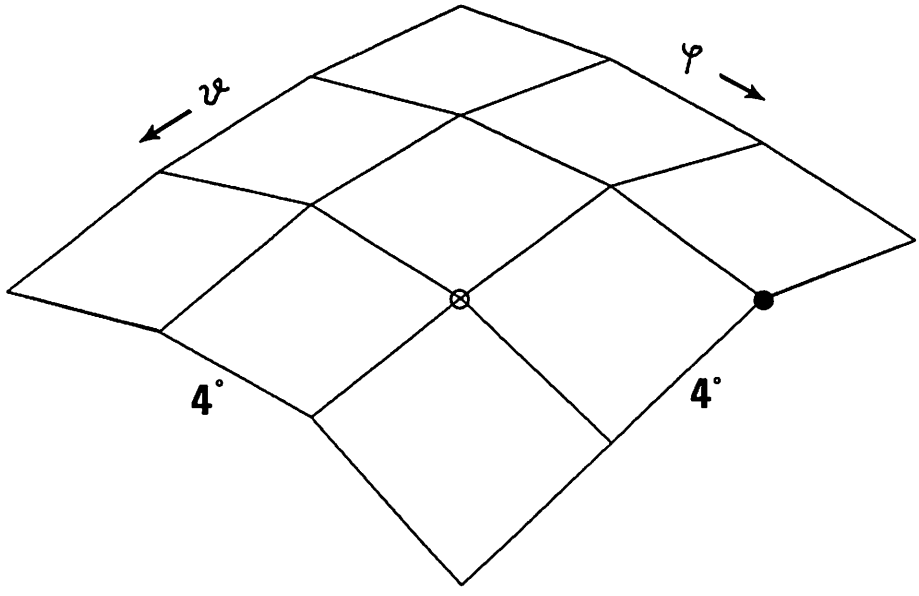
4.1.3 Fine sampling of the error surface. We now proceed with the description of the finer search for the error minimum, i.e., steps two and three. In the step two, the error surface is sampled in a small neighborhood around the minimum found during the coarse search. Then, this patch of the error surface is smoothed and the minimum of the smoothed error surface is found (step 3). The second and the third step are then repeated in the neighborhood of half the size of the previous neighborhood. These steps are shown in Figures 10 through 13, in which unsmoothed and smoothed local error surfaces are shown for motion 30° off the line of sight. These results are discussed in more detail below; the search procedure could be characterized as a *hierarchical search with smoothing.*

Results from coarse sampling of the error surface over the entire parameter range, shown in Figure 7, suggested that the minimum of the error surface is



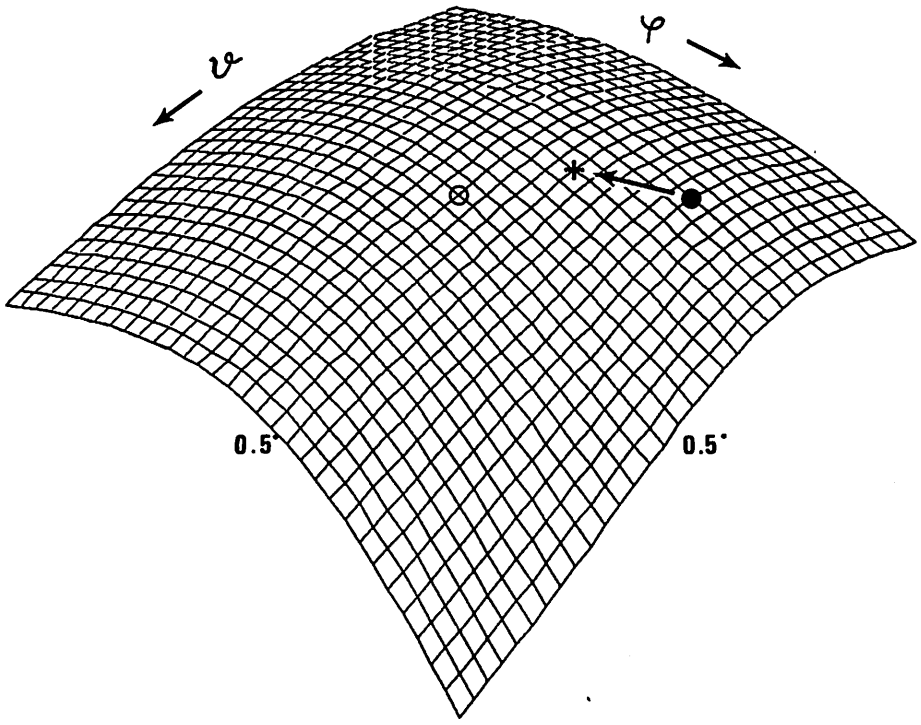
Unsmoothed error surface, near the minimum for eight feature points (inverted). ϑ changes from $40°-102°$ in increments of $2°$. φ changes from $40°-102°$ in increments of $2°$. The minimum (•) corresponds to the error of **0.1153**, at $\vartheta = 70°$, $\varphi = 60°$. The correct direction (o) is $\vartheta = 62.4°$, $\varphi = 64.8°$. A local minimum (+) is seen at $\vartheta = 56°$, $\varphi = 64°$ with the error value of **0.1228**, and it could have been promoted in the global minimum, if the coarser search missed the first minimum.

Figure 9: Densely sampled error surface near the minimum.



Unsmoothed error surface, near the minimum; the surface is shown inverted. ϑ changes from $22° - 34°$ in increments of $4°$. φ changes from $82° - 94°$ in increments of $4°$. The minimum (•) corresponds to the error of **0.2703**, at $\vartheta = 26°$, $\varphi = 94°$. The correct (o) direction of camera translation is for $\vartheta = 30°$, $\varphi = 90°$.

Figure 10: Unsmoothed error surface near minimum for 30° translational motion off the line of sight.



\+ *minimum*

o *correct minimum*

• *minimum from Figure* 10

Smoothed version of the error surface in Figure 10 (inverted). ϑ changes from $22° - 37.5°$ in increments of $0.5°$. φ changes from $82° - 97.5°$ in increments of $0.5°$. The minimum (+) corresponds to the error of **0.2739**, at $\vartheta = 27.5°$, $\varphi = 91.5°$. The correct (o) direction of translation is for $\vartheta = 30°$, $\varphi = 90°$. The regularization parameter is $\beta = 0.1$.

Figure 11: Smoothed error surface near minimum for 30° translational motion off the line of sight.

somewhere in the vicinity of the direction $\vartheta = 33.75°$, $\varphi = 90°$. Thus, a finer sampling of the error surface in the range $\vartheta \in [22°, 38°]$, $\varphi \in [82°, 98°]$ is performed; this surface is presented in Figure 10. (The actual surface is an array of 5×5 real values, here it is displayed only as a 4×4 surface.) The resolution in the parameter space was $\delta\vartheta = \delta\varphi = 4°$. The minimum **0.2703** found in this step was at $\vartheta_{min} = 26°$, $\varphi_{min} = 92°$. The quantity $\Delta\Omega \equiv \sqrt{\Delta\vartheta^2 + \Delta\varphi^2}$, where $\Delta\vartheta \equiv \vartheta_{min} - \vartheta_{correct}$; $\Delta\varphi \equiv \varphi_{min} - \varphi_{correct}$, was used as a measure of the deviation from the correct values of the position of the FOE/C. In this case we have $\Delta\Omega = 5.7°$, which represents a significant deviation from the correct direction of translation. The position of the known (correct) minimum is marked with a circle, while the minimum found by a simple search among the error surface values is marked with a solid circle (disk).

Figure 11 shows the error surface from Figure 10, after smoothing with the regularization parameter equal to 0.1. The resolution of the smoothed error surface was $\delta\vartheta = \delta\varphi = 0.5°$ and the range of the angular parameters remained the same as that in Figure 10. This implies an eightfold increase in the resolution, for a fraction of the computational cost. The minimum **0.2739** of the smoothed error surface, marked with + in Figure 11, at $\vartheta = 27.5°$, $\varphi = 91.5°$, represents an improvement in the position since the total error is $\Delta\Omega = 2.9°$. The improvement is marked by the arrow demonstrating the change in the minimum position after the smoothing.

The second and the third step were then repeated around the minimum of

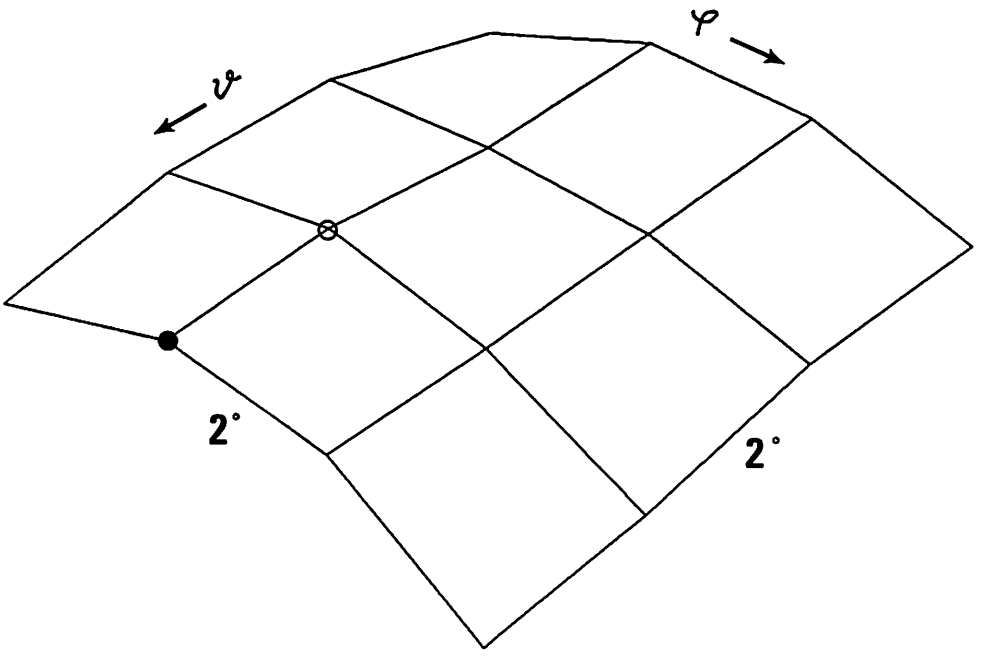
the smoothed error surface (i.e., around the point $\vartheta = 27.5°$, $\varphi = 91.5°$). The error surface was sampled in the range $\vartheta \in [26°, 34°]$, $\varphi \in [86°, 96°]$ with an angular resolution of 2° for both parameters. This neighborhood is half the size of the neighborhood from the second step. The result is shown in Figure 12.

The search in this smaller neighborhood (see Figure 12), found a better minimum, which was missed in the second step (see Figure 10). The smoothing, shown in Figure 11, guided by the overall shape of the error surface indicated that this might happen. The directional error $\Delta\Omega$ was further reduced from $\Delta\Omega = 2.9°$ to $\Delta\Omega = 2°$.

Finally, another smoothing was performed with the angular resolution of 0.25°. The result is shown in Figure 13. The minimum values were $\vartheta = 31.75°$, $\varphi = 89.25°$, the error being **0.2698**. The final error, after this smoothing, with which the direction of translation is found was $\Delta\Omega = 1.9°$. This error is obtained after a total of five steps:

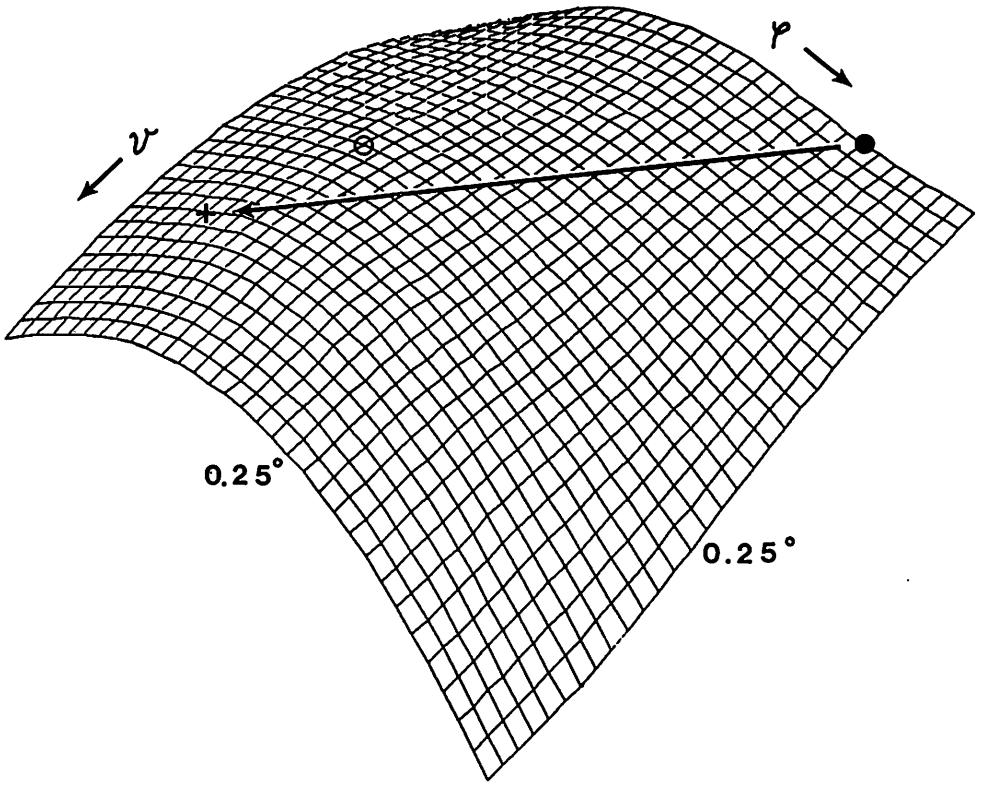
1. A coarse search over the entire parameter space $\vartheta \in [0°, 180°)$, $\varphi \in [0°, 360°)$,
2. A fine search in 16° × 16° neighborhood $\vartheta \in [22°, 38°]$, $\varphi \in [82°, 98°]$,
3. A smoothing in the same neighborhood,
4. A still finer search in a 8° × 8° neighborhood $\vartheta \in [26°, 34°]$, $\varphi \in [86°, 96°]$,
5. A smoothing in the last neighborhood.

We felt that further increase of the angular resolution (the last angular resolution being 0.25°) was beyond the precision of the FOE/C search with 16 interesting features. This was also in agreement with earlier experiments described in Chapter



Unsmoothed error surface, over a smaller domain around the minimum than in Figure 10. ϑ changes from $26° - 32°$ in increments of $2°$. φ changes from $86° - 94°$ in increments of $2°$. The minimum (•) corresponds to the error of **0.2687**, at $\vartheta = 32°$, $\varphi = 90°$. The correct (o) parameter values are $\vartheta = 30°$, $\varphi = 90°$. It is important to notice that the value of this minimum is smaller than the value of the minimum found in Figure 10, and that the present minimum was missed during the first search.

Figure 12: Unsmoothed error surface near minimum for 30° translational motion off the line of sight.



+ *minimum*

o *correct minimum*

• *minimum from Figure* 10

Smoothed version of the error surface in Figure 12. ϑ changes from $26° - 33.75°$ in increments of $0.25°$. φ changes from $88° - 95.75°$ in increments of $0.25°$. The minimum (+) corresponds to the error of **0.2698**, at $\vartheta = 31.75°$, $\varphi = 89.25°$. The correct (o) parameter values are $\vartheta = 30°$, $\varphi = 90°$. The regularization parameter is $\beta = 0.1$.

Figure 13: Smoothed error surface near minimum for 30° translational motion off the line of sight.

2.

The incorporation of smoothing into the translational algorithm results in an algorithm which is more reliable (using smoothing to guide us towards the correct minimum) and faster (concentrating on computation of error values only in a small neighborhood around the current minimum). The total number of points on the error surface computed was 323 (= 273 (coarsest) + 25 (medium) + 25 (finest)), where 273 of them (i.e., 84.5% of the total 323) were due to the lack of any knowledge about the direction of translation. These coarse calculations become unnecessary once the approximate direction of motion is established and in subsequent frames it is most likely that only finer, local search (which needed 50 surface points) would be needed.

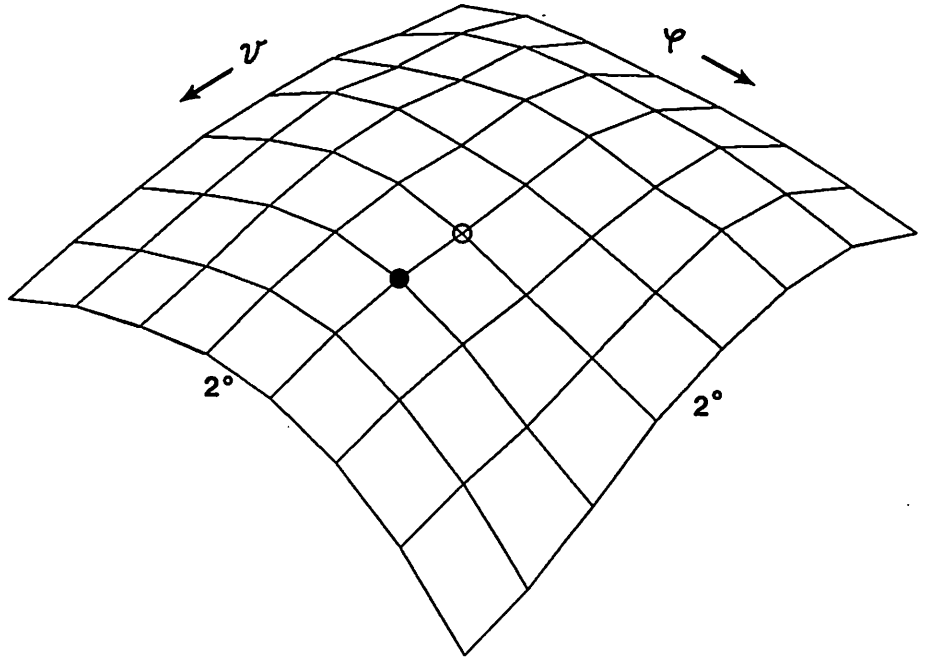
4.1.4 The advantage of finer error sampling. It is possible that a slight increase in the density of error surface sampling might actually save repetition of steps two and three from the previous subsection (i.e., steps 4 and 5 from the list above). Figure 14 shows the same error surface as in Figure 10, but sampled with twice the density, i.e., at the angular resolution of 2°. The minimum found at (32°, 90°) is a much better prediction than that of (26°, 94°), as found in Figure 10, as is the directional error $\Delta\Omega = 2°$. After smoothing the error surface in Figure 14 (shown in Figure 15) the minimum (31°, 89.5°) was found, with $\Delta\Omega = 1.1°$. This result is even better than that obtained in the two passes depicted in Figures 10 to 13. The latter, one-pass approach, needed 81 error surface computations

versus 50 needed for the two-pass approach. Both procedures, the one described in the previous subsection and the one in this subsection, seem to have about equal performance.

4.1.5 A need for larger neighborhood. A sampling in a larger neighborhood around the minimum might be required after the first step for motions that are more off the line of sight, as those shown in Figures 16 and 17 for the translation in direction (62.4°, 64.8°), and in Figures 18 and 19 for the translation in direction (90°, 90°). Note the difference in the angular separation in these figures as compared to 30° motion. The reason for sampling in a larger neighborhood is that the error surface is more shallow. The search did not fail, as the local hill-climbing method did, and the smoothing procedure always improved the position of the previously found minimum.

4.2 Overview of Experimental Results

Until now we have concentrated our attention on experiments for specific axes of translation. Now, we present a global overview of all the experiments. The numerical values for the error minima before and after smoothing and their positions for many possible translations are given in Appendix B. The details of how to interpret these tables are given in the introduction to Appendix B. From Tables 1 and 2 in Appendix B, where the results for 16 feature points are presented, we can see that the precision of the method remains in the range of 1-2 degrees for axes inside a 45° cone around the line of sight, and 5-10 degrees outside that

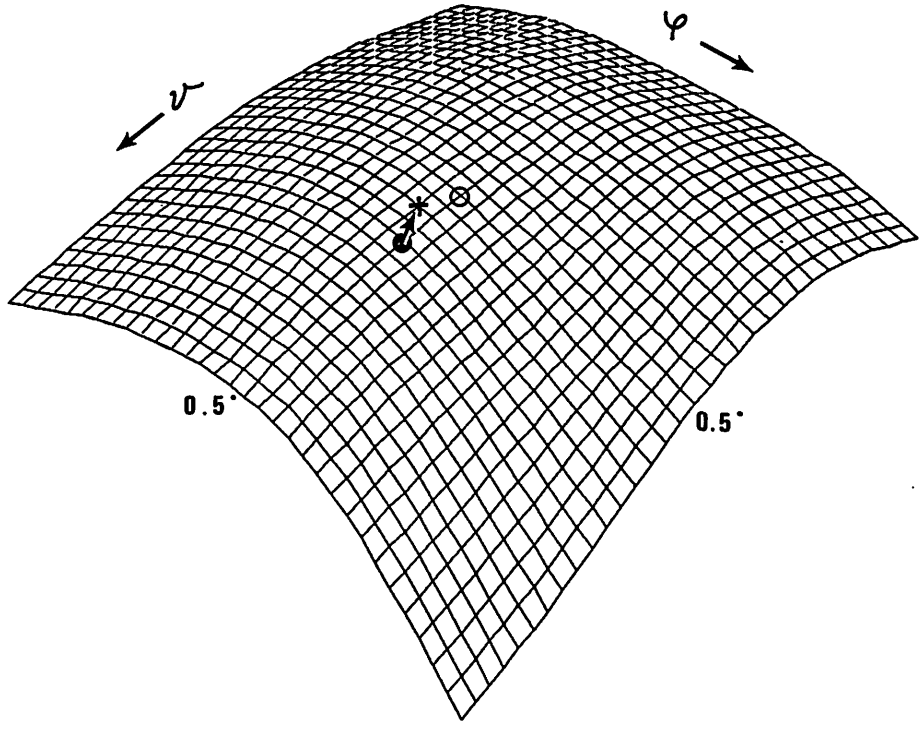


• *minimum*

○ *correct minimum*

Unsmoothed error surface near the minimum with finer sampling than in Figure 10. ϑ changes from $22° - 36°$ in increments of $2°$. φ changes from $82° - 96°$ in increments of $2°$. The minimum (•) corresponds to the error of **0.2687**, at $\vartheta = 32°$, $\varphi = 90°$. The correct (○) values are $\vartheta = 30°$, $\varphi = 90°$.

Figure 14: Unsmoothed error surface near minimum for 30° translational motion off the line of sight.



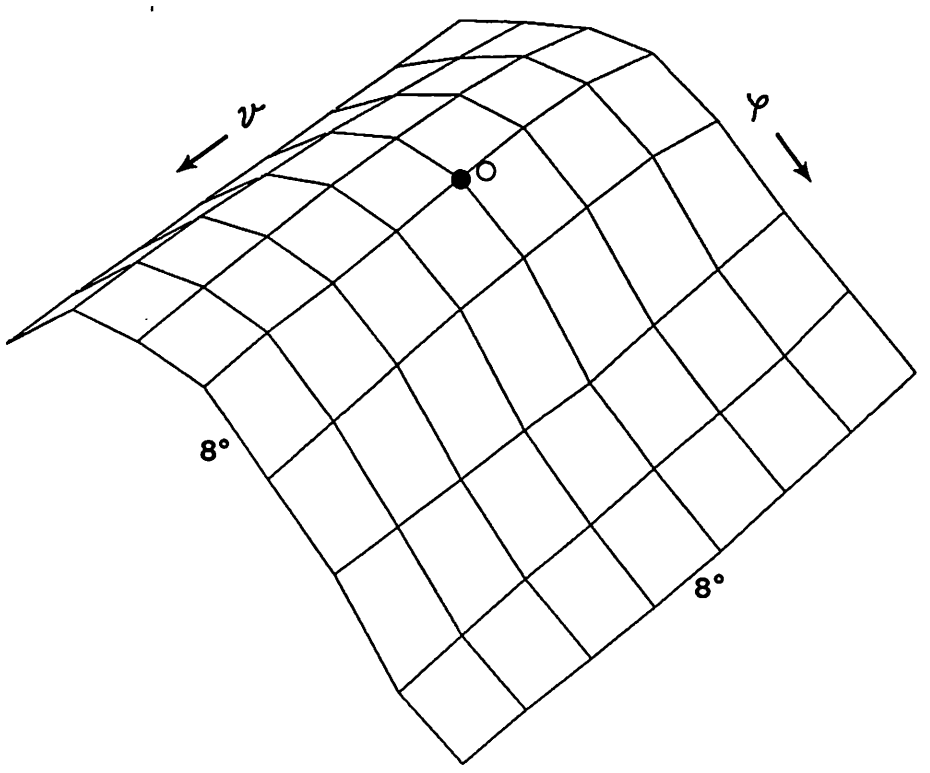
\+ *minimum*

o *correct minimum*

• *minimum from Figure* 14

Smoothed error surface near the minimum (see Figure 14). ϑ changes from 22° – 35.5° in increments of 0.5°. φ changes from 82° – 97.5° in increments of 0.5°. The minimum (+) corresponds to the error of **0.2709**, at $\vartheta = 31°$, $\varphi = 89.5°$. The correct (o) parameter values are $\vartheta = 30°$, $\varphi = 90°$. The regularization parameter is $\beta = 0.1$.

Figure 15: Smoothing of the error surface from Figure 14.

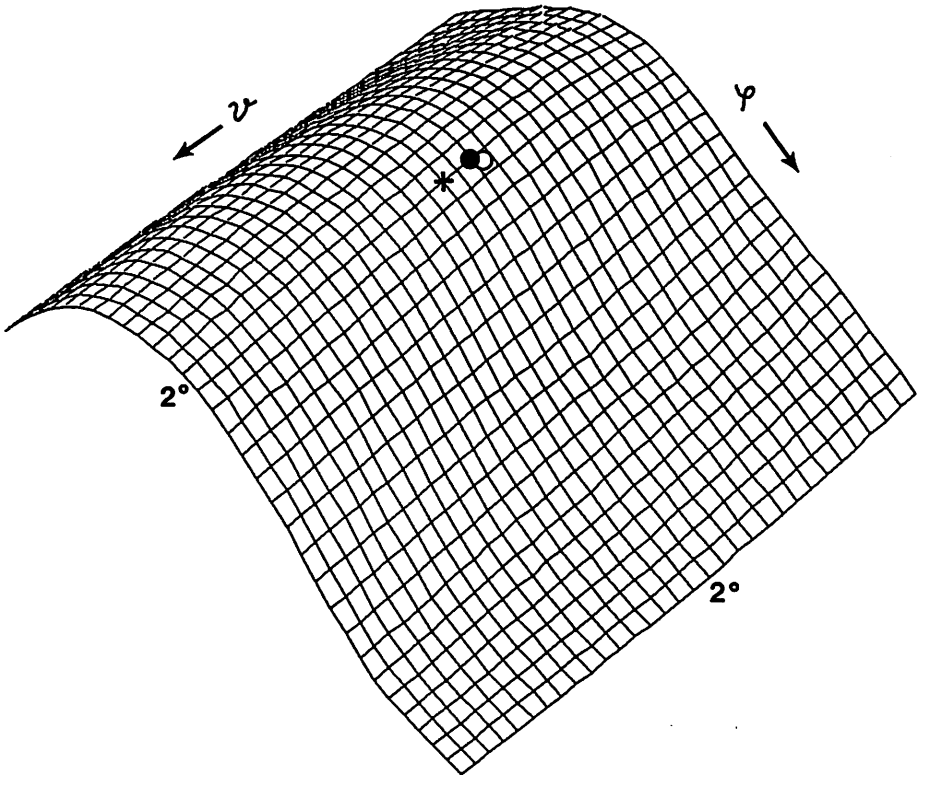


• *minimum*

○ *correct minimum*

Unsmoothed error surface, near the minimum (inverted). ϑ changes from $40° - 96°$ in increments of $8°$. φ changes from $40° - 96°$ in increments of $8°$. The minimum (•) **0.3782** is at $\vartheta = 64°$, $\varphi = 64°$. The correct (○) parameter values are $\vartheta = 62.4°$, $\varphi = 64.8°$.

Figure 16: Error surface near minimum for (62.4°, 64.8°) translational motion off the line of sight.



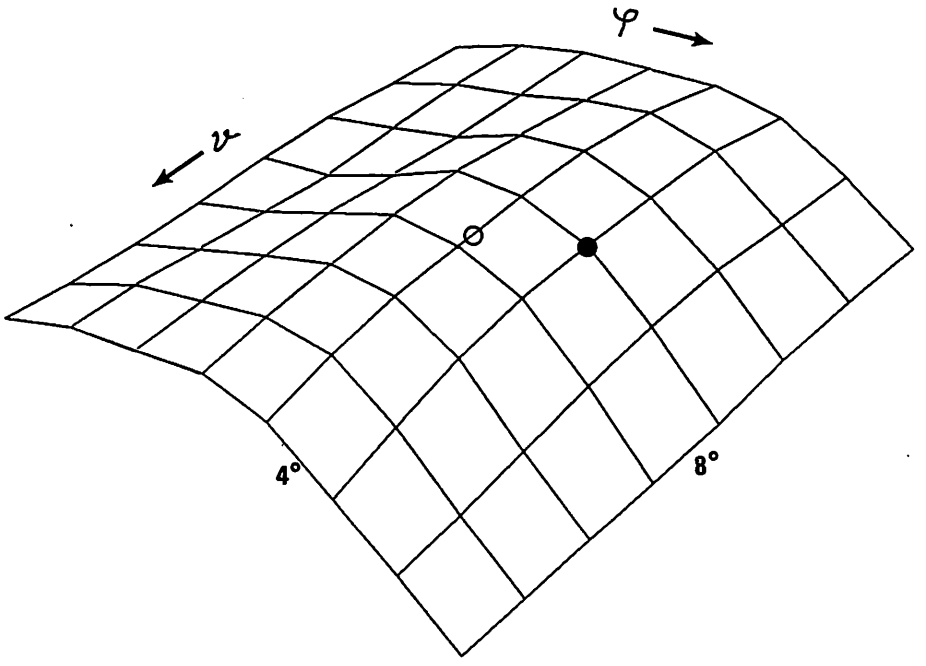
\+ *minimum*

o *correct minimum*

• *minimum from Figure* 16

Smoothed error surface, near the minimum. ϑ changes from 40° -- 102° in increments of 2°. φ changes from $40° - 102°$ in increments of 2°. The minimum (+) corresponds to the error of **0.3938**, at $\vartheta = 68°$, $\varphi = 64°$. The correct (o) values are $\vartheta = 62.4°$, $\varphi = 64.8°$. The regularization parameter is $\beta = 1$.

Figure 17: Smoothed error surface near minimum for (62.4°, 64.8°) translational motion off the line of sight.

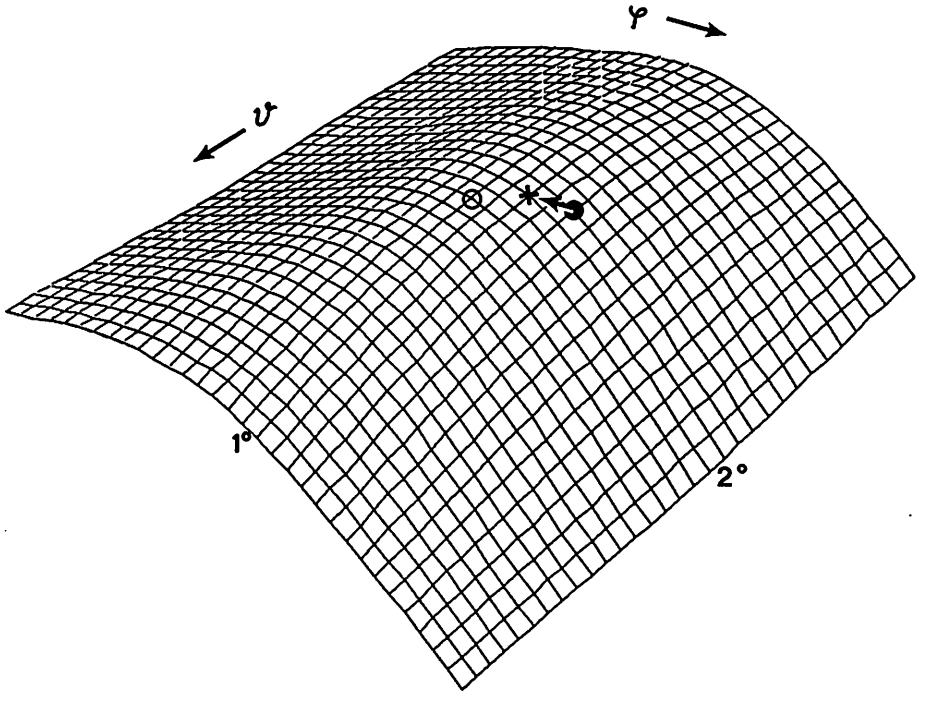


• *minimum*

○ *correct minimum*

Unsmoothed error surface, near the minimum. ϑ changes from 60° – 116° in increments of 8°. φ changes from 74° – 102° in increments of 4°. The minimum (•) corresponds to the error of **0.2238**, at $\vartheta = 84°$, $\varphi = 94°$. The correct (○) parameter values are $\vartheta = 90°$, $\varphi = 90°$.

Figure 18: Unsmoothed error surface near minimum for 90° translational motion off the line of sight.



\+ *minimum*

o *correct minimum*

• *minimum from Figure* 18

Smoothed error surface, near the minimum. ϑ changes from $60° - 122°$ in increments of $2°$. φ changes from $74° - 105°$ in increments of $1°$. The minimum (+) **0.2213** is at $\vartheta = 86°$, $\varphi = 92°$. The correct (o) parameter values are $\vartheta = 90°$, $\varphi = 90°$. The regularization parameter is $\beta = 0.1$.

Figure 19: Smoothed error surface near minimum for 90° translational motion off the line of sight.

cone. The search for the minimum is not influenced by the shape of the surface and there were no ambiguous results. The smoothing process improves the location of the recovered minimum because the position of the minimum is determined by the shape of the whole surface near the minimum. We now discuss some of the topics in the algorithm that are introduced by the new approach and that are not mentioned in Chapter 2.

4.2.1 Choice of the correlation measurement. We tested the influence of the choice of the correlation function on the shape of the error surface (see Section 2.2). The normalized correlation function and the Moravec correlation function gave similar results, both in terms of precision and in terms of efficiency. The absolute difference correlation function was less precise. As in the experiments in Chapter 2, decreasing the step size of feature displacements did not result in significant improvements. The most notable change in the error surfaces occurred when the number of features increased from eight to 16. This can be seen by comparing the Tables 7 and 8 with Tables 1 and 2, Appendix B. The latter choice resulted in much smoother error surfaces.

4.2.2 Influence of noise. The influence of white noise on the error surface was also examined. The first frame in the image sequence was corrupted with white noise whose amplitude was 60% of the range of intensity values in the noiseless image. The shape of the error surface in the image sequence with white noise showed only small change, which is not noticeable on the graph. Since is is difficult to judge

the influence of noise from the change of the shape of the error surface, we give a numerical example. For the translation (45°, 90°) the error values for images with added noise where $E_{min} = 2.2142$ at (46°, 90°) before smoothing, and $E_{min} = 2.2157$ at (48°, 90.5°) after smoothing. A similar result is found in Table 7, Appendix A. E_{min} for the same motion and image without noise was $E_{min} = 0.1980$ as seen in Table 2, Appendix B. Although the absolute values of the error surface increased more than tenfold after the noise was added, the directional error $\Delta\Omega = 3°$ remained in the vicinity of the directional error $\Delta\Omega = 1.1°$ for noise-free image. Thus, the FOE/C approach itself has good stability against noise, as we concluded in the earlier experiments.

4.2.3 How coarsely should the error surface be sampled? One of the advantages of this search procedure is that the sampling of the error surface can be relatively sparse. We tested the influence of the sparsity of the error surface around the minimum on the accuracy of the method. Results are given in Tables 3, 4, and 5 in Appendix B. An example of the error surface sampled every 4° for the translation along (30°, 90°) has been already shown in Figures 10 and 11. The performance deteriorates as the number of sampled points decreases. An error surface of dimension of 5×5 still gives acceptable resolution. Five points for one parameter can still determine the curvature of the error surface with enough precision. A 3×3 surface gives too crude results. By computing the error surface with $2^n + 1$ points, and smoothing it with $2^m + 1$ points ($m > n$) per dimension, the speed is increased by a factor of approximately 2^{m-n}. In our cases the speed increased by a factor

between 4 and 64.

4.2.4 Results for eight feature points. At the end of Appendix B we give results of experiments for 8 feature points in the image. As expected, the results are worse than those with 16 features, though still in acceptable error ranges.

5. CONCLUSIONS AND FUTURE WORK

We have examined a procedure for detection of translational motion of a sensor moving in a static environment by using global constraints on feature dynamics. It has shown a great robustness and quite impressive accuracy. It is the unique combination of the local and global properties of the FOE/C method that results in an optimal performance. We have improved the method by introducing a smoothness constraint on the error surface associated with the position of the FOE/C on the image plane. The smoothing process is superior to a local search process since it increases the speed and robustness of the method while it does not degrade the accuracy of the recovered minimum.

We have found that the performance of the FOE/C algorithm does deteriorate as translations of the camera become more perpendicular to the line of sight. We have shown that this deterioration is directly connected to the insensitivity of the correlation function for cases where the FOE/C is far from the center of the image plane.

The measurements were performed in a highly controlled environment with well-defined interesting points. Real-world images could produce less accurate results. The use of more frames [BHA85] could compensate for the inaccuracy in the position of the FOE/C found from only two frames. We have not analyzed

the effect of the uncertainty in the position of the FOE/C and feature points on error propagation in subsequent frames. This problem is addressed in [SNY86]. Higher resolution images and larger feature displacements would certainly increase the precision of the method.

The major stumbling block in terms of speed remains the bilinear interpolation and the correlation matching. As noted, the use of symbolic features would further increase the speed of the method, and others in our environment are examining this issue.

Note that the surface smoothing which we have introduced is not only a technical device for eliminating noisy error surfaces, but also a form of a regularization constraint. Thus, the surface smoothing reflects our assumption about visual recognition, which is that the overall shape of the error surface is more important in perception of motion than its detailed structure. The choice of the regularization functional is not unique, but the one we used proved to be satisfactory. There are methods to find the optimal regularization parameter, based on known (or assumed) knowledge of the noise in the images. In the case of the Wiener optimal filter [WIE50], for example, it is possible to find the optimal regularization parameter.

The natural extension of this work is the measurement of camera motion parameters for more complicated motions. Two immediate choices are motion in a known plane and rotation around an axis. Although in parameter spaces of higher dimensions ($dim > 2$) the methods we used will be less intuitive, it is possible to

extend the conjugate gradient algorithm to any finite-dimensional parameter space. Perhaps one can introduce a regularization constraint (of the same type we have used) for each pair of the parameters and assign different weights to corresponding functionals.

Another extension of this work is the construction of the environmental depth map using FOE/C and the extent of feature displacements. The depth map can be an intermediate-level knowledge base for higher-level vision algorithms. It can be used for both recognition of the environment and as a feedback for the evaluation of the camera motion parameters.

ACKNOWLEDGMENTS

We would like to thank those people who contributed their time and effort to both the intellectual and system support environment within which this research was done. Especially warm thanks go to Daryl Lawton, P. Anandan, Gilad Adiv, Robert Heller, Michael Boldt, Brian Burns, and Seraj Bharwani for many useful discussions.

APPENDIX A

In this appendix a summary of experimental runs reported in Chapter 2 is presented. Tables showing the results of the evaluation of the FOE/C algorithm for translational motion are presented. The results are discussed in greater detail in Section 2.4. The minimum of an error surface is obtained by a global search followed by a local hill-climbing search, (i.e., the original Lawton algorithm [LAW84]).

Each table is characterized by a change of the value of one of the parameters from a default set of parameter values. These default parameter values are:

1. The step size in the correlation matching along radial lines emanating from the FOE/C between the two windows around interesting points is one pixel.
2. The maximum distance a feature is assumed to move in the search for the best correlation is 10 pixels.
3. The size of the correlation window is 7×7 pixels.
4. The precision with which the translational axis was determined in the (ϑ, φ) space is $\delta_\infty = 0.005$ radians (0.29°).
5. Density of the initial global search for the minimum of the error surface is roughly 45°. The number of steps in the local search is a function of δ_∞.

The results with these default parameter values are shown in Table 1 of this appendix. The other tables summarize results obtained from multiple runs of the algorithm under varying conditions, where usually only one parameter from the set of default parameter values is changed. Tables 1 through 6 are images without noise, while Table 7 presents results with uncorrelated uniform white noise. The

following information is relevant for all tables:

- The parameter that is changed from the default value, given above, is specified in the title of the table.
- The second column in the tables represents correct values (known from the setup of the synthetic images) for the direction of motion. We have used left-handed spherical coordinate space. ϑ is the angle between the translational axis and the Z-axis; φ is the angle between the projection of the translational axis on the X-Y plane and the X-axis.
- Other columns give experimental results for **deviation** (in degrees) of angles ϑ and φ from the correct values.
- Symbol A stands for the "ambiguity" in the search for the right axis, a phenomenon discussed in Section 2.4.
- The quantity $\Delta\Omega = \sqrt{\Delta\vartheta^2 + \Delta\varphi^2}$ is a measure of a total deviation (in degrees) of the experimentally recovered axis from the correct axis.
- At the end of a table the number of correlation matches between the two images in given. This number is the total number of matches during both global and local search.

Table 1: Coarse Resolution Global Sampling of FOE/C Positions.

Correct and experimental values for the translational axis, using the initial (default) set of parameters in the algorithm. (Global sampling every 45°.) Note that the parameter ϑ is the key angular parameter since it measures deviation of motion from the line of sight. The results are discussed in Section 2.4.

Exp.	Correct	4 features		8 features		16 features	
No.	(ϑ, φ)	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$
1	(0, 0*)	(0.7, *)	0.7	(1.1, *)	1.1	(1.4, *)	1.4
2	(15, 90)	(13.7, -4.0)	14.2	(0.8, -4.2)	4.3	(1.2, 1.7)	2.1
3	(30, 90)	A		(3.4, -2.3)	4.1	(1.6, -0.5)	1.7
4	(45, 90)	A		A		A	
5	(60, 90)	A		A		A	
6	(75, 90)	A		A		A	
7	(90, 90)	A		A		A	
8	(120, 90)	(38.2, 0.0)	38.2	A		A	
9	(150, 90)	(8.2, 0.0)	8.2	(8.2, 0.0)	8.2	(8.2, 0.0)	8.2
10	(180, 90)	A		A		A	
11	(90, 45)	A		A		A	
12	(90, 0)	A		A		A	
13	(62.4, 64.8)	A		A		A	
No. of matches		2760		5520		11040	

* (any value of φ is acceptable in this case)

Table 2: Intermediate Resolution Global Sampling of FOE/C Positions.

Increased density of points during global sampling over the unit sphere (approximately every 22.5°, other parameters are the same as in Table 1).

Exp.	Correct	4 features		8 features		16 features	
No.	(ϑ, φ)	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$
1	(0, *)	(0.7, *)	0.7	(1.1, *)	1.1	(1.4, *)	1.4
2	(15, 90)	(13.7, -4.0)	14.2	(0.8,-4.2)	4.3	(1.2, 1.7)	2.1
3	(30, 90)	(20.8, -3.4)	21.0	(3.4, -2.3)	4.1	(1.7, -0.8)	1.9
4	(45, 90)	(16.7, -2.7)	16.9	(10.7, -2.3)	11.0	(-0.1, -0.3)	0.3
5	(60, 90)	A		A		A	
6	(75, 90)	A		A		A	
7	(90, 90)	A		A		A	
8	(120, 90)	(21.3, 0.0)	21.3	(21.3, 0.0)	21.3	(21.3, 0.0)	21.3
9	(150, 90)	(18.7, 0.0)	18.7	(-8.7, 0.0)	8.7	(8.2, 0.0)	8.2
10	(180, 90)	A		A		A	
11	(90, 45)	A		A		A	
12	(90, 0)	A		A		A	
13	(62.4, 64.8)	A		A		(-5.8, 25.2)	25.9
No. of matches		6680		13360		26720	

* (any value of φ is acceptable in this case)

Table 3: Fine Resolution Global Sampling of FOE/C Positions.

Increased density of global sampling (approximately every 11.5°, other parameters are the same as in Table 1). This table is the reference table for the set of runs with smoothing of the error surface presented in Appendix B. Compare also Section 2.4.

Exp.	Correct	4 features		8 features		16 features	
No.	(ϑ, φ)	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$
1	(0, *)	(0.7, *)	0.7	(1.1, *)	1.1	(1.4, *)	1.4
2	(15, 90)	(13.8, -4.0)	14.3	(0.8, -5.0)	5.0	(1.4, 2.4)	2.8
3	(30, 90)	(20.8, -3.5)	21.0	(3.4, -2.3)	4.1	(1.4, -0.2)	1.4
4	(45, 90)	(16.7, -2.7)	16.9	(5.9, -0.2)	5.9	(-0.1, -0.3)	0.3
5	(60, 90)	A		(18.1, -4.3)	18.6	(3.7, 0.0)	3.7
6	(75, 90)	A		A		A	
7	(90, 90)	A		A		A	
8	(120, 90)	(21.3, 0.0)	21.3	(9.8, 0.0)	9.8	(9.8, 0.0)	9.8
9	(150, 90)	(18.7, 0.0)	18.7	(-8.7, 0.0)	8.7	(8.2, 0.0)	8.2
10	(180, 90)	A		A		A	
11	(90, 45)	A		A		A	
12	(90, 0)	A		A		A	
13	(62.4, 64.8)	A		A		(-5.8, 25.2)	
No. of matches		15200		30400		60800	

* (any value of φ is acceptable in this case)

Table 4: Increased Correlation Resolution for Feature Matching.

Displacements between the windows are 1/3 of the pixel size. Other parameters are the same as in Table 1.

Exp.	Correct	4 features		8 features		16 features	
No.	(ϑ, φ)	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$
1	(0, *)	(1.2, *)	1.2	(1.7, *)	1.7	(2.1, *)	2.1
2	(15, 90)	(8.5, -4.4)	9.5	(1.6, -7.1)	7.2	(1.2, 1.0)	1.6
3	(30, 90)	A		(-0.4, -2.1)	2.1	(-2.3, 0.6)	2.4
4	(45, 90)	A		A		A	
5	(60, 90)	A		A		A	
6	(75, 90)	A		A		A	
7	(90, 90)	A		A		A	
8	(120, 90)	(38.2, 0.0)	38.2	A		A	
9	(150, 90)	(8.2, 0.0)	8.2	(8.2, 0.0)	8.2	(8.2, 0.0)	8.2
10	(180, 90)	A		A		A	
11	(90, 45)	A		A		A	
12	(90, 0)	A		A		A	
13	(62.4,64.8)	A		A		A	
No. of matches		8280		16560		33120	

* (any value of φ is acceptable in this case)

Table 5: Small Correlation Window for Feature Matching.

Correlation window size is 3×3. Other parameters are the same as in Table 1.

Exp.	Correct	4 features		8 features		16 features	
No.	(ϑ, φ)	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$
1	(0, *)	(2.4, *)	2.4	(2.3, *)	2.3	(2.2, *)	2.2
2	(15, 90)	(-1.8, 0.0)	1.8	(2.6, -6.9)	7.4	(-1.3, 8.7)	8.8
3	(30, 90)	(-6.4, 0.0)	6.4	(-6.1, 0.5)	6.1	(-4.2, 4.4)	6.1
4	(45, 90)	(1.7, -0.2)	1.7	A		(0.5, -1.0)	1.1
5	(60, 90)	(-3.9, -0.7)	4.0	(-8.5, -0.6)	8.5	A	
6	(75, 90)	A		A		A	
7	(90, 90)	(43.6, 0.2)	43.6	A		A	
8	(120, 90)	(13.6, 0.2)	13.6	A		A	
9	(150, 90)	(8.2, 0.0)	8.2	(-11.7, 0.0)	11.7	(8.2, 0.0)	8.2
10	(180, 90)	A		A		A	
11	(90, 45)	A		A		A	
12	(90, 0)	A		A		A	
13	(62.4, 64.8)	A		A		A	
No. of matches		2760		5520		11040	

* (any value of φ is acceptable in this case)

Table 6: Accuracy of Local Search.

Initial guess (the exact value) for the translation axis is supplied as an input parameter to the algorithm. Other parameters are the same as in Table 1.

Exp.	Correct	4 features		8 features		16 features	
No.	(ϑ, φ)	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$
1	(0, *)	(0.7, *)	0.7	(1.0, *)	1.0	(1.5, *)	1.5
2	(15, 90)	(13.8, -4.0)	14.3	(0.6, -4.6)	4.7	(1.3, 1.7)	2.2
3	(30, 90)	(21.1, -3.5)	21.3	(3.5, -2.4)	4.3	(-2.8, 0.8)	2.9
4	(45, 90)	(16.1, -2.7)	16.4	(10.7, -2.3)	11.0	(0.2, -0.4)	0.4
5	(60, 90)	(15.9, -2.7)	16.1	(8.7, -3.0)	9.2	(3.9, 0.0)	3.9
6	(75, 90)	F**		(25.2, -1.5)	25.2	(2.8, 0.9)	2.9
7	(90, 90)	(13.2, -0.4)	13.2	F**		F**	
8	(120, 90)	(20.7, 0.2)	20.7	(8.0, 2.1)	8.3	(7.5, 1.2)	7.6
9	(150, 90)	(0.1, 0.0)	0.1	(0.1, 0.0)	0.1	(0.1, 0.0)	0.1
10	(180, 90)	A		A		A	
11	(90, 45)	A		A		A	
12	(90, 0)	A		A		A	
13	(62.4, 64.8)	(11.4, -2.6)	11.7	(-6.7, -0.4)	6.7	(10.0, -4.5)	10.9
No. of matches		1720		3440		6880	

* (any value of φ is acceptable in this case)

**(fails to converge)

Table 7: Influence of Noise on Accuracy of Algorithm.

Addition of uniform, **uncorrelated**, white noise. Other parameters are the same as in the experiments in Table 3, with 16 features. For more information refer to Section 2.4.

Exp.	Correct	Experimental values					
No.	(ϑ, φ)	12%*	$\Delta\Omega$	30%*	$\Delta\Omega$	60%*	$\Delta\Omega$
1	(0, **)	(1.8, **)	1.8	(2.2, **)	2.2	(11.5, **)	11.5
2	(15, 90)	(2.1, 1.0)	2.3	(-0.7, -3.9)	4.0	(1.0, 7.3)	7.4
3	(30, 90)	(-3.6, 1.4)	3.9	(-2.8, 1.8)	3.3	(-3.4, 4.2)	5.4
4	(45, 90)	(0.5, -0.3)	0.6	(-6.2, 2.7)	6.7	(2.7, -2.4)	3.6
5	(60, 90)	(4.6, -0.3)	4.6	(4.6, 0.0)	4.6	(9.8, -2.1)	10.0
6	(75, 90)	A		A		A	
7	(90, 90)	A		A		A	
8	(120, 90)	(9.8, 0.0)	9.8	(9.8, 0.0)	9.8	(9.8, 0.0)	9.8
9	(150, 90)	(-1.0, 0.0)	1.0	(-1.0, 0.0)	1.0	(14.8, 18.3)	1.0
10	(180, 90)	A		A		A	
11	(62.4, 64.8)	(1.0, 25.2)	25.2	(-6.1, 25.2)	25.9	(1.0, 25.2)	25.2
No. of matches		60800		60800		60800	

* (percent of uncorrelated noise)

**(any value of φ is acceptable in this case)

APPENDIX B

In this appendix the summary of experimental runs using the improved algorithm developed in Chapters 3 and 4 is presented. The minimum of the error surface is first obtained by a global search over the entire parameter space. The position of the minimal error is returned. Then, the search for the global minimum is continued around the previously found minimum with an increased angular resolution in a smaller parameter domain. Finally, an increase in the resolution is achieved by a surface interpolation process using the conjugate gradient algorithm as described in Chapters 3 and 4.

All the runs have the same default set of algorithm parameters. This is a set which is similar to the set of parameters judged to give accurate results from the first set of experiments (Appendix A, Table 3). The set of default parameter values in this appendix is:

1. The step size along radial lines emanating from the FOE/C in the correlation matching between two windows is one pixel.
2. The maximum distance a feature is assumed to have moved between frames in the search for the best correlation is 10 pixels.
3. The size of the correlation window is 7 x 7 pixels.
4. The step size of the initial global search for the minimum of the error surface is 11.25° ($\pi/16$) for ϑ, and 22.5° ($\pi/8$) for φ.
5. The Moravec correlation function is used for matching between features.

The tables summarize the results obtained from multiple runs of the algorithm under varying conditions. Some of the results are displayed in Figures 9 through 19 in Chapter 4. The results are discussed in a greater detail in Sections 4.1 and 4.2. The regularization parameter β is a new variable in this set of experiments.

The first few tables represent runs with 16 interesting points. Table 1 represents results of a coarse sampling of the error surface for various directions of motion (compare Subsection 4.1.1). Table 2 gives data for a finer search around the minima found in Table 1 and is discussed in Subsection 4.1.2. Tables 3, 4, and 5 contain results with decreased sampling rate of the error surface (Subsection 4.2.3).

Tables 6 and 7 represent runs with eight feature points (Subsection 4.2.4). The results are mostly inferior to those with 16 interesting features. The results of finer search for eight feature points are given in Table 7.

The following information is relevant for the tables:

- The (ϑ, φ) column of the tables represents correct values (known from the setup for synthetic images) for the direction of motion. ϑ is the angle between the translational axis and the Z-axis. φ is the angle between the projection of the translational axis on the X-Y plane and the X-axis. We used left-handed spherical coordinate system. All the angles are in degrees.
- The column with the header $(\vartheta_L, \vartheta_H, \delta\vartheta)$ gives the range of values of ϑ, where ϑ_L, ϑ_H are the minimum and the maximum value (both included), respectively. $\delta\vartheta$ is the increment of ϑ.
- The column with the header $(\varphi_L, \varphi_H, \delta\varphi)$ gives the range of values of φ, where φ_L, φ_H are the minimum and the maximum value, respectively. $\delta\varphi$ is the increment of φ.
- The two columns $(\vartheta_L, \vartheta_H, \delta\vartheta)$ and $(\varphi_L, \varphi_H, \delta\varphi)$ are omitted for the initial runs (Tables 1 and 6) because the parameters are constant for these tables. They are $(\vartheta_L, \vartheta_H, \delta\vartheta) = (0°, 180°, 11.25°)$ and $(\varphi_L, \varphi_H, \delta\varphi) =$

$(0°, 360°, 22.5°)$.

- The E_{min} column gives the minimum value of the error computed directly using the FOE/C technique (i.e. the correlation matches).
- The column with the header $(\Delta\vartheta, \Delta\varphi)$ (not $(\delta\vartheta, \delta\varphi)$!) gives deviations from the correct translational axes, of the translational axis determined by the algorithm (i.e., the position of the minimum of the error surface). $\Delta\vartheta = \vartheta_{min} - \vartheta$ is the deviation of the angle ϑ_{min} (obtained from the error minimum) from the correct value (given in the first column). Similarly, $\Delta\varphi = \varphi_{min} - \varphi$ is a deviation for φ.
- The column with the header $\Delta\Omega$ contains the quantity $\Delta\Omega = \sqrt{\Delta\vartheta^2 + \Delta\varphi^2}$ (in degrees) which is a measure of the total discrepancy of the position of the translational axis from the correct position.
- Other columns represent results with smoothing of the error surface. The superscript CG stands for the values obtained after the conjugate gradient method is used to smooth the error surfaces.
- The total number of the correlation matches (window correlations) between the features in two images of the motions sequence is given at the bottom of the table. These numbers have the same meaning as the numbers given at the bottom of tables in Appendix A.

Table 1: Coarse Resolution Global Sampling of Error Surface.

Results of the runs with 16 feature points. The error surface is computed over the entire parameter range, i.e., $\vartheta \in [0°, 180°]$, $\delta\vartheta = 11.25°$; $\varphi \in [0°, 360°]$, $\delta\varphi = 22.5°$. After that the error surface is smoothed using the conjugate gradient method from Chapter 3. The first major column stands for exact axis of translation, the second major column are data for unsmoothed error surface and the third major column are data for smoothed error surface. For details of notation see the introduction to this appendix. The regularization parameter is $\beta = 1$.

Correct	Error Surface			Smoothed Error Surface		
(ϑ, φ)	E_{min}	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	E_{min}^{CG}	$(\Delta\vartheta, \Delta\varphi)^{CG}$	$\Delta\Omega^{CG}$
(0, 0)	0.1427	(0.0,*)	0.0	0.1227	(0.0,*)	0.0
(15, 90)	0.2931	(-3.8, 0.0)	3.8	0.3826	(7.5, 0.0)	7.5
(30, 90)	0.2734	(3.8, 0.0)	3.8	0.4017	(3.8, 0.0)	3.8
(45, 90)	0.1979	(0.0, 0.0)	0.0	0.4744	(0.0, 0.0)	0.0
(60, 90)	0.3091	(7.5, 0.0)	7.5	0.6503	(7.5, 0.0)	7.5
(75, 90)	0.3297	(3.8, 0.0)	3.8	0.7309	(15.0, 0.0)	15.0
(90, 90)	0.2323	(11.3, 0.0)	11.3	0.7031	(11.3, 0.0)	11.3
(150, 90)	0.3200	(7.5, 0.0)	7.5	0.4149	(-3.8, 0.0)	3.8
(62.4, 64.8)	0.4422	(-6.2, 2.7)	6.8	0.7368	(-6.2, 2.7)	6.8

* φ is arbitrary

Total number of correlation matches 20,480.

Table 2: Very Fine Resolution Error Surface around the Minimum.

Results for 16 feature points, near the error surface minima from Table 1. The regularization parameter is $\beta = 1$. The resolution of the smoothed error surface is equal **twice** the resolution of the unsmoothed error surface. For example, the unsmoothed error surface for translation (15°, 90°), i.e., 15° off the line of sight, is computed over the parameter range $\vartheta \in [5°, 21°]$, $\delta\vartheta = 1°$; $\varphi \in [82°, 98°]$, $\delta\varphi = 1°$. The smoothed error surface is computed for the same parameter range, but with the angular resolution $\delta\vartheta = \delta\varphi = 0.5°$. These numbers can be seen in the second row.

Correct	Parameter space		Error Surface		Smoothed Surface	
(ϑ, φ)	$(\vartheta_L, \vartheta_H, \delta\vartheta)$	$(\varphi_L, \varphi_H, \delta\varphi)$	E_{min}	$\Delta\Omega$	E_{min}^{CG}	$\Delta\Omega^{CG}$
(0, 0)	(0, 16, 1)	(0, 16, 1)	0.1347	2.0	0.1350	2.0
(15, 90)	(5, 21, 1)	(82, 98, 1)	0.2125	2.2	0.2124	1.8
(30, 90)	(22, 38, 1)	(82, 98, 1)	0.2680	2.2	0.2683	1.8
(45, 90)	(30, 62, 2)	(82, 98, 1)	0.1980	0.5	0.1983	0.5
(60, 90)	(44, 76, 2)	(82, 98, 1)	0.2990	8.1	0.2996	8.1
(75, 90)	(60, 124, 4)	(74, 106, 2)	0.3297	5.0	0.3265	11.2
(90, 90)	(60, 124, 4)	(74, 106, 2)	0.2149	2.8	0.2110	4.5
(150, 90)	134, 166, 2)	(82, 98, 1)	0.2880	7.2	0.2874	8.3
(62.4, 64.8)	(40, 104, 4)	(40, 104, 4)	0.3468	10.7	0.3674	6.1

Total number of correlation matches 23,120.

Table 3: Fine Resolution Error Surface around the Minimum from Table 1.

Results for 16 feature points, near the error surface minima from Table 1. The resolution of the smoothed error surface is four times the resolution of the unsmoothed error surface. The regularization parameter is $\beta = 0.1$.

Correct	Parameter space		Error Surface		Smoothed Surface	
(ϑ, φ)	$(\vartheta_L, \vartheta_H, \delta\vartheta)$	$(\varphi_L, \varphi_H, \delta\varphi)$	E_{min}	$\Delta\Omega$	E_{min}^{CG}	$\Delta\Omega^{CG}$
(0, 0)	(0, 16, 2)	(0, 16, 2)	0.1347	2.0	0.1366	2.0
(15, 90)	(5, 21, 2)	(82, 98, 2)	0.2135	2.0	0.2153	1.6
(30, 90)	(22, 38, 2)	(82, 98, 2)	0.2687	2.0	0.2709	1.6
(45, 90)	(30, 62, 4)	(82, 98, 2)	0.1983	1.0	0.2011	1.1
(60, 90)	(44, 76, 4)	(82, 98, 2)	0.3009	8.2	0.3032	0.5
(75, 90)	(60, 124, 8)	(74, 106, 4)	0.3295	9.0	0.3321	7.1
(90, 90)	(60, 124, 8)	(74, 106, 4)	0.2238	7.2	0.2213	9.1
(150, 90)	134, 166, 4)	(82, 98, 2)	0.2896	8.9	0.2944	4.5
(62.4, 64.8)	(40, 104, 8)	(40, 104, 8)	0.3782	1.8	0.3918	12.5

Total number of correlation matches 12,960.

Table 4: Coarse Resolution Error Surface around the Minimum from Table 1.

Results for 16 feature points, near the error surface minima from Table 1. The regularization parameter is $\beta = 0.1$. The resolution of the smoothed error surface is eight times the resolution of the unsmoothed error surface. For example, the unsmoothed error surface for translation (62.4°, 64.8°), i.e.,62.4° off the line of sight, is computed over the parameter range $\vartheta \in [40°, 104°]$, $\delta\vartheta = 16°$; $\varphi \in [40°, 104°]$, $\delta\varphi = 16°$. The smoothed error surface is given for the same parameter ranges, but with the angular resolution equal $\delta\vartheta = \delta\varphi = 2°$. These numbers can be seen in the last row.

Correct	Parameter space		Error Surface		Smoothed Surface	
(ϑ, φ)	$(\vartheta_L, \vartheta_H, \delta\vartheta)$	$(\varphi_L, \varphi_H, \delta\varphi)$	E_{min}	$\Delta\Omega$	E_{min}^{CG}	$\Delta\Omega^{CG}$
(0, 0)	(0, 16, 4)	(0, 16, 4)	0.1427	0.0	0.1383	1.5
(15, 90)	(5, 21, 4)	(82, 98, 4)	0.2135	2.0	0.2212	1.5
(30, 90)	(22, 38, 4)	(82, 98, 4)	0.2703	5.7	0.2739	2.9
(45, 90)	(30, 62, 8)	(82, 98, 4)	0.1983	1.0	0.2109	0.5
(60, 90)	(44, 76, 8)	(82, 98, 4)	0.3089	0.0	0.3157	7.1
(75, 90)	(60, 124, 16)	(74, 106, 8)	0.3334	1.0	0.3701	5.1
(90, 90)	(60, 124, 16)	(74, 106, 8)	0.2294	2.0	0.2766	6.3
(150, 90)	(134, 166, 8)	(82, 98, 4)	0.2896	8.9	0.3088	6.1
(62.4, 64.8)	(40, 104, 16)	(40, 104, 16)	0.4412	13.0	0.4746	10.0

Total number of correlation matches 4,000.

Table 5: Very Coarse Resolution Error Surface around the Minimum from Table 1.

Results for 16 feature points, near the error surface minima from Table 1. The regularization parameter is $\beta = 0.1$. The resolution of the smoothed error surface is **sixteen** times the resolution of the unsmoothed error surface.

Correct	Parameter space		Error Surface		Smoothed Surface	
(ϑ, φ)	$(\vartheta_L, \vartheta_H, \delta\vartheta)$	$(\varphi_L, \varphi_H, \delta\varphi)$	E_{min}	$\Delta\Omega$	E_{min}^{CG}	$\Delta\Omega^{CG}$
(0, 0)	(4, 30, 13)	(4, 30, 13)	0.1523	4.0	0.1522	0.0
(15, 90)	(5, 21, 8)	(82, 98, 8)	0.2508	2.0	0.2508	1.1
(30, 90)	(18, 34, 8)	(78, 94, 8)	0.2703	5.7	0.2764	5.0
(45, 90)	(20, 52, 16)	(80, 96, 8)	0.2218	7.3	0.2394	7.3
(60, 90)	(48, 80, 16)	(80, 96, 8)	0.3088	4.5	0.3342	6.2
(75, 90)	(60, 124, 32)	(80, 112, 16)	0.3999	16.2	0.4556	15.8
(90, 90)	(70, 134, 32)	(80, 112, 16)	0.4154	13.4	0.4407	13.4
(150, 90)	(140, 172, 16)	(78, 94, 8)	0.2880	7.2	0.3086	3.6
(62.4, 64.8)	(40, 104, 32)	(40, 104, 32)	0.6096	12.0	0.7213	9.2

Total number of correlation matches 1440.

Table 6: Coarse Resolution Global Sampling of Error Surface (8 features).

Results of the runs with eight feature points. The error surface is computed over the entire parameter range, i.e., $\vartheta \in [0°, 180°]$, $\delta\vartheta = 11.25°$; $\varphi \in 0°, 360°$, $\delta\varphi = 22.5°$. The regularization parameter is $\beta = 0.1$. The results of smoothing are shown here to emphasize the dependence of the results on β (here they are bad). More appropriate choice would be $\beta \approx 1$. In general, 8 features is performing worse.

Correct	Error Surface			Smoothed Error Surface		
(ϑ, φ)	E_{min}	$(\Delta\vartheta, \Delta\varphi)$	$\Delta\Omega$	E_{min}^{CG}	$(\Delta\vartheta, \Delta\varphi)^{CG}$	$\Delta\Omega^{CG}$
(0, 0)	0.0717	(0.0, 0.0)	0.0	0.0172	(0.0, 0.0)	0.0
(15, 90)	0.0964	(7.5, 0.0)	7.5	0.1455	(30.0, 0.0)	30.0
(30, 90)	0.0872	(3.8, 0.0)	3.8	0.1597	(20.6, -11.2)	23.4
(45, 90)	0.0645	(11.3, 0.0)	11.3	0.2422	(11.3, -11.2)	15.9
(60, 90)	0.1245	(-3.8, 0.0)	3.8	0.3370	(-3.8, -11.2)	11.8
(75, 90)	0.1014	(26.3, 0.0)	26.3	0.3381	(48.8, -11.2)	50.0
(90, 90)	0.0846	(0.0, 0.0)	0.0	0.4162	(33.8, -11.2)	35.6
(150, 90)	0.1124	(-3.8, 0.0)	3.8	0.1404	(-20.6, 0.0)	20.6
(62.4, 64.8)	0.1429	(-6.2, 2.7)	6.8	0.2416	(-11.8, 2.7)	12.1

Total number of correlation matches 10,240.

Table 7: Very Fine Resolution Error Surface around the Minimum from Table 6 (8 features).

Results for eight feature points, near the error surface minima from Table 6. The regularization parameter is $\beta = 0.1$. The resolution of the smoothed error surface is equal twice the resolution of the unsmoothed error surface. Here, β is smaller than needed, but it does not influence the smoothing as much as in Table 6 due to the lower curvature of the surface.

Correct	Parameter space		Error Surface		Smoothed Surface	
$(\vartheta,\ \varphi)$	$(\vartheta_L,\ \vartheta_H,\ \delta\vartheta)$	$(\varphi_L,\ \varphi_H,\ \delta\varphi)$	E_{min}	$\Delta\Omega$	E_{min}^{CG}	$\Delta\Omega^{CG}$
(0, 0)	(0, 16, 1)	(0, 16, 1)	0.0696	1.0	0.0681	0.5
(15, 90)	(5, 21, 1)	(82, 98, 1)	0.0705	4.1	0.0730	6.2
(30, 90)	(22, 38, 1)	(82, 98, 1)	0.0823	4.5	0.0825	4.3
(45, 90)	(30, 62, 2)	(82, 98, 1)	0.0576	11.2	0.0582	10.2
(60, 90)	(44, 76, 2)	(82, 98, 1)	0.1060	8.5	0.1073	10.4
(75, 90)	(60, 124, 4)	(74, 106, 2)	0.0958	21.1	0.0986	31.1
(90, 90)	(60, 124, 4)	(74, 106, 2)	0.0813	10.2	0.0786	4.1
(150, 90)	(134, 166, 2)	(82, 98, 1)	0.1041	6.7	0.1050	7.5
(62.4, 64.8)	(40, 104, 4)	(40, 104, 4)	0.1154	10.7	0.1323	3.7

Total number of correlation matches 11,560.

APPENDIX C - Conjugate Gradient Method

In this appendix we present in more detail the conjugate gradient method, without proofs of equations, but with enough rigor to help the reader understand the basic concepts and steps. For further details, the reader is referred to the Hestenes [HES80].

Fundamental Concepts

The problem with which we are concerned here is the minimization of the positive-definite quadratic functional

$$f(x) = \frac{1}{2}x^* Ax - h^* x + C. \qquad (C.1)$$

x is an n-dimensional vector of unknown values, A is an $n \times n$ positive-definite matrix, and h is a vector of given data. The symbol $*$ denotes the transposition of matrices and vectors. C is an arbitrary scalar. The solution of the problem x_o satisfies the equation

$$Ax_o - h = 0. \qquad (C.2)$$

To find the minimum of $f(x)$ in Eq. (C.1), one has to solve the set of linear equations, Eq. (C.2). When the matrix A is sparse (i.e., it has only few elements) the conjugate gradient methods are the best choice. From the previous

two equations it follows that

$$f(x) = f(x_\circ) + \frac{1}{2}(x - x_\circ)^* A\,(x - x_\circ). \qquad (C.3)$$

In the last equation we recognize the generalization of a simple quadratic equation to n dimensions. For $n = 1$ the solution can be obtained, for example, by Newton's iterative method. Given x_k as a guess for the solution $x_\circ$, a better guess for the solution is

$$x_{k+1} = x_k - f'(x_k)/f''(x_k) \qquad (k = 0, 1, \ldots) \qquad (C.4)$$

For $n > 1$ the equations are similar, and only a generalization of the functional derivatives has to be introduced.

Equal level surfaces in Eq. (C.3)

$$f(x) = Const, \qquad (C.5)$$

are $(n-1)$–dimensional ellipsoids having $x_\circ$ as their center. The derivatives $f'(x)$ and $f''(x)$ become gradient and Hessian in n dimensions, respectively

$$f'(x) \rightarrow \left(\partial f(x)/\partial x^i\right) \qquad (i = 1, 2, \ldots, n), \qquad (C.6)$$

$$f''(x) \rightarrow \left(\partial^2 f(x)/\partial x^i \partial x^j\right) \qquad (i, j = 1, 2, \ldots, n). \qquad (C.7)$$

The Taylor formula becomes

$$f(x + z) = f(x) + f'(x)^* z + \frac{1}{2} z^* f''(x)\, z + R(x, z), \qquad (C.8)$$

where the function $R(x, z)$ is the remainder.

An important step toward the generalization to n dimensions is the use of the concept of minimization of f along a line L through a point x_1 in the direction of a non-zero vector p:

$$L: \qquad x = x_1 + \alpha p, \tag{C.9}$$

where α is a parameter. Thus, we search for an α such that the function

$$\phi(\alpha) = f(x_1 + \alpha p) \tag{C.10}$$

has the minimum. α is found as a value for which $x_2 = x_1 + \alpha_1 p$ satisfies

$$\phi'(\alpha) = f'(x_2)^* p = 0 \tag{C.11}$$

and

$$\phi''(\alpha) = p^* f''(x_2)\, p > 0. \tag{C.12}$$

x_2 is called the critical point and we see from Eq. (C.11) that the gradient at that point is *orthogonal* to the line L. Another way of writing the solution is

$$x_2 = x_1 + \alpha_1 p, \tag{C.14a}$$

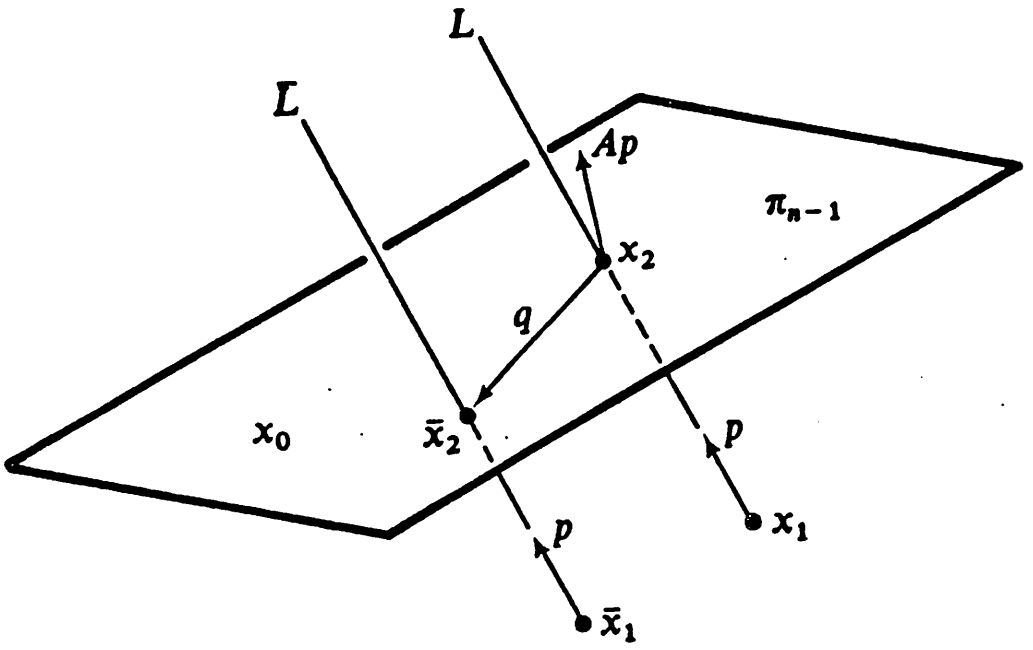
where

$$\alpha_1 = \frac{c}{d}; \qquad c = p^* r_1; \qquad d = p^* A\, p \tag{C.14b}$$

and

$$r_1 = -f'(x_1) = h - A x_1. \tag{C.14c}$$

r_1 is called the residual (negative gradient) of f at the point x_1 and points in the direction of the steepest descent.



Starting from a point x_1 in the direction $-f'(x_1)$ the minimum of $f(x)$ on line L is found to be x_2. Vector Ap is normal on the hyperplane $\hat{\pi}_{n-1}$. Search continues in the hyperplane $\hat{\pi}_{n-1}$, starting from the point x_2 in the new direction q.

Figure 1: Search for the minimum x_o.

The equation

$$p^*(Ax - h) = 0 \qquad (C.15)$$

is one constraint on a set of n values of x and defines an $(n-1)$–dimensional hyperplane $\hat{\pi}_{n-1}$. The hyperplane $\hat{\pi}_{n-1}$ passes through the solution x_o and the vector Ap is normal to it. The situation is depicted in Figure 1.

Another way of specifying the normality of $A\,p$ on the hyperplane $\hat{\pi}_{n-1}$ is by requiring that for

$$\forall q, \quad q \in \hat{\pi}_{n-1}: \qquad p^* A\, q = 0. \tag{C.16}$$

p is said to be conjugate (A–orthogonal) to $\hat{\pi}_{n-1}$. Observe that if A is the identity matrix the conjugacy becomes the usual orthogonality of vectors.

By finding the minimum of f along the line through x_1 the dimensionality of the problem is reduced by one. The search can now continue from the point $x_2 \in \hat{\pi}_{n-1}$, since the solution x_0 is somewhere in that hyperplane. In what direction should the search for x_0 proceed from x_2, i.e., what is a good choice for q (see Figure 1)? If such a q is found, then starting from x_2 one finds the minimum x_3 along the line specified by q and reduces the problem to the search for x_o in the hyperplane $\hat{\pi}_{n-2}$, and so on. The conjugate gradient method specifies that the new direction should be along the vector $q = p_k$ which is conjugate to *all* previous directions of the steepest descent:

$$p_k^* A p_i = 0 \qquad i = 1, 2, \ldots, (k-1). \tag{C.17}$$

It can be shown that the residual r_{k+1} is orthogonal (in the common sense of the word) to the direction vectors $p_1, \ldots, p_k$, signifying that x_{k+1} minimizes f on the k-plane

$$\hat{\pi}_k: \qquad x = x_1 + \alpha_1 p_1 + \cdots + \alpha_k p_k.$$

All these conclusions lead to the following CG-Algorithm.

Initial Step. Select a point x_1 and compute

$$p_1 = r_1 = -f'(x_1) = h - Ax_1. \qquad (C.18a)$$

Iterative steps. Having obtained x_k, r_k, and p_k compute x_{k+1}, r_{k+1}, and p_{k+1} by the formulas

$$a_k = \frac{c_k}{d_k}, \quad d_k = p_k^* A\, p_k, \quad c_k = p_k^* r_k, \qquad (C.18b)$$

$$x_{k+1} = x_k + a_k p_k, \quad r_{k+1} = r_k - a_k A\, p_k, \qquad (C.18c)$$

$$b_k = -p_k^* A r_{k+1}/d_k, \quad \textit{or} \quad b_k = \frac{|r_{k+1}|^2}{c_k}, \qquad (C.18d)$$

$$p_{k+1} = r_{k+1} + b_k p_k. \qquad (C.18e)$$

Termination. Terminate the m-th step if $r_{m+1} = 0$. Then $m \leq n$ and $x_{m+1} = x_o$, the minimum point of f.

The termination step is, according to the last statement, guaranteed to be reached by the n–th step (n being the dimensionality of the problem). The magnitude of the residual $|r_k|$ is not necessarily monotonically decreasing toward zero, thus it is never known how far away the solution is. Round-off error can cause $|r_k| \neq 0$ for all $k \leq n$. This is the problem of great importance for ill-conditioned matrices (for which the ratio of the largest and the smallest eigenvalue is a large number) where the convergence is slow. To avoid such problems, take the following steps:

- Extend the computation just one more step $(n + 1)$. This gives much better accuracy [HES80] with no significant increase in the computational cost.

- Compute the initial approximation for the solution by fitting planar patches between the sparse data (done through bilinear interpolation). The sparse data are assumed to be known on a regular grid of a lower level of resolution. A good initial guess assures faster convergence and fewer problems with round-off error.
- If the norm of the residual after $n + 1$ steps (exit condition for this iterative procedure) is not satisfactorily small, repeat the computation with the initial approximation for the solution being the previously found "solution". Starting with an approximation so close to the solution guarantees success.

This completes our exposition of the conjugate gradient method used in this paper.

REFERENCES

[ABR68] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions.* Dover Publications, Inc., New York, 1968.

[ADI84] G. Adiv, *Determining 3-D Motion and Structure from Optical Flow Generated by Several Objects.* Proceedings of Image Understanding Workshop, DARPA, October 1984, pp. 113-129.

[AGG81] J. K. Aggarwal, and W. N. Martin, *Correspondence Processeses in Dynamic Scene Analysis* Proceedings of the IEEE, 69, No. 5, May 1981, pp. 562-572.

[BAL81] D. H. Ballard, *Parameter Networks: Towards a Theory of Low-Level Vision.* IJCAI-7, Vancouver, British Coloumbia, August 1981, pp. 1068-1078.

[ANA85] P. Anandan, *A Review of Motion and Stereopsis Research.* Technical Report 85-52, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, December 1985.

[ANA86] P. Anandan, *Computing Optical Flow From Two Frames of an Image Sequence.* Technical Report 86-16, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, April 1986.

[BAL82] D. H. Ballard and C. M. Brown, *Computer Vision.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.

[BHA85] S. Bharwani, E. Riseman, and A. Hanson, *Refinement of Environment Depth Maps Over Multiple Frames.* Proceedings of Image Understanding Workshop, DARPA, December 1985, pp. 413-420.

[COU53] R. Courant and D. Hilbert, *Methods of Mathematical Physics.* Interscience Publishers, Inc., New York, 1953.

[DAV83] L. S. Davis, L. Kitchen, F. P. Hu, and V. Hwang, *Image Matching Using Generilized Hough Transform.* CAR-TR-27, Center for Advanced Automation Research, University of Maryland, College Park, MD 20742, October, 1983.

[FAN82] J. Q. Fang, and T. S. Huang, *A Corner Finding Algorithm for Image Analysis and Registration.* Proceedings of the AAAI-82, pp. 46-54.

[GIB50] J. J. Gibson, *The Perception of the Visual World.* Riverside, Cambridge, Mass, 1950.

[GRI81] W. E. L. Grimson, *A Computational Theory of Visual Surface Interpolation.* MIT, A.I. Memo 613, 1981.

[GRI82] W. E. L. Grimson, *A Computational Theory of Visual Surface Interpolation.* Phil. Trans. R. Soc. Lond., 1982, **B 298**, pp. 395-427.

[HAN84] A. R. Hanson and E. M. Riseman, *A Summary of Image Understanding Research at the University of Massachusetts.* Technical Report 83-35, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, October 1983.

[HES80] M. Hestenes, *Conjugate Direction Methods in Optimization.* Springer-Verlag, 1980.

[HIL84] E. C. Hildreth, *Computation of the Velocity Field.* Proc. R. Soc. Lond., 1984, **B 221**, pp. 189-220.

[HOR80] B. K. P. Horn and B. G. Schunck, *Determinining Optical Flow.* MIT, A.I. Memo 572, April 1980. Also, Artificial Intelligence, 1981, **17**, pp. 185-203.

[IKE81] K. Ikeuchi and B. K. P. Horn, *Numerical Shape from Shading and Occluding Boundaries.* Artificial Intelligence, 1981, **17**, pp. 141-184.

[LAW83] D. T. Lawton, *Processing Translational Motion Sequences.* Computer Graphics and Image Processing, 1983, **22**, pp. 116-144.

[LAW84] D. T. Lawton, *Processing Dynamic Image Sequences from a Moving Sensor.* Ph. D. Dissertation. University of Massachusetts, Amherst, MA 01003. Also, Technical Report 84-05, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, February 1984.

[MED85] G. Medioni and Y. Yasumoto, *Robust Estimation of 3-D Motion Parameters from a Sequence of Image Frames Using Regularization.* Proceedings of Image Understanding Workshop, DARPA, December 1985, pp. 117-128.

[MOR77] H. P. Moravec, *Towards Automatic Visual Obstacle Avoidance.* Proceedings of the 5th IJCAI, MIT, Cambridge, MA, 1977, p. 584.

[MOR81] H. P. Moravec, *Robert Rover Visual Navigation.* UMI Research Press, Ann Arbor, Michigan, 1981.

[ORO81] J. O'Rourke, *Motion Detection Using Hough Techniques.* Proceedings of PRIP, 1981, pp. 82-87.

[PAV85] I. Pavlin, E. Riseman, and A. Hanson, *Analysis of an Algorithm for Detection of Translational Motion.* Proceedings of Image Understanding Workshop, DARPA, December 1985, pp. 388-398.

[POG84] T. Poggio and V. Torre, *Ill-Posed Problems and Regularization Analysis in Early Vision.* A.I. Memo **773**, MIT, Cambridge, MA, 1984.

[POG85] T. Poggio et al., *MIT Progress in Understanding Images.* Proceedings of Image Understanding Workshop, DARPA, December 1985, pp. 25-39.

[SMI86] F. Smid, L. Kitchen and E. Riseman, *A Study of Interest-Point Operator for Token-Based Image Matching.* Technical Report, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, to be published.

[SNY86] M. Snyder, *The Accuracy of 3D Parameters in Correspondence-Based Techniques.* Technical Report 86-28, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003, July 1986.

[STU80] F. Stummel and K. Hainer, *Introduction to Numerical Analysis.* Scottish Academic Press, Edinburgh, 1980.

[TER82] D. Terzopoulos, *Multi-Level Reconstruction of Visual Surfaces.* A.I. Memo **671**, MIT, Cambridge, MA, 1982.

[TER84] D. Terzopoulos, *Computation of Visible-Surface Representations.* Ph. D. Dissertation, Massachusetts Institute of Technology, January 1984.

[TIK63] A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method.* Soviet Math. Dokl., 1963, **4**, pp. 1035-1038.

[TIK77] A. N. Tikhonov and V. Y Arsenin, *Solution of ill-posed problems.* Winston & Sons, Washington, D.C., 1977.

[ULL79] S. Ullman, *The Interpretation of Structure from Motion.* Proc. R. Soc. Lond., 1979, **B 203**, pp. 405-426.

[ULL81] S. Ullman, *Analysis of Visual Motion by Biological and Computer Systems.* IEEE Transactions on Computers, August 1981, pp. 57-69.

[WHI80] T. Whitted, *An Improved Illumination Model for Shaded Display.* Communications of the ACM, 1980, **23**(6), pp. 343-349.

[WIL80] T. D. Williams, *Depth from Camera Motion in Real World Scene.* IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-2**(6), 1980, pp. 511-516.

[WIE50] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* Wiley, New York, 1950.