Reconstruction of a Real Object Using Stereo Vision

Ignatios E. Vakalis
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses
Part of the Computer Sciences Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/1188

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
RECONSTRUCTION OF A REAL OBJECT
USING STEREO VISION

by
Ignatios E. Vakalis

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Science
Department of Computer Science

Western Michigan University
Kalamazoo, Michigan
April 1988
RECONSTRUCTION OF A REAL OBJECT
USING STEREO VISION

Ignatios E. Vakalis, M.S.
Western Michigan University, 1988

The problem of reconstructing a three dimensional object from its two dimensional image is the most critical problem which any powerful visual system must solve. The human visual system is known to solve this problem with impressive range and resolution. Computer solutions have fallen far short of human performance mainly because the solution process is poorly understood.

The reconstruction problem is solved using stereo vision. The distance of points on the object from the viewing device and the orientation of specific patches of the visual surface with respect to the viewing device are computed. Discontinuities in distance and orientation are directly labelled and identified before the full representation is completed. A different approach to the interpolation problem is given, preserving the "wobbling" of the reconstructed visual surface as the viewpoint changes. The algorithm presented is able to reconstruct the whole surface of the object, even if it is not initially visible by the system.
ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Dionysios Kountanis for his helpful advice during the preparation of this thesis. I would also like to thank Kathleen Gordon for her expertise during the writing stage and finally, a big thank you to Marianne, for her patience, understanding and unfailing support, always.

Ignatios E. Vakalis
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106
Reconstruction of a real object using stereo vision

Vakalis, Ignatios E., M.S.
Western Michigan University, 1988
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .......................................... ii
LIST OF FIGURES ............................................ vi

CHAPTER

I. INTRODUCTION ...................................... 1
II. REVIEW OF THE LITERATURE ....................... 11
III. METHOD ............................................ 18
    Introduction .................................... 18
    Constraints and Acquisition Device ............... 19
    Algorithm ....................................... 22
IV. CORRESPONDENCE PROBLEM ............................. 26
    General Overview ................................ 26
    False Target Problem ............................. 28
    Detect and Represent Intensity Changes ........... 30
    Solution of the Problem ......................... 38
V. CALIBRATION PROBLEM ................................. 47
    Camera Model .................................... 47
    Solution ........................................ 50
VI. 3-D DISTANCE ...................................... 55
    Camera Geometry ................................ 56
    Perspective Transformation ....................... 58
    Inverse Perspective Transformation ............... 61
    Distance Calculations ........................... 64

iii
Table of Contents—Continued

CHAPTER

Error Analysis .................................. 68
Inverse Rotation Transformation ................. 69

VII. ORIENTATION PROBLEM . . . . . . . . . . . . . . . . 70
Surface Orientation ............................ 70

VIII. RECONSTRUCTION OF VISIBLE SURFACE . . . . . . . . 73
Surface Consistency Constraint ................. 74
The Computational Problem ..................... 78
Possible Functionals ............................ 80
What Has Been Achieved ......................... 82
Discontinuities ................................ 84
Loose Constraints .............................. 85
Energy of Loose Plate .......................... 86
Sensitivity in Discontinuity Detection .......... 88
Viewpoint Invariance in Surface Reconstruction .... 90

IX. SUMMARY AND CONCLUSIONS ............. 97
Extensions: Open Problems ....................... 100

APPENDICES ............................................. 103
A. DOGs and V^2G ................................. 104
B. Transformation Matrices ....................... 106
C. Distance Using Glimbal Center ................ 109
Table of Contents—Continued

APPENDICES

D. Rotationally Symmetric Operators ..................... 112
E. Energy Calculations ..................................... 114
F. Derivation of Special Case Energies for a Continuous 1-D Rod ................................................. 117
G. Loose Membrane ....................................... 120

BIBLIOGRAPHY .............................................. 122
LIST OF FIGURES

1. Stages of Visual Processing ......................... 3
2. Overview of Early Visual Processing ................. 6
3. Representation of the Stereo Process ................. 8
4. Three Dimensional Acquisition System ............... 21
5. One Dimensional Detection of Image Changes ........... 32
6. Cross Section of Rotationally-Symmetric Operator .... 37
7. Match Over w/2 Range ................................ 42
8. Matching With Shifting ................................ 43
9. Ambiguous Matches .................................. 45
10. Reference Frame for Camera and Reference Frame for Perspective Image Plane ............... 48
11. Camera Geometry .................................... 54
12. Perspective Transformation With Two Reference Frames ... 57
13. Geometry of Stereopsis .............................. 65
14. Error Analysis for Stereo Depth Measurement ........ 68
15. Geometry of Surface Orientation ..................... 71
16. Fitting A Smooth Curve to Data With Discontinuities ... 85
17. Bi-Infinite Data d(x) With An Isolated Crease ........ 90
18. Invariant Distance Measurement ..................... 93
CHAPTER I

INTRODUCTION

Computer vision is the information-processing task of "understanding" a scene from its projected images. The process uses well-defined input and output. The input consists of a number of arrays of brightness values representing projections of a three-dimensional scene recorded by a camera or comparable imaging device. Several input arrays may provide information from several spectral bands (color) or from multiple viewpoints (stereo or time sequence).

The desired output is a concise description of the three-dimensional scene depicted in the image, the exact nature of which depends upon the goals and expectations of the observer. It generally involves a description of objects and their interrelationships, but may also include such information as the three-dimensional structure of surfaces, their physical characteristics (shape, texture, color), the location of shadows, light sources, etc.

Applying a computational approach to the vision problem, the study of the process is divided into three different levels: the computational theory, the algorithm to test the theory and the underlying implementation of the algorithm.

At the level of the computational theory, the physical
constraints that restrict the problem of describing the three-dimensional scene depicted in the images must be determined to allow the process to do what it does. The constraints may be treated as assumptions that are always taken as true, with or without verification. Also at this level the computational device is characterized as a mapping from one kind of information to another and the abstract properties of this mapping are defined precisely.

Having developed a computational theory of the processing involved in a visual task, one can then turn to the design of a particular algorithm to achieve the task. Many questions have to be answered. How can the computational theory be implemented? What is the representation for the input and output? How can an algorithm for the transformation be devised?

The third level of description is that of the implementation. How can the algorithm and the representation will be realized physically? What kind of computer architecture will be used?

It is important to mention that for a particular computational theory there may be many algorithms, even with the same overall structure and many more implementations of the algorithms. In this work only one method is presented for constructing three-dimensional representations of an object from one pair of images.

The overall structure of the solution to the vision problem is a two-stage process: early and late visual processing. In early processing the goal is to get useful information from the raw image and every part of the image is processed the same way. In the late processing the goal is to recognize the object(s) in the scene from
the useful information. Figure 1 suggests that the early processing is parallel and local; that is, lots of identical processing of small regions of the image are carried out. Its results are revealing enough that higher level processes can walk through them sequentially and extract a symbolic description of the three dimensional scene.

![Figure 1. Stages of Visual Processing.](image)

Taking a closer look at Figure 1 three representations in the course of visual processing can be identified: the primal sketch, the 2 1/2-D sketch and the 3-D model.

The primal sketch consists of a representation of significant intensity changes of the raw images to be processed. This representation was first introduced by Marr (1982), but in the years that followed his original work, many changes have taken place especially in the methods for detecting and analyzing intensity
changes in an image. In this work the method of zero-crossings is presented for the detection of changes, where the image is convolved by a Laplacian of a two-dimensional Gaussian filter. As Figure 1 shows, low level visual processing operates on the primal sketch representation, which is obtained by an initial process on the raw image, and creates as output a new representation which is called the 2 1/2-D sketch.

The 2 1/2-D sketch is the destination of the early processing and the starting point of the late processing. It is known that the final goal of a complete visual process is to identify and recognize objects, mainly three dimensional, from their projections as two dimensional images. Past experience has shown that the goal is impossible to be achieved if the visual process operates directly on the primal sketch without generating any other intermediate representations before the final goal is reached. This necessity brought up what it is called 2 1/2-D sketch. It is a representation of some useful information about the visual surfaces of a three dimensional object such as orientation, distance from the viewing system, reflectance, illumination, etc. The 2 1/2-D sketch is divided into two parts. The raw 2 1/2-D sketch consists of a representation of the useful information at points which are explicitly represented in the primal sketch. The second part, namely full 2 1/2-D sketch, is obtained from the first using interpolation techniques through the points of the raw 2 1/2-D sketch. The final outcome is a representation of the useful information for every point of the image in process.
The 3-D model is an object-centered representation of the three-dimensional object(s) in the scene, and of the organization of its viewed shape. The 3-D model contains structural information about the object(s) and its parts, if any, as well as their shapes and their spatial relationships. This description and representation is used in the tasks of visual object recognition and visually-guided manipulation. It has to be noted here that the extraction of such a representation is very difficult compared to the primal and 2 1/2-D sketch. This is because the further away from the raw image the visual process operates, the greater the uncertainty about the nature of the specific steps. This is one of the main reasons why most of the research in the literature has been done on the primal and 2 1/2-D sketches.

In this work only the first two representations are investigated by creating a description of image locations at which the corresponding scene location undergoes a change in a physical property. By processing the primal sketch, a description of the surface geometry of an object is obtained. In other words only part of the early visual processing is presented.

Figure 2 illustrates some of the modules of the early visual processing as seen in past literature.
Figure 2. Overview of Early Visual Processing.

The first step of the processing is the production of the primal sketch which mainly consists of points or contours, i.e., edges, where the image intensity undergoes a radical, abrupt change. This is done by processing the raw images or more precisely, the information stored in the gray levels of the image.

The following four modules of Figure 2, motion analysis, texture analysis, stereo analysis, and photometry, operate on the primal sketch representation. Texture analysis finds surface patches that consist of similarly organized elements and extracts information from changes in these elements. Motion analysis finds the features that appear to be moving together and extracts the
shape of the surface to which they belong. Photometry derives the shape of the visual surface from shading information.

The last module of the early visual processing is the stereo module which will be analyzed extensively in this work. The stereo problem can be defined as follows. Assume that images of a scene are available from two nearby points at the same horizontal level, i.e., the analogy of the images that play upon the retinas of the left and right human eyes. The goal of the visual process by a computer is the determination of the depth and orientation of every point on the surface of an object from the viewing device.

A fundamental assumption is being made at the level of the computational theory, that the human visual system is inherently modular, allowing the study of the process of stereo vision in isolation. It is generally believed that the human visual system is the best and so in this work an attempt is made to reconstruct the visual surface of a three dimensional object(s) by simulating what is known about the human visual process.

Figure 3 gives an outline of the process to be followed for the reconstruction of the whole surface of the object.
Figure 3. Representation of the Stereo Process.

Assume that images of a scene are available from two nearby points (i.e., 6 to 7 inches apart) at the same horizontal level. The images in the left and right cameras of the imaging device are somewhat different as a consequence of slight difference in vantage. This geometrical difference is called disparity. Assume further that a particular location of the surface in the scene is chosen from one image. The first part of the stereo problem is the identification of a point on the other image which corresponds to the same location of the scene. The correspondence cannot be achieved by matching gray-level values between the two images because of the high ambiguity in the matching process. A new representation the primal sketch, is developed for each image of the stereo pair, consisting of those positions at which a change in the image intensity takes place. The above step is accomplished using a method suggested by Marr, Poggio and Grimson, where each image is
convolved by a series of different size filters. Each filter generates a convolution of the image intensity with a Laplacian of a Gaussian operator. The matching process introduces the false target problem due to the large number of candidates for a correct match. During the matching process the computation of explicit surface information at the demarked locations of the primal sketches is accomplished.

In this work, a complete and thorough analysis is presented on how to compute the following useful information: the distance of object points from the viewing device and the orientation of specific patches of the visual surface with respect to the viewing device. During the derivation of the above computation a solution to the calibration problem is presented which is essential for the determination of the camera parameters. After the raw 2 1/2 - D sketch representation is obtained, the interpolation method is presented which will compute explicit surface information for all points in the scene. This will yield a complete representation of the surface shape of the object(s). It should be mentioned here that all the previous work in solving the stereo problem results in the representation of only the visible surface from a pair of cameras. In this work, this is extended and the goal is the reconstruction of the whole surface even if it is not initially visible by the two cameras. Through the process of solving the interpolation problem, this work presents a slightly different approach in that the discontinuities in distance, of the visual surface from the viewing device, and its orientation, are directly
labelled and identified before the full representation is completed. An analytic solution is also presented for the interpolation problem which preserves the "wobbling" of the reconstructed visual surface as the viewpoint changes.
CHAPTER II

REVIEW OF THE LITERATURE

The two most important and most difficult problems in the process of the complete reconstruction of the surface of a visible object using stereo vision techniques are the correspondence problem, and the selection of the best surface to be fitted through the scattered depth data produced by the stereo module. As a result most of the research in the area of the early visual processing addresses these problems. Many techniques have been proposed but unfortunately none of them succeeds in giving a general solution to the problems. The effectiveness and efficiency of the methods are directly influenced by the types of the objects to be examined or more generally by the types of the scenes (indoor, outdoor, natural images, polyhedral objects, etc.). In this chapter the most important techniques will be reviewed as well as the specific methods which have been chosen in this work to solve the above problems, will be presented.

The first major problem of every stereo system lies in the way of handling the matching process. This introduces the false target problem occurring because of the large number of candidates for a correct match. A lot of methods have been proposed for the minimization of the false target problem and the solution of the
overall correspondence problem.

The most important approaches for the solution of the matching problem are briefly presented here and fall into five categories:

1. Correlation techniques: The correlation of regions approach (Babu & Nevatia, 1980) tries to find "similar" regions in the two images without an analysis of the contents of these images. The choice of regions is based on the likelihood of finding corresponding regions for them and not necessarily on their expected importance for the analysis of the scene. The comparison tries to determine whether the two subimages (small regions of the stereo images) are formed by the same part of the surface of the object.

Yakimovky & Cunningham (1978) suggest a correlation algorithm in which specially configured masks in the shape of a concentric diamond are used instead of the more usual rectangular subregions of the stereo image, that is, the rectangular windows.

2. Relaxation technique: Bernord (1980) suggests an algorithm where a set of candidate matching points are selected independently in each image. These points are the locations of small, distinct features which are likely to be detected in both images. A network of possible matches between the two sets of candidates is constructed. An initial estimate of the probability of each possible disparity is made, based on the similarity of subimages surrounding these points. These estimates are iteratively improved by a relaxation labeling technique making use of the local continuity property of the disparity, which is a consequence of the
continuity of the real world surfaces.

3. Structural constraints technique: Hwang (1982) suggests a method where both the geometric and structural information of the segmented features in the two images are used for three dimensional scene matching. The segmented features such as regions, edge segments and vertices are initially labelled using a symbol set. Then the structural relationships among these labels are tabulated in a relational table. The consistent labels between two related tables associated with two given images are searched using a relaxation labelling process. In this process the matching line equation between the two images is used as a constraint function to remove the ambiguous labels from the two relational tables. This process is applied iteratively until two isomorphic relational tables are deduced.

4. Prediction-verification technique: Ayache & Paverjon (1985) has proposed to solve the stereo matching problem by a method of prediction and verification of the hypotheses applied to an edge based description of an image. First, both images are reduced to segment adjacency graphs. Segments come from a polygonal approximation of edges in the images. Tentative matches are hypothesized between segments which intersect a common epipolar plane, whose disparity lies within the maximum potential disparity range. Then the hypothesis is propagated in a recursive manner into its neighborhood by assigning new matches between nearby segments which intersect a common epipolar plane, whose disparity is close to the disparity computed between the previously matched segments.
5. Edge-matching technique: Baker (1980) describes a stereo matching technique based on edge data in the images. The use of edge data fulfills the basic requirement of visual busyness, (i.e., measurement of the textural and line structure of the image) for reliable matching and reduction of the computational cost. The correspondence problem is attacked one horizontal line at a time, using edge correlation procedures for finding the best association of the first and second image edges.

The technique which is used in this work for the solution of the correspondence problem is based on a model of the human stereovision process and also in the work of Marr & Poggio (1976). The two most important reasons which lead to the adoption of this technique are:

First, the best visual system is the human system. So it is logical to believe that a technique which simulates the best visual system will have many advantages over all of the previously mentioned approaches.

Second, one of the biggest problems in any visual system as well as in any computational task is the reduction of the real-time that a process takes until its completion. The approach that is proposed in this work could be implemented in parallel, in order that the minimum possible time is achieved. More details about the solution of the correspondence problem are presented in Chapter IV.

There are a number of applications in computer vision and robotics in which one desires to know the distance from all points of an object but the only information available is the depth (and
possibly surface orientation) of a sparse set of points. This is
the visual surface interpolation problem which stands as the last
step of the complete stereo problem. Researchers have used a number
of different approximations and representations for the
reconstructed surface, some of which will be mentioned in this
chapter.

The formulation of the problem leads to finding the surface
from a class of surfaces, using the one that minimizes a functional
(norm). Even when a number of constraints are applied, the approach
is faced with an infinite number of [class, norm] pairs from which
to choose. In general to make the best choice, one might appeal to
physical analogies (e.g., the norm should measure physical bending
energy of an ideal thin plate), personal biases (e.g., the space and
the norm should be rotationally invariant and the mathematical
definition of the class should be simple) or other adhoc assumptions
(e.g., the [class, norm] should be such that the optimal surface
reconstructs low degree polynomials).

However most of these assumptions still leave infinitely many
[class, norm] pairs satisfying them.

A number of alternate classes and their associated norms are

1. A simple family of classes $D^{-m} L^2$ using an isotropic and
   physically meaningful $m$th Sobolev semi-norm defined over
   $R^2$ (Boult, 1984).

2. A family of functions (space of functions is $D^{-m} R^2$) using
   an isotropic, physically meaningful semi-norm $Y$ (*)
   defined over $R^2$ (Boult, 1986).

3. A class of functions using the second Sobolev semi-norm

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
where $D^i_x$ represents the differential operator $\partial^i(./) \partial x^i$ and $D^j_y$ the $\partial^j(./) \partial y^j$, over a flat disk (Boult, 1985).

In this work, the approach of Grimson (1981) is followed and extended for the interpolation problem. The problem can be solved by minimizing a suitably defined functional and finding the most consistent surface. The first step in the solution is the reduction of the number of surfaces which interpolate the sparse depth points which are obtained from the output of the correspondence problem. This is done by using additional constraints and by applying the surface consistency constraint theorem, which states that between known depth values the surface cannot change in a radical manner. A series of mathematical calculations on the variation of the surface orientation which measures the change of a surface leads to the definition of two most appropriate functionals. One of those is chosen based on the evaluation of many criteria, and its minimization defines the best visible surface.

Terzopoulos (1983) has extended the initial work of Grimson (1981) in a different way by generalizing the interpolating surface to be a mixture of membrane and thin plate. In this work it is shown that his extension does not coincide with the ideas presented in Chapter VIII and it is established that the mixture mentioned above performs very poorly.

However, there are two fundamental problems with the above
optimization approaches suggested by Grimson (1981) and Terzopoulos (1983). First, the schemes do not calculate and label discontinuities, for surface and its orientation explicitly. The above methods do not clearly locate them and it is suggested that a two pass algorithm is required, which of course is an inefficient method to solve this problem. There are two main reasons for explicit detection of discontinuities. One is that such a representation will allow higher level processes, such as recognition and extraction of axes for 3-D shape analysis, to operate more easily since the process serves to make implicit information explicit. A second reason is the creation of a more accurate surface representation. In this work it is shown that discontinuities can be included explicitly if the optimization formulation uses "loose constraints."

The second problem is that the representations suggested in the past are not invariant. So the reconstructed scheme (surface) will "wobble" as the viewpoint varies. This is because instead of using invariant quantities of surface area and surface curvature, the schemes use approximations to these quantities which are not invariant to changes of the viewpoint. This problem is also solved by the approach incorporated in this work, the method of which is explained in Chapter VIII.
CHAPTER III

METHOD

Introduction

In this chapter concepts will be addressed concerning the process through which the goal of a complete stereo vision system will be reached. This consists of the reconstruction of the visible surface of the object(s) in the scene by a pair of stereo cameras.

At the beginning of this chapter the domain of the objects is clearly defined for which the method (algorithm) presented in this work, demonstrates satisfactory results.

Then the constraints as well as the necessary assumptions are listed which are essential for the correctness and efficiency of the proposed algorithm.

Also a general pictorial representation is given for the three dimensional acquisition system which will be used in this problem for the complete reconstruction of the whole surface of the object(s) in the scene to be examined.

At the end of this chapter the general steps of the particular algorithm are presented. A general overview is outlined for each step and the corresponding chapter is given where the technical details about each step are explained.
Constraints and Acquisition Device

Before the general algorithm for solving the stereo problem for specific cases of shapes of objects is outlined, it is essential to explicitly define the domain of objects and the nature of the scenes to be examined. This domain consists of polyhedral objects such as cubes, parallelepipeds, pyramids, prisms, or a pile of them. Also compound objects like a staircase (i.e., stairs "glued" together) do not affect negatively the performance of the algorithm. It must be mentioned here that the proposed method will correctly and effectively work even if the objects belonging to the restricted domain do not have planar faces as a pure geometrical definition of them would require. Their faces can have all kinds of textual irregularities (cracks, scratches, etc.) as well as any type of distribution of peaks and valleys present on the surface of each face. Actually the presence of these irregularities which naturally cause discontinuities in the depth and in the orientation of every visible face, cause the algorithm to demonstrate its power in the reconstruction of the visible surface by preserving the discontinuities.

In order for the algorithm to work properly, the following three constraints must be applied because of the limited usage of the method and of the geometry of the objects under consideration.

1. The object is assumed to be on a table that rotates around an axis perpendicular to its base.
2. The stereo pair of cameras must always be oriented in such a way so that only one face is visible during the execution of the main loop of the algorithm which reconstructs a side of the object. Points which belong to successive sides or faces of the object must have different distances from the midpoint of the baseline of the two cameras. This difference is a direct consequence of the physical geometry of the objects under consideration.

3. The object(s) is supposed not to be extremely white without any texture irregularity. It is clear that this constraint is satisfied in almost all cases.

Figure 4 illustrates the setup of the three dimensional acquisition system which implements the algorithm outlined in the next section. The system consists of a pair of stereo cameras guided by the control processor which also handles all the processing of images taken by the cameras. It is also connected to the turntable on which the object to be examined is positioned. This is so that the angle through which the table turns can be measured at each pass through the main loop of the algorithm. Note that the two cameras are oriented in such a way so that only one face of the object on the turntable is visible by both cameras at the same time.
Figure 4. Three Dimensional Acquisition System.
Algorithm

The successive steps for solving the particular problem described above are presented in algorithmic terms:

1. Adjust the cameras so that both aim at the front face (visible surface of the object from a specific point) of the object.

2. Convolve the two images using the Laplacian of a Gaussian operator \( (V^2G) \), extracting the zero-crossings. Using an iterative method to refine the resolution and to increase the accuracy, match the corresponding zero-crossings from the two images. Measure the disparity. These calculations are presented in Chapter IV.

3. Solve the calibration problem for the stereo pair of cameras by calculating geometrical parameters and constants for the stereoscopic device.

   These calculations are presented in Chapter V.

4. Using image coordinates, world coordinates and disparity information, calculate depth information from the baseline of the two cameras for every point in the scene which corresponds to a matched zero-crossing contour.

   Minimize the error caused by the fact that the two lines of sight do not intersect at a point in three dimensional space. This lack of intersection is caused by noise in images and accuracy errors in previous calculations. Find the best approximation.

   These calculations are presented in Chapter VI.

5. Store in memory the above calculated distances of the "specific" points (matched zero-crossing contours) as depth
information of the "old" face of the object. Use an array structure.

6. Calculate the orientation of the small regions on the surface in the scene which correspond to the matched zero-crossings. These are presented in Chapter VII.

7. Reconstruct the visible surface by interpolating the measured object ("specific") points. Look for the best solution which can be found by minimizing the energy functional of a thin plate. Use "loose constraints" and other parameters in order for the reconstructed surface to include discontinuities and to be viewpoint invariant. These calculations are presented in Chapter VIII.

8. Initialize a counter which measures the number of rotations.

9. Rotate the object by an appropriate angle (in case of cube or parallelepiped the angle is 90°) in order for another face to be visible. Find the matched zero-crossings as in Step 2.

10. Loop starts. Calculate the real distances from the baseline of the cameras to the specific points of the "new" visible face as in Step 4 (above), also taking into account the inverse rotation transformation of the coordinate system. This calculation is included in Chapter VI.

11. Compare these new distances (real distances of the "new" face) with those that have been stored in the array structure and which represent the "old" face.

Reject the points ("new" points) for which an equality holds
(or an approximate equality with some tolerance). This is because points on the "new" face cannot have the same distance with points on the "old" face.

Replace the distances and points of the "old" face by those of the "new" face in the array structure after the "rejection" step.

12. Repeat Step 6 but at the end of the calculations add the quantity "counter x appropriate angle" (i.e., 90° for the case of a cube) due to the rotation. Increment counter by 1.

13. Repeat step 7.

14. Loop ends. Repeat Steps 9 through 12 until you reach a rotation angle of 360°.

On close inspection of the steps of the above algorithm, it can be seen that the method reconstructs the sideways surface of an object which does not include its base and/or its top face. This is because they are never visible by the two cameras at the same time during the predefined rotation of the table.

It is essential to highlight the most important steps of the algorithm and emphasize its strong points as well as to mention its limitations.

One of the most important steps of the method is step 2. This is the step that is implemented in parallel fashion and the primal sketch for each image of the stereoscopic device is developed. The correspondence problem is solved by matching zero-crossings (which are the places that the second derivative of image intensity takes a zero value) using a multiresolution technique for more accurate measurements.
In step 4, the three dimensional distance of specific points on the surface of the object(s) in the scene from the baseline of the cameras is calculated using disparity measurements that have been developed in step 2. The accuracy of the depth measurements is directly related to the accuracy of the disparity measurement.

Step 7 could be characterized as the most important step of the algorithm. A brief description is given in the algorithm about the reconstructed surface which interpolates the sparse set of depth values preserving discontinuities and being also viewpoint invariant. This is actually the step that demonstrates the power of the algorithm to reconstruct a visible surface regardless of its shape.

The only limitation of the algorithm which directly influences the general usage of the method is the lack of determination of the angle, mentioned in step 9, for an object outside of the domain examined in this method. This is the angle by which the table has to turn in order for a complete new face of the object to be visible by the stereo cameras. This limitation emerges from the incapability of clearly and geometrically being able to define the term "side" or "face" for an object which does not have some type of regularity in its shape.
CHAPTER IV

CORRESPONDENCE PROBLEM

The initial part of the stereo process is faced with two major problems: (1) the extraction of a description of the elements of an image, corresponding to physically identifiable locations in the scene, and (2) the determination of the corresponding descriptions from each processed image. The first problem is solved by convolving each image with a set of filters of the form $V^2G$. For each size filter the zero-crossings of the convolved image are localized. The second is solved using the matching process which matches corresponding zero-crossing segments and measures the disparities using a coarse-to-fine resolution technique. The process of the derivation to the solution of these problems will be analyzed in this chapter.

General Overview

If two objects are separated in depth from the viewing system, then the relative positions of their images will differ in the two cameras. For example, one could hold his thumb at arms length, and view it first with the left eye closed, then with the right eye closed. He will notice that the background shifts relative to the position of the thumb in the two images. The process of the stereo
vision in essence, measures the difference in the relative positions, called disparity, and uses it to compute depth information for visible surfaces in the scene.

The relationship between distance and disparity suggests that one could determine the distance from the objects in scene by measuring their disparity. But how can the disparity be measured? The following three steps introduce a general solution:

1. A particular location on a surface in the scene is selected from one image.
2. The same location is found in other image.
3. The disparity between the two corresponding image points is measured.

The heart of the problem lies in identifying elements in each image which correspond to the same element in the scene. If one could identify a location beyond doubt in the two images, for example by illuminating it with a spot of light, the first two steps could be avoided and the problem would be easier. Since this method is not practical for all cases and since the human visual system solves the problem without using this technique, another more realistic method must be found.

So, the problem of identification of corresponding image points cannot be avoided. There are two parts to the correspondence problem: determining what elements from an image are to be matched; and given an element from one image evaluating candidates from the other image according to some matching criteria.

To solve the correspondence problem, one has to decide on the
nature of the elements to be matched. At one end of the spectrum object recognition could be performed before the stereo matching. From each monocular view various objects are identified. Since in most cases the probability of having two or more identical objects is small, but not negligible, the matching problem becomes simple. To determine the disparity, a particular object would be located in each of the two images and the disparity would be computed. Since experiments with random-dot stereograms, which do not contain any objects, have shown that human stereopsis could solve the matching problem and since the approach in this work tries to simulate it, the above extreme case will not be applicable.

On the other end of the spectrum, if random-dot stereograms are considered, the system will have great difficulty to match the dots based on their colors. In this case the extraction of features to be matched is very easy, but the matching itself very complicated.

A solution in the middle of the spectrum must be applied, taking into account some special features which must be extracted from the pair of images.

False Target Problem

In this section the false target problem which must be solved by the matching process is presented. Physical constraints are listed which will minimize the complexity of the problem. Before they can be used by the matching process, they must be mapped onto matching constraints in order to restrict the allowable ways of a given match.
The reason that the task of identifying corresponding locations in the two images is complicated, is because of the false target problem (Marr, 1976). So, for every feature of the left image that is to be matched, there are many candidates from the right image competing for a correct match. The problem now can be stated as: Which of the candidates must be chosen for a correct unambiguous match?

To answer this question, additional information is needed to decide which matches are correct, by constraining them in some way. To do that, one has to examine the basis in the physical world for making a correspondence between the two images.

The constraints needed are deceptively simple: (a) a given point on a physical surface has a unique position in space at a given time, and (b) matter is cohesive, it is separated into objects and the surfaces of the objects are generally smooth in the sense that the surface variation due to the roughness and cracks or other sharp differences that can attribute to changes in distance from the viewer, are small compared to the overall distance from the viewer.

The above observations are properties of the physical surfaces and they constrain the behavior of the surface position. In order to use these observations for establishing a correspondence between two images of a surface, one must ensure that the items to which applies them, are in one-to-one correspondence with well-defined locations on the physical surface. To do that, image predicates must be used, that correspond to surface markings, shadows,
 discontinuities in surface orientation and so forth.

The physical constraints must be rewritten as matching constraints, which restrict the allowable ways of matching two primitive symbolic descriptions one from each image. More details about this propagation of constraints will be presented later in this chapter.

Detect and Represent Intensity Changes

Initially in this section the exact types of features are presented which will serve as the input to the matching module. It is established that these features, zero-crossings, form a rich description of the image from which they are extracted. They contain enough information to enable the reconstruction of the visible surface to which they correspond. It is also shown that the usage of a non-directional operator for the convolution of the image and the extraction of the zero-crossings, results in big computational time savings and performs equally well as a directional second order operator. At the end of this section a detailed mathematical explanation is given about the optimum type of filter that is to be used to convolve the raw images taken by the stereo pair of cameras.

In most cases scratches, texture, markings or boundaries between surfaces will give rise to sharp changes in the irradiances of the image. As a consequence, there should be a direct correspondence between changes in irradiances and specific surface locations. If the irradiance changes could be reliably located, the
two physical constraints mentioned in the previous section could be
applied. It can be seen from Figure 5 that sharp changes in
irradiance give rise to an extremum in the first derivative and a
zero-crossing in the second derivative. This is true for one
dimensional arrays of irradiance. Figure 5(a) illustrates an one
dimensional step change in irradiance. The change is associated
with an extremum in the first derivative Figure 5(b) and a zero-
crossing in the second derivative in Figure 5(c). For two
dimensional arrays of irradiance such as an image, the same
observations hold true. A sharp change in the irradiances give rise
to an extremum in the directional derivative and zero-crossing in
the second directional derivative.
Figure 5. One Dimensional Detection of Image Changes.

It is known that the zero-crossings form an extremely rich description of the image and under certain conditions contain sufficient information to enable a reconstruction of the original signal. This argument relies on extensions of Logan's theorem (Logan, 1977) which essentially states that under certain conditions all the information needed to reconstruct a signal is contained in the zero-crossings of that signal. Also, as mentioned before, under general viewing conditions, inflections on the surface are preserved by zero-crossings. These two facts together imply that the
information needed to reconstruct the surface shape is encoded in the depth values along the zero-crossing contours of the primal sketch. So zero-crossings which are special features for the stereo images will be used to solve the correspondence problem.

As mentioned earlier, for the two dimensional arrays, a sharp change in the irradiance of the image will create a zero-crossing in the second directional derivative. To choose in which direction to take the derivative, one has to observe that the underlying motivation for detecting changes in intensity is that they will correspond to useful properties of the physical world, like changes in reflectance, illumination, surface orientation or distance from the viewer. Such properties are spatially continuous and can almost everywhere, be associated with a direction that projects to an orientation in the image. So the orientation of the directional derivative that is chosen to be used is that which coincides with the orientation formed locally by its zero-crossings. A question that has to be answered here is: under what conditions does this direction coincides with that in which the zero-crossing has the maximum slope? The answer is given by the condition of linear variation, Marr & Hildreth (1980), which is as follows:

The intensity variation near and parallel to the line of zero-crossings should locally be linear. This condition will be approximately true in smoothed images, which are the kind of images examined in this work.

Although it is possible to design an algorithm to determine the appropriate directional derivative, applying it to the image and
locating the zero-crossings in practice would require by the algorithm to convolve the images with a set of operators one for each direction. Convolutions are computationally expensive and it would be useful if their number could be reduced, by using a non-directional operator.

The only orientation-independent, second order differential operator is the Laplacian $\nabla^2$ and Theorem 2, Marr & Hildreth (1980) makes explicit the conditions under which it could be used. They are weaker than the condition of the linear variation, stating that, if the intensity variation is linear along but not necessarily near to a line of zero-crossings, then the zero-crossing contours will be detected and located from the zero values of the Laplacian. The only difference, using the Laplacian instead of a directional derivative, lies in the position and not in the existence of such contours. For example, at a corner, both the Laplacian and the directional derivative will give rise to zero-crossings, the only difference being in the exact location, which has a negligible consequence in the reconstruction of the three dimensional surface of any object.

One could apply the Laplacian to the images as they are but for practical reasons, reduction of noise, quantization of irradiance, it is better to smooth the images slightly by convolving them with a special function. So the key point here is to find the formula of that function, or in other words, the type of the filter to be applied on the images.

A major difficulty with natural images is that changes occur
over a wide range of scales. No single filter could be optimal simultaneously at all scales, so one has to find a way of dealing separately with the irradiance changes at different scales. This requirement leads to the basic idea in which one has to take local averages of the image at various resolutions and then detect the changes in intensity that occur at each one.

There are two physical considerations, the combination of which determines the appropriate smoothing filter. First, the motivation for filtering the image is to reduce the range of scales over which intensity changes takes place. The filter's spectrum should therefore be smooth and roughly band-limited in the frequency domain. This condition could be expressed by requiring that the variance should be small.

The second consideration is best expressed as a constraint in the spatial domain and is known as the constraint of spatial localization. The things in the world that give rise to intensity changes in an image are: (a) illumination changes, which include shadows, visible light sources and illumination gradients; (b) changes in the orientation or distances from the viewer to the visible surfaces; and (c) changes in surface reflectance. The critical observation here is that, at their own scale, all the above changes can be thought of as spatially localized. The consequence of the second constraint is that the contributions to each point in the filtered image should arise from a smooth average of nearby points. Hence the filters should also be smoothed and localized in the spatial domain and in particular, the spatial variance \( \Delta x \) should
also be small.

Unfortunately, the above localization requirements, the one in the spatial and the other in the frequency domain, are conflicting. They are in fact related by the uncertainty principle which states that $\Delta x \cdot \Delta \omega \geq 1/4\pi$

There is only one distribution that optimizes the above relation, namely the Gaussian distribution function. In two dimensions:

$$G(r) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r^2 = x^2 + y^2$$

and $\sigma$ is defined as the variance.

So the filter $G$ provides the optimal trade-off between conflicting requirements.

Wherever an intensity change occurs, there will be a corresponding zero-crossing for the Laplacian. In other words one has to find the zero-crossings of the following function:

$$f(x,y,\sigma) = \nabla^2 [G(x,y,\sigma) \ast I(x,y)],$$

where $I(x,y)$ is the image and $\ast$ is the convolution operator.

According to the derivative rule of convolutions:

$$f(x,y,\sigma) = \nabla^2 G(x,y,\sigma) \ast I(x,y)$$

So zero-crossings can be detected economically in the image at each given scale, by searching for the zero values of the convolution $G \ast I$. In two dimensions:

$$G(r) = -\frac{1}{\pi \sigma^4} \left[1 - \frac{r^2}{2\sigma^2}\right] \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

The last equation expresses a rotationally symmetric function with one free parameter $\sigma$, which determines the spatial size of the function. Its cross section is shown in Figure 6 where the central
negative region (width) is denoted by \( w_{2D} \) and given by: \( w_{2D} = 2\sqrt{2} \sigma \).

The size of the filter may be specified by the value of \( \sigma \) or equivalently by the value of \( w_{2D} \).

![Figure 6. Cross Section Of Rotationally-Symmetric Operator.](image)

The \( \nabla^2 G \) looks like a Mexican hat operator and it clearly resembles the difference of two Gaussians (DOG) and has been proven by Wilson (1977) (see Appendix A) that in fact \( \nabla^2 G \) is the limit of the DOG function as the sizes of the two Gaussians tend to one another.

As a conclusion of the above analysis, it is essential to mention again that the changes in irradiance can be determined by locating the zero-crossings of the output of the convolution of the
image with a Laplacian of a Gaussian operator. Equivalently, one has to find points $(x_0, y_0)$ such that $f(x_0, y_0, \sigma) = 0$.

Solution of the Problem

In this section the details of the matching process are explained. The matching constraints which have been deduced from the two physical constraints presented in the second section of this chapter are listed and their influence on the operation of the matching module is presented. At the end of this section, an explanation of the multiresolution technique is given which is used for the accurate matching of the corresponding zero-crossing contours and the measurement of the disparity values.

It has been indicated, that the basic symbolic descriptors to be matched by the stereo process must be on the level of zero-crossings of a $V^2G$ operator applied to the image. The problem to be addressed is how the matching between two primitive symbolic descriptions can be done. The two physical constraints mentioned in an earlier section of this chapter must be translated into computational rules that will govern how the right and left descriptions are to be matched. In making this translation it is necessary to ensure that the items to which the rules will apply, in the image, are in one-to-one correspondence with well-defined locations on a physical surface. The computational analogues of the two physical constraints, Marr & Poggio (1976) are:

1. Uniqueness. Except for the rare cases, each item from either image will be assigned at most one disparity value. This
statement relies on the assumption that an item corresponds to something that has a unique physical position. The exceptions can occur when two features lie along the line of sight (line defined by the actual object and the center of lens of the imaging device) from one camera but are separately visible from the other.

2. Continuity. Disparity will vary smoothly everywhere. This condition, which is a consequence of the cohesiveness of matter, states that only a small fraction of the area of an image is composed of boundaries that are discontinuous in depth and in disparity values also.

The basic problem to be overcome in binocular fusion is the elimination or the avoidance of false targets, whose difficulty is determined by two factors: (1) the abundance of matchable features in an image and (2) the disparity range over the matches are sought. If a feature occurs only rarely in an image the search for a match can cover quite a large disparity range before false targets are encountered but if the feature is common false targets can occur within quite small disparities.

For a given disparity range, the matching problem can be simplified by first decreasing the incidence of matchable feature pairs, that is, by making features rare. There are two ways to do that. One; is to make them quite complex or specific, so that even if their density in the image is high there would be so many different kinds that there would seldom be a compatible pair. The second way, is to reduce drastically the density of all features in the image by decreasing the spatial resolution at which it is
The second way is more promising. Indeed the existence of independent spatial-frequency-tuned-channels in binocular human stereo fusion suggests that several copies of the image, obtained by successively fine filtering, can be used during the fusion providing increasing and at the limit, very fine disparity resolution at the cost of decreasing disparity range. The most precise disparity values are obtained from the small size filters and therefore, it is essential that each part of the scene can ultimately be brought into the small disparity range, within which small size filters can operate.

The above observations suggest the following scheme for solving the fusion problem: (a) Each image is analyzed through filters of varying coarseness and matching takes place between corresponding images from the two cameras for disparity values of the order of the filter size. (b) Matching, using coarse resolution, obtains crude estimation about disparity which helps in its more fine measurement in fine resolution.

The idea of matching widely separated features first and then with the information so obtained, repeating the matching process at successively finer scales of resolution, appears promising. The question though is what features of the zero-crossings should be matched at different resolutions?

Before the criteria for matching is presented, it is essential that the following definitions must be mentioned:

Sign of a zero-crossing. This is computed by noting whether
the convolution values change from positive to negative or negative to positive, while scanning horizontally from left to right across a zero-crossing.

Orientation of zero-crossing on an image plane. This is found by computing the gradient of the convolution values across the zero-crossing and taking the orientation of the projection of the gradient onto the image as the orientation of the zero-crossing.

The following criteria are used for matching zero-crossings in the left and right images for each filter:

1. The zero-crossings must arise from convolutions with the same filter size.
2. The zero-crossings must have the same sign.
3. The zero-crossing segments must have roughly the same orientation.

The important point here is that zero-crossings cannot, on the average, occur too close together and this is true for any band-pass filter. It has been established in previous section that \( \nabla^2 G \) is a band-pass filter and so, its zero-crossings are usually separated by some minimum distance.

The matching process will be better understood by looking to the following series of figures. Figure 7 shows a zero-crossing from the left image marked as \( L \), which matches another of the same sign in the right image, which is displayed by a disparity range \( d \). The correct match is labeled \( R \), and a possible false target \( F \), shown dotted, appears nearby. Provided that only the disparity range \( w/2 \) is considered, even if \( R \) is right at one end of the range, which is
\[ d = w/2, \text{ statistical analysis assures that with probability of 95\%,} \]

it will be the only zero-crossing of this type within a disparity range that extends over \( w \). This assumes that \( R \) is the correct match, which means that the correct match lies in the range of \( w/2 \) that the procedure examines.

![Figure 7. Match Over \( w/2 \) Range.](image)

In general however, there will be an ambiguity and mismatch because if the width \( w \) is small enough to give good resolution there will be too much detail to match well.

Figure 8 gives the zero-crossings for filters of width \( w \) and \( w/2 \). Pair A matches with a disparity of 11; pair B matches with a disparity of 9; and pair C matches with a disparity of 4.
Figure 8. Matching With Shifting.

Assume now, that a narrow filter with width w/2 produces the zero-crossings shown in pair B of Figure 8. Since the width is different, the zero-crossings are in slightly different places. Note that the dotted line of pair B is equally close to two solid lines, which creates difficulty.

Now comes the key step. Before matching the zero-crossings using width w/2, the matching procedure will shift the zero-crossing represented by the solid line, as much as the disparity value obtained at width w. This provides the zero-crossings at the bottom of Figure 8. The initial disparity estimate introduced by shifting, places the zero-crossings close together making the matching possible according to the criteria described earlier in this section. So if $\delta_{\text{init}}$ is the initial disparity at the lower resolution associated with filter width w, and if $\delta_{\text{new}}$ is the
disparity within the reduced matching tolerance w/2, which can be obtained after a shifting of the initial zero-crossing by $\delta_{\text{init}}$, then the accurate disparity is given by $\delta = \delta_{\text{new}} \cdot \delta_{\text{init}}$. Note that $\delta_{\text{new}}$ could be negative.

The key idea of the above method was to use results from intentionally blurred images to guide work on less blurred ones. Sometimes this is called multiscale image analysis. Thus, by combining the processing of coarse and fine filtered images, fine resolution disparity can be detected over a large disparity range while minimizing the false targets.

The matching method described above is valid only in the cases that an unambiguous match takes place. But this does not happen always and so ambiguous matches could arise from the matching process.

Consider Figure 9 which gives an initial idea to these special cases.

Figure 9. Ambiguous Matches.
The matching process can be carried out from either of the two images of the two cameras. In Figure 9 for example, if matching is indicated from the left image, the match for $L_1$ is ambiguous but for $L_3$ is unique. From the right image, matching is unique for $R_1$ but ambiguous for $R_3$. Together the two unique matches provide the correct solution. The reason that the two unique matches should be correct rather than contradictory is a consequence of the uniqueness property embedded in one of the fundamental assumptions of stereopsis. As a result the matching process can be designed to accept unambiguous matches by starting from either image. However, this design does have fascinating consequences because it means that the uniqueness assumption is no longer internally verifiable by the process (algorithm) whereas the continuity assumption is.

To speed up the process of finding the solution to the correspondence problem, matching can be carried out from both images of the two cameras. Situations like that introduce ambiguity from both images. In this cases the ambiguity can be resolved by consulting the signs of the neighboring matches and choosing the match with the same sign. There is however, an important distinction between the two obvious ways to do this. Either the signs of the neighboring matches are considered that were unambiguous from the start or the signs of the neighboring matches that have not so far been assigned. This way of disambiguating a match is called the pulling effect.

The final output of the matching process is a disparity map with disparities assigned along most portions of the zero-crossing
contours obtained from the smallest filter used. The accuracy of the disparities obtained depends on how accurately the zero-crossings have been localized, which may be to a resolution much finer than the initial array of intensity values that constitute the image. It is worth noting here that horizontally oriented segments of the zero-crossing contours may be ignored since they do not have a well defined disparity. The last statement implies that horizontally oriented line segments in the image will not be assigned depth values. The last observation is still an open problem in two camera stereo vision and will be discussed more in the last chapter.
CHAPTER V

CALIBRATION PROBLEM

The problem of calibrating a stereo system is extremely important in practical applications where accurate depth measurements must be made. In this chapter an approach for developing a solution is described. At first the camera model is briefly presented. Then it is shown how this model can be effectively used to compute the geometrical parameters in the perspective transformation.

To achieve that, it is assumed that the 3-D coordinates of a number of reference points is given, and an efficient optimization technique is derived, based on the use of physical constraints, to find the parameters from the entries of the perspective transformation matrix.

Camera Model

In this section the camera model is presented and some rotation is introduced. It is clear that in order to solve the calibration problem to its full extent, both cameras of the stereo imaging system have to be calibrated. But since the solution is almost identical for each camera, only one will be calibrated by calculating the necessary parameters.
Suppose that a camera is represented by a pinhole lens together with an image plane lying a distance $f$, in front of the lens. This is done in order to avoid the minor annoyance that images are inverted left to right and top to bottom. The transformation that is discussed is called the perspective transformation and the general case is considered where $f$ is not equal to the focal length $F$ of the lens.

Figure 10 illustrates an overview of the different coordinate systems used and of the whole geometry.
The (xyz) is the world coordinate system. The camera lens are translated by a vector \( v_a = (x_1, y_1, z_1) \), panned through an angle \( \theta \) and tilted through \( \phi \).

Suppose \( \dot{v}_{iu} \) are the homogeneous coordinates (Haralik, 1980) of the image point \( v_{iu} \) with respect to the image's coordinate system. Let \( \hat{v} \) represent the homogeneous coordinates of the object points \( v \), to the world system. Let \([P]\) be the perspective transformation matrix.

The transformation of the coordinates in order to find the coordinates of the image points \( v_{iu} \), given the coordinates of the object point \( v \), is a mapping between the global coordinate system and the image coordinate system. This mapping is given by the product of the matrices: \([G]\), \([R]\), \([T]\) where,

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\cos \theta & \sin \theta & 0 & 0 \\
-\cos \phi \sin \theta & \cos \phi \cos \theta & \sin \phi & 0 \\
\sin \phi \sin \theta & -\sin \phi \cos \theta & \cos \phi & 0 \\
0 & 0 & 1 & 0 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & 0 & 0 & -x_1 \\
0 & 1 & 0 & -y_1 \\
0 & 0 & 1 & -z_1 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

The matrices \([G]\) and \([T]\) are translation matrices and \([R]\) is a rotation matrix. So the complete mapping will be given by:

\[
\dot{v}_{iu} = [P] \ [G] \ [R] \ [T] \ \hat{v}
\]

where \([P]\) is the perspective matrix. From the last equation it can be derived that the object point lies along the vector

\[
\dot{v} = [T]^{-1} \cdot [R]^{-1} \cdot [G]^{-1} \cdot [P]^{-1} \cdot v_{iu}.
\]

This equation denoting the inverse perspective transformation will be derived in the next chapter.
To calculate the distance of an object from the stereo device given the coordinates of each image (which are calculated from the solution of the correspondence problem), the entries of the matrix $[K]=[T]^{-1}•[R]^{-1}•[G]^{-1}•[P]^{-1}$ must be determined, or in an equivalent way the entries of the matrix $[H]=[P]•[G]•[R]•[P]$ have to be found.

Solution

Suppose $\{V_i\}$, $i=1,...,n$ is the set of $n$ three dimensional points and suppose that the corresponding image points are $\{V_{iu}\}$, $i=1,...,n$. The direct perspective transformation gives the picture coordinates of an image point in terms of an object point and also the geometrical parameters of the transformation. The functional form of this transformation is given by $V_{iu}=f(V,P)$ where $f$ is the vector-valued function giving the computed picture coordinates of the image of $v$, and $P$ is a vector composed of the geometrical parameters specifying the translation and rotation of the camera. In the following discussion it is assumed that the second component of both $V_{iu}$ and its homogeneous representation $\hat{V}_{iu}$, have been detected, since only two coordinates $u_1$ and $u_3$ are enough to determine an image point.

The vector $f$ is sufficiently complicated to make it difficult to solve directly for the components of $P$, given a set of corresponding object and image points. Also it may be the case that $n$ is large enough to result in $P$ being over-determined. Keeping this in mind, the solution of the calibration problem is found by solving an equivalent optimization.
Suppose that the exact positions of \( n \) object points as well as the positions of the corresponding image points are known precisely. The homogeneous form of the general perspective transformation will be: \( \hat{v}_i = P \cdot G \cdot R \cdot T \cdot v \).

A 4\( \times \)n matrix \([V]\), is defined to have as its columns the \( n \) vectors \( \hat{v}_i, i=1,\ldots,n \) and a 3\( \times \)n matrix \([V']\) to have as its columns \( n \) vectors \( v_i \). In other words, \([V]\) is composed of the set of \( n \) object points and \([V']\) of the set of \( n \) observed points. The 3\( \times \)4 matrix \([H]\) = \([P]\) \([G]\) \([R]\) \([T]\) is defined with the second row being deleted as mentioned above. Now the calibration problem is mapped into the problem of finding the matrix \([H]\) such that \([H]\) \([V]\) = \([V']\) or

\[
\begin{bmatrix}
H_{11} & H_{12} & H_{13} & H_{14} \\
H_{21} & H_{22} & H_{23} & H_{24} \\
H_{31} & H_{32} & H_{33} & H_{34}
\end{bmatrix}
\begin{bmatrix}
\hat{v}_{11} & \hat{v}_{12} & \cdots & \hat{v}_{1j} \\
\hat{v}_{21} & \hat{v}_{22} & \cdots & \hat{v}_{2j} \\
\hat{v}_{31} & \hat{v}_{32} & \cdots & \hat{v}_{3j}
\end{bmatrix}
= 
\begin{bmatrix}
\hat{v}_{1411} & \hat{v}_{1412} & \cdots & \hat{v}_{141j} \\
\hat{v}_{1421} & \hat{v}_{1422} & \cdots & \hat{v}_{142j} \\
\hat{v}_{1431} & \hat{v}_{1432} & \cdots & \hat{v}_{143j}
\end{bmatrix}
\]

Unfortunately the solution of the above matrix equation is quite complicated.

First, although \([H]\) has twelve components there are only five independent parameters; namely, the three translation and the two rotation parameters of the camera. Each of the twelve components of \([H]\) is a function of three parameters. So, in principle, one must not choose any arbitrary \([H]\).

The second complication results from the fact that it is impossible to simultaneously specify the scale of both an object point (in homogeneous coordinates) and its image point (in
homogeneous coordinates). In particular, if \( v = (x, y, z)^t \) is an object point and \( v_{1,2} = (v_{1,1}, v_{1,2}) \) an image point, one cannot freely write their homogeneous representations as \( v = (wx, wy, wz, w)^t \) and \( v = (wv_{1,1}, wv_{1,2}, w)^t \) since this would force the perspective transformation \([H]\) to produce the same number, \( w \), for the last homogeneous coordinate of both the object point and each corresponding image point.

To avoid the last pitfall, a diagonal matrix \([D]\) is introduced, which allows one to scale each of the homogeneous image vectors by an arbitrary factor. So the original problem can now be stated as follows. Find the 3x4 matrix \([H]\) such that \([H][V] = [V'][D]\). Also the nxn matrix \([D]\) has to be found.

There are \( n + 12 \) unknowns in this matrix equation, \( n \) contributed by \([D]\) and 12 by \([H]\) and also 3n equations. Therefore, a solution can be found for \( n \) having a numeric value of six since \( n + 12 = 3n \iff n = 6 \).

The mean squares method is the most appropriate for finding a solution to the calibration problem in order to find matrices \([H]\) and \([D]\) such that the sum of the square differences between the entries of \([H][V]\) and \([V'][D]\) is minimum.

The square error sum is given by

\[
E = \Sigma_{i=1}^{n} \left[ \Sigma_{j=1}^{3} \left[ \Sigma_{k=1}^{4} H_{ijk} V_{ki} - \Sigma_{s=1}^{3} V'_{s} D_{s,i} \right] \right]
\]

To minimize the error with respect to matrix \([H]\) the following derivative is set to zero:
\[
\frac{\partial \Sigma}{\partial H_{jk}} = \sum_{i=1}^{n} \left( \sum_{k=1}^{4} H_{jk} V_{x_i} V_{m_i} - \sum_{l=1}^{4} V_{i} D_{jk} V_{m_i} \right) = 0
\]

for \( m = 1,2,3,4 \) and \( j=1,2,3. \)

It can be shown that the above equation is equivalent to:

\[
[H] [V] [V]^t = [V'] [D] [V]^t
\]

from which it follows that:

\[
[H] = [V'] [D] [V]^t ([V] [V]^t)^{-1}
\]

Once \([H]\) has been determined from the above equation where \([D]\), \([V]\), \([V']\) are given or have been already calculated, the values of the geometrical parameters can be found by equating the corresponding components of \([H]\) with the components of the product \([P] [G] [R] [T]\). For example, by direct multiplication, it can be found easily, that the \((1,1)\) component of the product \([P] [G] [R] [T]\) is \(\cos \theta\) and so the value of the camera pan angle can be calculated.

The last parameter which must be measured is the distance \(f\) that the image plane is assumed to be in front of the lens. Many times this distance is confused with the focal length \(F\) of the lens. Its actual value can be determined from elementary camera geometry as illustrated by figure 11.
Suppose that $F$ is known, as well as the size of the image $I$ and the size of the negative, which the camera creates from an object. Suppose that $D$ is the distance behind the lens where the film is located. The focal length equation is given by:

$$\frac{1}{D} + \frac{1}{f} = \frac{1}{F}$$

Let $M$ be the magnification such that $M = \frac{I}{N} = \frac{f}{D}$. By using the last equation, the focal length equation gives:

$$\frac{M}{f} + \frac{1}{f} = \frac{1}{F} \quad \Rightarrow \quad f = (M+1)F$$

Therefore, the magnification of the image and the focal length of the camera lens can determine the distance $f$, in front of the lens, where the image plane hypothetically is located.
CHAPTER VI

3-D DISTANCE

With only few exceptions, work in image understanding has not dealt with the specific problem of distance determination. The matching problem has been considered the hardest as well as one of the most significant problems in computational stereo. Once accurate matches have been found, the determination of the distance will be a matter of triangulation. Nevertheless this step presents difficulties, especially if the matches are somewhat inaccurate or unreliable.

In this chapter we present a method for obtaining an accurate calculation of the distance of an object from the baseline (line segment connecting the center of the lenses) of the two cameras.

The model of the imaging device as well as the camera geometry will initially be presented. After a series of geometrical calculations the distance of the object from the stereoscopic device is obtained.

To a first approximation, the error in the stereo distance measurements is directly proportional to the positional error of the matches and inversely proportional to the length of the stereo baseline. Lengthening the stereo baseline complicates the matching problem by increasing both the range of disparity (the area that
must be searched) and the difference in appearance of the features being matched. The coarse-to-fine resolution strategy is considered to be the best to overcome this problem.

The goal is to minimize the error involved in distance calculations due to the inaccurate matches of zero crossings which results in the two lines of sight not intersecting in the three dimensional space. An error analysis of the depth measurements is presented towards the end of this chapter.

Camera Geometry

A perspective transformation is considered to be a first order approximation to the process of taking a picture. The cameras in this analysis, are represented by pinhole lenses, together with image planes lying at a distance of \( f \) in front of the lenses in order to avoid the annoyance that images are inverted left to right and top to bottom. The perspective transformation is not perfect and so the distance \( f \) is not equal to the focal length \( F \) of the lens of a camera but \( f = (M + 1) F \), as mentioned in the previous chapter.

It is true that a single reference frame, the image frame, is very convenient for locating picture points because it is centered at the center of the image plane. Using the same argument one frame is inconvenient for locating object points, since it constrains the measurements of distances to a set of axes whose position is determined by the camera. So, to produce more realistic measurements two coordinate systems are needed: a picture coordinate system, in which picture points are located and a global or world
coordinate system in which everything else concerning the camera and the object are measured.

Figure 12 illustrates a version of the cameras' geometry with the two image planes and the two global systems, one for each camera.

![Diagram of perspective transformation with two reference frames]

The global reference systems are denoted with the unprimed coordinates \((x_1, y_1, z_1)\) and \((x_2, y_2, z_2)\) and are used to locate the cameras and the object point \(v\), whose distance must be calculated. The cameras are both translated from the origin, by vectors \(v_{c1}\) and...
v_1 panned through angles \( \theta_1 \) and \( \theta_2 \) respectively, and tilted through angles \( \phi_1 \) and \( \phi_2 \) respectively. The image points \( v_{x1} \) and \( v_{x2} \) of the object point \( v \), are measured with respect to the corresponding image coordinate system. The symbols \( v \) and \( v_n \) represent the same physical point expressed in world (global) and image coordinate system respectively.

It is known that with each Cartesian reference frame, there is an associated homogeneous coordinate representation. So if \( v = (x, y, z) \) are the Cartesian coordinates of the object, its homogeneous coordinates are \( v = (wx, wy, wz, w) \) where \( w \) is an arbitrary constant. Clearly the actual Cartesian coordinates of a point are obtained from its homogeneous coordinates by dividing each of the first three components by the fourth one. A selection of such a system is made because the perspective transformation of a point in \( \mathbb{R}^3 \) to a point in \( \mathbb{R}^2 \), becomes a non-linear one when Cartesian coordinates are used. Generally the asterisk (*) notation is used to indicate the homogeneous coordinates of a point.

Perspective Transformation

This section outlines the derivation of the perspective transformation using two reference systems. The coordinates of an image point are calculated as a function of the camera parameters and the global coordinates of the corresponding physical point.

The overall transformation of the coordinates is a mapping between the global coordinate system and the image coordinate system. In order to define this mapping, the position and
orientation of the image reference frame must be defined with respect to the global frame. This can be done by the following three steps: (1) the global system at each camera is translated to the center of its lens; (2) each camera is panned and tilted, so that the y-axis is aligned with the corresponding optical axis; and (3) the center of lens is translated to the center of the image frame.

Let \( \mathbf{v}_{01} = (x_{11}, y_{11}, z_{11}) \) the vector from \( O_1 \) to the center of the lens of camera 1 and \( \mathbf{v}_{02} = (x_{12}, y_{12}, z_{12}) \) the vector from \( O_2 \) to the center of the lens of the second camera.

For the case of Figure 12, it is assumed that the distance between the lens of the first camera and the plane \( D \) is \( f_1 \). For the general case, where the distance is slightly greater refer to Appendix C.

The transformation of the coordinates is especially elegant if homogeneous coordinates are used. In this setting, pure translation which is a non-linear operation in ordinary Cartesian coordinates, becomes linear, Varga (1962). The matrix \( [\mathbf{T}] \) maps the homogeneous representation of the vector \( (x, y, z)^T \) into the homogeneous representation of vector \( (x-x_{11}, y-y_{11}, z-z_{11})^T \):

\[
\begin{bmatrix}
1 & 0 & 0 & -x_{11} \\
0 & 1 & 0 & -y_{11} \\
0 & 0 & 1 & -z_{11} \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

A pure rotation operation, Foley & Van Dam (1983) which is a
linear transformation in three dimensional Cartesian coordinates, remains linear in homogeneous coordinates. The following matrix \( \mathbf{R} \) represents the rotation operator that pans through an angle \( \theta \) and tilts through an angle \( \phi \):

\[
\mathbf{R} = \begin{bmatrix}
-\cos \theta & \sin \theta & 0 & 0 \\
-\cos \phi \cdot \cos \theta & \cos \phi \cdot \cos \theta & \sin \phi & 0 \\
\sin \phi \cdot \sin \theta & -\sin \phi \cdot \cos \theta & \cos \phi & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

Finally the matrix \( \mathbf{G} \) for performing the translation by \( f_1 \) is given by:

\[
\mathbf{G} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & -f_1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

The derivation of the matrices \( \mathbf{T} \), \( \mathbf{R} \), \( \mathbf{G} \) are presented in Appendix B.

The transformation which maps an object point expressed in homogeneous global coordinates into an image point expressed into homogeneous picture coordinates is given by:

\[
\hat{\mathbf{v}}_n = \mathbf{P} \mathbf{G} \mathbf{R} \mathbf{T} \mathbf{v}, \text{ where } [\mathbf{P}] \text{ is the following perspective transformation matrix:}
\]
Suppose that the object point is given by the vector \( v = (x, y, z)^t \). The image coordinates \( v_{1u1} \) and \( v_{1u3} \) of the image \( v_{1u} \) on the plane \( U \) are given by the following two equations:

\[
\begin{align*}
    v_{1u1} &= f_1 \frac{(x-x_{s1})\cos\theta_1 + (y-y_{s1})\sin\theta_1}{-(x-x_{s1})\cos\phi_1\sin\theta_1 + (y-y_{s1})\cos\phi_1\cos\theta_1 + (z-z_{s1})\sin\phi_1}
    \\
    v_{1u3} &= f_1 \frac{(x-x_{s1})\sin\phi_1\sin\theta_1 - (y-y_{s1})\sin\phi_1\cos\theta_1 + (z-z_{s1})\cos\phi_1}{-(x-x_{s1})\cos\phi_1\sin\theta_1 + (y-y_{s1})\cos\phi_1\cos\theta_1 + (z-z_{s1})\sin\phi_1}
\end{align*}
\]

It is clear that all the above equations have been derived for cameral. Equations of identical form hold for the other camera.

**Inverse Perspective Transformation**

In this section, the equation of the line of sight of cameral will be found using the perspective transformation equation. The object point lies on the line in a location still to be determined. From the equation of the perspective transformation it can be derived that:

\[
\tilde{v}_{1u} = P R G T \hat{v} \Rightarrow \hat{v} = T^{-1} R^{-1} G^{-1} P^{-1} \tilde{v}_{1u}
\]
This equation describes the inverse perspective transformation using homogeneous coordinates and two reference systems. Note that:

\[
T^{-1} = \begin{bmatrix}
1 & 0 & 0 & x_1 \\
0 & 1 & 0 & y_1 \\
0 & 0 & 1 & z_1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
G^{-1} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & f \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
R^{-1} = R^e
\]

Making the necessary calculations and transforming back to Cartesian coordinates, the previous equation can be rewritten as:

\[
v = \frac{1}{f_1-v_{4u2}} \begin{bmatrix}
f_1V_{4u2} \cos \theta_1 - f_1^2 \cos \phi_1 \cos \theta_1 + f_1V_{4u2} \sin \phi_1 \sin \theta_1 \\
f_1V_{4u2} \sin \theta_1 + f_1^2 \cos \phi_1 \cos \theta_1 - f_1V_{4u2} \sin \phi_1 \cos \theta_1 \\
+f_1^2 \cos \phi
\end{bmatrix} + \begin{bmatrix}
x_1 \\
y_1 \\
z_1
\end{bmatrix}
\]

The equation of the line of sight, (line through the center of the lens and the image point) as indicated by Figure 12, will be \(v - v_{1i} = \lambda (v_{1i} - v_{1i})\), where \(v_{1i}\) is the vector specifying the world coordinates of the lens and \(v_{1i}\) the world coordinates of the image of
v in U plane.

The expression for $v_+^-$ can be found from equation (1) by letting $v_{1+2}$ to be zero:

$$
V_1 = \begin{bmatrix}
    x_1 \\
    y_1 \\
    z_1 \\
\end{bmatrix} + \begin{bmatrix}
    V_{1+1} \cos \theta_1 - f_1 \cos \phi_1 \sin \theta_1 + V_{1+3} \sin \phi_1 \sin \theta_1 \\
    V_{1+1} \sin \theta_1 + f_1 \cos \phi_1 \cos \theta_1 + V_{1+3} \sin \phi_1 \cos \theta_1 \\
    f_1 \sin \phi_1 + V_{1+3} \cos \phi_1 \\
\end{bmatrix}
$$

(2)

The expression for $v$ can be found from equation (1) by letting $v$ to approach minus infinity:

$$
V_{*1} = \begin{bmatrix}
    x_{*1} \\
    y_{*1} \\
    z_{*1} \\
\end{bmatrix}
$$

(3)

Combining equations (1), (2), (3), one can obtain a final expression giving the projected ray in terms of the picture points $(V_{1+1}, V_{1+3})$ and a free non-negative parameter:

$$
v = \begin{bmatrix}
    x \\
    y \\
    z \\
\end{bmatrix} = \begin{bmatrix}
    x_{11} \\
    y_{11} \\
    z_{11} \\
\end{bmatrix} + \lambda \begin{bmatrix}
    V_{1+1} \cos \theta_1 - f_1 \cos \phi_1 \sin \theta_1 + V_{1+3} \sin \phi_1 \sin \theta_1 \\
    V_{1+1} \sin \theta_1 + f_1 \cos \phi_1 \cos \theta_1 - V_{1+3} \sin \phi_1 \cos \theta_1 \\
    f_1 \sin \phi_1 + V_{1+3} \cos \phi_1 \\
\end{bmatrix}
$$

To calculate the actual distance of the object from the imaging device one has to express the line of sight of camera2 in a similar way and then take the intersection of the lines. Things in reality are not so simple and a new problem arises because the lines of sight do not intersect due to errors in the matching process of zero crossing contours. A solution to this problem is presented in the
next section. As a result of the analysis given it should be clear that the perspective transformation maps an object point into a picture one but the inverse perspective transformation maps a picture point to a corresponding straight line (line of sight) using $v_{1,2}$ as a free parameter.

Distance Calculations

The analysis presented in the previous section shows that the perspective transformation is many-to-one. It follows that a given image point does not uniquely determine the location of its corresponding object point. In order to uniquely measure the three dimensional distance of an object point from the baseline of the stereoscopic device, the intersection of the lines of sight from the two cameras must be found. A mathematical analysis is presented in this section giving the best estimation to the depth measurements of a point in the physical scene under consideration.

Figure 13 is a simplification of Figure 12, and quite useful for the calculations of the three dimensional distance of the object point.
Suppose that \( U \) and \( U' \) are the image planes and \( L_1, L_2 \) the centers of the lenses and also \( r_1, r_2 \) the projected rays of an object \( v \). The vector \( \Delta = L_1 - L_2 \) is called the baseline of the stereoscopic device.

The problem of determining the three dimensional distance is now mapped into the trigonometric problem of determining the intersection of the two projecting rays. Assuming that they intersect, there exist two numbers \( a \) and \( b \), such that \( w_2 = \Delta + b \cdot w_1 \). The location of the object point is given by \( v = L_2 + a \cdot w_2 \).

In practice, however, one can assume that the two projected rays will fail to intersect due to various errors and especially due to inaccurate disparity measurements. A reasonable approach is to place the object point \( v \) midway between the two projected rays at
their point of closest approach. Formally:

\[ V = \frac{a'w_2 + (A + b'w_2)}{2} + L_2 \]  

(4)

where \( a' \) and \( b' \) are the values of \( a, b \) which minimize the:

\[ k = ||aw_2 - (\Delta + bw_2)||^2 \]

It can be verified that the quantity \( K \) is minimized by setting:

\[ a' = \frac{w_2\Delta - (w_1w_2)(w_1\Delta)}{1 - (w_1w_2)^2} \]  

(5)

\[ b' = \frac{(w_1w_2)(w_2\Delta) - (w_1\Delta)}{1 - (w_1w_2)^2} \]

Equation (4) gives the summation of the distances from object point to the center of the lens of camera plus the distance from the lens to the origin of the global system. Usually the distance of the object from the baseline is calculated as:

\[ V_1 = \frac{a'w_2 + (\Delta + b'w_1)}{2} - \frac{\Delta}{2} - \frac{a'w_2 + b'w_1}{2} \]

The coordinates of the midpoint of the baseline are given by the vector
The summation of the distances from the object point to the midpoint of the baseline plus the distance from this point to the origin of the global coordinate system is given by:

\[
V = \frac{a'w_2 + b'w_1 + L_1 + L_2}{2} + \frac{L_1 + L_2}{2}
\]  

(6)

In the previous equations the vectors \( w_i \) and \( w'_i \), denoted on Figure 13, are unit vectors, \( ||w_1|| = ||w_2|| = 1 \) and can be expressed as:

\[
W_i = \frac{V_i - L_i}{||V_i - L_i||}, \quad i = 1, 2
\]

where \( V_i \), \( L_i \) are taken from equations (2) and (3) respectively.

As a result of the above analysis, it can be noted that the location of the object point \( v \) which corresponds to two image points \( v_1 \) and \( v_2 \), is given by equation (4) or (6), where \( a' \) and \( b' \) are given by equations (5). The two unit vectors can be found from equations (7). Equations (4) and (6) are called triangulation equations of stereopsis.
Error Analysis

The error in range measurement of a point corresponding to an error in the distance measurement is derived in this section. Consider a point on the optical axes of the two views. Suppose that \( b \) is the baseline of the two cameras and \( R \) the distance of the point from the lens centers. Let \( \phi \) be an angle subtended by \( b \) at this point.

\[ \frac{dR}{d\phi} = \frac{R}{\phi} \]  

or

\[ \frac{\Delta R}{R} = \frac{|\Delta \phi|}{\phi} \]  

Figure 14. Error Analysis for Stereo Depth Measurement.

From the geometry of Figure 14 it can be seen that

\[ b = 2R \sin(\phi/2) \]

For small \( \phi \),

\[ R = \frac{b}{\phi} \quad \text{and so} \quad \left| \frac{dR}{d\phi} \right| = \frac{R}{\phi} \]  

or

\[ \Delta R = R \frac{|\Delta \phi|}{\phi} \]  

or

\[ \frac{\Delta R}{R} = \frac{|\Delta \phi|}{\phi} \]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
which is the expression for the relative error.

Inverse Rotation Transformation

As it has been described in the algorithm of Chapter III, in order for depth information for all sides of the object to be observed a rotation of the turntable is essential. Suppose that for each view the \((x,y,z)\) coordinates of the sampled points are computed and are denoted as \(x', y', z'\).

For the first view, before any rotation takes place, the coordinates which are computed are the same as the actual coordinates of the object point. However, for other views where the object has been rotated, in order to restore the real three dimensional coordinates, an inverse rotation transformation must be applied.

The rotation according to the algorithm, is about the \(z\)-axis. So the real \((x,y)\) coordinates of the object point will be given by:

\[
x = x \cos \theta - y \sin \theta \quad \text{and} \quad y = x \sin \theta + y \cos \theta
\]

where \(\theta\) is the angle of rotation.
CHAPTER VII

ORIENTATION PROBLEM

The next step in the complete representation of the surface of a real object is the computation of the orientation of small patches on the surface, which correspond to the matched zero crossings.

In this section the normal vector to the surface of the object in the scene is calculated as a function of the distance of the visible surface from the global coordinate system. The orientation of the visible surface can be measured assuming that the depth values of a large set of object points have been established. In the discussion following the calculations in the next section, it is established that the previous assumption is not always realistic and so the determination of the orientation must be postponed until the interpolation problem is solved.

Surface Orientation

The geometry of the computation of the orientation is illustrated in the next figure.
In this system, the vector corresponding to a particular point on the object in the scene is given by:

\[ \vec{r} = d(\cos \theta \cos \omega, \cos \theta \sin \omega, \sin \omega) \] or

\[ \vec{r} = d \vec{v}, \text{ where } \vec{v} = (\cos \theta \cos \omega, \cos \theta \sin \omega, \sin \omega), \text{ and } ||\vec{v}|| = 1. \]

In other words, \( \vec{r}(\theta, \omega) = d(\theta, \omega) \ (\cos \theta \cos \omega, \cos \theta \sin \omega, \sin \omega) \).

The partial derivatives of this vector are:

\[ \vec{r}_\theta = d_\theta \vec{v} + d \cos \omega (-\sin \theta, \cos \theta, 0) \] or

\[ \vec{r}_\omega = d_\omega \vec{v} + d \cos \theta \vec{v}_\theta, \text{ where } \vec{v}_\theta = (-\sin \theta, \cos \theta, 0). \]

In a similar way:

\[ \vec{r}_\omega = d_\omega \vec{v} + d \vec{v}_\omega, \text{ where } \vec{v}_\omega = (-\sin \omega \cos \theta, -\sin \omega \sin \theta, \cos \omega). \]

The normal vector \( \vec{N} \) at this point, is given by the cross product of the two partial derivatives:
\[
\vec{N} = \frac{\partial \vec{r}}{\partial w} \times \frac{\partial \vec{r}}{\partial \theta} = d d_w \cos w \vec{v}_w + d d_u \vec{v}_u - d^2 \cos w \vec{v}
\]

or \(\vec{N} = \{-d \cdot d\theta, d \cdot d_w \cos w, d^2 \cdot \cos w\}\)

To simplify the calculations the coordinate system is rotated around the x-axis so that the angle \(w\) is set to zero. In this case the normal vector is expressed as:

\[
\vec{N} = \{-d \cdot d\theta, d \cdot dw, d^2\}
\]

It should be mentioned here, that the algorithm of Chapter III explicitly determines the depth, only along small patches (contours) and so, one is confronted with the task of determining the derivatives of the distance from a sparse array of depth values. This is impossible for the cases that the disparity map represented in the 2 1/2-D sketch is too sparse. Note that the depth information have been established from the disparity map using the triangulation equation. A solution to that problem would be to first solve the reconstruction problem, presented in the next chapter, and then to calculate the surface orientation. Consequently the array of depth information would be very dense leading to an accurate and efficient determination of the depth derivatives.
CHAPTER VIII

RECONSTRUCTION OF VISIBLE SURFACE

Computational theories about stereo vision only specify computation of three dimensional surface information for special points in the image. Yet the visual perception is clearly of complete surfaces. To account for this, a computational theory of the interpolation (or approximation) of surfaces from visual information must be derived.

The problem is constrained by the fact that the surface must agree with the information derived from the output of the stereo correspondence procedure and also must not vary radically between points of known depth. In other words, the surface must obey the "surface consistency constraint" which will be defined in the next section.

To determine which of two possible surfaces is most consistent with the derived constraints, one must be able to compare the surfaces. To do this, a functional from the space of all possible surfaces to real numbers must be defined. In this way, the most consistent surface with the available visual information, will be that which minimizes the functional. To ensure that the functional has a unique minimum it must be, as explained later, a semi-inner product of a semi-Hilbert space.
In this chapter an extensive overview of the interpolation problem and its solution is presented. The drawbacks of the solution suggested by Grimson (1983) are emphasized, and the extensions suggested by Terzopoulos (1983) are also criticized. Two important and essential requirements that the reconstructed surface must meet, presented in this chapter, are: (1) it must explicitly define discontinuities on itself and on its orientation (2) must be viewpoint invariant. A detailed analysis is presented on how the solution given by Grimson (1981) can be extended so that the reconstructed surface to satisfy the two new requirements.

Surface Consistency Constraint

One of the initial steps which lead to the overall solution of the stereo problem is the accurate and efficient matching of corresponding zero crossing contours from two images. The three main steps, towards the solution of the matching problem, as been mentioned in a previous chapter are: (1) identify a location of the physical scene in one image; (2) identify the corresponding location in the second image and (3) compute a three dimensional surface value, representing the distance of the point, from the baseline of the stereoscopic device, based on the difference in the positions of the two corresponding points in the image.

It is essential to turn to the problem of determining computational constraints involved in the process of complete surface specifications by interpolating or approximating the known depth points. The zero crossings of a convolved image as well as
the depth information along these contours will be used as input to
the interpolation (approximation) procedure. It is obvious that an
infinite number of surfaces would fit through the depth points. Yet
there must be some way of deciding which surface or at least, which
small family of surfaces would give rise to the zero crossing
descriptions. In other words, there must be some additional
information available from the visual process which when taken into
account, will identify a class of nearby indistinguishable surfaces
that represent the visible surface of the scene.

To determine what information is available from the visual
process one must carefully consider the process by which the zero
crossings are generated. It has been mentioned in a previous
chapter that changes in reflectance of a surface caused by surface
scratches or texture markings, will give rise to zero crossings in
the convolved image. Sudden or sharp changes in orientation or
shape of the surface also give rise to zero crossings. This fact
can be used to constrain the possible shapes of surfaces.

Suppose a close zero crossing contour within which there are no
other zero crossings. An example in this case would be a circular
contour, along which the depth is constant. It is obvious, that
there are many surfaces that could fit this set of boundary
conditions, one of them being a flat disk. However, more surfaces
can be found to satisfy the given conditions. For example, the
surface formed by \( \sin(x^2 + y^2)^{1/2} \) would be consistent with the known
values. Such a rapidly varying surface should give rise to other
zero crossings. This follows from the observation that if the
surface orientation undergoes a periodic variation, then it is likely that the irradiance values will also undergo such a variation. Since the only evident zero crossings are at the borders of the object, this implies that the surface \( \sin(x^2 + y^2)^{1/2} \) is not a valid representative for the given set of boundary conditions. The new knowledge established from the previous example, is that the set of zero crossings contain implicit as well as explicit information about the surface to be constructed.

The implicit information contained in the image irradiances can be formalized as the surface consistency constraint, suggested by Grimson (1983), namely: The absence of zero crossings constrain the possible surface shapes.

Just as the presence of a zero crossing implies that some physical property is changing at a given location, the absence of them implies the opposite, that is, no physical property is changing and in particular the surface topography does not vary in a radical manner. Therefore the surface shape cannot change without informing the image by creating zero crossings on it.

To make explicit any constraints on the shape of the surface for locations in the image not associated with a zero crossing, one must carefully examine the image formation process from a physics point of view. Many factors are involved in the formation of image irradiances. Changes in any of these can cause a change in image irradiance and hence a zero crossing in the convolved image. For example, a change in surface material can correspond to a change in albedo and hence to a zero crossing in the image. Also a
Discontinuity in depth can correspond to a change in the illumination striking the surface and hence to a zero crossing. Finally, a discontinuity in surface orientation can create a change in the amount of illumination reflected towards the viewer, and as a consequence, to create a zero crossing.

From the above analysis, it is clear that if the albedo and illumination are constant, then the statement: changes in surface shape, create zero crossings, is equivalent with absence of zero crossings implies no changes in surface shape. The above statements are equivalent since one is the contrapositive of the other. In other words, the basic result, corresponding to the intuitive argument given above, is that the probability of a zero crossing to appear from regions where the illumination and albedo are roughly constant, is a monotonic function of the variation of the orientation of the normal. This means that the probability of a zero crossing, increases as the variation in the surface orientation increases. By inverting the argument it is evident that, the best surface to fit the known data is that which minimizes the variation in surface orientation.

In the literature, there is a confusion between two different construction problems. The visual surface interpolation and the visual surface approximation and both are referenced as the interpolation problem. A naive formulation of the visual surface interpolation approach would be to find the best approximation to a smooth surface, using only knowledge from the given points and requiring the surface to be interpolatory (i.e., passing through all
given data). A similar formulation of the visual surface approximation approach would be to find the best approximation to a smooth surface using the knowledge of a number of given points, but allowing the surface to be chosen freely.

It is obvious that in the second case the problem can be relaxed somewhat by requiring that the surface should approximately fit the known data and be smooth in some sense. Throughout the rest of this chapter the visual surface approximation problem as well as its solution are presented. It is evident, from the mathematical point of view that one may generate smoother surfaces by approximating three dimensional data, rather than interpolating them. Also the error in the approximation method is smaller, if one takes under consideration that the known data points are not so reliable, since they are obtained from the output of the raw 2 1/2 - D sketch, and may include errors, i.e., from disparity measurements which have been accumulated from previous stages of the visual process.

The Computational Problem

Given the initial boundary conditions of the known depth values along the zero crossing contours, there is an infinite set of possible surfaces each of which fits through those points. An essential requirement is the ability to compare pairs of members of this set and the ability to determine the most consistent one. A traditional way to compare surfaces, is to assign a real number to each surface. Then one may compare the surfaces by comparing the
corresponding real numbers. The assignment of real numbers to possible surfaces is accomplished by defining a functional, mapping the space of possible surfaces into the real numbers \( \Theta: X \rightarrow \mathbb{R} \). The functional should be such that, the more consistent the surface the smaller the real number assigned to it. Also, in order for the surface to satisfy the consistency constraint, the functional should measure the variation of the surface orientation. In this case, the most consistent surface will be the one that is minimal under the functional.

The key mathematical difficulty is to guarantee the existence and the uniqueness of a solution. In other words, there is at least one surface that minimizes the consistency constraint, and also, that any other surface passing through the known points, for which the functional has the same value, is indistinguishable from the first. This issue is not just a mathematical requirement, but is essential to the solution of many computational problems. Suppose that an iterative algorithm has been devised to solve the problem. What happens if the existence of the solution cannot be guaranteed? The iterative process could be set up to solve an equation and never converge to an answer. Further, suppose that a solution exists, but it is not unique. Then the iterative process may converge to one solution when started from one point and converge to another solution when started from a different point. Thus, in visual surface reconstruction, the real trick is to find the functional that accurately measures the variation of surface orientation, as well as guarantees the existence of a unique best surface.
Possible Functionals

The functionals that could be chosen as possible candidates must measure some factor of the second derivative of the surface, since it is strongly related to the variation of surface orientation.

One possibility is to measure the curvature of the surface, which implicitly reflects variation in the orientation. For any point on the surface, consider the intersection of the surface with a plane containing the normal vector at that point. This intersection defines a curve, the curvature of which can be measured as the arc rate of the rotation of its tangent. For any point there are infinitely many normal sections, each defining a curve. As the normal section is rotated through $2\pi$ radians, all possible normals will be observed. There are two sections of particular interest, one of them having the maximum curvature and the other the minimum. It can be shown that the directions of the normal sections are orthogonal, De Carmo (1976). These directions are the principal directions and the curvatures of the normal sections are the principal curvatures $k_a$ and $k_b$. The mean curvature of a surface is defined as $J = k_a + k_b$.

For a surface defined as $[x, y, s(x, y)]$ the mean curvature is:

$$J = \frac{\partial}{\partial x} \left[ \frac{s_x}{(1 + s_x^2 + s_y^2)^{1/2}} \right] + \frac{\partial}{\partial y} \left[ \frac{s_y}{(1 + s_x^2 + s_y^2)^{1/2}} \right]$$
So a possible functional could be

\[ \theta_1(s) = \iint (s - d)^2 \, dx \, dy + \iint J^2 \, dx \, dy \]

where \( s(x, y) \) is the smooth surface that is to be constructed, \( d(x) \) the given data points. The first integral, measures the faithfulness of the data. In the case of interpolation it has a zero value. Note that a continuous version of the approximation problem is considered here, since the known depth points are given as a function \( d(x) \). After a series of algebraic manipulations and by considering that \( s_x \) and \( s_y \) to be small, the functional is approximated

\[ \theta_2(s) = \iint (s - d)^2 \, dx \, dy + \iint (V^2s)^2 \, dx \, dy \]

Another possibility for a functional, is the quadratic variation in each of the surface variables. The quadratic variation in \( p = s_x \) is given by \( \iint (p_x^2 + p_y^2) \, dx \, dy \) and the quadratic variation in \( q = s_y \) is given by \( \iint (q_x^2 + q_y^2) \, dx \, dy \). If the surface is continuously differentiable, then \( p_y = a_x \) and by combining these two variations, a new functional could be defined as:

\[ \theta_3(s) = \iint (s - d)^2 \, dx \, dy + \iint (s_{xx}^2 + 2s_{xy}^2 + s_{yy}^2) \, dx \, dy \]

Still the question remains about the guarantee of the existence and uniqueness of the solution. In the case of interpolation and even the approximation, the calculus of variations can be used to determine a system of equations that the most consistent surface
must satisfy. While this system of equations characterizes the minimal surface, it does not guarantee uniqueness.

It can be shown that if the functional is a semi-inner product of a semi-Hilbert space of possible surfaces, then the most consistent surface is unique up to possibly an element of the null space of the functional, Grimson (1981). The null space is the set of surfaces that cannot be distinguished by the functional from the surface that is zero everywhere. In this way, the family of the most consistent surfaces can be found.

Although the condition of the semi-inner product is a mathematical requirement needed to guarantee a solution, it does not restrict in any unreasonable way the kinds of surfaces that can be considered, but it gives rise to at least two very natural functionals that have been derived before: (1) the integral of the square Laplacian and (2) the quadratic variation of the orientation.

Given that there are two possible functionals, are there any others? It can be shown that if the functional is: (a) monotonic function of the variation of the orientation, (b) semi-inner product and (c) rotationally symmetric (Appendix D), then there is a vector space of possible functionals, spanned by the square Laplacian and the quadratic variation, Grimson (1981). In other words, there is a family of possible functionals, given by all linear combinations of these two basic functionals.

What Has Been Achieved

As has been stated in the previous section for the general
reconstruction problem, there are mainly two constraints on the possible functionals. One is that the functional must measure a monotonic function of the variation of the surface orientation. The other is that the functional should satisfy the conditions of a complete parallelogram semi-norm, or equivalently, a semi-inner product. If the functional satisfies these conditions, then it is proven that there will be a unique family of surfaces that minimize that functional, and hence there exists a family of best possible surfaces to fit through the known depth values.

There are several points still to consider. Are there other possible functionals? How do they differ? What criteria can be applied to determine which is best? Since long mathematical proofs are required to answer the above questions, only the final results are presented here. There are many other possible functionals but they are all linear combinations of the two basic functionals the square Laplacian and the quadratic variation. Concerning the criteria used to determine the best functional, it can be proven that the square Laplacian and the quadratic variation both satisfy the same Euler equation, which implies that they are identical in the interior of the surface, and they only differ on the boundary conditions that they must satisfy. This criteria of boundary conditions is not the only one that must be considered, because, as it will be explained in a later section, the choice of the best functional also depends on which of them provides the most efficient solution, in the final discretization version of the approximation problem, where \( s(x) \) and \( d(x) \) are not continuous functions, but are
Given as sparse matrices.

Another criterion that must be used to choose the best functional is the size of the null space, which determines the level at which the functional cannot distinguish between two surfaces. It has been proven that the null space of the square Laplacian is the set of all harmonic functions and that the null space of the quadratic variation, is the set of linear functions. In the special case of the interpolation problem, (case of which is considered by mainly by Grimson), where \[ \int (s - d)^2 \, dx \, dy = 0, \] the size of the null space is a sufficient criterion to determine the best functional. Therefore, only in the case of the visual surface interpolation problem, the best surface can be found by simply minimizing the quadratic variation functional, because if the stereo algorithm provides at least three non-linear points (which is true most of the cases) then the element of the null space is uniquely determined to be the null surface (the surface that is zero everywhere). As will be explained later, this is not the case when the approximation problem is considered in its full extensions, preserving discontinuities and constructing a viewpoint invariant surface.

Discontinuities

One of the implicit assumptions in the previous discussions was that the pieces of the surface are in fact pieces of a single surface. Of course, this will frequently not be so. Consider for example, the front face of a staircase. It is obvious that the solution presented above must be modified to account for
discontinuities. One of the problems associated with the failure to make the surface discontinuities explicit, is that information about the shape of one surface affects the shape of the adjacent surface. The following figure gives an example of the Gibbs effect (oscillations of surface when applied around discontinuities in the data, for the case of a 1-D staircase). This phenomenon is not surprising since such reconstructions makes sense only over coherent parts of a signal, and applying a smooth reconstruction across a discontinuity implies that the shape of one surface can influence the shape of a second, distinct surface, an implication which is clearly incorrect.

![Diagram](image)

Figure 16. Fitting A Smooth Curve to Data with Discontinuities.

**Loose Constraints**

In the following sections of this chapter a reconstruction scheme for surface approximation is discussed which incorporates loose constraints. For a loose constraint, a penalty is charged
each time a constraint is broken. The penalty is weighted against certain other costs. If breaking a constraint leads, somehow, to a total saving in other costs that exceeds the penalty, then breaking that constraint is characterized as worthwhile. The energy of the other costs must be invariant, otherwise the existence and position of discontinuities will depend on the viewpoint.

Invariant schemes incorporating loose constraints are difficult to minimize because the cost function or the energy functional, are not convex.

The above reconstruction problem, as well as proposed solutions will be discussed in the following sections.

Energy of Loose Plate

As it has been explained in previous chapters, stereo image pairs can be matched to generate an array of sparsely distributed points of known depth. It is useful though, to reproduce a dense map from the data and particularly to mark discontinuities. These may either be steps or creases corresponding to either occlusions in the visible surface or to connections of edges (discontinuities in surface orientation). It is proposed that a plate rather than a membrane be used since both creases and step discontinuities must be recovered. The plate has the disadvantage, however, of requiring much more computation time. A mixture of plate and membrane has been shown to perform very poorly and so, for the general case, only a thin plate must be used.

The energy functional of a weak rod (1-D plate) is given by
each time a constraint is broken. The penalty is weighted against certain other costs. If breaking a constraint leads, somehow, to a total saving in other costs that exceeds the penalty, then breaking that constraint is characterized as worthwhile. The energy of the "other costs" must be invariant, otherwise the existence and position of discontinuities will depend on the viewpoint.

Invariant schemes incorporating loose constraints are difficult to minimize because the cost function or the energy functional, are not convex.

The above reconstruction problem, as well as proposed solutions will be discussed in the following sections.

Energy of Loose Plate

As it has been explained in previous chapters, stereo image pairs can be matched to generate an array of sparsely distributed points of known depth. It is useful though, to reproduce a dense map from the data and particularly to mark discontinuities. These may either be steps or creases corresponding to either occlusions in the visible surface or to connections of edges (discontinuities in surface orientation). It is proposed that a plate rather than a membrane be used since both creases and step discontinuities must be recovered. The plate has the disadvantage, however, of requiring much more computation time. A mixture of plate and membrane has been shown to perform very poorly and so, for the general case, only a thin plate must be used.

The energy functional of a weak rod (1-D plate) is given by
Courant and Hilbert (1963):

\[ E = \int (s - d)^2 dx + k^4 \int s'^2 dx + P \]

This expression differs from the energy of a weak string (1-D membrane) because it includes the second derivative of the surface \( s \), rather than the first. In the above expression the \( \int (s - d)^2 dx \) measures the faithfulness of the data. The term \( k^4 \int s'^2 dx \) is called a regularized term.

The last term in the above expression, the penalty term, consists of two factors: \( P_1 = b/dl \) which adds a penalty \( a \) per unit length of discontinuity and which adds a penalty \( \gamma \) for each crease discontinuity in \( s' \). The parameter \( k \) is called the scale factor.

The two dimensional equivalent in the above formula, for the plate, (Appendix E), comes in two varieties:

**Quadratic variation functional:**

\[ E = \int \int \{(s - d)^2 + k^2(s_{xx}^2 + 2s_{xy}^2 + s_{yy}^2)\} dxdy + P \]

**Square Laplacian functional:**

\[ E = \int \int \{(s - d)^2 + k^4(s_{xx} + s_{yy})^2\} dxdy + P \]

In fact, any linear combination of these two is a possible energy functional.

Since the energy is second order (i.e., contains second derivatives of \( s \)) it is possible to include penalties for both steps and creases.
Sensitivity in Discontinuity Detection

The performance of the rod as well as the plate are described by three parameters: \( k \) (scale), \( h_0 \) (contrast threshold) and \( g_0 \) (gradient difference threshold). The last of these concerns sensitivity to the detection of creases. An analysis of the different kinds of steps and creases as well as the necessary conditions to detect them, will be presented in this section for the case of 1-D surface. The same analysis applies to the 2-D surface.

The first case that is examined is an isolated step. Assume that a bi-infinite step of height \( h \) exists. The minimal energy with a step discontinuity in \( s(x) \) at \( x=0 \), is \( b \) because of the penalty factor \( P \). For a continuous \( s(x) \), the minimum energy can be shown to be

\[
E = \frac{1}{2\sqrt{2}} h^2 k \tag{Appendix F}
\]

The contrast threshold \( h_0 \) for the detection of an isolated step can be found from

\[
\frac{1}{2\sqrt{2}} h_0^2 k = 2b \Rightarrow h_0 = 2^{3/4} \sqrt{b/k}
\]

So if \( h > h_0 \) the lowest energy solution contains a discontinuity at the step.

The second case that has to be considered, is the interaction
of adjacent steps. Suppose two step discontinuities in \( s(x) \) are \( x=a/2 \) and \( x=-a/2 \). In this case the minimum energy will be \( 2b \) since \( P = b + b \). With continuous \( s(x) \) the minimum energy is:

\[
E = h^2 k \left( \frac{1}{\sqrt{2}} - \exp \left( -\frac{a}{\sqrt{2k}} \right) \cos \left( \frac{a}{\sqrt{2k}} + \frac{\pi}{4} \right) \right) \quad \text{(Appendix F)}
\]

Taking the limit for \( a \gg k \) the top hat (adjacent steps of height \( h \) and separation \( a \)) is treated as two isolated steps. In the case that \( a \ll k \) there is an iteration between the two steps.

The third case, that is examined is an isolated crease. Suppose a bi-infinite data \( d(x) \) with an isolated crease. This is an example where a rod (or a plate) exhibits behavior beyond that of a string (or that of a membrane). Suppose that the data \( d(x) \) consists of two semi-linear portions: \( d(x) = g_1 x \) if \( x<0 \) and \( d(x) = g_2 x \) otherwise, as can be seen in Figure 18. The data \( d(x) \) may either be approximated by a continuous rod (plate for the 2-D surface) or fit exactly by a rod \( s(x) \) with a crease coinciding with the crease in \( d(x) \).
The gradient difference is defined as \( g = |g_2 - g_1| \). If the \( s(x) \) has a crease at \( x = 0 \) then it fits \( d(x) \) perfectly with an energy (penalty factor \( P = \gamma \)). But if there is no crease in \( u(x) \), the minimum energy is given by:

\[
E = \frac{1}{2\sqrt{2}} g^2 k^3 \tag{Appendix F}
\]

Comparison of these two energies gives a threshold \( g_0 \) for detection of an isolated crease. Therefore it is detected if the gradient difference \( g \) satisfies

\[
g > g_0 = \left( \frac{\gamma}{a} \right)^{\frac{1}{2}} \frac{h_0}{k}
\]

The last case concerns the interaction of two adjacent
creases. The energy \( s(x) \) with two creases coinciding with those of \( d(x) \) is just \( E = \mathcal{P} = 2j \). The energy of a continuous rod (similarly for the plate) is given by:

\[
E = g^2k^3 \left( \frac{1}{\sqrt{2}} - \exp \left( \frac{a}{\sqrt{2k}} \right) \sin \left( \frac{a}{\sqrt{2k}} + \frac{\pi}{4} \right) \right) \quad \text{(Appendix F)}
\]

When \( a \gg k \) the two creases are treated as independent and are detected if \( g > g_0 \). But when \( a \ll k \) there is an interaction. The energy \( E \) (above) becomes

\[
E \equiv \frac{1}{2\sqrt{2}} (ga)^2k
\]

It can be proven that in this case the two adjacent creases (rod) or crease contours (plate) are interpreted as a step of height \( h = ga \).

Viewpoint Invariance in Surface Reconstruction

The plate and/or membrane are useful for construction of a visible surface either from stereo, laser rangefinder or other surface data. But the second order energies, already discussed, are not entirely suitable for this purpose. This is because they are not fully invariant in three dimensions. They are, of course, invariant to rotation in the \((x,y)\) plane, as has been shown in a previous section, but not to changes in viewpoint. This means that the optimal surface (the one that minimizes the energy functional subject to constraints) may wobble, as the viewpoint varies. This,
of course, is a big disadvantage, especially if motion-stereo is considered. It should be pointed out that the reconstruction from dense data is affected by non-invariance somewhat differently than reconstruction from sparse data. When data are dense, the reconstructing surface follows the data closely so that there is not much latitude for wobbling.

The cure to the above problem is to use an invariant energy based on surface area and curvature. The plate, discussed above, is in fact an approximation to an invariant energy. The approximation is better when all surface normals are nearly parallel to the line of sight. Hence wobble is more severe when some surface normals are nearly orthogonal to the line of sight.

The invariant equivalent of the square Laplacian functional is:

\[ E = \iint \left\{ (s - d)^2 \cos^2 \sigma + k^2 \left( k_1 + k_2 \right)^2 \right\} dS + \mathcal{P} \]

where \( k_1 \) and \( k_2 \) are the principal curvatures and \( dS \) is the surface element:

\[ dS = \sec \sigma dx dy = \sqrt{1 + s_x^2 + s_y^2} \, dx dy \]

where \( \sigma \) is the surface slant.

The square sum of the curvatures is given by (Monfredo de Carmo, 1976):

\[ (k_1^2 + k_2)^2 = \left( \frac{AS_{xx} - 2BS_{xy} + S_{yy}}{D^3} \right)^2 \]

where \( A = 1 + s_x^2 \), \( B = s_x s_y \), \( C = 1 + s_y^2 \), \( D = \sec^2 \sigma \). Note that \( A \), \( B \),
C, D are all functions of the first derivative of s only.

If, instead of the square Laplacian, the quadratic variation functional is to be used, then:

\[ k_1^2 + k_2^2 = (k_1 + k_2)^2 - 2k_1 k_2 \]

The justification for the \( \cos^2 \sigma \) term, which is an imperfect attempt to make \( (s - d)^2 \) term invariant, is illustrated in the next figure.

![Figure 18. Invariant Distance Measurement.](image)

Such a criterion, in the term \( (s - d)^2 \) is appropriate, under the assumption that "noise" (wobbling) derives from the surface.
However, it is inappropriate to the modification, if the "noise" derives from the sensor itself, because the "noise" then "lives" in the viewer (sensor) frame.

It can be shown, that expressing the energy $E$ as:

$$E = \int \bar{E}(s_x, s_y, s_{xx}, s_{yy}, s_{xy}) dx dy + P$$

the functional $\bar{E}$ is a non-convex function of $s_x$, $s_y$, $s_{xx}$, $s_{yy}$, $s_{xy}$. So it is not known, if there is a uniquely optimal $s$ (unique visible surface) to be found.

A method is suggested now which overcomes the problem of non-convexity. An estimation for $s_x(x,y)$ and $s_y(x,y)$ must first be obtained, by fitting an invariant membrane to the sparse stereo data (see Appendix G). The energy of an invariant membrane, is simply proportional to the surface area:

$$\lambda^2 \int dS = \lambda^2 \sqrt{1 + s_x^2 + s_y^2} dx dy$$

whose integrand, is a convex function of $s_x$ and $s_y$. So the loose membrane has energy:

$$E_a = \int \{(s - d)^2 \cos^2 \sigma + \lambda^2\} dS + P$$

where $P = b x$ (total length of discontinuities), as usual.

By minimizing $E$, one can obtain estimates for $s_x$ and $s_y$, and by insertion to the expression for $\bar{E}$ the functional becomes convex (unique minimum exists) with respect to $s_{xx}$, $s_{xy}$, $s_{yy}$.

Also the factor $P$ in the above formulae, must be made invariant. This could be done by incorporating the slant and tilt
of the surface. Suppose $\sigma$ is the surface slant and $T$ is the tilt. The penalty factor $P = b \times (\text{length of discontinuities}) = b \times (\text{length of contour of discontinuity})$. So it suffices to express the contour in terms of slant and tilt. Suppose $R$ is the contour of discontinuity:

$$L = \phi, \quad \| du \|, \text{where } \vec{u} \text{ is a vector.}$$

To obtain a plane with slant $\sigma$ and tilt $T$, one has to rotate plane $(x, y)$ by $T$ and then by $\sigma$ around the new $y$-axis. So, if $\vec{k}$ and $\vec{n}$ are respectively the unit normals to the image plane and to the rotated plane, then:

$$\cos \sigma = \vec{k} \cdot \vec{n}.$$ 

A vector $\vec{r}$, on the image plane satisfies the $\vec{r} \cdot \vec{k} = 0$, and a vector $\vec{u}$ on the rotated plane $\vec{u} \cdot \vec{n} = 0$. So:

$$\vec{r} = \vec{k} \times (\vec{u} \times \vec{k}) \quad \text{or}$$

$$\vec{u} = \frac{\vec{n} \times (\vec{r} \times \vec{k})}{\vec{n} \times \vec{k}} \quad \Rightarrow$$

$$\vec{u} = \frac{(\vec{n} \times \vec{k}) \vec{r} - (\vec{k} \times \vec{r}) \vec{k}}{(\vec{n} \times \vec{k})}.$$ 

By substitution:
\[ L = \oint \left\{ (dr)^2 + \frac{(n \cdot dx)^2}{\cos^2 \sigma} \right\}^{1/2} \]  

It is known that a unit tangent on a contour, is defined as:

\[ t = \frac{dr}{ds} \]

where \( s \) is the arc length of the contour. So:

\[ dr = t \, ds \]  \hspace{1cm} (3)

Also, the tangent vector \( \bar{t} \) and the normal \( n \) can be written as:

\[ \bar{t} = \cos \sigma \, \bar{i} + \sin \sigma \, \bar{j} \]  \hspace{1cm} (4)

\[ \bar{n} = \sin \sigma \, \cos \tau \, \bar{i} + \sin \sigma \, \sin \tau \, \bar{j} + \cos \sigma \, \bar{k} \]  \hspace{1cm} (5)

By substituting (3), (4) and (5) into equation (2), an expression of the length of the contour of discontinuity is obtained in terms of slant and tilt.
CHAPTER IX

SUMMARY AND CONCLUSIONS

The computational process for constructing the complete visual surface of an object(s) from a pair of stereo images, has been presented. Special techniques have been applied, so that the reconstructed surface will (a) be smooth between the fixed depth values, (b) to include discontinuities and (c) to be viewpoint invariant.

The first step in the analysis was to transform the stereo pair of images into the primal sketch, i.e., a representation of those image locations which correspond to positions in the scene at which some physical property of the surface changes. The initial raw images had to be convolved with the Laplacian of Gaussian operators ($\nabla^2 G$) of different sizes, and the zero-crossing contours had to be extracted. In other words, the raw images had to be processed and smoothed using special type of filters. Also the points where the second derivative of the image intensity had a zero were marked. A multi-resolution method was applied, so that unambiguous matches, from left and right images among zero-crossings could be obtained with high accuracy. The false target problem, due to the large number of possible candidates, was minimized. As a result, a wide range of disparity measurements have been calculated.
even in very fine resolution.

The next step involved the solution of the calibration problem, that is the optimum estimation of the elements of matrix \([H]\) from the equation \([H] \cdot [V] = [V']\), using the least square method, where \([V]\) gives the homogeneous coordinates of an object point in a global system and \([V']\) gives the homogeneous coordinates of the corresponding image point in the image system.

After the disparity had been calculated, the distance between every point on the surface corresponding to matched image points and the mid-point of the baseline of the imaging device was calculated using the triangulation equation. The errors in depth estimations, due to ambiguous or false matches was minimized by considering the object point lying in the middle of the shortest distance between the lines of sight.

The orientation of small regions on the surface in the scene which corresponded to the matched zero-crossing was obtained using equations relating the angular disparity to the distance of the small region from the two cameras.

The next step in the computational process was the transformation of the raw 2 1/2-D sketch to the full 2 1/2-D sketch. In other words, the problem of interpolation, or more precisely the approximation problem was solved. The reconstructed surface among the fixed depth points was modeled as a thin plate, using techniques from the theory of elasticity and was established by minimizing the energy functional of the plate. Loose constraints had to be taken into account and a penalty factor was added in the
expression of the energy function of the thin plate for detection of discontinuities. Also the problem of "wobbling" of the reconstructed surface was solved by using invariant quantities in the expression of the energy. Since the resulting functional was a non-convex one, a preprocess of fitting a thin membrane through the sparse depth points was required to transform the functional to a convex one, so that a unique minimum solution was guaranteed to exist.

After the object was rotated in order for a new face to be visible by the imaging device, the same process was repeated again to reconstruct the new surface. Real distances of the side face from the base line of the two cameras were calculated by using the inverse rotation transformation. Rotating the object a full 360 degrees the reconstruction of the whole surface was possible preserving also the discontinuities on it and on its orientation.

The main contributions of this work could be summarized as follows:

1. A presentation of a complete stereo visual process that receives as input a raw image of a scene and produces as output a 3-D representation which preserves discontinuities and is viewpoint invariant.

2. A detailed explanation of the matching process based on the work of Marr, Poggio (1976) and Grimson (1981) that gives a mathematical reasoning for the selection of the \(\nabla^2 G\) as the optimum filter to convolve the raw images.

3. After the presentation of the solution to the calibration
problem of the stereoscopic device and the determination of the 3-D depth values for a sparse set of points, a thorough derivation is given for the solution of the interpolation problem. A different approach to Grimson’s (1981) work is followed so that the reconstructed surface actually approximates the sparse matrix of depth values. Loose constraints have been used so that the surface labels explicitly the discontinuities in depth and orientation.

Extensions: Open Problems

It is known that one of the biggest subproblems in the stereo vision process is the correspondence problem, which still remains unsolvable in its full extent. In this work some implicit primitives, the zero crossings, were matched to obtain a solution with high accuracy. But the determination of these kind of features from images is not the right one in all kind of scenes, and especially when images are very complex with a highly repetitive structure (i.e., images taken at a distance of an office building with identical windows). In this case many ambiguous matches will appear. So, new kinds of features must be discovered, so that, the performance of an algorithm that solves the correspondence problem to be independent of the nature of the examined scenes.

Using image representations of multiple spatial scale, as part of a coarse-to-fine control strategy, has proven extremely useful in solving the correspondence problem inherent in stereo. Unfortunately, execution times of implementations on sequential machines are too long for real-time performance. It should be
emphasized that, in order to develop a real-time vision system, the implementation of the correspondence algorithm must be such that it would run on massively parallel machines as the NON-VON at Columbia, the connection machine at M.I.T., etc. In other words, the algorithm must run on a mesh-connected computer.

In the conventional two camera stereo, no disparity values have been calculated for horizontal edges or horizontal segments of zero crossings in the primal sketch of the two images. A possible solution to that problem, is to consider a three or a four camera stereo, where the cameras could be arranged on orthogonal axes or on a square grid or in a triangular configuration. For the last setup, which seems to be more promising, suppose the three images of the object, are called base-image, horizontal-image (image from the camera on the same level) and vertical-image. A point will be unambiguously matched if a unique candidate could be found among all three images. Matches among horizontal edges could easily be done, between the base and vertical image. Also the correspondence problem would be solved in less time, since a three or four camera stereo imposes more strict conditions on matching almost all points (matching must be done on all three images) except the ones that correspond to horizontal edges and occlusions.

In this work, an algorithm has been suggested for the complete construction of a regular object (i.e., cube, parallelepiped, staircase, pyramid, or a pile of them, but not with necessarily flat sides). The algorithm could also be extended in order to work with any object whose shape does not allow the explicit identification of
what we call an object's side or an object's face. For example, consider spherical or cylindrical objects or combinations of them. The only problem would be the correct determination of the number of rotations, during each pass through the main loop of the algorithm, as well as the angle for each rotation step.

A very important extension that must be mentioned here is the integration of other possible visual sources in order to obtain a real-time vision system. The big difficulty in this step, is the coupling of information from sources which, in a first approximation, have been assumed as independent. Processes or sources, such as stereopsis and analysis of motion, naturally generate local information about distances of surfaces from the viewer, while processes that analyze texture, contours and shading are better suited to provide information about local surface orientations. Therefore, the subsequent process which computes visible surface representations must integrate these multiple sources of scattered depth and orientation constraints. For example, a vision system, i.e., a mobile robot, could be developed by a motion-shading-stereo approach for the solution of the reconstruction problem, where the system could optimally update its perception (visible surface) of the object(s) as new parts of the scene are made visible during the movement in time and space and being able to perform equally well for a wide variety of scenes.
APPENDICES
Appendix A

DOGs and V2G
The Limit of A DOG Function

Wilson's DOG function may be written in one dimension as:

$$\text{DOG}(\sigma_a, \sigma_i) = \frac{1}{(2\pi)^{1/2}\sigma_a} \exp\left(-x^2/2\sigma_a^2\right) - \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left(-x^2/2\sigma_i^2\right)$$

where $\sigma_a$ and $\sigma_i$ are the excitatory and inhibitory space constants.

Writing $\sigma_a = \sigma$ and $\sigma_i = \sigma + \delta$, the right-hand side of the above equation varies with:

$$\frac{1}{\sigma} \exp\left(-x^2/2\sigma^2\right) - \frac{1}{\sigma + \delta} \exp\left(-x^2/2(\sigma + \delta)^2\right) = \delta \frac{\partial}{\partial \sigma} \left( \frac{1}{\sigma} \exp\left(-x^2/2\sigma^2\right) \right)$$

This derivative is equal to: $$-\frac{1}{\sigma^2} \exp\left(-x^2/2\sigma^2\right) - \frac{x^2}{\sigma^3} \exp\left(-x^2/2\sigma^2\right),$$

which is equal to $G^\nu$ up to a constant. So the limit of the DOG is the $V^2G$.

Approximation of $V^2G$ by A DOG

The function

$$\text{DOG}(\sigma_a, \sigma_i) = \frac{1}{(2\pi)^{1/2}\sigma_a} \exp\left(-x^2/2\sigma_a^2\right) - \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left(-x^2/2\sigma_i^2\right)$$

has the following Fourier transformation:

$$\text{DOG}(w) = \exp\left(-\sigma_a^2 w^2/2\right) - \exp\left(-\sigma_i^2 w^2/2\right).$$

Notice that $\text{DOG}(w)$ behaves like $w^2$ for small values of $w$ and so approximates a second derivative operator.
Appendix B

Transformation Matrices
Assume that \((x_n', y_n', z_n')\) are the new coordinates after a geometrical transformation that takes place on a point \(V\). We denote \((x_o, y_o, z_o)\) the initial coordinates of the point \(V\).

A translation of the point \(V\) from the origin is expressed by:

\[
\begin{align*}
X_n &= x_o - x_s; \\
y_n &= y_o - y_s; \\
z_n &= z_o - z_s.
\end{align*}
\]

In homogeneous coordinates: \(V_n = TV_0\) where the matrix \([T]\) is given by:

\[
T = \begin{bmatrix}
1 & 0 & 0 & -x_s; \\
0 & 1 & 0 & -y_s; \\
0 & 0 & 1 & -z_s; \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

We examine now the case of a combined rotation, initially by an angle \(\theta\) and then by an angle \(\phi\). The rotation matrix \([R]\) can be derived by multiplication of the matrix \([R_1]\) which represents a rotation about the \(X_1\) axis, by a matrix \([R_2]\), which represents a rotation about the \(Z_2\) axis. So \(V = R_1R_2V_0\) and \(R = R_1R_2\). The matrices \([R_1], [R_2]\) are given by:

\[
R_1 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos\phi_1 & \sin\phi_1 & 0 \\
0 & -\sin\phi_1 & \cos\phi_1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
\[ R_x = \begin{bmatrix}
\cos \theta_1 & \sin \theta_1 & 0 & 0 \\
-sin \theta_1 & \cos \theta_1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} \]
Appendix C

Distance Using Glimbal Center
It has been assumed in Chapter VI that the center of rotation of the camera, called glimbal center, coincides with the lens. Here a generalization is presented where the glimbal point and the center of the image plane define a vector \( l = (l_1, l_2 + f, l_3) ^ T \). In other words, the vector \( l \) expresses the distance between the center of rotation of each camera, of the stereoscopic device, and the corresponding image plane. In this case the transformation matrices \([T], [R], [G] \) could be written as:

\[
T = \begin{bmatrix}
1 & 0 & 0 & -X_1 \\
0 & 1 & 0 & -Y_1 \\
0 & 0 & 1 & -Z_1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
R = \begin{bmatrix}
\cos \theta_1 & \sin \theta_1 & 0 & 0 \\
-cos \phi_1 \sin \theta_1 & \cos \phi_1 \cos \theta_1 & \sin \phi_1 & 0 \\
\sin \phi_1 \sin \theta_1 & -\sin \phi_1 \cos \theta_1 & \cos \phi_1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
G = \begin{bmatrix}
0 & 0 & 0 & -l_1 \\
0 & 1 & 0 & -(l_2 + f) \\
0 & 0 & 1 & -l_3 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

Using the generalized form of \([T], [R], [G] \) and taking into account that \( \vec{v}_{ia} = (wv_{i+1}, wv_{i+2}, wv_{i+3}, w) ^ T \), the equation of the
The line of sight for the first camera becomes:

\[
\begin{align*}
\mathbf{v} &= \frac{1}{f_1 - v_{1u2}} \begin{bmatrix}
(f_1 v_{i1} - l_1, v_{i1} + l_1 f_1) \cos \theta_1 + (l_2 v_{i2} - l_2 f_1, -f_1^2) \cos \phi_1 \sin \theta_1 \\
-(l_2 v_{i2} - l_2 f_1, v_{i2} - l_2 f_1) \sin \phi_1 \sin \theta_1 \\
((f_1 v_{i1} - l_1, v_{i1} + l_1 f_1) \sin \theta_1 + (l_2 v_{i2} - l_2 f_1, f_1^2) \cos \phi_1 \cos \theta_1 \\
+(l_2 v_{i2} - l_2 f_1, v_{i2} - l_2 f_1) \sin \phi_1 \cos \theta_1 \\
-(l_2 v_{i2} - l_2 f_1, v_{i2} - l_2 f_1) \sin \phi_1 -(l_3 v_{i3} - f_1, v_{i3} - l_3 f_1) \cos \phi_1
\end{bmatrix}
\end{align*}
\]

The parametric equation of the line of sight becomes:

\[
\begin{align*}
\mathbf{v} &= \begin{bmatrix}
x \\
y \\
z
\end{bmatrix} = \begin{bmatrix}
x_1 + l_1 \cos \theta_1 - l_2 \cos \phi_1 \cos \theta_1 + l_3 \sin \phi_1 \sin \theta_1 \\
y_1 + l_1 \sin \theta_1 + l_2 \cos \phi_1 \sin \theta_1 - l_3 \sin \phi_1 \cos \theta_1 \\
z_1 + l_1 \sin \phi_1 + l_3 \cos \phi_1 + \lambda (v_{i1} \cos \phi_1, v_{i1} \sin \phi_1 + v_{i2} \sin \phi_1, v_{i2} \cos \phi_1 - v_{i3} \sin \phi_1, v_{i3} \cos \phi_1)
\end{bmatrix}
\end{align*}
\]
Appendix D

Rotationally Symmetric Operators
In this appendix, a series of propositions is presented that leads to the conclusion that the square Laplacian and the quadratic variation are rotationally symmetric operations.

Proposition 1

A function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ is rotationally symmetric, if its polar form is only dependent on radial distance $r$ and not on the direction angle $\theta$.

Proposition 2

Up to scaling, the only rotationally symmetric quadratic form of $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$ is: $V(x,y)$.

Proposition 3

The rotationally symmetric quadratic forms $\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial x \partial y}, \frac{\partial^2}{\partial y^2}$ form a vector space.

From the above propositions we can conclude that a rotationally symmetric quadratic form, forms a vector space that has the quadratic variation $(\frac{\partial^2}{\partial x^2})^2 + 2(\frac{\partial^2}{\partial x \partial y})^2 + (\frac{\partial^2}{\partial y^2})^2$ and the square Laplacian $\bullet (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})^2$ as basis.
Appendix E

Energy Calculations
Energy for the rod

A finite rod has energy functional:

\[ E = \int_{a}^{b} \{ (s(x) - d(x))^2 + k^4 (s''(x))^2 \} \, dx \]

The function \( s(x) \) which minimizes the \( E \) is to be found using calculus of variations. It can be shown that \( s(x) \) satisfies the Euler equation:

\[ \frac{d^4 s}{dx^4} + k^4 s = d \]  \hspace{1cm} (1)

with boundary conditions:

\[ s''(-a) = 0 = s''(b) \text{ and } s'''(-a) = s'''(b) = 0 \] \hspace{1cm} (2)

The solution of the differential equation (1) which satisfies the boundary conditions (2) is given by:

\[ s(x) = \int_{a}^{b} G(x, x') \, d(x') \, dx' \]

where \( G(x, x') \) is the Green's function:

\[ G(x, x') = \frac{1}{2\mu} \exp \left( \frac{-|x-x'|}{\sqrt{2} k} \right) \cos \left( \frac{|x-x'|}{\sqrt{2} k} - \frac{\pi}{4} \right) \]

It can be proven by integration of the initial functional that for finite and infinite regions:

\[ E = \int d(x) \, (d(x) - s(x)) \, dx \]

In the case of infinite domain, the energy can also be calculated using the Fourier transformation of \( d(x) \):

\[ \hat{d}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d(x) e^{-i\omega x} \, dx \]

So,
Energy for the plate

It can be proven that the energy functionals for the plate given by the quadratic variation and the square Laplacian, have the same Euler-Lagrange equation, namely:

\[ s + k^4 \nabla^4 s = d \]

The energy for the plate is obtained from:

\[ E = \int_A d(x) (d(x) - s(x)) dA \]

A complete mathematical derivation of the above formulas is presented by Blake, A. & Zisserman, A. (1985).
Appendix F

Derivation of Special Case Energies for a Continuous 1-D Rod
The equation for a continuous 1-D rod is given by:

\[ E = K^2 \int_{-\infty}^{\infty} \frac{\omega^4}{1 + \kappa^2 \omega^4} \left| \hat{d}(\omega) \right|^2 d\omega \]

where \( \hat{d}(\omega) \) is the Fourier transformation of \( d(x) \):

\[ \hat{d}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d(x) e^{-ix\omega} dx \]

**Isolated Step**

In the case of an isolated step of height \( h \), the Fourier transformation of \( d(x) \) is given by:

\[ \hat{d}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{i\omega} \frac{2h \sin(\omega a/2)}{\sqrt{2}} \]

Substituting in the general formula for the energy of 1-D rod the minimum energy is obtained to be:

\[ E = \frac{1}{h^2} \frac{h^2 k}{\sqrt{2}} \]

**Adjacent Steps**

Consider a step discontinuity of the \( s(x) \) function at the points \( x=a/2 \) and \( x=-a/2 \). In this case, the Fourier transformation of \( d(x) \) is:

\[ \hat{d}(\omega) = \frac{2h}{\sqrt{2\pi}} \frac{\sin(a\omega/2)}{\omega} \]

The minimum energy is:
\[ E = h^2 k \left( \frac{1}{\sqrt{2}} - \exp \left( \frac{-a}{\sqrt{2} k} \right) \cos \left( \frac{a}{\sqrt{2} k} + \frac{\pi}{4} \right) \right) \]

**Isolated Crease**

Suppose a bi-infinite data function denoted by \( d(x) \), with an isolated crease. In this case, the function \( d(x) \) is denoted as:

\[ d(x) = \begin{cases} g_1 x & \text{if } x < 0 \\ g_2 x & \text{if } x > 0 \end{cases} \]

The Fourier transformation of \( d(x) \) is given by:

\[ \tilde{d}(\omega) = \frac{-g}{\sqrt{2} \pi} \frac{1}{\omega^2}, \text{ where } g = |g_2 - g_1| \]

The minimum energy is given by:

\[ E = \frac{1}{2} g^2 k^3 \]

**Adjacent Creases**

It can be proven that in the case of adjacent creases, the Fourier transformation is:

\[ \tilde{d}(\omega) = \frac{-g}{\sqrt{2} \pi} \frac{1 - e^{-i\omega a}}{\omega^2} \]

where \( a \) is the separation between creases. The minimum energy of a continuous 1-D rod with two creases is:

\[ E = g^2 k^3 \left( \frac{1}{\sqrt{2}} \exp \left( \frac{-a}{\sqrt{2} k} \right) \sin \left( \frac{a}{\sqrt{2} k} + \frac{\pi}{4} \right) \right) \]
Appendix G

Loose Membrane
The energy functional of a membrane is given by:

\[ E = \int \left\{ (s-d)^2 + \lambda^2 (\nabla^2 s)^2 \right\} dA \]

where \( s \) satisfies the Euler-Lagrange equation \( s^2 - \lambda^2 \nabla^2 s = d \). The boundary condition is expressed by \( n \nabla s = 0 \) on \( C \), where \( C \) is the boundary of the domain and \( n \) its normal.

The energy functional of a loose membrane where a penalty factor \( l = b \int dl \) is applied, is given by:

\[ E = \int \left\{ (s-d)^2 + \lambda^2 (\nabla s)^2 \right\} dA + b \int dl \]

The coefficient \( b \) denotes the penalty for a discontinuity of unit length. To calculate \( s(x) \) from its associated energy, the Green's function of the differential equation \( s^2 - \lambda^2 \nabla^2 s = d \) must be obtained. It can be proven, that this function has the form:

\[ G(x, x') = \frac{1}{2\pi \lambda^2} K_0 \left( \frac{|x-x'|}{\lambda} \right) \]

where \( K_0 \) is the modified Bessel function. Hence the surface function \( s(x) \) is:

\[ s(x) = \int G(x, x') \, d(x') \, dA \]
BIBLIOGRAPHY


