THE UNIVERSITY OF CALGARY

Stereo Image Matching and Auto-DEM

by

Feng Tan

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

APRIL, 1998

ABSTRACT

Image matching using conventional aerial photographs has been intensively researched during the past two decades to solve various photogrammetric tasks like parallax measurement, point transfer, orientation of cameras, DEM surface reconstruction and other applications. A few groups have developed state-of-the-art photogrammetric workstations to make digital mapping production nearly automatic. The reason that DEM generation has not been fully automated is because the matching process is not yet 100% reliable. Blunders and missing points in DEM should not be permitted when making topographic maps and it is labor intensive to manually edit these data.

The goal of this research is to investigate, develop and test methodologies and procedures for solving the image matching problem and to estimate its potential and accuracy for extensive topographic mapping applications. In order to achieve a higher level of reliability, the combined method of signal and feature matching as well as pyramid image structures have been researched and implemented. After some testing with conventional aerial photographs, the results illustrate that the computer matching software VCOR can compete with human operators in terms of reliability, convenience and efficiency.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my supervisor, Dr. M. A. Chapman, for his continuous guidance, teaching, support and encouragement throughout my graduate studies. I am also grateful to Dr. J. A. R. Blais and his graduate DTM course which gave me a more in-depth understanding about terrain surface modeling. The staff and graduate students in the Department of Geomatics Engineering are thanked for supporting my studies. The research was carried out at and supported by VTA Photogrammetric Consultants Ltd. Calgary, AB. The cooperation of Mr. Vass Tsivos, President of VTA, and my co-worker Mr. Doug Sanden is very much appreciated.

DEDICATION

To my partner, David Wordsworth, for his unconditional love, care and support.

To my daughter and my parents in China for their understanding and sacrifices.

TABLE OF CONTENTS

APPROVAL PAGE ii
ABSTRACT iii
ACKNOWLEDGEMENTS iv
DEDICATION v
TABLE OF CONTENTS vi
LIST OF TABLES ix
LIST OF FIGURES x

CHAPTER **Page**

1. INTRODUCTION 1

1.1 The Advance of Automated Photogrammetric mapping 1
1.2 Research Motivation 2
1.3 The Scope of This Research 3

2. DIGITAL ELEVATION MODEL AND STEREO IMAGE MATCHING 5

2.1 Overview 5
2.2 Mathematical Description of Image Matching and Surface Reconstruction 7
2.3 Surface Sampling Techniques and Accuracy 12
2.3.1 Systematic Terrain Sampling 13
2.3.2 Progressive Terrain Sampling 14
2.4 Theoretical Accuracy Consideration of Terrain Sampling 15
2.5 Image Matching and Auto-DEM 18
2.6 General Consideration on Image Matching 19
2.7 Initial DEM Modeling 22
2.7.1 Polynomial 23
2.7.2 Test Results 24
2.8 Implementation of Auto-DEM 26

3. IMAGE GEOMETRY, EPIPOLAR LINES AND ORIENTATION 27

3.1 Geometry Optimization 27
3.1.1 Collinearity Equations 27
3.1.2 Stereoscopic Parallax 31
3.1.3 Elevations by Parallax Differences 33
3.1.4 Pushbroom Sensor with Line-scan Imagery 36
3.2 Epipolar Geometry 37
3.2.1 Basic Geometry of Epipolar Lines 38
3.2.2 Photo Coordinates and Scanning Coordinates 41
3.2.3 Image Coordinates in Homologous Epipolar Lines 42
3.3 Image Orientation 44

3.3.1 Overview 44
3.3.2 Interior Orientation 45
3.3.3 Relative Orientation 46
3.3.4 Absolute Orientation 47
3.3.5 Utilization of GPS Data in Block Bundle Adjustment 47

4. SIGNAL MATCHING 49

4.1 Correlation Basics 49
4.1.1 The Principle of Correlation 49
4.1.2 The Correlation Function 52
4.1.3 The Maximum Value of Cross Correlation 54
4.2 Area-based Signal Matching 56
4.3 Correlation Computation Method 58
4.3.1 Covariance Function 58
4.3.2 Correlation Coefficient 60
4.3.3 Weighted Correlation Coefficient 61
4.3.4 Absolute Values of Grey Level Differences 62
4.3.5 The Squares of Grey Level Differences 63
4.3.6 Grey Value Differences of Neighboring Pixels 63
4.4 Two Dimensional and One Dimensional Correlation 64
4.4.1 Two Dimensional Correlation 65
4.4.2 One Dimensional Correlation (Epipolar Line Correlation) 67
4.5 Least Squares Matching 69
4.6 Discussion of Signal Matching 70

5. FEATURE MATCHING (INTEREST OPERATOR) 73

5.1 Overview 73
5.2 Concept of Feature Based Matching 73
5.2.1 Selecting Distinct Points with an Interest Operator 75
5.2.2 Preliminary Correspondence Based on Similarity 76
5.2.3 Achieving Consistence 78
5.3 Interest Operator 79
5.4 Test Results 79

6 DATA STRUCTURING AND PYRAMID IMAGE MATCHING 81

6.1 Overview 81
6.2 Pyramid Image Formation 81
6.3 Pyramid Image Matching Procedure 83
6.3.1 The Unconstrained Pyramid Matching 85
6.3.2 The Constrained Pyramid Matching 86
6.4 Analysis of Multiresolution Image Correlation 87
6.5 Improvement of Correlation Result 91
6.6 Reliability of Correlation 93

6.7 Test Results 95

7. ORTHOPHOTOS AND DEM QUALITY CONTROL 99

7.1 Orthophotos and Digital Differential Rectification 99
7.1.1 Basic Concepts 99
7.1.2 Principle of Digital Rectification 99
7.1.3 Digital Rectification by Direct Method (Forward Solution) 101
7.1.4 Digital Rectification by Indirect Method (Backward Solution) 103
7.2 DEM Quality Control (QC) 105
7.2.1 Manual 105
7.2.2 Iterative Orthophoto Refinement (IOR) 106
7.2.2.1 Overview 106
7.2.2.2 DEM Error Computation 107
7.2.2.3 The Advantage of IOR 109
7.2.2.4 Implementation of IOR 110
7.3 Surface Fitting and Contour Map Generation 111
7.3.1 Surface Fitting 111
7.3.2 Contour Map Generation 112

8. CONCLUSION AND RECOMMENTATION 113

BIBLIOGRAPHY AND REFERENCES 115

APPENDIX I. THEORY OF PROJECTIVE TRANSFORMATION 121

1 Projective Transformation 121
2 Homogeneous Coordinates 124
3 Theory of Transformation for Photo Rectification 130
4 DLT and Collinearity Equations 130

APPENDIX II. IMAGE RESAMPLING 133

1 The Nearest Neighbor Method 134
2 Bi-linear Interpolation 134
3 Bicubic Convolution 137

LIST OF TABLES

No. **Page**

2.1 DEM initial approximation using polynomial method 25

5.1 Different window sizes using gradient operator 80

5.2 Success rates for variance and weighted gradient operators 80

6.1 Overall statistics 96

6.2 Statistics between image levels and blunders 97

6.3 Statistics between image level reduction and blunders 98

LIST OF FIGURES

No.		Page
2.1	The resultant grid meshes consists of *nxn* elements	8
2.2	The ground and the stereo image relationship	9
2.3	Four main factors responsible for image grey value	21
2.4	Auto-DEM flow-chart	26
3.1	Relationship between photo and ground	28
3.2	An overlapping pair of vertical photographs above datum	32
3.3	Geometry relationships on the vertical photo pair	34
3.4	SPOT pushbroom line scan imagery	36
3.5	Basic geometry of epipolar lines	38
3.6	Coordinate systems between image point and object point	39
3.7	Epipolar lines on real photograph and vertical photograph	41
3.8	Epipolar plane	43
3.9	Flow chart of image orientation	45
4.1	Signal A and B represent the portion of stereopair on the images	50
4.2	Signal A and B correlation curve	51
4.3	The real correlation curve	52
4.4	Correct position of correlation searched for	55
4.5	Area-based signal matching	57
4.6	Epipolar line correlation	67

6.1 Pyramid image 82

6.2 High frequency correlation curve changes 88

6.3 Different frequency channels 89

6.4 Improvement of correlation result 92

7.1 Digital rectification by direct method 102

7.2 Digital rectification by indirect method 104

7.3 Relationship between orthophoto mismatches and DEM errors 108

7.4 Flow chart of IOR 110

II.1 Bi-linear interpolation 135

II.2 Bi-cubic convolution interpolation 138

CHAPTER ONE

INTRODUCTION

1.1 The Advance of Automated Photogrammetric Mapping

The role of photogrammetry in the mapping community has been appropriately expressed in the following quote: "During the past few decades, photogrammetry has grown tremendously as an art and a science, and become a dominant mapping tool all over the world. We have observed the rise and decline of 'optical train' stereo plotters, the birth of analytical triangulation and block adjustment, and the ascent of analytical plotters to their current position as preferred photogrammetric instruments. Photogrammetric cameras have become highly advanced. They soar into space to map the moon and the earth from the highest possible vantage points. All in all, photogrammetry has served mankind well, in mapping and in industry, in the air and in space" (Helava, 1992)

We are now witnessing the advent of a new era: that of digital photogrammetry. Its roots are in computer technology, computer imaging and analytical photogrammetry. The softcopy photogrammetric workstation is a new type of analytical photogrammetric plotter. The application of digital photogrammetry has great potential to bring about a bigger revolution than any instrument or methodology development in photogrammetry

up to now. When fully developed, its effects will be felt in neighboring fields and photogrammetry itself will be dramatically changed.

1.2 Research Motivation

Surface representation plays an important role in geomatics for several reasons. The topographic surface of the Earth itself, in the form of a Digital Elevation Model (DEM), is frequently the end result of photogrammetric processes. Perhaps, more importantly, surfaces are an intermediate and essential step toward image understanding and object recognition. In this regard, stereo image matching is the main tool used to reconstruct the DEM in automated digital mapping.

Computer matching methods are still weak in comparison with matching performed by a human operator. Due to the image quality and the complexity of the terrain surface, matching on digitized conventional aerial photographs is still in a semi-automatic stage. While matching reliability is improved by using signal matching, many false matches still exist in automatic DEM generation. Blunders and missing points in DEM generated by image matching should not be permitted when making topographic maps. It is labor intensive to edit these data manually. The search for more advanced matching algorithms in order to improve reliability has become the main concern in the development of softcopy photogrammetric workstations and this study. The combined method of signal and feature matching as well as the pyramid image structures can improve the matching

reliability and can reduce blunders to a more manageable level which allow the dream of fully automated DEM generation to become closer to reality.

1.3 The Scope of the Research

The main objective of this research is to investigate, develop and test methodologies and procedures for automatically solving the image matching problem with an estimation of its potential and accuracy for extensive mapping applications. Besides that, this research serves for production and commercial purposes. This study was based on implementing, modifying and optimizing existing matching algorithms published in the various scientific papers thereby solving the actual mapping production problems. This research has been carried out in a real production environment and emphasizes practical aspects of the advanced matching algorithms, for example, reliability, convenience and efficiency. The images used in this research are digitized conventional aerial photographs. It is shown that computer matching software can compete with human operators in terms of reliability, convenience and efficiency.

This thesis contains eight chapters. In this Chapter, a review of the technology advances in automated photogrammetric mapping and an outline of the research motivation and objectives are given. Chapter 2 highlights the main considerations in DEM and surface reconstruction using stereo image matching. Chapter 3 provides an overview of basic mathematics and procedures of analytical photogrammetry. Digital techniques are introduced in Chapter 4, which addresses area-based signal matching. Feature extraction

and the feature-based matching concept is the subject of Chapter 5. Data structuring and multiresolution matching techniques, which have an impact on efficiency and reliability of the matching results, are addressed in Chapter 6. The important issues of DEM quality control and the iterative orthophoto refinement (IOR) for detecting matching blunders as well as the generation of contour maps are dealt with in Chapter 7. Finally, conclusions and recommendations for future research are presented in Chapter 8.

CHAPTER TWO

DIGITAL ELEVATION MODEL AND STEREO IMAGE MATCHING

2.1 Overview

A digital elevation model is often the end result of mapping. The major goal of this thesis is to use image matching as a tool to generate DEM automatically. Therefore, it is necessary to introduce some of the basics about the DEM and the relationship between image matching and surface reconstruction. Digital image matching techniques have the potential of modernizing, accelerating, and facilitating the task of data acquisition for deriving up-to-date, accurate digital terrain models (DTMs).

The term Digital Terrain Model (DTM) was first introduced by Miller and la Flamme who worked at the Massachusetts Institute of Technology during the late 1950s. The definition given by them was as follows: "The digital terrain model (DTM) is simply a statistical representation of the continuous surface of the ground by a large number of selected points with known X, Y, Z coordinates in an arbitrary coordinate field". Since then, several other terms - e.g. Digital Elevation Model (DEM), Digital Height Model (DHM) and Digital Ground Model (DGM) - have been coined to describe this, and other closely related processes. Although in practice these terms are often presumed to be synonymous, in reality they often refer to quite distinct products.

In this thesis following definition is used: "The expression of the elements of basic information such as topography, resources, geology, environments, land use, population distribution, etc. in a certain two-dimensional area which form the basic content of a complete geospatial information system. A digital elevation model is defined as the expression of pure topographic information, which is only a component part of the DTM. The DEMs are created by photogrammetric or cartographic methods" (Wang, 1990).

A DEM has wide applications, such as the production of contour maps, orthophoto maps, and perspective maps, etc. In engineering, they can be used in the route planning and location in the construction of railways and highways. All engineering projects in their design and construction stages can make use of a DEM in the computation of areas, volumes, and slopes. A DEM can be applied in the determination of the visibility between two ground points and in the definition and plotting of an arbitrary profile. In addition, a DEM can also be used in close-range photogrammetry in plotting the shape and size of an object with a complex surface structure. In remote sensing, DEM data can be used in conjunction with remote sensing data in remote sensing mapping, stereo satellite image production, and land classification. In military engineering, DEM data can be exploited in air navigation, and the guidance of long-range missiles (Wang, 1990, p. 324).

2.2 Mathematical Description of Image Matching and Surface Reconstruction

"A requirement for photogrammetric object surface reconstruction is the identification of homologous points in different images. A human operator uses the stereo effect to find the position of these points at maximum accuracy. Image matching allows for automatic determination of the homologous points. The image information has to be present in digital form, i.e. in the form of density value matrices" (Ebner and Heipke, 1988)

The object surface can be described using a geometric model $z(x, y)$ and a radiometric model of density values $g(x, y)$. The geometric modelling is carried out according to the finite element approach, for example. For the radiometric description, object surface elements are defined forming a square grid in the x, y plane which in size corresponds approximately to the pixels of the digital images multiplied by the image scale factor. A density value is assigned to every resulting object surface element.

The parameters of the geometric model, the assigned density values of the object surface elements, and the orientation parameters of all images are treated as unknowns and are estimated from the pixel density values and control information.

For reasons of simplicity, the DEM shall be restricted to bilinear surface elements, forming a square grid in x, y. Each of the resultant grid meshes consists of nxn object surface elements (Fig. 2.1). The elevation of an object surface element can be represented as a linear function of the z coordinates of the four neighbouring grid points.

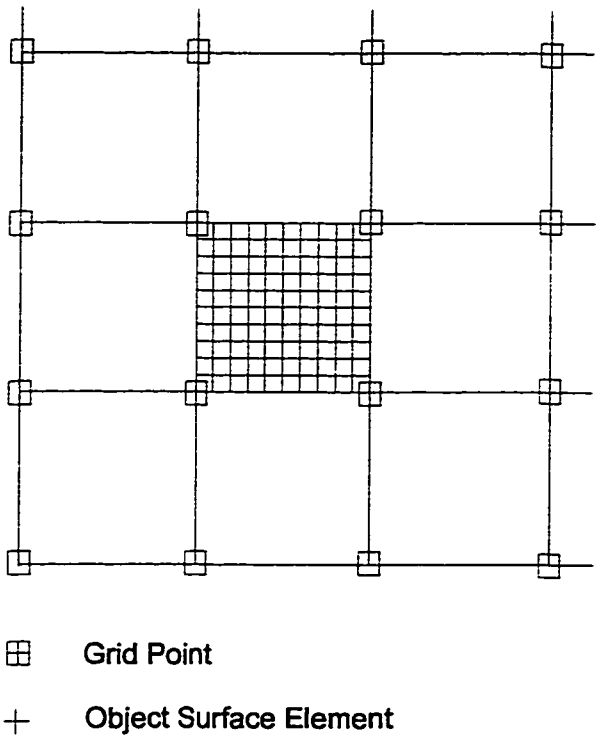


Fig. 2.1 The resultant grid meshes consists of nxn elements

(Adapted from Ebner and Heipke, 1988)

From the x, y, z coordinates of the center of the object surface elements and the orientation parameters of the surface elements and the orientation parameters of the relevant images, the image coordinates x_I, y_I can be computed (Fig.2.2). The position x_s, y_s of each point x_I, y_I in the scanning coordinate system is calculated using an affine transformation with the fiducials as reference points.

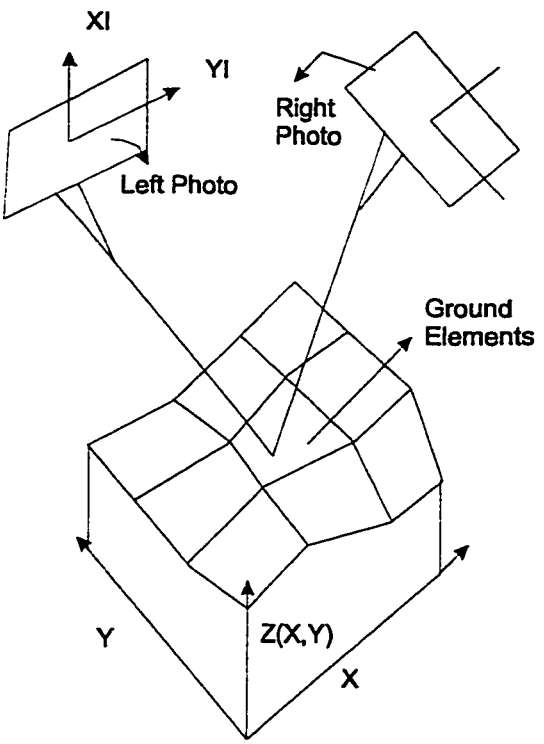


Fig. 2.2 The ground and the stereo image relationship

(Adapted from Ebner and Heipke, 1988)

Because x_s, y_s are non-integer coordinates the corresponding image density value g must be computed from the neighbouring density values possibly using bilinear interpolation. Thus, g is a function of the unknown coordinates $\hat{z}$ of the grid points of the geometric surface model and the unknown orientation parameters $\hat{p}$ of the images.

The measurement of the fiducials in the scanning coordinate system is carried out using image matching techniques. If for example, the considered object surface element is

transformed into four images, the algorithm yields four image density values, which are then regarded as observations in a least squares adjustment. Four observation equations of the following type can be formulated:

$$\hat{v} = \hat{g} - g(\hat{z}, \hat{p}) \tag{2.1}$$

Where $\hat{g}$ is the unknown density value assigned to the object surface element and $\hat{v}$ the difference between $\hat{g}$ and the corresponding image density value g.

For all nxn object surface elements in every grid mesh the corresponding observation equations are formulated. The system is completed by adding the observation equations for the control information.

The equation (2.1) is nonlinear in unknowns $\hat{z}$ and $\hat{p}$, hence a Taylor expansion is necessary. The linearized observation equations read:

$$\hat{v} = \hat{g} - (\frac{\partial g}{\partial z})_0 \Delta\hat{z} - (\frac{\partial g}{\partial p})_0 \Delta\hat{p} - g(z^0, p^0) \tag{2.2}$$

where z^0 and p^0 stand for initial values and $\Delta\hat{z}, \Delta\hat{p}$ for the corrections of the unknowns $\hat{z}$ and $\hat{p}$ of the adjustment. The image density values $g(z^0, p^0)$ are computed from the

initial values z^0 and p^0 using the collinearity equations, an affine transformation and a given density value interpolation (Ebner and Heipke, 1988):

Collinearity equations:
$$\begin{cases} x_l = x_l(z^0, p^0) \\ y_l = y_l(z^0, p^0) \end{cases} \tag{2.3}$$

Affine transformation:
$$\begin{cases} x_s = x_s(x_l, y_l) \\ y_s = y_s(x_l, y_l) \end{cases} \tag{2.4}$$

Density value interpolation:
$$g(x_s, y_s) = \sum_k \sum_l \alpha_{k,l}\, g(x_k, y_l); \quad x_k, y_l \in nxn \quad \text{where} \sum_k \sum_l \alpha_{k,l} = 1 \tag{2.5}$$

Thus

$$g(x_s, y_s) = g(x_l, y_l) = g(z^0, p^0) \tag{2.6}$$

Expanding the coefficients $(\frac{\partial g}{\partial z})_0$ and $(\frac{\partial g}{\partial p})_0$ of the design matrix yields

$$\begin{cases} (\frac{\partial g}{\partial z})_0 = (\frac{\partial g}{\partial x_s}\frac{\partial x_s}{\partial z})_0 + (\frac{\partial g}{\partial y_s}\frac{\partial y_s}{\partial z})_0 \\ (\frac{\partial g}{\partial p})_0 = (\frac{\partial g}{\partial x_s}\frac{\partial x_s}{\partial p})_0 + (\frac{\partial g}{\partial y_s}\frac{\partial y_s}{\partial p})_0 \end{cases} \tag{2.7}$$

$\frac{\partial g}{\partial x_s}$ and $\frac{\partial g}{\partial y_s}$ represent density value gradients of the digital image at the location $x_s(z^0, p^0); y_s(z^0, p^0)$. In the most simple case the weight matrix of the density value observations is represented by the identity matrix:

$$P = I \tag{2.8}$$

The solution of the adjustment is found iteratively applying the least squares method. The standard deviations for the unknowns are computed, too.

This approach combines digital image matching and object surface reconstruction to form a general model in digital photogrammetry. The pixel density values substitute the image coordinates as observations and the object surface is reconstructed as a whole. During the matching procedure, both orthophoto and DEM can be created at the same time. The grey value being assigned to a given location is available twice (in both images). This information may offer a greater flexibility because the mean value of the grey level can be used to reach a more realistic solution (Larouche, 1995).

2.3 Surface Sampling Techniques and Accuracy

The position and elevation accuracy which can be achieved using phtogrammetric methods, the scale of the final maps or terrain model, the possible contour interval, etc., are dependent on various interrelated factors, but chiefly:

(a) the scale and resolution of the aerial photograph;

(b) the flying height at which the photograph was taken;

(c) the base to height ratio (i.e., the geometry) of the overlapping photographs;

(d) digitizer (e.g. CCD Scanner).

2.3.1 Systematic Terrain Sampling

A systematic pattern of spot heights may be measured in a regular geometric (e.g. square, rectangular, triangular) pattern as specified by the users. Such an approach is favored in any type of photogrammetric operation which is either fully or partially automated where the location of the required grid node points can be pre-programmed and driven to under computer control. Thus the grid can be preset to a specific interval, which has the consequence that the finer but perhaps significant terrain features will not be measured specifically. Therefore, the obvious shortcoming of the sampling pattern to data acquisition is that the distribution of data is not related to the characteristics of the terrain itself. If the data point sampling is conducted on the basis of a regular grid, then the density must be high enough to portray accurately the smallest terrain features present in the area being modeled. If this is done, then the density of data collected will be too high in many areas of the model, in which case, there will be an embarrassing and unnecessary data redundancy in these areas. In this situation, filtering of the preprocessing activity before the DEM can be defined (Wang, 1990).

2.3.2 Progressive Terrain Sampling

Since this grid based approach is most easily implemented in computer-controlled photogrammetric machines such as analytical plotters, a solution to the above shortcoming has come from photogrammetrists in the form of progressive sampling, and its development, composite sampling, originally proposed by Makarovic (1973) of ITC. Instead of all the points in a dense grid being measured, the density of the sampling is varied in different parts of the grid, being matched to the local roughness of the terrain surface.

The basis of the method is that one starts with a widely spread (low-resolution) grid which will give a good general coverage of height points over the whole area of the model. Then a progressive increase in the density of sampling (or measurement) takes place on the basis of an analysis of terrain relief and slope using the on-line computer attached to the photogrammetric machine. Thus the basic grid is first defined by halving the size of the grid-cell in certain limited areas based on the results of the proceeding terrain analysis. Measurements of the height points at the increased density are carried out under computer control only in these predefined areas. A further analysis is then carried out for each of these areas for which the measured data has been increased or identified. Based on this second analysis, an increased density of points may be prescribed for still smaller area. Normally, three such runs or iterations are sufficient to acquire the terrain data necessary to define a satisfactory model.

In this way, the progressive sampling technique attempts to optimize automatically or semi-automatically the relationship between specified accuracy, sampling density and terrain characteristics.

2.4 Theoretical Accuracy Consideration of Terrain Sampling

"The accuracy of DEM is affected by the density of sampling and observation errors of the reference points. These two disturbing sources are independent of each other and influence the root mean square errors or fidelity of regenerated data in different degrees" (Wang, 1990, p. 329).

One method of assessing the way in which sampling density affects accuracy is the use of a transfer function whereby you can make a theoretical analysis of the fidelity of the data regenerated by reference points, i.e., to understand the amount of information transferred. A transfer function is used to express how a certain linear system responds to the sinusoid inputs of different frequencies. Here frequencies are used to express the shape of a terrain surface, whereas any profile or the shape of a surface can be expressed by the Fourier series, which are composed of an infinite number of sine and cosine waves with different multiple frequencies, that is:

$$Z(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} A_k \cos(k\omega_0 x + \varphi_k)$$

$$= \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\omega_0 x + b_k \sin k\omega_0 x) \tag{2.9}$$

where Z, x represent respectively the vertical and horizontal coordinates of a curve. Here they indicate the height and length of a terrain profile; $w = 2\pi / T = 2\pi f_0$ is the angular frequency of the fundamental wave; T is the period of the wave; $f_0 = 1/T$ is the frequency; k is an integer defining the number of harmonic waves.

$$a_k = A_k \cos\varphi_k ; b_k = -A_k \sin\varphi_k ; A_k = \sqrt{a_k^2 + b_k^2} ; \varphi_k = arctg\left(\frac{-b_k}{a_k}\right) \tag{2.10}$$

Theoretically, there is an infinite number of frequencies in the total summation. However, in image processing the high frequency f_{max} existing in images is in fact finite. In accordance with the Nyquist Sampling theorem, the interval of sampling Δx should be selected by the following relationship, i.e.,

$$\Delta x \leq 1/2 f_{max} \tag{2.11}$$

Data acquisition is an important step in building a DEM. It is clear that a sparse distribution of reference points will affect the accuracy of the DEM, whereas a dense distribution will increase the amount of work in data acquisition, processing and storage. The density of reference points can be determined according to the Shannon sampling theorem, that is, the interval of sampling should meet the relationship

$$\Delta \leq 1/2f_c \tag{2.12}$$

Δ is the local sampling interval;

f_c is the local area frequency

Therefore, the density of sampling points should be related to the change of terrain undulations.

For "vertical" photographs, the collinearity equations are expressed as follows (derivation is shown in Chapter 3):

$$\begin{cases} x_a = -f\dfrac{\Delta X}{\Delta Z} = -f\dfrac{X_A - X^0}{Z_A - Z^0} \\ y_a = -f\dfrac{\Delta Y}{\Delta Z} = -f\dfrac{Y_A - Y^0}{Z_A - Z^0} \end{cases} \tag{2.13}$$

If taking the differential of x_a to Z_A , the following exists:

$$dx_a = f\frac{X_A - X^0}{(Z_A - Z^0)^2}dZ_A \tag{2.14}$$

From Eq.(2.14), the resolution or interval of digitizing can be qualitatively analyzed. It is clear that the interval of sampling relates to several factors: the scale of the photo; the relative position of digitized area to the principal point; and the accuracy of the DEM, i.e.,

a. If recording a small change in elevation in our DEM (small dZ_A), we should use a small sampling interval or high digitizing resolution.
b. If the scale of the photo is small (or the height of flight is high) the interval of digitizing should be small.
c. The nearer the position of the area to be digitized is to the photograph center, the smaller the necessary sampling interval.

2.5 Image Matching and Auto-DEM

"In stereo and motion vision, matching refers to identifying corresponding visual phenomena in image sequences, caused by the same phenomena in object space. The purpose of stereo vision is surface recovery of 3-D object space from conjugate image pairs. 3-D surface description lies at the basis of a structural description of the real world. Digital photogrammetry applies matching techniques for the determination of DEM and point transfer in aerotriangulation" (Lemmens, 1988).

The problem of stereo analysis consists of the following main stages:

1.Space extraction of phenomena in both images;

2.Space selection of corresponding phenomena and computation of their object space coordinates from triangulation;

3.Space interpolation to arrive at a full 3-D surface description.

Image matching can be classified to three categories:

- signal matching (area-based matching)
- feature matching (interest operator)
- relational matching (structured matching)

To reduce the search space and find matches more reliably, multiresolution, coarse-fine pyramid image structures are frequently used. But the three matching methods can be viewed in a pyramid way, too: relational matching, to find rough (i.e. global) matches, next feature matching will give precise (i.e. local) matches, and signal matching is more detailed information (i.e. non-local) matches. For high precision measurements these feature matches may be considered as approximation values for signal matching.

2.6 General Consideration for Image Matching

The foundation of stereo vision is the recording of an object space from slightly different view points. The difference in positions causes disparities and from triangulation, using the exterior orientation, 3-D object coordinates are computed. By interpolation the 3-D surface structure can be recovered. So, the three main steps in any stereo vision algorithm are (Lemmens, 1988):

1. detection of items or phenomena;
2. matching and calculating depth;
3. surface recovery.

The phenomena may be:

1. non-local (i.e. neighborhoods of pixels).
2. local (e.g. edges and blobs).
3. locally extended (e.g. line segments and areas).

A non-local item consisting of the grey values of a neighborhood of pixels is a target area, a local item is a feature and a locally extended item a structure. So the above subdivision of phenomena defines broadly the three types of matching methods.

The correspondence problem is constrained by three conditions (Lemmens, 1988):

1. Compatibility constraint: if two elementary groups of pixels could have arisen from the same object item, they can match, or else they can not match;
2. Uniqueness constraint: each elementary group of pixels in the one image can only match one element in the other;
3. Continuity constraint: disparities vary smoothly almost everywhere.

There are four main factors responsible for the grey values in an image:

- illumination;
- reflectivity;
- geometry;
- viewpoint.

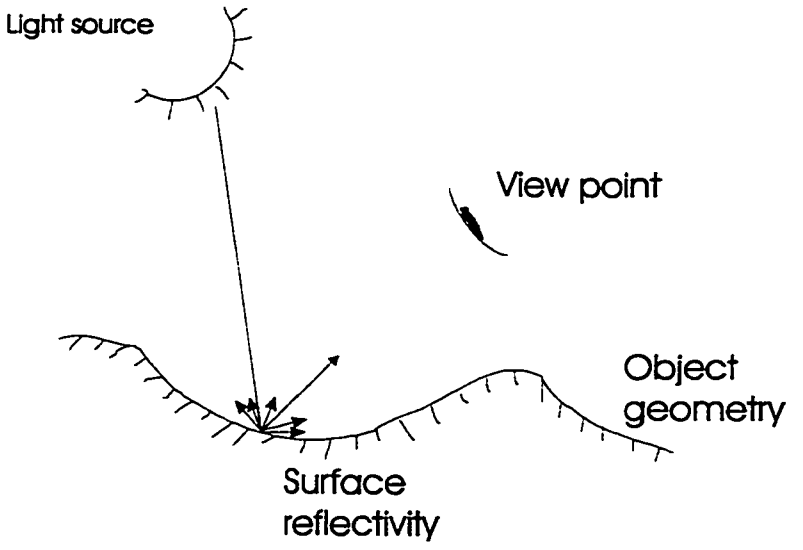


Fig. 2.3 Four main factor responsible for image grey value

(Adapted from Lemmens, 1988)

To achieve unambiguous matches, all four factors should be known in advance. However, only the position and attitude (i.e. exterior orientation) of the viewpoints and the illumination source(s), are in general known, and the geometry of the object space surface is what we are looking for. The reflectivity of the surface is generally also unknown. Because, except in cases of perfect Lambertian, the geometry affects the amount of light that arrives at the view point, even when the surface reflectivity is known, no unique image grey value distribution is determinable.

2.7 Initial DEM Modeling

"The raw object space is an important intermediate representation. Moreover, many of the points are precisely known in object space, giving rise to a first surface approximation.

The raw object space contains basic information about the surface: breaklines, depth, and surface normals. Unlike the images, the representation of the raw object space is not iconic; it is already a step toward the geometric and semantic description of the desired end result" (Schenk and Toth, 1992).

The initial DEM approximation models can be used to estimate starting position for automatic image correlation based on the limited known object space information such as Ground Control Points (GCPs), and break lines. A good approximation DEM is the key for image matching. It can not only reduce the search range, avoid the mismatching, but also give a precise initial approximation for least-square matching to save computation time.

A polynomial math model is used in this research to estimate the initial value of Z at all grid points based on raw object space information before actually performing image matching.

2.7.1 Polynomial

"The simplest way to describe gradual long-range variations is to model them by polynomial regression. The idea is to fit a polynomial surface by least squares through the known data points (GCPs). It is assumed that the spatial coordinates X,Y are the independent variables, and that Z, the property of interest, is the dependent variable. In two dimensions the polynomials are surfaces of the form" (Burrough, 1986, p. 149)

$$f\{(X,Y)\} = \sum_{r+s\le p}(b_{rs}X^{r}Y^{s}) \tag{2.15}$$

in which the first three are:

b_0	flat
$b_0 + b_1X + b_2Y$	linear
$b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2$	quadratic

The integer p is the order of the trend surface. There are $P = (p+1)*(p+2)/2$ coefficients that are normally chosen to minimize

$$\sum_{i=1}^{n}\{Z(V_i) - f(V_i)\}^2 \tag{2.16}$$

where V_i is the vector notation for (X,Y). So a horizontal plane is zero order, an inclined plane is first order, a quadratic surface is second order, and a cubic surface with 10 parameters is third order.

The advantage of trend surface analysis is that it is a technique that is superficially easy to understand, at least with respect to the way the surfaces are calculated. The disadvantage of trend surfaces is that polynomial functions are smoothing functions, rarely passing

exactly through the original data points unless these are few and the order of the surface is large. But less GCPs can not get enough redundancy.

2.7.2 Test Results

Table 2.1 shows that the polynomial method with third order needs estimated search range equals to 6 times r.m.s based on of GCP residuals in cooperation with the computation time.

Accepted: correlation coefficient 0.6~1.0; Rejected: correlation coefficient <0.6

Total points: 2749

	Polynomial	
Search Range	Accepted	Rejected
10 (pixels)	1080	1669
27 (pixels)	1697	1052
54 (pixels)	2162	587
76 (pixels)	2329	420
108 (pixels)	2366	383
162 (pixels)	2375	374
216 (pixels)	2389	360

Table 2.1 DEM initial approximation using Polynomial method.

The results are from processing (dataset) with VCOR (a VTA matching software)

2.8 Implementation of the Auto-DEM

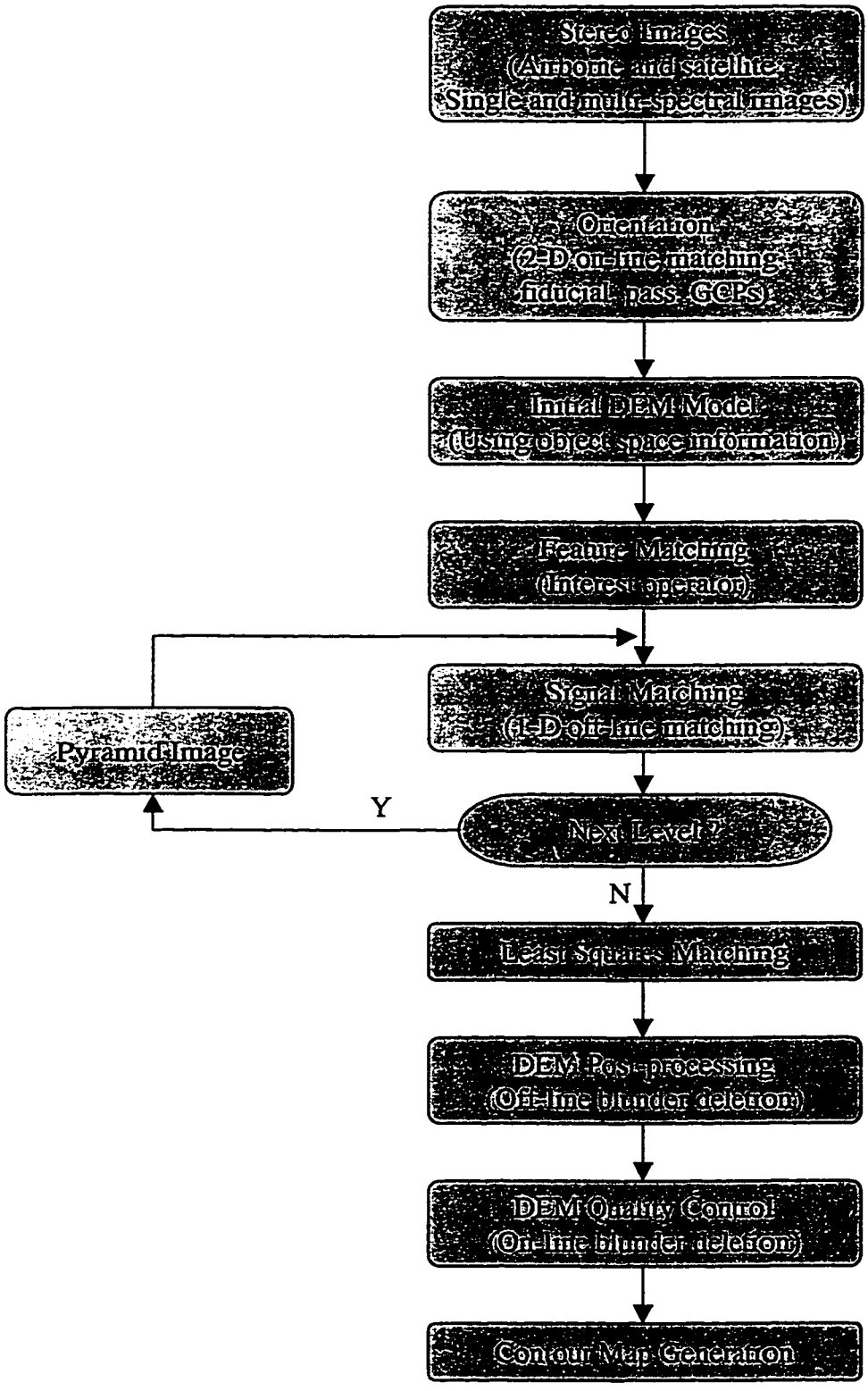


Fig. 2.4 Auto-DEM flow-chart

CHAPTER THREE

IMAGE GEOMETRY, EPIPOLAR LINES AND ORIENTATION

3.1 Geometry Optimization

Before discussing image matching techniques on stereo images, we need to give a simple review of the basic geometry of an image and procedures in analytical photogrammetry in order to reconstruct the terrain surface by precise mathematical models.

3.1.1 Collinearity Equations

Collinearity assumes that object point $A(X_A, Y_A, Z_A)$, perspective center $P(X^0, Y^0, Z^0)$ and image point $a(x_a, y_a, f)$ are lie on the same line, here f is focal length (Fig.3.1) (Chapman, 1992)

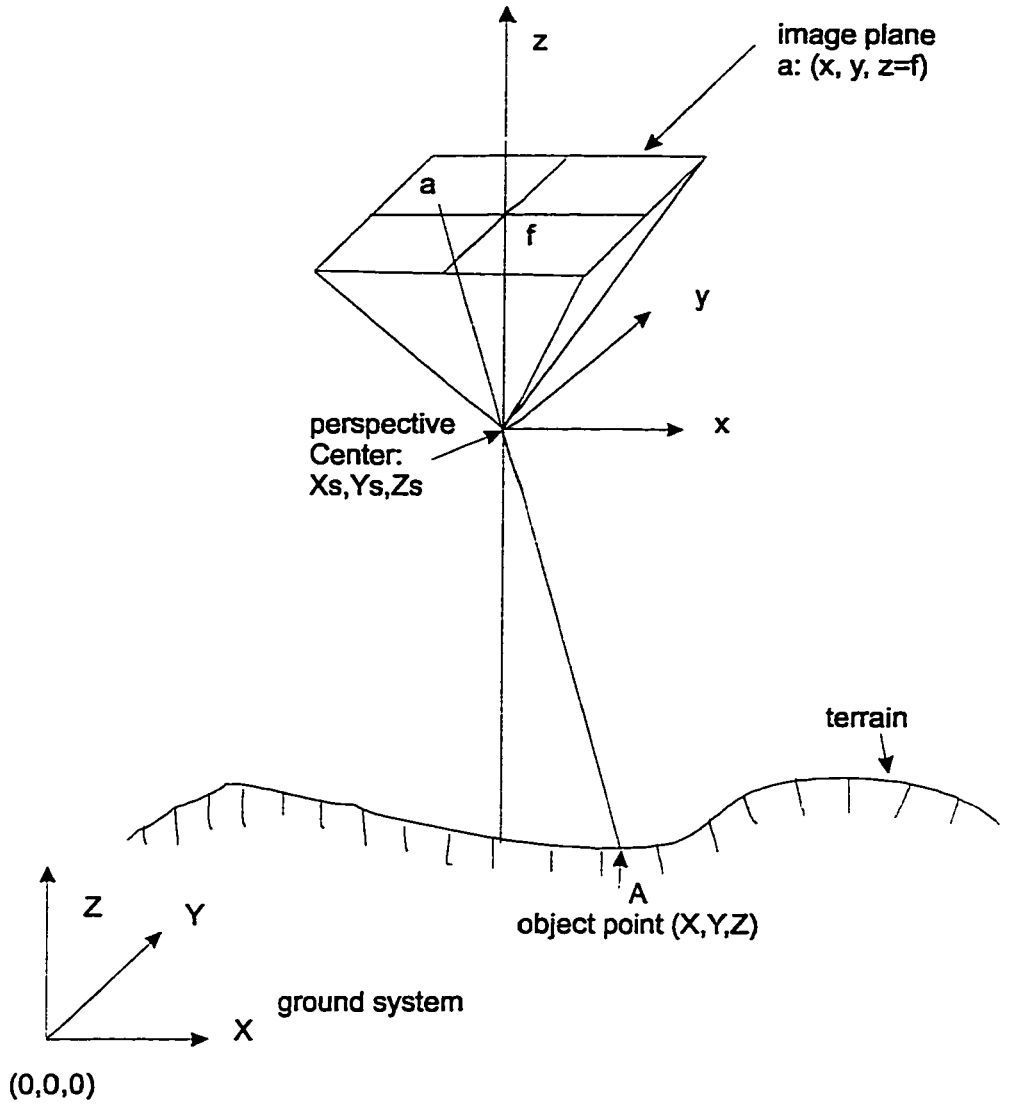


Fig. 3.1 Relationship between Photo and Ground

(Adapted from Chapman, 1992)

They meet the image forming equations of the central projection with orientation matrices:

$$\begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} = f(X^0, Y^0, Z^0, \omega, \varphi, \kappa, X_A, Y_A, Z_A) = \lambda R_\kappa R_\varphi R_\omega \begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} \tag{3.1}$$

where

$$R_\omega = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{pmatrix}$$

$$R_\varphi = \begin{pmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{pmatrix}$$

$$R_\kappa = \begin{pmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} = \lambda R \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} \tag{3.2}$$

where

$$R = R_\kappa R_\varphi R_\omega = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

$$\begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} = \begin{pmatrix} X_A - X^0 \\ Y_A - Y^0 \\ Z_A - Z^0 \end{pmatrix}$$

That is:

$$\begin{cases} x_a = \lambda(r_{11}\Delta X + r_{12}\Delta Y + r_{13}\Delta Z \\ y_a = \lambda(r_{21}\Delta X + r_{22}\Delta Y + r_{23}\Delta Z \\ -f = \lambda(r_{31}\Delta X + r_{32}\Delta Y + r_{33}\Delta Z \end{cases}$$

or

$$\begin{cases} x_a = -f\dfrac{(r_{11}\Delta X + r_{12}\Delta Y + r_{13}\Delta Z)}{(r_{31}\Delta X + r_{32}\Delta Y + r_{33}\Delta Z)} \\ \\ y_a = -f\dfrac{(r_{21}\Delta X + r_{22}\Delta Y + r_{23}\Delta Z)}{(r_{31}\Delta X + r_{32}\Delta Y + r_{33}\Delta Z)} \end{cases} \tag{3.3}$$

For "vertical" photographs flown east-west, the rotation angles

$$\omega = \varphi = \kappa = 0$$

Therefore, the rotation matrix should be an identity matrix, that is,

$$R = R_\kappa R_\varphi R_\omega = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The collinearity equations are as follows:

$$\begin{cases} x_a = -f\dfrac{\Delta X}{\Delta Z} = -f\dfrac{X_A - X^0}{Z_A - Z^0} \\ y_a = -f\dfrac{\Delta Y}{\Delta Z} = -f\dfrac{Y_A - Y^0}{Z_A - Z^0} \end{cases} \tag{3.4}$$

3.1.2 Stereoscopic Parallax

Parallax is the apparent displacement in the position of an object, with respect to a frame of reference, caused by a shift in the position of observation (Wolf, 1983).

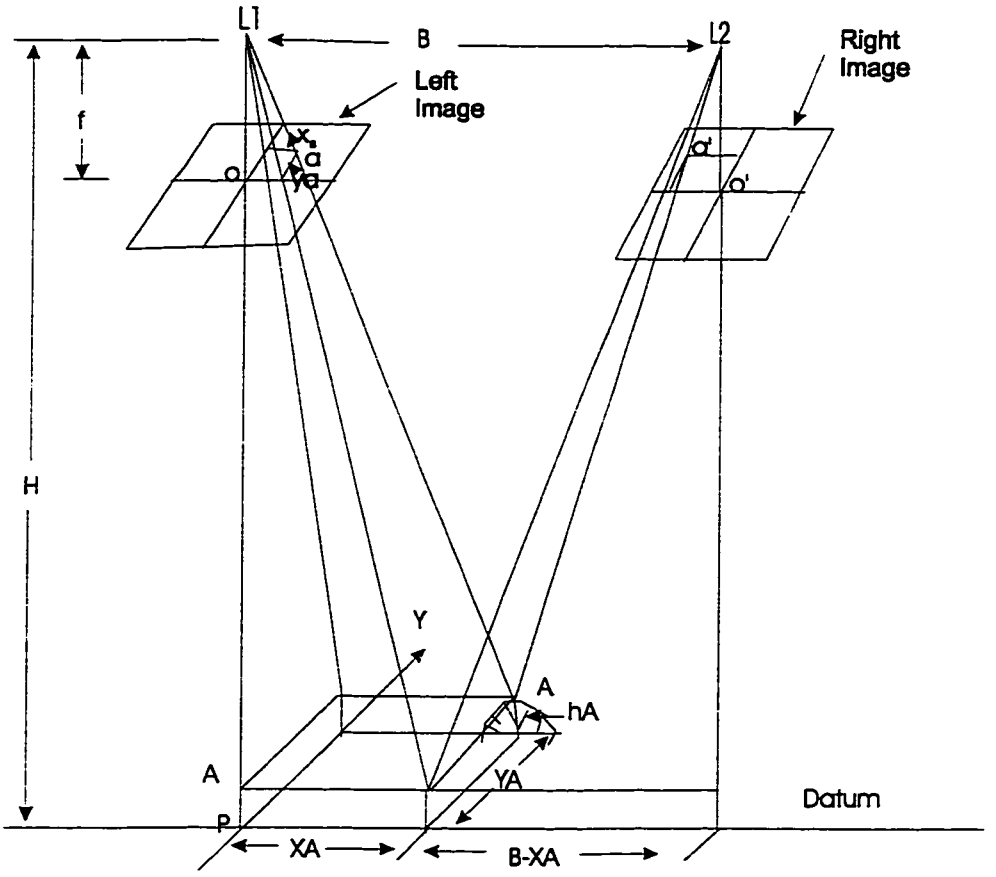


Fig. 3.2. An overlapping pair of vertical photographs above datum

(Adapted from Wolf 1983, p. 170)

Based on the measurements of their parallaxes on the stereopair X, Y, and Z ground coordinates can be calculated. Fig.3.3 illustrates an overlapping pair of vertical photographs which have been exposed at equal flying heights above datum.

$$h_A = H - \frac{Bf}{p_a} \tag{3.5}$$

$$\begin{cases} X_A = B\dfrac{x_a}{p_a} \\ \\ Y_A = B\dfrac{y_a}{p_a} \end{cases} \tag{3.6}$$

In Eqs. (3.5) and (3.6), h_A is the elevation of point A above datum, H is the flying height above datum, B is the air base, f is the focal length of the camera, p_a is the parallax of point A, X_A and Y_A are ground coordinates of point A in the previously defined unique arbitrary coordinate system, and x_a, y_a are the photocoordinates of point a measured with respect to the flight line axis on the left photo.

Equations (3.5) and (3.6) are commonly called the parallax equations. They are among the most useful equations to the photogrammetrist. These equations enable a complete survey of the overlap area of a stereopair to be made, provided the focal length is known and sufficient ground control is available so the airbase B and flying height H can be calculated.

3.1.3 Elevations by Parallax Differences

Parallax difference between one point and another are caused by different elevations of the two points. While parallax Eq. (3.5) serves to define the relationship of stereoscopic parallax to flying height, elevation, air base, and camera focal length, parallax differences are more convenient for determining elevations (Wolf, 1983).

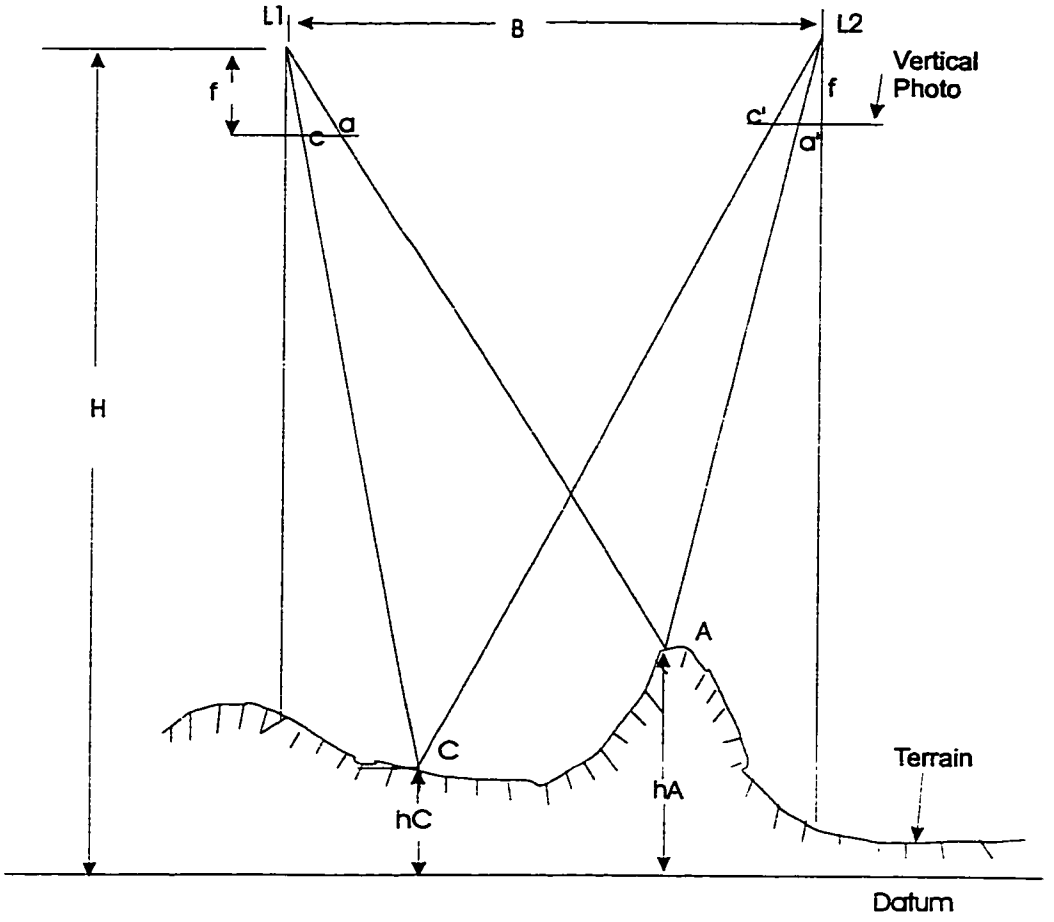


Fig. 3.3 Geometry relationships on the vertical photo pair

(Adapted from Wolf 1983, p. 173)

In Fig. 3.3, object point c is a control point whose elevation above datum is known. The elevation of object point A is desired.

The difference in parallax,

$$p_a - p_c = \frac{fB(h_A - h_C)}{(H - h_A)(H - h_C)} \tag{3.7}$$

Let the parallax differences $\Delta p = p_a - p_c$. The following expression for elevation h_A is obtained:

$$h_A = h_C + \frac{\Delta p(H - h_c)}{p_a} \tag{3.8}$$

The approximate equation for elevation from parallax differences is obtained from Eq. (3.8) by (1) substituting the photo base b of the stereopair for p_a; (2) substituting average flying height above ground, $H' = H - p_c$; and (3) letting $dh = h_A - h_C$:

$$dh = \frac{\Delta p H'}{b} \tag{3.9}$$

In Eq. (3.9), dh is the difference in elevation between two points whose parallax difference is Δp. For moderate relief, parallaxes for all points are approximately equal to b, so that the substitution of b for p_a is valid. Furthermore, if flying height is not extremely low and if relief is moderate, the substitution of average flying height above ground, H' for $H - h_C$ is valid. For very low flying heights or in areas of significant relief, or both, the assumptions of Eq. (3.9) are not met; in these cases Eq. (3.8) should be used. Equation (3.9) is especially convenient in photo interpretation where rough elevations, building and tree heights, etc., are often needed.

3.1.4 Pushbroom Sensor with Line-Scan Imagery

A linear array sensor is composed of a row of CCD elements perpendicular to the orbiting (or flying) direction. It is used in the SPOT [Chapman and Tam, 1990], and in airborne sensors, such as MEIS and CASI.

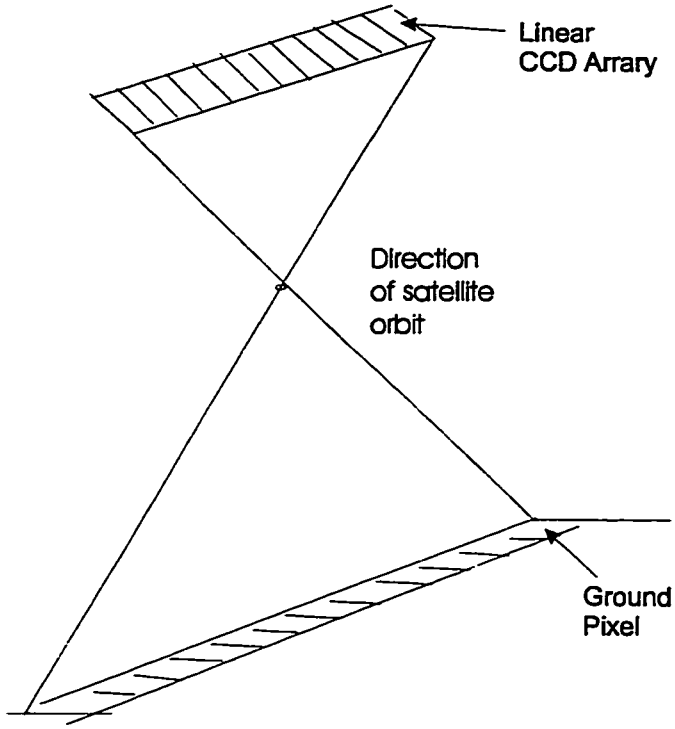


Fig. 3.4 SPOT pushbroom line scan imagery (Adapted from Wang 1990, p. 464)

As shown in Fig. 3.4, the line-scan imagery, the geometric relationship of the line-scan imagery is the same as that in slit photograph. The mathematical relations are as follows:

$$\begin{pmatrix} 0 \\ y \\ -f \end{pmatrix} = \lambda R \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} \tag{3.10}$$

which may be written as:

$$\begin{cases} 0 = -f \dfrac{(r_{11}\Delta X + r_{12}\Delta Y + r_{13}\Delta Z)}{(r_{31}\Delta X + r_{32}\Delta Y + r_{33}\Delta Z)} \\ y = -f \dfrac{(r_{21}\Delta X + r_{22}\Delta Y + r_{23}\Delta Z)}{(r_{31}\Delta X + r_{32}\Delta Y + r_{33}\Delta Z)} \end{cases} \tag{3.11}$$

where the orientation elements and exposure center all refer to corresponding values at the instant t_j, the value of which is computed from the image point coordinate x_j calculated on the basis of the initial time t_0.

3.2 Epipolar Geometry

Determined by the geometry of epipolar lines, homologous points must be located on homologous epipolar lines. Thus by use of the concept of epipolar lines, a two-dimensional correlation problem of searching for homologous points along the x, y directions will be reduced to a one-dimensional correlation problem of searching along epipolar lines (Wang 1990).

3.2.1 Basic geometry of epipolar lines

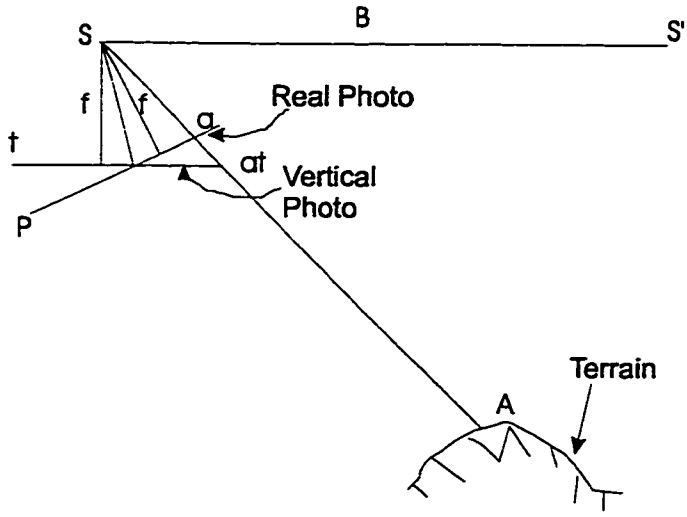


Fig. 3-5. Basic geometry of epipolar lines (Adapted from Wang, 1990, p. 445)

Let it be assumed that Fig.3.5 represents a plane passing through the photographic baseline SS'=B and a certain projected ray SA, namely, the epipolar plane passing through image point a. In the figure, P denoted the left aerial photograph; t represents a normal photograph (ideal photograph) with respect to the photographic baseline, and a is the corresponding image point in the normal photograph.

According to the principles of analytical photogrammetry, the space coordinates u, v, and w of image point a are (see Eq. 3.1):

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ -f \end{pmatrix} \tag{3.12}$$

where $a_1, a_2, \ldots, c_3$ represent the cosines of the angles between the axial lines, which are functions of the elements of orientation of the photograph with respect to the photographic baseline; x, y are the coordinates of the image point on the real photograph; f is the focal length.

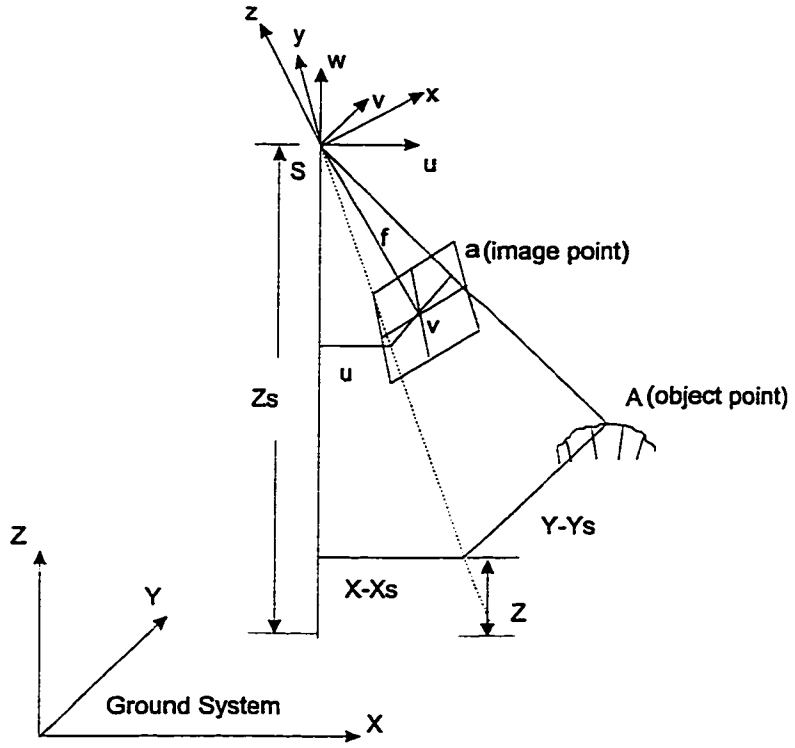


Fig. 3.6 Coordinate systems between image point and object point

(Adapted from Wang, 1990, p. 4)

What we want to solve now are the image point coordinates (u_t, v_t) of the corresponding image point a_t in the normal photograph t. according to Fig.3.6, they should be

$$\begin{cases} u_t = u\left(\dfrac{-f}{w}\right) = -f\dfrac{a_1x + a_2y - a_3f}{c_1x + c_2y - c_3f} \\ v_t = v\left(\dfrac{-f}{w}\right) = -f\dfrac{b_1x + b_2y - b_3f}{c_1x + c_2y - c_3f} \end{cases} \tag{3.13}$$

Or to compute inversely, the following equations are achieved:

$$\begin{cases} x = -f\dfrac{a_1u_t + b_1v_t - c_1f}{a_3u_t + b_3v_t - c_3f} \\ y = -f\dfrac{a_2u_t + b_2v_t - c_2f}{a_3u_t + b_3v_t - c_3f} \end{cases} \tag{3.14}$$

Obviously, in the P plane the epipolar lines are convergent, whereas in the t plane, the epipolar lines are parallel with one another and also parallel with the x coordinate of the photograph, as shown in Fig.3-7 (a) and (b) respectively. For "vertical" photographs flown east-west, we can assume $\omega = \varphi = \kappa = 0$, so that we can consider the "vertical" photographs as normal photographs, that is, the epipolar lines are parallel with x coordinate on the original photograph.

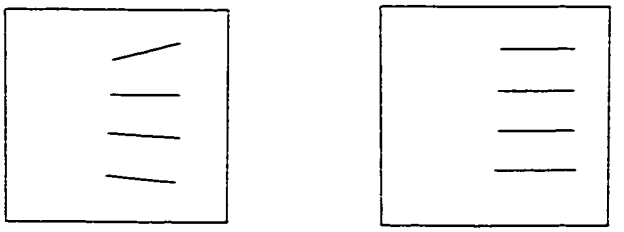


Fig. 3.7. Epipolar lines on real photograph and vertical photograph

(Adapted from Wang , 1990, p. 446)

3.2.2 Photo Coordinates and Scanning Coordinates

In all theoretical discussions, the coordinates are taken in the image coordinate system. Since in the process of digitization by photo scanning, the axes of the scanning coordinate system are generally not parallel to those of the image coordinate system. Therefore, it is necessary to perform transformation from the scanning coordinates $x^{'}, y^{'}$ to the image point coordinates x, y. For this purpose, we may use the following equations of affine transformation:

$$\begin{cases} x = h_0 + h_1 x^{'} + h_2 y^{'} \\ y = k_0 + k_1 x^{'} + k_2 y^{'} \end{cases} \tag{3.15}$$

The parameters $h_0, h_1, h_2, k_0, k_1, k_2$ in the above equations can be computed by adjustment using the error equations formulated from the known image coordinates of the fiducial marks of the photograph (or other marks) and their corresponding scanning coordinates.

In order to avoid the difficulty that the point (x, y) is not on the grid point, we can use the inversion equations of Eq. (3.15). That is:

$$\begin{cases} x' = h_0' + h_1' x + h_2' y \\ y' = k_0 + k_1' x + k_2' y \end{cases} \tag{3.16}$$

The grey values of the new pixels can be obtained through interpolation (resampling techniques) of the known grey values of its neighboring pixels.

3.2.3 Image Coordinates in Homologous Epipolar Lines

A basic problem involved in arranging the grey levels of pixels in the direction of epipolar lines is to start from an arbitrary image point on the photograph (say, point a on the left photograph in Fig. 3.8), and solve for the positions of the other image point b on the epipolar line of the left photograph and the two points a' and b' on the homologous epipolar line of the right photograph.

According to the coplanarity condition of the three lines SS', Sa, Sb, we have

$$SS' \bullet (SaXSb) = 0 \tag{3.17}$$

We obtain:

$$\begin{vmatrix} B & 0 & 0 \\ u_a & v_a & w_a \\ u_b & v_b & w_b \end{vmatrix} = B \begin{vmatrix} v_a & w_a \\ v_b & w_b \end{vmatrix} = 0 \tag{3.18}$$

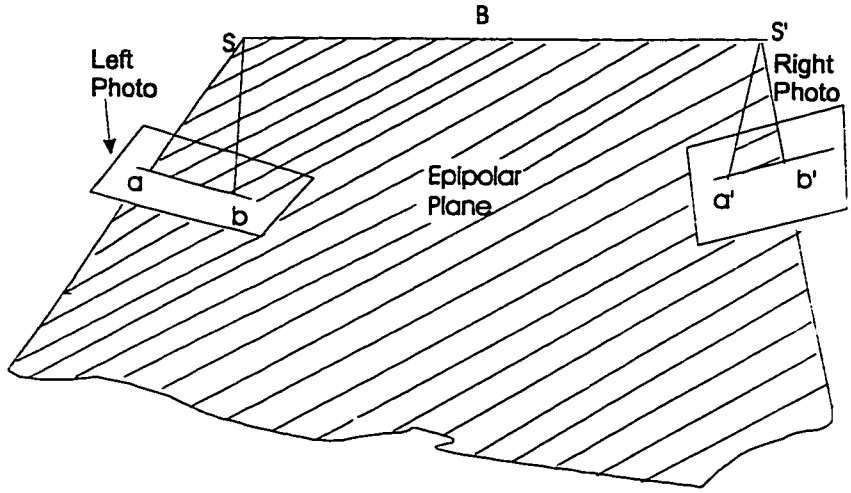


Fig. 3.8 Epipolar plane (Adapted from Wang , 1990, p. 446)

The space coordinates u, v, w of an image point in the above equation are obtained from its image coordinates x, y, and the image coordinates can be derived from its scanning coordinates x', y' by computation of Eq. (3.16). That is to say, by means of Eq. (3.18) and other relations, we can, with the knowledge of the scanning coordinates x', y' of an

arbitrary image point a, reasonably assume any $x_b^{'}$ coordinate of image point b, say $x_b^{'}$=100mm, whence the corresponding $y_b^{'}$ can be obtained.

Similarly, if we assume arbitrary values of the scanning coordinate x' of two points a' and b' on the homologous epipolar line on the right photograph, say, $x_a^{'}=0, x_b^{'}=100mm$, the other two coordinates $y_a^{'}$ and $y_b^{'}$ of the two points can also be obtained by use of the orientation elements of the right photograph in the same manner.

3.3 Image Orientation

3.3.1 Overview

The procedure of orientation includes interior, relative and absolute orientation. For the computation of the orientation elements it is necessary to know the x and y coordinates in both fiducial marks and those of some selected pass points as well as those of a number of ground control points. The goal is to find the coordinates of those points semi-automatically through 2-D image correlation. The main contents are follows:

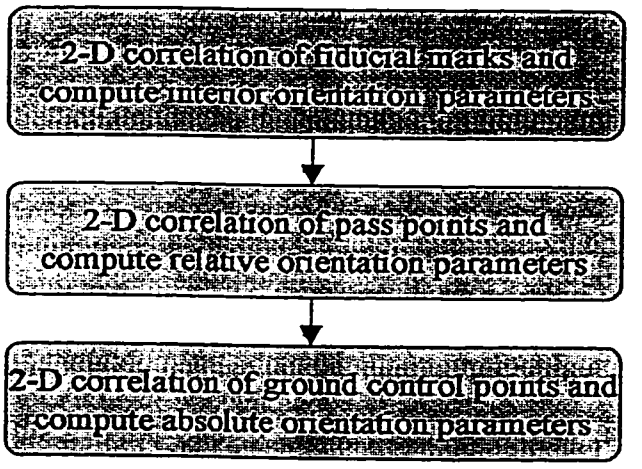


Fig. 3.9 Flow chart of image orientation

3.3.2 Interior Orientation

Measurement of photographs is a very important step in establishing analytical aerotriangulation (Wang, 1990). Usually the measurement is performed on a stereocomparator or a mono-comparator and is oriented by means of fiducial marks. Depending on the structure of an instrument, readings on a stereocomparator may be (x_L, y_L, x_R, y_R) or (x_L, y_L, p, q) or (x_R, y_R, p, q) and the final results are all (x_L, y_L, x_R, y_R). Their relationships are:

$$\begin{aligned} x_R &= x_L - p \\ y_R &= y_L - q \end{aligned} \tag{3.19}$$

The coordinates x_R, y_R of an image point on the right photo of a stereopair should be equal to the coordinates x_L, y_L of the left photo of the next stereopair, and the difference between their corresponding measured valued can be utilized to check the precision of the measurements.

Errors involved in setting and orienting photos in a comparator will lead to deviations between the measured coordinates of the instrument and their corresponding coordinates of the image point. These kinds of errors can be partially eliminated through reduction by means of measuring fiducial marks of the photograph, as is the case with the second and the third schemes of photo setting and orientation.

Other systematic errors of photo coordinates caused by film distortion lens distortion, atmospheric refraction, Earth curvature etc. should be taken into consideration.

3.3.3 Relative Orientation

One of the basic principles underlying stereo photogrammetry is to create a stereo model with two overlapping images. This principle can be applied to a photogrammetric network consisting of a series of image pairs as well as an independent pair. Since relative orientation of photos can be determined through the condition that the two homologous rays in a pair intersect each other, it has nothing to do with ground control points. In the process of relative orientation, the image pair can be placed in an arbitrary absolute orientation. The two cases shall be elaborated on that are commonly used in

relative orientation, i.e., (1) the independent photo pair and (2) the consecutive photos. Relative orientation is based on the coplanarity condition and needs at least 5 tie points to solve for 5 unknowns.

3.3.4 Absolute Orientation

The inverse computation for the exterior orientation elements, $X^0, Y^0, Z^0, \omega, \varphi, \kappa$ of a photographic bundle of rays when the space coordinates of some ground points are known is a problem of single-image resection in space. When it is used for the calibration of a camera or for some other precise measurements, the interior orientation elements $x_0, y_0,$ and f can also be derived simultaneously (Wang, 1990). Normally, there are 7 unknowns for frame camera images, 3 translations, 3 rotations, 1 scale in order to orient a photograph reference to the ground. If you consider 3 interior orientation elements, 10 unknowns should be solved by ground control points.

3.3.5 Utilization of GPS Data in Block Bundle Adjustment

It has been proved (Ackermann 1988) that the internal positioning accuracy by the Satellite Global Positioning System (GPS) is very high, and can reach 1cm or less. All the large errors that have occurred are systematic errors and errors caused by different kinds of deviation. We can take the parameters of these errors as unknowns in accordance with their theoretical regularities and add them to the combined block bundle adjustment

solutions of photogrammetric data. In this case, each strip should be given different parameters to be determined.

Experiments of combined block adjustment with GPS determined camera stations have been carried out by computer simulation (Fries, 1986). It has been shown that the camera positions as determined by GPS can be used as ground control points, potentially replacing conventional geodetic ground control points almost completely. The GPS test "Flevoland" conducted by institutions of F.R.Germany and the Netherlands (Ackermann 1988) consists of a well-controlled photogrammetric block of large scale (1:3800). The block was aerotriangulated, giving the "true" camera stations (accurate to 4.3cm, 4.2cm, 1.9cm, standard deviations in X, Y, Z respectively) to be compared with GPS-kinematic positioning. The positioning was based on two receivers, one stationary on a known point on the ground, the other in the aircraft. The comparison of the GPS coordinates of the GPS antenna stations with the "true" camera air stations was done on several levels; the result from the highest level showed that by fitting the GPS positions into the camera stations by linear transformation (linear regression) the remaining r.m.s. coordinate differences of all the strips (a block covering 7 strips in this experiment) were 5.2cm (or 3.8cm after subtracting the photogrammetric error). When the procedure was repeated using only the GPS receiver in the aircraft, i.e., without considering the stationary receiver, the corresponding result was 6.0cm (or 4.8cm after subtracting the photogrammetric error).

CHAPTER FOUR

SIGNAL MATCHING

4.1 Correlation Basics

Signal matching was the earliest and simplest method introduced in the image matching process. It is based on the image grey shade values only. Grey value correlation employs statistical, e.g. covariance values, as similarity measures. This method has been very well researched.

4.1.1 The principle of correlation

As shown in Fig. 4.1, A and B represent the portion of a stereopair on the images, and there exist parallaxes in the small areas.

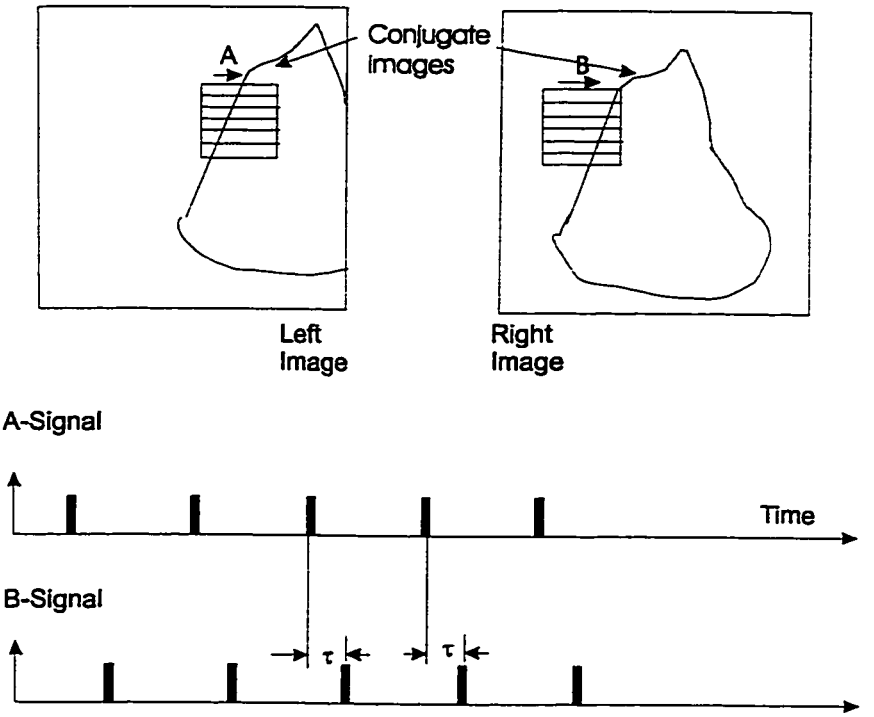


Fig. 4.1 Signal A and B represent the portion of stereopair on the images

(Adapted from Wang , 1990, p. 246)

Since the A on the left image is shifted more to the left compared with its conjugate part B on the right image, signal B will be slightly more delayed than signal A. It can be seen from the graph of output signals that there is a constant τ of time delay between two output signals. The correlation function is to determine automatically this time difference. When the delay time is just equal to the time difference between the signal A and signal B, then the output from the correlation function is the largest. The change of the delay of the output signal from the correlation is described by a curve called the curve of correlation function. The simplest form of the curve can be illustrated by the simplest signal in Fig. 4.2 (Wang, 1990)

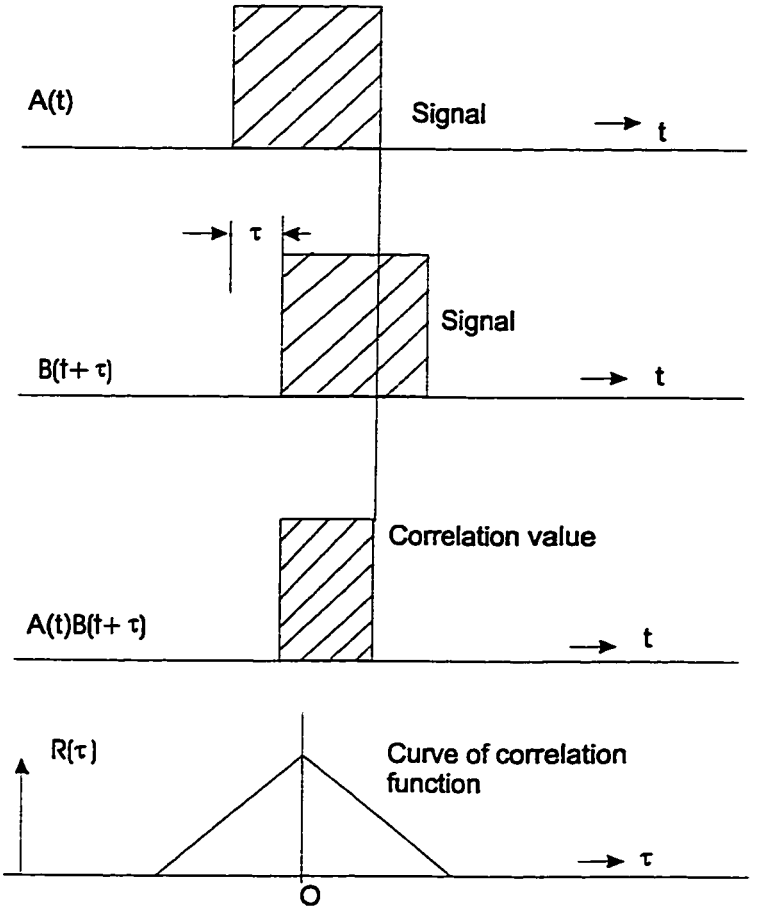


Fig. 4.2 Signal A and B correlation curve (Adapted from Wang , 1990, p. 247)

It can be seen that when the delay τ of signal B is equal to zero, then the correlation value $R(\tau)=A(t)B(t+\tau)$ is a maximum. When the delay τ changes from zero to the width equivalent to that of the signal, then the maximum of $R(\tau)$ reduces linearly to zero. Thus the shape of the curve of the correlation function is shown as in Fig. 4.2. Since in reality the shape of an image is complex and the density of an image changes gradually, the curve of its correlation function usually takes the shape shown in Fig. 4.3.

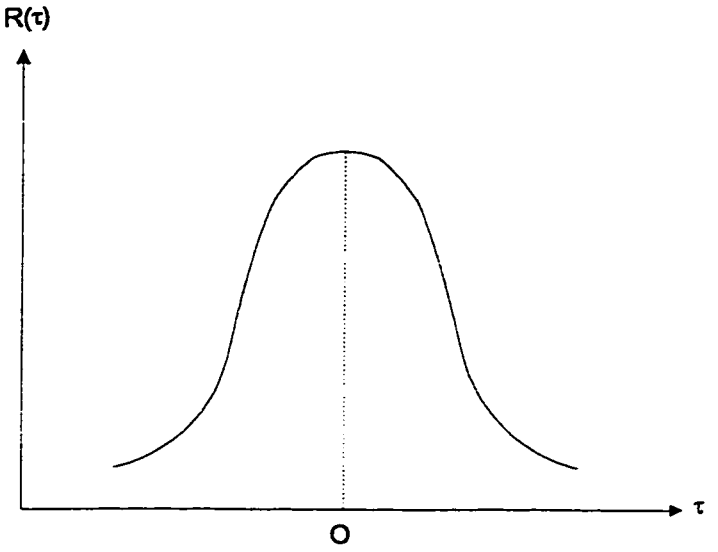


Fig. 4.3. The real correlation curve (Adapted from Wang, 1990, p. 247)

4.1.2 The correlation function

The correlation function of two signals A and B, now replaced by symbols $x(t)$ and $y(t)$, are expressed as follows:

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)y(t+\tau)dt \tag{4.1}$$

Theoretically, the definition of the correlation function states that T should be infinite. Obviously, this cannot be realized in actual measurements. It is then necessary to change the infinite length of data into a finite one, namely, T is large to a reasonable degree so

that the statistical variance thus formed is small to an acceptable extent. The practical equation for use can be written as:

$$R_{xy}(\tau) = \frac{1}{T}\int_0^T x(t)\,y(t+\tau)dt \tag{4.2}$$

The above correlation function is a cross correlation, which refers to the correlation between two signals $x(t)$ and $y(t)$. When two random signals are identical, that is, when $x(t) = y(t)$, then the above equation becomes:

$$R_{xx}(\tau) = \frac{1}{T}\int_0^T x(t)x(t+\tau)dt \tag{4.3}$$

which is called the autocorrelation function.

The problems to be dealt with in digital correlation in photogrammetry are concerned with homologous points or lines on two photographs. Therefore, the correlation in our discussion always refers to cross correlation. However, since the curves of grey levels of homologous points or lines (conjugate epipolar lines) are similar to each other, the investigation of the autocorrelation function may also provide a basis for the rational design of digital image correlation systems and to make estimations about such problems as the sampling interval, the image resolution, and the sharpness of a correlation function, etc.

4.1.3 The maximum value of cross correlation

In photogrammetry the grey value distributions of two correlated images are dissimilar to different degrees. Therefore, the correlation is always cross correlation. The main reasons for the dissimilarity between the grey values of correlated images are:

(a). With respect to radiometric differences: radiation from objects related to the spatial direction where it is located, degeneration caused by the atmosphere and the imaging system, differences in film processing, possible differences in grey shade values, etc.

(b). With respect to geometric differences: differences between the elements of camera orientation, image displacements due to terrain relief, different image distortions, etc. In terms of high precision correlation, these factors should be sufficiently compensated for in advance.

Let us assume that the images of two points A and B on the ground are A_1, B_1 and A_2, B_2 (Fig. 4.4) on the left (x_1) and right (x_2) photographs respectively. The effect of photo tilt, and particularly that of terrain slopes results in $A_1B_1 \neq A_1B_1$, the difference of which is the horizontal parallax Δp within the photo pair.

Now let us search for the conjugate point A_2 on the right image using the point A_1 on the left image by means of one-dimensional correlation.

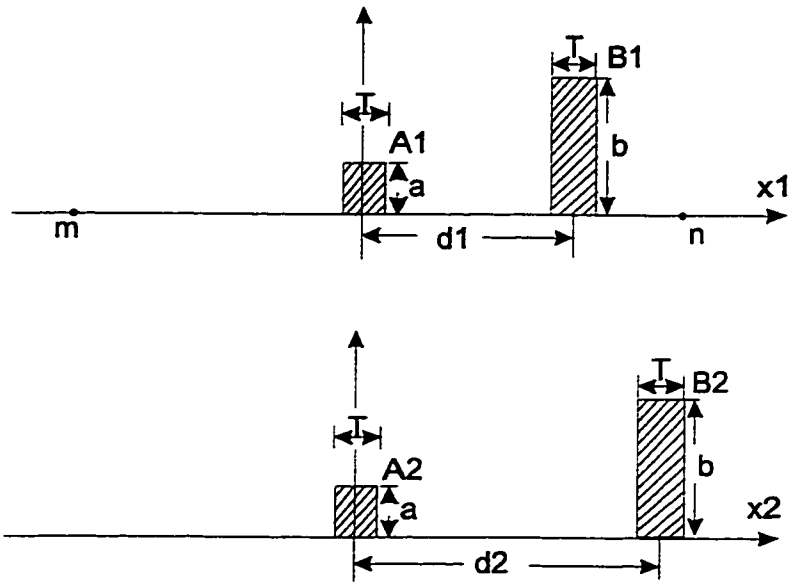


Fig. 4.4 Correct position of correlation searched for

(Adapted from Wang , 1990, p. 439)

Suppose we first take a target area (Fig. 4-4), i.e., a grey value array surrounding a specific point (A_1 in this case) symmetrically in one image, to search for its conjugate grey value array in the search area on the right image. Apparently, the relative position shown in Fig. 4-4 is the correct position of correlation searched for with respect to the ground point A. Here $\tau = 0$, and a peak value occurs. The value of its correlation function should be, according Eq. (4.2):

$$R_{xy}(0) = a^2$$

However, in the process of searching for the position of correlation, when the target area moves further to the right until the two images B_1 and B_2 coincide, another peak value will result:

$$R_{xy}(d_2 - d_1) = b^2$$

Since the signal is stronger at point B (here, assuming b>a), the last peak value is mistaken for the position of correlation, and the resulting error of correlation is $d_2 - d_1$. However, if the target area (m, n) is smaller $2d_1$, then the image point B_1 is already excluded from the target area, and we can certainly obtain correct correlation result. Nevertheless, if the target area is too small, then the sample has a very limited amount of information, and hence the reliability of the obtainable correlation coefficient will be decreased accordingly.

4.2 Area-based Signal Matching

In area based signal matching, neighborhoods of pixels are used. The similarity is indicated by the resemblance between grey values.

Cross correlation is grey value based correlation which employs statistical covariance measures as similarity measures. Taking a target area in one image (normally the left image), and a search area on another image (normally the right image), we shift the same

size 2-D window as target area over the search area on the right image, and then compute the correlation coefficient (Fig. 4.5). The center pixel of the target area and that of the most similar search area are taken as corresponding points when the correlation coefficient is the highest among the all search area. The similarity measure should exceed a threshold. The above measures have to be maximized to find the best correlation.

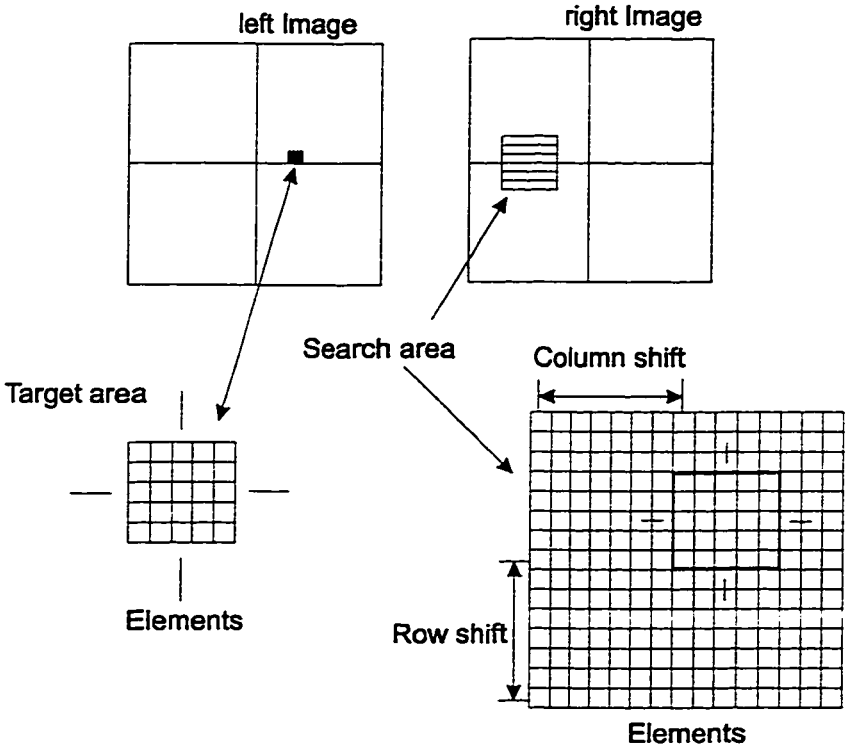


Fig. 4.5 Area-based signal matching (Adapted from Wang, 1990, p. 435)

Matching in this case is indicated by the degree of similarity of grey level between the target object and search object. Covariances within searching windows are used to calculate the correlation coefficient. The value of the correlation coefficient ranges between +1 to -1, where +1 means, theoretically, the maximum match. The maximum

correlation coefficient practically obtained indicates the best matching between the search and target images. The acceptability of a match, however, is decided according to a threshold imposed on the correlation coefficient. If the correlation is higher than the threshold then matching is successful, otherwise it fails. Deciding the value of the threshold is largely a subjective matter, depending on experience, image quality, etc.

Matching with correlation is implemented in the spatial-radiometric space or in the frequency domain. For the latter, some appropriate transformation is used, such as a Fourier transform.

4.3 Correlation Computation Method

Following are some correlation computation methods which are often used in signal matching (Wang, 1990).

4.3.1 Covariance function

The values of the covariance function of the two data groups in correlation are computed and the largest covariance obtained is taken as the position of correlation being searched for.

Assuming that x_i represents the grey level of a point i in the first data group, and y_i represents the grey level of its corresponding point in the second data group, then the

averages $\bar{x}$ and $\bar{y}$ of grey levels of the two data groups are respectively (averaged over grey levels of n points for each data group):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{4.4}$$

The variances σ_{xx}, σ_{yy} of the two data groups are respectively:

$$\sigma_{xx} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad \sigma_{yy} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{4.5}$$

Or after reduction, they may be written as:

$$\sigma_{xx} = \frac{1}{n}\sum_{i=1}^{n} (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \frac{1}{n}(\sum_{i=1}^{n} x_i^2) - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n} (x_i) + \bar{x}^2 \tag{4.6}$$

Namely,

$$\sigma_{xx} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2, \qquad \sigma_{yy} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 \tag{4.7}$$

The covariance σ_{xy} of the two data groups is:

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy} \tag{4.8}$$

Thus the correlation relationship between images can be expressed by the covariance function expressed by Eq. (4.8). A covariance function is similar to the correlation function except that a covariance function demands that the averaged grey levels $\bar{x}$ and $\bar{y}$ should be subtracted from their signals x_i, y_i. The covariance function and the correlation function have the same characteristics. Since the average values have been eliminated from the signals in the former case, the correlation error due to the difference between the average grey levels of the two images can be compensated for. Therefore, the use of covariance function should be better than that of the correlation function.

4.3.2 Correlation Coefficient

A correlation relationship may also be represented by a correlation coefficient ρ or its square ρ^2. The correlation coefficient is defined as:

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_{xx}\sigma_{yy}}} \tag{4.9}$$

In the two-dimensional case, the correlation equations Eqs. (4.4) through (4.8) become respectively:

$$\bar{x} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}\,, \qquad \bar{y} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}$$

$$\sigma_{xx} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}^2 - \bar{x}^2, \qquad \sigma_{yy} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}^2 - \bar{y}^2 \tag{4.10}$$

$$\sigma_{xy} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} y_{ij} - \overline{xy}$$

The correlation coefficient ρ is a normalized covariance function. A covariance function provides a relative criterion for comparison. Since it is subject to interference of variance changes of an image, it would be more reasonable to make it into a correlation coefficient through normalization.

The value of the correlation coefficient is variable between +1 and -1. The closer its value is to 1, the stronger the correlation between the two signals, i.e., the more similar they are to each other. Then it follows that what they correspond to are very likely homologous objects.

4.3.3 Weighted cross correlation coefficient

The procedure is the same as above except that the weights w_{ij} are added. In the two-dimensional case, the correlation equations Eqs. (4.10) become respectively:

$$\bar{x} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_{ij}, \qquad \bar{y} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} y_{ij}$$

$$\sigma_{xx} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_{ij}^2 - \bar{x}^2, \quad \sigma_{yy} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} y_{ij}^2 - \bar{y}^2 \tag{4.11}$$

$$\sigma_{xy} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_{ij} y_{ij} - \overline{xy}$$

The correlation coefficient ρ is a weighted cross correlation coefficient. And you can give higher weights for central pixels to favor the center and give lower weights for further pixels away from the center to decrease the influence of the edge pixels. Because the ordinary correlation coefficient does not include any geometry information of the matched windows, introducing the weight factor can add some geometry information into the distribution of the pixels in the windows indirectly.

4.3.4 Absolute values of grey level differences

This is not a scheme of correlation, but a substitute method for the comparison of two images for the degree of their similarity. The method is first to eliminate the average grey values $(\bar{x}, \bar{y})$ from the values (x_i, y_i) of their respective elements of each of the two data groups, and then compute the grey value differences between the corresponding elements and add their absolute values, i.e.,

$$d_1 = \sum_{i=1}^{n} \left| (x_i - \bar{x}) - (y_i - \bar{y}) \right| \qquad (4.12)$$

Then after translating one data group in relation to the other, repeat the above operation until the place is found where the sum is a minimum. This should be the place where the

maximum correlation is obtained. The amount of computational work involved in this method is comparatively small, and useful results can also be obtained.

4.3.5 The squares of the grey level differences

The procedure is the same as the above except that the squares of the grey level differences are added, i.e.,

$$d_2 = \sum_{i=1}^{n} ((x_i - \bar{x}) - (y_i - \bar{y}))^2 \tag{4.13}$$

and the place where its value is a minimum is searched for.

4.3.6 Grey value differences of neighbouring pixels

In this method, a difference treatment of grey values is performed at first on the original digital array of images. For example, we have:

The first difference: $\Delta x_i = x_i - x_{i+1}$

Or the second difference: $\Delta^2 x_i = \Delta x_i - \Delta x_{i+1}$

Thus substituting these differences (either the first or the second ones) for the original grey values, we can proceed with the solution for the maximum value of the correlation function, i.e.,

$$R_{\Delta x \Delta y} = \sum_{i=1}^{n-1} \Delta x_i \Delta y_i = max \tag{4.14}$$

Or for the minimum value of the sum of the absolute values of the grey value differences (the differences of differences in this case), i.e.,

$$d_{\Delta x \Delta y} = \sum_{i=1}^{n-1} \left| \Delta x_i - \Delta y_i \right| = \min \tag{4.15}$$

4.4 Two-Dimensional and One-Dimensional Correlation

Since digital image is a two-dimensional record of data, image correlation is essentially a two-dimensional task. However, stereo image correlation can be reduced to a one-dimensional epipolar line case, and thus the computational work involved can be greatly reduced. For general photographic images, we need to take two steps. In the first step, a sufficient number of well distributed orientation points in the overlapping area of the photographs are chosen, so that their vertical parallaxes are determined by means of two-dimensional correlation and then the parameters of relative orientation can be computed. The second step is to compute the geometric relationships of epipolar lines between the image planes. In the subsequent point determination by correlation, we may use the correlation within the corresponding conjugate epipolar lines and thus reduce the correlation procedure into a one-dimensional search. Before we perform the one

dimensional image correlation, we need to resample the pixels along epipolar lines for both the left and right images.

4.4.1 Two-Dimensional Correlation

An image is a two-dimensional representation of reflectance or intensity. In an overlap area, matching is basically a two-dimensional operation. The correlation of two continuous image functions can be represented as follows (Saleh and Scarpace, 1992):

$$C(x,y) \Rightarrow f(g,h) \tag{4.16}$$

where C: correlation coefficient

g,h: image function

f: function of

$$C(x,y) = \iint g(T+x, S+y)h(T,S)\,dT\,dS \tag{4.17}$$

where x, y: window center in search image space

T,S: coordinate in search window space

dT,dS: searching increments

This equation assumes continuous image function. Digital imagery, however, are discrete

functions. Hence, correlation equation can be written as follows:

$$C(x,y)=\sum\sum g(T+x,S+y)h(T,S) \tag{4.18}$$

where g and h are the discrete pixel values in both images. For simplicity, equation (4.18) can be written as:

$$C=\sum\sum gh \tag{4.19}$$

which can be normalized by the means of the samples, yielding

$$C=\sum\sum(g-g_m)(h-h_m) \tag{4.20}$$

where g_m: mean of g

h_m: mean of h

or by the second moments of the samples,

$$C=\sum\sum gh/(\sum g^2\sum h^2) \tag{4.21}$$

or by both,

$$C=\sum\sum(g-g_m)(h-h_m)/(\sum\sum(g-g_m)^2\sum\sum(h-h_m)^2)^{1/2} \tag{4.22}$$

In this case, correlation has an absolute value less than or equal to one. Although more expensive, normalized correlation is designed to handle first, a constant offset, by subtracting the means, and second, a constant gain, by dividing by the variances.

The correlation approach has been attractive to many researchers. Extension and optimization to one dimension is an important development.

4.4.2 One-Dimensional Correlation (Epipolar Line Correlation)

The search time required for matching is an important factor with respect to computer implementation of the technique.

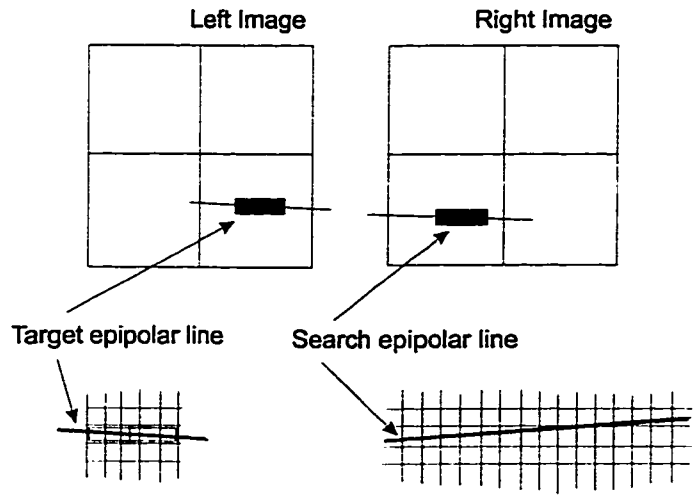


Fig. 4.6 Epipolar line correlation (Adapted fromWang, 1990, p. 435)

The most important concept of epipolar geometry to be realized in image matching is that conjugate imagery is always found along conjugate epipolar lines (as shown in Fig. 4.6), regardless of photographic orientation. So points situated on epipolar lines are void of y-parallax. What remains along any epipolar line is x-parallax, which can determine elevation. Based on the character of epipolar geometry, when a pair of epipolar lines are known, the conjugate point in one image can be found in the conjugate epipolar line on another image. In this case, the two-dimensional search of a conjugate image point reduces to a one-dimensional search. It improves both efficiency and reliability in matching calculation. Usually, the direction of scanned lines can be set to be parallel to epipolar lines. Sometimes, it needs geometric transformation and image resampling to meet this requirement. In this case, we must have knowledge of the relative orientation parameters. It makes the process of matching much more convenient and efficient. Furthermore, the pair of images are normalized to have approximately the same mean shade.

In photogrammetry, the epipolar condition allows reducing the search from two dimensions to one dimension. The one-dimensional form of (4.17) is:

$$C(x) = \int g(T + x)h(T)dT \tag{4.23}$$

The discrete form is:

$$C(x) = \sum g(T+x)h(T) \tag{4.24}$$

The final normalized one dimensional correlation is:

$$C = \sum(g - g_m)(h - h_m) / (\sum(g - g_m)^2 \sum(h - h_m)^2)^{1/2} \tag{4.25}$$

The epipolar search saves considerable time, but assumes that the two images are already relatively oriented.

4.5 Least-Squares Matching (LSM)

The LSM takes into account geometric distortion and radiometric shifts between the left image chip and right image chip. In this approach, the intensity differences between the target image and the search image are treated as noise. The LSM method is then applied to find the minimum sum of the squares of the noise. The location of this minimum represents the best match. The general function is:

$$n(x, y) = g(x, y) - h(x, y) \tag{4.26}$$

where n: noise or difference in pixel values

While the cross correlation compensates for the radiometric differences within the algorithm itself, LSM requires normalization before the matching takes place. This compensation is performed by geometric and radiometric transformations.

$$\begin{cases} x = a_1 + a_2 x' + a_3 y' \\ y = b_1 + b_2 x' + b_3 y' \\ g_t(x,y) = r_1 + r_2 g_s(x,y) \end{cases} \tag{4.27}$$

where a,b: geometric transformation coefficients

r: radiometric transformation coefficients

g_s: geometrically transformed image

g_t: normalized image

Several extensions in this approach have been developed as a result of the flexibility of LSM (Forstner, 1982; Halave, 1988; Rosenholm, 1987b; Wrobel,1988). Various constraints are introduced in these extensions to handle featureless areas, and to reduce the effects of slope, curvature, etc. With sufficient iterations, a window size big enough is required to cover the target point. Rosenholm (1987a) reported that 20X20 and 30X30 are optimal window sizes for precision. On the other hand, nonconvergency would result in infinite iterations. To avoid this situation, accurate initial approximation is imperative. Such a requirement is a disadvantage of LSM. Experiments conducted by Chen (1989) show that higher success rates can be attained by LSM, with a trade-off in terms of speed.

4.6 Discussion of Area-based Signal Matching

Signal matching suffers from the following limitations:

(1) the presence of detectable texture or edges is required; in areas with a smooth grey value function no optimal match will appear;

(2) repetitive micro structures will cause several equally likely matches;

(3) linear edges will cause many pronounced matches along the edges;

(4) surface discontinuities can not be handled;

(5) the target area may have no counterpart in the search space because of occlusion;

(6) they are computationally expensive;

(7) they are not rotational and scale invariant;

Ambiguities in the match result caused by low structure contents and linear edges may be avoided by examining the target area in advance. For instance, the presence of pronounced texture or corners indicate suitable targets. Repetitive structures will cause a sequence of good correlation values. We should select target areas by examining the maximum grey value variances in four or more directions.

Computation time may be reduced by:

(1) Limitation of the search space by:

- the epipolar geometry reduces the 2-D search space to 1-D;
- general knowledge about the object space;
- prediction of the next point from previous matches;

(2) Coarse-fine correlation by a multiresolution approach, matches at higher levels, guide match examination at low levels;

(3) Utilization of less costly similarity measures, like sum of the absolute values of differences.

LSM can reach precision up to 1/20 sub-pixel accuracy. The disadvantages of LSM are:

(1) it needs a time-consuming resampling;
(2) it needs very accurate approximations.

Due to LSM requiring an approximate surface model, the approximate values have to be very precise, even such that for many applications the approximations are good enough as the final matching result. So, LSM is actually a fine correlation method to bring the known matches to a higher precision level.

CHAPTER FIVE

FEATURE MATCHING (INTEREST OPERATOR)

5.1 Overview

Contrary to signal matching procedures, features are detected in both images, leading to the primal sketch. Features can be points, lines and areas. The border detection operator for line feature extraction is often used in the techniques in digital image processing. Point features such as definite points, corner points have rather high accuracy which might meet image matching requirements. The operator for point feature extraction is usually called the interest operator.

5.2 Concept of Feature Based Matching

Area-based cross-correlation cannot respond appropriately to a number of facts that are related to 3D objects. The conjugate images created under the laws of perspective projection might differ considerably from each other. Cross correlation works fast and well if the patches to be matched contain enough signal without too much high frequency content, and if geometrical and radiometric distortions are kept at a minimum. Both conditions are often not encountered in standard aerial photos. Although grey-level-based correlation methods can reach very high accuracy in matching (say, 0.1 pixel), wrong

correlation may possibly occur. Enlightened by the experience of the human visual system in viewing objects which goes from global to local, and from the general to the detailed, people set up feature-based matching research from the extraction of image features so as to improve the reliability of the matching, reduce man-machine intervention, and increase the efficiency of automation. Therefore, great efforts have been made in recent years to apply feature based matching that are more able than signal matching (Forstner, 1986).

Feature based matching (FBM) procedures contrast to grey level or area based methods, like classical image correlation or least squares matching (LSM). FBM is superior to image correlation with respect to speed and versatility and is superior to LSM with respect to range of convergence, speed and versatility. Especially the high requirements for approximate values of LSM. FBM procedures are widely used in pattern recognition and computer vision and find increasingly interest also in photogrammetry (Lam, 1990; Larouche, 1995).

There are three stages in feature matching:

- selection of distinct features by an interest operator (distinction check);
- selection of candidate features which may form possible matches, using one or more similarity measures (similarity check);
- thinning of the list of candidate points, until unique matches remain, consistent with an object model, i.e. determination of the correct matches (consistency check).

5.2.1 Selecting Distinct Points with an Interest Operator

In FBM instead of matching all pixels in an image, only selected points with certain features are to be matched. The selection principle should fulfill the following requirements (Forstner, 1988):

(1) Distinctness: The points should be distinct, i.e. be different from neighboring points. e.g. points on edges should not be selected if the epipolar geometry constraint is not used; also points in flat areas should not be selected.

(2) Invariance: The selection as well as the selected position should be invariant with respect to the expected geometric and radiometric distortions. This, besides the distinctness, probably is the most important requirement. The degree of invariance directly influences the precision and the reliability of the matching

(3) Stability: The selected points should be expected to appear in the other images. Thus the selection should be robust with respect to noise. In image sequence analysis the selected points should appear in long sequences of consecutive frames.

(4) Seldomness: Where distinctness guarantees local separability of points seldomness aims at global separability. This is essential in images with partially repetitive patterns. In order to avoid confusion elements of repetitive patterns should not be selected or at least should get a low weight. Thus the selection of seldom or interesting points leads to reliable results, explaining the notion "interest operator". A similar line of thought leads to the notion of salient features.

(5) Interpretability: The selection principle should be interpretable in some sense, e.g. looking for edges, corners, blobs or other simple but labeled features. This requirement is not essential from an engineering point of view, but may be essential if the interest operator is used for image analysis.

The result of this first step are two lists with the n' and n" points selected in the two images I' and I", their pixel coordinates and their description, e.g. in the form of the local grey level function in the selected window. The advantage of the selection is obvious: it leads to a great information reduction, as you only have do deal with the two lists not with all pixels. It explains the requirements for the selection principle as the selected points have to represent reliably the total image content with respect to the matching problem.

5.2.2 Preliminary Correspondence based on Similarity

From the n'xn" possible pairs of points only a few are pairs of corresponding points. In this step a preliminary list of candidate pairs is built, which is based on the similarity of the points. Points are said to be similar if their description is similar. The similarity measure should fulfill the following requirements (Forstner, 1988):

1) Invariance: The similarity measure should be invariant with respect to the expected geometric and radiometric distortions between the images. e.g. the correlation coefficient is invariant with respect to linear transformations of the grey values, but not with respect

to geometric distortions.The problem with similarity measures is the form of the window which usually is chosen to be square or circular and which in general is not invariant with respect to scale differences or even shears.

2) Seldomness: The similarity measure should be able to include the seldomness of the individual points. Thus the degree of seldomness of both points of concern should also decide whether they remain in the preliminary list of the corresponding point pairs or at least should influence the weight of the correspondence.

3) Heuristics: A priory knowledge may be incorporated in this step. E.g. the maximum parallax to be expected may be used to further reduce the number of candidate pairs. A special case would be the condition resulting from the known epipolar geometry reducing the search space by one dimension.

4) Metric: For a thorough analysis it is convenient if the similarity measure has metric properties, i.e. besides being a distance measure it fulfills the triangle equation $d_{ij} \leq d_{ik} + d_{kj}$. This holds for the sum of squares of the grey level differences between the selected windows. A large distance d may correspond to a small similarity s=1/d or s=1-d.

The preliminary list of candidate pairs, resulting from this step, is a further information reduction. Whereas in the previous step we still kept the full description of the individual points we now only need their position and the weight of the correspondence, unless it is needed for achieving consistency.

5.2.3 Achieving Consistency

The local one-to-one comparison using the similarity measure and the heuristics in general is not able to yield a globally consistent matching result. Consistency here is understood as the first fit of the data with respect to an object model or at least to a global model for the mapping function between the two images. In order to arrive at a final solution we therefore have to the following (Forstner, 1988):

(1) Provide a 3D model of our object. The strength of the model directly influences the quality of the solution. The model may also be setup for the mapping function between the images, which - using the inverse perspective relations - can then be interpreted as a 3D model.

(2) Choose a consistency measure which is able to determine the closeness of fit of the data with the model. The choice of the target function is difficult in cases where different types of deviations between data and model have to be balanced. A classical problem is the proper relative weighting of the measuring errors, the smoothness of the surface and the frequency of discontinuities. A common theoretical framework, which allows maximum likelihood estimation, seems to be of great advantage for a thorough setup.

(3) Choose an algorithm to find a solution of optimal or at least satisfying consistency. There are various algorithms in use: nearest neighborhood methods are very common. Relaxation schemes used by Barnard and Thompson (1980) are very popular, as they may incorporate quite different types of consistency conditions. The clustering

approach proposed by Stockman et. al. (1982) shows intuitively that a global solution is aimed at.

5.3 Interest Operator

The interest operator has to find points with the two requirements from point transfer.

a) it should be close to a circle and

b) it should be small.

The error ellipse can only be calculated using the grey levels within a certain window, which usually is chosen 5x5 or 7x7 pixels. The center of the window in general is not the best point for matching, as the transformation of this point is not invariant to geometric distortions. An optimal point within the window is the weighted center of gravity, which proves to have attractive features (Lu, 1988).

5.4 Test Results

1) Table 5.1 shows the 11x11 gets a better result than either 5x5 or 7x7 windows which are suggested by Lu in 1988.

5x5 (run time 42 min.)		7x7 (run time 44 min.)		9x9 (run time 49 min.)		11x11 (run time 46 min.)	
Accepted 0.6 ~ 1.0	Rejected < 0.6	Accepted 0.6 ~ 1.0	Rejected < 0.6	Accepted 0.6 ~ 1.0	Rejected <0.6	Accepted 0.6 ~ 1.0	Rejected <0.6
16246	7111	16677	6680	17054	6307	17261	6096

Table 5.1. Different window sizes using weighted gradient operator (total 23357 points)

The results are from processing (dataset) with VCOR (a VTA matching software)

2) Table 5.2 shows the weighted gradient gets better results than the variance operator. From the visual inspection, most improvements are in the linear features such as the roads, edge of the parcels, edge of the ploughed agriculture fields, which are parallel or near parallel to the epipolar line.

	Total Points	Blunders dz > 10 m.	Success Rate	RunTime (min.)
Variance operator	19179	364	98.1%	40
Weighted Gradient	16281	236	98.6%	37

Table 5.2. Success rate for variance and weighted gradient operators

The results are from processing (dataset) with VCOR (a VTA matching software) using PC 586 computer, the time unit is minute.

CHAPTER SIX

DATA STRUCTURING AND PYRAMID IMAGE MATCHING

6.1 Overview

A multiresolution coarse to fine matching strategy is frequently applied to reduce search space and to find matches more reliably. Pyramid image data structure is the best way to suit the needs.

6.2 Pyramid Image Formation

Multiresolution structures of image or the image pyramids are data structures which consist of sequences of images of the same object presented at successively reduced resolutions. Such a structure contains no information that is not implicitly presented in the finest version of the image in the sequence (the original image), but it has great potential of gaining computational efficiency in making some of this information explicit and improve the reliability of image matching. It also has the property that only a small overhead in memory space relative to the input image is required. These three characteristics make the mutltiresolution structure a very efficient tool in image analysis.

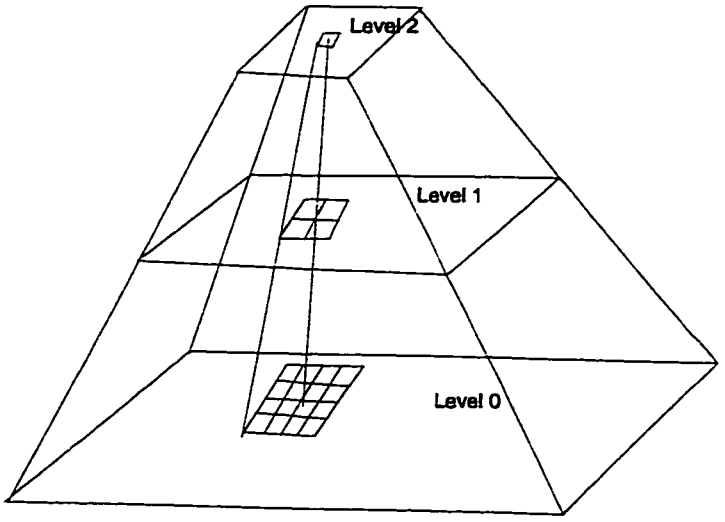


Fig. 6.1 Pyramid image (Adapted from Larsson 1984, p. 37)

To illustrate a multiresolution approach in spectrum estimation, consider a data sequence x(n), n=1, 2, 3, ..., to be decomposed into a pyramid structure of subsequences. Corresponding to different spacing, different levels of decomposition are generated (Blais, 1996):

level 0: $\{x(1), x(2), x(3), x(4), x(5), x(6), ..\}$

level 1: $\{x(1), x(3), x(5), x(7), x(9), x(11), ...\}$

$\{x(2), x(4), x(6), x(8), x(10), x(12), ...\}$

and generally,

$$\{x(1), x(1+2^k), x(1+2\cdot 2^k), x(1+3\cdot 2^k), \ldots\}$$

level k: .....

$$\{x(2^k), x(2\cdot 2^k), x(3\cdot 2^k), x(4\cdot 2^k), \ldots\}$$

6.3 Pyramid Image Matching Procedure

The first and most difficult step in recovering 3-D information (or generating DEM) from a pair of stereo images is that of matching points from one digital image of the pair to the corresponding points in the second image. Many computation algorithms have been used in attempts to solve this problem. These techniques primarily use area-based measures or multipoint matching techniques (Larouche, 1995).

The coarse-fine correlation used by a multiresolution pyramid approach (Hannah, 1988) can reduce computation time, which matches the images at higher level and calculating approximate solution to guide the matching at lower level. Beside that, the exertion of the epipolar geometry and rectified normal image can also limit the search space.

For area matching techniques, each image point to be matched is in fact the center of a small window of points in the first or reference image; this window is statistically compared with similarly sized windows of points in the second or target image of the stereo pair. Different matching measures are given in Section 4.3. Because comparison of a given reference window to every possible target window is computationally expensive, various heuristics have been developed to limit the area that must be searched. In addition

to the well-known epipolar constraint, these techniques have included extrapolation from already computed neighboring disparities, the use of image pyramids, and successive iterations of correlation and interpolation in a pyramid.

In order to match the image points, the first step can be to select a set of well scattered windows in one image, such that each window contains sufficient information to produce a reliable match. To accomplish this, an interest operator is passed over the image; this operator is a product of the image variance and minimum of ratios of direct differences over windows of the special size. Local peaks in the output of this operator are recorded as the preferred places to attempt the matching process. The motivation behind this operator is that it penalizes windows with low information and windows whose only information is contained in strongly linear edges, because either of these situations can cause difficulties in obtaining the correct match by means of area-based signal matching.

Whether or not point (x_1, y_1) in the first image I_1 is matched by point (x_2, y_2) in the second image I_2 is determined by computing the cross correlation, normalized by both mean and variance, over windows surrounding the points. The matching point is taken to be the point in I_2 with highest correlation, as located by one of several search algorithms.

The matching strategy begins with a few points that are highly likely to be matchable. These are matched by very global but very conservative searching algorithms. Each successive algorithm operates on less promising points, but uses more information from matches made at previous levels to constrain the search to smaller and smaller portions of

the epipolar line, until eventually all interesting points have been processed. Pixels in each reduced image of the hierarchy are produced by convoluting the parent image with Gaussian filter, then resampling; images are almost always reduced in size by a factor of 2 at each step of the pyramid.

There are two kinds of algorithms to start pyramid matching, without using the reference information or using limited known information. The first case is called unconstrained pyramid matching. The second case is called constrained pyramid matching.

6.3.1 The unconstrained pyramid matching

In the first case, nothing is assumed known about the relative orientations of the images, other than that they cover approximately the same area, at about the same scale, with no major rotation between the images nor any significant time-lapse changes. Each specified point is matched using an unguided pyramid matching technique. This technique begins with a point in the largest image (i.e., the highest-resolution image) and numerically traces that point back up through that image's hierarchy by repeatedly scaling down the coordinates of the point until it reaches an image that is approximately the size of the correlation window. It then uses a two-dimensional spiral search, followed by a hill-climbing search for the maximum of the correlation between the image windows. This global match is then refined back down the image pyramid; that is, the disparity at each level is used as a starting point for a hill-climbing search at the next level. The correlation window size remains constant at all levels of the pyramid, so the match is effectively

performed first over the entire image, then over increasingly local areas of the image. This technique permits the use of the overall image structure to set the context for a match; the gradually increasing detail in the imagery is then followed down through the pyramid to the final match.

In this matching technique, as in all the others used, matches must pass fairly strict tests in order to be considered correct, and only the successful matches are recorded for further use. At any level in the hierarchy, matches with very poor correlation are discarded, as are matches that fall outside the image.

Each match must also be confirmed by a technique of back-matching. Having found that point (x_1, y_1) in the first image I_1 is best matched by (x_2, y_2) in the second image I_2, then repeat the entire matching algorithm, this time starting with (x_2, y_2). If (x_1, y_1) and (x_1', y_1') differ by more than one pixel, the entire match is discarded as being unreliable.

6.3.2 The constrained pyramid matching

In the second case, besides the procedure of the unconstrained pyramid matching, further processing can make use of camera models (i.e., relative or absolute orientation data) and initial DEM model which uses the limited object space information as mentioned in Section 2.7. For imagery supplied with camera models, the given information is used which called constrained pyramid. The next technique to be applied is epipolar constraint. Having the camera parameters, you can know the manner in which a point in the first

image projects to a line in the second image - the epipolar constraint. This constraint allows us to cut the search range from two dimensions to one dimension at each level of the pyramid. In all other respects, epipolar constrained pyramid matching proceeds very much like unconstrained pyramid matching, with the additional match-evaluation criteria that matches must lie within a specified search range on the epipolar line. The initial DEM model gives the starting position on the epipolar line. The maximum search range is determined based on the maximum slope of the terrain in the stereopair. This technique is used on any unmatched points among the two most interesting points for each grid cell.

In conclusion, the basic procedure of multiresolution pyramid image matching method is: the matching starts from the most coarse level, and the solution is transferred to the next finer level as the initial approximation. On each of the image levels, a one-dimensional search with reduced search ranges along the epipolar line is performed in order to find the highest correlation coefficient point.

6.4 Analysis of multiresolution image correlation

It can be inferred that the fewer the high frequency components in a signal, the wider the curve of a correlation function is, as shown by the curve (a) in Fig. 6.2 Conversely, when there are more high frequency components in a signal, the curve will become more abrupt, as with the curve (b) in Fig. 6.2 (Wang, 1990).

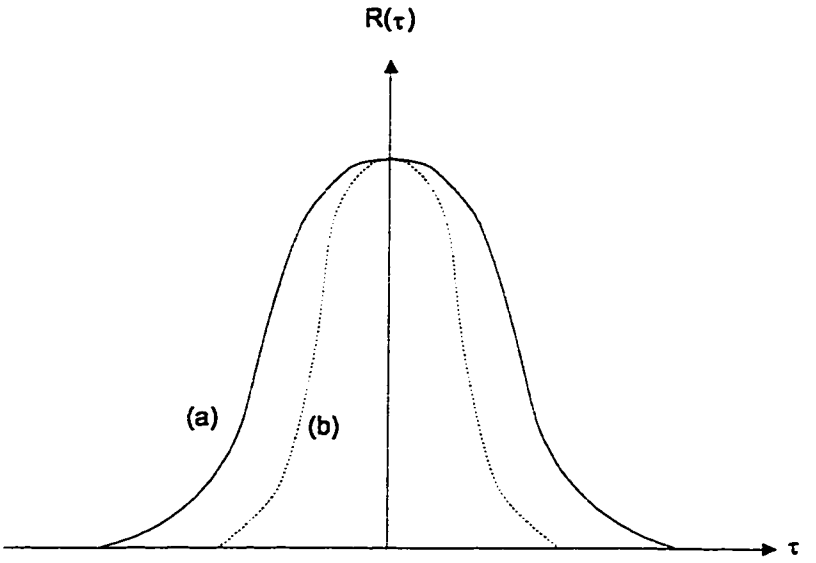


Fig. 6.2 High frequency correlation curve changes

(Adapted from Wang , 1990)

The variation of the correlation value for the latter is more sensitive to the changes in the displacement value between the target area and the search area. For the former, it is more appropriate to use coarse correlation, since it has a large search range; as for the latter, it would be best to exploit it for precise correlation.

Considering both, the following steps have to be adopted to perform coarse to fine multiresolution pyramid image matching.

(1) preliminary correlation to find the approximate positions of homologous points;

(2) then gradually adding components of higher frequencies (finally the original signal is used), and correlation is carried out in a gradually narrowing search area to acquire best accuracy.

In order to change the spectral band components of the original signal, filtering the signal may be needed in advance. The simplest method of low-pass filtering is to take the average grey value of every two neighboring pixels.

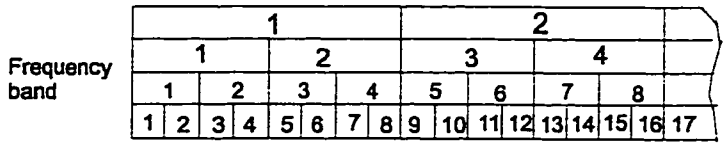


Fig. 6.3 Different frequency channels (Adapted from Wang, 1990)

For example, if the grey values whose original sampling interval is 0.025mm are taken as the first frequency channel (Fig. 6.3) then the grey values of the 0.050mm interval resulting from the averaging every two adjacent pixels form the second frequency channel, and its frequency now is lower than the first frequency channel by half. In a similar way, the third and fourth frequency channels can be constructed. This is equivalent every time to resampling with a rectangular aperture spot with a double side length for the next frequency channel and has the effect of low-pass filtering. Correlation starts from the lowest frequency channel of a signal and gradually goes to higher frequency ones until finally the first frequency (highest frequency) channel is reached. Since in processing, the signal of a frequency channel always uses the result of correlation of its preceding one as the initial value of searching, the range of searching is thus narrowed; reliability is improved, and accuracy of correlation is guaranteed.

Generally in digital correlation, the number of points taken in the target area are the same for all frequency channels. Since the sizes of pixels in different frequency channels differ, the lengths of the target areas of different frequency channels are not equal.

For example, the number of points selected from the target area is 21 for all frequency channels. When the sampling interval is 0.05mm, then the lengths of the first, second, third, fourth frequency channels are 1.05, 2.10, 4.20, 8.40mm respectively, and their corresponding pixels (namely, sampling intervals) are respectively 0.05, 0.10, 0.20, 0.40mm. Since correlation starts from the fourth frequency channel, and the range of its search area is dependent upon the maximum horizontal parallax (Δp)max, the size of the search area of channel four may be taken as (in pixels):

$$N_4 = \left[\frac{(\Delta p)\,\max}{0.40}\right] + 21 \qquad (6.1)$$

When search in a frequency channel is performed on the basis of correlation in its preceding lower frequency channel, its range of search may be confined within three points in its preceding lower frequency channel band in consideration of the possible error of one point (either on the left or right) in searching for homologous point in the immediately lower frequency channel.

In order to consider the effect of terrain undulations, you may utilize the result of the first correlation in the low frequency channel to ascertain the ground slope therein, and then make necessary geometric corrections of the image for the next correlation.

6.5 Improvement of Correlation Result

In accordance with the fact that correlation proceeds by comparing a row of pixels taken as the target area in the left image with its corresponding row of pixels of the same length in the search area of the right image, the correlation coefficient can be solved for the correlation relationship at the center of the pixel row. The same procedure is carried out for all the pixels which are taken as the central points one after another in the search area and then we shall obtain a series of correlation coefficients (as shown in Fig. 6.4) of which the address of the central one (point i in the figure) of the pixel row in the search area the maximum coefficient is taken as the conjugate point (homologous point) searched for (Wang, 1990).

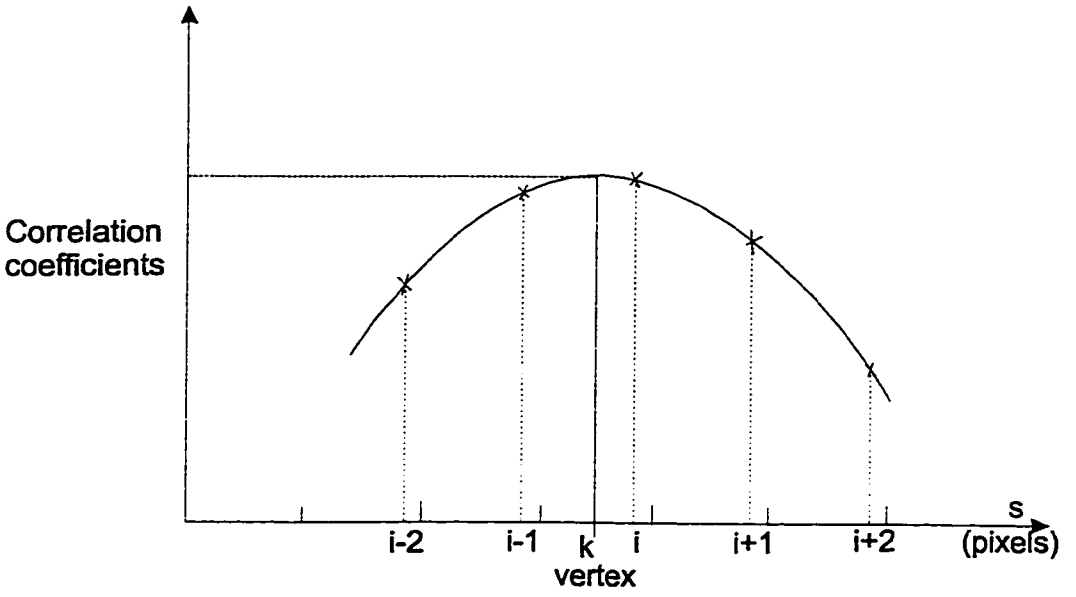


Fig. 6.4 Improvement of correlation result (Adapted from Wang, 1990)

To determine the position of the homologous point more precisely, it would be better to link the correlation coefficients obtained at several points located immediately to the left and right of the point i (assuming two points each on its left and right) with one adjustment function and take the maximum value of the function as the place (point k in the figure) where the homologous point searched for is located.

Assuming that there are five coefficients obtained at the adjacent pixels, as shown in Fig. 6.4, and they are approximated with a quadratic parabola equation. Its general form is taken as:

$$f(S) = A + BS + CS^2 \tag{6.2}$$

where the parameters A, B, C are computed by the method of adjustment of indirect observation. Here the address of the vertex k of the parabola should be:

$$k = i - \frac{B}{2C} \tag{6.3}$$

6.6 Reliability of Correlation

In order to improve the reliability of correlation, a number of measures may be taken, such as simultaneous use of two or more algorithms, or multiresolution correlation. Since reliability is of key importance and it is difficult to ensure a perfect determination, you should try every means possible to check it instead of depending entirely upon one criterion such as correlation. For example, at the same time when you use the method of the maximum correlation coefficient to search for a homologous point, you may also consider some other factors such as whether this coefficient is larger than a certain threshold; whether the interval between the maximum correlation and the predicted correlation is too large; whether the standard deviation of the grey levels of the pixels of the two correlated images is too small, whether the right image grids have the same sequences as the left image grids etc. You may set up a standard for reliability by combining these considerations in the assessment of the quality of matching so that you may discover promptly any unreliable points to be matched in the process of correlation and then make appropriate treatment on time (Wang, 1990).

The results of correlation reveal changes of horizontal parallaxes between homologous image points, which in turn indicate ground slopes and their variations. Since under normal circumstances there are always certain limitations for the value of ground slopes and their variations, they may be taken as grounds on which to judge the resulting parallaxes and make necessary smoothing. This kind of checking should be performed in the whole process of correlation.

The process of correlation described above is rather tedious. However, it is only necessary for the first several homologous epipolar lines to be correlated in such a way, whereas in the subsequent correlation of a large number of epipolar lines in the overlap area of the image pair. You may obtain much more correctly predicted values and corrections from the correlation results of the first few epipolar lines. Thus the range of search can be narrowed and some iterative steps can be saved.

Since the correlation results of the first few epipolar lines will be taken as a basis for subsequent correlation of the remaining epipolar lines, you should choose those image parts with plenty of details and less ground elevation differences from the image pair. The maximum length of the search area and the number of stages in multiresolution image correlation are also determined by the possible maximum horizontal parallaxes in these epipolar lines.

To secure better reliability of correlation results, the factor of ground features has been gradually introduced into current correlation technology. This relates to the advance of the feature-based matching approaches.

6.6 Test Results

(1). Table 6.1 shows overall statistics between different image levels and matching efficiency on model 412. You can see that as more levels are used, the coarsest level centered distance is smaller, and more points are rejected. This is because when more levels are used, less noise exists in higher level images and more criteria to pass. From finest center distance, you can see all of them are within 1 pixel except the only one level image result. This is proof of reliable matching by the multiresolution matching approach.

Level	1	2	3	4	5
Model	412	412	412	412	412
Ground Grid (m.)	20.	20.	20.	20.	20.
Correlation Threshold	0.7	0.7	0.7	0.7	0.7
Texture Threshold (variance)	2.0	2.0	2.0	2.0	2.0
Low Correlation Coefficient	0.4	0.4	0.4	0.4	0.4
Image Level Reduction	2	2	2	2	2
Coarsest Search Range (pixels)	119	119	119	119	119
Finest Search Range (pixels)	119	10	10	10	10
Coarsest Centered Distance-pix	15.623	14.417	13.985	13.396	13.068
Finest Centered Distance – pix	15.623	0.116	0.296	0.093	0.308
Total Attempted Points	669	669	669	669	669
Total Accepted Points	567	555	550	541	529
Total Rejected Points	102	114	119	128	140
Run Time (min.)	8	12	14	21	33

Table 6.1 Overall Statistics. The results are from processing (dataset) with VCOR (a VTA matching software)

(2). Table 6.2 is a comparison between the image levels and blunders. The test is based on model 7068. You can see that as more image levels are used, less blunders exist. Image reduction by 2 has less blunders than image reduction by 1. This is because less image level reduction more noise may be included in the matching chips and these noise may cause more blunders.

Image Levels	Run Time (PC-586)	Accepted (0.6~1.0)	Rejected (<0.6)	Blunders (dz>20m.)
5-4-3-2-1	2:33	19198	4159	189
5-3-1	2:18	18821	4536	166
4-3-2-1	1:27	19304	4053	224
4-2-1	1:19	19241	4116	206
3-2-1	1:00	19291	4066	332
3-1	0:55	19179	4178	255
2-1	0:47	19346	4010	672
1	0:39	19934	3417	2495

Table 6.2 Statistics between image levels and blunders

The results are from processing (dataset) with VCOR (a VTA matching software) using PC 586 computer.

Image resolution: 2000dpi; image size: 13000x13000 pixels; photo scale: 1:40,000

Terrain type: rolling; sampling interval: 40m.; image quality: good

(3). Table 6.3 shows the image level reduction by 2 gets the least blunders. This test is also based on the model 7068. The result is even better than when image level reduction by 1.

Image Levels	Image Level Reduction	Run Time (PC-586)	Accepted (0.6~1.0)	Rejected (<0.6)	Blunders (dz>20m.)
5-4-3-2-1	1	2:33	19198	4159	189
5-3-1	2	2:18	18821	4536	166
5-2-1	3	2:15	18580	4777	206
5-1	4	2:02	19696	3661	987

Table 6.3 Statistics between image level reduction and blunders

The results are from processing (dataset) with VCOR (a VTA matching software) using PC 586 computer.

CHAPTER SEVEN

ORTHOPHOTOS AND DEM QUALITY CONTROL

7.1 Orthophotos and Digital Differential Rectification

After completing the DEM by matching process, orthophotos can be made as one of the final digital mapping products. Another useful aspect of orthophotos is that they can be used as a tool to check the quality of the DEMs.

7.1.1 Basic Concepts

Rectification is to transform an aerial photograph or satellite image with tilt and relief displacements into one with correct planimetric point positions. A photograph that has the properties of an orthographic projection is called an orthophotograph. It is derived from a conventional perspective photograph by simple or differential rectification so that image displacements caused by camera tilt and relief of terrain can be removed (Wang, 1990).

7.1.2 Principle of Digital Rectification

If the orientation parameters which usually vary with the time are given and the DEM is also available, different remote sensing images can be processed by their corresponding

projective transformation equations. However, the practical method for geometrical processing of remote sensing images is the computation based on a mathematical model using control points. This method is mainly executed in digital form and is called digital differential rectification (Wang, 1990).

Prior to digital differential rectification, the imagery should undergo necessary image transformations depending upon situations, such as corrections for the error due to the homogeneous rotational speed of the scanning mirror, corrections for Earth curvature and atmospheric refraction, the error due to terrain undulations and the correction for image distortion due to the earth rotation, etc.

In the geometrical rectification of digital imagery, what is to be determined first of all is the geometric relationship between the original image and the image after rectification. Assuming that the coordinates of an arbitrary pixel p before and after image rectification are respectively (x_p, y_p) and (X_p, Y_p), then two reciprocal analytical expressions are formed:

$$X_p = F_x(x_p, y_p), Y_p = F_y(x_x, y_p) \tag{7.1}$$

$$x_p = f_x(X_p, Y_p), y_p = f_y(X_p, Y_p) \tag{7.2}$$

The former is the forward transformation equation, used in the so called direct method, while the latter is the backward transformation equation, called the indirect method.

Since the effect of terrain undulations is very small in space-borne remote sensing, the transformation functions F_x, F_y, f_x, f_y are basically transformations between two planes. The simplest function to use is a polynomial. However, in air borne remote sensing, or when high accuracy is required in image rectification, the effect of terrain undulations cannot be neglected. Therefore, besides known orientation elements of photography, knowledge of the heights of ground points (with DEM) is also necessary in digital differential rectification so that you may use corresponding projective transformation equations (generally collinearity equations) to perform transformation between the image point coordinates (x_p, y_p) of the original image, and the coordinates X_p, Y_p of the rectified image and the heights Z_p of the points. Since the computational work involved in this method is very large, generally the digital differential rectification of an area is to divide it into sub-areas. For the four corner points (sometimes called anchor points) of a sub-area, a method of computation with more rigorous projective transformation equations is used, whereas for the large number of points within the sub-area, we may use simple polynomials in transformations.

7.1.3 Digital Rectification by Direct Method (Forward Solution)

Fig. 7.1 illustrates the process of transformation by the direct method using Eqs. (7.1).

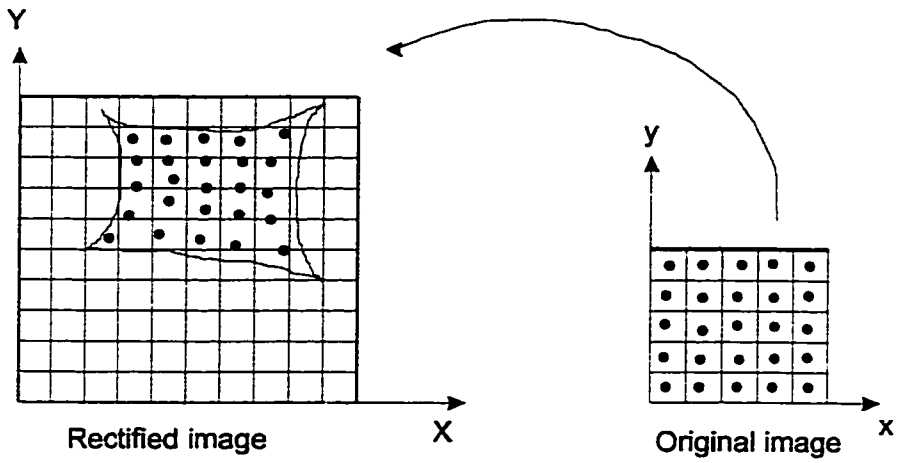


Fig. 7.1 Digital Rectification by Direct Method

(Adapted from Wang, 1990 p. 477)

Its advantage is that the pixels stored in a group of storage addresses on the tape may be used directly, but its disadvantage is that the output rectified image points are no longer those with equal intervals. In addition, the height Z_p of the points thus taken must be obtained by a certain method, generally by means of interpolation in the DEM of regular points. Since Z_p is a function of X_p and Y_p, successive approximation is required at this moment. The theoretical equations for the transformation in the direct method are varied. For example, for the projective transformation equations in frame camera photography, according to Eqs. (7.1), we have:

$$\begin{cases} (X_p - X_s) = (Z_p - Z_s)\dfrac{r_{11}x_p + r_{21}y_p - r_{31}f}{r_{13}x_p + r_{23}y_p - r_{33}f} \\ (Y_p - Y_s) = (Z_p - Z_s)\dfrac{r_{12}x_p + r_{22}y_p - r_{32}f}{r_{13}x_p + r_{23}y_p - r_{33}f} \end{cases} \tag{7.3}$$

Here iterations must be used, since Z_p is generally determined by its planimetric positions X_p, Y_p.

The simplified polynomial digital rectification equations are:

$$\begin{cases} X_p = a_0 + a_1 x_p + a_2 y_p + a_3 x_p y_p + a_4 x_p^2 + a_5 y_p^2 \\ Y_p = b_0 + b_1 x_p + b_2 y_p + b_3 x_p y_p + b_4 x_p^2 + b_5 y_p^2 \end{cases} \tag{7.4}$$

7.1.4 Digital Rectification by Indirect Method (Backward Solution)

The other method is the rectification process by the indirect method, as shown in Fig. 7.2

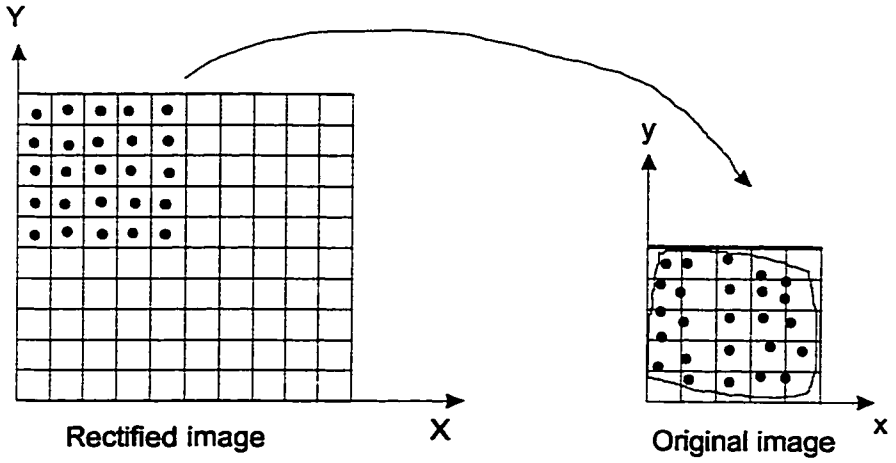


Fig. 7.2 Digital Rectification by Indirect Method

(Adapted from Wang 1990, p. 478)

In this case, you start from the node points (X_p, Y_p) of a certain output regular DEM to compute inversely for the positions (x_p, y_p) of their corresponding image points. Here the heights of the node points of the DEM are known, and can be used directly, but it is necessary to estimate the part of image to be stored with direct access.

The projective transformation equations for frame camera photography are (see Eqs. (3.1)):

$$\begin{cases} x_p = -f \dfrac{r_{11}(X_p - X_s) + r_{12}(Y_p - Y_s) + r_{13}(Z_p - Z_s)}{r_{31}(X_p - X_s) + r_{32}(Y_p - Y_s) + r_{33}(Z_p - Z_s)} \\ y_p = -f \dfrac{r_{21}(X_p - X_s) + r_{22}(Y_p - Y_s) + r_{23}(Z_p - Z_s)}{r_{31}(X_p - X_s) + r_{32}(Y_p - Y_s) + r_{33}(Z_p - Z_s)} \end{cases} \tag{7.5}$$

In the indirect method, simplified polynomial equations for digital rectification are more commonly used, such as:

$$\begin{cases} x_p = c_0 + c_1 X_p + c_2 Y_p + c_3 X_p Y_p + c_4 X_p^2 + c_5 Y_p^2 \\ y_p = d_0 + d_1 X_p + d_2 Y_p + d_3 X_p Y_p + d_4 X_p^2 + d_5 Y_p^2 \end{cases} \tag{7.6}$$

since the pixels in the original image obtained through the backward computation generally do not fall on the sampling points in digital rectification by the indirect method, the grey values of these points cannot be read directly but have to be obtained by interpolation. This procedure is called resampling.

7.2 DEM Quality Control (QC)

DEM quality control and manual checking is also very important. It guarantees that DEM data are clean before visualization and mapping.

7.2.1 Manual

In most DEM generation systems, matching and surface densification are truly automatic tasks, requiring human intervention only in the beginning to initialize the process. Despite all of the checking performed by the two tasks, it is essential that the DEM is now

checked by a human operator for accuracy and completeness, a process that may be called quality control (QC).

QC is implemented in an interactive environment comprising the two steps of displaying the DEM (visualization) and editing the data if necessary. The task is crucial for it does not only affect the quality of the DEM but also the economy of an automated approach.

The perspective view of DEM and rendering (shading of the DEM based on an artificial light source) are two effective ways for detecting errors, but they do not offer any accuracy assessment. One way to detect subtle mistakes in the DEM and to assess its accuracy is a true 3D display, superimposed on the images. In this case, the operator perceives the model from the stereopair and can compare it with the 3D view of the DEM. Softcopy photogrammetric workstations provide an ideal environment for this task. However, careful checking remains a tedious process. Norvelle reported (Norvelle, 1992) that the amount of data produced in 15 minutes of matching a stereopair, using the correlation method, can require up to 5 hours to check and edit.

7.2.2 Iterative Orthophoto Refinement (IOR)

7.2.2.1 Overview

The idea is to produce two orthophotos with the DEM and the two images of the stereopair. If the DEM is correct, then the two orthophotos should be identical, except for

radiometric differences. Geometric displacements incur from a wrong DEM. There are two steps in the IOR method. First, correlation is performed on a stereopair of images to determine a dense grid of conjugate points. These points are intersected to obtain a DEM. The DEM is then used in conjunction with the exterior orientation parameters for the stereoimages to generate an orthophoto of each image of the stereopair.

IOR is used in a second mode to automatically determine the geometric mismatches between the two orthophotos. The procedure is the same as in the first mode but now the conjugate image points determined by IOR should have identical orthophoto image coordinates. The extent to which they are not identical (mismatches) can be transformed into equivalent elevation errors that are then used to correct the current DEM values (Norvelle, 1992).

7.2.2.2 DEM Error Computation

Fig. 7.3 shows the geometric relationship between mismatches in the orthophotos and the DEM error (dh) that caused the mismatch.

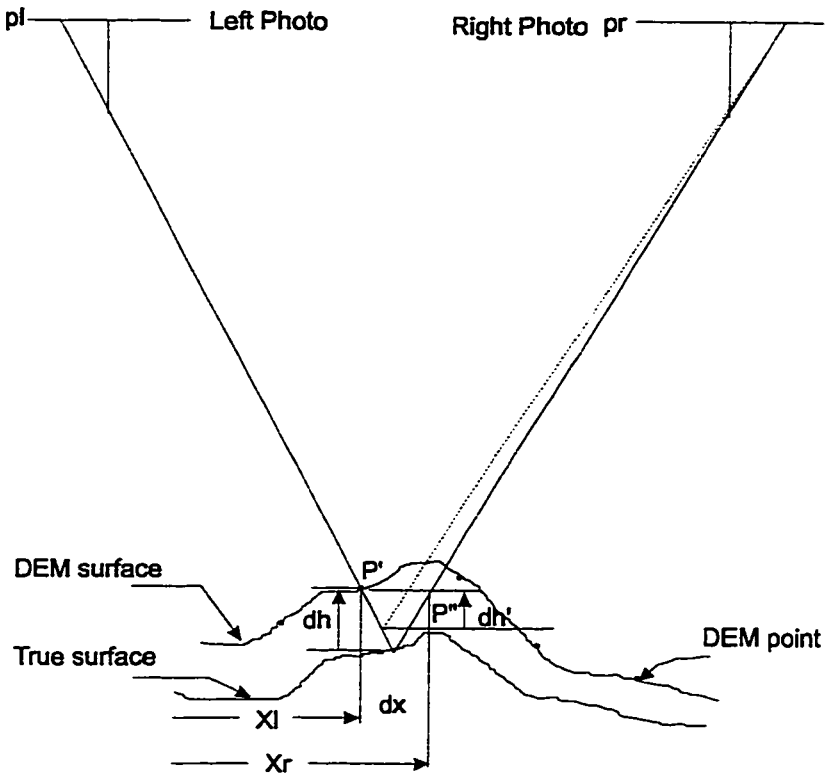


Fig 7.3 Relationship Between Orthophoto Mismatches and DEM Errors

(Adapted from Norvelle, 1992)

P' is a measured DEM point (show at the dot positions) and is in error by a value of dh. The orthophoto image of point P' falls at a ground position Xl on the left orthophoto and at Xr on the right. The mismatch between the two is dx. Since the coordinates of the camera stations and the ground positions of P'(Xl) and P"(Xr=Xl+dx) are known, it is possible to compute the position of P by intersection.

The error in elevation between P' and P can be computed and, if the slope of the true ground surface is known, converted to the error dh directly beneath the DEM point P'.

The true ground slope is not known, however, but can be approximated from the measured DEM surface. Alternatively, the approximate error dh' can be computed as dh'=dx(H/B), where H/B is the reciprocal of the base-height ratio between camera stations. The error dh (or dh') can then be used to update the DEM value of point P'. In the cases where the DEM errors are erratic (individual "spikes" in the elevation data), the slope of the ground are approximated by the slope of the DEM

7.2.2.3 The Advantage of IOR

Tests show that the IOR method is considered more accurate and faster than conventional correlation on original stereoimages followed by manual or interactive editing. It is faster because the orthophotos are more nearly identical and less searching space ("pull-in" range) is required to find the conjugate points. It is more accurate because dissimilarities between the original stereoimages, a potential cause of error in a correlation process, are essentially removed from the orthophotos.

Usually, two or three iterations are necessary to correct a DEM. Special cases, such as the conjugate images of the tall, individual trees because of the large x-parallax between tree images and the adverse affects of large tree shadows, may require more. Generally, if the mismatches are not removed from the orthophotos in 3 iterations, conditions (occlusions, shadows, etc.) probably exist which cause DEM errors that can only be corrected by the manual method.

7.2.2.4 Implementation of IOR

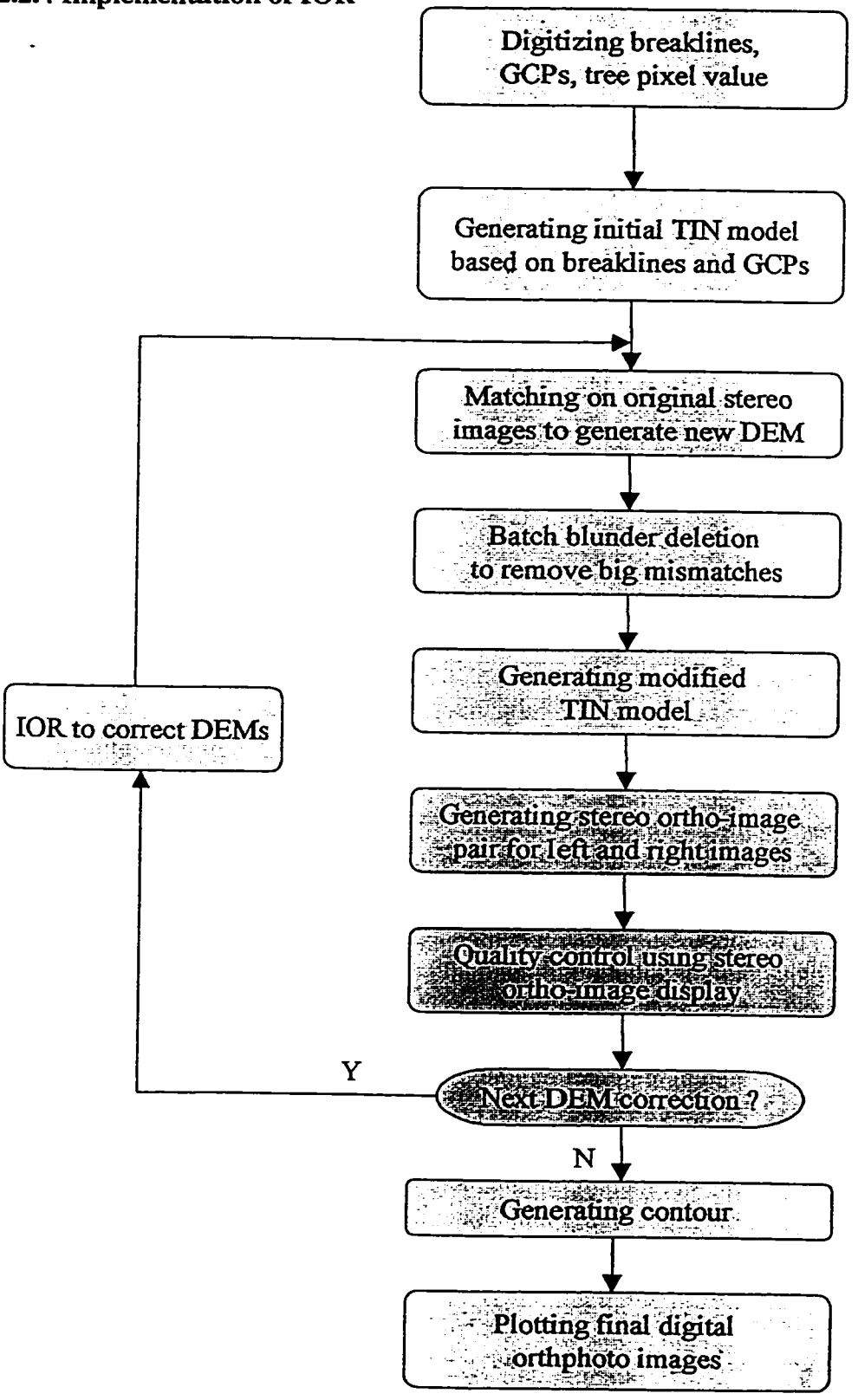


Fig. 7.4 Flow chart of IOR

7.3 Surface Fitting and Contour Map Generation

7.3.1 Surface Fitting

The points obtained by image matching are not evenly distributed and do not completely represent the terrain surface. Even if all pixels were selected as DEM grid point, there would be holes left because matching is not always successful. Thus, the 3-D points must be interpolated. The term surface fitting is more general as it includes interpolation as well as approximation methods.

The spline function is the most popular interpolation method. Another interpolation method needs to be mentioned here is using TIN (Triangulated Irregular Network) profiling to make irregular DEM grid to a regular DEM grid. So there will be no "holes" in DEM. Least-squares collocation surface fitting method aims at representing residual systematic errors (which cannot be expressed by mathematic models) by means of covariances or correlation functions in the form of signals.

After surface fitting, perspective view, rendering and other visualization tools can be used in softcopy photogrammetric workstations to present DTMs in three-dimensional space.

7.3.2 Contour Map Generation

Contour maps are generated from TIN models by intersecting horizontal planes with the network based on cleaned image matching results. The secondary data structure of ridges and channels can be used as a guide to the starting points of each contour envelope. The contour envelopes may need secondary processing to remove artifacts resulting from the edges of the triangles.

CHAPTER EIGHT

CONCLUSION AND RECOMMENDATIONS

8.1 Conclusion

Research efforts in digital photogrammetry have begun to bear fruit. Together with advances in hardware and system software, it is possible to develop operational systems. Furthermore, we are witnessing the advent of softcopy photogrammetric workstations which now deal with digital imagery.

As always when new technologies and methods emerge, expectations are raised that photogrammetric processes can be automated. This will be especially true for softcopy workstations: potential uses may push expectations beyond what can be delivered in the foreseeable future. We argue that major progress toward autonomous softcopy workstations depends more on advances on the conceptual level rather than on the refinement of system components such as hardware, system software, and low-level algorithms. This pushes us to adopt the computer vision for photogrammetric applications, understand the methods and tools that are available, and develop skills which enable researchers to analyze precisely how human operators perform their tasks to generate photogrammetric products.

8.2 Recommendations

Multiresolution pyramid image matching algorithm can greatly reduce blunders and increase the matching reliability. In order to make the process more productive, more research need to be done for automatically selecting optimal highest image levels based on photo scale, image resolution, terrain roughness and image contrast etc. Fractal assumption may help us to set up a proper model.

Softcopy photogrammetric workstation can be partly used in production to gradually replace the analogue machines. But as with new technology, the users should be aware of potential problems, such as the blunder detection, DEM quality control etc. Human interaction is still need especially in low texture terrain, forest, urban/ building area.

The problem of increasing the degree of automation is the problem of making implicit information, embedded in the raw image, explicit. Photogrammetrists are perhaps too focused in the sub-pixel world, deeply if not exclusively concerned with accurate point positioning. We have to shift our attention from data-driven, pixel-to-pixel operations to a more symbolic processing of abstract representations.

BIBLIOGRAPHY AND REFERENCES

Ackermann, F., 1984. Digital Image Correlation: Performance and Potential Applications in Photogrammetry, Photogrammetry Record, 11(64):429-439.

Ackermann, F., 1988. Impact of GPS on Photogrammetry. 3rd South East Asian Survey Congress, Bali, Indonesia.

Al-Tahir, R., Ch. Toth, and T. Schenk, 1990. Automatic Ground Control Point Recognition with Parallel Associative Memory. Proceedings of the ACSM-ASPRS Fall Convention, pp. b-10-B-18.

Blais, J.A.R, 1996. Spectrum estimation using maximum entropy and multiresolution considerations, Journal of Geodesy, 70: 349-356.

Burrough, P. A., 1986, Principles of Geographical Information Systems for Land Resources Assessment, Clarendon Press . Oxford.

Chapman, M. A. and Tam, A. 1990. A Rigorous Approach for the Estimation of Terrestrial Coordinates from Digital Stereo SPOT Imagery.

Chapman, M.A. 1992. Stereo Digital Image Processing. Lecture Notes, Department of Geomatics Engineering, The University of Calgary.

Chen, C., 1989, Derivation of Digital Elevation Models by digital Image Processing and Numerical Methods", Ph.D. Dissertation, Department of Civil and Environmental Engineering, University of Wisconsin-Madison.

Dowman, I. J., 1992. Overview of European Developments in Digital Photogrammetric Workstations, Photogrammetric Engineering & Remote Sensing, Vol. 58, No. 1, January pp. 51-56.

Ebner, H. and Reiss, P., 1981. Experiences with Height Interpolation by Elements. ASP-ACSM Fall Technical Meeting. San Francisco – Honolulu 1981

Ebner, H., and Heipke, C. 1988. Integration of Digital Image Matching and Object Surface Reconstruction, Proceedings of ISPRS Commission III, Kyoto, Japan.

Forstner, W., 1982, On the Geometric Precision of Digital Correlation, Int. Arch. of Photogrammetry and Remote Sensing, Vol. 24, Commission III, pp. 176-189, Helsinki.

Forstner, W., 1986. Digital Image Matching Techniques for Standard Photogrammetric Applications, Proceedings of the ACSM-ASPRS Annual Convention, Vol. 4, pp. 210-219.

Grun, A., and Baltsavias, E. P., 1988. Geometrically Constrained Multi-photo Matching, Photogrammetric Engineering & Remote Sensing, 55(5):581-586.

Hannah, M. J., Digital Stereo Image Matching Techniques, Commission III/4

Helava, U. V., 1988. Object-Space Least-Squares Correlation, Photogrammetric Engineering & Remote Sensing. 54(6):711-714.

Helava, U. V., 1992. Prospects In Digital Photogrammetry, Proceedings of ASPRS/ACSM - Technical Papers, Volume 2 - Photogrammetry and Surveying, pp.19-24.

Horn, B. K. P., 1983. Non-correlation methods for stereo matching, Photogrammetric Engineering & Remote Sensing, 49(4):535-536.

Hoshi, T., Matsushita, T. and Koishikawa, Y., 1988, Accuracy of Stereo Matching Using Color CCD Camera, Int. Arch. of Photogrammetry and Remote Sensing, Vol 27-B, pp. 273-282, Kyoto.

Lam, W. K., 1990, An Object Based Image Matching Procedure for Photogrammetric Applications. M.Sc. thesis, University of Calgary, UCSE Reports Number 20037.

Larouche, C., 1995, Automation of Photogrammetric Operations Using Advanced Digital Image Matching Techniques. Ph.D Thesis, University of Calgary, UCGE Reports Number 20085.

Larsson, R., Hierarchical Data Structures and Algorithms for Digital Stereocopical Mensuration, Papers of the XV Congress of the ISPRS, RIO De Janeiro 1984, Stockholm. pp. 34.

Li, M., 1991. Hierarchical Multipoint Matching, Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 8, August, pp. 1039-1047.

Lu, Y., 1988, Interest Operator and Fast Implementation. Int. Arch. of Phptogrammetry and Remote Sensing, Vol. 27, Kyoto.

Makarovic, B., 1973. Progressive sampling for digital terrain models. ITCJ. 1973-3, 397-416.

Miller, S. B., Helava, U. V. and Devenecia, K. H., 1992. Softcopy Photogrammetric Workstations, Photogrammetric Engineering & remote Sensing, Vol. 58, No. 1, January, pp. 77-83.

Norvelle, F. R., 1992. Stereo-Correlation: Window Shaping and DEM Corrections, Photogrammetric Engineering & Remote Sensing, 58(1):113-117.

Norvelle, F. R., 1992, Using Iterative Orthophoto Refinements to Correct Digital Elevation Models (DEM's), Proceedings of ASPRS/ACSM - Technical Papers, Volume 2 - Photogrammetry and Surveying, pp. 27-35.

Rosenholm, D., 1987. Multi-Point Matching Using Least-Squares Technique for Evaluation of Three-Dimensional Models, Photogrammetric Engineering & Remote Sensing, 53(6):621-626.

Rosenholm, D., 1987a, Empirical Investigation of Optimal Window Size Using the Least Squares Image Matching Method, Photogrammetria, Vol 42, pp. 113-125.

Rosenholm, D. 1987b, Least Squares Matching Method: Some Experimental Results, The Photogrammetric Record, Vol. 12(70), pp. 493-512.

Saleh, R. A., and Scarpace, F. L., 1992, Multi-spectral Matching Techniques for DTM Generation, Proceedings of ASPRS/ACSM - Technical Papers, Volume 2 - Photogrammetry and Surveying, pp. 36-46.

Schenk, T. and Toth, C. K., 1992. Conceptual Issues of Softcopy Photogrammetric Workstations, Photogrammetric Engineering &Remote Sensing, Vol. 58, No, 1, January, pp. 101-110.

Wang, Zhizhuo, 1990. Principles of Photogrammetry, Press of Wuhan Technical University of Survey and Mapping, Publishing House of Surveying and Mapping, Beijing, China

Wrobel, B. P., 1988. Least-Squares Methods for Surface Reconstruction from Images, Proceedings of the International Congress for Photogrammetry and Remote Sensing, Commission III, Kyoto, Japan.

Zong, J., J.C. Li, J. C. and Schenk, T., 1991. Application of Forstner Interest Operator in Automatic Orientation System, Proceedings of the ASPRS-ACSM Annual Convention, 5:440-448.

APPENDIX I. THEORY OF PROJECTIVE TRANSFORMATION

1. Projective Transformation (Wang , 1990)

The most general linear transformation is the projective transformation, also referred to as collineation, which, expressed in three-dimensional form, is:

$$\begin{cases} x = \dfrac{a_{11}X + a_{12}Y + a_{13}Z + a_{14}}{a_{41}X + a_{42}Y + a_{43}Z + a_{44}} \\ y = \dfrac{a_{21}X + a_{22}Y + a_{23}Z + a_{24}}{a_{41}X + a_{42}Y + a_{43}Z + a_{44}} \\ z = \dfrac{a_{31}X + a_{32}Y + a_{33}Z + a_{34}}{a_{41}X + a_{42}Y + a_{43}Z + a_{44}} \end{cases} \qquad (1)$$

This projective transformation consists of three linear fractional functions having the same denominator, and transforms a point (X, Y, Z) in a three-dimensional space into another point (x, y, z). One important property of projective transformation is that a curve of a given degree is transformed into a curve of that degree. Therefore, the points in a line remain to be the points in another line, and a plane expressed by a set of variables goes over into a plane expressed by a set of new variables. For example, suppose there is an equation of a plane:

$$Ax + By + Cz + D = 0 \tag{2}$$

Substituting Eqs (1) and after rearrangement, we have:

$$\begin{aligned}(Aa_{11} + Ba_{21} + Ca_{31} + Da_{41})X + (Aa_{12} + Ba_{22} + Ca_{32} + Da_{42})Y + \\ (Aa_{13} + Ba_{23} + Ca_{33} + Da_{43})Z + (Aa_{14} + Ba_{24} + Ca_{34} + Da_{44}) = 0\end{aligned} \tag{3}$$

which remains to be an equation of a plane.

Another important property of projective transformation is that there is no essential distinction between finite points of space and infinite points of space. For instance, when any point (X, Y, Z) satisfies the following condition:

$$a_{41}X + a_{42}Y + a_{43}Z + a_{44} = 0$$

from the transformation based on Eqs. (1-39) it can be seen that the plane of its corresponding x, y, z is infinite.

Affine transformation is a kind of projective transformation, the conditions for which are:

$$a_{41} = a_{42} = a_{43} = 0 \qquad a_{44} \neq 0$$

If a_{ij} / a_{44} in Eqs. (1) is replaced by the symbol a_{44}, then we have the following equations for affine transformation:

$$\begin{cases} x = a_{11}X + a_{12}Y + a_{13}Z + a_{14} \\ y = a_{21}X + a_{22}Y + a_{23}Z + a_{24} \\ z = a_{31}X + a_{32}Y + a_{33}Z + a_{34} \end{cases} \tag{4}$$

There still exist a correspondence between points, lines,and planes in affine transformation, but it differs from the general projective transformation in that finite points of space are still transformed into finite points of space. After affine transformation, parallelism between lines is retained, whereas perpendicularity is not.

When the condition of the following determinant is satisfied in an affine transformation:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 1 \tag{5}$$

then the transformation figure would keep its area, but not its shape.

When the transformation matrix A in an affine transformation

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{6}$$

satisfies the following condition:

$$A^T A = k^2 E \tag{7}$$

Then the transformation becomes similarity transformation, which is a transformation when only shape, and not size is retained, and k is a scale factor. Similarity transformation, when k is removed, becomes a transformation where both shape and size are retained, and is also called transformation of rigid motion.

There are fifteen independent parameters in the general linear projective transformation equations (Eqs. (1)). This is because when the numerator and the denominator on the right side of Eqs. (1) is each divided by any parameter which is not equal to zero, no change is effected in that equation. The affine transformation equation (4) comprises twelve independent parameters, whereas the similarity transformation has seven independent parameters and rigid motion transformation has six.

2. Homogeneous Coordinates

If the following relations are substituted for the Cartesian coordinates X, Y, Z and x, y, z of a space point,

$$X = \frac{U}{T}, \qquad Y = \frac{V}{T}, \qquad Z = \frac{W}{T} \tag{8}$$

$$x = \frac{u}{t}, \qquad y = \frac{v}{t}, \qquad z = \frac{w}{t}$$

then the homogeneous coordinates U, V, W, T and u, v, w, t of that point are obtained. The homogeneous coordinates are important in dealing with problems in terms of projective geometry.

Substituting Eqs. (8) into Eqs. (1), we have:

$$\begin{cases} u = a_{11}U + a_{12}V + a_{13}W + a_{14}T \\ v = a_{21}U + a_{22}V + a_{23}W + a_{24}T \\ w = a_{31}U + a_{32}V + a_{33}W + a_{34}T \\ t = a_{41}U + a_{42}V + a_{43}W + a_{44}T \end{cases} \tag{9}$$

which, written in matrix form, become:

$$\begin{bmatrix} u \\ v \\ w \\ t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} U \\ V \\ W \\ T \end{bmatrix} \tag{10}$$

Rewriting the linear fraction form of Eqs. (1) into the linear form of Eq. (10) to facilitate computation is one of the major advantages of the homogeneous coordinates.

Now the inversion formula of Eq. (10) can be written as:

$$\begin{bmatrix} U \\ V \\ W \\ T \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}^{-1} \begin{bmatrix} u \\ v \\ w \\ t \end{bmatrix} \tag{11}$$

It is apparent that if all the homogeneous coordinates are multiplied by a non-zero factor r, one obtains new homogeneous coordinates (ru, rv, rw, rt), which, according to Eq (11), would result in the same Cartesian coordinates (x, y, z) as before. This implies that the cartesian coordinates of a point are unique, whereas the homogeneous coordinates are unique only to a scale ratio. It follows that the homogeneous coordinates (4, 8, 2, 2) and (2, 4, 1, 1) represent the same point, with their Cartesian coordinates being (2, 4, 1). An infinite point in the homogeneous coordinate system can be expressed as (u, v, w, 0) with its origin being (0, 0, 0, t).

The premise for the operation of Eq. (11) is that the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \tag{12}$$

should have a corresponding inverse. In other words, the matrix must be non-singular. In many applications, the matrix is non-singular, but in photogrammetry where a three

dimensional object point should be projected onto a two-dimensional image point, the transformation is singular. Hence, the theory of generalized inverse must be used for the inversion of Eq. (12).

When Eq. (10) are applied in photogrammetry, since all the three-dimensional space of the object space is projected onto one image plane, the following plane constraint condition should be applied to Eq. (10):

$$au + bv + cw + dt = 0 \tag{13}$$

which makes all the image points coplanar. Substituting (10) into Eq. (11), we obtain:

$$(a_{11}a + a_{21}b + a_{31}c + a_{41}d)U + (a_{12}a + a_{22}b + a_{32}c + a_{42}d)V +$$
$$(a_{13}a + a_{23}b + a_{33}c + a_{43}d)W + (a_{14}a + a_{24}b + a_{34}c + a_{44}d)T = 0$$

Since (U, V, W, T) represent the point coordinates of a three-dimensional object space, it is apparent that if the above equation is to be satisfied, all the coefficients should be identically equal to zero. In order to get the non-zero solution for a, b, c, d, the following determinant

$$\begin{vmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{vmatrix}$$

must be equal to zero, which means Eq. (12) is singular.

Because the chosen coordinate system is rather arbitrary, we may assume the plane for Eq. (13) to be:

$$cw + dt = 0$$

where c is not zero. Then from Eq. (11), we have:

$$c\frac{w}{t} + d = c\frac{a_{31}U + a_{32}V + a_{33}W + a_{343}T}{a_{41}U + a_{42}V + a_{43}W + a_{44}T} + d$$
$$= \frac{a_{31}X + a_{32}Y + a_{33}Z + a_{34}}{a_{41}X + a_{42}Y + a_{43}Z + a_{44}} + d = 0$$

In order that this condition can be satisfied for the values of (X, Y, Z), it is necessary that all the a_{3i} / a_{4i} should be equal to a constant (i = 1, 2, 3, 4), which means that the expression for photographic projection in a homogeneous coordinate system is

$$\begin{bmatrix} u \\ v \\ t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} U \\ V \\ W \\ T \end{bmatrix} \tag{14}$$

In the linear fractional equation form, Eqs. (1) is:

$$\begin{cases} x = \dfrac{a_{11}X + a_{12}Y + a_{13}Z + a_{14}}{a_{31}X + a_{32}Y + a_{33}Z + a_{34}} \\ \\ y = \dfrac{a_{21}X + a_{22}Y + a_{23}Z + a_{24}}{a_{31}X + a_{32}Y + a_{33}Z + a_{34}} \end{cases} \tag{15}$$

where elevation coefficients out of twelve are independent. Dividing both the numerator and the denominator on the right side of Eqs. (15) with $a_{34}, (a_{34} \neq 0)$ result:

$$\begin{cases} x = \dfrac{b_{11}X + b_{12}Y + b_{13}Z + b_{14}}{b_{31}X + b_{32}Y + b_{33}Z + b_{34}} \\ \\ y = \dfrac{b_{21}X + b_{22}Y + b_{23}Z + b_{24}}{b_{31}X + b_{32}Y + b_{33}Z + 1} \end{cases} \tag{16}$$

These are the image forming equations in the form of so-called Direct Linear Transformation (DLT) which are used extensively in close-range photogrammetry.

3. Theory of transformation for Photo Rectification

Eqs. (15) express the projective transformation of the object points in a three-dimensional pace into a plane (the image plane). Now assuming that the object point (X, Y, Z) should also be located in a plane, we then have the transformation of one plane being projected into another, which represents the theoretical relationships of photo rectification.

Let $Z = AX + BY + C$

substitution of which into Eqs. (15), we have:

$$\begin{cases} x = \dfrac{(a_{11} + a_{13}A)X + (a_{12} + a_{13}B)Y + (a_{14} + a_{13}C)}{(a_{31} + a_{33}A)X + (a_{32} + a_{33}B)Y + (a_{34} + a_{33}C)} \\ y = \dfrac{(a_{21} + a_{23}A)X + (a_{22} + a_{23}B)Y + (a_{24} + a_{23}C)}{(a_{31} + a_{33}A)X + (a_{32} + a_{33}B)Y + (a_{34} + a_{33}C)} \end{cases} \tag{17}$$

which can be simplified into:

$$\begin{cases} x = \dfrac{b_{11}X + b_{12}Y + b_{13}}{b_{31}X + b_{32}Y + b_{33}} \\ y = \dfrac{b_{21}X + b_{22}Y + b_{13}}{b_{31}X + b_{32}Y + b_{33}} \end{cases} \tag{18}$$

4. Direct Linear Transformation (DLT) and Collinearity Equations

The most important and basic collinearity equations in photogrmmetry, which are the image forming equations of the central projection, are:

$$\begin{cases} \bar{x} = -f \dfrac{r_{11}(X - X_s) + r_{12}(Y - Y_s) + r_{13}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} \\ \bar{y} = -f \dfrac{r_{21}(X - X_s) + r_{22}(Y - Y_s) + r_{23}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} \end{cases} \tag{19}$$

Considering the elements of the interior orientation x_0, y_0, we have $\bar{x} = x - x_0, \bar{y} = y - y_0$, and Eq (19) becomes:

$$\begin{cases} x - x_0 = -f \dfrac{r_{11}(X - X_s) + r_{12}(Y - Y_s) + r_{13}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} \\ y - y_0 = -f \dfrac{r_{21}(X - X_s) + r_{22}(Y - Y_s) + r_{23}(Z - Z_s)}{r_{31}(X - X_s) + r_{32}(Y - Y_s) + r_{33}(Z - Z_s)} \end{cases} \tag{20}$$

which are also equations for the transformation from the three-dimensional object space point coordinates (X, Y, Z) to the two-dimensional image point coordinates (x, y).

While the meanings they express are the same as those of the DLT Eqs. (16), their derivation is based on the use of the reasonings from the elements of the interior and exterior orientations of the photographic bundle of rays. In these equations, the nine direction cosines $r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}, r_{33}$ are the functions of the three independent angular orientation elements ω, φ, κ. Therefore, the equations comprise nine parameters consisting of the interior and exterior orientation elements $(x_0, y_0, f, \omega, \varphi, \kappa, X_s, Y_s, Z_s)$ in contrast with the eleven parameters in the DLT Eqs (16). In order that the two sets of equations correspond to each other in a rigid manner, two additional parameters may be

added to Eqs. (20) when the number of parameters for Eqs. (16) is adopted. For instance, a scale change ratio of the y direction with respect to the x direction of the image point coordinates (affinity) as one parameter, and the non-perpendicularity of x and y axes (non-orthogonality) as a second parameter may be added. On the other hand, when the number of parameters in Eqs. (20) is adopted, it may be assumed that two condition equations can be established between the parameters in Eqs. (16). In practice, however, when either Eqs. (16) or Eqs. (20) are applied, the theoretical correspondence between them, as discussed above, is seldom taken into consideration.

APPENDIX II. IMAGE RESAMPLING

If we want to know the values of points not located at the matrix points (sampling points) of the original function f (x, y), interpolation should be carried out. This is what we refer to as resampling, meaning one more sampling performed on the basis of the original sampling. Each time when geometric changes are made with a digital image, this kind of problems will inevitably occur. Among its typical cases are image rotation, alignment of epipolar lines, or digital rectification, etc. It is apparent that one or more of such geometric transformations will be frequently encountered in the application of digital image processing in photogrammetry and remote sensing (Wang, 1990).

As is evident from the sampling theory, when the sampling interval Δx is equal to or smaller than 1/(2W), and there are no frequency spectral components in the image greater than w, then the original image can be recovered by the use of following equation:

$$f(x) = \sum_{k=-\infty}^{\infty} f(k\Delta x) \bullet \delta(x - k\Delta x) \bullet \frac{\sin 2\pi Wx}{2\pi Wx} = \sum_{k=-\infty}^{\infty} f(k\Delta x) \frac{\sin 2\pi W(x - k\Delta x)}{2\pi W(x - k\Delta x)} \qquad (1)$$

Which is the convolution of the original image and the sinc function, with the sinc function taken as the convolution kernel. Since this kind of computation is very complicated, simpler functions are often used to replace the sinc function. Described below are three methods of resampling commonly used in practice.

1. The nearest neighbor method

By taking the grey value of the nearest pixel N to the position of point P (x, y) directly as the sampled grey value of that point, that is:

$$I(P) = I(N)$$

With N being the nearest point, then its image coordinates are:

$$\begin{aligned} x_N &= Int(x + 0.5) \\ y_N &= Int(y + 0.5) \end{aligned} \tag{2}$$

Where "Int" indicates that an integer function is used.

2. Bi-linear Interpolation

The convolution kernel in bi-linear interpolation is a triangle function, whose expression is:

$$W(x) = 1 - (x), 0 \le |x| \tag{3}$$

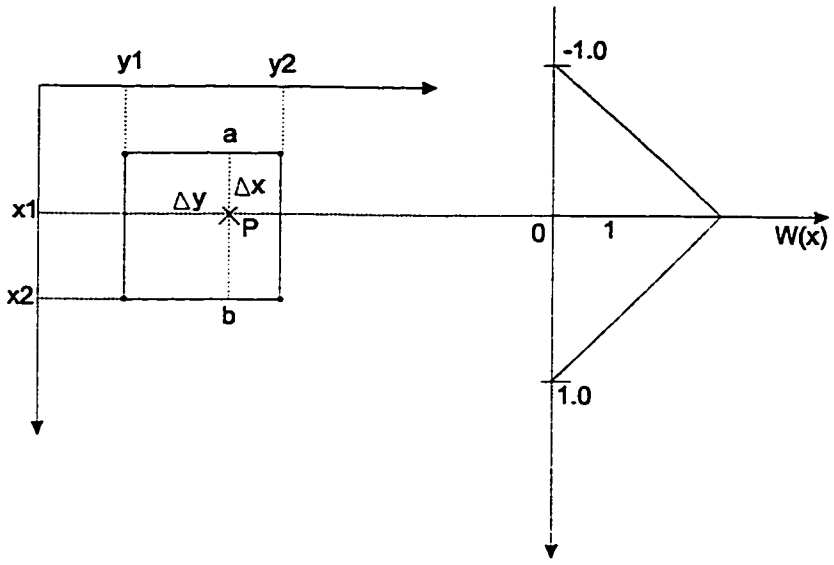


Fig. II-1. Bi-linear interpolation (Adapted from Wang, 1990, p 431)

It can be proved that resampling of any point by means of convolution by Eq. (3) is similar to the use of the sinc function to some extent. In this case, four original pixels in the vicinity of the resampling point P are required in the computation, as shown in Fig.II-1. Fig.II-1 (b) indicates the position where the convolution kernel graph of Eq. (1) should be placed when it performs resampling along the x direction.

Computations can be done separately along the x direction and the y direction, i.e., first resampling the grey levels of two points a, b along the y direction; then resampling point P along the x direction using these two points. In performing resampling computations in either direction, we should make the zero point of the convolution kernel even with point

P so as to read the corresponding values at the original pixels. In fact, the computations in the two directions can be combined into one, i.e., the weight values contributed by the

four original points to the point P can be directly computed after arrangement and reduction based on the above computational procedure to construct a 2x2 two-dimensional convolution kernel weight matrix W. Then performing Hadamard product operation on the 2x2 matrix **I** constituted by the grey values of the four original pixels, we can obtain a new matrix. Accumulating the elements of the new matrix will yield the grey value I (P) of the resampling point as follows:

$$I(P) = \sum_{i=1}^{2}\sum_{j=1}^{2} I(i,j) * W(i,j) \tag{4}$$

Where

$$I = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}; \qquad W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}$$

$$W_{11} = W(x_1) \bullet W(y_1);$$
$$W_{12} = W(x_1) \bullet W(y_2);$$
$$W_{21} = W(x_2) \bullet W(y_1);$$
$$W_{22} = W(x_2) \bullet W(y_2);$$

Here according to Eq. (4) and Fig. II-1, we have:

$$W(x_1) = 1 - \Delta x; W(x_2) = \Delta x; W(y_1) = 1 - \Delta y; W(y_2) = \Delta y$$
$$\Delta x = x - \mathrm{int}\,eger(x)$$
$$\Delta y = y - \mathrm{int}\,eger(y)$$

The resampling value of point P is obtained as follows:

$$\begin{aligned} I(P) &= W_{11}I_{11} + W_{12}I_{12} + W_{21}I_{21} + W_{22}I_{22} \\ &= (1-\Delta x)(1-\Delta y)I_{11} + (1-\Delta x)\Delta y I_{12} + \Delta(1-\Delta y)I_{21} + \Delta x \Delta y I_{22} \end{aligned} \tag{5}$$

3. Bicubic Convolution

We may also employ a kind of cubic function for the convolution kernel. The cubic spline function expressed in the following equation (6) is very similar to the sinc function.

$$\begin{cases} W_1(x) = 1 - 2x^2 + |x|^3, (0 \le |x| \le 1) \\ W_2(x) = 4 - 8|x| + 5x^2 - |x|^3, (1 \le |x| \le 2) \\ W_3(x) = 0, (2 \le |x|) \end{cases} \tag{6}$$

Resampling of any point by using Eqs. (6) as the convolution kernel requires sixteen original pixels in the vicinity of that point in the computation, as shown in Fig. II-2.

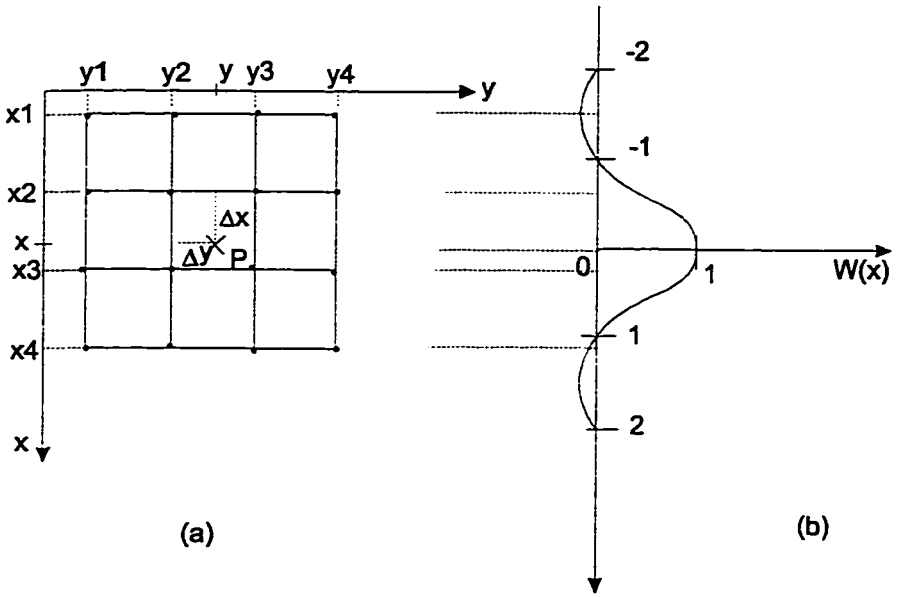


Fig. II-2. Bi-cubic convolution interpolation (Adapted from Wang, 1990, p 433)

Fig. II-2 (b) indicates the position where the convolution kernel graph in Eqs (7) should be placed when resampling is done along the x direction. Computations can be expressed separately along the x and y directions, or we may solve for the weight values contributed by the sixteen surrounding points to the resampling point P in one computation, Here we have:

$$I(P) = \sum_{i=1}^{4}\sum_{j=1}^{4} I(i,j)*W(i,j) \tag{7}$$

Where

$$I = \begin{bmatrix} I_{11} & I_{12} & I_{13} & I_{14} \\ I_{21} & I_{22} & I_{23} & I_{24} \\ I_{31} & I_{32} & I_{33} & I_{34} \\ I_{41} & I_{42} & I_{43} & I_{44} \end{bmatrix}$$

$$W = \begin{bmatrix} W_{11} & W_{12} & W_{13} & W_{14} \\ W_{21} & W_{22} & W_{23} & W_{24} \\ W_{31} & W_{32} & W_{33} & W_{34} \\ W_{41} & W_{42} & W_{43} & W_{44} \end{bmatrix}$$

Where $W_{ij} = W(x_i) \bullet W(y_j)$

According to Eqs (7) and Fig. II-2, we have the following relationships:

$$\text{x direction} \begin{cases} W(x_1) = W(1+\Delta x) = -\Delta x + 2\Delta x^2 - \Delta x^3 \\ W(x_2) = W(\Delta x) = 1 - 2\Delta x^2 + \Delta x^3 \\ W(x^3) = W(1-\Delta x) = \Delta x + \Delta x^2 - \Delta x^3 \\ W(x_4) = W(2-\Delta x) = -\Delta x^2 + \Delta x^3 \end{cases} \tag{8}$$

$$\text{y direction} \begin{cases} W(y_1) = W(1+\Delta y) = -\Delta y + 2\Delta y^2 - \Delta y^3 \\ W(y_2) = W(\Delta y) = 1 - 2\Delta y^2 + \Delta y^3 \\ W(y_3) = W(1-\Delta y) = \Delta y + \Delta y^2 - \Delta y^3 \\ W(y_4) = W(2-\Delta y) = -\Delta y^2 + \Delta y^3 \end{cases} \tag{9}$$

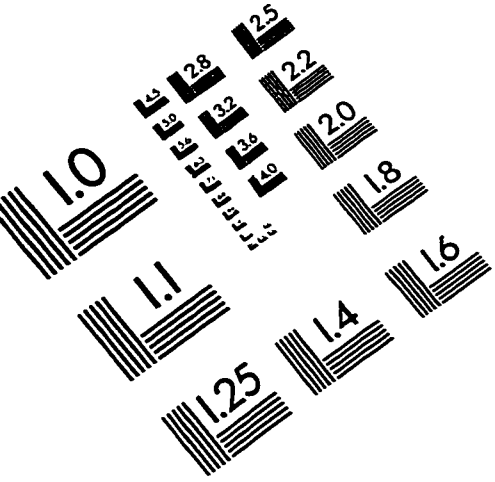
$$\begin{cases} \Delta x = x - \text{int}eger(x) \\ \Delta y = y - \text{int}eger(y) \end{cases} \tag{10}$$

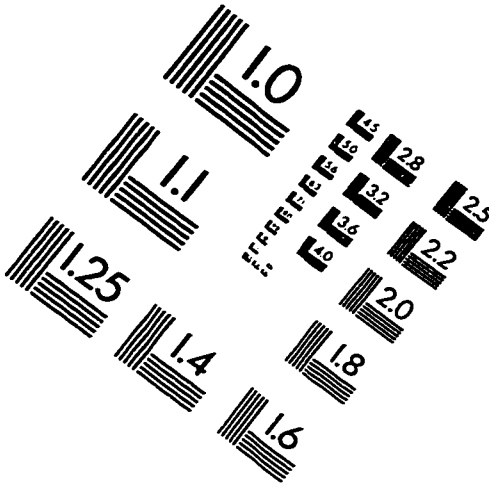
The mean square error of resampling by the cubic spline function is approximately 1/3 of that by bi-linear interpolation. However, this will involve more computational work.

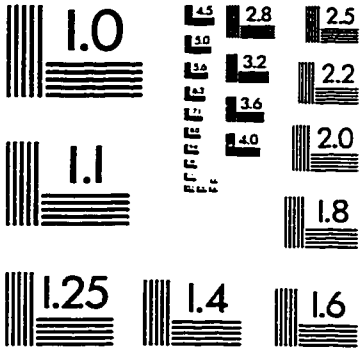
Of the three methods mentioned above, the method of nearest pixel is the simplest. It is distinguished for its fast computation speed, but its geometric accuracy is rather weak,

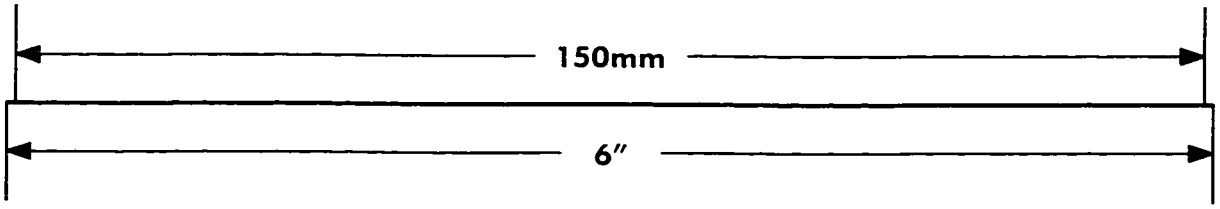
with the maximum accuracy attaining to 0.5 pixel. The other two methods have better geometric accuracy but longer computation time. The method of bicubic convolution is especially time consuming. In general cases, it seems appropriate to use the bilinear interpolation.

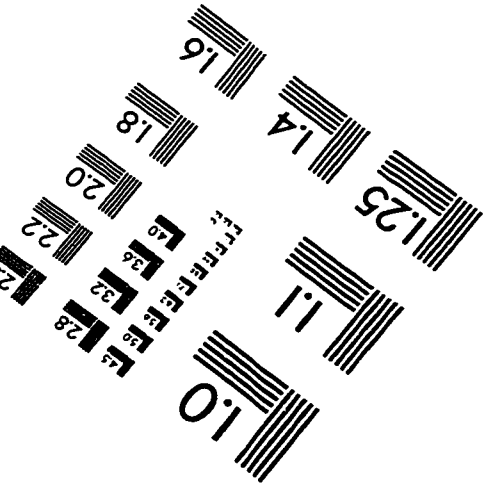
IMAGE EVALUATION TEST TARGET (QA-3)

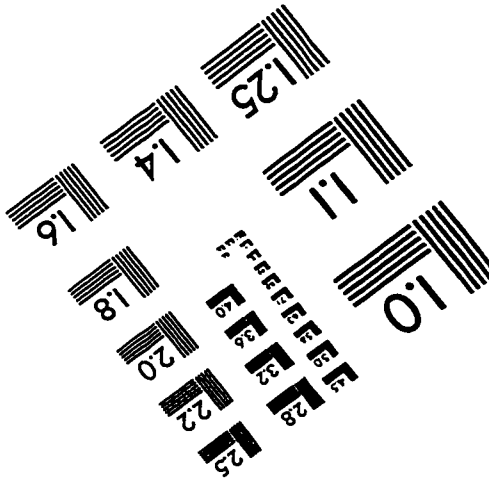












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989