

A UNIFIED APPROACH TO THE GEOMETRIC RECTIFICATION
OF REMOTELY SENSED IMAGERY

by

FRANK HAY-CHEE WONG

B.Eng., McGill University, 1969

M.Sc., Queen's University, 1971

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

IN

THE FACULTY OF GRADUATE STUDIES
(Department of Computer Science)

We accept this thesis as conforming
to the required standard.

THE UNIVERSITY OF BRITISH COLUMBIA

APRIL 1984



Frank Hay-Chee Wong, 1984

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Computer Science

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date 27 April, 1984.

ABSTRACT

Many applications of remotely sensed digital imagery require images that are corrected for geometric distortions. It is often desirable to rectify different types of satellite imagery to a common data base. A high throughput rectification system is required for commercial application. High geometric and radiometric precision must be maintained.

The thesis has accomplished the following tasks:

1. The sensors used to obtain remotely sensed imagery have been investigated and the associated geometric distortions inherent with each sensor are identified.
2. The transformation between image coordinates and datum coordinates has been determined and the values of the parameters in the transformation are estimated.
3. A unified rectification approach has been developed, for all types of remotely sensed digital imagery, which yields a high system throughput.
4. Use of digital terrain models in the rectification process to correct for relief displacement has been incorporated.
5. An efficient image interpolation algorithm has been developed. This algorithm takes into account the fact that imagery does not always correspond to sampling on a uniform grid.
6. The applications of rectified imagery such as mosaicking and multisensor integration have been studied.

7. Extension of the rectification algorithm to a future planetary mission has been investigated.

The sensors studied include TIROS-N, Landsat-1, -2 and -3 multispectral scanners, Seasat synthetic aperture radar, Landsat-4 thematic mapper and SPOT linear array detectors. Imagery from the last two sensors is simulated.

TABLE OF CONTENTS

	PAGE
ABSTRACT	ii
TABLE OF CONTENTS	iv
LIST OF TABLES	x
LIST OF FIGURES	xi
ACKNOWLEDGEMENT	xvi
1. INTRODUCTION	1
1.1 Thesis Overview	1
1.2 Problem Statement	3
1.3 Background of Remote Sensing Satellites	6
1.4 Thesis Outline	11
2. INVESTIGATION OF AN APPROACH	13
2.1 Introduction	13
2.2 Geometric Transformation / Navigation	16
2.2.1 Weighted Arithmetic Mean	17
2.2.2 Polynomial Transformation	18
2.2.3 Spline Functions	19
2.2.4 General Time Series Model	20
2.2.5 Time Series Model with Recursive Estimator	21
2.3 Comparison of Geometric Transformation / Navigation	
Methods	23
2.3.1 Distribution of GCPs	23
2.3.2 Transformation Accuracy	25
2.3.3 Unique Advantages of Time Series Approach	26
2.3.4 Unique Advantages of a Recursive Estimator	27

2.3.5	Summary	27
2.4	Remapping and Interpolation	28
2.4.1	Background	28
2.4.2	Bilinear Remapping	29
2.4.3	Two-Pass Processing	31
2.4.4	Three-Pass Processing	32
2.5	Interpolation Kernel	38
2.5.1	Outline of Interpolation Methods	38
2.5.1.1	Nearest Neighbour and Linear Interpolations	38
2.5.1.2	Cubic Spline	39
2.5.1.3	B-Spline	42
2.5.1.4	The Sampling Theorem	43
2.5.1.5	Cubic Convolution	44
2.5.1.6	Comparison of Interpolation Kernels	45
2.5.2	Preprocessing	47
2.5.2.1	Tabulation of Sinc Function	47
2.5.2.2	Kernel Size Determination	51
2.5.2.3	Windowing	52
2.5.3	Nonuniformly Spaced Samples	55
2.6	System Data Flow	59
2.6.1	Auxiliary Data Processing	61
2.6.2	Relief Displacement Correction	62
2.7	Chapter Summary	65
3.	DIGITAL TERRAIN MODELS	67
3.1	Introduction	67

3.2	Acquisition of DTMs	67
3.3	Types and Interpolation of DTMs	69
3.3.1	Irregular DTMs	70
3.3.2	Regular DTMs	71
3.3.3	Triangulated Irregular Network	72
3.4	DTM Accuracy	73
3.4.1	Accuracy Requirement in Interpolated Coordinates	73
3.4.2	DTM Interpolation in Scrolling Buffer	78
4.	SCANNER IMAGERY	83
4.1	Introduction	83
4.2	Overview of Geometric Effects	88
4.2.1	Common Distortions	88
4.2.2	Landsat-4 TM Imagery Specific Distortions	92
4.2.3	SPOT Imagery Specific Distortions	92
4.2.4	Relief Displacement	93
4.3	Rectification Transformation	95
4.3.1	Closed Form Solution	95
4.3.2	Block Size	97
4.4	Recursive Estimator in Navigation	98
4.4.1	State Vector Parameters	98
4.4.2	Extension to TIROS-N Imagery	101
4.4.2.1	Objective	101
4.4.2.2	TIROS-N Characteristics	102
4.4.2.3	Orbit Accuracy Analysis	105
4.4.2.4	Attitude Accuracy Analysis	106
4.4.2.5	Experimental Results and Discussions	107

4.5	Experiments — Interpolation in the Presence of Scan Gap	114
4.5.1	Objective	114
4.5.2	Theory of Proposed Method	117
4.5.3	Implementation Considerations	122
4.5.4	Experiments with Sine Waves	124
4.5.5	Scan Gap Simulation in Imagery	129
4.5.6	Experiments with Sine Wave Imagery	134
4.5.7	Synthesized TM Imagery	148
4.6	Experiments — Relief Displacement	154
4.6.1	Objective	154
4.6.2	Gaussian Terrain with Landsat-4 TM Sensor Parameters	155
4.6.3	Real Terrain with Landsat-4 TM Sensor Parameters	162
4.6.4	Real Terrain with SPOT PLA Sensor Parameters	166
4.7	Experiments with Landsat-2 Imagery	170
5.	SYNTHETIC APERTURE RADAR IMAGERY	181
5.1	Introduction	181
5.2	Rectification Transformation	183
5.3	Terrain Induced Distortions	184
5.3.1	Relief Displacement	184
5.3.2	Importance of Compression to Zero Doppler	189
5.3.3	Target Defocusing	191
5.3.4	Intensity Migration	191
5.4	Experiments with Seasat Imagery	194
5.4.1	Image Location Accuracy Study	194
5.4.2	Orbit Refinement	198

5.4.3	Rectification with DTM	201
5.5	Navigation by Automatic Focusing	208
6.	APPLICATIONS	213
6.1	Introduction	213
6.2	Multisensor Integration	213
6.3	Image Mosaicking	227
6.3.1	Linear Interpolation	227
6.3.2	Feathering	229
6.3.3	Filtering	231
6.3.4	Experiments	233
6.4	A Future Planetary Mission	237
6.4.1	Mission Description	237
6.4.2	Navigation	237
6.4.3	Venus Atmospheric Refraction	239
6.4.4	Use of Altimeter Data	242
6.4.5	Digital Terrain Model	243
7.	CONCLUSIONS AND RECOMMENDATIONS	245
7.1	Conclusions	245
7.2	Recommendations for Future Work	247
	BIBLIOGRAPHY	253
	APPENDIX A LANDSAT-4 SENSORS	259
	APPENDIX B SPOT SENSORS	273
	APPENDIX C RECURSIVE ESTIMATOR FILTER	277
	APPENDIX D BILINEAR AND AFFINE TRANSFORMATIONS	282
	APPENDIX E CLOSED FORM TRANSFORMATION FOR SCANNER IMAGERY	286

APPENDIX F	THEORY OF SAR AND SAR DATA PROCESSING	294
APPENDIX G	CLOSED FORM RECTIFICATION TRANSFORMATION	312
APPENDIX H	REFRACTION OF ELECTROMAGNETIC WAVE	316

LIST OF TABLES

TABLE	PAGE
1.1 Values of Relief Displacement	5
3.1 Resolution and Worst Relief Displacement Factor	75
3.2 DTM Resampling in Scrolling Buffer (I)	81
3.3 DTM Resampling in Scrolling Buffer (II)	82
4.1 TIROS-N Characteristics	103
4.2 Error Analysis of TIROS-N Navigation	115
4.3 RMS Error Versus Scan Gap Size Quantization	125
4.4 Attitude at 9 Points over the Image	178
5.1 Along Track and Across Track Errors	196
5.2 Sensitivity of FM Rate to Parameter Variations in Seasat Imagery	210
6.1 Sensitivity of FM Rate to Parameter Variations in VOIR Imagery	238

LIST OF FIGURES

FIGURE	PAGE
1.1 Typical Satellite Orbit and Ground Track	7
1.2 An Evolution Cycle of Satellite Sensor Definition	10
2.1 Recursive Estimation	24
2.2 Two One-Dimensional Remapping and Interpolation Processes	33
2.3 Three One-Dimensional Remapping and Interpolation Processes	36
2.4 Coupling of Passes	37
2.5 Nearest Neighbour and Linear Interpolation --- Kernels and Frequency Responses	40
2.6 RMS Error and Frequency Response for Cubic Convolution	46
2.7 RMS Error and Frequency Response for Sinc Function Interpolation	48
2.8 Approximation of Sinc Function by Step Function	50
2.9 Examples of Autocorrelation Function	53
2.10 RMS Interpolation Error Versus Kernel Size	54
2.11 Four-Point Sinc Function and Its Frequency Response	56
2.12 Four-Point Sinc Function with Kaiser Window and Its Frequency Response	57
2.13 Sixteen-Point Sinc Function with Kaiser Window and Its Frequency Response	58
2.14 System Data Flow	60
2.15 First Remapping and Interpolation Pass	64
3.1 DTM Accuracy Requirement for Landsat-4 TM Sensor	76
3.2 DTM Accuracy Requirement for SPOT Sensors, Nadir Looking	76
3.3 DTM Accuracy Requirement for SPOT Sensors, Off Nadir Looking Angle = 26°	77

3.4	DTM Accuracy Requirement for Seasat SAR Sensor	77
4.1	Scanner Sensor System	84
4.2	Typical Mechanical Scanning Sensor	85
4.3	Typical Electronic Scanning Sensor	86
4.4	Earth Rotation Effect	91
4.5	Relief Displacement in Scanner Imagery	94
4.6	Scan Vector Geometry	96
4.7	Accuracy Versus Block Size for Landsat Sensors	99
4.8	Accuracy Versus Block Size for SPOT Sensors	100
4.9	Resolution Versus Distance from Centre for TIROS-N Imagery	104
4.10	Uncertainties and Residual Error Versus Number of GCPs Processed, Verification Test 1	109
4.11	Verification Test 2	111
4.12	Operational Run	111
4.13	Interpolation Kernel of Yen's Method	119
4.14	RMS Error Versus Interpolation Position	127
4.15	Comparison of Methods — RMS Error Versus Scan Gap Size	130
4.16	Effect of Overlap and Underlap on a Roadway	132
4.17	Jitter Model	133
4.18	Sine Wave Image	135
4.19	Distorted Sine Wave Image with Overlap	136
4.20	Rectified Sine Wave Image (Overlap) with Sinc Interpolation	137
4.21	Rectified Sine Wave Image (Overlap) with Proposed Method	138
4.22	Radiometric Error in Rectified Sine Wave Image (Overlap) with Sinc Interpolation	139
4.23	Radiometric Error in Rectified Sine Wave Image (Overlap) with Proposed Method	140

4.24 Distorted Sine Wave Image with Underlap	142
4.25 Rectified Sine Wave Image (Underlap) with Sinc Interpolation	143
4.26 Rectified Sine Wave Image (Underlap) with Proposed Method	144
4.27 Radiometric Error in Rectified Sine Wave Image (Underlap) with Sinc Interpolation	145
4.28 Radiometric Error in Rectified Sine Wave Image (Underlap) with Proposed Method	146
4.29 Error Magnitude	147
4.30 MSS Image Used to Synthesize TM Image	149
4.31 Radiometric Error in Rectified Synthesized TM Image with Sinc Interpolation	150
4.32 Radiometric Error in Rectified Synthesized TM Image with Proposed Method	151
4.33 RMS Error Versus Scan Gap Size in Simulated TM Image	153
4.34 Gaussian Terrain	156
4.35 Orthographic Image (Gaussian Terrain)	157
4.36 Distorted Image (Gaussian Terrain)	158
4.37 Rectified Image (Gaussian Terrain)	159
4.38 Cross-Sectional Plot of a Hill	160
4.39 Error Image (Gaussian Terrain)	161
4.40 DTM of St. Mary Lake Area	163
4.41 Landsat-4 TM Orthographic Image	164
4.42 Landsat-4 TM Relief Distorted Image	165
4.43 Landsat-4 TM Rectified Image	167
4.44 Geometric Rectification Accuracy (Landsat-4 TM)	168

4.45 Error Image (Landsat-4 TM)	169
4.46 SPOT Orthographic Image	171
4.47 SPOT Relief Distorted Image	172
4.48 SPOT Rectified Image	173
4.49 Geometric Rectification Accuracy (SPOT)	174
4.50 Error Image (SPOT)	175
4.51 Landsat-2 MSS Raw Image	176
4.52 Residual Error Versus Number of GCPs	178
4.53 Rectified Landsat-2 MSS Image	180
5.1 Geometric Relation Between Slant Range and Ground Range	185
5.2 Plot of $\Delta g/\Delta s$ Versus Range	186
5.3 Relief Displacement in Radar Imagery	187
5.4 Relief Displacement and Critical Angle	190
5.5 Predicted and Actual Target Trajectories of an Elevated Target	192
5.6 Intensity Migration in Radar Imagery	193
5.7 GCP Error Vectors After Navigation	200
5.8 Block Size Versus Remapping Accuracy	202
5.9 Raw Seasat Image	204
5.10 Rectified Seasat Image with DTM	205
5.11 Rectified Seasat Image without DTM	206
5.12 Effect of Azimuth Misfocusing	209
5.13 Navigation Loop Using Automatic Focusing	212
6.1 Seasat Component in Landsat/Seasat Composite	215
6.2 Landsat Component in Landsat/Seasat Composite	216
6.3 Landsat/Seasat Composite (12 Km x 12 Km) of the Ottawa Area	217

6.4 Histogram of Product of two Random Variables	219
6.5 Regions of Interest in the Composite	221
6.6 Landsat/Seasat Composite (40 Km x 40 Km) of the Ottawa Area	223
6.7 Seasat Image of Vancouver, NTS Mapsheet 92G/3	224
6.8 Landsat/Seasat Composite of Vancouver, NTS Mapsheet 92G/3	225
6.9 NTS Mapsheet, 92G/3	226
6.10 Radiometric Mosaicking by Linear Interpolation	228
6.11 Filtering Method of Image Mosaicking	232
6.12 Mosaic with no Overlap	234
6.13 Mosaic by Linear Interpolation	235
6.14 Mosaic by Filtering	236
6.15 Refraction Geometry	240
6.16 Index of Refraction Versus Altitude	240

ACKNOWLEDGEMENT

I wish to thank my supervisor, Dr. Robert J. Woodham, for his valuable support, guidance, advice and encouragement throughout my years at the University of British Columbia. Dr. Woodham provided me with the digital terrain model of the St. Mary Lake area. He also imparted to me his knowledge of the synthesis of satellite borne imagery. Thanks are also due to Dr. Alan K. Mackworth, also of the University of British Columbia, for introducing me to Moire contourography and for widening my bird's eye view of artificial intelligence. Manual digitization of the digital terrain models was performed by Mr. James Little and Mr. Ezio Cantanzariti of UBC using software originally supplied by Dr. Thomas Peucker of Simon Fraser University. Mr. Little also provided me with a digital terrain model of the Shawnigan Lake area.

Sincere appreciation is extended to the following colleagues at MacDonald, Dettwiler and Associates Ltd. (MDA). I am extremely indebted to Dr. Robert Orth, without whose encouragement and permission I would never have undertaken this research program. He has guided me technically throughout this research work. In addition he has consistently shown deep concern for my career development at MDA and has taught me many other aspects of life. Dr. Orth is more than a friend and a colleague to me. I am equally indebted to Dr. John Bennett for his technical guidance in the Synthetic Aperture Radar (SAR) area. He has continuously provided me, even though I am not formally working in his section, with SAR projects which are relevant to the thesis topic. I am also grateful to the members of his SAR team, especially to Dr. Ian

Cumming, Mr. Ron Fielden and Mr. Peter McConnell for their valuable discussions on the processing of SAR data and for providing me with SAR imagery. Thanks are also due to Mr. Daniel Friedmann for his discussion on interpolation, and to Mrs. Patricia Palmer for developing part of the software used to correct TM imagery.

Last, but certainly not least, I wish to acknowledge the Graduate Research and Engineering Technology award received from the Science Council of British Columbia and to thank MDA for allowing me the free and extensive use of its research facilities. Consequently thanks are due to Dr. John MacDonald, president of MDA, for giving me the free privilege and for allowing me to pursue my academic interests while employed at MDA.

CHAPTER 1 INTRODUCTION

1.1 Thesis Overview

This thesis is concerned with the geometric rectification of remotely sensed digital imagery including the incorporation of terrain data from digital terrain models. The imagery studied includes that obtained from existing satellites in the Landsat series, the NOAA series and Seasat. It also includes that to be obtained from the proposed SPOT satellite. Landsat imagery and NOAA are obtained by mechanical scanning sensors while synthetic aperture radar (SAR) imagery is obtained by a radar sensor. SPOT imagery will be obtained by a linear array sensor.

The thesis develops a framework for an integrated image processing system with the flexibility to deal with different sensors, to yield high throughput, to retain high geometric and radiometric fidelity and to be operationally viable. The thesis combines a review and an evaluation of previous work, extending it when applicable, with the development of novel techniques to deal with problems previously not encountered. Specifically, the following is dealt with in this thesis:

- (a) The sensors used to obtain image data are investigated and the associated geometric distortions inherent with each sensor are identified. Particular attention is given to Landsat-4, SPOT and Seasat SAR. The geometric distortions in Landsat-4 Thematic Mapper (TM) imagery and in SPOT panchromatic linear array (PLA) and multispectral linear array (MLA) imagery are shown to be

different from those in Landsat-1, -2 and -3 multispectral (MSS) imagery. The geometric distortions in Seasat SAR imagery are also different from those in MSS and linear array imagery. These distortions depend on the method of processing the Seasat SAR data.

- (b) The transformation between image coordinates and datum coordinates is determined. The values of the parameters in the transformation are estimated. The orbit and attitude parameters are of particular importance. Their values can be determined with the aid of ground control points (GCPs). In SAR imagery, only the orbit parameters are important. A method to determine their values by automatic focusing is discussed.
- (c) A unified rectification approach is developed, for all types of remotely sensed digital imagery, which can be implemented in a conventional image processing system and which yields a high system throughput. High throughput is achieved with one-dimensional processing of imagery which is represented by a two-dimensional array of digital values.
- (d) Digital terrain models (DTMs) are incorporated into the rectification process to correct for relief displacement. Correction for relief displacement is also achieved by one-dimensional processing.
- (e) An efficient image interpolation algorithm is developed. This algorithm takes into account the fact that imagery does not always correspond to sampling on a uniform grid. This is required for Landsat-4 Thematic Mapper (TM) imagery where scan gaps are present.

- (f) Two applications of rectified imagery are presented: image mosaicking and multisensor integration. In image mosaicking, two images bordering on the same area are merged into a single image. In multisensor integration, two images from different sensors, Seasat SAR and Landsat-2 MSS, are combined into a single image.
- (g) Extensions to the rectification approach are considered for a future planetary mission.

1.2 Problem Statement

Geometric rectification is the warping of raw image data to a predefined datum such as a particular map projection. Many applications of remotely sensed imagery require rectification to a common datum. One example is the production of Seasat/Landsat composites (to be shown in Chapter 6) which combine the high resolution of Seasat with the spectral content of Landsat. Rectification is difficult since different satellite sensors have different orbit parameters and geometric attributes.

Rectification of satellite imagery involves two steps:

- (a) Geometric transformation --- determination of transformation between input imagery and datum. The transformation depends on the following parameters: satellite platform and orbit data, terrain height, earth curvature and rotation rate. The transformation is normally established with the aid of GCPs.
- (b) Interpolation --- sampling of image data using a specific

kernel; e.g., spline or sinc function. The method of interpolation depends upon the geometric transformation algorithm.

Rectification has been dealt with by many other authors [BAKE75, EBNE76, HORN79, KONE76, krat72, KRAU78, LARS80, LECK80, ORTH78a, ORTH78b, OTEP78 and SIMO75a]. Previous work collectively shares the following deficiencies when considered for the general case:

- (a) Relief displacement is not accounted for. Table 1.1 shows the pixel shifts that would appear due to relief displacement for various satellite imagery at maximum scan angle. While relief displacement is not significant in Landsat-1, -2, and -3 imagery, it is significant in Seasat SAR imagery and Landsat-4 TM imagery and will be significant in SPOT imagery. Terrain information is required for geometric transformation.
- (b) The transformation and interpolation algorithms presented do not address the issue of system throughput. Throughput is of prime importance in TM and SPOT imagery since each scene contains or will contain as much as 6000 lines by 6000 pixels per line compared to approximately 3000 lines by 3000 pixels per line in an MSS scene.
- (c) The rectification algorithms are sensor and satellite specific. Imagery from different sensors cannot be rectified in a single shared system.
- (d) The geometric transformation algorithm is often of generalized form (e.g., polynomial) and does not model the imaging geometry

TABLE 1.1. VALUES OF RELIEF DISPLACEMENT

Satellite sensor	Altitude	Maximum scan angle	Nominal pixel spacing	Target elevation for 1 pixel displacement
Landsat-1, -2 -3 MSS	902 km	5.78°	58 m	507 m
Landsat-4 MSS	705 km	7.47°	58 m	406 m
Landsat-4 TM (Bands 1, 2, 3, 4, 5 and 7)	705 km	7.47°	30 m	210 m
Landsat-4 TM (Band 6)	705 km	7.47°	120 m	840 m
SPOT MLA (nadir mode)	820 km	2.12°	20 m	472 m
SPOT MLA (side looking mode)	820 km	28.12°	26 m	44 m
SPOT PLA (nadir mode)	820 km	2.12°	10 m	236 m
SPOT PLA (side looking mode)	820 km	28.12°	13 m	22 m
Seasat SAR	820 km	20.00°	25 m	11 m

explicitly. Subpixel rectification accuracy requires high quality GCPs which are difficult to establish. Algorithms which require a small number of GCPs are preferable.

(e) The problem of scan gap in TM imagery is new.

This thesis addresses these deficiencies explicitly. Each step is considered and the effect of terrain, if any, is investigated.

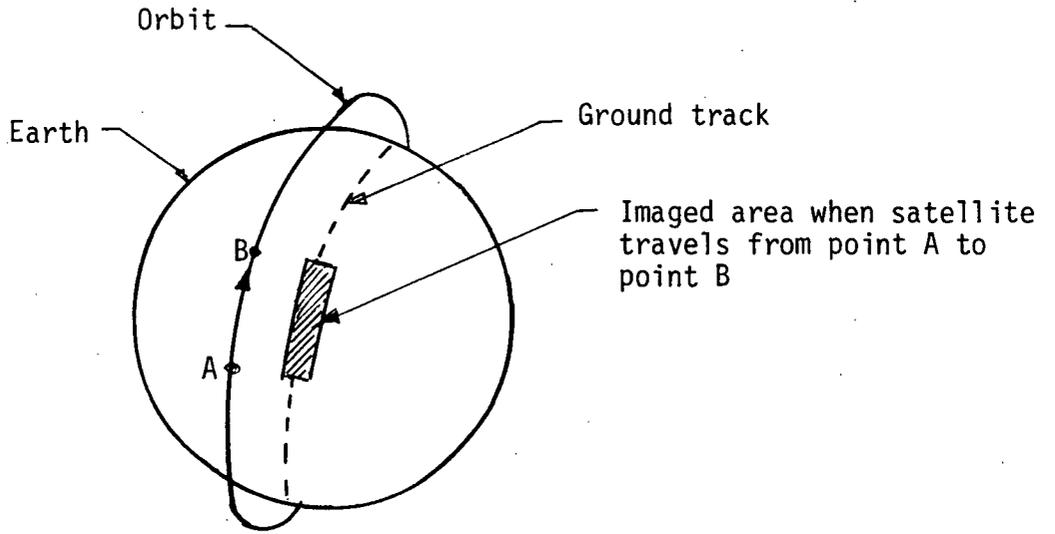
1.3 Background of Remote Sensing Satellites

Satellite imagery is obtained from a platform orbiting around the earth as shown in Figure 1.1. The combined effect of earth rotation and satellite velocity allows the satellite to give repetitive coverage of the entire globe. This produces a large volume of remotely sensed data in a reasonably short time, and above all, at a low cost.

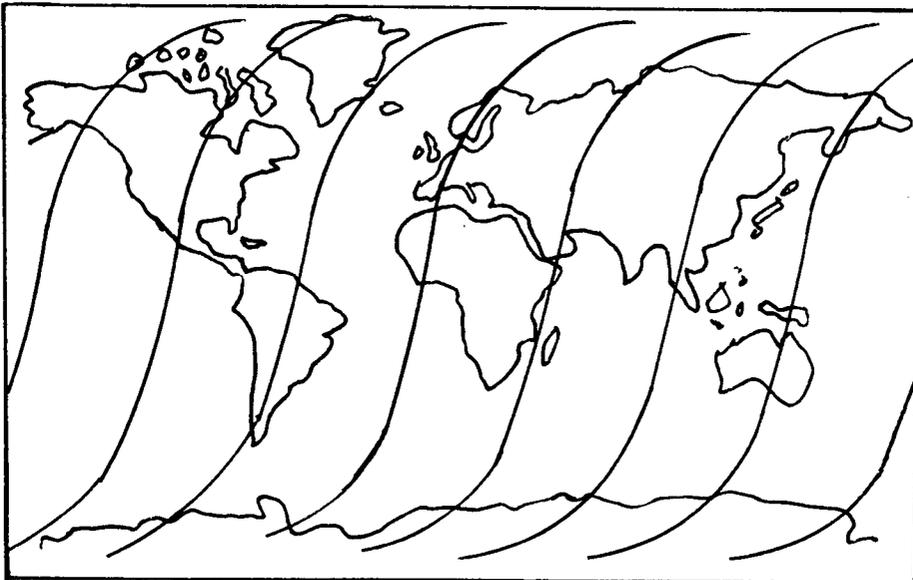
The first generation of remote sensing satellites began in the 1960s with the TIROS and NIMBUS series [NASA76a] which offered a spatial resolution of about 1 km. Despite the low spatial resolution, the imagery obtained demonstrated the potential of viewing the earth from space.

The second generation of satellites was the Landsat [NASA76b] series launched in the 1970s. Each satellite recorded imagery in four spectral bands with resolution¹ greater than 100 metres. Currently, data from these satellites are still being exploited. Landsat imagery has demonstrated the utility of remotely sensed satellite data in earth

¹ Resolution increases with a decrease in its numerical value.



(a) ORBITING PLATFORM



(b) GROUND TRACK

FIGURE 1.1 TYPICAL SATELLITE ORBIT AND GROUND TRACK

resources applications such as geology, agriculture, forestry and oceanography. Application is also found in cartography. The Landsat series, although not designed for mapping, was capable of generating imagery for presentation at a scale of 1:250,000. Precision processing of Landsat imagery has been a first step in mapping from satellite borne imagery.

Also in the late 1970s, a civilian SAR system (Seasat [MDA76]) was deployed and this gave a view of the earth in the radio frequency portion of the electromagnetic spectrum. Seasat SAR data has been precision processed to give 25 metres resolution. Although Seasat was operational for only four months, all the data collected has yet to be completely processed.

Landsat-D [GSFC82] was launched in July 1982 becoming Landsat-4. A key feature in Landsat-4 is its bidirectional scanning TM sensor¹ with seven spectral bands and a spatial resolution of 30 metres in six of these channels. The data should support mapping at a scale of 1:100,000 and will be used for thematic interpretation. Due to the bidirectional scanning of the TM sensor and its high resolution, the resulting imagery suffers from geometric distortion more severely than imagery obtained from the earlier Landsat satellites.

The next generation of satellites will include SPOT² [CNES81] which will provide a spatial resolution from 10 to 27 metres. SPOT's design also provides stereo capability. Large areas of the earth remain to be mapped precisely and SPOT imagery will be used to prepare and update topography maps at a scale of 1:100,000. SAR systems have been carried on

¹ See Appendix A.

² See Appendix B.

the space shuttles. In Canada, a SAR system (RADARSAT) is planned for ice study.

The trend to use satellite imagery in mapping in addition to conventional aerial photography is gaining support. Satellites are a more cost effective source of cartographic information compared to conventional aerial photography. The digital nature of satellite imagery can also render it easier to handle and to process. Colvocoresses [COLV79] has advocated a mapping satellite system (MAPSAT) to be deployed in the 1980s. It will also have the stereo capability. Both SPOT and the proposed MAPSAT will use solid state linear array detectors which can produce imagery of high geometric fidelity.

Another synthetic aperture radar system has been proposed by NASA to map the surface of the planet Venus [JAME82]. The proposed system will be launched in the late 1980s or in the 1990s.

In a short time span, satellite imaging systems have evolved from low spatial resolution to high spatial resolution, have increased the number of spectral channels and spectral resolution, have extended outside the visible portion of the spectrum to include near infrared, thermal infrared and radio frequencies, and have begun to use solid state linear array detectors. This evolution proceeds in cycles as illustrated in Figure 1.2. Each cycle starts with the user definition of sensor requirements, including the number of spectral bands, the spectral response and spatial resolution. Then a satellite sensor is developed and launched. Data acquired by the sensor are transmitted to a ground station where they

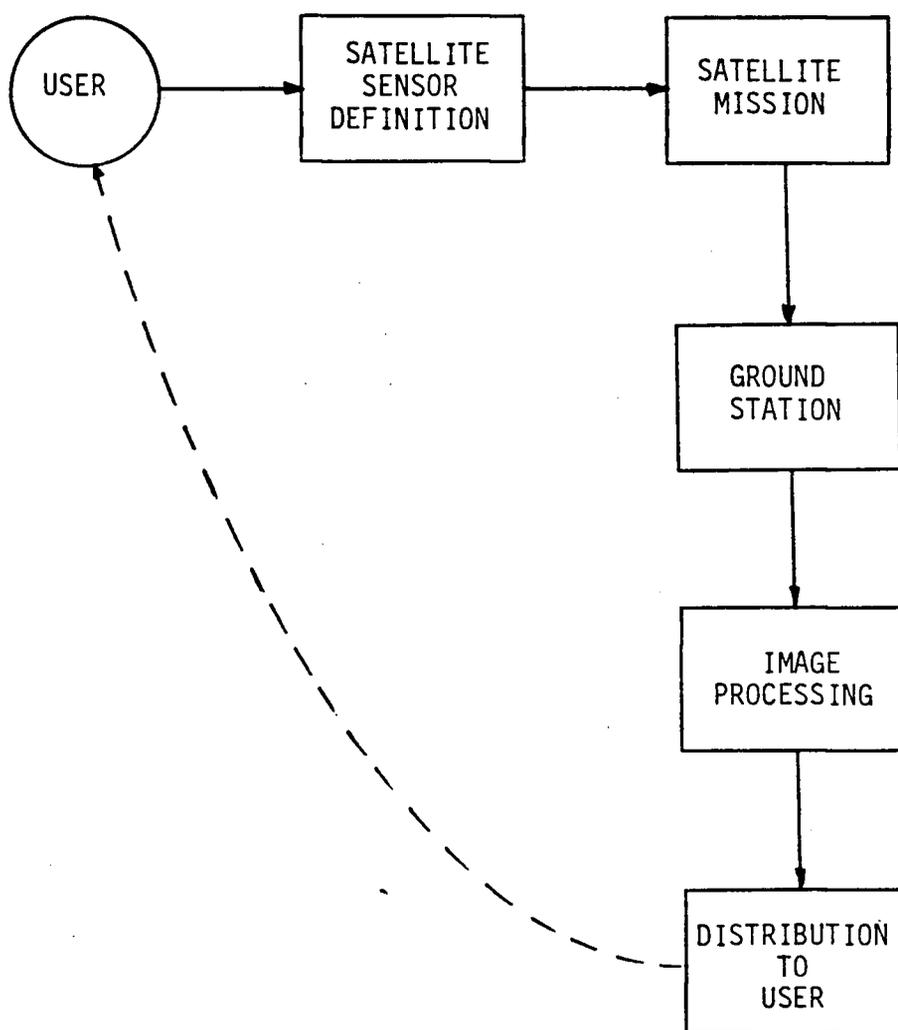


FIGURE 1.2 AN EVOLUTION CYCLE OF SATELLITE SENSOR DEFINITION

are received and processed. Finally, image products are distributed to the user. Experience gained using the imagery allows the user to define requirements for the next generation of satellites.

1.4 Thesis Outline

This introductory chapter has defined the problem. The rest of the thesis is organized as follows:

Chapter 2 gives the methodology for the rectification of remotely sensed imagery. It presents methods for geometric transformation and interpolation and demonstrates the use of DTMs in the rectification process. The methods yield high throughput. The overall geometric transformation is coefficient driven. Image interpolation is performed by one-dimensional processing. Different types of imagery are handled.

Chapter 3 describes DTMs, their acquisition, types, accuracies and methods of interpolation.

Chapter 4 discusses scanner and linear array imagery. It contains a review of related work in the area of image rectification and how GCPs are used. The attitude modelling method (the attitude information is in the rectification geometric transformation) is tested on both Landsat and TIROS-N imagery. Finally, synthesized Landsat-4 and SPOT imagery is used to show the specific geometric distortions including relief displacement effects. Rectifications using the proposed algorithm are performed on the synthesized imagery as well as on Landsat-2 MSS imagery.

Chapter 5 discusses SAR imagery. Unlike scanner and linear array imagery, the raw data of SAR imagery need to be processed to be presented in image form. This chapter starts with an introduction to the theory of SAR and then it outlines the geometric transformation. It also shows how GCPs are used to update the satellite orbit data. Finally, the problem of relief displacement is examined and results of rectifying a Seasat SAR scene using a DTM of the same area are presented.

Chapter 6 presents some applications with rectified imagery such as image mosaicking and multisensor integration. It also extends the method to imagery obtainable from a proposed planetary mission.

Chapter 7 concludes the thesis with discussions of the method presented, the experimental results and recommendations for future research work.

CHAPTER 2 INVESTIGATION OF AN APPROACH

2.1 Introduction

This chapter first discusses and compares various rectification approaches, then proposes and investigates one which can handle remotely sensed imagery from different satellites.

Rectification is the warping of an input image to an output grid. The input image is seen by a satellite sensor and hence contains inherent geometric distortions caused by earth curvature and rotation, satellite attitude variation across the image, panoramic and relief distortions. The output grid is a user defined map projection. Rectification involves two steps: geometric transformation and interpolation.

Since the final result of rectification is an image in the output coordinates, it is necessary for the geometric transformation to map the output image coordinates (x,y) to the input image coordinates (u,v) . The transformation is represented by the general form:

$$u = f_a(x,y) \tag{2.1a}$$

$$v = g_a(x,y) \tag{2.1b}$$

The functions f_a and g_a can be represented in many forms, the simplest being a pair of low order polynomials and the more elaborate being a pair of complicated algebraic functions explicitly taking parameters of the imaging geometry and earth rotation into account. Regardless of the representation, function parameters and coefficients need to be determined

and this normally requires the use of GCPs. In this thesis the determination of these parameters is called navigation¹. A typical example of parameters is the set of attitude angles of the satellite platform.

The output image may contains millions of data points. To perform the mapping defined by Equations 2.1a and b for all the data points may be too time consuming for practical purposes. Therefore it is useful to find a fast mapping:

$$u = f_b(x,y) \quad (2.2a)$$

$$v = g_b(x,y) \quad (2.2b)$$

such that f_b and g_b approximate f_a and g_a and this new form renders a fast computation. In this thesis the establishment of f_b and g_b is called remapping. It must be emphasized that f_b and g_b are not in the same functional form as f_a and g_a , especially when f_a and g_a are complicated algebraic functions. But f_a and g_a must be established first and they are used to determine f_b and g_b .

Interpolation determines an intensity value at each output image integral coordinate (x,y) . Integral values of (x,y) do not necessarily map into integral values of (u,v) . Interpolation determines the intensity at each mapped, but non-integral (u,v) . The convolution method is used in this thesis. If the input imagery data are band limited, then the convolution kernel (also called the interpolation kernel) should correspond to that of an ideal low pass filter [LATH65]. Such a filter can only be realized with an infinite kernel size and hence is not

¹ The term "navigation" is used in this thesis to avoid a lengthy phrase "Using GCPs to determine the parameter values in the geometric transformation given by f_a and g_a in Equation 2.1".

practical for direct implementation. A practical kernel has to be designed so that high radiometric accuracy is retained in the interpolation process.

Image interpolation is a two-dimensional convolution process (i.e., the convolution kernel is a two-dimensional spatial filter). This can be a slow process for two reasons: the number of arithmetic operations is large and many physical image read/write operations may be required. However by carefully choosing the remapping functions f_b and g_b , interpolation is reduced to several one-dimensional processes while retaining the same degree of radiometric accuracy as in the two-dimensional interpolation case.

This chapter presents a navigation algorithm which yields high geometric precision and which uses relatively few GCPs per image scene. This chapter also presents a fast interpolation algorithm, including the use of DTM data. Remapping is table (or coefficient) driven so that it has the flexibility to handle a variety of satellite imagery. Remapping and interpolation are implemented in several one-dimensional processing passes. The result is a high throughput system for rectifying various kinds of satellite borne imagery, including relief displacement correction. Furthermore, radiometric fidelity is preserved since one-dimensional processing allows the use of a large interpolation kernel. The image is stored in a conventional manner (i.e., successive lines of imagery are stored in consecutive records on disk or tape).

Sections 2.2 and 2.3 present and compare various navigation methods. Section 2.4 discusses remapping and interpolation. Section 2.5

develops a suitable choice of interpolation kernel.

Finally, Section 2.6 concludes with an illustration of the flow of data (image and DTM) through a typical system configured to use the algorithms developed. Though the geometry of different sensors is different, the methodology and software performing the rectification is the same.

2.2 Geometric Transformation / Navigation

Procedures developed for rectification of satellite imagery are based on the computation of geometric coefficients of mathematical models that characterize the distortions in the image. The distortions are represented by functions f_a and g_a in Equations 2.1a and b. After obtaining values for these coefficients with the aid of GCPs in the navigation process, the mapping function to be used in the remapping process is determined. In the literature, numerous algorithms have been reported [BAKE75, BERN76, EBNE76, HORN79, KONE76, KRAT72, KRAU78, LARS80, LECK80, LITT80, ORTH78a, ORTH78b, and SIMO75a]. In general, these algorithms either use a polynomial approach or set up an attitude time series model. This section describes some of the approaches, and identifies the particular approach adopted in this thesis.

The transformations in Sections 2.2.1 to 2.2.5 refer to f_a and g_a in Equation 2.1.

2.2.1 Weighted Arithmetic Mean

The weighted arithmetic mean method calculates for any interpolation point (x,y) in the output grid the displacements $\Delta x = u - x$ and $\Delta y = v - y$, each as a function of displacements of all GCPs over the entire image. Specifically:

$$\Delta x = \frac{1}{K} \sum_{i=1}^n W_i(x,y,x_i,y_i) \Delta x_i \quad (2.3a)$$

$$\Delta y = \frac{1}{K} \sum_{i=1}^n W_i(x,y,x_i,y_i) \Delta y_i \quad (2.3b)$$

where i varies over all GCPs, Δx_i and Δy_i are displacements of the i -th GCP, n is the total number of GCPs, W_i is a weighting function and should be a monotonically decreasing function of distance between the i -th GCP and the interpolation point (x,y) , and K is the normalization constant given by:

$$K = \sum_{i=1}^n W_i \quad (2.4)$$

This is known as Shepard's method [SCHU76a] in the interpolation literature. While this method is simple, it requires that the distance weight be independent of the direction and position of the interpolation point [BAKE75 and KONE76]. This requirement is not necessarily obeyed in satellite imagery.

2.2.2 Polynomial Transformation

The polynomial approach [BERN76, KONE76, LECK80] extrapolates the displacement made at the GCPs to all other points in the image by modelling the distortion function as two polynomial functions of position within the image. The coefficients of the polynomials are estimated by least square methods. In effect, this method is a "rubber sheet fitting" technique between the image and the reference projection. This approach is independent of satellite attitude, scanner geometry and earth geometry.

The polynomial can be written as :

$$u = \sum_{p=0}^N \sum_{q=0}^{N-p} a_{pq} x^p y^q \quad (2.5a)$$

$$v = \sum_{p=0}^N \sum_{q=0}^{N-p} b_{pq} x^p y^q \quad (2.5b)$$

where N is the order of the polynomial transformation. The GCPs are used to compute the coefficients a_{pq} and b_{pq} by a least squares method.

Sometimes the image is semi-corrected for system errors such as earth rotation, panoramic distortion, nonlinear sensor sweep and scan skew. Let the semi-corrected image coordinates be (x', y') . In this case, polynomials are applied to displacement errors between (x, y) and (x', y') :

$$\Delta x = x' - x = \sum_{p=0}^N \sum_{q=0}^{N-p} a_{pq} x^p y^q \quad (2.6a)$$

$$\Delta y = y' - y = \sum_{p=0}^N \sum_{q=0}^{N-p} b_{pq} x^p y^q \quad (2.6b)$$

The weighted arithmetic mean of Section 2.2.1 is a special case of the polynomial transformation as can be demonstrated by rewriting Equations 2.3a and b as follows:

$$u = x + \sum_{i=1}^n f_i(x,y) (u_i - x_i) \quad (2.7a)$$

and

$$v = y + \sum_{i=1}^n f_i(x,y) (v_i - y_i) \quad (2.7b)$$

with

$$f_i(x,y) = W_i / \sum_{i=1}^n W_i \quad (2.7c)$$

The powers of the polynomials in x and y then depend on the functional forms of $f_i(x,y)$.

2.2.3 Spline Functions

Transformation by polynomials over the entire image often has the disadvantage that higher order polynomials are needed to ensure good fit. It is possible to achieve the same degree of fit with lower order polynomials by first segmenting the image into separate regions. A polynomial is determined for each region. These polynomials can be made continuous up to order N-1 (N is the polynomial order) at the region boundaries. These polynomials define spline functions whose coefficients are determined using GCPs and a least squares fitting method.

2.2.4 General Time Series Model

The time series approach models the displacement terms mathematically based on platform and sensor geometry. This typically involves trigonometry and other complex mathematical functions. If all the parameters are known, the satellite data can be geometrically corrected by transformation equations of the form¹ :

$$x = x [p_1, p_2, \dots, p_n, t(u,v)] \quad (2.8a)$$

$$y = y [p_1, p_2, \dots, p_n, t(u,v)] \quad (2.8b)$$

where t is the imaging time of an input pixel having coordinates (u,v) and the p_i 's ($i=1$ to n) are parameters of scanner geometry, earth rotation, satellite orbit and attitude, mirror sweep rate and the like. Unfortunately, all of the parameters are not known a priori. Unknown parameters can be expressed as polynomials in time. For example, the i -th parameter in the group is represented by an N_i -th degree polynomial:

$$p_i = \sum_{l=0}^{N_i} k_{li} t^l \quad (2.9)$$

The coefficients k_{li} are estimated using GCPs in the following manner [KONE76] using Newton's method:

- (a) Assume initial values of the coefficients.
- (b) Compute the transformation error Δx and Δy for all GCP locations.
- (c) At the same GCP locations, evaluate the partial derivatives of x and y with respect to the coefficients.
- (d) Steps (b) and (c) determine a set of simultaneous linear equations in Δk_{li} which can be solved by a least squares method.
- (e) Update k_{li} by Δk_{li} .

¹ For the discussion of determining the p_i values, it is more convenient to express x and y as functions of u and v than vice versa.

(f) Repeat Steps (b) to (e) until the values of the coefficients in the time series no longer change enough to affect the overall geometric accuracy of the transformation.

2.2.5 Time Series Model with Recursive Estimator

Another approach [MDA78, CARO75 and ORTH78a] is similar to that of Section 2.2.4 except that the GCPs drive a recursive estimator to refine the time series coefficients. The recursive estimator is a special case of the Kalman filter [Appendix A, MDA78, DUPL67, MILL71, LEON70 and UCLA79] and is discussed in Appendix C.

The recursive estimate incorporates the GCPs one at a time into the estimator, updating the error model (time series) after each. The first GCP allows removal of some of the image errors, and each following GCP refines the geometric precision. More importantly, the uncertainty in the predicted location of each new GCP decreases due to the incrementally improved estimate. This has distinct operational advantages, both in searching for the next GCP location and in detecting pointing errors associated with the current GCP.

The estimator is a set of recursive equations for optimally, in the least squares sense, estimating the state vector containing variables which are obtained from noise corrupted measurement data. In the application here, the state vector consists of the time series coefficients. The recursive equations (derived in Appendix C) are summarized in what follows.

An earth centred rotating (ECR) coordinate system is chosen to reference a GCP on the earth ellipsoid. In this coordinate system the origin is at the earth's centre, the x-axis points to 0° longitude, the z-axis points to the north pole and the y-axis completes a right-handed coordinate system.

Let $\underline{\Psi}$ be the state vector, P be the covariance matrix of the state vector, E be the error covariance matrix projected on the ECR coordinates in the GCP measurement, H be the measurement matrix to be defined and I be the identity matrix. The recursive equations are then given by:

- (a) Compute optimal weight

$$b = P(i) H^T [H P(i) H^T + E]^{-1} \quad (2.10)$$

- (b) Make optimal estimate

$$\underline{\Psi}(i) = \underline{\Psi}(i-1) + b [\underline{R}_g(i) - \underline{R}_g(\underline{\Psi}(i-1))] \quad (2.11)$$

- (c) Update covariance matrix

$$P(i) = (I - b H) P(i) \quad (2.12)$$

Here, $\underline{R}_g(i)$ are the GCP's ECR coordinates (i.e., from a mapsheet), and $\underline{R}_g(\underline{\Psi})$ are the estimated ECR coordinates for the same GCP using $\underline{\Psi}$ in the transformation. The components of E represent GCP designation accuracies in the three ECR coordinate axes. The matrix E is taken to be diagonal since the designation accuracies in the axes are uncorrelated. Hence:

$$E(i,j) = \xi^2 \quad (2.13)$$

Therefore the value of ξ reflects the confidence level in the GCP

measurement, and this is determined by the operator's pointing accuracy on the displayed image and the GCP accuracy on the mapsheet. The components of H are the partial derivatives of R_{-g} with respect to the state vector parameters. The recursive estimation procedure is shown in Figure 2.1.

Landsat-1, -2 and -3 imagery is routinely rectified using this technique [ORTH77, ORTH78a, and ORTH78b] and subpixel accuracy is achievable using 3 to 12 GCPs in a Landsat scene. The number of GCPs required decreases with increased reliability in GCP map coordinates and GCP identification. Here the parameters to be estimated are satellite attitude angles (pitch, roll and yaw) and satellite altitude.

2.3 Comparison of Geometric Transformation / Navigation Methods

In this section comparison is made between the polynomial and the time series methods. In general, one concludes that the latter technique is superior to the former. Comparison along specific dimensions is presented in the subsections which follow.

2.3.1 Distribution of GCPs

Transformation by the polynomial method has the advantage that no prior information on the geometric distortions is included. Hence, no trigonometric or other complex algebraic transformation equations are involved. However, while polynomials may be made to fit well at the GCPs from which the coefficients are determined, they can deviate significantly in areas where no GCPs are present. Therefore, a uniform

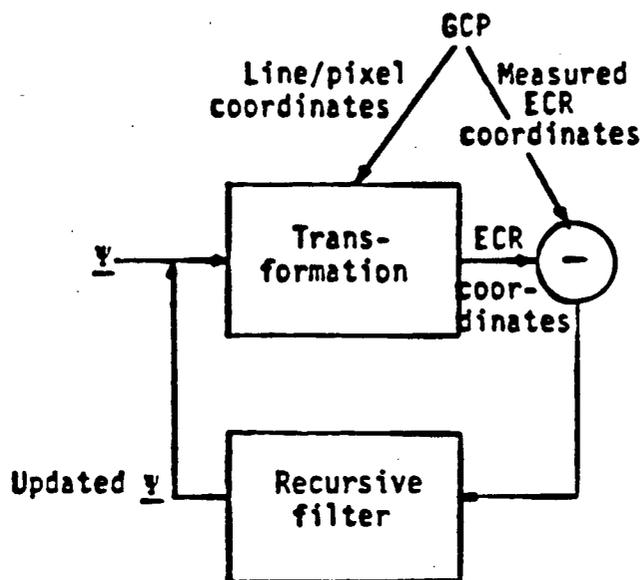


FIGURE 2.1 RECURSIVE ESTIMATION

spatial distribution of the GCPs is vital to the overall accuracy of the polynomial transformation method.

In the case of the attitude time series model, the GCP distribution is constrained only by the requirement that yaw is best estimated by GCPs at the frame edge, since yaw corresponds to rotation about the frame centre line. Similarly, the effect of roll is most noticeable at the edges of a scan line or frame. Finally the maximum effect of pitch is to be found at the top and bottom of a frame.

An equivalent polynomial order for the time series method is reflected in the complexity of Equations 2.8a and b. While this is undoubtedly a high order transformation, advantage is taken in the formulation of the interdependence of coefficients to reduce the total number of degrees of freedom. The fact that this interdependence is taken into account explicitly in the time series method guarantees mapping accuracy in areas with no GCPs, insofar as no discontinuous changes in the time series parameters can occur.

2.3.2 Transformation Accuracy

Image rectification accuracy is often required to subpixel precision, RMS, across the scene. Simon and Caron [SIM075] point out that the polynomial approach cannot be expected to reach this precision, with few (less than 20) GCPs. To achieve subpixel rectification accuracy for Landsat-1, -2 and -3 imagery, each attitude component should be known to the nearest tenth of a milliradian. To this end, Caron and Simon

developed a linear recursive estimator for the time series model. The transformation accuracy obtained is at the subpixel level with about 10 GCPs, under favourable circumstances (i.e., easy identification of GCP and reliable GCP map coordinates).

Knoecny's [KONE76] survey of the polynomial approach shows that to achieve the same degree of accuracy, a 12 degree polynomial and 64 GCPs would have to be used. Konecny's survey also shows that if an affine transformation is used, 15 GCPs would give an RMS error of 160 metres in Landsat imagery. By increasing the number of GCPs to 33, the RMS error drops to 150 metres.

2.3.3 Unique Advantages of Time Series Approach

The time series approach allows certain parameters to be determined explicitly (rather than being embedded in polynomial coefficients). This is advantageous when those parameters are required for other considerations.

In the case of scanner imagery, roll also affects relief displacement (see Section 4.4.4). Hence, for this correction, it is important to have the roll information in addition to terrain data. The polynomial approach does not give the roll angle whereas the time series modelling approach gives each attitude angle as a function of time. Therefore, for each pixel position in an image, the roll angle can be determined.

2.3.4 Unique Advantages of a Recursive Estimator

The recursive estimator approach has operational advantages derived from the sequential measurement of each GCP followed by an estimator update. This sequential structure allows an estimate of the accuracy of a GCP measurement before it is used to update the current attitude estimates. Erroneous GCP measurements can be rejected. Moreover, as errors diminish at each stage, it becomes increasingly easy to locate new GCPs in the image. No such information can be generated by the polynomial or time series model with an iterative approach. Also in the recursive estimator approach, the GCP deviations can be listed following the fitting procedures. Excessive deviations are noted by the operator, who then must edit the input data to the estimator to delete undesirable GCPs. The operator would also have to remeasure GCPs, if required.

2.3.5 Summary

A time series model with a recursive estimator is preferred for the following reasons:

- (a) It is superior to the polynomial approach in terms of overall geometric accuracy.
- (b) In the navigation process, fewer GCPs are required. This is of great operational importance since high quality GCPs are rare and difficult to obtain.
- (c) A recursive estimator processes one GCP at a time. A badly measured GCP is immediately noticeable and can be rejected.
- (d) In the case of scanner imagery, the roll angle across the

scene is "reconstructed". This angle is of importance in the relief displacement correction phase.

This thesis extends this technique to TIROS-N AVHRR (Section 4.3.2) which suffers from a much higher panoramic¹ distortion, covers a larger area for a scene and has a considerably lower resolution than Landsat imagery.

2.4 Remapping and Interpolation

2.4.1 Background

Remapping is the determination of a transformation, given by f_b and g_b in Equation 2.2, from the output grid (x,y) to the input grid (u,v) . This transformation, when combined with interpolation, must give a fast computation of the intensity at (x,y) and store it away without excessive disk traffic (both read and write). Remapping and interpolation are highly related. Therefore, when designing a remapping algorithm, one must take into account the eventual interpolation method.

Existing methods are processing by blocks [ZOBR82] or by strips [RAMA77]. In processing by blocks, the output grid is segmented into blocks and one block at a time is processed in core. The input area containing the output block is read into core while processing the block. In processing by strips all the input lines required to process an

¹ Since pixel sampling in a scan line proceeds in equal time intervals, the ground segment imaged is proportional to the tangent of the scan angle. This effect, combined with the earth's curvature, produces a panoramic distortion in the along scan direction.

output line are stored and processed in core. The core size requirement is then a function of output image size and of the rotation angle between input and output.

This section discusses a processing method in which remapping and interpolation are performed in one-dimensional passes over the image. It offers the following advantages:

- (a) For each pass, either all the column numbers or all the row numbers remain the same between input and output. This means that one row/column of input data corresponds to the same row/column of output data.
- (b) One-dimensional interpolation is implemented in each pass.
- (c) The algorithm can be performed in-place, hence reducing storage requirements.
- (d) Relief displacement correction can be incorporated into the first pass which operates in the along scan direction.

2.4.2 Bilinear Remapping

Fast transformation from the output image coordinates (x,y) to input image coordinates (u,v) can be achieved by segmenting the output image into blocks. The transformation for each block is approximated by two piecewise bilinear functions of the output coordinates. The coefficients for each block are determined by first mapping the four corner points from (x,y) to (u,v) . This, called inverse transformation, cannot be expressed in closed form. Derivation of this inverse transformation starts with the forward transformation from (u,v) to (x,y) which can be expressed in

closed form and is a function of:

- sensor geometry,
- satellite orbit ephemeris,
- satellite attitude (not applicable in SAR imagery),
- method of processing, and
- earth spheroid and earth rotation.

Once the transformation parameters are determined (from navigation), it is possible to transform from image line and pixel coordinates to ECR coordinates and eventually to any desired output map projection, all in closed form.

Piecewise bilinear functions are the simplest functions that guarantee continuity at the block boundaries. The block size is set by the required geometric accuracy. The bilinear functions are developed in Appendix D.

Alternatively, an affine transformation can be used instead of a bilinear transformation. But an affine mapping is in general discontinuous from one block to the next. This can introduce image artifacts and consequently is not discussed further here.

While remapping by the bilinear transformation between (x,y) and (u,v) is straightforward, a direct usage of the transformation between these coordinates will call for two-dimensional interpolation. The remainder of Section 2.4 shows that a bilinear transformation can be

established between the input and output of each one-dimensional processing pass, and that the bilinear transformation in each pass can be reduced to an equivalent one-dimensional linear transformation.

2.4.3 Two-Pass Processing

Two-dimensional interpolation is slow and expensive to implement. A further complication arises when the interpolation kernel increases in size in both directions. However, it may be possible to modify the process into two one-dimensional interpolation processes¹ which are easier to implement and more efficient with respect to interpolation time requirement. Also, it has the added advantage that relief displacement and other high frequency distortions can be corrected in the first pass.

Let p, q be the coordinates of an intermediate image as a result of the first remapping and interpolation pass. It is possible to split the bilinear transformation into two parts [TRW77] as follows:

First pass:

$$u = f_1(p, q) \quad (2.14a)$$

$$v = q \quad (2.14b)$$

Second pass:

$$p = x \quad (2.15a)$$

$$q = g_1(x, y) \quad (2.15b)$$

¹This is possible if the interpolation kernel $H(x, y)$ is separable. That is, $H(x, y)$ can be decomposed into two functions $H_1(x)$ and $H_2(y)$ where $H(x, y) = H_1(x) H_2(y)$. For example, if $H(x, y) = \text{sinc}(x) \text{sinc}(y)$, then $H(x, y)$ is separable with $H_1(x) = \text{sinc}(x)$ and $H_2(y) = \text{sinc}(y)$.

The intermediate image and the output image are partitioned into blocks, and f_1 and g_1 are bilinear polynomials. Since in f_1 the value of q is fixed and in g_1 the value of x is fixed, f_1 and g_1 degenerate to linear functions. Geometrically it means that all pixels are moved horizontally to their final x -position first. Also in this pass, all along scan high frequency errors, such as relief displacement, sensor offset, line length variation, earth rotation and pixel spacing nonlinearity can be corrected. Furthermore, it has the significant advantage that a two-dimensional interpolation kernel reduces effectively to a one-dimensional kernel since the kernel weight in the v -direction is zero at all values of v when $v = q$. In the second pass which is performed along columns of data, the kernel again reduces to one-dimension.

Lines of an image are stored in consecutive records in a typical system. Therefore, to perform the second pass remapping and interpolation, the intermediate image (p,q) should be rotated through 90° first. Fast matrix transpose methods for large files have been described by Eklundh [EKL72] and Goldbogan [GOLD81]. Note that matrix rotation through 90° can be achieved by matrix transposition and reading the resulting rows of data backwards.

Figure 2.2 illustrates the remapping and interpolation directions in each pass.

2.4.4 Three-Pass Processing

The output image (x,y) can assume any orientation which is user

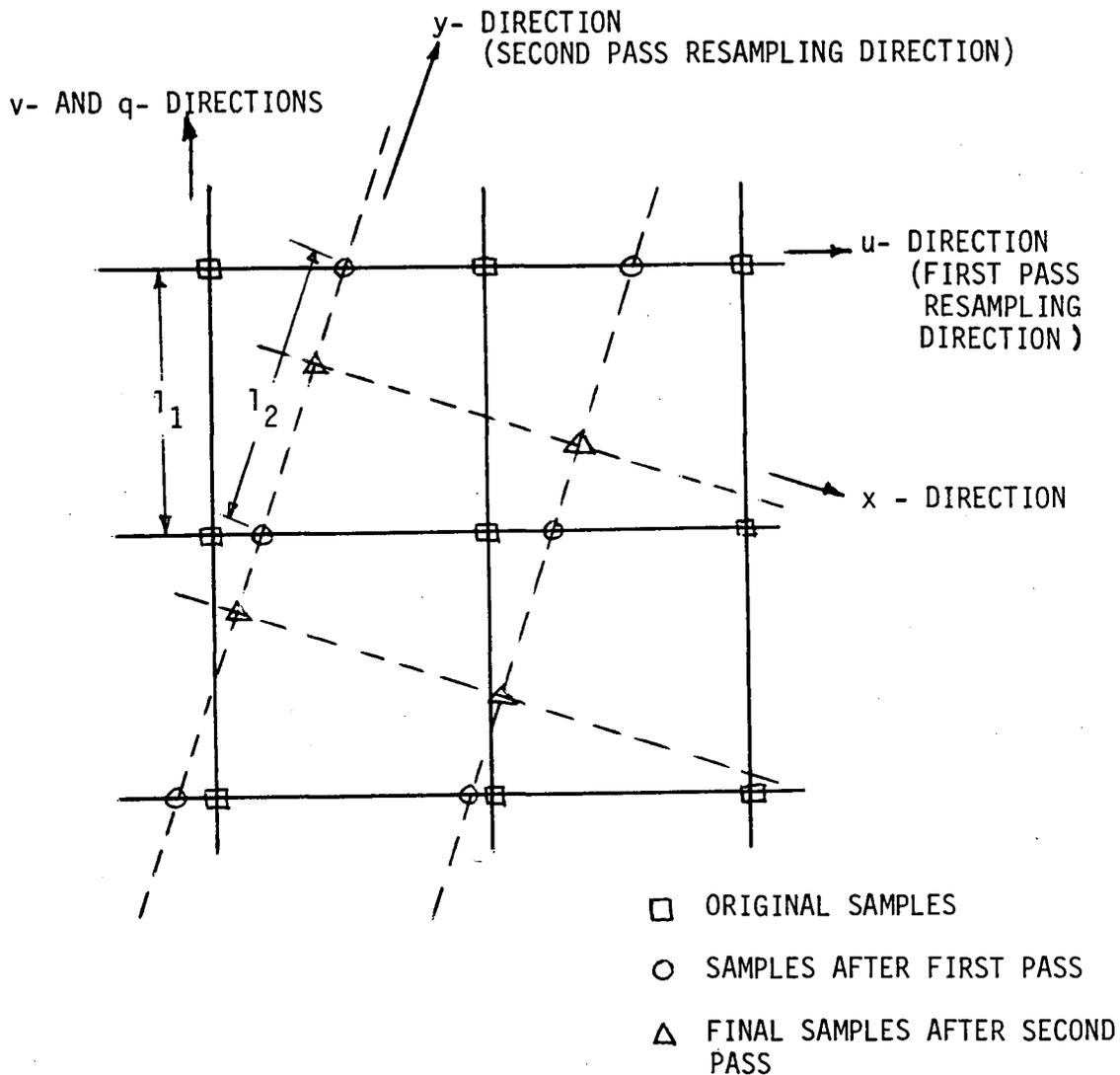


FIGURE 2.2 TWO ONE-DIMENSIONAL REMAPPING AND INTERPOLATION PROCESSES

specified. If there is a rotation with respect to the raw image (u,v) , aliasing is introduced in the two-pass processing method resulting in incorrect radiometric values. Aliasing is explained as follows. In the two-dimensional method, the intensity at a Δ -marked point is obtained by interpolating on the neighbouring \square -marked points. However, in the two-pass method, the intensity at the same Δ -marked point is obtained by interpolating on the 0 -marked points. Let l_1 and l_2 be the across scan spacing between the \square -marked points and 0 -marked points respectively as shown in the figure. Since $l_2 > l_1$, aliasing occurs as a result of interpolation.

Friedmann has solved the aliasing problem in the case of image rotation by modifying the two-pass one-dimensional processing method into a three-pass one-dimensional processing method [FRIE81]. An oversampling procedure is introduced in the first pass (along scan) to push apart the spectrum replica of the original image in the frequency domain. The oversampling s is given by:

$$s > 1 / (1 + \tan \theta) \quad (2.16)$$

where θ is the angle of rotation and $0 < s < 1$. The number of pixels per line in the first intermediate image is increased by a factor $1/s$.

The rotation matrix can be expressed as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.17)$$

where (x',y') is the unrotated image coordinate system (the y' -axis aligns

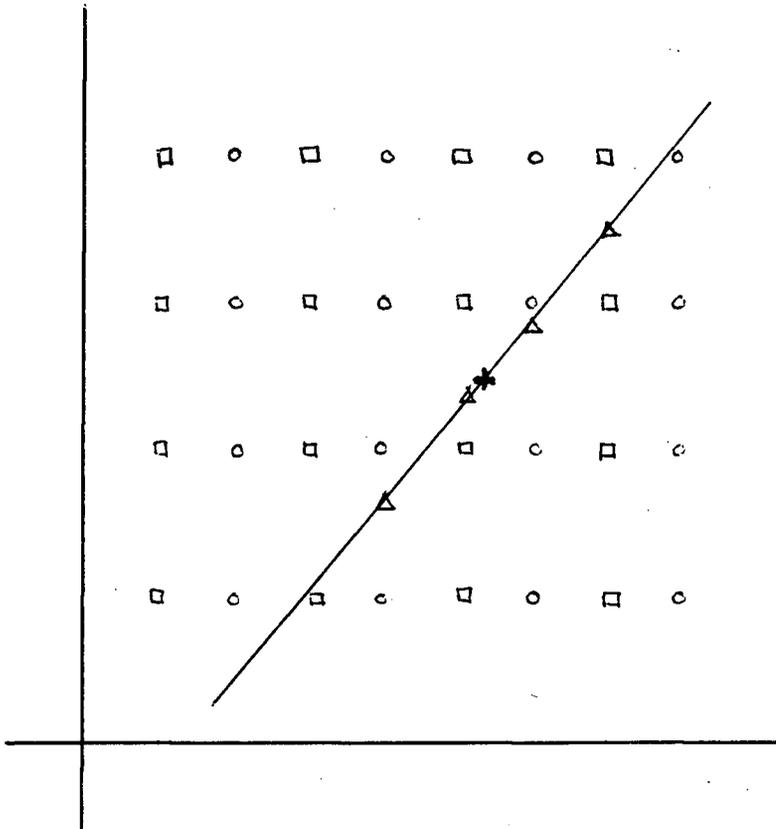
with satellite velocity) and θ is measured in the counter clockwise sense.

In Friedmann's three-pass method, image rotation can be performed properly with an oversampling operation in the first pass, and the whole process can be written as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \underbrace{\begin{bmatrix} s & 0 \\ 0 & 1 \end{bmatrix}}_{M_1} \underbrace{\begin{bmatrix} 1 & 0 \\ -s \tan \theta & \sec \theta \end{bmatrix}}_{M_2} \underbrace{\begin{bmatrix} \cos \theta / s & \sin \theta / s \\ 0 & 1 \end{bmatrix}}_{M_3} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.18)$$

where M_1 is a row direction (along scan) oversampling operation with scaling factor s along a row of data, M_2 is a column direction (across scan) operation and M_3 is another row direction operation. Operations M_1 , M_2 and M_3 are performed in order, resulting in an image after each operation. Figure 2.3 shows the three-pass rotation method.

This thesis extends Friedmann's idea to the remapping and interpolation of satellite imagery in which, in addition to rotation, along scan and across scan distortions have to be corrected. The whole process is conceptually depicted in Figure 2.4. At first, assume that there is no rotation and the two-pass processing given by Equations 2.14 and 2.15 (with x replaced by x' and y by y') will produce an image free of aliasing. Then the three-pass rotation is performed. The two steps give a five-pass process as shown in Figure 2.4a. Let the first pass operation be denoted by H_1 and the second pass operation by H_2 . The oversampling operation M_1 can be coupled with H_1 resulting in modifications to H_2 , M_2 and M_3 which are now H_2' , M_2' and M_3' as shown in



SAMPLES AFTER THE

FIRST PASS ◯ AND ◻
 SECOND PASS △
 THIRD PASS +

NOTE: ◯ ARE THE OVERSAMPLED PIXELS

FIGURE 2.3 . THREE ONE-DIMENSIONAL REMAPPING AND INTERPOLATION PROCESSES

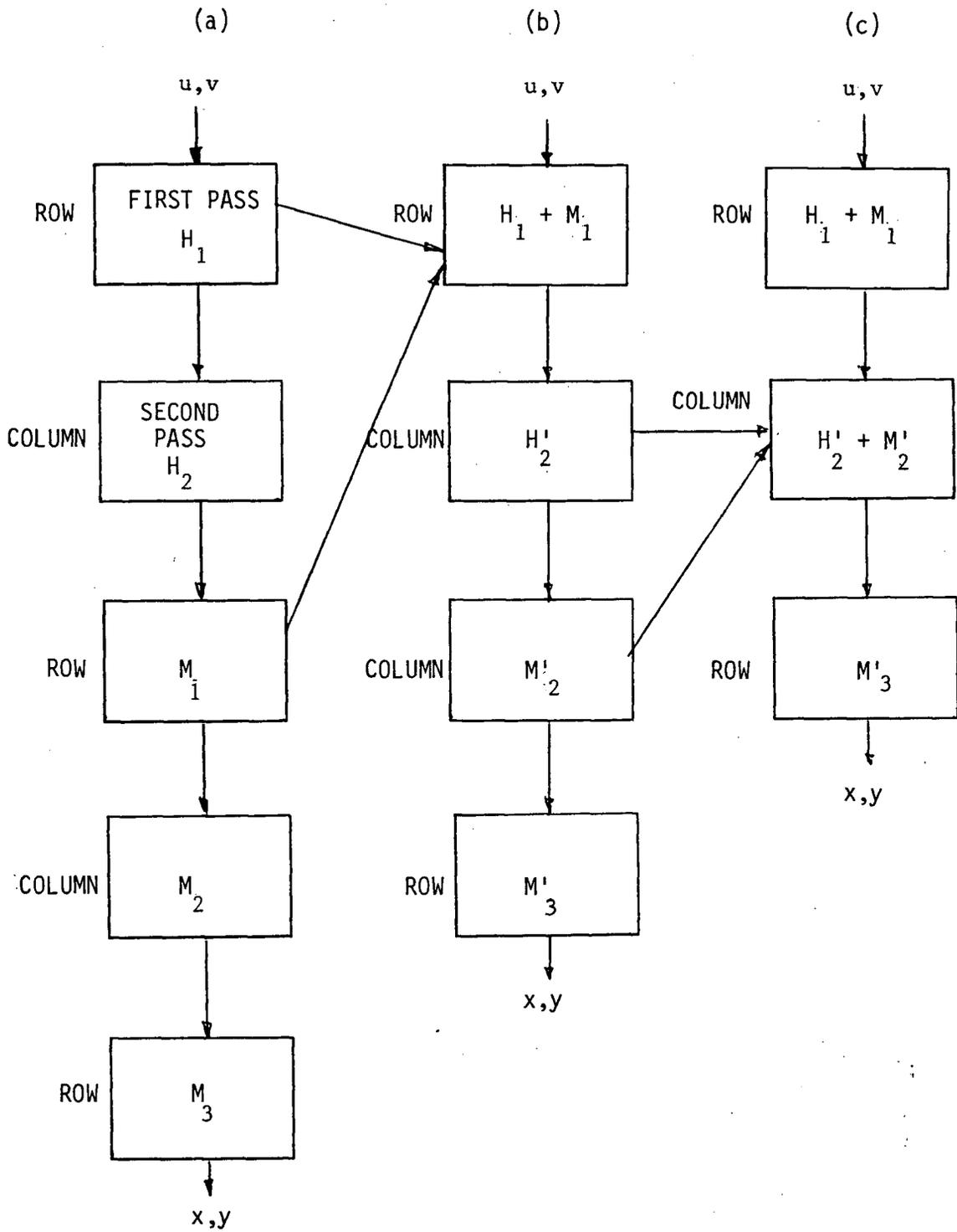


FIGURE 2.4 COUPLING OF PASSES

Figure 2.4b. Since H_2' and M_2' are neighbouring column operations, they can be coupled to form one operation. The final result is a three-pass process as shown in Figure 2.4c.

2.5 Interpolation Kernel

Interpolation is the computation of the intensity of each output pixel with an interpolation kernel. This section examines different kernels and proposes one which is most suited to the three-pass process presented in Section 2.4.

2.5.1 Outline of Interpolation Methods

2.5.1.1 Nearest Neighbour and Linear Interpolations

Nearest neighbour interpolation is the simplest form of all interpolation methods. The interpolation kernel is given by:

$$f_0(x) = \begin{cases} 1 & , \quad -1/2 < x \leq 1/2 \\ 0 & , \quad \text{elsewhere} \end{cases} \quad (2.19)$$

Its frequency response is a sinc function given by:

$$F_0(f) = \text{sinc}(\pi f) \quad (2.20)$$

where $\text{sinc}(x)$ is defined as $\sin(x)/x$.

Another simple form is linear interpolation given by:

$$f_1(x) = \begin{cases} 1 - x & , \quad x \leq 1 \\ 0 & , \quad \text{elsewhere} \end{cases} \quad (2.21)$$

Its frequency response is given by:

$$F_1(f) = [F_0^2(f)] = \text{sinc}^2(\pi f) \quad (2.22)$$

The interpolation kernels and their frequency responses are plotted in Figure 2.5. These two interpolation kernels perform poorly in terms of radiometric accuracy, as shown by their frequency response deviations from the ideal low pass filter, when applied to image interpolation. But they are adequate for DTM interpolation as will be discussed in Chapter 3.

2.5.1.2 Cubic Spline

A spline is a smooth curve passing through a set of discrete sample points which are not necessarily uniformly spaced. A common spline is the cubic spline which is defined as follows. Let the sample points be t_i where i is the sample point index. Then in the interval t_i to t_{i+1} , a cubic polynomial is fitted and the fitting function, its first and second derivatives are continuous at both end points. The interpolation in this interval is given by:

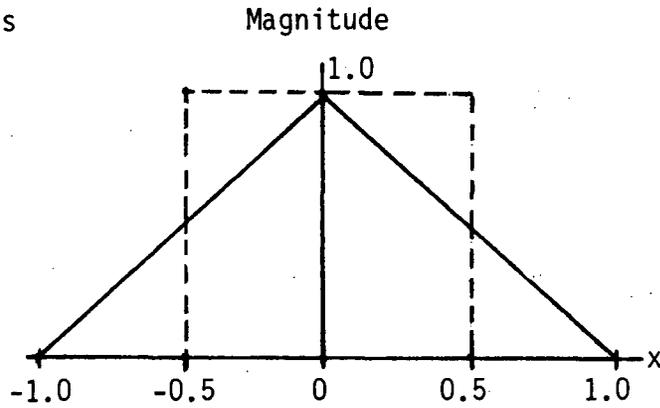
$$f(t) = C_{0i} + C_{1i}(t-t_i) + C_{2i}\frac{(t-t_i)^2}{2} + C_{3i}\frac{(t-t_i)^3}{6} \quad (2.23)$$

From the continuity conditions, the four coefficients C_{0i} to C_{3i} can be found [DEB079] and the result is presented below without proof here.

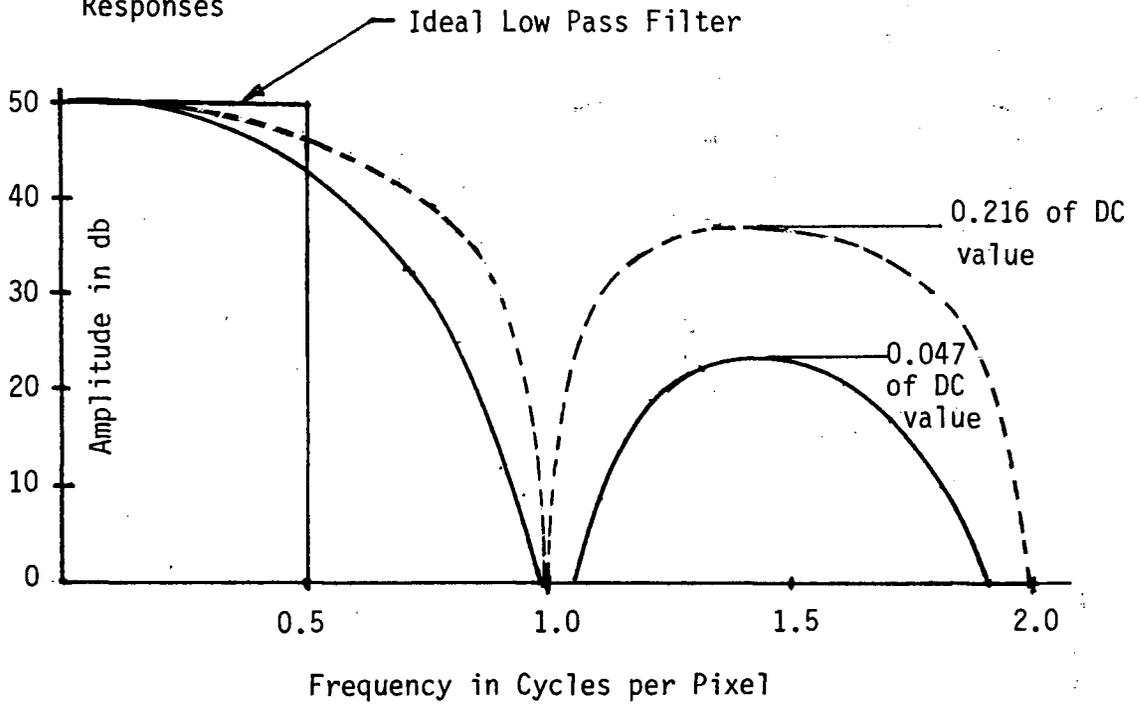
Denote the data points by F_i , the slope of $f(t)$ at t_i by s_i , the spacing between t_i and t_{i+1} by h_i ; i.e.,

$$h_i = t_{i+1} - t_i \quad (2.24)$$

(a) Interpolation
Kernels



(b) Frequency
Responses



----- Nearest Neighbour Interpolation
 _____ Linear Interpolation

FIGURE 2.5 NEAREST NEIGHBOUR AND LINEAR INTERPOLATION
 - KERNELS AND FREQUENCY RESPONSES

values of s_i . This can be done by taking s_i as the slope at t_i of the quadratic polynomial which agrees with F at t_{i-1} , t_i and t_{i+1} . Then:

$$s_i = \frac{h_i F[t_{i-1}, t_i] + h_i F[t_i, t_{i+1}]}{h_i + h_{i+1}} \quad (2.30)$$

The error in interpolation is given by:

$$|e| < (5/384) h_i^4 |f^{(4)}| \quad (2.31)$$

where $f^{(4)}$ is the fourth derivative of $f(t)$.

2.5.1.3 B-Spline

Another class of splines is the set of B-splines. A B-spline is a polynomial fit in each interval t_i to t_{i+1} , and the 0-th to $(n-1)$ th derivatives are continuous where n is the polynomial order. Interpolation is performed with the aid of a set of basis function B :

$$f(t) = \sum_{i=1}^n \alpha_i B_{i,k}(t) \quad (2.32)$$

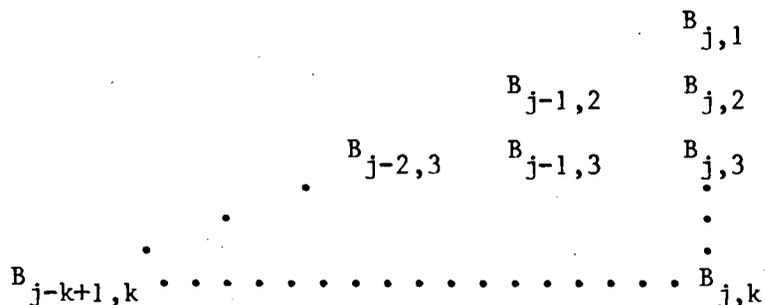
where n is the total number of points, $k-1$ is the polynomial order, α_i is a set of coefficients to be determined, and $B_{i,k}(t)$ belongs to the set B . The set B is defined recursively:

$$B_{i,1}(t) = \begin{cases} 1 & , \quad t_i < x < t_{i+1} \\ 0 & , \quad \text{elsewhere} \end{cases} \quad (2.33a)$$

$$B_{i,k}(t) = \frac{t - t_i}{t_{i+k-1} - t_i} B_{i,k-1}(t) + \frac{t_{i+k} - t}{t_{i+k} - t_{i-1}} B_{i+1,k-1}(t) \quad (2.33b)$$

with $k > 1$ and $0/0 = 1$. The evaluation of the B-spline by the

recurrence relation is very stable since only additions of non-negative quantities are involved. For a given t say $t_j > t > t_{j+1}$, then only the following B-splines are non-zero:



It remains to determine α_i . This is done by solving a set of simultaneous linear equations:

$$\sum_{i=1}^n \alpha_i B_{i,k}(t) = F_i \tag{2.34}$$

In matrix form this is:

$$\begin{bmatrix} \theta \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} F_1 \\ \vdots \\ F_n \end{bmatrix} \tag{2.35}$$

The matrix θ is a function of $B_{i,k}$ and is diagonally dominant.

2.5.1.4 The Sampling Theorem

Let F_i be an infinite string of pulses sampled on a function $f(t)$ at equal time intervals iT where T is the time spacing between samples. The original function $f(t)$ and its spectrum can be recovered completely if the following conditions are satisfied:

- (a) The signal $f(t)$ is band limited with bandwidth W ; i.e., its highest significant frequency is less than W cycles/second.
- (b) The signal is sampled at more than the Nyquist rate of $2 W$ samples per second.

This is known as the sampling theorem [LATH65, JERR77 and SAKR68].

For a given signal, the highest frequency¹ that can be represented is 0.5 cycle/pixel and the interpolation is given by:

$$f(t) = \sum_{i=-\infty}^{\infty} F_i \operatorname{sinc} [\pi(t-i)] \quad (2.36)$$

2.5.1.5 Cubic Convolution

The interpolation given by Equation 2.36 requires an infinite number of data points. Cubic convolution uses a truncated sinc function which extends over four points. The kernel is modified so that the first derivative, as a result of interpolation, is continuous. The kernel is [BERN76]:

$$f(x) = \begin{cases} 1 - 2|x| + |x|^3 & , 0 < |x| < 1 \\ 4 - 8|x| + 5|x|^2 - |x|^3 & , 1 < |x| < 2 \\ 0 & , \text{elsewhere} \end{cases} \quad (2.37)$$

Cubic convolution is advocated by Bernstein [BERN76] and Simon [SIM075b] because the kernel size in two-dimensional interpolation is only 4×4 .

¹ For ease in discussion, frequency here is expressed as cycles/pixel. This can easily be converted to cycles/second if the sampling rate is known.

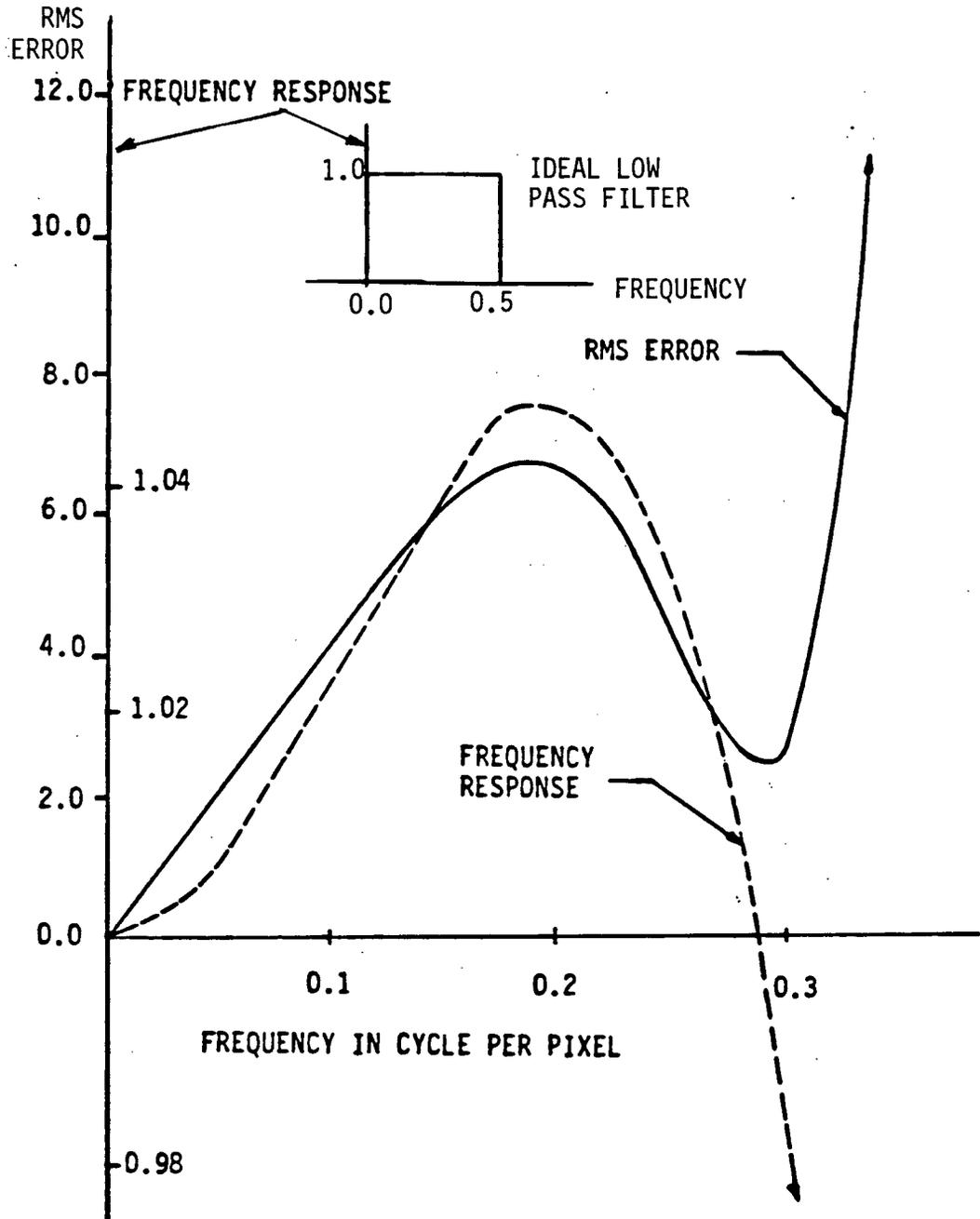
Shlien [SHLI79] has also investigated other splines with different kernel sizes, which are approximations to the ideal sinc function.

2.5.1.5 Comparison of Interpolation Kernels

The spline methods are not pursued further here. The more traditional engineering approach of using the convolution of sample points with the sinc function (Equation 2.36) is used in the remainder of the thesis.

Figure 2.6 shows the frequency response of the cubic convolution kernel. The ideal low pass filter, which the sinc function yields, is inserted in the figure for comparison purposes. Interpolation experiments have been performed with sine waves of different frequencies and the RMS interpolation error is also shown in the figure. Since the frequency response crosses 1 at 0.28 cycle per pixel, a dip occurs in the RMS error around this frequency. Beyond this frequency, the response deviates drastically from the ideal low pass filter, and this results in a rapid growth in the RMS error. With frequency below 0.28 cycle per pixel, the RMS error ranges from 0 to 7 levels out of 256 levels (assuming 8-bit image data).

Note that RMS error and frequency are related as follows. Let $R(f)$ be the frequency response for frequency f . Then the response of a sine wave with magnitude of 1 unit will have a magnitude of $R(f)$ units after interpolation. The RMS error for frequency f is measured between the



Note: The RMS error includes the effect of tabulating the interpolation kernel at every 1/32 data point interval. See Section 2.5.2.1.

FIGURE 2.6 RMS ERROR AND FREQUENCY RESPONSE FOR CUBIC CONVOLUTION

output and input sine waves, which have the same frequency and phase angle but different amplitudes.

The RMS error can be reduced by expanding the filter size. Since the three-pass process only calls for one-dimensional interpolation, the number of arithmetic operations increases only linearly with kernel size. Figure 2.7 shows the frequency response of a 16-point sinc function (see Section 2.5.2.2) modified by a Kaiser window (see Section 2.5.2.3) and the RMS error as a result of interpolation on sine waves of different frequencies. A comparison with Figure 2.6 shows that this 16-point sinc function gives a much lower RMS interpolation error. In particular, the MSS and TM modulation transfer functions have significant frequency responses up to 0.16 cycle per pixel. Below this frequency, the overall RMS interpolation error is about 1 level. Based on these results, this kernel is used in the interpolation phase.

2.5.2 Preprocessing

Preprocessing of the sinc function consists of the following three steps:

2.5.2.1 Tabulation of Sinc Function

For an interpolation of $f(t)$, it is necessary to compute the value of the sinc function at the data points. Such computation can be avoided if the sinc function is replaced by a table and the filter weight can be obtained by a simple table look up technique. This requires the sinc

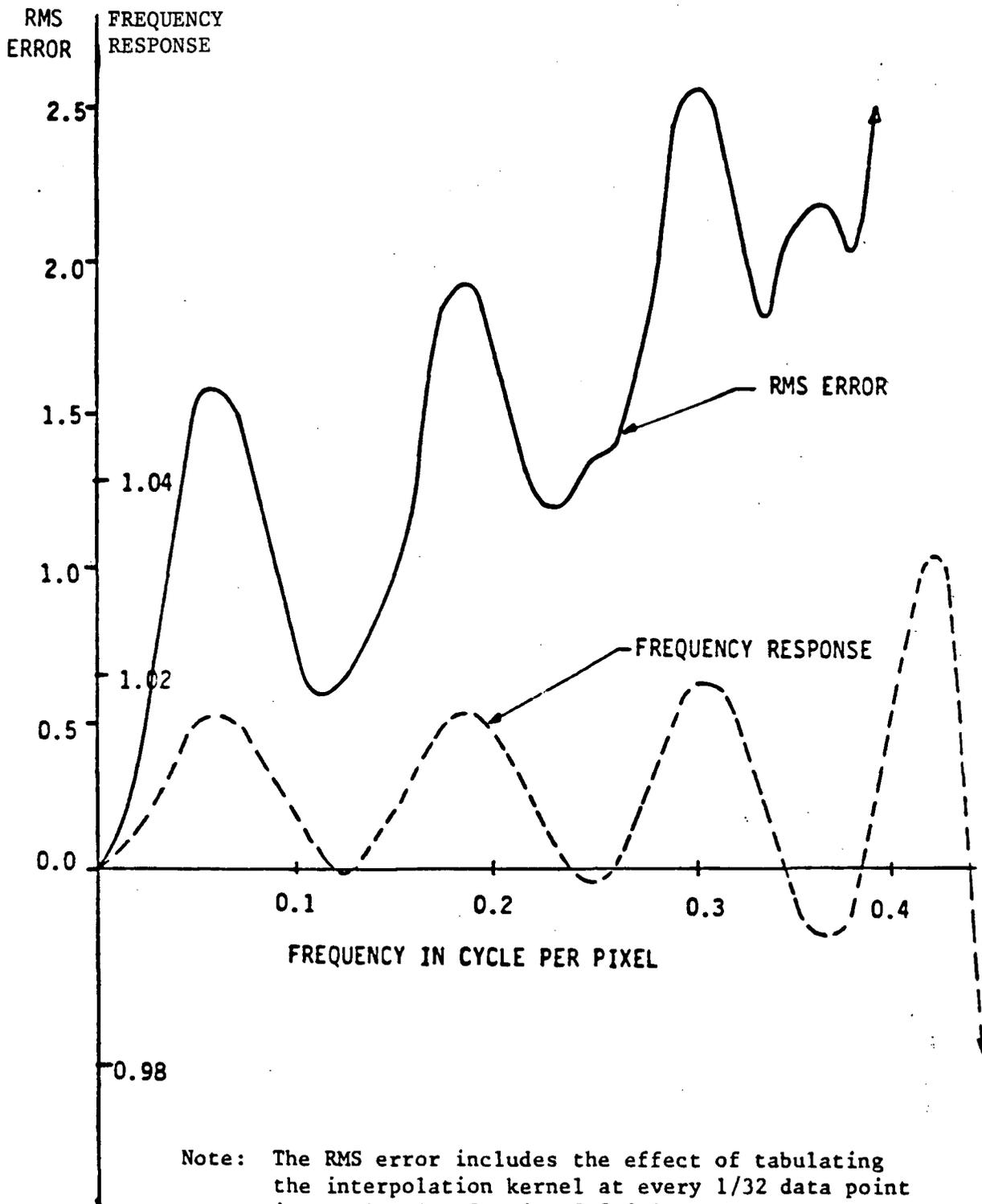


FIGURE 2.7 RMS ERROR AND FREQUENCY RESPONSE FOR SINC FUNCTION INTERPOLATION

function to be divided into strips and a step function is used in each strip to approximate the sinc function as shown in Figure 2.8a. Therefore, preprocessing is required to compute the table. The interpolation accuracy is then a function of the strip size. The worst case is that each F_i lies at the edge of a strip. The result is that the entire function is shifted by $1/2$ strip as shown in Figure 2.8b.

The strip width can be determined as follows. The modulation transfer function (MTF) of a typical mechanical scanner dictates that it takes about three pixels to respond from one intensity extreme to the other (this corresponds to about 0.16 cycle per pixel). For this worst case on 8-bit imagery, the error is:

$$\text{Error} = 0.5 \times \text{strip width} \times 256 / 3 \text{ levels,}$$

where $256/3$ is the slope of the response in levels per pixel. Therefore:

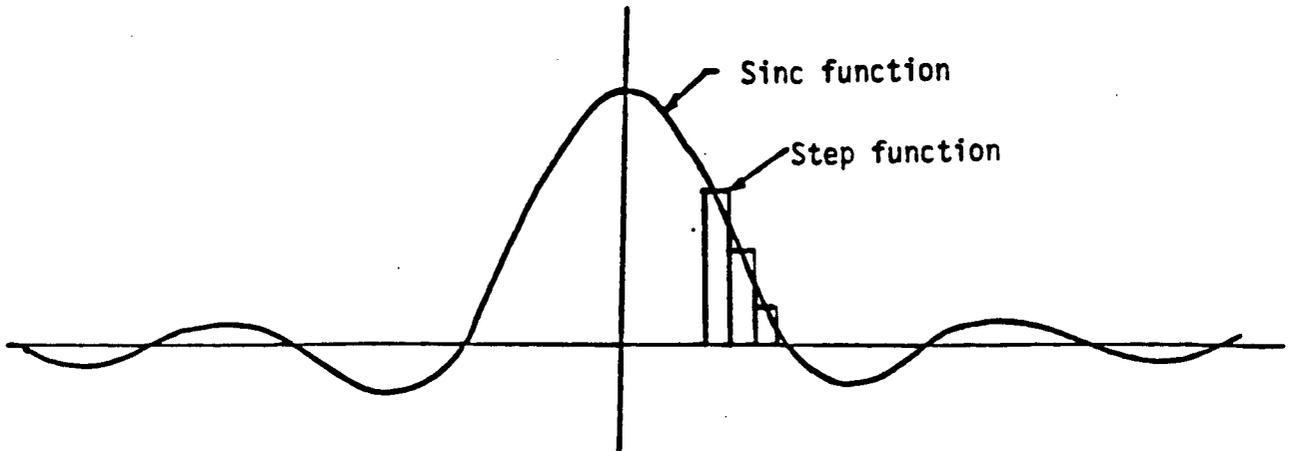
$$\text{Error} = \text{strip width} \times 43 \text{ levels.}$$

Assuming uniform distribution in intensity slope, the RMS error is then:

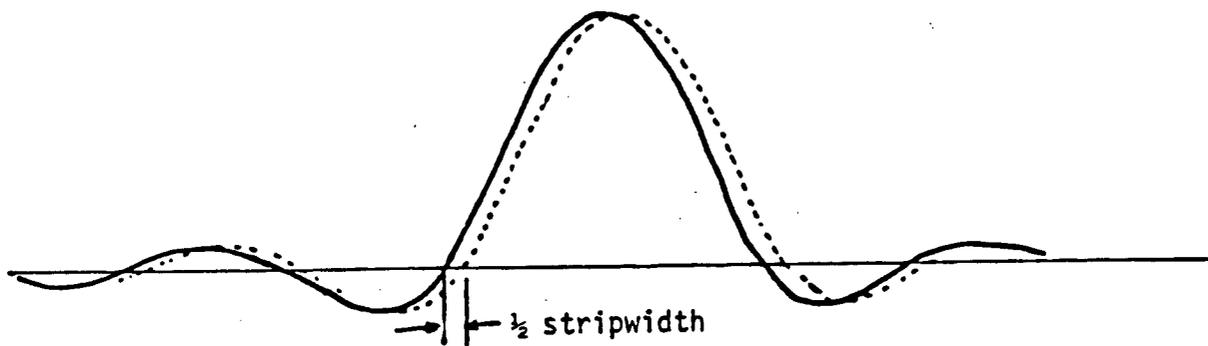
$$\text{RMS error} = \text{strip width} \times 43 / 3 \text{ levels.}$$

If the strip width is $1/32$ pixel, the RMS error is then below one level which is acceptable in many applications.

The MTFs of SPOT's linear array sensors are not known at present, but the strip width can be determined in the same manner once the MTFs become known.



(a) Approximation



(b) Worst error

FIGURE 2.8 APPROXIMATION OF SINC FUNCTION BY STEP FUNCTION

2.5.2.2 Kernel Size Determination

It takes an infinite number of points to synthesize the ideal low pass filter. This is reflected in Equation 2.36 in which the summation is over all data points. However, the maximum magnitude of each sidelobe¹ of $h(t)$ decreases as a function of distance away from the centre. For example, in the 8-th sidelobe the maximum magnitude is about 1/27 of the magnitude at the centre. This suggests that the ideal but infinite sinc function can be approximated by a truncated sinc function.

Any approximation would introduce error in $h(t)$ in the frequency domain where the spectrum of the truncated sinc function is no longer an ideal low pass filter. The most important tradeoff in performing image interpolation is speed versus accuracy. The larger the filter, the greater the interpolation accuracy and the slower the computation speed. The accuracy obtainable with any given kernel size can be computed in a statistical sense [SHLI79].

Let the function $F(t)$ have autocorrelation R :

$$E [F(t) F(t+\tau)] = R(\tau) \quad (2.38)$$

where E denotes expected value. Let $f(t)$ be an estimate of $F(t)$ given by:

$$f(t) = \sum_i F_i h(t-i) \quad (2.39)$$

where $h(t)$ is the sinc function multiplied by a window. Then the RMS

¹ Sidelobes in the time domain must not be confused with sidelobes in the frequency domain. It should be clear which domain is referred to in the text.

error in interpolation is given by:

$$e^2(t) = E\{[F(t) - f(t)]^2\} \quad (2.40)$$

which after expanding gives:

$$e^2(t) = R(0) - 2 \sum_i R(t-i)h(t-i) + \sum_{ij} R(i-j)h(t-i)h(t-j) \quad (2.41)$$

Thus the RMS error can be calculated for any kernel size provided the autocorrelation function R of the data is known.

The autocorrelation function is scene and terrain type dependent. A Seasat scene (Orbit 230) and Landsat scene (I.D.2921-18025) are used to study this function. Both scenes were imaged over the Vancouver area and cover different terrain types such as mountain, sea and city. Figure 2.9 illustrates the autocorrelation function for a few terrain types. Using the function for various terrain types, the RMS interpolation error versus kernel size is derived and is shown in Figure 2.10. From this figure it is seen that for a kernel size of 16, the maximum RMS error is less than 0.5 level out of 256 for both scenes. The kernel size used in the remaining part of the thesis is 16 unless otherwise stated.

2.5.2.3 Windowing

The sinc function interpolation kernel has alternate algebraic signs for successive weights in the main lobe. This, together with truncation in the sinc function, produces a "ringing" effect, i.e., Gibb's phenomenon, for sharp bends in the data values. In the frequency domain, the truncated sinc function no longer corresponds to an ideal low

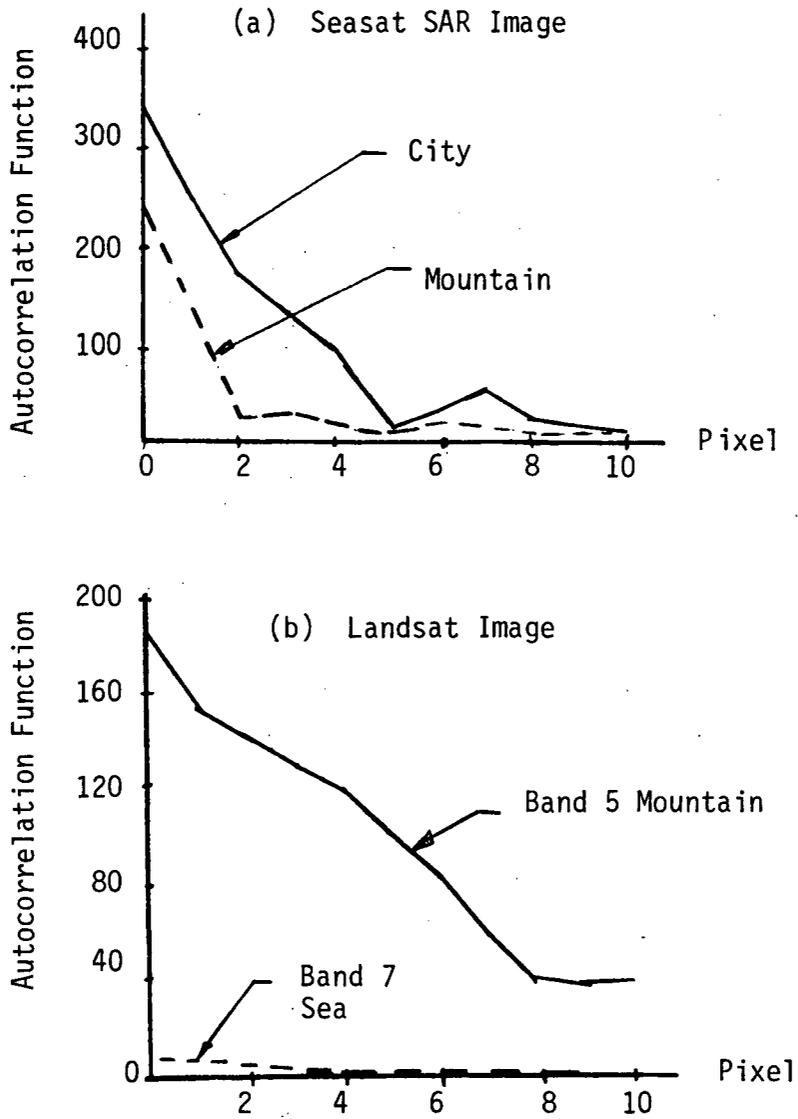


FIGURE 2.9 EXAMPLES OF AUTOCORRELATION FUNCTION

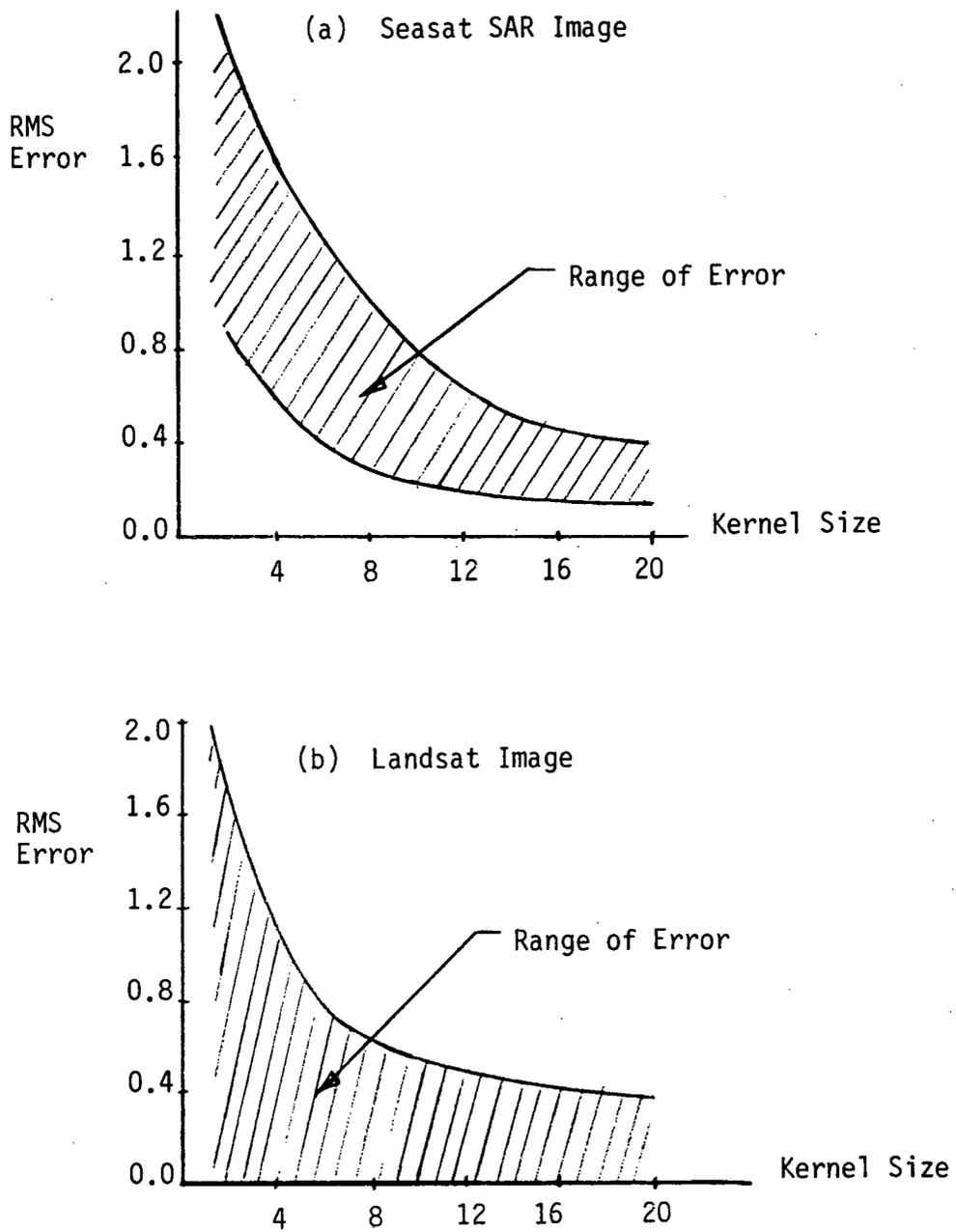


FIGURE 2.10 RMS INTERPOLATION ERROR VERSUS KERNEL SIZE

pass filter, but a filter with sidelobes. A windowing function can be used to minimize the ripples in these sidelobes, at the expense of narrowing the main lobe. This will reduce ringing in the time domain. One example is the Kaiser window given by:

$$w(t) = \frac{J_0 \left(\beta \sqrt{1 - [2t/(N-1)]^2} \right)}{J_0(\beta)} \quad (2.42)$$

where N is the total number of points in the sinc function, $J_0(\beta)$ is the zeroth order Bessel function [KORN61] and β controls the tradeoff between sidelobe peaks and width of the main lobe.

The interpolation kernel is multiplied by the Kaiser window prior to the interpolation process. Figures 2.11¹ and 2.12 are plots of the 4-point truncated sinc function without and with Kaiser windowing. The effect on the sidelobes is immediately obvious by comparing these two figures. Figure 2.13 shows the frequency response of the 16-point sinc function with Kaiser windowing.

Another example is a Dolph-Chebyshev window which makes the sidelobe peaks all equal [RABI75].

When the sinc function is divided into strips, the multiplication by a window imposes very little overhead in the preprocessing.

2.5.3 Nonuniformly Spaced Samples

The above interpolation kernels are only applicable to samples

¹ Figures 2.11 to 2.13 were provided by Mr. D. Friedmann at MDA.

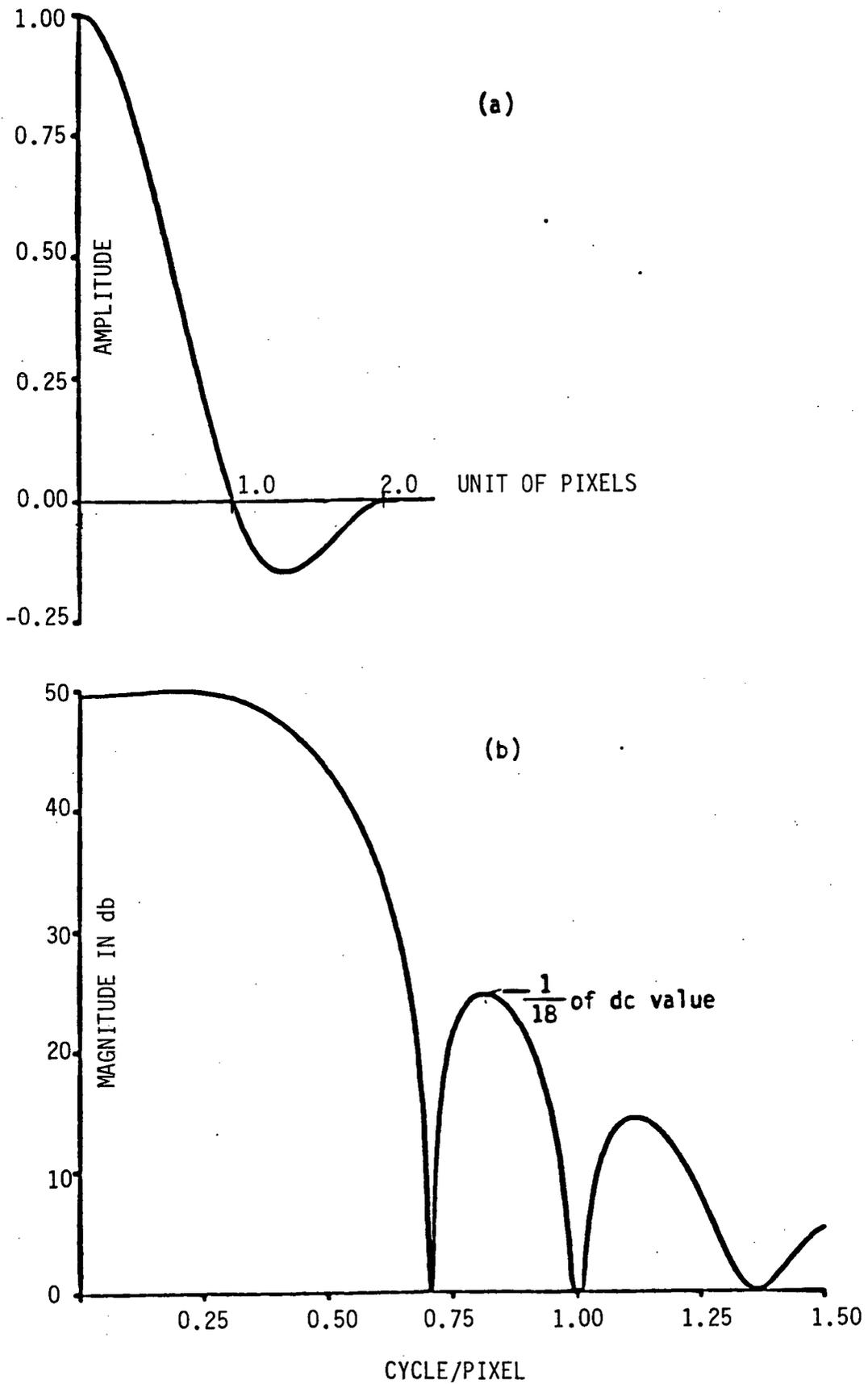


FIGURE 2.11 FOUR-POINT SINC FUNCTION AND ITS FREQUENCY RESPONSE

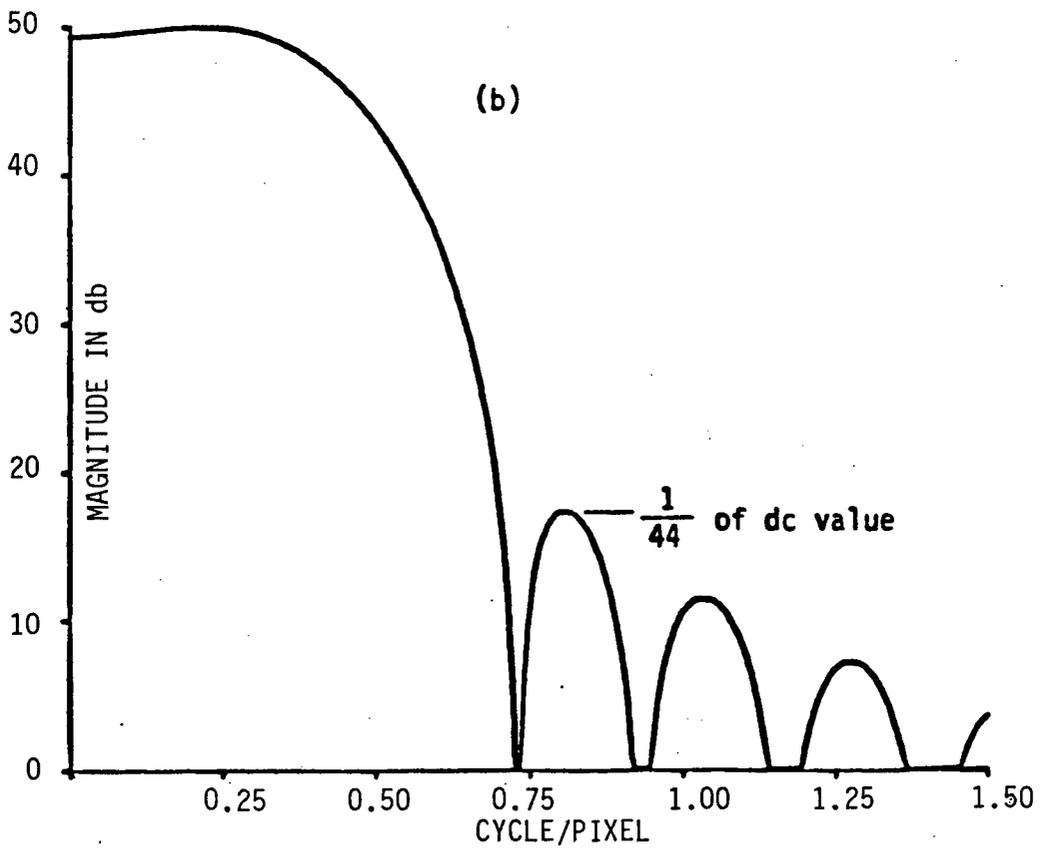
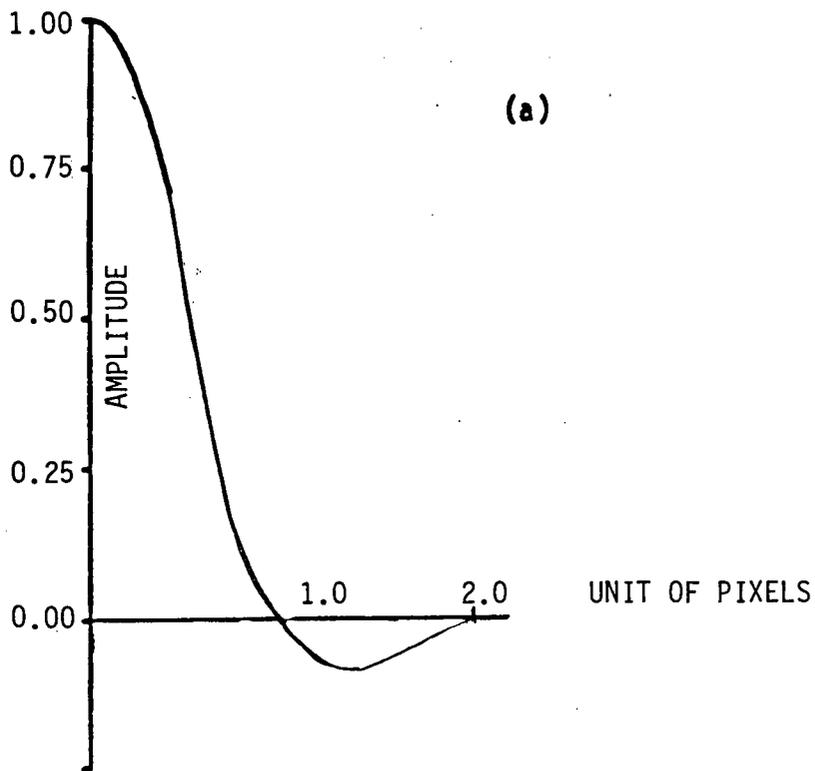


FIGURE 2.12 FOUR-POINT SINC FUNCTION WITH KAISER WINDOW AND ITS FREQUENCY RESPONSE

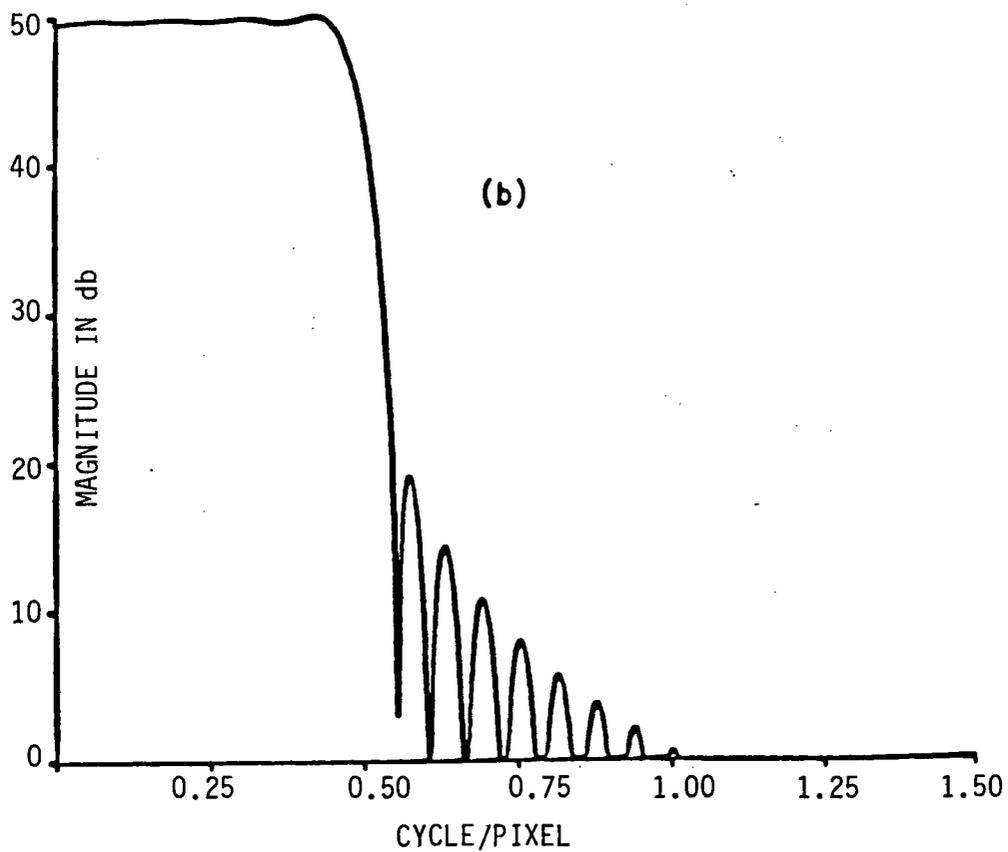
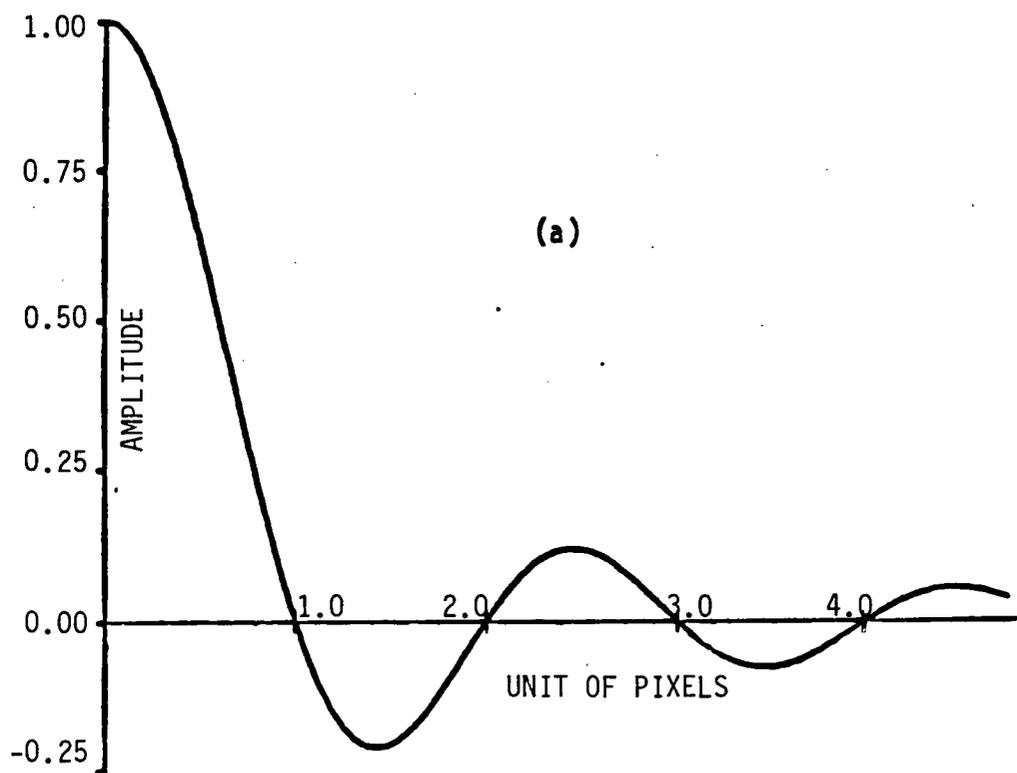


FIGURE 2.13 SIXTEEN-POINT SINC FUNCTION WITH KAISER WINDOW AND ITS FREQUENCY RESPONSE

which are uniformly spaced. However, for Landsat-4 TM imagery, the data samples are no longer equally spaced in the across scan direction due to the presence of jitter, variations in satellite altitude and ground speed, earth's obliqueness and bowtie effect [GSFC82]. All these distortions will create "scan gaps" which result in deviations from nominal sample spacing. The sinc kernel is no longer suitable for the across scan interpolation.

YEN [YEN56] has shown that the signal can still be reconstructed from the discrete samples. In the case of no scan gap, the kernel he proposes degenerates to the sinc function. A modification of Yen's method is applied to the present problem. **Since** this is specific only to Landsat-4 TM imagery in the across scan pass, detailed discussion on this method is deferred to Chapter 3.

2.6 System Data Flow

Figure 2.14 depicts overall system data flow. It is divided into three parts:

- (a) Processing of auxiliary data so that the useful data are stored in disk files.
- (b) Generation of the three-pass remapping and interpolation coefficients.
- (c) Use of remapping coefficients and the DTM.

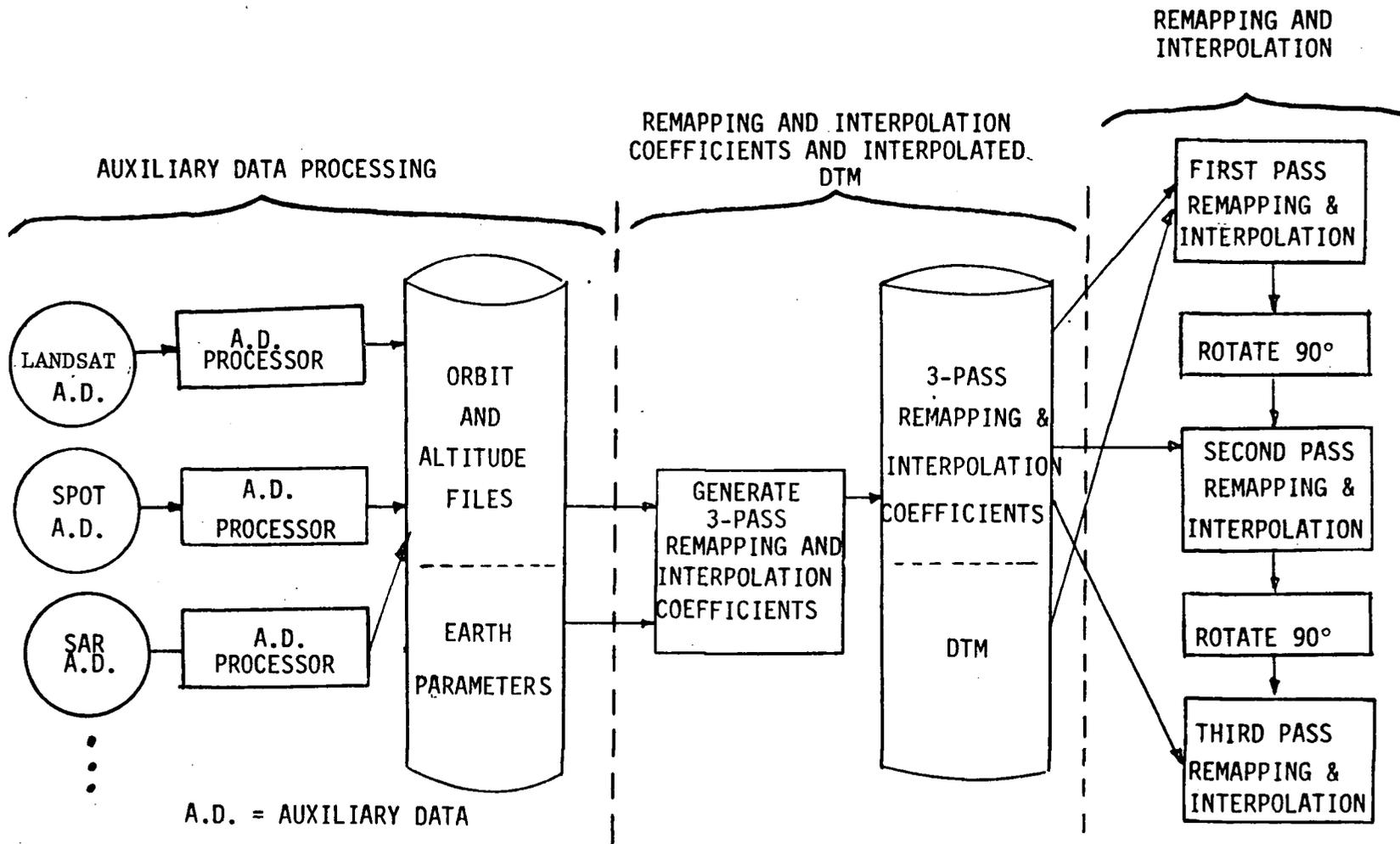


FIGURE 2.14 SYSTEM DATA FLOW

Generation and usage of remapping and interpolation coefficients have been discussed above. This section discusses auxiliary data processing and the relief displacement correction in the first pass of remapping and interpolation.

2.6.1 Auxiliary Data Processing

Auxiliary data contains various kinds of correction information that the user can apply to produce a geometrically and radiometrically corrected image. The contents of auxiliary data can be divided into three areas:

- satellite data for geometric correction,
- sensor data for geometric and radiometric correction, and
- earth and projection dependent data for placing the image on a map grid.

Satellite data contains orbit information and can be in the form of orbital elements (which are parameters to describe the orbit) and/or sampled satellite position and velocity. Orbit information is used to determine the satellite position as a function of time. Satellite data also contains attitude (pitch, roll and yaw) information. Sensor geometric data contains information which, when combined with satellite data, can be used to determine the geolocation of any pixel in the image. Both sets of data are obtained from telemetry design specifications or telex from the appropriate agencies. These data have to be processed so that the orbit and attitude information are stored in

data files which will be used in the generation of remapping coefficients. In this way the system has the flexibility to handle imagery from various satellites since the transformation is table driven and has become satellite independent.

Remapping accuracy depends upon the accuracies of the orbit and attitude information. These data can usually be refined by ground control points (GCPs). The transformation accuracy will have an impact on the eventual interpolation of the DTM. How their accuracies affect the use of the DTM is discussed in Chapter 3.

Earth parameters are used to determine where an instantaneous view vector intersects the earth spheroid. Projection parameters contain information for mapping a point on the earth spheroid to a two-dimensional surface. The map projection parameters depend on the chosen map projection. Earth parameters also include terrain fluctuations. A DTM must be available in order to correct for relief displacement.

2.6.2 Relief Displacement Correction

The usual geometric errors in satellite borne imagery are slow varying and thus the output image can be segmented into blocks as has already been argued. Relief displacements are fast varying and, therefore, it is essential to separate relief displacement correction from the bilinear model. To correct for relief displacement, relief information must be available in the form of a DTM.

A simple, fast and operationally feasible correction for relief displacement can be performed on the raw image data. Here, the distortion is only in the along scan direction, because it is in this direction only that the displacement due to central perspective occurs. The bilinear model for geometric correction can be implemented in three passes over the same image as shown in Section 2.4. The first pass is in the along scan remapping and interpolation process which is illustrated in Figure 2.15. Relief displacement correction is incorporated into this pass. Other fast varying errors such as line length variation, sensor offset and earth rotation are also incorporated into the first pass.

As shown in the figure, the raw image (Box A) is read one line (Box B) at a time into core. It is desired to correct for all the along scan distortions and hence map this line into an output buffer (Box C). The mapping consists of both low frequency and high frequency components. The low frequency component is modelled by the along scan coefficients (Box D). The high frequency components, relief displacement (Box E), line length variation and sensor offset (Box F) depend on pixel location.

The DTM is used to correct for relief displacement in the following manner. To prevent a two-dimensional interpolation on the DTM, the ideal case would be to interpolate it as an off-line process to the image coordinates defined after the first pass. In this way the terrain height of the point being processed can be read easily and the relief displacement can then be computed. However, this is operationally unwise since the interpolated DTM is only specific to one particular scene. Furthermore, in the case of TM imagery where scan gaps occur, it

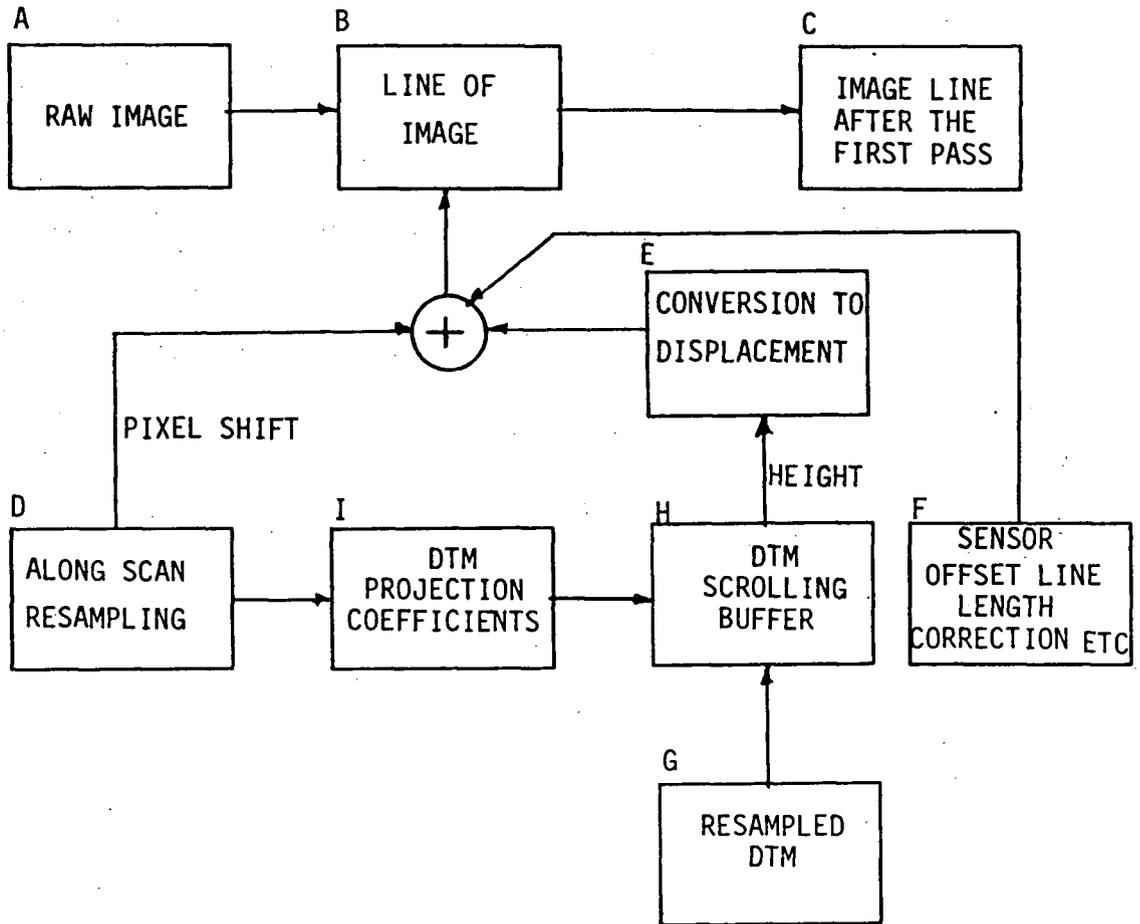


FIGURE 2.15 FIRST REMAPPING AND INTERPOLATION PASS

is difficult to get a DTM to such a projection. A more feasible way is to interpolate the DTM to a map projection close to the input image coordinates (Box G); e.g., assume nominal orbit parameters and zero attitude angles. Then a scrolling buffer (Box H) is used in main memory during the along scan remapping and interpolation process. The buffer should contain enough lines so that the heights of all the pixels of the current image line can be found in the buffer. Indexing to the buffer is done by predetermined coefficients (Box I) which vary from scene to scene. This is operationally feasible since the orientations of all scenes over the same area do not differ significantly. Hence, the interpolated DTM can be used in all these scenes and only the indexing coefficients have to be changed.

Interpolation of the DTM can also be done by the three-pass processing method since the DTM can be treated as an image with the height values representing the pixel intensities. The remapping coefficients can readily be determined since each point in the DTM represents a certain geographic location and each line, pixel coordinate in the "nominal" image can be mapped into a geographic location. Therefore, it is possible to develop a transformation between the DTM and the raw "nominal" image coordinates. The size of the DTM scrolling buffer is then a function of the yaw angle.

2.7 Chapter Summary

A unified approach to the rectification of remotely sensed imagery has been developed. The key features are:

- (a) Geometric transformation parameters, e.g., satellite attitude angles, are refined in the navigation process from a priori estimates using a recursive estimator and GCPs. Only a reasonable number of GCPs per scene are required to achieve a high geometric accuracy.
- (b) Remapping and interpolation is performed with three one-dimensional passes over the image. The one-dimensional processing is the key to yield a high throughput.
- (c) Remapping is performed via a pair of piecewise bilinear polynomials and fast computation can be achieved.
- (d) A reasonable interpolation kernel size (say 16) is used to retain radiometric accuracy. The kernel is tabulated, again for speed in computation.
- (e) Relief displacement and other along scan distortions can be corrected in the first pass over the image.

CHAPTER 3 DIGITAL TERRAIN MODELS

3.1 Introduction

In order to correct remote sensing data for relief displacement, elevation information must be available in digital form. A DTM is an ordered array of numbers that represents the spatial distribution of terrain characteristics. The usual characteristic is the terrain height at the given x,y position (when this is the case, the DTM is also referred to as DEM or Digital Elevation Model). The x,y coordinate system is usually in some map projection or in latitude/longitude.

This chapter describes the types of DTMs, their acquisition methods, spatial accuracies and interpolations. Finally their accuracy requirements and interpolation methods with respect to the rectification process are investigated.

3.2 Acquisition of DTMs

DTM data may be acquired by several methods [DOYL78]. The usual and least expensive method is to manually digitize existing mapsheets. This is done on a commercially available digitizing table. However, this method is very laborious and demands a lot of patience by the operator and is thus susceptible to human error. An improved method is to use automatic contour line following instruments. In this case, a contour map must be prepared by removing all contour labels and filling in the gaps. This is definitely an improvement over the manual method except

that the instruments may have difficulty with lines of different weights such as indexed contours. Furthermore, there is also difficulty in automatically assigning elevations to each contour.

Another automatic DTM acquisition method is to use a scanning device on existing maps. Depending on the type of scanner used, one line or several lines can be scanned at a time. The contour sheet is placed on a drive which either rotates under a fixed light source or has the light source rotating with respect to the drive. Each time the scanner crosses a contour, the x and y coordinates are recorded. Elevation assignment is still a problem in this method.

DTMs acquired by the above approaches may leave large empty spaces between contours. To render these DTMs useful, time consuming interpolation has to be applied to fill in the spaces.

The typical accuracy of data from topographic maps is:

<u>Scale</u>	<u>Class</u> ¹	<u>Spatial Accuracy</u>		<u>Elevation Accuracy</u>	
		<u>R₉₀</u> ²	<u>σ</u>	<u>R₉₀</u>	<u>σ</u>
1:50000	A1	25 m	16 m	10 m	6 m
1:50000	B2	50 m	31 m	20 m	13 m
1:250000	A1	125 m	78 m	50 m	31 m
1:250000	B2	250 m	156 m	100 m	63 m

where σ is the standard deviation and R_{90} is the value which 90% of the map falls within. To obtain better DTM accuracies, a photogrammetric stereo model can be used.

¹ NATO standards, also adopted in Canada.

² $R_{90} \approx 1.6 \sigma$ assuming Gaussian distribution.

An example of a photogrammetric stereo model is the Gestalt Photomapper described by Kelly et al [KELL77]. It is a highly automated photogrammetric system designed to produce DTMs by using electronic image correlation to measure parallax between the left and the right images. The measured parallax can be transformed into height. The spacing between the rows and columns in the DTM is 182 μm on the original photographs. On typical 1:50,000 aerial photographs, the pixel spacing corresponds to 9 metres. From these photographs, the resulting RMS errors in the X, Y and Z axes in the DTM are 0.6, 0.85 and 1 metre respectively. The errors are even smaller in the newer version of the Gestalt Photomapper (GPM II-1).

Another possible source of DTMs is data obtained from altimeters on board spacecraft. Examples include the altimeter on board Seasat.

At the present time terrain data is very sparse. The U.S. Defence Mapping Agency has digitized contour data from 1:250,000 maps for the whole U.S., and USGS plans to do the same from 1:24,000 maps. Other mapping organizations of Canada, U.K. and Australia are producing some DTMs. In the future it is expected that a world data base with good accuracy will be built up. However, it has been estimated that this would require 2×10^{15} bits of data to cover every square metre on land where each elevation is digitized to 16 bits.

3.3 Types and Interpolation of DTMs

Different types of DTMs may be distinguished by their coordinate

spacings. Since DTM interpolation is required (as an off-line process) in the rectification algorithms, this section presents different DTM types and their interpolation methods. For a detailed description, refer to [SCHU76b, ALLA78, PEUC78 and HEIL78].

3.3.1 Irregular DTMs

In an irregular DTM, reference points are not equally and regularly spaced. In some cases they are selected at random positions, while in other cases they are selected at positions where there is a change in slope or some distinct feature. The latter method gives an adequate description of the topography with the smallest number of reference points.

The contour DTM, acquired by digitizing contour maps, gives equally spaced elevation points. However, the spatial distribution of reference points is not regular, but is selected at equal intervals along the contour or spaced more closely along contours in areas where the terrain is hilly than when it is flat. Prior to using an irregular DTM, the general practice is to transform it to a regular DTM; i.e., to compute terrain heights at the nodes of a square grid.

One method [SCHU76] is to define the terrain height by:

$$h = a_0 + a_1x + a_2y + b_1x^2 + b_2xy + b_3y^2 \quad (3.1)$$

For the interpolation of a point, the coordinates and the height of each of the surrounding reference points are substituted into this

equation. Also each reference height is given a weight that is a monotonically decreasing function of the distance to the reference point. Then the parameters in the equation are solved by a method of least squares. Taking the origin of the coordinate system to be the interpolated point, where its height is required, then only a_0 needs to be computed. To keep the computation time within reasonable bounds, only reference points within a specified maximum distance from an interpolated point will be used. Generally, going from one interpolated point to an adjacent one, the surface defined by Equation 3.1 will change its orientation and, possibly, its shape. For this reason, this method is also referred to as the "moving surface" method.

Another method [HARD71] is to construct around each reference point a fixed surface which has a vertical axis of symmetry at that point. Interpolation is performed by summing the heights of all surfaces computed at that point.

3.3.2 Regular DTMs

In a regular DTM, spacings of the spatial coordinates are regular. Hence, reference points are at the nodes of a regular grid. Since the regular DTM consists of an easily addressable array of elevation values, it is simpler than the irregular DTM to store and access in computer storage. According to the sampling theorem, terrain variations which correspond to a wave length less than twice the grid spacing are not represented.

If the elevation of a point other than the nodes is required, it can be obtained by the interpolation methods discussed in Section 2.5, such as nearest neighbour, bilinear, cubic convolution or convolution with a larger kernel. If the DTM is dense enough to represent all significant changes in slope, then bilinear or even nearest neighbour interpolation is adequate.

3.3.3 Triangulated Irregular Network

Usually, terrain is not regular but changes from one land area to another. If a regular DTM is used to represent the terrain, it has to be adjusted to be adequate for the roughest terrain. But then the data is highly redundant in smooth terrain.

A triangulated irregular network [PEUC78 and HEIL78] tries to represent terrain adequately with the smallest number of reference points. This is done by covering the terrain with a network of triangles. Each triangle is a "faceted" representation of the terrain surface. The continuity of the surface representation is maintained by restricting the construction such that each triangle has only one other triangle sharing a side. Vertices are chosen at local maximum or minimum points. The sides of the constructed triangles follow ridges and channels.

Interpolation of a point inside a triangle follows from the condition that the point lies on the three-dimensional triangle formed by its surrounding vertices:

$$\text{Det} \begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0$$

where (x_i, y_i, z_i) , $i = 1$ to 3 , are the vertices of the triangle.

3.4 DTM Accuracy

3.4.1 Accuracy Requirement in Interpolated Coordinates

The DTM is warped as an off-line process close to the input image coordinates (u,v) . A DTM has three coordinates: x , y and z . Inaccuracies in x , y and z will cause an error in the interpolation position in the input image. The acceptable error is less than $1/2$ pixel. This subsection deals with the accuracy requirements in x , y and z for the interpolated DTM, as a function of sensor and satellite parameters.

The displacement d on ground is related to terrain height h by:

$$d = fh \tag{3.2}$$

where the factor f is satellite and sensor specific and is also a function of pixel position. Then:

$$\Delta d = f \Delta h \tag{3.3}$$

Let the pixel grid spacing be r . The usual requirement is that $d < 1/2 r$. Hence:

$$f \Delta h < r/2$$

and

$$\Delta h < r/(2f) \quad (3.4)$$

But the contributions to Δh are: spatial accuracy $\Delta p = \sqrt{\Delta x^2 + \Delta y^2}$, and elevation accuracy Δz . Let θ be the slope. Then Δh can be expressed in the RMS sense by:

$$\Delta h = \sqrt{\Delta p^2 \tan^2 \theta + \Delta z^2} \quad (3.5)$$

Substituting into Equation 3.4, we have

$$\sqrt{\Delta p^2 \tan^2 \theta + \Delta z^2} < r/(2f) \quad (3.6)$$

To obtain the most stringent requirement, we have to allow for the worst possible case of relief displacement; i.e., maximum f . The maximum value of f and the value of r for various sensors are given in Table 3.1.

Plots of the accuracy requirement for various sensors are shown in Figures 3.1 to 3.4 for terrain slopes of 20° , 30° and 45° . With a decrease in slope, the accuracy Δx (and hence Δy) can be relaxed as anticipated.

In Figures 3.1 and 3.2 the points A and B, for class A1 and B2 maps, lie to the left of the $\theta = 45^\circ$ line. This shows that DTMs obtained by digitizing 1:50,000 class A1 or B2 maps are adequate for $\Delta d < 1/2$ pixel position accuracy for these two types of imagery.

For SPOT side looking (26° off nadir) and Seasat SAR imagery, high resolution DTMs made by devices such as the Gestalt photomapper have to be

TABLE 3.1 RESOLUTION AND WORST RELIEF
DISPLACEMENT FACTOR

Sensor	Displacement $d = fh$	Resolution r
Landsat-4 TM	$d = 0.14 h$	30 m
SPOT MLA (nadir looking)	$d = 0.042 h$	20 m
SPOT MLA (side looking ¹)	$d = 0.60 h$	26 m
SPOT PLA (nadir looking)	$d = 0.042 h$	10 m
SPOT PLA (side looking ¹)	$d = 0.60 h$	13 m
Seasat SAR	$d = 2.3 h$	25 m

¹ In the SPOT side looking mode, the off nadir angle is 26°.

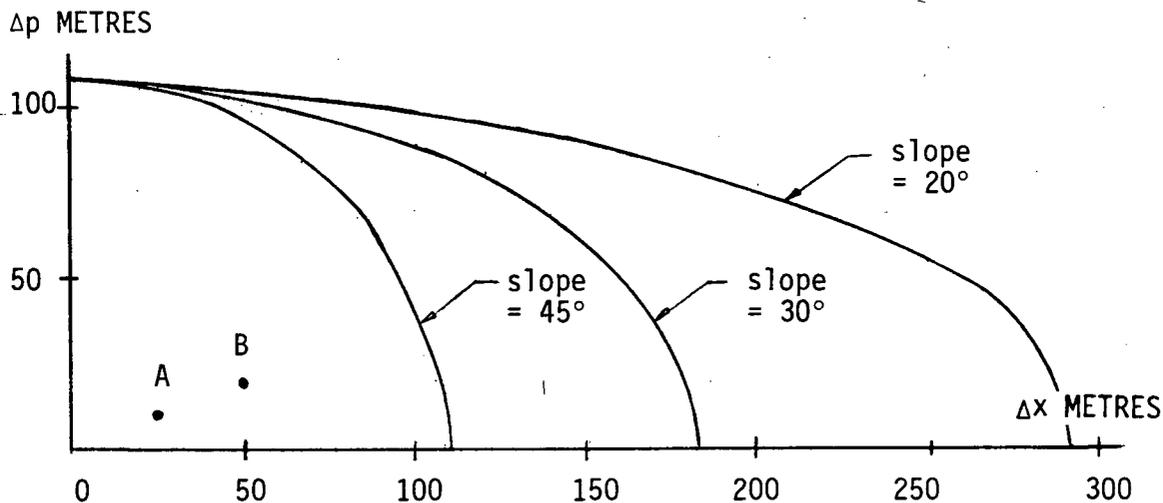


FIGURE 3.1 DTM ACCURACY REQUIREMENT FOR LANDSAT-4 TM SENSOR (A AND B ARE TYPICAL Δx AND Δp FOR A ACCURACY AND B ACCURACY 1:50,000 MAPS RESPECTIVELY)

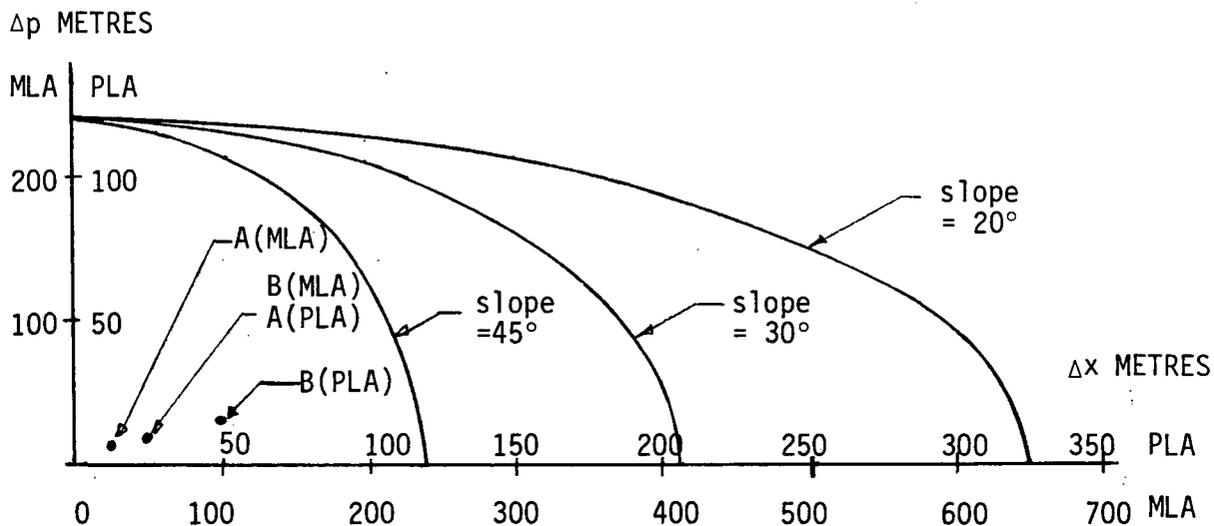


FIGURE 3.2 DTM ACCURACY REQUIREMENT FOR SPOT SENSORS, NADIR LOOKING (A AND B ARE TYPICAL Δx AND Δp FOR A AND B ACCURACY MAPS RESPECTIVELY)

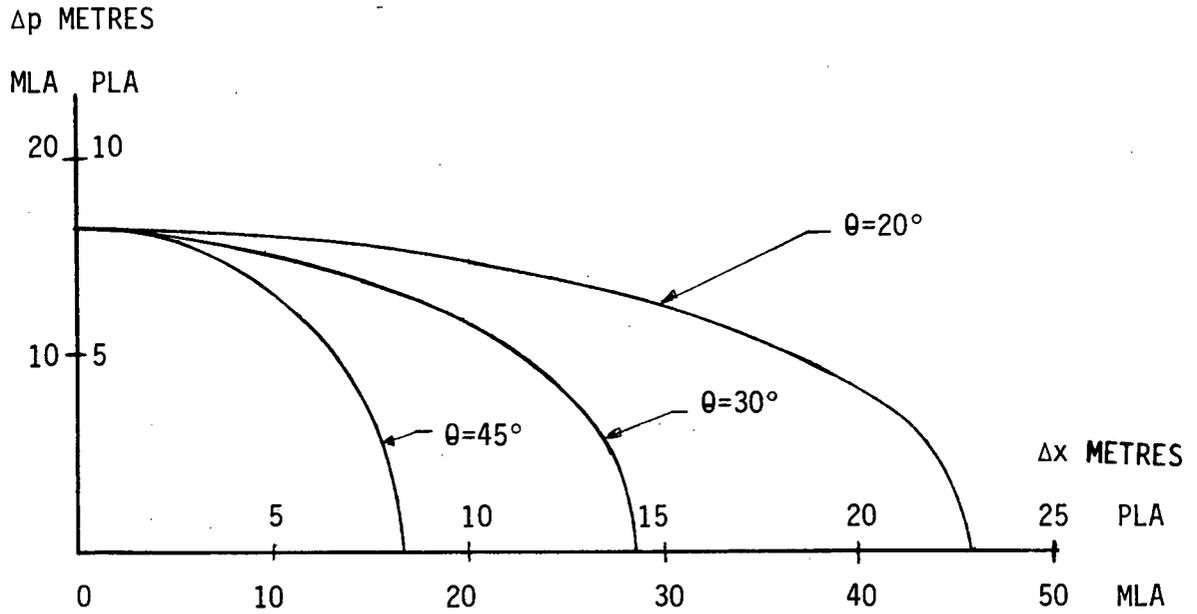


FIGURE 3.3 DTM ACCURACY REQUIREMENT FOR SPOT SENSORS,
OFF-NADIR LOOKING ANGLE = 26°

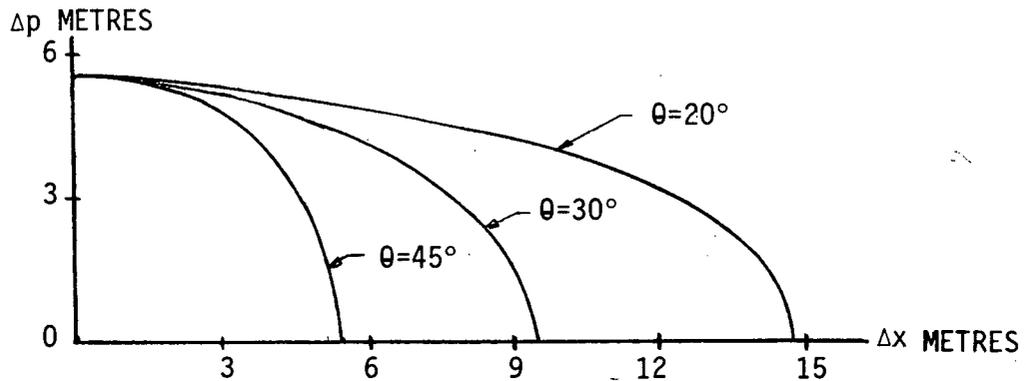


FIGURE 3.4 DTM ACCURACY REQUIREMENT FOR SEASAT SAR SENSOR

used if half a pixel accuracy is desired. This is shown in Figures 3.3 and 3.4 where the corresponding points A and B shown in the other two figures are now to the right of the $\theta = 20^\circ$ line.

Using a good approximation (say a 16×16 point kernel) of the sinc function to sample the raw DTM to the output coordinates, the inaccuracies in Δx , Δy , and Δz should be about the same in both coordinate systems. Hence, the accuracy requirements just discussed should also apply to the raw DTM.

3.4.2 DTM Interpolation in Scrolling Buffer

The DTM scrolling buffer shown in Figure 2.15 contains enough lines to ensure that all the heights of an output line can be found. However, the DTM pointer for each output point is usually **not** an integer and therefore, DTM interpolation is required to compute the heights. If nearest neighbour interpolation is used, the height extraction process will be very fast. A method is presented here, with examples, to determine when nearest neighbour interpolation can be used without significantly affecting the geometric mapping accuracy. (Caution: If the DTM is also used for radiometric correction, nearest neighbour interpolation may be inadequate.)

The method consists of two steps:

- (a) Determine the error due to nearest neighbour interpolation.

If this error is less than a pixel, then nearest neighbour

interpolation can be used. Let this error be e_1 .

- (b) If error e_1 is greater than a pixel, compare it with map induced error e_2 . If e_2 masks out e_1 , then use nearest neighbour interpolation, otherwise use bilinear interpolation.

The error e_1 , can be computed as follows. The maximum error made by using nearest neighbour interpolation in x and y is:

$$\Delta x = \Delta y = r' / 2 \quad (3.7)$$

where r' now is the resolution of the interpolated DTM. Therefore maximum e_1 is, from Equation 3.2, given by:

$$e_1 = f \sqrt{(\Delta x)^2 + (\Delta y)^2} \tan \theta$$

In terms of pixels:

$$e_1 = f \frac{\sqrt{2}}{2} \frac{r'}{r} \tan \theta \quad (3.8)$$

In the first example, assume that:

- (a) $r' = r$, i.e., the interpolated DTM has the same pixel spacing as the raw image.
- (b) Δx and Δy have uniform distributions.
- (c) θ has a maximum value of 45° and $\tan \theta$ has uniform distribution.

Then:

$$\begin{aligned} e_1 &= f \frac{\sqrt{2}}{2} \frac{r'}{r \sqrt{3}} \frac{\tan 45^\circ}{\sqrt{3}} \\ &= f \sqrt{2} / 6 \text{ pixels RMS} \end{aligned}$$

The computation of e_2 is given by Equations 3.3 and 3.5 in which $\Delta p = \sqrt{\Delta x^2 + \Delta y^2}$, and Δz as a function of map class is given in Section 3.1.

Two DTM sources are used, namely, photogrammetric stereo and 1:50000 Class A1 maps. The results of e_1 and e_2 are shown in Table 3.2. Here nearest neighbour interpolation can be used for all the imagery mentioned since e_1 is less than 1 pixel RMS. Note that in the last row, subpixel rectification accuracy can no longer be achieved for Seasat imagery due to the magnitude of e_2 .

A second example is presented here with the following changes:

- (a) $r' = 2r$
- (b) 1:50,000 Class B2 maps are used instead of Class A1 maps.

The results are presented in Table 3.3. Of particular interest are the Seasat SAR cases. If the DTM source is photogrammetric stereo, nearest neighbour interpolation should not be used since the error e_1 is now more than 1 pixel RMS. However if the DTM source is from the Class B2 map, then nearest neighbour can be used since error e_2 masks out e_1 .

Similar analysis can be performed for various slopes and DTM resolutions.

TABLE 3.2 DTM RESAMPLING IN SCROLLING
BUFFER (I)

$$r' = r$$

DTM Source	Satellite Sensor	e_1 Pixels RMS	e_2 Pixels RMS	Total Error Pixels RMS	Recommended DTM Interpolation Method
Photo-grammetric stereo	Landsat-4 TM	0.03	negligible	0.03	Nearest neighbour
	SPOT (nadir looking)	0.01		0.01	Nearest neighbour
	SPOT (side looking)	0.14		0.14	Nearest neighbour
	Seasat SAR	0.54		0.54	Nearest neighbour
1:50,000 Class A1 maps	Landsat-4 TM	0.03	0.08	0.09	Nearest neighbour
	SPOT MLA (nadir looking)	0.01	0.036	0.037	Nearest neighbour
	SPOT PLA (nadir looking)	0.01	0.072	0.073	Nearest neighbour
	SPOT MLA (side looking)	0.14	0.39	0.41	Nearest neighbour
	SPOT PLA (side looking)	0.14	0.78	0.79	Nearest neighbour
	Seasat SAR	0.54	1.56	1.65	Nearest neighbour

TABLE 3.3 DTM RESAMPLING IN SCROLLING
BUFFER (II)

$$r' = 2r$$

DTM Source	Satellite Sensor	e_1 Pixels RMS	e_2 Pixels RMS	Total Error Pixels RMS	Recommended DTM Interpolation Method
Photo-grammetric stereo	Landsat-4 TM	0.06	Negligible	0.06	Nearest neighbour
	SPOT (nadir looking)	0.02		0.02	Nearest neighbour
	SPOT (side looking)	0.28		0.28	Nearest neighbour
	Seasat SAR	1.08		1.08	Bilinear
1:50,000 Class B2 Maps	Landsat-4 TM	0.06	0.16	0.10	Nearest neighbour
	SPOT MLA (nadir looking)	0.02	0.072	0.04	Nearest neighbour
	SPOT PLA (nadir looking)	0.02	0.14	0.14	Nearest neighbour
	SPOT MLA (side looking)	0.28	0.78	0.83	Nearest neighbour
	SPOT PLA (side looking)	0.28	1.56	1.58	Nearest neighbour
	Seasat SAR	1.08	3.10	3.28	Nearest neighbour

CHAPTER 4 SCANNER IMAGERY

4.1 Introduction

Scanner imagery is of two types: mechanical and electronic. Examples of the former are Landsat multispectral scanner (MSS) and TIROS-N imagery while an example of the latter is SPOT imagery.

A multispectral scanner (MSS) system consists essentially of three parts: a mechanical, an optical and an electronic system as illustrated in Figure 4.1. The line scan is performed in the mechanical system by a rotating or oscillating mirror. The optical system consists of a beam splitting device which divides the incident light into several spectral bands. The electronic system is an array of detectors. Each spectral band is focused onto a set of detectors. The satellite moves forward during the scanning process, hence producing a two-dimensional image. Figure 4.2 shows a typical way in which imagery is obtained.

An electronic scanning sensor employs a large number of detectors (in the thousands), as shown in Figure 4.3, in the focal plane so that no oscillating mirror is required. The detectors are sampled electronically in sequence. For this reason, the electronic scanner is also called a "push-broom" scanner. The main difference between the electronic scanner and the mechanical scanner lies in the physical arrangement of the detectors. In the former, each detector "sees" only one

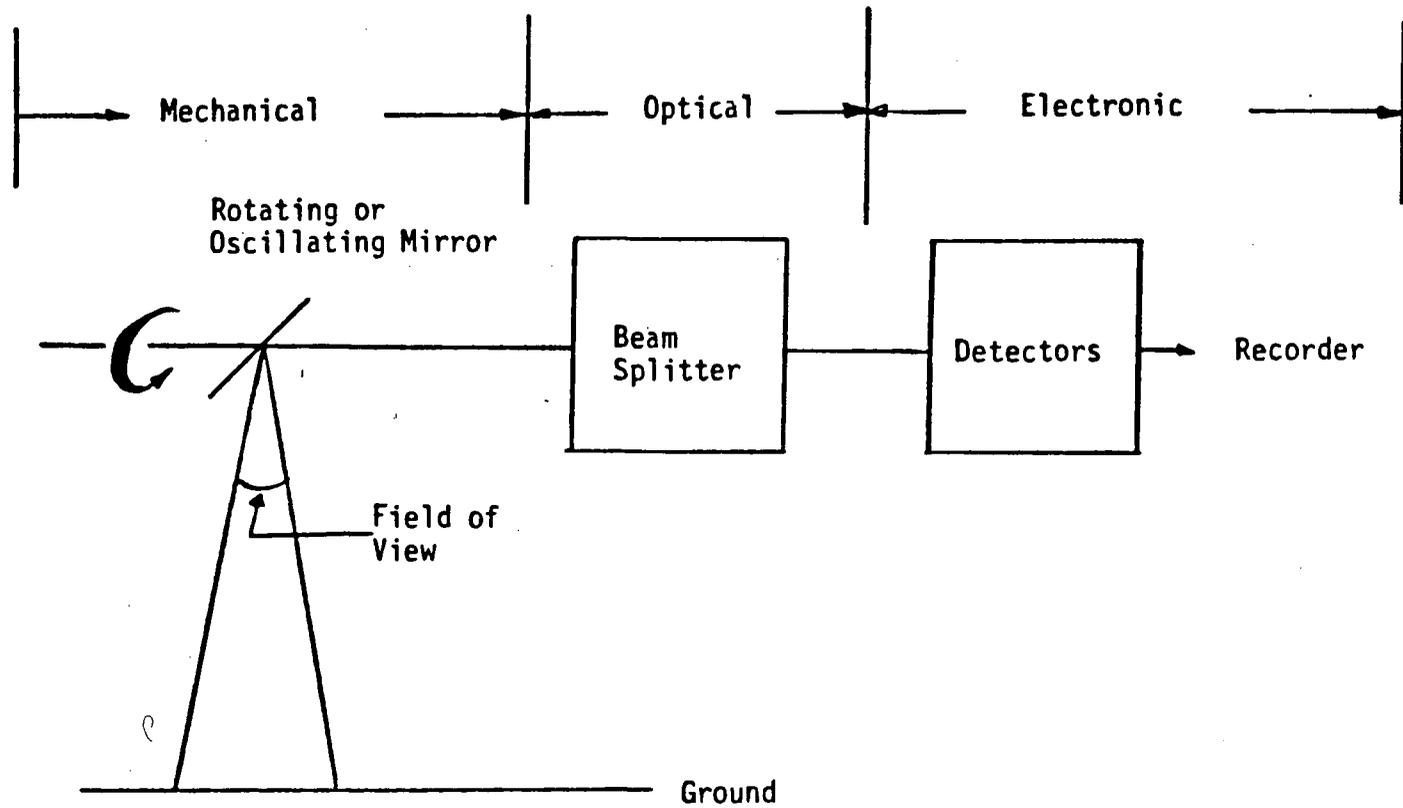


FIGURE 4.1 SCANNER SENSOR SYSTEM

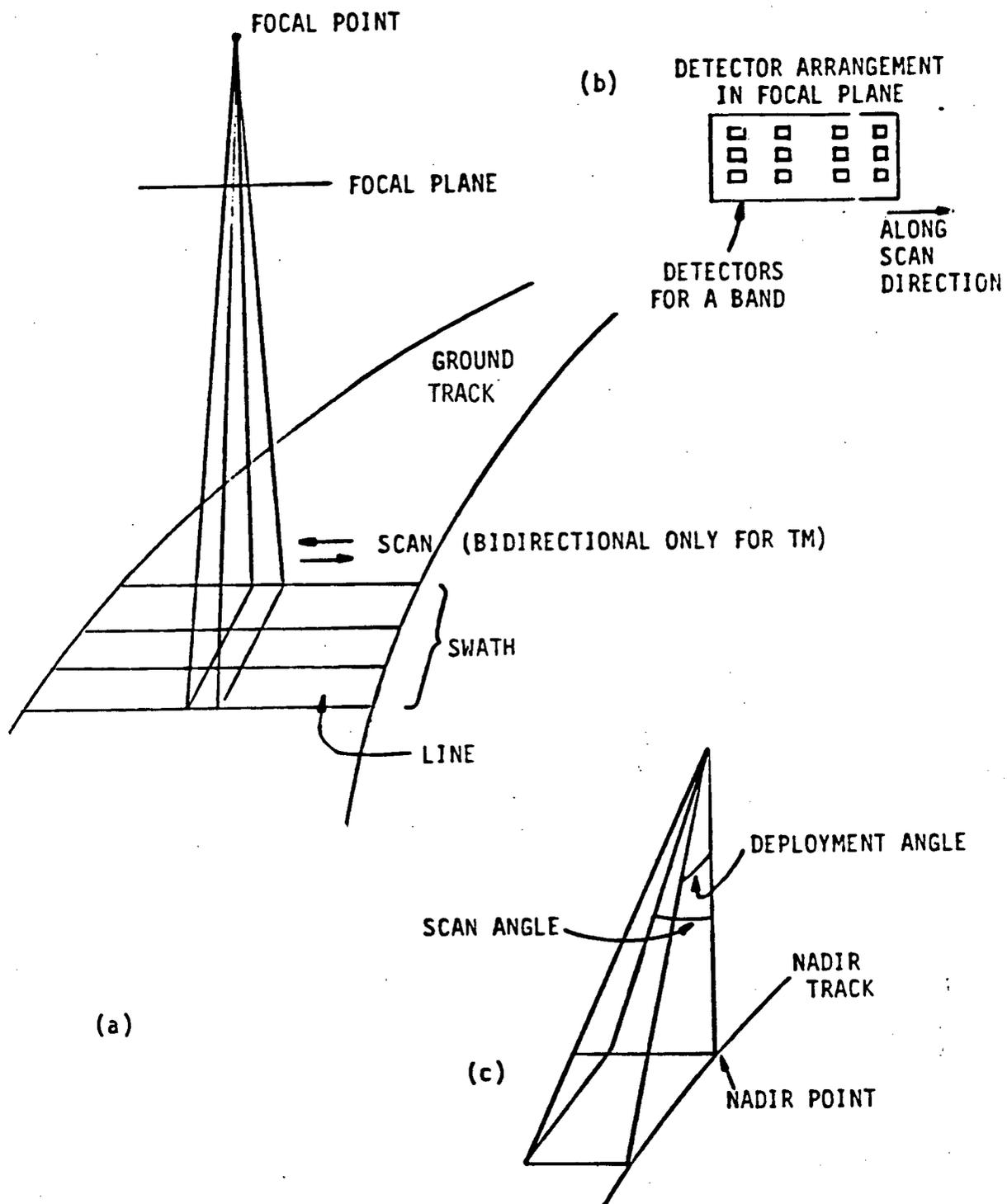


FIGURE 4.2

TYPICAL MECHANICAL SCANNING SENSOR

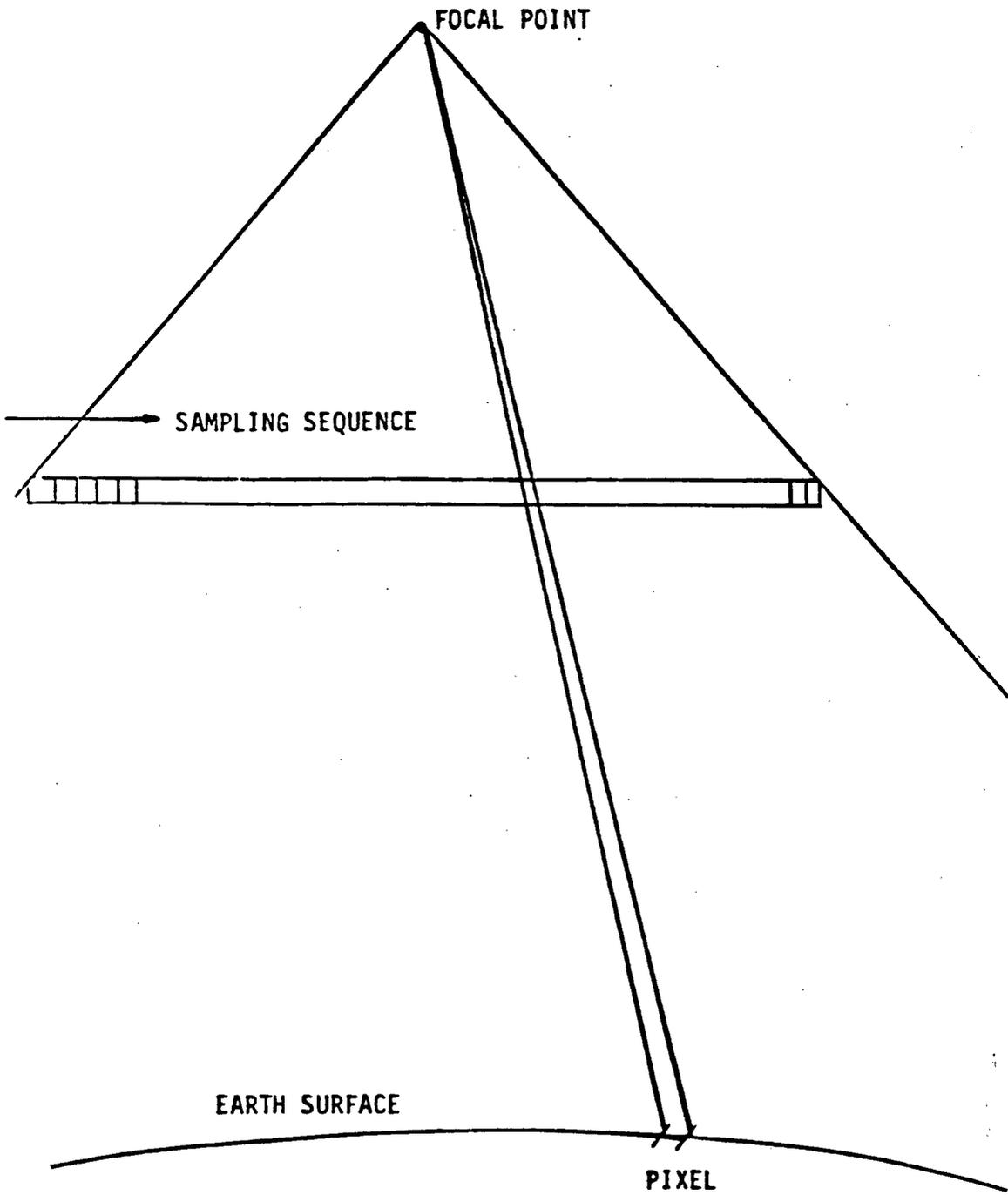


FIGURE 4.3 TYPICAL ELECTRONIC SCANNING SENSOR

spot on the ground along a line of the image, while in the latter, a detector "sees" spots on the ground along a column of the image.

Raw scanner data do not need to be processed to be rendered as an image. However without processing, the raw imagery will be geometrically distorted due to earth rotation, nonlinear sensor sweep, panoramic distortion, variation in satellite attitude angles (pitch, roll and yaw), scan skew and the like. Furthermore, relief displacement is significant in Landsat-4 TM imagery and will be significant in SPOT imagery.

Digital rectification corrects these distortions using information determined from a priori data and measurement data such as GCPs. Panoramic distortion and distortions due to nonlinear scanner sweep and scan skew are systematic; i.e., their effects can be predicted in advance. Distortions due to earth rotation are a function of satellite latitude and orbit and thus can be predicted from tracking data. While distortions due to attitude and altitude changes are also systematic, the parameters required to account for them are not known. If these distortions are to be corrected, their effects must be measured for each image. The measurement technique involves identifying in the raw image recognizable geographic features (GCPs) whose map coordinates are known. Using the image locations of these points, the image distortion caused by the sensor platform can be determined.

The purposes of this chapter are as follows:

- (a) Review and compare various rectification algorithms pertinent to

MSS imagery. The study shows that attitude modelling is superior to other methods in terms of accuracy with a reasonable number of GCPs.

- (b) Establish how GCPs can be used to refine the attitude angles and orbit parameters.
- (c) Identify low and high frequency distortions common to all scanner imagery.
- (d) Identify geometric distortions specific to Landsat-4 TM and SPOT imagery.
- (e) Identify attitude and orbit parameters to be used in the recursive estimator in the navigation phase. The recursive estimator is tested with Landsat and TIROS-N imagery. These two types of imagery have different resolutions and panoramic distortions.
- (f) Investigate the problem of relief displacement, especially in Landsat-4 TM and SPOT imagery. Synthesize relief distorted imagery and rectify the resulting distorted imagery.
- (g) Develop an interpolation algorithm to handle imagery for which pixel spacings are not uniform. This is applicable to Landsat-4 TM imagery where scan gaps are present.

4.2 Overview of Geometric Effects

4.2.1 Common Distortions

Raw scanner imagery is geometrically distorted due to earth shape,

scanner geometry and satellite motion. These characteristic distortions can be classified into two types: low frequency and high frequency.

The low frequency distortions are:

- (a) Panoramic. The pixel size increases with the scan angle which the imaging sensor makes with the vertical.
- (b) Attitude. Pitch causes a line shift, roll causes a pixel shift while yaw causes an image rotation.
- (c) Earth rotation. Earth rotation causes an image skew.
- (d) Mirror scan. The scanning rate of a mechanical scanning sensor is nonuniform, hence the sensor pointing angle cannot be determined as a simple linear function of the pixel number. For electronic scanners, a similar problem may occur because scanning is not done at a rate corresponding to equal increments in scan angle.
- (e) Band offset. There exists a time delay in sampling between bands and the sensors of different bands are positioned differently on the focal plane. These factors contribute to an offset between bands.

The high frequency components are:

- (a) Sensor offset. In MSS and TM the detectors are deployed in a matrix pattern so that one swath contains more than one line of imagery data. There exists a time delay in sampling the corresponding pixel between the detectors. In addition, the

corresponding detectors are positioned differently on the focal plane. These factors contribute to sensor offset which is different from line to line within a swath.

- (b) Line length variation. The line length in a mechanical scanner varies due to variations in the scanning mirror speed. This problem does not exist in linear array imagery which is obtained by electronic scanners. But, in this case, a similar problem arises from the fact that gaps occur between groups of detectors. Thus the output pixel spacing is regular only within each group of detectors.
- (c) Earth rotation. Earth rotation between detectors within a swath has to be taken into consideration. Earth rotation is not linear from line to line as illustrated in Figure 4.4. This problem only occurs in imagery with more than one line per swath.
- (d) Satellite altitude and velocity. Satellite altitude variation produces a change in the scan line spacing within a swath, while satellite velocity variation produces a change in swath centre spacing. The two factors when coupled together cause irregular line spacings in the along track direction.
- (e) Relief displacement. This causes a pixel shift in the along scan direction and is discussed in more detail in Section 4.4.4.

The bilinear coefficients used in remapping can only account for low frequency distortions. High frequency components are added to the result of the bilinear transformation in the first and second image

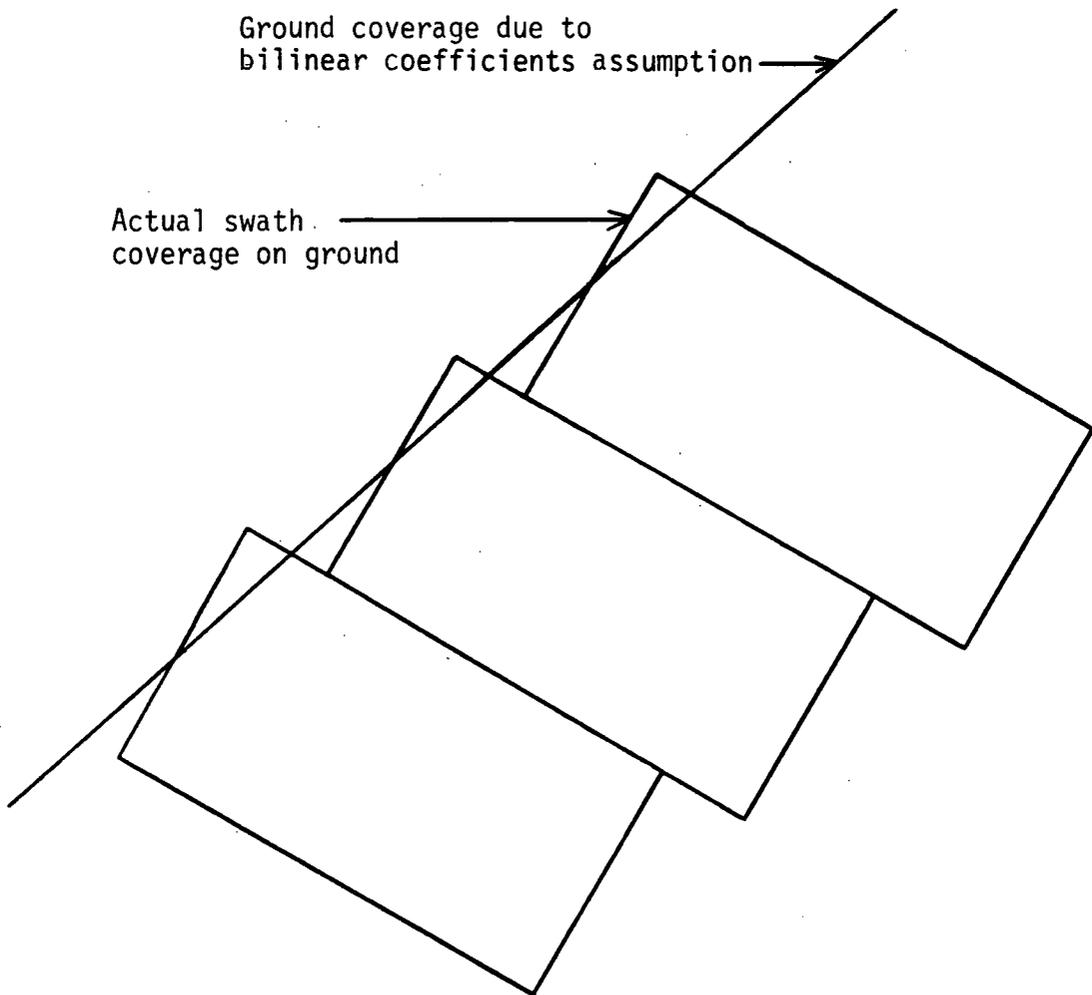


FIGURE 4.4 EARTH ROTATION EFFECT

processing passes.

4.2.2 Landsat-4 TM Imagery Specific Distortions

The major TM imagery distortion, in addition to the ones mentioned in Section 4.2.1, is the underlap and overlap of swath data. This is caused by variations in satellite altitude (from 685 to 740 km) and velocity, jitter, bowtie effect and by earth rotation as described in Appendix A. The overlap and underlap is known as scan gap and is also described and analyzed in detail in Appendix A. The scan gap size varies between ± 2 data points and this has a significant impact on interpolation (in the second pass) since the data samples are no longer uniformly spaced. A special interpolation kernel to minimize the radiometric error in interpolation will be discussed in Section 4.5.

4.2.3 SPOT Imagery Specific Distortions

The SPOT sensor is described in Appendix B. The geometric distortions associated with SPOT are as mentioned in Section 4.2.1, but the following points should be noted:

- (a) Since each swath contains only one line of imagery data, the high frequency earth rotation effect in the along scan direction does not exist. Also altitude and velocity variations no longer produce nonuniform line spacings in the across scan direction.
- (b) Pixel resolution in the along scan direction is a function of off nadir angle as demonstrated in Appendix B.

SPOT imagery suffers no high frequency distortions other than relief displacement.

4.2.4 Relief Displacement

Any imaging system introduces relief displacement, the magnitude of which depends on the imaging parameters.

The exact formulation for relief displacement in units of raw input pixels can be derived with the aid of Figure 4.5. The displacement in scan angle β can be expressed explicitly by:

$$\tan \Delta\beta = \frac{Hz}{R} \frac{\sin \beta}{H \cos \beta - (R+z) \sqrt{1 - (H \sin \beta / R)^2}} \quad (4.1)$$

where z is the terrain height, R is the local earth radius and H is the local orbit radius. Note that $R \gg z$ and $\Delta\beta$ is small enough so that $\tan \Delta\beta \approx \Delta\beta$. Then:

$$\frac{\Delta\beta}{z} = \frac{H}{R} \frac{\sin \beta}{H \cos \beta - R \sqrt{1 - (H \sin \beta / R)^2}} \quad (4.2)$$

where $\frac{\Delta\beta}{z}$ is in units of angular measure per unit height. This can easily be converted into pixels per unit height by:

$$\frac{\Delta p}{z} = \frac{\Delta P}{\Delta\beta} \frac{\Delta\beta}{z}$$

where Δp is the displacement in pixels and $\Delta p / \Delta\beta$ is the angular resolution.

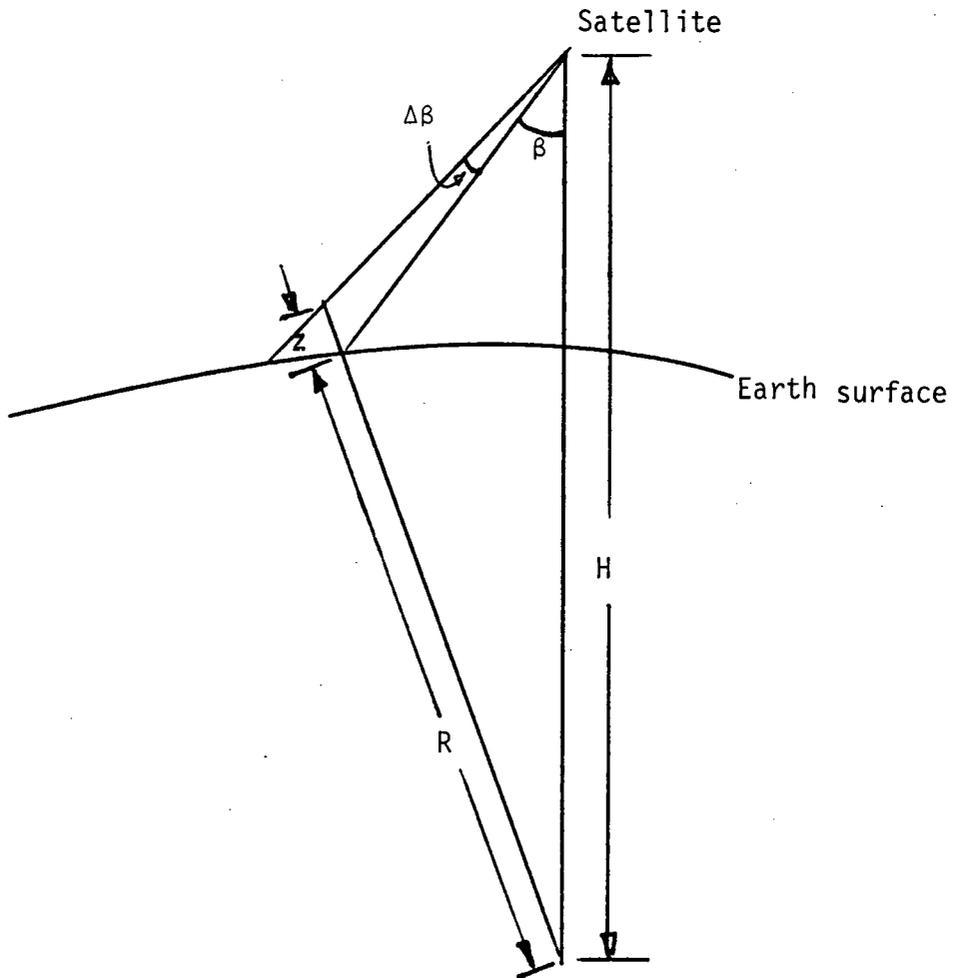


FIGURE 4.5 RELIEF DISPLACEMENT IN SCANNER IMAGERY

The angle β is a function of pixel number, mirror scan nonlinearity and roll angle. The roll angle cannot be ignored especially in the case of SPOT PLA imagery when the off nadir angle is at its maximum, i.e. 26° . In this case a roll angle of 0.5° would introduce a 10 metre shift per km height.

4.3 Rectification Transformation

4.3.1 Closed Form Solution

This section develops a closed form transformation between the instantaneous sensor viewing angle and the geographic coordinates of any point on the earth's surface. This is referred to as the forward transformation. Derivation of the transformation, detailed in Appendix E, starts by postulating a unit scan vector viewing a GCP in the sensor platform coordinate system. This vector, shown in Figure 4.6, is then transformed into the satellite coordinate system and finally to the ECR system. Let \underline{u} be the unit vector pointing from the satellite to a point on the earth, \underline{R}_g the ground point vector and \underline{R}_s the satellite position vector as shown in Figure 4.6. The transformation from the scan vector $D\underline{u}$ to the ground point at \underline{R}_g is given by:

$$\underline{R}_g = \underline{R}_s + D\underline{u} \quad (4.3)$$

where D is the distance between the satellite and the image point.

Appendix E shows that D can be explicitly expressed as:

$$D = \underline{u} \cdot \underline{R}_s - e' R_z u_z + \sqrt{T} \quad (4.4)$$

where $e' = (a^2 - b^2) / b^2$

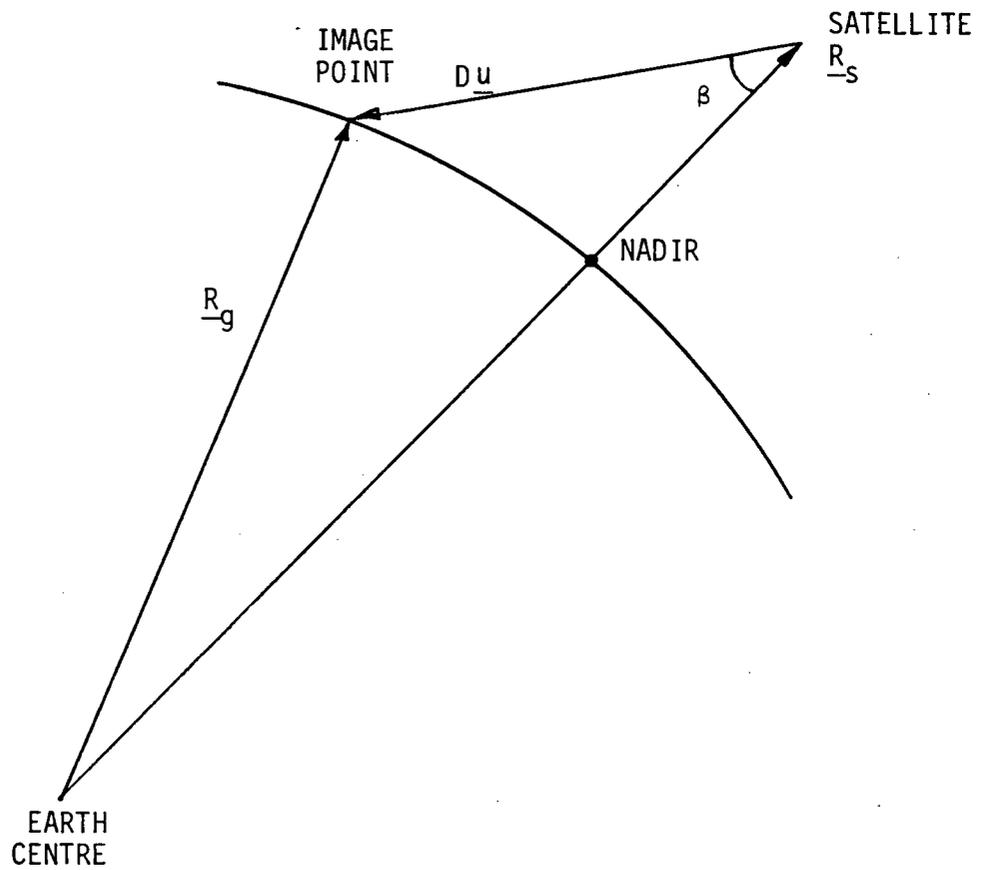


FIGURE 4.6 SCAN VECTOR GEOMETRY

a = semi-major axis of the earth

b = semi-minor axis of the earth

R_z = z-component of \underline{R}_s

u_z = z-component of \underline{u}

$$\text{and } T = (u_z^T \cdot \underline{R}_s + e' R_z u_z)^2 - (1 + e' u_z^2)(|\underline{R}_s|^2 + e' R_z^2 - a^2) \quad (4.5)$$

The functional form of \underline{u} is developed in Appendix E where it is expressed in terms of satellite orbit position, satellite attitude angles, sensor deployment angle and scan angle.

4.3.2 Block Size

Once the attitude and orbit parameters are known it is quite straightforward to transform from line, pixel coordinates to ECR coordinates. But the inverse transformation is required to perform the remapping which is performed into three passes. The transformation in each pass is performed by a pair of piecewise bilinear polynomials. The first and second intermediate images are segmented into blocks and a set of bilinear remapping coefficients is associated with each block. The remapping accuracy decreases with increased block size. A trade-off study has been performed between remapping accuracy and block size and this section gives the results of the study.

In the investigation, nominal sensor and orbit parameters were used. The block size and RMS error quoted in the figures to be presented are in ground distance in metres, which when combined with the

required output pixel size would give the block size and RMS error in terms of output pixel and output line spacings. The graphs were determined experimentally.

Figure 4.7 shows the results for Landsat-1, -2 and -3 MSS imagery with resolution of about 79 metres X 79 metres. Hence a reasonable output pixel size would be 50 metres X 50 metres. From the figure, if a 0.1 pixel (5 metres) RMS error is required then a block size of 160 X 160 (8 km X 8 km) has to be used. Figure 4.7 also shows the corresponding results for Landsat-4 MSS and TM imagery. The two graphs illustrate that the block size required for Landsat-4 MSS can be larger than that for Landsat-1, -2 and -3 for the same RMS error. This is because in Landsat-1, -2 and -3 the drift in each attitude angle is 10^{-2} degree per second whereas in Landsat-4 it is 10^{-6} degree per second. The attitude angles are thus a significant factor in the block size.

Figure 4.8 shows the results for SPOT PLA and MLA imagery (Section 4.2.3). In this figure the block size is also a function of the off nadir pointing angle. From this the block size for any off nadir pointing angle can be determined for a fixed RMS error.

4.4 Recursive Estimator in Navigation

4.4.1 State Vector Parameters

Rectification accuracy of satellite borne line scan imagery is a function of satellite orbit and attitude accuracies. Satellite orbit

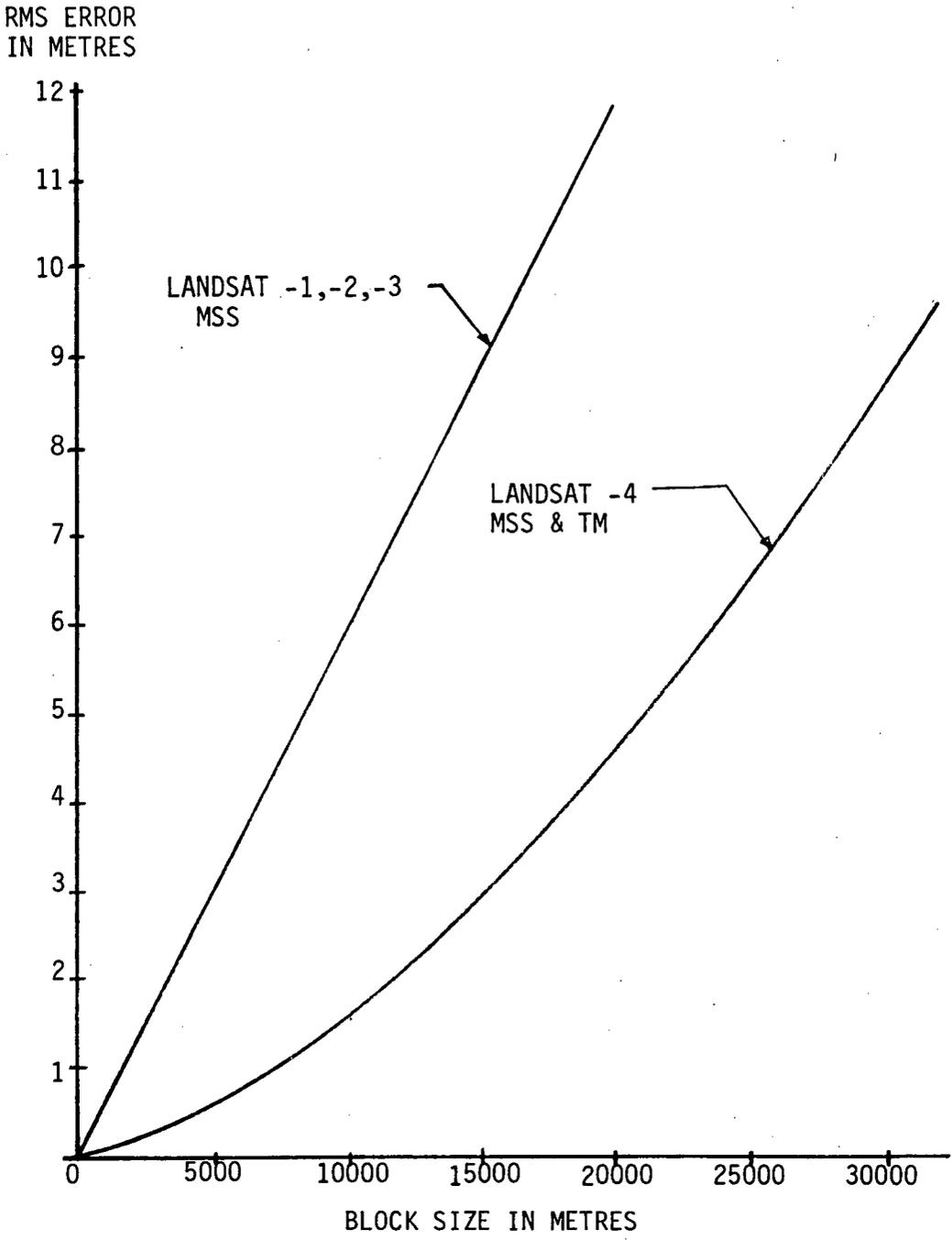


FIGURE 4.7 ACCURACY VERSUS BLOCK SIZE FOR LANDSAT SENSORS

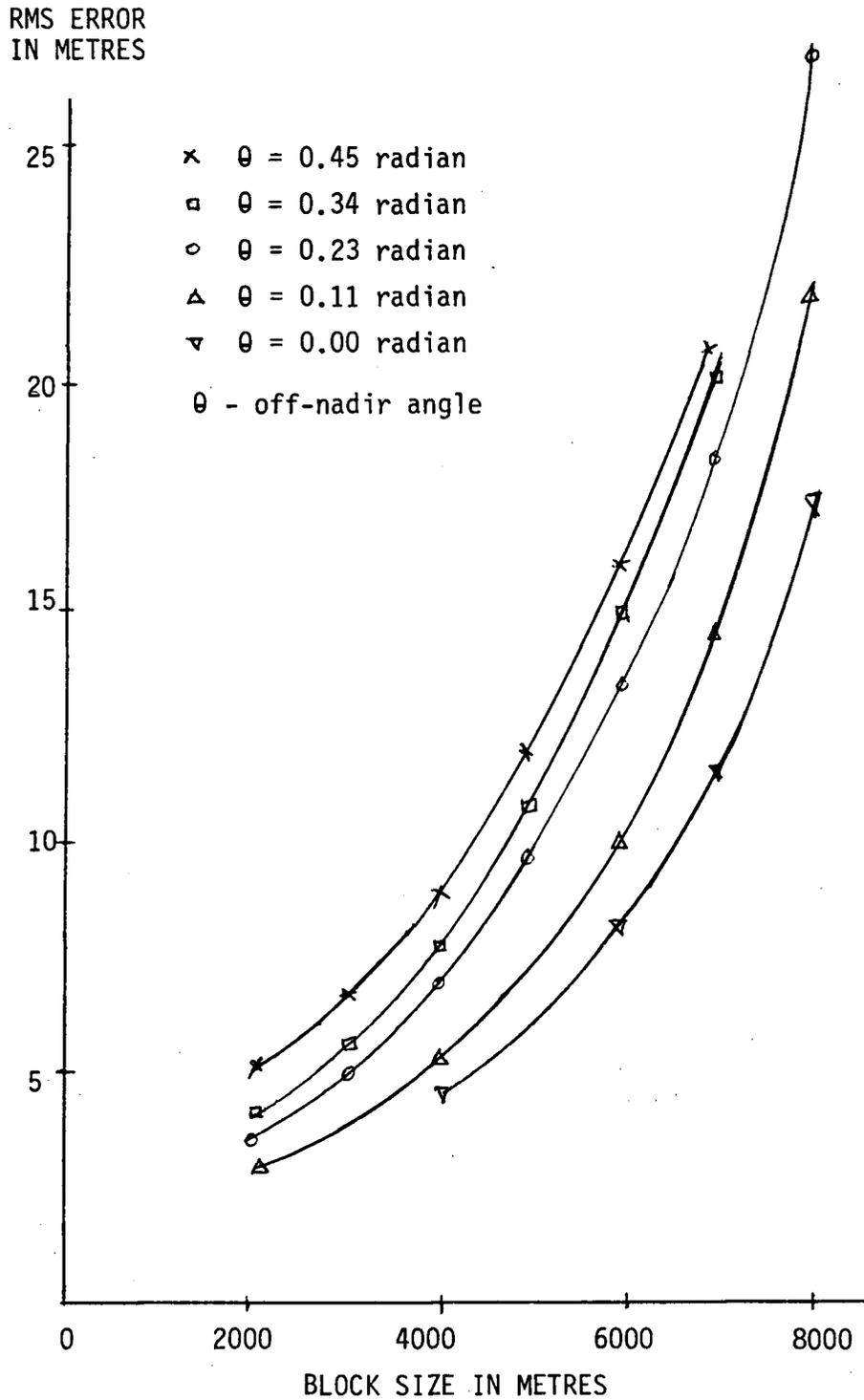


FIGURE 4.8 ACCURACY VERSUS BLOCK SIZE FOR SPOT SENSORS

information can be determined from ground based tracking measurements and attitude data are derived from onboard measurements and are telemetered to the ground. When such attitude measurements are made, they are not precise enough for a priori subpixel level rectification. The state vector $\underline{\Psi}$ (see Section 2.2.5) then contains the time series coefficients of the attitude angles. A further variable in the state vector is the satellite altitude above the earth's surface.

Landsat-1 and -2 imagery has been rectified to subpixel accuracy, with a cubic polynomial in each attitude angle and a constant in satellite altitude in the state vector.

4.4.2 Extension to TIROS-N Imagery

4.4.2.1 Objective

Guided by Konecny's survey [KONE76] and the findings in the rectification of Landsat-1 and -2 imagery, the recursive estimator technique has been extended in this thesis to the rectification of TIROS-N imagery. Such imagery suffers from a much higher panoramic distortion, covers a much larger area for a typical scene and has a considerably lower resolution than the Landsat imagery. The objective is to verify this method independent of resolution and panoramic distortion.

As an example, a TIROS-N Advanced Very High Resolution Radiometric (AVHRR) image was chosen for study. Simulation results have shown that the technique could yield subpixel rectification geometric accuracy

with about 10 to 20 GCPs under favourable conditions.

4.4.2.2 TIROS-N Characteristics

In order to analyze the transformation accuracy, it is important to understand the satellite orbit and its AVHRR sensor characteristics. The nominal values of the orbit and sensor parameters for the AVHRR are shown in Table 4.1.

Transformation accuracy with a time series model was shown by Caron and Simon [CARO75] to be within one pixel (79 metres) for Landsat MSS imagery. However, orbit and sensor specifications for the TIROS-N AVHRR are somewhat different from those for Landsat MSS. The main difference relating to transformation accuracy is the viewing angle of 110.8° (compared to 11.5° in Landsat MSS). Therefore, a scan line in a TIROS-N image covers about 3000 km compared to 185 km in a Landsat MSS image. The wider area of coverage introduces more panoramic distortion in an AVHRR image than in a Landsat MSS image. The effect of earth curvature is also considerably larger than in Landsat. Typically for AVHRR imagery a pixel at the scene centre covers 0.78 km compared with 4.5 km at the edge, giving a ratio of about 5.8 to 1; the corresponding ratio for Landsat imagery is 1.01 to 1. A graph of resolution versus pixel coordinates is shown in Figure 4.9. The line spacing is about 1.1 km as compared to 79 metres in Landsat MSS.

TABLE 4.1 TIROS-N CHARACTERISTICS

Semi-major axis	7211 km
Eccentricity	0.001
Inclination	98.7°
Nodal period	101.58 minutes
Field of view	110.8°
Scan period	1/6 second
Number pixels per scan	2048
Pixel sampling rate	39936 samples per second

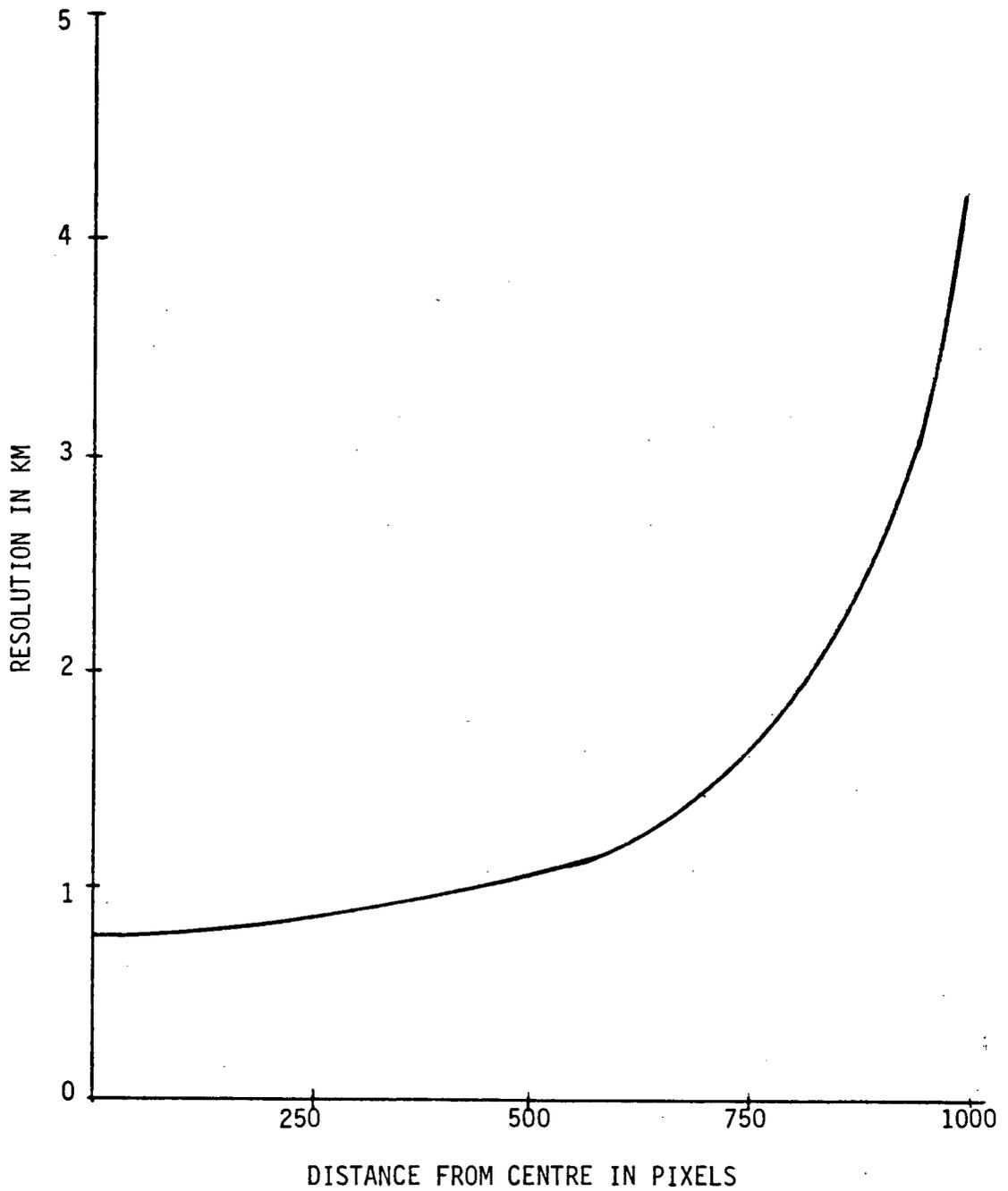


FIGURE 4.9 RESOLUTION VERSUS DISTANCE FROM CENTRE FOR TIROS-N IMAGERY (ASSUMING ZERO ATTITUDE ANGLES)

4.4.2.3 Orbit Accuracy Analysis

Geometric transformation is a function of the following parameters:

- (a) Satellite inclination — Inaccuracy will give an image rotation.
- (b) Nodal period and altitude — These give a scaling change in the image. These two parameters are related by Kepler's laws of motion.
- (c) Eccentricity — Since this value is about 0.001, the satellite altitude can vary by about 14 km, taking the earth centre as a focus of the orbit.
- (d) Equatorial crossing — This is the longitude at which the satellite crosses the equator. Any inaccuracy will give an along scan error.
- (e) Elapsed time between equatorial crossing and image acquisition — Inaccuracy will give an along track error.

The equatorial crossing has to be known to within 0.005° resulting in no more than 0.5 pixel error in the along scan direction. Also, the elapsed time has to be known to within 0.08 seconds resulting in no more than 0.5 line error in the along track direction.

The altitude is assumed to be the major contribution to geometric transformation error in the orbital parameters and requires an update in the recursive filter.

4.4.2.4 Attitude Accuracy Analysis

The recursive filtering technique estimates the attitude angles only to a certain accuracy. It can be shown that in the absence of yaw (κ) and pitch (ϕ), roll (ω) will have to be determined to a precision of

$$|\Delta\omega| \leq 0.48 \text{ mrad} \quad (4.6a)$$

in order for the along scan pixel coordinate to be accurate to ± 0.5 pixel. Similarly for pitch (ϕ),

$$|\Delta\phi| \leq 0.66 \text{ mrad} \quad (4.6b)$$

will result in ± 0.5 line accuracy in the across scan direction. Finally for yaw (κ),

$$|\Delta\kappa| \leq 0.45 \text{ mrad} \quad (4.6c)$$

will result in ± 0.5 line accuracy. If these accuracies for the attitude angles are attainable, then the geometric accuracy as a result of the combined effect will be $\sqrt{3} \times 0.5 = 0.86$ pixels, thus subpixel accuracy will be achieved.

It is possible to relate the RMS deviations $\Delta\omega$, $\Delta\phi$ and $\Delta\kappa$ to the accuracy of designating GCPs, and thus arrive at an estimated mapping accuracy and the number of GCPs required. Assume that any GCP can be located to within 1 pixel (i.e., 1 instantaneous field of view, or IFOV) for designation purposes. Then for equally probable roll, pitch and yaw, a GCP measurement results in $\Delta\omega \approx \pm 0.96$ mrad, $\Delta\phi \approx \pm 1.32$ mrad and $\Delta\kappa \approx \pm 0.90$ mrad. For n GCP measurements accurate to ± 1 pixel, the individual error of each attitude angle must be combined by the Gaussian rule so

that:

$$\Delta\omega_n = +0.96 / \sqrt{n} \text{ mrad}$$

$$\Delta\phi_n = +1.32 / \sqrt{n} \text{ mrad}$$

$$\text{and } \Delta\kappa_n = +0.90 / \sqrt{n} \text{ mrad}$$

To attain the attitude accuracies shown in Equations 4.6a, b and c for subpixel geometric accuracy, it follows that $n \geq 4$. That is, in this example at least 4 GCPs are required to reduce the attitude angles to acceptable levels. In practice, GCP designation of 1 pixel accuracy may be difficult to achieve. If each GCP measurement of m IFOV is attained, $4m^2$ GCPs will be necessary for the acceptable attitude angles. Thus, if in practice ± 2 IFOVs precision per GCP measurement is made, a total of 16 GCPs will be needed.

4.4.2.5 Experimental Results and Discussions

Experiments using the recursive estimator technique were performed with a TIROS-N image covering the western part of the United States. In the experiments, error is defined:

$$\text{Residual RMS error} = \sqrt{\frac{\sum_{i=1}^n (\Delta x_i^2 + \Delta y_i^2)}{n}} \quad (4.7)$$

where n is the number of GCPs processed, Δx_i and Δy_i are the differences in pixel and line coordinates for a particular GCP i , between the actual and estimated (using the updated state vector after processing n GCPs) values.

Verification Test 1

In this experiment each attitude angle was allowed to vary between $\pm 0.5^\circ$ as a cubic function of time. The time duration was selected to be about 200 seconds which corresponds to 1,200 lines of imagery data and to about 1,300 km in the along track direction. Nineteen GCPs were simulated. The object of the experiment was to study the performance as a function of the number of GCPs. The selection of the initial covariance matrix $P(0)$ is based upon the following assumptions:

- (a) The components of $\underline{\psi}$ are uncorrelated.
- (b) The uncertainty in the zeroth order component is the same for each attitude angle and is bounded by the attitude measurement uncertainty given in the TIROS-N specification. This is 0.5° .
- (c) The uncertainty in the j -th order component is the same for each attitude angle and its contribution at any time to the attitude angle does not exceed the contribution made by the $(j-1)$ th component.
- (d) The uncertainty in the semi-major axis follows from the TIROS-N specification. This is 5 km.

In the experiment the confidence value of GCP measurement ξ is first set equal to 0.5 pixel. Figure 4.10 shows the result of the experiment. The final residual error reduces to less than 0.02 pixel. The value of ξ was then set equal to 4 pixels and the result is also shown in Figure 4.10. After processing the GCPs, the residual error reduces to about 0.7 pixel. This error reduces further by processing more

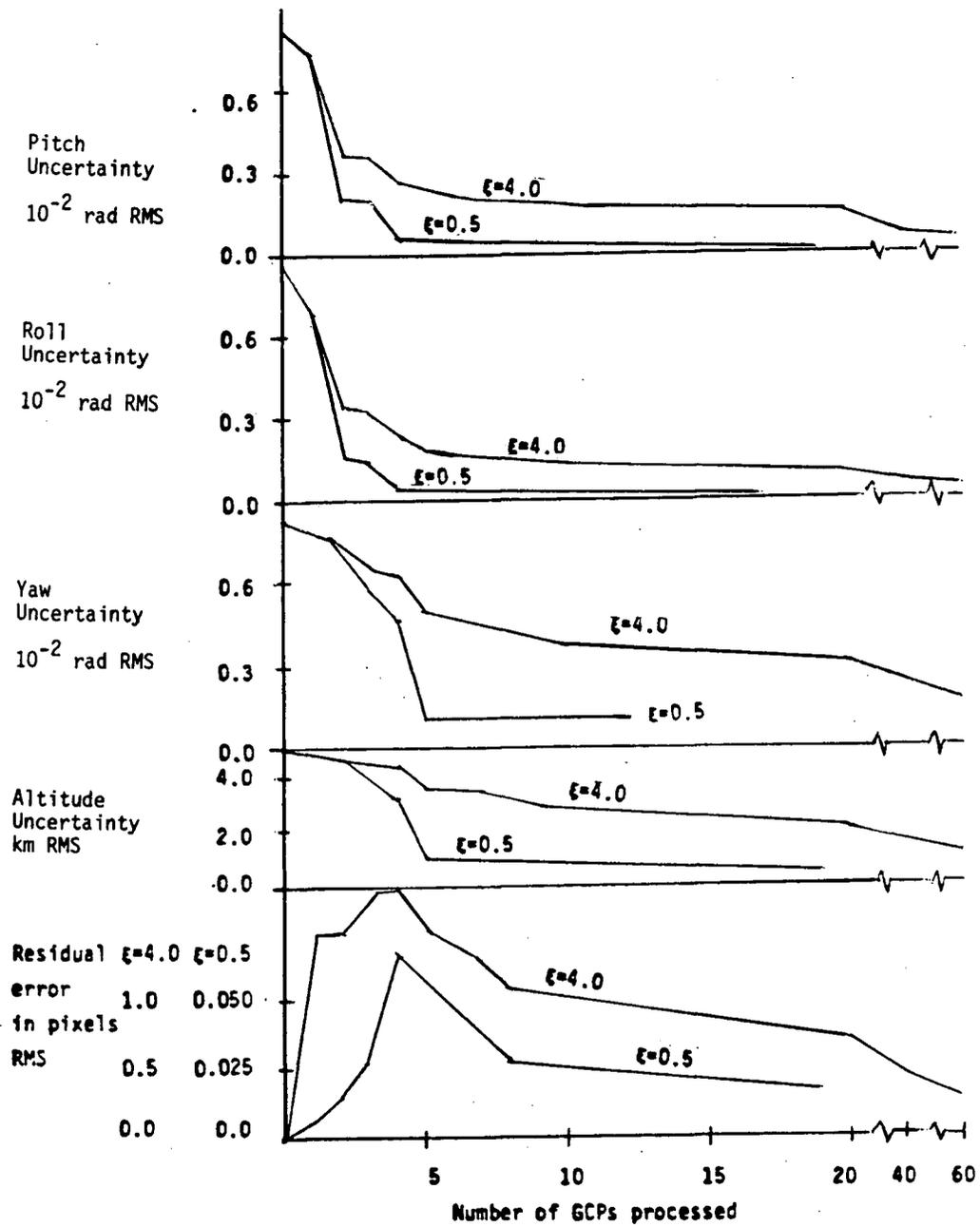


FIGURE 4.10 UNCERTAINTIES AND RESIDUAL ERROR VERSUS NUMBER OF GCPs PROCESSED, VERIFICATION TEST 1

simulated GCPs as shown in the figure. Also in both cases, it was found that the attitude time series and the semi-major axis after processing these GCPs agreed with the simulated values.

A further test was performed in the following manner. The geographic coordinates of the GCPs were corrupted by random noise equivalent to 3 pixels RMS. This simulates combined GCP designation error and measurement error. The results show that the attitude time series and semi-major axis obtained after processing all the GCPs agreed well with the simulated values and the residual error was about 3 pixels RMS. This further shows that although the uncertainty of the attitude angles can be reduced to any desired limit by processing a large enough number of GCPs (shown in Equations 4.6a, b and c), the residual error as defined in Equation 4.7 may not necessarily reduce to zero due to measurement error.

Verification Test 2

Verification test 1 was repeated but a bias of 0.1° was introduced into the equatorial crossing and a shift of 2 seconds was introduced into the elapsed time from equatorial crossing. These perturbations correspond to 10 pixels and 12 lines shift, or to an additional pitch of 0.015 rad and an additional roll of 0.0096 rad. Figure 4.11 shows the RMS residual error as a function of the number of GCPs processed. After processing the 19 GCPs the RMS residual error of about 1.6 pixels was obtained and the additional roll and pitch were correctly reflected in the output attitude time series.

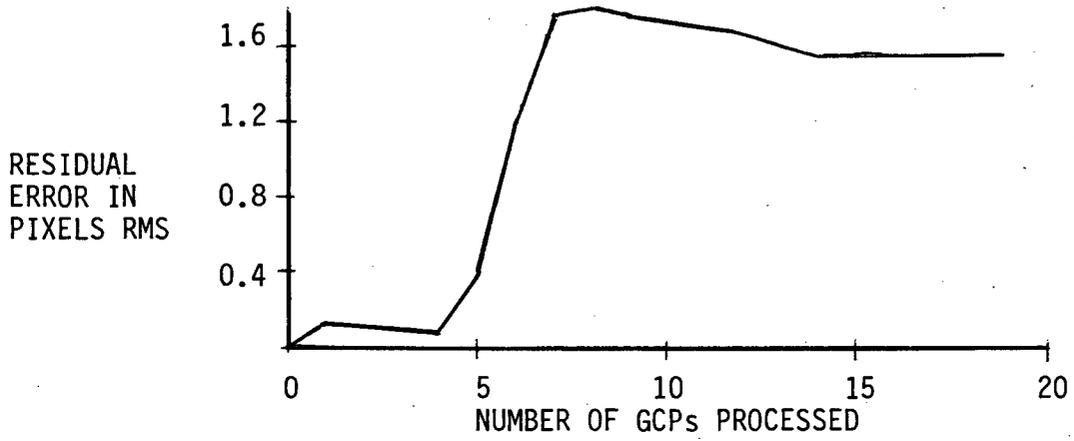


FIGURE 4.11 VERIFICATION TEST 2

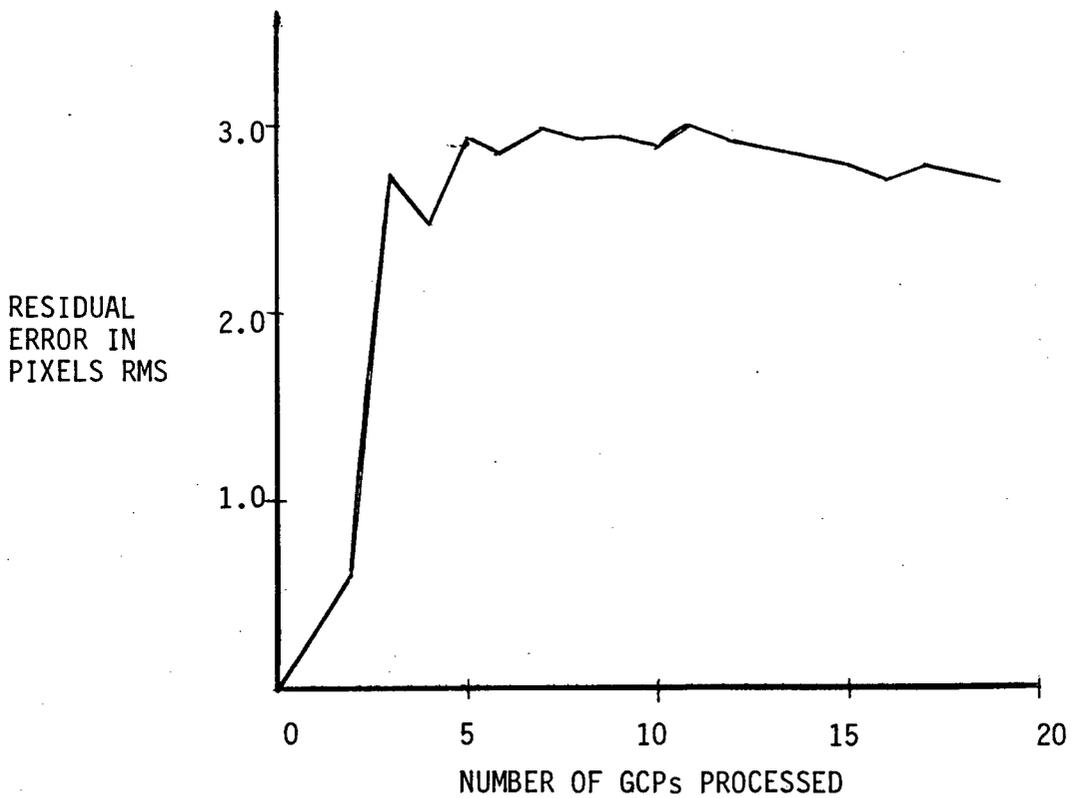


FIGURE 4.12 OPERATIONAL RUN

The equatorial crossing was then shifted by 0.2° and the elapsed time from equatorial crossing shifted by 4 seconds. The RMS residual error increased to 4 pixels.

These experiments show that if the measurement geometry, i.e., matrix H in the recursive filter, is in error it is not possible to absorb all the errors into the state vector elements. One solution is to place all the uncertain terms which will affect the transformation accuracy into the state vector.

Experiments with Landsat MSS imagery show that such perturbations can be absorbed into the pitch, roll and yaw components with very little effect on the transformation accuracy. This is due to the fact that a Landsat image covers a much smaller area than a TIROS-N image. An experiment was performed with a Baltimore/Washington scene (I.D. 20598-14590) in which the scene centre was shifted by 30 pixels and 22 lines and the satellite heading was assumed to be constant over the entire scene. The residual error thus obtained was less than 50 metres RMS with 15 GCPs. This is quite acceptable if the user is not interested in how the attitude angles vary. However, the yaw angle variation over the scene was found to be 0.25° . This value exceeds the specification and is due to the fact that the satellite heading was erroneously assumed to be a constant. The satellite heading was then allowed to vary as it should, and as a result the yaw variation dropped to an acceptable value of 0.017° . The residual error was also less than 50 metres RMS. This suggests that the attitude angles cited by Caron and Simon are the attitude angles plus perturbation effects due to other parameters, such as

satellite inclination, equatorial crossing, elapsed time and orbit semi-major axis.

Operational Run

In this experiment 19 GCPs were marked and distributed randomly on the image from 38°N to 48°N in latitude. This area corresponds to about the same 1,200 lines as in the first two experiments. Mapsheets of scale 1:2,500,000 were used to mark the geographic coordinates of the GCPs (better resolution maps were not available).

The GCPs were marked at river confluences, points on lakes and reservoirs, that is all on water/land boundaries. It should be noted that water levels, especially in the U.S. midwest, are susceptible to seasonal changes. More deterministic and time invariant GCPs such as highway intersections could not be identified due to the resolution of TIROS-N AVHRR imagery.

The value of $P(0)$ was initialized in the same way as in Verification Test 1. Results in Figure 4.12 show that the residual error is 2.8 after 19 GCPs. The value of ξ lies between 0.5 and 4 pixels and each value depended on the confidence in designating the GCP. Also in designating each GCP the image was zoomed by cubic convolution by 7 times. Furthermore in an additional test the image was extended to 2,000 lines and the RMS residual error remained at 2.8 pixels.

Error Analysis

The estimated error sources noted in Table 4.2 can be combined to fully account for the measurement error of 2.8 pixels RMS in the operational run. The table shows a combined error of 2.6 pixels RMS. In addition there could be two more sources of error. It is doubtful if the equatorial crossing of the satellite and the elapsed time were accurate to the limit as discussed in Section 4.4.2.3. It has been shown that errors in these two parameters cannot be absorbed into the current state vector. Also each attitude angle may be of a higher degree than a cubic polynomial as a function of time¹.

In future work, the state vector should contain more elements such as equatorial crossing, elapsed time and more coefficients for the attitude angles and altitude.

Although the residual error in the operational test is about 2.8 pixels RMS, it is felt that the technique with high quality ground control points and maps has the potential for measurement errors of about 1 to 2 pixels RMS.

4.5 Experiments — Interpolation in the Presence of Scan Gap

4.5.1 Objective

Appendix A shows that data samples for Landsat-4 TM imagery are not uniformly spaced in the across scan direction. If the sinc

¹At the time of writing we are still waiting for the results of the attitude measurements on board the satellite.

TABLE 4.2 ERROR ANALYSIS OF TIROS-N NAVIGATION

<u>Error Source</u>	<u>RMS Error</u>	<u>Comments</u>
Map accuracy	2.0 km	1:2,500,000 scale
GCP marking on map sheet	1.3 km	0.5 mm measurement error
GCP marking on image	0.5 km	0.5 pixel error in image
Altitude error	1.0 km	Due to constant height assumption over 1300 km

interpolation kernel is used for image interpolation, then the corresponding low pass filter will operate on a spectrum which does not accurately represent the image irradiance. Another interpolation kernel will have to be used. This kernel must be able to yield a high throughput in image interpolation.

One technique, based on cubic convolution and on creating uniformly spaced samples has been proposed by Prakash and Beyer [PRAK81] and simultaneously by Avery and Hsieh [AVER81].

In this method, as in the case of uniformly spaced samples, cubic convolution is used. Therefore the kernel covers four samples and is a weighted and truncated sinc function. When these four pixels span two swaths, at most two samples are not uniformly spaced. Two, at most, uniform samples are then created by using two pixels on each swath. A 4-point cubic spline is then used to performed the interpolation to give the values of the created samples. Then cubic convolution is applied to the uniform samples which include the created samples.

Their method suffers the following drawbacks:

- (a) A 4-point cubic convolution does not give an accurate interpolation compared to a 16-point truncated sinc function with Kaiser windowing (as demonstrated in Section 2.5.1.6).
- (b) The 4-point cubic spline makes no attempt to preserve the signal spectrum.

This section presents a new method of interpolating Landsat-4 TM imagery in the presence of scan gaps. Experiments are performed on sine waves with various frequencies and scan gap sizes, on synthesized sine wave imagery and on synthesized TM imagery. Experimental results have verified that this method yields an RMS interpolation error of one level out of 256 levels in 8-bit imagery data for the case of overlap.

Failure to use a method such as the one presented here can lead to significant radiometric errors, thus greatly reducing the utility of geometrically corrected imagery data.

4.5.2 Theory of Proposed Method

Let a band limited signal $f(t)$ be sampled at the Nyquist rate and divided into two uniformly sampled groups. Specifically, let the sample positions be denoted by:

$$\tau_m = \begin{cases} -pT & , t \leq 0 \\ \Delta t + pT & , t > 0 \end{cases} \quad (4.8)$$

where T is the sample spacing and p is a non-negative integer. For convenience, separation of the two groups by an amount Δt is assumed to occur at $t = 0$. Following Linfoot and Shepherd's work [LINF39], Yen [YEN56] has shown that the signal $f(t)$ can be accurately constructed by:

$$f(t) = \sum_{m=-\infty}^{\infty} f(\tau_m) \psi_m(t) \quad (4.9)$$

where

$$\psi_m(t) = \frac{(-1)^m \Gamma(\Delta t/T + p)}{p! \Gamma(t/T) \Gamma[(\Delta t - t)/T]} \begin{cases} \frac{1}{p + t/T} & , \tau_m = \frac{p}{T} \\ \frac{1}{p + (\Delta t - t)/T} & , \tau_m = \Delta t + \frac{p}{T} \end{cases} \quad (4.10)$$

Therefore the kernel $\psi_m(t)$ can be expressed as Gamma functions [KORN61] of Δt , scan gap where Δt occurs, sample number m and sampling point. The Gamma function has the recursive property:

$$\Gamma(n+1) = n \Gamma(n) \quad (4.11)$$

For $\Delta t = 0$, i.e., the sample points are uniformly spaced, $\psi_m(t)$ degenerates to the sinc function as expected.

This method is applicable here for the following reason. Conforming to the size of the 16-point sinc function in the case of uniformly spaced samples, the kernel size in this case also spans 16 points. Since each swath contains 16 samples, the kernel covers at most one scan gap. The sample further away from the interpolation position usually carries less weight. Also, the contribution by samples not under the kernel is insignificant. Therefore the sequence of sample points under the kernel can be viewed to consist of two groups separated by a gap. Each group can be assumed to extend infinitely since points after the 16th sample from the interpolation location are not under the kernel. The two infinite groups assumption then satisfies the condition cited in Yen's method.

Figure 4.13 shows plots of the kernel for various values of sample

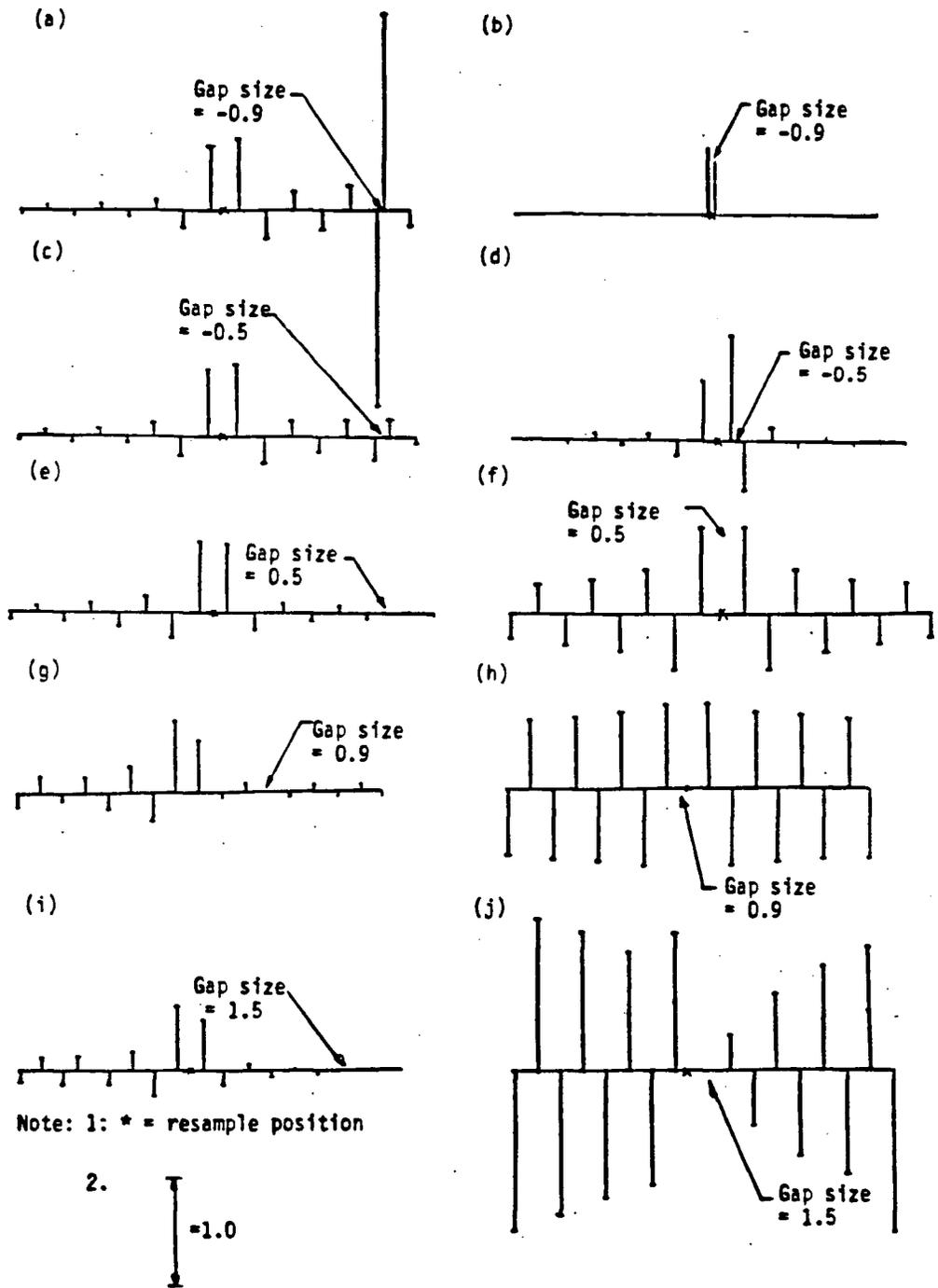


FIGURE 4.13 INTERPOLATION KERNEL OF YEN'S METHOD

number and scan gap size. The following observations are made:

- (a) When the scan gap is negative, the kernel weight decreases with distance of sample point away from the interpolation position except at the scan gap (e.g., Figure 4.13a).
- (b) When the scan gap is negative and interpolation is performed in the gap, the two nearest sample points weigh more and all the other sample points weigh less than the sinc function kernel, as expected.
- (c) When the scan gap is larger than 1 sample and interpolation is performed in the gap, the kernel weights do not decrease as a function of sample distance within the 16 samples.

The figure shows that two modifications have to be made:

- (a) When the gap is positive, extra samples have to be created so that the adjusted scan gap is reduced to a size when the kernel weights decrease appreciably as a function of sample distance. The adjusted scan gap is then always less than a predetermined size (see Section 4.5.4 for experimental results).
- (b) When the scan gap is negative and the gap is close to either end of the 16 sample points, the sinc interpolation is used. This is because, in Yen's method, the kernel weights are heavy at both ends of the gap and the kernel weights are still significant for a few more samples beyond the gap. Hence, if the kernel is truncated to 16 points, significant kernel weights will be eliminated. This has been proved in the experiment

(Section 4.5.4) to yield a larger interpolation error than the sinc function interpolation. Section 4.5.4 shows the experimental results which determine the scan gap position where switching to the sinc function interpolation should occur.

Extra samples can be created by a local cubic fit using the four nearest points, i.e., two points on each side of the scan gap. Let these data points be F_1, F_2, F_3 and F_4 at positions t_1, t_2, t_3 and t_4 respectively. Then interpolation can be done by the divided differences method:

$$f(t) = \sum_{i=1}^4 f[t_1, \dots, t_i] \prod_{j=1}^{i-1} (t - t_j) \quad (4.12)$$

where $F[]$ is the divided difference which can be defined recursively by:

$$F[t_i] = F_i$$

$$F[t_i, \dots, t_{i+k}] = \frac{F[t_{i+1}, \dots, t_{i+k}] - F[t_i, \dots, t_{i+k-1}]}{t_{i+k} - t_i}$$

The cubic polynomial fit can equivalently be represented by a Lagrange polynomial:

$$f(t) = \sum_{i=1}^4 F_i \phi_i(t) \quad (4.13)$$

$$\text{where } \phi_i(t) = \prod_{j=1, j \neq i}^4 \frac{t - t_j}{t_i - t_j} \quad (4.14)$$

The additional interpolation error due to creating new samples is determined as follows. Let X_i' , $i = 1$ to n , be the values of the n new samples and X_i be the actual values of the function to be interpolated at the same locations. Also let W_i , $i=1$ to n , be the kernel weights at the n positions. Then additional interpolation error is:

$$\sum_{i=1}^n w_i (X_i - X_i')$$

This error is larger when interpolating in the scan gap than away from the scan gap because the kernel weights are heavier when the interpolation position is closer to the gap.

Experimental results (Section 4.5.4) also show that a Kaiser window is necessary for a better performance of the method.

4.5.3 Implementation Considerations

The two groups of uniformly spaced samples are symmetrical about the midpoint of the scan. The symmetrical property reduces the computation of the kernel weights by a factor of two.

It is not necessary to compute the kernel weight for every sample each time the interpolation is called. The kernel weights can be stored in a table the size of which is determined as follows. The kernel is a function of sample number, t and scan gap size. If:

(a) sample spacing is divided into 32 subintervals and the kernel

size is 16 samples,

- (b) a new sample is created when the gap size is larger than 0.3 line (discussed in Section 4.5.4),
- (c) a sample is deleted when the gap size is less than or equal to -1 line,
- (d) uniform sample spacing is assumed when the gap size is between -0.1 and 0.1 line (discussed in Section 4.5.4),
- (e) no interpolation is necessary if the interpolation position is within 1/64 of any sample point, and
- (f) scan gap size is varied in steps of 0.05 line (see discussion below),

then the number of coefficients, taking advantage of the symmetrical property, is

$$\underbrace{(1/0.5) \times 10}_{(I)} \times \underbrace{31}_{(II)} \times \underbrace{16}_{(III)} \times \underbrace{8}_{(IV)} = 79360$$

where the term:

- (I) = number of quantized gap sizes,
- (II) = number of subpixel intervals minus 1,
- (III) = number of data points, and
- (IV) = number of scan gap size positions (taking advantage of the symmetrical property).

The scan gap quantization step of 0.05 line is obtained by the following experiment. Scan gaps of ± 0.5 line were chosen for a sine wave with frequency of 6 pixels per cycle and interpolation was done

with erroneous scan gap sizes. The results, presented in Table 4.3, show that the RMS error in interpolation due to a 0.05 line error in scan gap size is still acceptable (about 1 level for underlap and 0.7 for overlap). Note that in Canada overlap occurs more often than underlap and the scan gap size is expected to be less than 0.3 line.

The above calculation assumes that the new samples are actually created when required. By examining Equation 4.13, it is obvious that the new samples need not be physically created, but the kernel weights in the two samples on each side of the scan gap need to be adjusted by ϕ_1 as shown in the equation. Therefore, the scan gap size for 12 of the 16 kernel weights varies from -0.9 to 0.3 line, and it varies from -2 to 2 lines for the remaining four kernel weights. Therefore, the number of coefficients, if the new samples are not physically created, is about:

$$(1/0.5) \times 31 \times 8 \times (10 \times 16 + 40 \times 4) = 158720$$

The required number of coefficients cited above will be reduced by 20% if switching to sinc kernel is adopted under certain conditions as discussed in Section 4.5.2.

4.5.4 Experiments with Sine Waves

Sine waves are first tested for the following reasons:

- (a) The frequency and scan gap can be controlled easily and hence the interpolation can be evaluated as a function of frequency

TABLE 4.3 RMS ERROR VERSUS SCAN GAP
SIZE QUANTIZATION

ACTUAL SCAN GAP SIZE	ERRONEOUS SCAN GAP SIZE	SCAN GAP SIZE ERROR	RMS ERROR
-0.5 (overlap)	-0.50	0.00	0.67
	-0.51	0.01	0.67
	-0.52	0.02	0.68
	-0.53	0.03	0.69
	-0.54	0.04	0.69
	-0.55	0.05	0.70
0.5 (underlap)	-0.50	0.00	0.60
	-0.51	0.01	0.63
	-0.52	0.02	0.71
	-0.53	0.03	0.83
	-0.54	0.04	0.99
	-0.55	0.05	1.17

Note: Test was performed on sine wave with a frequency of 6 pixels per cycle.

and scan gap size.

- (b) Using the sine waves, it can be shown how other methods (e.g., interpolation by sinc function kernel¹) compare to the chosen method.
- (c) In this part of the experiment, the following parameters in the chosen method can be determined:
- the scan gap position for switching to the sinc kernel interpolation, and
 - the scan gap size above which new samples are created.

Each sine wave in this part of the experiment consists of 32 data points. The first 16 and the last 16 sample points are equally spaced (both groups have the same spacing Δ_0). The spacing between the 16th and 17th data point (Δ_1) reflects the magnitude of the scan gap and scan gap size is defined as $\Delta_1 - \Delta_0$. In the experiment, the scan gap size varies from -0.9 to 2.0 samples. Scan gap size less than or equal to -1.0 sample represents redundant data and will not affect the experimental result.

The TM modulation transfer function (MTF) has a significant frequency response up to 0.16 cycle per pixel (the maximum is 0.5 cycle per pixel). The experiment hence places a strong emphasis around this particular frequency.

Performance of the sinc interpolation and Yen's method are shown in Figure 4.14. Frequency of the sine wave is 0.16 cycle per pixel. The

¹ Here the sinc function is referred to as the 16-point truncated sinc function with Kaiser windowing.

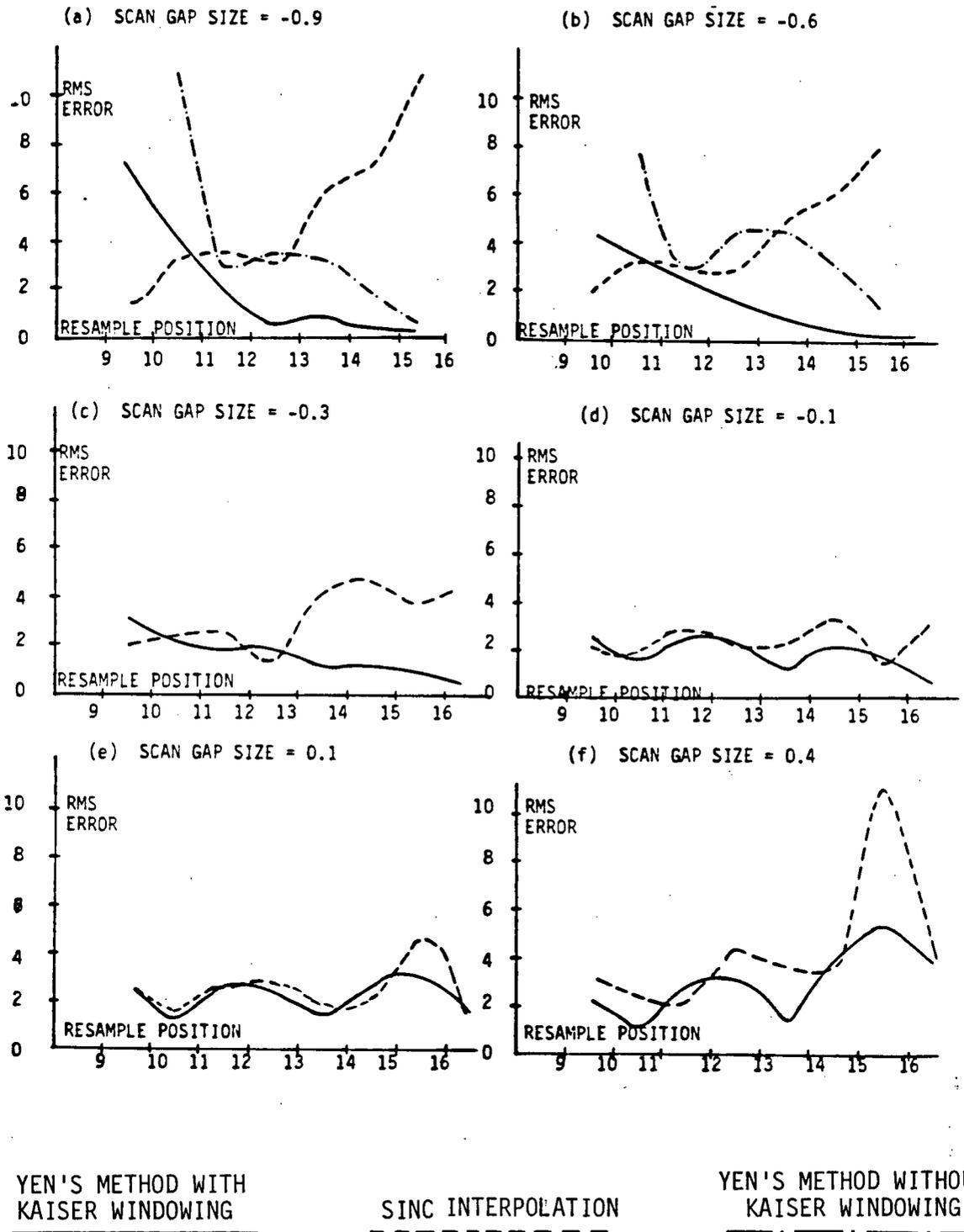


FIGURE 4.14 RMS ERROR VERSUS INTERPOLATION POSITION

abscissa of each plot is the interpolation position from 8 to 16 + half of scan gap. The kernel weights of other interpolation positions are replicas of the weights in this segment. Each plot is the envelop¹ of the RMS error.

The following conclusions can be drawn from the figure:

- (a) Yen's method without Kaiser windowing gives a higher RMS error than Yen's method with Kaiser windowing. Therefore, the former method is discarded. This is an expected result since this is also true for the sinc function interpolation and the overall shape of the Yen's method kernel is similar to that of the sinc function.
- (b) The sinc function interpolation and Yen's method give approximately the same RMS error when the scan gap size is between ± 0.1 line. Hence, for this range of gap size, the sample spacings can be assumed to be uniform and the sinc interpolation is used.
- (c) When the interpolation position is such that the scan gap is close to the end of the kernel and its size is negative, Yen's method gives a very large error because the kernel weights after the gap are still significant. In this case the sinc interpolation performs better. The point of switching is when the scan gap is at the following (marked with x) positions in the kernel:



¹ The RMS error is zero when the interpolation position falls on a data point. The envelop is the locus following the maximum error between pairs of data points.

It remains to determine when new samples have to be created. Experimental results show that one sample should be created when the scan gap is between 0.3 and 1.3 samples, and two samples should be created if the scan gap is larger than 1.3 samples. It is not necessary to create more than two samples because the scan gap is always less than 2.0 samples.

Figure 4.15 gives comparisons between the sinc interpolation, Yen's method and the proposed method (i.e., Yen + Kaiser windowing + creating of new samples) for four different frequencies. The figure shows that the chosen method is superior to the other two methods. Sharp turns exist at the gap size equal to 0.3 and 1.3 samples for the chosen method before new samples are added. Without adding the new samples the RMS error increases rapidly due to the fact that the kernel requires more than 16 points, i.e., the kernel weights have not subsided yet at the end of the 16-point kernel.

4.5.5 Scan Gap Simulation in Imagery

The scan gaps were assumed to be due to the following factors: satellite altitude variation, jitter and scan skew. This section describes how these scan gap constituents were simulated. The simulated scan gap is used in the synthesis of distorted imagery (Section 4.5.6).

1. Altitude Variation

This causes a constant scan gap size, in the absence of jitter and scan skew, given by:

$$G = \Delta_1 - \Delta_0$$

where Δ_0 denotes the interline spacing within a swath and Δ_1 denotes the spacing between the two swaths. The value of G in the experiment varied between ± 2.0 lines. A slanted road due to a positive and negative value of G is shown in Figure 4.16. A negative G means the satellite is at a higher orbit than nominally, and hence the sensor images a wider portion of the road with its 16 detectors. As a result, the road is segmented as shown in Figure 4.16a. This corresponds to overlap. On the contrary, a positive value of G shows underlap and the slant road becomes segmented in the way shown in Figure 4.16b.

2. Jitter

Jitter has frequencies in the range from 2 to 125 Hertz, it can be modelled as a series of sine waves at the output of a typical band pass filter centred at 60 Hertz. The 3 db points at 30 and 90 Hertz are shown in Figure 4.17. Hence, the scan gap is:

$$d(t) = c \sum_{i=1}^N \{ \exp[(-60 + 1200i/N)^2/1300] \} \sin[2\pi(120it/N)]$$

where t = time,

N = total number of sine waves, and

c = multiplicative constant to adjust the gap size.

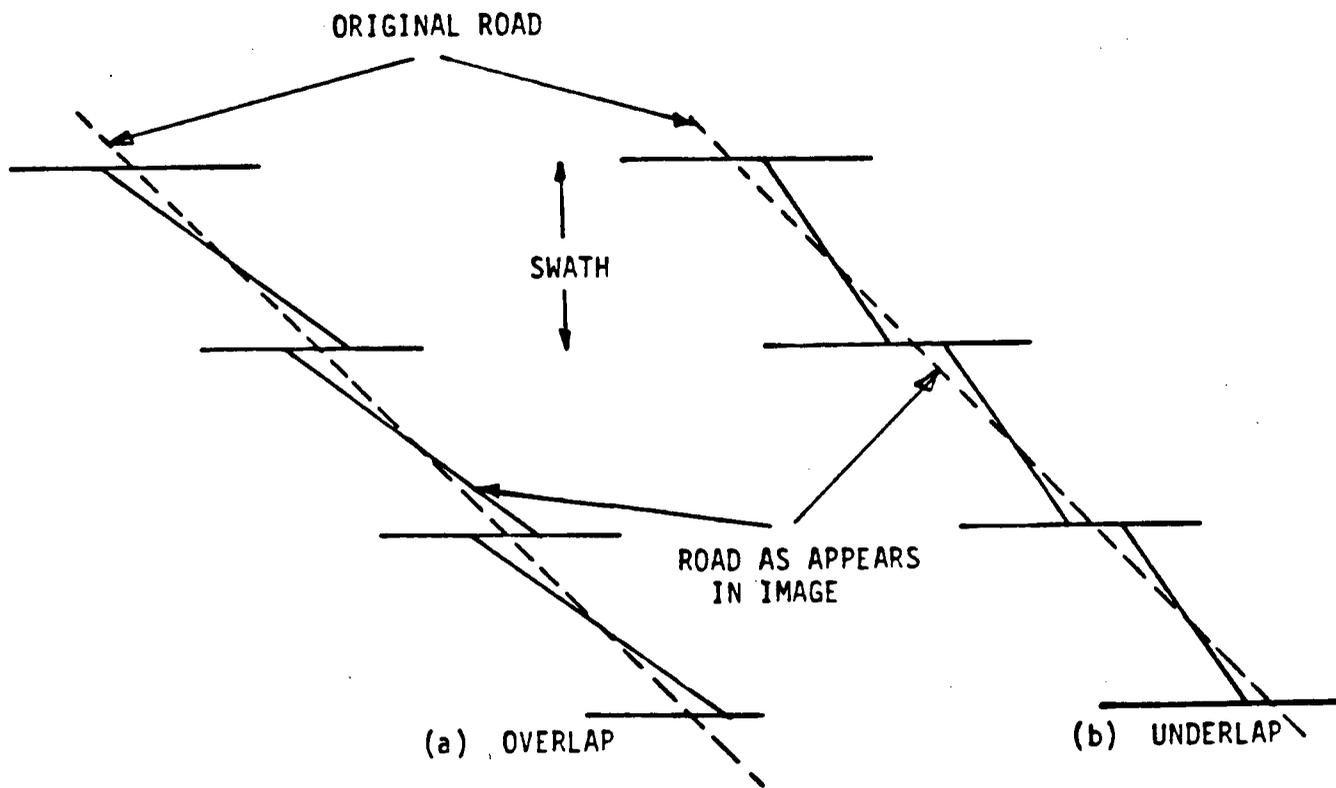


FIGURE 4.16 EFFECT OF OVERLAP AND UNDERLAP ON A ROADWAY

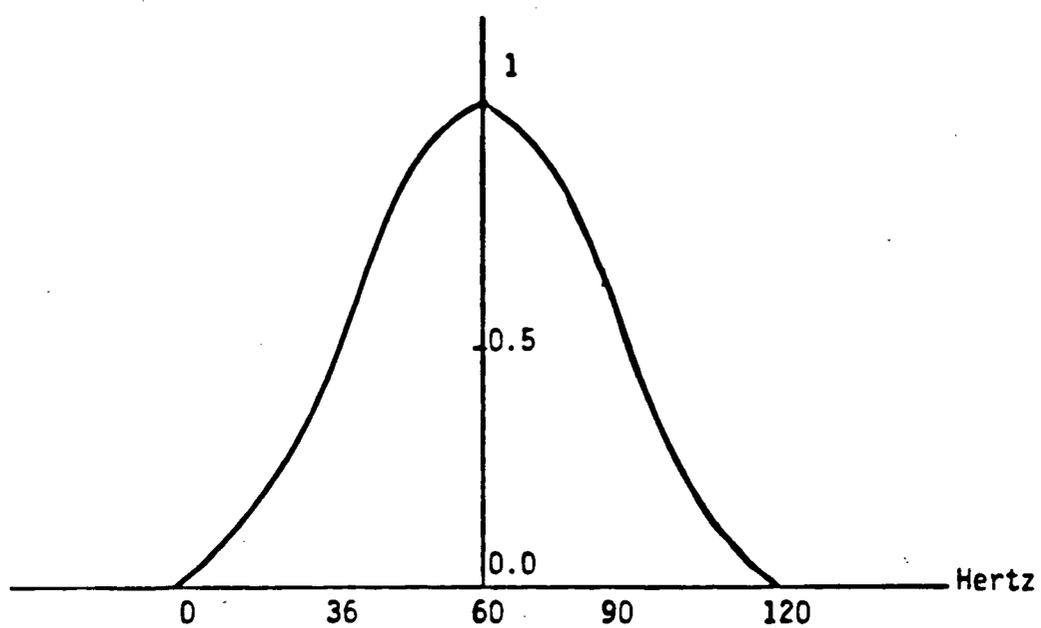


FIGURE 4.17 JITTER MODEL

In the experiment, the value of c was chosen such that the jitter contribution to the scan gap varied from -0.5 to 0.5 line.

3. Scan Skew

The scan skew was chosen so that its contribution to the scan gap was between ± 0.1 line.

4.5.6 Experiments with Sine Wave Imagery

Sine wave patterns were used in the simulation of imagery because they can synthesize linear features such as roadways and hence the effect of scan gap on these features can be visualized.

Figure 4.18 depicts linear features on the ground running at an angle of 45° to the vertical. The sine wave has a frequency of 6 pixels per cycle. This frequency was chosen because the MTF of the TM sensor dictates that it takes about 3 pixels to respond from one intensity extreme to the other. The simulated TM image is shown in Figure 4.19, where the scan gap ranges from -1 to 1 line. Hence, this corresponds to overlap and the sine patterns are segmented in a manner similar to Figure 4.16a. Two methods of reconstruction were used — sinc interpolation and the proposed method. The results are shown in Figures 4.20 and 4.21. The difference images, each multiplied by a factor of 4, are shown in Figures 4.22 and 4.23.

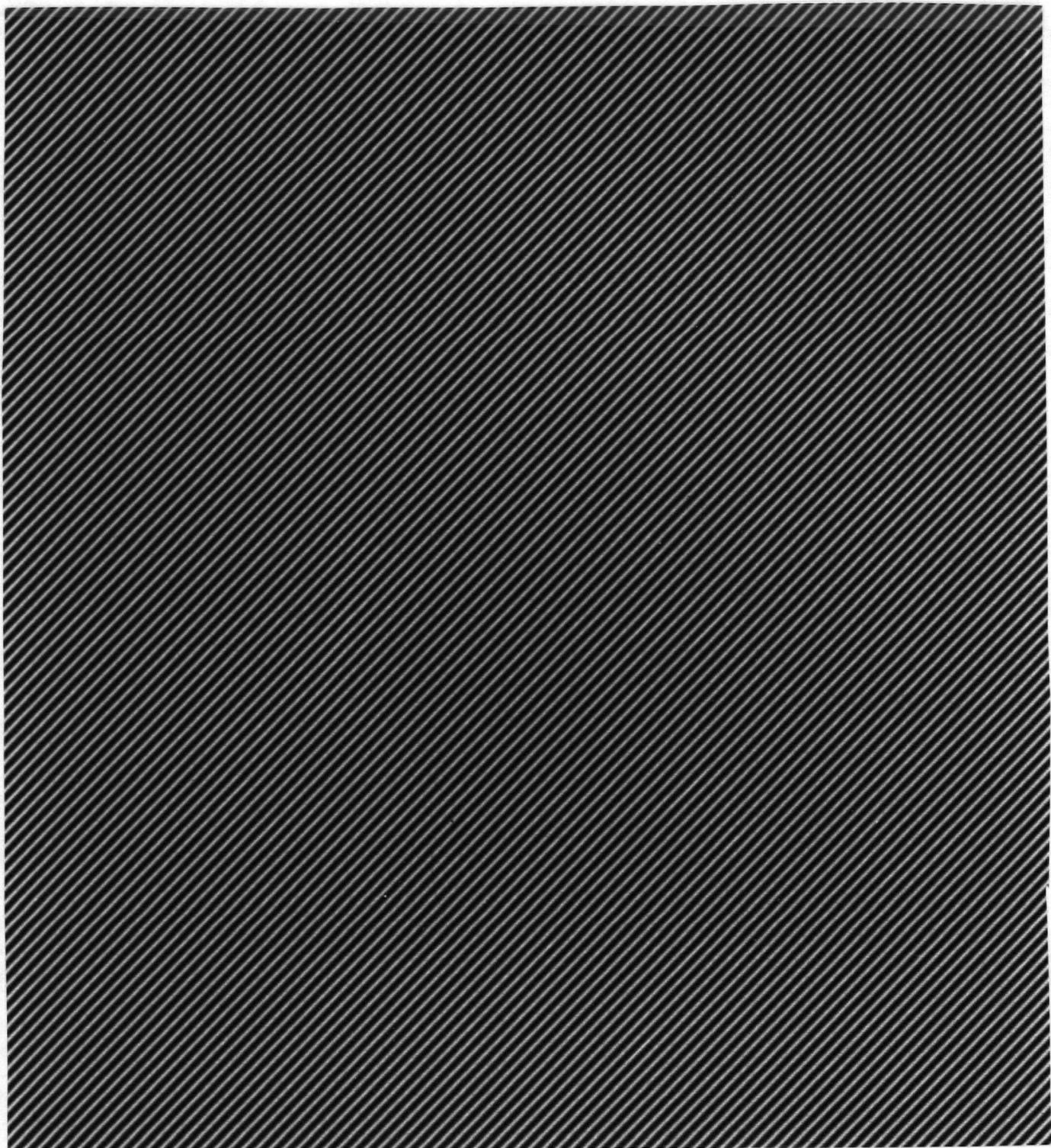


FIGURE 4.18 SINE WAVE IMAGE

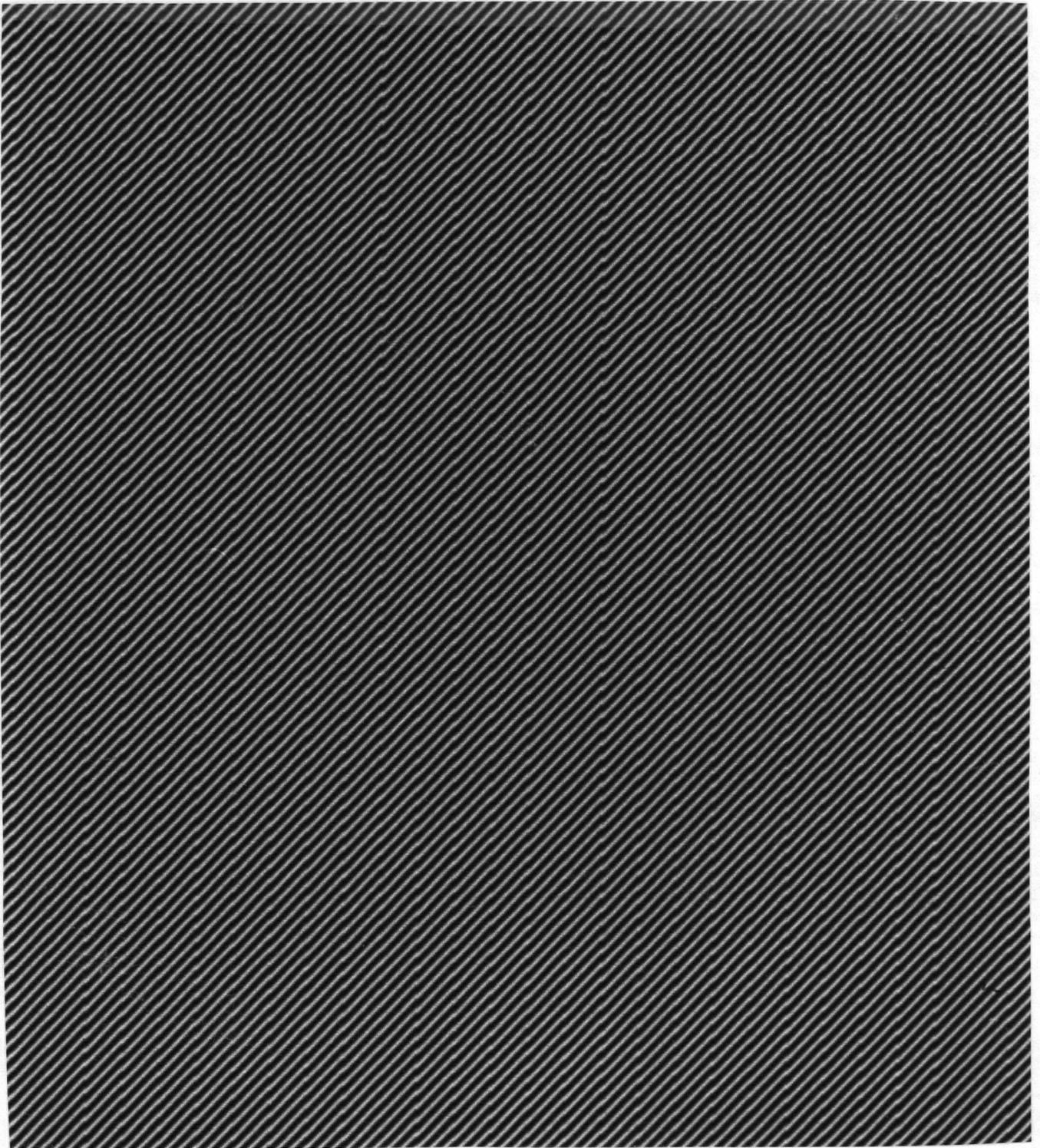


FIGURE 4.19 DISTORTED SINE WAVE IMAGE WITH OVERLAP

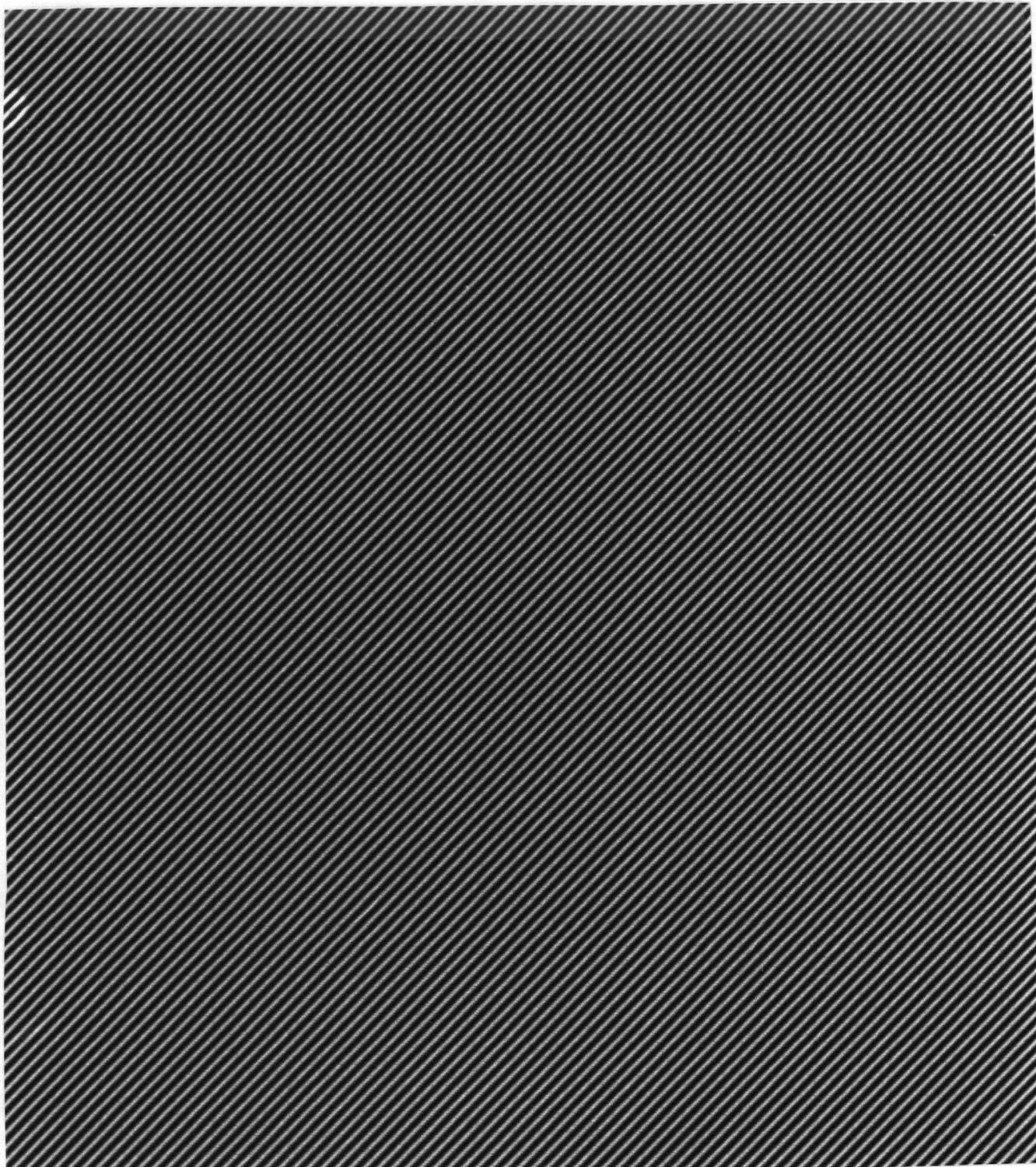


FIGURE 4.20 RECTIFIED SINE WAVE IMAGE (OVERLAP) WITH SINC INTERPOLATION

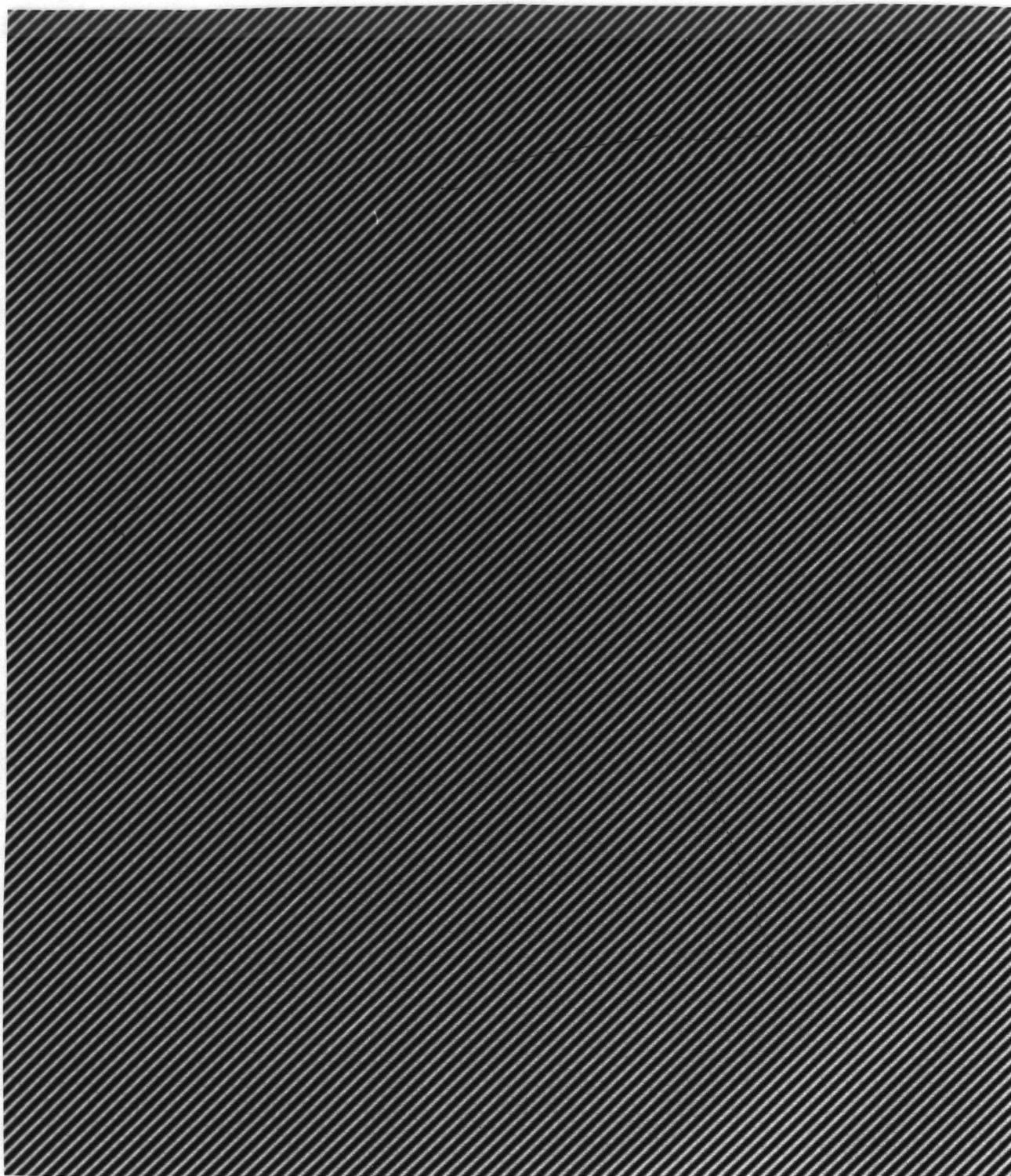


FIGURE 4.21 RECTIFIED SINE WAVE IMAGE (OVERLAP) WITH PROPOSED METHOD



FIGURE 4.22 RADIOMETRIC ERROR IN RECTIFIED SINE WAVE IMAGE
(OVERLAP) WITH SINC INTERPOLATION

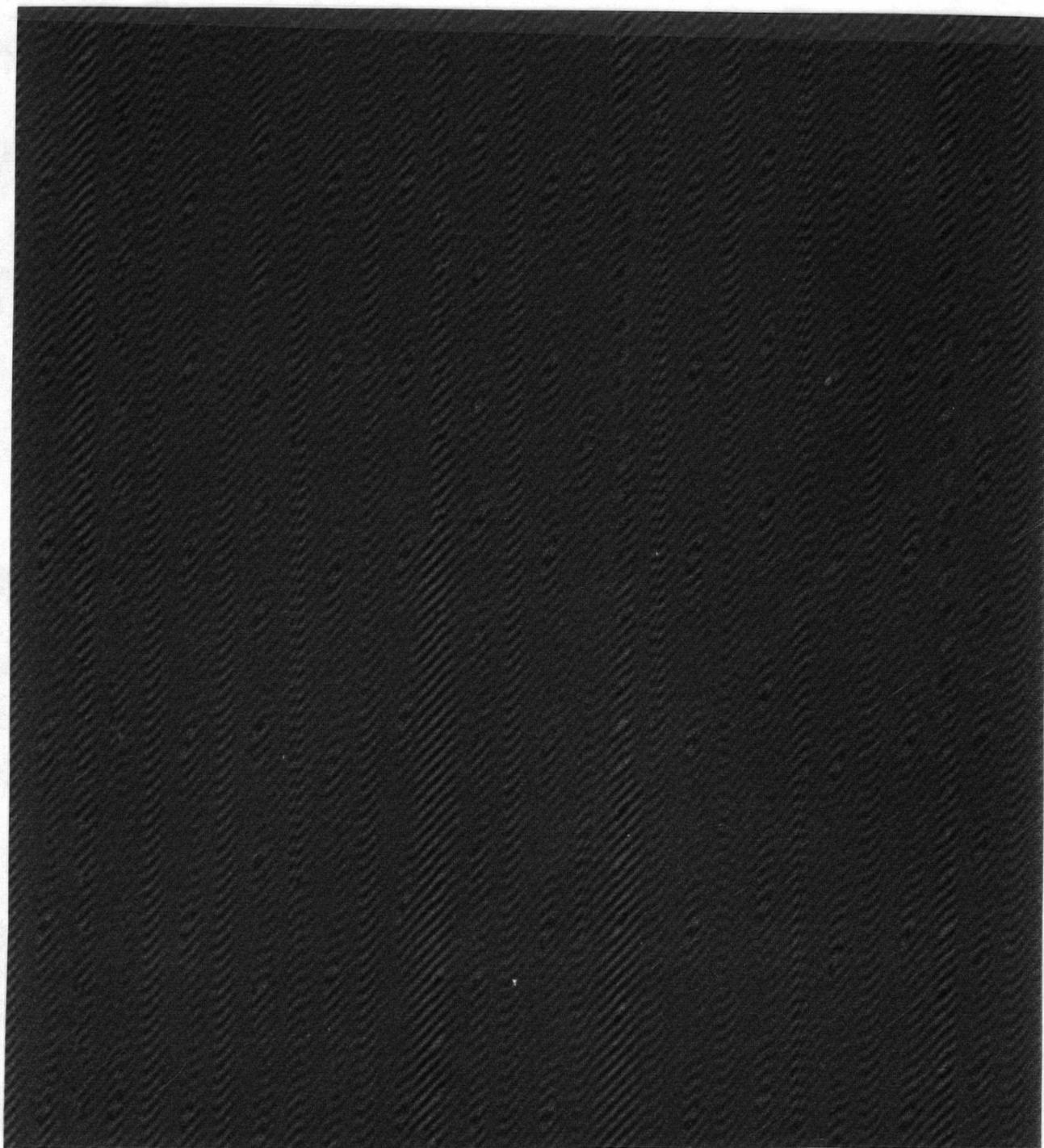


FIGURE 4.23 RADIOMETRIC ERROR IN RECTIFIED SINE WAVE IMAGE
(OVERLAP) WITH PROPOSED METHOD

The same image, but with underlap, was also simulated with scan gap sizes from 0.3 to 1.4 lines. The result is shown in Figure 4.24 in which the sine patterns are segmented in a manner similar to Figure 4.25 to 4.28. Figure 4.25 (sinc interpolation) shows more obvious seams than Figure 4.26 (proposed method).

The difference images clearly show that the proposed method is better than the sinc interpolation for both the overlap and underlap cases. The quantitative results are given by the RMS value in the difference images:

	Overlap	Underlap
Sinc interpolation	2.8	6.9
Proposed method	1.1	2.6

Thus, the proposed method is about 2.5 times better than the 16-point sinc interpolation in terms of RMS error.

A plot of absolute error for a sequence of data points in the difference images is given in Figure 4.29. This figure clearly demonstrates the superiority of the proposed method. The method performs about two to three times better than the sinc interpolation. However, by just considering the neighbourhood of the scan gaps, the proposed method is even better. The sinc interpolation will introduce considerable artifacts even in the case of overlap.

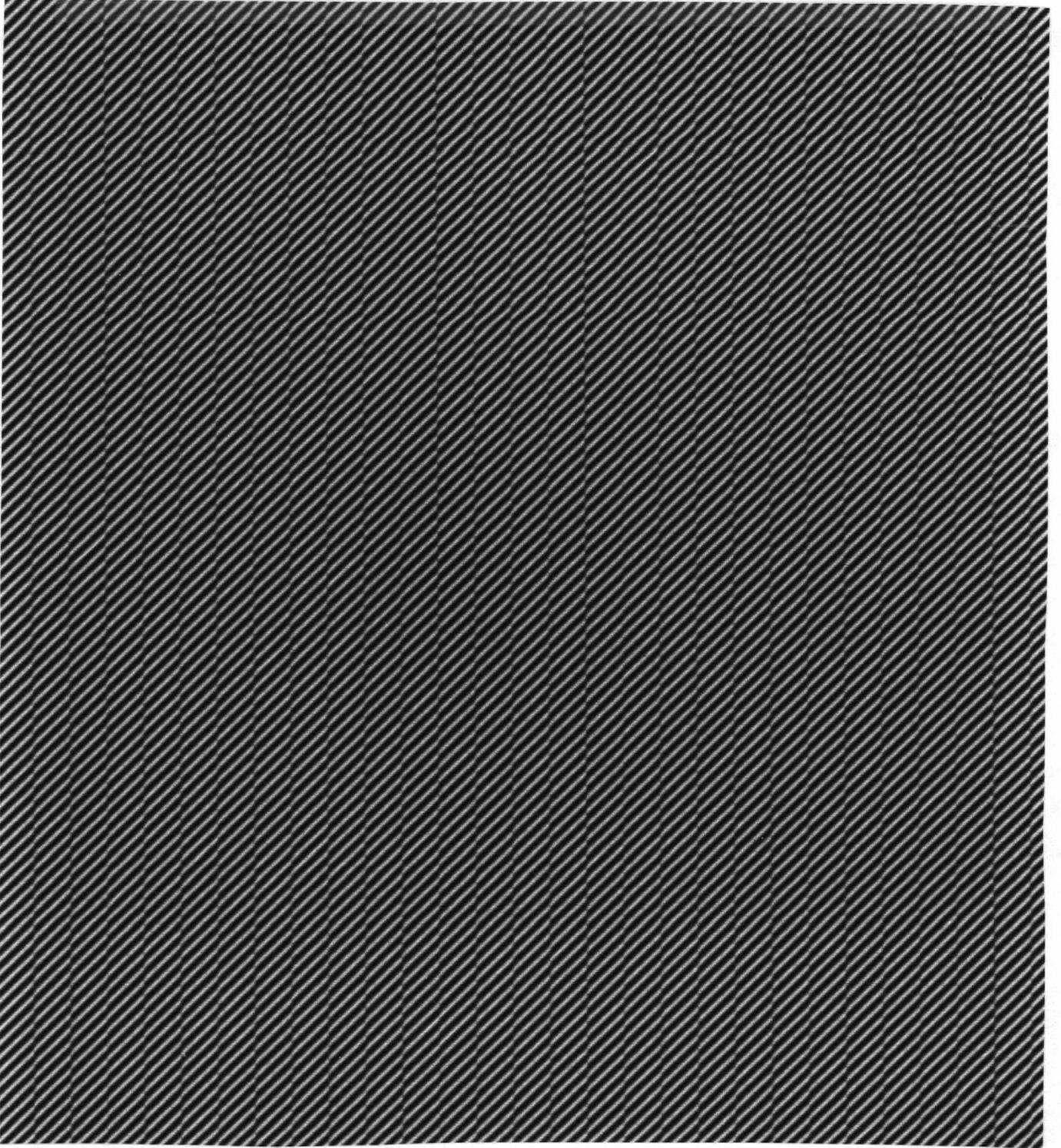


FIGURE 4.24 DISTORTED SINE WAVE IMAGE WITH UNDERLAP

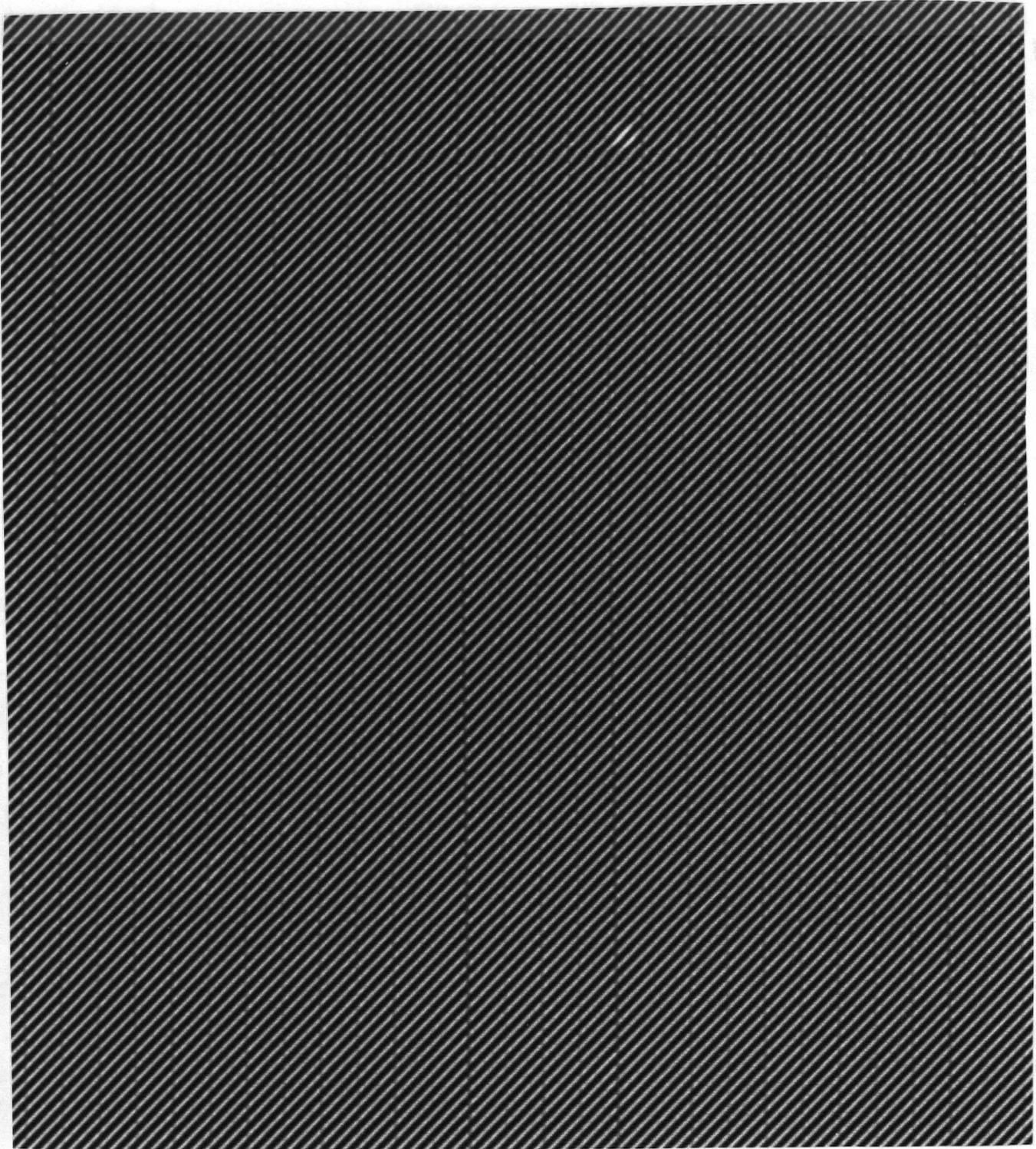


FIGURE 4.25 RECTIFIED SINE WAVE IMAGE (UNDERLAP) WITH SINC INTERPOLATION

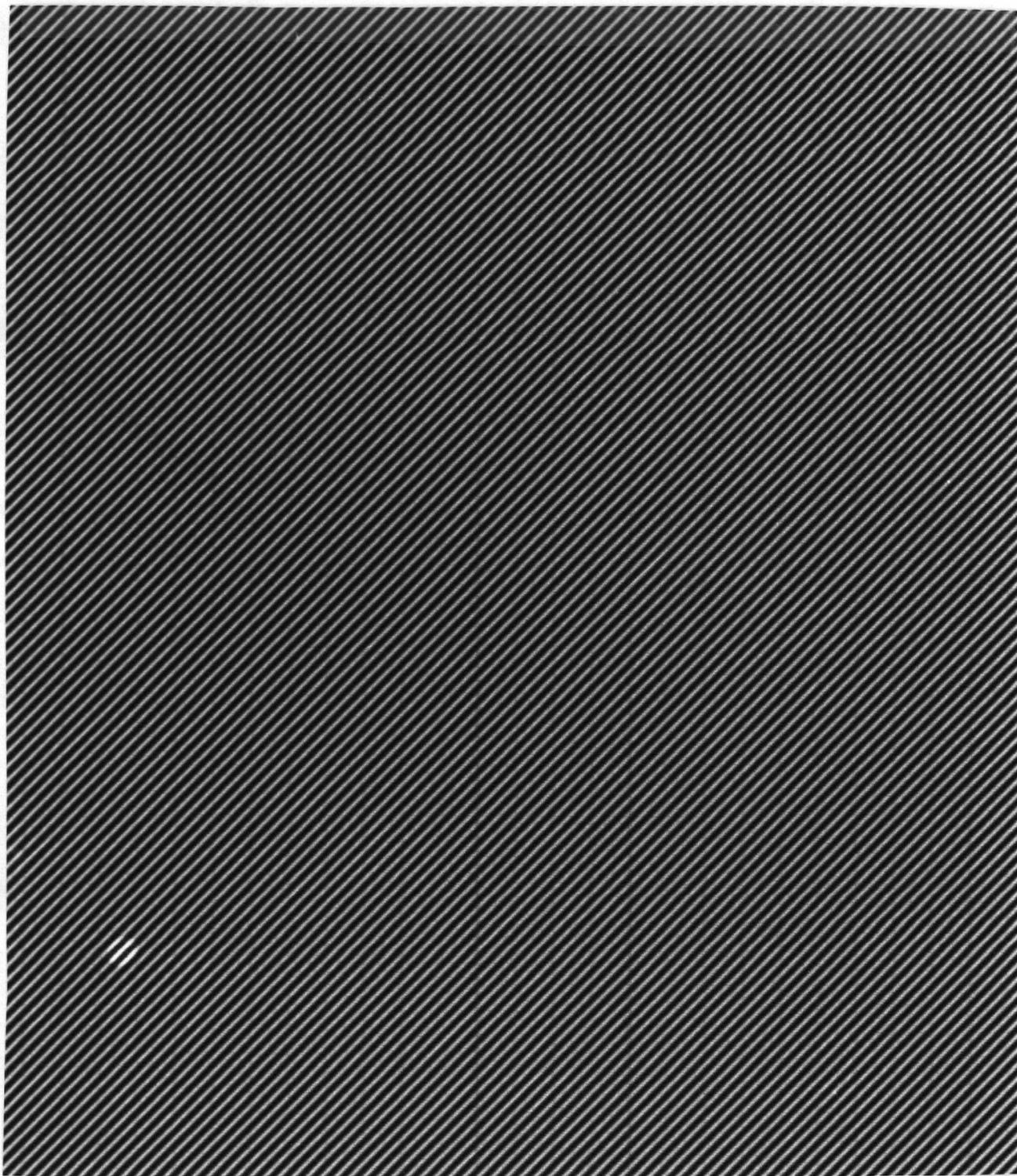


FIGURE 4.26 RECTIFIED SINE WAVE IMAGE (UNDERLAP) WITH PROPOSED METHOD

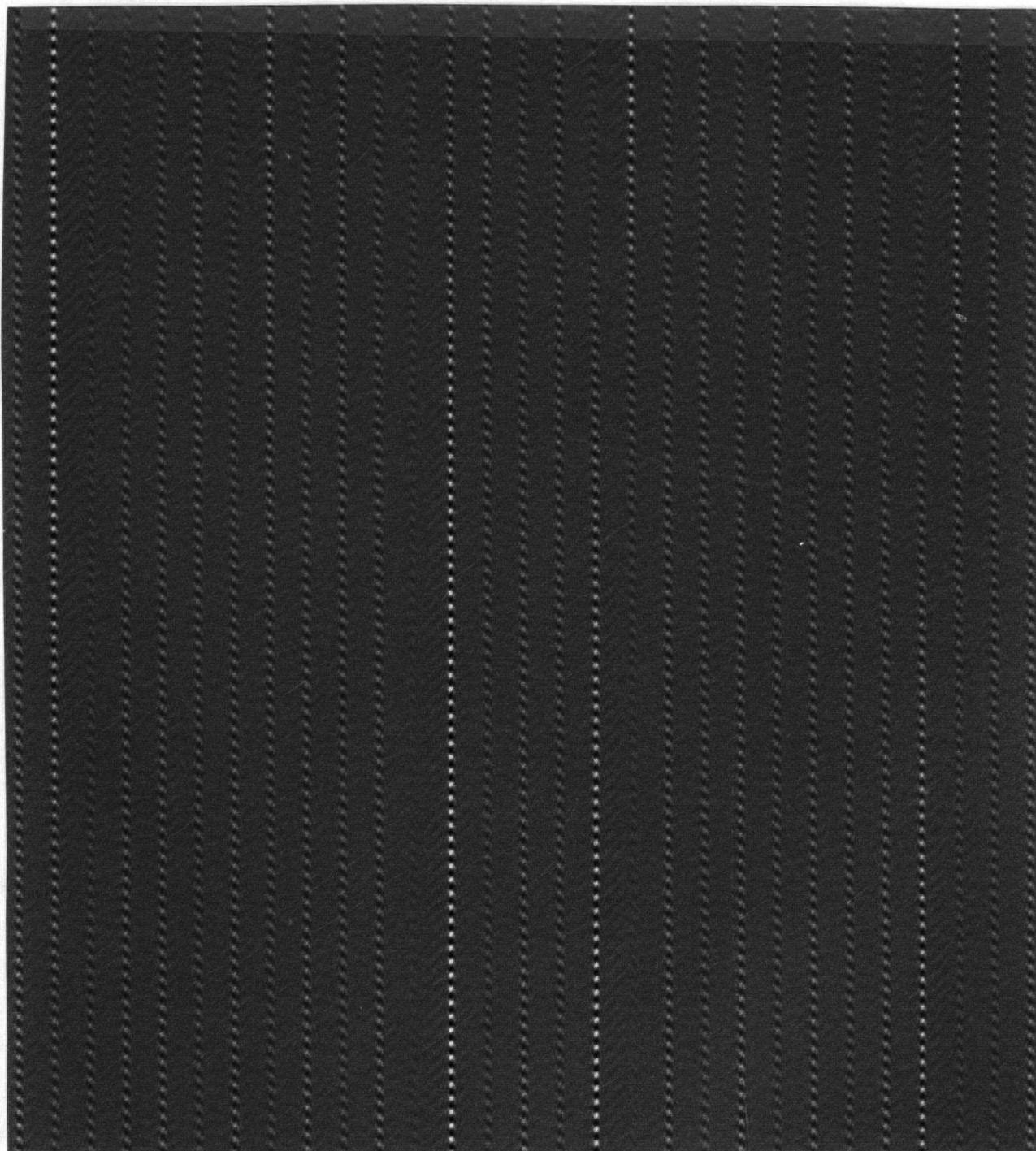


FIGURE 4.27 RADIOMETRIC ERROR IN RECTIFIED SINE WAVE IMAGE
(UNDERLAP) WITH SINC INTERPOLATION

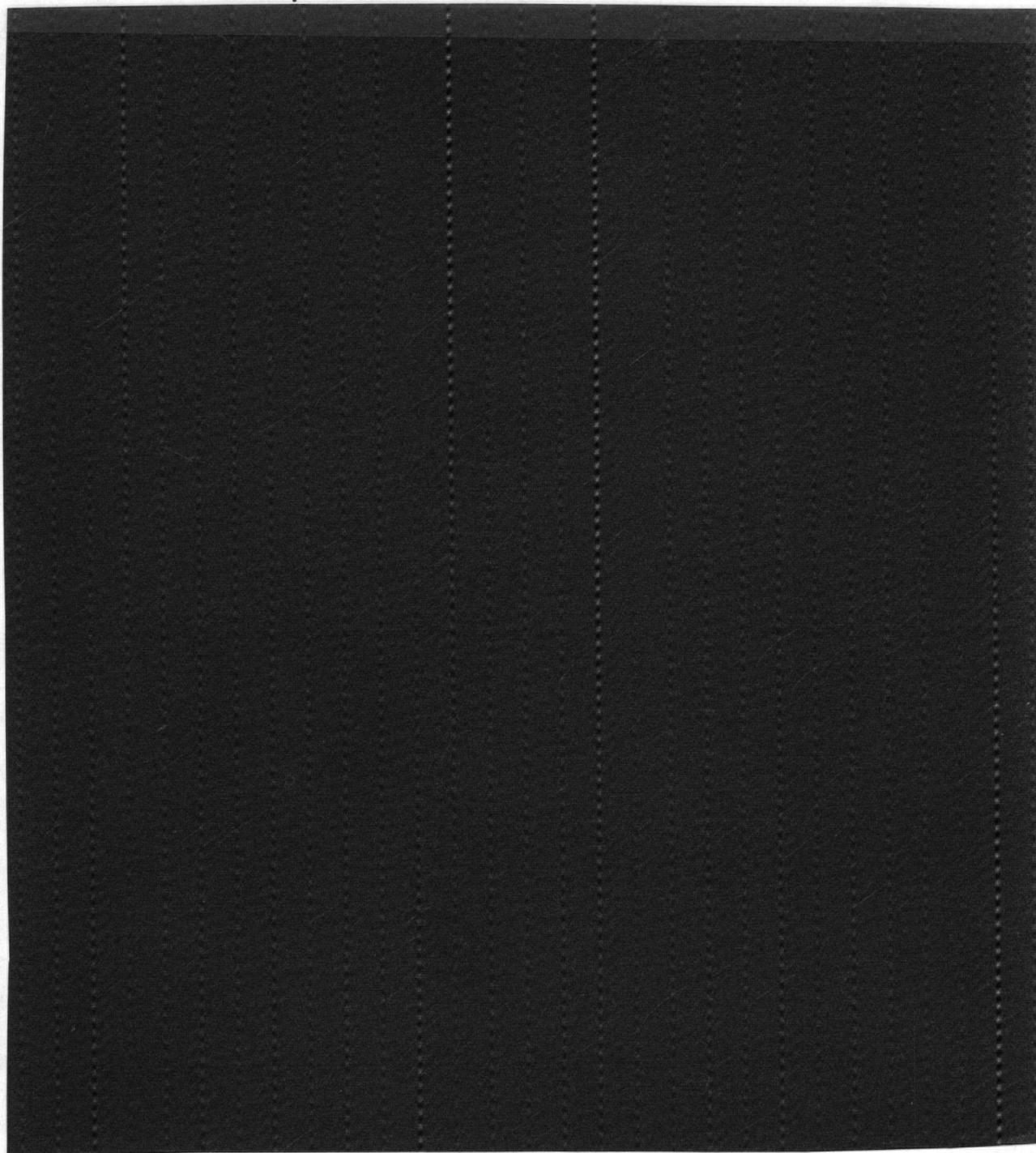
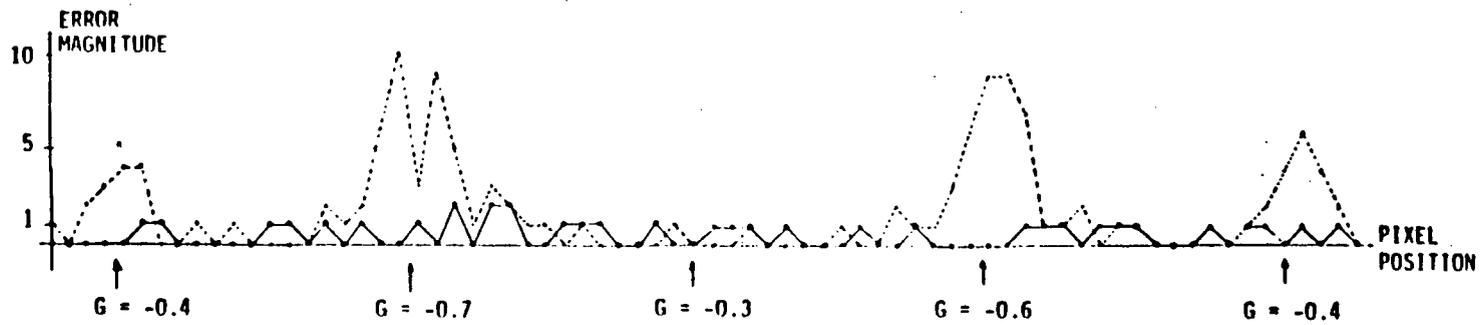
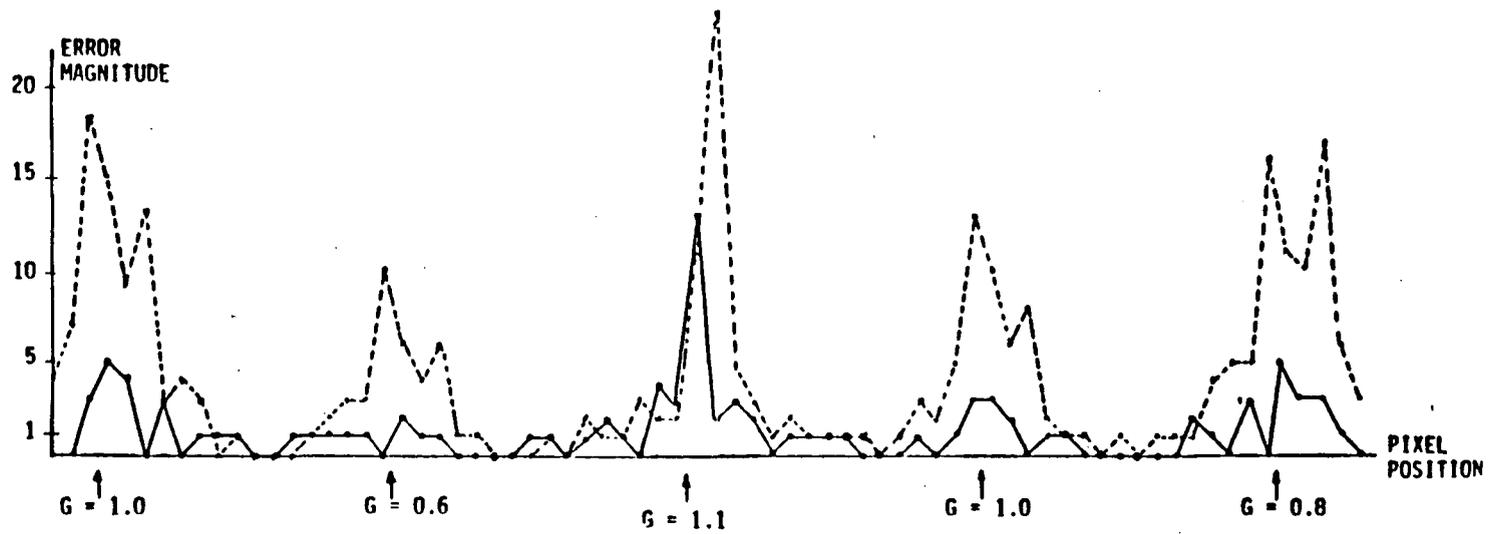


FIGURE 4.28 RADIOMETRIC ERROR IN RECTIFIED SINE WAVE IMAGE
(UNDERLAP) WITH PROPOSED METHOD



(a) OVERLAP



(b) UNDERLAP

----- Sinc Interpolation
 _____ Proposed Method

FIGURE 4.29 ERROR MAGNITUDE

4.5.7 Synthesized TM Imagery

TM imagery with scan gaps can be synthesized using existing Landsat MSS imagery. A 512 x 512 Landsat-2 MSS image over Vancouver Island, B.C., Canada was selected for experiment.

Figure 4.30 shows the scene which is imaged over rugged terrain. A scan gap of 0.0 to -1.0 line corresponding to overlap was used to simulate the jittered image. In the simulation, a 16-point sinc interpolation was used. Reconstruction was done by both the sinc interpolation method and the proposed method. Figures 4.31 and 4.32 show the error images, each multiplied by a factor of 4. The error images show that the proposed method yields less interpolation error than the sinc interpolation. The RMS error for the proposed method is 0.8 level and that for the sinc interpolation is 1.7 levels. RMS error also includes the error introduced in simulating the jittered image by a 16-point sinc interpolation. The error in the simulation is about 0.5 level RMS. When this simulation error is subtracted (in the Gaussian sense), the proposed method performs two to three times better than the sinc interpolation and this is consistent with the sine wave experimental results. For the sinc interpolation, the worst errors occur in the scan gap neighbourhoods; while in the proposed method (overlap case), the error is independent of positions relative to scan gaps. This again agrees with the sine wave experimental results. The larger errors in the scan gap neighbours are due to the following:

- (a) Kernel weights in larger scan gaps do not taper off as fast as those in smaller scan gaps.
- (b) Error is attributed to creating new samples. When interpolating



FIGURE 4.30 MSS IMAGE USED TO SYNTHESIZE TM IMAGE

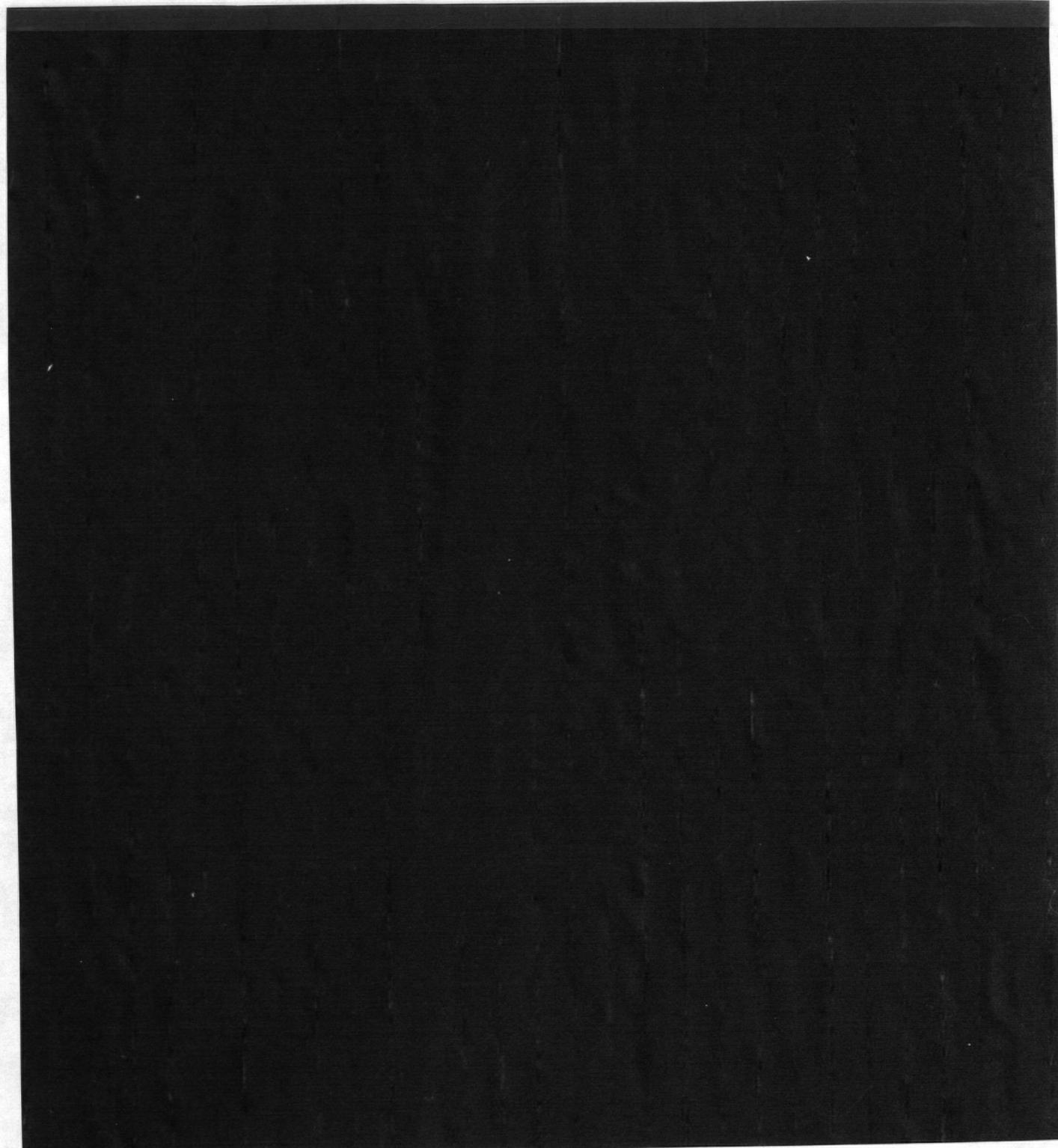


FIGURE 4.31 RADIOMETRIC ERROR IN RECTIFIED SYNTHESIZED TM
IMAGE WITH SINC INTERPOLATION



FIGURE 4.32 RADIOMETRIC ERROR IN RECTIFIED SYNTHESIZED TM
IMAGE WITH PROPOSED METHOD

in a gap neighbourhood, the kernel weights on these new samples are heavy.

Different scan gap sizes are tested with the scene and the RMS errors are shown in Figure 4.33. The following should be noted:

- (a) The distorted images with scan gaps were simulated by using 16-point sinc interpolation which would introduce an RMS error of about 0.5 level.
- (b) In the overlap area, the proposed method yields an RMS error two to three times lower than that of the sinc interpolation method (after subtracting the error in the RMS sense introduced in simulating the distorted image). This agrees with the sine wave experimental results.
- (c) In the underlap area, the RMS errors for both methods are about equal. This is due to errors introduced in the simulation of the distorted imagery. In addition to the 0.5 level RMS error, another error source could be that the data spacings in the original MSS image are not uniform. A scan gap occurs once every 6 lines of data and can be as big as 0.1 line. In the interpolation process, creating new samples is another error source.

Nevertheless, the imagery experiments indicate that the proposed method would yield an interpolation error of less than 1 level RMS in the case of overlap.

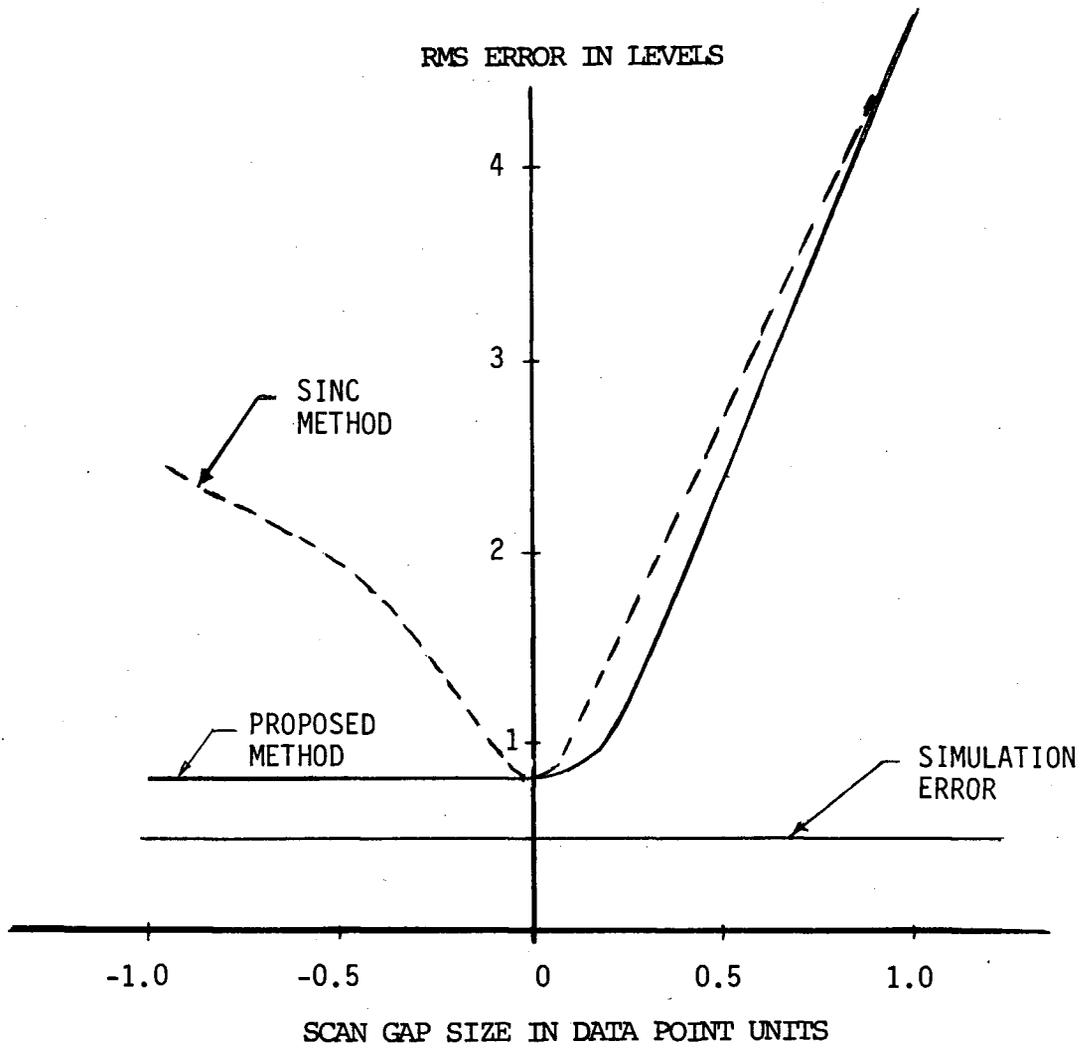


FIGURE 4.33 RMS ERROR VERSUS SCAN GAP SIZE IN SIMULATED TM IMAGE

4.6 Experiments — Relief Displacement

4.6.1 Objective

Rectification experiments of the first pass have been performed with synthesized imagery. The objective of this section is to present the experimental results which show that subpixel geometric accuracy can be achieved. Terrain models consisted of a simulated Gaussian terrain and a real DTM. Both Landsat-4 and SPOT sensor parameters are considered.

In synthesizing the imagery, a Lambertian surface reflectance model [WOOD80 and HORN81b] is assumed where the intensity of any particular point (x,y) is given by:

$$I(x,y) = \frac{p p_s + q q_s + 1}{1 + p^2 + q^2} \frac{1 + p_s^2 + q_s^2}{1 + p_s^2 + q_s^2} \quad (4.15)$$

where $p = \partial f(x,y) / \partial x$,

$q = \partial f(x,y) / \partial y$,

$f(x,y)$ = altitude of point (x,y) , and

$(-p_s, -q_s, 1)$ is a vector pointing in the direction of the sun.

Other surface reflectance models would yield different radiometric values for the synthesized imagery but would not significantly affect the study of geometric rectification accuracy.

The results obtained with simulated Gaussian terrain are presented first. The advantage of using this terrain model is that slopes p , q required by the Lambertian model, and the intersection of the sensor

viewing vector with the terrain, can be computed analytically. In the case with the real DTM, the slopes and the intersection points have to be estimated numerically.

4.6.2 Gaussian Terrain with Landsat-4 TM Sensor Parameters

The Gaussian terrain model consists of several Gaussian hills as illustrated in Figure 4.34. Here elevation has been shown directly proportional to image brightness. The terrain varies from 0 to 2,500 metres. A synthesized orthographic image using Landsat-4 TM sensor parameters of the same area is shown in Figure 4.35. Here, the sun is at elevation 46° and azimuth 122° , corresponding to sun position in a TM scene. The image area is about 15 km x 15 km. The synthesized image has a pixel size of 30 metres x 30 metres. The image as seen by the sensor is shown in Figure 4.36. This distorted image is synthesized by computing, analytically, the intersection of the sensor vector with the terrain and then the slopes at the point of intersection. Since the worst case terrain distortion is to be studied, the area is assumed to appear on the leftmost part of the imaged scene. Here, an elevation of 1 km is displaced by about 130 metres.

The distorted image is rectified by the system depicted in Figure 2.6 (Section 2.6.2) and the rectified image is shown in Figure 4.37. The geometric accuracy of the rectified image is illustrated in Figure 4.38. The terrain of one of the hills is shown in Figure 4.38a and the intensity of the same area in the orthographic image is plotted in Figure 4.38b. Figures 4.38c and d are identical except for minor radiometric errors.

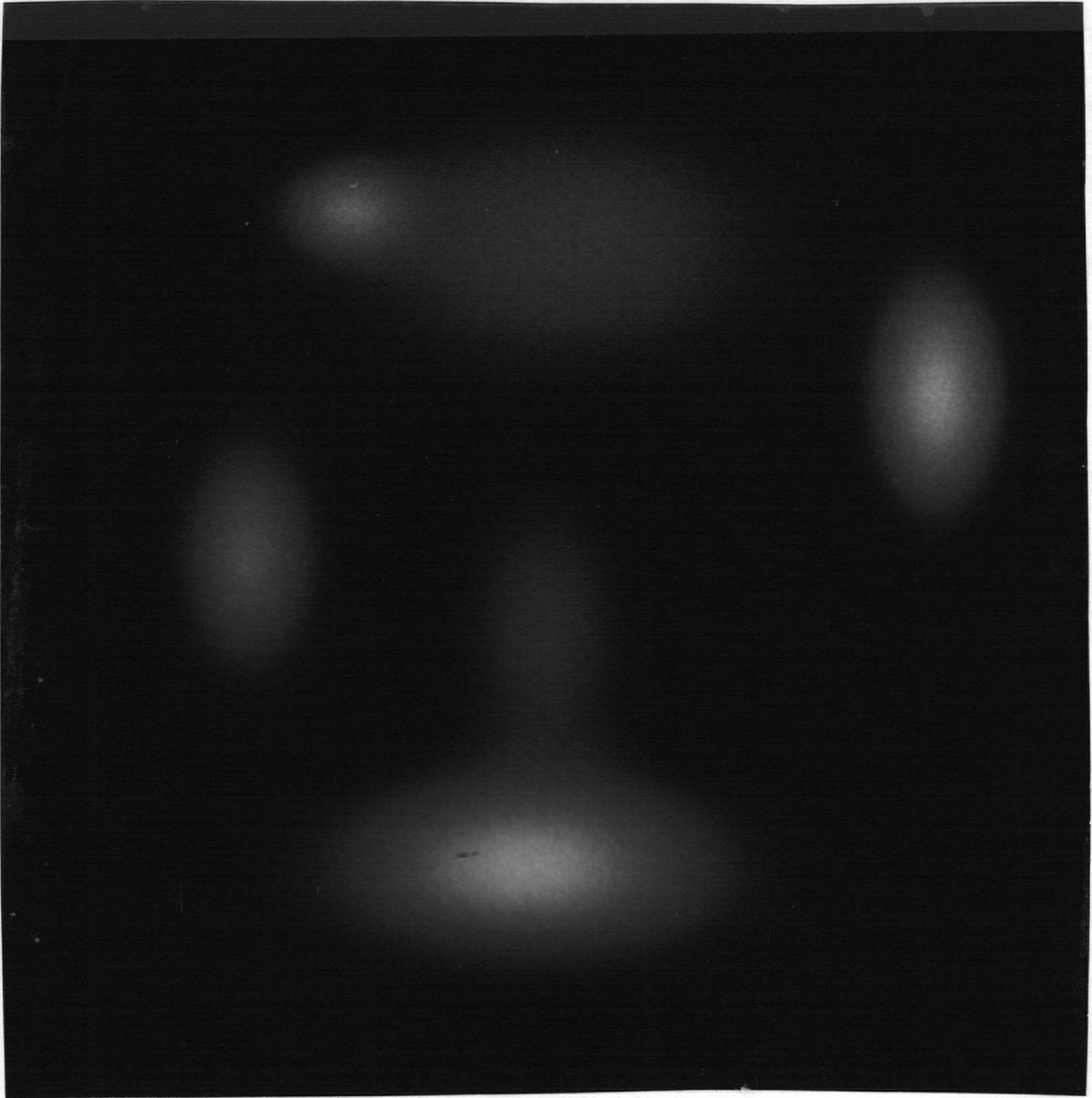


FIGURE 4.34 GAUSSIAN TERRAIN

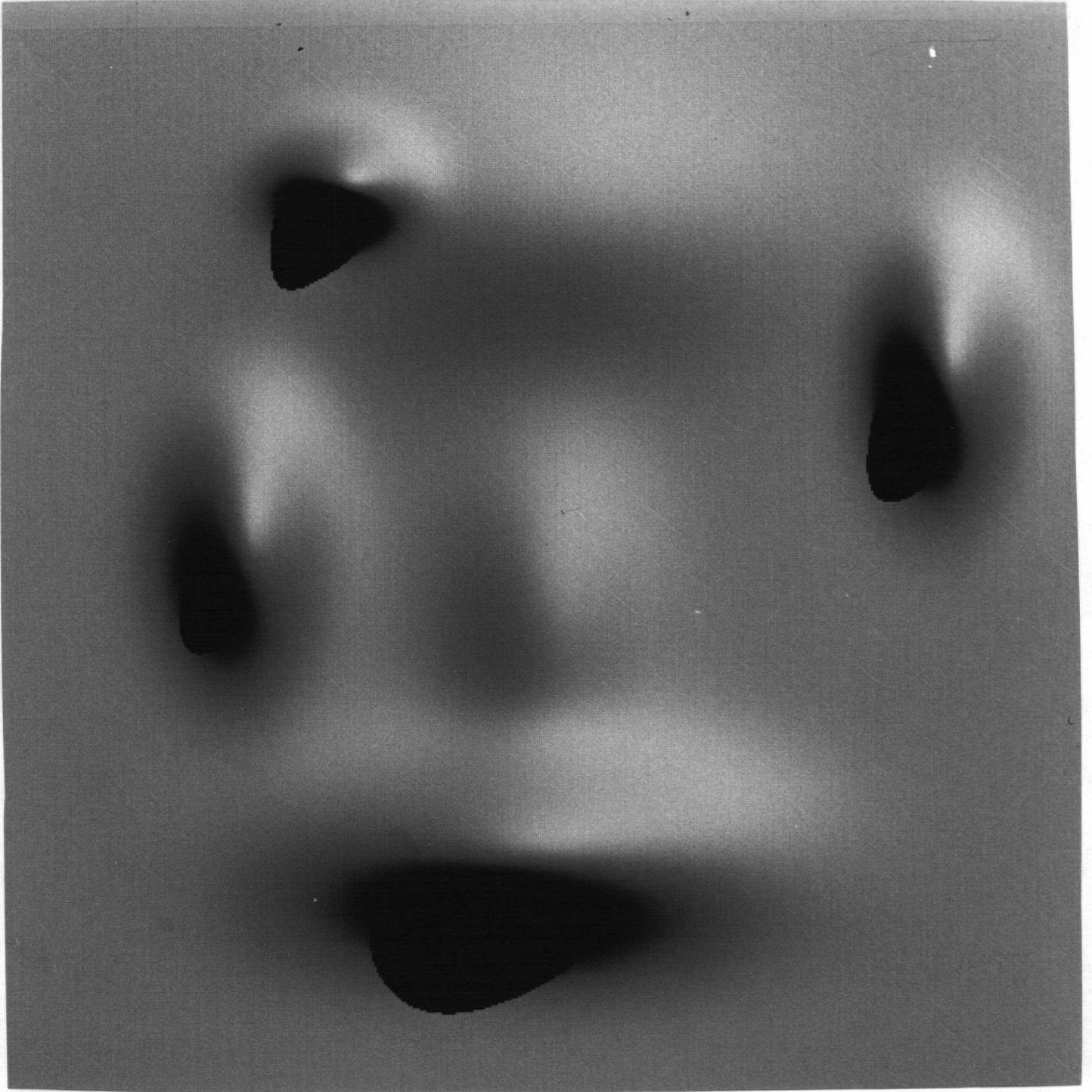


FIGURE 4.35 ORTHOGRAPHIC IMAGE (GAUSSIAN TERRAIN)

z ← | —



FIGURE 4.36 DISTORTED IMAGE (GAUSSIAN TERRAIN)

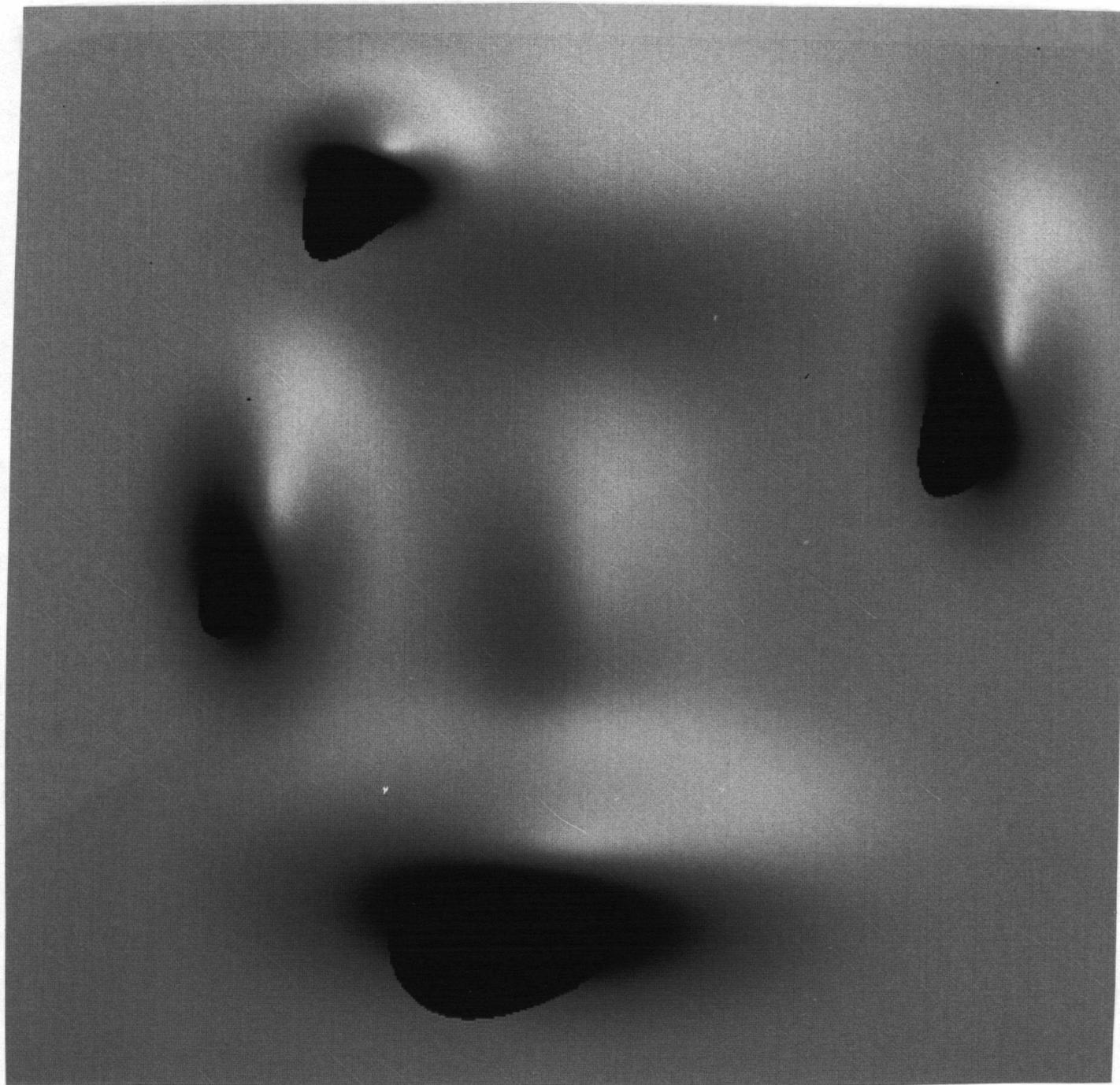


FIGURE 4.37 RECTIFIED IMAGE (GAUSSIAN TERRAIN)

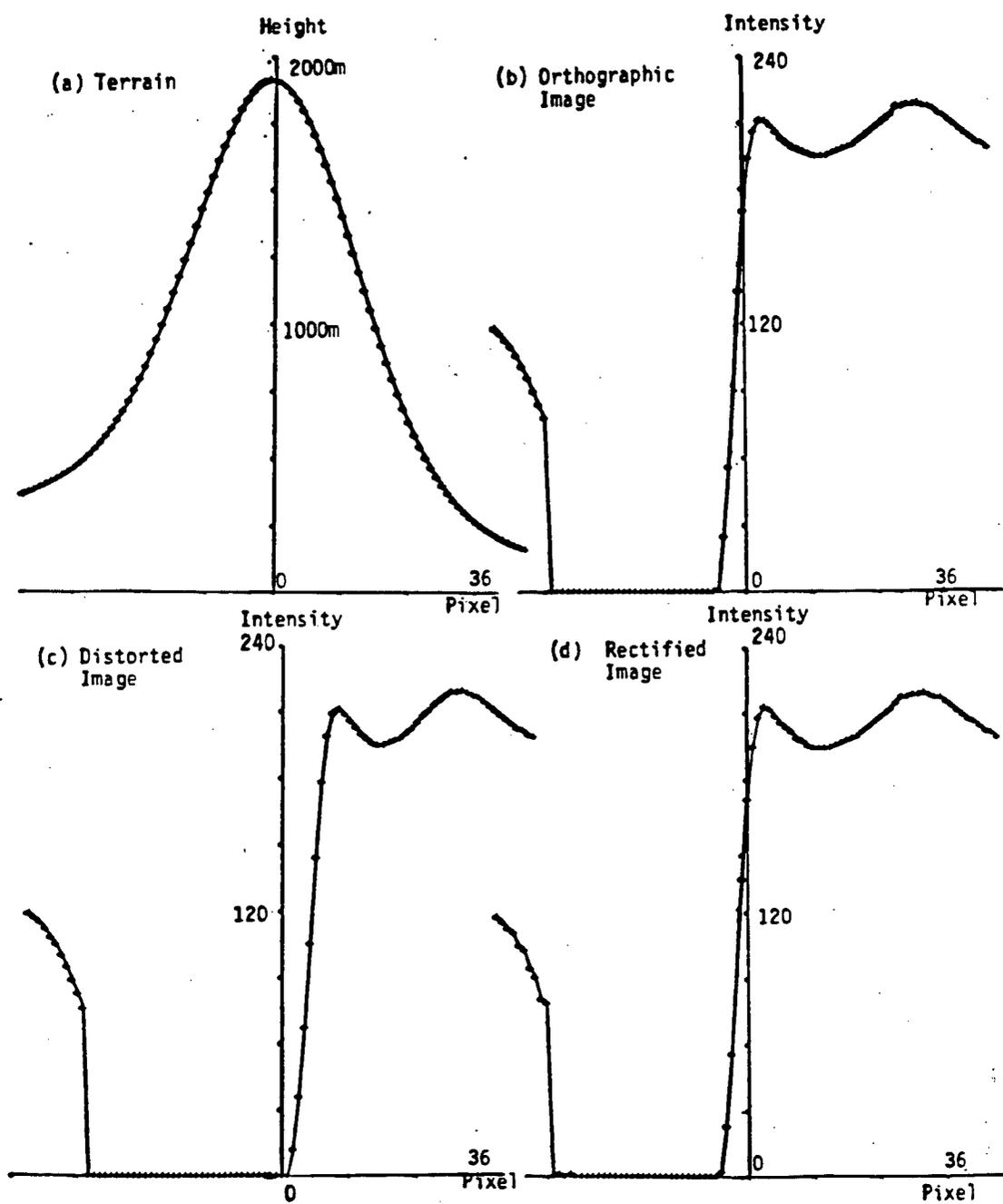


FIGURE 4.38 CROSS-SECTIONAL PLOT OF A HILL

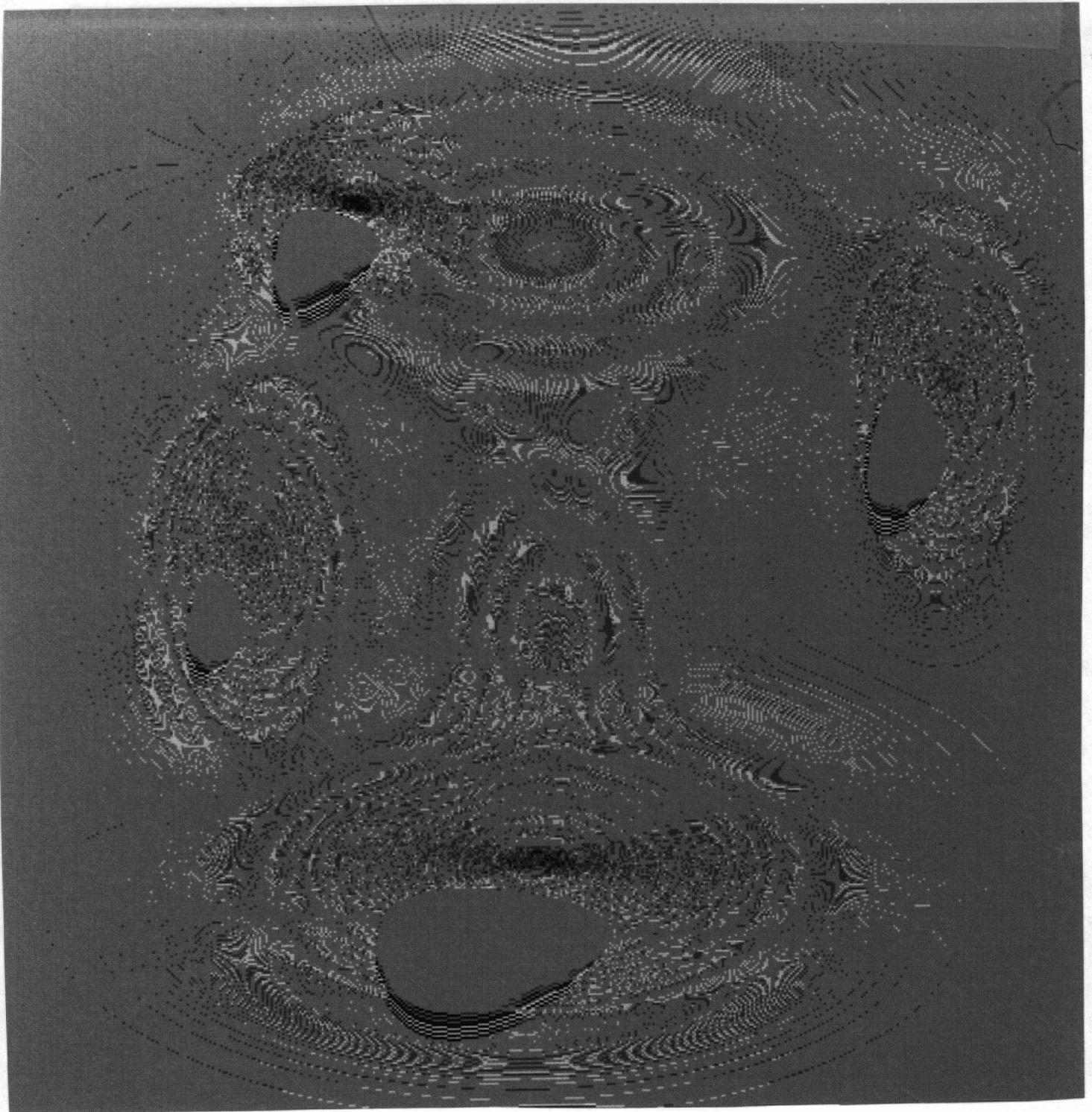


FIGURE 4.39 ERROR IMAGE (GAUSSIAN TERRAIN)

The relief displacement can easily be seen by comparing Figures 4.38b and c. The difference between the orthographic image and the rectified image shows the intensity of each Gaussian hill is accurate to within 0.5 level RMS out of 256. The difference, accentuated by a factor of 80, is shown in Figure 4.39. This difference is mostly due to an interpolation such as ringing (Gibb's phenomenon) at the shadow boundaries.

This simple experiment demonstrates the successful result obtained with the proposed rectification algorithm with an analytical terrain model.

4.6.3 Real Terrain with Landsat-4 TM Sensor Parameters

A regular gridded DTM of the St. Mary Lake area in B.C, Canada is used with a spatial resolution of 30 metres x 30 metres. The DTM covers an area of 15 km x 15 km. Figure 4.40 shows the terrain and Figure 4.41 shows the orthographic image. The terrain ranges from 950 metres in the valley to 2,860 metres at the peaks. Again, to obtain the worst case terrain distortion in the distorted image, the area is assumed to be at the leftmost part of the image. The distorted image is shown in Figure 4.42. The sun position is the same as in the Gaussian terrain case, i.e., from the southwest. (A more accurate sense of the terrain can be viewed by turning the picture upside down since human visual interpretation is more accustomed to having the illumination from the top left corner.)

In synthesizing the distorted image, the slope at a point in each

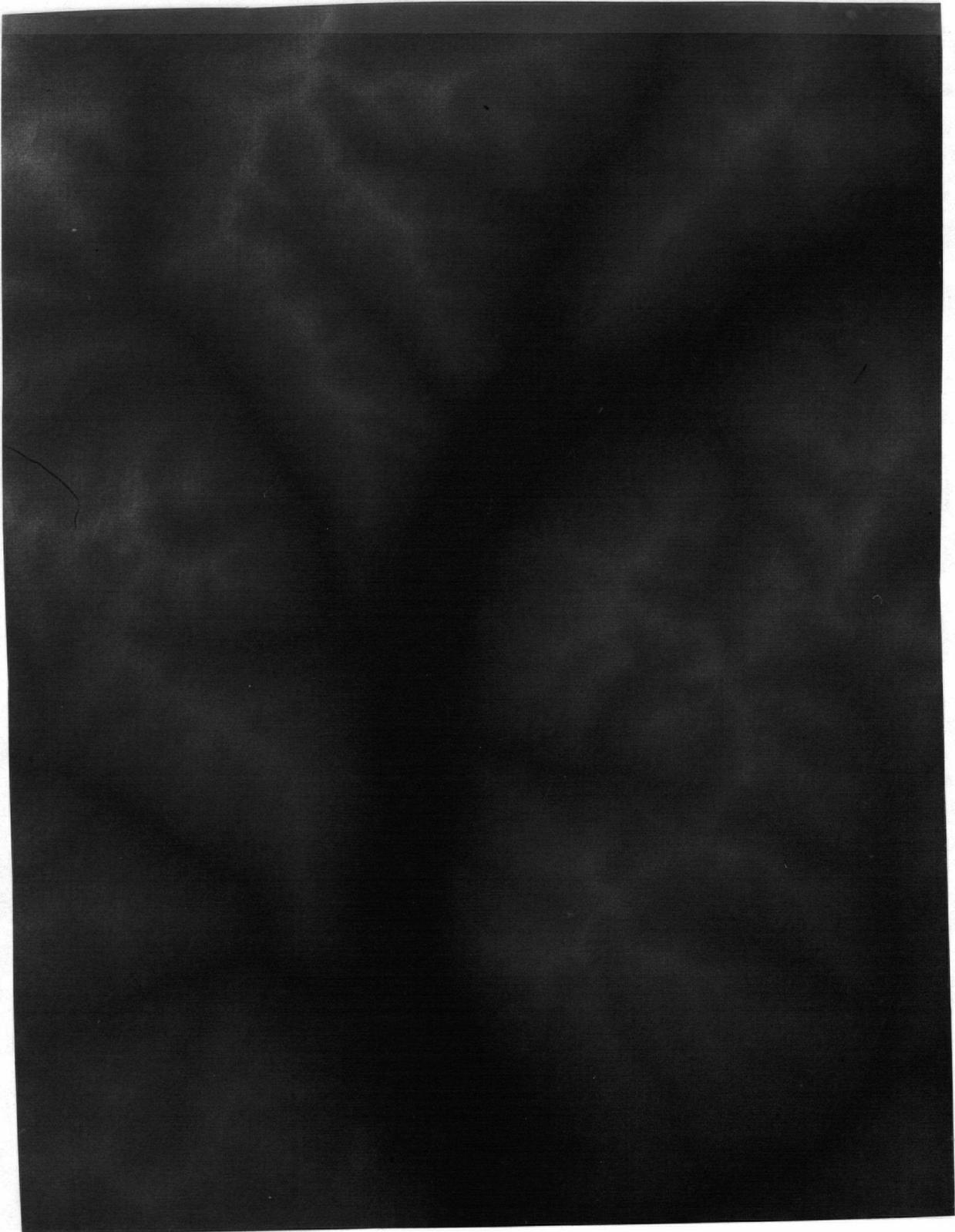


FIGURE 4.40 DTM OF ST. MARY LAKE AREA

z ←



FIGURE 4.41 LANDSAT-4 TM ORTHOGRAPHIC IMAGE



FIGURE 4.42 LANDSAT-4 TM RELIEF DISTORTED IMAGE

direction is determined by fitting a quadratic curve through the three closest points in that direction. Newton's form and divided differences can readily be applied here. Let the three closest points to t , where the slope is required, be (t_1, F_1) , (t_2, F_2) and (t_3, F_3) where the F 's are the DTM values. Then the slope at t is:

$$F'(t) = F_2 - F_1 + 0.5 (F_3 - 2 F_2 + F_1) (2t - t_1 - t_2)$$

which can be computed very easily.

The rectified image is shown in Figure 4.43. The geometric accuracy can best be illustrated by Figure 4.44 which is a plot of the displacement vectors as a result of correlation between the original orthographic image and the rectified image for 100 points. The mean error is only 0.023 pixel and the standard deviation is 0.077 pixel. The maximum error is only 0.22 pixel. Hence subpixel geometric rectification accuracy has been achieved with the DTM.

The RMS difference between the orthographic and the rectified images is about 6 levels RMS out of 256. This radiometric difference is caused by the fact that the intensity function of the synthesized distorted image is not band limited. The difference, accentuated by a factor of 20, is shown in Figure 4.45.

4.6.4 Real Terrain with SPOT PLA Sensor Parameters

The off nadir pointing angle for the SPOT sensor can be as high as 26° . A terrain height of 1,000 metres is displaced by about 700 metres at



FIGURE 4.43 LANSAT-4 TM RECTIFIED IMAGE

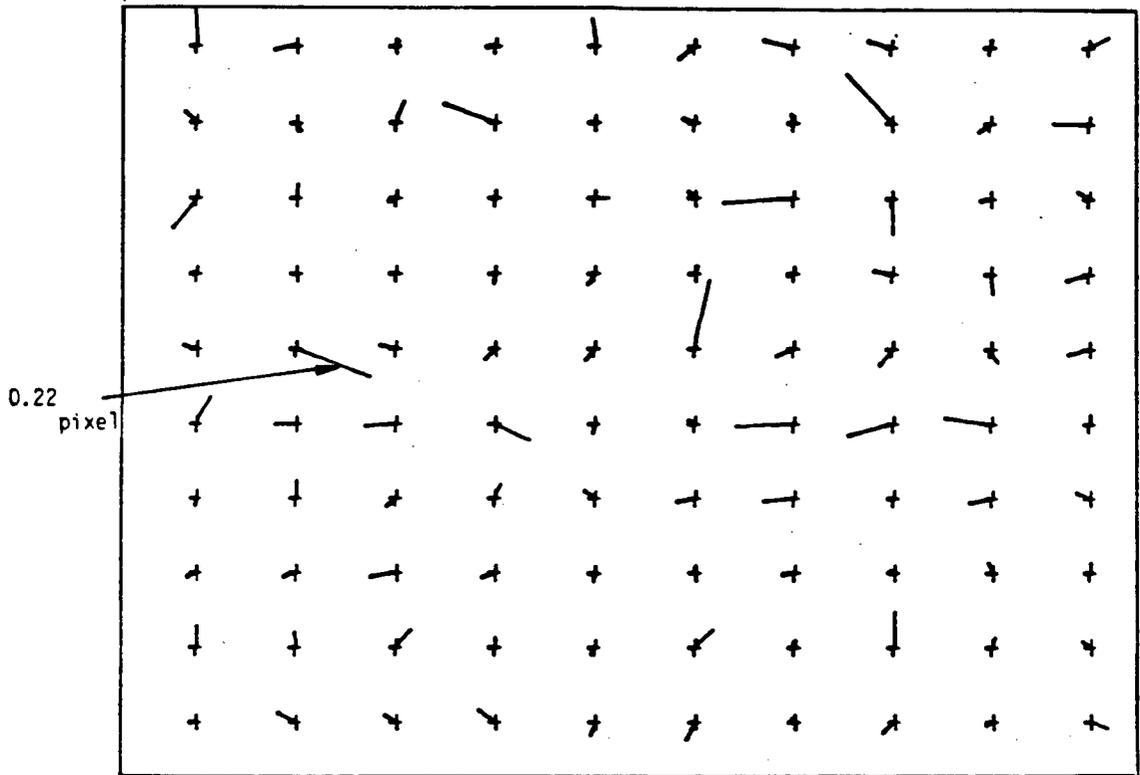


FIGURE 4.44 GEOMETRIC RECTIFICATION ACCURACY (LANDSAT-4 TM)

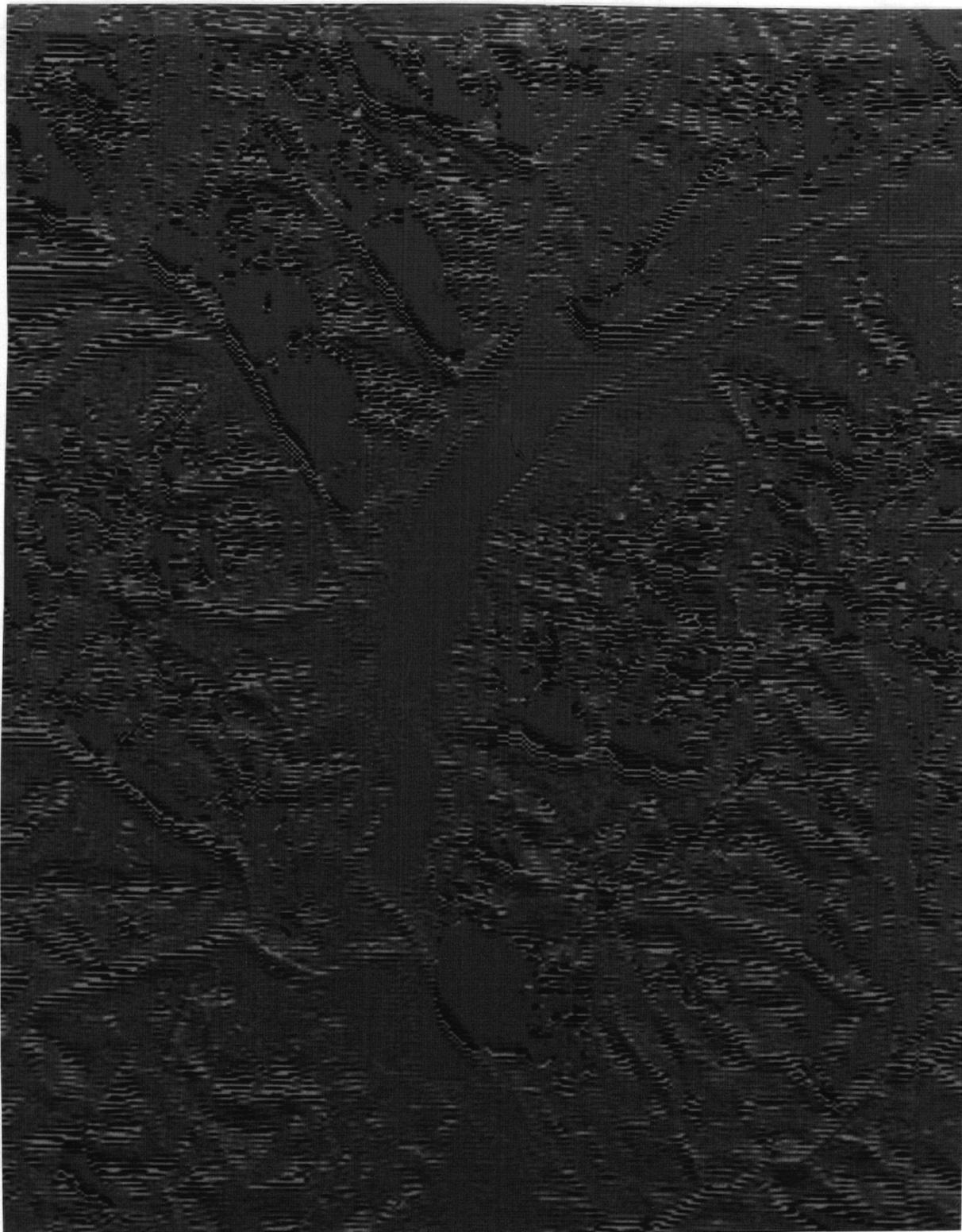


FIGURE 4.45 ERROR IMAGE (LANDSAT-4 TM)

Note: This results from both geometric and radiometric errors.

the far edge of the scene. The same regularly gridded DTM as in the Landsat-4 TM case is used, but the spatial resolution is now assumed to be 20 metres x 20 metres. Also, the altitude is multiplied by 2/3 to retain the same slopes. A realistic sun angle is taken to be at elevation 68° and azimuth 104° . The orthographic image is shown in Figure 4.46 and the distorted image is shown in Figure 4.47. It can be seen from Figure 4.47 how severely the mountain ranges are distorted. Moreover, since the sensor and the sun are at opposite sides of the image, the brighter sides of the slopes which face the sun are much narrower than the darker sides of the slopes which face away from the sun. This will undoubtedly introduce radiometric errors in the rectified image as shown in Figure 4.48. However, the displacement vectors for 100 points in Figure 4.49 show that the mean error is only 0.028 pixel and the standard deviation is 0.07 pixel. Also, the maximum error is 0.42 pixel. Hence, subpixel geometric rectification accuracy can also be achieved in SPOT imagery for the PLA sensor when operating in the side looking mode.

The RMS difference between the orthographic and the rectified images is again about 6 levels out of 256 due to reasons already mentioned for the Landsat-4 TM case. The difference, accentuated by a factor of 20, is shown in Figure 4.50.

4.7 Experiments with Landsat-2 Imagery

The Landsat-2 scene (I.D.2921-18025, imaged on 31st July 1977) used is an image of an area over Vancouver, B.C., Canada. The image is shown in Figure 4.51.

z ←



FIGURE 4.46 SPOT ORTHOGRAPHIC IMAGE

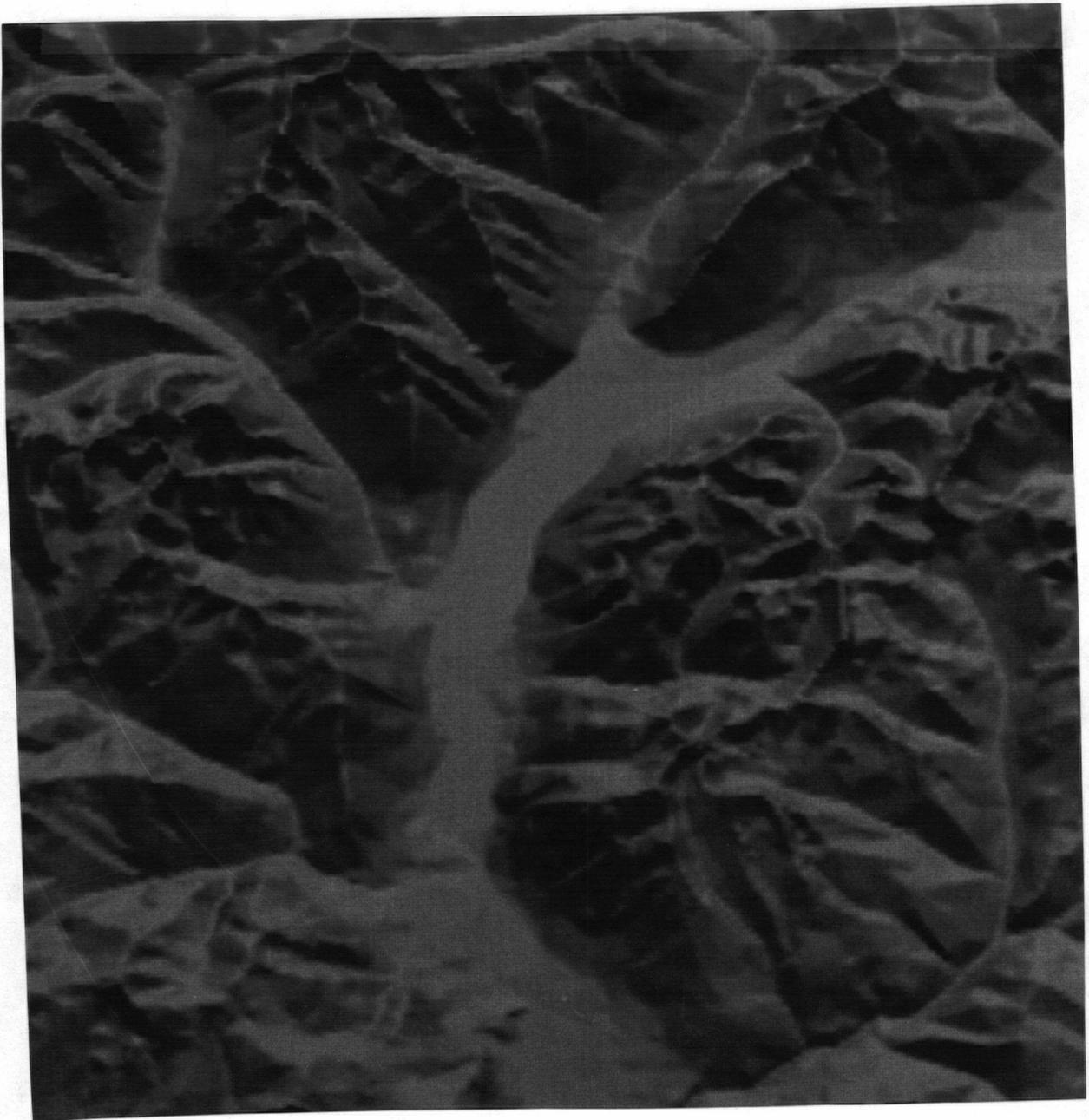


FIGURE 4.47 SPOT RELIEF DISTORTED IMAGE



FIGURE 4.48 SPOT RECTIFIED IMAGE

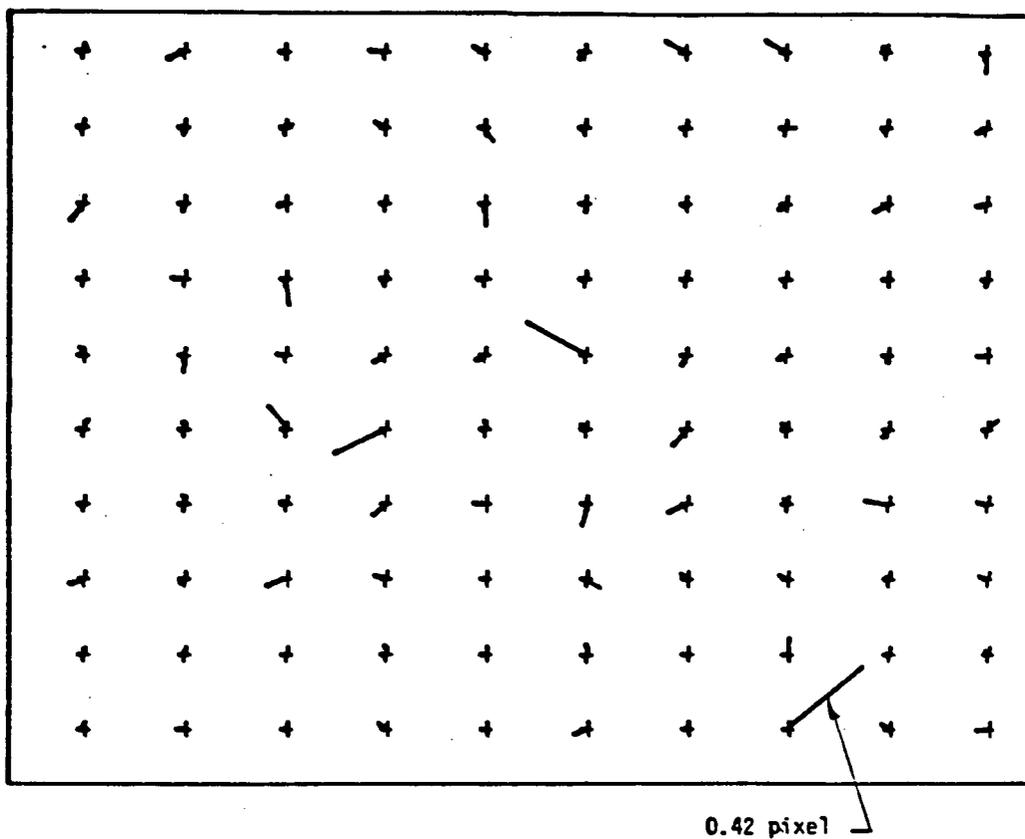


FIGURE 4.49 GEOMETRIC RECTIFICATION ACCURACY (SPOT)



FIGURE 4.50 ERROR IMAGE (SPOT)

Note: This results from both geometric and radiometric errors.

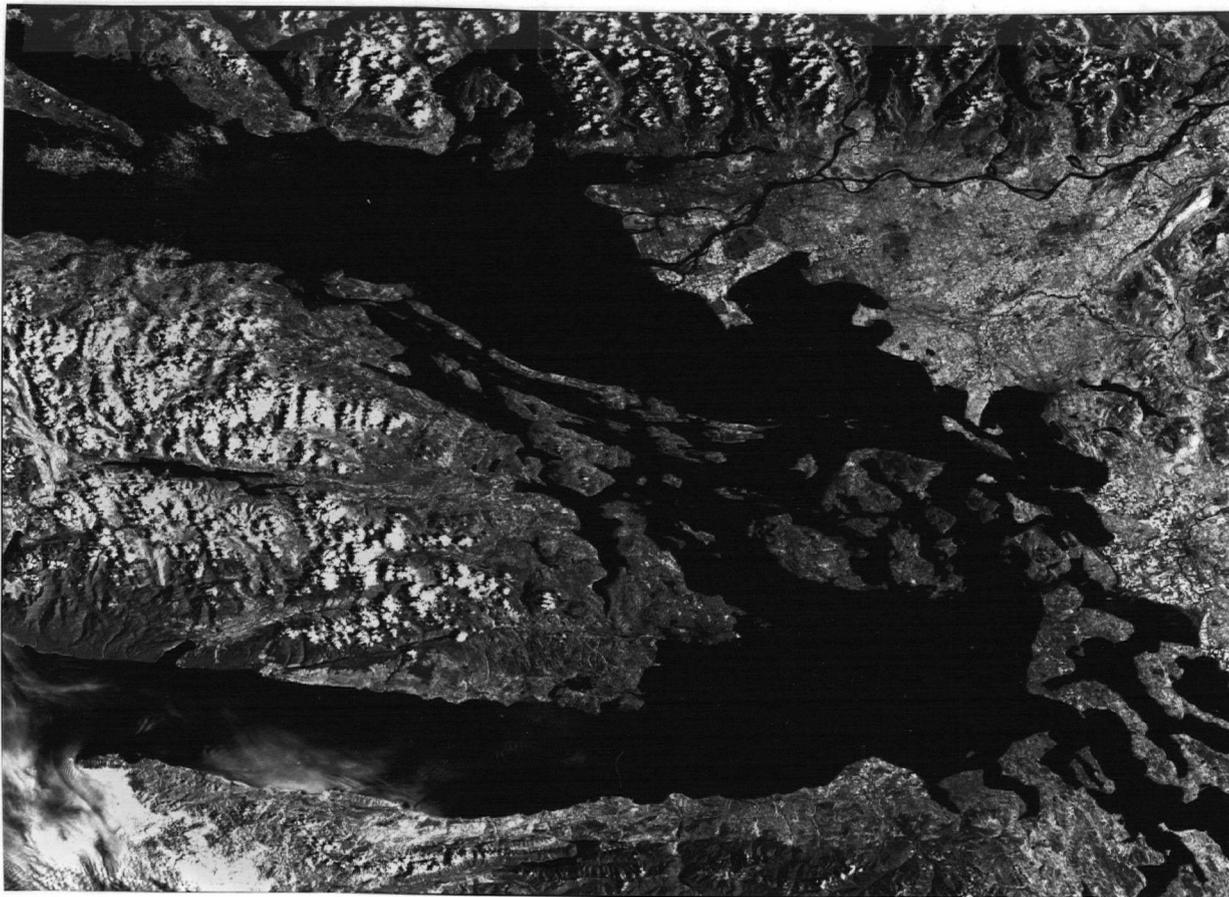


FIGURE 4.51 LANDSAT-2 MSS RAW IMAGE

Results of the navigation procedure are shown in Figure 4.52 and Table 4.4. In the figure, the RMS residual error stays relatively constant at about 70 metres. The error is mainly caused by:

- (a) Marking a target line, pixel location in the image from which a pixel corresponds to 58 metres x 79 metres. In the process the maximum error is estimated to be 1 pixel in the along and across scan direction. Assuming uniform distribution, the RMS error is:

$$\sigma_1 = \sqrt{\left(\frac{58}{\sqrt{3}}\right)^2 + \left(\frac{79}{\sqrt{3}}\right)^2} \text{ metres} = 57 \text{ metres}$$

- (b) Marking of target location on a mapsheet. The mapsheets used are 1:50,000 scale. A marking error of 1 mm corresponds to 50 metres. Assuming this is the maximum error, then the RMS error is:

$$\sigma_2 = \sqrt{\left(\frac{50}{\sqrt{3}}\right)^2 + \left(\frac{50}{\sqrt{3}}\right)^2} \text{ metres} = 21 \text{ metres}$$

- (c) Map accuracy. It is not known what class the mapsheets belong to. Assume that the best maps (Class A1) are used, then the RMS error in each direction is 16 metres. Denote this by σ_3 .

Therefore the total error budget is:

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = 72 \text{ metres}$$

which is close to the experimental residual error.

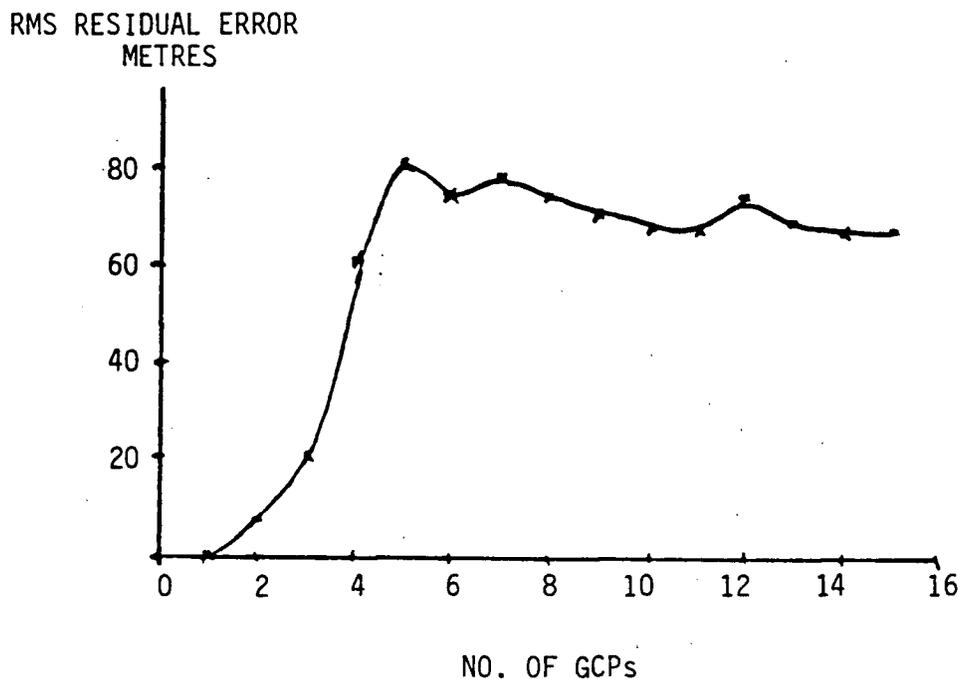


FIGURE 4.52 RESIDUAL ERROR VERSUS NUMBER OF GCPs

TABLE 4.4 ATTITUDE AT 9 POINTS OVER THE IMAGE

(NOTE: ALL VALUES IN DEGREES)

ROLL	PITCH	YAW
-0.309	0.124	-0.516
-0.315	0.120	-0.517
-0.319	0.116	-0.518
-0.323	0.120	-0.516
-0.325	0.109	-0.514
-0.326	0.107	-0.511
-0.325	0.107	-0.505
-0.322	0.111	-0.500
-0.319	0.119	-0.494

The rectified scene is on a UTM projection with the north axis aligned with the vertical. The angle of rotation (from raw image to rectified image) is about 14° . The rectified image is shown in Figure 4.53. The pixel spacing is 50 metres and the scale is 1:1,000,000. This choice of scaling proves to be very convenient if a subarea of the image is required and on a scale of $1:1,000,000n$ or $1:1,000,000/n$ where n is an integer. An example of using this rectified image on another scale will be given in the next chapter.

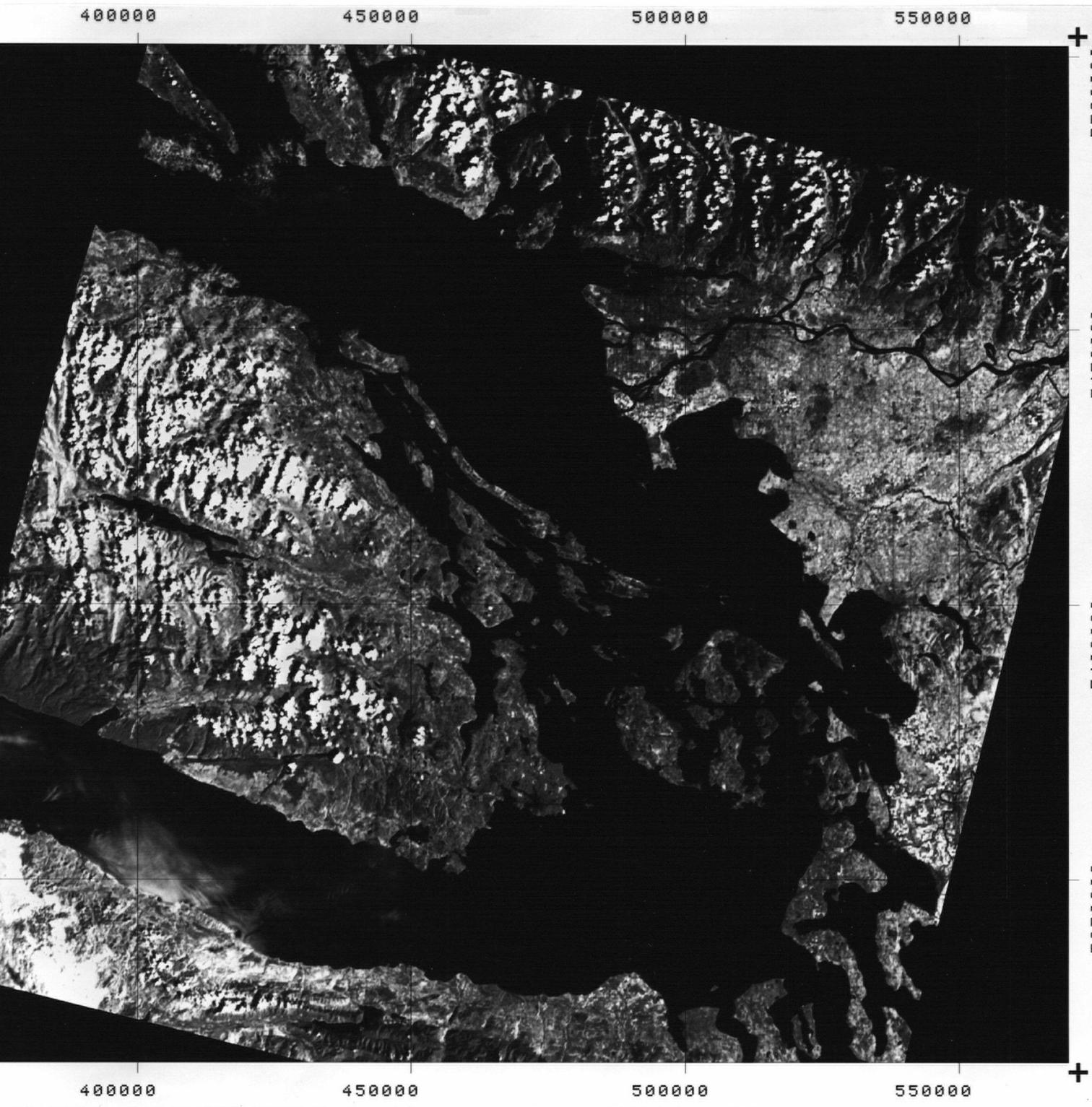


FIGURE 4.53 RECTIFIED LANDSAT-2 MSS IMAGE

CHAPTER 5 SYNTHETIC APERTURE RADAR

5.1 Introduction

Relief displacement is a major factor in synthetic aperture radar (SAR) imagery. It is necessary to understand how SAR works and how the SAR signals are processed before the rectification remapping can be derived and relief displacement correction can be formulated. The theory of SAR can be found in references [KOVA, MDA76, MDA77a, MDA77b and HOVA79] and is summarized in Appendix F. References [BENN78, BENN79, CUMM79, MDA79 and BENS80] show how the SAR signals are processed. This again is described in Appendix F. The basic properties of SAR and the range and azimuth resolutions attainable are summarized as follows.

A SAR system sends out finite pulses and receives their echoes at regular intervals; meanwhile the satellite carrying the SAR advances in position between pulses. The main characteristics of the processed SAR imagery is that it has high azimuth (along track direction) resolution which is satellite altitude independent, and that this resolution is numerically proportional to the antenna aperture (i.e., the smaller the antenna aperture, the higher the resolution). By virtue of the satellite motion in the azimuth direction, the return signal of a target is a chirp (linear frequency modulated) signal in this direction. By matched filtering this return with its replica, the azimuth resolution ρ_a obtained is

$$\rho_a = D/2 \quad (5.1)$$

where D is the antenna aperture.

High resolution in range (across track direction) can be obtained by transmitting a chirp signal and matched filtering the return by its replica. The range resolution ρ_r thus obtained is

$$\rho_r = 0.886 c / (2 K \tau) \quad (5.2)$$

where c is the velocity of propagation, K is the linear frequency modulation (FM) rate and τ is the radar pulse duration. With 4 looks (Appendix F) and sidelobe suppression the azimuth and range resolution for Seasat imagery is

$$\rho_a = 25 \text{ metres}$$

$$\rho_r = 25 \text{ metres}$$

Without matched filtering, the azimuth resolution is

$$q_a = \lambda R / D \quad (5.3)$$

and the range resolution is

$$q_r = c \tau / (2R) \quad (5.4)$$

where λ is the wavelength of operation and R is the target range. For Seasat they correspond to

$$q_a = 18.7 \text{ km}$$

and $q_r = 5.1 \text{ km}$

These results readily show that the matched filtering is important to obtain high resolutions in both azimuth and range.

The purposes of this chapter are to:

- (a) Develop the rectification transformation which depends on how the SAR data is processed.
- (b) Investigate the geometric distortions in which relief displacement is a major problem.
- (c) Develop orbit refinement methods. In SAR imagery, attitude angles are not parameters in the rectification transformation and hence these angles do not need to be refined.
- (d) Perform rectification experiments with Seasat data.

5.2 Rectification Transformation

The MDA Seasat Processor described in Appendix F is capable of producing a high resolution SAR image whose coordinates are slant range and azimuth. When converted to ground range, the resolution is 25 X 25 metres². To rectify this to a map projection (e.g., UTM), the geometric corrections must take into account satellite ephemeris data, Doppler frequency to which the input data has been compressed (as defined by the Seasat processor), terrain, earth curvature and earth rotation. These geometric corrections together with the output map projection parameters, define the transformation linking the input image to the output image. The transformation is detailed in Appendix G.

It should be noted that the slant range s is not linearly related to ground range g . They are related, from Figure 5.1, by:

$$g = R \theta$$

$$\text{where } \cos \theta = (H^2 + R^2 - s^2) / (2HR)$$

A plot of $\Delta g / \Delta s$ as a function of range for a typical Seasat image is shown in Figure 5.2.

An important result of the transformation is that it is independent of the satellite attitude angles. These attitude angles only affect the positioning of the SAR beam which can be determined by an independent method. However, the attitude angles and orbit must be determined to within a tolerance limit for focusing purposes.

5.3 Terrain Induced Distortions

5.3.1 Relief Displacement

Relief displacement is the major geometric distortion in radar imagery. It has the following attributes:

- (a) A target migrates towards the radar by an amount approximately equal to the product of its height and the tangent of the depression angle (see Figure 5.3). Hence, the displacement effect decreases with depression angle. This can also be explained by the fact that for a larger depression angle the slant range difference between the top and foot of an elevated target is

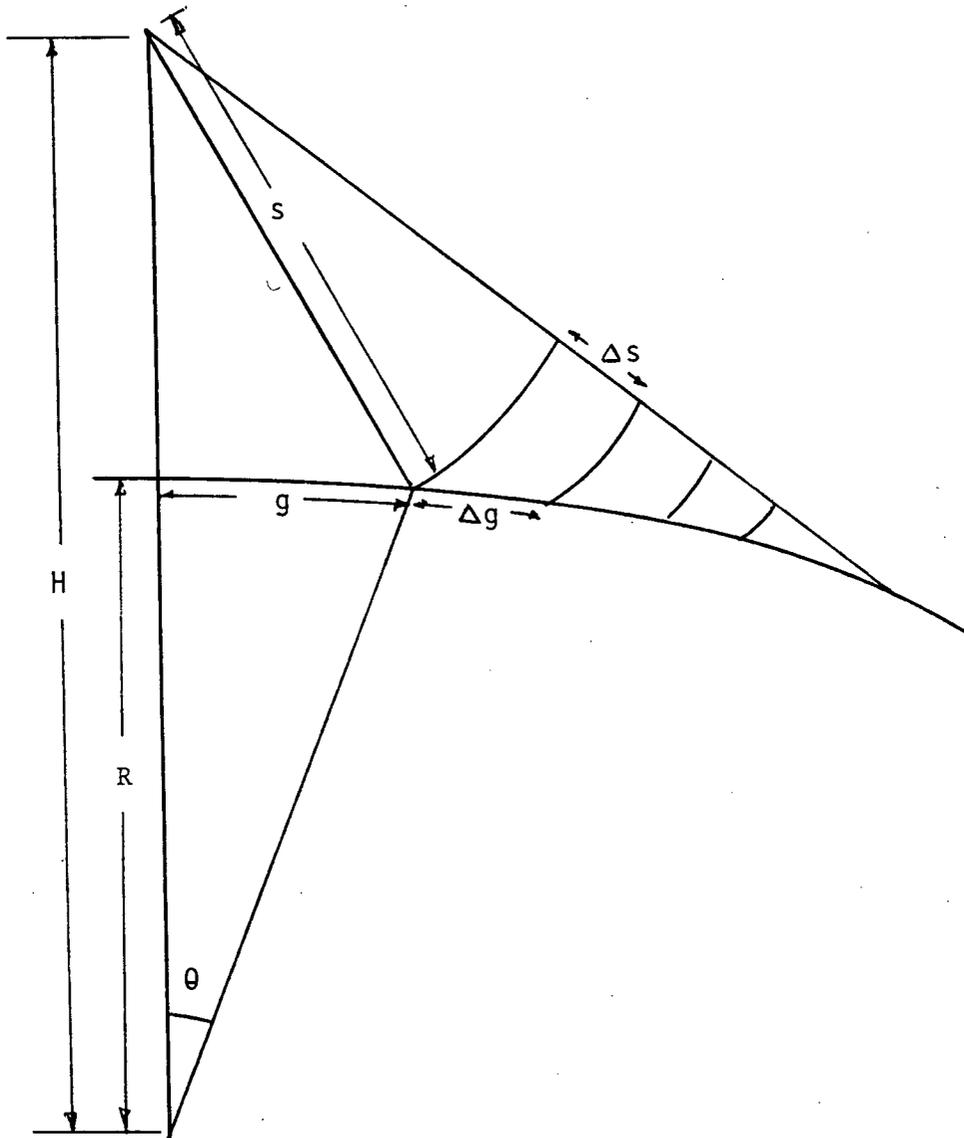
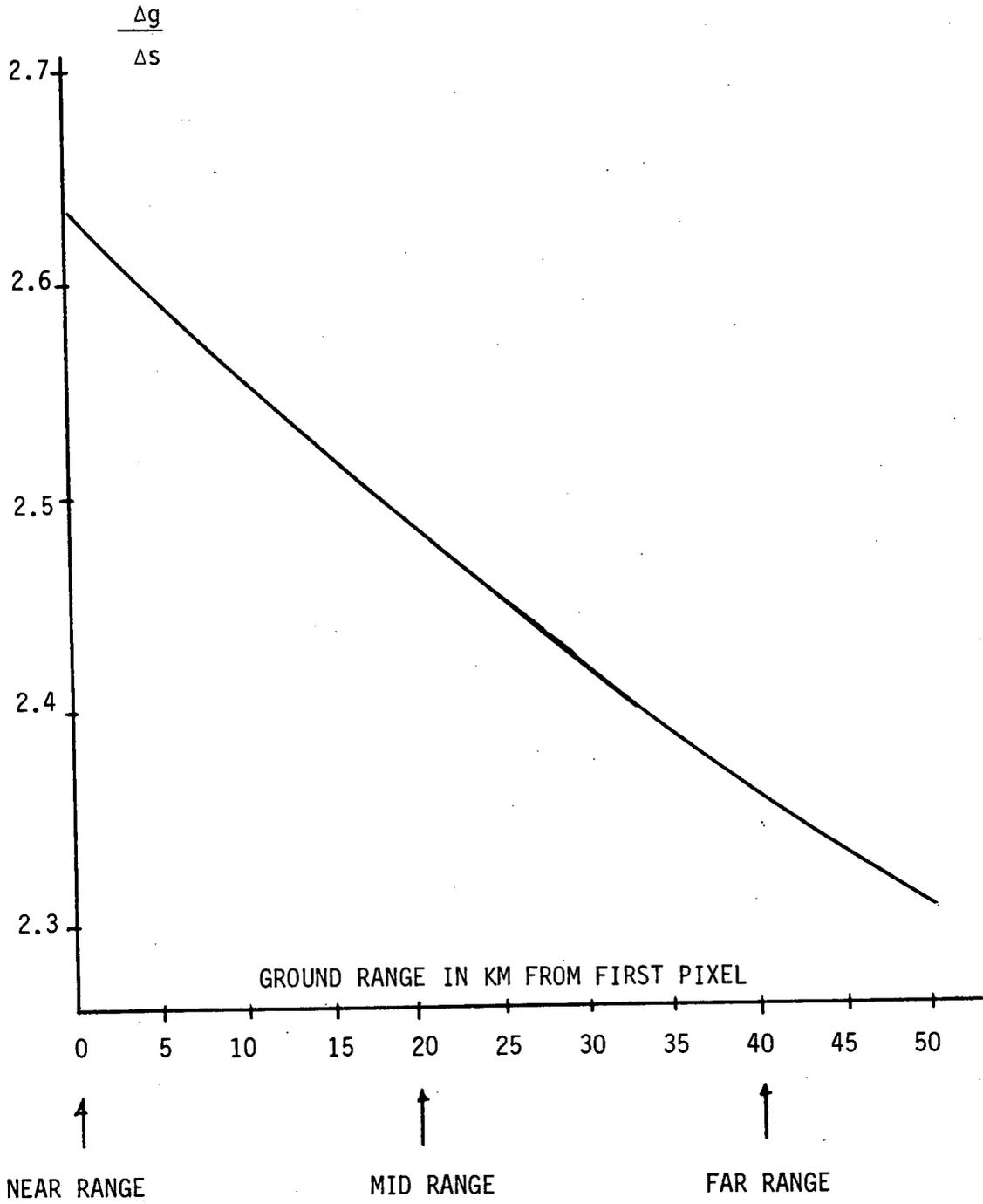


FIGURE 5.1 GEOMETRIC RELATION BETWEEN SLANT RANGE AND GROUND RANGE

FIGURE 5.2 PLOT OF $\Delta g/\Delta s$ VERSUS RANGE

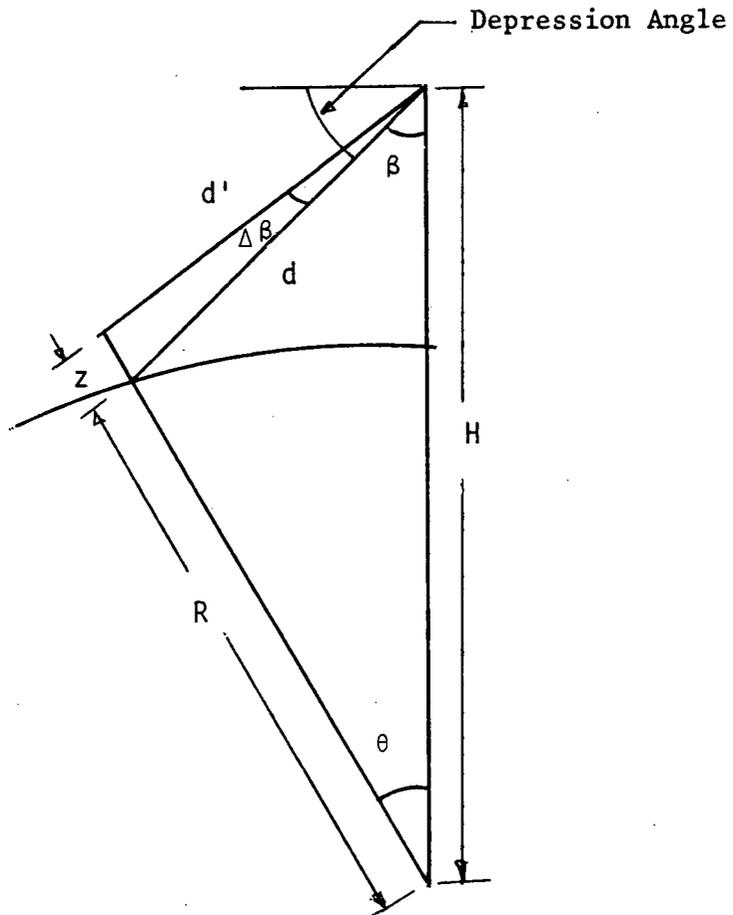


FIGURE 5.3 RELIEF DISPLACEMENT IN RADAR IMAGERY

- greater than the difference for a smaller depression angle.
- (b) Slopes facing the radar appear to be shorter in the image and slopes facing away from the radar appear to be longer. This is referred to as slope foreshortening.
- (c) Radar layover occurs when targets further away on the ground appear to be closer to the radar than nearer targets. This occurs when the slope exceeds a limit defined by the depression angle. Hence, in very mountainous areas, mountain tops and valleys can have the same pixel number (i.e., same slant range) in the image.

In radar imagery, sampling is done at equal increments in slant range. Hence, in the remapping process, it is preferable to define relief displacement in slant range units. From Figure 5.3, the displacement is given by $\Delta d = d' - d$. The value of d' is known to be a linear function of pixel number, and d can be determined from the geometry of the figure:

$$d = \sqrt{R^2 + H^2 - 2RH \cos \theta} \quad (5.5)$$

$$\text{where } \cos \theta = [H^2 + 2(R+z)H - (d')^2] / [2(R+z)H] \quad (5.6)$$

Hence Δd can be expressed in closed form by finding the difference of $d' - d$.

Since $z \ll R$, an approximation of d can be formulated as follows. From Equations 5.5 and 5.6 it can be shown that

$$\Delta d = d' - d \approx z (R - H \cos \theta) / d$$

From the geometry of the figure the above equation can be written as:

$$d = -z \sqrt{1 - (H \sin \beta / R)^2} \quad (5.7)$$

where β is the radar look angle of the vertical and the negative sign shows that displacement is towards the radar.

Using nominal values of Seasat orbit and look angle β , the displacement and the critical slope angle when layover occurs are shown in Figure 5.4.

Relief displacement is a high frequency geometric distortion and can be removed in the first processing pass as will be demonstrated later.

5.3.2 Importance of Compression to Zero Doppler

In the azimuth compression process, the matched filtered output is placed at location of zero Doppler. This has significant impact on the remapping process for the case that the antenna illumination is not broadside. The advantage of compression to zero Doppler is that relief displacement is always only along scan. If compression is to nonzero Doppler, then relief displacement also has a nonzero along scan component.

For example, the Seasat antenna is tilted at about 2° . If compression is performed to the Doppler frequency along the beam centre, then a 1,000 metre high target will be displaced by $2,300 \sin 2^\circ = 80$ metres in the azimuth direction, assuming that each metre in elevation is displaced by 2.3 metres on the ground.

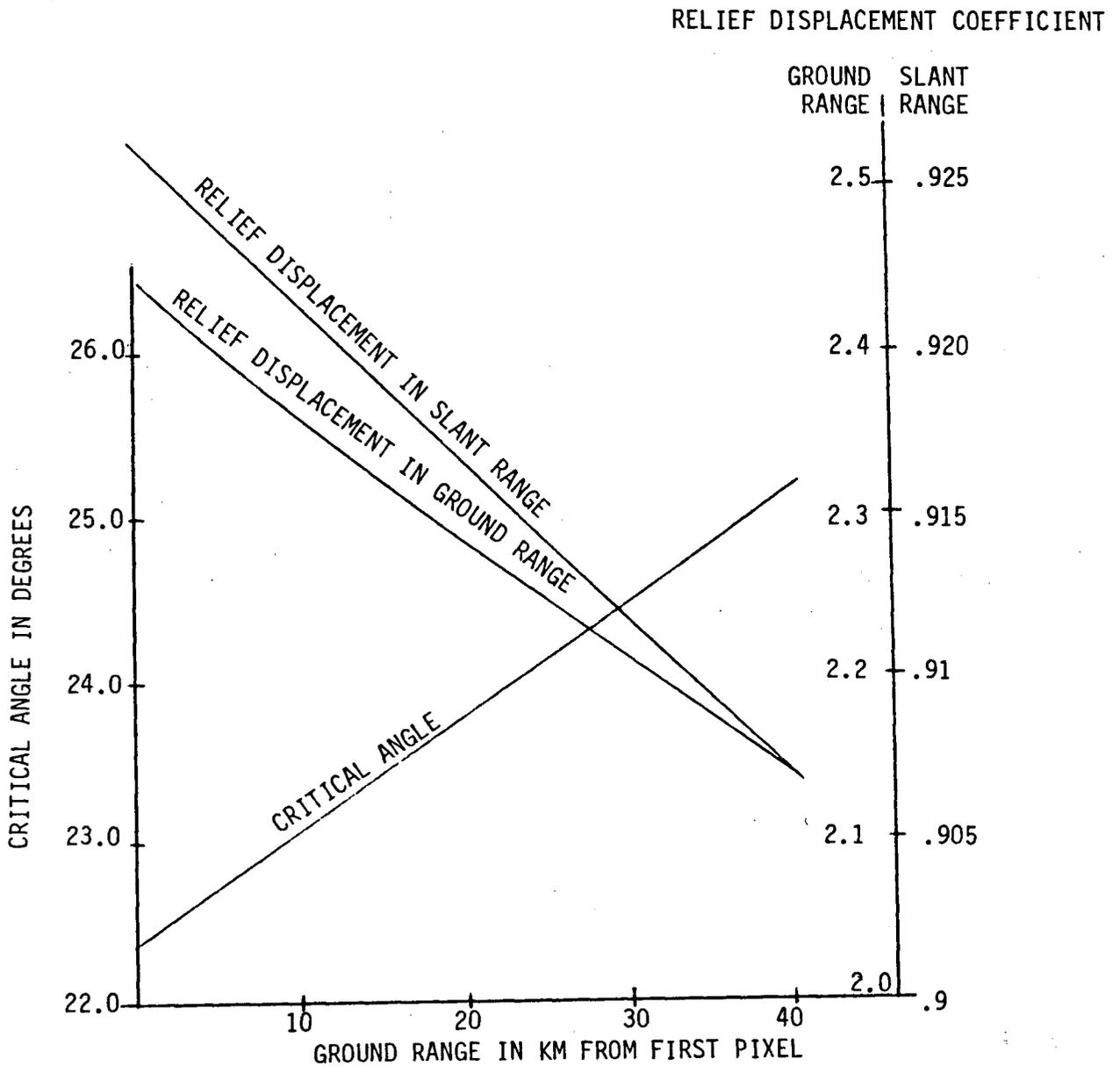


FIGURE 5.4 RELIEF DISPLACEMENT AND CRITICAL ANGLE

NOTE: RELIEF DISPLACEMENT COEFFICIENT IS THE RATIO OF DISPLACEMENT TO TARGET ELEVATION

5.3.3 Target Defocusing

Another problem with terrain is that the trajectory of an elevated point target is incorrect in the azimuth matched filtering process, hence, defocusing the target. The following examines how the trajectory deviates from zero elevation and hence assesses the amount of defocusing.

For Seasat imagery, the worst distortion in slant range r is shown in Figure 5.5 as Δr . This occurs at the tip of the illumination beam. Using the Seasat parameters as shown and nominal orbit parameters, it can be shown that $\Delta r \leq 0.1$ metre per km in elevation. This amount of Δr has no significant effect on range cell migration and hence, produces no significant target blurring.

5.3.4 Intensity Migration

Intensity migration occurs when the same area is imaged in more than one pass. It is caused by a change in incidence angle to the same location due to the relative position of the orbit in the passes. The change in incident angle is illustrated in Figure 5.6.

Intensity migration, although not strictly a geometric distortion, has an adverse effect on the registration between two radar scenes.

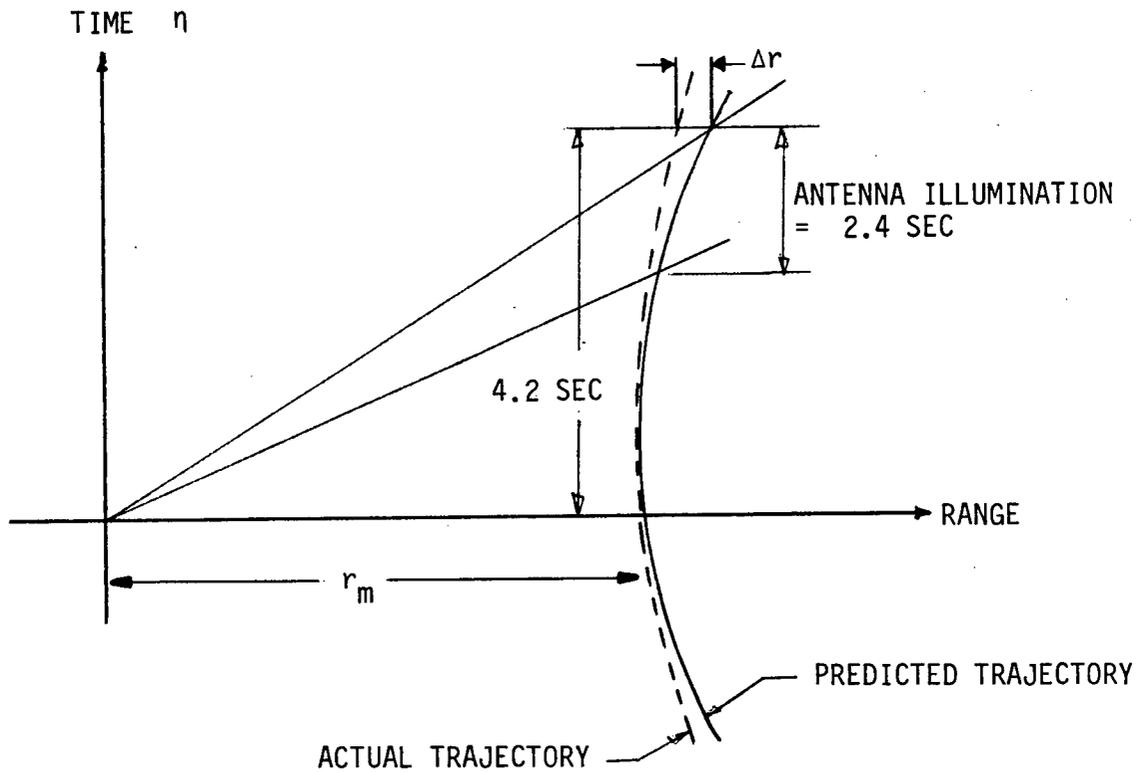
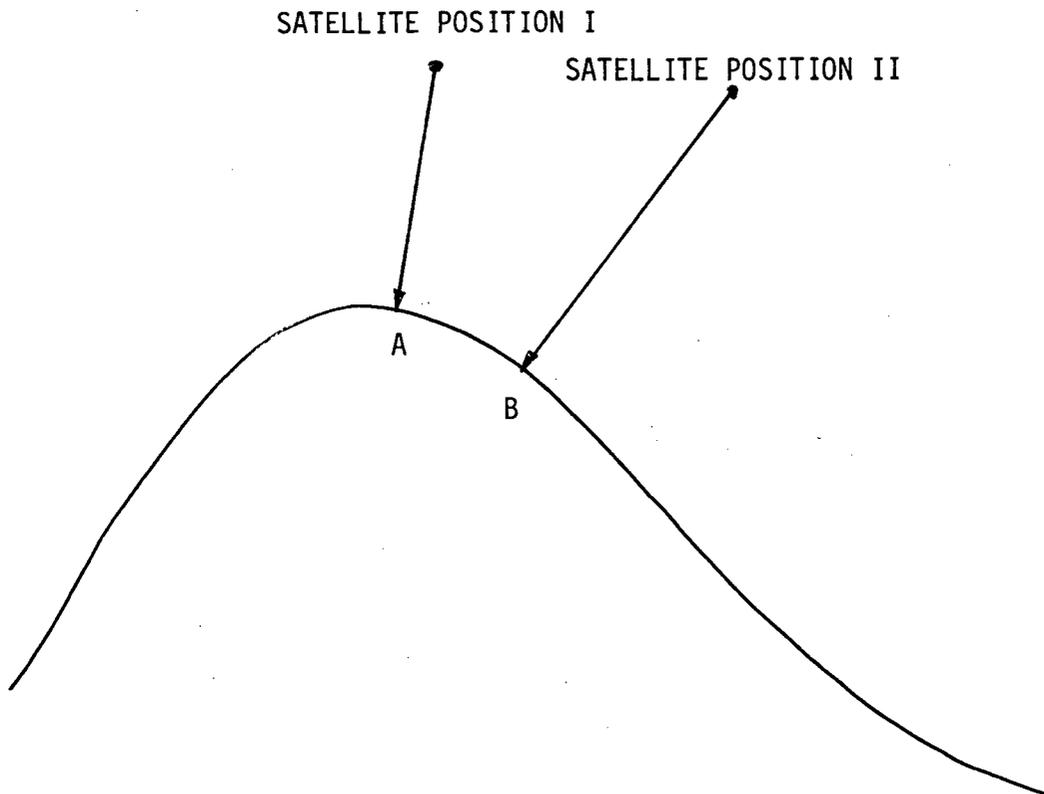


FIGURE 5.5 PREDICTED AND ACTUAL TARGET TRAJECTORIES OF AN ELEVATED TARGET



NOTE: POINT A IS OF MAXIMUM REFLECTIVITY TO
SATellite POSITION I

POINT B IS OF MAXIMUM REFLECTIVITY TO
SATellite POSITION II

FIGURE 5.6 INTENSITY MIGRATION IN RADAR IMAGERY

5.4 Experiments with Seasat Imagery

5.4.1 Image Location Accuracy Study

The purpose of the image location accuracy study is to determine the geometric accuracy in SAR imagery produced by a typical SAR processor (which in this case is the MDA SAR processor) and hence drive a navigation method. The experiment in the study consisted of the following procedures:

- (a) Select a number of clearly identifiable targets in several Seasat scenes (each scene is 40 km X 40 km).
- (b) From topographic maps, determine the location and elevation of these targets.
- (c) Using only the location of the target within an image, as well as Goddard Space Flight Centre supplied Definitive Orbit Record data [JPL77], predict the location of each of these targets, correcting for each target's elevation.
- (d) Determine the means and standard deviations of the differences derived from steps (b) and (c).

The scenes selected were as follows:

- Ottawa, Ontario, Orbit 472, descending,
- Pembroke, Ontario, Orbit 1218, descending, and
- Vancouver Island, B.C., Orbit 193, ascending.

The first two scenes were received at Shoe Cove, Newfoundland while the last one was received at Goldstone, California. The three chosen scenes vary from gently rolling terrain (e.g., Pembroke scene where the terrain

varies between 100 and 200 metres) to very rugged terrain (e.g., Vancouver Island scene where the terrain varies from 0 to 1,000 metres).

The maps used were 1:50,000 Class A1 maps from the National Topographic Series produced in Canada. Each map is on a UTM grid. The accuracy of A1 maps has already been discussed in Chapter 3.

Experimental results are summarized in Table 5.1 which gives the along and across track errors. The table shows two types of errors:

- absolute or mean error with bias in the along and across track directions; and
- relative error after bias removal and expressed in terms of standard deviation.

The across track bias lies below 54 metres for all three scenes. The along track bias varies drastically in the Pembroke scene from the other two scenes. In the Pembroke scene it exceeds 6,000 metres. This substantial bias is believed to result from a timing error in the receiving station clock of approximately one second.

The relative error results from a variety of sources :

- (a) Marking a target line and pixel coordinate in the image. A pixel and a line each corresponds to 12.5 metres on the ground.
- (b) Marking of target on a mapsheet. In a 1:50,000 scale mapsheet, a measurement error of 1/2 mm would correspond to 25

TABLE 5.1 ALONG TRACK AND ACROSS TRACK ERRORS

Scene	No. of Targets	Along track bias in Metres	Across track bias in Metres	Along track σ in metres Experimental/Budgeted	Across track σ in metres Experimental/Budgeted
Ottawa	17	47	35	23/30	38/27
Pembroke	19	6186	54	24/30	38/27
Vancouver Island	12	117	44	23/30	13/29

metres.

(c) Map accuracy. For Class A1 1:50,000 mapsheets, the spatial RMS error is 15 metres in easting and northing directions combined (equivalent to 15 metres in along track and across track combined).

(d) Orbit ephemeris data error. The Definitive Orbit Record data supplied by the Goddard Space Flight Centre contains satellite position information once every minute. The error for each position is [JPL77]:

σ in along track = 17 metres

σ in across track = 10 metres

σ in radial direction = 10 metres

For each scene, a quartic polynomial fit to the satellite position was used to define the satellite position. This can be shown to give an error of less than 1.2 metres in each direction. Hence the ephemeris interpolation error is negligible.

(e) Relief displacement. This has effect only in the across track direction. As has been mentioned, σ for elevation error is 0.3 contour interval for Class A1 maps. On gently rolling areas, the contour interval is 25 feet. On hilly areas, it can be 50 feet or 100 feet. Then for gently rolling areas, σ (Class A1 map) = 2.29 metres. This corresponds to 5 metres in ground displacement, since for every metre in height, relief displacement is approximately 2.3 metres on the ground.

(f) Contour Interval Interpolation error. This again has effect only in the across track direction. Usually a target does not

fall onto a contour line in the mapsheet and the target height has to be interpolated. An interpolation accuracy of 1/4 contour interval is estimated. Thus, for a contour interval of 25 feet and assuming uniform distribution in interpolation error, σ (Class A1 map) = 1.1 metres. This corresponds to 3 metres in relief displacement.

The budgeted overall σ was calculated for each scene and the result is also shown in Table 5.1. The contour intervals of various mapsheets have to be taken into account in the calculation. For the Ottawa and Pembroke scenes, the maps used were with both 25 and 50 foot contours; and for the Vancouver Island scene, the mapsheets were with 20 metre contour interval. The experimental errors agree fairly well with the predicted values for the first two scenes and lie within the predicted value for the third scene.

5.4.2 Orbit Refinement

The transformation described in Appendix G is a function of earth parameters and orbit ephemeris, and is attitude independent. Results in Section 5.4.1 show that the orbit ephemeris has to be refined for precision rectification of SAR imagery.

In Seasat SAR imagery, the orbit latitude ϕ and longitude λ are expressed as quartic polynomials:

$$\phi = \sum_{i=0}^4 \phi_i t^i \quad (5.8a)$$

$$\lambda = \sum_{i=0}^4 \lambda_i t^i \quad (5.8b)$$

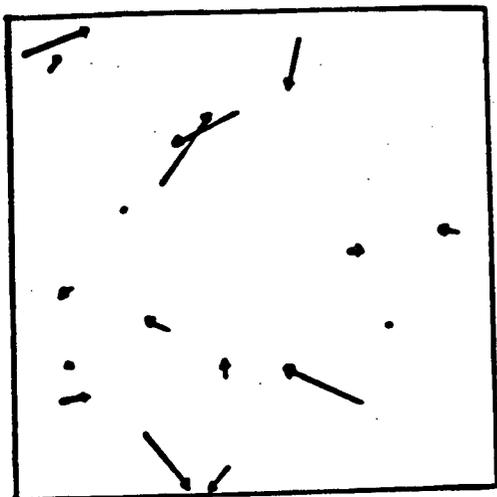
For convenience $t = 0$ is chosen as the image scene centre. Results in the image location accuracy study clearly show that by adding a bias to both ϕ_0 and λ_0 , the resulting rectified image has a residual error in the along and across track directions of about 40 metres in which measurement error is included. Therefore, the true rectification error, free from map induced measurement error, is less than 40 metres in the two directions.

An even simpler way is to translate the image with respect to the map projection grid lines. The worst translations occur in the Pembroke scene (about 6 km in the along track direction). The resulting image will still have the accuracy cited as above.

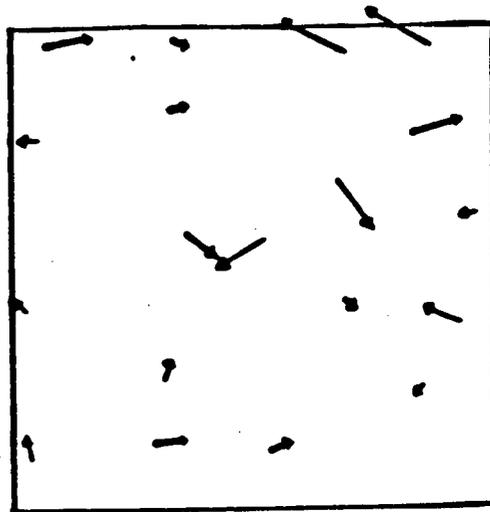
Although the two methods (adding biases ϕ_0 and λ_0 and image translation) are not mathematically equivalent, it can be shown that the differences are negligible over a 40 X 40 km² scene.

Such a method of rectification of Seasat SAR imagery is possible due to the high degree of accuracy in tracking data supplied by the Goddard Space Flight Centre. The along track error and across track error are most probably due to timing errors and hence can be removed by an image translation in the manner described above as shown in Figure 5.7. These plots do not exhibit any residual height error or

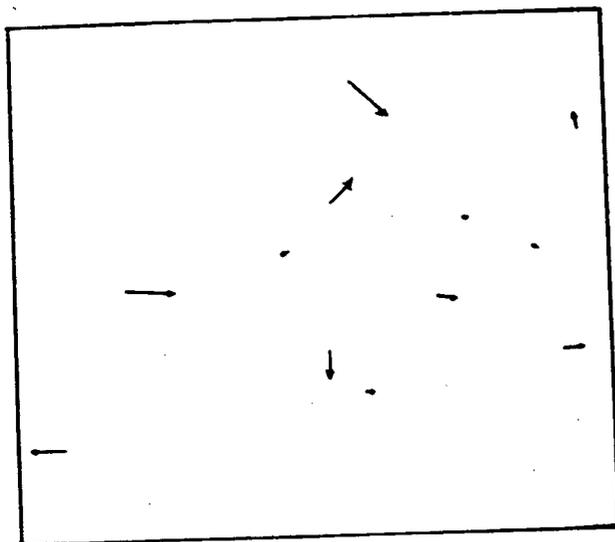
(A) OTTAWA SCENE



(B) PEMBROKE SCENE



(C) VANCOUVER ISLAND SCENE



50 m



FIGURE 5.7 GCP ERROR VECTORS AFTER NAVIGATION

residual image rotation, hence justifying the translation process.

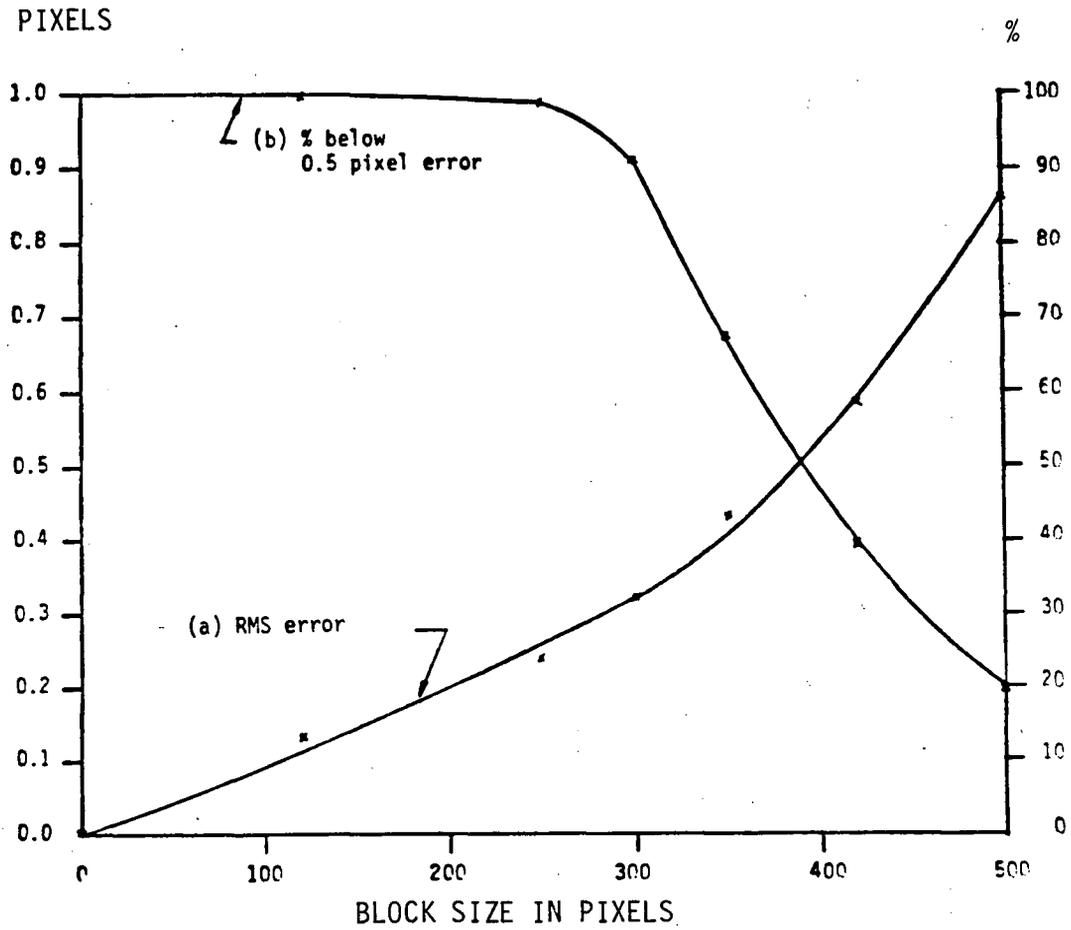
For SAR imagery in general, orbit refinement can be done by a least squares filtering technique similar to that presented in Section 4.3 for scanner imagery. However, attitude angles should not be entered in the state vector. For typical SAR imagery the parameters are satellite latitude, longitude and local orbit radius.

5.4.3 Rectification with DTM

The Vancouver Island scene used in the image location accuracy study is also used as an illustrative example of how a DTM can be used in the rectification. This scene is chosen because it is imaged over a very rugged area and a DTM is available. The DTM covers an area of 16 km X 20 km and has spatial resolution of 60 metres X 60 metres. It was obtained from a triangulated network [PENC78 and HEIL78] on a 1:50,000 Class A1 topographic map with a 20 metre contour interval. More data points were digitized in steep areas than relatively flat areas. About 5,000 points were manually digitized.

The navigation process has been performed in the image location accuracy study.

The remapping and interpolation was performed in three one-dimensional passes and the northing direction in the resulting image was to be aligned with the vertical. The block size versus remapping accuracy is shown in Figure 5.8. It was chosen to be 100 pixels X 100



NOTE: PIXEL SIZE IS 12.5 METRES

FIGURE 5.8 BLOCK SIZE VERSUS REMAPPING ACCURACY

lines in the first and second passes, resulting in a 0.1 pixel RMS remapping error in the final product and all pixels lie within 0.5 pixel remapping error.

The original image is shown in Figure 5.9. The image appears to be "turned around" since it was acquired in an ascending pass. Figure 5.10 shows the rectified image with the use of the DTM while Figure 5.11 shows the rectified image without using the DTM. The sharp boundaries in Figure 5.10 show where the DTM ends. Note that in the last two figures the north is not aligned with the northing axis due to a nonzero convergence angle [MALI73] in the UTM projection.

The rectification error, by measuring several control points on the lakes in the mountainous areas, was found to be 35 metres RMS. This is within the error budget considering GCP marking error and map accuracy.

A comparison of the last two figures clearly demonstrates the advantage of using a DTM. Many areas in the rectified image in Figure 5.10 appear to be orthographic while in Figure 5.11 the same areas appear to be shifted towards the radar.

The "noisy" nature in areas with steep slope is due to relief layover and slope foreshortening effects. A very narrow band of bright pixels (facing the radar) in the raw image has to be mapped to a wide area. Furthermore, speckles are present within this narrow band. The sinc function interpolant kernel does not take slope into account. The removal of radiometric effects due to slope is left as a further research topic.

H10W1Z0

GROUND RANGE

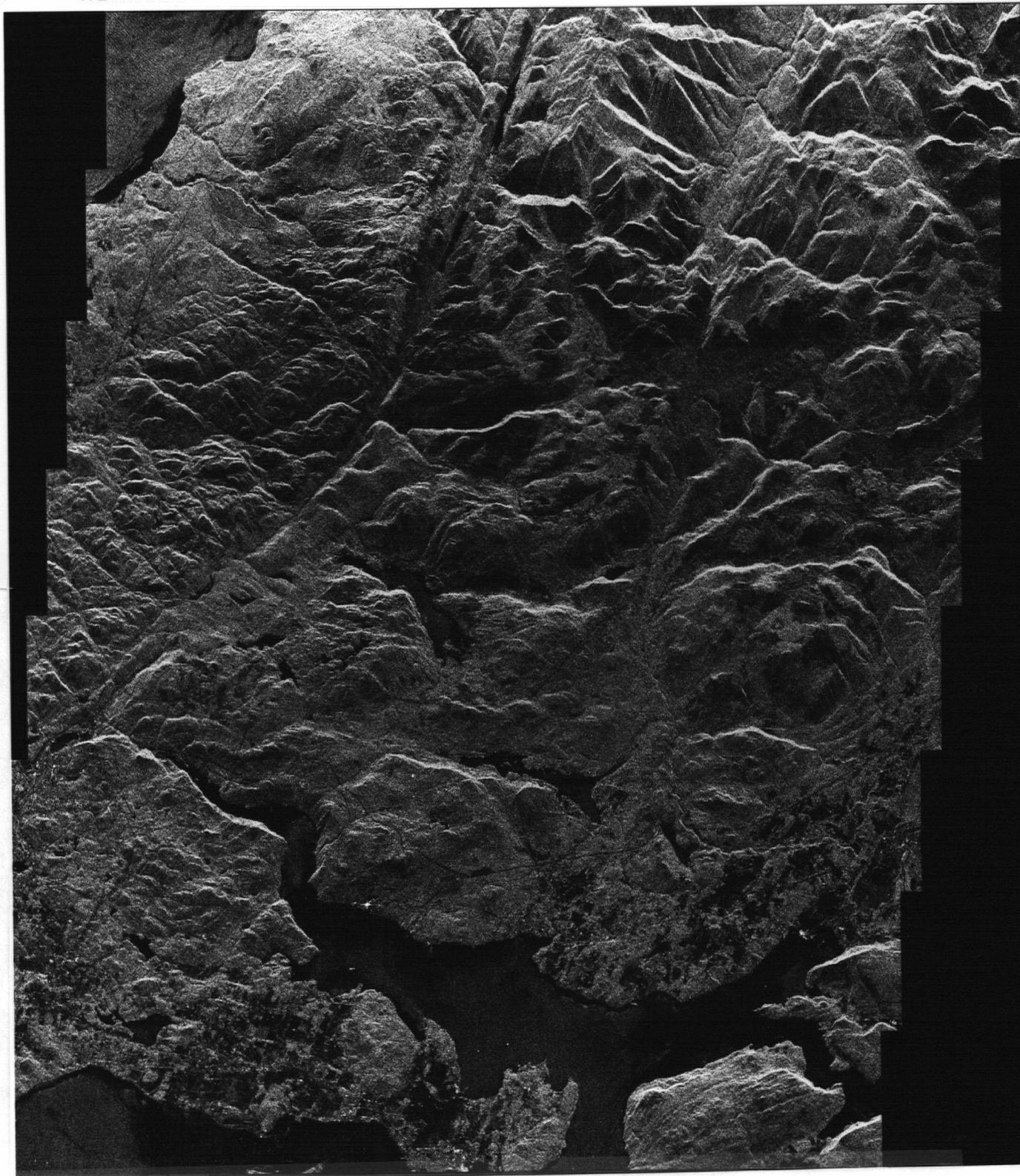


FIGURE 5.9 RAW SEASAT IMAGE

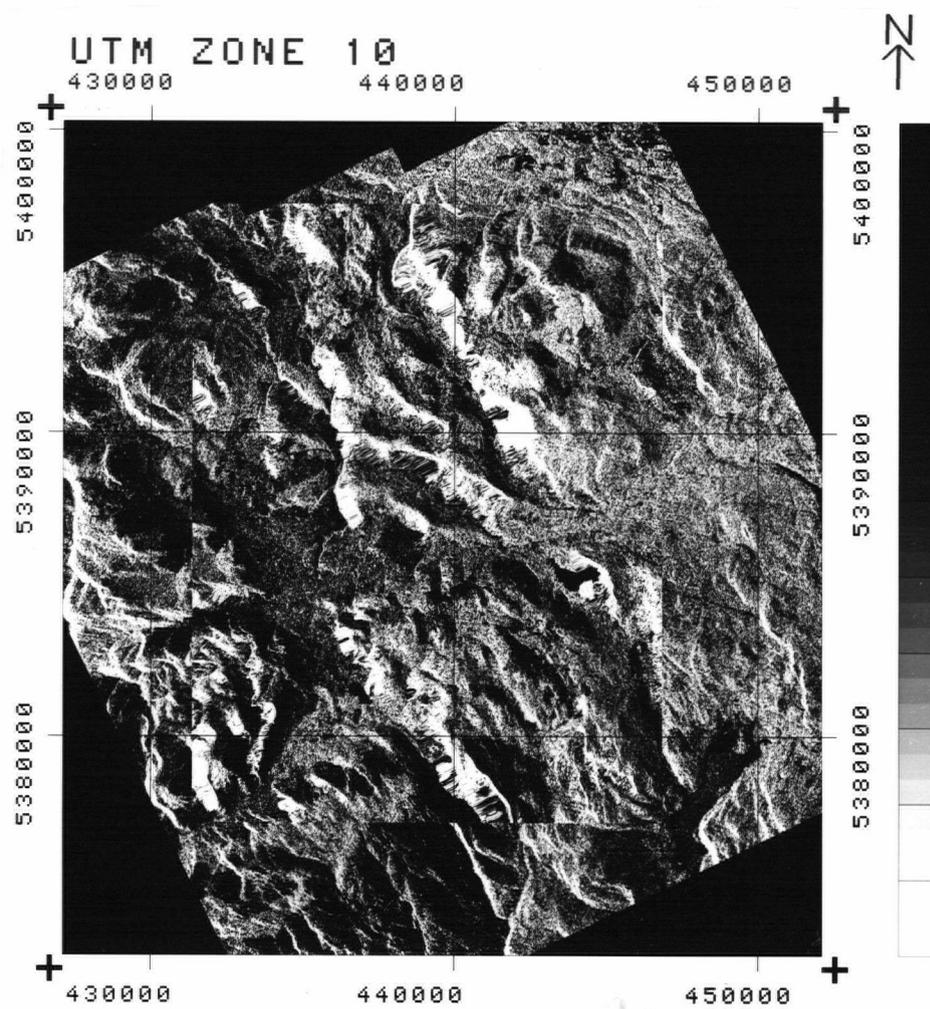


FIGURE 5.10 RECTIFIED SEASAT IMAGE WITH DTM

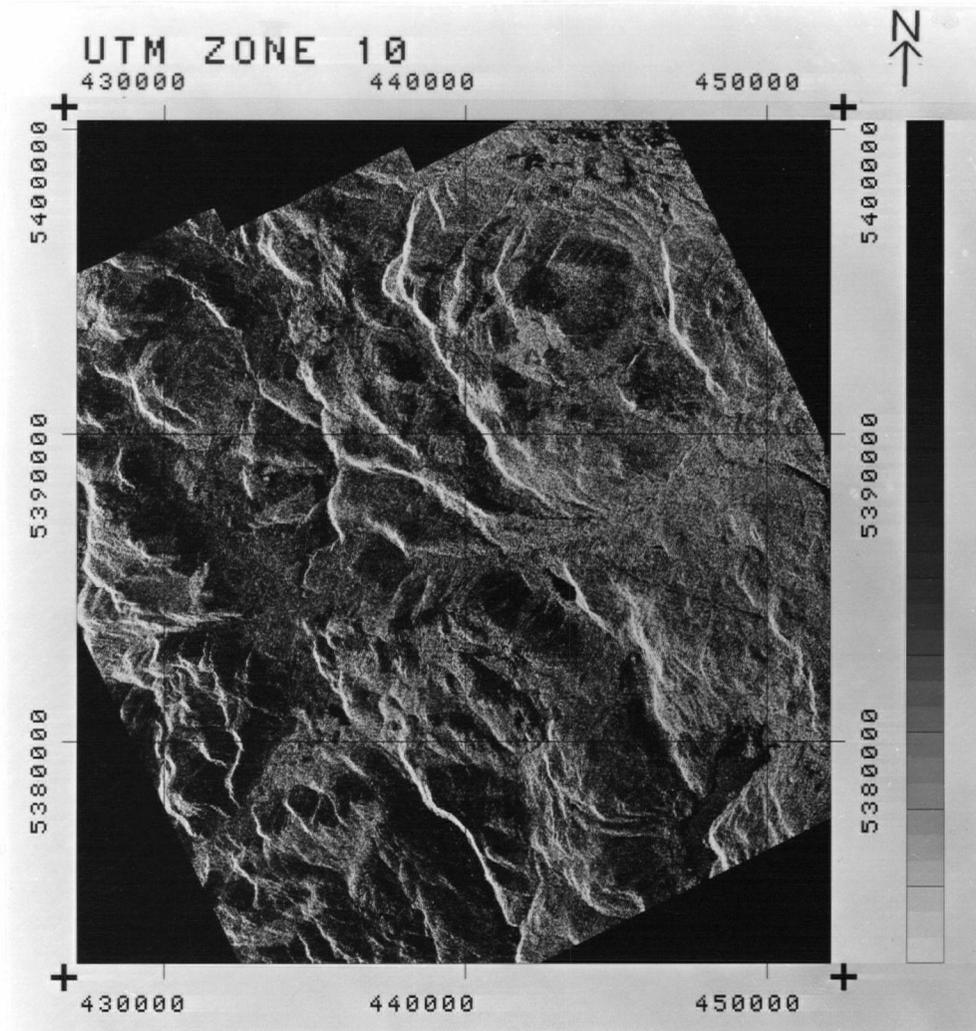


FIGURE 5.11 RECTIFIED SEASAT IMAGE WITHOUT DTM

An undesirable geometric error resulting from using the DTM in a steep area is illustrated by Grant Lake (coordinates: 443,000 Easting and 5,385,000 Northing). Its potato shape appearance in the raw image has changed to an L shape in the rectified image. The protruding part corresponds to an area of extreme high slope of 38° as measured from the given topographic map. This has far exceeded the critical slope angle shown in Figure 5.4, thus relief layover has occurred.

The rectification error in this part is due to DTM inaccuracy. Experimental results show that the remapping error is about 60 metres and this corresponds to an elevation error of 26 metres (since the relief displacement is about 2.3 metres per metre height). Indeed, some DTM values are found to differ from the topographic map by as much as 20 metres in the hills around Grant Lake. The DTM error is due to the following factors:

- (a) Digitization error. In a 1:50,000 map with 20 metre contour interval, 4 mm crosses 2 contour intervals at 38° slope. Hence, a 0.5 mm measurement error would correspond to 20 metres.
- (b) Interpolation error. The initial DTM was set up in a triangulated network and then interpolated to a regular DTM. In a steep area, the initial DTM must therefore be dense enough to retain accuracy in the regular DTM.
- (c) Water level. It is found that the width of the lake in the image differs from that in the map. This may be caused by seasonal changes in the water level.

5.5 Navigation by Automatic Focusing

A key feature in the rectification of SAR imagery is the possibility of refining satellite trajectory data without using GCPs. This method is automatic focusing.

In both range and azimuth, the accuracy of focusing is dependent on the precision to which the matched filter required for compression can be determined. In azimuth compression the FM rate is a function of the earth spheroid and satellite orbit parameters. The uncertainties in these parameters will introduce an error in FM rate and hence will not bring a target into focus. An example of azimuth misfocus is shown in Figure 5.12 where the SAR data is processed first with the correct FM rate and then with an FM rate off by 10%. Distortions in the FM rate with respect to various parameters have been analyzed for Seasat [MDA77] and are presented in Table 5.2. For two geographic locations. Focusing targets in the imagery can thus provide a way to refine the parameters.

One technique is to process the imagery with the nominal value of the FM rate, so that discrete strong targets can be identified. These targets are processed again by varying the FM rate aiming at minimizing the half power width of the responses and maximizing the peak magnitudes of the selected targets.

Another technique takes advantage of the fact that registration between azimuth looks, in multi-look processing, is very sensitive to



(A) PROCESSING WITH CORRECT FM RATE



(B) PROCESSING WITH FM RATE DISTORTED BY 10%

FIGURE 5.12 EFFECT OF AZIMUTH MISFOCUSING
(Courtesy of Mr. P. Hasan and Mr. P. McConnell of MDA)

TABLE 5.2 SENSITIVITY OF FM RATE TO
PARAMETER VARIATIONS IN
SEASAT IMAGERY

	% Change in FM Rate		Units
Target location	60°N 58°W	0°N 96°W	
Target latitude	4.62	3.46	/deg
Orbit semi-major axis	20.52×10^{-6}	15.62×10^{-6}	/m
Orbit eccentricity	121.73	0.0	
Satellite angular frequency	3250	3269	$\frac{(\text{deg})}{\text{sec}}$
Orbit inclination	0.32	0.22	/deg
Earth semi-major axis	8.85×10^{-6}	13.85×10^{-6}	/m

variation in the azimuth matched filter FM rate [BENN78]. Hence the FM rate can be tuned until the looks are registered.

From the refined FM rate at different positions, the orbit trajectory parameters can be refined by, again, a recursive estimator as shown in Figure 5.13. Detailed study of this method is left for future research.

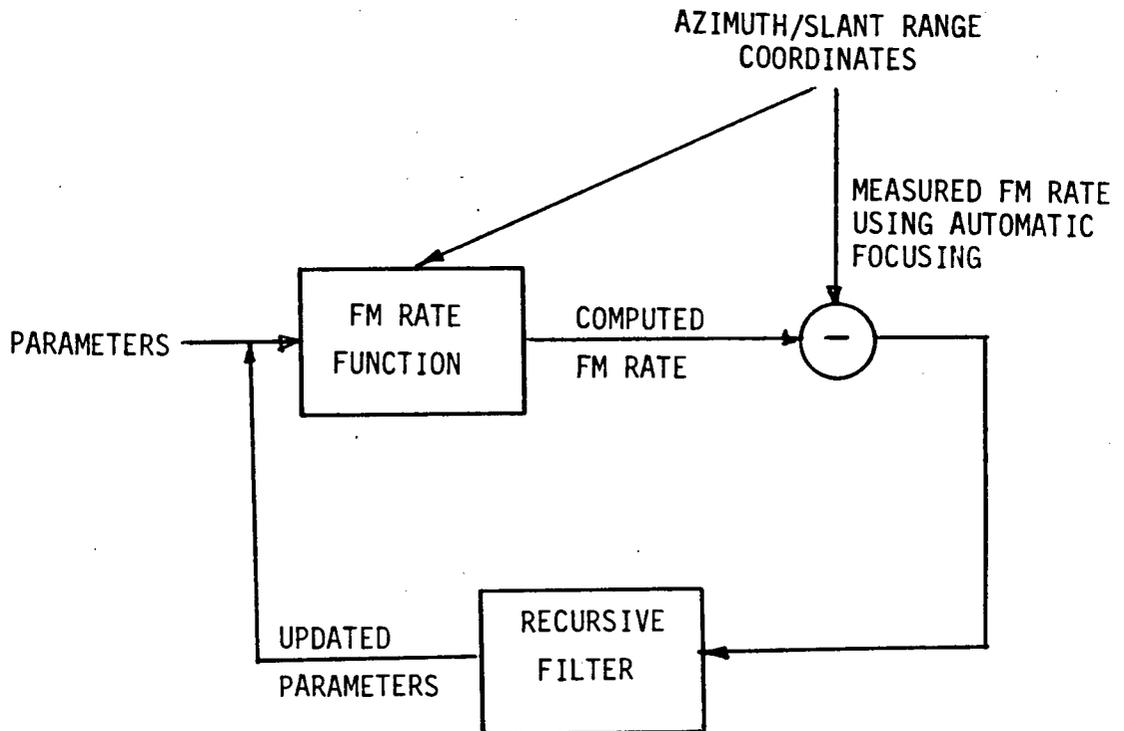


FIGURE 5.13 NAVIGATION LOOP USING AUTOMATIC FOCUSING

CHAPTER 6 APPLICATIONS

6.1 Introduction

This chapter presents some applications of precision rectified satellite imagery and an application of the rectification method to a future planetary mission.

Applications of precision rectified imagery presented here include multisensor integration and image mosaicking. These two examples merely illustrate the potential of rectifying imagery to a common data base, and are by no means exhaustive instances in the applications area.

The rectification method can be extended to a planned future NASA planetary exploration mission - Venus Orbiting Imaging Radar. The lack of GCPs will make the automatic focusing method promising in the rectification process.

6.2 Multisensor Integration

A Synthetic Aperture Radar (SAR) can produce very high resolution imagery which contains fine details in surface roughness. On the contrary, Landsat imagery is of lower resolution but contains information about surface reflectivity in different spectral bands. Since Landsat and SAR imagery complement one another in this way, their composite would contain very rich and useful information for image interpretation.

With the availability of spaceborne SAR imagery (Seasat), the production of Landsat/Seasat colour composites becomes feasible. Such products may be produced routinely by a future satellite system incorporating both an MSS and SAR on the same platform. In this thesis an example is shown by a Landsat/Seasat colour composite image of the Ottawa area in Canada, precision rectified to UTM coordinates. Basically, both the Landsat and Seasat images are registered to UTM coordinates, the procedures of which have been described in previous sections. Then the intensity of the black and white Seasat image is used to modulate the false colour Landsat image (made from bands 4, 5 and 7). The resulting image can be viewed on a video display or written to film by a film recorder.

Figures 6.1 to 6.3 are pictures taken directly from a 512 x 512 digital display. The dynamic range of the display is 8 bits. Each pixel is 25 x 25 metre².

The Seasat image is shown in Figure 6.1 and the Landsat image in Figure 6.2. The Seasat image is used to modulate the Landsat image. The modulation is justified by the fact that image intensity I_0 is related to the surface albedo ρ by a multiplicative rule:

$$I_0 = K \rho$$

where K is a function of imaging geometry, atmospheric parameters and wave length etc. The picture in Figure 6.3 is produced in the following modulation way:



FIGURE 6.1 SEASAT COMPONENT IN LANDSAT/SEASAT COMPOSITE

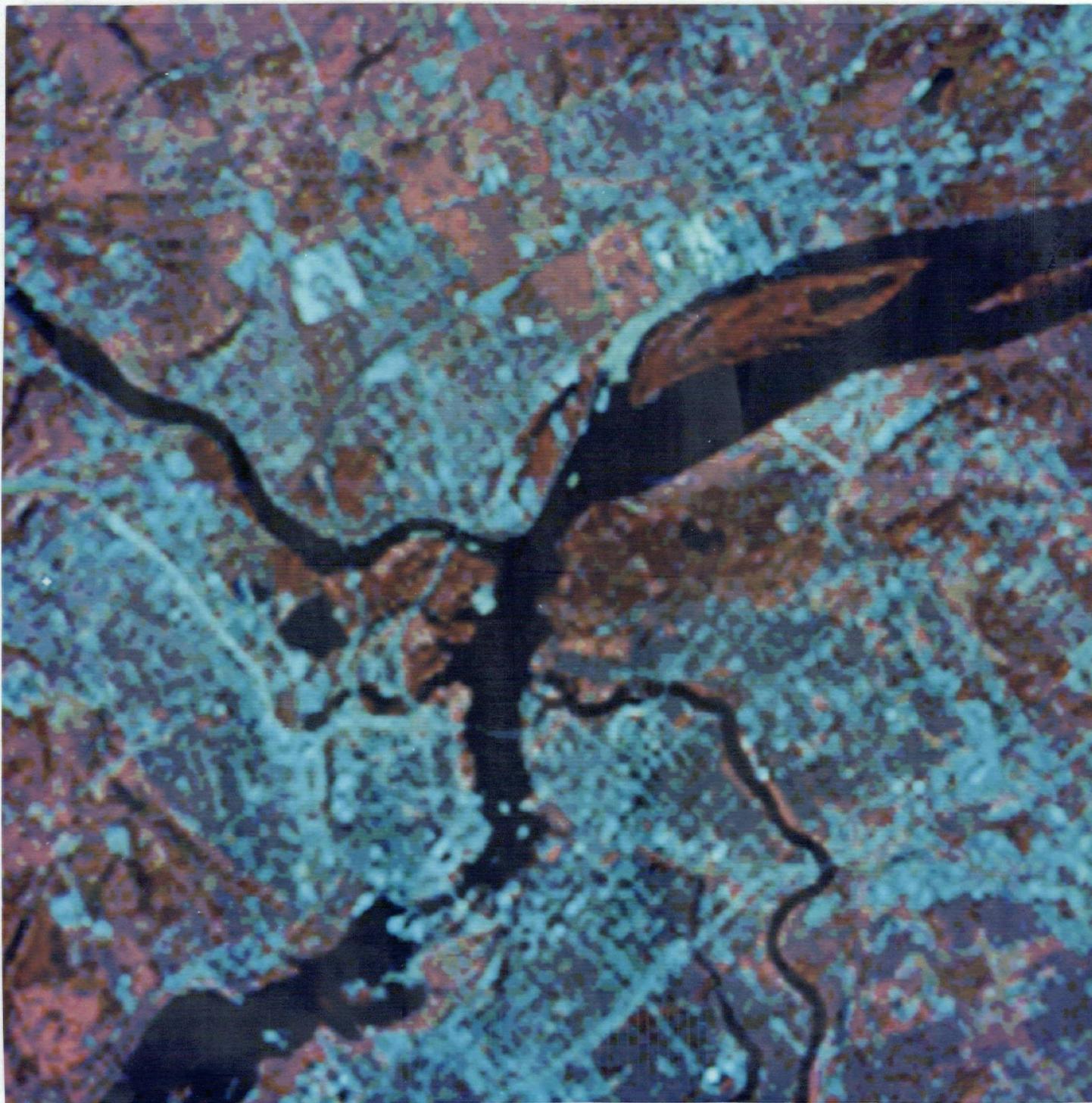


FIGURE 6.2 LANDSAT COMPONENT IN LANDSAT/SEASAT COMPOSITE

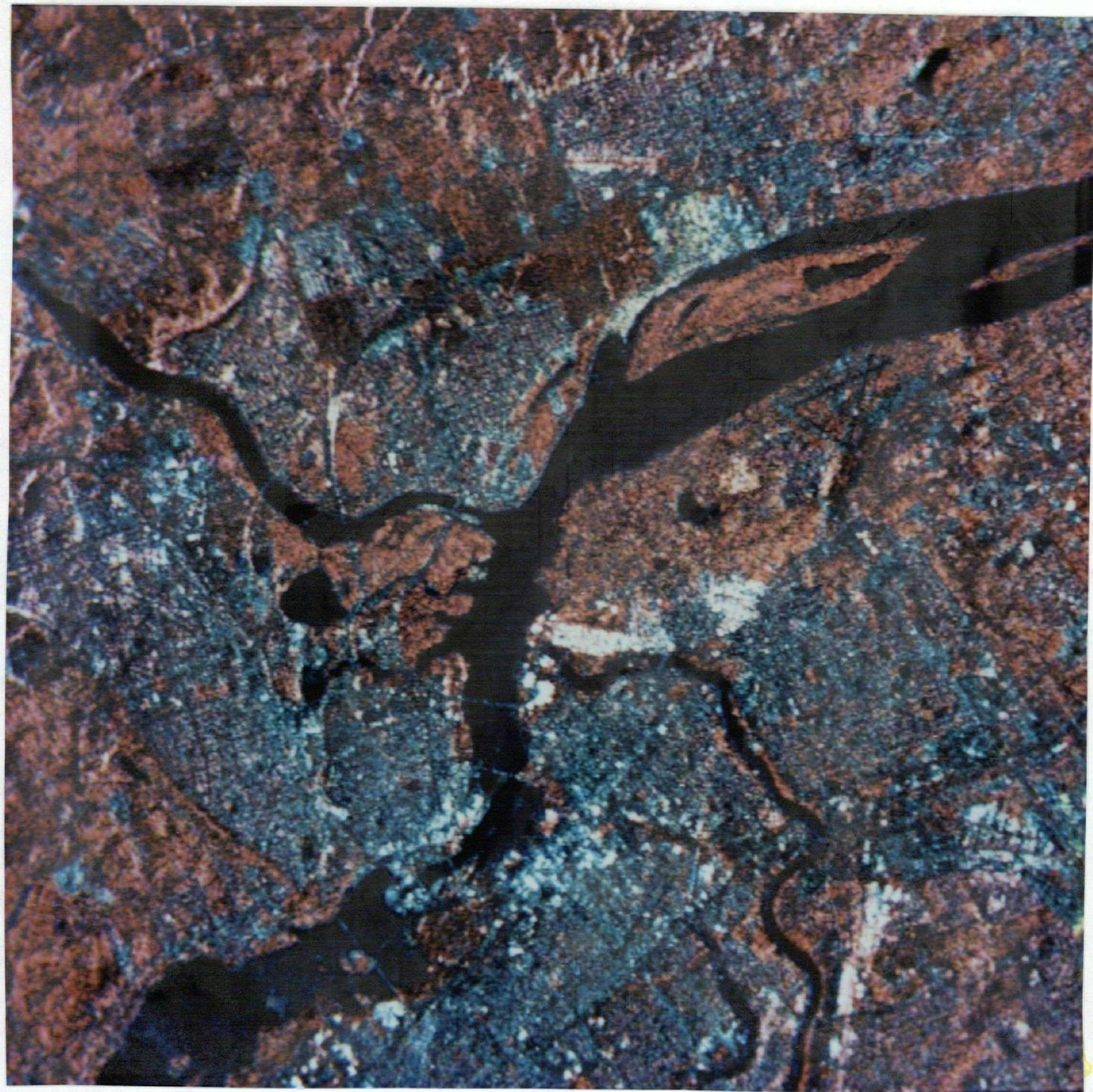


FIGURE 6.3 LANDSAT/SEASAT COMPOSITE (12 KM X 12 KM)
OF THE OTTAWA AREA

$$\text{Band } n = \sqrt{\text{Band } n \text{ of Landsat} \cdot \text{Seasat}}$$

where $n = 4, 5, 7$. Therefore, in each case 3 intermediate products were made. The colour composites were made with the following colour assignment in each case:

Band' 4 blue

Band' 5 green

Band' 7 red

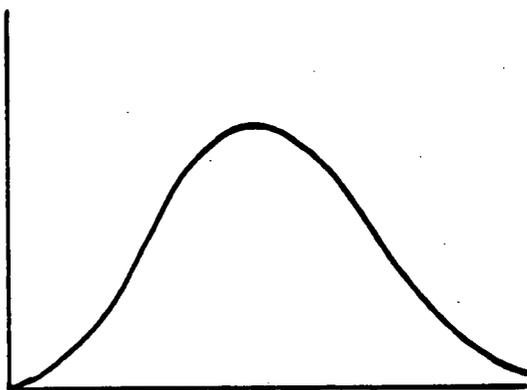
Also for each Band' n , its histogram is used as a guide to set up an intensity transfer function so that the maximum dynamic range can be utilized.

The geometric mean is used for the following reason. Let two random variables be of the same mean M_0 . Their product then has a mean $M_1 < M_0$, while the geometric mean has a mean of $M_2 = M_0$ as illustrated in Figure 6.4. The geometric mean composite then has the same "colour balance" as the Landsat. Note that the geometric mean is equivalent to summation in the logarithm domain:

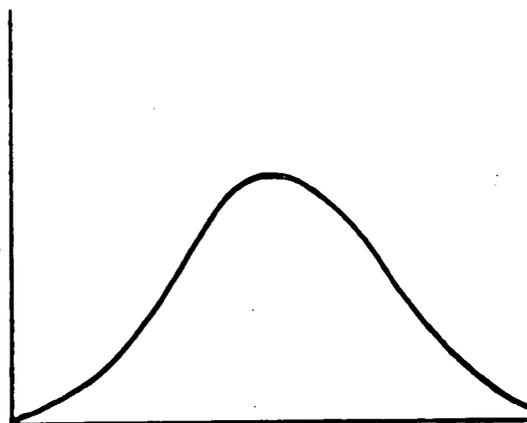
$$\log(\text{composite}) = 0.5 [\log(\text{Seasat}) + \log(\text{Landsat})]$$

Figures 6.1 and 6.2 give a comparison between the Seasat and Landsat images. The Landsat image is of low resolution but contains good surface reflectivity information since the Landsat sensors operate in a region of spectral bands. On the other hand, the Seasat image is of

HISTOGRAM
OF
VARIABLE X



HISTOGRAM
OF
VARIABLE Y



HISTOGRAM
OF
X . Y
AND
 $\sqrt{X.Y}$

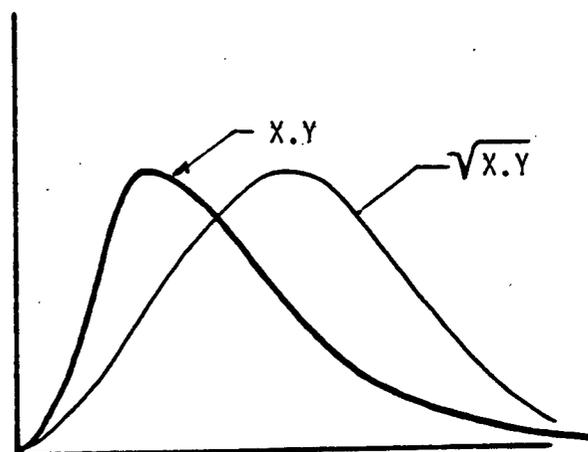


FIGURE 6.4 HISTOGRAM OF PRODUCT OF TWO RANDOM VARIABLES

high resolution and contains good surface texture information¹. By combining the two, as shown in Figure 6.3, the advantages of both sets of sensors are integrated into one image. The colour composite shows bridges, roads, highways and terrain prominently whereas these structures are not as clear in the Landsat image. In addition, the colour composite has spectral reflectance content not present in the Seasat image alone. To suit various applications such as forestry, geology or urban planning, the colour assignment and the dynamic range of individual components may have to be varied.

In particular several places of interest are (see Figure 6.5 for region number):

Region 1: The inner structure of the square can be seen here and the square as a whole is clearly noticeable in the composite, but this is not very noticeable in the Seasat image. While it is also noticeable in the Landsat image, it does not show high resolution.

Region 2: The Ottawa River here is not very clear in the Seasat component.

Region 3: The target here is not visible in the composite. This is expected since this target only appears in the Seasat image and not in the Landsat image.

Region 4: The Seasat image saturates in this area (due to corner reflectors formed by the downtown Ottawa area) but the Landsat image does not. The colour composite also saturates in this area because a linear stretch function

¹ One example is that Seasat imagery shows geological fault lines prominently.

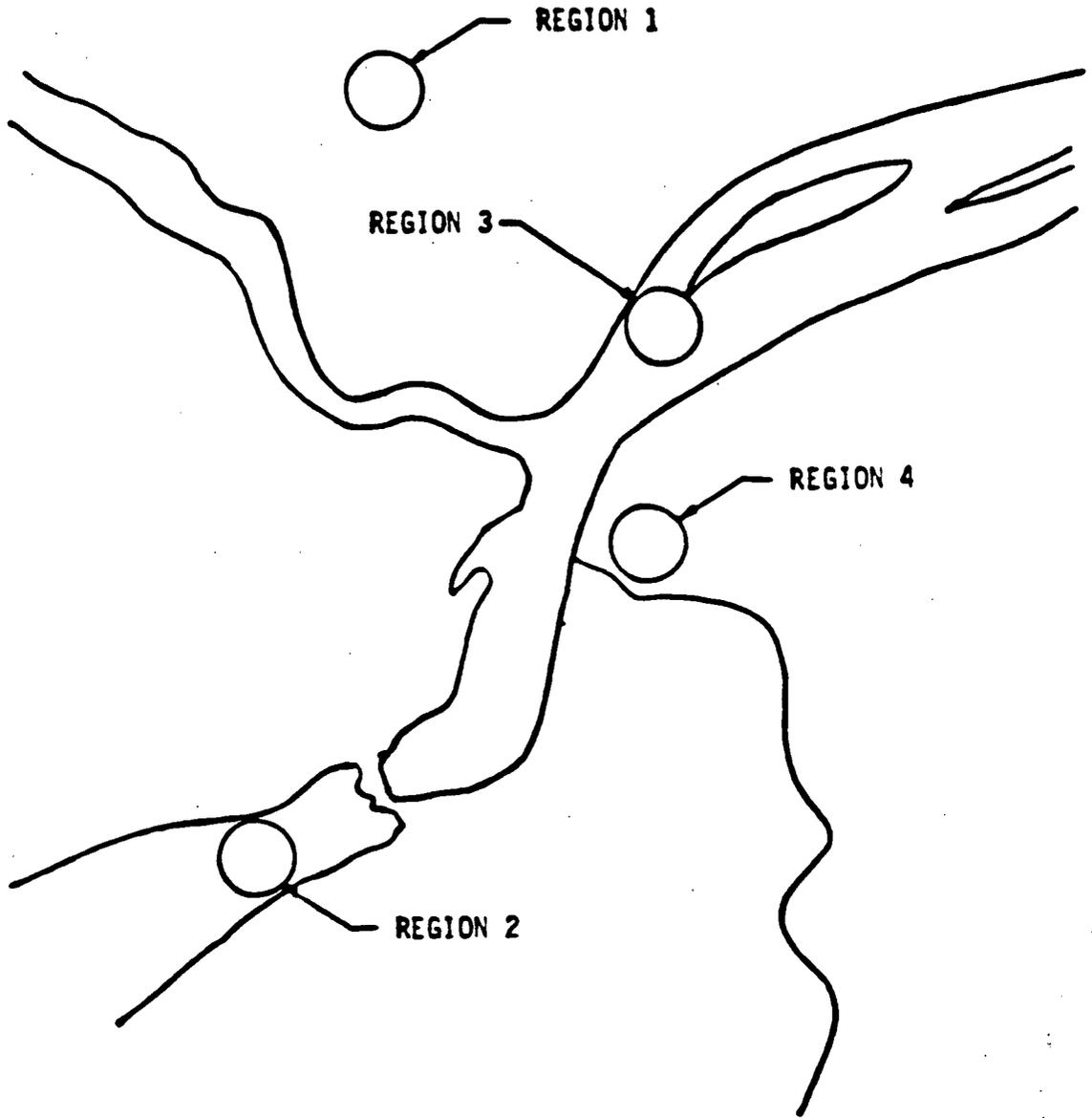


FIGURE 6.5 REGIONS OF INTEREST IN THE COMPOSITE

with clipping at high intensity levels is used in producing the composite. Saturation can be prevented in the composite by using the cumulative distribution function as the transfer function.

General: As has been mentioned, the composite shows bridges, roads, highways and downtown Ottawa very prominently.

"Disappearance" of targets in the geometric mean composite (Region 3) can be avoided if the composite is done by:

$$\text{Composite} = \sqrt{(a_1 \cdot \text{Seasat} + a_2) (a_3 \cdot \text{Landsat} + a_4)}$$

where a_i ($i = 1$ to 4) are suitably chosen coefficients. In other words, a linear stretch is applied to each image first before taking the product. By examining the terms of the expression, the composite contains multiplicative as well as additive terms.

A composite of a larger area around the Ottawa Valley is shown in Figure 6.6. Each pixel covers an area of 12.5×12.5 metres² in this composite.

Another example is the Vancouver Landsat/Seasat composite as shown in Figure 6.7 (Seasat) and Figure 6.8 (composite). The rectified Landsat image has been shown in Figure 4.53. The area corresponds to NTS mapsheet number 92G/3 which is shown in Figure 6.9. This example demonstrates the benefits of rectifying satellite imagery to a common data base, e.g., to a conventional mapsheet numbering system in the following

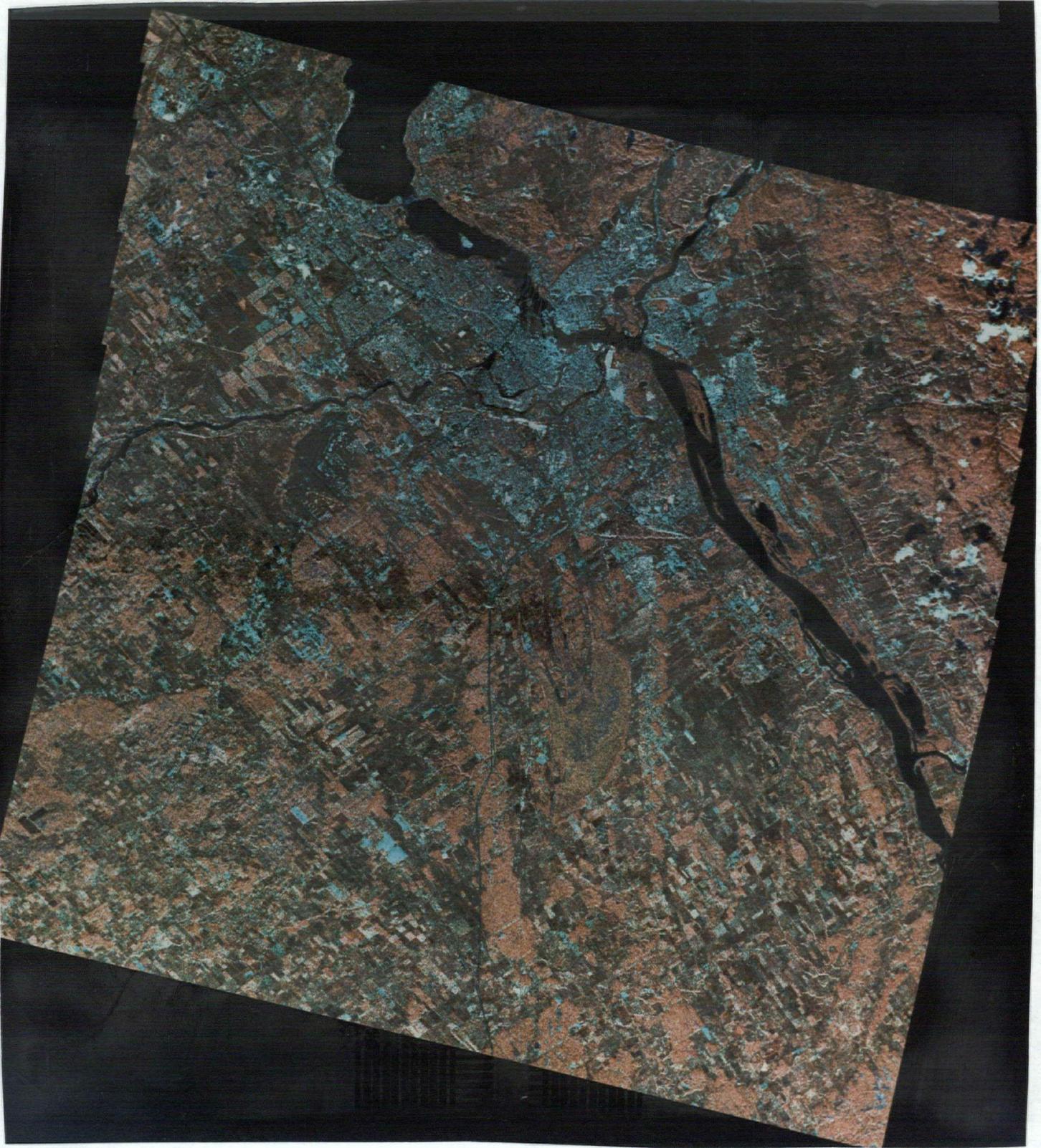


FIGURE 6.6 LANDSAT/SEASAT COMPOSITE (40 KM X 40 KM)
OF THE OTTAWA AREA



FIGURE 6.7 SEASAT IMAGE OF VANCOUVER, NTS MAPSHEET 92G/3



FIGURE 6.8 LANDSAT/SEASAT COMPOSITE OF VANCOUVER, NTS MAPSHEET 92G/3

sense. The rectified image can easily be overlaid to an existing mapsheet. In addition, an "image map" in which topographic information from a mapsheet is overlaid onto the image can be made.

6.3 Image Mosaicking

An application of rectified imagery in neighbouring passes to the same data base is the ability to form mosaics. This subsection discusses different methods of mosaicking and presents an example of image mosaicking.

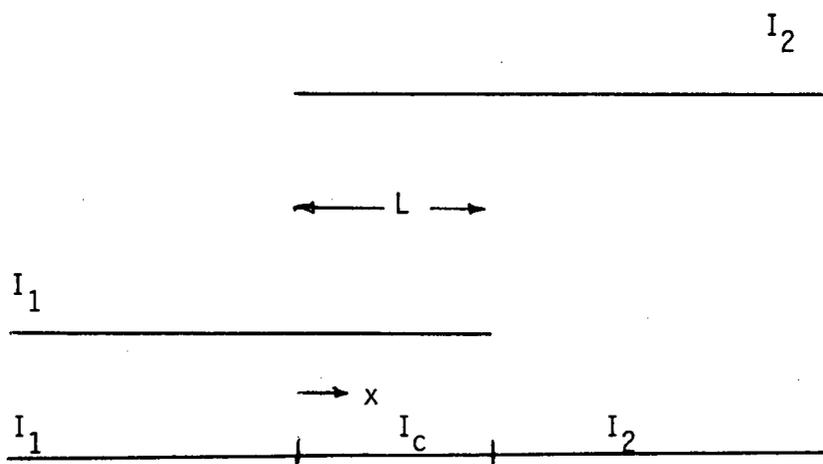
6.3.1 Linear Interpolation

Visual discontinuities will occur if two images are joined together without some kind of smoothing operation at the joint. This is because the local histograms for the two images at the joint are not balanced even though the histograms for both images may be balanced. Smoothing at the joint can be done by linear interpolation of the intensities in the overlap region.

Let the image intensities be denoted by I_1 and I_2 and the width of the overlap area denoted by L as shown in Figure 6.10. Then the intensity in the overlap area is given by:

$$I_c(x) = [(L - x) I_1(x) + x I_2(x)] / L \quad (6.1)$$

where x is defined in the figure. Equation 6.1 produces a gradual transition or ramp effect.



INTENSITY IN THE OVERLAP AREA GIVEN BY:

$$I_c = \frac{L - x}{L} I_1(x) + \frac{x}{L} I_2(x)$$

FIGURE 6.10 RADIOMETRIC MOSAICKING BY LINEAR INTERPOLATION

The disadvantage of this method is image degradation. The ramp slope determines how many points are affected by the smoothing process. A more gradual slope conceals the discontinuity better, but it also has the adverse effect of degrading information content in the vicinity of the joint. Another disadvantage is that features that appear in one image but not the other (e.g., clouds) seem to fade away miraculously. However, this method is fast and easy to implement and gives good visual results at the boundary.

6.3.2 Feathering

In this method, an "optimal" seam point is located for each image line, hence the seam edge zigzags across the mosaicked image. This method is proposed by Milgram [MILG75] and consists of three steps. The first step is to equalize the gray level statistics (mean and standard deviation) between the two images. The second step is to locate the seam point within the overlap region. In the mosaic, one image fits to the right of the seam point and the other image to the left of the seam point. The third step is to smooth across the seam using a ramp function.

To equalize the gray level statistics in the overlap area, those regions containing features that occur in the one image but not the other should be excluded in computing the gray level transform. For example, clouds, snow or ground water present in one image but not the other should be excluded.

The next step is to choose a seam point for each mosaic image

line. This choice is based on the minimum local difference in the two images in the overlap area. The seam point selection for each mosaic image line is based on the minimum local difference in the overlap area. A difference is formed at each pixel location in the overlap area, and the pixel at which the minimum difference occurs is chosen as the seam point.

The final step is seam smoothing. This is again linear interpolation with a chosen ramp function width.

Using this method of seam point selection, one finds a succession of points whose horizontal positions are unrelated to one another. This random positioning of the seam is known as "feathering", since it seems that one image is "feathered" into the other in the mosaic. This method has the advantage of being able to track around clouds or other features present in one but not in the other.

This method has the disadvantage of introducing discontinuities in the vertical direction. A refinement would be to restrict the range of candidate seam points depending on the magnitude of the minimum local difference on the previous line. Milgram also proposes that the range of candidate seam points should be restricted from line to line.

An extension to refine the seam point locations has also been proposed by Milgram [MILG77]. He defines a cost function as the sum of minimum local differences for all lines. The best seam through the mosaic is the least cost path.

This algorithm suffers from the fact that the operator has to input the value of overlap width where the minimum local difference is to be obtained, an expression for range of a candidate seam point and the smoothing ramp function width. An added disadvantage is that pixel intensities of one scene have to be adjusted first (for gray level balance) and this sometimes may not be desirable.

6.3.3 Filtering

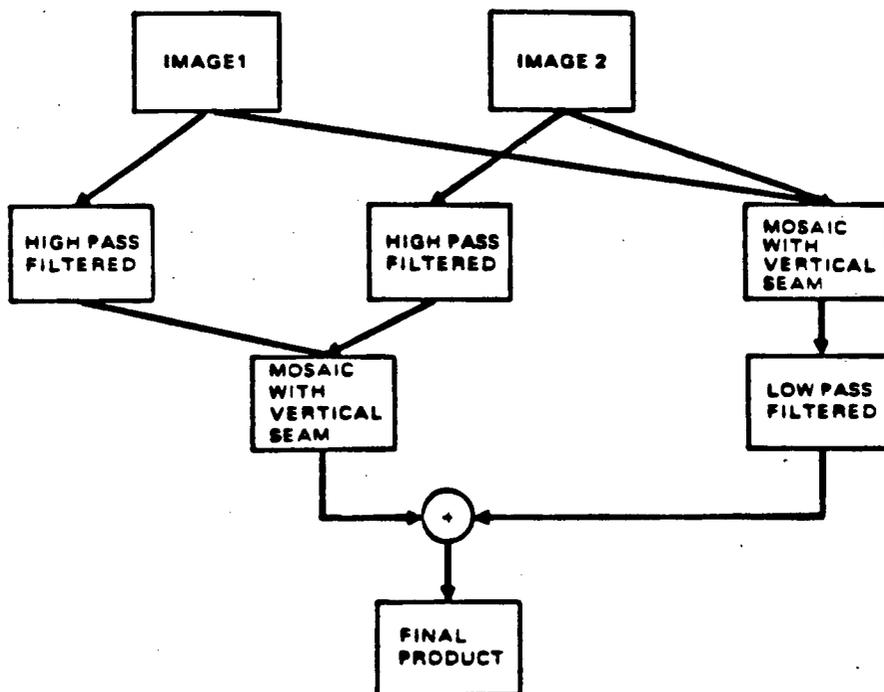
A filtering method has been employed by the USGS for several years. The steps for a two-frame mosaic, depicted in Figure 6.11a, are:

1. Mosaic the two images with a vertical seam.
2. Low pass filter the mosaicked image.
3. High pass filter the original two images and mosaic them with the same vertical seam as in Step 1. Size of the high pass filter should be equal to that of the low pass filter.
4. Add the outputs from Steps 2 and 3.

A geometric interpretation of this method is shown in Figure 6.11b. The low pass filter gives the effect of a ramp function and the high pass filter restores the high frequency components which are removed by the low pass filter.

This technique yields a mosaic similar to that of linear interpolation, but is computationally more expensive.

(A) METHOD



(B) GEOMETRIC INTERPRETATION

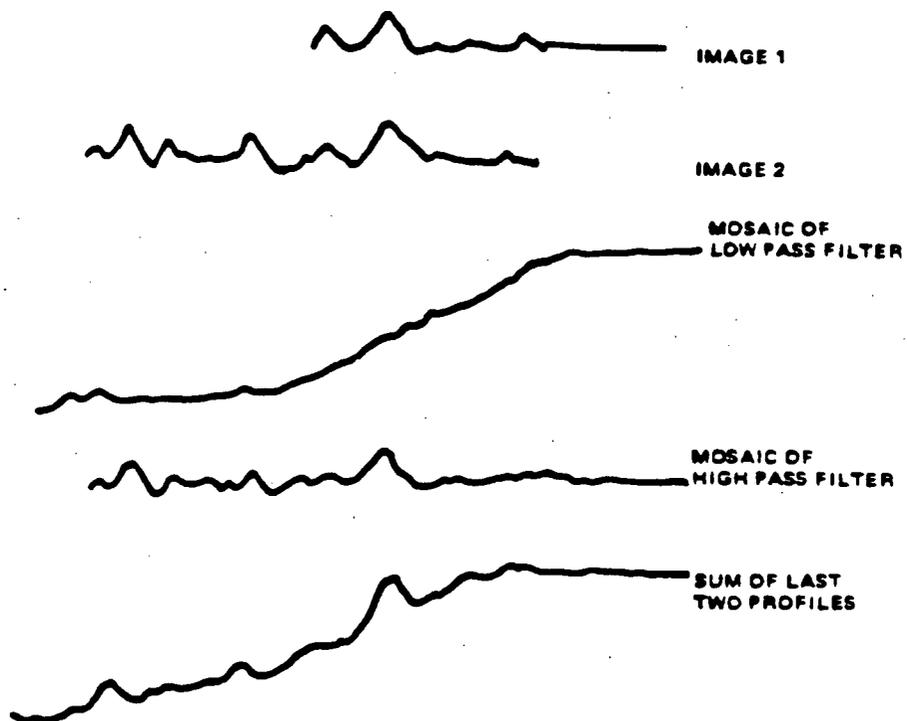


FIGURE 6.11 FILTERING METHOD OF IMAGE MOSAICKING

6.3.4 Experiments

The linear interpolation and filtering methods are considered in the experiment. The feathering method is not considered because of its "feathering" artifact. The test data used were two scenes imaged over an ice area of Cambridge Bay, Canada. In order to highlight the mosaicking effect, one image has been adjusted to have a higher mean than the other. The result is shown in Figure 6.12 where the two scenes were mosaicked with a vertical seam.

Figure 6.13 shows mosaicking by linear interpolation while Figure 6.14 shows mosaicking by filtering.

The mosaic frame is 240 pixels wide and the overlap area is 60 pixels in the middle of the mosaic frame. The mean of the left frame is 20 intensity levels (out of 256) higher than that of the right frame. The results show that the smoothing effect of these two methods is quite similar (ignoring the effect of the filter width). However, in the filtering method there appears to be a vertical seam through the lake in the overlap area. This is caused by the difference in contrast of land to lake in the frames. Such difference in contrast does not have any effect in the linear interpolation method. For this reason, linear interpolation is recommended and furthermore, it is computationally faster than the filtering method.

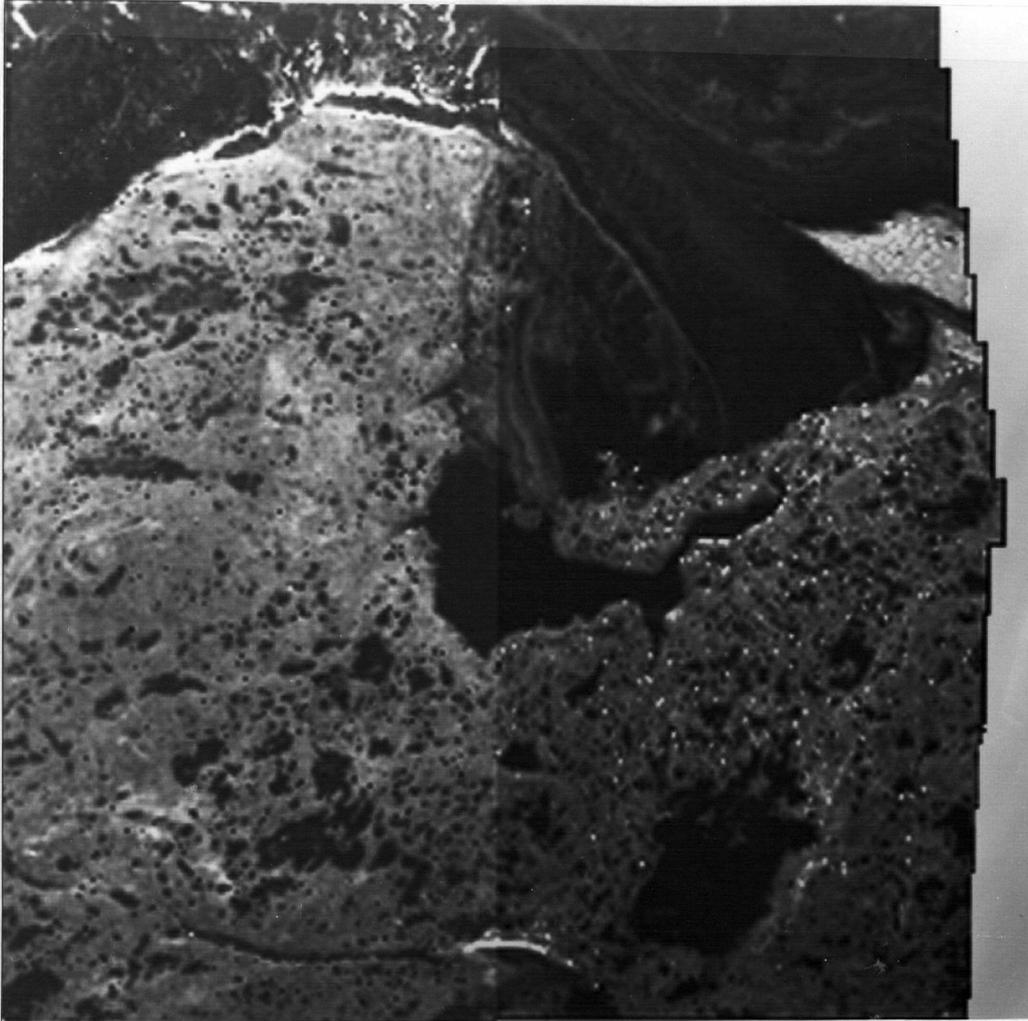


FIGURE 6.12 MOSAIC WITH NO OVERLAP



FIGURE 6.13 MOSAIC BY LINEAR INTERPOLATION

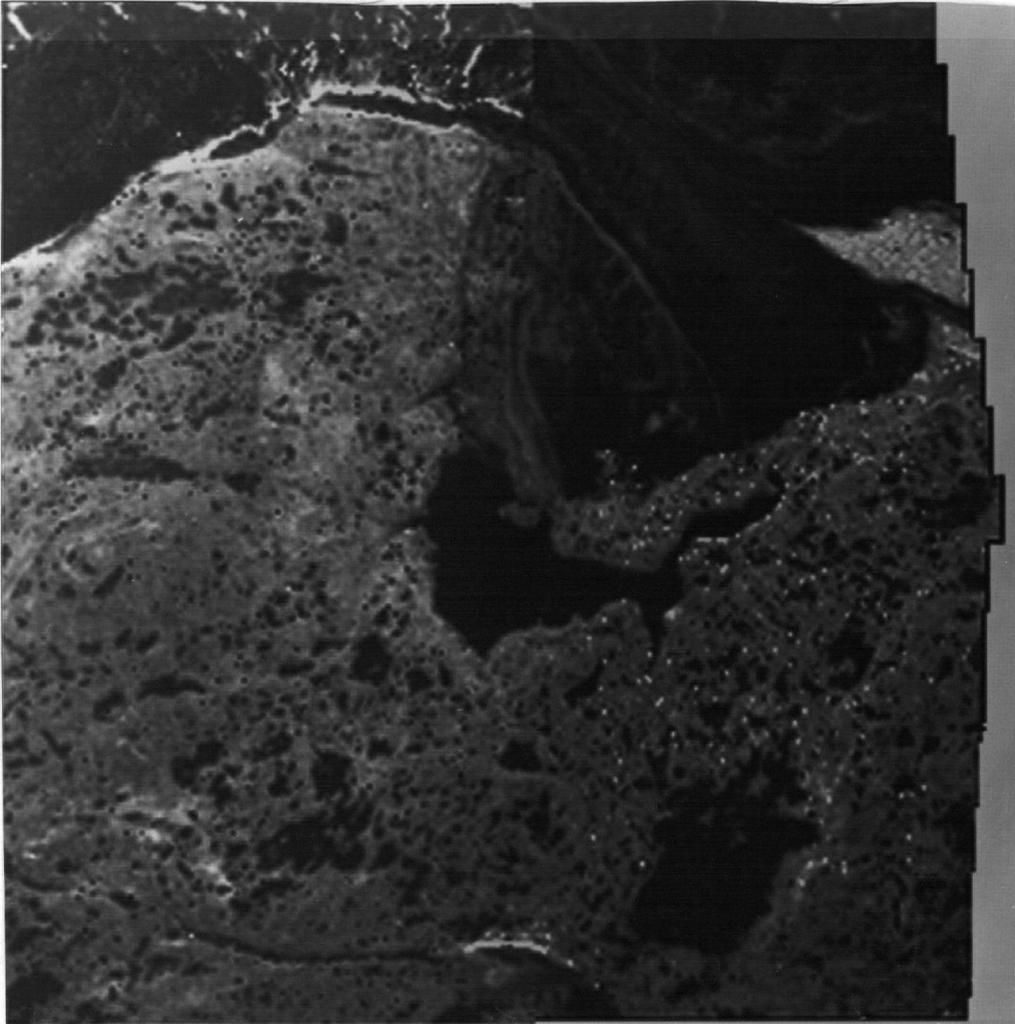


FIGURE 6.14 MOSAIC BY FILTERING

6.4 A Future Planetary Mission

6.4.1 Mission Description

The Venus Orbiting Imaging Radar (VOIR) due to be launched in the future under the auspices of NASA/JPL will carry a SAR System on board to chart the planet's surface. The satellite and its SAR system are designed to have the parameters presented in Table 6.1. Complementary to the SAR system, the satellite will also carry an altimeter on board.

In the rectification of SAR imagery obtained from the mission, special attention must be paid to:

- the navigation phase since no GCPs are available,
- the remapping transform which should also include the effect of atmospheric refraction in the thick Venus atmosphere, and
- the use of altimeter data.

6.4.2 Navigation

Due to the lack of GCPs, navigation will have to be performed by the automatic focusing method. The method has been outlined in Section 5.5. The sensitivity of FM rate to various parameters is shown in Table 6.1.

TABLE 6.1 SENSITIVITY OF FM RATE TO
PARAMETER VARIATIONS IN VOIR
IMAGERY

	% Change in FM Rate		Units
	0°N	71°N	
Target Location	0°N	71°N	
Target Latitude	0.51×10^{-4}	0.35×10^{-3}	/deg.
Orbit Semi-major Axis	0.016	0.016	/m
Orbit Eccentricity	2.0	2.0	
Satellite Angular Frequency	3117	3055	/ $\left(\frac{\text{deg}}{\text{sec}}\right)$
Orbit Inclination	0.055	0.095	/deg
Venus Semi-Major Axis	0.017	0.017	/m
Venus Rotation Rate	15.7	29.1	/deg

6.4.3 Venus Atmospheric Refraction

Since an electromagnetic wave refracts on entering Venusian atmosphere at nonnormal incidence, the SAR beam geometry departs from the linear. This effect changes the location of the swath by about 0.5 km, and causes the low resolution swath to be about 16 metres narrower than in the absence of refraction. Theoretically, the effect will contribute to inaccuracies in the overall target location equations, and can be taken into account by introducing a shift y_s , z_s (see Figure 6.15) in y_s and z_s . In the satellite coordinate system y_s points in the along track direction and z_s points from the satellite to the planet centre.

A description of the technique for computing these deviations follows. The index of refraction is calculated from the temperature, pressure density, and composition as a function of altitude. The dependence of index of refraction on height is shown in Figure 6.16 [MCCA80], where a static spherical symmetric multilayer model is used for the Venusian atmosphere.

To trace a ray path from the satellite to the surface, the atmosphere is assumed to consist of many concentric shells of finite thickness and constant refraction index. When the ray intersects a boundary between two layers, Snell's law is applied to calculate the change in direction. The ray then propagates in a straight line to the next boundary, where the process repeats. The mathematical treatment is presented in Appendix H.

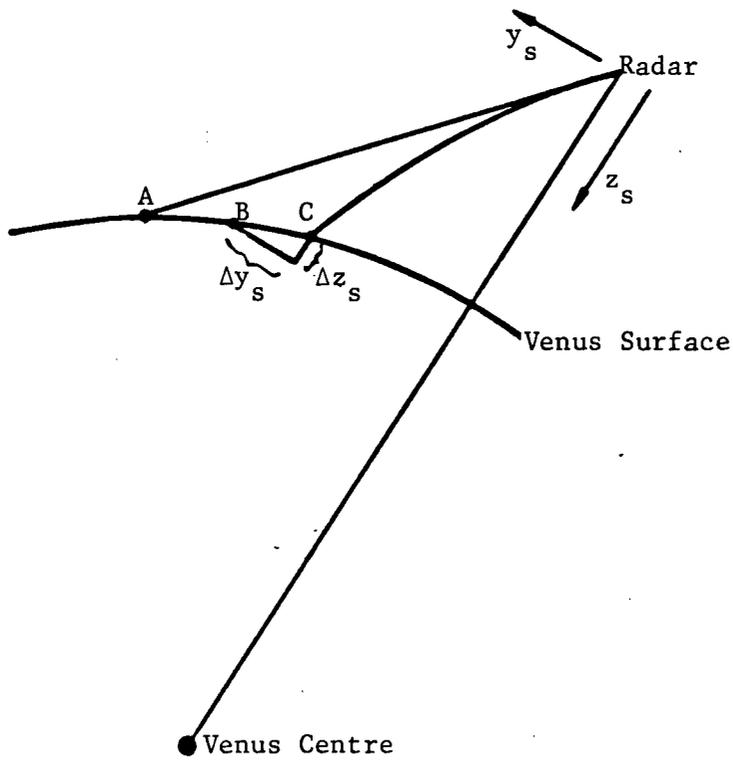


FIGURE 6.15 REFRACTION GEOMETRY

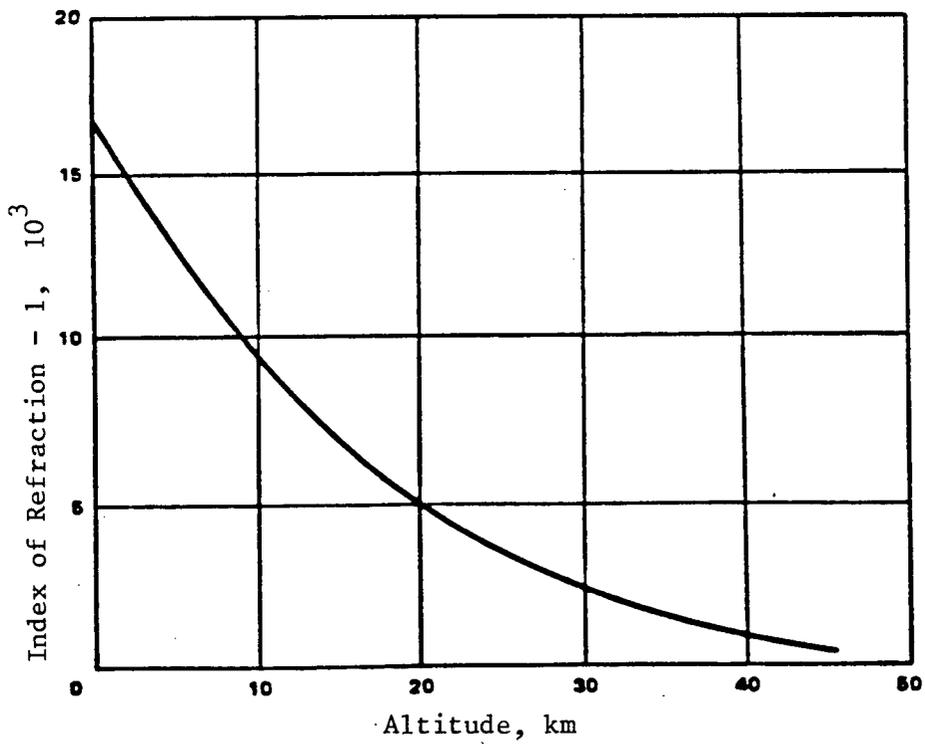


FIGURE 6.16 INDEX OF REFRACTION VERSUS ALTITUDE

The procedures to compute the deviations Δy_s and Δz_s are then (refer to Figure 6.15):

- (a) Trace the ray launched at the initial angle for free space until it intersects the planet surface at point C. Tracing is through the refracting atmosphere (Appendix H). Store the y_s and z_s coordinates of C.
- (b) Compute the two-way time τ taken by the beam illuminating the target at C.
- (c) Using the value of τ from Step (b) and free space value for the speed of light, solve for x_s and y_s . This gives point B.
- (d) Compute Δy_s , Δz_s which are the differences between the results of Steps (a) and (c).

Simulations for the model atmosphere of Venus give the following results:

Near range $\Delta y_s = 0.539$ km

$\Delta z_s = 0.026$ km

Mid range $\Delta y_s = 0.546$ km

$\Delta z_s = 0.028$ km

Far range $\Delta y_s = 0.555$ km

$\Delta z_s = 0.030$ km

Since the range of variation for the correction is small (16 metres maximum), the dependence of Δy_s and Δz_s on τ can be assumed to be negligible in the neighbourhood of a chosen target.

6.4.4 Use of Altimeter Data

In the rectification transformation (Appendix G), the satellite coordinates (x_s, y_s, z_s) of the target are a function of satellite distance H from the centre of the planet and planet radius r at target location. The accuracies of these two variables are:

$$H = 6301.4 \pm 1 \text{ km}$$

$$r = 6051.4 \pm 0.1 \text{ km}$$

Errors in H and r would introduce errors in the computation of x_s , y_s and z_s .

The VOIR altimeter measures the satellite altitude h from nadir, and its error is ± 0.24 km. Because the error in h is less than the error in H and since H can be written as $h + r$, the satellite coordinates, if expressed as a function of h and r , would be more accurately determined using altimeter data.

Referring to Appendix G, z_s is given by:

$$z_s = \frac{(c_{T_m}/2)^2 + H^2 - r^2}{2H} \quad (6.3)$$

and

$$\Delta z_s = \frac{\partial z_s}{\partial H} \Delta H + \frac{\partial z_s}{\partial r} \Delta r \quad (6.4)$$

Substituting $H = h + r$, z_s is given by:

$$z_s = \frac{(c\tau_m/2)^2 + H^2 + 2hr}{2(h+r)} \quad (6.5)$$

and

$$\Delta z_s = \frac{\partial z_s}{\partial H} \Delta H + \frac{\partial z_s}{\partial r} \Delta r \quad (6.6)$$

Table 6.2 compares the maximum error, using the nominal satellite orbit parameters, introduced in the computation of satellite coordinates of both representations. It is seen that the h representation (Equation 6.4), using the altimeter measurement data, is better than the H representation (Equation 6.3) in terms of rectification accuracy. Also, in the h representation, $(\partial z_s / \partial r) \Delta r$ is negligible.

6.4.5 Digital Terrain Model

Digital terrain data can be incorporated in the remapping process to correct for relief displacement as described in Section 2.6.2. The only DTM available for Venus has been compiled by Pettengill et al [PETT79] with a 200 km grid spacing. Note that even if relief displacement is not corrected, reasonable swath-to-swath registration can still be achieved because of the VOIR geometry, parallax of a target in two adjacent strips being negligible for most of the planet. In imaging from one strip to the next the orbit is displaced about 10 km, and it can be shown that the elevation needs to be about 7 km above the datum to produce a 200 metre parallax.

TABLE 6.2 H REPRESENTATION VERSUS
h REPRESENTATION

Error

$$\Delta D = \sqrt{(\Delta x_s)^2 + (\Delta y_s)^2 + (\Delta z_s)^2}$$

<u>H Representation</u>		Worst ΔD
	ΔD	$\Delta H = 1 \text{ km}$
		$\Delta r = 0.1 \text{ km}$
Near range	1.299 ($\Delta H - \Delta r$)	1.4289 km
Mid range	1.272 ($\Delta H - \Delta r$)	1.3992 km
Far range	1.248 ($\Delta H - \Delta r$)	1.3728 km
<u>h Representation</u>		Worst ΔD
	ΔD	$\Delta h = 0.24 \text{ km}$
Near range	1.299 Δh	0.3118 km
Mid range	1.272 Δh	0.3053 km
Far range	1.248 Δh	0.2995 km

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

The unified approach to the rectification of remotely sensed digital imagery discussed in this thesis provides the basis for a system to rectify various satellite imagery to a common data base, yielding a high throughput while retaining geometric and radiometric accuracy. (For a 5000 x 5000 Seasat image, remapping and using nearest neighbour interpolation requires about 6 hours of computer time. On the contrary, the suggested method takes only 1/2 hour.) Hence imagery with different sensor geometry and characteristics can be overlaid as has been illustrated by the Landsat/Seasat composite.

The thesis has investigated the two phases of the rectification process -- geometric transformation and interpolation, with geometric transformation further subdivided into navigation and remapping. The investigation was performed for various satellites with different imaging geometries (e.g., scanner imagery and SAR imagery). Regardless of the type of imagery, high throughput is achieved with one-dimensional interpolation of imagery data. Based on the foregoing investigations, the following conclusions can be drawn.

1. A DTM is required for the rectification of high resolution imagery. This has been verified in experiments with Seasat imagery and with synthesized TM and SPOT imagery. Terrain information is required in both the navigation phase and the remapping phase.

2. In the navigation phase, unknown parameters in the rectification transformation can be expanded as time series, the coefficients of which can be determined with the aid of GCPs and a recursive estimation technique. The number of GCPs required to attain a certain rectification accuracy depends upon the quality of the GCPs. This method has been investigated from high resolution and narrow swath width (Landsat MSS) to low resolution and wide swath width imagery (TIROS-N). It was shown that about 10 GCPs per scene were necessary for subpixel rectification accuracy.

3. For SAR imagery, another way of navigation is via automatic focusing. This method has been discussed. However, investigation of this method is left for future research.

4. To retain radiometric accuracy and high throughput using one-dimensional interpolation, three one-dimensional passes have to be applied. In each pass, a 16-point sinc interpolation kernel weighted by a Kaiser window will yield a maximum of 1 level RMS error out of 256 levels. This resulting accuracy is far better than cubic convolution which has been proposed by a few researchers.

5. The presence of scan gaps in TM imagery has led to the investigation of a new interpolation kernel. This kernel is able to yield an acceptable 1 level RMS error out of 256 levels in the case of overlap. The error in the case of underlap depends on the scan gap size.

6. The example of a Landsat/Seasat composite has succinctly demonstrated the benefits of multi-sensor integration. The artificially produced image composite combines the data acquired by two different sensors and satellites, one a passive system emphasizing tonal qualities of ground cover and the other an active radar system accentuating surface texture.

7.2 Recommendations for Future Work

As a result of the investigations in the thesis, the following recommendations are made for further research work:

1. In the navigation phase, it is desirable to reduce the number of GCPs. It is a very time consuming process to mark GCPs and in many areas of the world, high quality maps are not available. One way to reduce the number of GCPs is to model the satellite attitude by a physical law of satellite motion: rate of change of angular momentum (a function of rate of change in attitude angles) equals the total torque acting on the satellite. Hence if the total torque acting on the satellite is known, the variations in the attitude angles can be evaluated. This method requires an extensive knowledge of the satellite's physical parameters (such as moment of inertia, mass and dimension of solar panels, control parameters, etc.) [FRIE83].

2. Again on the topic of navigation, the process of marking GCPs can be automated with image to image correlation techniques [BARN72]. Here, a small section of an image around each GCP is stored for a particular

sensor. When another image with the same sensor or a different sensor is obtained over the same area, automatic correlation can be evoked. However, the following problems must be solved:

- How will geometric distortions such as those present in TM imagery affect the correlation technique?
- How should sensor to sensor (with different sensor resolutions) correlation be performed? Furthermore what should be the orientation of the GCP reference subimage?
- How would seasonal and sun position changes affect the correlation?
- Automatic registration for two radar scenes over the same area is difficult, let alone registration of radar to scanner imagery. The problems lie in the intensity migration, presence of speckles and polarization effects in radar imagery.

3. Automatic focusing appears to be promising for the navigation of SAR imagery. This method should be investigated with Seasat SAR imagery and simulated imagery. The following should be studied:

- What geometric parameters can be updated?
- What is the effect of "matched filter control point" distribution?
- Performance and limitations.

4. In the processing of SAR data, a one-dimensional remapping and interpolation pass has been evoked to correct for range cell migration. Can this be coupled with the three one-dimensional passes method to reduce

the total number of remapping and interpolation passes?

5. A radiometrically corrected or destriped Landsat scene always shows some residual striping. How does this residual striping affect the interpolation accuracy?

6. Is it possible to develop an interpolation kernel the size of which is adaptive to the frequency content of the subarea it is operating on? Then this filter must have built-in intelligence to know or predict the gray level variations in the subarea. For relatively flat areas (in terms of intensity), a kernel size of less than 16 is adequate to retain the required radiometric accuracy of 1 level RMS out of 256; hence interpolation time is reduced.

7. The following is the general problem of using terrain data to perform radiometric correction. In the correction of SAR imagery, the slope foreshortening effect causes a narrow band of high pixel intensities to be mapped to a wider area. While it is correct from the signal processing point of view to use a sinc function for interpolation, a different method which also accounts for the slope effect may yield better results for image interpretation purposes. An example is to average out the intensity of the narrow band to the mapped wider area.

8. The relief displacement for Mapsat and a possible high squint SAR imagery will also have an across scan component. How can the present approach be modified to accommodate the correction of such distortion?

9. How can the Landsat/Seasat overlay be used? An example will be to show how such overlay can improve classification accuracy. Another example will be to detect features which cannot be detected in only one component.

10. The resulting Landsat/Seasat composite has the high resolution of the Seasat image while retaining the spectral information of the Landsat image. This leads one to believe that in the design of a high resolution MSS system, not all the spectral bands need to be of the same resolution. This is for data reduction purposes. For a set of correlated spectral bands, it may be required only to have one band with the high resolution and all the other bands in the set with a relatively lower resolution. When the lower resolution bands are modulated by the corresponding high resolution band for all sets of spectral bands, the result will be a high resolution colour image. Hence data are reduced without significantly degrading the amount of information. Experiments can be performed with Landsat MSS or TM imagery. Simulation of the low resolution bands can be done by low pass filtering the imagery data in the corresponding bands.

11. A major area of research definitely lies in the production of DTMs. The following work should be investigated:

- The side looking mode of SPOT will allow a stereo pair to be obtained and hence terrain information can be extracted of the imaged area. However the relief displacement vectors in two neighbouring passes are not parallel (except at the equator). How would this affect the correlation process to obtain

- terrain height and the height accuracy?
- One technique to obtain a contoured image is by Moire contourography [PIRO82]. In this technique, an object has to be much closer to the interference grating than the illumination to the grating. Can this technique be modified to satellite borne imagery? Since in this application the terrain is much farther away from the grating than the illumination is to the grating, the problem of diffraction has to be overcome. A SAR system is a suitable choice of illumination, but will the SAR data processing annihilate the desired contours which are formed by the principle of interference? Another problem is that the contours thus obtained do not show concavity or convexity.
 - Another technique of obtaining a contoured image is mentioned by Graham [GRAH74]. It uses two antennas separated by a vertical distance on the spacecraft. Hence the two images obtained will create an interference pattern to produce contour lines. The positioning of the contour depends on the imaging geometry of the two antennas. However, the concavity and convexity of the contours are not immediately known and have to be inferred. Also the contours are displaced according to contour elevation.

Vast application of remotely sensed imagery has already been found in geology, agriculture, forestry, oceanography, cartography and many other earth resources management areas. With the increased resolution and

number of spectral bands in the present and future generations of satellite sensors, the rectified imagery will undoubtedly be more invaluable than ever in these application areas.

BIBLIOGRAPHY

- [ALLA78] Allan, M. M. "DTM Application in Topographic Mapping", Photogrammetric Engineering and Remote Sensing, Vol.44, No.12, Dec.1978, pp.1513-1520.
- [AVER81] Avery, J. E. and J. S. Hsieh, "Geometric Correction Resampling for the Landsat-D Thematic Mapper", International Geoscience and Remote Sensing Symposium, Washington, D.C., June 1981.
- [BAKE75] Baker, J. R. and E. M. Mikhail, "Geometric Analysis and Restitution of Digital Multispectral Scanner Data Arrays", LARS Information Notes 052875, Purdue University, 1975.
- [BARN72] Barnea, D.T. and H. Silverman, "A class of Algorithms for Fast Digital Image Registration", IEEE Transactions on Computers, Vol.C-21, Feb.1972, pp.179-186.
- [BENN78] Bennett, J.R. and I.G. Cumming, "Digital Technique for the Multilook Processing of SAR Data With Application to Seasat-A", Fifth Canadian Symposium on Remote Sensing, Victoria, B.C., Aug.1978, pp.506-516.
- [BENN79] Bennett, J.R. and I.G. Cumming, "A Digital Processor for the Production of Seasat Synthetic Aperture Imagery", Surge Workshop, Esrim, Frascati, Italy, July 1979.
- [BENS80] Benson, M., "Digital Processing of Seasat-A SAR Data using Linear Approximations to Range Cell Migration Curves", IEEE International Radar Conference, 1980, pp.176-181.
- [BERN76] Bernstein, R., "Digital Image Processing of Earth Observation Sensor Data", IBM Journal of Research and Development, Jan.1976, pp.40-57.
- [CARO75] Caron, R.H. and K.W. Simon, "Attitude Time Series Estimation for Rectification of Spaceborne Imagery", Journal of Spacecraft and Rockets, Vol.12, No.1, Jan.1975, pp.146-151.
- [CNES81] "SPOT to Ground Station Interface Document", CNES, France, May, 1981.
- [COLV79] Colvocoresses, A.P., "Proposed Parameters for Mapsat", Photogrammetric Engineering and Remote Sensing, Vol.45, No.4, Apr.1979, pp.501-505.
- [CUMM79] Cumming, I.G. and J.R. Bennett, "Digital Processing of Seasat SAR Data", IEEE International Conference on Acoustics, Speech and Signal Processing, Washington, D.C., Apr.1979, pp.710-718.

- [DEBO79] DeBoor, C., "A Practical Guide to Splines", Springer-Verlag, 1979, Chapter IV.
- [DOYL78] Doyle, F.J., "Digital Terrain Models, An Overview", Photogrammetric Engineering and Remote Sensing, Vol.44, No.12, Dec.1978, pp.1481-1485.
- [DUPL67] Du Plessis, R.M., "Poor Man's Explanation of Kalman Filtering", Northern American Rockwell Corporation, June 1967.
- [EBNE76] Ebner, H., "A Mathematical Model for Digital Rectification of Remote Sensing Data", XIII Congress of the International Society for Photogrammetry, Commission III, Helsinki, Finland, July 1976.
- [EKLU72] Eklundh, J.O., "A Fast Computer Method for Matrix Transposing", IEEE Transactions on Computers, Vol.C-21, June 1972, pp.801-803.
- [FRIE81] Friedmann, D.E., "Two-Dimensional Resampling of Line Scan Imagery by One-Dimensional Processing", Photogrammetric Engineering and Remote Sensing, Vol.47, No.10, Oct.1981, pp.1459-1467.
- [FRIE83] Friedmann, D.E., J.P. Friedel, K.L. Magnussen, R. Kwok and S. Richardson, "Multiple Scene Precision Rectification of Space Borne Imagery with Very Few Ground Control Points", Photogrammetric Engineering and Remote Sensing, Vol.49, No.12, Dec.83, pp.1657-1667.
- [GOLD81] Goldbogan, G.C., "PRIM: A Fast Matrix Transpose Method", IEEE Transactions on Software Engineering, Vol.SE-7, No.2, Mar.1981, pp.255-257.
- [GRAH74] Graham, L.C., "Synthetic Interferometer Radar for Topographic Mapping", Proceedings of IEEE, Vol.62, No.6, June 1974, pp.763-768. Also a selected paper in [KOVA76], pp.275-280.
- [GSFC82] "Landsat-D to Ground Station Interface Description", Revision 5, Goddard Space Flight Centre, Maryland, Aug.1982.
- [HARD71] Hardy, R.L., "Multiquadratic Equation of Topography and Other Irregular Surfaces", Journal of Geographical Research, Vol.76, No.8, 1971.
- [HEIL78] Heil, R.V. and S.M. Brych, "An Approach for Consistent Topographic Representation of Varying Terrain", DTM Symposium, St. Louis, Missouri, May 1978, pp.397-410.

- [HORN78] Horn, B.K.P. and B.K. Bachman, "Using Synthetic Images with Surface Models", Communications of the ACM, Vol.21, No.11, Nov.1978, pp.914-924.
- [HORN79] Horn, B.K.P. and R.J. Woodham, "Landsat MSS Coordinate Transformations", Proceedings of the Fifth International Symposium on Machine Processing of Remotely Sensed Data, 1979, pp.59-68.
- [HORN81] Horn, B.K.P., "Hill Shading and the Reflectance Map", Proceedings of the IEEE, Vol.69, No.1, Jan.1981, pp.14-47.
- [HOVA79] Hovanessian, S.A., "Introduction to Synthetic Array and Imaging Radars", Hughes Aircraft, Mar.1979.
- [JAME82] James, W., "Unveiling Venus with VOIR", Sky and Telescope, Feb.1982, pp.141-144.
- [JERR77] Jerri, A.J., "The Shannon Sampling Theorem - Its Various Extensions and Applications: A Tutorial Review", Proceedings of the IEEE, Vol.65, No.11, Nov.1977, pp.1565-1596.
- [JPL77] "Seasat-A Instrument Data Processing System Detail Functional Specification", Vol.1, JPL, Pasadena, California, June 1977.
- [KELL77] Kelly, R.E., P.R.H. McConnell and S.J. Mildenberger, "The Gestalt Photomapper System", Photogrammetric Engineering and Remote Sensing, Vol.43, No.11, Nov.1977, pp.1407-1417.
- [KONE76] Konecny, G., "Mathematical Models and Procedures for the Geometric Restitution of Remote Sensing Imagery", XIII Congress of the International Society for Photogrammetry, Commission III, Helsinki, Finland, July 1976.
- [KORN61] Korn, G. and T. Korn, "Mathematical Handbook for Scientists and Engineers", McGraw-Hill, 1961, Sections 21.4 and 21.8.
- [KOVA76] Kovaly, J.J., "Synthetic Aperture Radar", Artech, 1976.
- [KRAT72] Kratky, V., "Photogrammetric Solution for Precision Processing of ERTS Images", XII Congress of the International Society of Photogrammetry, Ottawa, June 1972.
- [KRAU78] Kraus, K., "Rectification of Multispectral Scanner Imagery", Photogrammetric Engineering and Remote Sensing, Vol.44, No.4, Apr.1978, pp.453-457.
- [LARS80] Larson, J., "Rectification of Digital Images for Remote Sensing Analysis", Department of Photogrammetry, Royal Institute of Technology, Stockholm, Sweden, 1980.

- [LATH65] Lathi, B.P., "Signals, Systems and Communications", John Wiley and Sons, 1965, Chapter 11.
- [LECK80] Leckie, D.G., "Use of Polynomial Transformation for Registration of Airborne Digital Line Scan Images", Sixth Canadian Symposium on Remote Sensing, Halifax, N.S., May 1980, pp.635-641.
- [LEON70] Leondes, C.T., "Theory and Applications of Kalman Filtering", University of California, Los Angeles, Feb.1970.
- [LINF39] Linfoot, E.H. and W.M. Shepherd, "On a Set of Linear Equations (II)". Journal of Mathematics, Vol.10, 1939, pp.84-98.
- [LITT80] Little, J.J., "Automatic Rectification of Landsat Images Using Features Derived from Digital Terrain Models", M.Sc. Thesis, U.B.C., Vancouver, 1980.
- [MALI73] Maling, D.H., "Coordinate Systems and Map Projections", George Philip and Son Limited, London, England, 1973, p.194.
- [MCCA80] McCarthy, J.F., "Summary of Atmospheric Refraction Study", Hughes Aircraft internal memo, June 1980.
- [MDA76] "Feasibility Study Signal Processing System for Satellite-Borne Synthetic Aperture Radar", MDA Technical Report for CCRS and CRC, Ottawa, Feb.1976.
- [MDA77a] "Considerations in the Implementation of the Range/Doppler Domain Approach to Seasat-A SAR Data Processing", MDA, Dec. 1977.
- [MDA77b] "A Theoretical Foundation for the Design of a Ground Data Processor for a Satellite Borne Synthetic Aperture Radar", MDA, Aug.1977.
- [MDA78] "The Development of an Ice Status System", MDA, Mar.1978.
- [MDA79] "SAR Seasat Processor Functional Description", MDA, Apr.1979.
- [MILG75] Milgram, D., "Computer Methods for Creating Photomosaics", IEEE Transactions on Computers, Vol.C-24, No.11, Nov.1975' pp.1113-1119.
- [MILG77] Milgram, D., "Adaptive Techniques for Photomosaics", IEEE Transactions on Computers, Vol.C-26, No.11, Nov.1977, pp.1175-1180.
- [MILL71] Millar, R.A., "The Principles of Kalman Filtering", CRC, Ottawa, Sept.1971.

- [NASA76a] "Real-Time Data Systems for the TIROS-N Spacecraft Series", NASA Document, Mar.1976.
- [NASA76b] "ERTS Data Users' Handbook", NASA Document No.71SD4249, Sept.1976.
- [ORTH77] Orth, R. and D.S. Sloan, "A Self-Contained Landsat Data Reception and Precision Cartographic Image Production System", Proceedings of the International Symposium on Image Processing, Interactions with Photogrammetry and Remote Sensing, Graz, Austria, Oct.1977, pp.189-196.
- [ORTH78a] Orth, R., K. Brydges and F. Wong, "Photomaps from Precision Rectified Landsat Imagery", Fifth Canadian Symposium on Remote Sensing, Victoria, B.C., Aug.1978, pp.292-295.
- [ORTH78b] Orth, R., F. Wong and J.S. MacDonald, "The Production of 1:250,000 Maps of Precision Rectified and Registered Landsat Imagery using the MDA Image Analysis System: Initial Results", Twelfth Symposium on Remote Sensing of Environment, Manila, Phillipines, Apr.1978, pp.2163-2176.
- [OTEP78] Otepka, G., "Practical Experience in the Rectification of MSS Images", Photogrammetric Engineering and Remote Sensing, Vol.44, No.4, Apr.1978, pp.459-467.
- [PETT79] Pettengill, G.J., E. Eliason, P.G. Ford, G.B. Loriot, H. Masursky and G.E. McGill, "Pioneer Venus Radar Results: Altimetry and Surface Properties", Journal of Geophysical Research, Dec.1979.
- [PEUC78] Peucker, T.K., R.J. Fowler, J.J. Little and D.M. Mark, "The Triangulated Irregular Network", DTM Symposium, St. Louis, Missouri, May 1978, pp.515-540.
- [PIRO82] Pirodda, L., "Shadow and Projection Moire Techniques in Absolute and Relative Mapping of Surface Shapes", Optical Engineering, Vol.21, No.4, Aug.1982, pp.640-649.
- [PRAK81] Prakash, A., and E.P. Beyer, "Landsat-D Thematic Mapper Image Resampling for Scan Geometry Correction", Proceedings of the Seventh International Symposium on Machine Processing of Remotely Sensed Data, June 1981, pp.189-200.
- [RABI75] Rabiner, L.R. and B. Gold, "Theory and Application of Digital Signal Processing", Prentice-Hall, 1975, Section 3.11.
- [RAMA77] Ramapriyan, H.K., "Data Handling for the Geometric Correction of Large Images", IEEE Transactions on Computers, Vol.C-26, No.11, Nov.1977, pp.1163-1167.

- [RICH72] Richardus, P., and R.K. Adler, "Map Projections for Geodesists, Cartographers and Geographers", North-Holland 1972, Section 7.5.
- [SAKR68] Sakrison, D.J., "Communication Theory: Transmission of Waveforms and Digital Information", Wiley, 1968, Section 2.7.
- [SCHU76a] Schumaker, L., "Fitting Surfaces to Scattered Data", Approximation Theory II, ed. G. Lorentz, Academic Press, 1976.
- [SCHU76b] Schut, G.H., "Review of Interpolation Methods for Digital Terrain Models", XIII Congress of the International Society for Photogrammetry, Commission III, Helsinki, Finland, July 1976.
- [SHLI79] Shlien, S., "Geometric Correction, Registration and Resampling of Landsat Imagery", Canadian Journal of Remote Sensing, Vol.5, No.1, May 1979, pp.74-89.
- [SIMO75a] Simon, K.W., S.S. Rifman and R.H. Caron, "Precision Digital Processing of Landsat MSS Data", Eleventh International Symposium on Space Technology and Science, Tokyo, Japan, July 1975.
- [SIMO75b] Simon, K.W., "Digital Image Reconstruction and Resampling for Geometric Manipulation", Proceedings IEEE Symposium on Machine Processing of Remotely Sensed Data, June 1975, pp. 3A-1 - 3A-11.
- [UCLA79] "UCLA Kalman Filtering Lecture Notes", UCLA, California, Mar.1979.
- [WOOD80] Woodham, R.J., "Using Digital Terrain Data to Model Image Formation in Remote Sensing", SPIE, Image Processing for Missile Guidance, Vol.238, July 1980, pp.361-369.
- [YEN56] Yen, J.L., "On Non-Uniform Sampling of Bandlimited Signals", IRE Transactions on Circuit Theory, Dec.1956, pp.251-257.
- [ZOB82] Zobrist, A., "Computational Aspects of Remapping Digital Imagery", Proceedings of the NASA Workshop on Registration and Rectification, NASA, July 1982, pp.358-366.

APPENDIX A LANDSAT-4 SENSORS

A.1 General Description

The Landsat-D satellite has been successfully launched in late 1982 to become Landsat-4. The sensors on board consist of a multispectral scanner (MSS) and a thematic mapper (TM). The satellite and the sensors have the following characteristics:

Orbit inclination:	99°
Nodal period:	99 minutes
Satellite altitude:	varies from 685 km to 740 km, 705 km at 40 N° latitude
Swath width:	185 km
Pixel resolution:	58 m x 79 m MSS 30 m x 30 m TM Bands 1,2,3,4,5 and 7 120 m x 120 m TM Band 6

The MSS sensor is similar, both in spectral bands and resolution, to the MSS sensors in the previous Landsat satellites. Only the TM high resolution bands are considered in this thesis due to the complex imaging geometry of the detectors and the high susceptibility to geometric distortions as a result of the high resolution. The TM is a bidirectional scanner, imaging the earth in the forward and reverse scanning directions across the ground track as shown in Figure A.1. The scanning is accomplished by an oscillating mirror. A scan line corrector (SLC) helps to align the ground track projections together. Without the SLC,

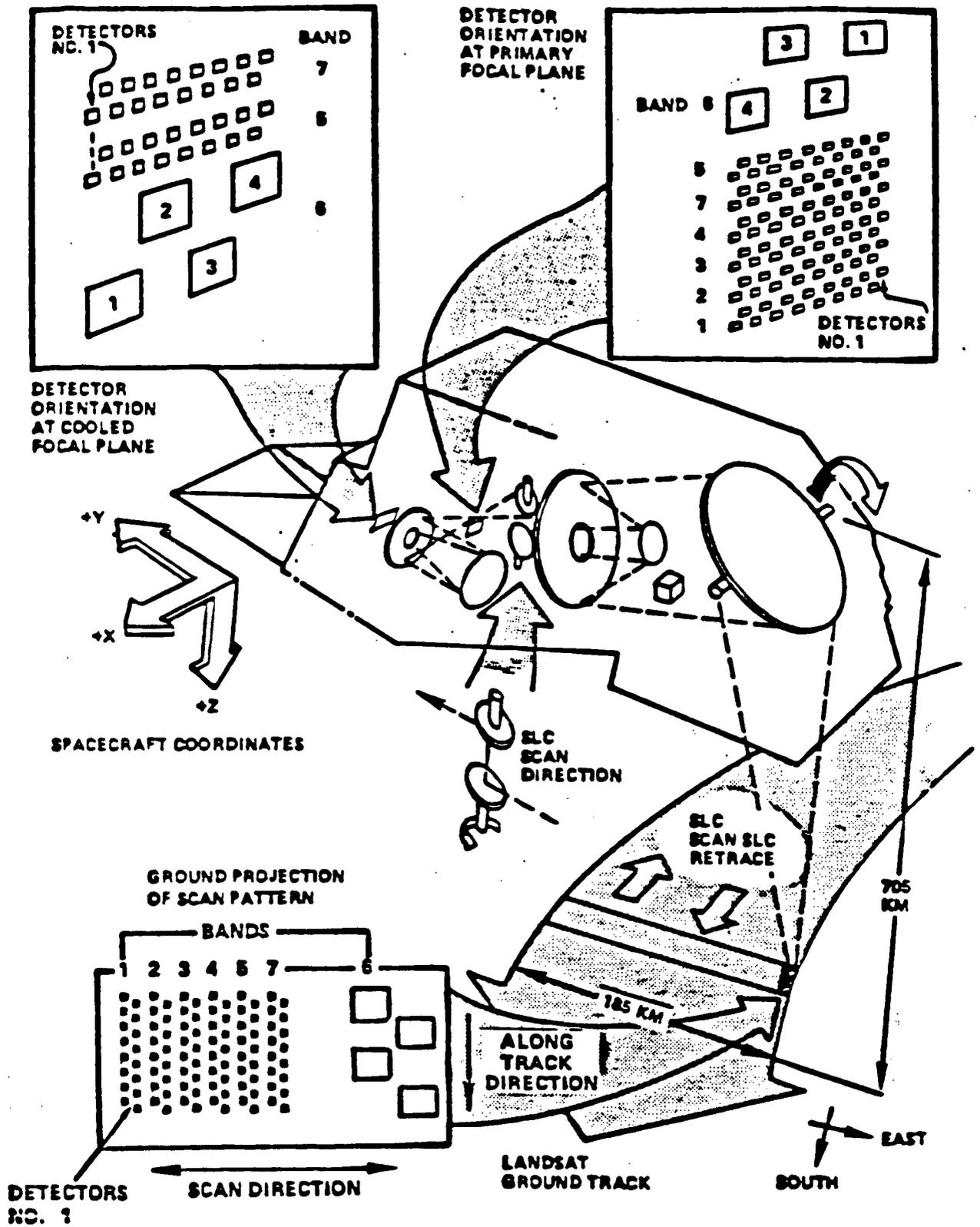


FIGURE A.1 DETECTOR ARRAY PROJECTIONS ON GROUND TRACK [GSFC 82]

there exists a significant amount of overlap and underlap from swath to swath. The functions of the SLC are shown in Figure A.2 which shows the ground projection of the scan with the satellite stationary. With the velocity vector of the satellite, the final and required projection is shown in Figure A.2c.

The characteristic of TM imagery is the presence of scan gaps to be described in what follows. The actual imagery obtained will not be as ideal and is shown in Figure A.3, because the SLC is set for nominal satellite ground speed. But the ground speed deviates from nominal because of:

- satellite altitude variation, and
- satellite velocity variation.

Furthermore, the TM sensors suffer from high frequency vibration jitter and from the bowtie effect (pixel projection on the ground is larger for a larger scan angle). These distortions introduce unevenly spaced samples when projected on the ground. This is shown in Figure A.3c in which Δ_0 denotes the interline spacing within a swath and Δ_1 denotes the spacing between the two end detectors in two swaths. Ideally, Δ_0 should be the same as Δ_1 . But due to the various distortions, Δ_0 may not be equal to Δ_1 . Note that Δ_0 is constant over the time interval of imaging a scene. As a result, there are 15 equal sample spacings followed by one unequal sample spacing since each swath contains 16 lines of imagery.

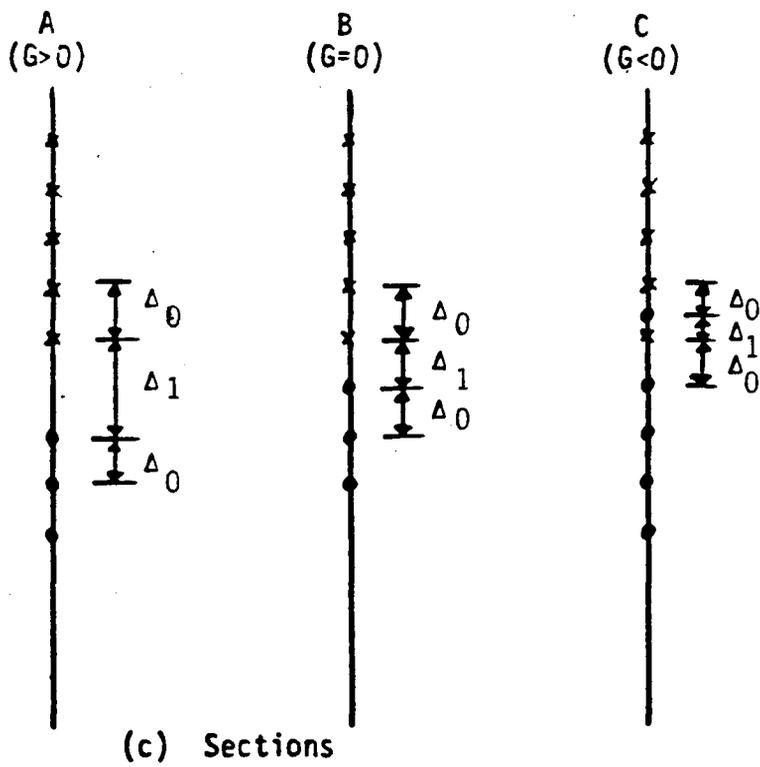
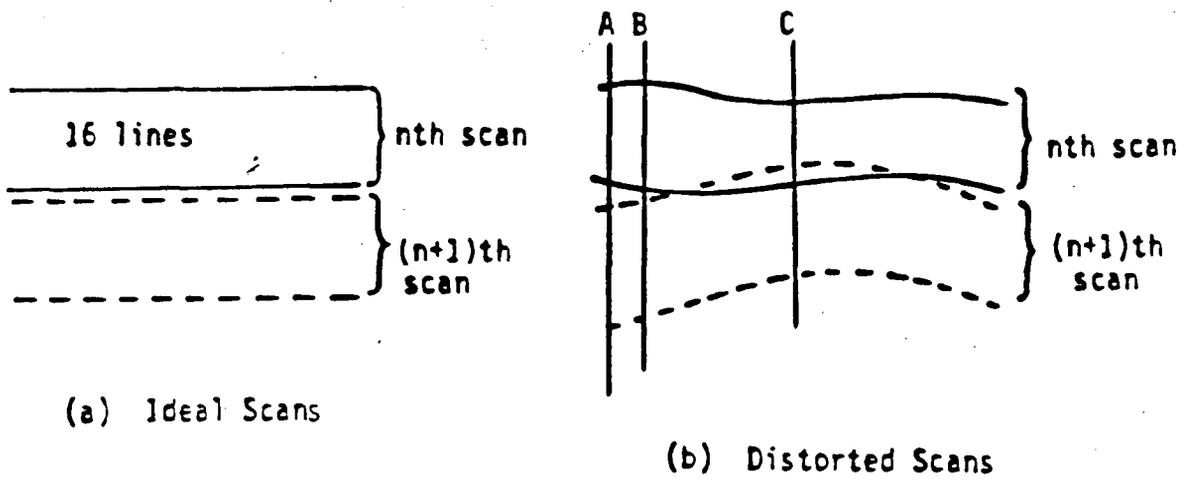


FIGURE A.3 TM SCAN GAPS

A.2 Scan Gap Analysis

A characteristic of TM imagery is the presence of scan gaps. This section analyzes the scan gap magnitude. Figure A.3 shows that the ground coverage contains a certain amount of underlap (missing data) and overlap (redundant data). The underlap/overlap is sometimes referred to as scan gap and is defined as:

$$G = \Delta_1 - \Delta_0 \quad (\text{A.1})$$

Therefore, when G is 0, all the samples are evenly spaced: when G is positive, underlap occurs and when G is negative, overlap occurs. For TM imagery, G lies between ± 2 lines.

The distortions that contribute to the scan gap are discussed in what follows.

1. Jitter

Jitter is caused by vibration in the mirror assembly. Anticipated RMS jitter error, referenced to the spacecraft coordinate system, is shown in Table A.1 [GSFC82]. Significant error is not expected to occur above 77 Hertz. The non-zero scan gap is partly caused by jitter of frequencies higher than 0.4 Hertz. For example, jitter of 7 Hertz or more corresponds to one cycle in less than 32 lines, which is two swaths of data. The across scan distortion, caused by jitter (3σ level) is thus between ± 6 metres. Hence, the maximum scan gap is 12 metres at the 3σ level in the worst case if the jitters for two consecutive swaths are exactly out of

TABLE A.1 ERROR MAGNITUDE CAUSED BY JITTER

Frequency Range (Hertz)	Number of Lines per Cycle	Number of Swaths per Cycle	Error Magnitude (3σ) in arc-sec	Across-Scan Error Magnitude (3σ) in metres
0.4 to 7	560 to 32	35 to 2	0.9 all axes	3.1
Greater than 7	Less than 32	Less than 2	2.79 roll 0.60 pitch 0.90 yaw	9.5 2.1 0.3

phase.

Jitter also introduces distortion in the along scan direction, but this is slow varying as shown below. Taking the highest significant frequency of jitter to be 77 Hertz and the pixel sampling rate 9.6×10^{-6} sec, one cycle corresponds to at least

$$1/(77 \times 9.6 \times 10^{-6}) = 1352 \text{ pixels}$$

2. Satellite Altitude Variation

The altitude of the Landsat-4 orbit, considering both orbit eccentricity and the earth shape, will vary between 685 and 740 km. At 705.3 km, the swath width in the along scan direction is 480 metres. Variation in the orbit altitude will produce a scale change in the line spacing within a swath but the distance between swath centres will not change (assuming constant satellite velocity). The contribution to the scan gap can be calculated by the relation:

$$G = 480 \left(1 - \frac{h}{705.3}\right) \quad (\text{A.2})$$

where h is the orbit altitude in km. For the specified altitude range, G varies from 13.8 to -23.6 metres.

3. Satellite Velocity Variation

Velocity variation causes a change in the spacing between swath centres but not within the swath and hence, affects the scan gap.

Assume that the major contribution to velocity variation is altitude variation, then the satellite velocity follows from Kepler's Law:

$$v = K \left(\frac{1}{R + h} \right)^{0.5} \quad (\text{A.3})$$

where K is a proportionality constant, h is orbit altitude, R is the local earth radius at the nadir point, and R + h is the local satellite radius. The ground speed is then

$$v_G = KR \left(\frac{1}{R + h} \right)^{1.5} \quad (\text{A.4})$$

and the swath centre spacing is

$$W = KRT \left(\frac{1}{R + h} \right)^{1.5} \quad (\text{A.5})$$

where T is the swath time. Since W is about 480 metres with an orbit altitude of 705.3 km, the value of W at any other orbit altitude h is then given by:

$$W = 480.0 \frac{R}{R_0} \left(\frac{R_0 + 705.3}{R + h} \right)^{1.5} \quad (\text{A.6})$$

where R_0 is the local earth radius at the subsatellite point when the orbit altitude is 705.3 km.

The maximum difference between R and R_0 is 22 km on the earth spheroid, and this difference gives an error of less than one metre in W if R is assumed to be equal to R_0 in the above equation.

When the satellite orbit is highest at 740 km, the value of W is 476.5 metres and when it is lowest at 685 km, the value of W is 482.1 metres. Hence, the contribution to the scan gap varies from -3.5 metres to 2.1 metres.

4. Scan Skew

This is a combined effect of the SLC rate, satellite altitude, ground velocity and earth rotation. To produce the ideal ground track as shown in Figure A.2c, the satellite altitude and ground velocity must exactly compensate for the SLC. This occurs only at 40°N in latitude where the altitude is 712.8 km and ground velocity is 6.821 km/sec [PRAK81]. At the other altitudes, the ground velocity changes according to Equation A.4. The scan skew components are shown in Figure A.4 in which

A = ground velocity x active scan time,

B = SLC rewind ground distance,

C = distance travelled during rewind,

E = earth rotation in along-track direction in one swath time, and

d_0, d_1 = ground distance between swath centres at both ends.

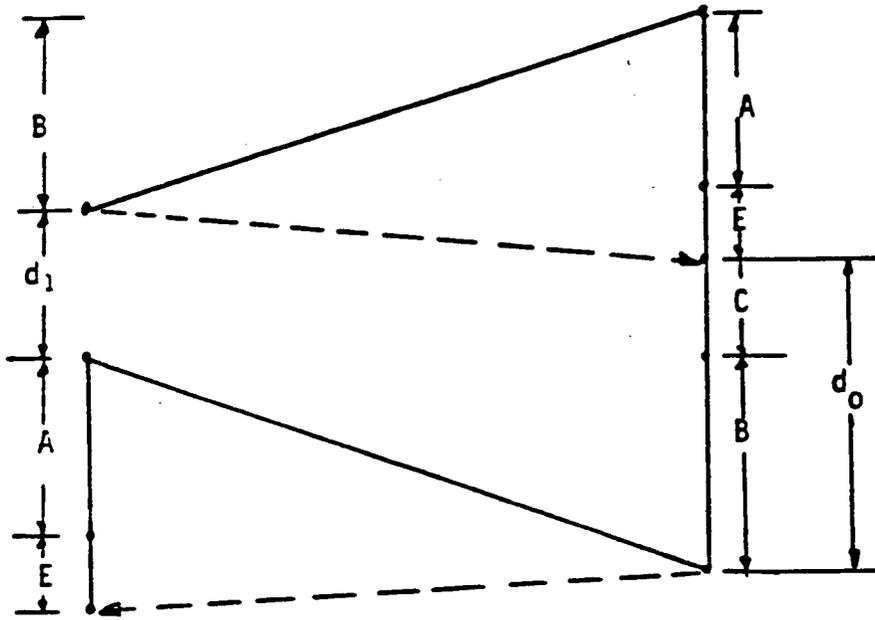
From the figure

$$d_0 = B + C$$

and
$$d_1 = 2(A + E) - B + C$$

At the extreme altitudes, the values are

Altitude km	d_0 metres	d_1 metres	$(d_0 - d_1)$ metres
685	473.3	516.7	-42.9
740	504.6	474.0	30.6



Note: Dotted line represents loci of swath centre.
See test for distances A,B,C and E.

FIGURE A.4 SCAN SKEW

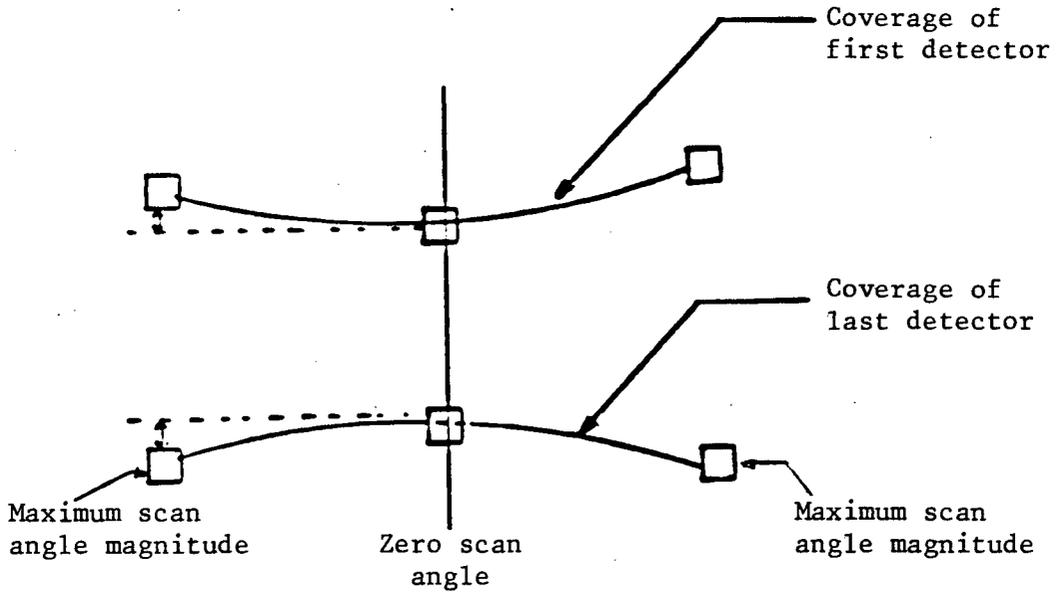


FIGURE A.5 BOWTIE EFFECT

If scan skew d is defined as the skew measured from the swath centre, then

$$d = (d_0 - d_1) / 2$$

and it varies between ± 21.5 metres at 685 km altitude and between ± 15.3 metres at 740 km.

5. Bowtie Effect

The bowtie effect is caused by the earth curvature and the way the detectors are positioned in the focal plane. This effect is illustrated in Figure A.5 which shows the positions of the first and last detectors of a particular band. The swath width in the along track direction is narrower in the centre than at both ends, hence causing a line displacement. The line displacement d , as shown in the figure, is 2.3 metres in the worst case. Hence, this gives a scan gap G of 0 to -4.6 metres, depending on the pixel number. Including the effect of altitude change, the range of gap G is from 0 to -4.8 metres.

This distortion is of high frequency which is obvious from the fact that the line displacements are in opposite directions for the first and last detectors.

6. Earth Rotation

Earth rotation causes the line spacing between two swaths to widen. The magnitude of scan gap G at the swath centre (nadir) is

$$G = T \omega_e \cos \theta_i \quad (\text{A.7})$$

where T is the swath time, ω_e is the earth rotation velocity at the equator and θ_i is the satellite inclination. For the northbound and southbound passes, G is always positive.

7. Swath Time Variation

The time taken to sweep one swath of imagery plus repositioning for the sensor to start imaging the next swath of data may vary. The magnitude of variation is not precisely known yet. The effect of this variation is that swaths of data are distorted in the across scan direction. Such distortion is of high frequency since the variation is different between swaths.

8. Summary

Table A.2 shows that the scan gap varies between $-2 \Delta_0$ and $2 \Delta_0$ where Δ_0 is the line spacing within a swath. In the table, the worst case for each scan gap contribution factor is assumed so that the factors all add constructively wherever possible. The total scan gap size in the table does not include swath time variation, the magnitude of which is not known at this moment.

TABLE A.2 WORST CASES OF SCAN GAP SIZE

Altitude	685 km	740 km
Line spacing within a swath $\Delta\theta$	29.14 m	31.48 m
SCAN GAP		
Jitter (3σ)	12.0 m	-12.0 m
Altitude Variation	13.8 m	-23.6 m
Velocity Variation	5.6 m	-7.3 m
Scan Skew	21.5 m	-15.3 m
Bowtie Effect	0 m (swath centre)	-4.8 m (swath edge)
Earth Rotation	5.2 m	5.2 m
Swath Time Variation	not known	not known
Total (not including swath time variation)	58.1 m = 2.0 $\Delta\theta$	-57.8 m = -1.8 $\Delta\theta$

APPENDIX B SPOT SENSORS

B.1 General Description

The SPOT satellite is designed to have two modes of operations : multispectral and panchromatic. The multispectral mode employs a multilinear array (MLA) of detectors. The panchromatic mode also employs a linear array of detectors and sometimes is referred to as panchromatic linear array (PLA). A key feature of the sensors is the provision for off nadir viewing, with the off nadir angle adjustable to $\pm 26^\circ$. This feature provides stereo viewing capability.

The satellite and sensors have the following characteristics:

Orbit inclination	:	98.7°
Nodal period	:	101.5 minutes
Satellite altitude	:	820 km nominal
Swath width	:	60 km nadir pointing 80 km off nadir angle = 26°
MLA pixel resolution	:	20 m nadir pointing 26 m off nadir angle = 26° 3 bands
PLA pixel resolution	:	10 m nadir pointing 13 m off nadir angle = 26°

The sensors have no mechanical moving parts; images are obtained using the "push-broom" scanning mode. Each line of the image is formed by a linear array of 6000 detectors in the panchromatic band, 300 in each of the other three bands; and line scanning is performed electronically.

Successive lines of the image are produced as a result of the satellite motion along its orbit.

B.2 Scan Angle

The detectors are spaced, except for between groups of detectors, equally apart. The scan angle can be expressed mathematically with the aid of Figure B.1 as follows:

Let β_0 = angle of centre pixel made with the vertical,
 β = angle of imaging pixel made with the vertical,
 p = imaging pixel number, the first pixel being 1,
 θ = angle subtended by sensor from the first pixel to
the last pixel after removing the detector gaps,
 r = spacing between two detectors in focal plane,
 f = focal length, and
 N = total number of detectors.

Then, the angle β is given by:

$$\beta = \beta_0 + \tan^{-1} \{ [\tan(\theta/2)] - (p-1)r/f \} \quad (\text{B.1})$$

B.3 Resolution

The resolution in the along scan direction is then a function of off nadir pointing angle and is given by:

$$\text{Resolution} = R \frac{d\theta}{d\beta} \Delta\beta$$

where θ and R are defined in Figure B.1. The resolution thus obtained is plotted in Figure B.2.

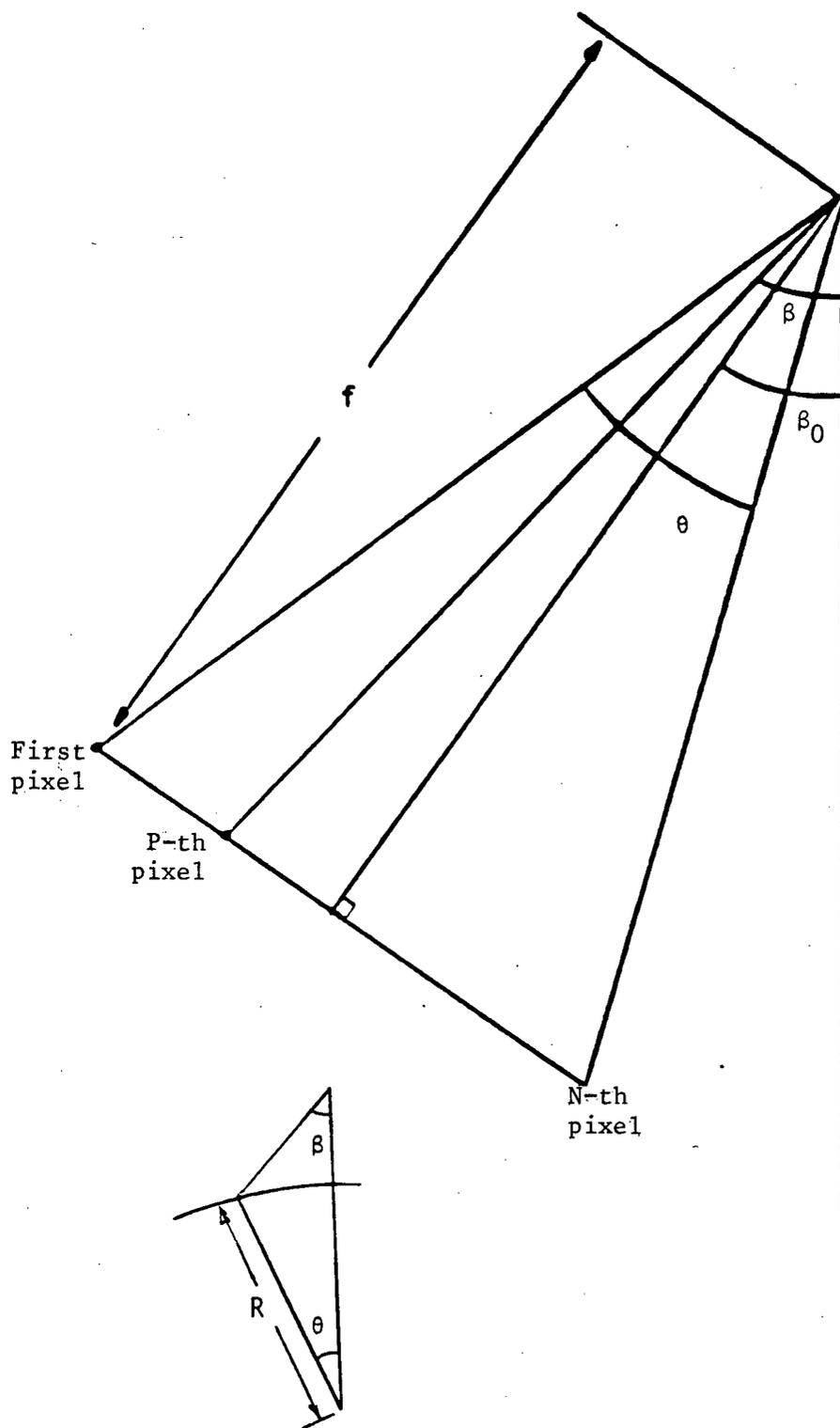


FIGURE B.1 IMAGING PIXEL AND ITS SCAN ANGLE

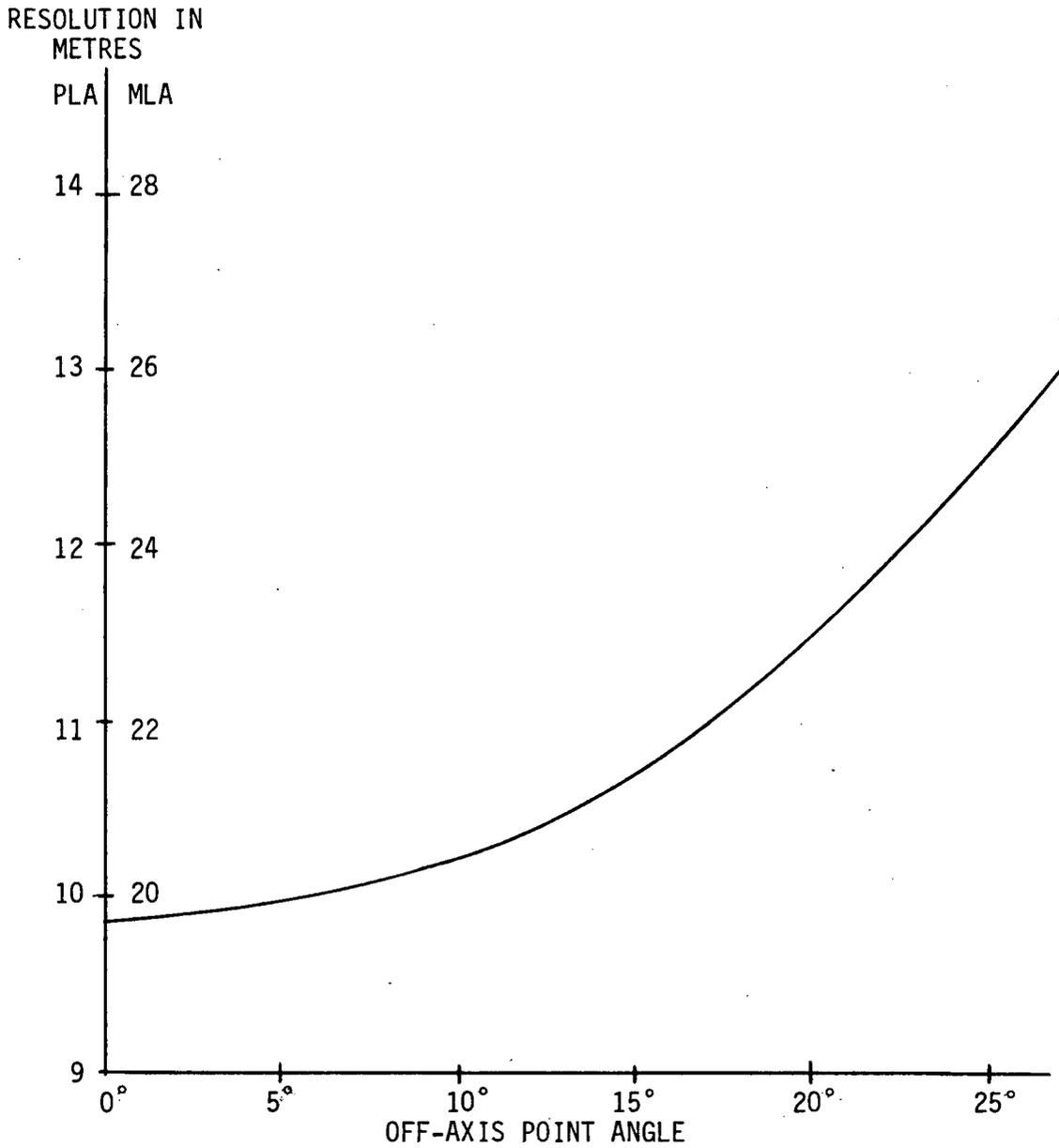


FIGURE B.2 RESOLUTION VERSUS OFF-AXIS POINTING ANGLE IN SPOT IMAGERY

APPENDIX C RECURSIVE ESTIMATION FILTER

C.1 Background

In 1960, R.E. Kalman introduced a set of matrix recursion relations for optimally estimating noise corrupted measurement data. This set of equations is known as the discrete time Kalman filter [LEON70]. The parameters to be estimated are collected in a state vector. If these parameters are time independent, the Kalman filter degenerates to a recursive estimator.

A recursive estimation procedure yields an estimate following each measurement. Specifically, it enables one to estimate an unknown quantity based on the last estimate and the present measurement. In general, the procedure can be formulated as:

$$\underline{x}(i) = k_y y(i) + k_x \underline{x}(i - 1) \quad (C.1)$$

where x is the estimate, y the measurement, i the incremental index number, and k_x and k_y are predetermined coefficient matrices. Such a procedure eliminates the need for storing, and performing calculations on, any amount of past data. Each new measurement is utilized as it is received.

This appendix first describes briefly the underlying principles of recursive filtering using the recursive estimation procedure. Then it shows how the recursive estimator can be applied to the state vector made up of the attitude time series coefficients using GCPs as measurement data. The method can be generalized to update SAR orbit trajectory data

using GCPs or possibly the FM rate as measurement data.

C.2 Recursive Estimator Equations

First the one-dimensional case is considered. Let x_1 and x_2 be two independent estimates of x with variances σ_1^2 and σ_2^2 respectively. Then the best estimate, or minimum variance estimate, can be written as:

$$\hat{x} = (1 - w) x_1 + w x_2 = x_1 + w(x_2 - x_1) \quad (C.2)$$

where the weight w can be obtained by minimizing the variance of x . The variance of x is given by:

$$\sigma^2 = E[(x - E(x))^2]$$

where $E(x)$ denotes the expected value of x . By setting $\partial\sigma^2/\partial w = 0$, the weight w is given by:

$$w = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (C.3)$$

$$\text{Hence } \hat{x} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} = x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (x_2 - x_1) \quad (C.4)$$

$$\text{and } \sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \sigma_1^2 (1 - w) \quad (C.5)$$

These equations can easily be extended to a multi-dimensional case. Furthermore, the two estimates are not of equal dimensions. Let \underline{x} and \underline{y} denote two variables related by:

$$\underline{y} = M \underline{x} \quad (C.6)$$

Let \underline{x}_1 and \underline{y}_2 be the corresponding estimates with covariance P_1 and P_2 respectively. Then the optimal estimate of \underline{x} is given by:

$$\hat{\underline{x}} = \underline{x}_1 + W'M (\underline{y}_2 - \underline{x}_1) \quad (C.7)$$

By minimizing the covariance of \underline{x} , it can be shown that

$$W' = P_1 M^T [M P_1 M^T + P_2]^{-1} \quad (C.8)$$

and the covariance of \underline{x} is given by:

$$P = (I - W'M)P_1 \quad (C.9)$$

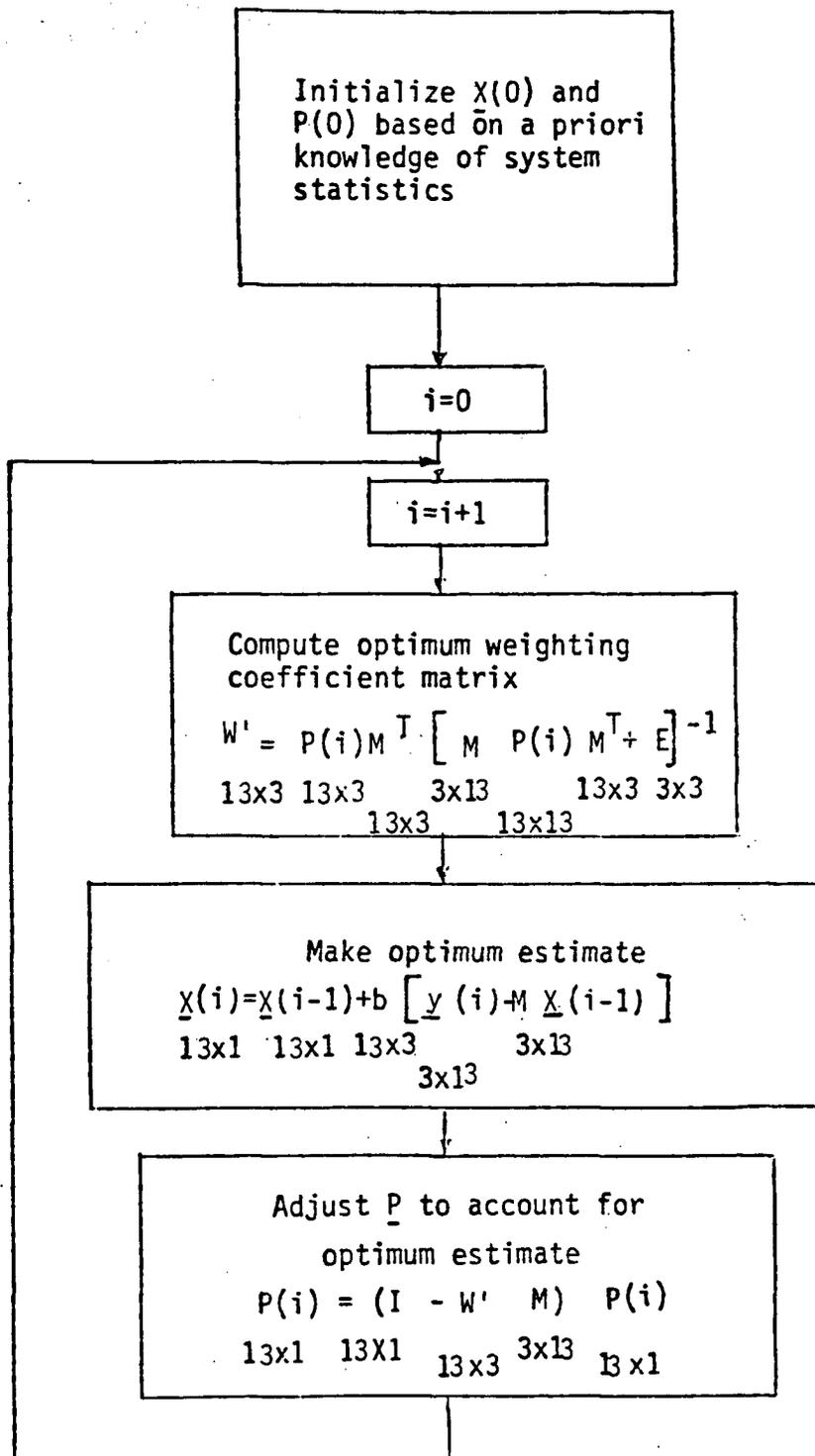
Equations C.7, C.8, and C.9 are the key recursive estimator equations.

C.3 Application to Attitude Time Series Coefficients

Recursive estimation is well suited to the rectification problem with the ECR coordinates of GCPs as measurement data, and the attitude time series and satellite altitude coefficients as the state vector.

The nominal values of pitch, roll, yaw and satellite altitude are first used in the transformation for the first GCP selected. The ECR coordinates of the GCP are computed and compared with the measured (from mapsheets or surveys) values. The difference is then fed to the recursive estimator which refines the coefficients.

Figure C.1 shows how the equations are used in the process. After each GCP is selected, the state vector and its covariance matrix are updated. Here \underline{x} denotes the state vector, P its covariance matrix, \underline{y}



NOTE: THE NUMBER BELOW EACH MATRIX DENOTES ITS DIMENSION. THERE ARE FOUR COEFFICIENTS (CUBIC POLYNOMIAL) FOR EACH OF THE THREE ATTITUDE ANGLES AND ONE FOR SATELLITE ALTITUDE

FIGURE C.1. LEAST SQUARES RECURSIVE FILTER EQUATIONS FOR ATTITUDE TIME SERIES MODEL

the measured ECR coordinates of the current GCP, E the measurement error covariance matrix (the level of confidence in identifying the GCP in the image) and I the identity matrix.

APPENDIX D BILINEAR AND AFFINE TRANSFORMATIONS

This appendix derives the coefficients of the bilinear and affine transformations.

D.1 Bilinear Transformation

Let (x,y) be the output grid and (u,v) be the input grid coordinates; Let the output grid be segmented into rectangular blocks. The bilinear transformation for each block can be represented by:

$$u = a_0 + a_1 \Delta x + a_2 \Delta y + a_3 \Delta x \Delta y \quad (\text{D.1a})$$

$$v = b_0 + b_1 \Delta x + b_2 \Delta y + b_3 \Delta x \Delta y \quad (\text{D.1b})$$

where Δx and Δy are measured from the block's top left corner as shown in Figure D.1. If the mapping of the 4 corners of the block is known, then the coefficients a_i and b_i ($i = 0$ to 3) can be determined as follows:

$$a_0 = u_1$$

$$a_1 = (u_2 - u_1) / X$$

$$a_2 = (u_4 - u_1) / Y$$

$$a_3 = (u_1 + u_3 - u_2 - u_4) / (XY)$$

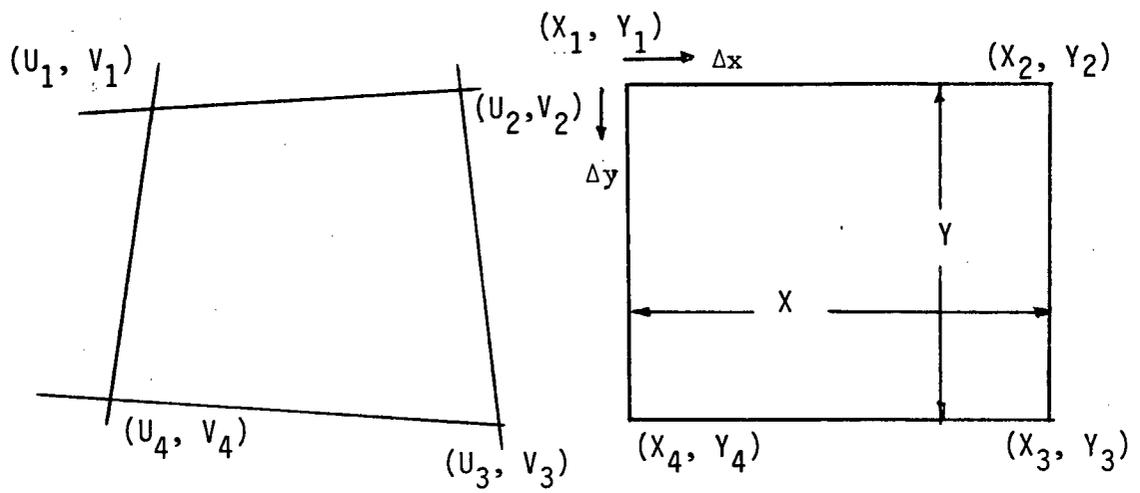
$$b_0 = v_1$$

$$b_1 = (v_2 - v_1) / X$$

$$b_2 = (v_4 - v_1) / Y$$

$$b_3 = (v_1 + v_3 - v_2 - v_4) / (XY)$$

where X and Y are the block dimensions and (u_i, v_i) , $i = 1$ to 4 , are



$$(X_i, Y_i) \xleftrightarrow{\text{MAPPING}} (U_i, V_i), \quad i = 1, 2, 3, 4$$

FIGURE D.1 REMAPPING BLOCK

coordinates of the mapped corner points in the input grid as shown in Figure D.1.

The mapping can easily be proved to be continuous from one block to the surrounding blocks. This can also be deduced from the fact that the same two corner points, common to two neighbouring blocks, are used to evaluate their respective bilinear coefficients.

D.2 Affine Transformation

If a_3 and b_3 are set equal to 0 in Equation D.1, the resulting transformation is affine. Then the other six coefficients are given by a least squares fit:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = (M^T M)^{-1} M^T \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \quad (\text{D.3a})$$

and

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = (M^T M)^{-1} M^T \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \quad (\text{D.3b})$$

where

$$M = \begin{bmatrix} 0 & 0 & 1 \\ X & 0 & 1 \\ 0 & Y & 1 \\ X & Y & 1 \end{bmatrix} \quad (\text{D.4})$$

Upon simplification,

$$a_0 = (-u_1 + u_2 - u_3 + u_4) / (2X)$$

$$a_1 = (-u_1 - u_2 + u_3 + u_4) / (2Y)$$

$$a_2 = u_1 + (-u_1 + u_2 + u_3 - u_4) / 4$$

$$b_0 = (-v_1 + v_2 - v_3 + v_4) / (2X)$$

$$b_1 = (-v_1 - v_2 + v_3 + v_4) / (2Y)$$

$$b_3 = v_1 + (-v_1 + v_2 + v_3 - v_4) / 4$$

(D.5)

This transformation has the restriction that parallel lines are always mapped into parallel lines, but the mapping is not continuous across blocks.

APPENDIX E CLOSED FORM TRANSFORMATION FOR SCANNER IMAGERY

This appendix is to derive a transformation between the sensor instantaneous viewing direction and the geographic coordinates of any image point.

The derivation starts by postulating a unit scan vector viewing an image in the platform coordinates. Then it is transformed into the satellite coordinates and finally into the ECR coordinates. All vectors are in ECR coordinates unless otherwise stated. Relative orientations of various coordinate systems are shown in Figure E.1.

E.1 Basic Formulation

Let \underline{u} be the unit vector pointing from the satellite to a point on the earth, \underline{R}_g the ground point vector and \underline{R}_s the satellite position vector as shown in Figure 4.6. Then

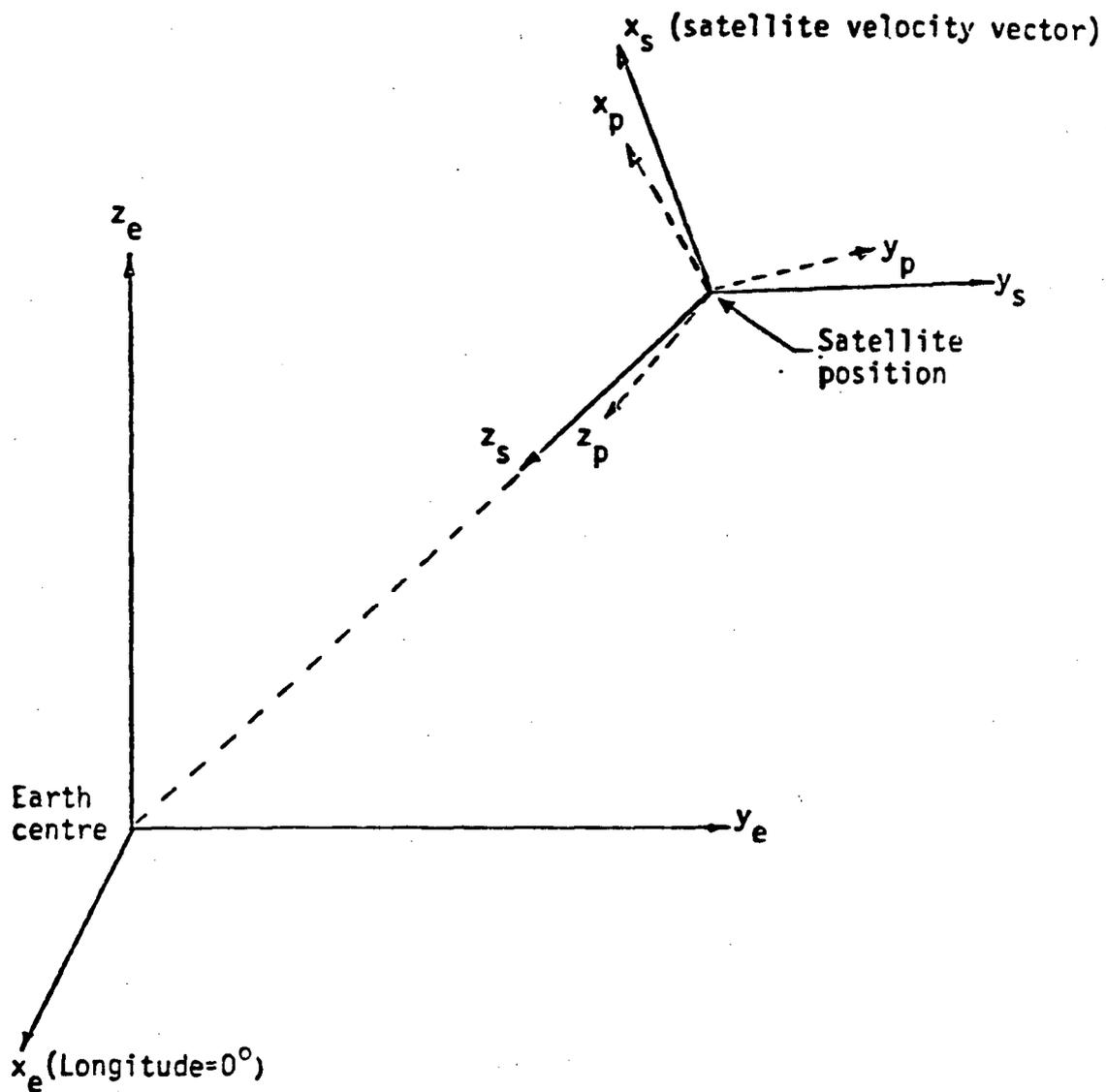
$$\underline{R}_g = \underline{R}_s + D\underline{u} \tag{E.1}$$

where D is the distance between the satellite and the ground point.

The first step is to solve for D. Let χ be the geocentric latitude of the ground point, a the earth's semi-major axis and e its eccentricity. Define e' by:

$$e' = \frac{e^2}{1-e^2} \tag{E.2}$$

It can be shown that



ECR coordinate system: (x_e, y_e, z_e)

Satellite coordinate system: (x_s, y_s, z_s)

Platform coordinate system: (x_p, y_p, z_p)

FIGURE E.1 RELATIVE POSITIONS OF COORDINATE SYSTEMS

$$\left| \underline{R}_g \right|^2 = \left| \underline{R}_s \right|^2 + 2 D \underline{R}_2 \cdot \underline{u} + D^2 \quad (\text{E.3})$$

and

$$\left| \underline{R}_g \right|^2 = a^2 / (1 + e' \sin \chi) \quad (\text{E.4})$$

Let $\underline{R}_g = (R_{gx}, R_{gy}, R_{gz})^T$

then $\sin \chi = R_{gz} / |\underline{R}_g| \quad (\text{E.5})$

Substituting $\sin \chi$ into Equation E.4 and simplifying gives

$$\left| \underline{R}_g \right|^2 = a^2 - e' R_{gz}^2 \quad (\text{E.6})$$

But from Equation E.1

$$R_{gz} = R_{sz} + D u_z$$

where $\underline{R}_s = (R_{sx}, R_{sy}, R_{sz})^T$ and $\underline{u} = (u_x, u_y, u_z)^T \quad (\text{E.7})$

Substituting Equation E.7 into E.6 and simplifying gives

$$\left| \underline{R}_g \right|^2 = a^2 - e' R_{sz}^2 - 2 e' D u_z R_{sz} - e' D^2 u_z^2 \quad (\text{E.8})$$

Combining Equations E.8 and E.3 gives a quadratic equation in D. The roots to this equation are summarized in Equations 4.3 to 4.5.

E.2 Scan Vector in Platform Coordinates

The two instantaneous viewing angles of a detector are deployment angle α and scan angle β as illustrated in Figure E.2. In platform coordinates, the scan vector \underline{D}_u is

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = D \underline{1} = \begin{bmatrix} \sin \alpha \\ \cos \alpha \sin \beta \\ \cos \alpha \cos \beta \end{bmatrix} \quad (\text{E.9})$$

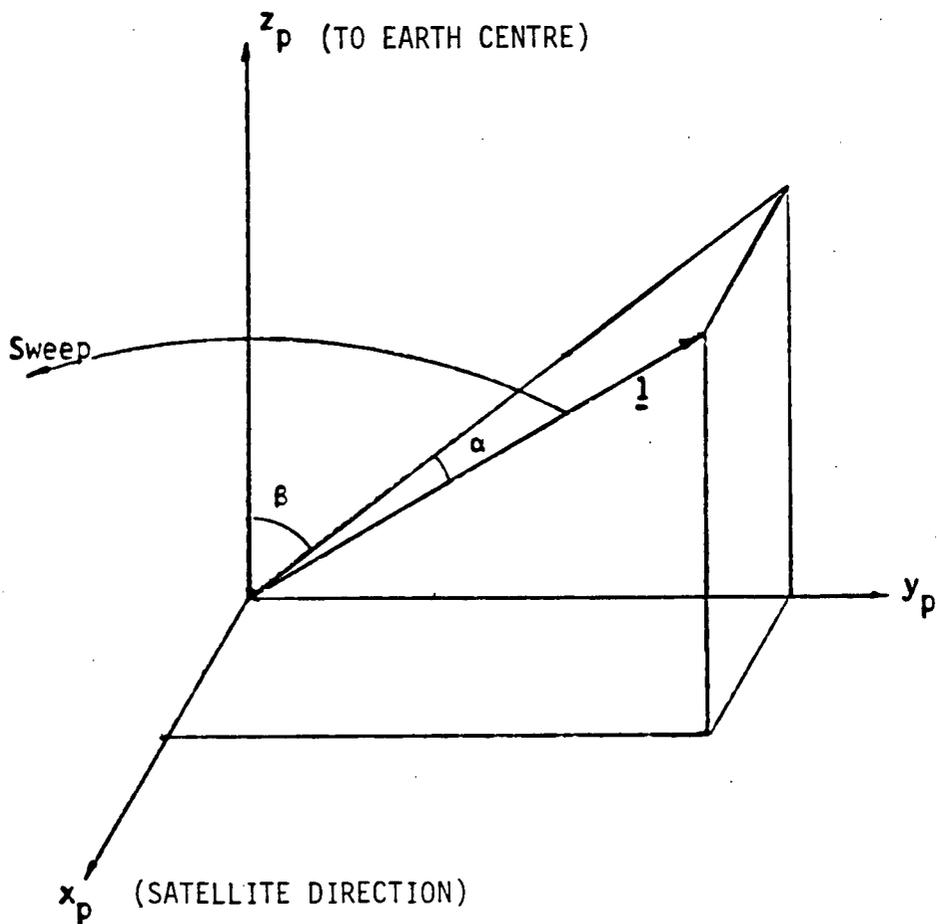


FIGURE E.2 SCAN VECTOR IN PLATFORM COORDINATES

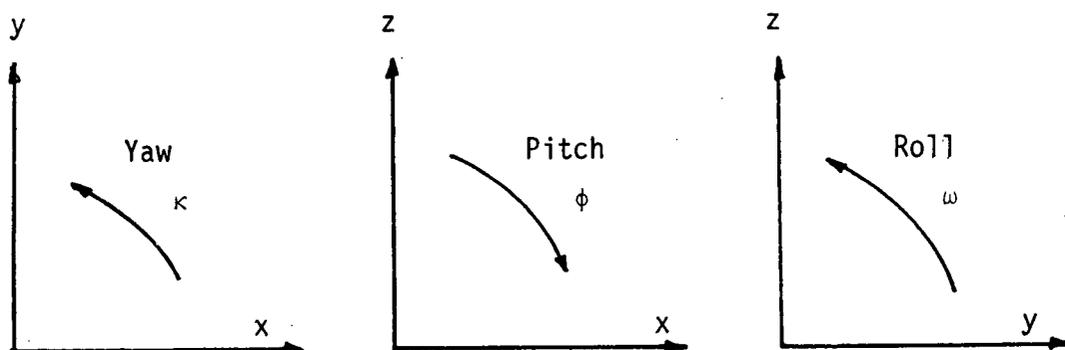


FIGURE E.3 SENSE OF ATTITUDE ROTATION

where $\underline{1}$ is the unit scan vector in platform coordinates.

E.3 Scan Vector in Satellite Coordinates

The platform has to be rotated through a yaw (κ), pitch (ϕ) and roll (ω), not necessarily in this order, to align its axes with those of the satellite coordinate system. This serves to define the platform coordinates with respect to the satellite coordinate system. To conform with photogrammetric convention the order of rotation is taken to be κ , ϕ and ω , and the sense of rotation is shown in Figure E.3. The scan vector $D\underline{u}$ in terms of satellite coordinates is then

$$\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = M_{ps} \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = D M_{ps} \underline{1} \quad (\text{E.10})$$

where

$$M_{ps} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & C\omega & -S\omega \\ 0 & S\omega & C\omega \end{bmatrix} \begin{bmatrix} C\phi & 0 & S\phi \\ 0 & 1 & 0 \\ -S\phi & 0 & C\phi \end{bmatrix} \begin{bmatrix} C\kappa & -S\kappa & 0 \\ S\kappa & C\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{E.11})$$

and where $C\theta = \cos \theta$ and $S\theta = \sin \theta$

Since the magnitude of κ , ϕ and ω are small (less than 1° is sufficient for the following approximation), M_{ps} can be expressed as :

$$M_{ps} = \begin{bmatrix} C\kappa C\phi & -\kappa & \phi \\ \kappa & C\omega C\kappa & -\omega \\ -\phi & \omega & C\phi C\omega \end{bmatrix} \quad (\text{E.12})$$

E.4 Scan Vector in ECR Coordinates

Before the scan vector can be expressed in ECR coordinates the satellite coordinate system has to be defined with respect to the ECR coordinate system. The following assumptions are made:

- (i) The z-axis points towards the earth's centre.
- (ii) The x-axis points in the direction of the satellite velocity vector which is given by the satellite heading, including earth rotation effect, with respect to the geographic north.

The transformation between ECR (x_e, y_e, z_e) and satellite coordinates (x_s, y_s, z_s) is given by:

$$\begin{bmatrix} x_e \\ y_e \\ z_e \end{bmatrix} = M_{se} \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} + \underline{B}$$

where M_{se} can be derived with the aid of Figure E.4 to give

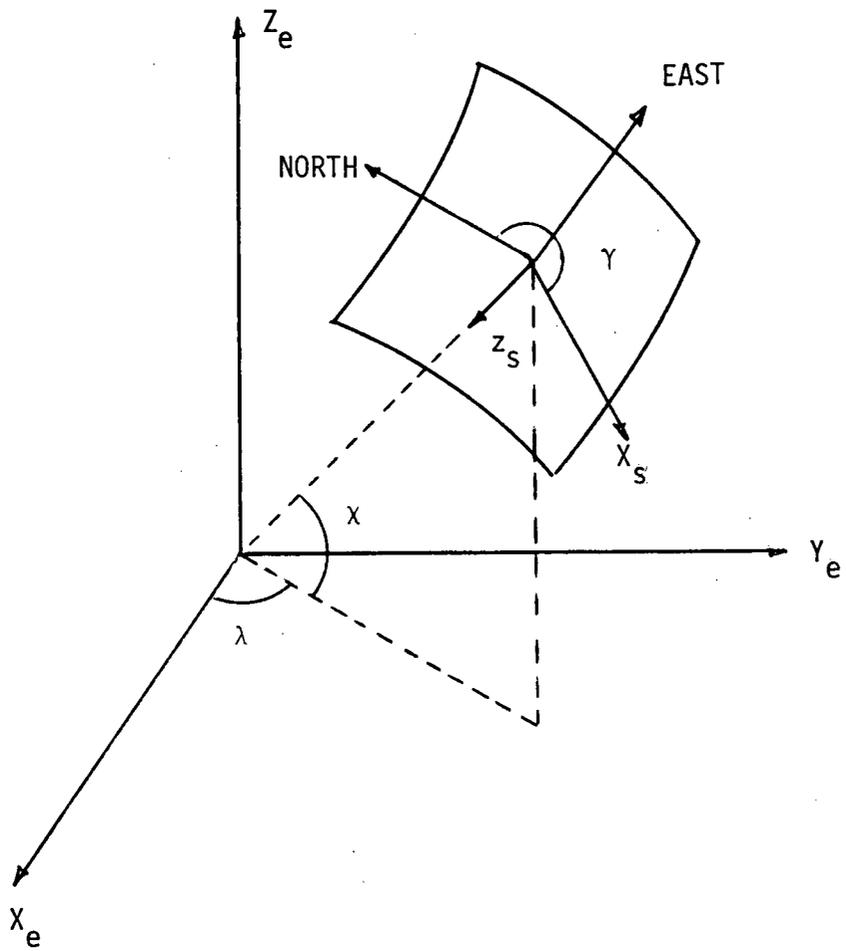


FIGURE E.4 SATELLITE COORDINATES TO ECR COORDINATES

$$\begin{aligned}
 M_{se} (1,1) &= -S\lambda S\gamma - S\chi C\lambda C\gamma \\
 M_{se} (1,2) &= M_{se} (2,3) M_{se} (3,1) - M_{se} (3,3) M_{se} (2,1) \\
 M_{se} (1,3) &= -C\chi C\lambda \\
 M_{se} (2,1) &= C\lambda S\gamma - S\chi S\lambda C\gamma \\
 M_{se} (2,2) &= M_{se} (3,3) M_{se} (1,1) - M_{se} (1,3) M_{se} (3,1) \\
 M_{se} (2,3) &= -C\chi S\lambda \\
 M_{se} (3,1) &= C\chi C\gamma \\
 M_{se} (3,2) &= M_{se} (1,3) M_{se} (2,1) - M_{se} (2,3) M_{se} (1,1) \\
 M_{se} (3,3) &= -S\chi
 \end{aligned}
 \tag{E.13}$$

where γ is the satellite heading angle, χ is the geometric latitude and λ is the longitude of the satellite as shown in Figure E.4. The vector \underline{B} is the displacement vector between the satellite coordinate system and the ECR coordinate system.

APPENDIX F THEORY OF SAR AND SAR DATA PROCESSING

This appendix summarizes principles of SAR and SAR data processing.

F.1 Formation of SAR

Figure F.1 depicts a radar system carried by a spacecraft. The function of the radar is to obtain a high resolution map of the terrain as shown in the figure. High range resolution can be obtained by transmitting pulsed chirp radar signals and matched filtering the return signals.

Resolution in azimuth is in principle limited by the half power beamwidth (HPBW) of the antenna in the azimuth direction. The HPBW at range R is roughly $\lambda R/D$ where λ is the wavelength of operation and D is the antenna aperture. Hence the resolution improves with increase in aperture. However the physical size of the antenna that can be carried on board the spacecraft is limited, and hence limiting the azimuth resolution.

Azimuth resolution, higher than the conventional limit of $\lambda R/D$ can be attained by the synthetic aperture technique. If radar pulses are transmitted from a sequence of positions along the flight path, the return signal from a target will be a chirp waveform in the azimuth direction. If the chirp's frequency modulation (FM) rate is known, the signal can be matched filtered to achieve high azimuth resolution. The resolution is equivalent to synthesizing an aperture that can be thousands of metres long.

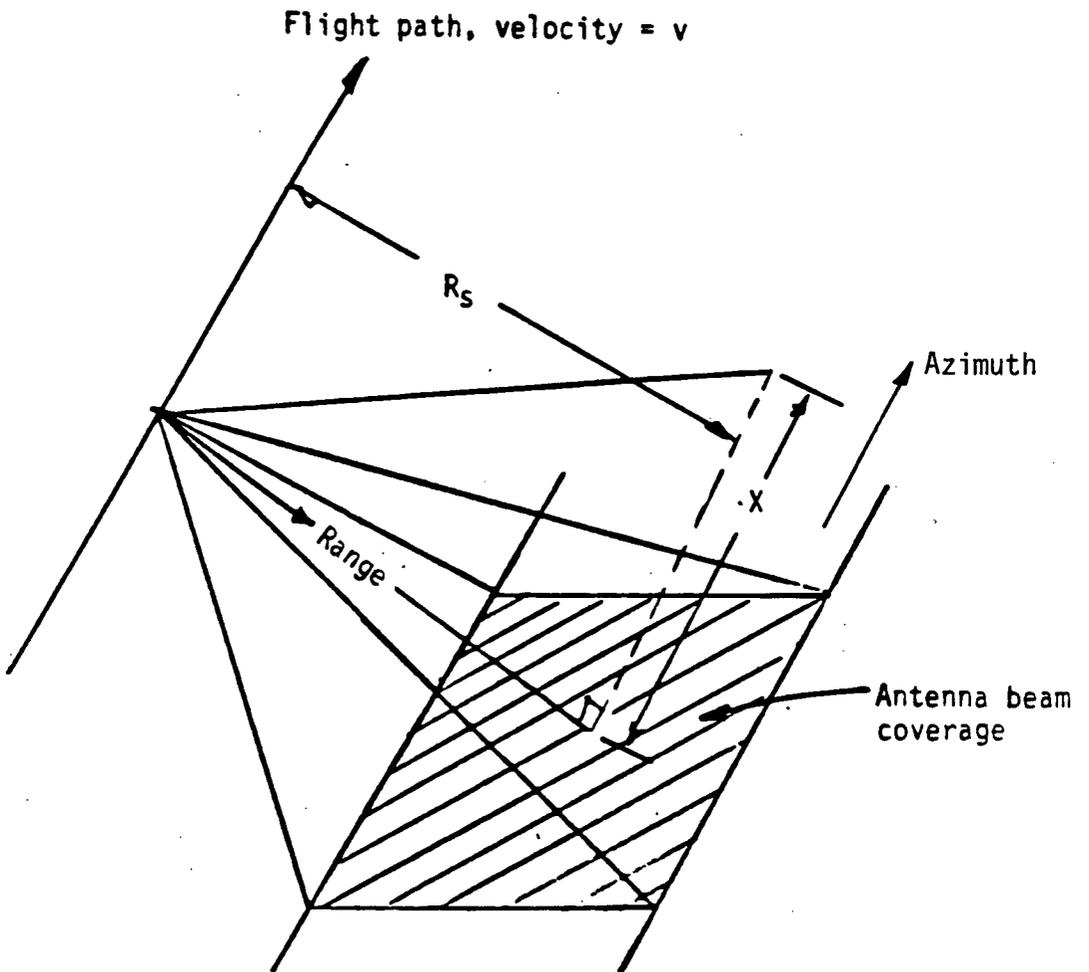


FIGURE F.1 SAR GEOMETRY

F.2 Resolution in Range

Consider the case in which an unmodulated pulse of length τ and fixed frequency is transmitted. The range resolution is given by

$$q_r = c \tau / 2 \quad (\text{F.1})$$

where c is the velocity of propagation.

An improvement in range resolution can be obtained by transmitting a short pulse which requires more transmitter peak power. Equipment peak power limitations therefore limit the obtainable range resolution.

The resolution limit for an unmodulated pulse is limited to the value given by Equation F.1 because the beginning and the end of the pulse are indistinguishable. If the frequency across the pulse is made to vary, then the returned pulse at various positions can be distinguished by virtue of the frequency received. This provides the motivation for studying frequency modulated pulses. For a chirp radar, the transmitted signal is a linear FM pulse:

$$f_0(t) = \text{rect}(t/\tau) \cos \left[2 \pi (f_0 t + K t^2 / 2) \right] \quad (\text{F.2})$$

where t = time

f_0 = carrier frequency

K = FM rate

$$\text{rect}(x) = \begin{cases} 1, & |x| < 1 \\ 0, & \text{elsewhere} \end{cases}$$

The instantaneous frequency of the pulse is $f_0 + K\tau$. Ignoring a multiplicative constant and the rect function the return signal of a target with slant range r is

$$f_r(t) = \cos 2\pi \left[\left[f_0 t - \frac{2r}{c} \right] + \frac{K}{2} \left(t - \frac{2r}{c} \right)^2 \right] \quad (\text{F.3})$$

The baseband signal, obtained by beating $f_r(t)$ with $\cos 2\pi f_0 t$ and followed by low pass filtering is

$$f_{rB}(t) = \cos 2\pi \left[\frac{k}{2} \left(t - \frac{2r}{c} \right)^2 \right] \quad (\text{F.4})$$

The matched filter output, ignoring a multiplicative factor, is a sampling function

$$f_{rm} = \text{sinc} [\pi K \tau (t - 2r/c)] \quad (\text{F.5})$$

where $\text{sinc } x = \sin x/x$. The function f_{rm} is plotted in Figure F.2. The effective pulse width is the -3db width of the main lobe and is given by

$$\tau_{\text{eff}} = 0.886/(K\tau) \quad (\text{F.6})$$

Range resolution is therefore

$$\rho_r = \frac{c \tau_{\text{eff}}}{2} = \frac{c}{2} \frac{0.886}{K \tau} \quad (\text{F.7})$$

For Seasat, $K = 562 \times 10^9/\text{sec}^2$ and $\tau = 33.9 \mu\text{sec}$, hence $\tau_{\text{eff}} = 0.0465 \mu\text{sec}$ and $\rho_r = 7$ metres (compare to $q_r = 5085$ metres). Therefore by transmitting the linear FM pulse, the gain is 730 in range resolution.

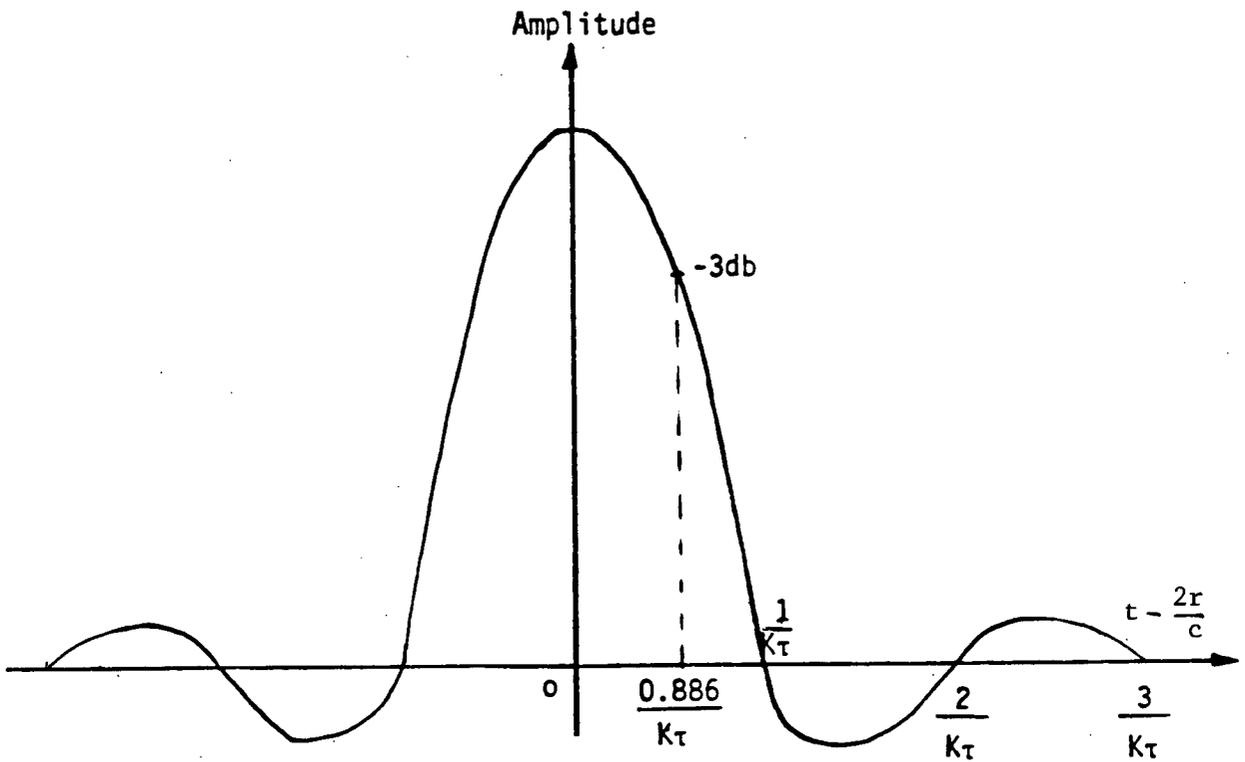


FIGURE F.2 MATCHED FILTER OUTPUT

F.3 Azimuth Resolution

For simplicity, let the transmitted wave form be:

$$f_1(t) = \cos 2 \pi f_0 t \quad (\text{F.8})$$

then the return signal is

$$f_a(t) = \cos 2 \pi f_0 (t - 2r/c) \quad (\text{F.9})$$

and the baseband signal is

$$f_{aB}(t) = \cos 2 \pi f_0 (2r/c) \quad (\text{F.10})$$

Figure F.1 shows the SAR geometry in which

$$r = \sqrt{R_s^2 + x^2} \quad (\text{F.11})$$

$$\text{For } R_s \gg x, r \approx R_s + x^2 / (2R_s) \quad (\text{F.12})$$

$$\text{Also } x = vt \quad (\text{F.13})$$

where v is the velocity of spacecraft. Hence

$$f_{aB}(t) = \cos 2\pi f_0 [2R_s + v^2 t^2 / (cR_s)] = \cos [\theta + 2\pi f_0 v^2 t^2 / (cR_s)] \quad (\text{F.14})$$

where θ is a constant phase angle which is irrelevant to the subsequent development. Hence the return is a linear FM signal by virtue of vehicle motion. Comparing this with Equation F.4, the azimuth FM rate is

$$K_a = \frac{2f_0 v^2}{cR_s} \quad (\text{F.15})$$

For Seasat $K_a = 520/\text{sec}^2$. The time duration in which the target stays in the antenna beam (within the HPBW) is

$$T_s = R_s \lambda / (D v) \quad (\text{F.16})$$

The matched filter output is again a sampling function

$$f_{am} = \text{sinc} (\pi K_a T_s t) \quad (\text{F.17})$$

The plot of this function is similar to that of f_{rm} in Figure F.2. The function f_{rm} then shows that a high resolution can be obtained if the target remains illuminated for a longer period of time.

The azimuth resolution, similar in form to the range resolution given by Equation F.7 is

$$\rho_a = \frac{0.886 v}{K_a T_s} \quad (\text{F.18})$$

Substituting Equation F.15 and F.16 into F.17 and noting that $c = \lambda f_0$, the azimuth resolution is

$$\rho_a = 0.886 D / 2 \quad (\text{F.19})$$

For the Seasat antenna, $D = 10.7$ metres. Hence $\rho_a = 4.74$ metres. The conventional resolution ($\rho_a = \lambda R_s / D$) is 18.7 km. The gain is therefore 3943 times in azimuth resolution. The length of the array synthesized is 43.2 km.

Characteristics of SAR can be summarized by Equation F.19. They are:

- (a) Azimuth resolution is higher with a shorter antenna. This is due to the fact that the target stays illuminated for a longer period of time for a shorter antenna (larger HPBW).
- (b) The azimuth resolution is range independent. Normally the farther the sensor is from the target, the lower the resolution. However, this is exactly counter balanced by the fact that the target stays illuminated for a longer period of time when the sensor is further away.

F.4 Two-Dimensional Signal

If the transmitted pulse is $f_0(t)$ as given by Equation F.2 and is repeated once every T seconds, then the return will be a two-dimensional (nT and r , where n is an integer) signal given by:

$$f_r(nT) = \cos 2 \pi \left[f_0 \left(t - \frac{2r(nT)}{c} \right) + \frac{K}{2} \left(t - \frac{2r(nT)}{c} \right)^2 \right] \quad (\text{F.20})$$

Again a multiplicative constant has been ignored and in the equation assumes an integral value. The baseband signal is

$$g_B(\eta) = \cos 2 \pi \left[- \frac{2f_0 r(\eta)}{c} + \frac{K}{2} \left(t - \frac{2r(\eta)}{c} \right)^2 \right] \quad (\text{F.21})$$

where η has replaced nT which represents time in the azimuth direction. This is a linear FM signal in both directions. The matched filter output is the product of f_{am} and f_{rm} , i.e., a two-dimensional sampling function.

$$g_m(\tau, \eta) = \text{sinc} [\pi K \tau (t - 2r/c)] \text{sinc} [\pi K_a T_s \eta] \quad (\text{F.22})$$

where K_a is again given by Equation F.15. In the derivation of Equation F.2 it has been assumed that

$$r(\eta) = \sqrt{R_s^2 + x^2} \approx R_s + v^2 \eta^2 / (2 R_s) \quad (\text{F.23})$$

in which the closest approach of the target is assumed to be at $\eta = 0$.

An equal phase plot of g_B and g_m is shown in Figure F.3. Therefore the fundamental operations to obtain SAR imagery from raw data (or

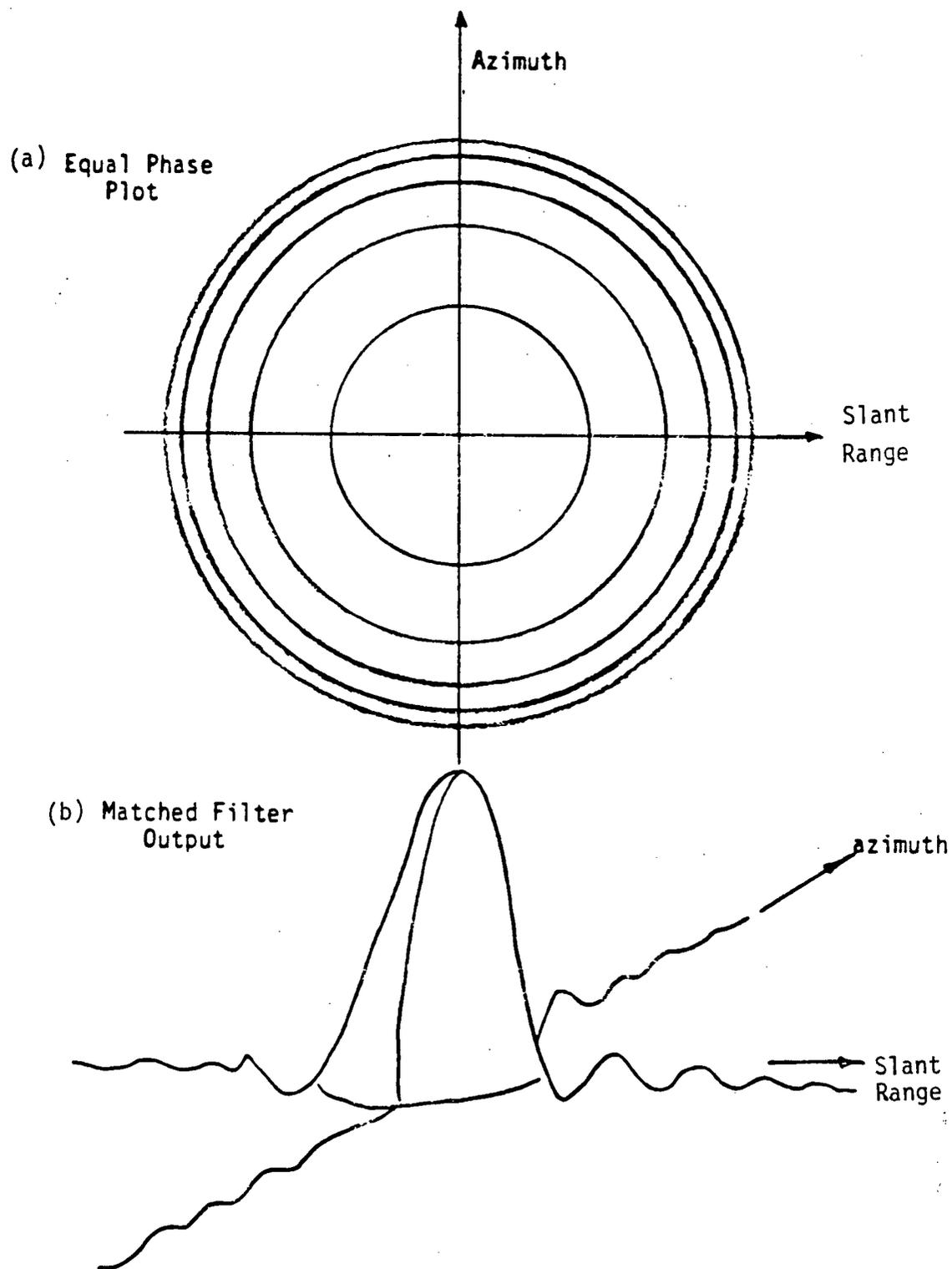


FIGURE F.3 TWO DIMENSIONAL SIGNAL AND MATCHED FILTER OUTPUT

baseband signal) is to match filter the data in range and azimuth. These operations are called range compression and azimuth compression respectively.

F.5 The Principal Problem in Satellite SAR Processing [MDA77b]¹

In Equation F.11, the Pythagorus theorem $r^2 = R^2 + x^2$ has been invoked, thus assuming that the ground has no curvature. Also, x has been expressed as v and this assumes that the ground is stationary. However, both of these assumptions are incorrect for any planet. Taking the shape of the earth and its rotation into account

$$r \simeq R_s + B \eta^2 / (2 R_s) \quad (\text{F.24})$$

where B is a function of trajectory data, earth curvature and rotation. The azimuth FM rate is then given by:

$$K_a = 2 B / (\lambda R_s)$$

Therefore for a flat and non-rotating earth, $B = v^2$.

The target trajectory, by Equation F.24, is a parabola and is plotted in Figure F.4. The equal phase plot of the baseband signal is also shown in Figure F.4.

The range compression is performed along lines of constant η and produces an output on each η line, a sampling function given by Equation F.5. The sampling function for all η lines, as shown in the equation, is distributed along the locus $t = 2r(\eta)/c$. As a result, the azimuth

¹ Sections F.5 and F.6 are abridged from this reference.

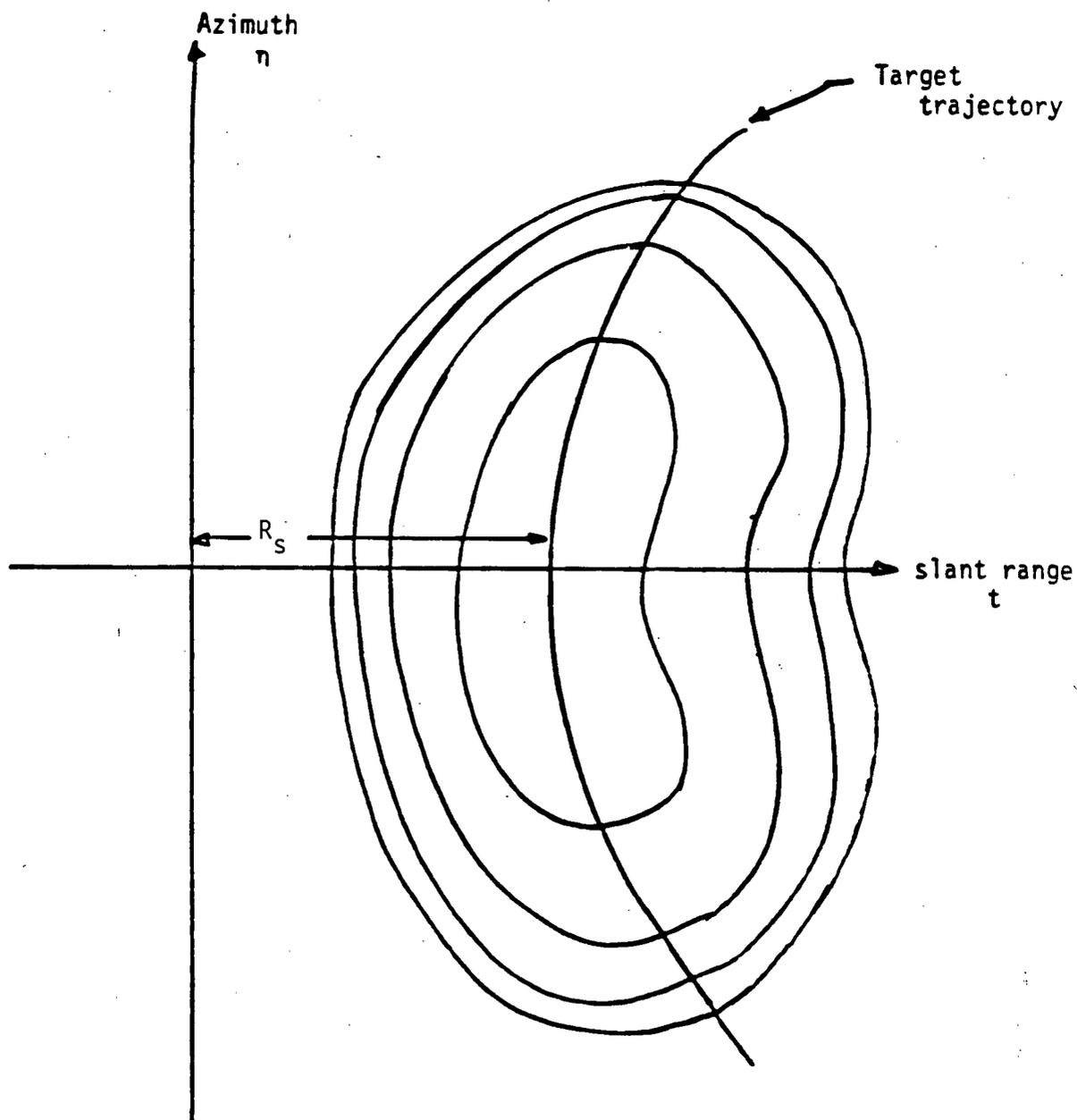


FIGURE F.4 TARGET TRAJECTORY AND EQUAL PHASE PLOT

compression must be performed along lines of constant $t - 2r(\eta)/c$ as shown by the trajectory in Figure F.4. While this poses little problem mathematically, it can impose considerable implementation difficulty depending upon the desired resolution and the vehicle / terrain geometry. In the case of airborne SAR, the maximum duration of $t - 2r(\eta)/c$ is normally less than 2 resolution cells and azimuth compression can, in practice, be performed along lines of constant range. However, in the satellite borne case, this deviation can be up to 100 resolution cells. The azimuth compression must then be performed along lines that "wander" through the two-dimensional data. This problem is referred to as the range cell migration (RCM) problem and is the principal problem in satellite borne SAR data processing.

F.6 Matched Filtering

The procedures to process SAR data are range compression, RCM correction and azimuth compression.

The data is received a range line at a time and is immediately range compressed. This can readily be done since the FM rate of the transmitted pulse is a function of system design. Once range compressed, data is then corner turned so that it can be read an azimuth line at a time.

The RCM correction is then performed by an interpolation process as shown in Figure F.5. Azimuth matched filtering is performed on the interpolated data array which effectively represents a target point locus.

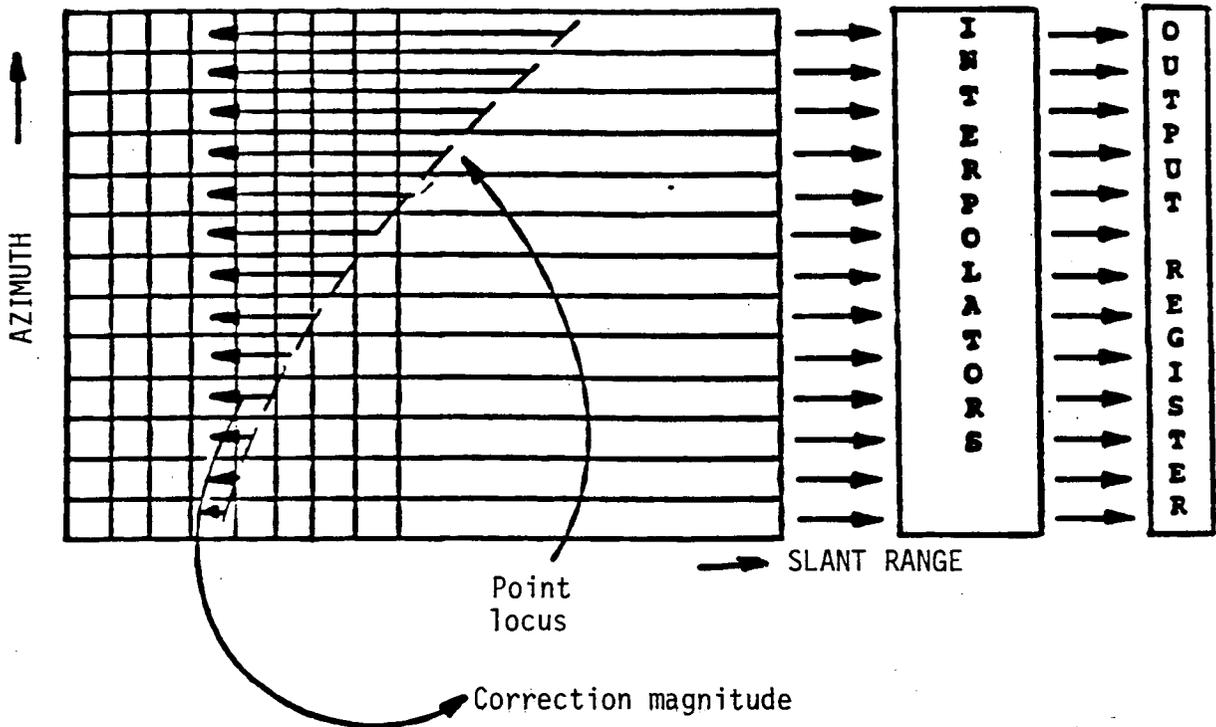


FIGURE F.5 RANGE CELL MIGRATION CORRECTION [MDA77b]

The azimuth FM rate is a function of target position, satellite position and velocity.

The antenna illumination pattern is not necessarily exactly broad-side (i.e., not exactly right angle to the satellite velocity vector) due to a combination of earth rotation velocity and satellite pitch, roll and yaw. The beam centre line lies along the $\eta = 0$ line (time closest approach to target) only in the fictional case of a stationary earth and zero pitch and yaw or in the rare case where the three effects cancel. A more typical case is depicted in Figure F.5. It is necessary to restrict the impulse response duration of the matched filter as much as possible, confining it to the illuminated position only. As shown in Figure F.6, targets (a) and (b) are matched filtered at the illuminated portions of their trajectories and the outputs are placed at locations A and B respectively. This is frequently referred to as compression to zero Doppler since at $\eta = 0$ all targets exhibit zero Doppler velocity. Benson [BENS80] at CCRS, Ottawa, has approximated the target trajectory by a straight line. This reduces the number of arithmetic operations in processing the SAR data, but at the expense of degrading the final resolution.

The beam centre position can be obtained by Doppler centroid tracking [MDA79] which estimates the centroid of the antenna pattern from Doppler spectral analysis. It can be shown that the peak of the mean spectral distribution occurs at the Doppler frequency corresponding to the peak of the antenna azimuth pattern. The Doppler frequency is a linear function of azimuth distance.

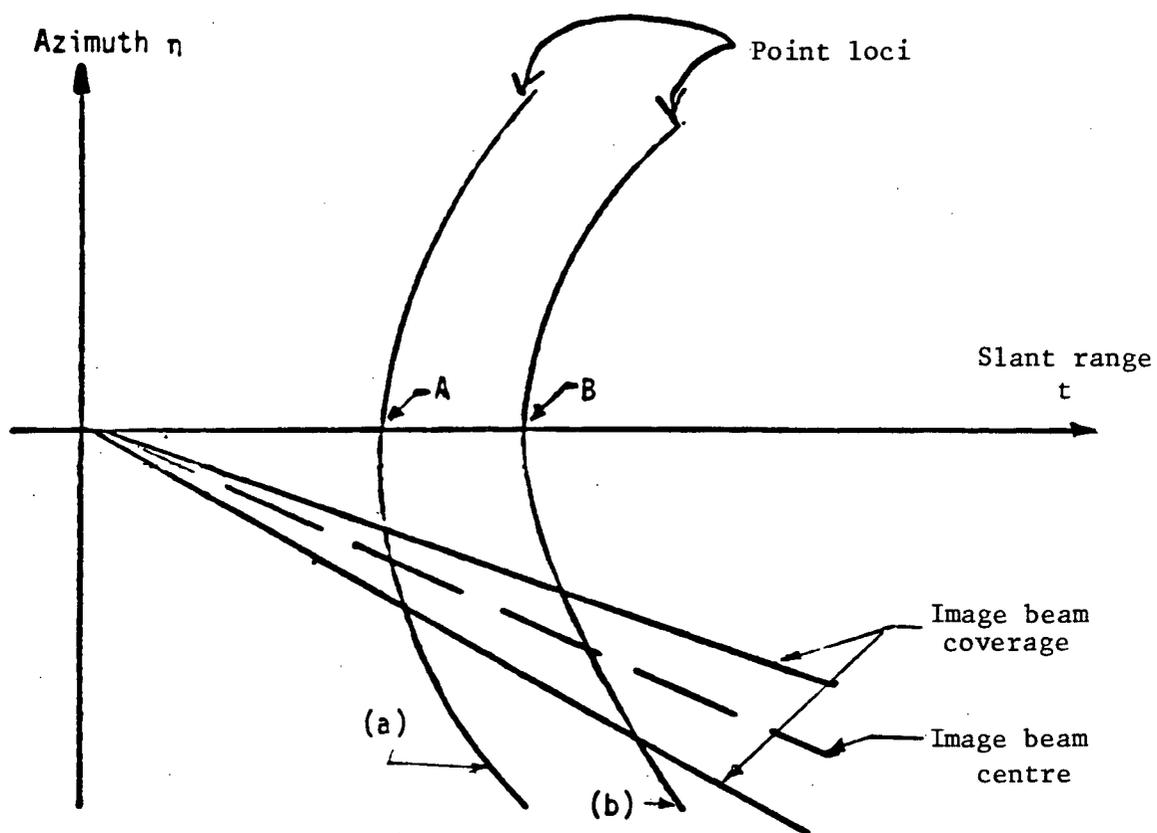


FIGURE F.6 A TYPICAL BEAM ORIENTATION

F.7 Multi-Look Processing

High resolution SAR imagery suffers from a 'scintillation' problem. A reduction of the scintillation can be achieved by superimposing images taken from different angles to the reflecting surface (different "look" angles). In effect, a number of coherently generated subimages are incoherently summed to reduce the scintillation in the final image. A system of this nature is called a "multi-look" system. The multiple looks can be achieved by separately azimuth compressing the returns from different portions of the antenna horizontal beamwidth and incoherently summing the results in a "post detection integrator". This reduces the Doppler bandwidth available for each look by a multiplicative factor equal to the inverse of the number of looks. This degrades the final resolution but reduces scintillation by a factor in the order of \sqrt{N} where N is the number of looks.

Figure F.7 shows the antenna beam is divided into subbeams in a 4-look situation and the method can be generalized to any number of looks. Matched filtering is done for each subbeam. Let their outputs be I_1 , I_2 , ..., I_N , where N is again the number of looks. Then the looks are summed incoherently to give an output:

$$I(t) = I_1 + I_2 + \dots + I_N \quad (\text{F.26})$$

If the absolute magnitude of each output is not taken, the effect is coherent summation corresponding to 1-look processing. It can be shown that the incoherent summation gives

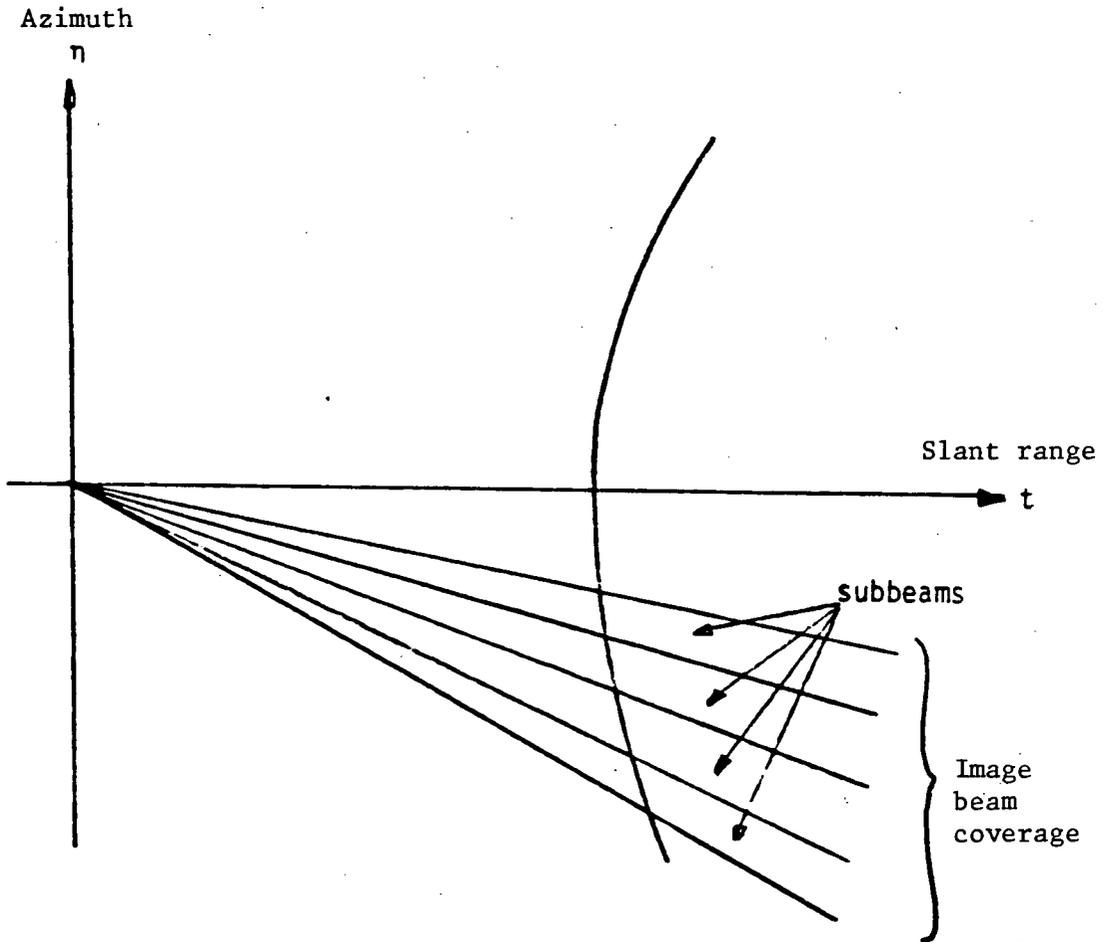


FIGURE F.7 FOUR-LOOK PROCESSING

$$I(t) \propto N \operatorname{sinc}(\pi K_a T_s t / N) \quad (\text{F.27})$$

Hence the resolution for N looks decreases to N times that for 1-look.

For Seasat imagery, the resolution for 1-look is 4.74 metres. Hence for 4-looks it is $4.74 \times 4 = 19$ metres. Due to a sidelobe suppression technique which degrades the resolution¹ somewhat, the final azimuth resolution is about 25 metres.

¹ Range resolution is 7 metres in slant range and this corresponds to 19 metres in ground range. Sidelobe suppression in the range direction degrades the range resolution to about 25 metres.

APPENDIX G CLOSED FORM RECTIFICATION TRANSFORMATION OF SAR IMAGERY

This appendix summarizes the geometric rectification transformation of SAR imagery from line, pixel coordinates to map projection coordinates.

G.1 Slant Range/Azimuth to Map Projection Transformation

The process of transforming from line/pixel coordinates to the output map projection coordinates is performed by a series of discrete transformations, beginning with the line/pixel coordinates in the input slant range image, as follows:

u, v	Line/pixel in input image coordinates
τ_m, t_r	Slant range, azimuth coordinates
x_s, y_s, z_s	Satellite coordinates
x_e, y_e, z_e	Earth centred rotating (ECR) coordinates
x, y	Map projection coordinates

Note that τ_m and t_r are in units of time.

The transformations are the same as those presented in Appendix E except for input image coordinates to satellite coordinates. These discrete transformations are specified in order below.

G.2 Input Image to Satellite Coordinate System

Given a target's line and pixel coordinates in image memory, its slant range and azimuth coordinates can be computed from sampling rates in both directions. If this target has a Doppler frequency f_D , the target's coordinates (x_s, y_s, z_s) in the satellite coordinate system can be solved from the intersection of three surfaces:

Doppler surface

$$m_x x_s + m_y y_s + m_z z_s = f(f_D) \quad (G.1)$$

Wave front surface (free space, $c = \text{constant}$)

$$x_s^2 + y_s^2 + z_s^2 = (cT_m/2)^2 \quad (G.2)$$

Planet surface

$$x_s^2 + y_s^2 + (z_s - H)^2 = r^2 \quad (G.3)$$

where

$f(f_D)$ = a function of Doppler frequency,

H = satellite distance from centre of earth,

r = planet radius at target location,

c = velocity of propagation in free space, and

m_x, m_y, m_z = known functions of spacecraft velocity and position.

It can be shown that

$$\begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} = -M_{se} \frac{dR}{dt}$$

Here M_{se} is the rotation matrix from satellite coordinates to ECR coordinates presented in Equation E.11 and \underline{R}_s is the satellite position vector in ECR coordinates (see Section E.1).

In the satellite coordinate system, the z_s axis points to the centre of the earth and x_s axis lies along the satellite velocity vector. By solving the equations representing the three surfaces, it can be shown that

$$\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} [f(f_D) - m_x x_s - m_z z_s] / m_x \\ \pm \sqrt{(c\tau_m/2)^2 - y_s^2 - z_s^2} \\ [(c\tau_m/2)^2 + H^2 - r^2] / (2H) \end{bmatrix} \quad (G.4)$$

The value of Y_s is positive for a radar clock angle of 270° and negative for a clock angle of 90° . The value of r is target position dependent and requires the use of a planetary datum, plus the planetary coordinates of the target, and incorporates a DTM.

G.3 Satellite Coordinates to Earth Centred Rotating Coordinate System

The coordinates of the target in the ECR coordinate system are given by:

$$\begin{bmatrix} x_e \\ y_e \\ z_e \end{bmatrix} = M_{se} \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} + \underline{B} \quad (G.5)$$

where the rotation matrix M_{se} shown in Equation E.12 is a function of satellite position. The vector \underline{B} is the displacement vector between the

satellite coordinate system and the ECR coordinate system.

G.4 ECR to Map Projection Transformation

The map projection coordinates (x, y) are functions of (x_e, y_e, z_e) :

$$x = g_1(x_e, y_e, z_e)$$

$$y = g_2(x_e, y_e, z_e)$$

The functions g_1 and g_2 depend on the map projection specified.

APPENDIX H REFRACTION OF ELECTROMAGNETIC WAVE

H.1 Introduction

This appendix describes a method to trace an electromagnetic wave path when entering a planet's atmosphere, taking refraction into account. Refraction is caused by the difference in speed of propagation at different altitudes due to change in atmospheric density. The velocity of propagation is modelled by

$$c' = c_0 \exp(\beta z_s) \quad (\text{H.1})$$

where c_0 is the velocity of propagation at satellite altitude and β is the velocity attenuation coefficient due to the planet's atmosphere.

Referring to Figure H.1, expressions are derived for z_s , x_s and R for a given θ_0 , and the relation between θ_0 and θ_n is determined. The values of x_s and z_s are used in Section 6.4.3.

H.2 Expression for z_s

If the atmosphere is divided into infinitesimally thin layers, the law of refraction states that (Figure H.2):

$$\frac{\sin \theta_0}{c_0} = \frac{\sin \theta_1}{c_1} = \dots = \frac{\sin \theta_i}{c_i} = \dots \quad (\text{H.2})$$

Therefore at any layer i

$$\sin \theta_i = (c_i / c_0) \sin \theta_0 \quad (\text{H.3})$$

and vertical displacement is given by:

$$dz_s = (dr) \cos \theta_i$$

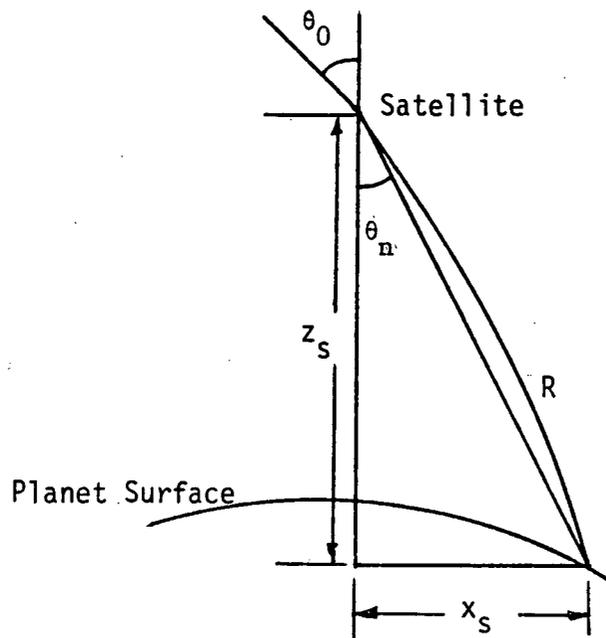


FIGURE H.1 REFRACTION OF ELECTROMAGNETIC WAVE

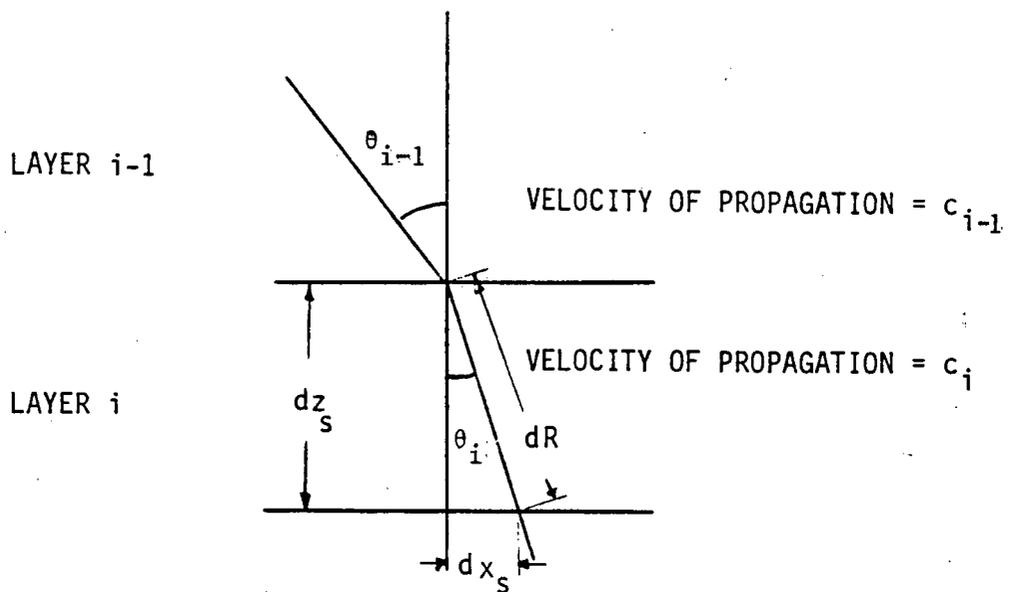


FIGURE H.2 REFRACTION IN A THIN LAYER

But $dr = c_i dt$, therefore

$$dz_s = c_i \cos \theta_i dt$$

From Equations H.1 and H.2,

$$\begin{aligned} \cos \theta_i &= \sqrt{1 - [c_i / c_0] \sin \theta_0]^2} \\ &= \sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0} \end{aligned} \quad (\text{H.4})$$

Therefore,

$$dz_s = c_0 \exp(-\beta z_s) \sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0} dt$$

and

$$dt = \frac{\exp(\beta z_s) dz_s}{c_0 \sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0}}$$

so that

$$\begin{aligned} t &= \int_0^{z_s} \frac{\exp(\beta z_s) dz_s}{c_0 \sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0}} \\ &= \frac{1}{c_0} \exp(\beta z_s) \sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0} - \cos \theta_0 \end{aligned} \quad (\text{H.5})$$

$$\text{Hence } z_s = \frac{1}{2\beta} \ln [(t c_0 + \cos \theta_0)^2 + \sin^2 \theta_0] \quad (\text{H.6})$$

H.3 Expression for x_s

Referring to Figure H.2, the lateral displacement in layer i is

$$dx_s = (dz_s) \tan \theta_i$$

From Equations H.1 and H.2 and Figure H.2, it can be shown that

$$\begin{aligned} dx_s &= (dz_s) \tan \{\sin^{-1} [\exp(-\beta z_s) \sin \theta_0]\} \\ &= (dz_s) \frac{\exp(-\beta z_s) \sin \theta_0}{\cos \{\sin^{-1} [\exp(-\beta z_s) \sin \theta_0]\}} \end{aligned}$$

$$= (dz_s) \frac{\exp(-\beta z_s) \sin \theta_0}{\sqrt{1 - \exp(-\beta z_s) \sin \theta_0}}$$

Therefore,

$$\begin{aligned} x_s &= \int_0^{z_s} \frac{\exp(-\beta z_s) \sin \theta_0}{\sqrt{1 - \exp(-\beta z_s) \sin \theta_0}} \\ &= \frac{1}{\beta} \{ \theta_0 - \sin^{-1} (\exp(-\beta z_s) \sin \theta_0) \} \end{aligned} \quad (\text{H.6})$$

H.4 Expression for R

Referring to Figure H.2, the path length is

$$dR = dz_s / \cos \theta_i$$

$$dR = \frac{dz_s}{\sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0}}$$

Therefore,

$$\begin{aligned} R &= \int_0^{z_s} \frac{dz_s}{\sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0}} \\ &= \frac{1}{\beta} \ln \frac{1 + \sqrt{1 - \exp(-2\beta z_s) \sin^2 \theta_0}}{\sqrt{1 + 1 - \sin^2 \theta_0}} \exp(\beta z_s) \end{aligned} \quad (\text{H.7})$$

H.5 Relation Between θ_n and θ_0

Referring to Figure H.2,

$$\sin \theta_n = (c_n/c_0) \sin \theta_0 = \exp(-\beta z_s) \sin \theta_0$$

From Equation B.7,

$$\begin{aligned}\exp(\beta z_s) &= (t c_0 + \cos \theta_0)^2 + \sin^2 \theta_0 \\ &= 2 t \beta c_0 \cos \theta_0 + 1 + t^2 \beta^2 c_0\end{aligned}$$

Therefore,

$$\sin \theta_n = (2 t \beta c_0 \cos \theta_0 + 1 + t^2 \beta^2 c_0) \sin^2 \theta_0$$

PUBLICATIONS

R. Orth, F.Wong and J.S. MacDonald,

"The Production of 1:250,000 Maps of Precision Rectified and Registered Landsat Imagery Using the MDA Image Analysis System: Initial Results", Twelfth International Symposium on Remote Sensing of Environment, Manila, Philippines, April 1978.

F.Wong and R.Orth,

"Registration of Seasat/Landsat Composite Images to UTM Coordinates", Sixth Canadian Symposium on Remote Sensing, Halifax, N.S., May 1980.

F. Wong, D.E. Friedmann and R. Orth,

"The Use of Digital Terrain Model in the Rectification of Satellite Borne Imagery", Fifteenth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, May 1981.

F. Wong and R. Orth,

"Orbit Refinement for the Rectification of SAR Imagery", ISPRS, Symposium of Commission III, Helsinki, Finland, June 1982.

F. Wong and R. Orth,

"Rectification of TIROS-N Imagery with Recursive Filtering Technique for Time-Series Estimator", ISPRS, Symposium of Commission III, Helsinki, Finland, June 1982.

F. Wong,

"Resampling of Landsat-D Thematic Mapper Imagery in the Presence of Scan Gaps"

In Preparation