

**UCGE Reports
Number 20182**

Department of Geomatics Engineering

**Spatial Enhancement of Digital Terrain Models Using
Shape from Shading with Single Satellite Imagery**

(URL: <http://www.geomatics.ucalgary.ca/links/GradTheses.html>)

by

Mohammad A. Rajabi

August 2003



ABSTRACT

In most of the geoscientific, environmental, engineering and military related research and applications, among others, Digital Terrain Models (DTMs) play a ubiquitous role. Depending on the region of interest, one may have access to a DTM with a reasonable accuracy and resolution or not. Regardless of what is available, there is always a demand for a denser, hence more accurate DTM.

This research work is an attempt to use Shape from Shading (SFS) with a single (as opposed to stereo) satellite imagery to enhance the interpolation accuracy of DTMs. The motivation is availability of relatively inexpensive yet globally available multiresolution, multispectral single satellite imageries.

Different SFS formulations as well as their corresponding solutions are studied and reviewed. Based on the characteristics of the specific problem of this investigation, the SFS is formulated and the general variational approach is selected as the solution method.

The main deficiency of the standard variational technique is its tendency to over smooth the recovered surface. The over smoothing is due to the smoothing constraint in the formulation of the SFS solution, presence of noise in the image and sudden discontinuities in the surface heights. A more intelligent way of handling the smoothness constraint, which can also distinguish the noise from sudden surface discontinuities, is to use robust statistics.

Using robust statistics, the smoothness constraint in the SFS is reformulated. Moreover, the brightness constraint is applied using the ambiguity cone concept. In addition to the numerical stability, this makes the SFS solution independent of the choice

of the regularization factor. Furthermore, calibration of the albedo factor in the SFS formulation is done through classification of pixels using multispectral imageries. It is shown that the new formulation retrieves the original shape of the object much better than the standard variational method, especially in the presence of noise in the image and sudden discontinuities in the surface heights.

This research is of both theoretical and practical value in the context of its topic, as it not only develops a framework for the SFS formulation and solution, but also provides some valuable practical considerations.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the Ministry of Science, Research, and Technology of Iran for providing me the financial support to further my education to this level. Obviously, without their help, this journey would not have been realized.

My sincere appreciation and thanks go to my distinguished supervisor Dr. J.A. Rod Blais for his continuous guidance, discussions, advice and also support throughout my graduate studies. His patience and constant availability to me play a key role in reaching the goals of this research. He is also acknowledged for his partial financial support to this research. I will never bear to forget his kindness and will be thankful to him forever. As a real scientist, he took me on the journey of unknown lands with himself. The memories of this wonderful trip have been engraved in my heart and will never be forgotten.

I would like also to thank my co-supervisor and the head of department of Geomatics engineering, Dr. Gérard Lachapelle for his thoughtful suggestions and advice throughout the course of my program, especially in the regular supervisory meetings we had together. He is sincerely appreciated for his partial financial support for this research too. As a very successful scientist who has been able to bring theory and practice together, he is surely admired by both faculty members as well as students. It was my great honour to have his regular and close supervision in spite of his tight schedules.

Dr. Naser El-Sheimy, the other member of my supervisory committee, is also greatly acknowledged for providing me his valuable insight and experience during our regular meetings. His inputs and precious critics have been very helpful to me during this

research.

I wish also to express my deep gratitude to Dean Allen Provins for the valuable discussions that we had together during my program. As a computer expert in fields of programming as well as Linux operating system, he has always been helpful to me and for this I sincerely appreciate him. He is also acknowledged for proof reading of the first three chapters of this dissertation.

Dr. Doug Phillips from Information Technologies, University of Calgary, is sincerely thanked for the time and efforts he put on running my C++ codes on the Multimedia Advanced Computational Infrastructure (MACI) Alpha Cluster. If it were not because of his help, I would still have been busy with running my codes on PCs. To me, his interest in his job and the passion that he has toward his work is very remarkable.

SPACE IMAGING as well as Canadian Geomatic Solution Ltd. are greatly thanked for providing this research the IKONOS, IRS-1, and Landsat 7 ETM + imageries of the Crowsnest area, southern Alberta, Canada. My sincere gratitude specially goes to Dr. Steve Adam for making the process of accessing the imageries faster and easier.

The Library of University of Calgary, specifically Ms. Laurie Schretlen from Maps, Academic Data, Geographic Information Centre (MADGIC), is gratefully acknowledged for providing SPOT and Landsat 7 ETM + imageries of Waterton area as well as all of the provincial DTM data used in the numerical analysis of this dissertation.

Last but not least, I would like to emphasize that many other people helped me with this research directly and indirectly. Unfortunately, I cannot thank everyone explicitly here. Nevertheless, hereby I would like to extend my sincere gratitude to them as well

TABLE OF CONTENTS

APPROVAL PAGE	ii
ABSTARCT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF SYMBOLS AND NOTATIONS	xvi
LIST OF ABBREVIATIONS	xx
CHAPTER ONE: INTRODUCTION	1
1.1 Background	1
1.2 Digital Terrain Model	2
1.2.1 Application of DTM	3
1.2.2 DTM Generation	5
1.2.3 DTM Quality Assessment	7
1.2.4 Publicly Available DTMs	9
1.3 Remote Sensing	10
1.3.1 Characteristics of Remote Sensing Systems	12
1.4 Shape from Shading	18
1.5 Objectives	21
1.6 Thesis Outline	24

CHAPTER TWO: SHAPE FROM SHADING AND REMOTE SENSING.....26

2.1 Shape from Shading.....27

2.1.1 Image Formation.....27

2.1.1.1 Microstructure of Surfaces.....28

2.1.1.2 Simplified Equation30

2.1.2 SFS Solution Techniques.....34

2.1.2.1 Characteristic Strip Approach.....35

2.1.2.2 Iterative Solution on a Regular Grid of Points.....39

2.1.2.2.1 Existence and Uniqueness.....47

2.1.2.3 Local Approach.....47

2.1.2.4 Photometric Stereo Approach50

2.1.2.5 Linear Approach51

2.1.3 General Comments54

2.2 Characteristics of Remotely Sensed Imageries.....55

2.2.1 Remote Sensing Systems55

2.2.2 The Source of EM Radiation57

2.2.3 The Effect of Propagation Media.....59

2.2.4 The Object’s Surface.....63

2.2.5 Bidirectional Reflectance Distribution Function (BRDF)64

2.2.6 The Imaging Sensors.....67

CHPATER THREE: SPATIAL ENHANCEMENT OF DTM GRIDS USING

SFS.....70

3.1	Basic Assumptions.....	71
3.2	The Choice of Solution Method.....	72
3.3	Recapitulation of the Standard Variational Method	74
3.4	Deficiencies of the Standard Variational Method.....	76
3.5	Modifications to the standard Variational Algorithm.....	77
3.5.1	Adaptive Regularizer Kernels from Robust Statistics	79
3.5.2	The error Kernel.....	82
3.5.3	The Width Parameter of the error Kernel	89
3.5.4	Geometrical Application of Brightness error Constraint.....	92
3.5.5	Data Fusion and Albedo Factor	94
 CHAPTER FOUR: EXPERIMENTAL RESULTS AND ANALYSIS		97
4.1	Numerical Considerations.....	98
4.2	Analysis of SFS Performance and Choice of Patch Size.....	103
4.2.1	The Simulated Data Set	104
4.2.2	Simulation with a Real DTM Data Set	110
4.3	The Choice of error Kernel Function.....	114
4.3.1	Simulated Data Set.....	115
4.3.2	Real DTM Data Set and Satellite Imagery	119
4.4	Modification to the Albedo Factor.....	125
4.4.1	Simulated Data Set.....	125
4.4.2	Real Data Set.....	127
4.4.3	Further Modification of the Albedo Factor.....	129

4.5	More Experiments with Real Data Sets	132
4.5.1	Landsat 7 ETM+ and the Waterton Lake Area DTM.....	132
4.5.2	SPOT and the Waterton Lake Area DTM from GTOPO30	136
4.5.3	IKONOS, IRS-1, Landsat 7 ETM + and the Crowsnest Area DTM .	139
4.6	Phase Dependency of the SFS Solutions	145
4.7	Roughness of the Terrain.....	147
CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS.....		149
5.1	Conclusions.....	150
5.2	Contributions.....	156
5.3	Recommendations.....	158
REFERENCES.....		160
APPENDIX A: MINIMIZING FUNCTIONS.....		169
APPENDIX B: ENFORCING CONSTRAINTS.....		174
APPENDIX C: PENALTY TERMS BASED ON CONSTRAINTS.....		175
APPENDIX D: MINIMIZING THE SFS COST FUNCTION.....		176
APPENDIX E: DATA FUSION IN SPECTRAL DOMAIN.....		179
APPENDIX F: CLASSIFICATION		181
APPENDIX G: CHARACTERISTICS OF THE REMOTELY SENSED IMAGERIES AND DTM DATA SETS USED FOR NUMERICAL EXPERIMENTS.....		183

LIST OF TABLES

Table 1.1: Existing Global Digital Terrain Model Data	9
Table 1.2: Current, planned, and proposed land observation satellites.....	15
(a) Frequent global coverage, multispectral classification focus.....	15
(b) Radar and radar plus optical.....	15
(c) High resolution, small area coverage (PAN and VNIR Only).....	16
(d) High resolution, multispectral, hyperspectral application tests.....	16
Table 3.1: Examples of re-descending influence functions	84
Table 3.2: Examples of sigmoidal influence functions.....	85
Table 3.3: Quadratic influence function	86
Table 4.1: The convex hemisphere with the 3 x 3 point grid patch.....	106
Table 4.2: The convex hemisphere with the 7 x 7 point grid patch.....	109
Table 4.3: The real DTM data set with the 3 x 3 point grid patch.....	111
Table 4.4: The real DTM data set with the 7 x 7 point grid patch.....	113
Table 4.5: The convex hemisphere	116
Table 4.6: The convex hemisphere with noise and discontinuity in curvature.....	118
Table 4.7: The Waterton Lake area DTM data set and SOPT imagery	122
Table 4.8: The rough part of Waterton Lake area DTM data set and SOPT imagery	124
Table 4.9: Simulated data set with constant albedo factor.....	126
Table 4.10: Simulated data set with variable albedo factor	127
Table 4.11: Real data set with variable albedo factor	129

Table 4.12: Simulated data set with the second modification to albedo factor	131
Table 4.13: Real data set with the second modification to albedo factor	131
Table 4.14: Panchromatic band of Landsat 7 ETM + and the Waterton Lake DTM	134
Table 4.15: ETM bands of Landsat 7 ETM + and Waterton Lake area DTM.....	135
Table 4.16: SPOT and Waterton Lake area GTOPO30 DTM.....	139
Table 4.17: IKONOS and Crowsnest DTM data set.....	142
Table 4.18: IRS-1 and Crowsnest DTM data set	143
Table 4.19: Landsat panchromatic and Crowsnest DTM data set	143
Table 4.20: Landsat ETM and Crowsnest DTM data set.....	144
Table 4.21: The Waterton Lake area DTM with SPOT imagery.....	146
Table 4.22: IKONOS and Crowsnest DTM data set.....	146
TABLE 4.23: Correlation between the roughness of terrain and rate of improvement.	148
Table G.1: Characteristics of Landsat MSS and TM sensors	185
Table G.2: Landsat 7 ETM+ characteristics	188
Table G.3: Some characteristics of HRV sensor in SPOT 1, 2, and 3	190
Table G.4: HRVIR and vegetation sensors of SPOT 4.....	191
Table G.5: Characteristics of SPOT 5 and 5A sensors	192
Table G.6: Some characteristics of the IRS-1A and IRS-1B sensors	193
Table G.7: IRS-C and IRS-D characteristics	193
Table G.8: Sensor characteristics of IKONOS	194

LIST OF FIGURES

Figure 1.1: Full scale model of Sputnik I	11
Figure 1.2: Spatial and temporal characteristics of commonly used remote sensing systems and their sensors	14
Figure 2.1: Reflection in non-specular surfaces	28
Figure 2.2: Reflection in surface layers with inhomogeneities in refractive index	29
Figure 2.3: Reflection in compound surfaces	30
Figure 2.4: A simple model of image geometry. A distant point source illuminant at direction \mathbf{L} , a patch of surface with surface normal \mathbf{N} , and a viewer in direction \mathbf{V}	33
Figure 2.5: The ambiguity cone	34
Figure 2.6: Geometric interpretation of the characteristic strip equations. (a) Iso-brightness map (b) Reflectance map	38
Figure 2.7: Remote sensing system components	56
Figure 2.8: Energy from perfect blackbodies	58
Figure 2.9: Atmospheric absorption for different parts of EM spectrum	60
Figure 2.10: The behavior of Rayleigh and Mie scattering for a few selected wavelengths	61
Figure 2.11: Atmosphere and the EM radiation	62
Figure 2.12: Specular, near-perfect specular, near-perfect diffuse and diffuse reflection	64
Figure 2.13: The concept and parameters of BRDF	66
FIGURE 2.14: Field of view in a radiometer imaging system	68

Figure 3.1: Three re-descending influence functions (row (a)), error kernels (row (b)), and derivative of the error kernels (row (c)) all corresponding to Table 3.1	84
Figure 3.2: Three sigmoidal influence functions (row (a)), error kernels (row (b)), and derivative of the error kernels (row (c)) all corresponding to Table 3.2.....	85
Figure 3.3: Quadratic influence function (row (a)), error kernel (row (b)), and derivative of the error kernel (row (c)) all corresponding to Table 3.3	86
Figure 3.4: The ambiguity cone	92
Figure 3.5: The surface normal before and after mapping on the ambiguity cone.....	93
Figure 4.1: (a) A 3 x 3 grid point patch. Squares are the grid points with known heights and the circles are those with interpolated heights, (b) The corresponding 3 x 3 pixel patch of the imagery.....	104
Figure 4.2: a) The original object; b) The corresponding image; c) The differences between the original object and the IGS; d) The differences between the original object and the SFS solution; and e) same as (d) but with a 7 x 7 patch.....	107
Figure 4.3: (a) A 7 x 7 grid point patch. Squares are the grid points with known heights and the unmarked ones are the points with interpolated heights. The circles are those points with interpolated heights for which enhancement in accuracy is sought through SFS. (b) The corresponding 7 x 7 pixel patch of the imagery.....	109
Figure 4.4: a) The original object; b) The corresponding image; c) The differences between the original object and the IGS, and d) The differences between the original object and the SFS solution. The arrow signs show the North direction.....	112
Figure 4.5: Image of the simulated object with a discontinuity in curvature	118

Figure 4.6: Waterton Lake area DTM data set. The dark spots are the lakes	120
Figure 4.7: SPOT image of the Waterton Lake area.....	121
Figure 4.8: The rough part of the real DTM data set	123
Figure 4.9: Landsat 7 ETM +, Waterton Lake area	132
Figure 4.10: Two views of the Crowsnest area DTM data set.....	140
Figure 4.11: Resampled IKONOS imagery of the Crowsnest area	141
Figure A.1: Different Paths in the one dimensional extremum problem.....	169
Figure G.1: Landsat MSS	186
Figure G.2: The HRV sensor of SPOT	190
Figure G.3: GTOPO30 tiles	197
Figure G.4: GTOPO30 data source around the world	198

LIST OF SYMBOLS AND NOTATIONS

The list of symbols and notations provided here is not exhaustive but it contains those which are used more frequently in the text.

(x,y,z)	The components of 3D Cartesian coordinate system
$z(x,y)$	The height of a surface patch above xy-plane at location (x,y)
$\mathbf{N}(x,y) = (N_1, N_2, N_3)^T$	The surface normal vector at location (x,y)
$p(x,y) = \partial z / \partial x$	The gradient of the surface $z(x,y)$ in x direction at location (x,y)
$q(x,y) = \partial z / \partial y$	The gradient of the surface $z(x,y)$ in y direction at location (x,y)
σ	The slant of the surface (steepest descent measured from zenith)
τ	The tilt of the surface (steepest descent measured from north)
σ_s	The slant of the light source direction
τ_s	The tilt of the light source direction
$R(p,q)$	The scene radiance (intensity or brightness) at location (x,y)
ρ_a	The albedo of the surface for ambient illumination
ρ	The albedo of the surface for point source illumination
ρ_s	The albedo of the surface for specular reflectance
$\mathbf{L} = (l_1, l_2, l_3)^T$	The light source direction
\mathbf{H}	The vector that bisects the angle formed by \mathbf{L} and the line of sight

m	The scattering of the specular reflectance
\mathbf{V}	The viewer direction
$E(x,y)$	The image irradiance (intensity, brightness) at location (x,y)
$\delta x, \delta y, \delta z$	Steps in x, y and z directions
$\delta p, \delta q$	Steps in p and q directions
$r = z_{xx}$	The second partial derivative of the surface z
$s = z_{xy} = z_{yx}$	The second partial derivative of the surface z
$t = z_{yy}$	second partial derivative of the surface z
R_p, R_q	The partial derivatives of the scene radiance with respect to p and q
E_x, E_y	The partial derivatives of the image irradiance (brightness) with respect to x and y
$\delta \xi$	Step in characteristic strip direction (ξ)
$e(x,y)$	The brightness error at location (x,y)
$I(p,q)$	The energy function
Ω	The image domain
\hat{p}, \hat{q}	The estimated surface gradient
z_x, z_y	The surface gradients along x and y directions
Δ	The Laplacian operator
\hat{p}_x, \hat{q}_y	The partial derivatives of the estimated surface gradients along

	x and y directions
i,j	The cell location
n	The iteration number
$\bar{\cdot}$	Mean value
$\bar{\cdot}_{i,j}^n$	Mean value at location (i,j) , iteration n
p_x, p_y, q_x, q_y	The partial derivatives of the surface gradients along x and y directions
$\ \cdot \ $	The norm of a vector
N_x, N_y	The partial derivatives of the unit surface normal along x and y directions
$\beta(x, y)$	Lagrangian multiplier
λ	The regularization parameter (weight)
R_x, R_y	The partial derivative of reflectance map in x and y directions
$E_{xx}, E_{x,y}, E_{yy}$	The second order partial derivatives of image irradiance
r	The radius of curvature
F_z	Fourier Transform of z
(ω_1, ω_2)	The angular velocity in x and y directions, respectively
$\psi_w(v)$	The influence function
v	Residual
w	Width of the influence function

$\Psi_w(\mathbf{v})$	The error kernel
$\alpha_{i,j}^n$	The normalizing factor for cell (i,j) at iteration n
Φ	The shape index
$\mathbf{M}_{i,j}^{n-}$	The smoothed normal vector at step n and location (i,j)
$R(\theta)$	Rotation matrix with rotation angle of θ

LIST OF ABBREVIATIONS

AVHRR	Advanced Very High Resolution Radiometer
CGIS	Canada Geographic Information System
CGVD28	Canadian Geodetic Vertical Datum 1928
CIESIN	The Consortium for International Earth Science Information Network (CIESIN)
CZCS	Coastal Zone Color Scanner
DED	Digital Elevation Data
DEM	Digital Elevation Model
DIME	Dual Independent Map Encoding
DTD	Digital Terrain Data
DTED	Digital Terrain Elevation Data
DTM	Digital Terrain Model
FGDC	Federal Geographic Data Committee
GIS	Geographic Information System
HCMM	Heat Capacity Mapping Mission
IFSAR	InterFerometric Synthetic Aperture Radar
ITHD	Interferometric Terrain Height Data
JPL	Jet Propulsion Laboratory
LIDAR	LIght Detection And Ranging (LIDAR)
MSL	Mean Sea Level

NAD27	North American Datum 1927
NAD83	North American Datum 1983
NAVD88	North American Vertical Datum 1988
NCGIA	National Center for Geographic Information and Analysis
NGVD29	National Geodetic Vertical Datum 1929
NSSDA	National Standard for Spatial Data Accuracy
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SFS	Shape from Shading
SFT	Shape from Texture
SIR	Shuttle Imaging Radar
SPOT	Système Probatoire pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
TIN	Triangulated Irregular Network
TM	Transverse Mercator
UTM	Universal Transverse Mercator

CHAPTER ONE

INTRODUCTION

1.1 Background

According to Foresman [1998], the first Geographic Information System (GIS) in the world was the Canada Geographic Information System (CGIS). It was developed in the mid-1960s and designed mainly as a computerized map measuring system. In the late 1960s, the US Bureau of Census started the Dual Independent Map Encoding (DIME) program to create digital records of all US streets and to support automatic referencing and aggregation of census records. By the late 1970s, the Computer Graphics and Spatial Analysis laboratory of Harvard University developed the first general-purpose GIS software called ODYSSEY GIS. However, the modern history of GIS dates from the early 1980s, when the price of sufficiently powerful computers fell below a critical threshold.

GIS is a technology which can be used for scientific investigations, resource management, and development planning. Broadly speaking, a GIS consists of four different components: 1) data collection, 2) data storage, 3) data analysis and processing, and last but not least 4) information retrieval. Among many types of data, Digital Terrain Models (DTMs) as well as remote sensing play a key part in the development of GIS. This chapter provides a general review of DTM and remote sensing. It also discusses the Shape from Shading (SFS) which is used as a main tool in this research. It defines the objectives of this research work and outlines the thesis.

1.2 Digital Terrain Model

A DTM has multiple definitions. In some countries, it represents the bare terrain with uniformly spaced height values. However, in some other countries, DTMs may also incorporate the elevation of significant topographic features on the land, plus mass points and breaklines that are irregularly spaced so as to better characterize the true shape of the terrain. Regardless of the definition, the aim of a DTM is to define the terrain's features more clearly, locate them more precisely and ultimately approximate the shape of the terrain more closely. Other commonly used terms for this kind of data are: Digital Terrain Data (DTD), Digital Elevation Model (DEM), Digital Terrain Elevation Data (DTED), and Digital Elevation Data (DED) [Aronof, 1989; Petrie and Kennie, 1991; El-Sheimy, 1998; Burrough and McDonnell, 1998; Maune, 2001; Longley et al, 2001].

DTMs are important types of input for a GIS. They play an essential role in many different scientific and engineering as well as military applications and are used as an important tool for modeling and analysis of spatial topographic information. Whether the problem is to find the best location to build a new road, tower, housing development, or identify the best place to find a particular animal or plant species, the analysis of terrain data is almost always a key ingredient. Most information derived from analysis of terrain data is created as input to site selection and analysis models in GIS.

DTMs are increasingly becoming the focus of attention within the larger realm of digital topographic data. This is due to the fundamental nature of the data, and the insight that the elevation information can provide. The insight that DTM data can add is becoming extremely valuable in numerous applications within the fields of Earth,

environmental science and engineering. This is due, at least in theory, to the fact that DTMs can be used to simulate the true elevations, slopes and aspects of the Earth's surface.

1.2.1 Applications of DTM

The earliest use of DTMs dates back to the 1950s when the U.S. Air Force first experimented with aircraft simulator technology. Since that time, DTMs have proved to be an important ingredient for all types of geographic modeling and the analysis of spatial topographic information.

Broadly, there are four main application fields where DTMs are used: 1) civil engineering, 2) Earth science applications, 3) planning and resource management, 4) military applications. Civil engineers are mainly interested in using DTMs for the cut-and-fill computations involved with road design, site planning, and volumetric calculations for building dams, reservoirs, the area that would be flooded by a hydroelectric dam and the like.

DTMs can be used to analyse and delineate areas that can be seen from a location on terrain. Such intervisibility analyses are necessary for planning route locations of radar antennas or microwave towers, and defining viewsheds. It may be appropriate to point out that owing to such overt concerns with volume and design, calling a DTM a "terrain model" has more relevance to a civil engineer than other DTM users.

The Earth science applications center mainly around specific functions for modeling, analysis and interpretation of the unique terrain morphology. These may include

drainage basin network development and delineation, hydrological run-off modeling, geomorphological simulation and classification, and geological mapping. Generating slope and aspect maps, and slope profiles for creating shaded relief maps is an area where DTMs are used routinely. Research and application of global change science is also another area where the need for quality topographic data is essential to creating Earth based simulated models of real world situations.

The application of terrain data in planning and resource management is composed of diverse fields including remote sensing, agriculture, soil science, meteorology, climatology, environmental and urban planning, as well as forestry, whose central focus is the management of natural resources. Examples include site location, DTM production from remote sensing, geometric and radiometric correction of remote sensing images, soil erosion models, crop suitability studies, wind flow, and pollution dispersion models. Simulation and virtual reality applications are also another area of study that requires terrain elevation data

The military is not only a leading consumer but also a significant producer of DTMs. Most military operations depend on a reliable and accurate understanding of the natural and manmade terrain. This includes a detailed modeling of elevation, slope, and aspect of the land surface. The military's use of DTMs employs a combination of the methods used by all the previous applications. Examples would include simple visualization, intervisibility analysis of the battlefield, three-dimensional displays for weapons guidance systems and flight simulation, and radar line-of-sight analysis.

Some other examples where DTMs are needed are as follows [Blais, 1997; Maune,

2001]:

- Representation and storage of elevation data for digital topographical maps in national databases as well as statistical analysis and comparison of different types of terrain.
- Visualization of terrain and topography using hillshade techniques and also as a background for displaying thematic information or for combining relief data with other data layers as soils, land use or vegetation.
- Computation of terrain corrections for geophysical, gravity and other field observations.
- Investigation on erosion processes, land slides, contaminations by dangerous chemicals, and other environmental issues.
- Computation of different hydrological parameters such as flow direction across the terrain, creation of stream networks and watersheds.

1.2.2 DTM Generation

DTMs have traditionally been generated by ground surveying, from existing contour maps, by photogrammetric analysis of stereo aerial photographs, or more recently by automated analysis of stereo satellite data. Elevation data derived from field survey operations provide the most accurate method of obtaining data. This is done by physically taking measurements along a grid of positions on the ground. The values obtained during the field survey are either input directly into computer systems through data recorders coupled to field instruments or subsequently stored in a digital form. Very

high levels of precision and resolution are attained using this method of digital elevation data collection. However, because of the amount of time required to obtain this level of detail, this is the least cost-effective and the most time-consuming method of producing digital elevation data. Traditionally, because of the time required, it is only feasible to use such ground surveys for small engineering projects. This information is usually not available publicly but must be obtained directly from survey and civil engineering firms which have worked in the region of interest.

DTMs derived from existing contour maps are the most common of all sources, as analog maps have been around for a long time in comparison to the other data sources. Since the mid-1980s scanning technology along with raster-to-vector conversion algorithms have been used extensively in the digital data capture arena. Considering the availability of different scale and quality of paper maps, this method of generating DTMs turns out to be relatively inexpensive. Nevertheless, depending on the quality of the map, one can expect a low to medium elevation resolution. Taking into account all the error sources involved in the process of making maps as well as the process of extracting DTM from them, it is not surprising that the level of accuracy of DTMs coming out of this process is relatively low.

Photogrammetric methods have been extensively used to generate accurate medium to high-resolution DTMs. These methods are based on the stereoscopic interpretation of aerial photos or recently stereo satellite imageries, using suitable photogrammetric equipment. As examples, Kok [1986] explained the automated information extraction from stereo aerial photos, while Gagan and Dowman [1988], Simard et al [1988], Tam

[1990], and Cosandier [1999] discussed the automated generation of elevation data from SPOT (Système Probatoire pour l'Observation de la Terre) stereo satellite imagery. The ability to exploit detailed and accurate elevation information from stereo photography is not a new process. Sound mathematical procedures have been in existence for many years. It is a relatively recent development though, that allows this type of processing to be performed on inexpensive workstations. Moreover, it is worth mentioning that there are few satellites which are designed for stereo image acquisition. Consequently, it is not always easy to have access to stereo satellite images without any special planning and preparation.

There are a number of newer sources that are also being developed for elevation data extraction. These include radar, microwave altimetry, Synthetic Aperture Radar (SAR) interferometry, and LIght Detection And Ranging (LIDAR). Although these forms of imagery may not come under the classical photogrammetric discipline, the mathematical approach to elevation extraction is derived directly from classical photogrammetry.

1.2.3 DTM Quality Assessment

The quality of DTMs is assessed by a number of parameters which are usually described in their associated description file or metadata. The structure of this file can meet either the specific needs of a company or a certain user group or the Federal Geographic Data Committee (FGDC) standards [FGDC, 1998]. Regardless of format, the more complete this file is, the more useful the associated DTM will be.

Horizontal and vertical datums used for DTMs are two key parameters to characterize

them. As an example, the North American Datums of 1927 (NAD27) and of 1983 (NAD83) are two official horizontal datums in U.S. and Canada. Furthermore, the heights in DTMs are usually orthometric heights referred to the Mean Sea Level (MSL) or geoid. The National Geodetic Vertical Datum of 1929 (NGVD29) used to be the official vertical datum in U.S., but at present, the North American Vertical Datum of 1988 (NAVD88) is officially in use. The Canadian Geodetic Vertical Datum of 1928 (CGVD28) is the official vertical datum in Canada which is practically the same as NGVD29. Due to some complications [Blais et al, 1997], the datum transformation process from CGVD28 to NAVD88 has not been done in Canada yet and therefore, CGVD28 is still in use as the official vertical datum.

The map projection and planar coordinate system are other important specifications for DTMs. The coordinate system used to represent the DTMs is either geodetic coordinates (latitude/longitude) or rectangular coordinates such as Universal Transverse Mercator (UTM), 3-degree Transverse Mercator (3TM), State Plane coordinates, or others. In metadata, complete specifications of the map projection system are supposed to be mentioned. As an example, when UTM is used as the projection system, the zone number, false easting, false northing, the longitude of central meridian, and the scale factor along the central meridian are among the important data which should be specified in the metadata [FGDC, 1998].

Horizontal and vertical accuracies are other parameters to assess the DTMs. The standard procedures for accuracy assessment are described within National Standard for Spatial Data Accuracy (NSSDA) [FGDC, 1998]. It specifies a statistical and testing

methodology for estimating the positional accuracy of points on maps and in digital geospatial data, including Triangulated Irregular Networks (TINs) and DTMs, with respect to georeferenced ground positions of higher accuracy. Most of the time the accuracy is provided as a confidence level percentage which itself is a factor of Root Mean Square Error (RMSE). For example, the 95 percent confidence level error in horizontal and vertical components are $1.7308 \times \text{RMSE}$ and $1.9600 \times \text{RMSE}$ respectively [Maune, 2001].

Spatial resolution or the density of rectangular grid in x and y directions is the other parameter to assess the DTMs. It simply shows the degree of smoothing that the terrain features are subject to. In other words, features smaller than the resolution of the DTM will be reduced to a single data point or be lost altogether.

1.2.4 Publicly Available DTMs

Table 1.1 provides a list of global DTMs with different resolutions (from 10 arc minutes to 3 arc seconds) which are freely available on the internet. The National Center for Geographic Information & Analysis (NCGIA) [2002a and 2002b] provides links to the DTMs as well as the freeware and shareware programs available on internet to manipulate, reformat, and view different types of DTM files.

Table1.1: Existing Global Digital Terrain Model Data [NCGIA, 2002a]

Elevation Data Set	Format	Spacing	Approximate Linear Distance	Area of Coverage
US Navy 10'	Grid	10 arc min.	18.4 km	Global
ETOPO5	Grid	5 arc min.	9.2 km	Global
GTOPO30	Grid	30 arc sec.	921 m	Global
DTED Level 1	Grid	3 arc sec.	100 m	Global w/Gaps

After the successful Shuttle Radar Topography Mission (SRTM) in February 2000, a global DTM with high spatial resolution is planned to become available for the first time. During the 11-day Shuttle Radar Topography Mission, two SAR-IN sensor were used: an American C-Band System run by NASA, and a X-Band system run by the German Aerospace Center (DLR). The second antenna of the systems was mounted on a mast of 60 metre length. The whole construction had to be balanced by steering the Shuttle "Endeavour" into the optimum viewing geometry.

The objective of the mission was to use Interferometric Synthetic Aperture Radars (IFSARs) to acquire topographic data over 80% of Earth's land mass (between 60° N and 56° S). The digital topographic map products are supposed to meet Interferometric Terrain Height Data (ITHD)-2 specifications (30 m x 30 m spatial sampling with less than 16 m absolute vertical (linear) accuracy, 10 m relative vertical (linear) accuracy and 20 m absolute horizontal circular accuracy). All accuracies are quoted at the 90% level, consistent with U.S. National Map Accuracy Standards [SRTM, 2003]. The US Geological Survey's EROS Data Center has started distribution of the unedited data covering U.S. to the public. By the end of 2003, the 90 m topographic data for outside of U.S. will be ready for public distribution.

1.3 Remote Sensing

Remote sensing in geomatics refers to the acquisition of information about the Earth's (or other planets') surface including oceans and atmosphere without coming in physical contact with them. Among many different sources of data, remote sensing image data

have a special role in GIS [Ehlers et al, 1989]. It began in the 1840s as balloonists took pictures of the Earth's surface using the newly invented photo-camera. Perhaps the most novel platform at the end of the 19th century was the famed pigeon fleet that operated in Europe. However, the space program is held by many historians to truly have begun with the launch of Sputnik I by the Soviets on October 4, 1957. Figure 1.1 shows a full-scale model of the first Sputnik (about the size of a basketball, weighing 83 kg [182 lb]), with radio and one scientific instrument, on display at the National Air and Space Museum in Washington, D.C.

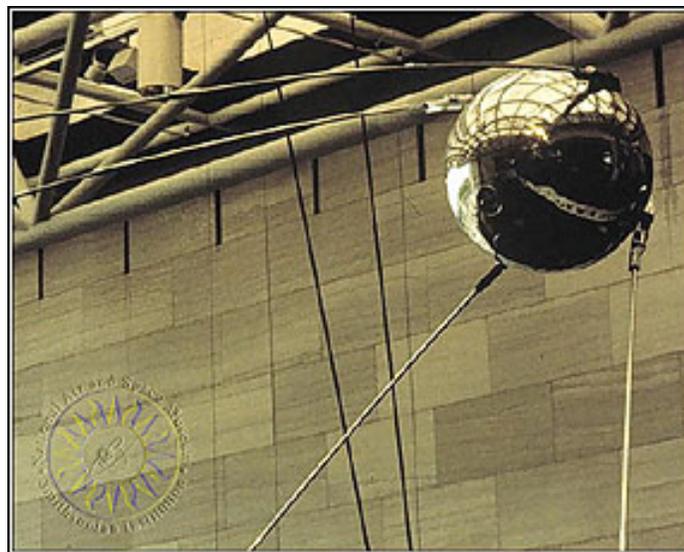


Figure 1.1: Full scale model of Sputnik I [Goddard Space Flight Center, 2002]

The first military satellites of the 1950s were developed and deployed in great secrecy to gather intelligence. However, in the 1970s the remote sensing industry was revolutionized specially by Landsat, the first satellite dedicated to mapping natural and

cultural resources on land and ocean surfaces. The commercial use of remote sensing started in 1980s when the Landsat was privatized. Since then, there has been an increase in different geoscientific application of remote sensing especially in the field of Earth resources management.

1.3.1 Characteristics of Remote Sensing Systems

The key characteristic of any remote sensing system is the resolution. There are three types of resolution in remote sensing: 1) spatial, 2) spectral, and 3) temporal. Spatial resolution refers to the size of objects that can be resolved and is usually measured in terms of picture elements or pixels. The spatial resolution of satellite-based systems varies from 1 m for the new high-resolution sensors such IKONOS, to 1 km for weather satellite satellites such as AVHRR. The resolution of aerial photographs is something in the range of 0.1 m to 5 m. The coverage of remote sensing images is something between 10x10 to 200x200 km.

On the other hand, the part of electromagnetic spectrum that the sensor is working in is called spectral resolution. Obviously, different objects respond differently when they are excited in different parts of electromagnetic spectrum. Depending on the applications, sensors with different spectral resolutions are selected for remote sensing task.

Temporal resolution refers to the repeat cycle and describes the frequency by which the images are acquired of the same area. From the temporal resolution point of view, there are basically two types of satellites: 1) Earth orbiting, and 2) geostationary. The

Earth orbiting satellites are sun synchronous and have a polar (or close to polar) orbit with a fixed altitude. Geostationary satellites are usually used for communication purposes rather than remote sensing. Figure 1.2 shows the spatial and temporal characteristics of commonly used remote sensing systems and their sensors [Longley et al, 2001; Jensen and Cowen, 1999]. Chapter 4 gives more detailed information about the characteristics of the satellites, imageries of which have been used in this research work.

The main disadvantage of satellite imageries (in the visible band) is cloud coverage and lack of standardized support data/sensor models. The cloud coverage is a fact of life and low cloud cover requirements must be judiciously balanced against acquisition time frames that are affected by localised weather and seasonal patterns.

Tables 1.2(a)-(d) provide lists of 46 satellites from eleven countries with four major sensor types that either have been placed in operation since 1985 or will be launched by 2009 [Ames Research Center, 2002]. This impressive list convinces anyone that remote sensing has become a major technological and scientific tool for monitoring planetary surfaces and atmospheres. In fact, the budgetary expenditures on observing the Earth and other planets, since the space program began, now exceed \$150 billion. Much of this money has been directed towards practical applications, largely focused on environmental and natural resource management. The Consortium for International Earth Science Information Network (CIESIN) [2002] provides several examples of scientific applications of remote sensing in environmental studies.

Table 1.2: Current, planned, and proposed land observation satellites**(a) Frequent global coverage, multispectral classification focus**

Country	Owner	Program	Instrument(s)	Launch	Sensor
U.S.	Gov.	Landsat 5	TM	'85	M
India	Gov.	IRS-1B	LISS-2, (LISS-1)	'91	M
France	Gov.	SPOT 3	HRV	'93	M,P
India	Gov.	IRS-P2	LISS-2	'94	M
India	Gov.	IRS-1C	LISS-3, PAN, (WIFS)	'95	M,P
Germany	Gov.	PRIRODA	MOMS-02	'96	M,P
Japan	Gov.	ADEOS	AVNIR	'96	M,P
India	Gov.	IRS-1D	LISS-3, PAN, (WIFS)	'97	M,P
China-Brazil	Gov.	CBERS	CCD, IRMSS	'97	M,P
France	Gov.	SPOT 4	HRVIR, VEGITATION	'98	M,P
India	Gov.	IRS-P5	LISS-4, LISS-3	'98	M
U.S.	Gov.	Landsat 7	ETM+	'99	M,P
U.S.	Com.	Resource 21	XXX	'99	M
India	Gov.	IRS-2A	LISS-4, LISS-3, (WIFS)	'00	M
France	Gov.	SPOT 5A	HRG, VEGITATION	'02	M,P
India	Gov.	IRS-2B	LISS-4, LISS-3, (WIFS)	'04	M
France	Gov.	Spot 5B	HRG, VEGITATION	'04	M,P
U.S.	Gov.	EOS AM-2	LATI (MODIS)	'04	M,P

(b) Radar and radar plus optical

Country	Owner	Program	Instrument(s)	Launch	Sensor
Russia	Gov.	Almaz 1	SAR	'91	R
ESA	Gov.	ERS-1	SAR	'91	R
Japan	Gov.	JERS-1	OPS, SAR	'92	M,R
U.S.-Ger.-It.	Gov.	Shuttle Radar	SIR-C/X-SAR	'94	R
ESA	Gov.	ERS-2	SAR	'94	R
Canada	Gov.	Radarsat	SAR	'95	R
ESA	Gov.	Poem	ASAR	'98	R
Russia	Gov.	Almaz 1B	1 SLR, 3 SARs, 4 SCANNERS*	'98	M,P,R
Russia	Gov.	Almaz 1B	1 SLR, 3 SARs, 4 SCANNERS*	'01	M,P,R
Japan	Gov.	ALOS	AVNIR-2, VSAR	'02	M,P,R
Russia	Gov.	Almaz 2	1 SLR, 3 SARs, 4 SCANNERS*	'04	M,P,R
Japan	Gov.	ALOS-A1	AVNIR-3, VSAR	'04	M,P,R
India	Gov.	IRS-3	MULTI FREQ POL SAR	'04	R

Japan	Gov.	ALOS-B1	A-SAR	'06	R
Japan	Gov.	ALOS-A2	AVNIR-4, VSAR	'07	M,P,R
Japan	Gov.	ALOS-B2	A-SAR	'09	R

(c) High resolution, small area coverage (PAN and VNIR Only)

Country	Owner	Program	Instrument(s)	Launch	Sensor
Russia	Gov.	SPIN-2	KVR-1000, TK-350	'96	P(f)
U.S.	Com.	Earthwatch	EarlyBird	'96	M,P
U.S.	Com.	Space-Imaging	SIS	'97	M,P
U.S.	Com.	EarthWatch	QuickBird	'98	M,P
U.S.	Com.	Orbimage	Orbview	'98	M,P
U.S.	Com.	GDE	XXX	'99	P
India	Gov.	IRS-P6	PAN	'99	P

(d) High resolution, multispectral, hyperspectral application tests

Country	Owner	Program	Instrument(s)	Launch	Sensor
U.S.	Gov.	CTA Clark	W-VIEW	'97	M,P
U.S.	Gov.	TRWLewis	HIS	'97	H,P
U.S.-Japan	Gov.	EOS AM-1	ASTER, (MODIS)	'98	M
U.S.	Gov.	EO-1	XXX	'99	M,H
U.S.	Gov.	Warfighter 1	XXX	'99	M or H

M: Multispectral, H: Hyperspectral, P: Panchromatic, R: Radar, (f): Film

**: SLR-3, SAR-3, SAR-10, SAR-70, OES, MSU-E (MSU-SK, SROSM)*

XXX: Wide Field of view, Lower resolution sensors

It is readily apparent from the data characteristics of the satellites that there are wide variations in the image features available from the different systems. Generally speaking, the major features of satellite imagery, which make remote sensing so special, are:

- The images are readily available in digital format. This means that computers can process them for either machine assisted information extraction or enhancement before an image product is formed.
- An image frame covers a wide area (scene) in comparison to airphotos. This reduces

both interpretation time and time needed to assemble the separate interpretations from each photo. Moreover, since a satellite scene is imaged at essentially the same time, it provides a consistency which is difficult or impossible to achieve for large areas with aerial photography.

- The satellite imagery comes in different spatial resolution. In practice the systems intended for the identification of land cover and land use and thus broad area coverage have focused on moderate resolutions between 5 and 30 metres and swaths of 100 to 200 km. On the other hand, the high resolution satellites are designed with 1 to 3 metre resolution and 4 to 40 km swaths [SpaceImage, 2002; Richards, 1993].
- The satellite sensors differ in spectral coverage. The major use of the different spectral bands is automated classification of land cover types. The lower infrared band has been shown to be nearly essential in such analysis and thus is featured in the global coverage type systems. The upper infrared has proven of great value in the discrimination of surface mineralogy. The thermal infrared band measures the radiation from heat sources and has not been generally used in land classification projects, partially because of its lower, 60-120 m resolution. The high-resolution systems carry Visible Near InfraRed (VNIR) bands only because of the extra costs of adding bands above that region. Since their major use will be in mapping and the identification of manmade objects, they are not as concerned with having the capability of computer classification of the natural environment.
- Temporal coverage is another characteristic of satellite imageries. There are two types of temporal coverage. Global repeat coverage is the frequency with which all

locations can be visited and is a direct function of the swath imaged by the sensor. Site revisit period is the time required to revisit any given site and is a function of the side pointing capability of the satellite and or instrument. All of the satellites except Landsat have cross tracks pointing and thus revisit times of 2 to 3 days at the equator. Because all the satellite orbital paths cross near the poles, both global repeat and site revisit times are halved at 60 degrees latitude and get rapidly better above that latitude. The temporal coverage of satellite imageries helps the scientist to study the fourth dimension (time) of spatial features on the Earth. Due to availability of multiple satellites, one can have access to nearly every other day coverage now. This is beneficial to many potential users who require data to be taken over very large areas as frequently as possible.

Having access to satellite imagery by itself is not useful unless one can extract the information of interest. The interpretation and analysis of remotely sensed data to extract meaningful information is one of the main applications of satellite remote sensing. Information extraction comprises all methods that aim to make the information that is contained in remote sensing data usable. In most cases, the objective is to derive useful spatial information that can be used in conjunction with other geographical data or in thematical maps. As in this research work shape from shading (SFS) technique is used for information extraction purpose, the next section reviews this method very briefly.

1.4 Shape from Shading

A single two-dimensional image is an ambiguous representation of the three-dimensional

world as many different scenes could have produced the same image. Yet, the human visual system is extremely capable in recovering the correct depth model from this type of representation. However, inverse rendering or the procedure for recovering three-dimensional surfaces of unknown objects from two-dimensional images is still an important task in computer vision research. A robust procedure, which can correctly reconstruct surfaces of an object, is important in various applications such as visual inspection, autonomous land vehicle navigation, and surveillance and robot control.

In the past three decades, there have been extensive studies on inverse rendering. Shape from defocusing [Pentland, 1987; Hwang et al, 1989], shape from stereopsis (3D vision) [Medioni and Nevatia, 1985], shape from motion [Waxman and Gurumoothy, 1988], shape from texture (SFT) [Witkin, 1981; Kender, 1979], and shape from shading (SFS) [Horn, 1990; Wei and Hirzinger, 1997; Zhang et al, 1994] are examples of techniques used for inverse rendering.

The human visual system has the ability to infer the shape of a three-dimensional object using only two-dimensional information. One of the cues the human visual system uses to perform this inference is shading information. An object's shading is its variation in brightness over a given domain. Understanding how the shape of a three-dimensional object may be recovered from shading in a two-dimensional image of the object, i.e., SFS, is one of the most important - and still unresolved - problems in machine vision.

Shading has sometimes been described as a “weak” cue in three-dimensional vision, particularly when compared to the other cues like stereo as it is the most reliable one which overrides all others. However, stereo vision is not always available. In that case,

shading will be an important cue to shape. This is significant, for example, when one is viewing a smooth surface without any surface markings. If there were no shading, one could only guess at the shape. Moreover, when one looks at a still photograph or in situations when the observer is too far away from objects for stereo to provide useful cues, shading can supply important cue to shape recovery.

The recovery of shape from shading is by no means trivial. One cannot simply associate a given image brightness with a particular surface orientation. The problem is that there are two degrees of freedom to surface orientation. In other words, it takes two quantities to specify the direction of a unit vector perpendicular to the surface. But there is only one brightness measurement/equation at each pixel (picture cell) with two unknowns at every point in the image.

The earliest work on the quantitative use of shading information appears to have been in the mid-1960s in recovering the shape of parts of the lunar surface in preparation for the human exploration of the moon. However, the basic quantitative SFS algorithms in computer vision have been developed primarily by Horn and his colleagues [Horn, 1977]. In these algorithms, a number of simplifying assumptions are made and addition of constraints, such as continuity and smoothness of the surface, have been implemented. Pentland [1984] showed that with no constraint or prior knowledge about the viewed scene, local shading information is insufficient in determining the precise surface or whether the surface is convex or concave.

Most of SFS algorithms rely on idealistic conditions such as an orthographic projection, a distant point light source, and Lambertian reflectance. It is known that most

real surfaces are neither perfectly diffuse (Lambertian) nor ideally specular (mirror-like); however, most SFS algorithms assume Lambertian reflectance. It is necessary to develop new techniques to recover the shapes of objects whose surfaces are not necessarily Lambertian.

During the past three decades SFS has received a lot of attention in the literature of computer vision research and there have been many significant developments in this area. However, it is still a challenging problem to solve. Fine tuning the basic assumptions and parameters involved in the solution of the problem for the specific application in hand are still areas which one needs to take carefully into consideration.

1.5 Objectives

As mentioned in the previous sections, DTM is an important tool for modeling and analysis of spatial topographic information. It is used in all branches of Earth and environmental science, engineering, and military applications as well. However, in spite of all the available global and local DTMs, there are always applications for which the resolution of available DTMs is not sufficient.

Obviously, the first solution to densify DTM is by using direct techniques which can be either ground surveying, photogrammetry or remote sensing stereo measurements, as well as any other new measurement techniques. Although these methods provide the most accurate results, they are both time consuming and expensive. Moreover, in the case of photogrammetry or remote sensing stereo measurements, there are some parts of the world that due to many different reasons there is no access to complete or up to date

airphoto or stereo satellite imagery coverage. Also, in absence of sufficient, non-repetitive image texture, stereo measurement techniques fail to produce correct and reliable results.

The second solution to the DTM densification problem is to use simple interpolation methods which turn out to be very fast, however, reviewing the numerical methods literature, one realises that there are many interpolation techniques available, each with advantages and limitations. Deciding on the type of interpolation model and fine tuning its related parameters have always been challenging issues. Furthermore, the estimation accuracy is relatively low in comparison to direct measurement methods. This is a more serious problem when one tries to model a rough surface with interpolation methods.

On the other hand, single satellite imagery with virtually complete, most up to date global multiresolution and multispectral coverage is relatively (in comparison to the other data sources with the same amount of information) inexpensive and publicly available. Even in some cases, satellite imagery might be the only source of data readily available. Considering the availability of high-resolution satellite imageries nowadays, there is another approach to densify the DTM which is the objective of this thesis. This solution relies on using the SFS technique with a single satellite imagery with a better resolution (at least one dyadic order of magnitude) than the DTM. The main goal of this research is to use SFS as a means to refine, enhance, and improve the accuracy of DTM interpolation using a single satellite imagery of the corresponding area. In order to implement these ideas, the specific objectives of this thesis are:

- Reviewing different SFS solution methods and analysing advantages and

- disadvantages of each method;
- Analysing the weaknesses of the general minimization solution to the SFS problem and applying some modifications to the solution to overcome the difficulties;
 - Fine tuning the general SFS solution for applying it to satellite imageries with different characteristics;
 - Analysing the correlation between the roughness of terrain and the rate of improvement in standard deviation of interpolation solution using SFS techniques;
 - Developing an algorithm which can be easily adapted for parallel processing methods to facilitate faster solution to this specific SFS problem.

The original contributions of this research work encompass the derivations of algorithms to increase the DTM interpolation accuracy using SFS techniques. These algorithms were implemented through the development of software packages in MATLAB and Visual C++ environments. The Visual C++ codes are also modified so that they can be run on the parallel processing facilities of the University of Calgary.

All experimental analyses carried out in this thesis were based on the available satellite and DTM data sets. The first DTM data have the 25 m grid spacing in UTM map projection system from Waterton Lake area (South Alberta, Canada) which belong to Alberta's provincial DTM data set. The second DTM data are the publicly available GTOPO30 data set with 30 arc seconds (approximately one km) spacing available from the U.S. Geological Survey website. The third set is the 25 m spacing in UTM map projection system from Crowsnest/Blairmore area (south Alberta) which belongs to Alberta's provincial DTM data set too. The Library of the University of Calgary

provided the provincial DTM data sets to this research.

SPOT, Landsat 7 ETM+, IRS, and IKONOS are the four types of satellite imagery used in this work. The library of the University of Calgary provides the SPOT and Landsat 7 ETM + imageries of the Waterton Lake area while Space Imaging and the Canadian Geomatic Solution Ltd (CGSL) provide the Landsat 7 ETM +, IKONOS and IRS imageries of the Crownsnest/Blairmore area.

1.6 Thesis Outline

Chapter 2 starts with definition of the shape from shading problem. It discusses the general image formation process and provides a detailed explanation of the parameters involved in shape from shading context. In order to establish a solid basis for this thesis work, detailed explanations of different solutions to the SFS problem are given and the advantages and disadvantages of the methods are discussed. This chapter concludes with the specific characteristics of satellite imageries as in this research work they are inputs to the SFS process.

Chapter 3 covers the derivation of the method applied to the SFS problem in accordance with the objectives of this thesis by pointing out the deficiencies of the existing solutions to the SFS problem and explaining the modified solution. It shows how one can take advantage of the local geometry in addition to statistics to modify the SFS solution. Moreover, the method applied to fine tuning the albedo factor in this special application of SFS solution is explained.

As the heart of this thesis work, Chapter 4 concentrates on the numerical testing of

the derived method. A detailed explanation of both the synthetic and real data sets is provided here. This chapter explains the general scheme of the implementation procedures involved in solution as well as the data processing steps. An in depth discussion on the obtained results in support of the developed theory in Chapter 3 are given in this section.

Last but not least, in Chapter 5, conclusions based on investigations carried out in the previous chapters are listed. Recommendations specifically related to the methods derived and implemented in this thesis work are also given.

CHAPTER TWO

SHAPE FROM SHADING AND REMOTE SENSING

The image of the world projected onto the retina is essentially two dimensional. However, from this image one can recover information about the shapes of objects in a three-dimensional world. Besides stereo and the other cues available for recovering the missing dimension, one can use the variation of brightness, or shading. The fact is that human's visual system tries to interpret the brightness patterns on the retina, as shading is mainly due to spatial fluctuations of surface orientation and spatial variations in the reflecting properties of the surface.

The direct problem which is the mapping from surface orientation to image brightness is unique and can be determined for different surface materials and illuminating conditions. However, as mentioned before, the inverse problem, i.e., the recovery of shape from shading (SFS), is still a challenging subject. The first part of this chapter reviews the SFS problem closely.

In this research work, satellite imageries are used as inputs for the SFS problem. In order to be able to analyse and judge the test results obtained for the developed mathematical models, one should be aware of the quality of the data used as input. Therefore, the second part of this chapter deals with the general specifications and characteristics of the remote sensing systems. Appendix G has more detail information about the real data sets used in these investigations.

2.1 Shape from Shading

This section covers the image formation process and the basic assumptions that are made to simplify the SFS analysis. Based on the simplifying assumptions, the reflectance map is introduced and discussed. Finally, in the last part of this chapter different SFS solutions are studied.

2.1.1 Image Formation

In order to formulate the SFS problem in an appropriate way, one should have a thorough understanding of the image formation process. Although it has been known for a long time that shading provides an important depth cue, only relatively recently has the SFS problem been properly formulated.

Monochrome images of smoothly curved surfaces with homogeneous reflecting properties commonly exhibit a variation in the measured image brightness (also called image irradiance) or shading. This shading provides important information about the object's shape and has been exploited in machine vision [Van Diggelen, 1951; Rindfleisch, 1966; Horn, 1975; Horn, 1977; Horn et al, 1978; Woodham, 1977, 1978a, 1978b]. The change in image brightness or shading is due to the interaction of four principal factors: 1) the illumination (radiance or intensity and the distance and extension of the source), 2) the shape of the surface, 3) the reflecting characteristics of the material, and 4) the image projection.

The reflecting characteristic of the material is the most critical factor in the image formation process and has great importance in understanding vision. This characteristic depends on the microstructure of the surface material as it determines how much of the

incoming light will be re-emitted in various directions. Because of its importance in image formation and therefore SFS formulation, the following section reviews the reflecting characteristic of the material more closely.

2.1.1.1 Microstructure of Surfaces

When a ray of light strikes the surface of an object it may be absorbed, transmitted or reflected. If the surface is flat and the underlying material homogenous, the reflected ray will lie in the plane formed by the incident ray and the surface normal and will make an angle with the local normal equal to the angle between the incident ray and the local normal. This is referred to as specular, metallic or dielectric reflection. Objects with surfaces of this kind form virtual images of surrounding objects.

However, many surfaces are not perfectly flat on a microscopic scale and thus scatter parallel incident rays into a variety of directions (Figure 2.1). If deviations of the local surface normals from the average are small, most of the rays will lie near the direction for ideal specular reflection and contribute to a surface shine or gloss.

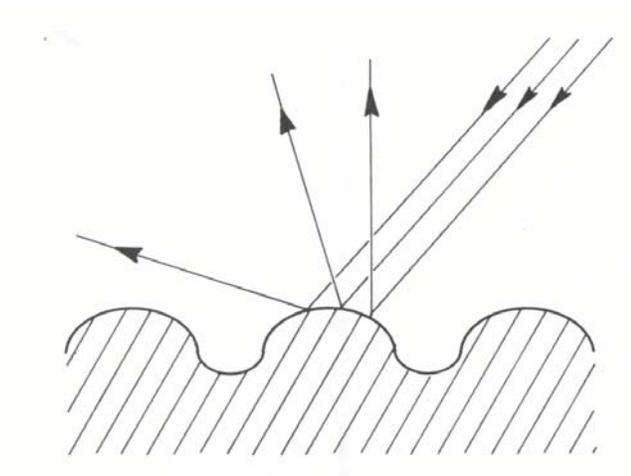


Figure 2.1: Reflection in non-specular surfaces

Moreover, other surface layers might be inhomogeneous on a microscopic scale and thus scatter light rays that penetrate the surface by refraction and reflection at boundaries between regions with differing refractive indices (Figure 2.2). Scattered rays may re-emerge near the point of entry with a variety of directions and so contribute to diffuse, flat or matte reflection. Snow and layers of white paint are examples of surfaces with this kind of behaviour. Frequently both effects occur in surface layers, with some rays reflected at the nearly flat outer surface of the object, while others penetrate deeper and re-emerge after multiple refractions and reflections in the inhomogeneous interior.

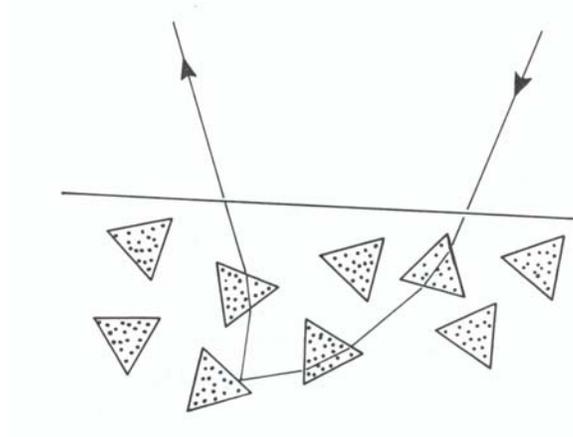


Figure 2.2: Reflection in surface layers with inhomogeneities in refractive index

The distribution of reflected light in each case above depends on the direction of incident rays and the details of the microstructure of the surface layer. Naturally, what constitutes microstructure depends on one's point of view. Surface structures not resolved in a particular imaging situation are taken here to be microstructure. When viewing the moon through a telescope, for example, smaller hillocks and craterlets are

part of this microstructure. This consideration leads to more complicated models of interaction of light with surfaces than those discussed so far. For instance, it is possible to consider an undulating surface covered with a material, that in itself already has a complicated reflecting behaviour (Figure 2.3).

Considerable attention has been paid to the reflective properties of various surface layers. Some researchers have concentrated on the experimental determination of surface reflectance properties while others have developed models for surface layers based on some of the considerations presented above. Unfortunately, models are often too simple to be realistic, or too complicated to yield closed form solutions.

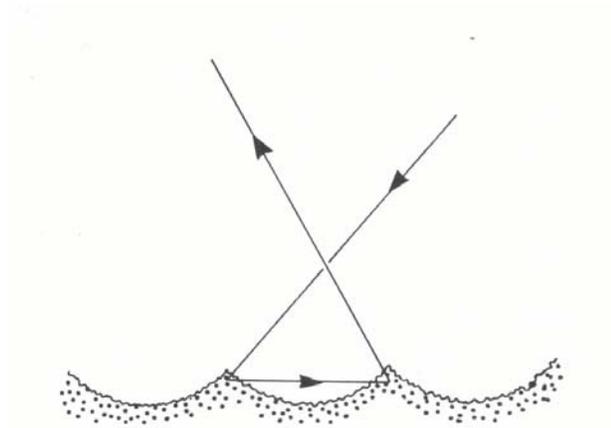


Figure 2.3: Reflection in compound surfaces

2.1.1.2 Simplified Equation

To make the fairly complex SFS problem more manageable, most researchers simplify it with a number of assumptions. The principal simplification arises from the assumption that the viewer and the light sources are far enough away from the object's surface being viewed, i.e., the incident direction and the direction to the viewer are constant. Moreover, it is assumed that the surface is uniform in its reflecting properties and

furthermore it is Lambertian, i.e., the brightness of an oriented portion of the surface is independent of its spatial position. This means that brightness depends only on the orientation of the surface patch irrelevant of its position in space. It is also assumed that the projection system is orthographic rather than perspective, something that simplifies most vision problems.

There are a number of ways to express the orientation of a surface patch. Let $z(x,y)$ define the height of the surface patch above the xy -plane in a Cartesian coordinate system. One way to specify the orientation of the surface at location of (x,y) is to determine the unit normal vector $\mathbf{N}(x,y) = (N_1, N_2, N_3)^T$ where N_i , $i = 1,2,3$ are the components of the unit normal vector along x , y , and z directions, respectively. Another way is to specify the surface gradient components $p(x,y) = \partial z / \partial x$ and $q(x,y) = \partial z / \partial y$. As it is seen, p and q are the slopes of the surface $z(x,y)$ in the x and y directions at location (x,y) , respectively. The surface gradient components p and q are related to the unit surface normal \mathbf{N} by

$$\mathbf{N}(x,y) = (N_1, N_2, N_3)^T = \frac{1}{\sqrt{1+p^2+q^2}}(p, q, 1)^T. \quad (2.1)$$

A third way of specifying surface orientation is to give the slope and the direction of the steepest descent. The terms slant (σ) and tilt (τ) have been introduced for angles used in this fashion, which is unfortunate, since these terms are not mnemonic and are frequently confused. The slant and tilt are related to the unit normal vector by

$$\mathbf{N}(x,y) = (\sin \sigma \cos \tau, \sin \sigma \sin \tau, \cos \sigma)^T. \quad (2.2)$$

Blinn [1977] showed that the image formation equation or reflectance map consists of

three terms. The first term is the effect of the ambient illumination on a surface, the second one is the effect of the diffuse reflectance of a point source's illumination at a surface neighbourhood, and the third is the specular reflectance at the same neighbourhood. In mathematical language, the reflectance map equation which relates the scene radiance (also called intensity or brightness) to the surface geometry is

$$R(p, q) = \rho_a + \rho(\mathbf{N} \cdot \mathbf{L}) + \rho_s(\mathbf{N} \cdot \mathbf{H})^m \quad (2.3)$$

where $R(p, q)$ refers to the brightness of the object patch at point (x, y) , ρ_a , ρ , ρ_s are albedo factors of the surface for ambient illumination, point source illumination, and specular reflectance, respectively, $\mathbf{L} = (l_1, l_2, l_3)^T$ is the unit light source vector, \mathbf{H} is the unit vector that bisects the angle formed by \mathbf{L} and the line of sight, while m describes the scattering of the specular reflectance which would be an infinity large number for a perfect mirror, and “ \cdot ” is the inner product symbol.

Assuming for now that there is no ambient illumination ($\rho_a = 0$) and the object has a near specular reflectance (i.e., $m \rightarrow \infty$) with a distant point light source and a Lambertian surface, then the brightness of the patch is given by the following simplified reflectance map equation [Van Diggelen, 1951; Rindfleisch, 1966; Horn, 1975]

$$R(p, q) = \rho \mathbf{N} \cdot \mathbf{L} = \rho \frac{l_1 p + l_2 q + l_3}{\sqrt{p^2 + q^2 + 1}} \quad (2.4)$$

Obviously, for a given surface material and known illumination direction, the brightness of the patch will depend only on surface orientation. Assuming orthographic projection, Figure 2.4 shows a simple model of image generation.

The general solution of the SFS problem revolves around the so-called image

irradiance equation relating image intensity (shading) to scene radiance (brightness)

$$E(x, y) = R(p, q) = \rho \frac{l_1 p + l_2 q + l_3}{\sqrt{p^2 + q^2 + 1}} \quad \forall x, y \in \Omega \quad (2.5)$$

where $E(x, y)$ is the measured or observed image brightness at point (x, y) in the image domain Ω . The image irradiance equation is a non-linear first-order partial differential equation, as p and q are the derivatives of the normal vector with respect to x and y directions, respectively. On an image with $K \times K$ pixels, with the assumption of known albedo factor ρ and light source direction \mathbf{L} , one should solve for $2K^2$ samples of p and q with K^2 shading observations.

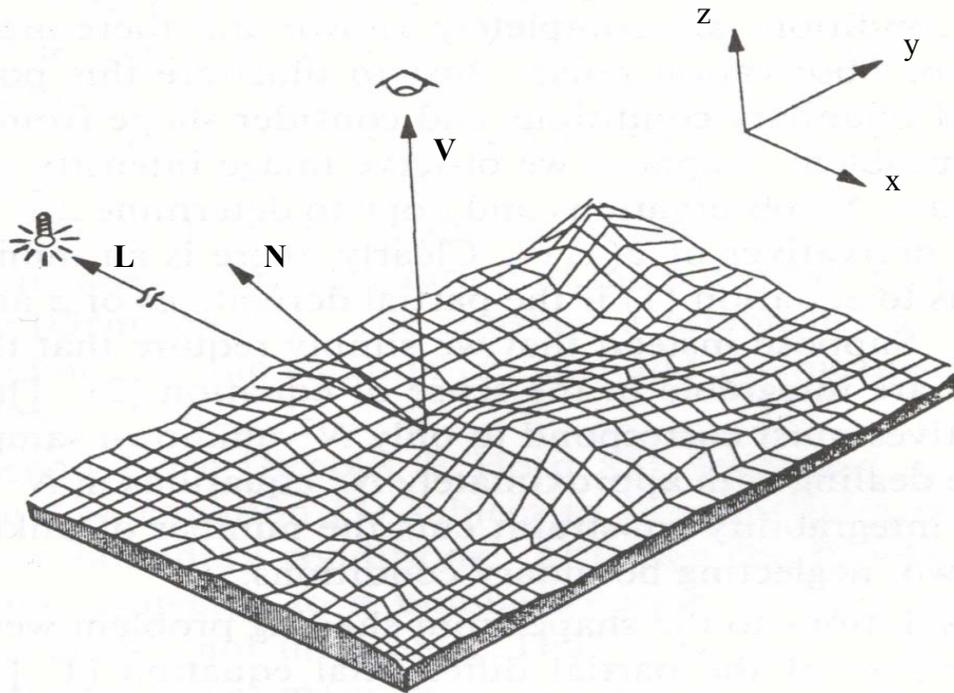


Figure 2.4: A simple model of image geometry. A distant point source illuminant at direction \mathbf{L} , a patch of surface with surface normal \mathbf{N} , and a viewer in direction \mathbf{V}

From a geometrical point of view it is worth mentioning that the irradiance equation (2.5) defines an ambiguity cone about the light source direction (Figure 2.5). In other words, the individual surface normals for which the irradiance equation is valid must assume directions that fall on this cone.

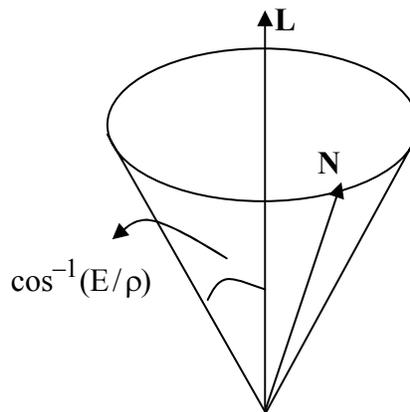


Figure 2.5: The ambiguity cone

2.1.2 SFS Solution Techniques

There have been two general classes of algorithms developed for the SFS problem: global algorithms, which propagate information across a shaded surface starting from points with known surface orientation, and local algorithms, which attempt to estimate shape information from local variations in image intensity.

Global algorithms, primarily due to Horn and his students [Horn, 1977; Ikeuchi and Horn, 1981; Horn and Brooks, 1986], usually take two steps to estimate the surface shape. In the first step, the measured image brightness $E(x,y)$ is used to estimate the surface orientations (slopes) p 's and q 's at each location of x and y . In this process, there

is usually some assumption about the surface shape. Most of the time the assumption is that the surface is smooth in some sense. In the second step, a subsequent integration is normally required to convert the estimated surface orientation into an estimate of surface shape ($z(x,y)$).

On the other hand, local algorithms originally suggested by Pentland [1984] use assumptions about local surface shape in order to extract estimates of surface orientation from the shading information within a small image neighborhood. As with the global algorithms, integration is normally required to obtain the surface shape. The following subsections review different SFS solution techniques more closely.

2.1.2.1 Characteristic Strip Approach

In this global approach, the image irradiance equation (2.5) is reduced to an equivalent set of five coupled ordinary differential equations called the characteristic strip equations. In order to explain the idea, assume a point at location (x,y) on the surface of the object with height $z(x,y)$. It is required to extend the solution a small distance in some arbitrary direction by taking steps δx in x and δy in y . Knowing the components of the surface gradients, i.e. p and q , one can write

$$\delta z = p\delta x + q\delta y . \quad (2.6)$$

So, to explore the surface one needs to keep track of p and q in addition to x , y , and z . In other words, the changes in p and q should be computed along the step. This can be done using

$$\begin{aligned} \delta p &= r\delta x + s\delta y \\ \delta q &= s\delta x + t\delta y \end{aligned} \quad (2.7)$$

where $r = z_{xx}$, $s = z_{xy} = z_{yx}$, and $t = z_{yy}$ are the second partial derivatives of the height.

On the other hand by using the image irradiance equation (2.5) one can find the gradient of image intensity as

$$\begin{aligned} E_x &= rR_p + sR_q \\ E_y &= sR_p + tR_q. \end{aligned} \tag{2.8}$$

where R_p and R_q are the partial derivatives of the scene radiance R with respect to p and q , E_x and E_y are the partial derivatives of the observed image brightness E with respect to x and y . As there is no limitation in choosing the direction of the step $(\delta x, \delta y)$, one can pick the arbitrary direction $\delta \xi$ and have

$$\begin{aligned} \delta x &= R_p \delta \xi \\ \delta y &= R_q \delta \xi. \end{aligned} \tag{2.9}$$

Then, substituting eq. (2.9) into eq. (2.7) and simplifying the result by using eq. (2.8), one can write

$$\begin{aligned} \delta p &= E_x \delta \xi \\ \delta q &= E_y \delta \xi. \end{aligned} \tag{2.10}$$

Finally, the eqs. (2.6), (2.9) and (2.10) can be summarized in the set of ordinary differential equations as

$$\begin{aligned}
\frac{dx}{d\xi} &= R_p \quad , \quad \frac{dy}{d\xi} = R_q, \\
\frac{dz}{d\xi} &= pR_p + qR_q, \\
\frac{dp}{d\xi} &= E_x \quad \text{and} \quad \frac{dq}{d\xi} = E_y,
\end{aligned}
\tag{2.11}$$

where ξ is the parameter which varies monotonically along a particular curve named characteristic strip [Horn, 1977; Bruss, 1979].

A particular solution of these equations generates a so-called characteristics curve on the surface, along with surface orientation on the curve. The projection of such a curve into the image is called a base characteristic. The characteristic curve, along with the orientation, defines a characteristic strip on the surface. Many closely spaced strips define the shape of the surface.

When the characteristic strip is extended, one moves to a new point (x,y) in the image plane and a new point (p,q) in the reflectance map plane (Figure 2.6) [Horn 1977; Woodham, 1977]. The orientation of the surface is known all along the strip. One has to move in a particular direction, $(\delta x, \delta y)$, in the image plane in order to be able to compute the orientation at the new point (Figure 2.6a). The equations above show that this direction is along the steepest ascent, (R_p, R_q) , at the corresponding point in the reflectance map (Figure 2.6b). At the new image point so determined, a different value of image intensity is found, corresponding to a new point in the reflectance map. This point in the reflectance map also lies in a well-defined direction, $(\delta p, \delta q)$, from the previous point in the reflectance map. From the equations above one knows that this direction is along the steepest ascent, (E_x, E_y) , at the corresponding point in the image.

It turns out that direct numerical implementations of the above equations do not yield particularly good results, since the paths of the characteristics are affected by noise in the image brightness measurements and errors tend to accumulate along their length. In particularly bad cases, the base characteristics may even cross, which doesn't make any sense in terms of surface shape. It is possible, however, to grow characteristic strips in parallel and use a so-called sharpening process to keep neighboring characteristics consistent by enforcing conditions like either $dz = p dx + q dy$ and $E(x, y) = R(p, q)$ along curves connecting the tips of characteristics or smoothness of the surface between strips advancing in parallel [Horn, 1970, 1975]. This greatly improves the accuracy of the solution, since the computation of surface orientation is tied more closely to image brightness itself rather than to the brightness gradient. This also makes it possible to interpolate new characteristic strips when existing strips spread too far apart, and to remove some when they approach each other too closely.

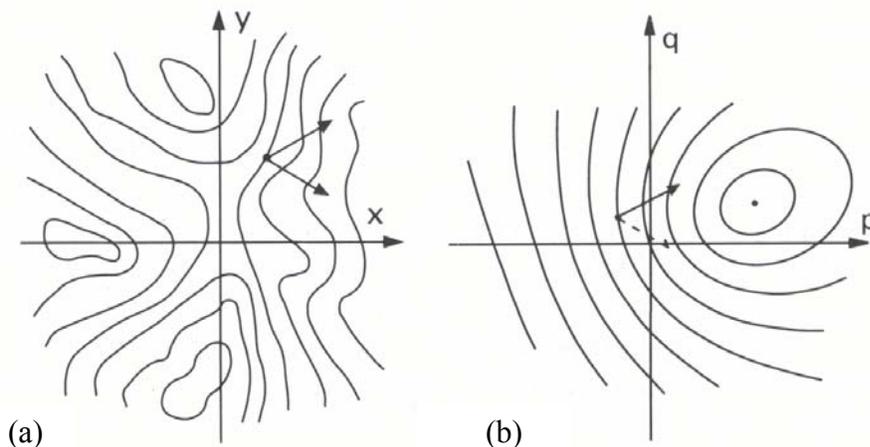


Figure 2.6: Geometric interpretation of the characteristic strip equations. (a) Iso-brightness map (b) Reflectance map [Ikeuchi and Horn; 1981]

Strips may be started near a so-called singular point and grow outwards from there. Because of the directionality of the progress of the solution, information from the ends of the characteristic strips cannot be exploited by this method. This is an important shortcoming since crucial information is obtained from occluding boundaries. Normally, these would be reached at the end of the computations, with no means available to influence the completed solution.

Algorithms based directly on the numerical solution of the discrete approximations of the characteristic strip equations are inherently sequential in nature and as emphasised before, have difficulty with unavoidable noise in the image data.

2.1.2.2 Iterative Solution on a Regular Grid of Points

The method of characteristic strip expansion suggests neither a reasonable solution scheme, nor efficient and robust computational methods for the SFS problem. Alternative global solutions for the SFS equation are the iterative methods which can be implemented in two steps on a grid of locally interconnected cells. The first step consists of estimating the surface orientations using shading information, while the second step involves an integration of the estimated orientations to get an estimate of the surface shape (heights).

As the first step, iterative approaches repeatedly make adjustments to surface orientations until the predicted shading, based on the estimated shape, matches the actually observed shading in the image. In most of these methods, the shape is specified not by height above a reference plane, but by surface orientation. Orientation estimates are stored for every point in a dense grid of points, usually one point for every pixel.

However, the resulting field of surface orientations will most likely not correspond to a continuous surface.

As mentioned in the previous section, eq. (2.5) has only a theoretical value. In practice, modeling errors such as reflectance map mismatch, imperfect knowledge of the light source, spatial and radiometric quantization errors, observation noise, and albedo variations are inevitable. These factors all influence the existence and uniqueness of a solution to eq. (2.5) and the estimation of a good solution in the case that a unique one does not exist. In order to take the errors into account, one can rewrite eq. (2.5) as follows

$$E(x, y) = R(p, q) + e(x, y) = \rho \frac{l_1 p + l_2 q + l_3}{\sqrt{p^2 + q^2 + 1}} + e(x, y) \quad \forall x, y \in \Omega \quad (2.12)$$

where $e(x, y)$ denotes the brightness error at location (x, y) .

The iterative SFS methods are based on a variational formulation in which the surface normals (or surface gradients) are determined by minimizing an energy function over the entire (or part of the) image. The very first thing that the energy function contains is the brightness error which is derived directly from the image irradiance equation (2.12). In mathematical notation, the energy function to be minimized is

$$I(p, q) = \iint_{\Omega} |e(x, y)|^2 d\Omega = \iint_{\Omega} (E(x, y) - R(p, q))^2 d\Omega \quad (2.13)$$

where the domain of integration Ω corresponds to the entire or part (or the union of parts) of the image. The minimization of this energy function with respect to the unknowns p and q results in a surface such that the difference between the observed image brightness and the reconstructed one is minimized in the least-squares sense.

After estimating the surface orientation \hat{p} and \hat{q} , the second step in the iterative SFS methods is to recover the surface $z(x,y)$ from the estimated orientation. Given \hat{p} and \hat{q} , a line integration method or an approximate minimization of the functional as

$$\iint_{\Omega} \{(z_x - \hat{p})^2 + (z_y - \hat{q})^2\} d\Omega \quad (2.14)$$

is used, where z_x and z_y are the unknown surface gradients along x and y directions, respectively. This leads to the following equation

$$\Delta z = z_{xx} + z_{yy} = \hat{p}_x + \hat{q}_y \quad (2.15)$$

where Δ is the Laplacian operator, \hat{p}_x and \hat{q}_y are the derivatives of the estimated surface gradients along x and y directions, respectively. A discrete three-point approximation of the Laplacian operator is [Engeln-Mullges and Uhlig, 1996]

$$\{\Delta T\}_{i,j} \approx \frac{4}{\varepsilon^2} (\bar{T}_{i,j} - T_{i,j}) \quad (2.16)$$

where T is a function, indices i and j refer to the location of the cell in x and y directions respectively,

$$\bar{T}_{i,j} = \frac{1}{4} (T_{i,j+1} + T_{i,j-1} + T_{i+1,j} + T_{i-1,j}), \quad (2.17)$$

and ε is the spacing between cells. Using this approximation, the following iterative update equation for the surface depth is obtained

$$z_{i,j}^{n+1} = \bar{z}_{i,j}^n - \frac{\varepsilon^2}{4} (\{\hat{p}_x\}_{i,j}^n + \{\hat{q}_y\}_{i,j}^n) \quad (2.18)$$

where $\bar{z}_{i,j}^n = \frac{1}{4}(z_{i+1,j}^n + z_{i-1,j}^n + z_{i,j+1}^n + z_{i,j-1}^n)$, the terms in braces are numerical estimates of the indicated derivatives at the picture cell (i,j) and the superscript n denotes the iteration number.

Unfortunately, the minimization problem in eq. (2.13) is an ill-posed one. An ill-posed problem is one that does not have any solution, has infinitely many solutions, or has a solution that is very sensitive to the given data. The SFS problem is inherently ill-posed, since there are infinitely many surfaces that yield the same shading pattern. However, the problem is not necessarily ill-posed if information from the objects is taken into account. In this case, a regularization method is used to obtain a unique solution. Basically, a particular solution is selected out of an infinite number of possible solutions.

As information from the objects, one should remember that the neighbouring orientations have to correspond to some underlying surface, therefore, they cannot be chosen independently. Therefore, most of the times, it is the integrability, and/or smoothness of the surface which are used as the constraints for the surface orientation.

The smoothness constraint ensures a smooth surface in order to stabilize the convergence to a unique solution. Ikeuchi and Horn [1981], Brooks and Horn [1985], and Horn [1990] have used the smoothness constraint. This constraint is applied by minimizing the following functional

$$\iint_{\Omega} \{p_x^2 + p_y^2 + q_x^2 + q_y^2\} d\Omega \quad (2.19)$$

where the subscripts x and y correspond to the respective partial derivatives. Another less restrictive version of the smoothness constraint is obtained by minimizing a

functional which contains the change of slope only in the x and y directions

$$\iint_{\Omega} \{p_x^2 + q_y^2\} d\Omega. \quad (2.20)$$

The smoothness constraint can also be described in terms of the surface normal \mathbf{N} by minimizing the functional

$$\iint_{\Omega} \{\|\mathbf{N}_x\|^2 + \|\mathbf{N}_y\|^2\} d\Omega. \quad (2.21)$$

where $\|\cdot\|$ is the norm and again the subscripts x and y show the derivative directions.

This means that the surface normal should change gradually.

Introducing a Lagrangian multiplier $\beta(x,y)$ to enforce the constraint, one can minimize the following cost function

$$\iint_{\Omega} [\{p_x^2 + p_y^2 + q_x^2 + q_y^2\} + \beta(x,y)(E(x,y) - R(p,q))] d\Omega. \quad (2.22)$$

The corresponding Euler equations are (see Appendix A for details)

$$\begin{aligned} \Delta p + \beta(x,y)R_p &= 0 \\ \Delta q + \beta(x,y)R_q &= 0 \end{aligned} \quad (2.23)$$

where Δ is the Laplacian operator. After elimination of the Lagrangian multiplier $\beta(x,y)$, the following pair of equations are left [Horn, 1990]

$$\begin{aligned} R_q \Delta p &= R_p \Delta q \\ E(x,y) &= R(p,q). \end{aligned} \quad (2.24)$$

Unfortunately, no convergent iterative scheme has been found for this constrained variational problem [Horn and Brooks, 1986].

A quite different way of setting up the cost function is to use the smoothness

constraint as a regularization (penalty) term [Ikeuchi and Horn, 1981] and looking instead for a minimum of the following function with respect to the unknowns p and q

$$\iint_{\Omega} ((E(x, y) - R(p, q))^2 + \lambda \{p_x^2 + p_y^2 + q_x^2 + q_y^2\}) d\Omega \quad (2.25)$$

where λ is a scalar that weighs the relative importance of the regularization term. Note that λ here is not a Lagrangian multiplier, but a factor that balances the relative contributions of the brightness error term and the term measuring departures from smoothness. It establishes a trade-off between smoothness of the solution and the mean-square value of the brightness error $E(x,y)-R(p,q)$. That is, there is no absolute constraint imposed here, only a penalty term added that increases with departure from smoothness. Since the smoothness constraint stabilizes the solution in the sense of optimization, analog to the penalty functional, it is also called a stabilizing functional.

The regularization parameter λ adjusts the relation between the error minimization requirement and the smoothness constraint and it controls the roughness of the reconstructed surface model. It is clear if a large value of λ is selected the algorithm will produce a smoothed surface and if a small value of λ is selected the minimization process will face numerical instability. An appropriate choice of λ is not always an easy or a trivial job.

It should be pointed out as well that a solution to this regularized problem is not a solution of the original problem (eq. 2.13), although it may be close to some solution of the original problem. The solution for this regularized problem comes from the calculus of variations and leads to the following coupled pair of second-order partial differential equations (see Appendix A for details)

$$\begin{aligned}\lambda\Delta p &= -(E(x, y) - R(p, q))R_p(p, q) \\ \lambda\Delta q &= -(E(x, y) - R(p, q))R_q(p, q)\end{aligned}\quad (2.26)$$

where the subscripts p and q show the derivative directions. Using a discrete three-point approximation of the Laplacian operator given in eq. (2.16), one can immediately arrive at the set of the following iterative update equations

$$\begin{aligned}p_{i,j}^{n+1} &= \bar{p}_{i,j}^n + \frac{\varepsilon^2}{4\lambda} (E_{i,j} - R(p_{i,j}^n, q_{i,j}^n))R_p(p_{i,j}^n, q_{i,j}^n) \\ q_{i,j}^{n+1} &= \bar{q}_{i,j}^n + \frac{\varepsilon^2}{4\lambda} (E_{i,j} - R(p_{i,j}^n, q_{i,j}^n))R_q(p_{i,j}^n, q_{i,j}^n)\end{aligned}\quad (2.27)$$

where

$$\begin{aligned}\bar{p}_{i,j}^n &= \frac{1}{4} (p_{i+1,j}^n + p_{i-1,j}^n + p_{i,j+1}^n + p_{i,j-1}^n) \\ \bar{q}_{i,j}^n &= \frac{1}{4} (q_{i+1,j}^n + q_{i-1,j}^n + q_{i,j+1}^n + q_{i,j-1}^n),\end{aligned}\quad (2.28)$$

(i, j) is the cell location and the superscript n is the iteration number. Due to the fact that the reflectance map $R(p, q)$ is typically nonlinear and also because of the dimension of this system of equations, they are solved iteratively. Moreover, it is worth mentioning that there are several methods for approximating the Laplacian operator (eq. 2.16), including four-point, five-point and nine-point approximations [Engeln-Mullges and Uhlig, 1996]. However, it is well known that, the more points involved in the approximation, the more computations are involved, while the larger the lower-order error is [Horn, 1986].

In general, this approach produces solutions that are too smooth, with the amount of distortions depending on the choice of the parameter λ . For related reasons, this

algorithm does well only on simple smooth shapes, and does not perform well on complex, wrinkled surfaces.

On the other hand the integrability constraint ensures valid surfaces for which $z_{xy} = z_{yx}$. The functional describing this constraint can be

$$\iint_{\Omega} \{p_y - q_x\}^2 d\Omega \quad (2.29)$$

again where the subscripts show the derivative directions. Like the smoothness constraint one can set up a cost function similar to eq. (2.25) and by its minimization come up with the corresponding coupled partial differential equations [Horn and Brooks, 1986] by which the iterative scheme for computing the second derivatives of p and q is derived. This solution avoids the excessive smoothing of the one described earlier, but appears to be less stable, in the sense that it diverges under a wider set of circumstances.

Zheng and Chellappa [1991] applied the intensity gradient constraint instead of a smoothness regularization. The intensity gradient constraint requires that the intensity gradient of the reconstructed image be close to the intensity gradient of the observed image in both the x and y directions. This constraint can be described by minimizing the functional

$$\iint_{\Omega} \left\{ (R_x - E_x)^2 + (R_y - E_y)^2 \right\} d\Omega. \quad (2.30)$$

Generally speaking, the convergence of iterative solutions is questionable. Moreover, the solutions are typically not very accurate. There are two reasons for this: methods that enforce integrability can get stuck in local minima in their search for the global extremum, and methods that do not enforce integrability inherently trade off increased

surface smoothness against departures from an exact match of the shading information.

2.1.2.2.1 Existence and Uniqueness

Questions of existence and uniqueness of solutions of the SFS problem have still not been completely resolved yet. Assuming known reflecting properties of the surface and the light source vector, the question is if there is always a surface shape that will generate, under these conditions, any given (arbitrary) image brightness pattern. The answer is not known. It may be that there are patterns that could not have been produced as the result of shading on any three-dimensional shape. Then in this case, the SFS problem has no solution. People often appear to be able to tell that a particular pattern is not due to shading but to spatial variations in the reflecting properties of the surface.

The question of uniqueness is more difficult to answer. This is due to the fact that SFS is inherently ill-posed. However, there will be some unique surface orientation for which the brightness is a maximum (or a unique minimum, in some unusual cases). A point in the image where this maximal brightness is observed is called a singular point. These image points have particular importance since the surface orientation at the corresponding point on the surface is immediately known (provided, of course, that the reflectance map is given).

2.1.2.3 Local Approach

As mentioned in the previous sections, brightness information in an arbitrary image patch is, in general, infinitely ambiguous. However, if sufficient strong assumptions are made about the surface, some useful information can be recovered from the first- and second-

order variations of brightness within a patch. One assumption that leads to interesting results is that the surface is everywhere locally spherical or cylindrical. The catch is any point on these surfaces has equal, and locally constant principal curvatures. Such local methods have shown promising results.

The idea behind the local methods comes from the fact that the measured image brightness in addition to its first and second derivatives, i.e., $E, E_x, E_y, E_{xx}, E_{xy}, E_{yy}$, provide six independent measurements by which one can determine at most six image-formation parameters. On the other hand, six parameters are required to specify the image brightness in the neighborhood of a spherical (cylindrical) point on a Lambertian surface [Pentland, 1989]. These six parameters are two for surface normal (tilt (τ) and slant (σ) for example), one for the radius of curvature (r), two for the direction of unit illumination vector (l_1 and l_2), and one for surface albedo (ρ). Obviously, additional parameters would be required for any more complex family of shapes.

Pentland (1989) showed that under the assumption of a Lambertian spherical (cylindrical) surface, there would be a unique (up to a mirror reversal) combination of six image formation parameters based on any six values of $E, E_x, E_y, E_{xx}, E_{xy}, E_{yy}$ with $E_{xx}E_{yy} \geq 0$. It has also shown that one can compute the six image formation parameters based on the following equations

$$\begin{aligned}
\tau &= \tan^{-1} \frac{-(E_{xx} - E_{yy}) \pm \sqrt{(E_{xx} - E_{yy})^2 + 4E_{xy}^2}}{2E_{xy}} \\
\sigma &= \cos^{-1} \sqrt{\frac{xy \Delta E - (x^2 + y^2)E_{xy}}{xy \Delta E + (x^2 + y^2)E_{xy}}} \\
\rho l_1 &= -\frac{E_{xy}N_3^2r^2}{N_2} + E_x r \\
\rho l_2 &= -\frac{E_{xy}N_3^2r^2}{N_1} + E_y r \\
r &= \frac{N_1N_2}{2E_{xy}N_3^2} \left((N_1E_x + N_1E_y) \pm \sqrt{(N_1E_x + N_2E_y)^2 - \frac{4EE_{xy}N_3^2}{N_1N_2}} \right) \\
\rho &= \sqrt{\left(-\frac{E_{xy}N_3^2r^2}{N_2} + E_x r \right)^2 + \left(-\frac{E_{xy}N_3^2r^2}{N_1} + E_y r \right)^2 + \left(-\frac{E_{xy}N_3^3r^2}{N_1N_2} \right)^2}.
\end{aligned} \tag{2.31}$$

where N_1, N_2 , and N_3 are the components of the unit surface normal vector along x , y , and z directions.

The only ambiguity in these equations is in determining the tilt τ . Pentland [1989] also showed that the direction of maximum d^2E is also the tilt direction, by which the tilt ambiguity is resolved.

Lee and Rosenfeld [1985] later improved this technique so that it requires only first derivatives of image brightness, thereby reducing sensitivity to noise. Two difficulties with this local approach are the local sphericity assumption and the reliance on intensity derivatives, which can be very noisy. A third difficulty is that no provision is made for enforcing global consistency (such as integrability) of the local slopes so that it is difficult to construct a surface from the slope estimates.

2.1.2.4 Photometric Stereo Approach

The basic idea in this method is to get around the ambiguity inherent in a single measurement of image brightness by taking more measurements under different lighting conditions, rather than exploiting surface continuity or smoothness. In spite of the fact that there is seldom any opportunity to change the lighting at will, it leads to a method of great importance. The reason is that the recovery of surface orientation is completely local and very simple, involving little more than a lookup table without smoothness assumption. Calibration for different surface materials and different lighting conditions is also straightforward, requiring only observations of an object of known shape, such as a sphere, and construction of the lookup table.

Since the images are obtained from the same position, a particular point on the object will appear at the same spot in each image. This means that one does not have the problem of identifying projections of a particular surface feature in multiple views, as happens in ordinary stereo. A different reflectance map applies to each image, however, since the lighting is different for each view. For a given point in the image there is one brightness value corresponding to each of these reflectance maps. Suppose, for example, that a particular point, (x, y) , two measurements of image brightness are available. Then, there are two (nonlinear) equations for p and q

$$\begin{aligned} E_1(x, y) &= R_1(p, q) \\ E_2(x, y) &= R_2(p, q) \end{aligned} \tag{2.32}$$

where R_1 and R_2 are the reflectance maps corresponding to the two lighting situations, while E_1 and E_2 are the observed image brightness values at location (x, y) . The

intersection of the corresponding gradient-space contours provides one with the sought after surface orientation. This is the essential idea of photometric stereo. Naturally, the above pair of nonlinear equations in two unknowns may have more than one solution, in which case additional information (such as a third image) may be needed to find the unique answer.

Ikeuchi [1981] and Silver [1980] have come up with the lookup table idea. In this table, the quantized brightness values are used as indices, while the corresponding surface orientations are the entries in the table. Detection of errors is facilitated (if more than two images are used) by blank entries that represent incompatible combinations of brightness values. Values of p and q are found for every point in the image, yielding the surface orientations for the corresponding surface patches.

2.1.2.5 Linear Approach

Linear methods reduce the nonlinear problem into a linear one through the linearization of the reflectance map. The idea is based on the assumption that the lower order components in the reflectance map dominate. Therefore these algorithms only work under this assumption.

Pentland [1988] used the linear approximation of the reflectance map in p and q . The image irradiance equation for a Lambertian surface can be expressed as follows

$$E(x,y) = R(p,q) = \frac{\cos \sigma_s + p \cos \tau_s \sin \sigma_s + q \sin \tau_s \cos \sigma_s}{\sqrt{1 + p^2 + q^2}} \quad (2.33)$$

where σ_s and τ_s are the slant and tilt of the light source respectively. By taking the Taylor series expansion of the reflectance function about $p = p_0$, $q = q_0$, and ignoring

higher order terms, one can have

$$E(x, y) = R(p_0, q_0) + \left. \frac{\partial R(p, q)}{\partial p} \right|_{(p_0, q_0)} (p - p_0) + \left. \frac{\partial R(p, q)}{\partial q} \right|_{(p_0, q_0)} (q - q_0). \quad (2.34)$$

Considering eq. (2.33), the above equation at $p_0 = q_0 = 0$ reduces to

$$E(x, y) = \cos \sigma_s + p \cos \tau_s \sin \sigma_s + q \sin \tau_s \cos \sigma_s. \quad (2.35)$$

Next, Pentland takes the Fourier transform of both sides of the equation. Since the first term on the right is a constant term, it can be dropped. Using the identities

$$\begin{aligned} \frac{\partial}{\partial x} z(x, y) &\leftrightarrow F_z(\omega_1, \omega_2)(-i\omega_1) \\ \frac{\partial}{\partial y} z(x, y) &\leftrightarrow F_z(\omega_1, \omega_2)(-i\omega_2) \end{aligned} \quad (2.36)$$

where F_z is the Fourier transform of $z(x, y)$, and $i = \sqrt{-1}$, eq. (2.35) can be rewritten as

$$F_E = F_z(\omega_1, \omega_2)(-i\omega_1) \cos \tau_s \sin \sigma_s + F_z(\omega_1, \omega_2)(-i\omega_2) \sin \tau_s \cos \sigma_s \quad (2.37)$$

where F_E is the Fourier transform of the image $E(x, y)$. The surface height $z(x, y)$ can be computed by rearranging the terms in the above equation and then taking the inverse Fourier transform.

This algorithm gives a non-iterative, closed-form solution using Fourier transforms. The problem lies in the linear approximation of the reflectance map, which causes trouble when the nonlinear terms are large. As pointed out by Pentland, when the quadratic terms in reflectance map dominate, the frequency doubling occurs, in this case, the recovered shape will not be consistent with the illumination conditions.

Tsai and Shah [1994] employed the discrete approximations of p and q using finite differences in order to linearize the reflectance map in terms of z . Using $p = z_{i,j} - z_{i-1,j}$

and $q = z_{i,j} - z_{i,j-1}$ as the discrete approximations for p and q where indices of (i,j) are the discrete values of (x,y) location, the image irradiance equation can be written as

$$f(E_{i,j}, z_{i,j}, z_{i-1,j}, z_{i,j-1}) = E_{i,j} - R(z_{i,j} - z_{i-1,j}, z_{i,j} - z_{i,j-1}) = 0. \quad (2.38)$$

For a fixed point (i,j) and a given image brightness E , a linear approximation (Taylor series expansion up to the first-order terms) of the function f (eq. 2.38) about a given height z^{n-1} where n is the iteration number is

$$\begin{aligned} f(E_{i,j}, z_{i,j}, z_{i-1,j}, z_{i,j-1}) &\approx f(E_{i,j}, z_{i,j}^{n-1}, z_{i-1,j}^{n-1}, z_{i,j-1}^{n-1}) + \\ &(z_{i,j} - z_{i,j}^{n-1}) \frac{\partial}{\partial z_{i,j}} f(E_{i,j}, z_{i,j}^{n-1}, z_{i-1,j}^{n-1}, z_{i,j-1}^{n-1}) + \\ &(z_{i-1,j} - z_{i-1,j}^{n-1}) \frac{\partial}{\partial z_{i-1,j}} f(E_{i,j}, z_{i,j}^{n-1}, z_{i-1,j}^{n-1}, z_{i,j-1}^{n-1}) + \\ &(z_{i,j-1} - z_{i,j-1}^{n-1}) \frac{\partial}{\partial z_{i,j-1}} f(E_{i,j}, z_{i,j}^{n-1}, z_{i-1,j}^{n-1}, z_{i,j-1}^{n-1}) = 0. \end{aligned} \quad (2.39)$$

For a $K \times K$ image, there are K^2 such equations, which will form a linear system. This system can be solved easily using the Jacobi iterative scheme, which simplifies eq. (2.39) into the following equation

$$f(z_{i,j}) \approx f(z_{i,j}^{n-1}) + (z_{i,j} - z_{i,j}^{n-1}) \frac{\partial}{\partial z_{i,j}} f(z_{i,j}^{n-1}) = 0. \quad (2.40)$$

Then, for $z_{i,j} = z_{i,j}^n$, the surface height at the n th iteration, can be solved directly as

$$z_{i,j}^n = z_{i,j}^{n-1} + \frac{-f(z_{i,j}^{n-1})}{\frac{\partial}{\partial z_{i,j}} f(z_{i,j}^{n-1})}. \quad (2.41)$$

The implementation of this algorithm is straightforward. Based on eq. (2.41), the depth is

computed here by a simple division without any matrix inversion. This is a simple but efficient algorithm. However, special care is needed to avoid division by zero.

2.1.3 General Comments

The assumptions required for the simplified SFS problem are very strong. They would not in practice be met except in a strictly controlled environment. Now the question is how likely is it that the assumptions are met in an uncontrolled environment.

A known reflectance map or Lambertian, is a fair assumption, although few surfaces are precisely Lambertian [Harrison, 1945] so some errors in shape would arise. More importantly, many surfaces exhibit specular behaviour. Modelling such behaviour in SFS has not proved very successful in practice [Grimson, 1984].

Constant albedo may be a reasonable assumption for manufactured parts, but not for natural textured surfaces. Fortunately, if a surface is textured, many features will be found there for stereo matching. In this respect, stereo and SFS are neatly complementary.

Known illumination is another questionable assumption. Admittedly, if it is known that there is just one point light source then the estimation technique of Pentland [1982] should be able to discover its location. Obviously, in case of more than one illumination source, it is inconceivable that any manageable model will take account of secondary reflection and shadowing which would inevitably lead to gross errors in the reconstructed surface.

Even if all the assumptions are met, the uniqueness of the solution is still questionable. Non uniqueness can be present either at the level of the problem, or of the

algorithm. At the problem level, it may be that there are multiple surfaces consistent with the given data (as, for example, with inversion of the surface). However, at the algorithmic level, there may be several possible outputs, not all of which need be true solutions to the stated problem. Unfortunately, no entirely satisfactory scheme has yet been proposed that guarantees a unique solution. Nevertheless, imposing additional constraints can help one to get a unique solution. However, how close this solution is to the true surface depends on the degree to which the constraints are realistic.

2.2 Characteristics of Remotely Sensed Imageries

This section discusses the general characteristics and details of remote sensing systems, images of which are used as the input for the specific SFS problem of this investigation. The imaging systems that have particularly been used in this research work, i.e. Landsat MSS and TM, Landsat 7 ETM+, SPOT, IKONOS, and IRS are briefly reviewed in Appendix G.

2.2.1 Remote Sensing Systems

From a geomatics perspective, remote sensing is a radiometric and/or radar imaging technology for extracting physical information about the Earth and the oceans without having any physical contact with the surfaces under investigation [Collins, 2002]. Today, there are several satellite and airborne remote sensing systems being used in the world. Like our own visual system, any remote sensing imagery provides three primary types of information: 1) spatial (shapes and locations of the objects), 2) spectral (colors of the objects) and 3) radiometric (number of perceptible colors). However, one would not be

able to use the images in an intelligent way unless the details of many necessary corrections and preprocessings applied to them are known.

Generally speaking, any remote sensing system consists of the following four essential components: 1) a source of electromagnetic (EM) radiation, 2) a propagation medium between the source and the object as well as the object and the sensor, 3) an object or surface of interest, and 4) an instrument which measures the reflected and/or emitted radiation from the object or surface (Figure 2.7). Obviously, a good understanding of these four basic components is necessary for efficient use of any remote sensing data.

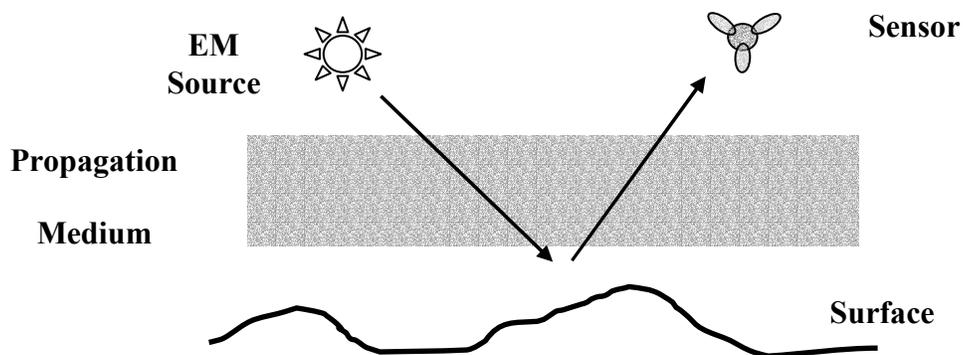


Figure 2.7: Remote sensing system components

Based on the types of the sensors used, remote sensing systems are classified into two different groups: 1) passive imagers or radiometers and 2) active imagers or radars. Radiometer imaging are those systems in which the energy radiated by the Sun, is reflected by the object and received by the sensor. In other words, in these systems the source of EM energy is the Sun. These systems operate in the visible, infrared and microwave regions of the EM spectrum. On the other hand, radar imaging systems are those in which the sensor works as both the source and the receiver. Since radar

imageries have not been used in this research, therefore, there is no discussion about these systems here.

2.2.2 The Source of EM Radiation

As mentioned before, the Sun is the primary source of EM radiation for all passive or radiometer imagers. The concept of blackbody radiation can perfectly explain its radiant exitance at any wavelength. A blackbody is defined as an idealized material that absorbs all incident radiation at all frequencies and all temperatures without reflecting any. Based on Planck's model, the spectral radiance from a blackbody follows [Richards, 1993]

$$L_{bb} = \frac{2hc^2}{\lambda^5} \left[\frac{1}{\exp(ch/\lambda kT) - 1} \right] \quad (2.42)$$

where, h is Planck's constant (6.63×10^{-34} joules), k is Boltzmann's constant (1.38×10^{-23} joules/Kelvin), c is the speed of light in vacuum (2.9979×10^8 m/s) and T is the temperature of the blackbody in degrees Kelvin. The spectral radiance is typically measured in Watts per square metre per metre. As an example, Figure 2.8 shows the energy radiation from perfect blackbodies as a function of wavelength and temperature.

The energy that the Sun radiates travels in space isotropically in an inverse square law fashion. This power density is called irradiance, and can be used to describe the strength of any emitter of EM energy. After hitting the Earth's surface, assuming that the surface is perfectly diffuse, it scatters uniformly into the upper hemisphere. The amount of power density scattered in a particular direction is defined by its density per solid angle, since equal amounts are scattered into equal cones of solid angles. This quantity is called radiance and has units of Watts per square metre per steradian.

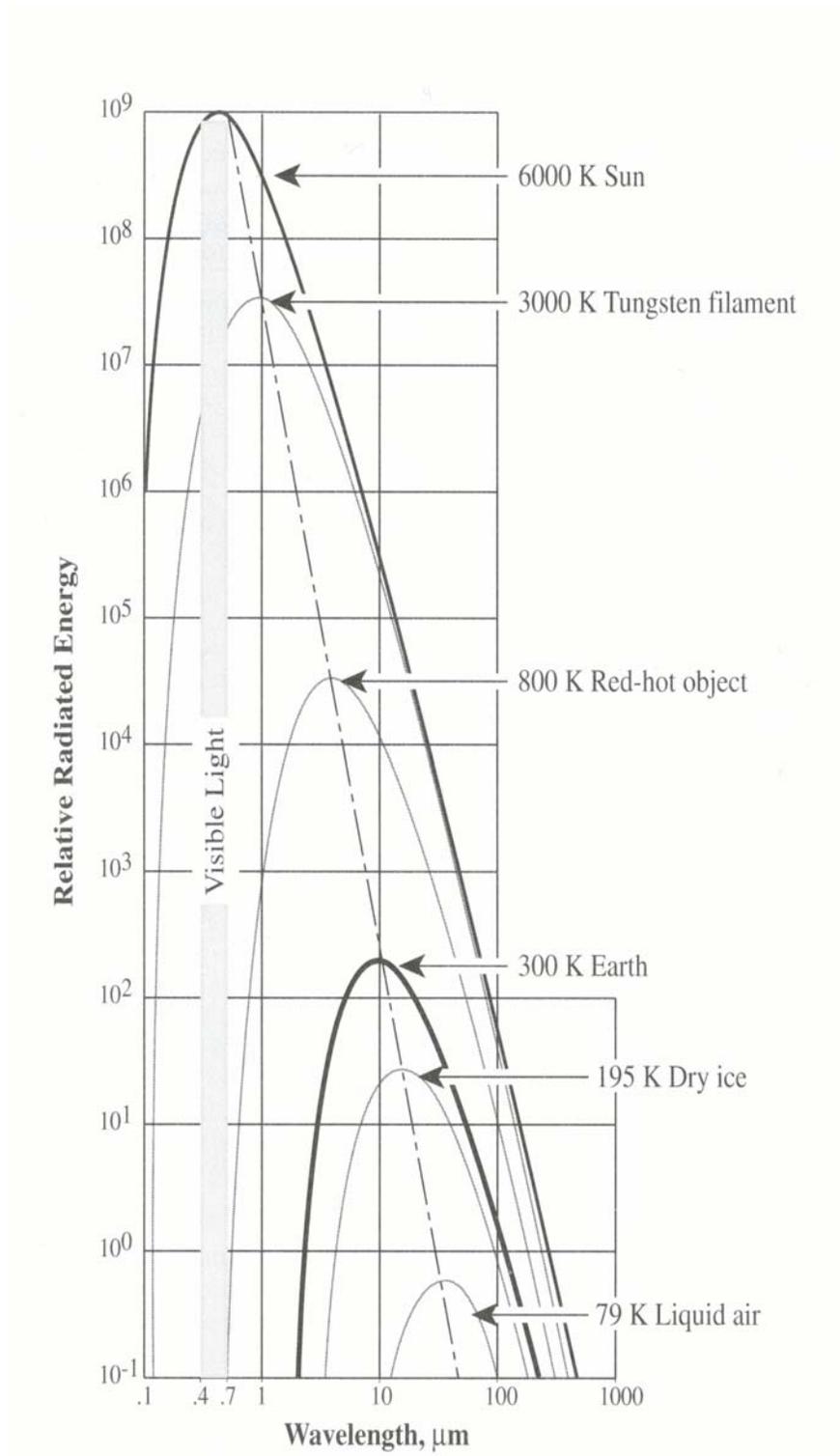


Figure 2.8: Energy from perfect blackbodies [Jensen, 2000]

2.2.3 The Effect of Propagation Media

In the absence of an atmosphere, the energy measured by the sensor will be simply a function of the level of energy from the Sun incident on the surface of the Earth as well as the reflectance properties of the surface itself. Assume the solar spectral irradiance at wavelength of λ is E_λ . Moreover, assume that the Sun is located at the zenith angle of θ . Then the spectral irradiance measured on the Earth's surface would be

$$E_G = E_\lambda \cos \theta. \quad (2.43)$$

Now assume that the Earth's surface has a reflectance ρ . Therefore the total amount of energy reflected by the surface to the upper hemisphere would be

$$L = \rho E_\lambda \cos \theta. \quad (2.44)$$

The sensors usually measure the spectral irradiance between two wavelengths, say λ_1 and λ_2 . Therefore the total spectral irradiance between these two wavelengths reflected by the surface and measured by the sensor is

$$I = \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} \rho E_\lambda \cos \theta d\lambda \quad (2.45)$$

where the division by π accounts for the upper hemisphere of solid angle. Knowing that in remote sensing the difference between the wavelengths is very small, with the assumption of $\Delta\lambda = \lambda_2 - \lambda_1$ one can approximate eq. (2.45) with

$$I = \rho E_{\Delta\lambda} \cos \theta \Delta\lambda / \pi \quad (2.46)$$

In the presence of the atmosphere the situation is drastically changed and therefore eq. (2.46) should be modified to take the atmosphere interaction on the EM propagation

into consideration. Generally speaking, the atmosphere has two effects on the EM propagation: 1) absorption, and 2) scattering. It is known that the absorption is a wavelength dependent phenomenon and sensors in remote sensing are designed so that they operate in wavelength intervals where there is no absorption by the atmosphere. This leaves scattering to have the most pronounced effect on EM propagation. Figure 2.9 demonstrates the behaviour of atmospheric absorption as a function of the electromagnetic spectrum.

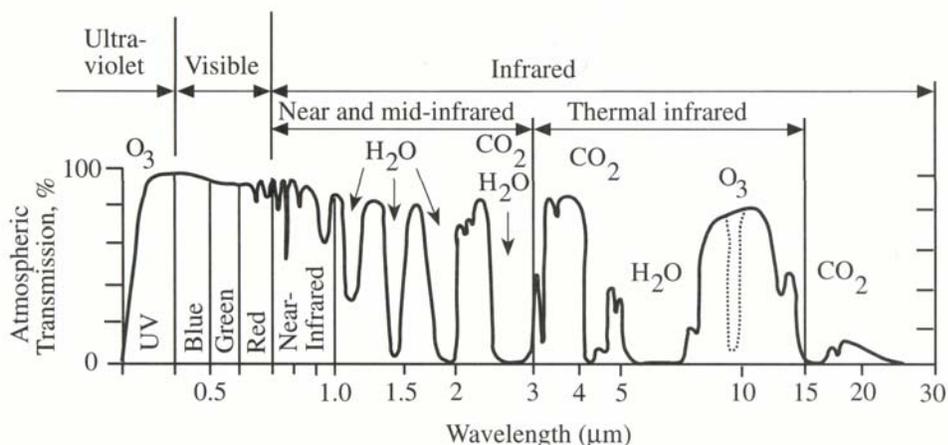


Figure 2.9: Atmospheric absorption for different parts of EM spectrum [Jensen, 1996]

Generally speaking there are two types of scattering. Rayleigh scattering is the one due to air molecules and is an inverse fourth power function of the wavelength used. The second scattering, aerosol or Mie scattering is caused by larger particles (of order of one-tenth to ten of the wavelengths used) such as smoke, haze and fumes. When the particles are not so large, the Mie scattering shows a wavelength dependency, although, the dependency it is not as strong as Rayleigh's. However, when the particles are quite large such as those in fogs, clouds, and dust, the dependency on the wavelength disappears.

Generally speaking, the shorter the wavelength is, the larger the Rayleigh scattering is. Furthermore, the amount of scattering depends on the length of the path through the atmosphere as the longer the path is, the more scattering there is. Figure 2.10 shows both Rayleigh and Mie scattering for a few selected wavelengths as functions of Sun elevation angle.

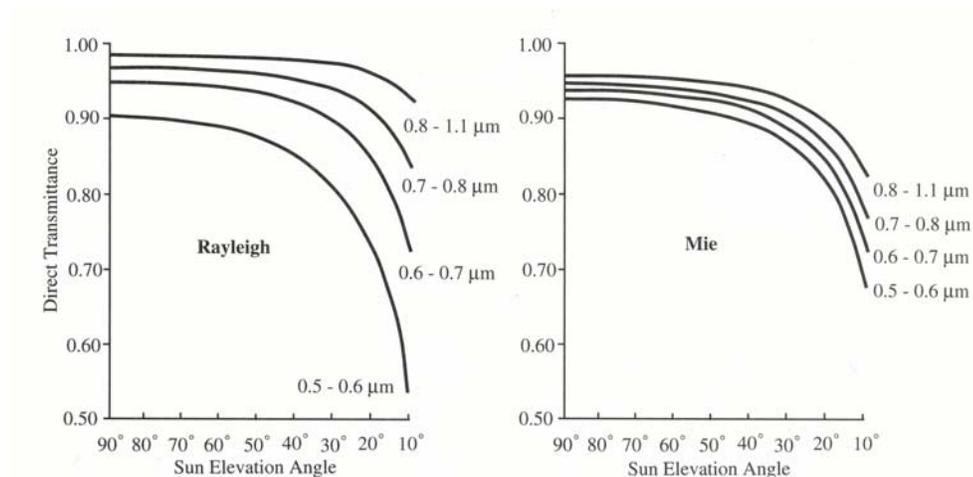


Figure 2.10: The behaviour of Rayleigh and Mie scattering for a few selected wavelengths [Jensen, 1996]

In the ideal case where there is no atmosphere there would be no absorption or scattering of EM power. However, due to the presence of the atmosphere some of the energy is reduced. Figure 2.11 summarizes the effect of the atmosphere on the EM radiation and irradiance. The actual amount of irradiance which reaches the ground in comparison to the case where there is no atmosphere is called transmittance T . Obviously, the transmittance depends on the length of the path through the atmosphere as the longer the path is, the more attenuation of EM power there is.

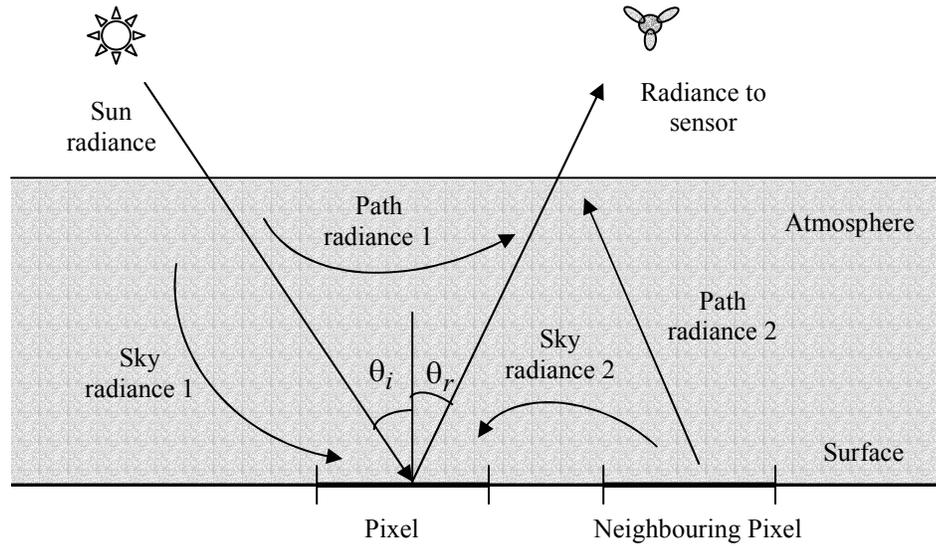


Figure 2.11: Atmosphere and the EM radiation [Richards, 1993]

Taking the effect of transmittance into consideration, one can rewrite eq. (2.43) as

$$E_G = T_{\theta_i} E_\lambda \cos \theta_i \quad (2.47)$$

where T_{θ_i} is the transmittance of the Sun irradiance. However, in addition to the direct solar irradiance, each pixel on the ground can receive some radiation from the scattered energy in the atmosphere or the sky irradiance. Therefore, the irradiance measured on the ground is:

$$E_G = T_{\theta_i} E_\lambda \cos \theta_i + E_s \quad (2.48)$$

where E_s is the amount of sky irradiance.

Analogously, the total amount of energy reflected by the surface to the upper hemisphere would be

$$L = \rho T_{\theta_r} \{T_{\theta_i} E_\lambda \cos \theta_i + E_s\} \quad (2.49)$$

where T_{θ_r} is the transmittance of the reflected energy. Taking the scattered or so called path irradiance into account and recalling the fact that sensors measure an interval of spectral irradiance rather than a single wavelength, therefore, the total spectral irradiance measured by the sensor is [Jensen, 1996; Richards, 1993]

$$I = \rho T_{\theta_r} \{T_{\theta_i} E_{\Delta\lambda} \cos\theta_i \Delta\lambda + E_s\} / \pi + E_p \quad (2.50)$$

where E_p is the path irradiance.

2.2.4 The Object's Surface

The parameter ρ or the reflectance used in the above equations models the interaction of the incident energy and the surface of the object. Actually, the process is more complicated and has a couple of fundamental characteristics which are important in remote sensing. The first important characteristic is that the incident radiation, the reflected radiation, and the normal to the surface all lie in the same plane. The second one is the fact that the angle of incidence and the angle of reflection are almost equal.

Reflecting surfaces can be classified into four different groups: 1) specular, 2) near-perfect specular, 3) near-perfect diffuse, and 4) diffuse. If the average surface height is several times smaller than the wavelength of the radiation hitting the surface, specular radiation happens. However, if the surface is not completely smooth and has few ripples, the reflection would be near-perfect specular. On the other hand, surfaces that have large height differences relative to the size of wavelength of the incident energy will reflect the energy in many directions. This type of reflection is called near-perfect diffuse reflection. A perfectly diffuse reflection happens on so called Lambertian surface where

the reflected energy has a constant power for any angle of reflectance (Figure 2.12).

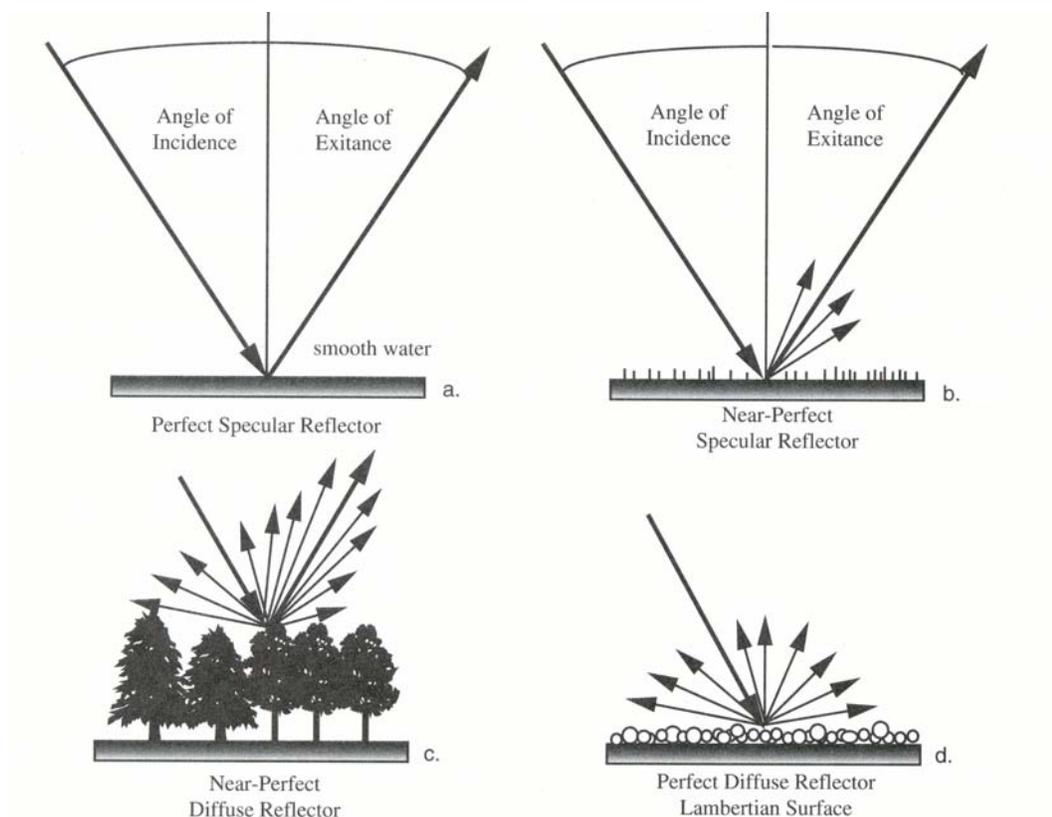


Figure 2.12: Specular, near-perfect specular, near-perfect diffuse and diffuse reflection [Jensen, 2000]

2.2.5 Bidirectional Reflectance Distribution Function (BRDF)

A true Lambertian surface where the surface reflects the same amount of radiant flux toward the sensor irrespective of the solar incidence and azimuth angles as well as the sensor viewing geometry is an ideal surface which is rarely the real case scenario. Bidirectional reflectance (Sun and/or sensor geometry relationship) measurements obtained over various vegetation and soil surfaces demonstrate that most terrestrial surfaces exhibit non-Lambertian (anisotropic) reflectance characteristics [Kimes, 1983;

Sandmeier and Itten, 1999]. Despite the fact that the non-Lambertian characteristics of the Earth's surface are known, their impact should still be further studied, as they are not well understood yet. In remote sensing, this is especially important in the realm of sensors with large fields of view or off-nadir viewing capabilities.

Fortunately, there are instruments which measure the reflectance distribution function of various surfaces. One of the most useful instruments is the goniometer designed by S. Sandmeier [Sandmeier, 1999]. It is a transportable instrument which basically consists of a computer-controlled GER 3700 spectroradiometer sensitive to the wavelength interval from 300-2,450 nm in 704 bands with a spectral resolution of 1.5 nm for the interval 300-1,050 nm and 8.4 nm for the interval 1,050-2,450 nm. It consists of three major parts: a zenith arc rail, an azimuth rail, and a motorized sled where the radiometer (sensor) is mounted.

The goniometer can be used to measure the bidirectional reflectance distribution function (BRDF) for a specific wavelength λ . This function is defined as the ratio of the radiance dL_r , reflected in one direction (φ_r, θ_r) to the corresponding Sun's incident irradiance dE_i from direction (φ_i, θ_i) . In mathematical notation one can write [Sandmeier, 1999; Sandmeier and Itten, 1999; Collins, 2002]

$$f_r(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda) = \frac{dL_r(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda)}{dE_i(\varphi_i, \theta_i)} \quad (2.51)$$

where f_r is the measured BRDF (Figure 2.13).

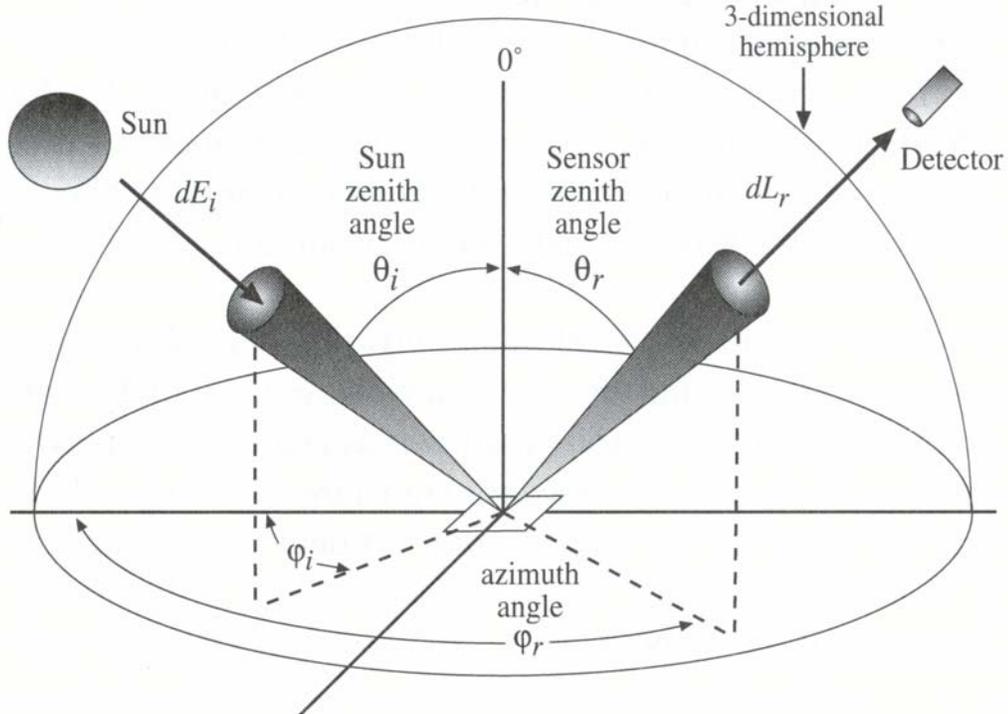


Figure 2.13: The concept and parameters of BRDF [Jensen, 2000]

In addition to BRDF, another parameter called bidirectional reflectance factor (BRF, also referred to as R) is computed. This parameter is the ratio of the radiance dL_r reflected from a surface in a specific direction to the radiance dL_{ref} reflected from a lossless Lambertian reference surface under same illumination geometry [Sandmeier and Itten, 1999; Collins, 2002]

$$R(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda) = \frac{dL_r(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda)}{dL_{ref}(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda)} R_{ref}(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda) \quad (2.52)$$

where R_{ref} is a calibration coefficient determined for the spectral reflectance surface.

Another parameter called anisotropy factor (ANIF) is calculated to analyse the spectral

variability in BRDF. This parameter is calculated by normalizing bidirectional reflectance data R to nadir reflectance R_0 [Sandmeier and Itten, 1999]

$$R(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda) = \frac{dL_r(\varphi_i, \theta_i, \varphi_r, \theta_r, \lambda)}{dL_{ref}(\varphi_i, \theta_i, \lambda)} . \quad (2.53)$$

Using the computed ANIF one can make the necessary radiometric adjustment to the brightness.

Despite the fact that numerous models are available for predicting the amount of energy radiation in specific wavelengths, there is still a lot to be done in this respect. Unfortunately, calibration of the models is usually very hard because of so much information needs to be known beforehand. This is still an active area of remote sensing research.

2.2.6 The Imaging Sensors

Radiometers collect image data either through scanning or using one- or two-dimensional CCD arrays. In any case, the radiometer sensor collects reflection from a well-defined patch on the ground which is called field of view. The measurement of field of view itself is a collection of smaller field of views which is called instantaneous field of view (IFOV). In fact the measurement associated with any image pixel is nothing else but the projection of IFOV on the ground (Figure 2.14).

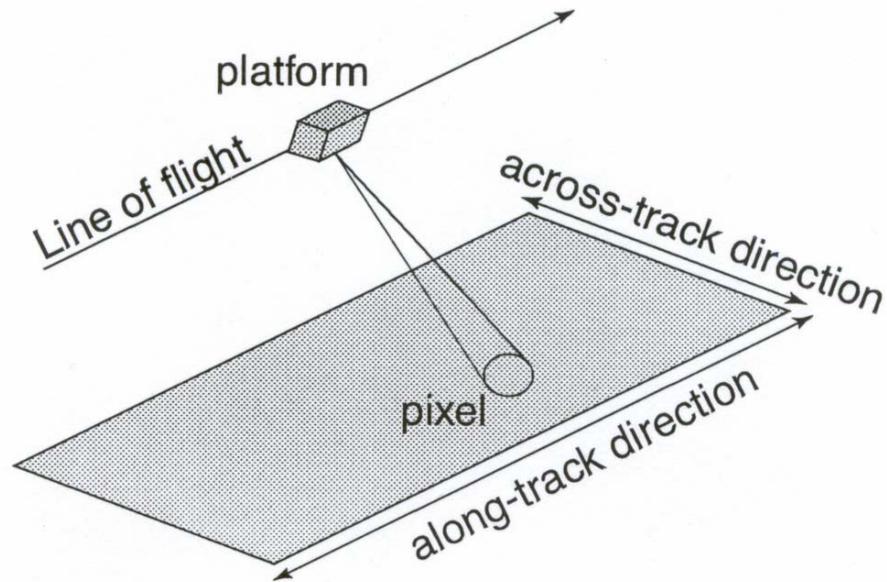


Figure 2.14: Field of view in a radiometer imaging system [Collins, 2002]

Most of the today's instruments that collect remotely sensed data are multispectral sensors. In other words, they collect the reflected, emitted or backscattered energy from an object or area of interest in multiple bands (or regions) of the EM spectrum. The remote sensor system first detects the received EM energy and records it as an analog electrical signal which is subsequently converted into digital values through an analog-to-digital (A-to-D) conversion. The A-to-D converter assigns each pixel of the image a brightness value in the range of 6 to 12 bits. This is generally called the quantization level of the sensor. Obviously, the greater the range of possible brightness values, the more precise would the radiance measurement.

According to the type of technology used, the multispectral radiometers are classified into three different groups: 1) multispectral imaging using discrete detectors and scanning

mirrors, 2) multispectral imaging using linear arrays, and 3) multispectral imaging using linear and area arrays. The first two categories are studied in Appendix G in more detail as the satellite sensors, images of which are used in this research work fall into these two groups.

CHAPTER THREE

SPATIAL ENHANCEMENT OF DTM GRIDS USING SFS

In this context DTM is a grid of elevations on the surface of the Earth. At present, depending on the area or region of interest, DTMs with different specifications and resolutions are available. However, no matter what the spatial resolutions of the available DTMs are, there is always a demand for a denser and more accurate elevation data set. Evidently, by default, one may think of field measurements as the first option for the purpose of accurate DTM densification. Unfortunately, in practice this is not always feasible. Obstacles like technical, political, financial as well as time considerations may raise serious questions about the possibility of any field measurements.

On the other hand, one may think of using interpolation techniques to densify the existing DTMs. Despite the fact that this solution is not only fast but also inexpensive in comparison to the field measurements, it has its own inherent problems. Selection of the interpolation model and its related parameters is not always a trivial task and often, one needs to spend a lot of time and effort to investigate them for every single and specific application. Moreover, the accuracy of interpolation techniques is always questionable especially when the surface under study is rough.

The objective of this research is to investigate the application of the SFS technique to a single (as opposed to stereo) satellite imagery (or aerial photograph) which has a higher resolution than the corresponding DTM to enhance the accuracy of the DTM

interpolation. In other words, the problem to investigate here is the enhancement of the DTM interpolation accuracy using the shading information in the corresponding satellite imagery or aerial photograph. The motivation for this research is the availability and ease of access to relatively inexpensive high resolution single remotely sensed imageries for almost all the world land areas.

The first section of this chapter reviews the basic assumptions and facts specific to the problem in hand. The second section discusses the choice of SFS solution method. The third section recapitulates the standard variational solution to the SFS problem. The fourth section explains the deficiencies of the solution in more detail. Last but not least, the fifth section deals with the modifications of the standard variational method and discusses the steps taken to enhance the SFS solution.

3.1 Basic Assumptions

As mentioned in the previous chapter, shading in images is due to the interaction of four principal factors: 1) the illumination (radiance or intensity and the distance and extension of the source), 2) the shape of the surface, 3) the reflecting characteristics of the material, and 4) the image projection. The main goal of SFS is to come up with a solution for the shape of the surface while assuming the other three factors to be somehow known or properly parameterized. Specifically, the very basic assumptions made for these three factors in this research are as follows.

The only source of illumination for remotely sensed images (or aerial photography) used in this investigation is the Sun. As the time and date of exposure are either recorded or known from the remote sensing system design, one can easily compute the Sun's

position vector at the time of exposure without any difficulty. Moreover, because the Sun is so far away from the Earth, it is considered as a distant point source. These facts about the source of illumination in remotely sensed images put this factor of shading into the category of completely known parameters.

As mentioned before, the reflecting property of a surface depends on its microstructure. However, what constitutes microstructure depends on one's applications. Obviously, the microstructure in remote sensing refers to an area on the ground which is equivalent to the size of the spatial resolution of the sensor. For the time being and in order to simplify the SFS problem, it is assumed that the Earth has a Lambertian surface with a known reflecting property. Evidently, these assumptions are too simplistic as the Earth's surface behaves in a much more complex way. Later on, necessary modifications will be applied to the SFS solution to take these simplifications into account.

It is also assumed that the remotely sensed imageries are in an orthographic projection system. Obviously, the specifications of sensors used in remote sensing vary widely. Some of them have imaging systems for which the orthographic projection assumption is true (or almost true) while for others, it is not. Those images with perspective projection can go through a well known process called orthorectification or registration by which the imageries and the corresponding DTMs are registered together and the displacements due to tilt and relief are removed and corrected for [Wolf, 1983].

3.2 The Choice of Solution Method

The very first step toward the goal of this study is to choose a general solution method for the SFS problem in hand. As discussed in the previous chapter, in addition to the

photometric stereo method, there are two main classes of solutions for the SFS problem, local and global methods. Application of the photometric stereo approach to the problem in hand is out of question because the very first assumption in this research is the availability of only single (not stereo) satellite imageries or aerial photographs which may come in different spectral or spatial resolutions.

On the other hand, in spite of their relative simplicity, local methods make a very strong assumption about the shape of the surface and use the first (and second) derivatives of image brightness to derive the shape parameters and approximate shape of the surface. Local sphericity assumption, reliance on brightness derivatives which can be very noisy, as well as lack of any mechanism to enforce global consistency (such as integrability) of the local slopes make the local methods inappropriate for the specific application of this study.

Global methods, however, are more complex, but generally provide more accurate shape. In the previous chapter, three different global methods, i.e., characteristic strip, variational, and linear, were briefly discussed. As with the local methods, the characteristic strip suffers from noise in the image. Furthermore, directional dependency of the characteristic strip solution and consequently accumulation of errors, in addition to the sequential process of the solution make this method neither efficient nor robust for the SFS problem especially within the realm of the objectives of this research. Although the linear methods give a non-iterative, closed-form solution to the SFS problem, they suffer from linear approximation of the reflectance map specially when the quadratic terms in reflectance map are significant.

However, variational methods seem to provide the most robust results among the other different SFS solutions if suitable constraints and parameters are selected in the formulation of the energy function. Due to this reason, the variational method is selected as the general solution method to the SFS problem in this investigation. Nevertheless, the formulation of the energy function needs to be manipulated and fine tuned in a way that it is more compatible with the facts and assumptions made for this specific SFS problem. Obviously, these modifications should guarantee the efficiency of the algorithm both from computational and robustness points of view. The following sections explain these modifications in detail.

Before getting into the details, it should be remembered that the variational SFS techniques estimate the shape of the objects in two steps. In the first step, the orientation of the surface is estimated over a grid, while, in the second step the height of the surface is estimated using the estimated orientations and a line integration process. The main focus of this chapter is on the first step where the orientations are estimated.

3.3 Recapitulation of the Standard Variational Method

Based on the simplified assumptions made in the previous chapter as well as earlier in this chapter, the reflectance map equation is

$$R(p,q) = \rho \mathbf{N}(x,y) \cdot \mathbf{L} = \rho \frac{l_1 p + l_2 q + l_3}{\sqrt{p^2 + q^2 + 1}} \quad (3.1)$$

where $R(p,q)$ is the reflectance map, $p(x,y)$ and $q(x,y)$ are the surface gradients at location (x,y) in x and y directions, respectively, ρ is the albedo, $\mathbf{N} = (p,q,1)^T$ is the surface

normal vector, and $\mathbf{L} = (l_1, l_2, l_3)^T$ is the light source position vector. In the context of this research work, all the parameters except the surface normal vector are assumed known. To solve for the unknown normal vector components, the image irradiance equation is written as

$$E(x, y) = R(p, q) + e(x, y) \quad \forall x, y \in \Omega \quad (3.2)$$

where $E(x, y)$ is the observed image brightness at location (x, y) in the image domain Ω and $e(x, y)$ is the corresponding error due to the noise in the observations as well as the incompleteness of the reflectance map model.

It was also discussed that based on the variational method, the SFS solution is obtained by minimization of the following cost function which is basically eq. 2.25 but in vector notation and with the assumption that both $E(x, y)$ and $R(p, q)$ are normalized with respect to the albedo factor

$$\iint_{\Omega} \left\{ (E(x, y) - \mathbf{N} \cdot \mathbf{L})^2 + \lambda \left\{ \|\mathbf{N}_x\|^2 + \|\mathbf{N}_y\|^2 \right\} \right\} d\Omega \quad (3.3)$$

where the integration domain Ω corresponds to the entire (or some part of the) image, the subscripts of x and y derivatives along x and y directions respectively, and λ is a scalar that weighs the relative importance of the smoothness (regularization) term. Consequently, based on the calculus of variations, the following second-order partial differential equation (Euler equation) is obtained (eq. 2.26)

$$\lambda \Delta \begin{Bmatrix} p \\ q \end{Bmatrix} = -(E(x, y) - \mathbf{N} \cdot \mathbf{L}) \begin{Bmatrix} R_p \\ R_q \end{Bmatrix} \quad (3.4)$$

where Δ is the Laplacian operator and (p, q) are the first two components of the surface

normal vector. Using a discrete three-point approximation of the Laplacian operator given in eq. (2.16), the following set of iterative update equation is obtained (eq. 2.27 in vector notation):

$$\mathbf{N}_{i,j}^{n+1} = \bar{\mathbf{N}}_{i,j}^n + \frac{\varepsilon^2}{4\lambda} (\mathbf{E}_{i,j} - \mathbf{N}_{i,j}^n \cdot \mathbf{L}) \mathbf{R}'_{i,j}^n \quad (3.5)$$

where (i,j) is the cell location, ε is the cell size, the superscript n is the iteration number, \mathbf{R}' is the gradient vector of the reflectance model with respect to p and q as its components, and

$$\bar{\mathbf{N}}_{i,j}^n = \frac{1}{4} (\mathbf{N}_{i+1,j}^n + \mathbf{N}_{i-1,j}^n + \mathbf{N}_{i,j+1}^n + \mathbf{N}_{i,j-1}^n). \quad (3.6)$$

The update eq. 3.5 for the surface normal vector consists of two terms. The first term, which comes from the smoothness constraint, is the local mean average of the unit surface normal vectors in the vicinity of the cell (i,j) . The second term is called data closeness and comes from the brightness error constraint.

3.4 Deficiencies of the Standard Variational Method

The principal obstacle of the variational algorithm formulated in the form of eq. 3.3 and similar approaches is the tendency to over smooth the recovered surface. Since the smoothness constraint is formulated in terms of the directional derivatives of the surface normal vector, it is obviously minimized by a flat surface. Thus, any conflict between the observed brightness $E(x,y)$ and the reconstructed one, i.e., the reflectance map model $R(p,q)$, leads to a strongly smoothed surface and the loss of fine detail.

In other words, the quadratic regularizing term absorbs sudden changes in the surface

normal direction across a surface. Applying the smoothness constraint homogeneously everywhere leads to over smoothing the surface at discontinuities. The main consequence is to blur high-curvature surface features which is undesirable when highly structured surfaces are being analysed. Moreover, the implied global smoothness assumption is at odds as most real objects are only smooth in the piecewise sense [Marr, 1982].

Evidently, the amount of smoothness depends on the value of regularization factor λ . Unfortunately selecting a proper value for this factor is not a trivial task. A small value for the regularization factor results in numerical instability while a large value makes the smoothness term dominate the energy function and therefore more fine details would be lost. In practice, in order to be confident that the scheme will remain numerically stable for all expected input images, a conservative value is often selected that intensifies the over smoothing problem.

Another deficiency of the standard variational method is about the update equation (3.5). As mentioned in the previous chapter, the individual surface normal vectors satisfying the irradiance equation should assume a direction on the ambiguity cone. However, based on the update equation (3.5), the surface normal vector is free to move away from the cone under the action of the smoothness constraint which obviously creates numerical instability.

3.5 Modifications to the Standard Variational Algorithm

To deal with the over smoothing problem of the variational algorithm, Horn [1990] discussed the problem of selecting a proper value for the regularization factor. He

explained that the smoothness penalty term is needed only for damping out numerical instabilities especially when the initial condition is far away from the solution. This is particularly important with real data, where one cannot expect to find an exact solution. However, it is also discussed that when the final solution is approached, one can reduce the value of the regularization factor so that the penalty term drops out. This reduces the model dominance problem by reducing the influence of the smoothness constraint.

On the other hand, Gültekin and Gökmen [1996] suggested an adaptive algorithm for dealing with the over smoothing problem in the standard variational method. In this method the regularization factor is a function of spatial coordinates. At each iteration, the space varying regularization parameter at location (x,y) is determined by the following function

$$\lambda_{\text{new}}(x,y) = \begin{cases} G(x,y,\lambda_{\text{old}}) & \text{if } c(x,y) > 0 \text{ and } \lambda_{\text{old}} > \lambda_0 \\ \lambda_{\text{old}}(x,y) & \text{otherwise} \end{cases} \quad (3.7)$$

where

$$G(x,y,\lambda_{\text{old}}) = (1 - \exp^{-c(x,y)/\mu})\lambda_0 + \exp^{-c(x,y)/\mu} \lambda_{\text{old}}(x,y), \quad (3.8)$$

$c(x,y)$ is the signal control parameter which is defined as $c(x,y) = |E(x,y) - R(p,q)|$, μ is a constant which controls the rate of exponential decrease, and λ_0 is a pre-selected minimum value that $\lambda(x,y)$ may have. As it is seen, $G(x,y,\lambda_{\text{old}})$ is an exponentially decreasing function so that the regularization parameter is only allowed to decrease during iterations.

The adaptive regularization factor scheme, discussed above, is based on the differences between the observed and reconstructed image brightness. This scheme is not

so effective particularly if the image is noisy or is contaminated with outliers. Moreover, this method has difficulty when there is a discontinuity in the surface under investigation. Obviously, discontinuities in the surface violate the smoothness constraint, therefore a more powerful discontinuity-adaptive mechanism is needed to avoid over smoothing of the surface while detecting and rejecting the outliers [Li, 1991; Li, 1995].

3.5.1 Adaptive Regularizer Kernels from Robust Statistics

Robust statistics have been used in many applications such as image restoration, smoothing and segmentation [Kashyap and Eom, 1988; Jolion et al, 1991; Meer et al, 1991], and surface and shape fitting [Besl et al, 1988; Stein and Werman, 1992], where outliers are an issue. The goal of this subsection is to illustrate how the smoothness constraint can be controlled more effectively using error estimator functions suggested by robust statistics. In order to develop the background on how to use calculus of variations along with this concept, the following section provides a short review on robust statistics.

According to Huber [1981] the word “robust” is loaded with many, sometimes inconsistent, connotations. However, in this context, robustness signifies insensitivity to small deviations from the assumptions. The concept of modern robustness began in 1960 with the papers by J.W. Tukey on sampling from contaminated distributions and F.J. Anscombe on rejection of outliers. Tukey’s paper drew attention to the dramatic effects of seemingly negligible deviations from the model. Anscombe introduced a seminal insurance idea: sacrifice some performance of the model in order to insure against ill effects caused by deviations from it. Most of the basic ideas, concepts, and methods of robustness were invented in quick succession during the subsequent years and were in

place by about 1970.

Several measures of robustness are used in the literature. Most common is the influence function or influence curve which describes the change in an estimate caused by insertion of a single data outlier as a function of the distance of the data from the uncorrupted estimate [Hampel et al, 1986; Huber, 1981]. As an example, the influence function of the least squares is simply proportional to the magnitude of the residual. To achieve robustness, the influence function should tend to 0 with increasing distance from the estimate.

The influence function $\psi_w(v)$ is defined for the residual v while w is the parameter which controls the width of the function. In spite of the fact that the influence function is widely used in robust estimation literature, it is the form of its error estimate function or so called error kernel $\Psi_w(v)$ which is of primary importance in smoothness regularization. Formally, the influence function is related to the error estimate function by [Huber, 1981]

$$\psi_w(v) = \frac{1}{v} \frac{\partial \Psi_w(v)}{\partial v}. \quad (3.9)$$

Using the error kernel $\Psi_w(v)$ with the magnitudes of the surface curvature in both x and y directions as the residual errors, one can reformulate the SFS cost function as

$$\iint_{\Omega} \left\{ (E(x, y) - \mathbf{N} \cdot \mathbf{L})^2 + \lambda \left(\Psi_w(\|\mathbf{N}_x\|) + \Psi_w(\|\mathbf{N}_y\|) \right) \right\} d\Omega. \quad (3.10)$$

Applying calculus of variations to this cost function yields the following Euler function [Appendix D]

$$\begin{aligned}
& (E(x, y) - \mathbf{N} \cdot \mathbf{L}) \mathbf{R}' + \\
& \lambda \left(\frac{\partial}{\partial x} \left(\|\mathbf{N}_x\|^{-1} \Psi'_w \left(\|\mathbf{N}_x\| \right) \mathbf{N}_x \right) + \frac{\partial}{\partial y} \left(\|\mathbf{N}_y\|^{-1} \Psi'_w \left(\|\mathbf{N}_y\| \right) \mathbf{N}_y \right) \right) = \mathbf{0}
\end{aligned} \tag{3.11}$$

where Ψ' is the derivative of the error kernel with respect to the residuals. As a result, the fixed-point iterative equation for updating the components of the surface normal vector is [Appendix D]

$$\mathbf{N}_{i,j}^{n+1} = \frac{1}{2\alpha_{i,j}^n} \left(\bar{\mathbf{M}}_{i,j}^n + \frac{2\varepsilon^2}{\lambda} (E_{i,j} - \mathbf{N}_{i,j}^n \cdot \mathbf{L}) \mathbf{R}'_{i,j}^n \right) \tag{3.12}$$

where

$$\alpha_{i,j}^n = \Psi'_w \left(\|\mathbf{N}_{x,i,j}^n\| \right) \|\mathbf{N}_{x,i,j}^n\|^{-1} + \Psi'_w \left(\|\mathbf{N}_{y,i,j}^n\| \right) \|\mathbf{N}_{y,i,j}^n\|^{-1} \tag{3.13}$$

and

$$\begin{aligned}
\bar{\mathbf{M}}_{i,j}^n = & \left[\|\mathbf{N}_x^n\|^{-1} \Psi'_w \left(\|\mathbf{N}_x^n\| \right) (\mathbf{N}_{i+1,j}^n + \mathbf{N}_{i-1,j}^n) + \varepsilon^2 \left[\|\mathbf{N}_x^n\|^{-2} \Psi''_w \left(\|\mathbf{N}_x^n\| \right) - \right. \right. \\
& \left. \left. \|\mathbf{N}_x^n\|^{-3} \Psi'_w \left(\|\mathbf{N}_x^n\| \right) \right] (\mathbf{N}_x^n \mathbf{N}_{xx}^n) \mathbf{N}_x^n \right] + \\
& \left[\|\mathbf{N}_y^n\|^{-1} \Psi'_w \left(\|\mathbf{N}_y^n\| \right) (\mathbf{N}_{i,j+1}^n + \mathbf{N}_{i,j-1}^n) + \varepsilon^2 \left[\|\mathbf{N}_y^n\|^{-2} \Psi''_w \left(\|\mathbf{N}_y^n\| \right) - \right. \right. \\
& \left. \left. \|\mathbf{N}_y^n\|^{-3} \Psi'_w \left(\|\mathbf{N}_y^n\| \right) \right] (\mathbf{N}_y^n \mathbf{N}_{yy}^n) \mathbf{N}_y^n \right] .
\end{aligned} \tag{3.14}$$

The normal vector update (eq. 3.12) is entirely general as many error kernel $\Psi_w(v)$ can be inserted into this equation. Therefore, the performance of the update equation in view of a solution to the SFS problem is critically dependent on the choice of error estimate function.

Before getting into the details of how to choose the right error estimate function for

the SFS updating scheme, it would be instructive to compare the general SFS solution (eq. 3.5) with the one obtained by applying the robust statistics (eq. 3.12). As it is seen, apart from the factor which can be considered as a normalization factor, both of the equations consist of two parts. The first terms come from the smoothness constraint and have some similarities. The second terms come from the image brightness error which tends to zero as the normals are updated towards a solution. As the brightness error is insensitive to rotation of the recovered normal around an axis defined by the light source direction, an infinite number of potential solutions exist. Since the primary goal is the minimization of this term over the image, the smoothness constraint helps to select a particular solution. In fact this particular solution is nothing else but the smoothest solution which minimizes the brightness error.

Comparing the smoothness constraint of (eq. 3.6) with (eq. 3.14), one can see obvious similarities. However, the behaviour of the latter is complicated by weighting terms as well as the presence of higher order terms.

3.5.2 The Error Kernel

As mentioned above, the update equation of the surface normal vector obtained in (eq. 3.12) is entirely general. Many error kernels can be inserted into the update equation and yield a SFS solution. However, the performance is critically determined by the choice of the kernel.

In the robust estimation literature, influence functions are classified based on the behaviour of the derivative of their error kernel functions [Li, 1995]. Generally speaking, there are three classes of influence functions: 1) re-descending function where the

derivative of the error kernel asymptotically approaches zero, 2) sigmoidal function where the derivative of the error kernel becomes asymptotically constant, and finally, 3) quadratic prior where the derivative of the error kernel is asymptotically monotonically increasing [Worthington and Hancock, 1999].

As examples, Table 3.1 shows three re-descending influence functions (first row) with their corresponding error kernels (second row) and the derivative of the error kernels (third row). Figure 3.1 shows the corresponding graphs for special case of $w = 1$. Analogously, as examples, Table 3.2 shows three sigmoidal influence functions with their corresponding error kernels and derivatives of their error kernels. Figure 3.2 shows their corresponding graphs.

Table 3.1: Examples of re-descending influence functions

	Tukey' bi-weight (1)	Li's Adaptive (2)	Li's Adaptive (3)
$\psi_w(v)$	$\begin{cases} 1 - ((v/w)^2) & \text{if } v < w \\ 0 & \text{otherwise} \end{cases}$	$\exp(-v^2/w)$	$(1 + v^2/w)^{-2}$
$\Psi_w(v)$	$\begin{cases} w(1 - (1 - (v/w)^2)^3) & \text{if } v < w \\ w & \text{otherwise} \end{cases}$	$-w \exp(-v^2/w)$	$-w(1 + v^2/w)^{-1}$
$\Psi'_w(v)$	$\begin{cases} (1 - ((v/w)^2))^3 & \text{if } v < w \\ 0 & \text{otherwise} \end{cases}$	$v \exp(-v^2/w)$	$v(1 + v^2/w)^{-1}$

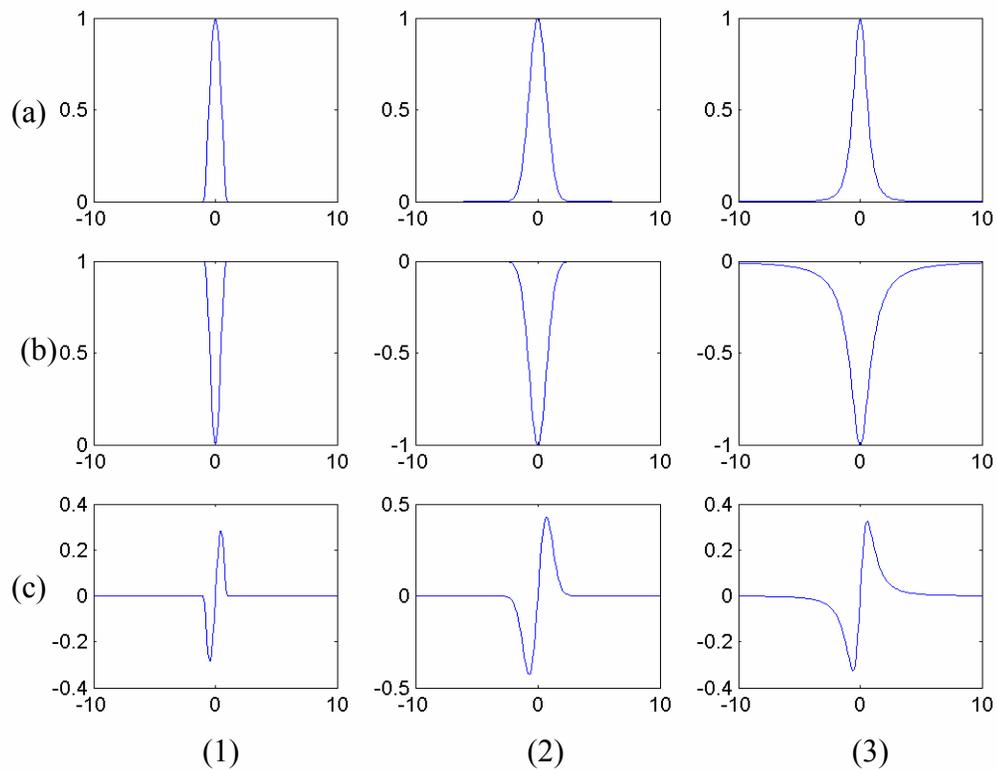
**Figure 3.1: Three re-descending influence functions (row (a)), error kernels (row (b)), and derivative of the error kernels (row (c)) all corresponding to Table 3.1**

Table 3.2: Examples of sigmoidal influence functions

	Hyperbolic Tangent (1)	Huber's (2)	Li's Adaptive (3)
$\psi_w(v)$	$1/v \tanh(\pi v/w)$	1 if $ v < w$ $w/ v $ otherwise	$(1 + v /w)^{-1}$
$\Psi_w(v)$	$w/\pi \log \cosh(\pi v/w)$	v^2 if $ v < w$ $2w v - w^2$ otherwise	$w v - w^2 \log\left(1 + \frac{ v }{w}\right)$
$\Psi'_w(v)$	$\tanh(\pi v/w)$	$2v$ if $ v < w$ $2w \text{Sgn}(v)$ otherwise	$wv\left(1 + \frac{ v }{w}\right)^{-1}$

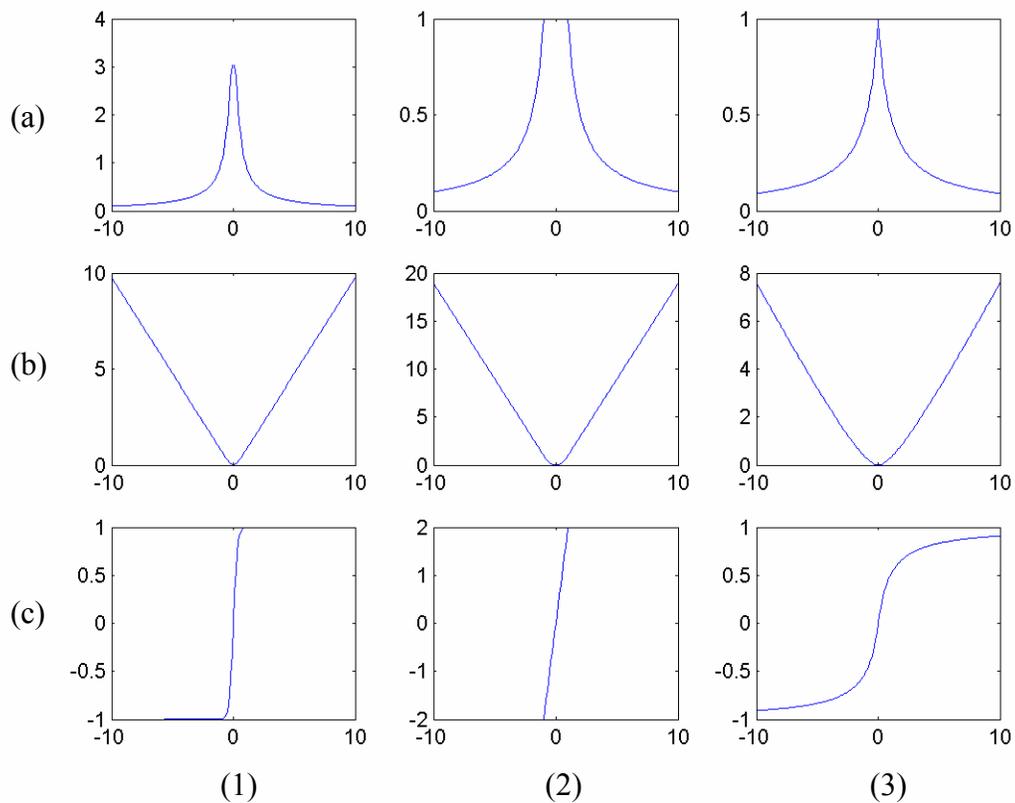
**Figure 3.2: Three sigmoidal influence functions (row (a)), error kernels (row (b)), and derivative of the error kernels (row (c)) all corresponding to Table 3.2**

Table 3.3 shows a quadratic influence function in addition to its error function and derivative of its error function. This quadratic influence function is used in the standard variational solution of SFS problem. Figure 3.3 shows the graphs corresponding to the influence function, error function and the derivative of the error function.

Table 3.3: Quadratic influence function

	Quadratic
$\Psi_w(v)$	1
$\Psi_w(v)$	v^2
$\Psi'_w(v)$	$2v$

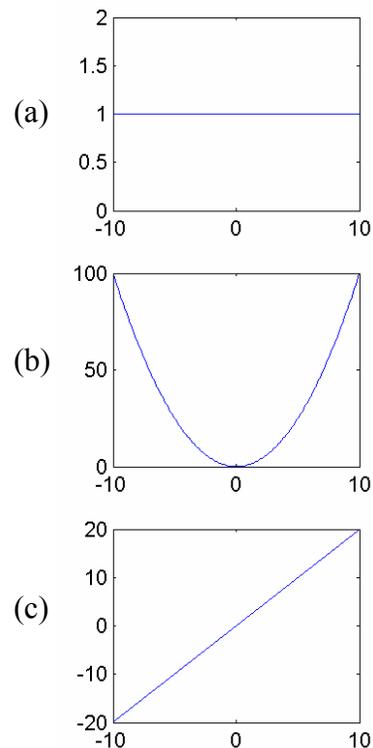


Figure 3.3: Quadratic influence function (row (a)), error kernel (row (b)), and derivative of the error kernel (row (c)) all corresponding to Table 3.3

It is instructive to investigate the effect of using re-descending and sigmoidal influence functions on the SFS solution. Among re-descending functions, Tukey's bi-weight can't be used in this application because of the piecewise nature of the function.

However, Li's adaptive functions are both continuous and as an example, the following influence function is used here

$$\psi_w(v) = \exp\left(-v^2/w\right). \quad (3.15)$$

Using this function one can evaluate that part of update equation which comes from the smoothness constraint (eq. 3.14) as (assuming $\varepsilon = 1$)

$$\begin{aligned} \bar{\mathbf{M}}_{i,j}^n = & \left[(\mathbf{N}_{i+1,j}^n + \mathbf{N}_{i-1,j}^n) + \frac{2}{w} (\mathbf{N}_x^n \mathbf{N}_{xx}^n) \mathbf{N}_x^n \right] \exp\left(-\frac{1}{w} \|\mathbf{N}_x^n\|^2\right) + \\ & \left[(\mathbf{N}_{i,j+1}^n + \mathbf{N}_{i,j-1}^n) + \frac{2}{w} (\mathbf{N}_y^n \mathbf{N}_{yy}^n) \mathbf{N}_y^n \right] \exp\left(-\frac{1}{w} \|\mathbf{N}_y^n\|^2\right). \end{aligned} \quad (3.16)$$

Comparing eq. (3.16) with eq. (3.6) shows two major differences. The first difference comes from the fact that in the latter the neighbouring normal vectors have been weighted with the exponential function. The second difference is the contribution of the second derivatives in (3.16). Interestingly, when the magnitudes of the surface curvature in x and y directions are small, (3.16) will be similar to (3.6). However, when they are large, then the exponential terms appearing in (3.16) reduce the effect of smoothness constraint.

Among the sigmoidal influence functions the hyperbolic tangent is used here as an example. Using the influence function $\psi_w(v) = \frac{1}{v} \tanh\left(\frac{\pi v}{w}\right)$, one can come up with the following update equation for that part which comes from smoothness constraint (assuming $\varepsilon = 1$) [Worthington and Hancock, 1999]

$$\begin{aligned}
\bar{\mathbf{M}}_{i,j}^n = & \left[\|\mathbf{N}_x^n\|^{-1} \tanh\left(\frac{\pi}{w} \|\mathbf{N}_x^n\|\right) (\mathbf{N}_{i+1,j}^n + \mathbf{N}_{i-1,j}^n) + \left[\frac{\pi}{w} \|\mathbf{N}_x^n\|^{-2} \operatorname{sech}^2\left(\frac{\pi}{w} \|\mathbf{N}_x^n\|\right) - \right. \right. \\
& \left. \left. \|\mathbf{N}_x^n\|^{-3} \tanh\left(\frac{\pi}{w} \|\mathbf{N}_x^n\|\right) \right] (\mathbf{N}_x^n \mathbf{N}_{xx}^n) \mathbf{N}_x^n \right] + \\
& \left[\|\mathbf{N}_y^n\|^{-1} \tanh\left(\frac{\pi}{w} \|\mathbf{N}_y^n\|\right) (\mathbf{N}_{i,j+1}^n + \mathbf{N}_{i,j-1}^n) + \left[\|\mathbf{N}_y^n\|^{-2} \operatorname{sech}^2\left(\frac{\pi}{w} \|\mathbf{N}_y^n\|\right) - \right. \right. \\
& \left. \left. \|\mathbf{N}_y^n\|^{-3} \tanh\left(\frac{\pi}{w} \|\mathbf{N}_y^n\|\right) \right] (\mathbf{N}_y^n \mathbf{N}_{yy}^n) \mathbf{N}_y^n \right] . \tag{3.17}
\end{aligned}$$

Comparing eq. (3.6) with eq. (3.17) shows that in the latter, the neighbouring normal vectors are weighted by a function of the form $\frac{1}{v} \tanh\left(\frac{\pi v}{w}\right)$. Obviously, considering the shape of this function, one realizes that the effects of neighbouring normal vectors are more pronounced when the magnitude of curvature is small. The remaining contribution to the smoothness process is of the form $v^2\left(1/\|v\|^2 \operatorname{sech}^2 v - 1/\|v\|^3 \tanh v\right)$ which vanishes at the origin and tends toward zero for large values of v , and pronounced only at intermediate error values.

Generally speaking, influence functions with monotonically increasing error kernel derivatives tend to over smooth genuine discontinuities in image brightness, as such discontinuities lead to large values of smoothness errors. Conversely, re-descending influence functions do not penalise sharp changes in surface orientation. Although this may lead to improved treatment of discontinuities, but it would be at the expense of increased noise sensitivity and numerical instability. Sigmoidal influence functions represent a compromise between the dual aims of recovering discontinuities and rejecting noise artifacts.

3.5.3 The Width Parameter of the Error Kernel

The error kernel used in robust statistics has one variable v and one parameter w . As mentioned before, v is the residual error and w is the width of the error kernel. Moreover, it was shown above that in the SFS context the curvature along x and y directions are used as the residual error parameter of the error kernel. However, nothing was mentioned about the width of the kernel.

Obviously, when the curvature is used as the residual error parameter, the width of the function should somehow reveal the dispersion or consistency of the curvature. Based on the differential geometry literature, the differential structure of a surface can be analyzed using an Hessian matrix which is given by

$$H = \begin{pmatrix} (\mathbf{N}_x)_1 & (\mathbf{N}_x)_2 \\ (\mathbf{N}_y)_1 & (\mathbf{N}_y)_2 \end{pmatrix} \quad (3.18)$$

where H is the Hessian matrix, \mathbf{N} is the surface normal vector, the subscripts x and y are the derivative directions and the indices 1 and 2 denote the first and second components of the parenthesized vectors respectively. The eigenvalues (h_1, h_2) of the Hessian matrix define the principal curvatures of the surface by which the mean $((h_1 + h_2)/2)$ and Gaussian $(h_1 h_2)$ curvatures of the surface are computed.

There are two complications about using the mean and Gaussian curvatures to classify a surface. The first problem, which is quite obvious, arises from the fact that one should work with two parameters to study the shape of a surface. The second problem is that one has to set four different thresholds to classify a surface.

To overcome these two problems, Koenderink and van Doorn [1992] defined the

shape index. The shape index is a continuous measure which encodes the same curvature class information as mean and Gaussian curvatures but in angular form. It is defined as [Koenderink and van Doorn, 1992]

$$\Phi = \frac{2}{\pi} \arctan \frac{(\mathbf{N}_x)_1 + (\mathbf{N}_y)_2}{\sqrt{((\mathbf{N}_x)_1 - (\mathbf{N}_y)_2)^2 + 4(\mathbf{N}_x)_2 (\mathbf{N}_y)_1}}. \quad (3.19)$$

The shape index can have values between -1 and 1 which correspond to the convex and concave spherical caps, respectively. Obviously, the surfaces with other combination of curvatures fall in between these shape index values.

Having defined the shape index, one can use it to measure the consistency or inconsistency of the curvature around a particular point. This measure in turn can be used to define the width of the error kernel. The variance dependent width can be defined as [Worthington and Hancock, 2001]

$$w = w_0 \exp \left(- \left(\frac{1}{M} \sum_{i,j} \frac{(\Phi_{i,j} - \Phi_c)^2}{\Phi_d^2} \right)^{1/2} \right) \quad (3.20)$$

where w is the width of the error kernel, w_0 is a reference width, M is the number of points used to calculate the width, $\Phi_{i,j}$ is the shape index at cell location (i,j) , Φ_c is the shape index with which the inconsistency is measured, and Φ_d is the shape index difference between two adjacent classes.

Eq. 3.20 shows that if the shape index varies significantly over a neighborhood, the value of width will be small which in return results in a bigger effect of smoothing. On the other hand, if the shape indexes around the point of interest are more or less similar,

the width will be a large value. This means a smaller value for the error kernel and therefore less smoothing.

The measure introduced by (eq. 3.20) is based only on the consistency of the curvature of the object over a neighborhood. No information about the consistency of the measured and the reconstructed brightness is used to decide about the width of the error kernel. To do so, one can define the so-called gradient consistency measure by either

$$w = w_0 \exp \left[- \left(\frac{1}{M} \sum_{i,j} \left(\left(E_{x_{i,j}} - \mathbf{N}_{x_{i,j}} \cdot \mathbf{L} \right)^2 + \left(E_{y_{i,j}} - \mathbf{N}_{y_{i,j}} \cdot \mathbf{L} \right)^2 \right) \right)^{1/2} \right] \quad (3.21)$$

or

$$w = w_0 \exp \left[- \left(\frac{1}{M} \sum_{i,j} \left(\left(E_{xx_{i,j}} - \mathbf{N}_{xx_{i,j}} \cdot \mathbf{L} \right)^2 + \left(E_{yy_{i,j}} - \mathbf{N}_{yy_{i,j}} \cdot \mathbf{L} \right)^2 \right) \right)^{1/2} \right] \quad (3.22)$$

where w is the width parameter of the error kernel, w_0 is the reference width, M is the number of points used to calculate the width, E is the measured brightness normalized with the albedo factor, subscripts x and y are the derivative directions, indices i and j are the cell location, and \mathbf{N} and \mathbf{L} are the surface normal and light source direction vectors, respectively.

Similar to (eq. 3.20), if there is no gradient consistency over a neighborhood, the value of the width parameter resulting from (eqs. 3.21) or (3.22) will be small which in return results in a bigger effect of smoothing. On the other hand, if there is a high gradient consistency around the point of interest the width will be a large value. This means a smaller value for the error kernel and therefore less smoothing.

3.5.4 Geometrical Application of Brightness Error Constraint

As mentioned in Chapter 2, from a geometrical point of view the irradiance equation (2.5) defines an ambiguity cone about the light source direction. In other words, the individual surface normals for which the irradiance equation is valid must assume directions that fall on this cone (Figure 3.4). However, in each iteration the updated normal vectors can assume any direction and are free to move away from the cone under the action of the smoothness constraint.

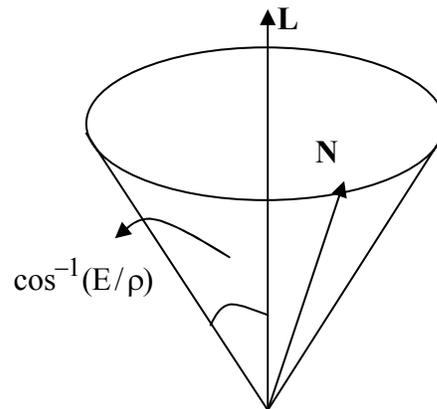


Figure 3.4: The ambiguity cone

The geometric constraint mentioned above can be imposed on the updated normal vector at each iteration. This is done by mapping the updated normals back to the closest normal lying on the cone. From a mathematical point of view, the resulting update equation can be written as

$$\mathbf{N}_{i,j}^{n+1} = R(\theta) \bar{\mathbf{M}}_{i,j}^n \quad (3.23)$$

where $\mathbf{N}_{i,j}^{n+1}$ is the updated normal vector at cell (i,j) and iteration n+1, $R(\theta)$ is the rotation matrix with rotation angle of θ (Figure 3.5) which maps the updated normal to the closest normal lying on the ambiguity cone, and $\bar{\mathbf{M}}_{i,j}^n$ is the intermediate surface normal at cell (i,j) and iteration n which comes out of the smoothness constraint.

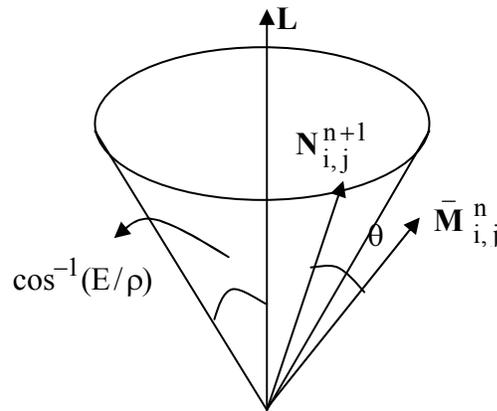


Figure 3.5: The surface normal before and after mapping on the ambiguity cone

The rotation is done around an axis perpendicular to the intermediate surface normal vector $\bar{\mathbf{M}}_{i,j}^n$ and the light source vector \mathbf{L} . This axis is found by taking the cross-product of the intermediate update vector with the light source direction

$$(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3)^T = \bar{\mathbf{M}}_{i,j}^n \times \mathbf{L}. \quad (3.24)$$

The angle of rotation is the difference between the angle subtended by the intermediate update and the light source, and the apex angle of the cone of ambiguity, i.e., (with the assumption of normalized vectors)

$$\theta = -\cos^{-1}\left(\bar{\mathbf{M}}_{i,j}^n \cdot \mathbf{L}\right) + \cos^{-1}(E_{i,j}/\rho). \quad (3.25)$$

Finally, the rotation matrix is given by [Bronshtein and Semendyayev, 1998]

$$\mathbf{R}(\theta) = \begin{pmatrix} c + g_1^2 c' & g_3 s + g_1 g_2 c' & -g_2 s + g_1 g_3 c' \\ -g_3 s + g_1 g_2 c' & c + g_2^2 c' & g_1 s + g_2 g_3 c' \\ g_2 s + g_1 g_3 c' & -g_1 s + g_2 g_3 c' & c + g_3^2 c' \end{pmatrix} \quad (3.26)$$

where $c = \cos\theta$, $c' = (1 - \cos\theta)$, and $s = \sin\theta$.

3.5.5 Data Fusion and Albedo Factor

Among different assumptions made at the beginning of this chapter to simplify the SFS problem the only one related to the reflectance property of the surface or the albedo factor is not completely valid. Up to now it has been assumed that the Earth's surface is a Lambertian one with a constant and known albedo factor. But, it is completely clear that the real Earth's surface behaves in a much more complex manner. It is neither perfectly Lambertian nor has a constant albedo. This section tries to manipulate the reflectance property assumption to make it a more realistic one.

The advantage of remotely sensed imageries is that they come in different spatial, spectral and temporal resolutions. Obviously, the more data available, the more information can be extracted out provided one uses proper tools to fuse the data.

However, the very basic assumption for this research is that there are no imageries for either stereo measurements or SFS using photometric approach for the region of interest. Nevertheless, there is no limitation imposed on the availability of imageries with different spatial or spectral resolutions coming from the same sensor with the similar

date and time of exposure.

To have a better understanding of the Earth's surface reflectance behavior some researchers have come up with the idea of measuring the reflectance in the field using the Field GOniometer System (FIGOS) [see Chapter 2 for more details]. But, obviously it is not practical to do the measurement for every single pixel of the image. On the other hand, others have used the look up table idea or some other simplistic mathematical model which is valid just for some certain and specific conditions and not any general case.

The basic idea for improving the performance of the assumed reflectance model in this research comes from the fact that the reflectance characteristics of the Earth's surface materials are not only different but also vary as a function of the wavelength of the light source. Having access to multispectral imageries, one can study the reflectance properties of the Earth surface using classification methods [see Appendix F for details].

After assigning each pixel to the predefined classes, one can apply a principal component transformation (PCT) to the spectral bands of the satellite imagery to remove the correlation between them and get the one with the highest energy to work with [see Appendix E for details].

The first approximate value of the albedo factor can be obtained by averaging the brightness values of the pixels falling under the same class. The basic statistical assumption in this process is that the direction of the surface normal vectors corresponding to pixels falling under a specific class has a normal distribution with mean zero with respect to the light source vector. Since satellite imageries cover relatively a

large area in comparison to their resolutions, therefore, the statistical assumption made above is not far away from the reality. Furthermore, Chapter 4 explains how one can further improve the estimated albedo factor if there is any violation in the statistical assumption made above.

The whole process mentioned here is to improve upon the idea of having constant albedo. In other words, there will be different albedo factors for pixels falling under different classes. The number of different albedo factors corresponds to the number of the classes. Obviously, the accuracy of the estimated albedo factors depends on the quality of the classification as well as the precision of the statistical assumption made above.

CHAPTER FOUR

EXPERIMENTAL RESULTS AND ANALYSIS

This chapter explains the numerical experimentation carried out to justify the proposed methodology and discusses the results in detail. The first section explains the general practical considerations taken into account during the analysis. It mainly consists of a discussion about the processing steps, the computation environments as well as the quality control methods used for this research.

The second section discusses the results of analysis on the general application of SFS to the spatial enhancement of DTM, while the smallest patch size or computation unit is investigated too. Both simulated and real data sets are used to carry out this analysis.

The third section consists of a discussion on the choice of the error kernel function (from robust statistics) used in the SFS solution. Three different types of error kernels with both simulated and real data sets are examined to investigate the best choice of error kernel for the specific application of this research. Results are discussed in detail in this section.

The fourth section is about the modification of the albedo factor. Using both simulated as well as real data sets, it is explained how a variable albedo factor can help the quality of the SFS solution.

The fifth section is an analysis on the phase dependency of the SFS solution. Two real data sets are processed with different known points to see the effect of changing the known points on the SFS solution.

The sixth section has a discussion about the correlation between the roughness of the terrain and the quality of the SFS solution. It is shown numerically that one can expect a better solution using SFS than the simple interpolation methods.

4.1 Numerical Considerations

The main goal of this investigation is to improve the accuracy of the DTM grid data interpolation by applying SFS techniques to the corresponding single satellite imagery or aerial photography.

The basic assumption is that the satellite imagery has one dyadic order better resolution than the original DTM data. Even though for the real data sets the exact one dyadic order is seldom the case, one can always change the resolution of the imageries using resampling techniques. Obviously, the quality of the resampled data depends on the resampling method. Moreover, it goes without saying that down sampling (increasing the sampling interval) is more reliable than up sampling (decreasing the sampling interval) and because of this fact, up sampling should be avoided whenever possible.

Furthermore, it is assumed that the reflectance characteristics of the surface are known. To start with, the analysis is carried out with the assumption of Lambertian surface with known constant albedo factor. As the experimentation goes on, more effective reflectance models with variable and more accurate albedo factors are used.

As mentioned before, with regards to the specific problem of this research, the illumination source is the Sun which is a distant point source with a known orientation. The position vector of the Sun is computed using a computer function called

“SUNPOS.FOR” downloaded from the “Mini Image Processing System (MIPS)” library of USGS [USGS, 2003].

The original computer function was in Fortran 77, but was converted to C++ to suit the Microsoft Visual C++ environment. The input for this function is the date (Julian), local time (hours, minutes, seconds), local time zone, and local geographic coordinates (in degrees). The MIPS also provides another computer function called “JULIAN.FOR” which converts the Gregorian date to the Julian and vice versa.

Considering the above mentioned facts and assumptions, it is seen that the only unknowns in the SFS formulation are p 's and q 's which are the surface normal vector components at each grid point. However, for the specific application of this research these estimated normal vector components are used as intermediate parameters to estimate the Earth's surface height which is sought in this research.

To ensure the integrability of the surface determined by the estimated surface normal components and also to take the known surface height grid points from the DTM data set into consideration, the estimated p 's and q 's are passed into a least-squares adjustment process. The results of this adjustment are the surface heights of the grid points for which enhancement in the interpolation accuracy are sought.

Recalling the fact that satellite imageries cover wide areas, one can easily encounter computational problems if appropriate measures are not implemented ahead of time. One of these considerations is the number of unknowns which grows rapidly as the size of image increases. Therefore, the first computational measure taken into consideration is the smallest patch or processing unit for which the SFS solution can provide reasonable

results within the realm of this investigation. In addition to the computational considerations, the smallest patch size is important when one is looking for the local behaviour or features of the surface. If the patch size is too large, the SFS solution will result in a surface which is relatively too smooth. Obviously, this is not what this investigation is about. The next section of this chapter justifies the choice of a 7 x 7 point grid patch.

The computational strategy adapted here is based on the patch idea mentioned above and consists of three stages: 1) pre-processing, 2) processing, and 3) post-processing. In the pre-processing stage two things are checked for each patch before anything else. The first check is if due to any reason (such as the existence of rivers or lakes) there is a gap in the DTM grid. The second check is if any of the corresponding grey levels in the image is either zero or the maximum possible value (depending on the dynamic range of the image this value can be different). In any of these two cases the whole patch is left unprocessed and the interpolated heights are considered as the final height values.

Otherwise the pre-processing stage continues with two steps. In the first step, the heights of the unknown points in the patch are approximated using interpolation techniques (bilinear or cubic convolution depending on the grid spacing and roughness of the surface). Thereafter, the relative orientation of the inner most square (see Section 4.2 for more detail) in the patch with respect to the light source is estimated. If this relative orientation implies that the corresponding grey level of the inner most square of the patch is zero, then there would be no shading information to improve the accuracy of the interpolated heights. Therefore, in this case too the interpolated heights are considered as

the final values.

The processing stage for each patch consists of two main steps. In the first step, the components of the unit surface normal vectors at the grid points are estimated using SFS techniques. In the second step, the surface normals are passed to an overdetermined linear adjustment process to solve for the heights. This is simply done by approximating p 's and q 's with their corresponding finite differences in terms of heights. The control goes back to the first step of this stage unless the average difference between the calculated and original image grey values of the pixels corresponding to the innermost square of the patch is less than a predetermined threshold.

The last stage, post-processing, takes care of different solutions for the unknown heights coming from the neighbouring patches. It consists of taking arithmetic means of two solutions for the unknown heights located on the boundary of the innermost square in each patch coming from the neighboring patches, except for the outsides of the peripheral patches.

To analyse the results, most of the SFS literature compares the estimated surface with the original surface graphically. However, some of them take one step further and compare the surfaces in a quantitative manner by computing the means and standard deviations of the differences between the original (if it is available at all) and the estimated surfaces. Recalling the specific application objectives of this research, one realizes that the graphical assessment of the results is not an option here. Therefore, the means and standard deviations of the differences between the original and the estimated heights using either interpolation method or SFS techniques are used here to assess the

results. Obviously, those points for which there is no solution from SFS techniques are not used in the assessment.

It is worth mentioning that technically, in this research work one could use the Root Mean Square Errors (RMSEs) instead of the means and standard deviations because the computed means do not refer to the true expected values. However, as the means show possible biases in the nonlinear SFS solutions, i.e., using mean values and standard deviations (from the estimated means) have advantages over the RMSEs in being more explicit.

The computer codes for all simulations and prototyping the solutions for this research are written in Matlab environment. However, to deal with the real data sets where a large number of points should be processed, the codes are converted into Microsoft Visual C++ 5.0 environment. To take advantage of the parallel processing facility available at the University of Calgary, i.e., Multimedia Advanced Computational Infrastructure (MACI) alpha cluster, the Microsoft C++ codes are modified to suit the gcc compiler on the cluster. The parallel processing reduced the computation time very considerably. Regardless of the environment used to do the main SFS computations, Matlab is always used for visualization and statistical analysis purposes.

These general assumptions and discussions apply to most of the analysis in this chapter. However, if at any stage they are altered or modified in one way or another, there will be an explicit explanation about them.

4.2 Analysis of SFS Performance and Choice of Patch Size

The first set of experiments is to test the effectiveness of SFS application in the context of this research, while evaluating the proper size of the smallest patch or computational unit. As mentioned in the previous section, the smallest patch size is important from two points of view. First the number of unknowns and equations involved in the system of equations, second, the smoothness of the solution. Obviously, the larger the patch size is, the smoother the solution will be. Because in this research the local behaviour of the surface is sought, it is necessary to work with the smallest possible patch. To find out about the optimum size of the patch as well as the effectiveness of the SFS solution, a number of numerical experiments are carried out with one synthetic object and one real DTM data set together with their corresponding synthetic images.

The experiments begin with the patch size that is considered the smallest one, i.e., a 3 x 3 point grid with its centre located at position (i,j) . Four of these nine grid points have known heights (squares in Figure 4.1) and the other five are the unknown points (circles in Figure 4.1) for which interpolation accuracy improvement is sought.

The first processing stage of this analysis consists in estimating the surface normal components using the SFS techniques. The method used here is the standard variational algorithm with the smoothness constraint explained in Chapter 2. However, the surface normal vectors computed from the smoothness constraint are projected back onto the ambiguity cone (explained in Chapter 3). This makes the SFS solution independent of the choice of the weight factor used in the SFS technique to weigh the smoothness constraint versus the brightness error.

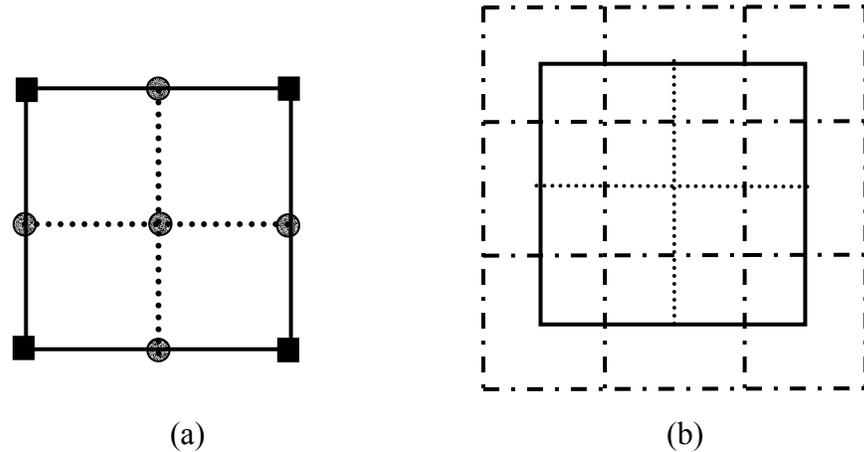


Figure 4.1: (a) A 3 x 3 point grid DTM patch. Squares are the grid points with known heights and the circles are those with interpolated heights, (b) The corresponding 3 x 3 pixel patch of the imagery. The grey levels at each pixel correspond to the centres of grid units.

The second processing stage involves solving an overdetermined system of equations in terms of the heights. This is done by approximating the estimated p 's and q 's (the surface normal vector components) from the SFS solution with their corresponding finite difference values. Doing that for each patch leads to nine linear equations corresponding to each grid point in the patch with four known and five unknown heights.

4.2.1 The Simulated Data Set

The synthetic object under investigation is a 33 x 33 point grid convex hemisphere with the radius of 6 units, sampled at each 0.5 unit. The corresponding less dense object (one dyadic order, i.e., 17 x 17 point grid) is extracted out from the object's height file. The 17 x 17 point grid object is interpolated using the bilinear interpolation method to get a 33 x 33 point grid approximation of the original object. The goal is to use the SFS technique to increase the accuracy of this approximate object.

The corresponding image of the object is created using the Lambertian reflectance model with the constant albedo factor of 255. A much (5 times) denser version of the object is used to generate the image. The resulting image is passed through a smoothing filter to get an image with the same density or pixel size as the original object under study. By changing the elevation angle of the light source, seven different images are created.

The differences between the original object, the Interpolated Grid Solution (IGS) and the SFS solution are analyzed. Table 4.1 summarizes these results (means and standard deviations of the differences). As mentioned before, the statistics shown in this table are computed with those points for which the SFS method is able to update their height values. The last row of the table shows the percentages of the patches which are updated by the SFS method.

In this analysis, the patches are not updated mainly due to two reasons. Either the relative position of the patches with respect to the light source is such that they are located in the shadow or the SFS solution didn't converge. Figure 4.2 shows the wiremesh of the object, the corresponding image and the wiremesh of differences between the original object and IGS as well as the SFS solution where the light source is located at azimuth and elevation angles of 135° and 45° , respectively.

Studying the means in the Table 4.1 shows that while there is no significant difference between the IGS and the SFS solutions, neither of them is biased either. Moreover, Table 4.1 shows that the improvement in the standard deviations of the differences is about 37% which is quite significant within 95% confidence interval.

Meanwhile, for this synthetic object it is seen that the optimum elevation angle of the light source is 60 degrees where there is about 41% improvement in the standard deviations while 90% of the patches are updated by the SFS solution. Obviously, for an object like this convex hemisphere, the number of points illuminated by the light source rises with increasing the elevation angle of the light source. The last two rows of Table 4.1 confirm that by increasing the light source elevation angle, the numbers of updated patches are increased as well as the rate of improvement in the standard deviations by SFS solution.

Table 4.1: The convex hemisphere with the 3 x 3 point grid patch

Elevation		30°	35°	40°	45°	50°	55°	60°
Object - IGS	Mean	-0.05	-0.03	-0.04	0.02	0.00	-0.02	0.03
	Std	0.28	0.29	0.31	0.32	0.32	0.35	0.37
Object - SFS	Mean	-0.04	-0.03	-0.06	-0.03	-0.02	-0.04	-0.05
	Std	0.19	0.20	0.20	0.20	0.20	0.21	0.22
Improvement in Std		32%	32%	34%	38%	39%	41%	41%
Patches updated out of 196		78%	80%	84%	85%	88%	88%	90%

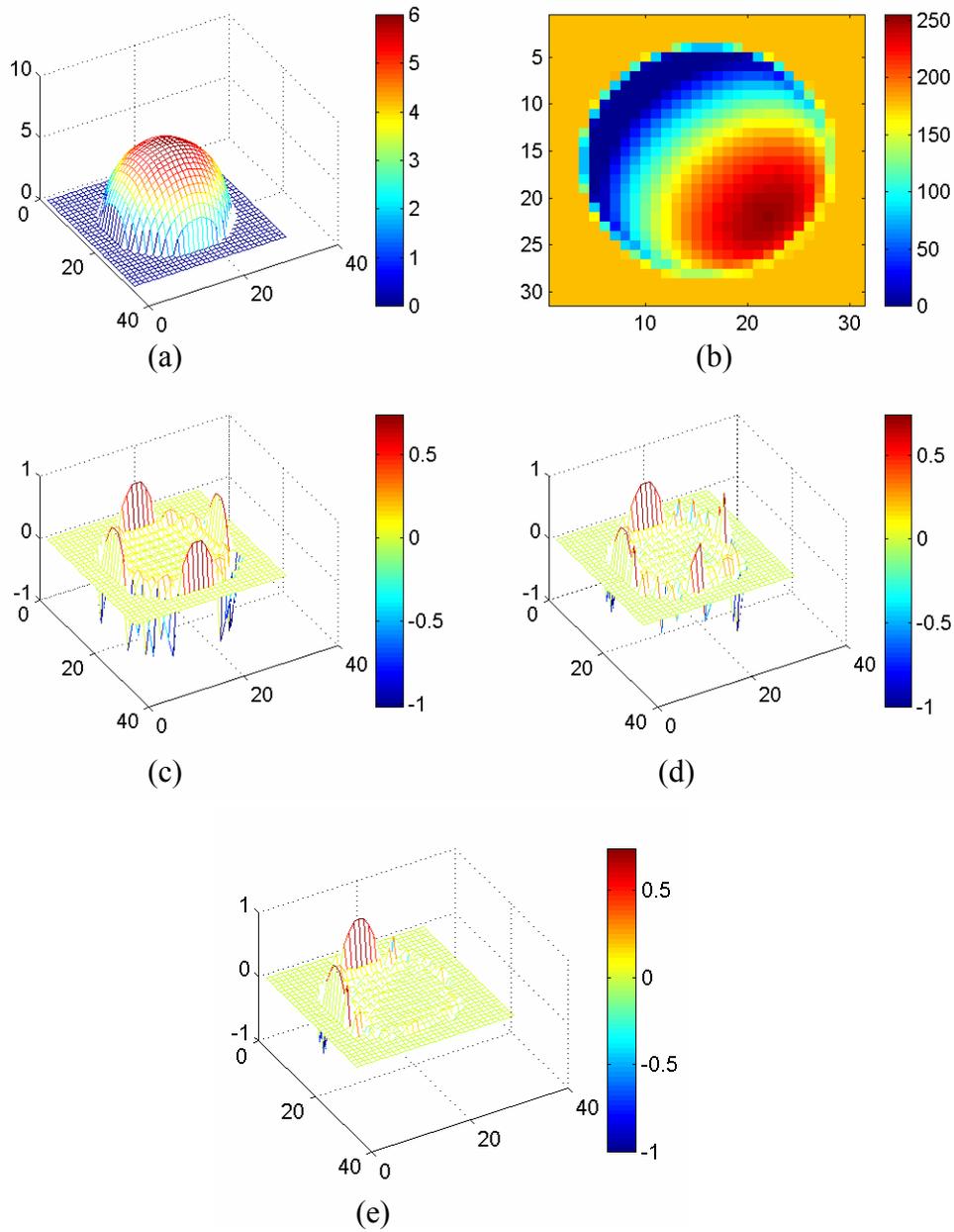


Figure 4.2: a) The original object; b) The corresponding image; c) The differences between the original object and the IGS; d) The differences between the original object and the SFS solution; and e) same as (d) but with a 7 x 7 point grid patch

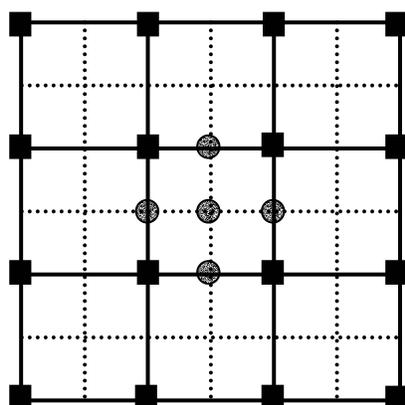
Table 4.1 says nothing about the azimuth of the light source as the symmetry of the synthetic object under investigation makes the process independent of the light source azimuth. Studying Figures 4.2-a to 4.2-d reveals that the methodology used in this analysis has not converged to a meaningful solution to recover the shape of the object at the base of the sphere where there is an abrupt change of slope. The reason for this is believed to be the size of the patch which is too small.

To investigate this, the patch size is changed to a 7 x 7 point grid (Figure 4.3) and the computations are repeated. Out of 49 points in this new patch, 16 of them (squares in Figure 4.3) have known heights, while the heights of the other 33 (circles in the Figure 4.3) points are approximated using interpolation techniques. In other words, using the new patch one has to solve 49 linear equations corresponding to the 49 grid points with 33 unknown and 16 known heights. Figure 4.2-e shows the differences between the original object and the SFS solution corresponding to Figure 4.2-d. Table 4.2 summarises the results of computations with the new patch size.

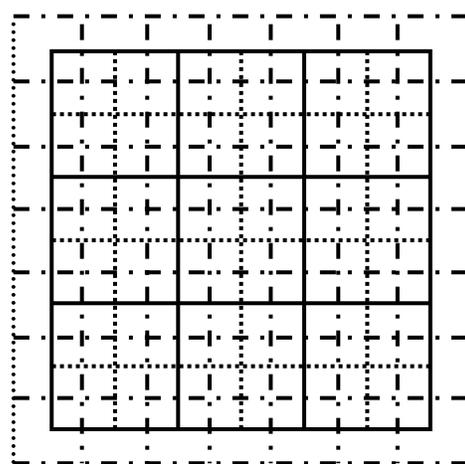
Results in Table 4.2 show that in comparison to Table 4.1, the means have not changed. However, Table 4.2 clearly confirms that changing the patch size from a 3 x 3 point grid to a 7 x 7 one has increased the number of patches which have been updated in the SFS computations. The number of patches which have not been updated in this SFS solution are approximately equal to the number of patches which are located in the absolute dark part of the image for which there is no SFS solution. Tables 4.2 shows that with the new patch size while more patches have been updated the rate of improvement in the standard deviation has also slightly increased from an average of 37% to 40%.

Table 4.2: The convex hemisphere with the 7 x 7 point grid patch

Elevation		30°	35°	40°	45°	50°	55°	60°
Object - IGS	Mean	-0.04	-0.03	-0.01	0.01	0.02	0.01	0.01
	Std	0.33	0.35	0.35	0.38	0.40	0.41	0.42
Object - SFS	Mean	-0.05	-0.04	0.02	-0.02	-0.03	0.01	-0.03
	Std	0.21	0.22	0.21	0.22	0.23	0.23	0.23
Improvement in Std		36%	37%	39%	41%	42%	43%	45%
Patches updated out of 196		82%	84%	85%	87%	90%	91%	93%



(a)



(b)

Figure 4.3: (a) A 7 x 7 point grid patch. Squares are the grid points with known heights and the unmarked ones are the points with interpolated heights. The circles are those points with interpolated heights for which enhancement in accuracy is sought through SFS. (b) The corresponding 7 x 7 pixel patch of the imagery. The grey levels at each pixel correspond to the centres of grid units.

4.2.2 Simulation with a Real DTM Data Set

The second test object used in this analysis is a real height data set from Waterton Lake area, southern Alberta, Canada, with 25 metre spacing in UTM zone 12 coordinate system. The original data set is a 1024 x 1024 point grid with more than 1300 metre height difference which is subsequently down sampled to 125 metres. These sampled height data are considered as the object under investigation and the corresponding one dyadic order coarser object, i.e., 250 metre grid, is extracted from it. Meanwhile, the corresponding synthetic image is generated using a Lambertian reflectance model with the original 25 metre height data. The albedo factor is set to the constant value of 255. The resulting image is passed through a smoothing filter to get an image with the same resolution as the object under study, i.e., 125 metres. By changing the azimuth and elevation angles of the light source nine different images are created.

Table 4.3 summarizes the results of the experiment with this data set using a 3 x 3 patch. It shows that in comparison to the IGS solution, the SFS solution is slightly closer to the object under study. Furthermore, in comparison to the other two cases the means have slightly higher values when the light source is at the azimuth of 180°. It is believed that these biases in the solutions are due to the morphology of the terrain. Table 4.3 also shows that the SFS solution has improved the standard deviation of the mean differences with the average amount of 35%. The results show a clear dependency of the SFS solution on the relative orientation of the light source. As it can be seen, the worst results are for the light source at the azimuth of 180 degrees. Obviously, this is due to the morphology of the terrain.

Table 4.3: The real DTM data set with the 3 x 3 point grid patch

Azimuth		135°			180°			225°		
Elevation		30°	45°	60°	30°	45°	60°	30°	45°	60°
Object – IGS	Mean (m)	-0.04	0.12	-0.21	0.32	0.41	0.63	0.09	0.17	0.19
	Std (m)	12.8	13.2	13.7	14.2	16.9	18.3	12.5	12.8	13.4
Object - SFS	Mean (m)	0.16	-0.06	0.09	0.24	0.36	0.19	0.11	0.14	0.13
	Std (m)	8.7	8.4	8.4	9.8	11.1	11.5	8.4	8.2	8.3
Improvement in Std		32%	36%	39%	31%	34%	37%	33%	36%	38%
Patches updated out of 9409		83%	88%	90%	75%	79%	82%	84%	88%	89%

Figure 4.4 shows the wiremesh of the object, the corresponding image and the wiremesh of the differences between the original object, IGS, and the SFS solution, where the light source is located at the azimuth and elevation angles of 135° and 45° , respectively.

Similar to the previous case, from Figure 4.4 it is clear that SFS has not been able to improve the interpolation values where there is a sudden change of height. The main reason for this is believed to be the size of the patch. To verify it, the computations are repeated with the bigger patch of 7×7 point grid. The result is summarized in Table 4.4.

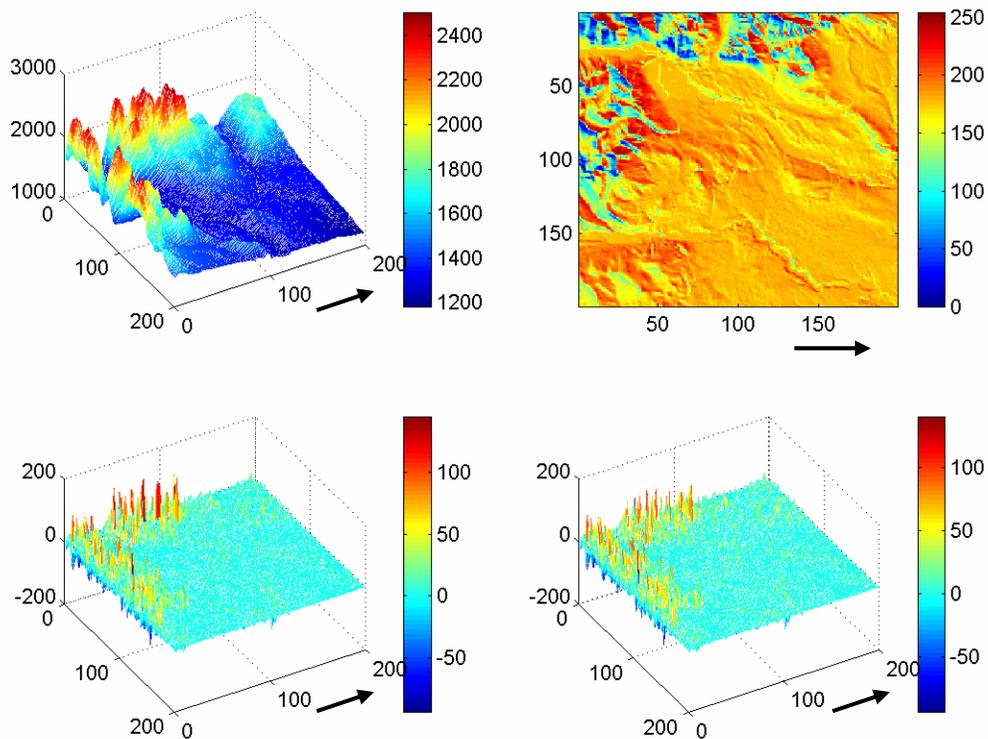


Figure 4.4: a) The original object; b) The corresponding image; c) The differences between the original object and the IGS, and d) The differences between the original object and the SFS solution. The arrow signs show the North direction

Table 4.4: The real DTM data set with the 7 x 7 point grid patch

Azimuth		135°			180°			225°		
Elevation		30°	45°	60°	30°	45°	60°	30°	45°	60°
Object – IGS	Mean (m)	0.04	0.11	0.22	0.42	0.53	0.69	0.15	0.21	0.25
	Std (m)	13.8	14.2	14.6	15.3	18.2	19.5	13.3	13.5	14.1
Object – SFS	Mean (m)	0.12	0.05	-0.03	0.29	0.39	0.47	0.09	0.13	0.18
	Std (m)	9.0	8.8	8.6	10.4	11.6	11.9	8.8	8.5	8.3
Improvement in Std		35%	38%	41%	32%	36%	39%	34%	37%	41%
Patches updated out of 9409		89%	93%	95%	80%	82%	84%	90%	91%	95%

Tables 4.3 and 4.4 show that the numbers of patches which have been updated in the SFS process have increased. This has been accompanied by a slight increase in the rate of improvement in standard deviation from 35% to 37% as well. It is worth mentioning that all the statistics shown above are computed with the grid points which SFS has been able to update their height values.

The results of the simulation analysis in this section show that using shading information even with the simplest formulation of the SFS can improve the accuracy of the interpolation quite significantly. Moreover, it is shown that the patch size has an obvious impact on the convergence of the SFS solution, especially in cases where there is

a discontinuity in the surface height.

The next smallest possible patch size is an 11 x 11 one. In this case, one should solve for 242 unknown p's and q's or 85 unknown heights in comparison to 49 unknown p's and q's or 33 unknown heights in case of 7 x 7 patch. As it is seen, the number of unknowns and in turn the computational effort to solve for these unknowns are significantly increased.

Moreover, comparing the results obtained for the convex hemisphere case using a 3 x 3 and a 7 x 7 point grid patch (Figure 4.2-d and 4.2-e) shows that the 7 x 7 point grid patch has been practically able to update all the patches for which shading information is available. In other words, based on the simulation results, it is seen that a 7 x 7 patch is good enough to provide a convergent solution to the specific SFS problem of this research.

Furthermore, one may think of using an 11 x 11 point grid patch with a larger innermost zone (7 x 7 instead of 3 x 3). As this larger innermost zone would undoubtedly result in a smoother surface in comparison to the patches with a 3 x 3 grid point innermost zone, it is not used here considering the objectives of this research. Therefore, based on the results of the analysis in this section, the 7 x 7 point grid patch is used for the rest of the analyses in this chapter.

4.3 The Choice of Error Kernel Function

Chapter 3 contains an extensive discussion about the application of error kernel functions from robust statistics in the SFS solution. Despite the fact that the choice of error kernel

functions is discussed from a theoretical point of view, a number of numerical tests are done here to verify the choice of best type of error kernel. As mentioned in Chapter 3, there are three different classes of error kernel functions, i.e., quadratic, re-descending, and sigmoidal. The standard variational solution to the SFS problem is used here as an example for the quadratic error kernel function. As a re-descending error kernel, $-w \exp(-v^2/w)$, one of the Li's adaptive functions, is used. Finally, the hyperbolic tangent error kernel function, i.e., $\frac{w}{v} \log \cosh\left(\frac{\pi v}{w}\right)$ is used here as a sigmoidal kernel.

The widths of the functions are initially set to one ($w = 1$) and the curvature is considered to be the error parameter of the functions. The derivations and formulations of SFS with these error kernels are explained in Chapter 3. Here, only the corresponding numerical examples are discussed.

4.3.1 Simulated Data Set

A 1024 x 1024 point grid convex hemisphere with the radius of 250 units, sampled at 0.5 units is created to test the behaviour of different error kernels. The corresponding one dyadic order coarser of this object, i.e., a 512 x 512 point grid, is obtained by taking every other elevation point out of the simulated elevation file of the object. Assuming a Lambertian model with a distant point source at three different elevation angles and also a constant albedo factor of 255, three different synthetic images corresponding to the object are generated. To make the simulated images as similar as possible to the real

images, a much denser (five times) version of the object is used in the image generation process. The dense images are passed through a smoothing filter to get images with the same density or pixel size as the original object.

The SFS algorithm with three different error kernels is applied to the simulated data set to improve the accuracy of interpolation. To check the performance of the kernels, similar to the previous cases, the standard deviation of the IGS is compared with the SFS solutions. Table 4.5 shows the results of this analysis.

Table 4.5: The convex hemisphere

Type of Kernel		Standard Variational Method (Quadratic)			Re-descending			Sigmoidal		
Elevation		30°	45°	60°	30°	45°	60°	30°	45°	60°
Object – IGS	Mean	0.07	-0.05	-0.08	-0.03	0.04	0.06	0.04	0.02	-0.05
	Std	0.27	0.35	0.41	0.28	0.34	0.41	0.30	0.37	0.43
Object – SFS	Mean	-0.06	-0.05	0.04	0.03	0.02	0.05	0.01	0.05	0.03
	Std	0.18	0.22	0.25	0.18	0.21	0.25	0.19	0.23	0.26
Improvement in Std		35%	36%	39%	35%	37%	39%	36%	37%	40%
Patches updated out of 257049		81%	85%	92%	83%	86%	92%	85%	88%	94%

Table 4.5 shows while there is no significant difference in means between the IGS and SFS solutions, they are not biased either. Table 4.5 also reveals a slight difference

between the solutions both from rate of improvement in standard deviation and number of updated patches, however, the difference is not that much significant. The reason can be justified based on discussions in Chapter 3 where the theoretical difference between the solutions is explained. According to that discussion the functionality of the kernels is different in the presence of noise and/or discontinuities. Considering this simulated case, there is neither noise in the data nor discontinuities except at the base of the sphere where there is an abrupt change in the slope of the object.

To experiment with the behaviour of the kernels in presence of noise and discontinuities, a convex hemisphere similar to the previous case is generated again but this time with a discontinuity of 10 units in its radius. Images of this object with three light source elevations are generated too. Moreover, normally distributed random errors with mean zero and variance three are added to the images. Figure 4.5 shows the simulated image of this object when the light source is at azimuth 135 degrees and elevation angle of 45 degrees. Table 4.6 summarizes the results of the SFS solutions with using different kernels. In this analysis the width parameter of the error kernel functions is considered as a variable function of shape index dispersion (explained in Chapter 3). Again, it should be remembered that all the statistics shown in Table 4.6 are computed with those points for which SFS solution is available.

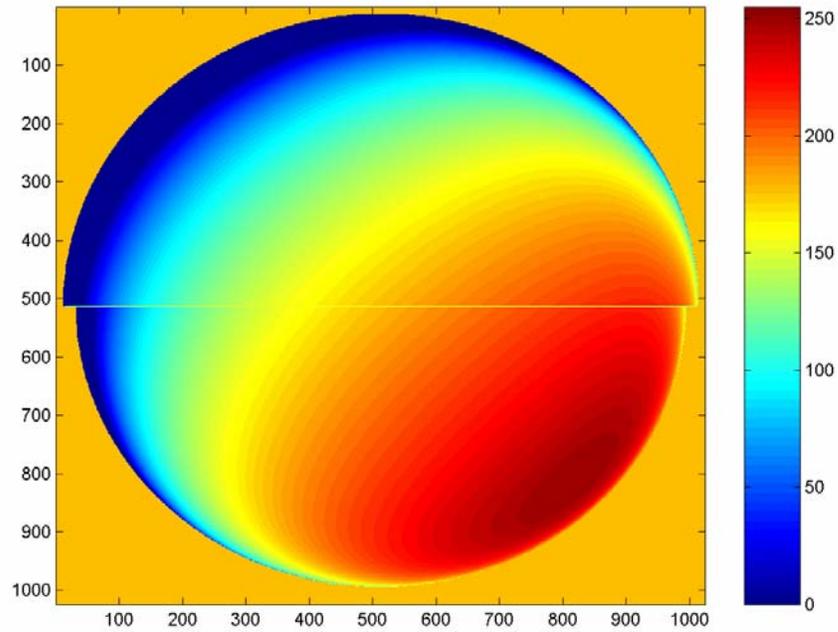


Figure 4.5: Image of the simulated object with a discontinuity in curvature

Table 4.6: The convex hemisphere with noise and discontinuity in curvature

Type of Kernel		Standard Variational Method (Quadratic)			Re-descending			Sigmoidal		
Elevation		30°	45°	60°	30°	45°	60°	30°	45°	60°
Object – IGS	Mean	1.25	1.27	1.24	1.30	1.31	1.35	1.28	1.23	1.27
	Std	2.15	2.19	2.18	2.19	2.23	2.24	2.21	2.22	2.25
Object – SFS	Mean	0.93	0.89	0.90	0.78	0.80	0.83	0.82	0.78	0.83
	Std	1.63	1.59	1.50	1.49	1.45	1.39	1.44	1.40	1.35
Improvement in Std		24%	27%	31%	32%	35%	38%	35%	37%	40%
Patches Updated out of 1521		74%	77%	82%	79%	82%	87%	83%	87%	90%

The results shown in Table 4.6 have been obtained by applying the SFS algorithms with different kernels to a narrow band around the discontinuity area. This is to make sure that the analyses clearly show the performance of the kernels in the presence of noise and discontinuity. Means in Table 4.6 show that both the IGS and the SFS solutions are biased. The bias obviously comes from the fact that the analysis has been done for a narrow strip along the discontinuity. However, the SFS solution shows a reduction in the bias by up to about 30%.

Table 4.6 also shows that in comparison to the other two kernels the standard variational method has come up with an object with less improvement in standard deviation and less number of updated patches than the other two kernels. While sigmoidal kernel gives the best matching shape in terms of the improvement in standard deviation as well as the number of the updated patches, a re-descending kernel acts somewhere in between the other two kernels.

4.3.2 Real DTM Data Set and Satellite Imagery

In order to further test the implementations in selecting the kernels, another set of experiments are done. The test object is a 1024 x 1024 point grid of real height data set from Waterton Lake area, southern Alberta, Canada, with 25 metre spacing in UTM zone 12 coordinate system (Figure 4.6). The maximum height difference is 1300 metres. The corresponding one dyadic order coarser object, i.e., 50 metre point grid, is extracted from the data set.

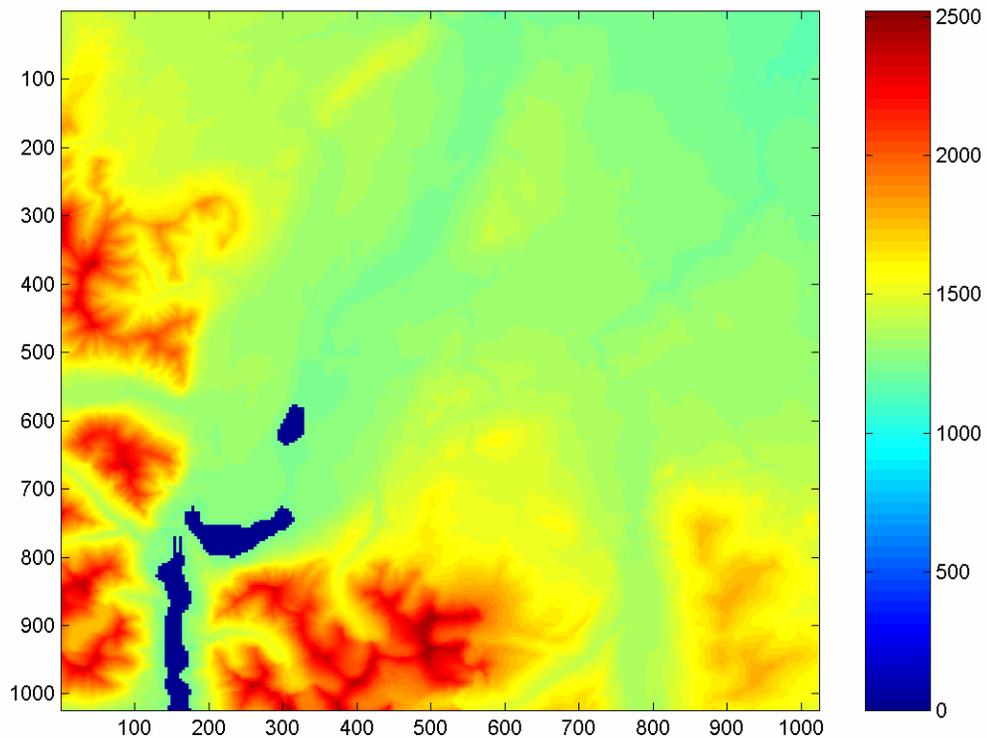


Figure 4.6: Waterton Lake area DTM data set. The dark spots are the lakes

The corresponding image is a three-channel multispectral SPOT imagery with 20 m resolution (Figure 4.7). PCI software along with 1/20,000 topographic maps are used for registration of the satellite imagery. With 23 feature points in the area of interest and a second order polynomial, the georeferencing has Root Mean Square (RMS) values of 2.96 m and 1.40 m in Easting and Northing directions, respectively. The principal component transformation (PCT) is applied to all three channels of the imagery to get one channel with the most energy. Finally, the first channel with 88.52% of energy is resampled to 25 metres and used in this analysis.

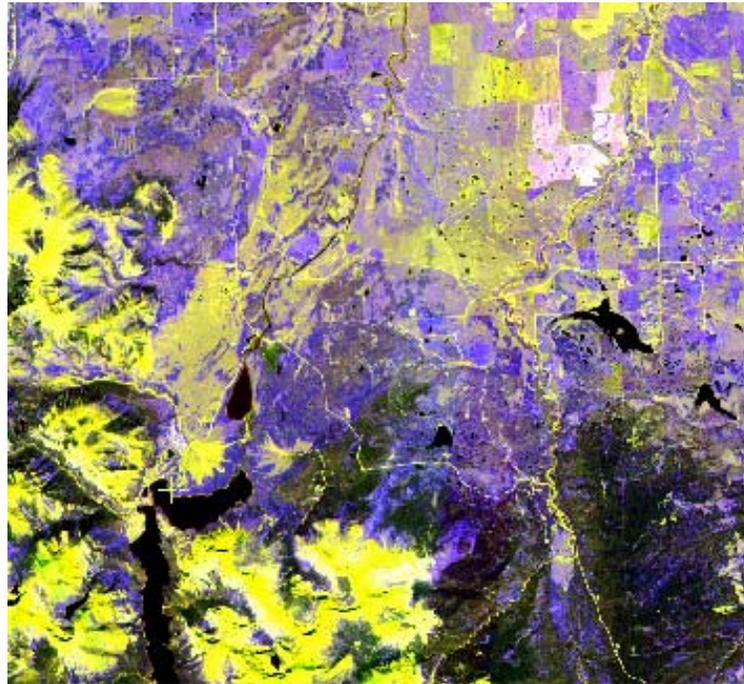


Figure 4.7: SPOT image of the Waterton Lake area

Using the SFS algorithm with three different kernels, the objective is to improve the interpolation accuracy of the reconstructed 25 m spacing object from its 50 m version. The width parameter of the kernels is considered as a variable function of the shape index dispersion in this analysis. The results which are computed by the points for which there is solution from SFS are summarized in Table 4.7.

Table 4.7: The Waterton Lake area DTM data set and SPOT imagery

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.03	0.05	-.007
	Std (m)	14.95	15.32	16.27
Object - SFS	Mean (m)	-0.04	-0.05	-0.08
	Std (m)	10.17	8.42	9.76
Improvement in Std		32%	45%	48%
Patches updated out of 257049		84%	89%	92%

The means in Table 4.7 shows that neither the IGS nor the SFS solution is biased. Moreover, according to the Table 4.7 the SFS solution with the sigmoidal kernel has the highest rate of standard deviation improvement. The number of updated patches in case of sigmoidal kernel is the highest in comparison to the other two kernels as well. The solution with the standard variational method with the quadratic kernel has the poorest result. However, the difference between the re-descending and the sigmoidal kernels is not that much significant. To justify this, one should remember that the difference between the performance of the kernels is highlighted when there are noises and discontinuities in the data. The Waterton Lake DTM data used in this analysis has both relatively smooth as well as rough areas where the smooth area is much larger than the rough area.

In order to focus more on the efficiency of the kernels another test is done with the same data set but this time for those areas which seem to have more discontinuities in the height values. These areas are the first 250 columns from the western boundary and the first 275 rows from the southern boundary of the DTM data set (Figure 4.8). The maximum height difference in this area is still more than 1300 metres. Table 4.8 shows the results for this test.

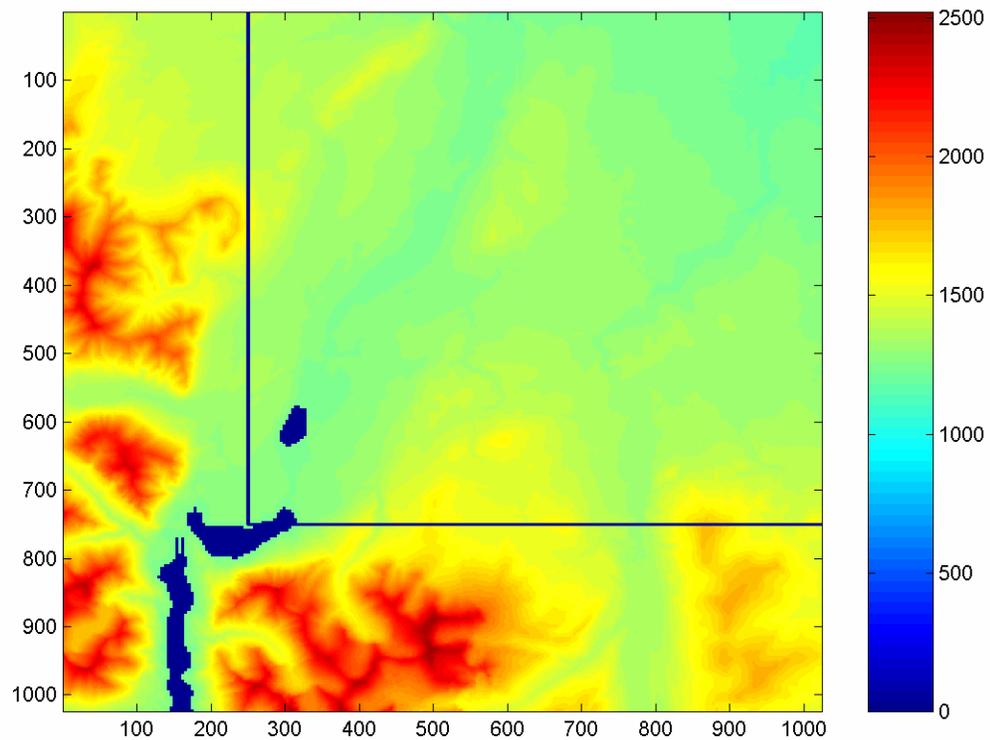


Figure 4.8: The rough part of the real DTM data set

Table 4.8: The rough part of Waterton Lake area DTM data set and SPOT imagery

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.12	0.17	0.21
	Std (m)	17.24	19.63	21.51
Object – SFS	Mean (m)	0.07	0.11	0.09
	Std (m)	12.59	10.21	9.46
Improvement in Std		27%	48%	56%
Patches Updated out of 110240		80%	85%	88%

Statistics in Table 4.8 have been computed with those points for which SFS has provided a solution. This Table shows that both IGS and SFS solutions are slightly biased. The bias is obviously due to the asymmetry of the terrain under study. However, as it is seen, the bias in the SFS solution is clearly less than the bias in the IGS solution. The Table also shows that SFS can considerably improve the accuracy of interpolation. However, the rate of improvement is different depending on the roughness of the terrain and the type of kernel used for regularizing the SFS solution. The rougher the terrain is, the larger the differences between the performances of different kernels are. While the re-descending kernel acts somewhere in between the other two kernels, the quadratic kernel has the poorest results as the shape coming from its algorithm tends to be a smooth one even when it is dealing with a rough terrain. The sensitivity of the sigmoidal kernel

to the changes in the curvature makes it ideal for application of SFS with rough surfaces.

4.4 Modification to the Albedo Factor

Having access to multispectral imageries, one can have an estimate for the albedo factor in the irradiance equation. As mentioned in Chapter 3, the estimation of the albedo factor is done through the classification of the pixels. The question to answer in this section is if manipulating the albedo factor can result in a better solution for the SFS. The following numerical experiments are carried out to investigate this.

4.4.1 Simulated Data Set

The simulated object and image used in this section are same as those used in Section 4.3.1. The object is a 1024 x 1024 convex hemisphere with 250 units radius sampled at every 0.5 unit with three simulated images corresponding to three different elevation angles (30, 45, and 60 degrees) for the light source. However, instead of using the constant 255 as the albedo factor, the grey values of the images are arbitrarily scaled with three different albedo factors to simulate a terrain with three different types of coverage. The albedo factors used are 75, 150, and 225 with some normally distributed random noise with mean zero and variance five.

A Mahalanobis classifier is used to classify the pixels in the images. Hundreds of randomly distributed pixels in each class are used as the training data set. Finally, after assigning each pixel to one of the three classes, the mean grey values of each class are computed and considered as the albedo factors for each class.

Tables 4.9 and 4.10 compare the interpolated solution to three different SFS

solutions. Table 4.9 summarizes the results for the case where the albedo factor is assumed to be constant all over the image, while, Table 4.10 shows the result for the case where the albedo factor is assumed to be variable for each pixel. It should be remembered that the statistics shown in these Tables have been computed using the points for which SFS has provided a solution.

Table 4.9: Simulated data set with constant albedo factor

Type of Kernel		Standard Variational Method (quadratic)			Re-descending			Sigmoidal		
Elevation		30°	45°	60°	30°	45°	60°	30°	45°	60°
Object – IGS	Mean	0.08	0.06	-0.04	0.05	-0.03	-0.07	0.06	-0.05	0.03
	Std	0.29	0.32	0.36	0.31	0.33	0.35	0.32	0.37	0.41
Object – SFS	Mean	0.03	-0.03	0.05	0.04	0.04	-0.03	0.03	-0.05	0.01
	Std	0.24	0.26	0.29	0.24	0.25	0.26	0.24	0.27	0.30
Improvement in Std		16%	18%	20%	23%	25%	25%	25%	26%	26%
Patches updated out of 257049		76%	78%	81%	79%	82%	86%	80%	83%	87%

Table 4.10: Simulated data set with variable albedo factor

Type of Kernel		Standard Variational Method (quadratic)			Re-descending			Sigmoidal		
Elevation		30°	45°	60°	30°	45°	60°	30°	45°	60°
Object – IGS	Mean	0.08	-0.09	-0.04	0.03	-0.04	-0.06	0.05	-0.03	-0.05
	Std	0.28	0.33	0.39	0.30	0.33	0.39	0.31	0.36	0.42
Object - SFS	Mean	0.05	-0.04	0.03	0.04	0.03	-0.05	0.04	-0.04	0.02
	Std	0.19	0.21	0.24	0.20	0.21	0.23	0.20	0.23	0.25
Improvement in Std		33%	35%	38%	34%	37%	40%	35%	36%	41%
Patches updated out of 257049		79%	82%	88%	81%	85%	89%	82%	85%	90%

The Tables show no bias in either of IGS or SFS solution. The perfect symmetry of the object under study justifies this fact. Even though the calculated albedo factors are not exactly the same as those used to scale the grey values, there is still a considerable improvement in SFS solutions when the variability of the albedo factor is taken into consideration.

4.4.2 Real Data Set

The idea of having a variable albedo factor is tested with a real data set here. The data set used for this test is same as the one used in Section 4.3.2. The object is a 1024 x 1024 real DTM data set from Waterton Lake area and its corresponding 3-channel

multispectral SPOT imagery. As mentioned in Section 4.3.2, the SPOT imagery is georeferenced in UTM map projection system using 1/20,000 topographic maps of the area. Then, the images are resampled from 20 m resolution to 25 m.

PCI Geomatics software is used to implement the principal component transformation (PCT) to the 3-channel multispectral SPOT imagery. The first transformed channel with 88.52% energy is selected to work with in this experiment. A Mahalanobis classifier is used to classify the image pixels into three different classes representing vegetation, soil, and water. About hundred pixels of each class are used as training data set for the classification.

The mean grey level of pixels in each class is used as the albedo factor for the corresponding class. Table 4.7 basically shows the result for the case of constant albedo factor and therefore it is not repeated here. Table 4.11, computed with those points for which there is a solution from SFS, summarizes the results for the variable albedo factor case.

Comparing the results in the last two rows of Tables 4.7 and 4.11 confirms that the improvement in the SFS solutions is quite significant when the albedo factor is considered as a variable factor. Moreover, similar to the simulated case, the SFS solution using the sigmoidal error kernel has the best solution among all of the SFS solutions.

Table 4.11: Real data set with variable albedo factor

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.05	-0.05	-0.09
	Std (m)	16.23	17.52	18.37
Object – SFS	Mean (m)	0.08	-0.12	0.07
	Std (m)	10.39	8.76	8.45
Improvement in Std		36%	50%	54%
Patches updated out of 257049		87%	91%	93%

4.4.3 Further Modification of the Albedo Factor

The idea of assigning the average grey levels of pixels in a class to the albedo factor corresponding to that class is based on the statistical assumption that pixels in each class assume all the possible directions. The fact that satellite imageries usually cover large areas shows that the assumption is not far away from the reality. This assumption is especially true for low-resolution satellite imageries which usually cover a larger area rather than high-resolution imageries. However, for high-resolution imageries one can apply a further modification to the albedo factor to take care of the possible small discrepancy between the assumption and the real situation.

The idea of further modification of the albedo factor comes from the fact that the spacing between DTM grid points are usually selected based on the roughness analysis of

the terrain. Therefore, one can assume that the interpolated heights are not too far away from the true heights. Consequently, based on this assumption, the mean difference between the observed image intensity and the reconstructed one using the estimated albedo factors from classification and the interpolated heights indicates the amount of modification to the albedo factors.

Tables 4.12 and 4.13, computed with those points for which SFS has a solution, show the results of testing this idea with the simulated and real data sets, respectively. The means show that neither of the IGS nor the SFS solutions is biased. Furthermore, as it can be seen, in case of the simulated data set and in comparison to Table 4.10, this further modification of the albedo factor has resulted in deterioration of the SFS solutions. However, in case of real data set and in comparison to Table 4.11, there is an improvement in SFS solutions. The reason for not having better results for the SFS solutions in case of the simulated data set is believed to be the perfect random distribution of the pixels in different classes all over the object surface. Moreover, the perfect symmetry of the simulated object surface, which is rarely the case with the real Earth's surface, can be another reason.

Table 4.12: Simulated data set with the second modification to albedo factor

Type of Kernel		Standard Variational Method (Quadratic)			Re-descending			Sigmoidal		
Elevation		30°	45°	60°	30°	45°	60°	30°	45°	60°
Object – IGS	Mean	-0.06	0.03	0.07	0.09	-0.02	-0.03	0.07	-0.05	0.06
	Std	0.29	0.33	0.38	0.31	0.32	0.38	0.32	0.35	0.39
Object – SFS	Mean	0.03	0.04	-0.02	0.06	-0.04	0.05	-0.03	-0.04	-0.03
	Std	0.24	0.26	0.30	0.23	0.23	0.26	0.24	0.25	0.27
Improvement in Std		18%	21%	22%	25%	28%	31%	26%	28%	32%
Patches Updated out of 257049		78%	80%	83%	81%	84%	88%	82%	84%	89%

Table 4.13: Real data set with the second modification to albedo factor

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object - IGS	Mean (m)	0.09	0.08	0.10
	Std (m)	17.83	19.19	20.01
Object – SFS	Mean (m)	-0.06	0.07	-0.07
	Std (m)	10.70	9.02	8.40
Improvement in Std		40%	53%	58%
Patches updated out of 257049		88%	93%	94%

4.5 More Experiments with Real Data Sets

Further tests are done to analyse and investigate the possibility of enhancing DTM interpolation accuracy using SFS techniques. The tests are done with a number of different satellite imageries along with their corresponding DTM grid points. The following sections explain and analyse the experiments in detail.

4.5.1 Landsat 7 ETM+ and the Waterton Lake Area DTM

A Landsat 7 ETM+ imagery was made available for this research by the Library of the University of Calgary. The imagery consists of eight channels including six ETM (Enhanced Thematic Mapper, 30 m resolution), one panchromatic (15 m resolution) and one thermal (60 m resolution). Figure 4.9 shows the scene around the Waterton Lake.

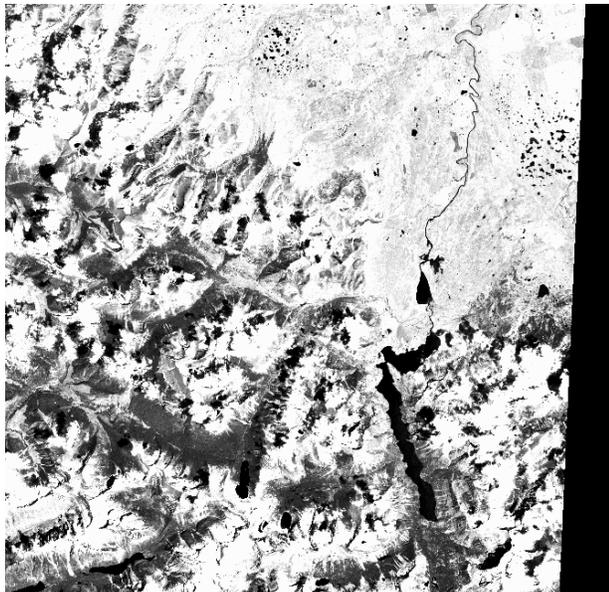


Figure 4.9: Landsat 7 ETM +, Waterton Lake area

The imagery was taken in July 1999. According to the metadata file which comes with the imagery all radiometric and systematic corrections were already applied to it. However, with regard to this research, as it can be seen in Figure 4.9, there are two main problems with the imagery. The first problem is the cloud cover over the rough part of the area of interest. Unfortunately, neither the area under the clouds nor those which are in the shadow of the clouds can be used in the analysis. The second problem with the imagery is the area covered. Because Waterton Lake area is located at the North East corner of the image scene, unfortunately, it doesn't cover all the area where the other analyses are done. Therefore, this makes it difficult to compare the solutions using this image with the other solutions.

To proceed with the experimentation, the first problem to solve is the registration of the image. Panchromatic band of the image (15 m resolution) with 17 control points extracted from 1/20,000 topographic maps are used for this task. PCI Geomatics software, with bilinear resampling method and second order polynomial as the parameters, is used for the georeferencing purpose. The RMS error in easting and northing are 1.79 m and 1.92 m, respectively. Thereafter, the ETM channels are also registered using image to image registration method.

As the first attempt the panchromatic band of the image is used to enhance the accuracy of 50 m to 25 m DTM interpolation. To do this, the image is resampled to 25 m resolution (ETM bands from 30 m and panchromatic band from 15 m). Afterward, the Mahalanobis classifier is used to classify the pixels and modify the albedo factor for each pixel. Table 4.14, computed with those points for which there is a solution from SFS,

summarizes the results.

Table 4.14: Panchromatic band of Landsat 7 ETM + and the Waterton Lake DTM

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.12	0.13	0.14
	Std (m)	4.36	4.71	4.78
Object – SFS	Mean (m)	0.13	0.13	0.13
	Std (m)	3.44	3.58	3.63
Improvement in Std		21%	24%	24%
Patches Updated out of 257049		28%	30%	31%

The means in Table 4.14 show that both IGS and SFS solutions are slightly biased. The biased is obviously due to the asymmetry of the terrain under study. The Table also shows that SFS solutions have slightly enhanced the accuracy of the interpolation solution. Moreover, the differences between different SFS solutions are very small. Furthermore, the rate of enhancement shows a relatively big difference in comparison to the case when SPOT imagery is used. And finally, the numbers of patches which are not updated with the SFS solutions are quite high.

The reason for having a small rate of enhancement by the SFS solutions is believed to be mainly the cloud coverage in the image. Unfortunately, the cloud has covered the

rough part of the area where the SFS solutions techniques perform the best. Moreover, the differences between different SFS solutions are more obvious if they are applied to an image corresponding to a rough surface.

Comparing Figure 4.9 with Figure 4.7 shows that almost 30% of the area is not covered by the Landsat 7 ETM + image. This fact in addition to the cloud coverage is the reason why there are few patches updated with the SFS techniques. Furthermore, poor image contrast and therefore poor classification may be another reason for these results.

To see if the contrast has had any effect on the results obtained in Table 4.14, the ETM bands of the Landsat image are used with the SFS algorithms. To do so, PCT is applied to the resampled ETM bands, and the first band with 92.21% of energy is used for the modification of albedo factor and SFS analysis. Table 4.15 shows these results.

Table 4.15: ETM bands of Landsat 7 ETM + and Waterton Lake area DTM

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.13	0.13	0.13
	Std (m)	4.54	4.79	4.80
Object – SFS	Mean (m)	0.12	0.13	0.13
	Std (m)	3.54	3.64	3.60
Improvement in Std		22%	24%	25%
Patches updated out of 257049		29%	30%	30%

Similar to Table 4.14, the means in Table 4.15 show that both IGS and SFS solutions are slightly biased. The bias is obviously due to the asymmetry of the terrain under study. The results in Table 4.15 are slightly better than the results in Table 4.14 where the panchromatic band is used with SFS techniques. There are basically two differences between the images used for these two experiments; therefore, the differences in the results must have originated from these. The very first difference between the images is in their resolutions. The panchromatic band has a resolution of 15 m while the ETM bands have 30 m resolution. Considering the fact that the resolution of the resampled image for SFS techniques is 25 m, one realizes that the effect of resampling, if there is any, should be in favour of the panchromatic image, while the results do not show that. The only other reason can be the effect of PCT. In the panchromatic mode only one image while in the ETM mode the result of six images are used for the SFS solutions. As the quality of the image has a direct effect on shading and therefore classification, it can be considered as the reason for having slightly better SFS solutions by using the ETM images.

4.5.2 SPOT and the Waterton Lake Area DTM from GTOPO30

The Waterton Lake area DTM is also extracted from GTOPO30 files to study the feasibility of improving DTM interpolation accuracy with SFS using much lower resolution DTM data set. The spacing of GTOPO30 DTM at latitude 50 degrees is about 598 m x 927 m. The extracted DTM is first projected to UTM map projection system and then resampled to 600 m resolution using cubic convolution interpolation method.

The registered 3-channel SPOT imagery is also resampled to 300 m resolution using a smoothing filter. Then, for the purpose of the albedo factor modification, all of the three bands are used with the Mahalanobis classifier to classify the pixels in three different categories representing vegetation, soil and water coverages. The PCT is applied to the imagery and the first band with 93.35% of energy is selected for modification of the albedo factor as well as further analysis.

The goal of this analysis is to enhance the accuracy of DTM interpolation which changes the resolution of the GTOPO30 DTM from 600 m to 300 m. In contrast to the previous cases, with GTOPO30 data set there are no known heights at the resolution of 300 m to check the quality of the SFS solutions with. However, one may think of using the 25 m resolution provincial DTM to check the results. But this cannot be of any help because if there is any systematic bias in any of the GTOPO30 or provincial DTMs it will bias the results, which obviously can not be separated from the others sources.

The solution to the quality control problem in this case is to use the direction of the slopes instead of the heights themselves. The slopes derived from both interpolation and the SFS solutions can be compared to the slopes derived from the provincial DTM. In this case, even if there is any systematic error in any of these DTM sets, it will be cancelled out in the process of deriving the slopes.

By applying the SFS algorithms to the data sets the heights and subsequently the slopes of the grid points in x and y directions are computed. On the other hand, the heights of the intermediate points are computed using cubic convolution interpolation. The interpolated heights are used to compute the slopes of the grid points in x and y

directions. Moreover, using the provincial DTM, the slopes between the points, heights of which are updated by SFS solutions are computed.

As mentioned above, the slopes of the terrain computed in x and y directions using the provincial DTM data set are considered as a reference for measuring the performance of the other methods. The mean difference between the orientation of the slopes computed from the provincial DTM and the interpolated DTM as well as the DTM from SFS solutions directions are computed. Along with the mean differences, standard deviations are computed too. Table 4.16, computed using the points for which there is a solution from SFS, summarizes the results.

Means in Table 4.16 show that both IGS and SFS solutions are biased. However, it is seen that SFS has provided a less biased solution than the IGS solution. The bias can be due to the asymmetry of the terrain at the resolution of 300 m. The Table clearly shows the effectiveness of the application of SFS to the DTM enhancement both in mean and standard deviation of the differences. Furthermore, Table 4.16 explicitly shows the differences between SFS algorithms with different kernels. It also demonstrates that the enhancement in the South-North component of the slope is much higher than in the East-West component. The reason goes back to the unsymmetrical resolution of the GTOPO30 data set which is about 600 m in the East-West and 930 m in the South-East direction, respectively. Obviously, resampling the DTM to a 300 x 300 m grid point is more accurate in the East-West direction.

Table 4.16: SPOT and Waterton Lake area GTOPO30 DTM (PD stands for Provincial Data)

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
PD – IGS Slope in E	Mean (Deg.)	2.01	2.35	2.41
	Std (Deg.)	3.98	4.16	4.23
PD – SFS Slope in E	Mean (Deg.)	1.85	1.53	1.46
	Std (Deg.)	3.14	2.54	2.20
Improvement in Std		21%	39%	48%
PD-IGS Slope in N	Mean (Deg.)	3.21	3.63	3.69
	Std (Deg.)	5.31	5.43	5.45
PD-SFS Slope in N	Mean (Deg.)	2.69	1.98	1.83
	Std (Deg.)	4.09	3.04	2.40
Improvement in Std		23%	44%	56%
Patches updated out of 1600		68%	83%	91%

4.5.3 IKONOS, IRS-1, Landsat 7 ETM+ and the Crowsnest Area DTM

The object under investigation in this section is a real terrain data set from Crowsnest/Blairmore area, southern Alberta, Canada with 25 metre spacing in UTM zone

11, NAD27 coordinate system. The goal is to enhance the accuracy of DTM interpolation process of 50 m to 25 m DTM (every other point in the original 25 m DTM is considered unknown) using SFS techniques. A 600 by 400 grid with more than 710 metre height difference is extracted from the four quadrants of NTS 82G08 and 82G09 DTM data file. Figure 4.10 shows two different snapshots of this data set.

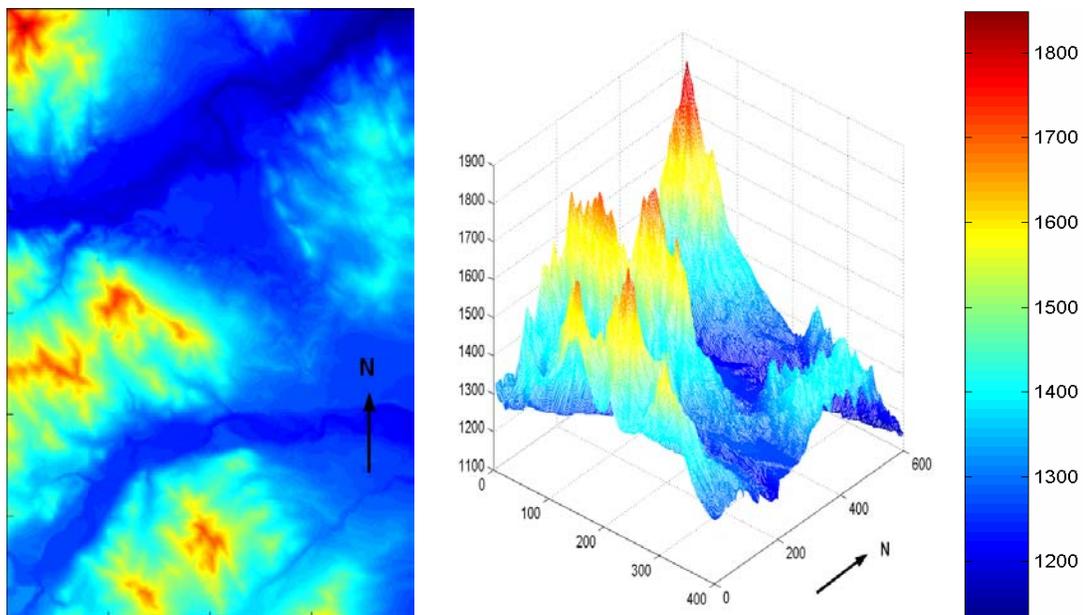


Figure 4.10: Two views of the Crowsnest area DTM data set

Three different satellite imageries including IKONOS 1-metre (panchromatic), IRS-1 5-metre (panchromatic) and Landsat 7 ETM+ (eight bands) corresponding to this region are made available to this research. The IKONOS and also IRS-1 were already orthorectified (using the 25 m DTM data set) and georeferenced to UTM zone 11 NAD83 coordinate system with the accuracy of 5 m. Because the DTM data set is referenced to

NAD27 datum, a datum transformation is needed to make the coordinate system of the imagery and the DTM data sets compatible with each other. The National Transformation software (NTv2) is used to transform the DTM grid data to the same coordinate system as the images. According to the report generated by the software, the accuracy of this transformation is within 0.5 m. Moreover, the IKONOS imagery was used to georeference the Landsat imagery, using an image-to-image registration method. Finally, all three imageries were resampled to 25 metres to match the resolution of the DTM. For IKONOS and IRS-1 imageries, the resampling method consists of passing the data through a smoothing filter, while for Landsat imagery, bilinear interpolation is used. Figure 4.11 shows the resampled IKONOS imagery of the region.

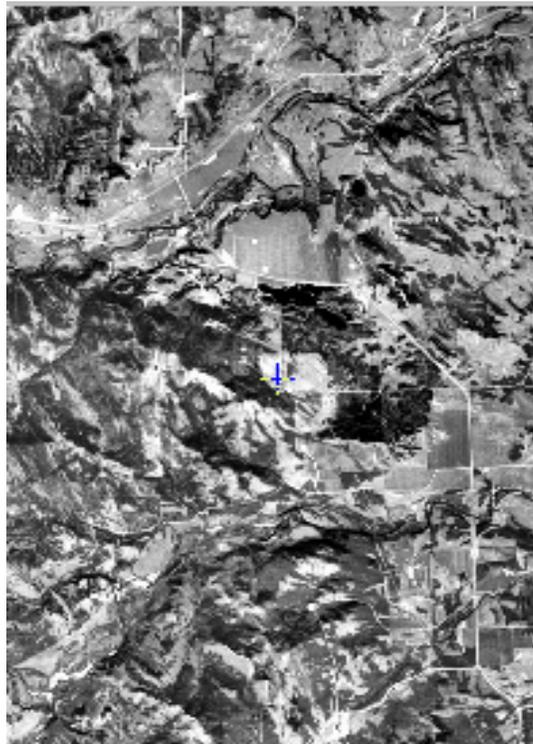


Figure 4.11: Resampled IKONOS imagery of the Crowsnest area

The resampled ETM bands of the Landsat 7 ETM+ imagery are used to assign each pixel of the image to one of three predefined classes, representing vegetation, soil and water. Based on the results of this classification, the IKONOS and IRS-1 imageries are classified too. Using PCI software, a principal component transformation are applied to the ETM bands of the Landsat 7 ETM+ imagery and the first channel with 93.16% energy is selected as the representative of the Landsat 7 ETM bands for further experiments. Based on the results of classification of the imageries, the mean albedo factors are computed for each image/class.

The difference between the original object and its interpolated solution (using cubic convolution interpolation method) as well as the corresponding SFS solutions are computed for imageries and the results are summarized in Tables 4.17 to 4-20.

Table 4.17: IKONOS and Crowsnest DTM data set

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.56	0.48	0.54
	Std (m)	9.36	10.13	10.56
Object – SFS	Mean (m)	0.41	0.53	0.46
	Std (m)	4.71	4.64	4.40
Improvement in Std		49.7%	54.2%	58.3%
Patches updated out of 58016		67%	74%	76%

Table 4.18: IRS-1 and Crowsnest DTM data set

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.61	0.65	0.62
	Std (m)	9.14	9.95	10.06
Object – SFS	Mean (m)	0.65	0.58	0.61
	Std (m)	4.62	4.62	4.22
Improvement in Std		49.4%	53.6%	58.1%
Patches updated out of 58016		68%	73%	76%

Table 4.19: Landsat panchromatic and Crowsnest DTM data set

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.59	0.54	0.61
	Std (m)	9.23	10.21	9.96
Object – SFS	Mean (m)	0.55	0.59	0.56
	Std (m)	4.73	4.81	4.26
Improvement in Std		48.7%	52.9%	57.2%
Patches updated out of 58016		64%	71%	74%

Table 4.20: Landsat ETM and Crowsnest DTM data set

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object – IGS	Mean (m)	0.63	0.58	0.53
	Std (m)	9.28	10.08	10.33
Object - SFS	Mean (m)	0.51	0.56	0.55
	Std (m)	5.09	5.13	4.90
Improvement in Std		45.1%	49.1%	52.6%
Patches updated out of 58016		64%	71%	73%

The results in Tables 4-17 to 4.20, computed with those points for which SFS has solution, show that the SFS has been able to improve the accuracy of interpolation up to the 58%. This is due to two facts which makes the improvement rate very considerable. First, the DTM represents very rough terrain. The bias seen in the means of the IGS and SFS solutions is believed to be due to the asymmetry of this rough terrain. Second, the interpolation method used here is cubic convolution which uses more information (16 grid points) for interpolation in comparison to the bilinear interpolation (4 grid points).

The Tables show, as well, that the differences between SFS solutions coming from IKONOS, IRS-1, and Landsat panchromatic imageries are more or less at the same level, while the solutions coming from the Landsat ETM imageries show slightly less

improvement in the results in comparison to the other images. This is believed to be due to process of resampling the ETM bands from 30 m to 25 m resolution. In other cases, the resampling has been from a higher resolution to a lower resolution.

The Tables clearly show that there is a difference between different SFS solutions. The SFS solution with standard variational method shows less improvement in comparison to the others, while the one with the sigmoidal kernel function has the best results. However, the differences between the solutions using re-descending and sigmoidal kernels are not so significant. It is believed that the interpolation method is sophisticated enough to approximate the terrain very closely.

4.6 Phase dependency of the SFS Solutions

Another question to answer is if there is any correlation between the SFS solutions and the location of the known points. In other words, it is required to investigate to see what will happen to the SFS solutions if different sets of points are assumed to be as the known points.

To study the effect of changing the location of the known points, two more experiments are done. The first one is with the Waterton Lake area DTM along with its 3-channel SPOT imagery and the second one is the Crowsnest area DTM together with its IKONOS image. Basically, the experiment corresponds to what is done in Sections 4.4.3 (Table 4.13) and 4.5.3 (Table 4.17) respectively, but in the new tests, the location of the known points are shifted with the amount of one grid unit in both directions. Tables 4.21 and 4.22 summarize the results for these two experiments.

Table 4.21: The Waterton Lake area DTM with SPOT imagery

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object - IGS	Mean (m)	0.06	0.03	0.07
	Std (m)	16.32	18.09	19.21
Object - SFS	Mean (m)	0.09	-0.05	-0.04
	Std (m)	9.96	8.32	8.26
Improvement in Std		39%	54%	57%
Patches updated out of 257049		87%	93%	94%

Table 4.22: IKONOS and Crowsnest DTM data set

Type of Kernel		Standard Variational Method (Quadratic)	Re-descending	Sigmoidal
Object - IGS	Mean (m)	0.61	0.55	0.59
	Std (m)	9.18	10.03	10.29
Object - SFS	Mean (m)	0.52	0.48	0.63
	Std (m)	4.60	4.60	4.30
Improvement in Std		48.9%	53.6%	58.7%
Patches updated out of 58016		66%	74%	77%

Comparing these two Tables with the Tables 4.13 and 4.17, it is seen that the results are practically the same. Even though that there is a slight discrepancy between the corresponding Tables, but they are not that much important from a practical point of view. The results show that among many other factors the SFS solutions are not dependent on the location of the known points. The quality of the image, the roughness of terrain, the classification of the image pixels are among those parameters which have a direct effect on the quality of the SFS solutions, but not the location of the known or unknown points.

4.7 Roughness of the Terrain

The analyses carried out in the previous sections show clearly that SFS could considerably improve the accuracy of DTM interpolation. As discussed above, the amount of improvement has a direct correlation with the type of kernel used in the SFS formulation as well as the roughness of the terrain. It was shown implicitly that the rougher the terrain is, the higher the improvement in accuracy is. This section tries to quantify this result.

The analysis in this section is based on the results mentioned in the third column of Table 4.13 where the SFS kernel is sigmoidal and the observed brightness in the corresponding SPOT imagery is normalized with the estimated albedo factor. The Waterton Lake provincial DTM is used in this analysis because in comparison to the Crowsnest area DTM it covers a larger area. Another advantage of this data set is that the surface terrain in this area covers terrain with different type of roughness whereas the

Crowsnest area DTM data set covers mostly rough terrain.

The statistics used here to classify the roughness of terrain are basically the variance of the shape index formulated in form of eq. 3.20. Based on these statistics, the terrain is classified into three different categories namely: 1) smooth, 2) rough, and 3) very rough. Table 4.23 refers to those points for which the SFS formulation with the sigmoidal kernel has provided a solution.

Table 4.23: Correlation between the roughness of terrain and rate of improvement

Roughness (w)	Smooth ($0.8 \leq w \leq 1$)	Rough ($0.2 \leq w < 0.8$)	Very Rough ($0 \leq w < 0.2$)
Improvement in Std	38%	55%	71%
Percent of points in each category	16%	48%	36%

Table 4.23 clearly shows that the relation between the rate of improvement in the accuracy of interpolation and the roughness of terrain in the SFS solution. It is worth mentioning that here smoothness is measured in terms of shape index variance which is basically a measure for variation of curvature. Obviously, for a flat terrain, the simple interpolation is good enough and there is no need to use SFS for interpolation accuracy improvement.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

The fact that DTMs play a vital role in many scientific, environmental, engineering as well as military applications shows how important an accurate and dense DTM is. Unfortunately, generating a dense and accurate DTM comes with the price of spending both time and money on the field measurements. Moreover, there are cases where in spite of the availability of these two factors, direct field measurements are not possible.

An alternative solution is to use interpolation techniques with the available global or local DTMs. Even though the interpolation is very fast and inexpensive, it has its own inherent problems especially with the selection of the mathematical model and required parameters. Meanwhile, the accuracy of interpolation is always questionable particularly when one is dealing with rough terrain.

This research study was an attempt to explore the feasibility of enhancing the DTM interpolation accuracy using the SFS algorithms with single (as opposed to stereo) satellite imageries. The motivation for this investigation is the availability of relatively inexpensive but up to date, multiresolution, multispectral single satellite imageries for almost the whole world.

The investigation included evaluating available SFS algorithms, studying their basic assumptions and also their simplified mathematical models. Moreover, the characteristics of both the available DTMs and satellite imageries within the realm of SFS application were assessed to improve the SFS models to best fit the interpolation

problem in hand. According to the results of these experiments, a strategy was proposed for using SFS to enhance the accuracy of DTM interpolation. The proposed SFS models were implemented in a software package and successfully tested.

In summary, the research clearly has both practical and theoretical significance for those who are in charge of densifying DTMs, since it not only presents a conceptual framework for accuracy enhancement of DTM interpolation, but also provides required implementation considerations. According to this dissertation, the following conclusions can be drawn.

5.1 Conclusions

- Shading in images is mainly due to interaction of the illumination, the reflecting property of the object, and the shape of the surface. In the context of this research, the illumination source is the Sun which has a known direction in the case of satellite missions. Assuming a Lambertian surface with known reflectance property, the only unknown in the image irradiance equation is the surface shape which is determined by the surface gradient components or equivalently the surface normal vector. This fact shows that the SFS problem is an ill-posed (an ill-posed problem is one that does not have any solution, does not have a unique solution, or has a solution that is very sensitive to the given data) one as for any pixel there is just one observed image brightness but with two unknown surface gradients then one unknown height
- According to the Lambertian model, the image intensity is the result of inner products of two vectors scaled by the albedo factor. The two vectors involved in the inner product process are the light source direction and the surface normal vectors.

Generally speaking, the image irradiance equation can provide only limited information about the surface shape because the image intensity depends only upon the surface normal, and all the surface normals on the ambiguity cone yield the same image intensity. The surface shape, therefore, is underconstrained by shading information.

- Since shading depends on the surface orientation, and hence first partial derivatives of the surface height, the lower spatial frequencies will have only a small effect on the image. Therefore, it should not be surprising that the largest errors in reconstruction of shape are in the lower-spatial-frequency components. In other words, SFS is good at recovering shape information corresponding to rapidly undulating surface features, while slow changes may be missed or misinterpreted.
- The assumptions required for the SFS scheme are very strong and it makes the robustness of the SFS solution questionable. The Lambert law is widely used in SFS algorithms for its simplicity, though no natural surface strictly obeys it, so some errors in shape arise. More importantly, many surfaces exhibit specular behaviour. Modeling such behaviour in SFS has not been very successful in practice. Furthermore, constant albedo may be a reasonable assumption just for manufactured parts, but obviously not for natural features. In case of this investigation, it was shown that an estimate for the magnitude of albedo factor could be obtained through classification of the pixels.
- Generally speaking, SFS algorithms compute the surface heights in two steps. In the first step, they come up with the surface gradient components, while in the second

step, these gradient components are used in an integration process to calculate the surface heights. In this research, the second step is used for two purposes: 1) applying the (discrete) surface integrability constraint, and 2) taking the known height grid points into account.

- Considering the first step, there are mainly two classes of solutions for SFS, global and local methods. In spite of their simplicity, local methods make very strong assumptions about the shape of the surface. Besides, they rely on the image brightness derivatives which can be very noisy. Moreover, they lack the mechanism to enforce global consistency. These characteristics make the local methods inappropriate for the specific problem of this investigation.
- On the other hand, global methods are more complex, but provide more accurate surface shapes. Three different global methods, i.e., characteristics strip, variational, and linear, were studied closely in this research project. It was shown that the characteristic methods are sequential processes in nature and suffer from noise in the image. Furthermore, because of their directional dependency, the errors are accumulated. These characteristics made these types of solutions neither efficient nor robust for the purpose of the problem of this project. Similarly, it was justified that linear methods are not appropriate for the application of this study as they suffer from linear approximation of the reflectance map. Last but not least, it was shown that among all the available methods, the variational solution is the best choice for the purpose of this investigation.
- It was shown that the standard variational method is suffering from the lack of a

mechanism to select a proper factor to weigh the smoothness constraint versus the brightness error. If the weight factor has a large value, SFS will provide with a very smooth surface resulting in losing the original surface details. On the other hand, if the weight factor has a small value, the SFS solution will encounter numerical problems and the numerical stability of the solution will be questionable. Generally speaking, choosing the right weight factor is not an easy task.

- Three modifications were applied to the SFS formulation to make it better suited to the problem of this investigation.
 1. The first modification is to use the error kernel function from robust statistics in formulation of the smoothness constraint. It was shown that robust statistics with the choice of sigmoidal error kernel prevents the SFS solution from losing the details and being too smooth especially when the surface under study is rough.
 2. The second modification is about the usage of the ambiguity cone and projecting the smoothed surface normals back onto the cone. This process frees the SFS solution from choosing any weight factor while guaranteeing the numerical stability of the SFS solution. This stability of the solution comes from the fact that at each iteration the normal vectors projected on the ambiguity cone are exact solutions of SFS model.
 3. The third modification is about the albedo factor. Using multispectral satellite imageries, one can apply classification methods to the imageries to get an idea of the albedo factor for each class. This estimated albedo factor can be used to normalize the image irradiance equation. Numerical experiments carried out in

Chapter 4 of this dissertation support the applicability of the idea.

- The motivation for using satellite imageries with SFS techniques comes from the fact that they are relatively inexpensive and easily accessible.. Considering today's technology, having access to up to date, multiresolution, and multispectral satellite imageries of any part of the world is taken for granted at present.
- The methodology adopted here is based on this assumption that the satellite imagery has one dyadic order better resolution than the original DTM data. Even though for the real data sets the exact one dyadic order is seldom the case, one can always change the resolution of the either DTMs or imageries using resampling techniques. Obviously, the quality of the resampled data depends on the resampling methods. Moreover, it goes without saying that down sampling is more reliable than up sampling and because of this fact, up sampling should be avoided whenever possible.
- The registration of the satellite imagery with the corresponding DTM has a key role in this special application of SFS. Obviously, if due to any reason the image grey levels do not agree the corresponding DTM heights, especially in a rough terrain, the computational procedures for the unknown terrain slopes will not converge.
- The image registration can be a problem with tilted satellite imagery or with large relief displacements (which are functions of ratios of the satellite height to the terrain height differences). Having access to a DTM as well as the specifications of the imaging platform, one can check if the orthorectification of the satellite imagery is sufficiently accurate for the intended purposes. If not, obviously, the methodology discussed in this research will fail to enhance the interpolation accuracy.

- The main advantage of the methodology developed in this research work is the fact that it needs only single imagery. It goes without saying that with stereo imagery, there is no need for SFS procedures, except possibly for auxiliary information extraction in areas where stereoscopic correlations fail.
- The computational effort grows rapidly with the project size, so if possible, parallel processing should be implemented. The developed SFS algorithm has been implemented in a way that it is appropriate for parallel processing. This goal has been achieved by coming up with the smallest independent patch size or computational unit. The computations involved in the developed procedure is linear in terms of the number of patches. Obviously, within each patch the computation effort is at least quadratic with the size of the patch.
- The smallest patch size has another importance as the patch size has a direct correlation with the smoothness of the solution. The bigger the patch size is the smoother the solution is. Considering the special application of this research, it was shown that a 7 x 7 point grid patch provides reasonable results.
- In general the SFS solution has slightly reduced the bias in comparison to the IGS solution.
- Considering all of these modifications, it was shown that SFS is able to improve the accuracy of DTM interpolation up to 38%, to 71% depending on the roughness of the terrain as well as the interpolation technique. Obviously the rougher the terrain is, the higher the rate of improvement is. In this research the roughness of terrain is measured by dispersion of the shape index.

- Questions of existence and uniqueness of solutions of the general SFS problem have still not been resolved entirely satisfactory. Assuming known reflecting properties of the surface and the light source vector, the question is if there is always a surface shape that will generate, under these conditions, any given (arbitrary) image brightness pattern? The answer is not known. It may be that there are patterns that could not have been produced as the result of shading on any three-dimensional object shape. Then in this case, the SFS problem has no solution. People often appear to be able to tell that a particular pattern is not due to shading but to spatial variations in the reflecting properties of the surface.
- The question of uniqueness is more difficult to answer. This is due to the fact that SFS is inherently ill posed. However, there will be some unique surface orientation for which the brightness is maximum (or a unique minimum, in some unusual cases). A point in the image where this maximal brightness is observed is called a singular point. These image points have particular importance since the surface orientation at the corresponding point on the surface is immediately known (provided, of course, that the reflectance map is given).

5.2 Contributions

As an ill-posed problem, SFS by itself is a quite challenging problem. In spite of the fact that many researchers have worked on this problem during the past 30 years, it is still an active research topic. Furthermore, the application of SFS in the field of geomatics and geoscience is quite new and practically still intact. In particular, in geomatics engineering, the topic of this research and the corresponding special application of SFS

are pioneering problems.

This thesis significantly contributed to the process of DTM densification which is a routine problem in all Earth related sciences and applications. It proposed both theoretical and practical frameworks for using SFS with single satellite imageries to increase the accuracy of DTM interpolation.

The proposed methodology is based on the analysis of different SFS solution approaches and their deficiencies within realm of the specific application of this research. Among different SFS methods, it was concluded that the general variational approach is best suited in this investigation. Meanwhile, the problem of this solution method, which is over smoothing the solution, was discussed in detail.

Specifically, the over smoothing problem in SFS solution apparently comes from the regularization factor which is used to weigh the smoothness constraint in variational method formulation versus the brightness error. Two solutions were proposed to overcome the over smoothness problem.

The first proposed solution was to use regularizing kernel from robust statistics as the proper tool to separate noise in data from any possible discontinuities. The second proposed solution in this respect is to use the ambiguity cone concept and map the intermediate smoothed normal vectors at each iteration back on the ambiguity cone. This not only makes the process of updating normal vector independent of the regularization factor, but also guarantees numerical stability as the solution at each update step is the solution to the SFS problem.

The last modification applied to the general variational SFS problem was the concept

of variable albedo factor. It was proposed to use a variable albedo factor which is basically obtained by classification of pixels through access to the multispectral satellite imageries.

From a practical point of view, the algorithm is developed such that parallel processing techniques can be implemented easily. This is achieved by introducing the smallest patch or computation unit which provides reasonable results. Using simulations, it was shown that the 7 x 7 point patch strategy is sufficient for optimal results. Considering the special application of this research where DTM and satellite imageries are involved, developing a parallel processing algorithm is a must.

5.3 Recommendations

- The SFS algorithm developed in this investigation should be tested with more data sets with different characteristics. Different data sets consist of different DTM grid points with different spacing and accuracy as well as different satellite imageries with different resolution and spectral characteristics. Moreover, terrain surfaces with other types of coverage than what is typically found in Canada is strongly recommended for testing purposes.
- The main deficiency of the experiments carried out in the previous chapters goes back to the lack of any solid reference object to test the quality of SFS solutions with. In computer science, the tests are done in a controlled laboratory environment where everything is known and under control. In the specific application of this research, the laboratory condition means having access to a terrain surface for which all the characteristics are known. Having access to a few test fields would make the analysis

results more robust and reliable.

- In this research the albedo factor needed for normalizing the image irradiance equation was estimated by using classification of the pixels. However, having access to the goniometer (see Section 2.2.5 for more detail), one can measure the reflectance distribution of various surfaces. It is suggested to test the performance of the developed framework using the measured albedo factor rather than the estimated one.
- In cases where direct measurement of the albedo factor is not possible one can still take advantage of existing landuse maps to come up with a better classification of pixels. Because of the correlation between classification quality and the process of albedo factor estimation, it is believed that using landuse maps will enhance the quality of the SFS solution.

REFERENCES

- Ames Research Center, 2002; <http://geo.arc.nasa.gov/esdstaff/landsat/wes2.html>. July 2002.
- Alberta Forestry, Lands and Wildlife, 1988; DEM Format Information. Land Information Services Division, Edmonton, Alberta, Canada, April
- Aronoff S., 1989; Geographic Information Systems: A Management Perspective. WDL Publications, Ottawa, Canada.
- Besl, P., Birch, J., Watson, L.T., 1988; Robust Window Operators, Proc. 2nd Int. Conf. Comput. Vision, Florida, pp. 591-600.
- Blais J.A.R., 1997; Digital Terrain Modeling and Applications. ENGO 653 Lecture Notes”, Dept. of Geomatics Eng., The University of Calgary.
- Blais, J.A.R., Gourdeau, D., Stewart, C., 1997; The Crown of the Continent Ecosystem Data Atlas Project – A Geomatics Perspective. Geomatica, Vol. 51, No. 3, pp. 235-245.
- Blinn J.F., 1977; Models of light Reflections for Computer Synthesised Pictures, Computer Graphics, Vol. 11, pp. 192-198.
- Bronstein, I.N. and Semendyayev K.A., 1998; Handbook of Mathematics. Springer, Berlin.
- Brooks M.J., Horn B.K.P., 1985; Shape from Shading, Proc. Int’l Joint Conf. Artificial Intelligence, pp. 932-936.
- Bruss A.R., 1979; Some Properties of Discontinuities in the Image Irradiance Equation, AI Memo 517, AI Laboratory, MIT, Cambridge, MA, April.

- Burrough P.A., McDonnell R., 1998; Principles of Geographical Information Systems for Land Resources Assessment. 2nd edition, Oxford, Clarendon Press.
- Collins M., 2002; Quantitative Remote Sensing, Principles, Processing, Analysis. Dept. of Geomatics Eng., University of Calgary.
- Cosandier D., 1999; Generating a Digital Elevation Model and Orthomosaics from Pushbroom Imagery. Ph.D. Thesis, Dept. of Geomatics Eng., University of Calgary.
- Courant R., D. Hilbert, 1953; Methods of Mathematical Physics, Vol. I, Wiley, New York, NY.
- DMA, 1986; Defense Mapping Agency product specifications for digital terrain elevation data (DTED) (2d ed.). Defense Mapping Agency Aerospace Center, St. Louis, Missouri, 26 p.
- DMA, 1990; Digitizing the future (3d ed.). Defence Mapping Agency, Washington, D.C., 105 p.
- Danko, D.M., 1992; The digital chart of the world. GeoInfo Systems, 2:29-36.
- Ehlers M., Edwards G., Bedard Y., 1989; Integration of remote sensing with GIS; a necessary evolution. Photogrammetric Engineering and Remote Sensing 55(11): 1619-27. A recent review of the relationship between the two technologies.
- El-Sheimy N., 1998; Digital Terrain Modeling. ENGO 573 Lecture Notes, Dept. of Geomatics Eng., The University of Calgary.
- Engeln-Mullges G., Uhlig F., 1996; Numerical Algorithms with C, Springer, Berlin.

- EOSAT, 1992; Landsat Technical Notes. Lanham, MD, EOSAT Inc.
- FGDC, 1998; Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy. Federal Geographic Data Committee, c/o USGS, Reston, VA, www.fgdc.gov/standards/documents/standards/accuracy/chapter3.pdf.
- Foresman, T.W., 1998; The History of Geographic Information Systems: Perspectives from the pioneers. Upper Saddle River, New Jersey, Prentice Hall PTR.
- Godard Space Flight Center, 2002; <http://rst.gsfc.nasa.gov/Front/overview.html>. July 2002.
- Goldstein, H., 1959; Classical Mechanics. Addison-Wesley, Reading, Massachusetts, U.S.A.
- Grimson W.E.L., 1984; Binocular Shading and Visual Surface Reconstruction, Computer Vision, Graphics and Image Processing, Vol. 28, pp. 19-43.
- Gugan, D.J., Dowman I.J., 1988; Topographic Mapping from Spot Imagery. Photogrammetric Engineering and Remote Sensing 54(10):1409-1414.
- Gültekin, A., Gökmen M., 1996; Adaptive Shape from Shading, ISCIS XI The Eleventh International Symposium on Computer and Information Sciences, pp. 83-92.
- Hampel, F.R., Rousseeuw, P. J., Ronchetti, E., Stahel, W. A., 1986; Robust Statistics: The Approach Based on Influence Functions, John Wiley, New York.
- Harrison V.G.W., 1945; Definition and Measurement of Gloss, Monograph, Printing & Allied Trades Research Association (PATRA)
- Horn B.K.P., 1970; Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View, Technical Report MAC-TR-79, Project MAC. Also

- Technical Report AI-TR-232, AI Laboratory, MIT, Cambridge, MA, November.
- Horn B.K.P., 1975; Obtaining Shape from Shading, *The Psychology of Machine Vision*, P.H. Winston Ed., McGraw-Hill, NY, pp. 115-155.
- Horn B.K.P., 1977; Understanding Image Intensities, *Artificial Intelligence*, Vol. 8, No. 11, pp. 201-231, April.
- Horn B.K.P., 1986; *Robot Vision*, MIT Press, Cambridge, MA; and McGraw-Hill, New York.
- Horn, B.K.P., 1990; Height and Gradient from Shading, *Int. J. Comput. Vision* 5(1), 37-75.
- Horn B.K.P., Brooks M.J., 1986; The Variational Approach to Shape from Shading, *Comput. Vision, Graph., Image Process.*, Vol. 33, pp. 174-208.
- Horn B.K.P., Woodham R.J., Silver W.N., 1978; Determining Shape and Reflectance using Multiple Images, A.I. Memo 490, MIT, Cambridge, MA, August.
- Huber, P.J., 1981; *Robust Statistics*, John Wiley, New York.
- Hwang T., Clark J.J., Yuille A.L., 1989; A depth recovery algorithm using defocus information. *IEEE Conference on computer vision and pattern recognition*, pp. 476-481.
- Ikeuchi K., 1981; Determination of Surface Orientation of Specular Surfaces Using the Photometric Stereo Method, Vol. 3, No. 6, pp. 661-669, November *IEEE PAMI*.
- Ikeuchi K., Horn B.K.P., 1981; Numerical Shape from Shading and Occluding boundaries, *Artificial Intelligence*, Vol. 17, pp. 141-185.
- Jensen, J.R., Cowen, D.C., 1999; Remote Sensing of Urban/suburban Infrastructure and

- Socio-Economic Attributes. Photogrammetric Engineering and Remote Sensing, Vol. 65, pp. 611-622.
- Jensen J. R., 1996; Introductory Digital Image Processing. 2nd Edition, Prentice Hall, New Jersey.
- Jensen J. R., 2000; Remote Sensing of the Environment. Prentice Hall, New Jersey.
- Jolion, J.M., Meer, P., Bataouche, S., 1991; Robust Clustering with Applications in Computer Vision, IEEE Trans. PAMI, Vol. 13, pp. 791-802.
- Kashyap, R.L., Eom, K.N., 1988; Robust Image Modeling Techniques with Their Applications, IEE Trans. ASSP, Vol. 36, No. 8, pp. 1313-1325.
- Kender J.R., 1979; Shape from texture. Proceedings, 6th International Journal Conference on Artificial Intelligence, Tokyo, pp. 562-570.
- Kimes D.S., 1983; Dynamics of Directional Reflectance Factor Distributions for Vegetations Canopies. Applied Optics, 22(9), pp. 1364-1372.
- Koenderink, J.J., van Doorn, A.J., 1992; Surface shape and Curvature Scales. Image and Vision Computing, Vol. 25, No. 9, pp. 1009-1026.
- Kok A.L., 1986; Preprocessing and Filtering of Digitally Correlated Gestalt Data. M.Sc. Thesis, Dept. of Civil Eng., Division of Surveying Eng., University of Calgary.
- Lee C.H., Rosenfeld A., 1985; Improved Methods of Estimating Shape from Shading Using the Light Source Coordinate System, AI, 26, pp. 125-143.
- Li, S.Z., 1991; Toward 3D Vision from Range Images: An Optimization Framework and Parallel Distributed Networks, PhD Thesis, University of Surrey, UK.
- Li, S.Z., 1995; On Discontinuity-Adaptive Smoothness Priors in Computer Vision, IEEE

- Trans. PAMI, Vol. 17, No. 6, June, pp. 576-586.
- Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., 2001; Geographic Information Systems and Science. John Wiley & Sons, LTD, New York.
- Maune, D.F., 2001; Digital Elevation Model Technologies and Applications: The DEM Users Manual. The American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland.
- Markham B., Barker J., Seiferth J., Kaita E., Barsi J., 2000; Landsat-7 ETM+ Radiometric Calibration. <http://www.eoc.csiro.au/millwshop/landsat7bm.pdf>.
- Marr, D.C., 1982; Vision, Freeman, San Francisco.
- Medioni J., Nevatia R., 1985, Segment-based stereo matching. *comput. Vision Graphics Image Process*, 31 July, 2-18.
- Meer, P., Mintz, D., Rosenfeld, A., Kim, D., 1991; Robust Regression Methods for Computer Vision: A Review, *Int. J. Comput. Vision*, Vol. 6, pp. 59-70.
- NASA, 2003; The Earth Science Enterprise Home Page, <http://www.earth.nasa.gov>.
- NCGIA, 2002a; www.ncgia.ucsb.edu/education/curricula/cctp/units/unit06/06_f.html. July 2002.
- NCGIA, 2002b; www.ncgia.ucsb.edu/education/curricula/cctp/units/unit09/09_f.html. July 2002.
- NOAA, 1975-1984; Landsat Data Users Notes. NOAA Landsat Customer Services, Sioux Falls, SD, 57198.
- Pentland A.P., 1982; Finding the illumination Direction, *Journal Optical Society of America*, Vol. 72, No. 4.

- Pentland A.P., 1984; Local Analysis of the Image, IEEE Trans. Pattern Anal. Mach. Recog., Vol. 6, No. 2, pp. 170-187.
- Pentland A.P., 1987; A new sense for depth of field. IEEE. Trans. Pattern Anal. Match. Intell. PAMI-9(4), 523-531
- Pentland A., 1988; Shape Information from Shading: A Theory about Human Preception, Proc. Int'l Conf. Computer Vision, pp. 404-413.
- Pentland A.P., 1989; Local Shading Analysis. Shape from Shading, MIT, pp. 473-479.
- Petrie G., Kennie T.J.M., 1991; Terrain Modeling in Surveying and Civil Engineering. McGraw-Hill, New York.
- Richards J. A., 1993; Remote Sensing Digital Image Analysis, An Introduction. 2nd Edition, Springer-Verlag, Berlin, Germany.
- Rindfleisch T., 1966; Photometric Method for Lunar Topography, Photogrammetric Eng., Vol. 32, pp. 262-276.
- Sandmeier S.R., 1999; Guidelines and Recommendations for the Use of the Sandmeier Field Goniometer for the NASA Stennis Space Center. Washington, DC, NASA.
- Sandmeier S., Itten K.I., 1999; Field Goniometer System (FIGOS) for Acquisition of Hyperspectral BRDF Data. IEEE Transactions on Geoscience & Remote Sensing, 37(2), pp. 978-986.
- Silver W., 1980; Determining Shape and Reflectance Using Multiple Images, S.M. Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, June.
- Simard R., Rochon G., and Leclerc A., 1988; Mapping with SPOT Imagery and

- Integrated Data Sets. Invited paper presented at the 16th congress of the International Society for Photogrammetry and Remote Sensing held July 1988 in Kyoto, Japan.
- SpaceImage, 2002; <http://www.spaceimage.com/default.htm>. July 2002.
- SRTM, 2003; <http://www.jpl.nasa.gov/srtm/index.html>. Shuttle Radar Topography Mission, March 2003.
- Stein, A., Werman, M., 1992; Robust Statistics in Shape Fitting, Proc. IEEE Conf. Comput. Vision Patt. Recogn., pp. 540-546.
- Tam A.P., 1990; Terrain Information Extraction from Digital SPOT Satellite Imagery. Dept. of Geomatics Eng., The University of Calgary.
- The Consortium for International Earth Science Information Network, 2002; <http://www.ciesin.org/TG/RS/RS-home.html>. July 2002.
- Tsai P.S., Shah M., 1994; Shape from Shading Using Linear Approximation, Image and Vision Computing J., Vol. 12, No. 8, pp. 487-498.
- Van Diggelen J., 1951; A Photometric Investigation of the Slopes and Heights of the Ranges of Hills in the Maria of the Moon, Bulletin of the Astronomical Instituted of the Netherlands, Vol. 11, July.
- USGS, 1993; Digital elevation models, data user guide 5. Reston, Virginia, 50 p.
- Waxman A.M., Gurumoothy N., 1988, Image flow theory: a framework for 3-D inference from time-varying imagery. Advances in Machine Vision (C Brown, Ed), pp. 165-224, Springer-Verlag, New York.
- Wei G-Q., Hirzinger G., 1997; Parametric Shape-from-Shading by Radial Basis Functions. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19,

- No. 4, April 1997.
- Witkin A.P., 1981; Recovering surface shape and orientation from texture. *Artif. Intell.* 17, 17-47.
- Wolf, P.R., 1983; *Elements of Photogrammetry with Air Photo Interpretation and Remote Sensing*. McGraw-Hill Book Company, New York.
- Woodham R.J., 1977; A Cooperative Algorithm for Determining Surface Orientation from Single View, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, MIT, Cambridge, MA, August 22-25, pp. 635-641.
- Woodham R.J., 1978a; Reflectance Map Technique for Analyzing Surface Defects in Metal Castings, Technical Report TR-457, A.I. Laboratory, MIT, Cambridge, MA.
- Woodham R.J., 1978b; Photometric Stereo: A Reflectance Map Technique for Determining Surface Orientation from Image Intensity, *Proceedings of SIPE's 22nd Annual Technical Symposium*, Vol. 155, pp. 136-143, August.
- Worthington P.L., Hancock, E.R., 1999; Needle Map Recovery Using Robust Regularizers. *Image and Vision Computing* 17 (1999), p.p. 545-557.
- Worthington P.L., Hancock, E.R., 2001; Surface Topography Using Shape from Shading. *Pattern Recognition* 34 (2001), p.p. 823-840.
- Zhang R., Tsai P.S., Cryer J.E., and M Shah, 1994; Analysis of Shape from Shading Techniques. *Proc. Computer Vision Pattern Recognition*, pp. 377-384.
- Zheng Q., Chellappa R., 1991; Estimation of Illumination Direction, Albedo, and Shape from Shading, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 7, pp. 680-702.

APPENDIX A

MINIMIZING FUNCTIONALS

One of the most important problems in calculus of variations is to find the curve for which some given integral is an extremum. Basically, the problem in the one-dimensional case is to find a path $y = y(x)$ between two values x_1 and x_2 such that the line integral of some function $f(x, y, y_x)$, where $y_x = dy/dx$, is an extremum. In mathematical terminology, the extremum of

$$I = \int_{x_1}^{x_2} f(x, y, y_x) dx \quad (\text{A.1})$$

is sought. Among different solutions to this problem, those paths are accepted for which $y(x_1) = y_1$ and $y(x_2) = y_2$ (Figure A.1).

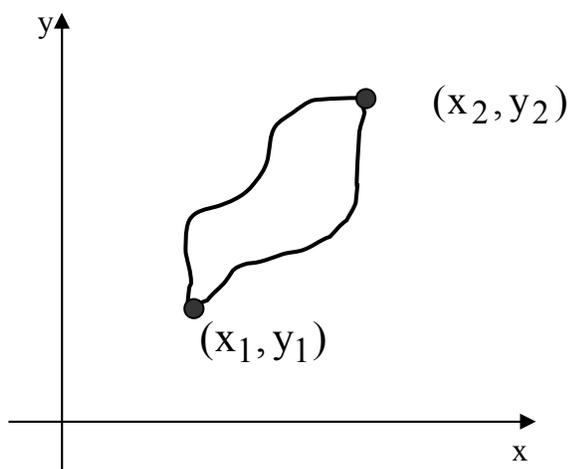


Figure A.1: Different paths in the one-dimensional extremum problem

The following parametric families of curves can represent the class of paths satisfying

the above mentioned condition

$$y(x, \alpha) = y(x, 0) + \alpha \eta(x) \quad (\text{A.2})$$

where $\eta(x)$ is any function of x which vanishes at $x = x_1$ and $x = x_2$, while for $\alpha = 0$ the curve would coincide with the path or paths giving an extremum for the integral.

Using this type of representation for y , one can rewrite eq. A.1 as

$$I(\alpha) = \int_{x_1}^{x_2} f(x, y(x, \alpha), y_x(x, \alpha)) dx. \quad (\text{A.3})$$

Obviously, to obtain the extremum of eq. A.3 the condition

$$\left(\frac{\partial I}{\partial \alpha} \right)_{\alpha=0} = 0 \quad (\text{A.4})$$

should be satisfied, where

$$\frac{\partial I}{\partial \alpha} = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} \frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial y_x} \frac{\partial y_x}{\partial \alpha} \right] dx. \quad (\text{A.5})$$

Considering the fact that $y_x = \partial y / \partial x$, one can rewrite the second term of the latter integral as

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y_x} \frac{\partial y_x}{\partial \alpha} dx = \int_{x_1}^{x_2} \frac{\partial f}{\partial y_x} \frac{\partial^2 y}{\partial x \partial \alpha} dx. \quad (\text{A.6})$$

Integrating the integral by parts yields

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y_x} \frac{\partial^2 y}{\partial x \partial \alpha} dx = \frac{\partial f}{\partial y_x} \frac{\partial y}{\partial \alpha} \Big|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d}{dx} \left(\frac{\partial f}{\partial y_x} \right) \frac{\partial y}{\partial \alpha} dx. \quad (\text{A.7})$$

From the conditions $y(x_1) = y_1$ and $y(x_2) = y_2$ imposed on the solution paths, it is clear that $\partial y / \partial \alpha = 0$ at x_1 and x_2 . Therefore, eq. A.7 can be rewritten as

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y_x} \frac{\partial^2 y}{\partial x \partial \alpha} dx = - \int_{x_1}^{x_2} \frac{d}{dx} \left(\frac{\partial f}{\partial y_x} \right) \frac{\partial y}{\partial \alpha} dx. \quad (\text{A.8})$$

This reduces eq. A.5 to

$$\frac{\partial I}{\partial \alpha} = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y_x} \right] \frac{\partial y}{\partial \alpha} dx. \quad (\text{A.9})$$

To obtain the extremum condition, eq. A.9 is multiplied by a differential $d\alpha$, and is evaluated at $\alpha = 0$. This results in

$$\left(\frac{\partial I}{\partial \alpha} \right)_{\alpha=0} d\alpha = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y_x} \right] \left(\frac{\partial y}{\partial \alpha} \right)_{\alpha=0} d\alpha dx. \quad (\text{A.10})$$

Recalling the fact that $\left(\left(\frac{\partial y}{\partial \alpha} \right)_{\alpha=0} d\alpha \right)$ represents some arbitrary variation of $y(x)$

obtained by variation of the arbitrary parameter α about its zero value, one concludes eq.

A.10 will be equal to zero, if

$$\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y_x} = 0. \quad (\text{A.11})$$

In other words, "I" in eq. A.1 is an extremum only for curves of $y(x)$ such that f satisfies the associated Euler differential equation, i.e., eq. A.11 [Goldstein, 1959]. In case of two-dimensional applications, as well as the other cases where f is a function of other parameters, the same logic can be used to find the extremum, and a corresponding Euler

equation must be satisfied.

Now let $F(x,y,z,z_x, z_y)$ measure the difference between a surface, $z(x,y)$, and a satisfactory solution at a point (x,y) . Assume that F is dependent not only on z , but also on the first partial derivatives z_x and z_y . Given that, the problem is to minimize the difference between the surface z defined over some region Ω in the plane, and the solution at the point (x,y) . In mathematical terms

$$I_1(z) = \iint_{\Omega} F(x, y, z, z_x, z_y) dx dy \quad (\text{A.12})$$

is an overall measure of error whose value is to be minimized. As it is seen the value of I_1 depends on the choice of the function z , and for this reason I_1 is termed a functional. Minimizing I_1 is again a problem in the calculus of variations.

As mentioned above, a fundamental result of the calculus of variations is that for the extrema of functionals, an associated Euler equation over the domain of interest must be satisfied. For the above form of the functional, the Euler equation is

$$F_z - \frac{\partial}{\partial x} F_{z_x} - \frac{\partial}{\partial y} F_{z_y} = 0. \quad (\text{A.13})$$

This is a necessary condition for the existence of an extremum, z , but not sufficient [Courant and Hilbert, 1953]. Note that local minima, global minima, global maxima, and inflection points are all examples of extrema.

It will prove useful to note the Euler equation corresponding to some other forms for F . In the event that F is dependent also on the second partial derivatives as in

$$I_2(z) = \iint_{\Omega} F(x, y, z, z_x, z_y, z_{xx}, z_{xy}, z_{yy}) dx dy \quad (\text{A.14})$$

the Euler equation expands to

$$F_z - \frac{\partial}{\partial x} F_{z_x} - \frac{\partial}{\partial y} F_{z_y} + \frac{\partial^2}{\partial x^2} F_{z_{xx}} + \frac{\partial^2}{\partial x \partial y} F_{z_{xy}} + \frac{\partial^2}{\partial y^2} F_{z_{yy}} = 0. \quad (\text{A.15})$$

Sometimes, a surface is parameterized not in terms of relative height, but in terms of surface normals. Two parameters are needed in this case. If the functions p and q are used to describe surface orientation and if the associated functional incorporates their first partial derivatives in x and y , the expression to be minimized then takes the form

$$I_3(p, q) = \iint_{\Omega} F(x, y, p, q, p_x, p_y, q_x, q_y) dx dy \quad (\text{A.16})$$

which has two corresponding Euler equations given by

$$F_p - \frac{\partial}{\partial x} F_{p_x} - \frac{\partial}{\partial y} F_{p_y} = 0 \quad (\text{A.17})$$

$$F_q - \frac{\partial}{\partial x} F_{q_x} - \frac{\partial}{\partial y} F_{q_y} = 0. \quad (\text{A.18})$$

In general, these constitute a pair of coupled partial differential equations in p and q . A pair of functions satisfying these equations will be necessary to generate an extremum of I_3 . As before, the extremum may be a minimum, maximum or inflection point [Horn and Brooks, 1986].

APPENDIX B

ENFORCING CONSTRAINTS

Sometimes one seeks a minimum of a functional subject to some independent constraint.

Suppose, for example, that it is required to minimize the following function:

$$I(z) = \iint_{\Omega} F(x, y, z, z_x, z_y) dx dy \quad (\text{B.1})$$

subject to the constraint that $g(x, y, z, z_x, z_y) = 0$. In this case one may use the

Lagrangian multiplier method in which the following augmented functional is minimized

$$I_1(z) = \iint_{\Omega} \left(F(x, y, z, z_x, z_y) + \mu(x, y)g(x, y, z, z_x, z_y) \right) dx dy \quad (\text{B.2})$$

where the Lagrangian multiplier μ is a function of x and y and must be treated as such when deriving the Euler equation.

Differentiating the functional with respect to $\mu(x, y)$, for a particular x and y , and setting the result equal to zero, one gets back the original constraint equation. This equation is required to help solve for μ , something one typically has to do in order to eliminate it from the Euler equation. At times, this may take some skill. More importantly, however, the equations that result often do not suggest convergent iterative schemes. In any case, a solution of the resulting Euler equation will have the property that $g(x, y, z, z_x, z_y) = 0$, with I_1 having an extremal value on the manifold $g(x, y, z, z_x, z_y) = 0$ [Horn and Brooks, 1986].

APPENDIX C

PENALTY TERMS BASED ON CONSTRAINTS

In view of the difficulties experienced when attempting to impose constraints exactly, one often considers an alternative method. In this approach, a penalty term derived from the constraint is employed. Thus one might rely on the Euler equation corresponding to the functional

$$I_2(z) = \iint_{\Omega} (F(x, y, z, z_x, z_y) + \lambda g(x, y, z, z_x, z_y)) dx dy \quad (C.1)$$

where λ is a scalar that aligns the arbitrary scales of F and g . Alternatively, it may be regarded as weighting of the relative importance of the components of the functional. It is, of course, necessary to square g if it is not already guaranteed to be non-negative over Ω for all functions z .

Solutions to the Euler equation for I_2 now specify surfaces that generate an extremal value of I_2 . However, these surfaces will not, in general, satisfy the constraint $g(x, y, z, z_x, z_y) = 0$ exactly. Rather, it will be the case that the value of g is small, along with the values of the other expressions being minimized. This is usually an acceptable compromise. More often than not, this approach proves more tractable than the Lagrangian method as there is no multiplier to be eliminated [Horn and Brooks, 1986].

APPENDIX D

MINIMIZING THE SFS COST FUNCTION

The SFS cost function which uses robust statistics with curvature as the constraint is as follows (eq. 4.11)

$$\iint_{\Omega} \left\{ (E(x, y) - \mathbf{N} \cdot \mathbf{L})^2 + \lambda \left(\Psi_w(\|\mathbf{N}_x\|) + \Psi_w(\|\mathbf{N}_y\|) \right) \right\} d\Omega \quad (\text{D.1})$$

where Ψ is the error kernel function with the first argument as the residual and w is the width of the kernel. Based on Appendix A, this cost function is of type $F(x, y, z, z_x, z_y)$ where the vectors \mathbf{N} , \mathbf{N}_x , and \mathbf{N}_y are z , z_x , and z_y . Therefore, applying the calculus of variations to the above mentioned SFS cost function would result in the following Euler equation

$$F_{\mathbf{N}} - \frac{\partial}{\partial x} F_{\mathbf{N}_x} - \frac{\partial}{\partial y} F_{\mathbf{N}_y} = 0. \quad (\text{D.2})$$

To compute the different components of the Euler equation, one should recall that

$$\mathbf{N}_x = \partial \mathbf{N} / \partial x, \quad \mathbf{N}_y = \partial \mathbf{N} / \partial y, \quad \|\mathbf{N}_x\| = \left(\sum \mathbf{N}_{xi}^2 \right)^{1/2}, \quad \text{therefore, } \partial \|\mathbf{N}_x\| / \partial \mathbf{N}_x = \mathbf{N}_x / \|\mathbf{N}_x\|,$$

$$\text{and } \partial \|\mathbf{N}_x\| / \partial x = \mathbf{N}_x \mathbf{N}_{xx} / \|\mathbf{N}_x\|, \quad \text{where } \mathbf{N}_{xx} = \partial^2 \mathbf{N} / \partial x^2.$$

Therefore, one can write:

$$\begin{aligned}
F_{\mathbf{N}} &= \partial F / \partial \mathbf{N} = -2(\mathbf{E} - \mathbf{N} \cdot \mathbf{L}) \mathbf{R}' \\
F_{\mathbf{N}_x} &= \partial F / \partial \mathbf{N}_x = \lambda \frac{\partial}{\partial \mathbf{N}_x} \Psi_w(\|\mathbf{N}_x\|) = \Psi_w'(\|\mathbf{N}_x\|) \frac{\mathbf{N}_x}{\|\mathbf{N}_x\|} \\
F_{\mathbf{N}_y} &= \partial F / \partial \mathbf{N}_y = \lambda \frac{\partial}{\partial \mathbf{N}_y} \Psi_w(\|\mathbf{N}_y\|) = \Psi_w'(\|\mathbf{N}_y\|) \frac{\mathbf{N}_y}{\|\mathbf{N}_y\|}.
\end{aligned} \tag{D.3}$$

Moreover, one can write:

$$\begin{aligned}
\frac{\partial}{\partial x} F_{\mathbf{N}_x} &= \frac{\partial}{\partial x} \left[\lambda \Psi_w'(\|\mathbf{N}_x\|) \frac{\mathbf{N}_x}{\|\mathbf{N}_x\|} \right] \\
&= \lambda \left[\Psi_w''(\|\mathbf{N}_x\|) \frac{\mathbf{N}_x \mathbf{N}_{xx}}{\|\mathbf{N}_x\|} \frac{\mathbf{N}_x}{\|\mathbf{N}_x\|} + \Psi_w'(\|\mathbf{N}_x\|) \frac{\mathbf{N}_{xx} \|\mathbf{N}_x\| - \mathbf{N}_x \mathbf{N}_x \mathbf{N}_{xx} \|\mathbf{N}_x\|^{-1}}{\|\mathbf{N}_x\|^2} \right] \\
&= \lambda \left[\|\mathbf{N}_x\|^{-1} \Psi_w'(\|\mathbf{N}_x\|) \mathbf{N}_{xx} + \left[\|\mathbf{N}_x\|^{-2} \Psi_w''(\|\mathbf{N}_x\|) - \|\mathbf{N}_x\|^{-3} \Psi_w'(\|\mathbf{N}_x\|) \right] (\mathbf{N}_x \mathbf{N}_{xx}) \mathbf{N}_x \right]
\end{aligned} \tag{D.4}$$

Analogously, $\frac{\partial}{\partial y} F_{\mathbf{N}_y}$ is:

$$\begin{aligned}
\frac{\partial}{\partial y} F_{\mathbf{N}_y} &= \frac{\partial}{\partial y} \left[\lambda \Psi_w'(\|\mathbf{N}_y\|) \frac{\mathbf{N}_y}{\|\mathbf{N}_y\|} \right] \\
&= \lambda \left[\|\mathbf{N}_y\|^{-1} \Psi_w'(\|\mathbf{N}_y\|) \mathbf{N}_{yy} + \left[\|\mathbf{N}_y\|^{-2} \Psi_w''(\|\mathbf{N}_y\|) - \|\mathbf{N}_y\|^{-3} \Psi_w'(\|\mathbf{N}_y\|) \right] (\mathbf{N}_y \mathbf{N}_{yy}) \mathbf{N}_y \right]
\end{aligned} \tag{D.5}$$

Substituting all terms into the Euler equation (eq. D.2) and considering

$$\begin{aligned}
\mathbf{N}_{xx} &\approx (\mathbf{N}_{i+1,j} - 2\mathbf{N}_{i,j} + \mathbf{N}_{i-1,j}) / \varepsilon^2, \\
\mathbf{N}_{yy} &\approx (\mathbf{N}_{i,j+1} - 2\mathbf{N}_{i,j} + \mathbf{N}_{i,j-1}) / \varepsilon^2
\end{aligned} \tag{D.6}$$

where ε is the spacing between cells, the fixed-point iterative scheme for updating the estimated normal at step $k+1$, using the previously available estimate from step k is:

$$\mathbf{N}_{i,j}^{k+1} = \frac{1}{2\alpha_{i,j}^k} \left(\bar{\mathbf{M}}_{i,j}^k + \frac{2\varepsilon^2}{\lambda} (\mathbf{E}_{i,j} - \mathbf{N}_{i,j}^k \cdot \mathbf{L}) \mathbf{R}'_{i,j}^n \right) \quad (\text{D.7})$$

where

$$\alpha_{i,j}^k = \Psi'_w \left(\|\mathbf{N}_{x,i,j}^k\| \right) \|\mathbf{N}_{x,i,j}^k\|^{-1} + \Psi'_w \left(\|\mathbf{N}_{y,i,j}^k\| \right) \|\mathbf{N}_{y,i,j}^k\|^{-1} \quad (\text{D.8})$$

and

$$\begin{aligned} \bar{\mathbf{M}}_{i,j}^k = & \left[\|\mathbf{N}_x\|^{-1} \Psi'_w \left(\|\mathbf{N}_x\| \right) (\mathbf{N}_{i+1,j}^k + \mathbf{N}_{i-1,j}^k) + \varepsilon^2 \left[\|\mathbf{N}_x\|^{-2} \Psi''_w \left(\|\mathbf{N}_x\| \right) - \right. \right. \\ & \left. \left. \|\mathbf{N}_x\|^{-3} \Psi'_w \left(\|\mathbf{N}_x\| \right) \right] (\mathbf{N}_x \mathbf{N}_{xx}) \mathbf{N}_x \right] + \\ & \left[\|\mathbf{N}_y\|^{-1} \Psi'_w \left(\|\mathbf{N}_y\| \right) (\mathbf{N}_{i,j+1}^k + \mathbf{N}_{i,j-1}^k) + \varepsilon^2 \left[\|\mathbf{N}_y\|^{-2} \Psi''_w \left(\|\mathbf{N}_y\| \right) - \right. \right. \\ & \left. \left. \|\mathbf{N}_y\|^{-3} \Psi'_w \left(\|\mathbf{N}_y\| \right) \right] (\mathbf{N}_y \mathbf{N}_{yy}) \mathbf{N}_y \right] . \end{aligned} \quad (\text{D.9})$$

APPENDIX E

DATA FUSION IN SPECTRAL DOMAIN

Principal Component Transformation (PCT) is a tool for the decorrelation of multispectral remote sensing data in the spectral domain. This section gives a brief review on the mathematical background of the transformation.

The multispectral or multidimensional nature of remote sensing data can be accommodated by constructing a vector space with as many axes or dimensions as these are spectral components associated with each pixel. In other words, the components of each vector in this vector space, say x_i , are the individual spectral responses in each band.

While the mean vector is useful to define the average or expected position of the pixels in a multispectral vector space, it is of value to have the covariance matrix available too. The covariance matrix is one of the most important mathematical concepts in the analysis of multispectral remote sensing data as it shows not only the scatter or spread of data, but also the correlation between data.

Most often the data in different spectral bands are highly correlated. Therefore, it is very fundamental to show the multispectral data in a vector space without correlations. The PCT is used to rotate the coordinate system in a way that the correlations vanish.

Assuming \mathbf{y} is the vector describing the pixels in the new uncorrelated coordinate system \mathbf{Y} which corresponds to the vector \mathbf{x} in the original vector space \mathbf{X} . Mathematically speaking, it is desired to find a linear transformation G of the original

coordinate system, such that

$$\mathbf{y} = \mathbf{G}\mathbf{x} \quad (\text{E.1})$$

subject to the constraint that the covariance matrix of the pixel in \mathbf{Y} is diagonal. In \mathbf{Y} , the covariance matrix is defined by

$$\mathbf{C}_y = E\left\{(\mathbf{y} - \mathbf{m}_y)(\mathbf{y} - \mathbf{m}_y)^T\right\} \quad (\text{E.2})$$

where \mathbf{C}_y is the covariance matrix, E is the expectation operator, and \mathbf{m}_y is the mean vector expressed in terms of the \mathbf{y} coordinates. It can be seen that

$$\mathbf{m}_y = E\{\mathbf{y}\} = E\{\mathbf{G}\mathbf{x}\} = \mathbf{G}E\{\mathbf{x}\} = \mathbf{G}\mathbf{m}_x \quad (\text{E.3})$$

where \mathbf{m}_x is the data mean in \mathbf{X} . Therefore, using eq. E.2 one can rewrite eq. E.3 as

$$\mathbf{C}_y = E\left\{(\mathbf{G}\mathbf{x} - \mathbf{G}\mathbf{m}_x)(\mathbf{G}\mathbf{x} - \mathbf{G}\mathbf{m}_x)^T\right\} = \mathbf{G}E\left\{(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^T\right\}\mathbf{G}^T = \mathbf{G}\mathbf{C}_x\mathbf{G}^T \quad (\text{E.4})$$

where \mathbf{C}_x is the covariance of the pixel data in \mathbf{X} . For \mathbf{C}_y to be diagonal, \mathbf{G} can be recognized as the transposed matrix of eigenvectors of \mathbf{C}_x , provided \mathbf{G} is an orthogonal matrix (by assumption that the eigenvalues are different). As a result, \mathbf{C}_y can then be identified as the diagonal matrix of eigenvalues of \mathbf{C}_x . In other words

$$\mathbf{C}_y = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_k] \quad (\text{E.5})$$

where λ_i $i = 1, 2, \dots, k$ are the eigenvalues of \mathbf{C}_x with dimensionality k . The elements of \mathbf{C}_y is arranged somehow that $\lambda_1 > \lambda_2 > \dots > \lambda_k$ so that the data exhibit maximum variance in y_1 , the next largest variance in y_2 and so on. The principal components transform defined by eq. E.1 subject to the diagonal constraint of eq. E.4 is also known as the Karhunen-Loève or Hotelling transform.

APPENDIX F

CLASSIFICATION

It is well known that the reflectance characteristics of the Earth's surface materials are not only different but also vary as a function of the wavelength of the light source. Based on this fact, different algorithms have been developed for both supervised and unsupervised classifications using multispectral imageries. As supervised classification is used in this research, it is explained here.

Generally speaking, supervised classification is the procedure most often used for quantitative analysis of remote sensing image data. It is based on using suitable algorithms to label the pixels in an image as representing particular ground cover types, or classes. Irrespective of the algorithm chosen, the essential practical steps are

- deciding on the set of ground cover types into which the image is to be segmented,
- choosing representative or prototype pixels (training data) from each of the desired set of classes,
- using the training data to estimate the parameters of the particular classifier algorithm, and
- using the trained classifier to label or classify the other pixels in the image.

There are many supervised classification methods among which maximum likelihood classification is the most common one used with remote sensing image data. The general idea in maximum likelihood classification is to divide the image into a definite number of different spectral classes. Most often, it is done based on the experience and knowledge of the operator or by using other sources of information like

landuse maps or aerial photographs. The probability distributions of these classes are assumed to be normal because of the simplicity and well-known properties of the normal distribution. The parameters such as means and variances are estimated using the training data set. Having this information, the last stage is to find out the probability that a pixel at a specific location belong to a specific class.

The effectiveness of maximum likelihood classification depends upon reasonably accurate estimation of probability distribution functions of each class. Obviously, this in turn depends on the number and quality of the training data set. In cases where the number of training data set is limited, the maximum likelihood classifier loses its effectiveness and it is better to use other classifiers which depend mostly on the mean positions of the classes rather than the covariance information.

Among different choices, the Mahalanobis classifier is the one which is very close to maximum likelihood classifier. Nevertheless, not only it is faster, but it also puts more emphasis on the mean of the classes, while it still uses the covariance information. Assuming \mathbf{x} is the vector of pixel values in a spectral space, \mathbf{C} is the covariance matrix of the classes (all class covariances are assumed equal), and \mathbf{m}_i is the mean of i th class, the Mahalanobis classifier is defined as

$$d(\mathbf{x}, \mathbf{m}_i)^2 = (\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}_i) \quad (\text{F.1})$$

where d is the distance, and T is the transpose.

Using the Mahalanobis distance classifier, obviously, one can calculate the distance between any pixel and the predefined classes. The pixel under study belongs to the class which has the shortest distance.

APPENDIX G

CHARACTERISTICS OF THE REMOTELY SENSED IMAGERIES AND DTM DATA SETS USED FOR NUMERICAL EXPERIMENTS

The real data sets used to test the performance of the developed algorithms consist of both satellite imageries as well as DTM. The first part of this appendix discusses the characteristics and details of satellites and multispectral sensors, images of which have been used in this research. This includes Landsat MSS and TM, Landsat 7 ETM+, SPOT, IKONOS, and IRS. The second part explains the specifications of the real DTM data sets used in the testings. In particular, the technical specifications of the Alberta provincial DTM and GTOPO30 data sets are discussed here.

G.1 Satellite Imageries

As discussed in Chapter 2, the multispectral radiometers are classified into three different groups: 1) multispectral imaging using discrete detectors and scanning mirrors, 2) multispectral imaging using linear arrays, and 3) multispectral imaging using linear and area arrays. The first two categories are studied here in more detail as the imageries of these satellite sensors are used in the numerical experiments.

G.1.1 Multispectral Imaging Using Discrete Detectors and Scanning Mirrors

In spite of the fact that the technology of these sensors goes back to mid-1960s, there are still new sensors which utilize it. Among many, one can mention Landsat Multispectral Scanner (MSS), Landsat Thematic Mapper (TM), and Landsat 7 Enhanced Thematic

Mapper Plus (ETM+). The Landsat program is the United States' oldest land-surface observation satellite system which has been collecting data since 1972 [NOAA, 1975-1984].

Landsats 1 to 3 were launched into circular orbits with inclination of 99 degrees at a nominal altitude of 919 km. The satellites orbited the Earth once every 103 minutes with a swath width of 185 km. The Sun-synchronous satellites had a temporal resolution of 18 days and crossed the equator at approximately the same local time (9:30 to 10:00 AM) on the illuminated side of the Earth. Landsats 4 and 5 have circular orbits with an inclination of 98.2 degrees at a nominal altitude of 705 km. With the swath width of 185 km, they orbit the Earth once every 98.9 minutes. Their temporal resolution is 16 days. The Sun-synchronous Landsats 4 and 5 cross the equator at 9:45 AM. Table G.1 summarizes the other characteristics of MSS and TM sensors [Richards, 1993].

Table G.1: Characteristics of Landsat MSS and TM sensors

	Band	Spectral Resolution (μm)	IFOV at Nadir (m)	Dynamic Range (bit)
Landsat (MSS)	1*	0.5-0.6	79 x 79	6
	2*	0.6-0.7	79 x 79	6
	3*	0.7-0.8	79 x 79	6
	4*	0.8-1.1	79 x 79	6
	6	10.4-12.6	240 x 240	6
Landsat 4, 5 (TM)	1	0.45-0.52	30 x 30	8
	2	0.52-0.60	30 x 30	8
	3	0.63-0.69	30 x 30	8
	4	0.76-0.90	30 x 30	8
	5	1.55-1.75	30 x 30	8
	6	10.40-12.5	120 x 120	8
	7	2.08-2.35	30 x 30	8

* These bands were originally numbered 4, 5, 6, and 7 respectively, because of Return-A Beam-Vidicon (RBV) sensor is also onboard the satellites recording energy in three bands labelled 1, 2, and 3.

G.1.1.1 Landsat Multispectral Scanner (MSS)

All Landsat satellites were equipped with a MSS multi-detector sensor. This optical-mechanical sensor scans the terrain perpendicular to the flight direction using a mirror. The terrain's EM energy received by the mirror is focused onto discrete detectors. Six parallel detectors sensitive to four spectral bands view the terrain simultaneously. The mirror oscillates through an angular displacement of $\pm 5.78^\circ$ off-nadir which in turn results in 185 km swath width. The MSS sensors scan each line across-track in a west-east direction, while the southward orbit of the satellites provides the along-track progression (Figure G.1).

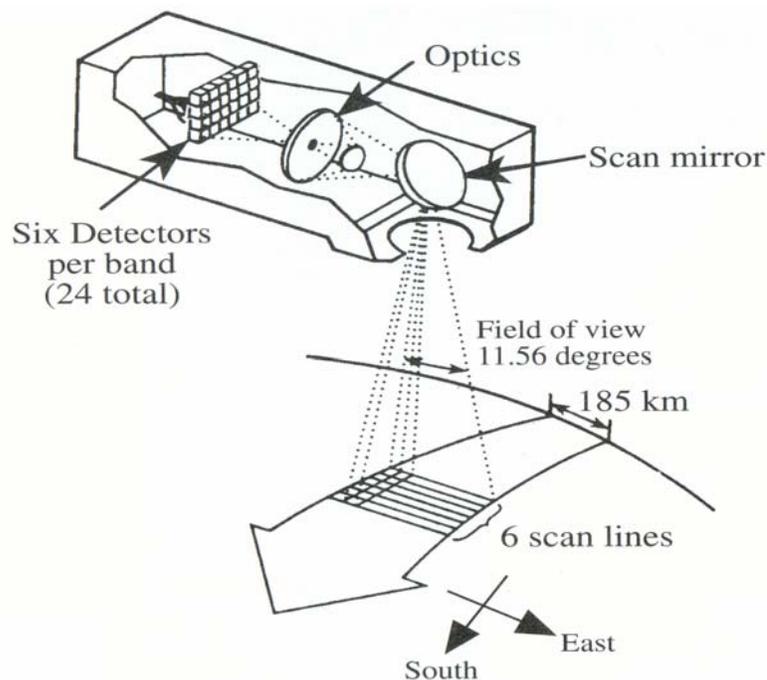


Figure G.1: Landsat MSS [Jensen, 2000]

G.1.1.2 Landsat Thematic Mapper (TM)

Like MSS sensors, TM is an optical-mechanical scanner sensor that records EM energy in the visible, reflective-infrared, middle-infrared, and thermal infrared regions of the EM spectrum. However, Table G.1 shows that TM has a higher spatial, spectral, temporal, and radiometric resolution than MSS sensors. Another important difference between TM and MSS sensors is in the choice of their bandwidths. The original MSS bandwidths were chosen based on their general use in vegetation and geological studies, while the refined bandwidths of TM were chosen after a long time study on behavior and characteristics of different elements affecting the imaging process and application.

EOSAT [1992] explained the detailed description of design and performance of the TM sensor.

G.1.1.3 Landsat 7 Enhanced Thematic Mapper Plus (ETM+)

NASA [2003] explains that Landsat 7 ETM+ has been designed to achieve three main goals. These objectives are:

- Maintain data continuity by providing data that are consistent in terms of geometry, spatial resolution, calibration, coverage and spectral characteristics with previous Landsat data.
- Generate and periodically refresh a global archive of substantially cloud-free, sunlit landmass imagery; and
- Continue to make Landsat-type data available to U.S. and international users at the cost of fulfilling user requests and to expand the use of such data for global change research and commercial purposes.

Landsat 7 ETM+ is in an orbit with similar characteristics of Landsats 4 and 5 orbits. The sensor used in Landsat 7 ETM+ is basically the same optical-mechanical TM scanner used in Landsats 4 and 5 but with some notable improvements. Apart from better onboard radiometric calibration equipment and facilities [Markham et al, 2000] in addition to a higher spatial resolution for thermal infrared (60x60 metre instead of 120x120 metre), the most notable enhancement is the addition of a new 15 x 15 metre resolution detector in panchromatic band. Table G.2 summarizes the characteristics of Landsat 7 ETM+ sensor [Jensen, 2000].

Table G.2: Landsat 7 ETM+ characteristics

Band	Spectral Resolution (μm)	IFOV at Nadir (m)	Dynamic Range (bit)
1	0.45-0.515	30 x 30	8
2	0.525-0.605	30 x 30	8
3	0.630-0.690	30 x 30	8
4	0.750-0.900	30 x 30	8
5	1.55-1.75	30 x 30	8
6	10.40-12.50	60 x 60	8
7	2.08-2.35	30 x 30	8
8 (pan)	0.52-0.90	15 x 15	8

Based on NASA [2003] Landsat 7 data are available to the user with several different levels of preprocessing including:

- Level 0R Data Products which consist of reformatted but not radiometrically corrected raw data and their corresponding metadata,
- Level 1R Data Products which consist of radiometrically corrected data and metadata describing calibrating parameters, payload correction data, mirror scan correction data, a geolocation table, and internal calibration lamp data, and
- Level 1G Data Products which consist of radiometrically and geometrically corrected data resampled to a user-specified map projection system and metadata describing the calibration parameters and a geolocation table.

G.1.2 Multispectral Imaging Using Linear Arrays

Linear array or so-called pushbroom sensors use very sensitive diodes staring constantly at the ground to record the reflected EM energy from the Earth's surface. In comparison to MSS or TM sensors, pushbrooms are more accurate as there is no scanning mirror

which in return results in a longer dwelling time for a specific part of the ground surface. SPOT, IRS, and IKONOS are examples of remote sensing imagers using pushbroom sensors and are explained in more details in the following subsections.

G.1.2.1 SPOT Earth Observation Satellites

As the pioneer in using pushbroom technology, the first SPOT satellite, SPOT 1, was launched in 1986. Since then five more satellites, SPOT 2, 3, 4, 5, and 5a were launched with the last one in 2002. From the very first day of operation these satellites have played an essential role as a reliable source of high-resolution Earth resources information. SPOT 1, 2, and 3 each carries two identical high-resolution visible (HRV) sensors through which stereoscopic view is possible. The HRV sensors receive the EM energy reflected back from the surface of the Earth and then project them onto two arrays of sensitive diodes or charge-coupled-devices (CCDs) where each CCD array consists of 6000 detectors arranged linearly. The linear array pushbroom sensor images a complete line of the Earth's surface in the cross-track direction as the satellite progresses downtrack (Figure G.2).

The sensors operate in two modes in the visible and reflective infrared regions of the EM spectrum. The first mode is the panchromatic mode which corresponds to observation over a broad spectral band, while the second mode is the multispectral mode corresponding to observation in three relatively narrow spectral bands.

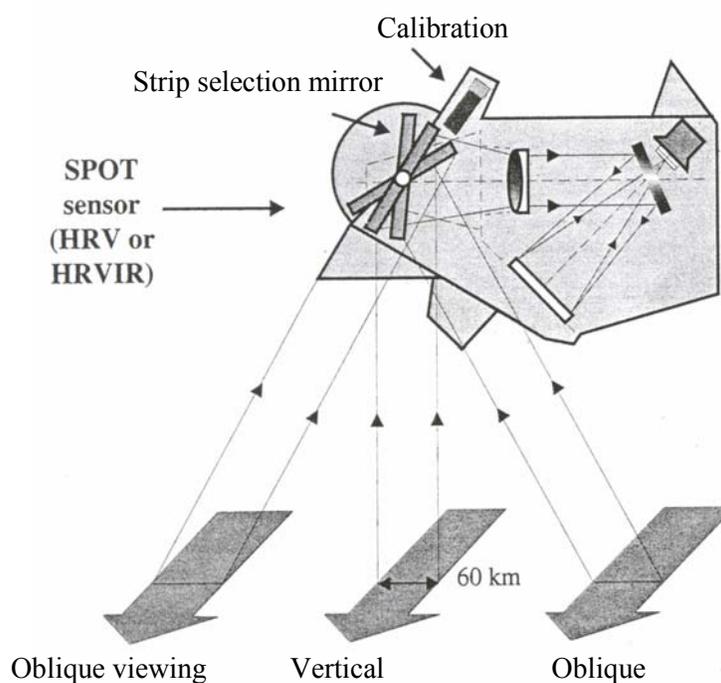


Figure G.2: The HRV sensor of SPOT [Jensen, 2000]

The satellites operate in a Sun-synchronous orbit with inclination of 98.2 degrees at an altitude of 832 km. They collect information with the nominal temporal resolution of 26 days and swath width of 60 km. Table G.3 provides the other characteristics of HRV sensor used in SPOT 1, 2, and 3 [Richards, 1993].

Table G.3: Some characteristics of HRV sensor in SPOT 1, 2, and 3

Band	Spectral Resolution (μm)	IFOV at Nadir (m)	Dynamic Range (bit)
1	0.50-0.59	20 x 20	8
2	0.61-0.68	20 x 20	8
3	0.79-0.89	20 x 20	8
Pan	0.51-0.73	10 x 10	8

In 1998 SPOT 4 was launched into an orbit with a similar specification as SPOT 1, 2,

and 3 orbits. However, in addition to the HRV sensors, SPOT 4 has another completely independent sensor on board called vegetation which is used for small scale vegetation, global change, and oceanographic studies. The swath width of the vegetation sensor is about 2250 km which enables this sensor to sweep the whole Earth's surface once per day. Additionally, each HRV sensor in SPOT 4 is also sensitive to short-wavelength infrared (SWIR) band and because of that they are referred to as HRVIR 1 and HRVIR 2. Table G.4 summarizes the characteristics of SPOT 4 sensors [Jensen, 2000].

Table G.4: HRVIR and vegetation sensors of SPOT 4

	Band	Spectral Resolution (μm)	IFOV at Nadir	Dynamic Range (bit)
HRVIR	1	0.50-0.59	20 x 20 m	8
	2	0.61-0.68	20 x 20 m	8
	Pan	0.61-0.68	10 x 10 m	8
	3	0.79-0.89	20 x 20 m	8
	SWIR	1.58-1.75	20 x 20 m	8
Vegetation	0	0.43-0.47	1.15 x 1.15 km	8
	2	0.61-0.68	1.15 x 1.15 km	8
	3	0.78-0.89	1.15 x 1.15 km	8
	SWIR	1.58-1.75	1.15 x 1.15 km	8

SPOT 5 and 5A, launched in 2000 and 2002 respectively, are in a similar orbits as SPOT 1, 2, 3, and 4. However, they carry a new imaging instrument called High Resolution Geometry (HRG). The HRG sensor consists of 12000-point linear sensor array which are made from very high dimensional stability materials. Table G.5 contains other characteristics of onboard SPOT 5 and 5A sensors [Jensen, 2000].

Table G.5: Characteristics of SPOT 5 and 5A sensors

	Band	Spectral Resolution (μm)	IFOV at Nadir	Dynamic Range (bit)
HRG	1	0.50-0.59	10 x 10 m	8
	2	0.61-0.68	10 x 10 m	8
	3	0.79-0.89	10 x 10 m	8
	Pan	0.51-0.73	5 x 5 m 2.5 x 2.5 m	8
	SWIR	1.58-1.75	20 x 20 m	8
Vegetation	0	0.43-0.47	1.15 x 1.15 km	8
	2	0.61-0.68	1.15 x 1.15 km	8
	3	0.78-0.89	1.15 x 1.15 km	8
	SWIR	1.58-1.75	1.15 x 1.15 km	8

G.1.2.2 Indian Remote Sensing (IRS) Satellites

Similar to SPOT, the sensors onboard the IRS satellites use the linear array sensor or so called pushbroom technology. The IRS-1A and 1B satellites were launched into Sun-synchronous orbits with inclination of 99.5 degrees and altitude of 904 km in 1988 and 1991, respectively. Both of them are equipped with Linear Imaging Self-scanning Sensors (LISS-I and LISS-II) which operate in four different spectral region each with swath width of 146 to 148 km. The temporal resolution of the satellites is 22 days. Table G.6 shows other characteristics of these satellites [Jensen, 2000].

Table G.6: Some characteristics of the IRS-1A and IRS-1B sensors

Band		Spectral Resolution (μm)	IFOV at Nadir (m)	Dynamic Range (bit)
1	LISS-I	0.45-0.52	72.5 x 72.5	8
	LISS-II		36.25 x 36.25	
2	LISS-I	0.52-0.59	72.5 x 72.5	8
	LISS-II		36.25 x 36.25	
3	LISS-I	0.62-0.68	72.5 x 72.5	8
	LISS-II		36.25 x 36.25	
4	LISS-I	0.77-0.86	72.5 x 72.5	8
	LISS-II		36.25 x 36.25	

IRS-1C and 1D were launched into Sun-synchronous orbits in 1995 and 1997, respectively. They revolve around the Earth in an orbital plane with inclination of 98.96 degrees and altitude of 817 km. The satellites carry three different sensors each, i.e., the LISS-III multispectral sensor, a panchromatic sensor, and a Wide Field Sensor (WiFS). The sensors have different characteristics which are summarized in Table G.7 [Jensen, 2000].

Table G.7: IRS-C and IRS-D characteristics

Band		Spectral Resolution (μm)	IFOV at Nadir (m)	Swath width (km)	Temporal Resolution (days)
LISS-III	1	-	-	-	24
	2	0.52-0.59	23 x 23	142	
	3	0.62-0.68	23 x 23	142	
	4	0.77-0.86	23 x 23	142	
	5	1.55-1.70	20 x 20	148	
Pan		0.50-0.75	5.8 x 5.8	70	5
WiFS 1		0.62-0.68	188 x 188	774	
WiFS 2		0.77-0.86	188 x 188	774	

G.1.2.3 Space Imaging, Inc. IKONOS

IKONOS is one of the first commercial satellites to use pushbroom technology to acquire very high-resolution data from the Earth's surface. Launched into a Sun-synchronous orbit with the altitude of 681 km, IKONOS started its mission in 1999. With a temporal resolution of less than 3 days and a swath width of 11 km, IKONOS is able to collect data from both cross-track and along-track in five different spectral regions. Table G.8 shows the characteristics of the sensor used in IKONOS [Jensen, 2000].

Table G.8: Sensor characteristics of IKONOS

Band	Spectral Resolution (μm)	IFOV at Nadir (m)	Dynamic Range (bit)
1	0.45-0.52	4 x 4	8
2	0.52-0.60	4 x 4	8
3	0.63-0.69	4 x 4	8
4	0.76-0.90	4 x 4	8
Pan	0.45-0.90	1 x 1	8

G.2 DTM Data Sets

Along with satellite imageries and aerial photographs, a number of real DTM data sets have been used to test the application of SFS solutions to the enhancement of interpolation accuracy. The real DTM data sets used in this research are basically from two different sources, i.e., Alberta provincial DTM and GTOPO30. In this section, the general characteristics of these data sets are discussed and explained.

G.2.1 Alberta Provincial DTM

The Alberta Environmental Protection Agency has originally generated the 100 m grid points as well as the 25-100 m feature points and breaklines using 1:60,000 aerial photographs and stereo-model procedures. However, these data are later used to generate the 25 m grid DTM using interpolation methods. This interpolated 25 m grid DTM is used to produce the contours on the 1:20,000 topographic maps of the province.

The horizontal datum of the Alberta provincial DTM is NAD27 while the vertical datum is CGVD28. UTM zone 11 (central meridian of 117° West) and 12 (central meridian of 111° West) are used as the map projection system of the DTM. However, the false easting of those points falling in UTM zone 12 is 1,500,000 m. The vertical accuracy of the DTM is estimated to be within 3.0 m at the 90% confidence level for open areas and 5.0 m for areas obscured by vegetation. Nothing has been mentioned about the horizontal accuracy of the DTM [Alberta Forestry, Lands, Wildlife; 1988].

The provincial DTM of Alberta comes in files which cover one 1:20,000 map sheet (15 minutes by 7.5 minutes) area with some overlaps (200 m) and follow the National Topographic System (NTS) tiling scheme. The NTS primary blocks are 8 degrees by 4 degrees, each divided into sixteen smaller 2 degrees by 1 degree blocks. The smaller blocks are divided into a 4 x 4 array with each 30 minutes by 15 minutes which are further subdivided into four 15 minute by 7.5 minute quadrangles.

The files, which are basically ASCII, are in Digital Map Data Format (DMDF) and contain both 25 m grid and feature points as well as breaklines. The DMDF was developed to permit transmission of digital map data in a format which is hardware

independent and contains the minimum amount of information required to convert the data to graphic files. The record types include point, line string, curve, annotation, and map header. Point record types list easting, northing, and height for both spot heights and regular grid points all in millimetres. Line record types provide coordinates for sharp and round breaklines and for structure lines such as lake shorelines. No height is recorded for water bodies.

G.2.2 GTOPO30 DTM Data Set

GTOPO30 is a global data set with a horizontal grid spacing of 30 arc seconds (or 0.0083333333333333 degrees which is approximately 1 kilometre). Completed in late 1996, it was developed over a three year period through a collaborative effort led by staff at the U.S. Geological Survey's EROS Data Center (EDC).

The horizontal coordinate system of the GTOPO30 is decimal degrees of latitude and longitude referenced to WGS84, while the vertical units represent elevation in metres above mean sea level. However, the ocean areas have been masked as "no data" and have been assigned a value of -9999.

To facilitate electronic distribution, GTOPO30 has been divided into 33 smaller tiles, six of them covering Antarctica. The Antarctica tiles cover 30 degrees of latitude and 60 degrees of longitude each, while the rest cover 50 degrees of latitude and 40 degrees of longitude each. There is no overlap between the tiles. Figure G.3 shows the tiles graphically.

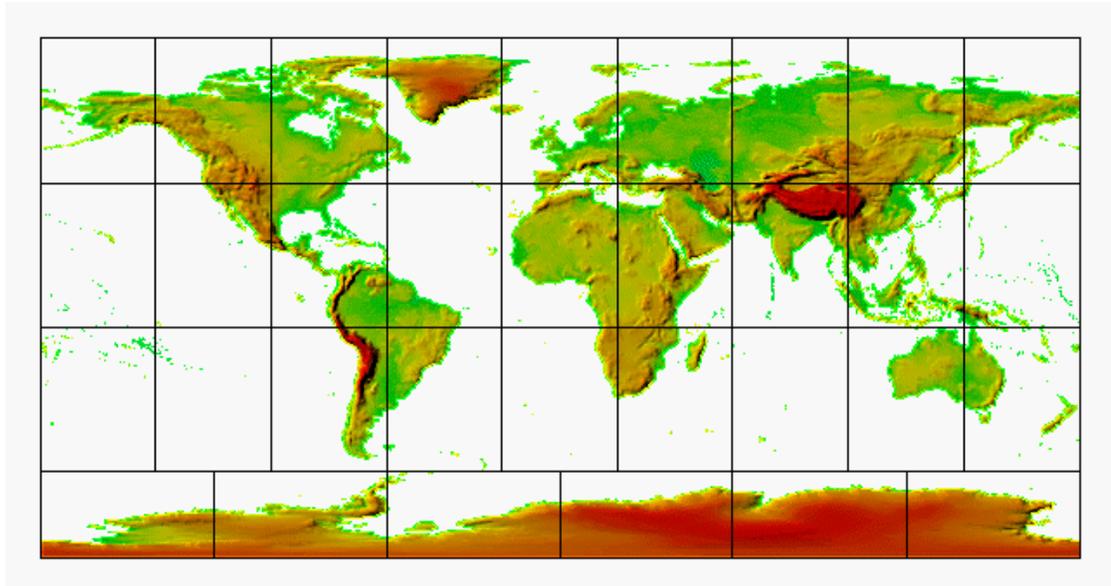


Figure G.3: GTOPO30 tiles

Eight different sources of elevation information have been used to generate the GTOPO30 data sets. These sources include

- 1) Digital Terrain Elevation Data (DTED),
- 2) Digital Chart of the World (DCW),
- 3) USGS 1-degree DEM's,
- 4) Army Map Service 1:1,000,000-scale maps,
- 5) International Map of the World 1:1,000,000-scale maps,
- 6) Peru 1:1,000,000-scale map,
- 7) New Zealand DEM, and
- 8) Antarctic Digital Database.

Figure G.4 is a color-coded map of the world showing the GTOPO30 source of

information around the globe.

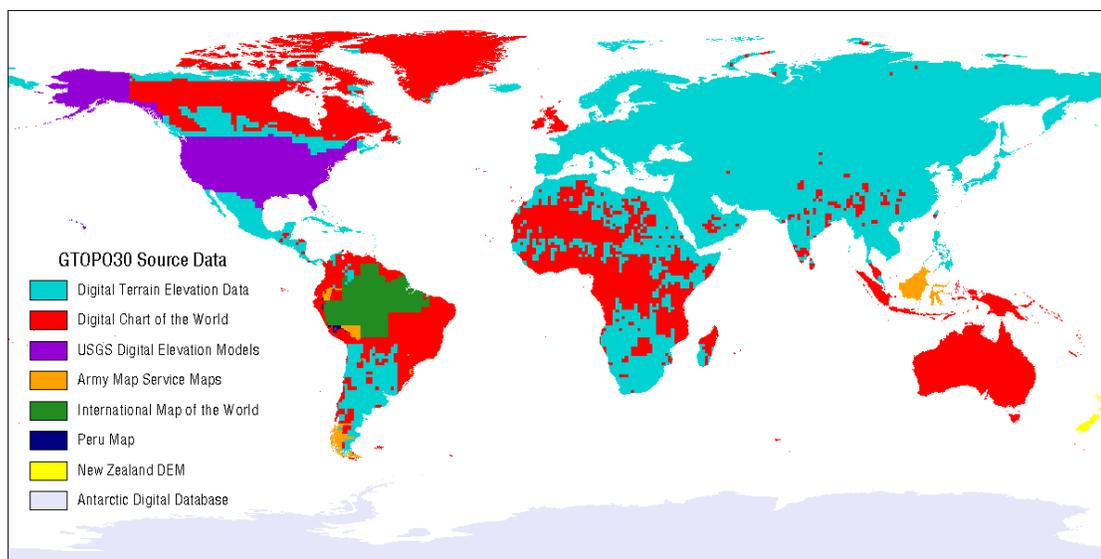


Figure G.4: GTOPO30 data source around the world

The first two sources, produced by the National Imagery and Mapping Agency (NIMA), provide 80% of the information for generating GTOPO30 data sets. The Digital Terrain Elevation Data (DTED) is a raster topographic data base with a horizontal grid spacing of 3 arc seconds (approximately 90 metres). The Digital Chart of the World (DCW) is basically a vector cartographic data set based on the 1:1,000,000-scale Operational Navigation Chart (ONC) series, the largest scale base map source with global coverage [Danko, 1992]. The DCW was used as the primary source for filling gaps in the DTED coverage, including all of Australia, most of Greenland, and large areas of Africa, South America, and Canada.

The absolute vertical accuracy of GTOPO30 varies with location according to the source data. Generally, the areas derived from the raster source data have higher

accuracy than those derived from the vector source data. The full resolution 3-arc second DTED and USGS DEM's have a vertical accuracy of 30 metres at the 90 percent confidence level [DMA, 1986; USGS, 1993].

The absolute vertical accuracy of the DCW is stated in its product specification as 650 metres at the 90% confidence level [DMA, 1990]. However, experience has shown that the grids derived from DCW data are in many areas be much more accurate than the 650-metre specification. The DCW grid has been compared to 30-arc second DTED, aggregated by averaging, for portions of southern Europe and the Mideast, and all of Africa. This comparison shows that the accuracy of DCW data set is about 160 metres at 90 percent confidence. This number compares favorably with an expected vertical accuracy (at 90 percent) of one-half of the primary contour interval of 1,000 feet (305 metres) for the topographic maps on which the DCW is based [DMA, 1990].

The accuracy of the areas of GTOPO30 based on the other sources can only be estimated based on what is known about each source. Using certain assumptions, the vertical accuracy of each source (and the derived 30-arc second grid) can be estimated from the contour interval. One assumption is that the original map sources meet the commonly used accuracy standard which states that 90% of the map elevations are within + or - one-half of the contour interval. It is unknown if any of these maps actually meet this standard. Also, map digitizing and elevation surface interpolation errors are unknown and therefore not included. However, one should bear in mind that although not specified for this data set, for many areas the relative accuracy is probably better than the estimated absolute accuracy.