INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Multiple-patch matching in the object space for aerotriangulation

Krupnik, Amnon, Ph.D.
The Ohio State University, 1994
MULTIPLE-PATCH MATCHING
IN THE OBJECT SPACE FOR
AEROTRIANGULATION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Amnon Krupnik, BSc., MSc.

* * * * *

The Ohio State University

1994
ACKNOWLEDGMENTS

I would like to thank my advisor Prof. Toni Schenk for introducing me to the field of digital photogrammetry, and for the scientific, moral and financial support during my studies.

I also thank:

- Prof. Kurt Novak for serving in my examination committee and for allowing me to use his bundle adjustment software for this research.

- Prof. Clyde Goad for serving in my examination committee.

- Dr. Charles Toth for his moral support and useful advises concerning various aspects of this research.

- André Templer for his technical help and for providing me with access to any available resource which could advance my research.

- Karen and John Angell for voluntarily devoting their time to proof reading my dissertation.

Finally, I offer my appreciation to my supportive family in Israel, and most of all I wish to express my deep gratefulness to my wife Michal and my children Orr and Ronnie who have stayed with me here through these long years away from home.
VITA

February 5, 1962 .................. Born - Beer Sheva, Israel

1985-1986 ...................... Teaching Assistant
Department of Mathematics
Technion–Israel Institute of Technology
Haifa, Israel

1986-1990 ...................... Teaching Assistant
Department of Geodetic Engineering
Technion–Israel Institute of Technology
Haifa, Israel

1988 .......................... BSc Geodetic Engineering
Technion–Israel Institute of Technology
Haifa, Israel

1988 .......................... BA Computer Sciences
Technion–Israel Institute of Technology
Haifa, Israel

1990 .......................... MSc Geodetic Engineering
Technion–Israel Institute of Technology
Haifa, Israel

1990-present .................. Research Associate
Department of Geodetic Science and Surveying
The Ohio State University

Fields of Study

Major Field: Geodetic Science and Surveying
# Table of Contents

**ACKNOWLEDGMENTS** ................................................................. ii

**VITA** ......................................................................................... iii

**LIST OF TABLES** ........................................................................ vi

**LIST OF FIGURES** ....................................................................... vii

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td><strong>INTRODUCTION</strong> ................................................................. 1</td>
</tr>
<tr>
<td></td>
<td>1.1 The Scope of This Work .................................................. 3</td>
</tr>
<tr>
<td></td>
<td>1.2 Definition of Terms ......................................................... 4</td>
</tr>
<tr>
<td></td>
<td>1.3 Organization of the Report ............................................... 5</td>
</tr>
<tr>
<td>II</td>
<td><strong>DIGITAL MATCHING</strong> ............................................................... 7</td>
</tr>
<tr>
<td></td>
<td>2.1 Accurate Matching Based on Gray Values .............................. 8</td>
</tr>
<tr>
<td></td>
<td>2.2 Obtaining Approximations by Feature-Based Matching ............. 16</td>
</tr>
<tr>
<td></td>
<td>2.3 Matching in the Object Space ............................................. 19</td>
</tr>
<tr>
<td></td>
<td>2.4 Multiple-Patch Matching ................................................... 22</td>
</tr>
<tr>
<td>III</td>
<td><strong>AUTOMATIC AEROTRIANGULATION</strong> ........................................ 26</td>
</tr>
<tr>
<td></td>
<td>3.1 Aerotriangulation—Overview ............................................ 26</td>
</tr>
<tr>
<td></td>
<td>3.2 Motivations ...................................................................... 34</td>
</tr>
<tr>
<td></td>
<td>3.3 Current Developments ...................................................... 36</td>
</tr>
</tbody>
</table>
## IV ACCURATE MATCHING FOR AEROTRIANGULATION

4.1 Outline of the Proposed Matching Strategy .................................................. 48
4.2 Image Patch Warping .................................................................................... 55
4.3 Matching Warped Images ........................................................................... 59
4.4 Reconstructing Object Space Surfaces by Hierarchical Matching .......... 64
4.5 Matching Multiple Image Patches ............................................................... 66

## V EXPERIMENTS AND RESULTS

5.1 Data Sets ..................................................................................................... 70
5.2 Reference and Approximate Coordinates ................................................... 73
5.3 Experimental Results .................................................................................. 77
   5.3.1 Multiple-Patch Matching with Simulated Data ..................................... 78
   5.3.2 Multiple-Patch Matching in the Object Space with Real Data .......... 86
      5.3.2.1 Tests with Stereo Models ............................................................ 87
      5.3.2.2 Tests with Blocks ........................................................................ 94
      5.3.2.3 Summary ..................................................................................... 95

## VI CONCLUDING REMARKS

.................................................. 97

## APPENDIX

MULTIPLE-PATCH MATCHING—DETAILED DERIVATION OF THE
MATHEMATICAL MODEL ........................................................................... 100

A.1 Reducing the Size of the Equation System ................................................. 100
A.2 Efficient Estimation of the Theoretical Gray Values of the Surface
   Elements ......................................................................................................... 105
A.3 The case of two image patches .................................................................... 107

## BIBLIOGRAPHY

.................................................. 110
## List of Tables

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Technical information about the data sets used in the experiments.</td>
</tr>
<tr>
<td>2</td>
<td>Accuracy and statistics of manually measured reference data.</td>
</tr>
<tr>
<td>3</td>
<td>Size of images, number of grid points and grid intervals at different resolution levels.</td>
</tr>
<tr>
<td>4</td>
<td>Results of the object space matching.</td>
</tr>
<tr>
<td>5</td>
<td>Results of the image space matching.</td>
</tr>
<tr>
<td>6</td>
<td>A comparison between the results of object space and image space matching for the same data sets.</td>
</tr>
<tr>
<td>7</td>
<td>Results of the multiple-patch object-space matching.</td>
</tr>
<tr>
<td>8</td>
<td>A comparison between the results of the multiple-patch matching in the object space and the manually measured points.</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Straightforward search for the optimum similarity.</td>
</tr>
<tr>
<td>2</td>
<td>Geometric discrepancies as a consequence of projective geometry.</td>
</tr>
<tr>
<td>3</td>
<td>Surface reconstruction in the object space</td>
</tr>
<tr>
<td>4</td>
<td>A typical block configuration in aerotriangulation.</td>
</tr>
<tr>
<td>5</td>
<td>Traditional aerotriangulation procedure.</td>
</tr>
<tr>
<td>6</td>
<td>Footprints of a photograph on the $XY$ plane.</td>
</tr>
<tr>
<td>7</td>
<td>Examples of two correlation curves.</td>
</tr>
<tr>
<td>8</td>
<td>A wrong match within a smooth surface.</td>
</tr>
<tr>
<td>9</td>
<td>Outline of the proposed matching scheme.</td>
</tr>
<tr>
<td>10</td>
<td>Warping and Matching Module.</td>
</tr>
<tr>
<td>11</td>
<td>Block Adjustment Module.</td>
</tr>
<tr>
<td>12</td>
<td>DEM Update Module.</td>
</tr>
<tr>
<td>13</td>
<td>Relations between image- and object- space pixels.</td>
</tr>
<tr>
<td>14</td>
<td>Image/Object space relief displacement relationship at occlusions.</td>
</tr>
</tbody>
</table>
15 A surface patch representation. ................................. 60
16 Schematic description of the concept of matching warped images. .... 61
17 Correcting the shift created by non-intersecting light rays. ......... 63
18 A scale space approach for matching small warped image patches. ... 65
19 Examples from the images used for the experiments...................... 72
20 Point distributions on the data sets. ................................ 74
21 An example of 0.3 noise level. ........................................ 79
22 The influence of matching parameters on the quality of the results. ... 83
23 Relations between grid points, patch resolution and matching windows. 90
CHAPTER I

INTRODUCTION

Aerotriangulation has been a well-established photogrammetric procedure for years. Its main goal is to obtain horizontal and vertical control points using photogrammetric means rather than field surveying. A minimum number of such points is required to perform an absolute orientation of a stereo model. Without aerotriangulation, the cost of obtaining these points is significantly higher.

With recent technological developments, two major opportunities for reducing the costs and efforts invested in aerotriangulation have opened. The first is by minimizing the number of required ground control points using airborne GPS for recording the coordinates of the exposure stations of each photograph (e.g., [2]). The second is by automating the tie point selection, transfer and measurement phases, using digital photogrammetric techniques. When both technological advantages are used, the role of a human operator in the aerotriangulation procedure is minimized and the efficiency is increased.

In digital photogrammetry, digital images are processed rather than analog photographs. Early experiments with image correlation were performed already in the late 1950's, but only recent progress in computer hardware has enabled the handling of large, high resolution digital images. Softcopy photogrammetric stations, which
are a direct result of this progress, are now gradually replacing analytical plotters in an increasing number of photogrammetric applications.

The main justification for the transition from analog photographs to digital images is the potential to automate photogrammetric procedures. The ultimate goal of a completely automated "map machine" may not be reachable in the foreseeable future, mainly due to a lack of solid theory about making maps [83]. However, there are many aspects for which automation is at the tip of our hands. Aerotriangulation is one of them.

Aerotriangulation is a rather complex task. Thus, one should not expect an automatic procedure in one step. Schenk and Toth [87] and Toth and Krupnik [93] described a general concept for automating the aerotriangulation procedure. These publications were mainly concerned with selecting tie points and obtaining approximate locations. The next step, accurately matching the tie points, is the main subject of the research described in this report.

Although accurate matching has been the subject of numerous research studies in the fields of photogrammetry and computer vision, its application for aerotriangulation adds some new aspects that have not been fully addressed:

- The exterior orientation parameters are not known. Therefore, constraining the matching by geometric conditions, e.g., the epipolar line, is not possible.
- The accuracy requirements are more demanding than in other applications.
- Only a single point is matched in each area. Therefore the matching should
be sufficiently reliable, and not dependent upon the surrounding surface for determining wrong solutions.

- More than two overlapping image patches are used at each matching location. The results obtained by simultaneously matching them are superior even to manual measurements. In the latter, simultaneous measurements are possible only on two images, and consistency is not guaranteed.

These issues and suggested ways to handle them are discussed in this work.

1.1 The Scope of This Work

A new method for reliably and accurately measuring tie points in aerotriangulation by matching gray values from multiple images is presented in this work. The method is specifically motivated by the aspects listed in the previous section. It is suggested that in order to increase the reliability, the matching will be performed in the object space, allowing the use of patches that are larger than those used in traditional area-based matching. In order to match in the object space, small object surfaces around each tie point are reconstructed, using a hierarchical approach.

The scope of the research presented in this dissertation is summarized in the following three steps:

- Analyzing the problem and designing the components of the matching method from a conceptual viewpoint.

- Practically implementing the matching scheme and testing it with real data.
• Analyzing the results, drawing conclusions, and offering suggestions for future research.

This work is concerned only with the last part of the Automated Aerotriangulation System, described in [92]. Therefore, automatically selecting and obtaining approximations for tie points are not included in the scope of this work. For the experimental part, these steps are performed interactively.

1.2 Definition of Terms

In this section, some of the terms used throughout the report are defined. These definitions are given here only for clarification purposes and refer to the applications described. More formal and complete definitions can be found in photogrammetric text books.

**Digital image:** 2-D matrix of pixels (picture elements), each of which has a gray value. Digital images are obtained either directly (by a digital camera), or by scanning analog photographs at a certain resolution. The latter is still more commonly used. The geometric transformation between the coordinate systems of the digital image and the analog photograph is obtained by measuring common points (fiducials) in both systems. In this work, it is always assumed that this transformation is known. Since the "image" is a digital copy of the "photograph," and the geometric transformation between them is known, these two terms are sometimes used interchangeably. Coordinates on a digital image are
not necessarily expressed by integer values, especially when sub-pixel accuracy is considered.

**Softcopy photogrammetric workstations (or softcopy stations):** Computer workstations that can handle large, high resolution digital images. In addition to traditional image processing capabilities, stereo images can be displayed, allowing the user to measure points and objects in 3-D space. Using appropriate software, softcopy stations are the equivalent to analytical plotters, with the difference that the operator analyzes digital images displayed on the computer monitor rather than looking at analog photographs through an optical system.

**Conjugate/corresponding points:** Points on different images which represent the same object space location. Although semantically the term “corresponding” is more general and “conjugate” refers to two images only, these terms are used interchangeably.

**Multiple overlapping image patches:** Small square (or rectangular) windows of the images, centered on a set of corresponding points. Except for radiometric and geometric distortions, these image patches represent the same object surface.

### 1.3 Organization of the Report

This report contains six chapters. In Chapter II, a general review of digital matching, which is a basic process for virtually every automatic photogrammetric procedure, is given. The issues of finding accurate locations of conjugate points (by area-based
matching), and obtaining approximations for these points (by feature-based matching) are discussed. In addition, discussions about matching in the object space and multiple-patch matching, which are the building blocks of the proposed matching scheme, are also included.

Chapter III discusses the concepts of aerotriangulation in general, and of automatic aerotriangulation in particular. A review of past research as well as current trends in this subject are included.

Chapter IV describes the proposed *Multiple-Patch Matching in the Object Space for Aerotriangulation*. The motivations for designing the new method as well as the limitations of the existing techniques are explained first. An outline of the method and an extensive description of its components follow. Additional mathematical derivations are given in the appendix.

In Chapter V the experiments that were performed to check the new method and to compare it to existing matching methods are described. The design of the experiments is explained and the results are shown and discussed.

Concluding remarks and suggestions for future research are given in Chapter VI.
CHAPTER II
DIGITAL MATCHING

One of the main goals of photogrammetry is to reconstruct the 3-D object space from 2-D images. Traditionally, quantitative determination of 3-D points and features has been performed interactively by a human operator, using analog, analytical or digital plotters. Semantic information has been attributed to these features by the operator, using his ability to fuse the implicit information embedded in the images, like gray values, color, texture etc. with his prior knowledge about the scene.

For an automated system, where a computer replaces a human operator, even extracting only the non-semantic information from the images is not a trivial task. Many methods have attempted to obtain quantitative 3-D information about the object surface from digital images. Examples are shape from shading, shape from texture, shape from shadows, and more (see, e.g., [13]). However, for obtaining direct depth information, stereo analysis is required [69, 72], while the other methods are used to support and enhance the information obtained from stereo [36, 40].

In order to calculate accurate object space location, conjugate points are identified on two or more images. By having such points, together with the exterior orientation parameters, intersection of the lines connecting these points with their respective projection center is enabled. The common name for finding conjugate points on
different images is *digital matching*, or *matching* for short. In the more specific case when only two images are involved, the term *stereo matching* is used.

Numerous research studies in photogrammetry and computer vision have been devoted to matching. The motivation for most of these studies has been the reconstruction of the object surface. Because extensive reviews exist (e.g., [14, 25, 26, 48, 60, 99]), this chapter will focus mainly on a review of matching concepts which are relevant for automatic aerotriangulation. The direct applicability of some of these concepts to aerotriangulation is further explained in Chapters III and IV.

### 2.1 Accurate Matching Based on Gray Values

The accuracy of a calculated point in the object space (assuming that the exterior orientation is known) is directly affected by the accuracy of the conjugate points. All the more so, very accurate image points are required if matching is performed to solve the orientation parameters. To make the results of digital matching compatible with those achieved in traditional photogrammetry, the accuracy should be at the few microns level or referring to digital images, a fraction of a pixel. Sub-pixel accuracy is usually obtained by matching gray values, a method called *area-based matching*.

Area-based matching is founded on the idea that image windows centered on conjugate points in different images have similar radiometric characteristics. Having a point in one image, its conjugate in the other\(^1\) is obtained by optimizing a certain similarity measure, defined over the gray values within the image window. Possible

---

\(^1\)The description here refers mainly to a stereo pair. The application of accurate matching methods to multiple images is discussed in Section 2.4 and further developed in Chapter IV.
similarity measures include cross-correlation coefficient, sum of squared differences and others (see [78] for details).

The simplest way to find the optimal solution is the following:

1. Define a template window in one image, centered on the point to be matched.

2. Slide the template within a search window in the other image, centered on the approximate location of the conjugate point (Figure 1a). At each location, a similarity value is calculated. This yields a discrete 2-D function of similarity values (Figure 1b).

3. If sub-pixel accuracy is sought, a continuous 2-D similarity function is fitted to the discrete values obtained in 2, and the location of its local extremum (either minimum or maximum, depending on the similarity measure used) is calculated to sub-pixel accuracy. Otherwise, the best value on the discrete similarity function is used.

Although this approach has been used in many research studies (e.g., [9, 18, 62], and to some extent, [48]), simply looking for maximum similarity between two image windows suffers from two main disadvantages:

- A priori assumptions about the characteristics of the similarity function fitted to the discrete values must be made. The location of the extremum therefore depends on the selected function.

- Images are usually perspective projections. Therefore, unavoidable geometric discrepancies between the two images lead to inaccurate matching results, since
Figure 1: Straightforward search for the optimum similarity: (a) Sliding a template within a search window; (b) A 2-D function of similarity values, for which an extremum is sought.
even at the correct location only the center pixel in the template will correspond exactly to its mate in the target window\(^2\). The discrepancies increase with increasing elevation differences in the object space surface within the matching window\(^3\). Figure 2 demonstrates the discrepancies caused by projective geometry for the case of a tall building.

The first problem is avoided by defining a mathematical model that calculates the similarity values and optimizes a target function simultaneously. In photogrammetry, the most common mathematical model is the Least-Squares Matching (or LSM), introduced by Ackermann [1] and Förstner [31]. In this approach, a certain type of geometric transformation between two image patches is introduced. The squared sum of gray-value differences between the image patches is minimized in a least-squares adjustment, and the results are estimates of the transformation parameters. These estimated parameters allow one to determine the correct conjugate location. In its simplest form, LSM assumes a one- or two-directional translation to model the geometric differences between the two image patches. The results in this case are similar to what is obtained with the straightforward search for the optimal location described earlier. However, the solution is unique and is not dependent upon the selection of the continuous function.

In order to solve the second problem (which is also apparent if only shift parameters are considered for LSM), the geometric differences between the two images are

\(^2\)Similar geometry is possible only when true vertical photographs are available, there are no elevation differences and the two image patches are located at the centers of the photographs. This however is not a common situation.

\(^3\)This problem is usually referred to, especially in computer vision, as "the foreshortening problem."
Figure 2: Geometric discrepancies as a consequence of projective geometry: On the right image, the square shape of the building is distorted, and one of its facades is visible.
modeled, and the image patches are warped (or "shaped") accordingly. In general (though not for the application described in this work), the model for warping is relatively simple if the ground surface is a horizontal plane. Knowing the exterior orientation parameters, one or both images are warped such that they have equal scale and their axes are parallel. In a more practical case however, where the surface is not a horizontal plane, a more complicated geometrical model is required.

In many implementations of LSM (e.g., [1, 42, 76, 79, 63, 65]), an affine transformation is selected to model the geometric discrepancies between the images. Such a transformation, assuming near vertical perspective photography, models the ground surface as a tilted plane, and does not require an explicit expression of the exterior orientation parameters. While this ground representation is sufficient for some applications, more realistic cases with discontinuities or gradient changes in the surface render unacceptable results. These situations are prominent in large scale images, especially when man-made features are present, and when large matching windows are used.

Grün and Baltsavias [42, 43, 44, 14] suggested overcoming inaccuracies in the matching process by constraining the solution to comply with the collinearity conditions. Although this yields some improvement in the $y$ direction (perpendicular to the base of the stereo model), no improvement is expected in the $x$ direction, since the collinearity constraint does not have any effect on the elevations. Furthermore, if the exterior orientation parameters are not known with sufficient accuracy, their errors, propagated to the intersected point, might be much larger than the matching errors
themselves. Introducing stochastic rather than fixed geometric constraints, provides only a partial solution, because there is no strict way to weight properly the geometric and radiometric parts of the equation system.

Ebner and Heipke [29], Helava [51] and Wrobel [99] describe a general approach for LSM, in which a rigorous geometrical model is formulated and embedded in the adjustment. In these theoretical global approaches, the orientation parameters, surface elevations and radiometric properties of the imaging process are determined simultaneously. Since it is impossible to estimate all these parameters from matching a single point, the entire surface is calculated. In practice, this model proved to be unstable mainly because of dependencies between the surface elevations and the orientation parameters. Furthermore, the amount of computations required for solving such a large equation system, which may contain thousands of unknowns, is enormous [30]. Even when some of the parameters (essentially, the orientation parameters) are known, erroneously modeled parameters are likely to lead to a matching procedure failure [28]. An alternative to the global approach is explained in Section 2.3.

Both the straightforward search for the maximum similarity and the least-squares techniques, even when the geometrical model is properly defined, require two preconditions to be fulfilled:

- The radius of convergence is rather small (on the order of few pixels), especially for the least-squares techniques. Therefore, the matching location should be approximately known.

- Convergence and correctness of the matching depend on the image content.
Both methods might fail if the image window does not contain sufficient gray-value variations, or if the window contains repetitive patterns.

Fulfilling these preconditions depends on the application. One matching application, for example, is assisting a human operator with accurately locating a measuring mark on the ground using a softcopy station [16, 95, 58]. In this case, the approximation is given by the user, and it is assumed that a noticeable feature is actually measured, so the images contain sufficient information. Another application is to obtain the object surface in areas where it is smooth, or where the image scale is sufficiently small, such that surface discontinuities are negligible. In such cases, the matching procedure is combined with a smoothness constraint, and matching points which do not comply with the constraint are marked as blunders (e.g., [79]).

These solutions however are not very common, as they do not provide a robust solution to a variety of possible applications. It is well accepted that alternative strategies to strict area-based matching techniques are required [55, 98].

Such strategies are usually based on the following ideas:

- Pure hierarchical matching where the disparity or the object space locations are obtained first at low resolution, and each level is used as approximation for a higher one (e.g., [30, 49, 65]).

- Feature-based matching (see Section 2.2).

- A combination of both the above strategies.
2.2 Obtaining Approximations by Feature-Based Matching

As mentioned in the previous section, approximate image locations in areas which contain sufficient information are required in order to ensure reliable results from an area-based matching process. Such locations are usually called *features*, and the process of finding corresponding features on different images is called *feature-based matching*.

Feature-based matching is originated in the computational theory of the human visual system. The ganglion cells on the retina are sensitive to the changes in brightness over the area covered by their receptive field [56]. The information from both eyes, in the form of edges (features) is then matched in higher levels of the human visual system.

Three steps are involved in feature-based matching [32]: Detecting the features, matching them, and if the final goal is to recover the entire surface, interpolating between them.

The main criterion for detecting the features is their distinction from other features [15]. Several types of features are most commonly used:

**Interest points:** are points that contain "interesting information" like corners, circles etc. These points are detected by interest operators (for example, Förstner [35], Moravec [71]). Since it is easy to extract interest points, and it is technically simple to compare points on different images (although high reliability is hard to obtain), interest points are widely used [15, 22, 33, 46, 47, 48, 59, 62, 63].
**Edge points (or edge elements):** are pixels where a significant change in gray values exist in one direction, pointing to possible features in the real world. Edge elements are extracted by convolving the original image with certain high-pass filters, or *edge detectors*. Although edge detection had been performed earlier, the *Laplacian of the Gaussian (LoG)* operator (or actually, a set of LoG operators, that differ from each other by a scale factor), presented by Marr and Hildreth in 1980 [68], undoubtedly marked a turning point in edge detection theories. Later edge detectors, such as Canny’s operator [21], or the OZCO operator [80] are based on the concepts of the LoG. Edge elements are used for matching in a similar way to interest points [22, 53]. An alternative option is to check the consistency of matched points by their connectivity to the other edge points in their neighborhood [12, 41, 75].

**Edge segments (or edges):** are linked edge elements. Long edges point to well-noticed geometric entities in the images, and therefore are more likely to appear on more than one image. Matching edge segments, although not straightforward to implement, is more robust and reliable than matching interest or edge points. Possible extended edge elements are straight segments [10, 54, 69] or more complex representations (e.g., [52, 38, 62, 85, 97]).

Once features are detected in both images, they are matched. No matter what features are used, “...the task is to find a one-to-one match in a many-to-many situation” [26]. A brute force approach is to base the search only on the similarity between features. Such an approach immediately leads to an exponential computation
complexity, especially if points or edge elements are used. Furthermore, in most cases it is virtually impossible to decide which solution is the best, as more than one candidate matching location may show similar measures. In order to reduce the computational complexity, the search space in which the correct solution is sought must be constrained.

One possibility, which is commonly used for edge points, is to assume epipolar geometry of the images (which can be obtained if the exterior and interior orientation of the images are known [24, 82]) and a bound for the relief displacements in the scene [12, 22, 41, 47, 48, 53, 69, 79]. The search is then constrained to a section of a scan line, rather than the entire image. Another possibility (which is not necessarily distinct from the former) is to assume that the object surface is smooth. Matching results which show sharp changes in the disparity values are considered to be incorrect. A more robust approach is to assume smoothness only between prospective discontinuities (or edges) [39, 67] and continuity along edges ("figural continuity") [41].

A formal, global approach to handle assumptions about features and their inter-relations is based on Shapiro and Haralick’s relational matching idea [89]. In this approach, features [19, 54, 97], or scan lines from either the original image or a filtered version [100] are represented symbolically. Rather than the original data, this symbolic representation of features, or peaks and valleys on the signal, and their inter-relations are then matched.
Matching symbolic (or relational) descriptions offers several important advantages over matching features or gray values directly:

- It provides a global solution for the correspondence problem.
- If proper representations are selected, it is invariant with respect to radiometric and geometric distortions.
- Partial occlusions or other disturbances affect the matching only locally, while the entire matching process is still possible.
- The required assumptions are less restrictive.

In general, feature-based and relational matching offer more reliable solutions for the matching problem. These methods are less sensitive to geometric and radiometric discrepancies between the matched images than methods for matching gray values described in Section 2.1. On the other hand, obtaining sub-pixel accuracy with feature-based matching, although possible, is not as common and easy to implement as with area-based techniques. Furthermore, a solution for matching features from more than two images simultaneously is not available. Therefore, the final refinement of the matched points to sub-pixel accuracy is left for gray-value matching techniques.

2.3 Matching in the Object Space

Section 2.1 discussed the requirement for a mathematical model which will consider the geometric differences between the matched images, especially those caused by the foreshortening problem. Schenk [84] and Doorn [27] presented a practical approach
for obtaining the surface by iteratively reconstructing its geometric and radiometric model.

According to this approach, images are rectified to the object space using the known exterior orientation parameters and an approximate surface. If the surface were perfectly known, this rectification would result in two orthophotos (see [70], p. 460 for definition). Since the surface is known only approximately, different images, termed as warped images are obtained. The geometric differences between the two warped images are much smaller than those of the original images. Therefore, a relatively simple matching model (possibly, with only shift parameters) can be used. When the entire warped images are matched (step by step, not simultaneously as in the global object space LSM mentioned in Section 2.1), a map of displacement vectors is obtained. The points, corrected according to these displacement vectors, are then projected back to the original images through the available surface. They are then intersected back to the object space and interpolated to a new, improved, model of the surface. The entire procedure is iterated. In the first matching level, the surface is not known. Therefore, the matching is performed in low image resolution (or a low level in the image pyramid) and the surface is refined along the procedure. One iteration level is demonstrated in Figure 3, and explained in more details (in the context of this work) in Sections 4.2 and 4.3.

This approach has three major advantages over the object space LSM approaches mentioned earlier:
Figure 3: Surface reconstruction in the object space (from [84]).
• It decouples the determination of the surface from the matching procedure itself, avoiding dependencies between parameters and direct influence of the geometric constraints on the gray-value matching procedure.

• Each point is determined separately, and therefore no large systems of equations need to be are solved.

• The mathematical model for matching includes only shift parameters. Furthermore, the matching is not constrained to the LSM methods. It can be used to improve the performance of any other method, including feature-based and relational matching.

The concept of directly matching warped images has shown very promising results. Norvelle [73] partially used this concept for a practical surface reconstruction system.

Quam [77] presented an approach which is similar to matching warped images. In his work, only one image was warped, rather than both. The warping was performed according to a disparity map without considering any explicit orientation parameters or surface elevations.

2.4 Multiple-Patch Matching

For simply obtaining 3-D locations in the object space, it is usually sufficient to have locations of points on two images only. This is one of the reasons why the issue of multiple-image matching was not considered in traditional methods. However, many applications require the matching of multiple images. Examples include:

**Image motion:** An object is tracked along a sequence of images.
Close range applications: Each object space area is usually covered by more than two images, each of which contribute some information which is not visible on other images.

Aerotriangulation: Model and strip tie points which appear on more than two images are measured.

Even for applications like surface reconstruction, adding more views of the surface significantly increases the robustness of the solution and reduces occluded areas.

Grünn and Baltsavias [42, 43, 44] suggested performing the matching simultaneously by adding geometric constraints that are based on the collinearity conditions to the adjustment (see Section 2.1). However, the actual gray-value matching is still performed in pairs, defining one image patch as a template, and comparing all the others to it.

Object space LSM methods [29, 51, 99] as well as other optimization methods (e.g., [36]) also allow the use of more than two image patches simultaneously. In these cases, the theoretical "gray values" of the ground are treated as unknowns, and the optimization process minimizes the squared differences between these values and the gray values of the participating image patches. While this is truly a multiple-patch matching approach, it still suffers from the limitations mentioned in Section 2.1.

Agouris [5] reported on initial experiments with two methods for multiple-patch matching for aerotriangulation. The first was based on matching multiple image patches for individual points. The unknowns of the mathematical model were affine transformation parameters between a template patch, selected from the participating
patches, and each of the other patches. By using a template rather than considering the transformation parameters between every pair of image patches, dependencies between the unknowns are avoided.

The observations for the least-squares adjustment were differences of gray values between each possible pair of patches. This poses the problem of linear dependencies between the observations themselves. The actual redundancy in this case is much lower than the theoretical one, and the values of the elements of the normal matrix outside the blocks around the diagonal are very small, and effectively negligible. The matching in this case is therefore performed practically in pairs.

The second method, called **Multiple-Image Multi-Point Least-Squares Matching**, was based on including geometric constraints in the least-squares matching model. Rather than using an affine transformation as the mathematical model for comparing one image patch to another, a rigorous formulation, based on the collinearity conditions was used. Since the bundle of rays formed around a small image patch is very narrow, the geometry of the intersection here is very poor. To compensate for this problem, a simultaneous adjustment, which included all the points in the block was used.

Multiple-Image Multi-Point Least-Squares Matching poses two problems. To avoid dependencies between the parameters in the adjustment process, an unknown scale factor within each image patch was assumed to be constant. In order not to violate this assumption, the matching is performed with relatively small image patches. This might cause a large rate of failure (see Chapter IV). The other problem is even
more severe. Simultaneous gray-value matching and block adjustment is an appropriate operation only if the errors resulting from the gray-value matching are assumed random. This is usually not the case, since matching errors are the result of either misidentification of a point (a blunder) or of an incorrect mathematical model. Furthermore, even if it is assumed that neither of the above is the case (i.e., the patch is sufficiently small such that the scale factor is constant, and therefore, the model is correct), there is still a risk that the geometrical constraints will dominate the adjustment such that the influence of the gray values will be negligible. Also not knowing the orientation parameters will lead to incorrect solutions.

The problems mentioned in this section have partially dominated the design of the multiple-patch matching method described in Chapter IV.
CHAPTER III

AUTOMATIC AEROTRIANGULATION

One of the most prominent applications of multiple-patch matching is in automatic aerotriangulation. In this chapter, conceptual issues of automatic aerotriangulation are reviewed and discussed. In Section 3.1, a brief overview of traditional aerotriangulation procedures is given. Section 3.2 describes the motivation for automating aerotriangulation, and Section 3.3 reviews some of the models and technological developments in this area.

3.1 Aerotriangulation—Overview


"The process for the extension of horizontal and/or vertical control whereby the measurements of angles and/or distances on overlapping photographs are related into a spatial solution using the perspective principles of photographs...".

In other words, aerotriangulation is aimed at increasing the number of control points beyond those available from the existing control network. Measurements are performed on the photographs rather than in the field, which reduces the cost significantly.
Aerotriangulation methods are classified into four groups (for an extensive review of aerotriangulation methods, see [11]):

- Graphical/mechanical methods
- Mechanical methods
- Semi-analytical methods
- Fully analytical methods

With the rapid increase of computing power and the use of analytical (and nowadays—digital) photogrammetric equipment, modern aerotriangulation is performed almost solely with analytical methods. Two major mathematical models are used: Block adjustment by independent models, where stereo models are tied to each other to form a block, and bundle block adjustment, where bundles of light rays are reconstructed simultaneously for the entire block. Both methods are described in detail in photogrammetry text book (e.g., [57, 70]). The summary presented in this chapter refers to both methods, with an emphasis on the bundle block adjustment.

First, some terms are defined:

**Strip (or flight line):** set of partially overlapping photographs, with the projection centers laying approximately on a straight line. The overlap between the photographs is 60%, but sometimes as much as 80%.

**Block:** set of parallel, partially overlapping strips. Here, the overlap (also referred to as sidelap) is usually 20% or sometimes 60%. In some cases, additional non parallel strips are also added. Figure 4 depicts a typical block configuration.
**Tie point:** point in object space which is visible on three or more photographs. Tie points are selected in appropriate locations (see Figure 4) so that their photo- (or model-) coordinates connect photographs (models) of a block to a stable geometric structure. Three types of tie points exist: natural points, signalized points (marked on the ground) or artificial points (marked on the photographs).

**Ground control point:** point for which the coordinates (either horizontal, or vertical or both) in the ground coordinate system are known from the control network.

The input to an aerotriangulation consists of a block of photographs and well-distributed control points. Four major steps are involved in conducting an aerotriangulation procedure [88], as shown in Figure 5:

1. **Tie point selection:** As a first step, the configuration of the photographs (i.e., their order and their approximate overlapping areas) is determined. A standard block configuration (60% forward, 20% side overlaps) requires the selection of at least six points per stereo model at the von-Gruber locations (nine points per photograph except for photographs along the margins of the block, where this number is reduced to four or six) [57]. In order to perform blunder detection easier and more reliably, clusters of points may be selected at the von-Gruber locations [34]. The points are selected in such a way that they will appear in as many photographs as possible. For the standard overlap configuration, this number ranges from two (at the margins of the block) to six. Areas with considerable deviation from the planned block configuration (e.g., shifts between
Figure 4: A typical block configuration in aerotriangulation.
photographs

ordered photographs, approx. locations of tie points

"pug-marked" photographs

tie and control point measurement

photo/model coordinates for tie and control points

block adjustment

orientation parameters ground coordinates of tie points

ground coordinates and description of control points

Figure 5: Traditional aerotriangulation procedure.
strips, or strips where the projection centers do not lie on a straight line) or where large portions are occluded, require special attention.

2. **Tie point transfer:** Each tie point must be measured at its exact conjugate location on all the overlapping photographs. For model points (which appear only on two images) this is a relatively easy task since stereo measurement is possible. A human operator finds conjugate points by his ability to reconstruct the object surface with stereo perception. However, for a tie point which appears on more than two photographs, simultaneous measurement is not possible. In traditional aerotriangulation, there exist two options to overcome this problem. The first is to use well-defined features (points) on the ground. However, such features are not always available, and marking them in the field (*signalized points*) is a rather expensive procedure. The second way which is usually preferred is to use a *point transfer device*. Here, points are marked by drilling tiny holes into the emulsion. This is not only a time consuming process, but it also reduces the quality of the photographs.

3. **Tie and control points measurement:** Tie and control points are measured in one of the following ways:

   - Monoscopic mode, where photo-coordinates of the tie and control points are measured separately on each image (for example, on a monocomparator).
• Stereoscopic mode, where photo-coordinates are measured simultaneously on two images using an analytical plotter. For each pair of photographs, a model is first formed by measuring model points, which are usually identical with prospective tie points. Then, control points are measured in 3-D mode. Usually, models are measured along strips. The first model is set with the first two photographs. Then, one photograph (the first) is replaced with the third, and so on. Advanced software packages (e.g., [94]) help the operator to preserve the consistency of the measurements.

The measured points are registered and then used in the block adjustment as observations.

4. **Block adjustment**: A block adjustment is the final step in an aerotriangulation project. The exterior orientation parameters of all the participating photographs, the ground coordinates of all the tie points and possibly other parameters (e.g., interior orientation parameters and unknown systematic errors [20]) are estimated. The input to the block adjustment includes photo- (or model-) coordinates of the tie and control points, ground coordinates of control points, and occasionally auxiliary data. An example for such additional data, are GPS coordinates of the projection centers [2, 17, 45], now increasingly used to reduce the required number of control points.

Performing aerotriangulation on softcopy stations offers several instrumental and procedural advantages over using analytical (or analog) equipment:
• A user friendly environment presents both the images and the software related information on a single display. With analytical plotters the photographs are viewed in the two eyepieces, and other information is displayed on a separate computer monitor.

• Points which were already measured are superimposed on the images. The operator can easily examine the location of the points already measured, their distribution in the model, etc. This option exists also in some analytical equipment (e.g., the Zeiss P1 analytical plotter), but since the field of view is relatively small, it is less effective. Furthermore, integrating such an option in analytical plotters requires hardware modifications and is rather expensive, while on softcopy stations it is included in the software.

• The interior orientation (i.e., the transformation between the photo- and image-coordinate systems) must be established only once.

• Once the images are stored, repeatedly using them does not reduce their quality.

• Images are easily rotated or flipped, allowing stereo measurements between photographs of different strips.

On the other hand, softcopy aerotriangulation has the following disadvantages:

• The resolution of digital images is lower than analog photographs due to computer storage limitations.

• Radiometric errors are unavoidable during the scanning process and thus make the quality of digital images inferior to analog photographs.
• The storage required for an entire block of photographs is enormous even when image compression techniques are used. Furthermore, compression of an image, by itself reduces the quality of the images.

The instrumental advantages of softcopy stations do not significantly increase the performance of aerotriangulation. In fact, Skalet et al. [90] suggested carrying an aerotriangulation on softcopy stations in the same fashion as on analytical plotters. However, digital photogrammetry offers additional advantages by automating some of the aerotriangulation steps. As described in the following sections, this leads to a new era of aerotriangulation, usually referred to as automatic aerotriangulation.

3.2 Motivations

The development of automatic aerotriangulation is motivated by three key factors:

Efficiency: Manual aerotriangulation, carried out by a human operator, is a time consuming procedure. Conducting it automatically in a batch mode, or even in a semi-interactive environment will save many human working hours.

Accuracy: Performing aerotriangulation on a softcopy station instead of on analytical plotters reduces the accuracy of the measurements as images with lower resolution are used (see Section 3.1). However, Ackermann and Schneider [3] showed that by applying digital matching techniques for aerotriangulation purposes, the accuracy of the results exceeds the accuracy of manual measurements with analytical plotters. Furthermore, digital matching makes it possible to simultaneously measure conjugate points on more than two photographs, leading
to better consistency of the measurements and avoiding the necessity of artificially marking points on the photographs.

**Reliability:** Compared to a human operator, an automated system is capable of finding a large number of tie points in a relatively short time. This in turn leads to more reliable results since the redundancy is very large and blunders are detected easily.

Although aerotriangulation is well-established and has been used and continuously developed by photogrammetrists over many years, major parts of the procedure are still performed manually. Among the four phases of aerotriangulation procedure (Section 3.1), only the forth one, block adjustment, is currently performed automatically (yet, intervention of a specialist is still required for analyzing and interpreting the results of the adjustment and the blunder detection). Clearly, this part of aerotriangulation is well-understood and mathematically formulated, which makes automation a relatively easy task.

The automation of the other three phases is studied in the context of digital image processing and low level vision techniques, especially matching as was described in Chapter II. The problem is generally divided into two major steps:

- Finding approximations for tie point locations, based only on a rough knowledge of the block configuration. These approximate locations should be fairly accurate to ensure that the matching methods will converge. The locations of the points ensure a proper coverage for block adjustment, and several points
are selected around each required location in order to enable a reliable blunder detection.

- Matching the tie points with an accuracy of few microns, in order to comply with the standards of aerotriangulation. The required accuracy and the reliability is exceeded by employing a multiple-patch (or "n-stage") matching procedure.

It should be noted that the automation of control point measurement, although feasible (see e.g., [6]), is a very difficult task. However, the effort involved in measuring these points interactively is not a major factor. Furthermore, when the GPS is used to determine the coordinates of the exposure stations, the required number of ground control points is reduced significantly, and the importance of measuring the remaining points automatically is minor.

3.3 Current Developments

The roots of automatic aerotriangulation are in the simpler procedure of orienting a pair of photographs. Automatic relative orientation was the subject of several research studies [46, 64, 85, 97]. However, relative orientation is only one of the steps required for aerotriangulation. Some other aspects are considered for the transition from a stereo model to an entire block. For example:

- Finding areas which are common to more than two photographs.

- Accurately matching more than two image patches.
• Selecting a sufficient number of points in appropriate locations such that a strong geometry is obtained.

In the past, the large amount of data involved, along with limited capabilities of computers, made it very difficult to approach the automated aerotriangulation problem. Recent hardware developments have lifted most of these limitations. Accessing and processing a large number of images is now feasible, thus paving the way for an automated system.

Few research studies have investigated the potential of automatic aerotriangulation and proposed solutions to the problems.

Ackermann and Schneider [3] applied matching for measuring tie points. Straightforward LSM between two image windows was used for accurate location of tie points on overlapping images. All possible combinations were matched and averaged. Manually measured points were used as approximations. The experiment showed that the matching accuracy is better than that obtained by manually measuring artificially marked points. The accuracy obtained is similar to what is expected when signalized points are used. These results are very encouraging. They verify the hypothesis that matching can overcome the need to artificially mark points on the images, and therefore, provides superior results. However, LSM between two images, with relatively small image windows might fail on many occasions, especially if the approximations are not good enough or the matching area is not sufficiently textured.

Helava [50] described the Digital Comparator Correlator System, which was aimed at automatic selection and measurement of tie points for aerotriangulation. Potential
locations of tie points were selected according to a predefined block configuration, and areas around these points were digitized with a CCD camera installed on an analytical plotter. Within these areas, interest points were detected and matched, and final refinement of the matching was performed with multi-image least-squares correlation in the object space.

Tsingas [96] and Ackermann and Tsingas [4] proposed another scheme for tie point selection and measurement within a block of aerial images. The method is based strictly on feature point matching, which is repeatedly used throughout the process. Starting with a rough approximation of the overlap configuration, a better overlap approximation is determined by initial matching with very coarse resolution images (2 mm pixel size). Windows from the overlapping images are then selected according to the approximate overlap, and a point matching procedure, based on graph theory and a heuristic search for the best fit, is applied for finding conjugate points. The procedure is repeated for each level of the image pyramid up to the required final resolution.

Two drawbacks are associated with this scheme. The first is the fact that only the overlap is determined at the first step. It forces the procedure to advance step by step through the image pyramid, rather than project the selected areas directly to the finest resolution. The other drawback is related to the interest-point matching which is also used at the final matching stage. Although interest points are accurately located, it is not guaranteed that exactly the same point is located on all the images [47].
Agouris [5] presented a different approach for finding approximate locations of tie points within a block. Rather than using the initial matching just to determine the approximate overlap, a photo-mosaic is formed. Based on feature matching at a low resolution (edges, extracted by the LoG operator), the relative orientation parameters of each overlapping pair of images are estimated. The result is that all the photographs in the block are registered to the coordinate system of the photo-mosaic. Therefore, a set of tie points in a desired configuration is easily selected. The approximate location of these tie points is better than what is obtained when only an initial overlap is known. These tie points are then matched by a multiple-patch matching scheme, as described in Section 2.4.

Schenk and Toth [87], Toth [91] and Toth and Krupnik [92] presented an improvement to the mosaic idea. This novel approach has three major phases, where the results are used as an input for a block adjustment:

**Strip and block formation:** Feature-based matching [38, 61] is applied to each pair of (low resolution) overlapping images. A dependent relative orientation is performed along each strip of photographs, similar to the procedure on analog instruments. Since it is not always possible to find model tie points in this stage, the scale is estimated from neighboring points within the triple overlap area. Once all the strips are formed, they are connected by strip tie points, found by matching across the strips.

The block formation brings all the points that are found during the matching process into a consistent 3-D coordinate system. Since many points are found
through the matching procedure, it is possible to interpolate a coarse DEM for the entire block. Using this DEM, the corners of the photographs are projected to the XY plane and generate the footprints of the photographs (Figure 6). The overlap at each area is then calculated. Based on these footprints, pixel coordinates on overlapping images are extracted for tie point selection. The size of the shifts among these points is approximately one pixel at the resolution used for the feature matching. This result is considerably better than the case where the image patches are extracted according to the overlap only, and where the shifts are as large as five pixels, or even more.

**Tie point selection:** The relatively good approximations that are obtained during the matching in the lowest resolution render small shifts between image patches that are extracted in the highest resolution. These shifts do not exceed 10% of the patch size. Image registration, based on edge matching, is performed between the overlapping image patches, and approximations to several tie points are found in each area.

**Tie point measurement:** Once approximations for tie points exist, multiple-patch matching takes place. The method used for multiple-patch matching is the main subject of this work, and is described in detail in Chapter IV.
Figure 6: Footprints of a photograph on the $XY$ plane (from [91]).
CHAPTER IV

ACCURATE MATCHING FOR AEROTRIANGULATION

This chapter describes the proposed Multiple-Patch Matching in the Object Space for Aerotriangulation. This procedure is the last step before the final adjustment in the automatic aerotriangulation concept (see [92]). The development of the matching method emerged from the need for accurate locations of tie points.

Three main goals motivated the design of the matching scheme which is outlined in Section 4.1:

1. Reliable matching with single points:
   A successful matching procedure is measured by its ability to provide correct results and to report failures. Decisions based on internal measures, like cross-correlation coefficient or the mean square error of LSM, do not always yield correct results. A simple example is shown in Figure 7. In chart (a) a distinct peak of a relatively low value is shown on a correlation curve. In chart (b), the correlation curve is almost flat, but has large values. It is obvious that the results in the former case, which reflect a distinct object in the image, are much more reliable than those in the latter.
Figure 7: Examples of two correlation curves: (a) low but reliable maximum; (b) high but unreliable maximum.
In DEM generation applications, constraints are used for obtaining reliable results (see Section 2.1). The object surface is assumed to be smooth, or at least smooth between discontinuities. If a location of a point (from the matching results) violates this assumption (as shown for example in Figure 8), it is usually eliminated and its elevation is interpolated from the neighboring points. When only a single point is required in each area, it is not possible to use the smoothness constraint to check the reliability of the results.

A possible solution is to use large matching windows, which are more likely to contain sufficient information. This however, leads to inaccuracies if image
space matching is used, since the object surface within the matching window cannot be approximated by a horizontal or a tilted plane (see Section 2.1).

The proposed matching method reconstructs the object surface around each point to be matched. The only purpose of this reconstruction is to have a better representation of the ground within the matching window. The surface is not used for constraining the matching results, however. For reconstructing the surface, a hierarchical approach is taken, where more points around the desired tie point are matched in lower resolution. The image patches are warped according to the reconstructed surface, and the matching is performed between the warped patches. Since the geometric differences are minimized, patches which are significantly larger than those used by traditional methods are matched. This considerably increases the reliability.

2. Avoiding numerical problems and dependencies among parameters:

Theoretically, the mathematical model for matching image patches could include the surface elevations and the exterior orientation parameters. However, if these parameters are introduced as unknowns in the LSM, numerical problems are to be expected which may lead to a weak solution. The cause for these problems are dependencies between the parameters. For example, such dependencies exist between the elevations and some of the exterior orientation parameters. Furthermore, the relatively small angle between the light rays makes it difficult to determine the orientation parameters from a single point. Schenk and Toth
[88] suggested a solution for the latter, where a separate orthogonal projection was used for each area around tie points.

The dependency problem here can be compared to the case where exterior orientation parameters and camera distortions are to be determined. The large correlation between the parameters leads to questionable results. A possible solution is to use a sequential adjustment procedure, alternating two groups of unknowns [23]. A similar idea is used in the proposed matching scheme. The image patches are warped, using the approximate surface and orientation parameters. Since the influences of the perspective projection and the surface elevations on the matched patches is eliminated, a simpler mathematical model in which the unknowns are only shifts can be used. Then, the entire set of tie points is used for determining the orientation parameters of all the photographs in the block by a standard block adjustment. By alternating these two steps while increasing the resolution of the images, the aerotriangulation solution is obtained.

3. Reducing correlation among observations

When more than two images are used for matching, a mathematical model, based on simultaneous adjustment, is used to take advantage of the additional information. The idea presented in [5] (see also Section 2.1) was based on using differences of gray values between each possible combination of a pair of images as observations. With this approach, there observations are correlated. For example, if three image patches are involved, only two observations are
independent for a certain pixel, e.g., the differences between the first and the second images, and between the first and the third\(^1\). The difference between the second and the third images depends on the other two, and does not add new information to the least-squares adjustment.

One can make an analogy between this case and measuring angles with a theodolite. If directions to three points are measured, only two horizontal angles are considered independent, while the third one is the complement of their sum to 360°. Therefore, using it as another observation will have no influence on the results.

The proposed matching method uses the gray values as observations, rather than gray-value differences. The “true” intensities of the surface (referred to as the gray values of the surface elements) are introduced as unknowns in the adjustment. Although there are dependencies between these observations, which result from the need to interpolate gray values during the warping procedure (see Section 4.2), these dependencies are very small and therefore negligible.

The novelty of the described strategy is noticed in three major aspects:

• Compared to the theoretical matching models proposed in other publications (e.g., [29, 51, 99], and to some extent, [88]): It separates the matching procedure from the determination of the orientation parameters. This separation strengthens the mathematical model by avoiding dependencies between unknowns.

\(^1\)These two observations are actually not completely independent, since the differences are derived from gray values, one of which is common to the two differences. However, this dependency is very low and therefore negligible.
• Compared to the warped image matching algorithm [27, 86]: Only a small object surface around each desired tie point is reconstructed rather than the entire surface. The algorithm is also extended from matching only a pair of warped images to multiple-patch matching. It will be shown (see Section 4.3) that matching warped images is feasible even when the orientation parameters are not accurately known, as their approximations are improved with each iteration.

• Compared to other accurate matching methods proposed for automated aerotriangulation methods [5], the proposed strategy reconstructs the object surface at the locations where tie points are required, allowing the use of larger matching windows, thus increasing the reliability.

4.1 Outline of the Proposed Matching Strategy

In this section, the proposed matching scheme is outlined. Figure 10 shows a schematic description of the iterative procedure. At the end of each iteration, the exterior orientation parameters for the entire block, as well as the surface around each tie point are improved as compared to the previous iteration. In each iteration, three main modules are used (Figure 9):

Warping and Matching Module (Figure 10): Image patches, centered on each tie point, are warped according to the available approximations of the orientation parameters and the surface around the tie point. During the final iteration, a multiple-patch matching (see Section 4.5) is performed with these warped image patches to determine the exact image coordinates of the tie point. In all
Figure 9: Outline of the proposed matching scheme.
Figure 10: Warping and Matching Module.
iterations but the last one, a grid is formed in the object space, centered on the approximate tie point. Multiple-patch matching is used (in the same manner as in the last iteration) to match each point on this grid, assuming again that an approximation of the surface is available from a previous iteration. Each matched grid point is projected back, through the available surface, to the image space (see Section 4.4), and the obtained photo-coordinates are used for the “Block Adjustment Module.”

**Block Adjustment Module** (Figure 11): The photo-coordinates found in the “Warping and Matching Module” constitute the input for a bundle block adjustment. Theoretically, all the matched grid points around each desired tie point could be used for the adjustment. However, these points are very close to each other. Therefore, the contribution to the block geometry by using all of them is negligible. On the other hand, the computation cost is considerably increased. Instead, one point (usually the center of the grid) is selected and used for the adjustment. The results of the block adjustment are new (improved) orientation parameters.

This module can be also used for eliminating erroneous matches. Since the automatic procedure for finding approximate locations of tie points provides a large number of tie points (see Section 3.3), a rather conservative approach can be taken by eliminating suspicious points.

**DEM Update Module** (Figure 12): Using the new orientation parameters estimated in the “Block Adjustment Module,” mini-DEMs are created by intersecting
new orientations (\( tp = \text{tie point} \))

Figure 11: Block Adjustment Module.

new tp, add. points photo 0

\( \cdots \)

new tp, add. points photo n-1

new orientations

(new tp, add. points photo 0 \( \cdots \) new tp, add. points photo n-1)

Improvement

\( (tp = \text{tie point}) \)

Improved DEM

Figure 12: DEM Update Module.
the photo-points (from the “Warping and Matching Module”) back to the object space and interpolating them into regular grids. The grid interval is usually half the interval that was used in the former iteration. In the first iteration, the DEM is created either by assuming a horizontal surface or by using available prior information, e.g., interpolated points from the feature-based matching (see Section 3.3).

Since each “Warping and Matching” phase yields better locations for the conjugate photo-points (as shown in Section 4.3), the results of the block adjustment render improved orientation parameters. These orientation parameters, together with the improved conjugate points, lead to better approximations for the object surfaces around each tie point. The process converges iteratively to the desired solution.

The scope of this work is limited to what is described earlier in this section. Therefore, some assumptions concerning the input to the procedure are made:

- The approximate photo-coordinates of the tie points for the first iteration are as accurate as few pixels.
- Approximate exterior orientation parameters, correspond either to the ground or to a relative coordinate system, and are known for all the photographs in the block. They can be computed by resection from the approximate photo-coordinates of the tie points.
- An approximation for the surface around each tie point is known. If no surface information is available, a horizontal plane at the elevation of the object
point intersected from the approximate photo-coordinates is assumed in the first iteration.

The proposed strategy for one iteration level is described as follows:

- **For each tie point:**
  
  1. **Use the “DEM Update Module” to obtain a DEM which covers the area.**
     
     The grid interval does not have to be as dense as the image resolution, as explained in Section 4.2.

  2. **With the “Warping and Matching Module,”**
     
     (a) Warp the image patches according to the DEM and the available orientation parameters.

     (b) Define a grid in the object space, and match its points by a multiple-patch matching procedure. This yields a map of 2-D displacement vectors (Section 4.2).

     (c) Project the new locations of the grid points (based on the displacement vectors) back to the image, using the existing DEM and orientation parameters.

     (d) Select one of the photo-points (close to the center of the area) to be part of the input to the block adjustment.

- **Use the selected set of points from all the areas as an input for the “Block Adjustment Module.”** The results of this adjustment are improved orientation parameters for all the photographs in the block.
4.2 Image Patch Warping

Image warping in the context of this work corresponds to rectifying patches from projective images onto an orthogonal coordinate system. In order to differentiate between orthophoto generation and the case discussed in this work, where neither the orientation parameters nor the surface elevations are accurately known, the term warped images is used. Rectifying photographs for obtaining orthophotos was being performed long before digital images were available. Novak [74] reviewed some of the common rectification methods. In this work described here, digital rectification is used.

In general, the following information is required for the rectification:

- An image (or an image patch).

- Exterior orientation parameters of the photograph to which this image patch belongs (or approximate parameters, if warped images are sought), as well as the transformation between the image and the photograph (interior orientation).

- A regular grid of points in the object coordinates system for the area covered by the image patch, and for which the elevation (or approximate elevation) at each grid point is known.

The density of the grid does not have to correspond to the resolution of the rectified image. However, if the grid is not sufficiently dense, distinct elevation differences in the object surface render large geometric distortions in the rectified image.
Each grid point is projected back to the image, using the collinearity equations. A distorted grid is formed on the image plane (Figure 13). Assuming (for the moment) that no occlusions exist in the image, every pixel location in the rectified image is associated with a specific location on the original image. The location on the image can be between pixel coordinates, which are expressed usually by discrete values. Therefore, the gray value assigned to the pixel of the rectified image is calculated by interpolation. Possible interpolation methods are “nearest neighbor,” “bilinear,” “cubic convolution” and others (see, e.g., [37], pp. 249–250; [81], pp. 55–58). The nearest neighbor interpolation may lead to considerable geometric distortions. Cubic convolution normally renders a better result. However, since sixteen image pixels contribute to the calculation of each warped image pixel, the computation is costly and large correlation exists between the pixels of the warped image. For the matching, it implies large correlation between observations. A compromise is to use the bilinear interpolation, which yields warped images with sufficient quality. Here, only four image pixels contribute to the value of a warped image pixel. The correlation between the pixels in this case is negligible.

In case of occlusions, some measures are taken prior to assigning gray values to the rectified image. Occlusions are areas in the object space that are not visible in the image. Figure 14 explains the criterion for deciding whether a point in the object space is occluded. Doorn [27] showed a practical implementation of this idea in a Cartesian coordinate system.
Figure 13: Relations between image- and object- space pixels.
$R_2 \geq R_1$ and $r_2 \geq r_1 \Rightarrow R_2$ is visible

$R_0 \geq R_1$ and $r_0 \geq r_1 \Rightarrow R_0$ is not visible

Figure 14: Image/Object space relief displacement relationship at occlusions (Based on [27]).
Gray values of a warped image within an occluded area are wrong. They reflect the radiometric values of the occluding object rather than those of the occluded area. Therefore, using these areas for matching yields inaccurate or, if the occluded area is large, invalid results. If the occluded area is small compared to the size of the matching window, its gray values are excluded from the set of observations for the matching procedure. Otherwise, matching is not possible in this particular location. For surface reconstruction applications, these missing data are completed by interpolation. For the context of this work, since there is much redundancy, these points are just eliminated.

4.3 Matching Warped Images

The warped image (Section 4.2), together with the DEM used to generate it, is a surface patch. Each element of this patch has, in addition to its elevation\(^2\), a set of \(\ell\) gray values, where \(\ell\) is the number of overlapping image patches (Figure 15). Keeping one patch fixed, a matching procedure is performed (as suggested in Section 4.5) which results in two shifts for each patch, except for the fixed one. By using the elevations at the new locations (after applying the shifts), the point is projected back to the image space.

The rationale of matching warped patches (or matching in the object space) was explained in detail in [27, 84, 86]. Here, the idea of matching one point is briefly

\(^2\)If the grid used for warping the image did not have an elevation value for each surface element, the values between the grid points are interpolated.
Figure 15: A surface patch representation.

described. For the sake of simplicity, the one dimensional case with two images is discussed. The extension to two dimensions with more than two images is obvious.

Figure 16 shows a schematic description of the geometric situation. For a given point $p_r$ on the left image, the conjugate point $p'_r$ on the right image is sought. If point $p_r$ is an approximation for $p'_r$, an object point $P$, which does not lie on the true surface, is obtained. Using the available surface elevations, the image patches, centered on $p_r$ and $p'_r$ are warped. The projection of the relevant image points onto the warped images is shown on the bottom of Figure 16. The corrected location found by the matching is then projected back to the image (based, again, on the available surface elevations), and a better approximation is obtained for the matched point. Note that as the differences between the approximate and the true surfaces become
Figure 16: Schematic description of the concept of matching warped images.
smaller, the correction will bring the location found by the matching closer to the true location.

In the two dimensional case, when the orientation parameters are not accurately known, the light rays that correspond to the given image points do not necessarily intersect. A common procedure is to calculate an adjusted location of the point in the object space. However, if we intersect the light rays in the object space, and use the adjusted coordinates to create the warped images as described in Section 4.2, the shifts between the images, which may be much larger than those caused by the incorrect elevations, might cause the matching procedure to diverge. These shifts, however, are almost always constant within the matching window. Moreover, they are easily modeled in the following way. If the adjusted object point is projected back to the images, two points, which are different from the original approximations are obtained (Figure 17). The differences between these points and the original ones are stored, and during the warping procedure each point on the distorted grid (see Section 4.2) is shifted by these stored values. This way, the warped patches are centered around the original pair of points and not around the points that were calculated by back projection of the adjusted object point.
Figure 17: Correcting the shift created by non-intersecting light rays.
4.4 Reconstructing Object Space Surfaces by Hierarchical Matching

Matching in the object space requires knowledge of the object surface. The more accurate this knowledge is, the better will be the matching results. However, in order to obtain the surface, matching should be employed. Therefore, an iterative hierarchical approach is taken.

Obtaining an approximation of the surface is possible if the elevations of more points around the desired tie point are known. If external information is not available, the elevations are obtained by matching. Since the approximations of points that are further away from the center point are inaccurate, lower resolution is used for their matching. Although this leads to a lower accuracy of the elevations, the surface model is still more valid than a horizontal or a tilted plane. The hierarchical approach is valid as long as the approximations of all the matched points (in the lowest matching resolution) are within the radius of convergence of the matching algorithm. The matching method avoids points on distinct surface discontinuities, like edges or corners of buildings. Such points are also not desired for the block adjustment. However, unlike with image space matching, different slopes in the terrain within the matching window are easily handled.

Figure 18 demonstrates the idea. It shows the grid points that are used for warping the image patches at four resolution levels. The grid interval is reduced with each iteration. The size of the area required for the whole process is larger than the size of the window used for the final matching. The reason is that if the surface within the
grid points used for warping at iteration:

Figure 18: A scale space approach for matching small warped image patches.
matching window in the highest resolution is required, the elevations at the margin points must be also known. In order to find these points by matching, another window, which only partially overlaps the final window, is used, and so on. This differentiates the hierarchical matching of small surface patches from hierarchical matching of the entire surface. In the latter, only the margins are affected by this problem, while in the former, the margins constitute a large portion of the matched area.

4.5 Matching Multiple Image Patches

In this section, the mathematical model for matching multiple image patches is described. It expands the classical LSM algorithm of using two image patches, to any number of patches. Actually, the proposed model is a simplified version of some theoretical extensions of LSM, discussed in Chapter II.

The model is viable for matching image patches in general. However, the emphasis of the derivations in this section is on matching warped image patches, where only shift parameters are considered.

Let $S$ be a set of $\ell$ warped image patches. These patches are assumed to be scaled and shifted copies of the projection of the ground surface on a horizontal plane (see Section 4.2). Let the patches have the same size, $n \times m$ pixels. If there were neither geometric nor radiometric distortions, these patches would be identical, and therefore, the following identity can be written for each pixel in each image patch in $S$:

$$g^i(r, c) - g^i(r, c) = 0$$

where $g^i(r, c)$ is the observed gray value of the pixel in row $r$ and column $c$ of patch
$i \in S$ and $g^i(r, c)$ is a value which is associated with the intensity of the ground at the corresponding location. It is referenced in the following as the theoretical gray value of the surface element.

In reality, there are geometric and radiometric distortions. Consequently, both distortions must be included in the mathematical model. The systematic radiometric distortions are significantly reduced by a preliminary radiometric correction (e.g., histogram equalization). Since no other information is available, the remaining distortions are assumed to be random. Considering the geometric model and the random radiometric distortions, Equation 4.1 is rewritten:

$$ g^i(T[r, c]) - g^i(r, c) = e(r, c) $$

where $T[r, c]$ is a certain geometric transformation on the pixel coordinates and $e(r, c)$ is the error vector which is assumed to be random. Unlike the general case of LSM, the image patches here are rectified. Therefore the only modeled geometric distortions are horizontal shifts, and Equation 4.2 is reduced to:

$$ g^i(r + \Delta r^i, c + \Delta c^i) - g^i(r, c) = e(r, c) $$

where $\Delta r^i$ and $\Delta c^i$ are the unknown horizontal shifts of patch $i$. Linearizing Equation 4.3 to the standard form of an observation equation gives:

$$ g_0^i(r, c) - g^i(r, c) = g_t^i(r, c)\Delta r^i + g_c^i(r, c)\Delta c^i - \Delta g^i(r, c) $$

where $g_0^i(r, c)$ is an approximation for the theoretical gray value of the surface element at location $(r, c)$, $g_t^i(r, c)$ and $g_c^i(r, c)$ are the intensity gradients across and along the
rows, respectively, and $\Delta g^i(r, c)$ is the unknown correction to the theoretical gray value of the surface element at $(r, c)$. Note that the actual observations in this model are the gray values themselves and not gray-value differences as used in traditional LSM.

Equation 4.4 will be the same for all the image patches but one. In order to avoid a datum defect, the location of one of the image patches must be fixed, and its shifts are set to zero. The observation equations for the pixels of this patch are reduced to

\[ g_0^i(r, c) - g^i(r, c) = -\Delta g^i(r, c) \]  

(4.5)

Note that fixing one patch to its original location is required only to prevent the model from being underdetermined. The solution however is independent on which patch is constrained.

The unknowns of the model presented in Equations 4.4 and 4.5 are $n \cdot m$ theoretical gray values of the surface elements, and $2(\ell - 1)$ shifts of the image patches. Each image patch contributes $n \cdot m$ equations, which brings the total number of equations to $\ell \cdot n \cdot m$. The redundancy of this model is always sufficient. It is also shown (see the appendix) that if only two image patches are used, this mathematical model is equivalent to the classical LSM approach.

The system of Equations 4.4 and 4.5 is formulated as a Gauss-Markov Model:

\[ y = A\xi + e \]  

(4.6)

where $A$ is the design matrix, $y$ is the observations vector, $\xi$ is the vector of unknowns and $e$ is the error vector:

\[ e \sim (0, \sigma^2_0 P^{-1}) \]  

(4.7)
where \( \sigma^2 \) is the variance component and \( P \) is a weight matrix. The solution is obtained by a least-squares adjustment, where the unknown vector is estimated by solving the normal system:

\[
(A^T PA) \hat{\xi} = A^T Py
\]  

(4.8)

Since the model is linearized, the solution is obtained iteratively.

It is obvious that the normal equations system to be solved here is relatively large. The dimension of the normal matrix is \( 2(\ell - 1) + n \cdot m \), which means a computational complexity (for the inversion) of \( O(N^3) \) (where \( N = n \cdot m \) is the number of pixels in an image patch). Since relatively large image patches are used here, the size of the matrix becomes critical. A careful analysis of the model shows that the estimation of the theoretical gray values of the surface elements can be separated from the calculations of the shift parameters. By partitioning the design matrix and the vector of unknowns (in a similar way to what is done in bundle adjustment, where the orientation parameters are calculated first, separately from the ground coordinates, e.g., [7]), the size complexity of the reduced normal matrix is \( O(\ell^3) \), hence depends only on the number of overlapping patches and not on the patch size. The theoretical gray values of the surface elements are estimated later by averaging each pixel across the image patches after resampling them according to the shift corrections obtained from the least-squares adjustment. This procedure is equivalent to a simultaneous solution of all unknowns. Detailed derivations for the model partitioning as well as a proof for the equivalence between the two ways for calculating the theoretical gray values of the surface elements are given in the appendix.
CHAPTER V

EXPERIMENTS AND RESULTS

The concept of *Multiple-Patch Matching in the Object Space for Aerotriangulation*, presented in Chapter IV, is implemented as a prototype computer program. The purpose of this implementation is to check the feasibility of the ideas presented.

Section 5.1 describes the data sets that are used in the experiments. In Section 5.2, the reference data ("ground truth," measured on an analytical plotter and a softcopy station) are explained and presented. Section 5.3 describes the actual experiments and discusses the results.

5.1 Data Sets

Four data sets are used in the experiments:

**OSU**: Stereo model which covers the area of the Ohio State University campus in Columbus. This model reflects a typical urban area, with tall buildings and many man-made features.

**SWISS**: Triplet of photographs, covering a built-up area in Switzerland. This data set is used twice. In one case, only the first two photographs are used. Three
Table 1: Technical information about the data sets used in the experiments.

<table>
<thead>
<tr>
<th></th>
<th>OSU</th>
<th>SWISS</th>
<th>WY</th>
<th>TEXAS</th>
</tr>
</thead>
<tbody>
<tr>
<td># of photographs</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Aerial camera</td>
<td>RMK</td>
<td>RC-20</td>
<td>RMK</td>
<td>RC-20</td>
</tr>
<tr>
<td>Focal length [mm]</td>
<td>~150</td>
<td>~150</td>
<td>~150</td>
<td>~150</td>
</tr>
<tr>
<td>Photo-scale</td>
<td>1:4000</td>
<td>1:2500</td>
<td>1:6000</td>
<td>1:4000</td>
</tr>
<tr>
<td>Scanning resolution [$\mu$m]</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>22.5</td>
</tr>
<tr>
<td>Photograph/image quality</td>
<td>average</td>
<td>excellent</td>
<td>average</td>
<td>good</td>
</tr>
</tbody>
</table>

photographs are used in the second case. To avoid confusion, these two cases are referenced as SWISS-2 and SWISS-3, respectively.

**WY:** Stereo model which covers a mountain area in the state of Wyoming. The area includes large elevation differences, but only a few man-made features. Therefore, the surface is relatively smooth within most of the covered area.

**TEXAS:** A mini-block of six images, organized in two strips of three photographs, covering an air field in Texas.

Figure 19 displays a representative image from each data set and Table 1 contains their technical information. The table shows that all sets are of relatively large scale. At such a scale, most matching procedures face problems because of foreshortening and surface discontinuities. Such problems are less acute at small scales where a solution is relatively easily obtained. The data sets chosen represent different possible scenes that automated aerotriangulation faces, from rural to heavily structured areas.
Figure 19: Examples from the images used for the experiments: (a) photo 192, OSU; (b) photo 3, SWISS; (c) photo 54, WY; (d) photo 67, TEXAS.
5.2 Reference and Approximate Coordinates

In order to obtain both reference (a "ground truth") and approximate coordinates, conjugate points were measured manually for each of the data sets mentioned in Section 5.1. In most cases, the locations of the points were selected around the von-Gruber locations. At each area, a few points were measured, in order to facilitate the detection of incorrect matching results. On the OSU data set, twenty four points, distributed over the entire model were used. Figure 20 shows the distributions of the points for all the data sets.

The OSU and SWISS data sets were measured on a Zeiss P1 analytical plotter, using the original diapositives. For the WY and the TEXAS sets the original photographs were not available. Therefore, measurements took place on an Intergraph softcopy station. The theoretical accuracy of measuring signalized points on an analytical plotter is $6 \mu m$ [57]. For non signalized points, as is the case here, one must assume a lower accuracy, for example $10 \mu m$. Only limited studies concerning the accuracy of measurements on softcopy stations are available. Madani [66] found experimentally that the accuracy depends on the zoom factor used during the measuring process. Based on the limited experience gained in this project, an accuracy of 0.5-1 pixels is assumed, which means approximately 7-15 $\mu m$ for the WY data set, and 11-23 $\mu m$ for the TEXAS data set.

No control points were measured nor used for the experiments because it is very difficult to separate matching errors from control point errors. The emphasis here is to check the matching accuracy.
Figure 20: Point distributions on the data sets.
Figure 20 (continued)
A bundle block adjustment was used to determine the block coordinates of the points, the orientation parameters and the consistency of the manual measurements. Since no control points were used, the datum defect was eliminated by fixing seven parameters. This involved the identification of the block coordinate system with the coordinate system of the first photograph. The scale of the block system is approximately equal to the scale of the photographs.

Table 2 summarizes the results of the bundle adjustment for all the data sets. The accuracy \((\sigma_0)\), and the maximum residual for each data set are shown, along with the statistical information about the configuration of the measured points. Type A refers to points that appear on two photographs, type B refers to three photographs and so on. Note that the accuracies \((\sigma_0)\) are better than the a priori accuracies of the measurements. These values are internal measures of the bundle adjustment. They depend on the configuration of the points within the block, and reflect a distribution of the inaccuracies among all the measurements. Nevertheless, since the experiments were performed with the same points that were measured, using \(\sigma_0\) to compare and estimate the quality of the results is valid.

As noticed from Table 2, the accuracy of the TEXAS data set is somewhat worse than it is for the other sets. There are two possible explanations for this. First, measuring points across strips without marking them on the images is difficult. An attempt was made to measure the points across strips as accurately as possible by creating a "model" with images from different strips. However, measuring exactly
Table 2: Accuracy and statistics of manually measured reference data.

<table>
<thead>
<tr>
<th></th>
<th>OSU</th>
<th>SWISS-2</th>
<th>SWISS-3</th>
<th>WY</th>
<th>TEXAS</th>
</tr>
</thead>
<tbody>
<tr>
<td># of points</td>
<td>24</td>
<td>24</td>
<td>36</td>
<td>24</td>
<td>68</td>
</tr>
<tr>
<td>A points</td>
<td>24</td>
<td>24</td>
<td>25</td>
<td>24</td>
<td>39</td>
</tr>
<tr>
<td>B points</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>C points</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>D points</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E points</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>$\sigma_0$ [$\pm \mu m$]</td>
<td>1.8</td>
<td>3.1</td>
<td>4.0</td>
<td>2.3</td>
<td>5.1</td>
</tr>
<tr>
<td>Max. residual</td>
<td>2.0</td>
<td>3.7</td>
<td>7.1</td>
<td>3.3</td>
<td>11.5</td>
</tr>
</tbody>
</table>

the same point on more than two images is not possible. The second reason is the relatively low resolution (22.5 $\mu m$) available for the images of this set.

5.3 Experimental Results

The experiments in this work are divided into two parts. In the first part (Section 5.3.1) the mathematical model of the multiple-patch matching is checked. Although this model is not new, adding unknown gray values to the mathematical model was done in the past only in the context of general approaches for least-squares matching (see Section 2.4). Here, this model is adopted for single point matching and the experiments examined the effects of radiometric and geometric distortions. The second part of the experiments is concerned with the effect of multiple-patch matching in the object space. Doorn [27] showed that matching in object space works for two images when the exterior orientation parameters of the photographs are known and when the matching is used for an entire stereopair. Here, the orientation parameters
are not known and only a small part of the object surface is reconstructed (around each point). The scheme is first checked with stereo models (Section 5.3.2.1) followed by small blocks to test the multiple-patch matching capability (Section 5.3.2.2).

5.3.1 Multiple-Patch Matching with Simulated Data

In this section, duplicated versions of the same image patch are matched. The patches are shifted with respect to each other and corrupted by random noise. The influence of the following parameters on the matching quality is investigated:

Noise level: An image patch is corrupted with noise. Each pixel of the corrupted image is assigned a new value \( g_n \) according to the weighted average:

\[
    g_n = (1 - w)g_0 + wu
\]

(5.1)

where \( w \) is a weight which ranges from 0 to 0.3, \( g_0 \) is the gray value of the original image and \( u \) is a random number in the range of 0 to 255. Note that a noise level of \( w = 0.3 \) is a relatively high noise level, as Figure 21 demonstrates.

Quality of approximations: The duplications of the patch are shifted in the row and column directions in order to check the radius of convergence. The shifts are randomly selected within a range of \( \pm S_m \) where \( S_m \) is assigned values between 0 and 3 pixels. Larger shifts cause the matching process to diverge in many of the cases.

Image patch size: Different sizes of the matching window are selected, ranging from \( 5 \times 5 \) to \( 55 \times 55 \) pixels.
Figure 21: An example of 0.3 noise level.
(a) original image; (b) the same image, corrupted by random noise (noise level=0.3)
Number of image patches: Different numbers of duplicated image patches are simultaneously matched, ranging between two and six.

With each set of parameters, the matched locations were compared to the true values (shifts applied when image patches were duplicated). This comparison is more meaningful than internal measures, like the mean square error of the LSM, which usually shows much smaller values. The differences for all the points are ordered by size, and the median, the 80% and the 90% values\(^1\) are shown in Figure 22.

The influence of the noise level on the results is shown in charts (a) and (b), where the noise level was ranged from 0 to 0.3. In chart (a) the image patches were not shifted, while in chart (b), shifts of up to 2 pixels were applied. The image patches size is 25 × 25, and six “overlapping” image patches were used at each location. It is clearly seen that even with noise levels of up to 0.2, and with shifts of up to 2 pixels, 90% of the image patches converged to a location, which deviates from the true location by less than 0.35 pixels.

The influence of the quality of the approximations on the results is shown in chart (c). In this case, the noise level was set to 0.2, and the same image patch configuration (six 25 × 25 overlapping patches) was kept. The graphs show that the number of wrong matches increases significantly when the shifts exceed 2 pixels. This can be interpreted as a general indication of the radius of convergence. A more elaborate analysis of the relationships between the content of the image patch, its size and the radius of convergence is required to derive a better understanding about the

\(^1\)with 100% considered to be the number of “overlapping” patches, subtracted by one (the fixed patch) and multiply by the number of points (30).
required accuracy of the approximations. This, however, is beyond the scope of this work. As a further indication of the relationship between noisy image patches and approximations, noiseless patches were matched. Except for one point, the matching in this case always converged exactly to the correct location even when shifts of up to \pm 3 pixels were introduced.

Charts (d) and (e) show the influence of the patch size on the matching results. With a size of $5 \times 5$ pixels, the matching processes diverges in many cases, because there is no sufficient information within the patch. In chart (d) the noise level was set to 0.1 and the maximum shift was set to 1 pixel. The respective values for chart (e) are 0.2 and 3. The higher values of noise and shifts have influenced the results for all patch sizes. However, it is clearly observable that while with a patch size of $15 \times 15$ pixels the value for the "90%" curve in chart (e) is approximately 9 times larger than that value on chart (d), this factor is only 3.5 for the $55 \times 55$ size patch. As expected, larger image patches yield more reliable results.

Chart (f) shows the influence of the number of overlapping patches on the matching. The changes in values on the graphs, especially in the 90% curve (as well as the same phenomena in charts (g) and (h)) are not fully understood. A possible explanation for that situation is the following. Each value on the graphs in this charts is derived from an observation set of different size. In the case of two overlapping image patches for example, 30 "observations" contributed to the values on the graphs; in the case of 6 overlapping patches this number is 150 and so on. Since different data sets are used, it is difficult to compare the values on the graphs on a one-to-one basis.
However, there is no significant trend in the graphs and the changes, even along the "90%" graph, are within 0.2 pixels. It means that the quality of the results does not change when increasing the number of overlapping image patches.

Charts (g) and (h) demonstrate the advantage of matching multiple image patches over matching in pairs. In chart (g), multiple-patch matching was used with shifts of up to 4 pixels and a noise level of 0.2. In chart (h), the same parameters and points were used, but one patch was selected as a template and each of the other patches were matched with it separately. The matching was performed with pairs. The graphs indicate that in the case where matching was performed in pairs, more points converged to wrong locations than in the case where multiple-patch matching was used. This does not come as a surprise because if the image patches have a weak signal, the resulting equations system is less robust. Using more patches simultaneously provides a stronger numerical solution.

The experiments shown in this section lead to the following conclusions:

- Good matching results are obtained even when the approximations are only as good as two pixels.

- In the presence of radiometric differences, better approximations for the locations are required. However, even with a noise level of 0.2 (which is relatively large, according to the definition given in this section), good results are obtained if the approximation is better than 2 pixels.

- With high noise level and bad approximations, larger matching windows increase the matching quality significantly.
Figure 22: The influence of matching parameters on the quality of the results.
Figure 22 (continued)

(d)

(e)
Figure 22 (continued)

(f)

(g)

(h)
• The number of the matched image patches does not influence the accuracy of the results. However, matching all the overlapping image patches simultaneously yields better results than matching them in pairs.

5.3.2 Multiple-Patch Matching in the Object Space with Real Data

The experiments described in this section are aimed at testing the Multiple-Patch Matching in the Object Space for Aerotriangulation. The iterative method is described extensively in Chapter IV. Briefly, three modules are involved:

Warping and Matching Module: where the image patches around the approximate conjugate points are warped, and then matched (in the object space) by the multiple-patch matching algorithm and projected back to the image. In all the iterations except for the last one, more than one point is matched in each area in order to reconstruct the surface.

Bundle Adjustment Module: where the results of the matching are used for adjusting the entire block. An improved set of exterior orientation parameters is then obtained.

DEM Update Module: where all the points matched in the "Warping and Matching Module" are intersected in the object space, using the new orientation parameters. They are then interpolated to create an improved object surface.

In Section 5.3.2.1, the method is tested with only two images at a time, while in Section 5.3.2.2 small blocks are considered.
5.3.2.1 Tests with Stereo Models

In this section the method of matching in the object space is tested independently from the multiple-image patch matching. Only stereo models are used here where at each model point two overlapping image patches are matched. Since the procedure does not involve multiple image patches, the matching itself is equivalent to traditional LSM (see appendix), with the difference that it is performed between warped rather than the original image patches.

As described in Section 4.4, matching in the object space requires the surface within the matching window. Since this surface is not known a priori, it is reconstructed through a hierarchical matching strategy. Three resolution levels were selected for this experiment, with pixel sizes of 240 µm, 120 µm and 60 µm. Higher resolution was not used (although available) in order to allow a comparison with the manually measured points. The selection of three levels is arbitrary, but it demonstrates the procedure. The matching can be performed on more hierarchical levels and on different resolutions.

An accurate surface at each point renders better matching results, since the warped patches have better quality. Obtaining the surface however depends on the size of the patch and the number of matched neighboring points at each resolution. Better matching results are obtained if the density of the points used for determining the surface is large. A large number of points, on the other hand, requires longer processing time. For this experiment, a relatively small number of points was selected. Nevertheless, these points represent the surface much better than a plane.
Using the available approximations of the relative orientation parameters and approximating the surface by a horizontal plane, conjugate image patches at the coarsest resolution level 240 μm, centered on the approximate locations of the conjugate points, were projected to the object space, creating “warped” image patches. Grids with a size of 5 × 5 and intervals of 1920 μm were then formed on both warped image patches. Corresponding points from both grids were matched, leading to the first approximation of the surface which is used for warping the images in the next resolution level. On the warped images at the 120 μm level, a 3 × 3 grid with intervals of 960 μm was defined, centered on the modified approximation point which was obtained during the first iteration. Each grid point was then matched, and a better approximation of the surface is used for the warping in the finest resolution level, where only one point is matched. Table 3 summarizes the statistics for the three iterations, and a graphic description is given in Figure 23.

Three data sets were used for this part of the experiments, OSU, SWISS-2 and WY. For each data set, three cases were tested to check radius of convergence. In the first, approximations for the points were taken directly from the reference coordinates (Section 5.2). In the second and third experiments, shifts of 60 μm and 120 μm, respectively, in random directions, were added to the matched coordinates.

The results are shown in Table 4. In the first row the standard deviations of the bundle adjustment are shown. The second row shows the maximum residual. Next comes the number of points that were rejected during the matching process due to

---

2Note that the scale of the block coordinate system is approximately equal to the scale of the photographs.
Table 3: Size of images, number of grid points and grid intervals at different resolution levels.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel size [μm]</td>
<td>240</td>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td>Size of matching windows [pixels]</td>
<td>13 × 13</td>
<td>23 × 23</td>
<td>45 × 45</td>
</tr>
<tr>
<td>Number of matched points</td>
<td>5 × 5</td>
<td>3 × 3</td>
<td>1</td>
</tr>
<tr>
<td>Grid interval [μm]</td>
<td>1920</td>
<td>960</td>
<td>—</td>
</tr>
</tbody>
</table>

either divergence or large residual in the adjustment results. The last two rows show the average difference $d$ between the resulted matched locations and the original measurements, calculated by:

$$
ar{d} = \frac{1}{n} \sum_{i=1}^{n} (y_{matched} - y_{measured})
$$

(5.2)

where $n$ is the number of accepted points, and the root mean square ($rms$, $\sigma_d$), calculated by:

$$
\sigma_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_{matched} - y_{measured})^2}
$$

(5.3)

Only the $y$ coordinates are compared here since no quality control was applied to the measured $x$ coordinates. In fact, an attempt to compare the $x$ coordinates showed a relatively large systematic error. Such errors can be rooted, for example, in the fact that no “index correction” was considered during the measurements. The matching procedure, however, does not have any preference to one of the axes. Therefore, if
Figure 23: Relations between grid points, patch resolution and matching windows.
Table 4: Results of the object space matching (for two images only).

<table>
<thead>
<tr>
<th>Shift [µm]</th>
<th>OSU</th>
<th>SWISS-2</th>
<th>WY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>σ₀ [±µm]</td>
<td>4.9</td>
<td>4.6</td>
<td>4.2</td>
</tr>
<tr>
<td>Max. residual [µm]</td>
<td>5.5</td>
<td>4.5</td>
<td>4.1</td>
</tr>
<tr>
<td>Number of blunders</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d [µm]</td>
<td>-2.9</td>
<td>-3.7</td>
<td>-2.9</td>
</tr>
<tr>
<td>σ₉ [±µm]</td>
<td>15.4</td>
<td>15.4</td>
<td>15.0</td>
</tr>
</tbody>
</table>

good matching results are obtained in the y direction, the results of the x direction are expected to be similar.

The results presented in Table 4 show that the method of matching in the object space yields results with an accuracy of 1/7 to 1/14 of a pixel. As expected, the matching failed for a certain number of points. This number is slightly larger for the cases where less accurate approximations were available. The experiments conducted here indicated that approximations with an accuracy of 2 pixels (in the highest resolution) still ensure convergence in most cases.

The average differences from the measurements (d) is always close to zero. That means that the matching has virtually no bias. The rms (σ₉) is larger than the standard deviation of the bundle adjustment (σ₀). This difference is explained by the following analysis. The accuracy of the reference data (See Section 5.2) is approximately 10 µm. For the SWISS-2 results for example, σ₉ (as shown in Table 4) is approximately 14 µm. Therefore, the matching accuracy is approximately 10 µm.
Table 5: Results of the image space matching.

<table>
<thead>
<tr>
<th></th>
<th>OSU</th>
<th>SWISS-2</th>
<th>WY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_0 \ [\pm \mu m]$</td>
<td>6.8</td>
<td>10.3</td>
<td>5.4</td>
</tr>
<tr>
<td>Max. residual $[\mu m]$</td>
<td>12.5</td>
<td>12.7</td>
<td>5.7</td>
</tr>
<tr>
<td>Number of blunders</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>$\bar{d} \ [\mu m]$</td>
<td>0.0</td>
<td>-4.8</td>
<td>-3.1</td>
</tr>
<tr>
<td>$\sigma_d \ [\pm \mu m]$</td>
<td>12.0</td>
<td>21.8</td>
<td>10.2</td>
</tr>
</tbody>
</table>

This value is almost twice as much as $\sigma_0$ from the adjustment. This difference is rooted in the fact that $\sigma_0$ is an internal measure of the bundle adjustment, which depends on the geometry of the problem, while the other estimate is a more realistic external measure. A similar situation is observed when comparing the accuracy results of the bundle adjustment with the manually measured data and the practical value for the accuracy of the measurements. Even if the measurement errors had been completely neglected, $\sigma_d$ still shows matching accuracies of $1/4$–$1/6$ of a pixel.

In order to compare the results of the object space matching with traditional image space matching approaches, the manually measured points were used as approximations for image space LSM. As with the last level of the object space matching, the window size was selected to be $45 \times 45$ pixels, and the matching resolution was $60 \ \mu m$. Table 5 shows the results.

Table 6 shows a comparison between matching in image space and object space. Since different sets of points were rejected in the two methods, the input to the bundle adjustment procedure which calculated the accuracies for this comparison did
Table 6: A comparison between the results of object space and image space matching for the same data sets.

<table>
<thead>
<tr>
<th></th>
<th>Object space</th>
<th></th>
<th>Image space</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OSU</td>
<td>SWISS-2</td>
<td>WY</td>
<td>OSU</td>
</tr>
<tr>
<td>$\sigma_0$ [±$\mu$m]</td>
<td>4.9</td>
<td>6.5</td>
<td>5.3</td>
<td>6.8</td>
</tr>
<tr>
<td>Max. residual [$\mu$m]</td>
<td>5.5</td>
<td>9.0</td>
<td>8.1</td>
<td>12.5</td>
</tr>
</tbody>
</table>

not include any of the rejected points. This is the reason for the slight difference between the numbers in this table and those of Tables 4 and 5.

The comparison shows a clear advantage of the object space matching on the image space matching for the first two data sets. Both $\sigma_0$ and the maximum residual are smaller for the object space matching. For the WY data set, there is no significant difference between the results of the two methods. Looking back at the image contents of the data sets, these results are expected. The WY data set covers a smooth rural area. Although there are elevation differences, no significant discontinuities and gradient changes exist within a matching window. Therefore, the mathematical model of the image space matching, i.e., approximating the object surface as a plane, still leads to satisfactory results. On the other hand, the two other data sets, OSU and SWISS-2, contain many man-made features which cause discontinuities and gradient changes in the surface within the matching window. Obtaining a better approximation of the surface and matching in the object space lead to much better results.
5.3.2.2 Tests with Blocks

In this section, the two components of the matching strategy, the multiple image patch and the object-space matching are combined. The experiments here are similar to those in the previous section, except that more than two images are involved. Consequently, more than two image patches are matched at some of the points.

Two data sets were used for this experiment: SWISS-3 and TEXAS. The size of the image patches, the number of grid points at each level, and the number of iterations were exactly the same as shown earlier in Table 3. For SWISS-3, also the resolution levels and the grid intervals remained unchanged. For TEXAS, the original scanning resolution is different (22.5 μm instead of 15 μm, see Table 1). Therefore, minor modifications were necessary. The resolution levels for this data set were 180 μm, 90 μm and 45 μm, and the grid intervals in the first two resolution levels were 1440 μm and 720 μm, respectively. The reference measurements served as approximations for the initial matching locations.

Table 7 shows the accuracy results of the bundle adjustment, based on the photo-coordinates obtained by the matching procedure. The first row shows the values of σ₀ for the two data sets, the second shows the maximum residual, the third lists the total number of observations, and the last row specifies the number of blunders detected in the block adjustment. In contrast to the case of a model, eliminating an observation here does not remove the object point from the adjustment. For example, when a point is eliminated on a particular photograph, the conjugate points on the other photographs are still preserved.
Table 7: Results of the multiple-patch object-space matching.

<table>
<thead>
<tr>
<th></th>
<th>SWISS-3</th>
<th>TEXAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_0$ [±μm]</td>
<td>5.6</td>
<td>4.3</td>
</tr>
<tr>
<td>Max residual [μm]</td>
<td>10.7</td>
<td>13.4</td>
</tr>
<tr>
<td>Original # of obs.</td>
<td>83</td>
<td>188</td>
</tr>
<tr>
<td># of rejected obs.</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

In order to compare the matching results with the results of the manual measurements, observations that were eliminated from the manual measurements or the matching results, were also eliminated in the other set. All in all, 9 observations were eliminated from the SWISS-3 data set (out of 83), and 19 (out of 188) observations were eliminated from the TEXAS data set. The comparison is summarized in Table 8. For the SWISS-3 data set, the results of the manual measurements are better than the matching results. However, the matching results of the TEXAS data sets, where two strips were involved, are comparable to the manual measurements. Note that in both data sets the finest resolution used for matching was relatively low (60 μm and 45 μm respectively), but the accuracy is in the order of 1/10 of a pixel.

5.3.2.3 Summary

The experiments described in Sections 5.3.2.1 and 5.3.2.2 checked the effect of multiple matching in the object space for aerotriangulation. From the experiments, the following conclusions are drawn:
Table 8: A comparison between the results of the multiple-patch matching in the object space and the manually measured points.

<table>
<thead>
<tr>
<th></th>
<th>matching results</th>
<th>manual measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SWISS-3</td>
<td>TEXAS</td>
</tr>
<tr>
<td>$\sigma_0$ [±$\mu$m]</td>
<td>5.6</td>
<td>4.3</td>
</tr>
<tr>
<td>Max residual [$\mu$m]</td>
<td>10.7</td>
<td>12.7</td>
</tr>
</tbody>
</table>

- The accuracy of the matching results is approximately $1/10$ pixel. Although relatively coarse resolution images were used (60 and 45 $\mu$m), the results compare favorably with manual aerotriangulation. If higher resolution images (which are available) were used, better results are expected.

- The proposed matching scheme performs better than traditional matching methods in areas that include surface discontinuities. The results are similar to image space matching when smooth surfaces are involved.

- Since multiple images are simultaneously matched, the method is superior to manual aerotriangulation because a human operator cannot measure on more than two images simultaneously.
CHAPTER VI

CONCLUDING REMARKS

A method for accurate and reliable image matching of points for aerotriangulation was developed and tested. In order to increase the reliability, large image patches that are more likely to contain significant information are necessary. Most matching methods use small image patches because the object surface around the matching area is approximated as a horizontal or tilted plane. In the method presented here, the matching is performed in the object space which minimizes the geometric differences between the matched patches. Consequently, much larger image patches can be used which increases accuracy and reliability.

In order to match in the object space, small surface patches around each point are reconstructed through an iterative procedure. A few points around the desired locations are matched first. Based on these points, the surface patches are iteratively reconstructed and used for warping the image patches. After each iteration, a block adjustment is used to determine improved orientation parameters with the matched points.

The method is derived theoretically and then implemented as a software prototype in order to show its effects on aerotriangulation. The results, presented and discussed in Chapter V, are very encouraging. They show considerable improvement
over matching in image space particularly in cases of surface discontinuities. With the image space approach, more points were either wrongly or inaccurately matched. The bundle block adjustment with points from object space matching consistently showed better results compared to image space matching.

Points that appear on more than two image patches were matched simultaneously. This is a distinct advantage over manual measurements because a human operator cannot measure points on more than two images at a time. The results discussed in Section 5.3.2.2 clearly reflect this advantage. The accuracy obtained from the block adjustment is approximately 1/10 of a pixel. Even with a relatively low resolution it exceeds the accuracy of the manual measurements.

Extensions to the research presented in this work are possible with respect to the following issues:

- Detecting occlusions in the object surface. If a considerable portion of the surface around a point is occluded it should be avoided.

- The sizes of the matching windows and the DEM should be determined dynamically, based on image content for example.

- Extending the tests to larger blocks and to higher resolution levels in order to further test the accuracy behavior.
• Using approximate points that are selected by an automated procedure (see Section 3.3) rather than interactive measurements. This will allow a more objective evaluation of the method, as well as a better automatic analysis of bad matching results.
Appendix

MULTIPLE-PATCH MATCHING—DETAILED DERIVATION OF THE MATHEMATICAL MODEL

A.1 Reducing the Size of the Equation System

In Section 4.5, the multiple patch matching algorithm was formulated as a Gauss-Markov model (Equation 4.6),

\[ y = A\xi + e \]  \hspace{1cm} (A.1)

Using the two types of observation equations (Equations 4.4 and 4.5 for the free and fixed patches respectively):

\[ g^0(r,c) - g^i(r,c) = g^i(r,c)\Delta r^i + g^i_c(r,c)\Delta c^i - \Delta g(r,c) \]  \hspace{1cm} (A.2)

\[ g^0_0(r,c) - g^i(r,c) = -\Delta g^i(r,c) \]  \hspace{1cm} (A.3)

the matrix \( A \) and the vectors \( y \) and \( \xi \) are easily derived.

Due to the relatively large number of unknowns (particularly the unknown theoretical values of the surface elements), the equation system is large and the inversion of the normal matrix is rather time consuming. In order to overcome this problem, a reduced set of equations, obtained by partitioning the model and eliminating
the unknown values of the surface elements is derived. These gray values are calculated once the geometric unknowns (shifts in this case) are obtained, as shown in Section A.2.

The design matrix is partitioned horizontally and gets the following form (with $\ell$ used for the number of overlapping image patches, and $n$ and $m$ are the numbers of rows and columns of each image patch, respectively):

$$ A \doteq \begin{bmatrix} A_1 & A_2 \end{bmatrix} \quad \text{(A.4)} $$

with

$$ A_1 \doteq \begin{bmatrix}
0 & 0 & \cdots & 0 & \cdots & 0 \\
A_{1,1} & 0 & \cdots & 0 & \cdots & 0 \\
0 & A_{1,2} & \cdots & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & A_{1,i} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \cdots & A_{1,\ell-1}
\end{bmatrix}, \quad A_2 \doteq \begin{bmatrix} I_{nm} \\ I_{nm} \end{bmatrix} \quad \text{(A.5)} $$

and where

$$ A_{1,i} \doteq \begin{bmatrix}
g_r^i(0,0) & g_c^i(0,0) \\
\vdots & \vdots \\
g_r^i(0,i) & g_c^i(0,i) \\
\vdots & \vdots \\
g_r^i(n-1,m-1) & g_c^i(n-1,m-1)
\end{bmatrix} \quad \text{(A.6)} $$

The vector of unknowns is partitioned vertically, where the first part contains $2(\ell - 1)$ shift parameters, and the second contains all the corrections to the unknown theoretical gray values of the surface elements:

$$ \xi \doteq \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \quad \text{(A.7)} $$
with

$$\xi_1 = \begin{bmatrix} \xi_{1,1} \\ \vdots \\ \xi_{1,i} \\ \vdots \\ \xi_{1,\ell-1} \end{bmatrix}, \quad \xi_2 = \begin{bmatrix} \Delta g^i(0,0) \\ \vdots \\ \Delta g^i(n-1, m-1) \end{bmatrix}$$  \hspace{1cm} (A.8)

and where

$$\xi_{1,i} = \begin{bmatrix} \Delta r^i \\ \Delta c^i \end{bmatrix}. \hspace{1cm} (A.9)$$

The "observation vector" is unaffected by partitioning the model and has the form:

$$y = \begin{bmatrix} y_0 \\ \vdots \\ y_i \\ \vdots \\ y_{\ell-1} \end{bmatrix}$$  \hspace{1cm} (A.10)

where

$$y_i = \begin{bmatrix} g_0^i(0,0) - g_i^i(0,0) \\ \vdots \\ g_0^i(r,c) - g_i^i(r,c) \\ \vdots \\ g_0^i(n-1,m-1) - g_i^i(n-1,m-1) \end{bmatrix}$$  \hspace{1cm} (A.11)

The observation equations system in its matrix form is therefore:

$$y = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + e$$ \hspace{1cm} (A.12)

and the normal equation system is:

$$\begin{bmatrix} N_{1,1} & N_{1,2} \\ N_{2,1} & N_{2,2} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$ \hspace{1cm} (A.13)

where $N_{i,j} = A_i^T A_j$ and $c_i = A_i^T y$. Solving for $\xi_2$ in the second equation of system A.13 and substituting in the first leads to the estimation of $\xi_1$:

$$\hat{\xi}_1 = (N_{1,1} - N_{1,2} N_{2,2}^{-1} N_{2,1})^{-1}(c_1 - N_{1,2} N_{2,2}^{-1} c_2)$$ \hspace{1cm} (A.14)
and the estimation for $\xi_2$ can be calculated by:

$$\hat{\xi}_2 = N^{-1}_{2,2}(c_2 - N_{2,1}\hat{\xi}_1) \quad (A.15)$$

or as shown in Section A.2.

In the remainder of this section, a solution for the unknown shift parameters is derived. When only shifts are considered, $A_2$ is a set of identity matrices, and $N_{2,2}$ in Equation A.13 gets the simple form of $\ell \cdot I_{nm}$, which leads to a trivial inverse

$$N^{-1}_{2,2} = \frac{1}{\ell} I_{nm} \quad (A.16)$$

The rest of the participating matrices and vectors are shown below:

$$N_{1,1} = A_1^T A_1 = \begin{bmatrix}
A_{1,1}^T & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & A_{1,i}^T A_{1,i} & \cdots & 0 & \cdots \\
\vdots & \vdots & \ddots & A_{1,\ell-1}^T A_{1,\ell-1} & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & A_{1,\ell-1}^T A_{1,\ell-1}
\end{bmatrix} \quad (A.17)$$

$$N_{1,2} = N_{2,1}^T = A_1^T A_2 = \begin{bmatrix}
A_{1,1}^T \\
\vdots \\
A_{1,i}^T \\
\vdots \\
A_{1,\ell-1}^T
\end{bmatrix} \quad (A.18)$$

$$c_1 = A_1^T y = \begin{bmatrix}
S_1^1 \\
S_1^2 \\
\vdots \\
S_i^1 \\
S_i^2 \\
\vdots \\
S_{\ell-1}^1 \\
S_{\ell-1}^2
\end{bmatrix} \quad (A.19)$$
where

\[
\begin{bmatrix}
S_1^i \\
S_2^i
\end{bmatrix} = A_{1,i}^T \hat{y}_i = \begin{bmatrix}
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_r^i(r,c) [g_0^i(r,c) - g^i(r,c)] \\
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_c^i(r,c) [g_0^i(r,c) - g^i(r,c)]
\end{bmatrix} \quad (A.20)
\]

\[
c_2 = A_2^T = \begin{bmatrix}
S_3(0,0) \\
\vdots \\
S_3(n-1,m-1)
\end{bmatrix} \quad (A.21)
\]

where

\[
S_3(r,c) = \sum_{i=0}^{t-1} [g_0^T(r,c) - g^i(r,c)] \quad (A.22)
\]

Simple matrix operations result in the required two components of Equation A.14:

\[
N_{1,1} - N_{1,2} N_{2,2}^{-1} N_{2,1} = \frac{1}{\ell} \begin{bmatrix}
(1 - \ell) A_{1,1}^T A_{1,1} & \cdots & A_{1,1}^T A_{1,j} & \cdots & A_{1,1}^T A_{1,t-1} \\
\vdots & & \vdots & & \vdots \\
A_{1,i}^T A_{1,1} & \cdots & (1 - \ell) A_{1,i}^T A_{1,i} & \cdots & A_{1,i}^T A_{1,t-1} \\
\vdots & & \vdots & & \vdots \\
A_{1,t-1}^T A_{1,1} & \cdots & A_{1,t-1}^T A_{1,2} & \cdots & (1 - \ell) A_{1,t-1}^T A_{1,t-1}
\end{bmatrix} \quad (A.23)
\]

where

\[
A_{1,i}^T A_{1,j} = \begin{bmatrix}
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_r^i(r,c) g_r^j(r,c) & \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_r^i(r,c) g_c^j(r,c) \\
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_c^i(r,c) g_r^j(r,c) & \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_c^i(r,c) g_c^j(r,c)
\end{bmatrix} \quad (A.24)
\]
and

\[
c_1 - N_{1,2}N^{-1}_{2,2}c_2 = \begin{bmatrix}
S^1_1 \\
S^1_2 \\
\vdots \\
S^1_t \\
S^t_{t-1} \\
S^t_{t-1}
\end{bmatrix} - \frac{1}{\ell} \begin{bmatrix}
S^t_1 \\
S^t_2 \\
\vdots \\
S^t_t \\
S^t_{t-1} \\
S^t_{t-1}
\end{bmatrix} \quad \text{(A.26)}
\]

where

\[
\begin{bmatrix}
S^t_4 \\
S^t_5
\end{bmatrix} = \begin{bmatrix}
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g^t(r,c)S_3(r,c) \\
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g^t_c(r,c)S_3(r,c)
\end{bmatrix} \quad \text{(A.27)}
\]

The solution for \( c_1 \) is obtained by inverting the reduced normal matrix (right hand side of Equation A.23) and multiplying it by the free elements vector (right hand side of Equation A.26).

**A.2 Efficient Estimation of the Theoretical Gray Values of the Surface Elements**

Since the solution of the multiple patch matching adjustment is obtained iteratively, the estimation of the theoretical gray values of the surface elements is required at each iteration. Expanding Equation A.15 using the original normal equations system
and $\xi_1$, we can rewrite the expression for $\hat{\xi}_2$:

$$\hat{\xi}_2 = \frac{1}{\ell} \left[ \begin{array}{c} S_3(0, 0) \\ \vdots \\ S_3(r, c) \\ \vdots \\ S_3(n - 1, m - 1) \end{array} \right] - \left[ \begin{array}{cccc} A_{1,1} & \cdots & A_{1,i} & \cdots & A_{1,\ell-1} \end{array} \right] \left[ \begin{array}{c} \Delta r^1 \\ \Delta c^1 \\ \vdots \\ \Delta r^{\ell-1} \\ \Delta c^{\ell-1} \end{array} \right]$$ (A.28)

Although a solution for each pixel can be obtained directly from each line of the equation system A.28, a simplified approach can be taken. For each pixel, the following identities hold:

$$\overline{\Delta g^t(r, c)} = \frac{1}{\ell} \left[ S_3(r, c) - \sum_{i=0}^{\ell-1} (g^i_t(r, c) \Delta r^i + g^i_c(r, c) \Delta c^i) \right]$$

$$= \frac{1}{\ell} \sum_{i=0}^{\ell-1} (g^i_0(r, c) - g^i_t(r, c) - g^i_r(r, c) \Delta r^i - g^i_c(r, c) \Delta c^i)$$ (A.29)

$$\approx g^i_0(r, c) - \frac{1}{\ell} \sum_{i=0}^{\ell-1} (g^i_t(r + \Delta r^i, c + \Delta c^i))$$

$$\Rightarrow g^i_0(r, c) - \overline{\Delta g^t(r, c)} = \frac{1}{\ell} \sum_{i=0}^{\ell-1} g^i_t(r + \Delta r^i, c + \Delta c^i)$$ (A.30)

In other words, the solution for $\hat{\xi}_2$ can be obtained by simply averaging the corresponding pixels from all image patches, after resampling them according to the estimated corrections $\Delta r^i$ and $\Delta c^i$. This resampling is done anyway for the succeeding iteration of the adjustment procedure, and therefore the cost of calculating the estimations for the gray values of the surface elements is minimal. This determination can also be explained intuitively. The solution for the geometric unknowns ensures that the new locations of the image patches are shifted to the locations where the differences from
the theoretical gray values of the surface element are minimal. Given a set of gray
values of corresponding pixels on different image patches, the value which minimizes
these differences is merely their average.

A.3 The case of two image patches

In this section, the equivalence between the multiple patch matching model and tra­
ditional LSM, with a pair of images, is shown.

Starting from Equation A.14, which is the reduced normal equations system for
calculating the geometric parameters, the derivations are rewritten for only two image
patches. Equation A.23 is reduced to

$$\bar{N} = N_{1,1} - N_{1,2}N_{2,2}^{-1}N_{2,1}$$

Assuming that the original approximation for the theoretical gray values of the surface
elements, \(g^0(r,c)\), were taken as the average gray values between the two images,
\((g^0(r,c) + g^1(r,c))/2\), Equation A.26 is reduced to

$$\bar{c} = c_1 - N_{1,2}N_{2,2}^{-1}c_2 = \frac{1}{2} \left[ \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g^1_r(r,c)(g^0_r(r,c) - g^1_r(r,c)) \right]$$

With traditional LSM, assuming only shift parameters, a linearized observation
equation gets the following form:

$$g^0(r,c) - g^1(r,c) = g^1_r \Delta r + g^1_c (r,c) \Delta c$$
where \( g^0(r, c) \) and \( g^1(r, c) \) are the pixel gray values in the template and in the moving patch, respectively; \( g^1_r(r, c) \) and \( g^1_c(r, c) \) are the gray value gradients in the moving patch along and across the rows respectively and \( \Delta r, \Delta c \) are the unknown shift parameters of the moving patch.

Using the Gauss-Markov Model in its matrix representation, the design matrix which corresponds to these observation equations is:

\[
A = \begin{bmatrix}
g^1_0(0, 0) & g^1_c(0, 0) \\
\vdots & \vdots \\
g^1_r(r, c) & g^1_c(r, c) \\
\vdots & \vdots \\
g^1_r(n-1, m-1) & g^1_c(n-1, m-1)
\end{bmatrix}
\]  

(A.34)

the unknown vector is

\[
\xi = \begin{bmatrix}
\Delta r \\
\Delta c
\end{bmatrix}
\]  

(A.35)

and the observation vector:

\[
y = \begin{bmatrix}
g^0_0(0, 0) - g^1(0, 0) \\
\vdots \\
g^0_r(r, c) - g^1(r, c) \\
\vdots \\
g^0_r(n-1, m-1) - g^1(r, m-1)
\end{bmatrix}
\]  

(A.36)

The estimation of the unknown vector \( \xi \) is obtained by

\[
\hat{\xi} = (A^T A)^{-1} A^T y
\]  

(A.37)

Letting

\[
N = A^T A = \begin{bmatrix}
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g^0_r(r, c) g^1_r(r, c) & \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g^0_r(r, c) g^1_c(r, c) \\
\sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g^0_c(r, c) g^1_r(r, c) & \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g^0_c(r, c) g^1_c(r, c)
\end{bmatrix}
\]  

(A.38)
and
\[ c = A^T y = \left[ \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_r^1(r, c)(g_0^0(r, c) - g_1^0(r, c)) \right. \\
\left. \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} g_c^1(r, c)(g_0^0(r, c) - g_1^0(r, c)) \right] \tag{A.39} \]

it can be clearly observed that
\[ N = 2N^0 \tag{A.40} \]
\[ c = 2\overline{c} \tag{A.41} \]

and therefore
\[ N^{-1}c = \overline{N}^{-1}\overline{c} \tag{A.42} \]

This result shows the equivalence between the multiple patch matching presented in this work to the traditional LSM. The advantage of the proposed scheme is its ability to include more than two image patches simultaneously.
BIBLIOGRAPHY


