Sub-pixel Registration In Computational Imaging And Applications To Enhancement Of Maxillofacial Ct Data

2006

Murat Balci

University of Central Florida

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Part of the Computer Sciences Commons, and the Engineering Commons

STARS Citation

https://stars.library.ucf.edu/etd/1084

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
SUB-PIXEL REGISTRATION IN COMPUTATIONAL IMAGING AND APPLICATIONS TO ENHANCEMENT OF MAXILLOFACIAL CT DATA

by

MURAT BALCI
B.S. Middle East Technical University, 1999
M.Sc. Middle East Technical University, 2001

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Electrical Engineering and Computer Science in the College of Engineering and Computer Science at the University of Central Florida Orlando, Florida

Fall Term
2006

Major Professor:
Hassan Foroosh
ABSTRACT

In computational imaging, data acquired by sampling the same scene or object at different times or from different orientations result in images in different coordinate systems. Registration is a crucial step in order to be able to compare, integrate and fuse the data obtained from different measurements. Tomography is the method of imaging a single plane or slice of an object. A Computed Tomography (CT) scan, also known as a CAT scan (Computed Axial Tomography scan), is a Helical Tomography, which traditionally produces a 2D image of the structures in a thin section of the body. It uses X-ray, which is ionizing radiation. Although the actual dose is typically low, repeated scans should be limited. In dentistry, implant dentistry in specific, there is a need for 3D visualization of internal anatomy. The internal visualization is mainly based on CT scanning technologies. The most important technological advancement which dramatically enhanced the clinician’s ability to diagnose, treat, and plan dental implants has been the CT scan. Advanced 3D modeling and visualization techniques permit highly refined and accurate assessment of the CT scan data. However, in addition to imperfections of the instrument and the imaging process, it is not uncommon to encounter other unwanted artifacts in the form of bright regions, flares and erroneous pixels due to dental bridges, metal braces, etc. Currently, removing and cleaning up the data from acquisition backscattering imperfections and unwanted artifacts is performed manually, which is as good as the experience level of the technician. On the other hand the process is error prone, since the editing process needs to be performed image by image.

We address some of these issues by proposing novel registration methods and using stone-cast models of patient’s dental imprint as reference ground truth data. Stone-cast models were
originally used by dentists to make complete or partial dentures. The CT scan of such stone-cast models can be used to automatically guide the cleaning of patients’ CT scans from defects or unwanted artifacts, and also as an automatic segmentation system for the outliers of the CT scan data without use of stone-cast models. Segmented data is subsequently used to clean the data from artifacts using a new proposed 3D inpainting approach.
to my parents Ayse and Sabri Kurtel
ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Hassan Foroosh for his valuable guidance, encouragement and friendship. His generous and sincere support helped me develop myself as an individual and scientist.

During my studies at University of Central Florida, I had great colleagues and friends, and beyond these great people I had the chance of personally coauthoring with Mais Alnasser and Xiaochun Cao. I want to thank them for sharing their time and intellect with me. I also want to thank Imran Junejo and Alex Cook for proof reading and contributing to this dissertation.

I would like to thank Prof. Charles Hughes, Dr. Taskin Kocak, and Dr. Kazumasa Yamazawa for their valuable advices and being part of my dissertation committee.
# TABLE OF CONTENTS

**LIST OF FIGURES** .......................................................... x

**LIST OF TABLES** .......................................................... xiv

**CHAPTER 1  INTRODUCTION** ............................................. 1
1.1 MOTIVATION ............................................................... 1
1.2 ORGANIZATION OF THIS THESIS ..................................... 3
1.3 BACKGROUND ............................................................. 4
1.4 CLASSIFICATION AND OVERVIEW OF EXISTING REGISTRATION ALGORITHMS IN MEDICAL IMAGING ................................. 8
  1.4.1 Dimensionality ....................................................... 9
  1.4.2 Nature Of Registration Basis .................................... 10
  1.4.3 Nature And Domain Of Transformation .......................... 17
  1.4.4 Interaction .......................................................... 18
  1.4.5 Optimization Procedure .......................................... 19
  1.4.6 Modalities Involved In The Registration ...................... 21
  1.4.7 Subject ............................................................. 22
  1.4.8 Object ............................................................. 23
1.5 REMARK ON SUB-PIXEL REGISTRATION .............................. 23
CHAPTER 2 SUB-PIXEL ESTIMATION OF SHIFTS

2.1 SUB-PIXEL ESTIMATION OF SHIFTS DIRECTLY IN THE FOURIER DOMAIN

2.1.1 Introduction

2.1.2 The Phase Correlation Method

2.1.3 Discrete Phase Difference

2.1.4 Solving For Image Shifts

2.1.5 Robustifying The Solution

2.1.6 Experimental Results

2.1.7 Conclusion

2.2 SUB-PIXEL REGISTRATION DIRECTLY FROM THE PHASE DIFFERENCE

BY USE OF HOUGH TRANSFORMATION

2.2.1 Problem Formulation

2.2.2 Proposed Solution

2.2.3 Experimental Results

2.2.4 Conclusion

CHAPTER 3 ALIGNMENT OF MAXILLOFACIAL CT SCANS TO STONE-CAST MODELS USING 3D SYMMETRY FOR BACKSCATTERING ARTIFACT REDUCTION

3.1 INTRODUCTION

3.2 MIRROR SYMMETRY

3.3 FRAMEWORK
LIST OF FIGURES

1.1 Classification of Medical Image Registration ........................................ 9
1.2 Examples of 2D transformations .......................................................... 19

2.1 (a) & (b) Two aerial images with some shifts, (c) phase difference matrix corresponding to shifts of (7.3,5.6) pixels, (d) phase difference matrix corresponding to shifts of (30.5,25.4) pixels, (e) & (f) one row of the phase matrices shown in (c) & (d), respectively ................................................................. 32
2.2 Some of the images used for simulation .................................................. 43
2.3 Simulation with additive noise: (a) image used for synthesizing shifted noisy pairs, (b) a typical phase difference matrix for a noisy pair  ......................................................... 44
2.4 Average SNR over 50 independent trials for the standard deviation of noise varying in [0, 5]: top shift along x-axis, bottom shift along y-axis .............................................. 45
2.5 (a) An image from the MRI data set (Courtesy of W.S. Hoge), (b) discrete phase difference matrix of the first pair in the data set ................................................. 46
2.6 The relative error in our registration compared to the physical calibration, Stone et al., and Hoge, for the MRI data set ................................................................. 47
2.7 (a) The Hough transform of the phase discontinuities, (b) the local maxima and the threshold, Peaks detected by thresholding .................................................. 52
2.8 Some of the images used for simulation .................................................. 57
2.9   (a) A phase difference matrix corrupted by aliasing and high-frequency noise, (b) average PSNR over 100 independent trials with varying window sizes and the standard deviation of noise $\sigma=5$. 

2.10 Pentagon stereo pair and the estimated local motion. 

2.11 Baseball stereo pair and the estimated local motion. 

2.12 Tsukuba stereo pair and the estimated local motion. 

2.13 A synthetic stereo pair and the PSNR as a function of the window size used for short length Fourier transform in local motion estimation. 

2.14 Top and middle: Stereo pair and their corresponding polar representations. Bottom: Their phase difference matrix and one row of the matrix. 

3.1   (a) 2D view for the computed plane of symmetry in the stone-cast CT scans (b) Original Patient Model (c) Corresponding Stone Cast Model. 

3.2   (a) Initial Models with planes of symmetry (b) Models after planes of symmetry are matched (c) Models after planes of symmetry are matched (Top View). 

3.3   (a) A section of zoomed Patient Data (b) Partial zoom view of corrected data by use of stone-cast matching (c) Patient Model after clean up. 

3.4   (a) Models matched together without any pruning of patient data (b) Stone Cast model and cleaned up patient data merged together. 

3.5   A slice of patient data. 

3.6   Gray level distribution of a slice. 

3.7   Single thresholding an axial view slice and its effect on segmentation. 

3.8   Gray level distribution histogram of complete 3D data set.
3.9 (a) Histogram of a complete CT Scan data set (b) Histogram after background is removed (c) histogram after tissues removed (d) histogram after final outlier threshold is applied. One may consider our approach as if focusing to the least occurring segment and looking for an automatic threshold to differentiate between outliers and the actual data. Please note that there are just a few hundred voxels marked as outlier in $\sim 70$ millions.

3.10 An example of sample data, corresponding outliers marked as green and a result of outlier removal is shown. Note that the color range is also improved by outlier removal as well.

3.11 After searching for a good threshold for the last block of original histogram, we are able to detect outliers. Examples of this kind of segmentation are shown in the figures.

3.12 Notation Diagram. Given the patch $\Psi_p$, $\mathbf{n}_p$ is the normal to the contour $\delta \Omega$ of the target region $\Omega$ and $\nabla I_p^\perp$ is the isophote (direction and intensity) at point $\mathbf{p}$. The entire image is denoted with $\mathbf{I}$ [CPT03].
3.13 Structure propagation by exemplar-based texture synthesis. (a) Original image, with the target region $\Omega$, its contour $\delta\Omega$, and the source region $\phi$ clearly marked. (b) Criminisi et al. synthesize the area delimited by the patch $\Psi_p$ centered on the point $p \in \delta\Omega$. (c) The most likely candidate matches for $\Psi_p$ lie along the boundary between the two textures in the source region. (d) The best matching patch in the candidates set has been copied into the position occupied by $\Psi_p$, thus achieving partial filling of $\Omega$. The target region $\Omega$ has shrank and its front has assumed a different shape.

B.1 (a) Discriminant of the characteristic equation and its linear approximation, (b) ratios of the singular values: dotted line: true ratio, solid line: approximated ratio.
## LIST OF TABLES

2.1 Table of results for shifts of the images in Figure 2.2 ......................... 43

2.2 Pairwise registration of the MRI data set. ........................................... 46

2.3 Table of results for global shifts of the images in Figure 2.8 .................... 56

3.1 Inpainting Algorithm Proposed by Criminisi[CPT03] ............................. 89

3.2 Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting 91

3.3 Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting 92

3.4 Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting 93

3.5 Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting 94

3.6 Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting 95

3.7 Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting 96

3.8 Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting 97
CHAPTER 1

INTRODUCTION

Medical imaging is the process by which the human body and the parts thereof are imaged for the purpose of evaluation and visualization. Medical imaging may be clinically motivated, seeking to diagnose and examine diseases in specific patients or it may be used by researchers in order to understand physiological processes in living organisms. [wik06]

Medical imaging techniques have made tremendous progress since the discovery of x-rays by Roentgen in 1895. The first medical application of x-rays was a radiograph of his wife’s hand and wedding ring [Par82]. Since then many other technological developments have been achieved that provide different imaging modalities such as ultrasound, nuclear medicine, thermography, and nuclear magnetic resonance providing alternative and often complementary information about human body.

1.1 MOTIVATION

In computational imaging, various data sets acquired by sampling the same scene or object at different times or from different orientations result in images in different coordinate systems. Image registration is the process of transforming different sets of data into a common coordinate system [wik06]. Registration is a crucial step in order to be able to compare, integrate, and fuse the data obtained from different measurements.

Tomography is the method of imaging a single plane, or slice, of an object. Computed tomography (CT), originally known as computed axial tomography (CAT) and body section roentgenography, is a medical imaging method employing tomography, where digital geometry processing is
used to generate a three-dimensional image of the internal elements of an object from a large series of two-dimensional X-ray images taken around a single axis of rotation. CT produces a series of axial images which can be manipulated through a process known as windowing, in order to recreate the image in a different plane. The CT system was invented by Godfrey Newbold Hounsfield using X-rays. Hounsfield conceived the idea in 1967, and it was publicly announced in 1972. Allan McLeod Cormack of Tufts University independently invented the same process and they shared a Nobel Prize in medicine in 1979 [wik06].

In dentistry, implant dentistry in specific, there is a need for 3D visualization of internal anatomy. The internal visualization is mainly based on CT scanning technologies. As discussed in [DMD01] the most important technological advancement which dramatically enhanced the clinician’s ability to diagnose, treat, and plan dental implants has been the CT scan. Advanced 3D modeling and visualization techniques permit highly refined, accurate assessment of the CT scan data in a manner which exceeds information gleaned from film alone [DMD01, SHS04, GP04]. Since the advent of the Conebeam CT technology, which is a low dose oral and maxillofacial scanner, the number of dental CT scanners has been growing exponentially. The strongest barrier in taking a CT scan is the dose of x-ray exposure, which is not always justifiable in oral surgery cases. Conebeam scanners have an average of 10 times less radiation than regular medical CT scanners [AEa01]. However, this comes at a price that the resulting scans are not free of defects.

In addition to imperfections of the instrument and the imaging process, it is not uncommon to encounter other unwanted artifacts in the form of bright regions, flares, and erroneous pixels due to, for instance, dental bridges or metal braces. Currently, removing and cleaning up the data from acquisition imperfections and unwanted artifacts is performed manually, which is as good as the
experience level of the technician. On the other hand the process is error prone, since the editing process needs to be performed image by image.

We propose new solutions for reduction and elimination of imperfections and scanning errors and other related issues by multiple techniques such as using stone-cast models of patient’s dental imprint, outlier analysis, 3D inpainting techniques, and applications of image processing algorithms.

Stone-cast models were originally used by dentists to make complete or partial dentures. The CT scan of such stone-cast models can be used to automatically guide the cleaning of patients’ CT scans from defects or unwanted artifacts. The key, however, to such applications would be an accurate registration of the data from two CT scans, i.e. the patient’s maxillofacial CT scans and that of the stone-cast model.

In the absence of stone-cast data, an automatic thresholding system which lets detection of outliers is required which is also applicable with the stone-cast approach for removal of the outliers that are in the boundary defined by the registered stone-cast. Removal of these outliers introduces one more complications since it requires detection of the correct intensity levels for those removed voxels.

1.2 ORGANIZATION OF THIS THESIS

The previous section provided a brief introduction and the underlying motivation. In this chapter we start with a background summary on image registration, steps of image registration, and classification of registration algorithms in medical imaging. Chapter 2 includes two specific algorithms developed for sub-pixel image registration. Chapter 3 focuses on application of image registra-
tion techniques in medical imaging, including alignment of maxillofacial CT scans to stone-cast models using 3D symmetry and automatic thresholding and inpainting.

The registration algorithms developed are universal and can be applied to a wide range of imaging modalities, including medical imaging as shown in Chapter 2. The core topics are alignment, segmentation, and post-processing of images under rigid transformations. As a result, the problem can be considered in two parts: the first is sub-pixel image registration and the second is registration of medical CT scans in order to enhance the medical data in hand.

1.3 BACKGROUND

Image registration is the process of establishing point-by-point correspondences between two images or image sequences of a scene [Gos05]. Although the classical literature has been mostly devoted to registration of 2D images, for medical imaging it is mostly required to register 3D data such as CT scans, MRI, or across different types of 3D modalities.

Image Registration can be used to extract information about an extra dimension such as extracting 3D depth information by registering 2D stereo images of a scene or registration of 3D CT scans of a patient to extract information about the process in time. Registration of different types of data can be used to extract complementary information by exploiting different modalities. For instance registration of MR and PET data of the brain will combine anatomy and functionality together.

Registration is a crucial step in the analysis and fusion of information between multiple images. Registration techniques typically assume that motion can be modeled using a given family of transformations such as projective, affine, or Euclidean. Registration is then performed by searching
for a particular transformation within the family that optimizes some similarity or redundancy cri-
terion, e.g. correlation coefficients or mutual entropy. Examples can be found in remote sensing,
robotics, and bio-medical imaging, among others [Bro92, MV98, ZF03]. In some applications, such as
coding and compression registration needs to be established locally, while other applica-
tions require only global registration.

In this section we will use the term “data set” instead of images to often indicate general imaging modalities as opposed to medical images. This chapter classifies registration algorithms based on the steps required as suggested by [Gos05]:

**Preprocessing** is the initial processing step of the source and destination data sets before feature extraction. It may include scale adjustment, noise removal and segmentation of the given source and destination data sets.

Scale differences, noise in the images, motion blur, haze, and sensor nonlinearities may arise in image registration.

**Feature Selection** is the process of finding interesting and useful feature types for a given domain and then the coordinates of those feature points in the given data sets. Some examples are corners, lines, curves, templates, regions, patches, etc. There is no optimal feature selection algorithm and the domain of the images affects this selection process. For instance, in man-
made scenes, line segments may work fine, but for satellite images, contours and regions may be more appropriate.

**Feature Correspondence:** at this step features extracted from the reference data set are searched in input data sets, and an optimal correspondence is determined. Matching the coordinates
of these features are used for determining the transformation.

**Determination of Transformation Functions** is the final step in finding an appropriate transformation model for matching feature points. To represent features a different metric is used according to the type of the feature. For templates we use the center of the template, for regions the center of gravity of that region, and for lines the intersections of corresponding line pairs. For curves locally maximum curvature points on corresponding curves may be used to represent corresponding transformation pair points. The transformation function then depends on the geometric differences between the reference and the sensed data sets.

As a result, overall registration algorithms can be classified as follows [wik06]:

**Area-based vs Feature-based** Image registration algorithms broadly fall within two categories: area based methods and feature based methods. The original image is often referred to as the reference image and the image to be mapped onto the reference image is referred to as the target image. For area based image registration methods, the algorithm looks at the structure of the image via correlation metrics, Fourier properties, and other means of structural analysis. However, most feature based methods, instead of looking at the overall structure of images, fine tune the mapping to the correlation of image features such as lines, curves, points, line intersections, boundaries, etc. [wik06]

**Transformation model** Image registration algorithms can also be classified according to the transformation model used to relate the reference image space with the target image space. The first broad category of transformation models includes linear transformations, which are a combination of translation, rotation, global scaling and shear components. Linear transformations are global in
nature, thus not being able to model local deformations. The second category includes “elastic” or “nonrigid” transformations. These transformations allow local warping of image features, thus providing support for local deformations.[wik06]

**Image nature** Another useful classification is between single-modality and multi-modality registration algorithms. Single-modality registration algorithms are those intended to register images of the same modality (i.e. acquired using the same kind of imaging device), while multi-modality registration algorithms are those intended to register images acquired using different imaging devices.[wik06]

There are several examples of multi-modality registration algorithms in the medical imaging field. Examples include registration of brain CT/MRI images or whole body PET/CT images for tumor localization, registration of contrast-enhanced CT images against non-contrast-enhanced CT images for segmentation of specific parts of the anatomy, and registration of ultrasound and CT images for prostate localization in radiotherapy.[wik06]

**Other classifications** Further ways of classifying registration algorithms consist of the amount of data they are optimized to handle, their application, and their underlying central theory. Image registration has applications in remote sensing (cartography updating), medical imaging (change detection, tumor monitoring), and computer vision. Due to the vast applications to which image registration can be applied, it is impossible to develop a general algorithm optimized for all uses.[wik06]

**Image similarity-based methods** Image similarity-based methods are broadly used in medical imaging. A basic image similarity-based method consists of a transformation model which is applied to reference image coordinates to locate their corresponding coordinates in the target image.
space, an image similarity metric which quantifies the degree of correspondence between features in both image spaces achieved by a given transformation, and an optimization algorithm which attempts to maximize image similarity by changing the transformation parameters.[wik06]

The choice of an image similarity measure depends on the nature of the images to be registered. Common examples of image similarity measures include Cross-correlation, Mutual Information, Mean-square difference and Ratio Image Uniformity. Mutual Information and its variants, e.g. Normalized Mutual Information, are the most popular image similarity measures for registration of multi-modality images. Cross-correlation, Mean-square difference, and Ratio Image Uniformity are commonly used for registration of images of the same modality.[wik06]

In the following section we focus specifically on the classification of medical image registration and the previous work in this area.

1.4 CLASSIFICATION AND OVERVIEW OF EXISTING REGISTRATION ALGORITHMS IN MEDICAL IMAGING

A registration procedure can be decomposed into three major parts: the problem statement, the registration paradigm, and the optimization procedure. The problem statement and the choice of paradigm and optimization procedure together provide a unique classification according to nine criteria which is based on the criteria formulated in Maintz et. al [MV98]. In this section we will introduce those classifications, as a summary of Maintz et. al [MV98]. Please refer to the related publication for more information.

Classification of a registration algorithm can be a difficult process due to many cross influences between parts and criteria. But we can say that: the “problem statement” is the major criteria for the classification criterions VII, VIII, and IX and has an effect on the criterions I and III. The
Figure 1.1: Classification of Medical Image Registration

“paradigm” imposes the criteria II, III, IV, and V, while the “optimization procedure” defines the classification decision based on criterions V and VI (Figure 1.1). It should be noted that the “problem statement”, “paradigm” and “optimization procedure” are presented in a compounded way and they are intertwined [FHM00].

### 1.4.1 Dimensionality

The spatial registration methods can be categorized according to whether all dimensions are just spatial or whether time is an added dimension. In either case, the problem can be further categorized depending on the number of spatial dimensions involved (2D, 3D, etc) [MV98].
1.4.1.1 Spatial Registration Methods

Current publications focus on the 3D-3D registration of two images [MV98]. 3D-3D registration corresponds to cases like registration of two tomographic three dimensional data sets or the registration of a single tomographic image to any spatially defined information such as a vector obtained from electroencephalogram (EEG) data. 2D-2D registration may apply to registration of separate slices from computed tomographic data. As a result, 2D-2D registration is less complex by an order of magnitude compared to 3D-3D registration. 2D/3D registration is applied for the direct alignment of spatial data to projective data, i.e a pre-operative CT image to an intra-operative X-ray image, or the alignment of a single tomographic slice to spatial data. Since most 2D/3D applications concern intra-operative procedures, they are heavily time constrained and consequently have a strong focus on speed issues related to the computational time of the paradigm and the optimization. The majority of applications allow for off-line registration, so speed issues need only be addressed as constrained by clinical routine.

1.4.1.2 Registration Of Time Series

Time series of images are acquired for various reasons, such as monitoring of bone growth in children, monitoring of tumor growth, or post-operative monitoring of healing, etc [MV98].

1.4.2 Nature Of Registration Basis

Image based registration methods can be divided into extrinsic, intrinsic, and non-image based methods [MV98]. These are described in the following sections.
1.4.2.1 Extrinsic Registration Methods

Extrinsic methods are based on artificial objects attached to the patient, which are designed to be well visible and accurately detectable in all of the modalities.[MV98]

Acquired in this way, the registration of these images is comparatively easy, fast, and can usually be automated. Moreover there is no need for complex optimization algorithms, since the registration parameters can often be computed explicitly. The main drawbacks of extrinsic registration are the prospective character (i.e. provisions must be made in the pre-acquisition phase), and the often invasive character of the marker objects. Non-invasive markers can be used, but as a rule are less accurate.[MV98]

A commonly used fiducial object is a stereotactic frame [Lun88, Van91, LKH94, LJ94, SAX94, HES95b, VCM95, PDM96] screwed rigidly to the patient’s outer skull table. As stereotactic frame is a device which until recently provided the best “gold standard” for registration accuracy. Such frames are used for localization and guidance purposes in neurosurgery. Since neurosurgery is one of the main application areas of registration, the use of a stereotactic frame in the registration task does not add an additional invasive strain to the patient. However, the mounting of a frame for the sole purpose of registration is not permissible. Sometimes other invasive objects are used, such as screw-mounted markers [GV93, LHM93, MMF93, LPC94, MAD94, MFG95, MAD95, SOB95, ETM96], but usually non-invasive marking devices are reverted to. Most popular amongst non-invasive markers are glued to the skin [EMT91, MNR91, MMG93, WFM94, WQC93, BSH94, EHH95a, EHH95b, LBM95, SCE95, WFM95, FWN96]. In addition, larger devices that can be fitted snugly to the patient such as individualized foam moulds, head holder frames, and dental adapters have also been used [GtD80, LLF85, SBS87, HHB92, EMC89, EMT91].[MV98]
Due to the rigid-transformation constraint they impose in most of the applications and due to various practical considerations, use of extrinsic 3D-3D methods is largely limited to brain and orthopedic imaging [SOB95, ETM96]. Markers can often be used in projective imaging of any body area. Non-rigid transformations can in some cases be obtained using markers such as in studies of animal heart motion, where markers can be implanted into the cardiac wall.[MV98]

1.4.2.2 Intrinsic Registration Methods

Intrinsic methods rely on patient generated image content only. Registration can be based on a limited set of identified salient points, i.e. landmarks; on the alignment of segmented binary structures such as object surfaces; or directly onto measures computed from the image grey values. These types of intrinsic registration methods are discussed in the following sections.[MV98]

Landmark Based Registration Methods

Landmarks can be anatomical or geometrical. The anatomical landmarks include salient and accurately locatable points of the morphology of the visible anatomy, usually identified interactively by the user [EMC89, EMT91, HHC91, HHH91, MNR91, ZTZ91, HCC92, Bij93, DSG93, FL93, GS93, HHH93b, MME93, NCH93, WQC93, GFV94, HVE94, MM94, PHF94, SAX94, EHH95a, EHH95b, GFK95, HSC95, HSe95a, LBM95, MLB95, MK95, SBC95, SGR95, SCE95, VCM95, ZSI95, CKM96, ESC96, ECN96, EKo96, FRR96, PDM96, RKP96]. The geometrical landmarks include the points at the locus of the optimum of some geometric property such as local curvature extrema, corners, etc. generally localized in an automatic fashion [HMC91, FCB93, AS94, EST94b, Thi94, AS95, UK95, AK96, CJ96, Thi96a].[MV98]

Landmark based registration is versatile since it can be applied to any image, no matter what the object or subject is. Landmark based methods are mostly used to find rigid or affine transfor-
mations. If the sets of points are large enough, they can theoretically be used for more complex transformations. The major drawback of landmark based registration is that user interaction is usually required for the identification of the landmarks.[MV98]

**Segmentation Based Registration Methods**

Segmentation based registration methods can be rigid model based [CPC87, LPC88, zA92, JRH92, AzT93, CrV93, Fri93, GRB93, GBB94, GBB95a, GBB95b, GH96, GH93, HHH93a, KCP93, MCA93, RTL93, TRG93, TJP93, ZYG93, CVS94, EGr94, EGL94, FA94, FPM94, FPC94, GrW94, GLI94b, GLI94a, HSP94a, HSP94b, HC94, HSD94, HSB94, HK94, Kan94, KMZ94, KHB94, eST94a, LPE94, MAD94, MCN94, rFe94, Phi94, PKF94, SHK94, SB94, SL94b, SL94c, SMD94, SAX94, SX94, THC94, WTZ94, ZAB94, ABH95, ASV95, And95, BF95, BFA95, BHC95, CKP95, FAB95, GEW95, HCZ95, HES95b, HSE95b, HNS95, HSC95, HSe95a, HeS95a, KB95, LS95, LLD95, MAD95, PBS95, PGB95, PG95, PT95, REL95, RGP95, SOB95, SHK95, SB95, SMZ95, SA95, TLV95, THJ95, VTL95, VCM95, XJ95, ZSI95, DFB96, ESC96, EGL96, FA96, GEW96, GVH96, GMF96, GDF96, HSE96, JZL96, eTS96, eSB96, QMH96, SL96, WCC96]. Anatomically the same structures are extracted from both images to be registered, and used as sole input for the alignment procedure. They can also be nonrigid deformable model based [BLR83, Zie93, Tau93, DP94, MAE94, SL94a, TEK94, Bro95, BCC95, MTB95, SL95, Thi95, CTM96, DPB96, Dav96, MT96, Thi96b], where extracted structures, such as surfaces or curves, from one image are elastically deformed to fit the second image [MV98].

The rigid model based approaches are the most popular methods currently in clinical use. This is probably for a large part due to the success of the “head-hat” method as introduced by Pelizzari and co-workers [CPC87, LPC88, PCS89, CP89], which relies on the segmentation of the skin
surface from CT, MR and PET images of the head [MV98]. Segmentation based methods remained popular, since it is fairly easy to perform and the computational complexity is relatively low. Therefore, many follow-up papers aimed at automating the segmentation step, improving the optimization performance, or otherwise extending the methods that have been published. Another cause for its popularity is the fast Chamfer matching technique for alignment of binary structures by means of a distance transform, introduced by Borgefors [Bor88]. A drawback of segmentation based registration is that the registration accuracy is limited to the accuracy of the segmentation step. In theory, segmentation based registration is applicable to images of many areas of the body. However, in practice the application areas have largely been limited to neuroimaging and orthopedic imaging. The methods are commonly automated. However, most of the times, the segmentation step is performed semi-automatically.[MV98]

The optimization criterion is different with deformable models. It is always locally defined and computed, and the deformation is constrained by elastic modeling constraints imposed onto the segmented curve or surface. Deformable curves appear in literature as snakes or active contours. Sometimes 3D deformable models are referred to as nets. Opposed to registration based on extracted rigid models, which is mainly suited for intrasubject registration, deformable models are in theory very well suited for intersubject and atlas registration, as well as for registration of a template obtained from a patient to a mathematically defined general model of the templated anatomy. A disadvantage of deformable models is that they often need a good initial position in order to properly converge, which is generally realized by rigid pre-registration of the images involved. Another disadvantage is that the local deformation of the template can be unpredictably erratic if the target structure differs sufficiently from the template structure. A typical error is that
the deformable model matches the anatomy perfectly, except in one interesting image area where a large tumor growth has appeared. [MV98]

**Voxel Property Based Registration Methods**

The voxel property based registration methods operate directly on the image grey values, without prior data reduction by the user or segmentation; therefore they stand apart from the other intrinsic methods. Intensity based methods approach registration in two ways. The first way is to immediately reduce the image grey value content to a representative set of scalars and orientations, whereas the second way is to use the full image content throughout the registration process. [MV98]

One of the prime examples of reductive registration methods are the Principal Axes and Moments Based Methods. Within these methods the image’s center of gravity and its principal axes are computed from the image zeroth and first order moments. Registration is then performed by aligning the center of gravity and the principal orientations [ABK90, BT94, EGr94, EGL94, aRC94, WF94, SHS95, DB96, WVL96]. The result is usually not very accurate, and the method is not equipped to handle differences in scanned volumes well. Because of the automatic and very fast nature of its use and the easy implementation, these methods are widely used in registration problems that require no high accuracy, such as in the re-alignment of scintigraphic cardiac studies and as a coarse pre-registration in various other registration areas. Moment based methods also appear as hybridly classified registration methods that use segmented or binarized image data for input. In many applications, pre-segmentation is mandatory in order for moment based methods to produce acceptable results. [MV98]

Voxel property based methods using the full image content are the most flexible of registration methods theoretically, since they do not start with reducing the grey valued image, but use all of the
available information throughout the registration process. Although voxel property based methods have been around a long time, their use in extensive 3D-3D clinical applications has been limited by the considerable computational costs. An increasing clinical call for accurate and retrospective registration, along with the development of ever-faster computers with large internal memories, has enabled full-image-content methods to be used in clinical practice. However, they have not yet been introduced in time-constrained applications such as intra-operative 2D/3D registration. Methods using the full image content can be applied in almost any medical application area, using any type of transformation. Many applications are tailored for rigid or affine global registration of 3D images of the head. [MV98]

Intensity-based algorithms for retrospective registration require much less pre-processing or user-interaction than for point-based or surface-based methods: these methods are relatively easy to automate. Although hybrid approaches are being suggested [FHM00], nearly all presented methods are automatic.

Intensity-based registration involves calculating a transformation between two images using the pixel or voxel values alone and finding a relationship. The term intensity is used to refer to the scalar values in image pixels or voxels. Simply the registration transformation is determined by iteratively optimizing some “similarity measure” calculated from all pixel or voxel values that satisfy the relationship between the images. Current literature reports on the following similarity measures [FHM00]:

- Cross-correlation (of original images or extracted feature images) [JMH90, BDC93, BGL93, EV93, Hil93, HF93, MR93, RRS93, BT94, CNP94, CPE94, Els94, EPS94, LJF94, MM94, MEV94, MEV96b, aRC94, RRS94, And95, ASV95, Cid95, CEH95, EMP95, HES95b, MK95,
MEV95, PWL95, SHH95b, SHH95a, DB96, BB96, HF96, LGS96, MEV96a].

- Fourier domain based cross-correlation, and phase-only correlation [CM87, LB87, Che93, LGS96, SBZ96a, WRH96].

- Minimization of variance of intensity ratios [Hil93, HHH93a, WMC93, ABK94, SHH95b, SHH95a, ZJB96].

- Minimization of variance of grey values within segments [CJ94, ABH95].

- Minimization of the histogram entropy of difference images [BW96].

- Histogram clustering and minimization of histogram dispersion [Hil93, HSH94, HH94, CVS95, HRH95, SHH95b, SHH95a, LGS96].

- Maximization of mutual information (relative entropy) of the histogram [CMD95, VI95, Vio95, IVK95, MCV96, Pok96, SHH96, VSS96, IVA96].

- Maximization of zero crossings in difference images [VGL83, VLR84, VL84, HF93, HDH93, VPW94, PWL95, BKS96].

- Cepstral echo filtering [BXL94].

- Determination of the optic flow field [BTH95, MGJ96].

- Minimization of the absolute or squared intensity differences [HDH93, LOT93, ZYG93, MM94, YYL94, CRM95, CMM95, HCJ95, HSO95, HSS95a, JR95, KB95, SHS95, UvL95, CKM96, EKF96, HCJ96].

- Matching local low-order Taylor expansions determined by the image grey values [SBS93].
Implicitly using surface registration by interpreting a 3D image as an instance of a surface in 4D space [FMD96].

1.4.2.3 Non-Image Based Registration

The registration of multimodal images can be non-image based when the imaging coordinate systems of the two scanners involved are somehow calibrated to each other. This calibration constraint usually requires the scanners to be brought in to the same physical location, and the assumption that the patient remains motionless between acquisitions. These are prohibitive prerequisites, but they can be met in applications involving the use of Ultrasound [HSD94, rCF95, EKo96], since Ultrasound systems can come as hand-held devices that are equipped with a spatial optical localization system, they are easily calibrated, and can be used while the patient is immobilized on the CT or MR. The technique of calibrated coordinate systems is also often used in registering the position of surgical tools mounted on a robot arm to images.[MV98]

1.4.3 Nature And Domain Of Transformation

1.4.3.1 Nature Of Transformation

An image coordinate transformation is called rigid, when only translations and rotations are allowed. If the transformation maps parallel lines onto parallel lines it is called affine. If it maps lines onto lines it is called projective. Finally, if it maps lines onto curves it is called curved or elastic. Each type of transformation contains in special cases the ones described before it. A composition of more than one transformation can be categorized as a single transformation of the most complex type in the composition. [MV98]
1.4.3.2 Domain Of Transformation

A transformation is called global if it applies to the entire image, and local if subsections of the image each have their own transformations defined. Examples of all transformation types mentioned can be seen in Figure 1.2. [MV98]

In most applications, rigid and affine transformations are global, and curved transformations are local. Applications are nearly always intrinsic, mostly deformable model based or using the full image content, and mostly semi-automatic, requiring a user-identified initialization, since local information of the anatomy is essential to provide an accurate local curved transformation. [MV98]

The global rigid transformation is most common case in registration applications. It is popular because in many common medical images the rigid body constraint is, at least to a good approximation, satisfied, and it has relatively few parameters to be determined, and many registration techniques are not equipped to supply a more complex transformation. The most common application area is the human head. [MV98]

1.4.4 Interaction

There are three levels of interaction for image registration: automatic, interactive, and semi-automatic. When the registration is automatic, the user only supplies the algorithm with the image data and possibly information on the image acquisition. When it is interactive, the user does the registration himself, assisted by a software supplying a visual or numerical impression of the current transformation, and possibly an initial transformation guess. When it is semi-automatic, the interaction required can be of two different natures: the user needs to initialize the algorithm or steer the algorithm. [MV98]
1.4.5 **Optimization Procedure**

The parameters that make up the registration transformation can either be computed directly or searched for. In the former case, the manner of computation is completely determined by the paradigm. Only general remark is that the use of computation methods is restricted almost completely to applications relying on very sparse information. In the case of searching optimization methods, most registration methods are able to formulate the paradigm in a standard mathematical
function of the transformation parameters to be optimized. This function attempts to quantify the similarity as dictated by the paradigm between two images given a certain transformation. [MV98]

Popular techniques are:

- Powell’s method [LPC88, HHH91, TRG93, EGr94, EGL94, HSD94, HK94, KHB94, LIF94, And95, ASV95, CMD95, LLD95, BKS96, GVH96, BB96, MCV96],
- the Downhill Simplex method [HHH91, GH93, HHH93a, HDH93, LHM93, HK94, KHB94, LPC94, MLB95, SHS95, EKF96],
- Brent’s method and series of one-dimensional searches [BDC93, MR93, AS94, PKF94, AS95, ABH95, MK95, HF96],
- Levenberg-Marquardt optimization [Tau93, HSP94a, HSP94b, SL94b, SL94c, BCC95, HSC95, HSe95a, LS95, UvL95, eSB96, SL96],
- Newton-Raphson iteration [FL93, WMC93, ZJB96],
- stochastic search methods [MCA93, VI95, Vio95, IVK95, VSS96, IVA96],
- gradient descent methods [ZAB94, PWL95, BW96, CKM96, CTM96],
- genetic methods [HHH93a, HSH94, HH94, SX94, KB95, CWH96],
- simulated annealing [LPE94],
- geometric hashing [zA92, AzT93, PG95],
- and quasi-exhaustive search methods [BGL93, EV93, HF93, CJ94, Els94, EPS94, MCN94, MEV94, MEV96b, EMP95, MEV95, DB96, MEV96a].
Many of these methods are documented in [PTV92]. Frequent additions are multi-resolution and multi-scale approaches to speed up convergence, to reduce the number of transformations to be examined, and to avoid local minima. Some registration methods employ non-standard optimization methods that are designed specifically for the similarity function at hand, such as the ICP algorithm [BM92, SHK94, FA94, MAD95, PG95, SHK95, BF95, BFA95, CKD95, FAB95, ETM96, FMD96, FA96, GDF96] created for rigid model based registration. Many applications use more than one optimization technique; frequently a fast but coarse technique followed by an accurate yet slow one. [MV98]

1.4.6 Modalities Involved In The Registration

The registration tasks can be categorized into four groups: monomodal applications, multimodal applications, modality to model, and patient to modality, using “modality” in a loose sense. [MV98]

In monomodal applications, the images to be registered belong to the same modality. In multimodal registration tasks, the images to be registered stem from two different modalities. In modality to model and patient to modality registration only one image is involved and the other “modality” is either a model or the patient himself. [MV98]

Monomodal tasks are well suited for growth monitoring, intervention verification, rest-stress comparisons, ictal-interictal comparisons, subtraction imaging (also DSA, CTA), and many other applications. The applications of multimodal registration are abundant and diverse, predominantly diagnostic in nature. A coarse division would be into anatomical-anatomical registration where images showing different aspects of tissue morphology are combined, and functional-anatomical where tissue metabolism and its spatial location relative to anatomical structures are related. The
patient to modality registration tasks appear almost exclusively in intra-operative [BSH94, HVE94, HSB94, LJF94, eST94a, LWM94, LPC94, SHK94, WTZ94, BFA95, BF95, BCC95, CKD95, EHH95a, EHH95b, HSe95a, HeS95a, LSP95, LWM95, MFG95, MGG95, REL95, SOB95, SHK95, ESC96, FWN96, eTS96, Lav96, PDM96] and radiotherapy applications [Bij93, GV93, LHM93, TLV95, VTL95, GVH96]. Modality to model can be applied to gather statistics on tissue morphology such as for finding anomalies relative to normalized structures, and to segmentation tasks [BLR83, RGP95, AK96, CTM96, JZL96]. [MV98]

1.4.7 Subject

When all of the images involved in a registration task are acquired of a single patient, we refer to it as intrasubject registration. If the registration is accomplished using two images of different patients or a patient and a model, then it is referred to as intersubject registration. If one image is acquired from a single patient, and the other image is somehow constructed from an image information database obtained using imaging of many subjects, it is known as atlas registration. [MV98]

Intrasubject registration is by far the most common of the three; used in almost any type of diagnostic and interventional procedure. Intersubject [BLR83, GRB93, MCA93, SL94c, SL94b, SL94a, CEH95, GFK95, HCJ95, SL95, Thi95, AK96, DFB96, FRR96, GH96, HCJ96, Thi96b] and atlas [CNP94, CPE94, DP94, MAE94, BTH95, CRM95, CMM95, SHS95, CKM96, CTM96, DPB96, FMD96] registration appear mostly in 3D-3D MR or CT brain image applications. Here, the nature of the registration transformation is mostly curved; these applications are always intrinsic, either segmentation based or voxel property based using the full image content. A manual
initialization is frequently desired. Some applications use rigid transforms, but their application is limited. Others use anatomical landmarks for a deformation basis of a curved transformation; unfortunately such applications often require the transformation in large image areas to be interpolated from the nearest landmark transformations, which may prove unreliable. The use of intersubject and atlas matching can notably be found in the areas of gathering statistics on the size and shape of specific structures, finding anomalous structures, and transferring segmentations from one image to another. [MV98]

1.4.8 Object

The list of objects is composed of those imaging areas encountered in recent literature. These include head, thorax, abdomen, pelvis and perineum, limbs and spine and vertebrae. The majority of papers concerns global head registration, which is possibly due to the fact that the head can be considered a rigid body in many applications, while such a constraint cannot be met in many thoracic, abdominal, pelvic, and spinal images. [MV98]

1.5 REMARK ON SUB-PIXEL REGISTRATION

An important example of applications that require registration at sub-pixel accuracy is multi-frame super-resolution [ABF91, AKS92, BW93, EF97, FRE04, IP91, KPB88, KBV90, LPK01, NY01, NMG01, PST97, PKS87, SS95, FC99, FBZ96, TH84], which aims to combine several degraded low-resolution images into a single high-resolution image in order to approach the Nyquist rate. The key to success in these multi-frame super-resolution techniques is the accurate registration with sub-pixel precision. Essentially, sub-pixel registration is the step that allows merging the samples of the low-resolution data in a denser grid. In the absence of sub-pixel registration and
fusion, the super-resolution problem would essentially reduce to that of classical deconvolution.

Among existing sub-pixel registration techniques Fourier domain methods [FZB02, Hog03, KS93, SOC01, FBZ96] and also closely related spatial domain variations [WZ00] are an important class of registration techniques that have gained popularity due to their remarkable accuracy. Most of these methods are in fact variations of the original phase correlation method [KH75]. For instance, [FZB02] Foroosh et al. demonstrated how the method can be extended to sub-pixel accuracy, [SOC01] stone et al. investigated the effect of aliasing error, and in [Hog03] Hoge describes how the shift parameters can be decoupled in the dominant left and right eigenvectors of the phase correlation matrix (not the phase difference matrix) using its rank-one constraint. The most commonly used approach for sub-pixel registration is based on interpolation. Interpolation is often performed explicitly via correlation interpolation [Dvo83, TH86], intensity interpolation [TH86], phase correlation interpolation [PHG77, TH86] or geometric methods [BKL87, GSP86]. Alternatively, interpolation can be applied implicitly via smoothing and regularization, like in optical flow techniques [HS81, HT81, IP91, Nag87, TH86, VGT90] or optimization-based methods [KS93, TRU98, TH84]. Other sub-pixel registration methods that use some form of interpolation include: the methods based on local normalized correlation [AWD97], polynomial regression [ZMW94], the discrete cosine transform [KL98], and model-based approaches using control points [EG94].

In the case of sub-pixel accuracy, we are interested in investigating the information contained in the phase domain, and particularly motivated by two factors:

- Applications that require registration at sub-pixel accuracy, e.g. examination of same-patient Magnetic Resonance Imaging (MRI) data in a clinical setting [EPV93a, HSS95b], or super-
resolution from multiple views [FBZ96, FC99, IP91, PKS87].

- Application to sensor modalities that sample directly the Fourier spectrum of the field of view, such as MRI [HBT99, Wri97], or Synthetic Aperture Radar (SAR) [Sko90].

In case of processing facial medical data sets, we made use of symmetry, automatic thresholding techniques, and inpainting algorithms and external information like stone case CT scans.
CHAPTER 2

SUB-PIXEL ESTIMATION OF SHIFTS

2.1 SUB-PIXEL ESTIMATION OF SHIFTS DIRECTLY IN THE FOURIER DOMAIN

2.1.1 Introduction

In this section, we establish the exact relationship between the continuous and the discrete phase-difference of two shifted images, and show that their discrete phase difference is a 2-dimensional sawtooth signal. Sub-pixel registration can thus be performed directly in the Fourier domain by counting the number of cycles of the phase difference matrix along each frequency axis. The sub-pixel portion is given by the non-integer fraction of the last cycle along each axis. The problem is formulated as an overdetermined homogeneous quadratic cost function under rank constraint for the phase difference, and the shape constraint for the filter that computes the group delay. The optimal trade-off for imposing the constraints is determined using the method of Generalized Cross Validation (GCV). Also, in order to robustify the solution, we assume a mixture model of inlying and outlying estimated shifts, and truncate our quadratic cost function using Expectation Maximization (EM).

The approach that we propose herein does not explicitly require interpolation, but assumes that bandlimited sampling and interpolation hold for the data. This makes the present work a natural extension to the approach in [FZB02]. Closely related recent works by other authors that provide very good solutions are reported in [Hog03] and [SOC01]. In [Hog03] the rank constraint of the phase correlation matrix has been exploited to derive a direct Fourier-domain algorithm, and in [SOC01] the authors focus on a specific task of reducing the influence of aliasing errors.
in estimating relative shifts. We derive analytic results describing the exact parametric model of the discrete phase difference, and its relation to the underlying continuous signals. An accurate solution is thus obtained by fitting the analytic model to the noisy data.

In the next section, we briefly describe the phase correlation method. We then derive in section 2.1.3 the relation between the discrete and the continuous phase differences of two shifted images. In sections 2.1.4 and 2.1.5, we present our approach for estimating the number of cycles of the phase difference matrix, and design a robust technique to estimate an accurate solution for the registration parameters. Section 2.1.6 provides the experimental results on both simulated and real data, with thorough comparisons with the methods in [FZB02],[Hog03], and [SOC01]. Finally, in section 2.1.7, we discuss the proposed technique, and provide some concluding remarks.

2.1.2 The Phase Correlation Method

Let \( f_1(x, y) \) and \( f_2(x, y) = f_1(x - x_o, y - y_o) \) be two square-integrable continuous signals with their relative shifts given by \((x_o, y_o)\). Their cross power spectrum is then given by

\[
\hat{c}(u, v) = \frac{\hat{f}_1 \hat{f}_2^*}{|\hat{f}_1 \hat{f}_2^*|} = \exp(ix_ou + iy_ov) \tag{2.1}
\]

where the hat sign as usual indicates the Fourier transform, and the asterisk stands for the complex conjugate. Hereafter, we will distinguish between the phase correlation and the phase difference, where the latter is the phase angle of the former. In other words, we refer to \( \exp(ix_ou + iy_ov) \) as the phase correlation function, and to \( x_ou + y_ov \) as the phase difference. Similarly, in the discrete case, we will refer to the discrete cross power spectrum as the phase correlation matrix, and its phase angle as the phase difference matrix.
As can be seen in (2.1), due to the Fourier shift property, the spatial translations lead to linear phase differences between the two signals along the two frequency axes, i.e. the phase difference between the two shifted continuous signals is given by the phase angle

\[ \hat{\phi}(u, v) = \angle \hat{c}(u, v) = x_o u + y_o v \]  

(2.2)

which is a planar surface through the origin.

Using discrete Fourier transform, it can be shown that equation (2.1) is also valid in the discrete case, which provides the basis for the well-known phase correlation method [KH75]. The classical approach to determine the shift parameters from (2.1) is by inverse transforming the discrete phase correlation matrix, which would lead to a unit impulse centered at \((x_o, y_o)\). This, of course, is true only for integer displacements. For non-integer (i.e. shifts with sub-pixel components) additional results are derived in [FZB02, SBZ96b], which show that inverse transforming the phase correlation matrix would yield a discrete Dirichlet function instead of a single impulse. The solution is then found by a least-squares fitting of the parameters to a set of coherent peaks of the resulting Dirichlet kernel.

As mentioned earlier, however, some sensor modalities such as MRI and SAR, provide directly the Fourier domain spectrum over the field of view. Therefore it would be interesting, from both computational and implementation viewpoints, to estimate the translations directly in the Fourier domain. A practical solution for this problem was first proposed by Hoge [Hog03], which requires the following steps in order to compute the shift parameters:

- A subspace approximation of the noisy phase correlation matrix (not the phase difference
matrix) to impose the rank constraint,

- Unwrapping of the dominant left and right eigenvectors of the approximated phase correlation matrix.

- Least-squares fitting of the shift parameters to the unwrapped eigenvectors.

It is worth noting that 2-dimensional phase unwrapping is a notoriously ill-posed problem. This is perhaps why Hoge proposed to perform the unwrapping step on the 1-dimensional dominant eigenvectors of the phase correlation matrix, rather than directly on the phase difference matrix itself. We will show below that due to the rank constraint of the unwrapped phase difference matrix, the unwrapping process becomes separable along the two frequency axes. In other words, it reduces to two 1-dimensional unwrapping steps. As a result a very good solution can also be found without subspace approximation [FB04]. We will also show that even phase unwrapping is an unnecessary step, since we will determine the exact parametric shape of the phase matrix.

### 2.1.3 Discrete Phase Difference

The relation between the continuous phase difference and the discrete phase difference matrix can be readily established from equation (2.2). Let $\varphi_{kl} \leftrightarrow \hat{\varphi}_{mn}$ denote the Fourier transform pair for the discrete phase difference. From bandlimited sampling theory $\varphi_{kl}$ is given by

$$\varphi_{kl} = -\frac{x_o}{\pi k} \left(2\text{sinc} \frac{\pi k}{x_o} - 2 \cos \frac{\pi k}{x_o}\right) - \frac{y_o}{\pi l} \left(2\text{sinc} \frac{\pi l}{y_o} - 2 \cos \frac{\pi l}{y_o}\right)$$

(2.3)
This latter equation can also be written as

\[
\varphi_{kl} = x_o \left( 2 \text{sinc} \frac{\pi k}{x_o} - 2 \cos \frac{\pi k}{x_o} \right) + y_o \left( 2 \text{sinc} \frac{\pi l}{y_o} - 2 \cos \frac{\pi l}{y_o} \right)
\]

which shows that $\frac{d\hat{\varphi}_{mn}}{dm}$ and $\frac{d\hat{\varphi}_{mn}}{dn}$ are given by the superposition of two terms: a pulse corresponding to the sinc function, and a pair of impulses at the two ends of the pulse corresponding to the cosine function. Therefore upon integration, we should obtain samples of a 2D sawtooth signal. We can demonstrate this formally as follows.

First, note that

\[
\varphi_{kl} = i x_o^2 \frac{\pi^2}{2k^2} \left( 2 \text{sinc} \frac{\pi k}{x_o} \cos \frac{\pi k}{x_o} - 2 \sin \frac{\pi k}{x_o} \right) + i y_o^2 \frac{\pi^2}{2l^2} \left( 2 \text{sinc} \frac{\pi l}{y_o} \cos \frac{\pi l}{y_o} - 2 \sin \frac{\pi l}{y_o} \right)
\]

On the other hand, it can be verified that

\[
x_o^2 \frac{\pi}{2k} \int_{-\frac{x_o}{2}}^{\frac{x_o}{2}} u \cos ku \, du + y_o^2 \frac{\pi}{2l} \int_{-\frac{y_o}{2}}^{\frac{y_o}{2}} v \cos lv \, dv = 0
\]
and similarly

\[
x_o \frac{2}{2\pi/x_o} \int_{-\pi/x_o}^{\pi/x_o} u\,du + y_o \frac{2}{2\pi/y_o} \int_{-\pi/y_o}^{\pi/y_o} v\,dv = 0
\]  

(2.9)

From (2.7), (2.8), and (2.9), and using the definition of the Discrete Fourier Transform (DFT) based on Fourier series [OSB89], it follows immediately upon substituting \( u = n \frac{2\pi}{N} \) and \( v = m \frac{2\pi}{M} \) that \( \varphi_{kl} \) is a DFT coefficient of the following discrete periodic signal

\[
\hat{\varphi}_{mn} = \begin{cases} 
2\pi \left( x_o \frac{n}{N} + y_o \frac{m}{M} \right) & \text{if } m' \frac{2\pi}{M} = m \frac{2\pi}{M} + j \frac{2\pi}{x_o} \\
\hat{\varphi}_{m'n} & \text{if } n' \frac{2\pi}{N} = n \frac{2\pi}{N} + k \frac{2\pi}{y_o} 
\end{cases}
\]  

(2.10)

where \( j \) and \( k \) are arbitrary integers.

This is a 2D sawtooth signal as opposed to the continuous phase difference in (2.2), which is a plane through the origin. In other words, the discrete phase difference matrix for a pair of shifted images is given by

\[
\Phi_{mn} = [\hat{\varphi}_{mn}]
\]  

(2.11)

where \( m = 0, \ldots, M - 1 \) and \( n = 0, \ldots, N - 1 \).

Figure 2.1 shows examples of noisy discrete phase difference matrices. The underlying 2-dimensional sawtooth signals are clearly visible. Figures 2.1(e) and 2.1(f) show one row of the figures in 2.1(c) and 2.1(d). The first observation that can be made from this result is that the unwrapping of a 2-dimensional sawtooth signal is separable, since its unwrapped matrix has to be
rank-2. This implies that a subspace approximation can be avoided. This observation led to the result that was reported in [FB04]. But even a more interesting conclusion that can be drawn from the above result is that unwrapping is also an unnecessary step for registration.

For this purpose, note how the period of the sawtooth signal along each axis determines the shifts along corresponding axis: the period along the $u$-axis is $\frac{2\pi}{x_o}$, and hence there are $x_o$ repeated cycles along each row of the phase difference, where $x_o$ may or may not be an integer. When $x_o$ is not an integer, the number of repeated cycles in a row is given by the integer part of $x_o$ plus a fraction of a cycle defined by the non-integer portion of $x_o$. A similar argument applies to the columns of $\Phi$. This process of counting the number of cycles along the rows and columns of the phase matrix is essentially all that is required to determine the shifts. The challenge of course is to determine the exact fractional portion of the repeated cycles. In the next two sections, we will design a robust estimator for this problem.

### 2.1.4 Solving For Image Shifts

As indicated above, the key to solve the problem is to find how many cycles of the sawtooth phase fit in the range $[0, 2\pi]$ along each frequency axis. The number of cycles i.e. $x_o$ and $y_o$ may or may not be integer values, and are given by

\[
x_o = \frac{\text{cycles}}{2\pi} = \frac{N}{2\pi} d\hat{\phi}_{mn} \tag{2.12}
\]

\[
y_o = \frac{\text{cycles}}{2\pi} = \frac{M}{2\pi} d\hat{\phi}_{mn} \tag{2.13}
\]

However, due to noise and other sources of error (see for instance [FZB02, Hog03, SOC01], for a thorough description of the sources of error), counting the number of cycles per $2\pi$ using
Figure 2.1: (a) & (b) Two aerial images with some shifts, (c) phase difference matrix corresponding to shifts of (7.3,5.6) pixels, (d) phase difference matrix corresponding to shifts of (30.5,25.4) pixels, (e) & (f) one row of the phase matrices shown in (c) & (d), respectively.
equations (2.12) and (2.13) may lead to inaccurate results. To overcome this problem, we need to use the fact that a total of $M \times N$ data points are available for regression. Therefore, a robust solution can be obtained by minimizing an appropriate error function, i.e. by a robust fitting of the parameters to the data.

For this purpose note that the 2D sawtooth signal $\Phi_{mn}$ has constant slopes for the vast majority of frequency coordinates along each row $r_m$ or each column $c_n$, except for a small number of frequencies, where discontinuities occur. Therefore, statistically the discontinuities along a row or a column may be treated as outliers. Below, we describe how a robust estimator can be designed to eliminate the influence of outliers. We will show the derivations for the columns of $\Phi_{mn}$. However, the approach is equally applicable to the rows.

Essentially, for any column $n$ of $\Phi_{mn}$, we need to design an optimal filter $h$ that can compute the slope of the noisy sawtooth signal along that column. We will model the filter as a finite impulse response (FIR) filter. Since a gradient filter is expected to be anti-symmetric, we will assume that $h$ is a type III FIR filter, of length $2L + 1$, i.e. $h = [h_1, \ldots, h_{2L+1}]^T$. The gradient of the $n^{\text{th}}$ column is then given by

$$Hc_n = c'_n$$

(2.14)

where $c'_n = \left[ \frac{d\hat{\Phi}_{mn}}{dm} \right]^T$, $m = L + 1, \ldots, M - L$ is the gradient vector truncated at both ends to avoid
border artifacts, and $H$ is a $(M - 2L) \times M$ matrix given by

$$H = \begin{bmatrix}
  h^T & 0 & \ldots & 0 \\
  0 & \ddots & \ddots & \vdots \\
  \vdots & \ddots & \ddots & 0 \\
  0 & \ldots & 0 & h^T
\end{bmatrix} \tag{2.15}$$

Upon rearranging equation (2.14), we can get

$$F_n h = c'_n \tag{2.16}$$

where $F_n$ is a $(M - 2L) \times (2L + 1)$ matrix given by

$$F_n = \begin{bmatrix}
  \hat{\varphi}_{1n} & \cdots & \hat{\varphi}_{(2L+1)n} \\
  \hat{\varphi}_{2n} & \cdots & \hat{\varphi}_{(2L+2)n} \\
  \vdots & \cdots & \vdots \\
  \hat{\varphi}_{(M-2L)n} & \cdots & \hat{\varphi}_{Mn}
\end{bmatrix} \tag{2.17}$$

For a general phase matrix, the problem would be a blind one, since we would need to determine both $h$ and $c'_n$. However, in our case, since $\Phi_{mn}$ is a sawtooth signal, except at a small number of discontinuities, we have

$$c'_n = \frac{2\pi}{M} y_n l \tag{2.18}$$

where $l = [1, \ldots, 1]^T$ is a vector of length $M - 2L$. 

36
As a result the equation in (2.16) can be written also in a homogeneous form as

$$\tilde{F}_n \tilde{h} = 0$$  \hspace{1cm} (2.19)

where

$$\tilde{F}_n = \begin{bmatrix}
\hat{\varphi}_{1n} & \ldots & \hat{\varphi}_{(2L+1)n} & -\frac{2\pi}{M} \\
\hat{\varphi}_{2n} & \ldots & \hat{\varphi}_{(2L+2)n} & -\frac{2\pi}{M} \\
\vdots & \ldots & \vdots & \vdots \\
\hat{\varphi}_{(M-2L)n} & \ldots & \hat{\varphi}_{Mn} & -\frac{2\pi}{M}
\end{bmatrix}$$  \hspace{1cm} (2.20)

and $\tilde{h} = [h_1 \ldots h_{2L+1} y_o]^T$.

This allows for a least-squares solution of (2.19), since $\tilde{h}$ lies on the right null-space of the matrix $\tilde{F}_n$. The least squares solution can thus be found by minimizing $\|\tilde{F}_n \tilde{h}\|$ subject to $\|\tilde{h}\| = 1$, which is also equivalent to

$$\tilde{h}_{\text{opt}} = \arg \min \|\tilde{F}_n \tilde{h}\| \|\tilde{h}\|$$  \hspace{1cm} (2.21)

where the norm constraint on $\tilde{h}$ is to avoid the trivial solution $\tilde{h}_{\text{opt}} = 0$.

The solution to (2.21) is the unit eigenvector corresponding to the smallest eigenvalue of $\tilde{F}_n$.

This least-squares solution, however, can be improved in two ways:

- By robustifying the solution to outlying data caused by the discontinuities of the discrete phase difference at $\hat{\varphi}_{mn} = \pm \pi$.

- By incorporating the a priori constraints on $\tilde{F}_n$ and $\tilde{h}$.
In particular, there are two important constraints that apply to $\tilde{F}_n$ and $\tilde{h}$. By inspection of (2.17), we can verify that $F_n$ should become a rank-2 matrix if we unwrap it. On the other hand, we know that for a noise-free phase matrix, the filter $h$ should be anti-symmetric, i.e. $h_{L+1+i} = h_{L+1-i}$, for $i = 1, ..., L$ and $h_{L+1} = 0$. Taking into account these considerations, we can formulate the problem as follows:

$$\tilde{h}_{\text{opt}} = \arg \min \|\tilde{F}_n\tilde{h}\| + \lambda\|Ah\| \tag{2.22}$$

where $\lambda$ is the regularization parameter, and $A = [a_{ij}]$ is a $L + 1 \times (2L + 1)$ matrix defined by

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \text{ or } i + j = 2L + 2 \\ 0 & \text{otherwise} \end{cases} \tag{2.23}$$

The formulation in (2.22) is a semi-norm Tikhonov-Arsenin [TA77] regularization of the least-squares problem defined in (2.21). The solution is given by $\tilde{h}_{\text{opt}} \sim [h^T \text{opt} \ 1]^T$, where

$$h_{\text{opt}} = (F_n^T F_n + \lambda A^T A)^{-1} F_n^T 1 \tag{2.24}$$

Note that, since our formulation in (2.22) is homogeneous, the solution is found only up to a scale factor as indicated by the $\sim$ notation. This scale ambiguity, however, can be readily resolved by assuming that $h$ is the discrete gradient of a smoothing kernel that preserves the first moment. This implies that the components of $h_{\text{opt}} = [h_i]_{i=1,\ldots,2L+1}$ should satisfy

$$\sum_{j=1}^{2L+1} \sum_{i=1}^{j} h_i = 1 \tag{2.25}$$
We now have the solution given by (2.24) and (2.25) up to an unknown regularization parameter. The optimal value of this parameter is given by the method of Generalized Cross Validation (GCV) [GH79, GM96], which amounts to minimizing

\[ \text{GCV}(\lambda) = \frac{\| (I - F_n(F_n^T F + \lambda A^T A)^{-1} F_n^T) I\|_2^2}{(\text{tr}(I - F_n(F_n^T F + \lambda A^T A)^{-1} F_n^T))^2} \]  

(2.26)

with respect to \( \lambda \).

The minimizer of (2.26) is usually obtained using numerical techniques by making simplifying assumptions, e.g. circulant \( F_n \) and \( A \), or by using numerical techniques such as quadrature rules and Lanczos algorithm [GH79]. However, in our case, by imposing a rank constraint on \( F_n \), we can find a simplified closed-form solution. For this purpose, let \( (A(F_n^T F_n)^{-1} A^T)^{-1} = U \Lambda \) be the eigen-decomposition of the positive symmetric matrix \( (A(F_n^T F_n)^{-1} A^T)^{-1} \). Define \( K_n = \Lambda^{-\frac{1}{2}} U^T \), and \( l_n = K_n A(F_n^T F_n)^{-1} F_n^T l + l_s \), where the vector \( l_s \) is on the right null space of \( K_n^T K_n \). Using \( K_n, l_n \) and the identity

\[ A(F_n^T F_n + \lambda A^T A)^{-1} A^T = \left( (A(F_n^T F_n)^{-1} A^T)^{-1} + \lambda I \right)^{-1} \]  

(2.27)

the GCV function can be written as [GM96]

\[ \text{GCV}(\lambda) = \frac{\| (I - K_n(K_n^T K_n + \lambda I)^{-1} K_n^T) l_n\|_2^2}{(\text{tr}(I - K_n(K_n^T K_n + \lambda I)^{-1} K_n^T))^2} \]  

(2.28)

\(^1\)In practice, \( l_s \) is taken as the eigenvector corresponding to the smallest eigenvalue of \( K_n^T K_n \).
If we now define \( P_n = I - K_n(K_n^T K_n + \lambda I)^{-1} K_n^T \), then

\[
GCV(\lambda) = \frac{\|P_n I_n\|^2}{(\text{tr}(P_n))^2}
\]  

(2.29)

Applying the matrix inversion lemma, we find

\[
P_n = \left( I + \frac{1}{\lambda} K_n K_n^T \right)^{-1}
\]  

(2.30)

Next, let

\[
K_n K_n^T = V \Sigma V^T = \sum_{j=1}^{N-2L} \sigma_j v_j v_j^T
\]  

(2.31)

be the spectral decomposition of \( K_n K_n^T \), where \( v_j \)'s are the columns of \( V \) that form a set of orthonormal basis, and \( \sigma_j \)'s are the corresponding eigenvalues. \( P_n \) can then be written as

\[
P_n = \sum_{j=1}^{N-2L} \frac{\lambda}{\lambda + \sigma_j} v_j v_j^T
\]  

\[
= \sum_{j=1}^{2} \frac{\lambda}{\lambda + \sigma_j} v_j v_j^T + \sum_{j=3}^{N-2L} v_j v_j^T
\]  

(2.32)

In order to simplify the GCV function in (2.29), we make a first order approximation of the matrix \( K_n K_n^T \) using its largest eigenvalue: \( K_n K_n^T \simeq \sigma_1 v_1 v_1^T \). In practice, we found that the dominant eigenvalue is usually orders of magnitude larger than the second eigenvalue. Therefore

\[
GCV(\lambda) \simeq \frac{\left( \frac{\lambda}{\lambda + \sigma_1} \right)^2 s_1^2 + \sum_{j=2}^{N-2L} s_j^2}{(N - 2L - 1 + \frac{\lambda}{\lambda + \sigma_1})^2}
\]  

(2.34)
where $s_j$ are the components of the vector $V^T l_n = [s_1 \ s_2 \ \ldots \ s_{N-2L}]^T$.

Differentiating this equation with respect to $\lambda$ and setting it equal to zero, we find after simplification that the optimal regularization parameter is given by

$$
\lambda^* = \frac{\sigma_1 \sum_{j=2}^{N-2L} s_j^2}{(N - 2L - 1)s_1^2 - \sum_{j=2}^{N-2L} s_j^2} 
$$

(2.35)

Using (2.24), (2.25), and (2.35), we can compute $h_{\text{opt}}$, which upon applying to the $n^{th}$ column will yield $(M - 2L) \times N$ values for $y_o$. These are the optimal values of $y_o$ obtained by minimizing a constrained cost function at different frequencies. The estimated values for $y_o$ may however have some error, and more importantly may be contaminated with outliers due to phase discontinuities.

In the next section, we use EM to cluster the inlying and the outlying values. This, in practice, is equivalent to truncating the quadratic error in (2.22), which is a well-known robust estimator [BA94].

### 2.1.5 Robustifying The Solution

We have a total number of $T = (M - 2L) \times N$ estimated values for $y_o$ (or equivalently $(N - 2L) \times M$ values for $x_o$). Our goal now is to use this highly redundant amount of information to estimate a robust solution for $y_o$ (or equivalently for $x_o$). Due to the inlying and the outlying values, the probability density function of $y_o$ is expected to be a mixture of two different densities. The problem is thus to distinguish between the samples drawn from the inlying and the outlying densities. Assuming that the two components of our mixture model are normally distributed $\mathcal{N} (\mu_i, \sigma_i^2)$, $i = 1, 2,$
our model can be written as

\[ p(y_o|\Theta) = \sum_{i=1}^{2} \alpha_i p(y_o|\theta_i) \]  \hspace{1cm} (2.36)

where \( p(\cdot) \) denotes the probability, and \( \alpha_i \) is the prior probability that \( y_o \) is drawn from the distribution \( i \), such that \( \sum_{i=1}^{2} \alpha_i = 1 \), \( \theta_i = [\alpha_i, \mu_i, \sigma_i^2] \), and

\[ p(y_o|\theta_i) = \frac{1}{\sqrt{2\pi \sigma_i}} \exp \left( -\frac{(y_o - \mu_i)^2}{2\sigma_i^2} \right) \]  \hspace{1cm} (2.37)

This clearly is a parametric model, and the parameter vector to be estimated is given by \( \Theta = [\alpha, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2]^T \), where \( \alpha = \alpha_1 = 1 - \alpha_2 \).

Our ultimate goal is to estimate the mean value of the inlying distribution, which is an unbiased estimator of the slope of the phase. Since optimizing our likelihood function is analytically intractable [McL97], a maximum-likelihood (ML) estimation of the parameters is built iteratively using the expectation maximization (EM) algorithm. Following [McL97], the expectation and the maximization steps for our problem reduce to the iterative updates of the unknown parameters as described below. The derivations of the following EM equations specifically for our mixture model in (2.36) are omitted, since they can be readily inferred from the general theory of EM algorithm. The \( t + 1 \) iteration is given by

- **E-step:** Estimate \( p(i|y_o^t, \Theta^t) \), given the current estimate of the parameter set \( \Theta^t \).
From Bayes’ law this is given by

\[ pr(i|y_o^j, \Theta^t) = \frac{\alpha_i^t pr(y_o^j|\theta_i^t)}{pr(y_o^j|\Theta^t)} = \frac{\alpha_i^t pr(y_o^j|\theta_i^t)}{\sum_{i=1}^{2} pr(y_o^j|\theta_i^t)} \quad (2.38) \]

\[ = \frac{\alpha_i^t pr(y_o^j|\theta_i^t)}{\sum_{i=1}^{2} pr(y_o^j|\theta_i^t)} \quad (2.39) \]

- **M-step**: Update the parameters of the model to maximize the likelihood of the data (see [McL97] for inferring the derivations of the following equations for our mixture model in (2.36))

\[ \alpha_i^{t+1} = \frac{1}{N} \sum_{j=1}^{T} pr(i|y_o^j, \Theta^t) \quad (2.40) \]

\[ \mu_i^{t+1} = \frac{\sum_{j=1}^{T} y_o^j pr(i|y_o^j, \Theta^t)}{\sum_{j=1}^{T} pr(i|y_o^j, \Theta^t)} \quad (2.41) \]

\[ \sigma_i^{t+1} = \left( \frac{\sum_{j=1}^{T} pr(i|y_o^j, \Theta^t)(y_o^j - \mu_i^{t+1})^2}{\sum_{j=1}^{T} pr(i|y_o^j, \Theta^t)} \right)^{1/2} \quad (2.42) \]

In order to apply the above two-step iterations, we need to initialize the parameter set. Since median has a breakdown point of 50%, a simple initialization close to the solution based on data and our prior knowledge can be done as follows:

- \( \mu_1^0 = \text{med}(y_o^1, \ldots, y_o^T) \), i.e. median of all \( y_o^j \).

- \( \sigma_1^0 = \text{med}|y_o^j - \text{med}(y_o^1, \ldots, y_o^T)| \), i.e. the median deviation of all \( y_o^j \).

- \( \mu_2^0 = \text{mean}(|y_o^j - \mu_1^0| > 3\sigma_1^0) \), i.e. mean of all \( y_o^j \) outside the support of the inlying density.
• $\sigma_2^0 = \text{std}(|y_{ij}^0 - \mu_1^0| > 3\sigma_1^0)$, i.e. standard deviation of all $y_{ij}^0$ outside the support of the inlying density.

• And finally, $a_0^0 = \frac{\text{no. of } y_{ij}^0 \text{ such that } |y_{ij}^0 - \mu_1^0| \leq 3\sigma_1^0}{T}$.

Applying the above steps to $y_{ij}^0$ (and $x_{ij}^0$), separates the inlying values from the outlying ones, from which the true shift parameters can be estimated as the mean values of the inlying distributions. This is essentially equivalent to truncating our quadratic cost function in (2.22) to mitigate the influence of outliers.

### 2.1.6 Experimental Results

We applied the technique to both synthetic and real data. For synthetic data, we used the approach described in [FZB02], to generate images with sub-pixel shifts, i.e. starting from a real high resolution image, we lowpass filtered and downsampled shifted versions of the image. Using appropriate downsampling rates, shifts with sub-pixel contents were produced. Two sets of synthetic simulations were performed. In the first set we applied the method to the same images used in [FZB02], which are shown in Figure 2.2. Results are shown in table 1 and are compared to those reported in [FZB02]. The accuracy in this set was predominantly higher than [FZB02].

In a second set of synthetic data, we generated some shifted images, but this time with additive zero-mean Gaussian noise. The input image used to generate the synthetically shifted and downsampled noisy data is an 8-bit gray level image shown in Figure 2.3-(a). Figure 2.3-(b) shows a typical phase difference matrix obtained from a pair of synthesized noisy and downsampled images. For this experimentation, we varied the standard deviation of the noise in the interval $[0, 5]$, and for each noise level, we computed the signal to noise ratio (SNR) of the estimated shift pa-
Table 2.1: Table of results for shifts of the images in Figure 2.2

<table>
<thead>
<tr>
<th>Image</th>
<th>True Shifts</th>
<th>Foroosh et al. 2001</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>(0.50, -0.50)</td>
<td>(0.486, -0.513)</td>
<td>(0.495, -0.496)</td>
</tr>
<tr>
<td></td>
<td>(0.25, 0.50)</td>
<td>(0.280, 0.491)</td>
<td>(0.256, 0.499)</td>
</tr>
<tr>
<td></td>
<td>(-0.25, -0.50)</td>
<td>(-0.255, -0.527)</td>
<td>(-0.250, -0.510)</td>
</tr>
<tr>
<td></td>
<td>(0.0, 0.75)</td>
<td>(0.013, 0.803)</td>
<td>(0.013, 0.745)</td>
</tr>
<tr>
<td>Pentagon</td>
<td>(0.167, -0.5)</td>
<td>(0.152, -0.492)</td>
<td>(0.161, -0.502)</td>
</tr>
<tr>
<td></td>
<td>(0.67, 0.25)</td>
<td>(0.692, 0.331)</td>
<td>(0.680, 0.244)</td>
</tr>
<tr>
<td></td>
<td>(-0.33, -0.167)</td>
<td>(-0.322, -0.154)</td>
<td>(-0.340, -0.161)</td>
</tr>
<tr>
<td></td>
<td>(0.33, 0.33)</td>
<td>(0.325, 0.323)</td>
<td>(0.333, 0.328)</td>
</tr>
</tbody>
</table>
rameters averaged over 50 independent trials. We then compared the variation of the average SNR in our technique with that of Hoge. Results are shown for the two shift parameters in Figure 2.4. Overall the two methods have similar performance, although it seems that on average, over the range of the noise levels of this experimentation, Hoge’s method was performing slightly better at the lower half of the noise interval, and our method was performing slightly better at the upper half of the interval. Of course, as can be seen in the plots, the dynamic range of the SNR is very small - i.e. the two methods are very close in terms of their performance over the entire interval.

![Figure 2.3: Simulation with additive noise: (a) image used for synthesizing shifted noisy pairs, (b) a typical phase difference matrix for a noisy pair.](image)

We also applied our technique to a set of real MRI data, used by Hoge in his experimentation. Details of the acquisition setup are described by Hoge in [Hog03]. Figure 2.5 shows a pair of MRI images from this data set, and their phase difference matrix. As can be seen the phase difference matrix is noisy and highly affected by aliasing or high frequency noise, which makes registration particularly challenging [RM04]. We repeated Hoge’s exact experimental procedure. There are
Figure 2.4: Average SNR over 50 independent trials for the standard deviation of noise varying in $[0, 5]$; top shift along x-axis, bottom shift along y-axis.

five images in this data set allowing for ten different combinations of pairwise registrations. Table 2.2 summarizes the shift parameters obtained by our approach compared to those obtained by physical calibration, Stone et al. and Hoge. As described in [Hog03], due to manual displacements the estimations based on the physical calibration may be more erroneous than the computed methods. Therefore similar to [Hog03], we compared the different algorithms using the relative error. Results are shown in Fig. 2.6. For the method proposed by Stone et al., we used the data provided in [Hog03]. Overall, given the very small variations of the relative error, the methods seem to be comparable in terms of performance.

2.1.7 Conclusion

The analyses and the experimentations we presented show that accurate results can be obtained for sub-pixel registration directly in the Fourier domain. Sub-pixel registration directly in the Fourier domain is particularly advantageous for some sensor modalities that provide directly the Fourier
Figure 2.5: (a) An image from the MRI data set (Courtesy of W.S. Hoge), (b) discrete phase difference matrix of the first pair in the data set.

Table 2.2: Pairwise registration of the MRI data set.

<table>
<thead>
<tr>
<th>Image Pairs</th>
<th>Physical (in pixels)</th>
<th>Stone et al.</th>
<th>Hoge</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2)</td>
<td>(-2.40,-4.00)</td>
<td>(-2.06,-5.97)</td>
<td>(-2.05,-4.02)</td>
<td>(-2.11,-4.00)</td>
</tr>
<tr>
<td>(1,3)</td>
<td>(-4.80,-8.00)</td>
<td>(-4.28,-9.99)</td>
<td>(-4.26,-8.01)</td>
<td>(-3.90,-7.49)</td>
</tr>
<tr>
<td>(1,4)</td>
<td>(-7.20,-4.32)</td>
<td>(-6.61,-5.67)</td>
<td>(-6.61,-4.33)</td>
<td>(-6.22,-3.93)</td>
</tr>
<tr>
<td>(1,5)</td>
<td>(-7.20,-12.00)</td>
<td>(-6.62,-13.97)</td>
<td>(-6.61,-12.03)</td>
<td>(-6.39,-11.42)</td>
</tr>
<tr>
<td>(2,3)</td>
<td>(-2.40,-4.00)</td>
<td>(-5.86,-3.85)</td>
<td>(-2.21,-3.98)</td>
<td>(-2.18,-3.87)</td>
</tr>
<tr>
<td>(2,4)</td>
<td>(-4.80,-0.32)</td>
<td>(-4.55,-1.71)</td>
<td>(-4.54,-0.32)</td>
<td>(-4.16,-0.31)</td>
</tr>
<tr>
<td>(2,5)</td>
<td>(-4.80,-8.00)</td>
<td>(-4.56,-10.00)</td>
<td>(-4.55,-8.01)</td>
<td>(-4.13,-7.73)</td>
</tr>
<tr>
<td>(3,4)</td>
<td>(-2.40,3.68)</td>
<td>(-2.34,2.34)</td>
<td>(-2.33,3.66)</td>
<td>(-2.34,3.55)</td>
</tr>
<tr>
<td>(3,5)</td>
<td>(-2.40,-4.00)</td>
<td>(-2.34,-5.97)</td>
<td>(-2.34,-4.03)</td>
<td>(-2.49,-3.83)</td>
</tr>
<tr>
<td>(4,5)</td>
<td>(0.00,-7.68)</td>
<td>(-0.81,-8.05)</td>
<td>(-0.01,-7.69)</td>
<td>(-0.03,-7.84)</td>
</tr>
</tbody>
</table>
Figure 2.6: The relative error in our registration compared to the physical calibration, Stone et al., and Hoge, for the MRI data set.

spectrum of the field of view. Examples of such imaging modalities are MRI and SAR.

We have shown that in the discrete case the phase matrix of the cross-power spectrum of two images is a 2-dimensional sawtooth signal. A specific consequence of this is that the shift parameters can be decoupled and estimated separately. This is similar to [Hog03], but we do not require a subspace approximation. The shift parameters (whether integer-valued or not) are determined by counting the cycles of the sawtooth phase matrix in any row or column. Furthermore the counting of the cycles is performed using a robust estimator to alleviate the sources of error.
Our method is essentially a two-step algorithm. The first step involves a constrained least-squares estimation of a set of shift values, and the second step involves truncating the quadratic error by clustering the computed shift values using EM. For MRI data (since no Fourier transformation is required)

- Like any direct least-squares solution, the time complexity of the first step is polynomial of 3rd order.

- The time complexity of finding the optimal regularization parameters is very small, since it only requires the computation of the largest eigenvalues over small matrices, i.e. $O(N(2L + 1)^2)$.

- The time complexity of the EM in the second step is linear per iteration. Its convergence is known to depend on the conditioning of the likelihood function [XJ96]. We found that due to our very close initialization, the second step typically converges within one or two iterations.

As in any registration technique the performance may degrade if the transformation model does not hold. We found that if the true motion is sufficiently close to translation then the degradation in performance may be small. Of course this point needs more detailed investigation.

### 2.2 SUB-PIXEL REGISTRATION DIRECTLY FROM THE PHASE DIFFERENCE BY USE OF HOUGH TRANSFORMATION

This section uses the concepts underlying the phase correlation method. We use the facts shown in previous section, but this time we formalize the problem by use of Hough Transformations as
a new approach compared to previous chapter. By use of the same background information, the formulation of the problem becomes straightforward as described in the subsequent sections.

Since by changing from Cartesian to polar coordinates both translation and rotation can be treated in the same manner, we discuss the background and our formulation in terms of translation.

### 2.2.1 Problem Formulation

As indicated before, the key to solve the problem is to find how many cycles of the discrete sawtooth phase difference fit in the range \([0, 2\pi]\) along each frequency axis. The number of cycles i.e. \(x_o\) and \(y_o\) may or may not be integer values, and are given by

\[
x_o = \frac{\text{cycles}}{2\pi} = \frac{N \, d\Phi(m, n)}{2\pi \, dn}
\]

\[
y_o = \frac{\text{cycles}}{2\pi} = \frac{M \, d\Phi(m, n)}{2\pi \, dm}
\]

Due to noise and other sources of error, however, counting the number of cycles per \(2\pi\) using equations (2.43) and (2.44) may lead to inaccurate results. To overcome this problem, we exploit the fact that a total of \(M \times N\) data points are available for regression. Therefore, an accurate solution can be obtained by minimizing an appropriate error function, i.e. by solving a largely overdetermined problem of fitting the parameters to the data set.

For this purpose note that the 2D sawtooth signal \(\Phi\) has constant slopes for the vast majority of frequencies along each row or each column, except for a small number of frequencies, where discontinuities occur. As depicted in Figure 2.1-(e), visually we use a landmark point of the cycle to count the number of cycles along each row or each column. This idea can be implemented as follows. A particular useful landmark point is the zero-crossing of the phase difference matrix,
where $\phi_{mn} = 0$. Since the same set of zero-crossings can be independently obtained by scanning along the rows or columns, we can fix $J = 0$ in equation

$$\phi_{mn} = \frac{2\pi}{N} \left( x_0 n + J \frac{N}{x_0} \right) + \frac{2\pi}{M} \left( m y_0 + K \frac{M}{y_0} \right)$$

(2.45)

where $J$ and $K$ are arbitrary integers, and after some algebraic manipulations obtain

$$n \cos \theta + m \sin \theta + \rho = 0$$

(2.46)

where

$$\tan \theta = \frac{N y_0}{M x_0}$$

(2.47)

and

$$\rho = K \frac{M}{y_0} \sin \theta$$

(2.48)

This shows that the zero-crossings are represented by a family of lines in the phase space parameterized by $K$ - i.e. each integer value of $K$ would give a different line along which the phase difference is zero. Each line itself is parameterized by the angle $\theta$ and its distance $\rho$ from the center frequency (i.e. the origin of the frequency domain). This set of zero-crossing lines form a function that is ideal for detection using Hough transform. The Hough transform basically maps these lines to a parameter space of $(\theta, \rho)$. As can be verified from the above derivations, $\theta$ remains invariant among all lines and $\rho$ varies as integer multiples of some other invariant parameter, i.e. $\rho = K \rho_0$. Therefore in the Hough-transform domain (i.e. in the Hough matrix), we expect to see a set of peak values situated at equal distances from each other, and parallel to the $\rho$-axis. Figure
2.7 shows an example of the Hough transform of the zero-crossing of the phase difference matrix of two shifted images, where the peaks can be clearly identified by a simple thresholding process. As is customary in Hough transform, we used the local maxima for finding a suitable threshold value. In our case, since all the peaks are known to be aligned parallel to the $\rho$-axis, we took the maximum of the Hough matrix for each $\rho$ as the local maximum. This yields a curve similar to the one shown in Figure 2.7(b). We then used the average of the local maxima curve as a threshold.

The problem now reduces to estimating $(x_o, y_o)$ from these peak values in the Hough transform domain. For this purpose note that by combining (2.47) and (2.48), we can obtain the following linear constraint on $x_o$ and $y_o$:

$$\frac{\rho}{\cos \theta} x_o + \frac{\rho}{\sin \theta} y_o = K(M + N)$$

(2.49)

Each peak point in the Hough-transform domain provides one such linear constraint on $x_o$ and $y_o$. Given a total of $t$ such peak values, we can construct a system of linear equations of the form

$$A r = b$$

(2.50)

where

$$A = \begin{bmatrix}
\frac{\rho_1}{\cos \theta_1} & \frac{\rho_1}{\sin \theta_1} \\
\vdots & \vdots \\
\frac{\rho_t}{\cos \theta_t} & \frac{\rho_t}{\sin \theta_t}
\end{bmatrix}$$

(2.51)

$r = [x_o \ y_o]^T$, and $b = (M + N)[K_1 \cdots K_t]^T$.

In the next section, we devise an approach to solve this overdetermined problem.
Figure 2.7: (a) The Hough transform of the phase discontinuities, (b) the local maxima and the threshold, Peaks detected by thresholding.
2.2.2 Proposed Solution

In order to solve the over-determined system of equations in (2.50), we formulate it as the following constrained optimization problem

\[
\mathbf{r}_{\text{opt}} = \arg \min \| \mathbf{A} \mathbf{r} - \mathbf{b} \| + \lambda \| \mathbf{L} \mathbf{r} \| \quad (2.52)
\]

where \( \lambda \) is the regularization parameter, and \( \mathbf{L} \) is such that

\[
\mathbf{L}^T \mathbf{L} = \begin{bmatrix}
2 & -1 \\
-1 & 2
\end{bmatrix} \quad (2.53)
\]

This choice of \( \mathbf{L} \) [TA77] implies that our \textit{a priori} belief is that our solution should be constant over all equations in the system, i.e. the equations should be consistent with each other. The formulation in (2.52) is basically a Tikhonov-Arsenin norm-regularization of the least-squares solution to the problem in (2.50). The first term imposes the faithfulness to data and the second term the regularity.

The solution is given by

\[
\mathbf{r}_{\text{opt}} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T \mathbf{b} \quad (2.54)
\]

We now have the solution given by (2.54) up to an unknown regularization parameter. The optimal value of this parameter is given by the method of Generalized Cross Validation (GCV), which amounts to minimizing

\[
\text{GCV}(\lambda) = \frac{\| (\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T) \mathbf{b} \|^2}{(\text{tr}(\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T))^2} \quad (2.55)
\]
with respect to $\lambda$, where $\text{tr}(\cdot)$ is the trace of a matrix.

The minimizer of (2.55) is usually obtained using numerical techniques by making simplifying assumptions, e.g. circulant $A$ and $L$, or by using numerical techniques such as quadrature rules and Lanczos algorithm [GH79]. However, in our case, due to the rank constraint of $A$, we can find a simplified closed-form solution. For this purpose, let $P = I - A(A^T + \lambda L^T L)^{-1}A^T$. The GCV function can then be written as

$$
\text{GCV}(\lambda) = \frac{\|Pb\|^2}{(\text{tr}(P))^2} \tag{2.56}
$$

Upon rearranging $P$ as follows

$$
P = I - AL^{-1}L(A^TA + \lambda L^TL)^{-1}L^TL^{-T}A^T
= I - AL^{-1}(L^{-T}A^TAL^{-1} + \lambda L^{-T}LTL^{-1})L^{-T}A^T
= I - K(K^TK + \lambda I)^{-1}K^T \tag{2.57}
$$

and applying the matrix inversion lemma, we find

$$
P = \left(I + \frac{1}{\lambda}KK^T\right)^{-1} \tag{2.58}
$$

where $K = AL^{-1}$.

Now, let

$$
KK^T = V\Sigma V^T = \sum_{j=1}^{t} \sigma_j v_j v_j^T \tag{2.59}
$$
be the eigen-decomposition of $KK^T$, where $v_j$’s are the columns of $V$ that form a set of orthonormal basis, and $\sigma_j$’s are the corresponding eigenvalues. $P$ can then be written as

$$P = \sum_{j=1}^{t} \frac{\lambda}{\lambda + \sigma_j} v_j v_j^T$$

$$= \sum_{j=1}^{2} \frac{\lambda}{\lambda + \sigma_j} v_j v_j^T + \sum_{j=3}^{t} v_j v_j^T$$

where the last equality follows from the fact that $A$ is rank-2.

In order to simplify the GCV function in (2.56), we make a first order approximation of the rank-2 matrix $KK^T$ using its largest eigenvalue: $KK^T \simeq \sigma_1 v_1 v_1^T$. In practice, we found that the dominant eigenvalue is usually orders of magnitude larger than the second eigenvalue (see the Appendix for more formal discussions). Therefore

$$\text{GCV}(\lambda) \simeq \frac{\left(\frac{\lambda}{\lambda + \sigma_1}\right)^2 s_1^2 + \sum_{j=2}^{t} s_j^2}{\left(t - 1 + \frac{\lambda}{\lambda + \sigma_1}\right)^2}$$

(2.62)

where $s_j$ are the components of the vector $V^T b = [s_1 \ s_2 \ldots s_t]^T$.

Differentiating this equation with respect to $\lambda$ and setting it equal to zero, we find after simplification that the optimal regularization parameter is given by

$$\lambda^* = \frac{\sigma_1 \sum_{j=2}^{t} s_j^2}{(t-1)s_1^2 - \sum_{j=2}^{t} s_j^2}$$

(2.63)

Using (2.54) and (2.63), we can compute the optimal solutions for $x_o$ and $y_o$. 

57
2.2.3 Experimental Results

We applied the technique to both the global sub-pixel registration problem and the local motion estimation. We also applied our approach to rotated images, using polar coordinate transformation. The experimentations included both synthetically generated shifts, and real data with and without ground truth. In all cases very good results were obtained. We used the synthetic simulations to evaluate our approach under various conditions.

For global registration, we synthetically generated the shifts by filtering and down-sampling shifted versions of a high-resolution image. Using appropriate combinations of these operations, shifts with sub-pixel contents were produced. Figure 2.8 shows some of the images to which the technique was applied. Results are shown in Table 1, and are compared to those reported in [FZB02]. The accuracy was predominantly higher than [FZB02], which we attribute to the largely overdetermined nature of the problem in the proposed approach. This of course introduces resiliency to random noise (assuming that random noise is zero-mean and has a well-behaved distribution).

Table 2.3: Table of results for global shifts of the images in Figure 2.8

<table>
<thead>
<tr>
<th>Image</th>
<th>True Shifts</th>
<th>Foroosh et al. 2001</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentagon</td>
<td>(0.50, -0.50)</td>
<td>(0.48,-0.51)</td>
<td>(0.496,-0.493)</td>
</tr>
<tr>
<td></td>
<td>(0.25, 0.50)</td>
<td>(0.28,0.49)</td>
<td>(0.255,0.498)</td>
</tr>
<tr>
<td></td>
<td>(-0.25, -0.50)</td>
<td>(-0.25,-0.52)</td>
<td>(-0.25,-0.52)</td>
</tr>
<tr>
<td></td>
<td>(0.0,0.75)</td>
<td>(0.0,0.80)</td>
<td>(0.0,0.744)</td>
</tr>
<tr>
<td>Paris</td>
<td>(0.167, -0.5)</td>
<td>(0.152,-0.49)</td>
<td>(0.163,-0.51)</td>
</tr>
<tr>
<td></td>
<td>(0.67, 0.25)</td>
<td>(0.69,0.33)</td>
<td>(0.679,0.242)</td>
</tr>
<tr>
<td></td>
<td>(-0.33, -0.167)</td>
<td>(-0.32,-0.15)</td>
<td>(-0.331,-0.159)</td>
</tr>
<tr>
<td></td>
<td>(0.33, 0.33)</td>
<td>(0.325,0.32)</td>
<td>(0.333,0.329)</td>
</tr>
</tbody>
</table>
As for random noise, it is worth noting that in many practical problems dealing with data obtained from real imaging instrumentation, one may typically have the high frequency portion of the phase difference matrix cluttered and highly corrupted with noise. Figure 2.9 shows an example of such phase difference matrix. This in practice does not introduce a major difficulty in our algorithm, since similar to [Hog03, SOC01] we can apply our algorithm within a window in the low-frequency portion, by cropping out the lower frequencies of the phase difference matrix. The important remark to realize is that, we would then be counting the number of cycles within less than $2\pi$. So for instance equations (2.43) and (2.44) (or other equations thereafter) would be adjusted accordingly.

When the data is contaminated with random noise or aliasing, the high frequency portion of the phase difference matrix is not useful, and in fact can introduce error in the results. In order to avoid this problem, similar to [Hog03, SOC01], we windowed the Fourier spectra of the images
to their low frequency portion. Therefore, in order to evaluate the performance of our approach under noise, we introduced additive Gaussian random noise with standard deviation in the range $[0, 5]$. We then computed the peak signal to noise ratio (PSNR) over 100 independent trials as we varied the size of the low-pass window, where we took the true shift value as the peak value of the signal. The results are shown in Figure 2.9 for $\sigma = 5$, and low-pass windows of width 10-100. It can be seen that the method is fairly stable over a wide range of window sizes. The SNR for the $y$ coordinate was better in this experimentation, because the shift along that direction was an integer value.

We then applied the technique to real data in a framework using short-length Fourier transform. The images used are stereo pairs from the CMU data set [CMU], and the Tsukuba pair. Results are shown in Figures 2.10, 2.11, and 2.12 for the Pentagon, baseball, and the Tsukuba image pairs, respectively. Of course, one challenge in the case of real data is that due to lack of ground truth the performance evaluation is somewhat subjective. For our real data, however, it is important to note that we used small windows (e.g. of width 15-25) to compute the short-length Fourier transform. This has a very important implication, since unlike global registration, we now have a very small set of linear equations to estimate the parameters.

Essentially, in global registration, we estimate two parameters using a huge number of pixels (i.e. for a $256 \times 256$ image, $t$ is typically a very large number). In the case of local motion estimation, if we for instance use window sizes of $15 \times 15$, $t$ is typically reduced by a factor of 300. As a result, it is quite natural to assume that the accuracy will drop. It should be however noted that this comment equally applies to all existing methods in the literature. In order to be able to evaluate how the method performs for varying lengths of short-length Fourier transform,
Figure 2.9: (a) A phase difference matrix corrupted by aliasing and high-frequency noise, (b) average PSNR over 100 independent trials with varying window sizes and the standard deviation of noise $\sigma=5$. 
we synthetically generated a stereo pair with known ground truth (see Figure 2.13), where the parallax for the house in the image is artificial. We then computed the PSNR for varying range of window sizes of short-length Fourier transform. The plot in Figure 2.13 shows the result of this experimentation. One can notice that the PSNR reaches the highest value for an optimal window size. Our interpretation is that for local motion, if the window size is too small, then the number of data points would be insufficient to get the optimal result, and if it is too large it may get dominated by the motion in the neighboring pixels. This is similar to what is known to as the aperture problem in optical flow. There is in fact an optimal window size for which we get the best trade-off between the two constraints.

Finally, we applied our method to estimate rotation in polar coordinates. For this purpose we used also a real stereo pair from the CMU data collection [CMU], for which the ground truth was known. Figure 2.14 shows the stereo pair, their polar coordinate representations, and the resulting phase difference matrix. We estimated the rotation as 4.737 degrees, which nicely matched the ground truth. Note that the number of the sawtooth cycles can be used to determine the rotation even if only a small portion is useful (i.e. not badly contaminated with noise). Also, note that the we can virtually get the correct solution regardless of number of steps used to sample the angular axis in the polar domain. For instance in Figure 2.14 the steps used (as can be seen from the horizontal axis) are every one degree. But our solution is not limited to one degree accuracy due to sub-degree (sub-pixel) accuracy provided by the proposed method. Therefore, for instance, we could sample the angular axis at every 2-degrees and still get sub-angle results as long as the aliasing can be avoided by low-pass windowing as discussed above.
2.2.4 Conclusion

The analyses and the experimentations presented in this chapter show that accurate results can be obtained for sub-pixel registration directly in the Fourier domain, by counting the number of cycles of the phase difference matrix. Since no inverse transforming is required the computational complexity is mostly dominated by the forward Fourier transform, which is $N \log N$. There are of course small overheads associated with computing the phase difference matrix and its Hough transform. In the context of super-resolution, sub-pixel registration directly in the Fourier domain is particularly advantageous and of interest for the super-resolution algorithms that work directly in the Fourier domain. It is also worth mentioning that some sensor modalities naturally provide the Fourier spectrum of the field of view. Examples of such imaging modalities are magnetic resonance imaging (MRI) [Wri97], and synthetic aperture radar (SAR) [Sko90].
Figure 2.10: Pentagon stereo pair and the estimated local motion.
Figure 2.11: Baseball stereo pair and the estimated local motion.
Figure 2.12: Tsukuba stereo pair and the estimated local motion.
Figure 2.13: A synthetic stereo pair and the PSNR as a function of the window size used for short length Fourier transform in local motion estimation.
Figure 2.14: Top and middle: Stereo pair and their corresponding polar representations. Bottom: Their phase difference matrix and one row of the matrix.
CHAPTER 3
ALIGNMENT OF MAXILLOFACIAL CT SCANS TO STONE-CAST MODELS USING 3D SYMMETRY FOR BACKSCATTERING ARTIFACT REDUCTION

3.1 INTRODUCTION

In this chapter, we present an approach to accurately align the CT scans of a patient to a stone-cast model of his/her mandible or maxilla, and use the result of the registration to clean up the patient’s scans from artifacts and defects. The proposed approach assumes that the maxillofacial features are roughly symmetric with respect to a 3D plane. Then 3D volumetric models of both the patient and the stone-cast are reconstructed from the input data using a marching cube algorithm. The planes of symmetry are extracted using an improved Extended Gaussian Images method. After an initial alignment of the two volumes guided by the plane of symmetry due to 3D homology, we minimize a global cost function that depends on the sum of square differences (SSD) of patient data with the stone-cast model to finally recover the rigid transformation between the two scans.

In dentistry, implant dentistry in specific, there is a need for 3D visualization of internal anatomy. The internal visualization is mainly based on CT scanning technologies. As discussed in [DMD01] the most important technological advancement which dramatically enhanced the clinician’s ability to diagnose, treat, and plan dental implants has been the CT scan. Advanced 3D modeling and visualization techniques permit highly refined, accurate assessment of the CT scan data in a manner, which exceeds information gleaned from film alone [DMD01, SHS04, GP04]. Since the advent of the Conebeam CT technology, which is a low dose oral and maxillofacial scanner, the number of dental CT scanners has been growing exponentially. The strongest barrier in
taking a CT scan is the dose of x-ray exposure, which is not always justifiable in oral surgery cases. Conebeam scanners have an average of 10 times less radiation than regular medical CT scanners [AE01]. However, this comes at a price that the resulting scans are not free of defects.

In addition to imperfections of the instrument and the imaging process, it is not uncommon to encounter other unwanted artifacts in the form of bright regions, flares, and erroneous pixels due to for instance dental bridges or metal braces. Currently, removing and cleaning up the data from acquisition imperfections and unwanted artifacts is performed manually, which is as good as the experience level of the technician. On the other hand the process is error prone, since the editing process needs to be performed image by image.

We attempt to address some of these issues by using stone-cast models of patients’ dental imprint. Stone-cast models were originally used by dentists to make complete or partial dentures. It is suggested herein that the CT scan of such stone-cast models can be used to automatically guide the cleaning of patient’s CT scans from defects or unwanted artifacts. The key, however, to such application would be an accurate registration of the two CT scan data, i.e. the patient’s maxillofacial CT scans and that of the stone-cast model.

The problem is rather challenging since, in some cases, it is a priori difficult to exactly identify which parts in the patient’s scans are soft tissues, nerves, or bones. Prior radiodensity scales defined by Hounsfield units often give only approximate solutions, and are not always available in the patient’s DICOM file. In order to solve this problem, we propose that the registration process can be drastically simplified and improved if we use the fact that the maxillofacial features are roughly symmetric. Symmetry is a very strong cue that has been used in the past in imaging problems for the somewhat opposite application of recovering 3D models [ZT98, AGR02]. Once
the two scans are accurately registered the cleaning process can be performed by simple masking
and morphological erosion.

In next section, Section 3.2 we discuss the geometry of a mirror-symmetric object, in Section
3.3 we explain the proposed framework for registration of CT scans of a patient to a stone-cast
model of his/her mandible or maxilla; Section 3.3.1 explains the technique used to extract plane
of symmetries, Section 3.3.2 focuses on matching of these planes and Section 3.3.3 explains local
optimization step of the proposed framework. Identifying the outliers of maxillofacial ct scan data
without using a stone-cast model is explained in section 3.4 and in Section 3.5 we explain the
inpainting approach used for the outlier voxels.

### 3.2 MIRROR SYMMETRY

Symmetry is a ubiquitous property of many natural and man-made objects, and is one of the
strongest cues used recently in computer vision for various applications such as 3D modeling
and image-based rendering. Symmetry imposes strong geometric constraints on objects that can
be exploited in the recovery of motion and structure. Symmetry may appear in objects in the form
of central symmetry, axial symmetry (surface of revolution), or the most frequently encountered
mirror symmetry. Examples include a face, a car, a chair, etc. In mirror symmetry, symmetry is
defined with respect to a plane $\Pi$ in the 3D space, so that for any point $\tilde{M} \sim [X \ Y \ Z \ 1]^T$
the object there exists a point $\tilde{M}' \sim [X' \ Y' \ Z' \ 1]^T$ on the object such that the line $MM'$ is
orthogonal to $\Pi$ and $d(M, \Pi) = d(M', \Pi)$, where $d(\cdot, \cdot)$ stands for the Euclidean distance and
indicates equality up to a non-zero scale factor. In the 3D space any such symmetric pair of points
are related via

\[ M' \sim \left( I_4 - 2 \frac{vv^T}{v^Tv} \right) \quad (3.1) \]

where \( v \sim \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T \), \( I_4 \) is an identity matrix and \( T \) stands for the transpose. Since CT scans are projections of planar slices of 3D volumetric data, they can be represented by a homography \( H \). Therefore, the corresponding image domain points would satisfy

\[ H^{-1}M' \sim \left( I_3 - 2 \frac{\ell \ell^T}{\ell^T \ell} \right) H^{-1}M \quad (3.2) \]

which would yield the following relation between the symmetric image points.

\[ m' \sim H \left( I_3 - 2 \frac{\ell \ell^T}{\ell^T \ell} \right) H^{-1}M \]
\[ \sim \left( I_3 - 2 \frac{\ell \ell^T}{\ell^T \ell} \right) m \]
\[ \sim H_i m \quad (3.3) \]

where \( I_3 \sim \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \).

This shows that under mirror symmetry the projections of symmetric points in the image plane are related via a homology.
3.3 FRAMEWORK

Our method includes the following steps:

1. Estimation of the plane of symmetry for both patient’s and stone-cast CT scans.

2. Alignment of the planes of symmetry.

3. Alignment of scans by searching among transformations that would fix the plane of symmetry.

4. Refinement using global minimization based on sum of square differences (SSD).

3.3.1 Extraction Of Planes Of Symmetry

A natural solution to estimate the planes of symmetry for the patients’ scan data and that of the stone-cast model might appear to be based on minimizing an error defined by 3D Homography. However, we found that such an approach would be computationally expensive and would require a manual initialization. Therefore in this section we propose an approach based on working directly on the volumetric data. Similar to [SS97] our method makes use of Extended Gaussian Images (EGI) [Hor64] to identify the 3D reflection plane. EGI have been proven to be very useful tools in such applications as pose estimation, orientation, and recognition [MR99, GSR04, Gos05, Rob63, EPV93b, Rob00]. The basic idea is exploiting the observation that in practice, if an object is symmetric then so is its EGI. Therefore, by sampling the EGI and representing its histogram one can directly identify a point on the plane of symmetry as well as the normal to the plane. Sun and Sherrah [SS97] proposed a discrete approach to this problem. The first step in their algorithm is to tessellate a sphere in order to use it as an orientation histogram. A search is then performed near
the principal axis directions for the strongest symmetries by performing correlation operations on
the histogram. A unique labeling for the triangular facets (bins) resulting from the tessellation is
required and thus also an efficient algorithm to determine the facet in which a given normal lies.

We avoid the tessellation, the labeling, and the need for an efficient search algorithm by simply
parameterizing the unit sphere uniformly. We then create a two-dimensional orientation histogram
by means of a simple table, where the indices $i$ and $j$ are determined as following:

$$
i = \text{floor}(u_1 n_y), \ j = \text{floor}(u_2 n_x)
$$

(3.4)

such that $0 \leq u_1, u_2 \leq 1$, $n_x$ is the x-resolution and $n_y$ is the y-resolution. Since we are dealing
with symmetric objects a resolution of less than square root of the number of normals is often
sufficient. Assuming that the CT scans are triangulated, with all normals computed and depicted
by two angles for each triangle in the facial model, the normal $\mathbf{n} = [n_x, n_y, n_z]$ is computed, and
hence its direction, depicted by the two angles $\varphi$ and $\theta$, would be given by:

$$
\theta = \arccos(n_z), \ \varphi = \arctan\left(\frac{n_y}{n_x}\right)
$$

(3.5)

and are then determined as following:

$$
u_1 = \frac{1 - n_z}{2}, \ u_2 = \frac{\varphi}{2\pi}
$$

(3.6)

Once the orientation histogram is populated, a search is conducted for the largest count of normals.
The corresponding indices $i$ and $j$ are retrieved and the xyz-coordinates of the most common normal
\( \mathbf{n} \) is determined using the following inverse relations:

\[
\begin{align*}
\begin{align*}
    u_1 &= \frac{i}{n_y}, \quad u_2 = \frac{j}{n_x} \\
    \varphi &= 2\pi u_2 \\
    n_z &= 1 - 2u_1 \\
    r &= \sqrt{\max(0, n_z^2)} \\
    n_x &= r \cos \varphi, \quad n_y = r \sin \varphi
\end{align*}
\]

Where \( r \) is the radius. The equation of the symmetry plane can now be found using the normal \( \mathbf{n} \) and the centroid \( \mathbf{c} \) of the model, as the set of points that satisfy:

\[
\mathbf{n}^T (\mathbf{p} - \mathbf{c}) = 0
\]

where \( \mathbf{p} \) is a point on the symmetry plane. We found excellent results by applying this approach.

### 3.3.2 Matching Symmetry Planes And 2D Search

Once the planes of symmetry are found in both the patient scans and the stone-cast scans, the two volumetric models can be pre-aligned. This is performed automatically by overlaying the two planes. In practice, this would provide an extremely good initialization, up to an unknown transformation with three degrees of freedom, i.e. two translation parameters and one rotation.

After the planes of symmetry are matched, we initially assume that the rotational misalignments are negligible. This reduces the search space for registration to pure 2D translation. The correct translation parameters are then obtained by searching for 2D translation parameters that would
Figure 3.1: (a) 2D view for the computed plane of symmetry in the stone-cast CT scans (b) Original Patient Model (c) Corresponding Stone Cast Model.
Figure 3.2: (a) Initial Models with planes of symmetry (b) Models after planes of symmetry are matched (c) Models after planes of symmetry are matched (Top View).
Figure 3.3: (a) A section of zoomed Patient Data (b) Partial zoom view of corrected data by use of stone-cast matching (c) Patient Model after clean up.
minimize

\[ T^* = \arg \min_T \sum_T \|D - TS\|^2 \]  

(3.9)

where \( D \) is the original patient CT scan data set, \( S \) is the related stone cast CT scan data set, \( T \) is a translation matrix, and \( T \) is the space of all possible translations within the planes of symmetry.

### 3.3.3 Local Optimization

The previous step would essentially provide an almost perfect solution to the problem. However, for medical applications, accuracy of alignment plays a very important role. Therefore after the initial alignment described above, we perform a local optimization in 3D, within some neighborhood to refine and adjust for minor rotational and translational errors. This is done by performing...
the following optimization step

\[ P^* = \arg \min_P \sum_P \| D - P T^* S \|^2 \]  \hspace{1cm} (3.10)\]

where \( T^* \) is calculated at the 2D search step, \( D \) is the original patient CT scan data set, \( S \) is the related stone cast CT scan data, \( P \) is the space of all possible transformations, \( P \) is a rigid transformation from the space of all transformations \( P \) in some neighborhood, and \( P^* T^* \) is the registration result.

### 3.4 IDENTIFYING THE OUTLIERS OF MAXILLOFACIAL CT SCAN DATA

After the registration step is completed, one can use the registration result in many ways; to prune the patient CT scan data, to overwrite the patient CT scan by corresponding stone-cast CT scans, etc. This usage of registration solves the problem of outliers which will be segmented out by the stone-cast registration. However for the outliers that are within the stone-cast region or in the absence of stone-cast models, there is a need for outlier detection. We can detect outliers, and we can use the stone-cast values to fill the gap, which will be caused by the removal of outliers.

We propose an automatic thresholding algorithm which extends the isodata thresholding algorithm [RC78, FHM00]. Originally, the isodata algorithm works on the principle of evaluating a unique threshold \( \tau \) for a given histogram assuming

\[ \tau = \frac{\mu_0 + \mu_1}{2} \]  \hspace{1cm} (3.11)\]
such that $\mu_0$ and $\mu_1$ are the means of components defined by the threshold. Since they are not known before having the threshold, an iterative algorithm was proposed:

- Select an initial threshold
- Calculate two means of the segments defined by current threshold
- Set new threshold to be average of calculated means
- Repeat until the threshold value converges

which corresponds to finding the threshold $\tau$ that minimizes the following error function $e^2$:

$$
e^2 = \int_0^\tau (i - A)^2 h(i)di + \int_\tau^N (B - i)^2 h(i)di$$

where $i$ indicates the gray level of the pixels (from 0 to N), $h(i)$ is the histogram weighting for every ray level, $A$ and $B$ are the binary values of each of the two components, and $\tau$ is the threshold. For proof, please refer to [FHM00].

The proposed thresholding algorithm is based on the idea that a good threshold should still be a good candidate for different neighborhoods defined by a window mask. So we propose to extend the isodata thresholding algorithm by applying a short length windowed version through the histogram of complete patient CT scan data. Each resulting short-length threshold candidate increases the corresponding gray level’s vote count. This approach is repeated with increasing window sizes as well to provide multi-scale solution. At the end, the threshold with the maximum votes is announced as the best current threshold for the given histogram. With the assumption that our data is composed of space (background), non-bone (tissue, etc), bones and outliers which
are generated due to imperfections or dental add-ons, the problem converts into finding 3 good thresholds by repeating the above process. Moreover due to the nature of the data, the first threshold which corresponds to the space is already known. After a threshold is found, the histogram is cropped accordingly and the next threshold is searched on the rest of the histogram. This approach led us to finding multiple good thresholds.

Algorithm 3.4.1: FINDOUTLIERTHRESHOLD(PD, bc, ϵ, wsize)

comment: PD is Patient Data, bc is bin count, wsize is minimum window size

[H T]← GenerateHistogram(PD, bc);

comment: First bin will correspond to space due to nature of our data

comment: so initialize first Threshold to be at bin 1

Threshold_index ← 1;

for each threshold

comment: We should look for the next threshold in the remainder of the histogram;

HH = H(Threshold_index * ϵ : end);

TT = T(Threshold_index * ϵ : end);

bin_count = bin_count − Threshold_index * ϵ;

votes = zeros(size(TT));

GenerateThresholdVoteMap();

Threshold ← find_threshold_with_maximum_votes;
Algorithm 3.4.2: \texttt{GENERATETHRESHOLDVOTE}\texttt{MAP}(PD, bc, \epsilon, wsize)

\textbf{comment}: PD is \textit{Patient Data}, bc is \textit{bin count}, wsize is \textit{minimum window size}

\begin{verbatim}
for window_size ← wsize to bin_count
\end{verbatim}

\begin{verbatim}
for window_start ← 1 to bin_count
\end{verbatim}

\begin{verbatim}
window_end ← window_start + window_size;
\end{verbatim}

\textbf{comment}: Find the threshold for the given search window neighborhood

\begin{verbatim}
Vtemp = HH(window_start : window_end);
Btemp = TT(window_start : window_end);
\end{verbatim}

\begin{verbatim}
window_threshold ← isodata(Vtemp, Btemp);
\end{verbatim}

\textbf{comment}: Increment the vote count of the related bin

\begin{verbatim}
i ← index of the bin which has window_threshold;
\end{verbatim}

\begin{verbatim}
votes[i] ← votes[i] + 1;
\end{verbatim}

\begin{verbatim}
Move search window towards the right
\end{verbatim}

\begin{verbatim}
window_size ← window_size + 1;
\end{verbatim}

3.5 INPAINTING

There are two classes of algorithms for removing objects and gap-filling, “texture synthesis” and “inpainting”. Texture synthesis algorithms are useful for generating large regions in source images, and inpainting algorithms are useful for filling small gaps in images [CPT03]. Texture synthesis is a way to fill large image regions with repetitive two-dimensional textural patterns,
Figure 3.5: A slice of patient data

Figure 3.6: Gray level distribution of a slice
Figure 3.7: Single thresholding an axial view slice and its effect on segmentation

Figure 3.8: Gray level distribution histogram of complete 3D data set
Figure 3.9: (a) Histogram of a complete CT Scan data set (b) Histogram after background is removed (c) histogram after tissues removed (d) histogram after final outlier threshold is applied. One may consider our approach as if focusing to the least occuring segment and looking for an automatic threshold to differentiate between outliers and the actual data. Please note that there are just a few hundred voxels marked as outlier in $\sim 70$ millions.
Figure 3.10: An example of sample data, corresponding outliers marked as green and a result of outlier removal is shown. Note that the color range is also improved by outlier removal as well.
Figure 3.11: After searching for a good threshold for the last block of original histogram, we are able to detect outliers. Examples of this kind of segmentation are shown in the figures.
based on a large body of texture synthesis research that seeks to replicate texture ad infinitum, given a small source sample of pure texture. Exemplar-based texture synthesis techniques generate new texture by sampling and copying color values from the source images [Ash01, FPC00, Gar81, Har01, Kan79]. A number of algorithms [BCV01, BBS01, BSC00, CS01, ZFC02] address the image filling issue for the task of image restoration. These image inpainting techniques fill holes in images by propagating linear structures into the target region via diffusion. They are inspired by the partial differential equations of physical heat flow, and work convincingly as restoration algorithms. Their drawback is that the diffusion process introduces some blur, which becomes noticeable when filling larger regions.[CPT03]

We adapt two different approaches for medical imaging: first, using the technique developed by [CPT03] but applying it to each slice of the 3D medical data. second, extending the work of Criminisi to 3D.

3.5.1 2D and 3D Inpainting

Criminisi et. al [CPT03] proposed a novel and efficient algorithm that combines the advantages of inpainting and texture synthesis, which optimizes the correct order of filling process by a best first algorithm in which confidence in the synthesized pixel values are propagated in a manner similar to the propagation of information in inpainting. The priority of a given patch $\Psi_p$ at point $p$ is defined as product of confidence $C(p)$ and data $D(p)$ terms as:

$$P(p) = C(p) \times D(p)$$  \hspace{1cm} (3.13)
where \( C(p) = \frac{\sum_{q \in \Psi_p \cap \overline{\Omega}} C(q)}{|\Psi_p|} \)

\[
D(p) = \frac{\|\nabla I_p^\perp \cdot n_p\|}{\alpha}
\]

\( \Omega \) as the target region to be removed

\( \phi \) as the source region

\( \delta \Omega \) as the contour

\( \Psi_p \) is the patch that delimits the area to be synthesized as shown in Figure 3.12

Their algorithm is summarized in Table 3.1.

Figure 3.12: Notation Diagram. Given the patch \( \Psi_p \), \( n_p \) is the normal to the contour \( \delta \Omega \) of the target region \( \Omega \) and \( \nabla I_p^\perp \) is the isophote (direction and intensity) at point \( p \). The entire image is denoted with \( I \) [CPT03].

After determining the outliers by use of the technique proposed in Section 3.4, we mark outliers
as target regions, and apply the inpainting approach proposed by [CPT03] to each segmented slice of medical imaging data. The results can be seen in the second column of Tables 3.2 to 3.8.

Table 3.1: Inpainting Algorithm Proposed by Criminisi [CPT03]

1. Extract the manually selected initial front $\delta\Omega^0$.

2. Repeat until done:
   (a) Identify the fill front $\delta\Omega^t$. If $\Omega^t = \emptyset$, exit.
   (b) Compute priorities $P(p) \forall p \in \delta\Omega^t$.
   (c) Find the patch $\Psi_p$ with the maximum priority,
   (d) Find the exemplar $\Psi_q \in \phi$ that minimizes $d(\Psi_p, \Psi_q)$ for the maximum priority patch found in step 2a.
   (e) Copy image data from source patch, $\Psi_q$, to target patch, $\Psi_p$.
   (f) Update confidence term, $C(p) \forall p \in \Psi_p \cap \Omega$

For 3D inpainting, we modify the priority calculations and searched for the exemplar patch to consider the complete 3D medical imaging data instead of just one slice. Our algorithm works as follows:
1. Load 3D Medical Data[Width x Height x Depth]

2. Reset 3D confidence, priority and data arrays;

3. Find delimiters for the 3D target region

4. In the cube covering target region, calculate the priorities of the voxels on the boundary.

5. While there is a target voxel do

   (a) Find the boundary patch with maximum priority in the covering cube.

   (b) Patch that cube with the patch which minimizes SDD.

   (c) Update boundaries, priorities and confidence terms.

This process is just an extension to [CPT03]. During the patching phase, we let the user decide on whether to patch the complete cube or the slice of the matching patch; the selection can be important when each slice of 3D medical data might have extremely different gray level distributions. The results can be seen in the third column of Tables 3.2 to 3.8.

### 3.6 EXPERIMENTAL RESULTS

We have evaluated our method on several cases of corresponding patient/stone-cast CT scans. We show one set of results below. Figure 3.1(b) shows a view of the 3D patient model. Figure 3.1(c) shows a view of the corresponding stone cast model used. While Figure 3.2(a) shows the initial situation after planes of symmetry are extracted, Figure 3.2(c) show the system after the transformation to match the symmetry planes between the stone-cast model and the patient’s CT scan. Once the two scans are registered the patient’s data can be improved by removing the unwanted
Table 3.2: Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting
Table 3.4: Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting
Table 3.5: Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting
Table 3.6: Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting
Table 3.7: Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting

<table>
<thead>
<tr>
<th>Input Image</th>
<th>2D Inpainting Result</th>
<th>3D Inpainting Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image28.png" alt="Image" /></td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Table 3.8: Input Image with marked outliers, Result of 2D Inpainting, Result of 3D Inpainting
artifacts or defects. Figure 3.3(a) shows defects of the input data in detail, and Figure 3.3(b) shows the same region cleaned up by use of our algorithm. The corresponding 3D view after clean up is shown in Figure 3.3(c). Merging these two models also gives us a better result for modeling purposes. Figure 3.4(b) shows two models merged after the registration and clean up processes. Inpainting results are presented from Table 3.2 to Table 3.8 as well.
CHAPTER 4
CONCLUSION AND FUTURE WORK

In this dissertation, we presented two important contributions to subpixel image registration. First, we derived the exact relationship between the continuous and the discrete phase-difference of two shifted images, and showed that their discrete phase difference is a 2-dimensional sawtooth signal. This led to key observations that the unwrapping of a 2-dimensional sawtooth signal is separable, since its unwrapped matrix has to be rank-2, and hence subspace approximation can be avoided [FB04]. More interesting conclusion was the finding that phase unwrapping is an unnecessary step for registration. As a result a robust solution is presented by minimizing an appropriate error function for a total of MxN data points (Tikhonov-Arsenin regularization of Least Squares Problem), Generalized Cross Validation (GCV), and Expectation Maximization. Second, we formalized the problem by the use of Hough transform, and we proposed an alternative solution by defining an over-determined system of equations, and applying Tikhonov-Arsenin norm-regularization and Generalized Cross Validation (GCV). We showed how our algorithms perform on noisy data, rotated images, and local translation searches, and compared our results to the state of the art.

In the latter part of this dissertation, we focused on the application of registration to enhancement of maxillofacial CT data. We proposed a framework for (1) accurately aligning the CT scans of a patient to a stone-cast model of his/her mandible or maxilla, (2) detecting outliers of the 3D CT scan data without using a stone-cast model, and (3) cleaning up the patient’s scans from artifacts and defects with or without the use of stone-cast models. We extended the classical isodata thresholding algorithm for determining multiple thresholds to identify different tissues and outlying data.
due to backscattering, and solved the problem of "gap filling" by use of inpainting techniques. As the core contribution of this framework, we present an efficient algorithm to register two CT scans by using the symmetry of the scanned areas. The approach is proven to be very useful for registering and removing defects of maxillofacial scans of patients that require oral surgery, but cannot be subjected to high dose radiation, in which case Conebeam scanners would play a very important role. The method is already being evaluated at low scale clinically by physicians, and is proven to produce extremely satisfactory results by the end users.

The areas covered by this thesis are open research areas with many future applications. Future work can focus on applications ranging from plastic surgeries, model generation of actual people, low-dose radiation based scanning system improvements and currently we also envision to propose a technique for dual-energy scanning approach.

We plan to extend our work to registration of non-rigid objects, multimodal images, and/or registration of time series data. The outlier detection algorithm developed herein can be applied to perception problems, and we plan to extend our research on high dynamic range image visualization.
This section is a short tutorial on the capabilities of the current UCFMed Software package developed on the Windows platform by use of Microsoft Foundation Classes, C++, and OpenGL. We also developed a 3D Visualization system which uses OpenGL extension for 3D textures.

**A.1 UCFMED PACKAGE**

When you run the executable the main dialog will pop up

![UCF MED CILab](image)

Click the “Select DICOM Files” button to load the files that you want to work on.
Select the data set and click Open. You may revise your selection before clicking the Ok button.

You will realize that there is a “Pre Process” check box. It will be checked by default after you load any data. If you uncheck it the DICOM files will be loaded into memory during runtime, and there won’t be any other views enabled. In short, 3D voxels won’t be generated.

After clicking Ok, please wait until the pre-process ends.

At this point, you will face with document type selection dialog. If you select UCFMed, you will start with the DICOM image viewer open. Otherwise you will start with the OpenGL view which is used to render 3 planes of data. Of course you may prefer to click cancel and open the views later in the same way.
Following will be a sample snapshot of what you will see if you select UCFMed. We show the toolbar of the mainframe and the controls under each view.

The attached toolbar is as follows:

Some functions included are Zoom, Pan, and Distance at the beginning; and at the end, the Panoramic View Buttons. To zoom in: in zoom mode, left click on the pixel where you want to
zoom in. To zoom out: in zoom mode, right click on the pixel where you want to zoom out. The controls under each view are as follows:

Current Image#, Row#, and Col# are shown at the bottom edit fields. By modifying these you can select which image you are viewing. Only one thing is set by key shortcuts. To view the axial image (default) press ‘1’ after selecting the view you want to use. For the other two views the ’2’ and ’3’ keys should be used. Also the mouse wheel is used to traverse between images. Select a view and scroll the mouse wheel forward and you will advance in the image list. Scroll backward and you will go to the previous images. When you move between images, please check the title of the related bar. It will show you the filename and actual image number for axial views; row and column numbers for other views. The first slider bar sets the window center and the second one the window that is used for masking. If you move one of those sliders to zero, they will be reset to the calculated average and range/2 relatively. If you press the “I” key, you will get the DICOM Info Dialog, which shows the properties of the DICOM file loaded. If you wish to create multiple views you can click the New button or select File→New. The same document type dialog will pop up and you can select between UCFMED_GL and UCFMED again. You may also split the current view into 4 and work using this view by using Window, and then the Split menu item.
Sample 3D plane view will be as follows. To rotate the view, click and hold the left or right mouse buttons and move the mouse. You will see the difference of the pressed mouse button easily.
To make measurements, select the Distance button (the ruler). Then click the first pixel on the image and move your mouse, and you will be able to see the distance in millimeters at left bottom.
section of the main window (on the status bar). For example:

![Panoramic View Example](image)

To generate the panoramic view, click the Panoramic view Button (Ps) and click the points that you want your line pass through. End the list by double left clicking.

To view the generated view, press 'P'. Sample line sequence and corresponding panoramic view examples are as follows.
The effect of the window center and window width based segmentation (masking) can be seen in the following two images. Use the slider bars to modify the setting. You can disable/enable the
window masking by use of the “Window Mask” button on the docking toolbar.

This docking toolbar will be used more frequently soon. For now it has the “window mask” button and a second button for test purposes.
Split View can be used as well:

Multiple Views (create as many as you want, click New, select the type, etc):
A.2 3D VISUALIZATION

For 3D Visualization, an OpenGL program which takes extracted raw textures as input was developed. Use of 3D texture extensions let us get better visualization results, without any need of 3D model extraction. Even though for the symmetry plane detection algorithm we had to implement a marching cube base modeling program on the same data, the final marching models are only used for the symmetry plane extraction. Sample results are already shown in Chapter 3 as in Figure 3.1(b) and Figure 3.1(c).
APPENDIX B

SINGULAR VALUES OF KK\(^T\)
In this appendix we show that the largest singular value of $\mathbf{K}\mathbf{K}^T$ is at least three times the second singular value. For this purpose we use the following results from linear matrix algebra:

- The non-zero singular values of $\mathbf{K}\mathbf{K}^T$ are equal to the singular values of $\mathbf{K}^T\mathbf{K}$.
- The singular values of a real square symmetric matrix are always real. See Theorem 6.6.3 in [Ste67].

It therefore follows that the two non-zero singular values of $\mathbf{K}\mathbf{K}^T$ (which are also the two largest singular values) are given by the solution of the characteristic polynomial of the $2 \times 2$ real square symmetric matrix $\mathbf{K}^T\mathbf{K}$, i.e.

$$\lambda^2 - \text{tr}(\mathbf{K}^T\mathbf{K})\lambda + \det(\mathbf{K}^T\mathbf{K}) = 0 \quad (B.1)$$

The two solutions are given by

$$\lambda = \frac{1}{2} \text{tr}(\mathbf{K}^T\mathbf{K}) \left(1 + \sqrt{[\text{tr}(\mathbf{K}^T\mathbf{K})]^2 - 4\det(\mathbf{K}^T\mathbf{K})}\right) \quad (B.2)$$

$$\lambda = \frac{1}{2} \text{tr}(\mathbf{K}^T\mathbf{K}) \left(1 - \sqrt{[\text{tr}(\mathbf{K}^T\mathbf{K})]^2 - 4\det(\mathbf{K}^T\mathbf{K})}\right) \quad (B.3)$$

Since $\mathbf{K}^T\mathbf{K}$ is a real square symmetric matrix, both singular values have to be real. This implies that

$$\frac{\text{tr}(\mathbf{K}^T\mathbf{K})^2}{\det(\mathbf{K}^T\mathbf{K})} \geq 4 \quad (B.4)$$
First order Taylor series approximations of these singular values are given by

\[
\lambda_1 = \frac{[\text{tr}(K^T K)]^2 - \det(K^T K)}{\text{tr}(K^T K)} + O^2
\] (B.5)

\[
\lambda_2 = \frac{\det(K^T K)}{\text{tr}(K^T K)} + O^2
\] (B.6)

The ratio of these two singular values is therefore given by

\[
\frac{\lambda_1}{\lambda_2} = \frac{[\text{tr}(K^T K)]^2}{\det(K^T K)} - 1
\] (B.7)

Combining (B.4) and (B.7) we conclude that

\[
\frac{\lambda_1}{\lambda_2} \geq 3
\] (B.8)

A question that may be raised with regards to this demonstration is the error involved in the Taylor series truncations and its effect on the results in (B.8). These truncations are basically based on expanding the discriminant of the quadratic characteristic equation of the matrix \(K^T K\). We now show that this truncation is practically un-consequential. Basically, although (B.5) and (B.6) have truncation errors, the error in their ratio is extremely negligible.

Instead of giving a formal proof of this claim we give a graphical proof. Figure B.1(a) shows the possible values of the discriminant of the characteristic equation and its linear approximation using first order Taylor series, where the discriminant of the characteristic equation is plotted against \(\frac{\det(K^T K)}{[\text{tr}(K^T K)]^2}\) (i.e. the independent variable in the Taylor series expansion). Although the errors grow for small values of the discriminant, the error in the ratio of the two singular values is
Figure B.1: (a) Discriminant of the characteristic equation and its linear approximation, (b) ratios of the singular values: dotted line: true ratio, solid line: approximated ratio.

practically negligible as shown in Figure B.1(b).
LIST OF REFERENCES


[CMU] CMU. “Vision and autonomous systems center’s image database, http://vasc.ri.cmu.edu/idb/.”


128


133


mann, and D. J. Schyler. “MR-PET image coregistration for quantitation of striatal 
dopamine D2 receptors. Journal of computer assisted tomography, 20(3):423428.” 
1996.


treatment evaluation. In H. U. Lemke, K. Inamura, C. C. Jaffe, and M. W. Vannier, 


registration for spiral CT angiography. In Computer vision and pattern recognition, 

Robb, editor, Visualization in biomedical computing, volume 1808 of Proc. e SPIE, 

M. H. Loew, editor, Medical imaging: image processing, volume 2167, pages 176187, 


[ZFC02] A. Zalesny, V. Ferrari, G. Caenen, and L. van Gool. “Parallel composite texture syn-
thesis. In Texture 2002 workshop - (in conjunction with ECCV02), Copenhagen, Den-
mark.” Jun 2002.

conference on computer vision, pages 650657, Los Alamitos, CA, IEEE e computer 


registration based on polynomial regression, Proc. Int. Conf. Automation, Robotics & 

[ZSI95] I. G. Zubal, S. S. Spencer, J. S. Khurseed Imam, E. O. Smith, G. Wisniewski, and 
P. B. Hoffer. “Difference images calculated from ictal and interictal technetium99m-
