UNIVERSITY of CALIFORNIA
Santa Barbara

Computational Methods for Automatic Image Registration

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy
in
Electrical and Computer Engineering

by

Marco Zuliani

Committee in charge:

Professor B. S. Manjunath, Chair
Professor S. Chandrasekaran
Professor A. Fusiello
Professor C. S. Kenney
Professor J. P. Hespanha

December 2006

The dissertation of Marco Zuliani is approved.

Professor S. Chandrasekaran

Professor A. Fusiello

Professor C. S. Kenney

Professor J. P. Hespanha

Professor B. S. Manjunath, Committee Chair

October 2006

Computational Methods for Automatic Image Registration

by

Marco Zuliani

To my family,

and to the memory of my grandmother, Anna Pia.

Acknowledgements

Completing my graduate studies has been an extremely enriching and rewarding experience both under a scientific and a human point of view. My doctorate is a team achievement, and in the next paragraphs I want to thank the people that contributed to this accomplishment.

First I want to thank prof. Manjunath for giving me the chance of joining his research group (I told you…I'll be back!), for directing my research leaving me a lot of freedom, for the constant confidence he placed in me and for all his support, at all levels.

I am extremely grateful to my doctoral committee members: to prof. Chandrasekaran for the uncountable discussions I had with him, to prof. Fusiello for sharing with me his expertise and rigor in many different fields of computer vision, to prof. Hespana for his interest in my research, to prof. Kenney for his informal, didactic, provoking, original and enthusiast attitude.

I would like to thank the Office of Naval Research (grant #N00014-04-1-0121) for supporting the work presented in this dissertation.

The suggestions and directions of prof. Rhodes and prof. Rose have been extremely valuable in completing this work. Thanks also to prof. Beghi and prof. Frezza who made it possible for me to start this experience. I am grateful to Dr. Bober for his guidance and support during my staying at the Mitsubishi Electric Visual Information Laboratory.

I have been honored to share the lab with great researchers and wonderful people: their support, acceptance, help and friendship have been a fundamental part of this experience. Anyndia, Baris, Dmitry, Emily, Ibrahim, Jelena,

Jiyun, Kaushal, Laura, Luca, Mike, Ken, Nhat, Pratim, Shawn, Sitaram, Thomas, Zhiqiang, Xinding, thank you all and to everybody else who has been a part of our research group! I also want to thank Guylene, John, Ken, Richard, Val who made my life as a grad student much easier and smooth.

During these years I shared countless wonderful moments and enriching experiences outside the lab with people that eventually became my "extended family": Marcelo (my agelong apt-mate who introduced me to cachaça) & Emily, the "sacred pint" man Gabriel, Rogerio, all the other members and co-funders of the V.P. society, Paolo M., Marco R., Ramesh, Vittorio, the family guys Jessica & Fernando, Francine & Hugo, Mylene & Marcelo, Luchino, Ibra, Dima, Max, Antoine, Sara S., N{a,e}da, Sandra, Jannelle, Nat, Sarah, Rimma, Elison, Desiree, Natalie, Daniel. My sincere gratitude goes to Fr. George, Fr. Joe and Fr. Paul for their friendship, guidance and support. Thanks also to all the international (actually mostly Italian. . .) visiting students or researchers that I met in the past few years: Ruggio, Stefano C. and Ale (V.P. Members), Antonio, Enrico, Marietto, Marina S., Marco A., Raffi, Corrado, Anna, Paola, Blandina, Gaia. All my friends from the glorious days in Padova also deserve to be acknowledged here: Cesco *Da Fogo*, Dry & Titti, Marco M., the Curto, Siro, Soa & Soetto, Luca & Silvana, Fabio, Ennio, Padu, Emilio, Luca M., Marina, *mami* Balla, *papi* Baretz, Poje, Angela, Ale & Stefano, Matteo, Lupo, Sara M., Eva, Regina, Mandrea, Emirasta, Lorenzo, Ruben, Paolo B., Davide *Reds*, Davide B., Sergio, Stefano A.. You guys paved the way for this achievement. Thanks also to Carlos, Giovanni and Raquel who made my staying in UK more pleasant and to David for his friendship throughout the years, since first grade.

I am forever indebted to my brother Francesco for his continuous support and encouragement (you are always able to make me smile), to my parents Luciana & Pierino for their teachings, guidance, patience and support to ensure I could have the best possible education. Thanks to my grandmothers Anna Pia & Nilde for being always present in my life and to my godfather, my godmother and all my close relatives for their caring support.

A special thanks to Elisa for her courage, her strength, her faith, her patience, her smile and her love. Bright, unique and special gifts you shared with me: *grazie cuore mio.*

Finally thank You, for Your gifts, for Your mysterious ways, for Your love.

Curriculum Vitæ
Marco Zuliani

July 2001	Laurea in Ingegneria Informatica Department of Information Engineering Università degli studi di Padova, Padova, Italy
July 2003	Master of Science Department of Electrical and Computer Engineering University of California, Santa Barbara
October 2006	Doctor of Philosophy Department of Electrical and Computer Engineering University of California, Santa Barbara

Fields of Study

Image analysis and pattern recognition.

Experience

2002-2006	Research Assistant
2005	Internship Mitsubishi Electric, Guildford, UK
2001-2006	Teaching assistant University of California, Santa Barbara
2002	Summer Internship FriulROBOT S.r.l, Udine, Italy

Publications

M. Zuliani, C. Kenney, and B. Manjunath, "Condition Theory for Point Neighborhood Characteristic Structure Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, In revision.

M. Zuliani, L. Bertelli, C. Kenney, S. Chandrasekaran and B. Manjunath, "Drums, Curve Descriptors and Affine Invariant Region Matching," *Image and Vision Computing*, Accepted for publication.

M. Zuliani, C. Kenney, and B. Manjunath, “The Multi-RANSAC algorithm and its application to detect planar homographies,” In *IEEE International Conference on Image Processing*, Genova, Italy, September 2005.

C. Kenney, M. Zuliani, and B. Manjunath, “An axiomatic approach to corner detection,” In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 191–197, San Diego, California, June 2005.

M. Zuliani, S. Bhagavathy, C. Kenney, and B. Manjunath, “Affine-invariant curve matching,” In *IEEE International Conference on Image Processing*, October 2004.

M. Zuliani, C. Kenney, S. Bhagavathy, and B. Manjunath, “Drums and curve descriptors,” In *British Machine Vision Conference*, Kingston-upon-Thames, UK, September 2004.

M. Zuliani, C. Kenney, and B. Manjunath. “A mathematical comparison of point detectors,” In *Proc. of the 2nd IEEE Workshop on Image and Video Registration*, Washington DC, June 2004.

C. Kenney, B. Manjunath, M. Zuliani, G. Hewer, and A. Van Nevel, “A condition number for point matching with application to registration and post-registration error estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1437–1454, November 2003.

Abstract

Computational Methods for Automatic Image Registration

by

Marco Zuliani

Image registration is the process of establishing correspondences between two or more images taken at different times, from different viewpoints, under different lighting conditions, and/or by different sensors, and aligning them with respect to a coordinate system that is coherent with the three dimensional structure of the scene. Once feature correspondences have been established and the geometric alignment has been performed, the images are combined to provide a representation of the scene that is both geometrically and photometrically consistent. This last process is known as image mosaicking.

The primary contribution of this research is the development of computational frameworks that tackle in a general and principled way the problems arising in the construction of an image registration and mosaicking system. Specifically, we present a general theory to detect image point features that are suitable for matching. Our theory generalizes and extends much of the previous work on detecting feature locations. We introduce a novel, physically motivated curve/region descriptor suitable to establish image correspondences in a geometrically invariant fashion. New methods to estimate robustly the image transformation parameters in presence of large quantities of outliers and of multiple models are also presented. Finally we present a fully automated registration and mosaicking system that can produce seamless mosaics from image pairs. Extensive experimental results with biological images, satellite images and consumer photographs are presented.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"Caminante, no hay camino,
se hace camino al andar.[1]*"*

A. Machado

Image registration is the process of aligning two or more images taken at different times, from different viewpoints, and/or by different sensors with respect to a coordinate system that is coherent with the three dimensional structure of the scene. Once feature correspondences have been established and the geometric alignment has been performed, the images are combined to provide a representation of the scene that is both geometrically and photometrically consistent. This process is known as image mosaicking.

For a long time, image registration and mosaicking have been two leading research themes in the image analysis community (as confirmed by three major

[1] Traveller, there is no road, you make your path as you walk.

from Proverbios y cantares XXIX

surveys [12, 137, 122] appearing in the span of 12 years). All the innovative and significant contributions to the registration problem have found immediate application in many disparate areas such as remotely sensed image processing, medical image analysis, scene reconstruction, surveillance, automatic navigation and augmented reality.

One of the reasons that image registration is an extremely challenging problem is the large degree of variability of the input data. The images that are to be registered and mosaicked may contain visual information belonging to very different domains and can undergo many geometric and photometric distortions such as scaling, rotations, projective transformations, non rigid perturbations of the scene structure, temporal variations, and photometric changes due to different acquisition modalities and lighting conditions. Figure 1.1 shows some examples of image pairs belonging to different domains that have been registered using the algorithms that will be described and analyzed in the next chapters.

Despite the large number of efforts made to construct efficient algorithms to solve different aspects of the image registration and mosaicking problem, there still exist a number of obstacles that need to be overcome and several open questions that need to be answered. In the next section we will discuss the motivations that lead us to tackle some of these obstacles and to answer some of these questions.

1.1 Motivation

An image registration system must be able to provide accurate and realistic results, to self assess the quality of its output and, at the same time, it should require minimal human intervention and reduced computational resources. It appears evident that the design of such a system requires the synergistic integration of the expertise coming from different fields such as: early vision, pattern recognition, robust statistics, 3D geometry, computer graphics and numerical analysis just to name a few. An immediate consequence of this observation is that the overall system will be composed of several modules that must interact robustly in a hierarchical fashion, where each unit is able to cope with the possibly noisy/inaccurate results produced in the earlier processing stages and to provide feedback to improve the quality of the final result.

The fundamental modules that compose the registration pipeline that we consider in this dissertation are shown in Figure 1.2. According to the taxonomy introduced in [137], we will focus our attention on *feature-based* approaches. The overall system first extracts a set of features from the images that are to be registered. Then, distinctive labels are associated with each feature to establish tentative image correspondences. These matches are further refined by pruning those correspondences that are incompatible with the underlying geometric model used to describe the transformation between the images. Finally the parameters of the models are estimated and the images are fused together to produce a coherent mosaic.

This thesis is motivated by the desire to study each of these modules in a rigorous and principled manner. In the following chapters we develop a framework to quantitatively analyze the problems to be solved and we design practical algorithms that are general enough to be applicable in a large variety of image registration scenarios. More specifically, for each module composing the registration

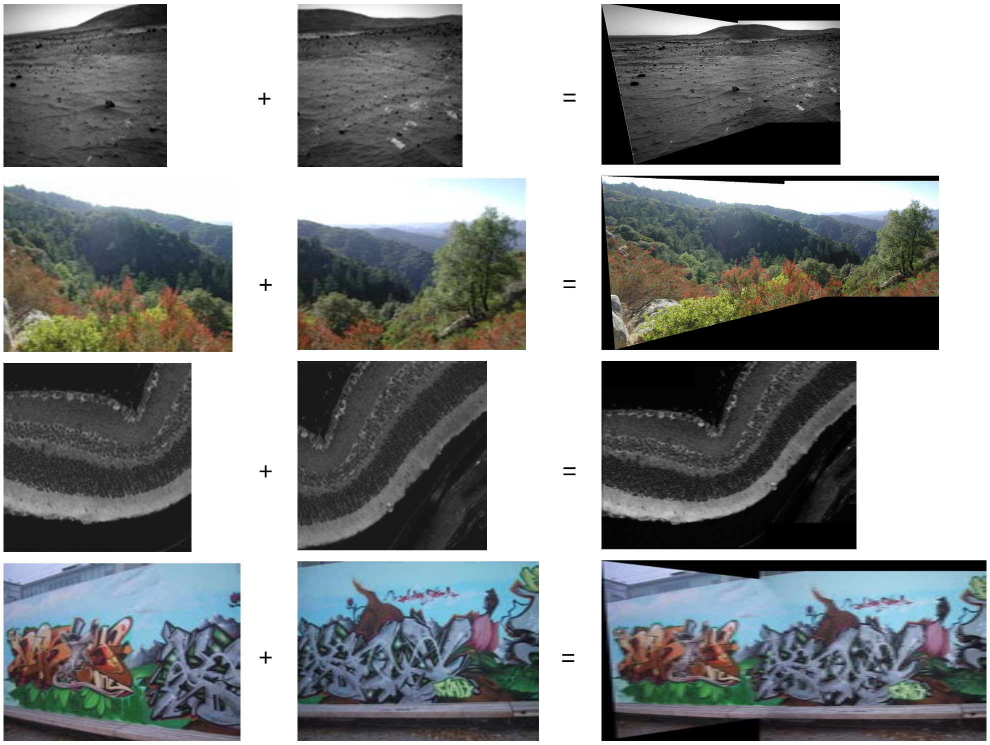


Figure 1.1: Some examples of image pairs that have been registered and mosaicked using the methods that will be described in the following chapters. First row: a pair of EDR (extreme dynamic range) images acquired by the right navigation camera of the Spirit rover during its mission to Gusev crater on Mars (courtesy of NASA). Second row: an image pair of a complex 3D outdoor scene taken with a consumer camera. Third row: a pair of retinal images acquired using a confocal microscope (courtesy of Dr. S. K. Fisher, Dr. G. Lewis and Dr. M. Verardo). Forth row: two images of a graffiti scene subject to a strong perspective distortion taken using a consumer camera.

system we will:

- state formally the generalized instances of the problem that is to be solved,
- establish connections with some algorithms already used by the image analysis community,
- develop models that limit the need to resort to empirical considerations to justify the design choices for the proposed algorithms,
- evaluate the impact of the approximations introduced to simplify both the theoretical analysis and the practical implementation of the algorithms, and
- quantify the strengths and limitations of the proposed algorithms and evaluate the accuracy and the quality of the results.

These modules are then implemented and combined to produce a registration system that is able to render photorealistic mosaics consistent with the 3D structure of the scene.

1.2 Thesis Organization and Contributions

We will now outline the structure of this dissertation and briefly summarize the contributions of each chapter.

1.2.1 Chapter 2: Point Feature Detectors: Theory

This chapter contains a thorough theoretical analysis of point feature detectors based on the Generalized Gradient Matrix (GGM) (also known as autocorrelation

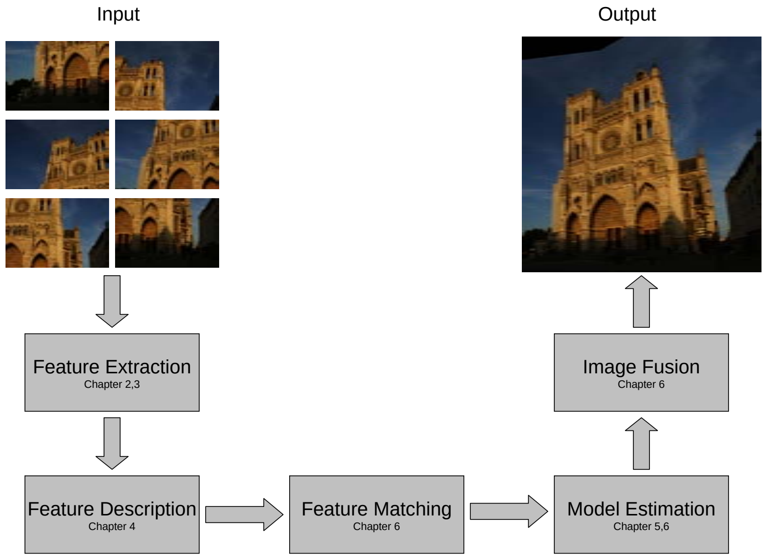


Figure 1.2: Overview of the registration system modules that have been studied in this thesis. The final mosaic of the images of the Cathedral of Our Lady of Amiens is obtained using the methods described in this disseration (image courtesy of J. Nieuwenhuijse, copyright by New House Internet Services BV, www.ptgui.com).

matrix or structure tensor). In this chapter:

- We introduce a *novel framework based on condition theory* that motivates the use of the autocorrelation matrix as a fundamental ingredient for point detection.
- We introduce a set of *generalized point detector functions* based on the spectral properties of the image GGM. Such detectors are defined for multichannel images with spatial dimension that can be greater than 2. For single channel images these generalized functions become equivalent to some of the commonly used point detectors.
- We establish in-depth connections among the detectors showing that certain commonly used detectors *are equivalent* modulo the choice of a specific matrix norm.
- We list a *set of analytical properties* of the generalized detectors that define bounds to their performance and suggest effective ways to reduce their computational complexity.

1.2.2 Chapter 3: Point Feature Detectors: Experiments

This chapter contains an exhaustive experimental evaluation of the point detectors studied in Chapter 2. More specifically:

- We experimentally validate the theoretical claims made in Chapter 1 regarding *detector equivalences*.

- We characterize the repeatability of the point detectors and find that they exhibit a behavior that is *almost linear* for a relevant set of scalings and projective distortions that are found in real life scenarios.
- Quite surprisingly we find that for natural images it is possible to *disregard the color information* and at the same time improve the detector performance.

1.2.3 Chapter 4: Drums, Curve Descriptors and Affine Invariant Region Matching

Motivated by the possibility of establishing image correspondences using curve features rather than interest points, in this chapter we introduce a novel curve/region descriptor based on the modes of vibration of an elastic membrane. In particular:

- We introduce and study the theoretical properties of a novel *physically motivated curve/region descriptor* based on the modes of vibration of a membrane. We revisit the problem of *curve isospectrality* within the image analysis domain.
- We develop a *normalization procedure* that allows us to characterize the shape of a curve independent of its affine distortions.
- We propose a method to couple the descriptor and the normalization procedure to robustly *match curves between images* taken from different points of view.

- We provide extensive experimental results to measure the performance of our descriptor using both synthetic and real images. We also compare our descriptor with state of the art curve/region descriptors.

1.2.4 Chapter 5: RANSAC Stabilization

Given the need to estimate the parameters of (multiple) geometric or photometric models in the presence of a large number of outliers, we develop a robustification framework that improves the results obtained using RANSAC. The novel contributions of this chapter are:

- The introduction of a *stabilization framework* that improves the quality of estimates obtained using RANSAC in the presence of large uncertainties of the noise scale and multiple instances of the model.
- The introduction of a *pseudo-distance* to quantify the dissimilarity between geometric transformations.
- The reduction of the problem of *grouping similar models* to the problem of identifying the largest maximal clique in a graph.
- The validation of the stabilization framework by means of extensive experiments using both synthetic and real data.

1.2.5 Chapter 6: Applications

This chapter contains an overview of the algorithms developed in the previous chapters integrated into a registration and mosaicking system. Using the framework developed in Chapter 2, we introduce the concept of *characteristic structure* of a point neighborhood and show how it can be used to improve the detection of matching points between image pairs related by large scale variations. We then devote our attention to the development of a set of techniques to obtain a seamless mosaic of the registered images. The contributions contained in this chapter can be summarized as follows:

- We apply the framework based on condition theory to identify the *characteristic structure* of a point neighborhood and show how this can be used to establish matches between images related by large scale variations.
- We explore the possibility of using *indexing* and *dimensionality reduction techniques* to speed the computation of tentative image correspondences.
- We introduce a *novel robust equalization procedure* to correct the photometric appearance of two images that are to be fused together.
- We present a *physically motivated* algorithm to calculate the best stitching line between registered images.

1.2.6 Summary

This thesis makes several new contributions to the classical problems of establishing correspondences between images, of robustly registering them and of producing geometrically and photometrically consistent mosaics. Practical, efficient and robust implementations of these methods have been developed and tested on large collections of images belonging to several different domains.

Chapter 2

Point Feature Detectors: Theory

> *"Basic research is what I'm doing*
> *when I don't know what I'm doing."*
>
> Attributed to W. von Braun

This chapter contains a thorough theoretical analysis of point feature detectors based on the Generalized Gradient Matrix (GGM) (also known as autocorrelation matrix or structure tensor). In this chapter:

- We introduce a *novel framework based on condition theory* that motivates the use of the autocorrelation matrix as a fundamental ingredient for point detection (Sections 2.3 and 2.4).

- We introduce a set of *generalized point detector functions* based on the spectral properties of the autocorrelation matrix. Such detectors are defined for multichannel images with spatial dimension that can be greater than 2. For single channel images these generalized functions become equivalent to some of the commonly used point detectors (see Section 2.5 and 2.6).

- We establish in-depth connections among the detectors showing that certain commonly used detectors *are equivalent* modulo the choice of a specific matrix norm (see Section 2.5).

- We list a *set of analytical properties* of the generalized detectors that define bounds to their performance and suggest effective ways to reduce their computational complexity (see Section 2.5).

2.1 Introduction

Corner detection in images is important for a variety of image processing tasks including tracking, image registration, change detection, determination of camera pose and position and a host of other applications. In the following, the term "corner" is used in a generic sense to indicate any local image feature that is useful for the purpose of establishing point correspondence between images.

Detecting corners has long been an area of interest to researchers in image processing. Some of the most widely used corner detection approaches (Harris-Stephens [50], Noble-Förstner [98, 38], Shi-Tomasi [116], Rohr [107]) rely on the properties of the averaged outer product of the image gradients:

$$L(\boldsymbol{x}, \sigma_D, I) = \left(G_{\sigma_D} * I\right)(\boldsymbol{x}) \tag{2.1a}$$

$$\mu(\boldsymbol{x}, \sigma_I, \sigma_D, I) = \left(w_{\sigma_I} * \nabla_{\boldsymbol{x}} L(\cdot, \sigma_D, I) \nabla_{\boldsymbol{x}}^T L(\cdot, \sigma_D, I)\right)(\boldsymbol{x}) \tag{2.1b}$$

In the previous equations $L(\boldsymbol{x}, \sigma_D, I)$ indicates the smoothed version of the single channel image I at the scale σ_D, whereas $\mu(\boldsymbol{x}, \sigma_I, \sigma_D, I)$ is a 2×2 symmetric and positive semi-definite matrix representing the averaged outer product of the image gradients (also known within the computer vision and image processing community as auto-correlation matrix, gradient normal matrix or structure tensor). The function w_{σ_I} weights properly the pixels about the point $\boldsymbol{x}$ at the scale σ_I. Note how the notion of scale is related to the shape of the Gaussian differentiation kernel G_{σ_D} (the smaller is σ_D the larger is the sensitivity to fine image details) and to the structure of the integration kernel (in general, the larger is the parameter σ_I, the larger is the averaging effect on the neighborhood about the point $\boldsymbol{x}$).

Förstner [38], in 1986 introduced a rotation invariant corner detector based on the ratio between the determinant and the trace of μ; in 1989, Noble [98] considered a similar measure in her PhD thesis. Rohr in 1987 [107] proposed a rotation invariant corner detector based solely on the determinant of μ. Combinations of first order image derivatives have also been used by Rohr et al. to locate point landmarks in 3D tomographic images [42, 108]. Harris and Stephens in 1988 [50] introduced a function designed to detect both corners and edges based on a linear combination of the determinant and the squared trace of μ, revisiting the work of Moravec [92] that dates back to 1980. This was followed by the corner detector proposed by Tomasi and Kanade in 1992 [124], and refined in 1994 in the well-known feature measure of Shi and Tomasi [116], based on the smallest eigenvalue of μ. All these measures create a value at each point in the image with larger values indicating points that are better for establishing point correspondences between images (i.e., better corners). Corners are then identified either as local maxima for the detector values or as points with detector values above a given threshold. All of these detectors have been used rather successfully to find corners in images but have the drawback that they are sometimes based on heuristic considerations.

Recently Kenney et al. in 2003 [63] avoided the use of heuristics by basing corner detection on the conditioning of points with respect to window matching under various transforms such as translation, Rotation Scaling and Translation (RST), and affine pixel maps. Along similar lines Triggs [129] proposed a generalized form of the multi-scale Förstner detector that selects points that are maximally stable with respect to a certain set of geometric and photometric transformations.

Methods to detect interest points in a scale invariant fashion have been developed by Lindeberg [71] using the tools made available by scale space theory [36, 70]. More recently Baumberg [5], Mikolajczyk [86] and Lowe [74] developed point detectors that are robust[1] with respect to affine transformations of the image. We want to emphasize how the approaches proposed by Baumberg and Mikolajczyk both depend on an initial step where candidate points are detected at different scales using the Harris detector. Therefore, rather than being truly affine invariant, such detectors are robust in the presence of affine transformations of the image; the degree of robustness is directly connected to the repeatability of the detector used to identify the candidate points. Similar considerations hold for Lowe's algorithm, that seeks for point candidates in correspondence of the local extrema of the scale space signature generated by the difference of Gaussians. Since images that are related via an affine transformation will not necessarily originate extrema at corresponding positions, the overall detector is robust but not invariant. In all the robust methods mentioned above, the auto-correlation matrix plays once again a fundamental role.

[1] In this context, the robustness of a detector refers to its capability of identifying corresponding points in images that are related by a certain geometric transformation. This property has been formalized quantitatively by Schmid et al. introducing the concept of ε-repeatability [113].

This chapter presents a theoretical analysis of corner detectors based on the image auto-correlation matrix. In this chapter we will reorganize and extend the ideas that were initially presented in the papers [63, 140, 64]. More specifically the contributions of this chapter can be summarized as follows:

- We will provide a justification for the central role that the gradient normal matrix plays in corner detection. We will motivate its importance using two different perspectives: the estimation of the optical flow and the characterization of the sensitivity of a point neighborhood with respect to noise perturbations. The novel mathematical tool that will be used is condition theory.

- We will provide generalized expressions for the some of the commonly used corner detectors, establish a relation between them and analyze and compare their relevant properties.

This chapter is structured as follows (see also Figure 2.1). We first introduce the auto-correlation matrix using two different perspectives, the first based on the computation of the optical flow (Section 2.3) and the second based on the characterization of the sensitivity of a point neighborhood with respect to noise perturbations (Section 2.4). In Section 2.5 we will introduce a set of generalized corner detector functions, establish relations between them and extensively discuss their theoretical properties. In Section 2.6 we will also show that some of the commonly used corner detector functions based on the auto-correlation matrix are just special instances of a specific generalized detector. Finally the conclusions and the discussion of some future research directions can be found in Section 2.7.

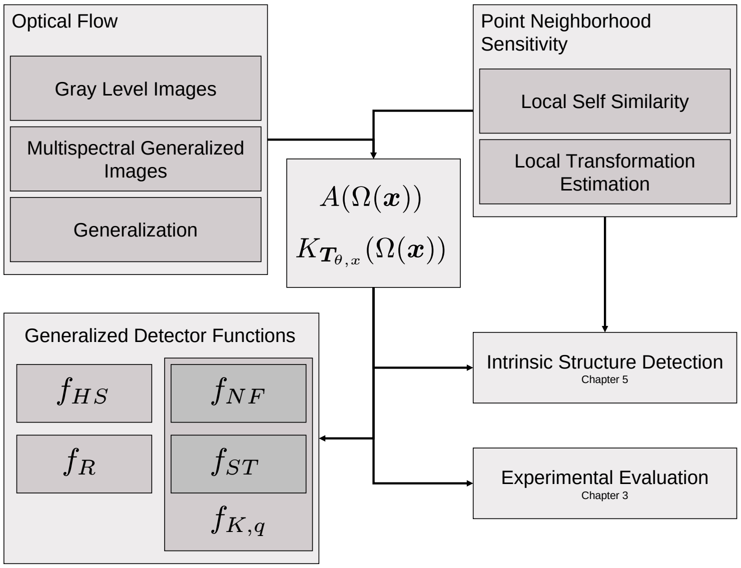

Optical Flow
Gray Level Images
Multispectral Generalized Images
Generalization
Point Neighborhood Sensitivity
Local Self Similarity
Local Transformation Estimation
$A(\Omega(\boldsymbol{x}))$
$K_{\boldsymbol{T}_{\theta,x}}(\Omega(\boldsymbol{x}))$
Generalized Detector Functions
f_{HS}
f_{NF}
f_R
f_{ST}
$f_{K,q}$
Intrinsic Structure Detection
Chapter 5
Experimental Evaluation
Chapter 3

Figure 2.1: Overview of the framework used to study the generalized corner detector functions.

This theoretical analysis will be supplemented in Chapter 2 by a set of experiments that will test the performance of the detectors with real imagery. In the next chapter we will also outline the connections between the experimental results and the theoretical properties of the detectors. Moreover in Chapter 6 we will introduce the notion of intrinsic neighborhood of an image point and describe an algorithm for the detection of such neighborhood using the tools made available by condition theory.

2.2 Preliminaries

First of all we will introduce a few notation conventions. Throughout the chapter boldface letters will indicate vectors. The image pixel dimension is indicated with the letter n. When $n = 2$ we are considering usual 2D images, but all the theoretical results will hold in cases where $n > 2$, for example in computed axial tomography (CAT) images, where the intensity signal is defined on a 3D lattice (in this case $n = 3$). We will refer to images with $n > 2$ as generalized images. The image intensity dimension is instead indicated by the letter m: $m = 1$ models a single channel image (such as graylevel image), $m = 3$ can model an RGB image and other values of m may be used to model arbitrary multichannel images.

2.2.1 The Gradient Matrix

We begin this section by introducing the gradient matrix in the special case of a 2D single channel image. This quantity will be generalized in the next sections. Let $I(\boldsymbol{x})$ be the intensity of a single channel image at the image point $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$. Let Ω be a window about the point of interest $\boldsymbol{x}$: the gradient matrix A over this window is defined as:

$$A(\Omega(\boldsymbol{x})) \overset{\text{def}}{=} \begin{bmatrix} I_{x_1}(\boldsymbol{y}_1) & I_{x_2}(\boldsymbol{y}_1) \\ \vdots & \vdots \\ I_{x_1}(\boldsymbol{y}_N) & I_{x_2}(\boldsymbol{y}_N) \end{bmatrix} \tag{2.2}$$

where subscripts indicate differentiation with respect to x_1 and x_2 and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \Omega(\boldsymbol{x})$. To simplify the notation we will omit the dependence of A on $\Omega(\boldsymbol{x})$ when this does not generate confusion.

The 2×2 *gradient normal matrix*[2] is given by:

$$A^T A \stackrel{\text{def}}{=} \begin{bmatrix} \sum_{i=1}^{N} I_{x_1}(\boldsymbol{y}_i)^2 & \sum_{i=1}^{N} I_{x_1}(\boldsymbol{y}_i) I_{x_2}(\boldsymbol{y}_i) \\ \sum_{i=1}^{N} I_{x_1}(\boldsymbol{y}_i) I_{x_2}(\boldsymbol{y}_i) & \sum_{i=1}^{N} I_{x_2}(\boldsymbol{y}_i)^2 \end{bmatrix}$$

where the summation is over the window Ω about the point of interest. As mentioned in the introduction, the gradient normal matrix $A^T A$ is the basis of many corner detectors that have been used by the computer vision and image processing community (Harris-Stephens [50], Noble-Förstner [98, 38], Shi-Tomasi [116], Rohr [107]). Note that this matrix can be obtained discretizing the auto-correlation matrix (2.1b) under the assumption that the weight has the form:

$$w_{\sigma_I}(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \boldsymbol{y} \in \Omega(\boldsymbol{x}), \\ 0 & \text{otherwise.} \end{cases}$$

Why should a corner detector just depend on A (or, equivalently, on $A^T A$)? Can we generalize the expression of $A^T A$ for multidimensional and multichannel images? What are the properties of this matrix? What is the relation among the corner detectors based on the gradient normal matrix? We will try to answer these questions by looking at the problem of estimating the optical flow and the sensitivity of a point neighborhood using the tools made available by condition theory, which are briefly introduced in the next section.

2.2.2 Condition Theory: A Brief Introduction

As early as 1987, with the work of Kearney et al. [62] it was realized that the normal matrix associated with locally constant optical flow is critical in de-

[2] A real square matrix M is normal if $MM^T - M^T M = 0$. It can be immediately verified that $M = A^T A$ is normal.

termining the accuracy of the computed flow. Kearney et al. also reported that ill-conditioning in the matrix $A^T A$ and large residual error in solving the equations for optical flow can result in inaccurate flow estimates. This was supported by the work of Barron et al. [3] who looked at the performance of different optical flow methods; see also [6]. More recently, Shi and Tomasi [116] presented a technique for measuring the quality of local windows for the purpose of determining image transform parameters (translational or affine). For local translation they argued that to overcome errors introduced by noise and ill-conditioning, the smallest eigenvalue of the normal matrix $A^T A$ must be above a certain threshold: $T_\lambda \leq \min(\lambda_1, \lambda_2)$ where T_λ is the prescribed threshold and λ_1, λ_2 are the eigenvalues of $A^T A$. When this condition is met the point of interest has good features for tracking.

The current viewpoint on condition estimation can trace its roots to the era of the 1950's, with the development of the computer and the attendant ability to solve large linear systems of equations and eigenproblems. The question facing investigators at that time was whether such problems could be solved reliably.

The solution of a system of equations can be viewed as a mapping from the input data $\boldsymbol{D} \in \mathbb{R}^n$ to the solution or output $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{D}) \in \mathbb{R}^m$. If a small change in $\boldsymbol{D}$ produces a large change in $\boldsymbol{D}(\boldsymbol{X})$ then $\boldsymbol{X}$ is ill-conditioned at $\boldsymbol{D}$. Following Rice [105], we define the δ-condition number of $\boldsymbol{X}$ at $\boldsymbol{D}$ by:

$$K_\delta = K_\delta(\boldsymbol{X}, \boldsymbol{D}) \equiv \sup_{\|\Delta \boldsymbol{D}\| \leq \delta} \frac{\|\boldsymbol{X}(\boldsymbol{D} + \Delta \boldsymbol{D}) - \boldsymbol{X}(\boldsymbol{D})\|}{\|\Delta \boldsymbol{D}\|}$$

where $\| \cdot \|$ denotes the vector 2-norm: $\|\boldsymbol{D}\|^2 = \sum_i |D_i|^2$. For any perturbation

$\boldsymbol{D}$ with $\|\Delta \boldsymbol{D}\| \leq \delta$, the perturbation in the solution satisfies:

$$\|\boldsymbol{X}(\boldsymbol{D} + \Delta \boldsymbol{D}) - \boldsymbol{X}(\boldsymbol{D})\| \leq \delta K_\delta$$

The δ-condition number inherits any nonlinearity in the function $\boldsymbol{X}$ and consequently is usually impossible to compute. For this reason the standard procedure is to take the limit as $\delta \to 0$. If $\boldsymbol{X}$ is differentiable at $\boldsymbol{D}$ we can define the local or differential condition number:

$$K = K(\boldsymbol{X}, \boldsymbol{D}) \equiv \lim_{\delta \to 0} K_\delta(\boldsymbol{X}, \boldsymbol{D})$$

Using a first order Taylor expansion, we have:

$$\boldsymbol{X}(\boldsymbol{D} + \Delta \boldsymbol{D}) = \boldsymbol{X}(\boldsymbol{D}) + \boldsymbol{X}_{\boldsymbol{D}}\, \Delta \boldsymbol{D} + O(\|\Delta\|^2)$$

where $\boldsymbol{X}_{\boldsymbol{D}}$ is the $m \times n$ gradient matrix with entries:

$$(\boldsymbol{X}_{\boldsymbol{D}})_{ij} = \frac{\partial \boldsymbol{X}_i}{\partial \boldsymbol{D}_j}$$

This expansion shows that the local condition number is just the norm of the matrix $\boldsymbol{X}_{\boldsymbol{D}}$:

$$K(\boldsymbol{X}, \boldsymbol{D}) = \|\boldsymbol{X}_{\boldsymbol{D}}\|$$

and:

$$\|\boldsymbol{X}(\boldsymbol{D} + \Delta \boldsymbol{D}) - \boldsymbol{X}(\boldsymbol{D})\| \leq K\|\Delta \boldsymbol{D}\| + O(\|\Delta \boldsymbol{D}\|^2)$$

Large values for $K(\boldsymbol{X}, \boldsymbol{D})$ indicate that $\boldsymbol{X}$ is *ill conditioned* in $\boldsymbol{D}$.

2.3 The Generalized Gradient Matrix: an Optical Flow Perspective

2.3.1 Optical Flow for Single Channel Images

Let $I = I(\cdot, t)$ be a single channel image sequence and suppose that a point of interest has time dependent coordinates $\boldsymbol{x} = \boldsymbol{x}(t)$. The optical flow problem is to discover the time evolution of $\boldsymbol{x}$. In the standard approach this is done by making the assumption of constant brightness:

$$I(\boldsymbol{x}(t), t) = I(\boldsymbol{x}(t) + d\boldsymbol{x}, t + dt) = c$$

where c is a constant with respect to t. If we expand this constraint about the point $\begin{bmatrix} x_1(t) & x_2(t) & t \end{bmatrix}^T$ and neglect higher order terms we obtain:

$$I_{x_1}(\boldsymbol{x}, t)\ dx_1 + I_{x_2}(\boldsymbol{x}, t)\ dx_2 + I_t(\boldsymbol{x}, t)\ dt = 0$$

where, as usual, subscripts denote differentiation.[3] The previous equation can be rewritten in matrix form as:

$$\begin{bmatrix} I_{x_1}(\boldsymbol{x}, t) & I_{x_2}(\boldsymbol{x}, t) \end{bmatrix} d\boldsymbol{x} = -I_t(\boldsymbol{x}, t)\ dt$$

where $I_t(\boldsymbol{x}, t)$ is the infinitesimal difference of successive frames and $d\boldsymbol{x} = \begin{bmatrix} dx_1 & dx_2 \end{bmatrix}^T$ is referred to as the optical flow vector. This is one equation for the two unknowns dx_1 and dx_2. To overcome this difficulty the standard approach is to assume that

[3] We will maintain the sign of equality even after neglecting the higher order terms of the Taylor expansions. However we should keep in mind that we are dealing with approximate relations.

dx_1 and dx_2 are constant in a region Ω about $\boldsymbol{x}$. This leads to the overdetermined set of equations:

$$\begin{bmatrix} I_{x_1}(\boldsymbol{y}_1,t) & I_{x_2}(\boldsymbol{y}_1,t) \\ \vdots & \vdots \\ I_{x_1}(\boldsymbol{y}_N,t) & I_{x_2}(\boldsymbol{y}_N,t) \end{bmatrix} d\boldsymbol{x} = - \begin{bmatrix} I_t(\boldsymbol{y}_1,t) \\ \vdots \\ I_t(\boldsymbol{y}_N,t) \end{bmatrix}$$

where we adopted a time scale in which $dt = 1$. More compactly we may write this as:

$$A(\Omega(\boldsymbol{x}))d\boldsymbol{x} = \boldsymbol{\eta}$$

where $\boldsymbol{\eta} = - \begin{bmatrix} I_t(\boldsymbol{y}_1,t) & \dots & I_t(\boldsymbol{y}_N,t) \end{bmatrix}^T$. The least squares solution to this set of equations is obtained by multiplying both sides by A^T to obtain a square system and then multiplying both members by $(A^TA)^{-1}$ to get:

$$d\boldsymbol{x}_{computed} = (A^TA)^{-1}A^T\boldsymbol{\eta} = A^{\dagger}\boldsymbol{\eta}$$

where $A^{\dagger}$ is also known as the pseudo-inverse of A. A major problem with this approach is that some points give better estimates of the true optical flow than others. For example, if the image intensities in the region about $\boldsymbol{x}$ are nearly constant (uniform illumination of a flat patch) then $A \approx 0$ and the least squares procedure gives bad results.

2.3.2 A Thought Experiment

We can assess which points are likely to give bad optical flow estimates by a simple ansatz: suppose that the scene is static so that the true optical flow is zero: $d\boldsymbol{x}_{exact} = 0$. If the images of the scene vary only by additive noise, then $\boldsymbol{\eta}$ (the difference between frames) represents the noise itself. The error in the optical flow estimate is given by $\boldsymbol{e} \overset{\text{def}}{=} d\boldsymbol{x}_{exact} - d\boldsymbol{x}_{computed}$, and we may write:

$$\|\boldsymbol{e}\| = \|d\boldsymbol{x}_{exact} - d\boldsymbol{x}_{computed}\| = \|0 - A^{\dagger}\boldsymbol{\eta}\| = \|A^{\dagger}\boldsymbol{\eta}\| \leq \|A^{\dagger}\| \; \|\boldsymbol{\eta}\|$$

Thus we see that the term $\|A^{\dagger}\|$ controls the error multiplication factor; that is the factor by which the input error (the noise $\boldsymbol{\eta}$) is multiplied to get the output error (the error in the optical flow estimate). Large values of $\|A^{\dagger}\|$ correspond to points in the image where we cannot estimate the optical flow accurately in the presence of noise at least for the static image case.

If we use the 2-norm together with Lemma A.2.2, then we have:

$$\|A^{\dagger}\|_2^2 = \frac{1}{\lambda_{min}(A^T A)}$$

(where $\lambda_{min}(A^T A)$ indicates the smallest eigenvalue of $A^T A$). We conclude that the error multiplication factor for the 2-norm in the optical estimate for the static noise case is equal to $\frac{1}{\sqrt{\lambda_{min}(A^T A)}}$. This motivates the use of the gradient normal matrix in point feature detection, since the ability to accurately determine optical flow at a point is intimately related to its suitability for establishing point correspondence between images (i.e. whether it is a good corner).

2.3.3 Optical Flow for Multichannel Generalized Images

The need to locate good points for tracking occurs in other settings besides images with pixel dimension two and intensity dimension one. For example we may want to consider good matching points in signals (pixel dimension is one) or tomographic medical images (pixel dimension is three) or color images (intensity dimension is three) or hyperspectral images (intensity dimension much greater than one). In order to set up a framework for discussing corner detection for images with arbitrary pixel and intensity dimensions let $\boldsymbol{x} \overset{\text{def}}{=} \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ denote the pixel coordinates and $\boldsymbol{I} \overset{\text{def}}{=} \begin{bmatrix} I_1 & \dots & I_m \end{bmatrix}^T$ the intensity vector for the image. We use the optical flow method described before to set up a corner detection paradigm. Let $\boldsymbol{x} = \boldsymbol{x}(t)$ be a point of interest in a time dependent image $\boldsymbol{I} = \boldsymbol{I}(\cdot, t)$. We assume that this point has constant brightness over time:

$$\boldsymbol{I}(\boldsymbol{x}(t), t) = \boldsymbol{I}(\boldsymbol{x}(t) + d\boldsymbol{x}, t + dt) = c \tag{2.3}$$

Expanding this constraint about the point $\begin{bmatrix} \boldsymbol{x}(t)^T & t \end{bmatrix}^T$ and neglecting higher order terms we obtain:

$$J\boldsymbol{I}(\boldsymbol{x}, t)\ d\boldsymbol{x} = -\boldsymbol{I}_t(\boldsymbol{x}, t) \tag{2.4}$$

where we once again assumed that $dt = 1$ and the Jacobian matrix $J\boldsymbol{I}(\boldsymbol{x}, t) \in \mathbb{R}^{m \times n}$ has entries $[J\boldsymbol{I}(\boldsymbol{x}, t)]_{i,j} = \partial I_i(\boldsymbol{x}, t)/\partial x_j$, and:

$$d\boldsymbol{x} = \begin{bmatrix} dx_1 & \dots & dx_n \end{bmatrix}^T \qquad \boldsymbol{I}_t = \begin{bmatrix} dI_1/dt & \dots & dI_m/dt \end{bmatrix}^T$$

As we did before let $A = J\boldsymbol{I}(\boldsymbol{x}, t)$ and $\boldsymbol{\eta} = -\boldsymbol{I}_t$. If A^TA is invertible then the least squares solution to (2.4) is given by:

$$d\boldsymbol{x} = A^{\dagger}\boldsymbol{\eta} \tag{2.5}$$

To illustrate this consider the problem for a signal (pixel dimension $n = 1$, intensity dimension $m = 1$). In this case the Jacobian is just the usual gradient of the signal: $JI(x, t) = dI(x, t)/dx$ and the matrix A^TA is invertible if the gradient is nonzero. Compare this with the case of an image (pixel dimension is two, intensity dimension is one). In this case the Jacobian is again the gradient $JI(\boldsymbol{x},t) = \nabla I(\boldsymbol{x},t) = \begin{bmatrix} \partial I(\boldsymbol{x},t)/\partial x_1 & \partial I(\boldsymbol{x},t)/\partial x_2 \end{bmatrix}$ and the matrix $A^T A = \nabla I(\boldsymbol{x},t)^T \nabla I(\boldsymbol{x},t)$ is the outer product of the gradient row vector. Consequently the 2×2 matrix $A^T A$ for a single channel image is rank deficient (its rank is at most 1) and so it is not invertible. This singularity disappears in the case of a multichannel image. For example if $\boldsymbol{I} = \begin{bmatrix} R & G & B \end{bmatrix}^T$, then the rows of the Jacobian are the gradients of the red, green and blue channels:

$$J\boldsymbol{I} = \begin{bmatrix} \frac{\partial R}{\partial x_1} & \frac{\partial R}{\partial x_2} \\ \frac{\partial G}{\partial x_1} & \frac{\partial G}{\partial x_2} \\ \frac{\partial B}{\partial x_1} & \frac{\partial B}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \nabla R \\ \nabla G \\ \nabla B \end{bmatrix}$$

In this case the 2×2 matrix $A^T A = \nabla R^T \nabla R + \nabla G^T \nabla G + \nabla B^T \nabla B$ is the sum of the outer products of the three color channel gradient row vectors. Consequently it is invertible if any two of the channels have independent gradient vectors.[4] In general we find that:

$$J\boldsymbol{I}^T J\boldsymbol{I} = \sum_{i=1}^{m} (\nabla I_i)^T \nabla I_i$$

From this we conclude that the gradient normal matrix $J\boldsymbol{I}^T J\boldsymbol{I}$ is $n \times n$ where n is the pixel dimension and has rank at most m where m is the intensity dimension. Consequently it is not invertible if the pixel dimension exceeds the intensity dimension ($n > m$). If the pixel dimension is larger than the intensity dimension then we may overcome the non-invertibility of $A^T A$ by making the additional constraint that the optical flow is locally (i.e. in a region) constant. In this case the equation (2.4) holds over the region $\Omega(\boldsymbol{x})$ composed of N points and the least

[4] We should note here that for natural images the RGB channels tend to be highly correlated and therefore the matrix $J\boldsymbol{I}$ is likely to be poorly conditioned. We will come back to this problem in the experimental section.

squares solution is obtained by stacking these sets of equations into a large system:

$$A(\Omega(\boldsymbol{x}))d\boldsymbol{x} = \begin{bmatrix} J\boldsymbol{I}(\boldsymbol{y}_1) \\ \vdots \\ J\boldsymbol{I}(\boldsymbol{y}_N) \end{bmatrix} d\boldsymbol{x} = - \begin{bmatrix} \boldsymbol{I}_t(\boldsymbol{y}_1) \\ \vdots \\ \boldsymbol{I}_t(\boldsymbol{y}_N) \end{bmatrix} = \boldsymbol{\eta}$$

Again the least squares solution has the form $d\boldsymbol{x} = A^{\dagger}\boldsymbol{\eta}$. If we now look at the static optical flow case $d\boldsymbol{x}_{exact} = 0$ and assume that the images in the time sequence differ only by additive noise then the vector $\boldsymbol{\eta}$ is the additive noise over the region and the error $\boldsymbol{e} = d\boldsymbol{x}_{exact} - d\boldsymbol{x}_{computed}$ satisfies:

$$\|\boldsymbol{e}\| = \|d\boldsymbol{x}_{exact} - d\boldsymbol{x}_{computed}\| = \|0 - A^{\dagger}\boldsymbol{\eta}\| = \|A^{\dagger}\boldsymbol{\eta}\| \leq \|A^{\dagger}\| \; \|\boldsymbol{\eta}\|$$

Thus we see that even in this general setting the term $\|A^{\dagger}\|$ controls the error multiplication factor; that is the factor by which the input error (the noise $\boldsymbol{\eta}$) is multiplied to get the output error (the error in the optical flow estimate). As in the case of single channel images, large values of $\|A^{\dagger}\|$ correspond to points where we cannot estimate the optical flow accurately in the presence of noise at least for the static image case. As said earlier for the single channel image case we have $\|A^{\dagger}\|_2^2 = \frac{1}{\lambda_{min}(A^TA)}$; this motivates the role of A^TA in corner detector for the general problem of arbitrary pixel and intensity dimensions.

Remark 2.3.1 *For the purposes of interpretation it is helpful to rewrite A^TA as:*

$$A(\Omega(\boldsymbol{x}))^T A(\Omega(\boldsymbol{x})) = \sum_{j=1}^{N}\sum_{i=1}^{m} \left(\nabla I_i(\boldsymbol{y}_j)\right)^T \nabla I_i(\boldsymbol{y}_j) = \sum_{j=1}^{N} \left(J\boldsymbol{I}(\boldsymbol{x}_j)\right)^T J\boldsymbol{I}(\boldsymbol{x}_j) \quad (2.6)$$

That is, A^TA is the sum over the points $\boldsymbol{y}_j$ in the region $\Omega(\boldsymbol{x})$ of the outer products of the gradient vectors of each intensity channel (since the gradient operator returns a row vector).

2.3.4 Optical Flow for Arbitrary Motion Models

Throughout the whole discussion concerning the estimation of the optical flow, we assumed that the motion model for the image region $\Omega(\boldsymbol{x})$ was a pure translation. In this section we will extend the previous discussion to general motion models (see [116] for a feature tracking approach with affine motion models and the exhaustive discussion in [2]). Consider a model that describes the motion of a point $\boldsymbol{y}$ in the region $\Omega(\boldsymbol{x})$:

$$\begin{aligned} \boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}} : \Omega(\boldsymbol{x}) \subseteq \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ \boldsymbol{y} &\mapsto \boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) \end{aligned}$$

and let $\overline{\boldsymbol{\theta}}$ represent the identity in the parameter space (i.e. $\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{y}$).

Example 2.3.2 *Consider the situation depicted in Figure 2.2. In this case the region $\Omega(\boldsymbol{x})$ is a circular neighborhood defined as:*

$$\Omega(\boldsymbol{x}) = \left\{ \boldsymbol{y} \in \mathbb{R}^2 : (y_1 - x_1)^2 + (y_2 - x_2)^2 \leq r \right\}$$

and the rotation, translation and scaling is represented by the transformation:

$$\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{x} + s \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} (\boldsymbol{y} - \boldsymbol{x}) + \begin{bmatrix} a \\ b \end{bmatrix}$$

Hence, the parameter vector is $\boldsymbol{\theta} = \begin{bmatrix} a & b & s & \phi \end{bmatrix}^T$ (and consequently the identity vector is given by $\overline{\boldsymbol{\theta}} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^T$). A more convenient representation for this transformation can be obtained letting $\boldsymbol{\theta} = \begin{bmatrix} a & b & C & S \end{bmatrix}^T$, where $C = s\cos\phi$, $S = s\sin\phi$. This is possible because any matrix $A = sR$ where $s \in \mathbb{R}$ and

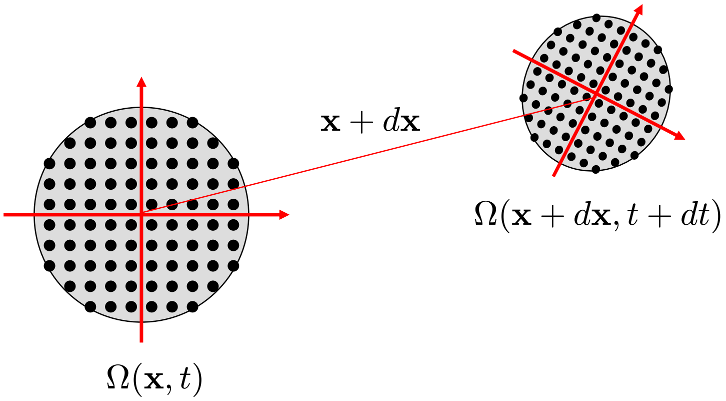

$\mathbf{x} + d\mathbf{x}$
$\Omega(\mathbf{x} + d\mathbf{x}, t + dt)$
$\Omega(\mathbf{x}, t)$

Figure 2.2: An example of a neighborhood $\Omega(\boldsymbol{x}) = \left\{\boldsymbol{y} \in \mathbb{R}^2 : (y_1 - x_1)^2 + (y_2 - x_2)^2 \leq r\right\}$ undergoing a rotation a translation and a scaling between time t and time $t + dt$.

$R \in SO(2)$ can be written in the form $\begin{bmatrix} C & -S \\ S & -C \end{bmatrix}$. Using this representation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ is linear in $\boldsymbol{\theta}$. Note also that $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{x}) = \boldsymbol{x} + \begin{bmatrix} a & b \end{bmatrix}^T$.

We can rewrite the brightness constraint equation (2.18) as:

$$\boldsymbol{I}(\boldsymbol{y}(t), t) = \boldsymbol{I}\left(\boldsymbol{T}_{\overline{\boldsymbol{\theta}}+d\boldsymbol{\theta},\boldsymbol{x}(t)}\left(\boldsymbol{y}(t)\right), t + dt\right) = c \tag{2.7}$$

The Taylor expansion of the second member yields:

$$\begin{aligned} \boldsymbol{I}\left(\boldsymbol{T}_{\overline{\boldsymbol{\theta}}+d\boldsymbol{\theta},\boldsymbol{x}(t)}\left(\boldsymbol{y}(t)\right), t + dt\right) = \boldsymbol{I}(\boldsymbol{y}(t), t) + J\boldsymbol{I}(\boldsymbol{y}(t), t) J_{\boldsymbol{\theta}} \boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}(t)}(\boldsymbol{y}(t)) d\boldsymbol{\theta} \\ + \boldsymbol{I}_t(\boldsymbol{y}(t), t) dt + \text{h. o. t.} \end{aligned}$$

and therefore, neglecting the higher order terms and plugging the previous expression in the brightness constraint equation we obtain:

$$J\boldsymbol{I}(\boldsymbol{y}(t), t) J_{\boldsymbol{\theta}} \boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}(t)}(\boldsymbol{y}(t)) d\boldsymbol{\theta} = -\boldsymbol{I}_t(\boldsymbol{y}(t), t) dt$$

If, similarly to what did before, we let $A(\boldsymbol{y}(t)) = J\boldsymbol{I}(\boldsymbol{y}(t),t)J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}(t)}(\boldsymbol{y}(t))$, $dt = 1$ and $\boldsymbol{\eta} = -\boldsymbol{I}_t(\boldsymbol{y}(t),t)$, we can write:

$$A(\Omega(\boldsymbol{x}))d\boldsymbol{\theta} = \boldsymbol{\eta} \tag{2.8}$$

To estimate the motion parameters $d\boldsymbol{\theta} \in \mathbb{R}^p$ we need at least p equations. Equation (2.8) can be solved in a least square sense only if $m \geq p$. If this condition is not met once again we stack the equations that describe the motion of every point belonging to the region $\Omega(\boldsymbol{x})$, obtaining a GGM that has the form:

$$A(\Omega(\boldsymbol{x})) = \begin{bmatrix} J\boldsymbol{I}(\boldsymbol{y}_1(t),t)J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}(t)}(\boldsymbol{y}_1(t)) \\ \vdots \\ J\boldsymbol{I}(\boldsymbol{y}_N(t),t)J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}(t)}(\boldsymbol{y}_N(t)) \end{bmatrix} \tag{2.9}$$

If also in this case we assume that the images in the time sequence differ only by additive noise (so that the vector $\boldsymbol{\eta}$ actually represents the additive noise over the region $\Omega(\boldsymbol{x})$) and we define the error vector to be:

$$\boldsymbol{e} = d\boldsymbol{\theta}_{exact} - d\boldsymbol{\theta}_{computed}$$

then the term $\|A^\dagger\|$ controls the error multiplication factor and if we consider the matrix 2-norm we still have that $\|A^\dagger\|_2^2 = \frac{1}{\lambda_{min}(A^TA)}$. Therefore we have shown how the generalized gradient normal matrix plays a central role in estimating the optical flow for generic motion models for generalized multispectral images. Finally note that (2.9) can be considered a generalization of the gradient matrix that was introduced in (2.2.1) in the presence of motion models more complicated than pure translations.

2.4 The Generalized Gradient Matrix: a Region Sensitivity Perspective

2.4.1 Condition Theory for Region Sensitivity

What is the sensitivity of an image point neighborhood $\Omega(\boldsymbol{x})$ to noise? To answer this question we shall measure how much the intensity pattern in the considered neighborhood looks like itself after it is perturbed by noise (in other words we are trying to quantify the degree of self-similarity of the neighborhood). To this purpose, the noise can be simply represented by an additive random signal that sums to the intensity. However, the quantitative measurement of the effects produced by the noise is a more complex task. Consider a point $\boldsymbol{y} \in \Omega(\boldsymbol{x})$. The expression for the image intensity $\boldsymbol{I}$ corrupted by noise $\boldsymbol{\eta}$ at point $\boldsymbol{y}$ is given by:

$$\widetilde{\boldsymbol{I}}(\boldsymbol{y}) \stackrel{\text{def}}{=} \boldsymbol{I}(\boldsymbol{y}) + \boldsymbol{\eta} \tag{2.10}$$

We choose to model the effect of the noise by a transformation parameterized by the vector $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}} + \Delta\boldsymbol{\theta}$ that describes the geometric distortion of the intensity pattern in $\Omega(\boldsymbol{x})$. More precisely:

$$\widetilde{\boldsymbol{I}}(\boldsymbol{y}) = \boldsymbol{I}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}}+\Delta\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})) \tag{2.11}$$

where:

$$\begin{aligned} \boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}} : \Omega(\boldsymbol{x}) \subseteq \mathbb{R}^n \quad &\rightarrow \quad \mathbb{R}^n & (2.12) \\ \boldsymbol{y} \quad &\mapsto \quad \boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) & (2.13) \end{aligned}$$

and $\overline{\boldsymbol{\theta}}$ represents the identity in the parameter space (i.e. $\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{y}$). It is clear that a neighborhood is more sensitive than another if the same amount of noise produces larger deviations from $\overline{\boldsymbol{\theta}}$ in (2.11). We now have all the ingredients to answer the question that opened this section: the sensitivity of a point neighborhood to noise will be measured using the notion of differential condition number introduced in (2.2.2):

Definition 2.4.1 *The condition number associated with the point neighborhood* $\Omega(\boldsymbol{x})$ *with respect to the transformation* $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ *is defined as:*

$$K_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}(\Omega(\boldsymbol{x})) \stackrel{\text{def}}{=} \lim_{\delta \to 0} \sup_{\|\boldsymbol{\eta}\| \leq \delta} \frac{\|\Delta\boldsymbol{\theta}\|}{\|\boldsymbol{\eta}\|} \tag{2.14}$$

The larger the condition number $K_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}$ is, the larger is the magnitude of the variation of the parameter vector $\Delta\boldsymbol{\theta}$ induced by the noise and consequently the larger is the sensitivity of the neighborhood to the noise (or more pictorially, the smaller is the condition number the more similar is the intensity in $\Omega(\boldsymbol{x})$ to itself after being perturbed by noise).

It is now worth noticing two things. First, the condition number becomes practically useful only if we are able to provide a closed form for its expression. Second, if the statistical distribution of the noise is fixed, we expect the derivatives of the image intensity pattern in $\Omega(\boldsymbol{x})$ to play a fundamental role in determining the sensitivity of $\Omega(\boldsymbol{x})$ (or equivalently in the calculation of the condition number). Along this line of thought, the following theorem provides a computable expression to estimate the condition number, which turns out to be intimately connected with the gradient matrix associated with the point neighborhood $\Omega(\boldsymbol{x})$ introduced in (2.2) and generalized in (2.9).

Theorem 2.4.2 *A first order estimate of the condition number* (2.14) *is given* *by:*

$$\hat{K}_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}(\Omega(\boldsymbol{x})) = \|A^{\dagger}\left(\Omega(\boldsymbol{x})\right)\| \tag{2.15}$$

where † *denotes the pseudo inverse of the matrix:*

$$A\left(\Omega(\boldsymbol{x})\right) \overset{\text{def}}{=} \begin{bmatrix} A(\boldsymbol{y}_1) \\ \vdots \\ A(\boldsymbol{y}_N) \end{bmatrix} \in \mathbb{R}^{mN\times p} \tag{2.16}$$

which is formed by the N *sub-matrices:*

$$A(\boldsymbol{y}_i) \overset{\text{def}}{=} w(\boldsymbol{y}_i - \boldsymbol{x}) J\boldsymbol{I}(\boldsymbol{y}_i)\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i) \tag{2.17}$$

obtained from a set of N *points that sample the neighborhood* $\Omega(\boldsymbol{x})$*. The scalar function* $w(\boldsymbol{y}_i - \boldsymbol{x})$ *denotes the weight associated with the point* $\boldsymbol{y}_i$*.*

Proof: In the limit for $\boldsymbol{\eta} \to \boldsymbol{0}$, we have that $\Delta\boldsymbol{\theta} \to \boldsymbol{0}$ and therefore (assuming that the necessary smoothness conditions are satisfied) we can expand the right hand side of equation (2.11) about the point $\overline{\boldsymbol{\theta}}$ (in the parameter space) via Taylor series, obtaining for each point $\boldsymbol{y}_i$ that samples the neighborhood $\Omega(\boldsymbol{x})$ the following expression:

$$\widetilde{\boldsymbol{I}}(\boldsymbol{y}_i) = \boldsymbol{I}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i)) + J\boldsymbol{I}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i))\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i)\Delta\boldsymbol{\theta} + \text{h. o. t.} \ = \boldsymbol{I}(\boldsymbol{y}_i) + \boldsymbol{\eta}_i$$

If we drop the higher order terms, we recognize that $\boldsymbol{I}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i)) \equiv \boldsymbol{I}(\boldsymbol{y}_i)$ and we multiply both members of the equation by a suitable weighting function $w(\boldsymbol{y}_i - \boldsymbol{x})$, we obtain the approximate equation:

$$w(\boldsymbol{y}_i - \boldsymbol{x}) J\boldsymbol{I}(\boldsymbol{y}_i)\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i)\Delta\boldsymbol{\theta} \approx w(\boldsymbol{y}_i - \boldsymbol{x})\boldsymbol{\eta}_i \tag{2.18}$$

where the $(h,k)^{th}$ entry of the Jacobian matrix $J\boldsymbol{I}(\boldsymbol{y}_i) \in \mathbb{R}^{m\times n}$ is given by $\partial I_h(\boldsymbol{y}_i)/\partial y_k$ and the matrix $J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i) \in \mathbb{R}^{n\times p}$ represents the Jacobian of the transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ with respect to the p dimensional parameter vector $\boldsymbol{\theta}$. If equation (2.18) holds for N points and if we indicate the overall weighted noise vector with $\boldsymbol{\eta}$, then we can group the resulting set of N equations into the linear system:

$$A\left(\Omega(\boldsymbol{x})\right)\Delta\boldsymbol{\theta} = \boldsymbol{\eta} \tag{2.19}$$

where $A\left(\Omega(\boldsymbol{x})\right)$ is obtained by stacking the matrices $A(\boldsymbol{y}_i)$ as written in (2.17). At this point we have been able to relate the displacement $\Delta\boldsymbol{\theta}$ of the parameter vector due to the noise. If $A(\Omega(\boldsymbol{x}))$ is full rank, then equation (2.19) can be inverted in a least square sense, yielding:

$$\Delta\boldsymbol{\theta} = A^{\dagger}\left(\Omega(\boldsymbol{x})\right)\boldsymbol{\eta} \tag{2.20}$$

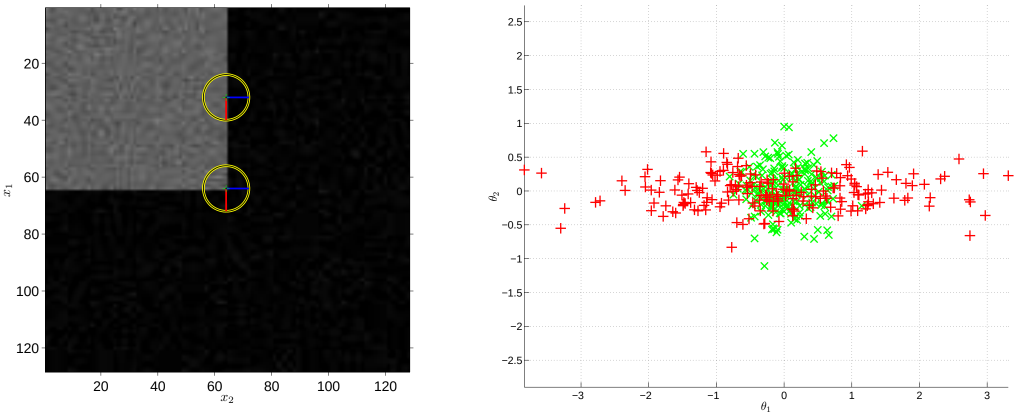
Since for any valid vector norm the positive homogeneity property holds, i.e. $\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$, then we can write:

$$\sup_{\|\boldsymbol{\eta}\|\le\delta}\frac{\|\Delta\boldsymbol{\theta}\|}{\|\boldsymbol{\eta}\|} = \sup_{\|\boldsymbol{\eta}\|\le\delta}\frac{\|A^{\dagger}\left(\Omega(\boldsymbol{x})\right)\boldsymbol{\eta}\|}{\|\boldsymbol{\eta}\|} = \sup_{\|\boldsymbol{\eta}\|=1}\|A^{\dagger}\left(\Omega(\boldsymbol{x})\right)\boldsymbol{\eta}\| = \|A^{\dagger}\left(\Omega(\boldsymbol{x})\right)\|$$

Therefore the condition number can be estimated as: $\hat{K}_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}(\Omega(\boldsymbol{x})) = \|A^{\dagger}\left(\Omega(\boldsymbol{x})\right)\|$.

■

Example 2.4.3 *We will illustrate the concepts introduced in this section using the single channel synthetic image shown in Figure 2.3(a). The original image is composed of a bright square in the top left corner (with intensity value equal to 128) placed over a dark background (with intensity value 0). We considered 200 images obtained by adding to the original image a different realization of Gaussian noise characterized by zero mean and standard deviation* $\sigma_\eta = 10$. *For each of these realizations we considered the circular neighborhoods* $\Omega_1(\boldsymbol{x}_1)$ *and*



(a)

(b)

Figure 2.3: Image (a) is realization of a synthetic image composed of a bright square in the top left corner (with intensity value equal to 128) placed over a dark background (with intensity value 0) corrupted by Gaussian noise characterized by zero mean and standard deviation $\sigma_\eta = 10$. The bright circles identify the neighborhood $\Omega_1(\boldsymbol{x}_1)$ (top) and the neighborhood $\Omega_2(\boldsymbol{x}_2)$ (center). Figure 2.3(b) shows the parameter vectors calculated for a specific instance of Figure 2.3(a) by solving (2.20) in a least square sense. The red crosses are associated with the region $\Omega_1(\boldsymbol{x}_1)$ and the green crosses to the region $\Omega_2(\boldsymbol{x}_2)$.

$\Omega_2(\boldsymbol{x}_2)$ *(represented as bright circles in Figure 2.3(a)) with radius 8 pixels and respectively centered at* $\boldsymbol{x}_1 = \begin{bmatrix} 32 & 64 \end{bmatrix}^T$ *and* $\boldsymbol{x}_2 = \begin{bmatrix} 64 & 64 \end{bmatrix}^T$. *Each marker in Figure 2.3(b) corresponds to the parameter vector calculated for a specific instance of Figure 2.3(a) by solving (2.20) in a least square sense. The red crosses are associated with the region* $\Omega_1(\boldsymbol{x}_1)$ *and the green crosses with the region* $\Omega_2(\boldsymbol{x}_2)$. *The transformation chosen to model the effects of noise is a pure translation:*

$$\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \begin{bmatrix} y_1 + \theta_1 \\ y_2 + \theta_2 \end{bmatrix}$$

It is clear that the spread of the parameter vector $\boldsymbol{\theta}$ *(and more specifically of its first component* θ_1*) is much larger for the neighborhood* $\Omega_1(\boldsymbol{x}_1)$. *This fact can be explained comparing the intensity pattern contained in the two neighborhoods: when* $\Omega_1(\boldsymbol{x}_1)$ *slides along the straight edge, its intensity content does not vary. Therefore small amounts of noise can be "compensated" by larger transformations. On the other hand the corner contained in* $\Omega_1(\boldsymbol{x}_1)$ *remains very distinctive even after it is perturbed by noise, and therefore the spread of the transformation parameters is smaller. Finally note that since the components of the vector* $\boldsymbol{\eta}$ *are i. i. d. Gaussian variables, then* $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2 I)$. *Moreover, since linear transformations of jointly Gaussian vectors are still jointly Gaussian vectors and* $A^\dagger \left(A^\dagger\right)^T = (A^T A)^{-1}$, *then from* $\boldsymbol{\Delta\theta} = -A^\dagger \boldsymbol{\eta}$ *it follows that* $\boldsymbol{\Delta\theta} \sim \mathcal{N}\left(\mathbf{0}, \sigma_\eta^2 (A^T A)^{-1}\right)$. *This observation explains the scatter of the parameter vector in Figure 2.3(b) and justifies the choice of defining the condition number (2.14) in terms of the supremum of the ratio between* $\|\Delta\boldsymbol{\theta}\|$ *and* $\|\boldsymbol{\eta}\|$.

2.4.2 Condition Theory for Local Transformation Estimation

In this section we study the conditions under which we can robustly estimate the parameters of a transformation that relates two image regions $\Omega(\boldsymbol{x})$ and $\Omega'(\boldsymbol{x}')$. The same approach was introduced in [63] in the case of gray-level images for a set of linear transformations.

Consider a transformation defined as in (2.12), so that for any $\boldsymbol{y} \in \Omega(\boldsymbol{x})$ there exists a point $\boldsymbol{y}' \in \Omega'(\boldsymbol{x}')$ such that $\boldsymbol{y}' = \boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})$ (technically speaking, $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ establishes a bijection between the sets $\Omega(\boldsymbol{x})$ and $\Omega'(\boldsymbol{x}')$). Let's also define the cost function:

$$C_{\boldsymbol{T}}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{\boldsymbol{y} \in \Omega(\boldsymbol{x})} w(\boldsymbol{y} - \boldsymbol{x}) \|\boldsymbol{I}(\boldsymbol{y}) - \boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}))\|^2 \tag{2.21}$$

where w is an appropriate weighting function. The cost function C measures the intensity discrepancy between two corresponding regions. In this case our goal is to estimate the parameter vector that minimizes (2.21), i.e. :

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathbb{R}^p}{\operatorname{argmin}}\, C_{\boldsymbol{T}}(\boldsymbol{\theta}) \tag{2.22}$$

Of course, we are interested in selecting point features (and their corresponding neighborhood) such that small amounts of noise will not bias the estimate (2.22). To decide whether or not a region $\Omega(\boldsymbol{x})$ can be used to reliably estimate the parameter vector $\boldsymbol{\theta}$, we resort once again to the notion of condition number. Consider a noise free case where $\boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})) = \boldsymbol{I}(\boldsymbol{y})$. The effect of the noise can be modeled as a variation of the vector that parameterizes the transformation between the image regions, i.e. :

$$\boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta}+\Delta\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})) = \boldsymbol{I}(\boldsymbol{y}) + \boldsymbol{\eta} \tag{2.23}$$

To relate the impact of $\boldsymbol{\eta}$ to the parameter variation $\Delta\boldsymbol{\theta}$ we will use the condition number defined in (2.14). The next theorem will show how the first order estimate of the condition number obtained in (B.1) is valid also for the problem defined in this section.

Theorem 2.4.4 *The expression for the condition number introduced in* (B.1)*, i.e. :*

$$\hat{K}_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}(\Omega(\boldsymbol{x})) = \|A^{\dagger}\left(\Omega(\boldsymbol{x})\right)\|$$

measures the stability of the parameter estimate $\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}\in\mathbb{R}^p} C_{\boldsymbol{T}}(\boldsymbol{\theta})$*.*

Proof: The first step of the proof is to relate the parameter variation $\Delta\boldsymbol{\theta}$ to the noise vector $\boldsymbol{\eta}$. Consider the Taylor expansion of the right hand side term in (2.23) (this is meaningful when $\boldsymbol{\eta} \to \mathbf{0}$ and therefore $\Delta\boldsymbol{\theta} \to \mathbf{0}$). Neglecting the higher order terms we can write:

$$\boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta}+\Delta\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})) = \boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})) + J\boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}))J_{\boldsymbol{\theta}}\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})\Delta\boldsymbol{\theta}$$

and therefore using expression (2.23) we get:

$$\boldsymbol{I}(\boldsymbol{y}) - \boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})) = J\boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}))J_{\boldsymbol{\theta}}\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})\Delta\boldsymbol{\theta} - \boldsymbol{\eta} \tag{2.24}$$

In the noisy case the cost function $C_{\boldsymbol{T}}$ is minimized by $\boldsymbol{\theta} + \Delta\boldsymbol{\theta}$: by plugging equation (2.24) in the expression of the cost function we obtain:

$$C_{\boldsymbol{T}}(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) = \frac{1}{2}\sum_{\boldsymbol{y}\in\Omega(\boldsymbol{x})} w(\boldsymbol{y} - \boldsymbol{x})\|J\boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}))J_{\boldsymbol{\theta}}\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})\Delta\boldsymbol{\theta} - \boldsymbol{\eta}\|^2$$

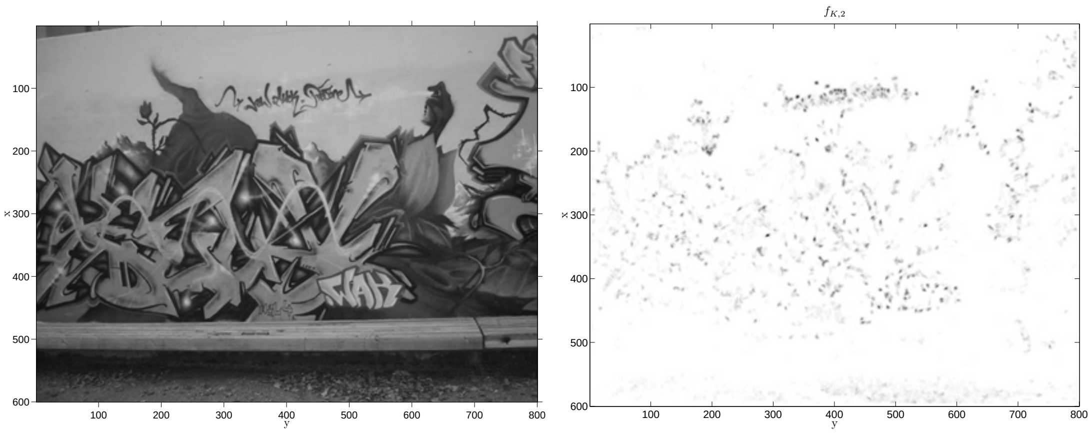
This expression is formed by a summation of positive terms and is minimized when each of these terms is simultaneously minimized. This happens when, for each $\boldsymbol{y}_i \in \Omega(\boldsymbol{x})$, we have that:

$$w(\boldsymbol{y}-\boldsymbol{x}) J\boldsymbol{I}'(\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}_i)) J_{\boldsymbol{\theta}} \boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}_i) \Delta\boldsymbol{\theta} = w(\boldsymbol{y}-\boldsymbol{x}) \boldsymbol{\eta}_i$$

The structure of this equation closely resembles the structure of equation (2.18). Therefore, from now on, we can follow the same line of thought used in the proof of Theorem 2.4.2, the only difference being the fact the second factor that composes the matrices $A(\boldsymbol{y}_i)$ depends on the Jacobian of the image $\boldsymbol{I}'$ about the point $\boldsymbol{y}_i' = \boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}_i)$. ■

From this discussion we conclude that $\boldsymbol{\theta}$ can be estimated reliably for regions $\Omega(\boldsymbol{\theta})$ where $K_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}$ is small.

The concepts introduced in this section share some commonalities with those described in [129]. In fact, the matrix M defined in [129], p. 5, Equation (5), corresponds to the constitutive block of the GGM (2.17). Moreover the saliency criterion defined by Triggs is essentially equivalent to the value of the estimate of the condition number (B.1) calculated using a specific matrix norm and neglecting the illumination model. Example 2.6.2 will summarize these considerations in the context of point detection for two dimensional graylevel images. Note also that the paper by Tommasini et al. [125] builds on the work of Shi and Tomasi [116] by examining the statistics of the residual difference between a window and a computed backtransform of the corresponding window in a second image with the goal of deriving conditions for rejecting a tentative match.



(a) (b)

Figure 2.4: Figure (a) shows a Graffiti scene and Figure (b) displays an example of the corresponding detector response map. Darker points indicate a stronger response. The response is larger for neighborhoods that contain highly structured (i.e. well conditioned) intensity patterns. Flat or poorly textured areas do not produce any response.

2.5 Generalized Corner Detector Functions

The main goal of the previous discussion was to explain why and how the GGM encodes the information necessary to identify set of points that can be reliably matched between images. In this section we will study a set of generalized detector functions that extract such information from the GGM and that produce a dense map that can be used to identify distinctive and stable tie points (see Figure 2.4). More specifically we will:

- review the properties of the GGM and study its behavior in the presence of geometric transformations of the images (Section 2.5.1),
- introduce a set of generalized corner detector functions (where the attribute generalized indicates the fact that they are defined for $n \geq 1$ and $m \geq$

1) based on the matrix $A(\Omega(\boldsymbol{x}))$, establish analytic relations between the generalized corner detector functions and find analytical bounds for different types of transformations associated with the GGM $A(\Omega(\boldsymbol{x}))$ (Section 2.5.2),

- enumerate a list of remarkable properties of the generalized corner detector functions (Section 2.5.3),
- specialize this results for 2-dimensional single channel images and show how some of the most widely used detector functions are nothing but specific instances of these generalized detectors (Section 2.6).

2.5.1 The Generalized Gradient Matrix: Recapitulation

In the previous sections we showed how the GGM $A(\Omega(\boldsymbol{x})) \in \mathbb{R}^{mN \times p}$ constructed for different purposes and at different levels of generalization contains the information that enables us to decide whether the region Ω about the point $\boldsymbol{x}$ can be tracked, identified or matched reliably. This analysis was carried out studying the effects of the noise and using the tools of condition theory. For the sake of convenience we recall that given a neighborhood $\Omega(\boldsymbol{x})$ and a set of points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ that sample such a neighborhood, the GGM is defined as:

$$A\left(\Omega(\boldsymbol{x})\right) \overset{\text{def}}{=} \begin{bmatrix} A(\boldsymbol{y}_1) \\ \vdots \\ A(\boldsymbol{y}_N) \end{bmatrix} \in \mathbb{R}^{mN \times p} \tag{2.25}$$

Such a matrix is formed by the N submatrices:

$$A(\boldsymbol{y}_i) = w(\boldsymbol{y}_i - \boldsymbol{x}) J_{\boldsymbol{\theta}} \boldsymbol{I}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i)) = w(\boldsymbol{y}_i - \boldsymbol{x}) J\boldsymbol{I}(\boldsymbol{y}_i)\; J_{\boldsymbol{\theta}} \boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_i)$$

stacked one on top of the other, and w denotes a suitable weighting function (usually chosen to be radially symmetric). If all the intensity channels of a generalized image are scaled by a factor ν, then the GGM is scaled by the same factor and so is its spectrum. For the sake of the analysis that will be developed in the reminder of the chapter we rewrite the matrix (2.25) as the product of three matrices. The first one is the diagonal matrix of the weights, the second one is the block diagonal matrix of the image Jacobians and finally the last one is the matrix composed of the transformation Jacobians:

$$A = \underbrace{\begin{bmatrix} w(\boldsymbol{y}_1 - \boldsymbol{x})I_m & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w(\boldsymbol{y}_N - \boldsymbol{x})I_m \end{bmatrix}}_{\mathbb{R}^{mN\times mN}} \underbrace{\begin{bmatrix} J\boldsymbol{I}(\boldsymbol{y}_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & J\boldsymbol{I}(\boldsymbol{y}_N) \end{bmatrix}}_{\mathbb{R}^{mN\times nN}} \underbrace{\begin{bmatrix} J_{\boldsymbol{\theta}}\boldsymbol{T}_{\bar{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_1) \\ \vdots \\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\bar{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}_N) \end{bmatrix}}_{\mathbb{R}^{nN\times p}}$$

It is worth remarking that the GGM inherits any dependence of $J_{\boldsymbol{\theta}}\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ on the parameters of the transformation. For this reason, if the transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ is affine, then $A\left(\Omega(\boldsymbol{x})\right)$ does not depend on $\boldsymbol{\theta}$.

On the Invariance of the Generalized Gradient Matrix

Before continuing, we would like to study what happens to the GGM when it is constructed for two *corresponding* neighborhoods that are related via the geometric transformation $\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}$ (defined, mutatis mutandis, as in (2.12)). We assume that such transformation establishes a bijection between $\Omega(\boldsymbol{x})$ and $\Omega'(\boldsymbol{x}')$ such that if $\boldsymbol{y}' = \boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y})$ then $\boldsymbol{I}(\boldsymbol{x}) = \boldsymbol{I}'(\boldsymbol{x}')$. At a first look one may expect that the models we presented in Section 2.4 should have the property of describing the sensitivity of a neighborhood despite the image transformation $\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}$. However it

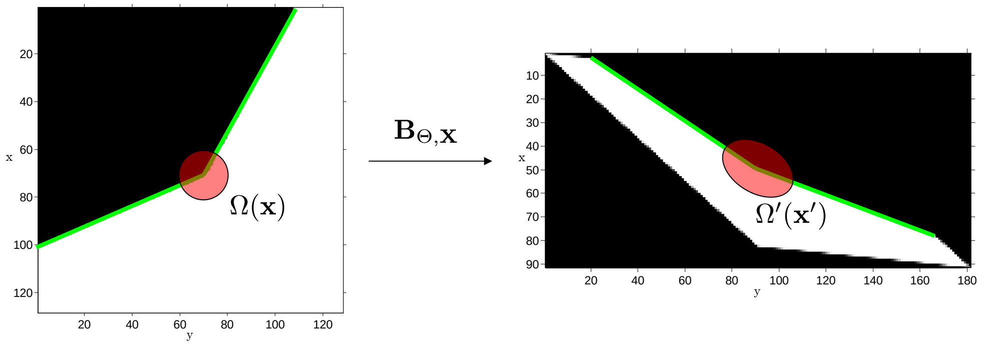


Figure 2.5: Two images related by a linear transformation. Note how the edges that define the corner structure in $\Omega(\boldsymbol{x})$ become almost collinear after the transformation (2.26) is applied to the image.

turns out that such an expectation cannot (and should not) hold in general. In fact it is possible that certain transformations modify the sensitivity of a neighborhood in the presence of noise. This is illustrated in the following example.

Example 2.5.1 *Consider the images in Figure 2.5 which are related via the linear transformation:*

$$\boldsymbol{B}_{\Theta,\boldsymbol{X}}(\boldsymbol{x}) = \begin{bmatrix} s_x & \gamma \\ 0 & s_y \end{bmatrix} \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \boldsymbol{x} \tag{2.26}$$

where $s_x = 0.4$, $s_y = 1.0$, $\gamma = 0.5$ and $\phi = -45°$. The intensity pattern that is contained in $\Omega(\boldsymbol{x})$ defines a corner-like structure. The transformation $\boldsymbol{B}_{\Theta,\boldsymbol{X}}$ has the effect of straightening the corner edges in $\Omega(\boldsymbol{x})$, making them almost collinear. Qualitatively it seems resonable that the intensity pattern contained in $\Omega'(\boldsymbol{x}')$ is more sensitive to the effects of noise.

We will now make the previous statements more precise by introducing a sufficient condition[5] that guarantees that the matrices $A(\Omega(\boldsymbol{x}))$ and $A'(\Omega'(\boldsymbol{x}'))$ have

[5] Another sufficient condition will be presented in Lemma 2.5.14.

the same spectrum (which is intimately connected to the structure and to the sensitivity of the associated neighborhood).

Theorem 2.5.2 (Spectrum Invariance Sufficient Conditions) *Let $\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}$ be a transformation that locally relates two images so that $\boldsymbol{I}'(\boldsymbol{x}') = \boldsymbol{I}(\boldsymbol{x})$ (where $\boldsymbol{x}' = \boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{x})$). If the sampling of the neighborhood is dense enough, and:*

$$J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}'}(\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y})) = J\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y})\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}) \tag{2.27}$$

then the spectra of the GGMs $A(\Omega(\boldsymbol{x}))$ and $A'(\Omega'(\boldsymbol{x}'))$ are the same.

Proof: The proof of the theorem starts by considering the blocks that compose the GGMs $A(\Omega(\boldsymbol{x}))$ and $A'(\Omega'(\boldsymbol{x}'))$. Without losing any generality we assume the weighting to be uniform (i.e. $w \equiv 1$). Using the notation introduced in Section 2.4 and summarized at the beginning of this section we can write:

$$\begin{aligned} A(\boldsymbol{y}) &= J_{\boldsymbol{\theta}}\boldsymbol{I}(T_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y})) \\ A'(\boldsymbol{y}') &= J_{\boldsymbol{\theta}}\boldsymbol{I}'(T_{\overline{\boldsymbol{\theta}},\boldsymbol{x}'}(\boldsymbol{y}')) \end{aligned}$$

Recalling that the vector $\overline{\boldsymbol{\theta}}$ parameterizes the identity transformation and expanding the previous expressions using the chain rule we obtain:

$$\begin{aligned} A(\boldsymbol{y}) &= J_{\boldsymbol{\theta}}\boldsymbol{I}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y})) \\ &= J\boldsymbol{I}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}))\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}) \\ &= J\boldsymbol{I}(\boldsymbol{y})\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}) \end{aligned} \tag{2.28}$$

and similarly:

$$\begin{aligned} A'(\boldsymbol{y}') &= J_{\boldsymbol{\theta}}\boldsymbol{I}'(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}'}(\boldsymbol{y}')) \\ &= J\boldsymbol{I}'(\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}'}(\boldsymbol{y}'))\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}'}(\boldsymbol{y}') \\ &= J\boldsymbol{I}'(\boldsymbol{y}')\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}'}(\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y})) \end{aligned} \tag{2.29}$$

Moreover, since $\boldsymbol{I}(\boldsymbol{y}) = \boldsymbol{I}'(\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y}))$, differentiating both members of the previous equation with respect to $\boldsymbol{y}$ yields:

$$J\boldsymbol{I}(\boldsymbol{y}) = J\boldsymbol{I}'(\boldsymbol{y}')\ J\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y}) \tag{2.30}$$

Plugging the previous expression in (2.28) we can rewrite the blocks that compose the matrix $A(\Omega(\boldsymbol{x}))$ as:

$$A(\boldsymbol{y}) = J\boldsymbol{I}'(\boldsymbol{y}')\ J\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y})\ J\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{x}}(\boldsymbol{y}) \tag{2.31}$$

Comparing (2.29) with (2.31) we see that the condition (2.27) is sufficient to guarantee that $A(\boldsymbol{y}) \equiv A'(\boldsymbol{y}')$. Finally, since we are considering corresponding neighborhoods, when the sampling is dense enough[6], the matrices $A(\Omega(\boldsymbol{x}))$ and $A'(\Omega'(\boldsymbol{x}'))$ are approximately related by via block permutations of their rows. Such an operation leaves the spectrum of the matrices unchanged. ■

There are cases where we are interested in *comparing the sensitivity* of two neighborhoods that are related via a transformation $\boldsymbol{B}_{\boldsymbol{\Theta},\boldsymbol{X}}(\boldsymbol{y})$. If the spectral invariance condition (2.27) is not met, one possible solution is to map the neighborhoods onto a normalized domain $\overline{\Omega}$ and evaluate $A(\overline{\Omega})$ (see Figure 2.6). We will now present an example where we adopt this strategy to estimate the local scaling factor between two corresponding neighborhoods.

Example 2.5.3 *Consider a RGB image* $\boldsymbol{I}'$ *obtained from the image* $\boldsymbol{I}$ *after applying a rotation of* 30° *degrees and a* 25% *scaling. The left (right) plot in Figure*

[6] Neither in the statement of Theorem 2.5.2 nor this its proof we formalized precisely how the sampling density actually modifies the spectrum of the A. Experiments with real imagery suggest that this issue has a limited impact in practical applications. An accurate analysis to derive analytical bounds for the spectrum fluctuations is beyond the scope of this work. However we believe that the Hoffman and Wielandt theorem (see [53], p. 365) is an essential tool to understand the relation between the density of the sampling and the fluctuation of the spectrum.

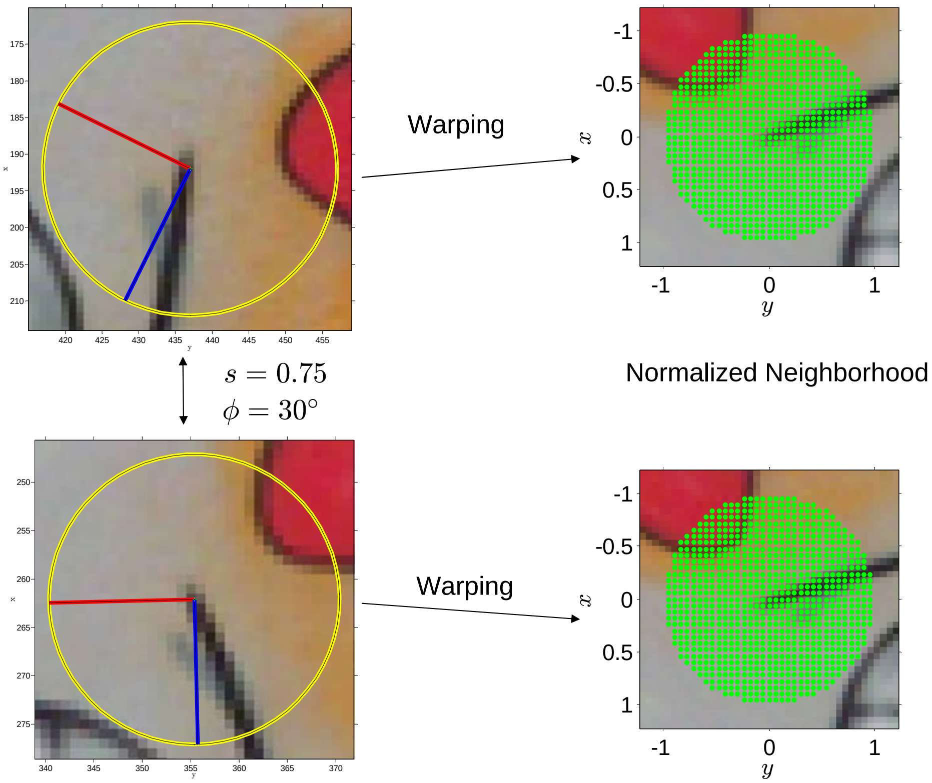


Figure 2.6: The image shows two corresponding circular regions and their average gradient direction defined as in (2.32). The regions are warped on a normalized neighborhood. The green dots represent the discretization lattice.

2.7 displays the value of the reciprocal of the condition number estimate of A (A') for a circular neighborhood parameterized by the radius r (recall that the condition number estimate depends only on the spectrum of A). The transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ that describe the effects of the noise models a rotation, a scaling and a translation, and therefore the condition (2.27) is not met. To cope with this problem, the neighborhoods are warped onto a circular normalized patch $\overline{\Omega}$ that is centered about the points $\boldsymbol{x}$ ($\boldsymbol{x}'$) and whose local coordinate system is aligned to the average

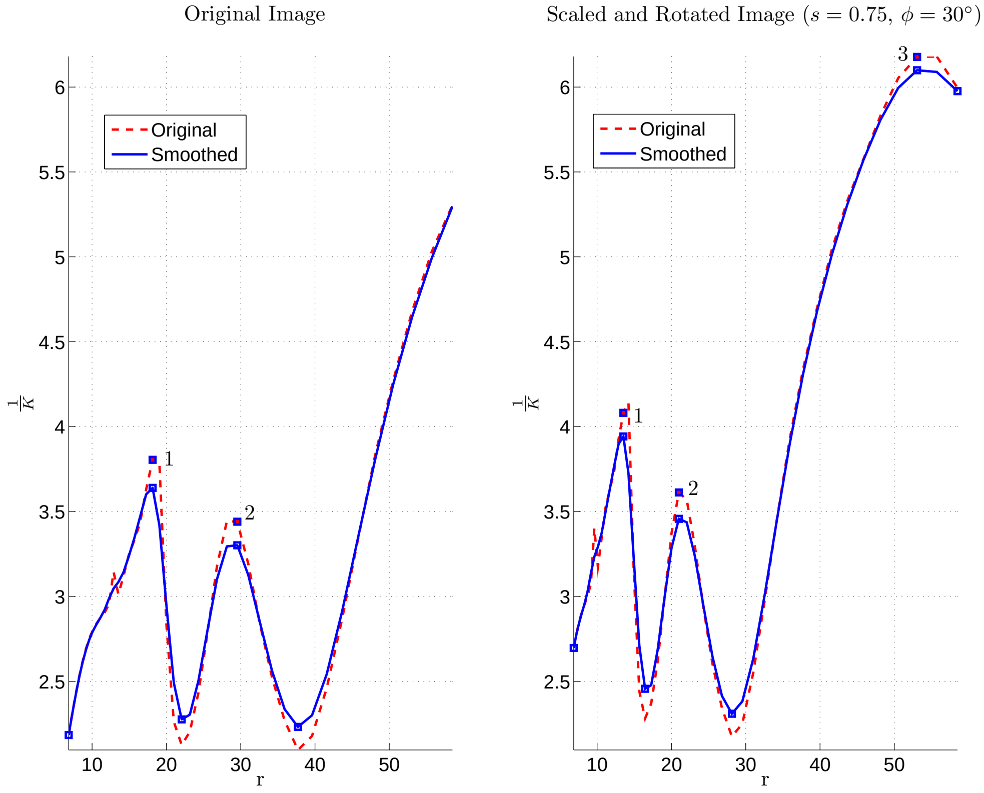


Figure 2.7: The plots display the reciprocal of the condition number of the GGM (2.25) for two neighborhoods centered about two corresponding points belonging to an image pair related by a rotation of 30° degrees and a 25% scaling (see Example 2.5.3). The considered transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ models a rotation, a scaling and a translation. The dashed line displays the original curve, whereas the continuous blue curve corresponds to the smoothed version.

image gradient (see Figure 2.6), which is calculated according to:

$$\boldsymbol{g} = \frac{1}{Nm} \sum_{i=1}^{N} \sum_{j=1}^{m} \nabla I_j(\boldsymbol{x}_i) \tag{2.32}$$

The plots of the reciprocal of the condition number associated with the normalized neighborhood are displayed in Figure 2.7. Such plots are essentially related by a uniform scaling along the radii axis. When the ratio between the radii corresponds to the scaling factor between the images, the inverse of the condition number ap-

proximately attains the same value. Indeed the ratio of the radii corresponding to the first local maximum of the curves returns the scaling factor between the images, as confirmed by the numerical values that can be retrieved from the plots in Figure 2.7:

$$\frac{\textit{first peak of } \frac{1}{K} \textit{ for image } \boldsymbol{I}}{\textit{first peak of } \frac{1}{K} \textit{ for image } \boldsymbol{I'}} = \frac{13.54}{18.14} \approx 0.75$$

The amplitude differences in the corresponding portions of the curves are to be attributed to the effects of finite precision arithmetic, to the discretization of the signals (in fact the curves appear to be more distorted for smaller neighborhoods) and to the fact that the neighborhoods are not sampled exactly at the same locations. See Section 6.1.1 in Chapter 6 for a thorough discussion regarding the possibility of using the condition number to detect the characteristic structure of a point neighborhood.

2.5.2 Generalized Corner Detectors Basics

Definition 2.5.4 *A (local) Generalized Corner Detector Function (GCDF) for an image with pixel dimension n and intensity dimension m associated with the transformation $T_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})$ is a real-valued function f of the GGM $A(\Omega(\boldsymbol{x}))$ defined as in (2.25).*

Definition 2.5.5 *Let $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_p(A)$ be the singular values of the GGM A (assuming $mN \geq p$). Then a Generalized Corner Detector Function (GCDF) that depends solely on the spectrum of the GGM is called a Spectral Generalized Corner Detector Function (SGCDF).*

- *Generalized Harris-Stephens (geometric/arithmetic mean SGCDF):*

$$f_{HS}(\Omega(\boldsymbol{x})) \overset{\text{def}}{=} \prod_{i=1}^{p} \sigma_i(A)^2 - \alpha \left[\sum_{i=1}^{p} \sigma_i(A)^2\right]^p \tag{2.33}$$

where $\alpha \in \left[0, \frac{1}{p^p}\right]$ *is a user supplied constant.*

- *Generalized Rohr (geometric mean SGCDF):*

$$f_R(\Omega(\boldsymbol{x})) \overset{\text{def}}{=} \left[\prod_{i=1}^{p} \sigma_i(A)^2\right]^{\frac{1}{p}} \tag{2.34}$$

- *Generalized Noble-Förstner (harmonic mean SGCDF):*

$$f_{NF}(\Omega(\boldsymbol{x})) \overset{\text{def}}{=} \begin{cases} \frac{1}{\sum_{i=1}^{p} \frac{1}{\sigma_i(A)^2}} & \textit{if } \sigma_i(A) \neq 0 \textit{ for every } i, \\ 0 & \textit{otherwise.} \end{cases} \tag{2.35}$$

- *Generalized Shi-Tomasi:*

$$f_{ST}(\Omega(\boldsymbol{x})) \overset{\text{def}}{=} \sigma_p(A)^2 \tag{2.36}$$

- *Kenney (condition number based SGCDF):*

$$f_{K,q}(\Omega(\boldsymbol{x})) \overset{\text{def}}{=} \begin{cases} \frac{1}{\left[\sum_{i=1}^{p} \frac{1}{\sigma_i(A)^{2q}}\right]^{\frac{1}{q}}} & \textit{if } \sigma_i(A) \neq 0 \textit{ for every } i, \\ 0 & \textit{otherwise.} \end{cases} \tag{2.37}$$

We will now present some general properties related to the analytical structure of the detector functions introduced above. This preliminary discussion serves to simplify the presentation of the results of Section 2.5.3, that target some relevant specific issues arising in the context of corner detection for multichannel generalized images.

Detector Structure

The Harris-Stephens, the Rohr and the Noble-Förstner Spectral Generalized Corner Detector Function (SGCDF) are closely connected to the so called Pythagorean means [49]:

$$\begin{array}{llll} M_A(x_1,\ldots,x_p) & \overset{\text{def}}{=} & \frac{1}{p}\sum_{i=1}^p x_i & \text{Arithmetic Mean} \\ M_G(x_1,\ldots,x_p) & \overset{\text{def}}{=} & \left(\prod_{i=1}^p x_i\right)^{\frac{1}{p}} & \text{Geometric Mean} \\ M_H(x_1,\ldots,x_p) & \overset{\text{def}}{=} & \frac{p}{\sum_{i=1}^p \frac{1}{x_i}} & \text{Harmonic Mean} \end{array}$$

In fact, from the previous definitions, we can derive the following identities:

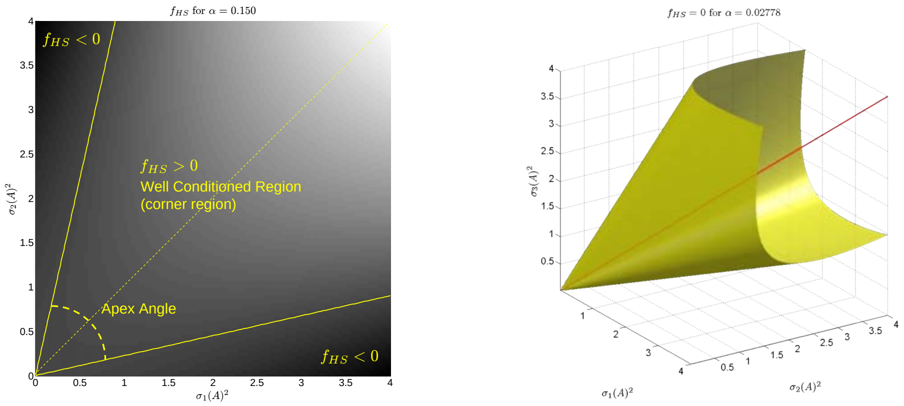
$$\begin{aligned} f_{HS}(\Omega(\boldsymbol{x})) &= M_G\left(\sigma_1^2,\ldots,\sigma_p^2\right)^p - \alpha p^p M_A\left(\sigma_1^2,\ldots,\sigma_p^2\right)^p \\ f_R(\Omega(\boldsymbol{x})) &= M_G\left(\sigma_1^2,\ldots,\sigma_p^2\right) \\ f_{NF}(\Omega(\boldsymbol{x})) &= \frac{1}{p}M_H\left(\sigma_1^2,\ldots,\sigma_p^2\right) \end{aligned}$$

When the matrix A is full rank, the Kenney's SGCDF is obtained by calculating the inverse of condition number estimate (B.1) of the GGM utilizing the Schatten q-norm (see (A.4)):

$$f_{K,q}(\Omega(\boldsymbol{x})) = \frac{1}{\|A^\dagger\|_{S,2q}^2}$$

The structure of the Rohr and of the Noble-Förstner SGCDFs will be motivated using the equivalence relations that will be presented in the next section.

We will now focus on the interpretation of the Harris-Stephens SGCDF. When $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ is a pure translation and $n = 2$ (and consequently $p = 2$), the space of the squared singular values is partitioned as shown in Figure 2.8. In his paper [50], Harris explains how the detector was designed to be invariant with respect to



(a) (b)

Figure 2.8: Figure (a) shows the amplitude of the Harris-Stephens SGCDF for a pure translation in the case $n = 2$. Brighter values indicate a larger response. The continuous yellow lines identify the loci defined by $f_{HS} = 0$. Figure (b) shows the level set surface $f_{HS} = 0$ of the Harris-Stephens SGCDF in the case $n = 3$. The red line is generated by the vector $\boldsymbol{d} = [1 \ \dots \ 1]^T$.

rotations, to have a small computational complexity[7] and to properly partition the space of the eigenvalues of $A^T A$:

> R [using our notation $R = f_{HS}$] is positive in the *corner region*, negative in the *edge regions* and small in the *flat regions*. Note that increasing the contrast (i.e. moving radially away from the origin) in all cases increases the magnitude of the response. The flat region is specified by Tr [the trace of $A^T A$] falling below some selected threshold.

From the more general point of view presented in Section 2.4, the Harris SGCDF partitions the space of the singular values of A so that the region around the axis of the cone defined by the condition $f_{HS} \geq 0$ is associated with neighborhoods that are well conditioned. We conjecture that these considerations can be extended to

[7] Calculating just the trace and determinant of the gradient normal matrix it is possible to avoid its explicit eigen-decomposition.

a generic transformation for an arbitrary dimension of the pixel space: well conditioned neighborhoods will be characterized by a set of singular values that lie near the axis of the cone defined by the condition $f_{HS} \geq 0$. Our conjecture is valid provided that the locus $F(\alpha) = \{\boldsymbol{\lambda} \in \mathbb{R}^p : f_{HS}(\boldsymbol{\lambda}) \geq 0\}$ defines a cone-like structure with apex at the origin and axis aligned with the vector $\boldsymbol{d} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T$ (for the case $p = 3$ see Figure 2.8(b)). Such locus is not empty if and only if:

$$\alpha \leq \frac{M_G\left(\sigma_1^2, \dots, \sigma_p^2\right)^p}{p^p M_A\left(\sigma_1^2, \dots, \sigma_p^2\right)^p} \tag{2.38}$$

This happens when α belongs to the interval $\left[0, \frac{1}{p^p}\right]$ (as specified in the formal definition of the Harris-Stephens SGCDF), as a consequence of the inequality between arithmetic and geometric means (see [49], p. 17). Such a result states that the arithmetic mean of a set of non-negative real numbers is greater or equal than the geometric mean of the same set of numbers and that equality is achieved if and only if all the number in the set are equal. This said, our conjecture can be formalized as:

Conjecture 2.5.6 (Relation Between α and ϕ) *Consider the hypercone* $C(2\phi) = \left\{\boldsymbol{\lambda} \in \mathbb{R}^p : \boldsymbol{d}^T\boldsymbol{\lambda} - \|\boldsymbol{\lambda}\|\|\boldsymbol{d}\|\cos(\phi) \geq 0\right\}$*, where* $\boldsymbol{d} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T$ *and* 2ϕ *is the apex angle. Then there exist* $\alpha^* \in \left[0, \frac{1}{p^p}\right]$ *(depending on* ϕ*) such that for any* $\varepsilon > 0$ *arbitrarily small we have:* $C(2\phi) \subseteq F(\alpha^*)$ *and* $C(2\phi + \varepsilon) \nsubseteq F(\alpha^*)$*.*

We believe that a proof of Conjecture 2.5.6 for $p > 2$ might be quite involved and that its discussion would take us too far afield. However we would like to gain a deeper understanding regarding the relation between the angle of the apex of the hypercone and the parameter α. To achieve this goal we will construct a point $\boldsymbol{\lambda}_C$ that lies on the surface of the cone $C(2\phi)$ and we will show how, when ϕ becomes larger, $F(\alpha)$ encloses such point only if α becomes smaller. Let's consider the point $\boldsymbol{\lambda}_C = \begin{bmatrix} \mu & t & \dots & t \end{bmatrix}^T$. This point belongs to the surface of $C(2\phi)$ if and only if $\boldsymbol{d}^T\boldsymbol{\lambda} - \|\boldsymbol{\lambda}\|\|\boldsymbol{d}\|\cos(\phi) = 0$. This constraint allows μ to take only two values:

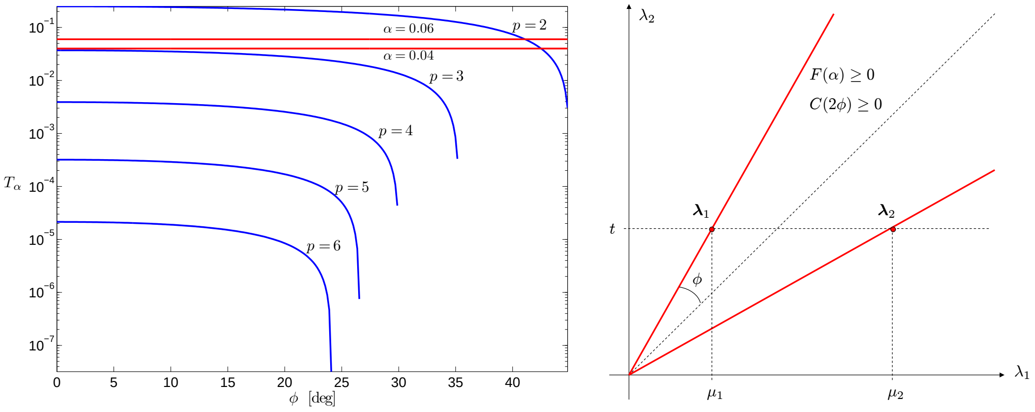
$$\mu_{1,2} = \frac{p-1 \pm p\sqrt{p-1}\cos(\phi)\sin(\phi)}{p\cos(\phi)^2 - 1}\, t$$

Interestingly enough, in the above expression t factors out, and therefore our considerations will be valid not just for the points $\boldsymbol{\lambda}_{C_{1,2}}$ but for all the points that have the same direction of $\boldsymbol{\lambda}_{C_1}$ or $\boldsymbol{\lambda}_{C_2}$. If we let $\gamma = p\sqrt{p-1}\sin(\phi)$ to simplify the notation, the condition (2.38) that enforces both $\boldsymbol{\lambda}_{C_{1,2}}$ to be contained inside $F(\alpha)$ can be written as $\alpha \leq T_\alpha$, where $T_\alpha = \min\{T_{\alpha_1}, T_{\alpha_2}\}$ and:

$$\begin{aligned}
T_{\alpha_2} &= \frac{[p-1+p\gamma\cos(\phi)]\left[p\,\cos(\phi)^2-1\right]^{p-1}}{p^p\left[p^2\cos(\phi)-p\cos(\phi)+\gamma\right]^p\cos(\phi)^p} \\
T_{\alpha_2} &= \frac{[p-1-p\gamma\cos(\phi)]\left[p\,\cos(\phi)^2-1\right]^{p-1}}{p^p\left[p^2\cos(\phi)-p\cos(\phi)-\gamma\right]^p\cos(\phi)^p}
\end{aligned}$$

Figure 2.9 shows the behavior of T_α as a function of the cone angle 2ϕ for different values of p. As suggested by our conjecture, when the cone angle grows larger the value of α must decrease to have $C(2\phi)$ contained in $F(\alpha)$. The graphs seem to suggest that when $p > 2$ the apex angle of the cones that can be contained in $F(\alpha)$ is limited by a value that is less than 90 degrees. However this is not true: when $\alpha = 0$ the locus $F(\alpha)$ contains all the points with non negative components and the detector response is always positive. Finally note that when $\phi = 0$, α is equal to $\frac{1}{p^p}$. Even if this analysis is not general, it definitely points to the validity of the conjecture.

Example 2.5.7 (Harris-Stephens Detector Sensitivity) *In this example we will specialize the previous observations for the case where $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ models a trans-*



(a) (b)

Figure 2.9: Figure (a) shows the curve that relate the threshold T_α to the cone apex angle 2ϕ. Figure (b) illustrates the situation studied in Example 2.5.7.

lation and $n = 2$. *It can be shown that the locus* $f_{HS} = 0$ *is indeed a cone and therefore there exists a one to one relation between* α *and* ϕ. *More specifically the values for* μ *are:*

$$\mu_{1,2} = \frac{1 \pm \sin(2\phi)}{\cos(2\phi)}\, t$$

and:

$$\alpha = T_{\alpha_1} \equiv T_{\alpha_2} = \frac{\cos(2\phi)}{4\cos(\phi)^2}$$

The above expression can be inverted yielding:

$$\phi = \arccos\left(\frac{1}{\sqrt{2}\sqrt{2\alpha+1}}\right)$$

The empirically arrived values suggested in the literature for α *are in the range* $[0.04, 0.06]$. *This approximately translates in a cone angle* $2\phi \in [82^\circ, 85^\circ]$. *Therefore smaller values of* α *make the Harris-Stephens detector more sensitive by enlarging the region that is associated with corner-like structures (i.e.* $f_{HS} \geq 0$, *see Figure 2.9(b)).*

Detector Equivalence Relations

The SGCDFs just introduced are not independent one from the other; it can be shown that two of them (the Noble-Förstner SGCDF and the Shi-Tomasi SGCDF) are actually equivalent to Kenney's SGCDF modulo an appropriate choice of the matrix norm used to estimate the condition number. Similar considerations hold for the Rohr's SGCDF, even though the relation with Kenney's SGCDF is more involved and is valid only in a limit sense. The following theorem expresses analytically such relations:

Theorem 2.5.8 (Generalized Detectors Equivalence Relations) *The following interesting relations hold among the* SGCDF (2.34)*,* (2.35)*,* (2.36) *and* (2.37)*:*

- *Generalized Rohr equivalence:* $\lim_{q\to 0} \sqrt[q]{p} f_{K,q} = f_R$
- *Generalized Noble-Förstner equivalence:* $f_{K,1} = f_{NF}$
- *Generalized Shi-Tomasi equivalence:* $f_{K,\infty} = f_{ST}$

Proof: We begin our proof showing the generalized Rohr equivalence. The following chain of equations hold:

$$\lim_{q\to 0} \sqrt[q]{p} f_{K,q} = \lim_{q\to 0} \frac{\sqrt[q]{p}}{\left[\sum_{j=1}^{p} \frac{1}{\sigma_j(A)^{2q}}\right]^{\frac{1}{q}}} = \frac{1}{\lim_{q\to 0} \frac{1}{p^{\frac{1}{q}}} \left[\sum_{j=1}^{p} \frac{1}{\sigma_j(A)^{2q}}\right]^{\frac{1}{q}}}$$

Because of the result presented in [49], (p. 15), which states that:

$$\lim_{q\to 0} \left(\frac{1}{p}\sum_{j=1}^{p} x_j^q\right)^{\frac{1}{q}} = \sqrt[p]{\prod_{j=1}^{p} x_j}$$

we can write:

$$\frac{1}{\lim_{q\to 0}\left[\frac{1}{p}\sum_{j=1}^{p}\frac{1}{\sigma_j(A)^{2q}}\right]^{\frac{1}{q}}} = \frac{1}{\sqrt[p]{\prod_{j=1}^{p}\frac{1}{\sigma_j(A)^2}}} = \left[\prod_{j=1}^{p}\sigma_j(A)^2\right]^{\frac{1}{p}}$$

which proves that $\lim_{q\to 0}\sqrt[q]{p}f_{K,q} = f_R$. As far as the generalized Noble-Förstner equivalence is concerned, note that $f_{K,1} = \frac{1}{\sum_{j=1}^{p}\frac{1}{\sigma_j(A)^2}}$. Finally the last equivalence follows from the fact that $\lim_{q\to\infty}\|A^\dagger\|_{S,2q}^2 = \sigma_{max}(A^\dagger)^2 = \frac{1}{\sigma_{min}(A)^2}$ (as explained in Appendix A.2.1). ■

Because of the equivalence relations established in the previous theorem, we will introduce the following definition:

Definition 2.5.9 *A (local) SGCDF that is derived from $f_{K,q}$ for any value of $q \geq 1$ is called condition number based (henceforth Condition Number Based (CNB)) SGCDF.*

Neither the Harris SGCDF nor the Rohr SGCDF are Condition Number Based (CNB) SGCDF. It is straightforward to verify that there does not exist a value of $q > 0$ such that $f_{HS} = f_{K,q}$ (in fact for any positive α the Harris-Stephens SGCDF can become negative). As far as f_R is concerned, recall that in the proof of Theorem 2.5.8 it was shown that:

$$\lim_{q\to 0}\frac{\sqrt[q]{p}}{\|A^\dagger\|_{S,2q}^2} = \frac{1}{\sqrt[p]{\prod_{j=1}^{p}\sigma_j(A^\dagger)^2}}$$

The quantity that appears at the denominators of the left hand side of the previous equation (i.e. the inverse of the geometric mean of the singular values of $A^\dagger$) is not a norm since it becomes zero when any of the singular values is zero. Moreover the triangle inequality is not satisfied.[8]

[8] Actually the geometric mean always satisfies the reversed triangle inequality: for any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ with positive components $M_G(\boldsymbol{u}+\boldsymbol{v}) \geq M_G(\boldsymbol{u}) + M_G(\boldsymbol{v})$.

Analytical Bounds

The GGM $A(\Omega(\boldsymbol{x}))$ depends on the type of geometric transformation that we are considering (see (2.17)). It seems reasonable to expect that transformations $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ that are more complex (i.e. described by more parameters) will produce SGCDF responses that have a smaller value. In fact, if we adopt the perspective introduced in Section 2.4.1, we expect that *all the parameters* of the transformation will be modified in the attempt to compensate for the effects of noise, hence producing a larger value for $\|\Delta\boldsymbol{\theta}\|$. Similarly, if we model the problem as we did in Section 2.4.2, it is intuitively clear that if we use the same amount of information, the stability of the estimate of a higher dimensional parameter vector will worsen. These empirical considerations are quantitatively formalized in the following theorem:

Theorem 2.5.10 (Generalized Detector Bounds) *Let n be the pixel dimension of the generalized image $\boldsymbol{I}$ and consider the translational transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{b} \in \mathbb{R}^n$, where $\boldsymbol{\theta} = \boldsymbol{b}$ and the affine transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{x} + B(\boldsymbol{y} - \boldsymbol{x}) + \boldsymbol{b}$, where $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{b}^T & \mathrm{vec}\,(B)^T \end{bmatrix}^T \in \mathbb{R}^{n(n+1)}$ and $\mathrm{vec}\,(B)$ returns the vector formed by columns of A stacked one upon the other. Then for any neighborhood $\Omega(\boldsymbol{x})$, any CNB SGCDF satisfies the inequality:*

$$f_{K,q}^{Translation} \geq f_{K,q}^{Affine} \tag{2.39}$$

Proof: This proof can be regarded as a constructive method to identify bounds for the response of a SGCDF for a given transformation. The Jacobian for the translational transformation is the identity matrix I_n whereas the Jacobian for the affine transformation (that is parameterized by a $p = n(n+1)$ dimensional vector) can be written as:

$$J_{\boldsymbol{\theta}}\boldsymbol{T}^{Affine}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}_i) = \begin{bmatrix} I_n & (y_{i,1}-x_{i,1})I_n & \dots & (y_{i,n}-x_{i,n})I_n \end{bmatrix} \in \mathbb{R}^{n\times p}$$

Consequently the i^{th} block of the GGM in the affine case becomes:

$$A^{Affine}(\boldsymbol{y}_i) = \begin{bmatrix} J\boldsymbol{I}(\boldsymbol{y}_i) & (y_{i,1}-x_{i,1})J\boldsymbol{I}(\boldsymbol{y}_i) & \dots & (y_{i,n}-x_{i,n})J\boldsymbol{I}(\boldsymbol{y}_i) \end{bmatrix} \in \mathbb{R}^{m\times p}$$

Therefore the matrix $A^{Translation}(\boldsymbol{y}_i)$ is obtained from $A^{Affine}(\boldsymbol{y}_i)$ by removing the last n^2 columns and the same consideration applies to the matrices $A^{Translation}(\Omega(\boldsymbol{x}))$ and $A^{Affine}(\Omega(\boldsymbol{x}))$. Using Inequality (A.6) from Theorem A.2.4 and passing to the reciprocals we can write:

$$\begin{array}{ccc} \dfrac{1}{\sigma_{1+n^2}} & \geq & \dfrac{1}{\sigma^{(n^2)}_{c,1}} \\ \dfrac{1}{\sigma_{2+n^2}} & \geq & \dfrac{1}{\sigma^{(n^2)}_{c,2}} \\ \vdots & & \vdots \\ \dfrac{1}{\sigma_{n+n^2}} & \geq & \dfrac{1}{\sigma^{(n^2)}_{c,n}} \end{array}$$

Summing the n corresponding members of the previous inequalities raised to the power $2q$ we get:

$$\sum_{i=1}^{n}\left(\frac{1}{\sigma_{i+n^2}}\right)^{2q} \geq \sum_{i=1}^{n}\left(\frac{1}{\sigma^{(n^2)}_{c,i}}\right)^{2q}$$

The left hand side of the previous equation can be augmented considering the first largest n^2 singular values of $A^{Affine}(\boldsymbol{y}_i)$:

$$\sum_{i=1}^{n^2}\left(\frac{1}{\sigma_i}\right)^{2q} + \sum_{i=1}^{n}\left(\frac{1}{\sigma_{i+n^2}}\right)^{2q} = \sum_{i=1}^{p}\left(\frac{1}{\sigma_p}\right)^{2q} \geq \sum_{i=1}^{n}\left(\frac{1}{\sigma^{(n^2)}_{c,i}}\right)^{2q}$$

It is straightforward to recognize that the previous inequality can be rewritten as:

$$\|A^{Affine,\dagger}(\Omega(\boldsymbol{x}))\|_{S,2q} \geq \|A^{Translation,\dagger}(\Omega(\boldsymbol{x}))\|_{S,2q}$$

and taking the reciprocals of both members we conclude that $f_{K,q}^{Affine} \leq f_{K,q}^{Translation}$.

■

This theorem holds neither for the Harris-Stephens SGCDF nor for the Rohr SGCDF. Using Inequalities (A.6) we can write:

$$\prod_{i=1}^{n} \sigma_{i+n^2} \geq \prod_{i=1}^{n} \sigma_{c,i}^{(n^2)}$$

but in general it is not possible to augment the left hand side of the previous inequality utilizing all the singular values of A^{Affine}. Similar considerations can be extended to the Harris-Stephens detector, where the situation is complicated by the presence of the difference between the geometric and the arithmetic mean.

Theorem 2.5.10 confirms the intuition that for any given region $\Omega(\boldsymbol{x})$, the higher is the complexity of the transformation used either to model the effects of noise (see Section 2.4.1) or to model the transformation of the intensity (as discussed in Section 2.4.2), the higher is the corresponding condition number and consequently the smaller the SGCDF response. This result is important from the computational viewpoint, in order to reduce the complexity associated with the calculation of the SGCDF when the transformation is characterized by a large number of parameters. In fact, as stated in (2.39), a neighborhood that yields a small SGCDF response with respect to a translation will have a smaller response for an affine transformation. Hence it makes sense to calculate the response of the SGCDF $f_{K,q}^{Affine}$ (high computational complexity task) only at the points where $f_{K,q}^{Translation}$ attains its maxima (small computational complexity task). We finally would like to emphasize how this theorem provides a partial answer to the following observation by Triggs (quoted from [129], p. 10):

> The main observation is that different models often select different keypoints, and more invariant models generate fewer of them, but beyond this it is difficult to find easily interpretable systematic trends.

Computational Complexity

All the SGCDFs relies on the calculation of the singular values of the GGM $A(\Omega(\boldsymbol{x}))$. In general this task is achieved using an iterative algorithm that requires about $2p^2(mN+p)$ floating point operations ([44], p. 254). As shown in Lemma A.2.1, $\sigma_j(A)^2 = \lambda_j(A^TA)$, and therefore when $p \ll mN$ we can take advantage of those algorithms that can diagonalize a symmetric matrix with a complexity $O(p^3)$. Moreover, $\text{trace}(A^TA) = \sum_{j=1}^{p} \sigma_j(A)^2$ and $\det(A^TA) = \left[\prod_{j=1}^{p} \sigma_j(A)\right]^2$, and therefore for the Harris-Stephens and Rohr SGCDF we can avoid the explicit calculation of the singular values of A at the price of computing the determinant of the generalized gradient normal matrix A^TA.

2.5.3 Properties of the Generalized Corner Detectors

In this section we will enunciate a set of relevant properties of the SGCDFs. Some of them are satisfied only in case $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ models a pure translation. Even with this limitation these properties are quite relevant, especially in the light of Theorem 2.5.10, that states that the responses of the SGCDFs associated with a pure translation provide an upper bound for the values that can be attained by the SGCDFs associated with more complex transformations.

Definition 2.5.11 *Given two matrices S_1 and S_2 compatibly dimensioned, we write $S_1 \leq S_2$ if $S_2 - S_1$ is positive semi-definite. That is, for any vector $\boldsymbol{v}$ we have $\boldsymbol{v}^T(S_2 - S_1)\boldsymbol{v} \geq 0$.*

Definition 2.5.12 *Let $\sigma_1, \ldots, \sigma_n$ be the eigenvalues of $A(\Omega(\boldsymbol{x}))$. We say that a set of points $\mathcal{X}$ in the image $\boldsymbol{I}$ has constant eigen-energy with respect to the q-Schatten norm if $\|A(\Omega(\boldsymbol{x})\|^2_{S,2q}$ is constant for every $\boldsymbol{x} \in X$.*

Definition 2.5.13 *A point $\boldsymbol{x}$ is isotropic (with respect to the image $\boldsymbol{I}$) if the singular values of the* GGM *are all equal: $\sigma_1 = \ldots = \sigma_p$.*

Rotation Invariance

A desirable property for a SGCDF is to be invariant with respect to rotations and reflections of a generalized image. The next lemma provides a sufficient condition for a SGCDF to be invariant with respect to orthogonal transformations (used to model rotations or rotoinversions, i.e. a rotation followed by a flip).

Lemma 2.5.14 (Rotation Invariance) *Any* SGCDF *that is associated with a translational transformation and that depends solely on the singular values of the* GGM *is invariant with respect to rotations and reflections of image.*

Proof: Consider two images $\boldsymbol{I}$ and $\boldsymbol{I}'$ related by a rotation and possibly a reflection, i.e. $\boldsymbol{I}(\boldsymbol{x}) = \boldsymbol{I}'(\boldsymbol{x}') = \boldsymbol{I}'(U\boldsymbol{x})$, where U is an orthogonal matrix in $\mathbb{R}^{n\times n}$. From (2.30) we can write $J\boldsymbol{I}(\boldsymbol{y}) = J\boldsymbol{I}'(\boldsymbol{y}')\ U$ and since $J\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}} = I_n$, the blocks (2.17) that compose the GGM can be written as:

$$A(\boldsymbol{y}) = w(\boldsymbol{y} - \boldsymbol{x})J\boldsymbol{I}(\boldsymbol{y}) = w(U^T(\boldsymbol{y} - \boldsymbol{x}))J\boldsymbol{I}'(\boldsymbol{y}')U = w(\boldsymbol{y}' - \boldsymbol{x}')A'(\boldsymbol{y}')$$

(assuming that the weighting function has a radial symmetry). Hence we can also write $A(\Omega(\boldsymbol{x})) = A'(\Omega'(\boldsymbol{x}'))U$ and the assertion follows from the observation that the singular values are invariant under the action of orthogonal matrices. ■

All the SGCDFs defined in 2.5.5 satisfy the previous lemma. This invariance property in general does not hold when the transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ is different from a pure translation. In this case $J\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ is different from the identity and therefore it is not possible to factor out from the GGM an orthogonal matrix.

Monotonicity

The spectrum of the gradient normal matrix encodes the information regarding the sensitivity of a neighborhood in the presence of noise: the larger the singular values the smaller the sensitivity to noise. This property is captured in the monotonic behavior of the SGCDFs.

Lemma 2.5.15 (Monotonicity) *The Rohr SGCDF and any CNB SGCDF are non decreasing in $\sigma_1(A), \dots, \sigma_p(A)$.*

Proof: By convenient abuse of notation let's make explicit the dependence of the SGCDF f from the singular values of the GGM; we need to show that the inequality:

$$f(\sigma_1, \dots, \sigma_i, \dots, \sigma_p) \leq f(\sigma_1, \dots, \sigma_i + \delta, \dots, \sigma_p) \tag{2.40}$$

holds for an arbitrary index i and for any $\delta > 0$. Let's start with the case $f = f_R$. It is simple to verify that for every i we have:

$$\frac{\partial f_R}{\partial \sigma_j} = \frac{2 f_R}{p \sigma_j} > 0$$

and therefore the non decreasing property holds with strict inequality. Let's now consider the case $f = f_{K,q}$. Since the q-Schatten norm of the GGM is defined as the q-norm of the vector composed of its singular values and since the vector q-norm is absolute (i.e. it depends only on the absolute values of the vector entries) then it is also monotone (see [53], p. 285). Therefore the assertion (2.40) follows from the inequality:

$$\left\| \begin{bmatrix} \frac{1}{\sigma_1} & \dots & \frac{1}{\sigma_i} & \dots & \frac{1}{\sigma_p} \end{bmatrix}^T \right\|_{2q} \geq \left\| \begin{bmatrix} \frac{1}{\sigma_1} & \dots & \frac{1}{\sigma_i + \delta} & \dots & \frac{1}{\sigma_p} \end{bmatrix}^T \right\|_{2q}$$

■

To understand the meaning of the previous lemma let's consider the SGCDF associated with a pure translation and let's recall that $\sigma_i(A)^2 = \lambda_i(A^T A)$. Then the matrix $A^T A$ provides a measure of both the strength of the intensity gradients and their independence. This can be encapsulated by the natural ordering on symmetric matrices formally defined in 2.5.11. Thus if we have two neighborhoods $\Omega_1(\boldsymbol{x}_1)$ and $\Omega_2(\boldsymbol{x}_2)$ such that $A_1^T A_1 > A_2^T A_2$ (where $A_1 = A(\Omega_1(\boldsymbol{x}_1))$ and $A_2 = A(\Omega_2(\boldsymbol{x}_2))$), then the condition expressed in Lemma 2.5.15 means that the gradient vectors at $\boldsymbol{x}_2$ are stronger and/or more independent than those at $\boldsymbol{x}_1$. Similar considerations hold when $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ models other transformations: in this case the matrix $A^T A$ encodes the level of strength/independence of the intensity gradients with respect to the transformation parameters. The Harris-Stephens SGCDF does not satisfy the non decreasing property as it can be shown by contradiction. If we assume that f_{HS} is non decreasing, then for every index i we have that

$\frac{\partial f_{HS}}{\partial \lambda_i} \geq 0$ (where $\lambda_i = \sigma_i^2$) and therefore:

$$\prod_{\substack{j=1 \\ j \neq i}}^{p} \lambda_j \geq \alpha p \left(\sum_{j=1}^{p} \lambda_j \right)^{p-1}$$

Since both the left and the right hand side of the previous equation are positive, we can multiply the left hand side and the right hand side of each inequality for all $1 \leq i \leq p$ obtaining:

$$\left(\prod_{j=1}^{p} \lambda_j \right)^{p-1} \geq \alpha^p p^p \left(\sum_{j=1}^{p} \lambda_j \right)^{p(p-1)} \tag{2.41}$$

and consequently:

$$0 < \alpha \leq \frac{\left(\prod_{j=1}^{p} \lambda_j \right)^{\frac{p-1}{p}}}{p \left(\sum_{j=1}^{p} \lambda_j \right)^{p-1}} = \frac{1}{p} \left[\frac{M_G(\lambda_1, \ldots, \lambda_p)}{M_A(\lambda_1, \ldots, \lambda_p)} \right]^{p-1} \tag{2.42}$$

From the previous equation[9] it follows that there exist values of α such that the partial derivatives of f_{HS} are not positive. This contradicts our initial assumption regarding the fact that the Harris-Stephesn SGCDF satisfies the non decreasing property.

Example 2.5.16 *In this example we will apply the previous analysis when $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ models a translation, $n = 2$ and $m = 1$ (i.e. we are dealing with a simple single channel image). Consider Figure 2.10 that shows two image neighborhoods containing a corner point. Let $\Omega_1(\boldsymbol{x}_1)$ and $\Omega_2(\boldsymbol{x}_2)$ be the circular neighborhoods (represented by the nodes of the discretization grid). Let Δ_i be the intensity variation and ϕ_i be the angle (between the segments that define the partition between*

[9] Note that the upper bound is not tight: this is a consequence of the method we used to construct the inequality (2.41).

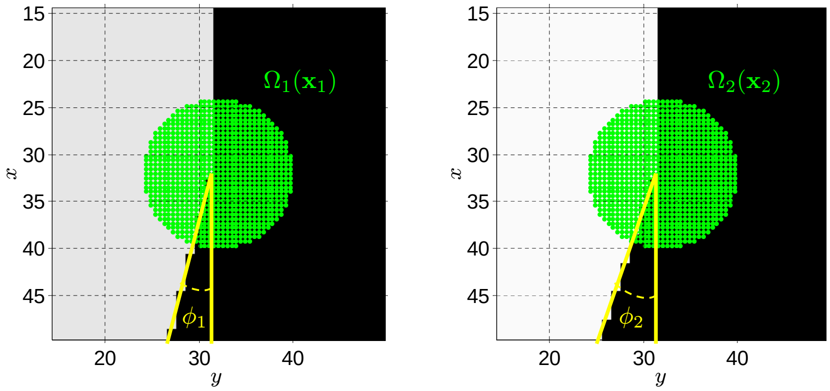


Figure 2.10: The images show two corner structures and the discretization lattice used to compute the SGCDF response. The left structure exhibits a "weaker" corner structure: in fact both the intensity difference and the angle ϕ_1 are smaller than those of the right structure.

the dark and the bright portions of the image) in the neighborhood $\Omega_i(\boldsymbol{x}_i)$ *(where* $i = 1, 2$*). Since* $\Delta_1 > \Delta_2$ *and* $\phi_1 > \phi_2$ *we can qualitatively say that the corner structure contained in the second neighborhood is more "strong" that the structure contained in the first one. This is confirmed quantitatively by the evaluation of the spectrum of the corresponding gradient normal matrices:*

$$\lambda(A_1^T A_1) = \{3.5804, 0.0674\} \qquad \lambda(A_2^T A_2) = \{4.1789, 0.1175\}$$

The eigenvalues of $A_2^T A_2$ *are larger than the corresponding eigenvalues of* $A_1^T A_1$*. In order for the Harris-Stephens* SGCDF *to be nondecreasing, the inequality* (2.42) *states that* α *must belong to the interval* $(0, 0.0673]$*. If we pick* α *out of this interval (say* $\alpha = 0.07$*) we obtain that:*

$$f_{HS}(\Omega_1(\boldsymbol{x}_1)) = -6.9014 \geq -8.0115 = f_{HS}(\Omega_2(\boldsymbol{x}_2))$$

In other words the Harris-Stephens SGCDF *yields a larger response for the neighborhood that contains a weaker corner structure.*[10] *This fact is undesirable, espe-*

[10] The negative value attained by the detector is in agreement with the plot displayed in Figure

cially in a context that requires a preliminary pruning of the points based on the magnitude of the SGCDF response. Finally we want to remark that this problem cannot be solved just by taking the magnitude of f_{HS}. As it can be inferred from Figure 2.8(a), considering $|f_{HS}|$ would cause points associated either to a small λ_1 or to a small λ_2 to produce responses with the same value of the responses produced by eigenvalue pairs inside the cone $f_{HS} > 0$ (in other words a corner would be confused with an edge).

Isotropy

There exist an infinite number of GGMs (i.e. an infinite number of neighborhoods) characterized by a set of singular values that have the same eigen-energy (see Definition 2.5.12). It turns out that the SGCDF introduced in this chapter attain their maximum value when *all* the singular values of the GGM *are the same.*

Lemma 2.5.17 (Isotropy) *The Harris-Stephens, Rohr and any CNB SGCDF calculated over a set of eigen-energy points (defined as in 2.5.12) attain their maximum value at a point of isotropy (defined as in 2.5.13).*

Proof: To prove the lemma we need to solve a constrained optimization that can be treated as a Lagrange multiplier problem (see A.3.1 and for a more thorough discussion of the topic see [9]). With the customary abuse of notation, we will express the dependence of the SGCDF and of the constraint in terms of the eigenvalues of the generalized gradient normal matrix A^TA, recalling that

2.8(a).

$\lambda_i(A^T A) = \sigma_i(A)^2$. Our goal is to maximize $f(\lambda_1, \ldots, \lambda_p)$ subject to the eigen-energy constraint $h(\lambda_1, \ldots, \lambda_p) = \left(\sum_{j=1}^p \lambda_j^q\right)^{\frac{1}{q}} = c$, where $c > 0$ is an arbitrary constant, and $\lambda_i > 0$ for every i. We seek for the candidate solutions at the stationary points of the Lagrangian $F(\lambda_1, \ldots, \lambda_p) = f + \gamma(h - c)$, where the unknown constant γ is the Lagrange multiplier .

The proof of the isotropy property for the Harris-Stephens is quite involved and therefore is carried out separately in Lemma A.3.3.

Let's now focus on the Rohr SGCDF. To simplify the notation, let's define the following quantities: $P \overset{\text{def}}{=} \prod_{j=1}^p \lambda_j$ and $S_q \overset{\text{def}}{=} \sum_{j=1}^p \lambda_j^q$, so that $\|\boldsymbol{\lambda}\|_q = S_q^{\frac{1}{q}} = c$. The components of the gradient of the Lagrangian are:

$$\frac{\partial F}{\partial \lambda_i} = \frac{P^{\frac{1}{p}}}{p\lambda_i} + \gamma S^{\frac{1-q}{q}} \lambda_i^{q-1}$$

and by equating all the gradient components to zero we obtain that:

$$\lambda_i = \frac{P^{\frac{1}{pq}}}{(-\gamma p)^{\frac{1}{q}} c^{\frac{1-q}{q}}}$$

Since the right hand side of the previous equations does not depend on the index i, then the gradient of the Lagrangian vanishes when all the eigenvalues of the generalized normal matrix are equal.

Finally let's consider the case of a CNB SGCDF, i.e. $f_{K,q} = \left(\sum_{j=1}^p \lambda_j^{-q}\right)^{-\frac{1}{q}}$. The components of the gradient of F are given by:

$$\frac{\partial F}{\partial \lambda_i} = S_{-q}^{-\frac{1+q}{q}} \lambda_i^{-q-1} + \gamma S^{\frac{1-q}{q}} \lambda_i^{q-1}$$

As we did before, by forcing the gradient components to zero we get that all the eigenvalues must be equal, since:

$$\lambda_i = \left(-\frac{S_{-q}^{-\frac{1+q}{q}}}{\gamma c^{1-q}}\right)^{\frac{1}{2q}}$$

■

Also in this case it is useful to initially consider $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ to be a pure translation and to analyze the generalized gradient normal matrix $A^T A$. If the matrix $A^T A$ has a large value of $\boldsymbol{v}^T A^T A \boldsymbol{v}$ for a vector $\boldsymbol{v}$ then it is well-conditioned for point matching with respect to translational shifts from $\boldsymbol{x}$ in the direction $\boldsymbol{v}$. As a directional vector $\boldsymbol{v}$ moves over the n-dimensional unit sphere the values of $\boldsymbol{v}^T A^T A \boldsymbol{v}$ pass through all the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A^T A$. This means that if one eigenvalue is smaller than the others, then the corresponding eigenvector $\boldsymbol{v}$ is a direction in which the corner is less robust than in the other eigenvector directions (in the sense of point matching conditioning, see Section 2.4.2). From this we see that Lemma 2.5.17 can be interpreted as the property that the SGCDF subject to the restriction of constant eigen-energy attains its maximum if the neighborhood identifies a corner that doesn't have a weak direction: all the unit norm directional vectors $\boldsymbol{v}$ yield the same value for $\boldsymbol{v}^T A^T A \boldsymbol{v}$. That is, we must have $\lambda_1 = \ldots = \lambda_n$. Similar considerations extend to the case where $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ models a transformation that is more general than a pure translation.

Neighborhood Restriction

The next lemma describes what happens to a corner detector when its response is calculated on a restriction of the original point neighborhood $\Omega(\boldsymbol{x})$.

Lemma 2.5.18 (Neighborhood Restriction) *Consider a neighborhood $\omega(\boldsymbol{x}) \subseteq \Omega(\boldsymbol{x})$. Then for the Rohr SGCDF and for any CNB SGCDF the following inequality holds:*

$$f(\Omega(\boldsymbol{x})) \geq f(\omega(\boldsymbol{x}))$$

Proof: Let's consider a set of points $\mathcal{Y} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_N\}$ that sample the region $\Omega(\boldsymbol{x})$ and suppose that the points that sample the region $\omega(\boldsymbol{x})$ form a subset of $\mathcal{Y}$ of cardinality N'. Because of this, the GGM associated with $\omega(\boldsymbol{x})$ can be obtained from the GGM associated with $\Omega(\boldsymbol{x})$ by retaining only the row blocks associated with the points that sample $\omega(\boldsymbol{x})$. The proof of this lemma follows very closely the proof of Theorem 2.5.10, the only difference being the fact that the interlacing property of the singular values is now a consequence of the removal of $N - N'$ blocks of m rows from the GGM $A(\Omega(\boldsymbol{x}))$. Using the notation introduced in Theorem (A.2.4) and Equation (A.7), it is readily inferred that the claim holds for Rohr's SGCDF. As far as the CNB SGCDFs are concerned, we start by writing the set of inequalities:

$$\begin{array}{ccc}\frac{1}{\sigma_1} & \leq & \frac{1}{\sigma_{r,1}^{(m(N-N'))}} \\ \frac{1}{\sigma_2} & \leq & \frac{1}{\sigma_{r,2}^{(m(N-N'))}} \\ \vdots & & \vdots \\ \frac{1}{\sigma_p} & \leq & \frac{1}{\sigma_{r,p}^{(m(N-N'))}}\end{array}$$

Summing the corresponding members of the previous inequalities after raising them to the power $2q$ and computing the q^{th} root we obtain:

$$\|A^{\dagger}(\Omega(\boldsymbol{x}))\|_{S,2q} \leq \|A^{\dagger}(\omega(\boldsymbol{x}))\|_{S,2q}$$

which allows us to conclude that $f_{K,q}(\Omega(\boldsymbol{x})) \geq f_{K,q}(\omega(\boldsymbol{x}))$. ■

This lemma sheds some light on the relation between the sampling density and the detector response. If the step of the discretization lattice is smaller, then the response of the SGCDF is smaller as well. The property just discussed does not apply to the Harris-Stephens detector, as it will be shown in the next numerical example.

Example 2.5.19 *Consider the following numerical example. The spectra of the* GGMs $A(\Omega(\boldsymbol{x})) \in \mathbb{R}^{3 \cdot 50 \times 4}$ *and* $A(\omega(\boldsymbol{x})) \in \mathbb{R}^{3 \cdot 38 \times 4}$ *(obtained from* $A(\Omega(\boldsymbol{x}))$ *removing 12 points, i.e. 12 blocks of* $m = 3$ *rows) are:*

$$\begin{aligned} \sigma\left(A(\Omega(\boldsymbol{x}))\right) &= \{8.5792, 3.5454, 3.2763, 3.1456\} \\ \sigma\left(A(\omega(\boldsymbol{x}))\right) &= \{7.2104, 3.1528, 2.8921, 2.6782\} \end{aligned}$$

and they satisfy the interlacing Inequalities (A.7)*. The respective responses of the Harris-Stephens detector for* $\alpha = 0.2\frac{1}{p^p}$ *are:*

$$\begin{aligned} f_{HS}(\Omega(\boldsymbol{x})) &= -3.3787 \cdot 10^3 \\ f_{HS}(\omega(\boldsymbol{x})) &= 2.8687 \cdot 10^3 \end{aligned}$$

and they violate the inequality of Lemma 2.5.18.

Neighborhood Reduction

Motivated by the result stated in the previous Lemma, we now discuss the behavior of a SGCDF when we reduce the dimensionality of the considered neighborhood. More specifically, let's consider the function:

$$\begin{aligned} \boldsymbol{p} : \mathbb{R}^{n_P} &\rightarrow \mathbb{R}^n \\ \boldsymbol{y}' &\mapsto \boldsymbol{x} + V(\boldsymbol{y}' - \boldsymbol{x}') \end{aligned}$$

where $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_P}$ are a set of orthonormal vectors and $V = \begin{bmatrix} \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_n \end{bmatrix} \in \mathbb{R}^{n \times n_P}$. In other words the function $\boldsymbol{p}$ maps points from $\mathbb{R}^{n_P}$ to the affine space generated by the columns of V and passing through the point $\boldsymbol{x}$. Therefore the restriction of the neighborhood $\Omega(\boldsymbol{x})$ can be defined as

$$\omega(\boldsymbol{x}) \overset{\text{def}}{=} \{\boldsymbol{x}' \in \mathbb{R}^{n_P} : \boldsymbol{p}(\boldsymbol{x}') \in \Omega(\boldsymbol{x})\}$$

In the next lemma we will show that the response of a SGCDF restricted to an n_P-dimensional neighborhood is always larger or equal than the response calculated over the original neighborhood $\Omega(\boldsymbol{x})$.

Lemma 2.5.20 (Neighborhood Reduction) *Let's consider the SGCDFs associated with a translational transformation for the neighborhoods $\Omega(\boldsymbol{x})$ and $\omega(\boldsymbol{x})$. Then*

$$f(\Omega(\boldsymbol{x})) \leq f(\omega(\boldsymbol{x}))$$

Proof: The effect of the noise under the mapping $\boldsymbol{p}$ can be modeled similarly to what we did in (2.10), i.e. $\widetilde{\boldsymbol{I}}(\boldsymbol{p}(\boldsymbol{y}')) \overset{\text{def}}{=} \boldsymbol{I}(\boldsymbol{p}(\boldsymbol{y}')) + \boldsymbol{\eta}$. Since we are considering a transformation that models a pure translation, we can write $\widetilde{\boldsymbol{I}}(\boldsymbol{p}(\boldsymbol{y}')) = \boldsymbol{I}(\boldsymbol{p}(\boldsymbol{T}_{\overline{\boldsymbol{\theta}}+\Delta\boldsymbol{\theta}',\boldsymbol{x}'}(\boldsymbol{y}')))$. Thus, adapting the reasoning used to obtain (2.17), we get:

$$A'(\boldsymbol{y}') = w(\boldsymbol{y}' - \boldsymbol{x}') J\boldsymbol{I}(\boldsymbol{p}(\boldsymbol{y}'))\ J\boldsymbol{p}(\boldsymbol{y}')\ J_{\boldsymbol{\theta}}\boldsymbol{T}_{\overline{\boldsymbol{\theta}'},\boldsymbol{x}'}(\boldsymbol{p}(\boldsymbol{y}))$$

From the hypothesis of the lemma we have that $J\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}} = I_n$ and therefore the blocks that form the GGM are related as:

$$A'(\boldsymbol{y}') = A(\boldsymbol{y})V$$

Consequently the GGM becomes $A'(\omega(\boldsymbol{x})) = A(\Omega(\boldsymbol{x}))V$. If we complete the basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_P}$ with a set of $n - n_P$ orthogonal vectors we have that

$$\sigma(A') = \sigma\left(A \begin{bmatrix} V & U \end{bmatrix} \begin{bmatrix} I_{n_P} & 0 \\ 0 & 0 \end{bmatrix}\right) \tag{2.43}$$

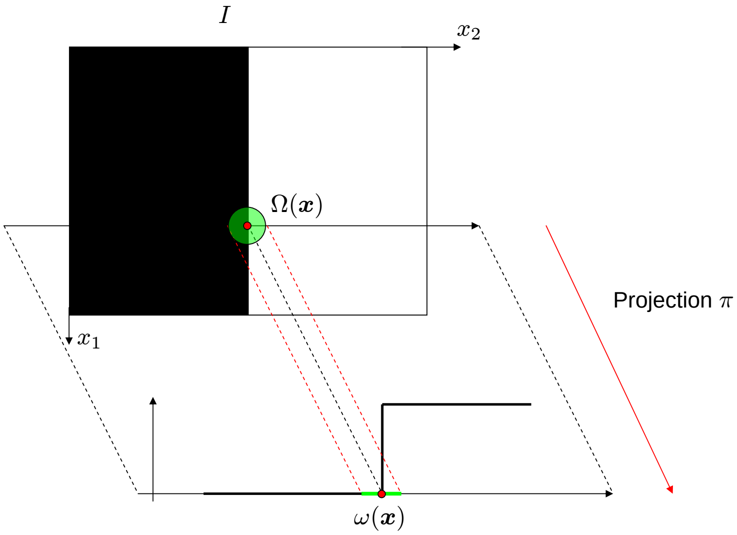


Figure 2.11: The picture illustrates the neighborhood reduction property presented in Lemma 2.5.20. The restriction to a lower dimensional subspace has improved the sensitivity properties of the neighborhood Ω.

The right multiplication by the matrix that has its diagonal composed of n_P ones and $n - n_P$ zeros acts as a selector of the columns of A'. The singular values of A' are thereby interlaced according to the inequalities (A.6). Finally, since $\sigma(A \begin{bmatrix} V & U \end{bmatrix}) = \sigma(A)$, we can conclude the proof the same way we did for Theorem 2.5.10. ■

Note that the Harris-Stephens and the Rohr detectors violate Lemma 2.5.20, because of the same reasons they violate Theorem 2.5.10.

To motivate this lemma, consider an image which is black to the left of the center line and white to the right of the center line (see Figure 2.11). Such an image has an aperture effect in that we may be able to determine left-right motion but not up-down motion. That is any point $\boldsymbol{x}$ on the center line is not suitable as a feature for full motion detection. This is seen in the eigenvalues of the gradient normal matrix $A^T A$: $\min(\lambda_1, \lambda_2) = 0$. Thus we get a zero value for the Rohr and any CNB SGCDF; the Harris-Stephens detector gives a negative value for this example. Now suppose that we pass a line through $\boldsymbol{x}$ and consider the signal of intensity values from the original image along this line. This signal is piecewise constant with a step as it crosses through $\boldsymbol{x}$. Thus it has a good feature for tracking at $\boldsymbol{x}$; the restriction to a lower dimensional subspace has improved the sensitivity properties of the neighborhood Ω.[11]

The consequences of this property are important if we want to attain efficiency of detection by for example using a 1D corner detector in say the x_1-direction; we could then cull the points which have a small response and then do a full detector evaluation at the remaining points in the image. The isotropy property and the definition of the SGCDF ensures that local maxima for the full detector were not eliminated during the preliminary 1D sweep.

Remark 2.5.21 *At first glance the neighborhood restriction and the neighborhood reduction properties seem to contrast with each other. However a careful look into such properties makes it clear that the considered scenarios are deeply different. The neighborhood restriction property deals with a neighborhood obtained as* a subset *of the points in* $\Omega(\boldsymbol{x})$. *On the other hand, the neighborhood reduction property describes the case where* $\omega(\boldsymbol{x})$ *is a* lower dimensional *subspace and therefore the number of parameters that are used to model the noise distortion is smaller than for* $\Omega(\boldsymbol{x})$. *In the example shown in Figure 2.11 the* restriction *to* $\omega(\boldsymbol{x})$ *produces a response that is zero both for the Rohr's and the* CNB SGCDF*s.*

[11] As a technical note, if the subspace line that we choose through $\boldsymbol{x}$ is vertical then no step will appear and the point is still not suitable as a feature point. However this does not violate the spirit of Lemma 2.5.20 since the neighborhood was already unsuitable as a corner in the original (higher dimensional) setting.

Intensity Projection

Similarly to what we did in the previous section, we now study the effects of a projection in the space of the intensities.

Lemma 2.5.22 (Intensity Projection) *Let $m_P \leq m$ and let $P \in \mathbb{R}^{m \times m}$ be the orthogonal projector that projects the intensity space onto a m_P-dimensional space. Then for Rohr's detector and for any* CNB SGCDF*, if we indicate with the superscript prime the response for the projected image, the following inequality holds:*

$$f'_{K,q}(\Omega(\boldsymbol{x})) \leq f_{K,q}(\Omega(\boldsymbol{x})) \tag{2.44}$$

Proof: Suppose that the intensity is projected onto a space spanned by the orthonormal basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ so that the orthogonal projector is defined as the rank m_P matrix $P = VV^T$, where $V = \begin{bmatrix} \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_{m_P} \end{bmatrix}$. The GGM (2.25) associated with the image whose intensity is projected onto the range of V can be written as:

$$A' = \underbrace{\begin{bmatrix} w(\boldsymbol{y}_1 - \boldsymbol{x})I_m & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & w(\boldsymbol{y}_N - \boldsymbol{x})I_m \end{bmatrix}}_{W \in \mathbb{R}^{mN \times mN}} \underbrace{\begin{bmatrix} P & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & P \end{bmatrix}}_{\Pi \in \mathbb{R}^{mN \times mN}} \underbrace{\begin{bmatrix} J\boldsymbol{I}(\boldsymbol{y}_1) J_{\boldsymbol{\theta}} \boldsymbol{T}_{\overline{\boldsymbol{\theta}}, \boldsymbol{x}}(\boldsymbol{y}_1) \\ \vdots \\ J\boldsymbol{I}(\boldsymbol{y}_N) J_{\boldsymbol{\theta}} \boldsymbol{T}_{\overline{\boldsymbol{\theta}}, \boldsymbol{x}}(\boldsymbol{y}_N) \end{bmatrix}}_{\overline{A} \in \mathbb{R}^{mN \times p}}$$

Since W is diagonal and Π is block diagonal and given that the two matrices are compatibly partitioned, it is possible to show that they commute, i.e. $W\Pi = \Pi W$. Thus we can write that $\sigma(A') = \sigma(W\Pi\overline{A}) = \sigma(\Pi W\overline{A})$. If we now consider the singular value decomposition of Π, i.e. $\Pi = U_\Pi \Sigma_\Pi V_\Pi^T$ and we recall that the singular values of a matrix are invariant with respect to unitary transformations, we

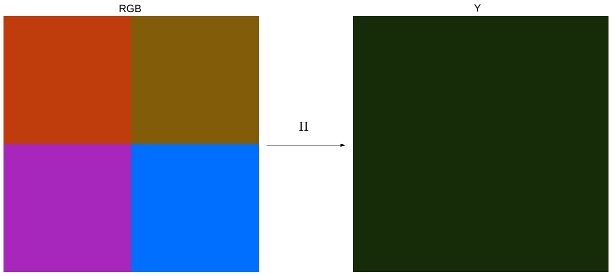


Figure 2.12: The left picture represents the RGB image $\boldsymbol{I}$ and the right picture I' shows the luminance component Y of $\boldsymbol{I}$. Whereas in the left image there is a well defined corner, the right image does contain any interest point.

have that $\sigma(\Pi W\overline{A}) = \sigma(\Sigma_\Pi V_\Pi^T W\overline{A})$. Using the same argument we can also write $\sigma(V_\Pi^T W\overline{A}) = \sigma(W\overline{A}) = \sigma(A)$. At this point we resort once again to the interlacing property of the singular values to relate the spectrum of A' to the spectrum of A. Lemma A.2.3 states that the spectrum of Π is given by $\sigma(\Pi) = \{\underbrace{1,\ldots,1}_{m_P N}, 0\ldots,0\}$ and consequently the left multiplication by the matrix Σ_Π selects a subset of the rows of the matrix $V_\Pi^T W\overline{A}$ that, as we saw before, has the same spectrum of A. The final part of this proof and of the proof of Lemma 2.5.18 run along similar lines. Inequalities (A.7) allow us to conclude that $f_R(\Omega(\boldsymbol{x})) \geq f'_R(\Omega(\boldsymbol{x}))$ and that:

$$\|A^\dagger(\Omega(\boldsymbol{x}))\|_{S,2q} \leq \|A'^\dagger(\Omega(\boldsymbol{x}))\|_{S,2q}$$

Consequently $f_{K,q}(\Omega(\boldsymbol{x})) \geq f'_{K,q}(\Omega(\boldsymbol{x}))$. ■

Example 2.5.23 *Let's consider an RGB image* $\boldsymbol{I}$ *and its grayscale version* I'*. Usually the grayscale information of a color image is obtained from the luminance component of the YIQ representation*[12] *of the same image. The linear relation that*

[12] The YIQ system is a color primary system constructed by a linear transformation of the RGB cube.

transforms the RGB representation into the YIQ representation is the following:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.229 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Consequently the luminance subspace is generated by the vector:

$$\boldsymbol{v} = \begin{bmatrix} 0.229 \\ 0.587 \\ 0.114 \end{bmatrix}$$

and since the norm of $\boldsymbol{v}$ is not one, the orthogonal projector must be written as $P = \frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}$. Any RGB intensity value which is orthogonal to the subspace generated by $\boldsymbol{v}$ will have the same luminance Y. As shown in Figure (2.12), it may happen that a prominent corner structure disappears when we consider its graylevel representation. Even though this example describes a simple worst case scenario, it provides a practical interpretation of the inequality (2.44).

2.5.4 Summary

Table 2.1 summarizes the properties of the SGCDFs introduced in this section. In the second column of the table it is possible to check whether such properties are valid for an arbitrary transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ or if they only hold in the special case where $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ models a translation.

Property	Only for Translations	Harris-Stephens	Rohr	CNB
Analytic Bounds	N. A.	No	No	Yes
Rotation Invariance	Yes	Yes	Yes	Yes
Monotonicity	No	No	Yes	Yes
Isotropy	No	Yes	Yes	Yes
Neighborhood Restriction	No	No	Yes	Yes
Neighborhood Reduction	Yes	No	No	Yes
Intensity Projection	No	No	Yes	Yes

Table 2.1: Summary of the fundamental properties of the SGCDFs.

2.6 Specialization for 2-Dimensional Single Channel Images

We will now restrict our attention to the important case of single channel, two dimensional images and specialize the results derived in the previous section. To model the effect of noise we will consider the following linear transformations:

- Translation:

$$\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{y} + \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{2.45}$$

  The Jacobian of this transformation coincides with the 2×2 identity matrix.

- RST:

$$\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{x} + \begin{bmatrix} \theta_3 & -\theta_4 \\ \theta_4 & \theta_3 \end{bmatrix} (\boldsymbol{y} - \boldsymbol{x}) + \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{2.46}$$

  If we let $\theta_3 = s \cos \phi$ and $\theta_4 = s \sin \phi$ then s is the scaling factor and ϕ the rotation angle (see Example 2.3.2). The corresponding Jacobian is:

$$J_{\boldsymbol{\theta}}\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}} = \begin{bmatrix} 1 & 0 & y_1 - x_1 & -(y_2 - x_2) \\ 0 & 1 & y_2 - x_2 & y_1 - x_1 \end{bmatrix}$$

- Affine:

$$\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{x} + \begin{bmatrix} \theta_3 & \theta_5 \\ \theta_4 & \theta_6 \end{bmatrix} (\boldsymbol{y} - \boldsymbol{x}) + \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{2.47}$$

The Jacobian of the transformation is:

$$J_{\boldsymbol{\theta}}\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}} = \begin{bmatrix} 1 & 0 & y_1 - x_1 & 0 & y_2 - x_2 & 0 \\ 0 & 1 & 0 & y_1 - x_1 & 0 & y_2 - x_2 \end{bmatrix}$$

All the Jacobians do not depend on the transformation parameters $\boldsymbol{\theta}$ but only on the geometry of the neighborhood $\Omega(\boldsymbol{x})$. The immediate consequence is that the GGM depends only on the functional form of the transformation and not on the transformation parameters.

Generalized Detectors Specialization

In the case of 2 dimensional ($m = 2$) and single channel images ($n = 2$), for a transformation that models a pure translation the expressions of the SGCDFs become:

- $f_{HS} = \lambda_1\lambda_2 - \alpha(\lambda_1 + \lambda_2)^2 = \det(A^TA) - \alpha\,\mathrm{trace}(A^TA)^2$

- $f_R = \sqrt{\lambda_1\lambda_2}$

- $f_{NF} = \frac{\lambda_1\ \lambda_2}{\lambda_1+\lambda_2} = \frac{\det(A^TA)}{\mathrm{trace}(A^TA)}$

- $f_{ST} = \lambda_2$

where $\lambda_1 \geq \lambda_2$ are the eigenvalues of the gradient normal matrix $A^T A$. The previous identities, combined with Theorem 2.5.8, shows how some of the commonly used corner detector functions based on the gradient normal matrix can be obtained *as special instances* of the Kenney's detector and therefore can be interpreted using the approaches based on condition theory discussed in Sections 2.4.1 and 2.4.2.

Given the structure of the function that models an RST transformation we can also derive the following analytical bounds:

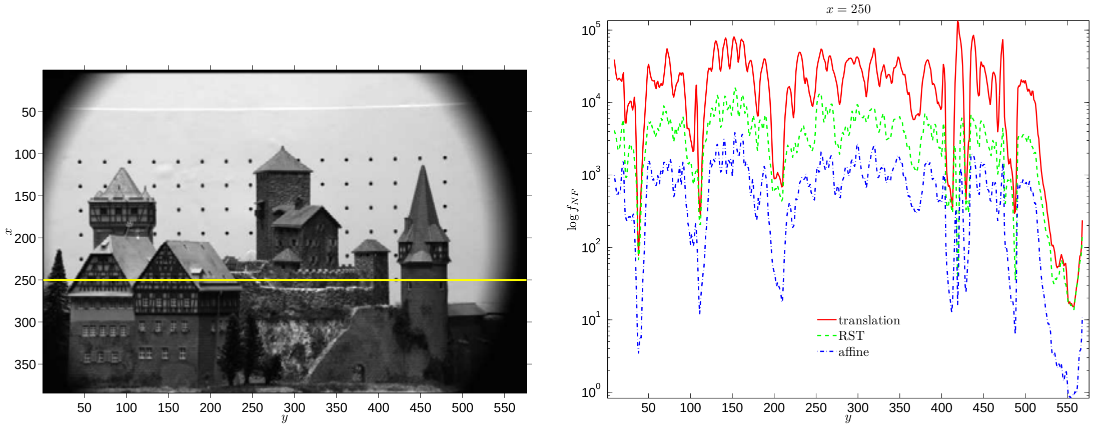
Lemma 2.6.1 (Analytical Bounds Specialization) *Consider the transformations* (2.45), (2.46) *and* (2.47). *Then:*

$$f_{K,q}^{Translation} \geq f_{K,q}^{RST} \geq f_{K,q}^{Affine}$$

Proof: The proof proceeds along the same lines as the proof of Theorem 2.5.10, after noting that the GGM associated with the RST transformation can also be obtained by removing the columns of the RST transformation can also be obtained by removing the columns of the gradient matrix associated with the affine transformation. ■

Figure 2.13 illustrate visually the result stated in the previous lemma.

Example 2.6.2 *Consider the castle scene shown in Figure 2.14(a) and the response maps of the Noble-Förstner detector when* $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})$ *models a translation (b), a rotation, scaling and translation (c) or an affine transformation (d). The neighborhood used to compute the gradient normal matrix* $A^T A$ *is circular and its radius is 6 pixels.*



(a) (b)

Figure 2.13: Plot (b) shows the logarithm of the response of the Noble-Förstner detector along the red scanline shown in (a) when $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})$ models a translation (2.45), a rotation, scaling and translation (2.46) or an affine transformation (2.47). Note that the plots satisfy the inequalities introduced in Lemma 2.6.1.

It is interesting to visually inspect how the corner detectors responses vary according to the transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})$*. In particular the analysis of the response map near the black points that appear in the background of the scene give rise to some interesting considerations. If we follow the approach introduced in 2.4.2, we can state that such points allow one to clearly estimate local translations but are not sufficient to estimate the parameters of a local rotation and scaling or even worse of a local affine transformation. This is why Figures 2.14(c, d) show a very weak response of the detector at such points. These remarks are in agreement with the analysis presented in [129] and summarized in Figure 1 of the same paper.*

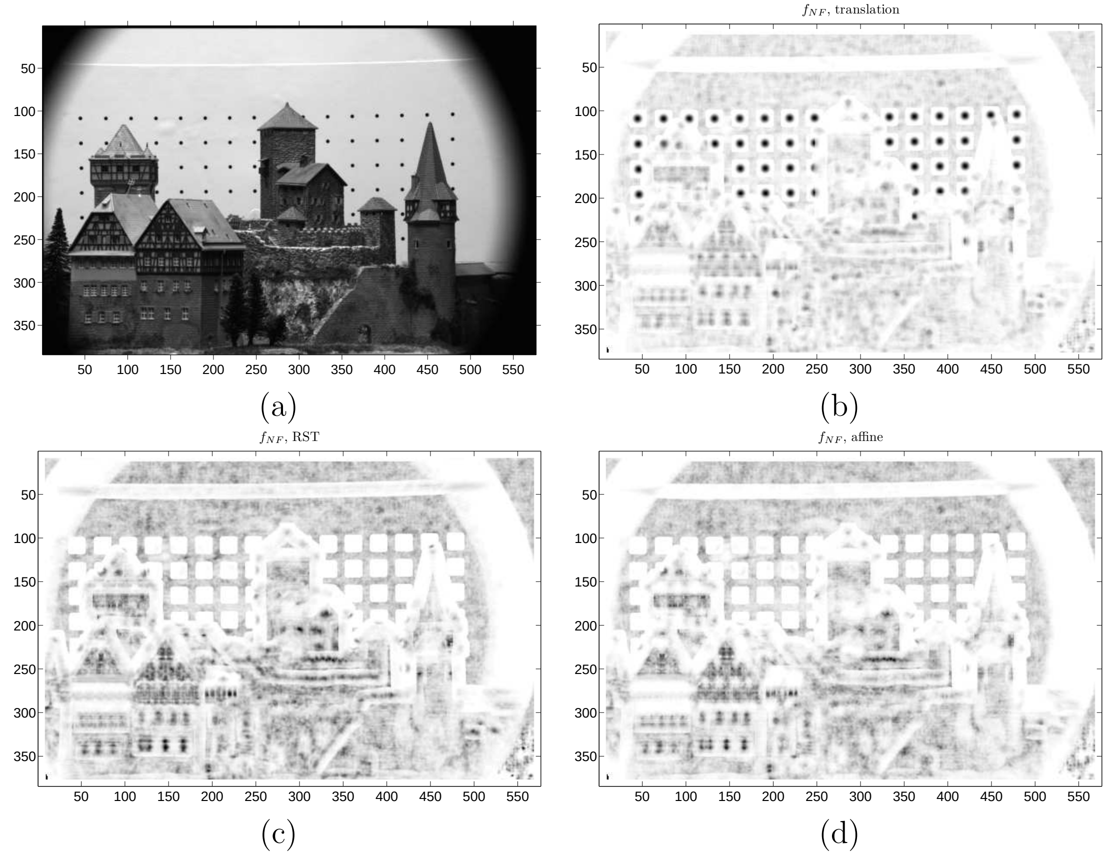


Figure 2.14: The response maps of the Noble-Förstner detector for the Castle scene in (a) when $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y})$ models a translation (b), a rotation, scaling and translation (c) or an affine transformation (d). Darker colors indicate larger values of f_{NF}.

2.7 Conclusions

In this chapter we developed a thorough theoretical analysis of point feature detectors based on the spectral properties of the GGM. We introduced a novel framework based on condition theory that motivates the use of the autocorrelation matrix as a fundamental ingredient for point detection. We then introduced a set of spectral generalized corner detector functions based on the spectral properties of the generalized matrix matrix. Such detectors are defined for multichannel images with spatial dimension that can be greater than 2. For single channel images these generalized functions become equivalent to some of the commonly used point detectors. Within this framework we established in-depth connections among the detectors showing that certain detectors are equivalent modulo the choice of a specific matrix norm and we listed a set of analytical properties of the SGCDFs that define bounds to their performance and suggest effective ways to reduce their computational complexity. The theory presented in this chapter will be supplemented by the experimental analysis in Chapter 3. The theoretical foundations laid in this chapter and the experiments carried out in the next chapter will be the starting point to design the detector module for the registration and mosaicking system described in Chapter 6. Finally we want to remark that the framework presented in this chapter is general enough to be applicable in other domains of image analysis, such as the identification of curve landmarks, as discussed in Appendix B.

Chapter 3

Point Feature Detectors: Experiments

"There's two possible outcomes:
if the result confirms the hypothesis, then you've made a discovery.
If the result is contrary to the hypothesis, then you've made a discovery."

E. Fermi

This chapter contains an exhaustive experimental evaluation of the point detectors studied in Chapter 2. More specifically:

- We experimentally validate the theoretical claims made in Chapter 2 regarding *detector equivalences* (see Section 3.4).
- We characterize the repeatability of the point detectors and find that they exhibit a behavior that is *almost linear* for a relevant set of scalings and projective distortions that are found in real life scenarios (see Section 3.4).

- Quite surprisingly we find that for natural images it is possible to *disregard the color information* and at the same time improve the detector performance (see Section 3.4).

3.1 Introduction

In this chapter we will supplement the theoretical analysis of the point detectors introduced previously by means of a thorough set of experiments. More specifically, we will study how the performances of the proposed detectors vary in the presence of geometric and photometric transformations that are commonly found in real life scenarios. We will consider both the cases when the detectors operate on RGB images and on graylevel images.

This chapter is structured as follows. In Section 3.2 we will discuss the fundamental issues arising in the practical implementation of the detectors introduced in the previous chapter, and the methodology that we use in our experimental analysis. Section 3.4 contains the description of the experiments as well as their discussion. Finally some directions for the design of point feature detectors based on the gradient normal matrix are presented in Section 3.5.

3.2 Implementation Details

In this section we turn our attention to some issues that arise in the implementation of the detector functions described in Section 2.6. First of all we need to decide which neighborhood we will consider to calculate the gradient normal matrix. We opted for a circular neighborhood whose radius is related to the standard

Parameter Description	Symbol	Value	Units
Standard deviation of the differentiation filter	σ_D	1	pixels
Neighborhood radius	r	$\lceil 3\sigma_D \rceil = 3$	pixels
Point features minimum distance	d_{min}	3	pixels

Table 3.1: Summary of the parameters used to implement the detectors described in Section 2.6.

deviation σ_D of the Gaussian filter used to compute the image derivatives. Thus we set $r = \lceil 3\sigma_D \rceil$. The image derivatives are computed convolving the image with the derivatives of Gaussian kernels, as suggested in [36, 70]. In our experiments we set $\sigma_D = 1$. Where needed, the spectral decomposition of the matrix $A^T A$ is performed as described in Appendix A.2.4. We used a uniform weight $w = 1$ for all the points inside the neighborhood. This choice is mainly dictated by computational efficiency constraints: in our implementation the components of the gradient normal matrix are updated recursively taking advantage of the SIMD (Single Instruction, Multiple Data) architecture of the CPU. After the response of the detector is calculated at each point of the image (except for the border portions which are discarded because of the artifacts due to the convolution) we check the 8-connected neighborhood of each pixel. A feature point is detected when the value of the response at the central pixel is *not smaller* than the response of the neighboring ones and at least one of the 8-connected neighbors has a response that is *strictly smaller* than the central one. Moreover we discard tie points that are closer than d_{min} pixels. Our current research-oriented implementation of the detector takes about 0.75 seconds to find about one thousand points on a 800×600 image with a 2.4GHz Intel Xeon CPU.

3.3 The Experimental Setup

3.3.1 Repeatability

The main tool that we will use for our analysis is deeply influenced by the work of Schmid et al. [113] and Mikolajczyk et al. [87, 89], where the performance of a point detector is measured in terms of *repeatability*. The repeatability quantifies the ability of a point detector to identify corresponding points in the presence of image distortions that modify both the geometry of the image (such as rotations, scalings or projective transformations) and its intensity (such as Gaussian noise, compression artifacts, motion blur, and light changes). Quoting Schmid et al. in [113]:

> [High] Repeatability signifies that detection is independent of changes in the imaging conditions, i.e. the parameters of the camera, its position relative to the scene, and the illumination conditions.

In our experiments we considered images related by a geometric transformation that can be modeled by a planar homography H.[1] Under this assumption, if we call $\mathcal{X}_1$ the set of tie points detected in the first image and $\mathcal{X}_2$ the set of tie points detected in the second image, we can define the *ε-corresponding sets* as:

$$\mathcal{C}_1(\varepsilon) \overset{\text{def}}{=} \left\{\boldsymbol{x}_1 \in \mathcal{X}_1 : \exists \boldsymbol{x}_2 \in \mathcal{X}_2 \text{ s.t. } \operatorname{dist}(\boldsymbol{x}_1, H^{-1}\boldsymbol{x}_2) \leq \varepsilon\right\} \tag{3.1}$$

$$\mathcal{C}_2(\varepsilon) \overset{\text{def}}{=} \{\boldsymbol{x}_2 \in \mathcal{X}_2 : \exists \boldsymbol{x}_1 \in \mathcal{X}_1 \text{ s.t. } \operatorname{dist}(H\boldsymbol{x}_1, \boldsymbol{x}_2) \leq \varepsilon\} \tag{3.2}$$

where $\operatorname{dist}(\cdot,\cdot)$ is the function that returns the Euclidean distance between two points expressed in homogeneous coordinates. We will also call *ε-corresponding*

[1] When the distortion only affects the intensity of the image, then the corresponding homography coincides with the identity matrix.

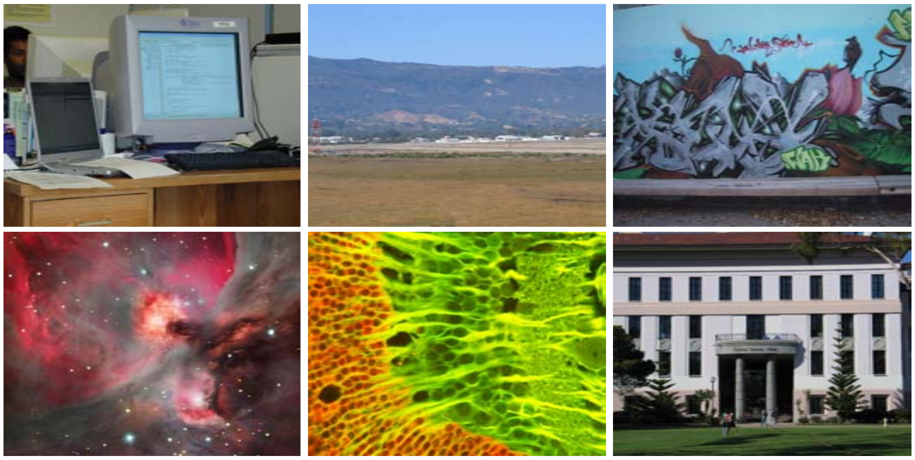


Figure 3.1: Test images used in the experiments.

two points $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $\text{dist}(\boldsymbol{x}, \boldsymbol{x}) \leq \varepsilon$. This said, the definition of the ε-*repeatability* is:

$$r_\varepsilon \overset{\text{def}}{=} \frac{|\mathcal{C}_1(\varepsilon)|}{\min(|\mathcal{X}_1|, |\mathcal{X}_2|)} = \frac{|\mathcal{C}_2(\varepsilon)|}{\min(|\mathcal{X}_1|, |\mathcal{X}_2|)} \tag{3.3}$$

where $|C_i(\varepsilon)|$ indicates the cardinality of the set $C_i(\varepsilon)$. Note that the ε-repeatability is normalized in the interval $[0, 1]$: when $r_\varepsilon = 1$ all the points detected in one image are also detected in the other image within a tolerance of ε pixels.

3.3.2 Image Distortions

The image dataset that we used in our experiments consists of a set of 6 RGB images belonging to different domains (see Figure 3.1). Their original resolution is 800 × 600 pixels. For each image we synthesized the following distortions.

- **Rotations**. We generated 18 images obtained by rotating the original image from 10° to 180° with rotation increments of 10°.

- **Scalings**. We generated 15 images obtained by scaling the original image from 5% to 75% with scaling increments of 5%.

- **Projective distortions**. We generated 8 images obtained by applying to the original image an homographic transformation that simulates a change in the position of the camera. Each homography is generated following the procedure that is explained pictorially in Figure 3.2. The rectangle $ABCD$, which represents the boundary of the original image, is transformed into the rectangle $A'B'C'D'$, which represents the boundary of the new image. The transformation is parameterized by a single positive scalar α such that $\overline{A'O} = (1+\alpha)\overline{AO}$, $\overline{B'O} = (1+\alpha)\overline{BO}$ and $\overline{C'O} = (1-\alpha)\overline{CO}$, $\overline{D'O} = (1-\alpha)\overline{DO}$. The values of α that we used are:

$$\{-0.20, -0.15, -0.10, -0.05, 0.05, 0.10, 0.15, 0.20\}$$

- **Intensity noise**. We generated 10 images by perturbing each channel of the original RGB image with zero mean Gaussian noise with standard deviation ranging from 2.55 to 25.5 with increments of 2.55 intensity units.

- **Blur**. We generated 7 images by applying a Gaussian low-pass filter to each channel of the original RGB image with standard deviation ranging from 1 to 4 pixels with increments of 0.5 pixels.

The images obtained applying the geometric transformations have been rendered using a bicubic interpolation method. Figure 3.3 displays some of the distortions considered in our experiments.

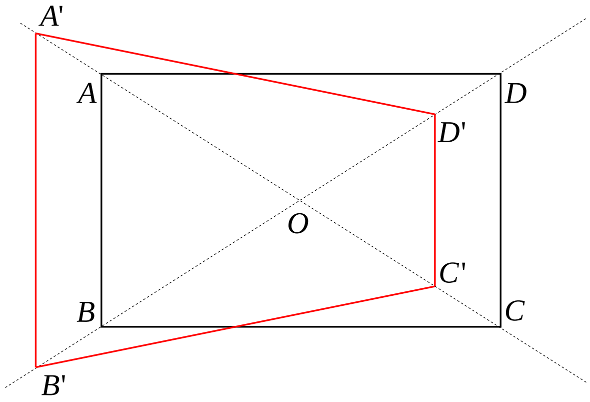
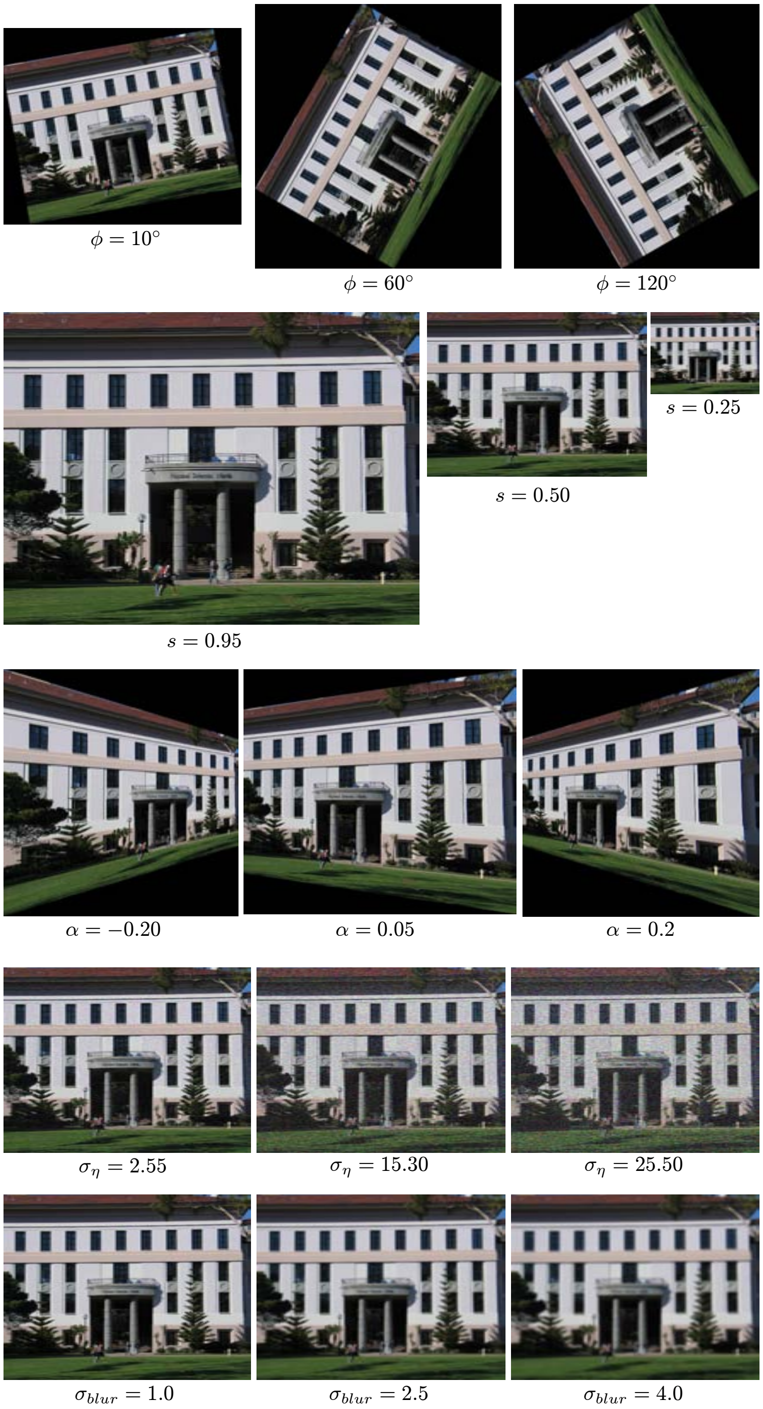


Figure 3.2: The picture illustrates the method used to generate the homographies that model the projective distortions.

3.4 Experimental Results

We carried out three groups of experiments considering five detector functions: Harris-Stephens, Rohr, Noble-Förstner (i.e. Kenney for $q = 1$), Kenney for $q = 2$ and Shi-Tomasi (i.e. Kenney for $q = \infty$). In all the experiments we compared the performance of the detectors when the input is either an RGB or graylevel image and for two values of the distance threshold: $\varepsilon = 1, 2$ pixels.

In Section 3.4.1 we analyzed how the number of corresponding points that are detected both in the original image and in its distorted version varies with respect to different image distortions. Section 3.4.2 presents the results of the average repeatability for all the geometric and photometric distortions, for all the considered detectors, for $\varepsilon = 1, 2$ pixels. Finally Section 3.4.3 discusses the rate of growth of the repeatability with respect to the number of detected points. All the plots for the experimental results can be found at the end of this chapter.



$\phi = 10^\circ$ $\phi = 60^\circ$ $\phi = 120^\circ$

$s = 0.95$ $s = 0.50$ $s = 0.25$

$\alpha = -0.20$ $\alpha = 0.05$ $\alpha = 0.2$

$\sigma_\eta = 2.55$ $\sigma_\eta = 15.30$ $\sigma_\eta = 25.50$

$\sigma_{blur} = 1.0$ $\sigma_{blur} = 2.5$ $\sigma_{blur} = 4.0$

Figure 3.3: Some examples of the images synthesized by considering specific instances of the following geometric distortions (from top to bottom): rotation, scaling, projective, intensity noise and blur.

Before continuing, we would like to emphasize how the comparison of our results with the results obtained by Schimd et al. in [113] requires some care. First of all the image dataset we are dealing with is not acquired changing the parameters of the imaging device (such as its pose or its position) or the scene conditions (such as the lighting), but instead it is synthesized starting from a set of real images. This allows us to have precise control over the considered distortions, at the price of disregarding some other distortions that occur when dealing with real images. The motivation behind this choice is to isolate the impact of *each* distortion on the performance of the detectors. We want also to emphasize that in [113] only two scenes were considered (namely the “Asterix” and the “Van Gogh” scene, whose pixel resolution is not specified in the paper) whereas our dataset covers a wider variety of images.

3.4.1 Average Percentage of Corresponding Points

Using the definition (3.2), we plotted the percentage $100\frac{|C_1(\varepsilon)|}{N_P}$ of corresponding point pairs that are detected within the given distance threshold ε (in our experiments one or two pixels) with respect to the parameters that define the image distortions. The total number of feature points detected in the reference image is denoted by N_P.[2] For these experiments the graphs displayed in Figure 3.4 to 3.7 contains five curves, one for each detector. Each point that defines a curve is obtained by taking the average over all the corresponding experimental instances for the images in the dataset.

[2] Note that in our experiments the transformed images contain *entirely* the original image, therefore our experimental setup is meaningful, since we do not have to worry about points that are potentially detected in one image but not in the other.

In general all the detectors perform better when the graylevel version of the image is considered, even though the fluctuations are never larger than 5%. When the distance threshold is lowered from two pixels to one pixel the percentage of the corresponding points decreases approximately by 20%, except in the case of the intensity perturbation, where the number of corresponding points decreases about 30%. Interestingly enough, the curves for the scaling and projective distortions exhibit an essentially linear behavior with respect to the transformation parameters. Even if the performance of the detectors is quite similar, the Harris-Stephens detector tends to behave slightly worse than the other detectors, whereas the Noble-Förstner and the Rohr detector consistently yield the best percentages. Note also that the image blurring produces a rapid drop in the performance of the detectors.

3.4.2 Repeatability for Geometric and Photometric Distortions

The graphs displayed in Figure 3.8 to 3.12 show the values of the ε-repeatability for different distortions. Also in this case the detectors perform better when they operate on the graylevel version of the image. The overall results for the graylevel images for the Harris-Stephens detector are analogous to those obtained in [113]. As expected from the equivalence Theorem 2.5.8, the other detectors are characterized by performances that are very close to each other, the only noticeable difference being the better behavior of Rohr's detector in the presence of image blur. Also for the ε-repeatability the curves relative to scaling and projective distortions are essentially linear. As expected the curves describing the average distance between ε-corresponding feature points are inversely correlated with the ε-repeatability. Moreover larger distortions cause the average corresponding point distance to shift towards the threshold ε.

3.4.3 Repeatability Rate of Variation

The goal of this experiment is to study the rate of variation of the repeatability with respect to the total number of points N_P detected in an image. The ordinate of the graphs tabulate the values $\frac{r_\varepsilon}{r_{\varepsilon_min}}$; as expected they are increasing with the number of considered points. This observation suggests that it is important to evaluate the benefit obtained by considering a larger number of local maxima in the detector response map (i.e. considering also the local maxima that have a small value). For this experiment we only considered the Noble-Förstner detector and for each type of distortion we plotted the curves corresponding to a representative subset of the parameters used to synthesize the distorted images. The points are sorted in order of decreasing detector response. The corresponding plots are shown in Figure 3.13 to Figure 3.16. In the presence of large scalings ($s \approx 0.5$) it is beneficial to consider a larger number of local maxima in the detector response map (i.e. to augment N_P). This is also true when dealing with projective distortions, whereas for rotations the benefit obtained by augmenting the value of N_P is negligible. Similar considerations hold for photometric distortions. In general the rate of change of the repeatability of the Noble-Förstner detector does not change considerably when considering RGB images rather than graylevel images and when changing the distance threshold from one pixel to two pixels.

3.4.4 Experiment Summary

The set of the experiments described in the previous sections support the following general statements.

- Experientially, the behavior of the Harris-Stephens detector, the Rohr detector and the Condition Number Based (CNB) detectors on synthetically distorted natural images is very similar. However the Harris-Stephens detector seems to perform slightly worse than the other detectors, whereas the Rohr and Noble-Förstner detectors perform slightly better.

- We were surprised to find out that all the considered detectors perform consistently better when they operate on the graylevel version of the image rather than on the corresponding RGB version. We explain this by recalling that the RGB channels *of natural images* tend to be highly correlated. This redundancy may cause the numerical stability properties of the Generalized Gradient Matrix (GGM) to deteriorate. This is because the GGM is composed of blocks of three rows that are almost "linearly" dependent.

- The number of ε-corresponding features roughly decreases by 20% when the distance threshold changes from two pixels to one pixel.

- The repeatability curves exhibit an approximately linear behavior for the scaling and projective distortions tested in the experiments. This makes it possible to predict a priori the number of expected ε-corresponding features for a given range of distortions.

The previous observations are in good accordance with the conclusions derived from the theoretical discussion presented in Chapter 2. The similarity of behavior of the tested detectors is predicted by Theorem 2.5.8, that established a set of analytical equivalence relations between the Condition Number Based (CNB) detectors and the Rohr detector. We also hypothesize that the slight decrease of performance for the Harris-Stephens detector can be attributed to the fact that it fails to satisfy the monotonicity property 2.5.15.

3.5 Prolegomena for the Design of SGCDFs

In light of the theoretical analysis presented in the previous chapter and of the experimental results discussed in this chapter, we will provide some prefatory remarks regarding the design of feature point detectors based on the spectral properties of the Generalized Gradient Matrix (GGM).

- The image derivatives should be computed in accordance with the directions provided by scale space theory, i.e. by means of convolution with the derivatives of Gaussian kernels. This is also supported by the discussion in [113].

- For natural images it is recommended to operate on their graylevel versions. This reduces the computational complexity and improves the performances of the detectors.

- The Noble-Förstner (i.e. Kenney for $q = 1$) detector is the suggested choice to detect features in natural images. Its experimental performances are extremely satisfactory, its computation does not require the explicit eigendecomposition of the gradient normal matrix $A^T A$, and it can be derived within a sensible mathematical framework. Moreover it does not require the introduction of a constant whose value can be guessed only by resorting to empirical considerations (like the Harris-Stephens detector).

- Before setting up a pyramidal decomposition approach such as the one described by Lowe [74], evaluate if it is worth coping with the computational burden associated with the image decomposition and with the generation of a large number of candidate feature points. Our experiments indicate that the average percentage of ε-correspondences and the repeatability associated with scaling and projective distortions remain acceptable also in presence of remarkable geometric and photometric distortions.

- It is not necessarily good practice to consider all the points that can be obtained from the local maxima of the detector map. Our experiments show that only the points that are characterized by a strong response maintain good repeatability properties. For example, for a projective distortion associated with the parameter value $\alpha = 0.15$, the 2-repeatability increases by 20% if the total number of considered points increases from 200 to 800.

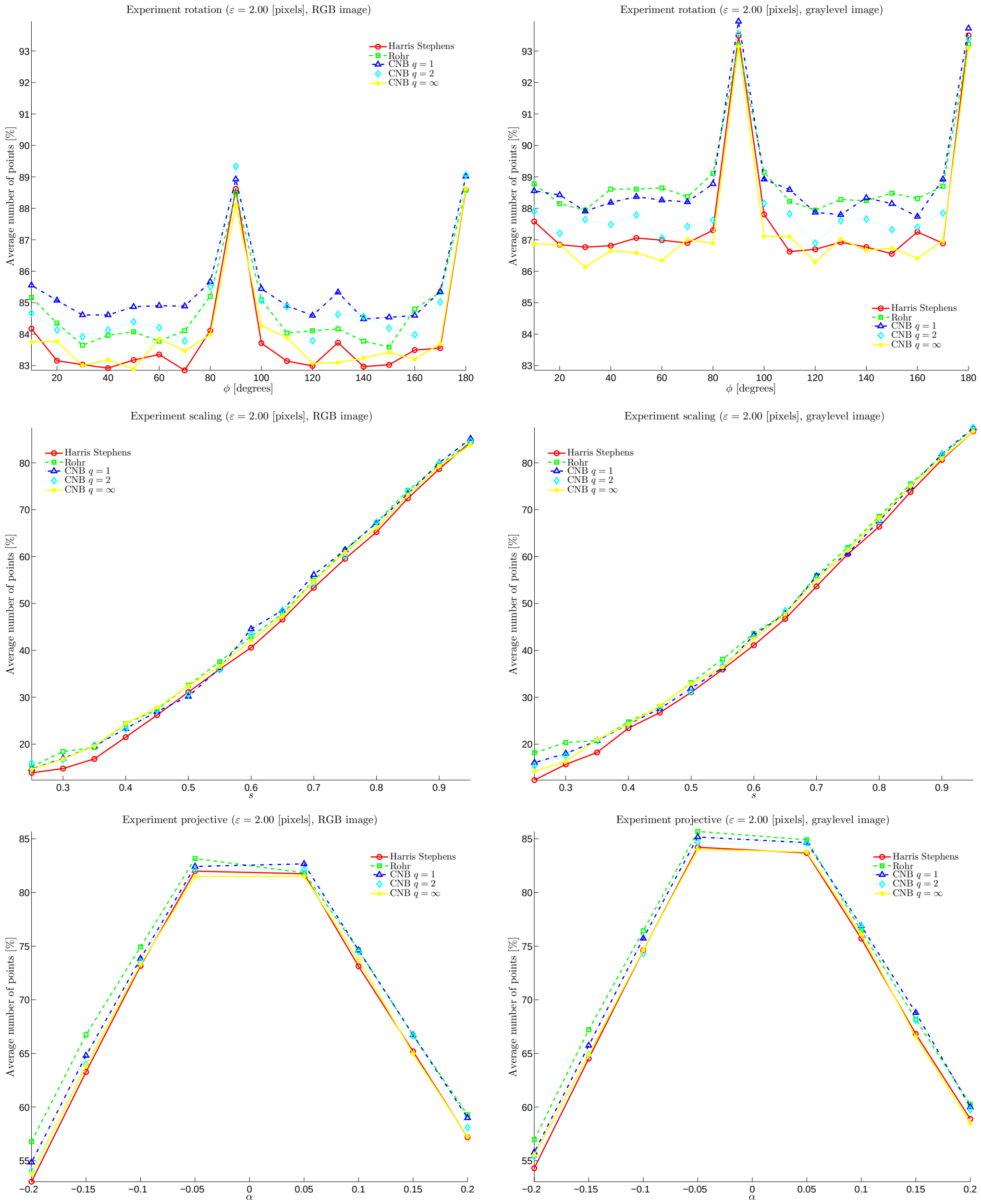


Figure 3.4: Percentage of detected points for geometric distortions for $\varepsilon = 2$ pixels.

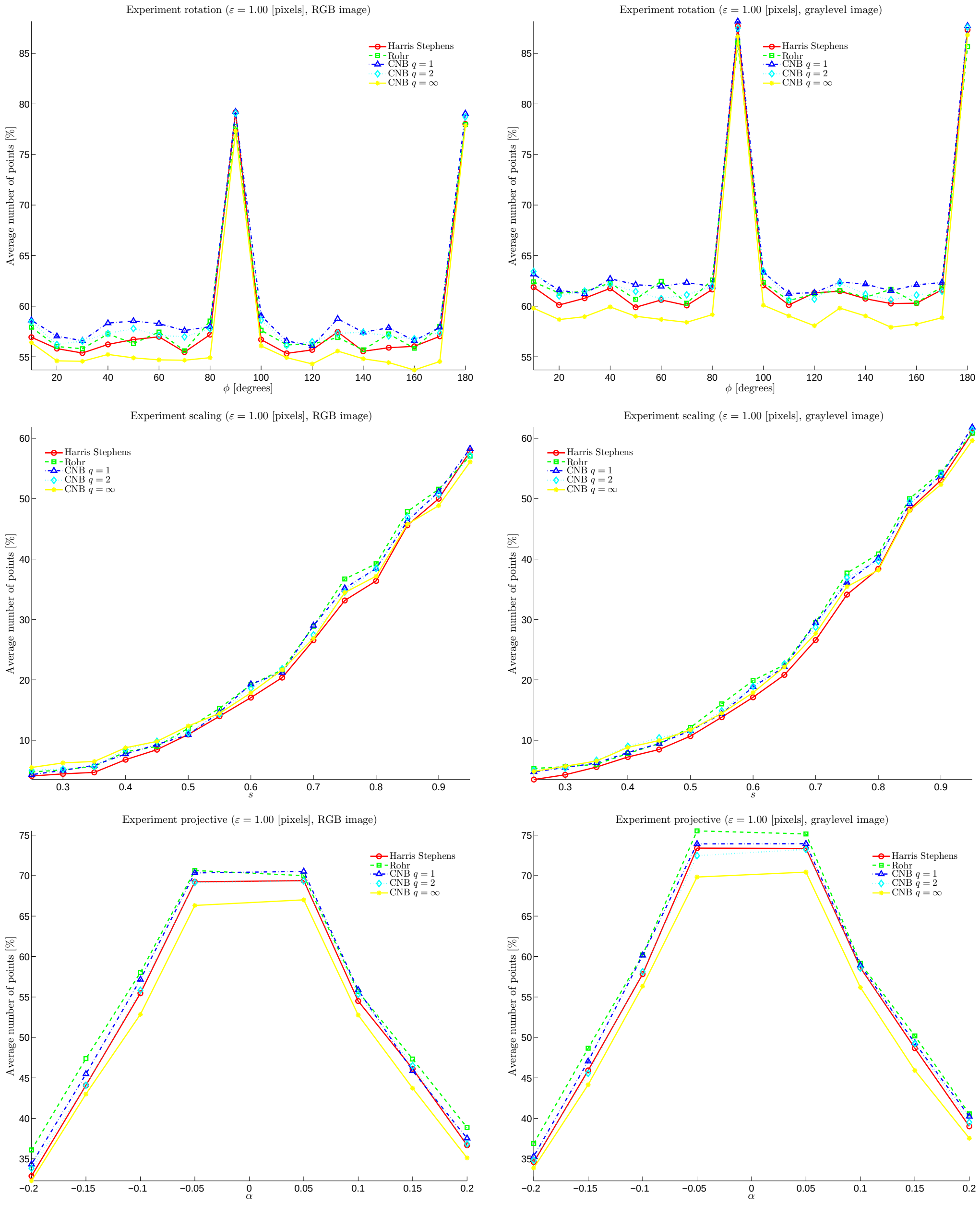


Figure 3.5: Percentage of detected points for geometric distortions for $\varepsilon = 1$ pixels.

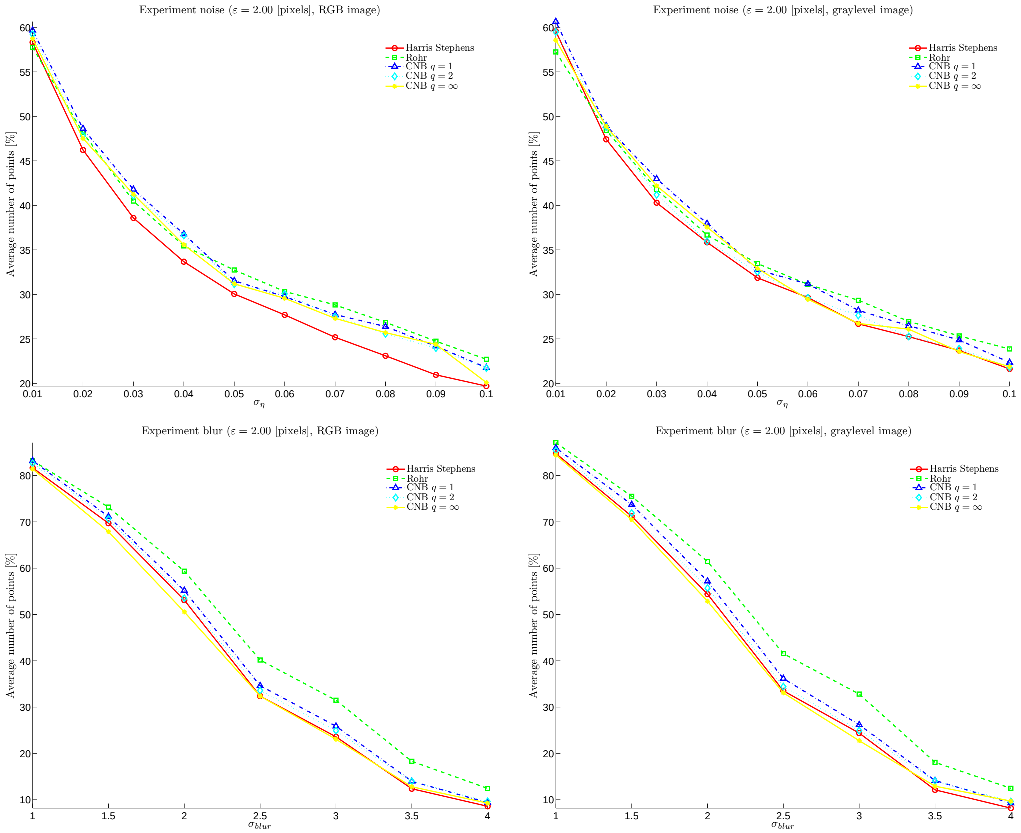


Figure 3.6: Percentage of detected points for photometric distortions for $\varepsilon = 2$ pixels.

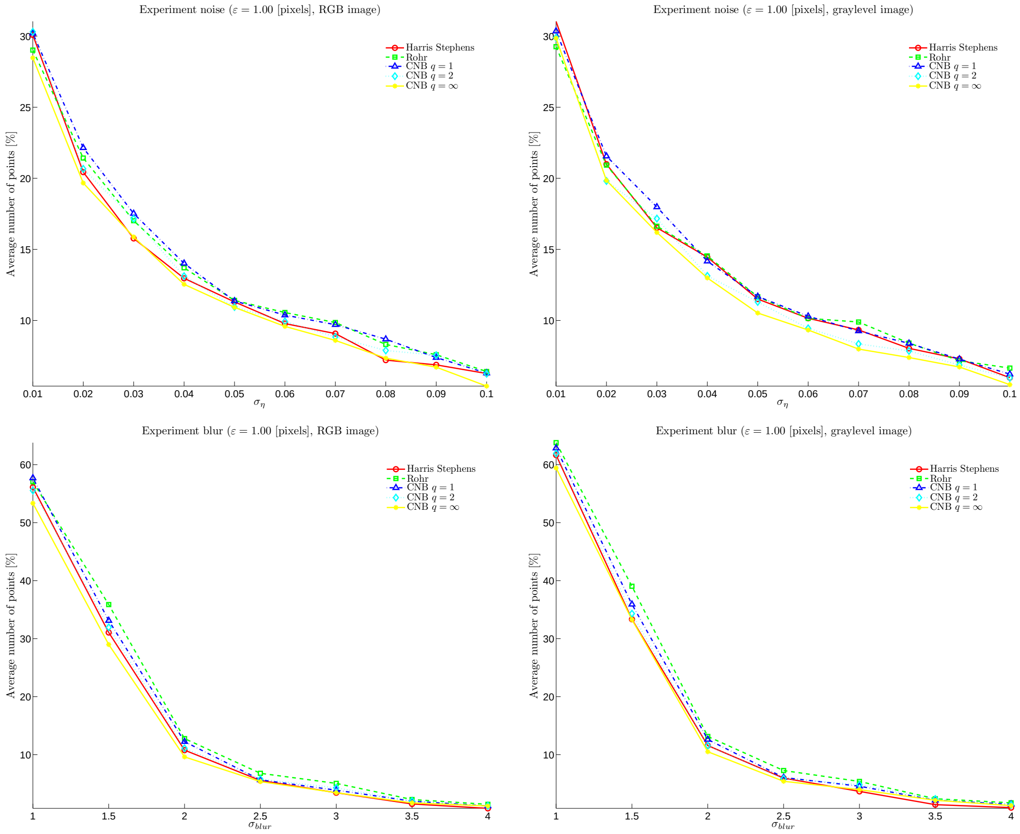


Figure 3.7: Percentage of detected points for photometric distortions for $\varepsilon = 1$ pixels.

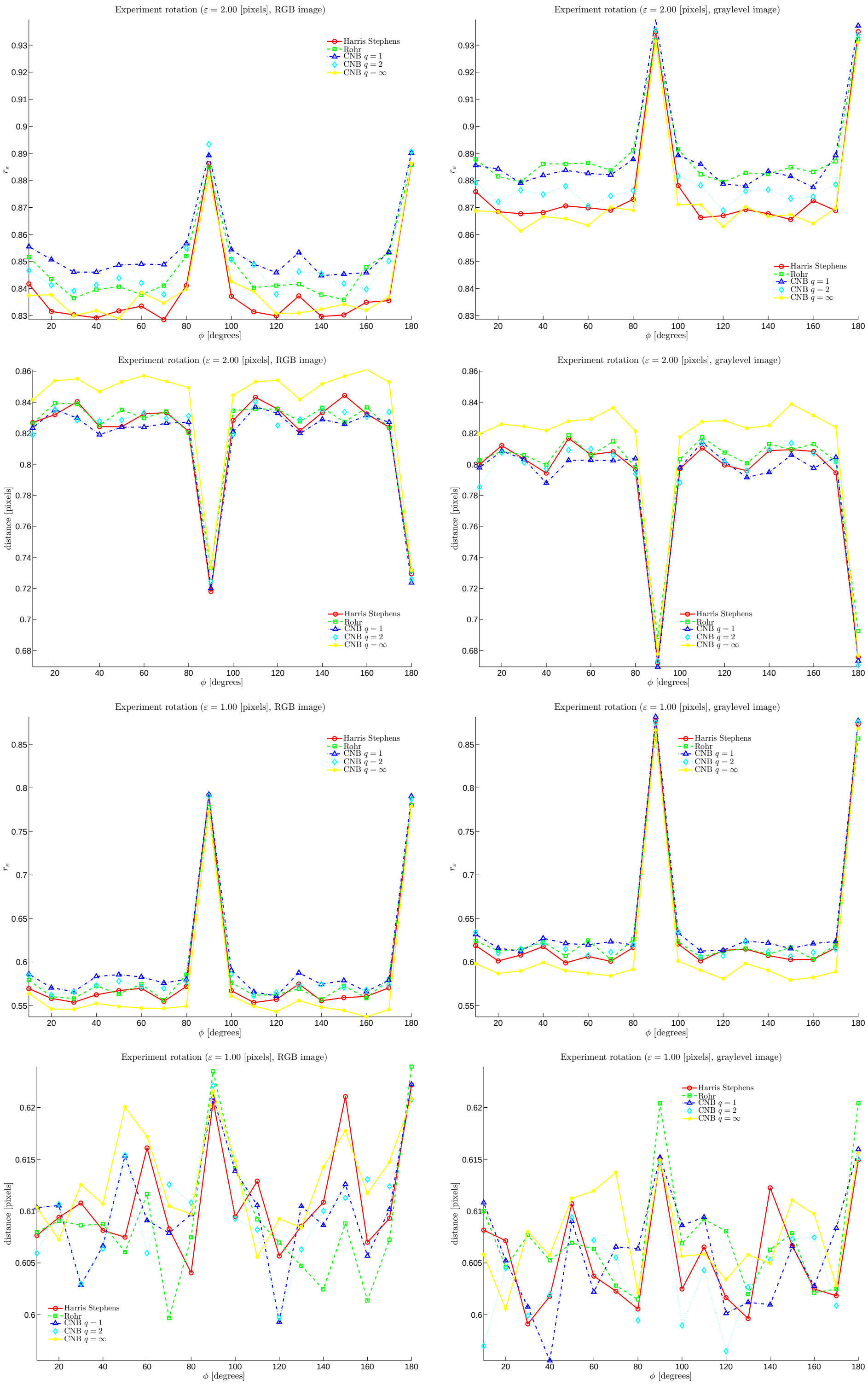


Figure 3.8: Repeatability and average correspondence distances for image rotations.

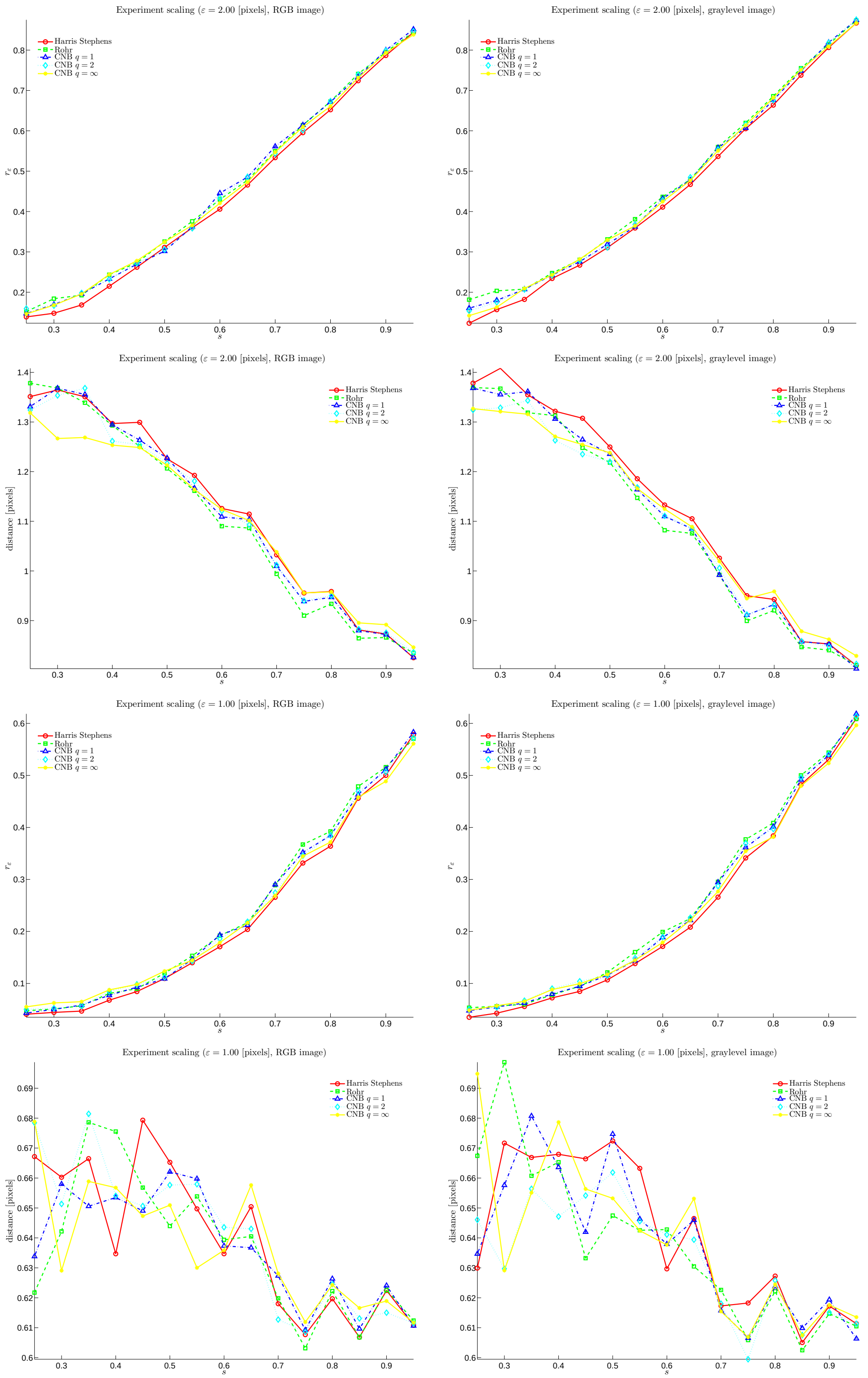


Figure 3.9: Repeatability and average correspondence distances for image scalings.

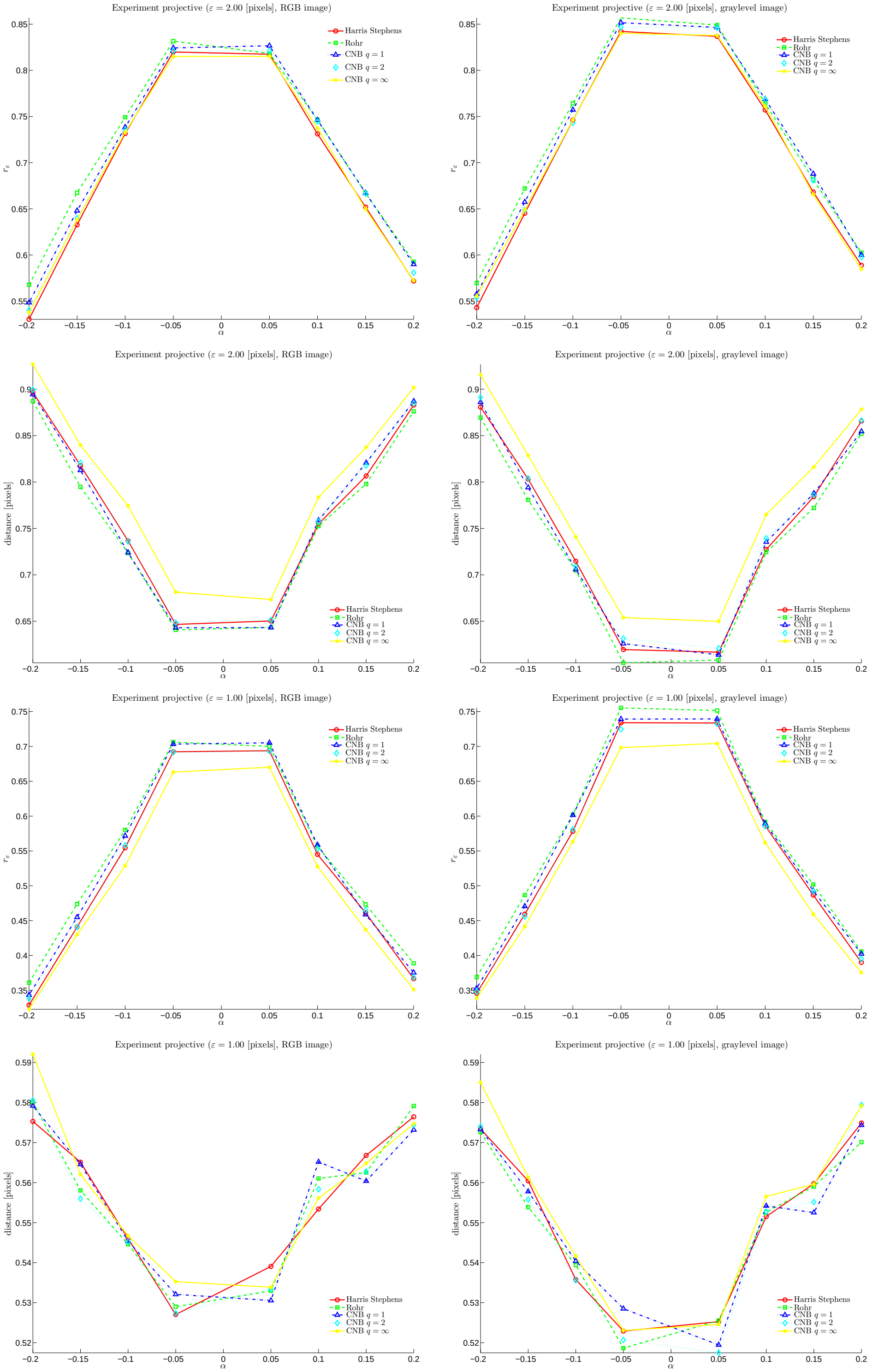


Figure 3.10: Repeatability and average correspondence distances for image projective distortions.

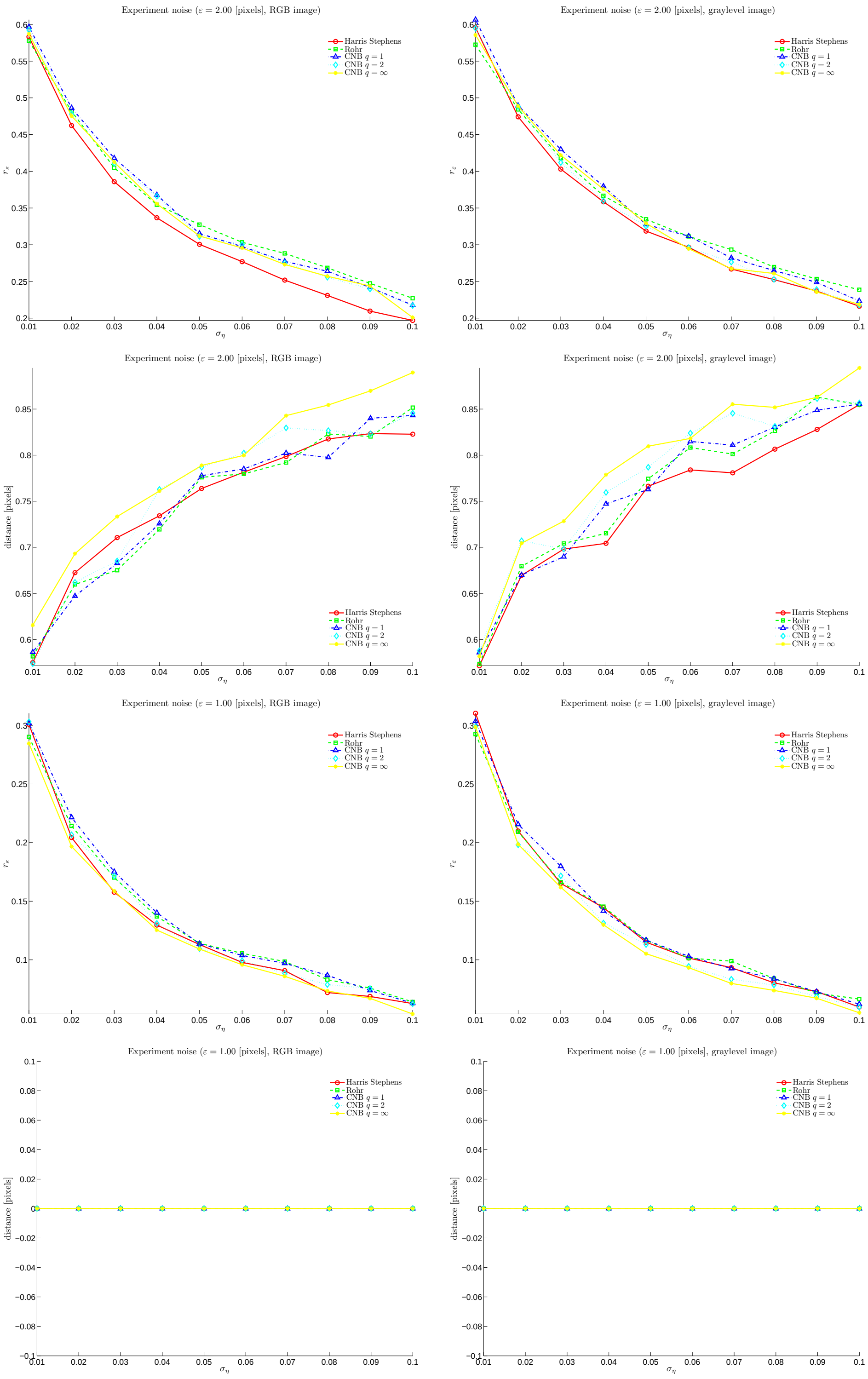


Figure 3.11: Repeatability and average correspondence distances for image intensity Gaussian noise.

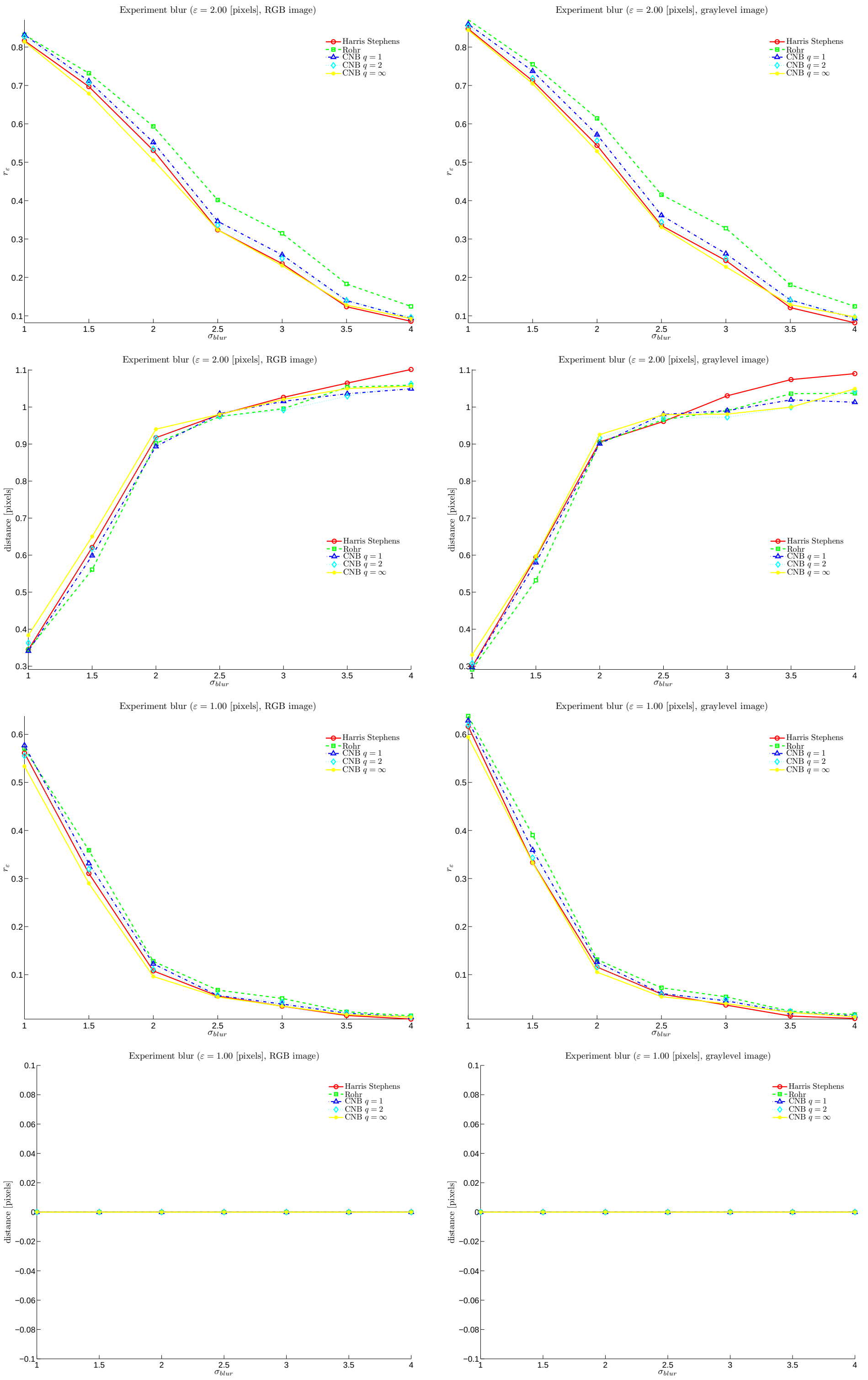


Figure 3.12: Repeatability and average correspondence distances for image blur.

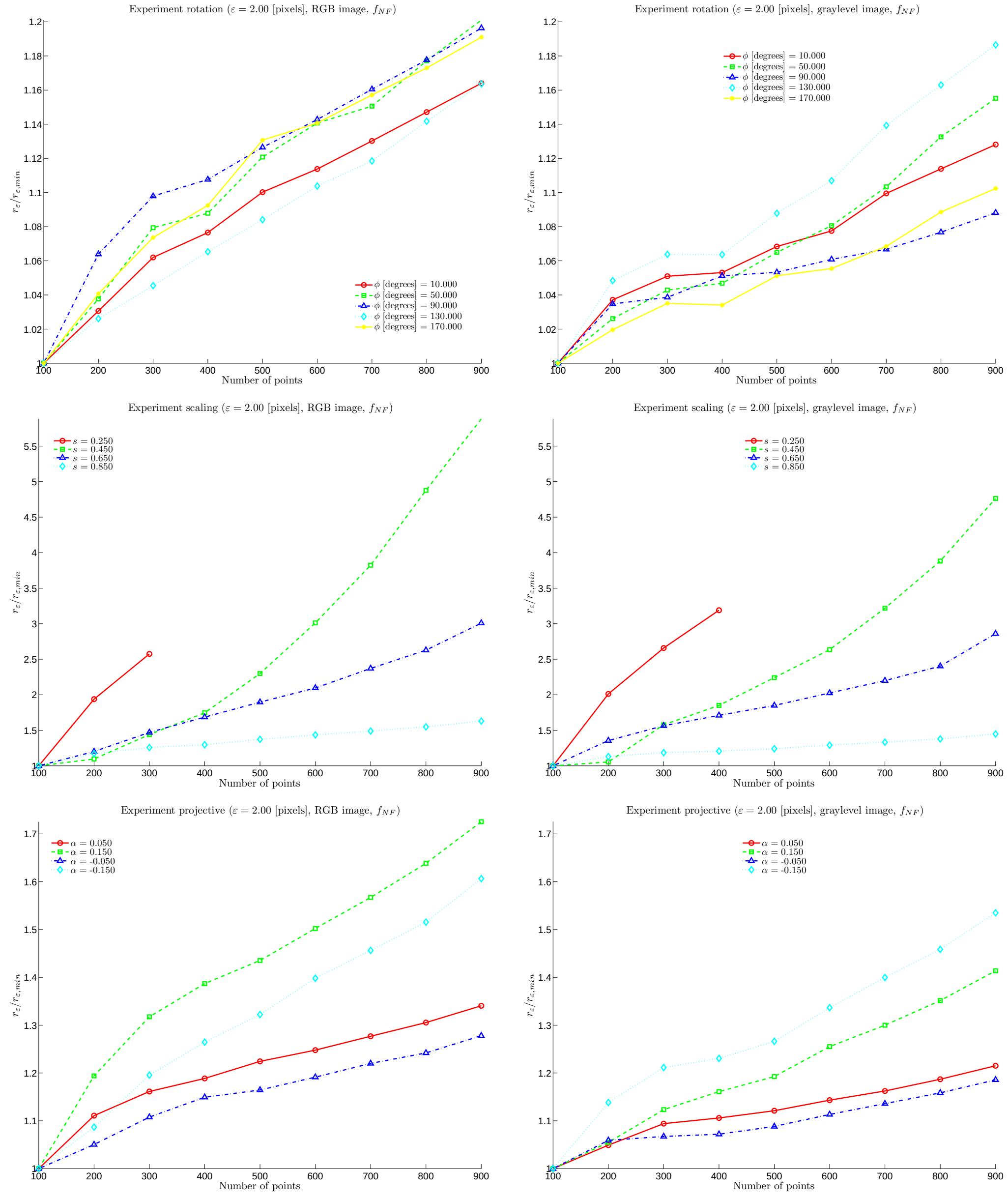


Figure 3.13: Repeatability variation for geometric distortions for $\varepsilon = 2$.

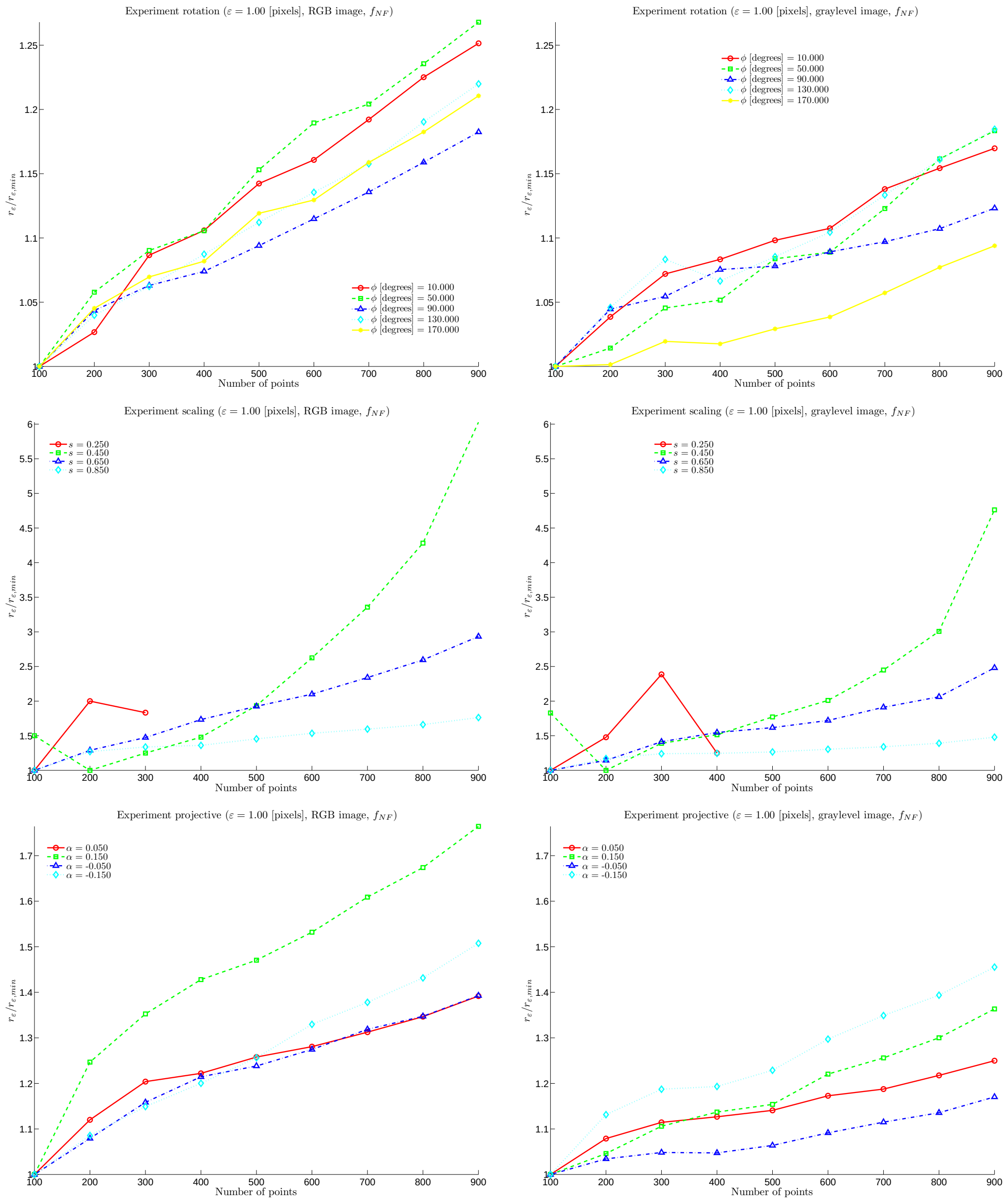


Figure 3.14: Repeatability variation for geometric distortions for $\varepsilon = 1$.

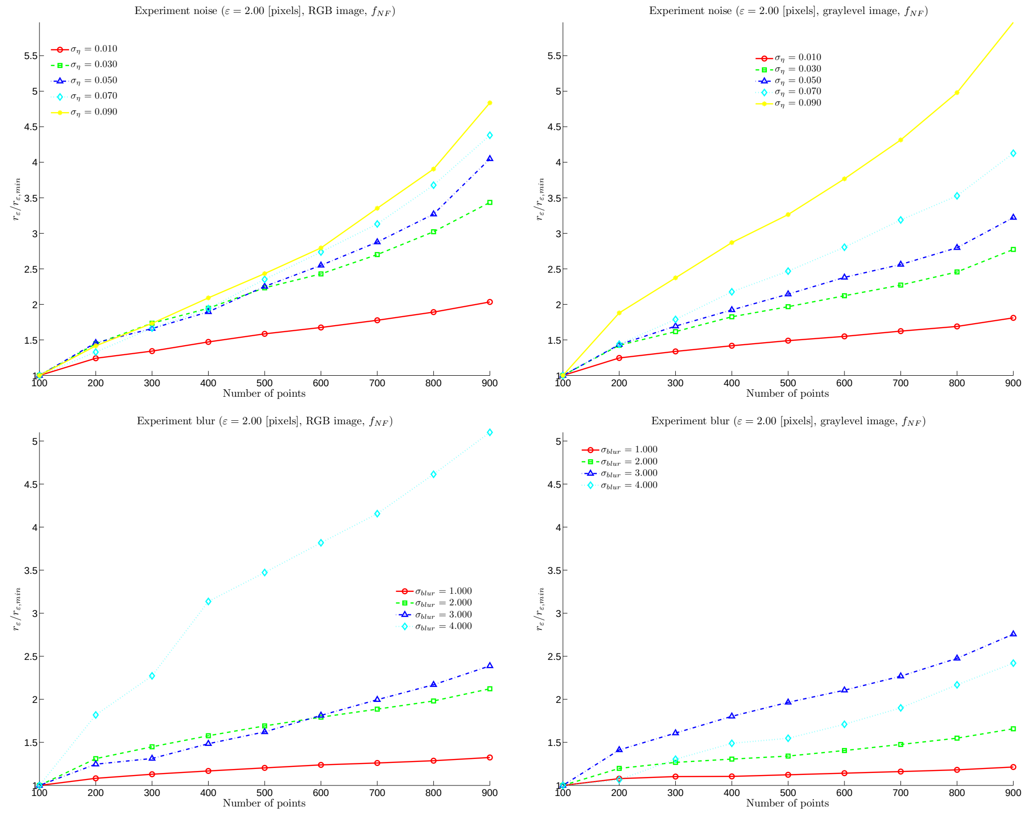


Figure 3.15: Repeatability variation for photometric distortions for $\varepsilon = 2$.

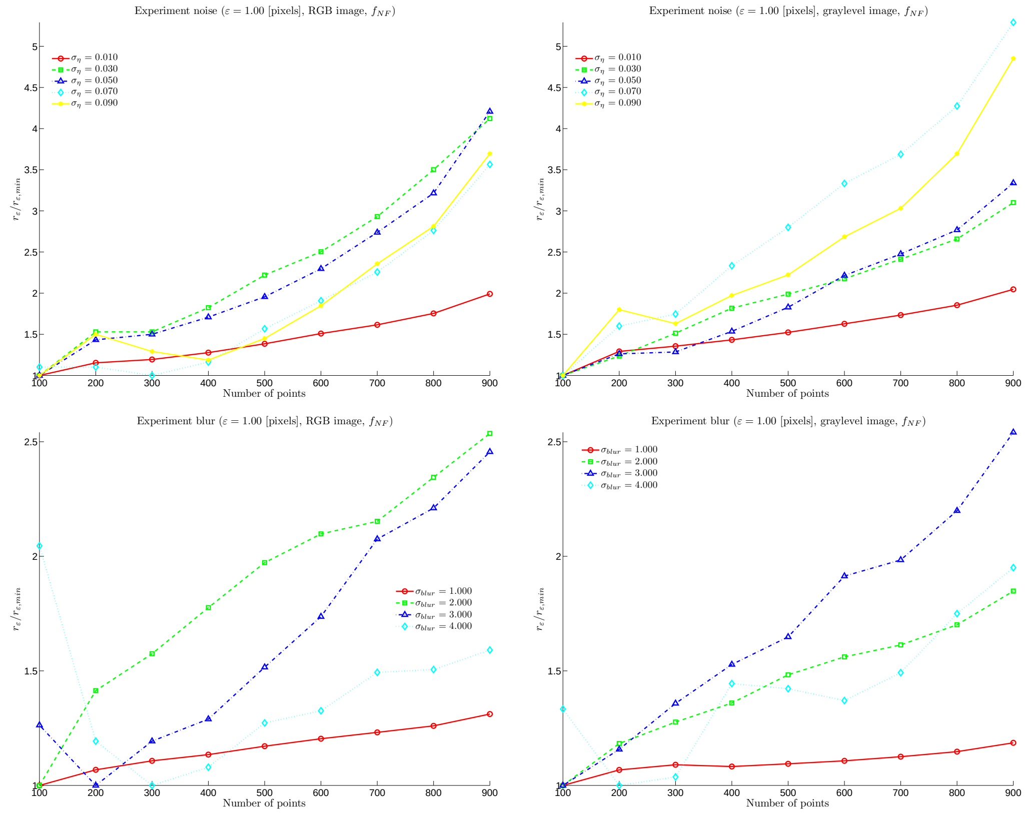


Figure 3.16: Repeatability variation for photometric distortions for $\varepsilon = 1$.

Chapter 4

Drums, Curve Descriptors and Affine Invariant Region Matching

"In mathematics you don't understand things.

You just get used to them."

J. von Neumann

Motivated by the possibility of establishing image correspondences using curve features rather than interest points, we introduce in this chapter a novel curve/region descriptor based on the modes of vibration of an elastic membrane. In particular:

- We introduce and study the theoretical properties of a novel *physically motivated curve/region descriptor* based on the modes of vibration of a membrane. We revisit the problem of *curve isospectrality* within the image analysis domain (see Section 4.2).

- We develop a *normalization procedure* that allows us to characterize the shape of a curve independent of its affine distortions (see Section 4.3).

- We propose a method to couple the descriptor and the normalization procedure to robustly *match curves between images* taken from different points of view (see Section 4.3).

- We provide extensive experimental results to measure the performance of our descriptor using both synthetic and real images. We also compare our descriptor with state of the art curve/region descriptors (see Section 4.4).

4.1 Introduction

The quest for efficient curve and region descriptors has been one of the leading themes in the image analysis community. In general, good descriptors should be invariant under an appropriate class of geometric transformations (e.g. rotation-scaling-translation or affine), robust in the presence of noise, efficient to compute and easy to compare. Zhang et al. [135] classified the curve description approaches into two groups: *contour-based* and *region-based* methods. Each of these groups is further subdivided into two subgroups containing *global* or *structural* approaches. Some of the recently proposed descriptors fall in the contour-based category, as the *curvature scale space* (CSS) descriptor [90] (which has been standardized within the MPEG-7 framework) and the *shape context matrices* [8]. Some others belong to the class of region-based methods, like the descriptors based on *moments* (geometric [37], Zernike, also standardized within the MPEG-7 framework, and Legendre [123]), on *region frequency representations* (Fourier descriptors [134]), on the *medial axis transform* [93] and on *shock graphs* [117]. Recently Gorelick et

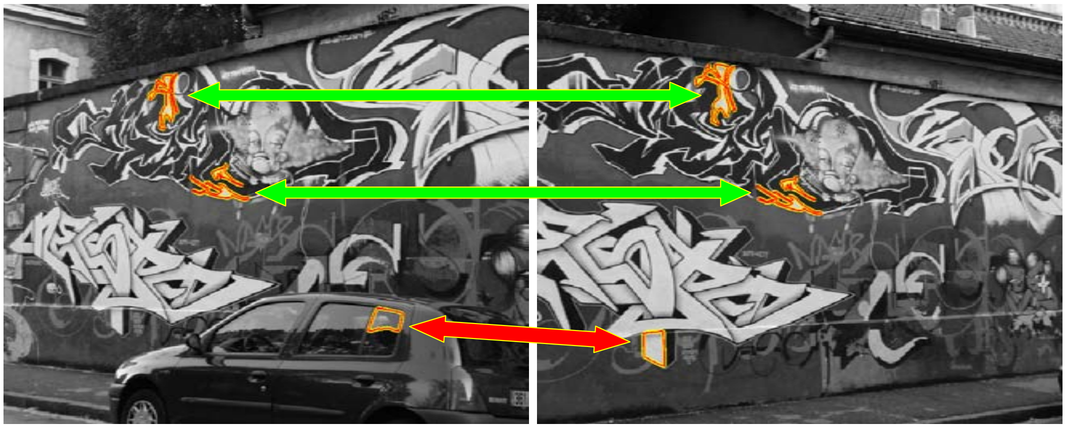


Figure 4.1: A set of curves extracted and matched in a pair of images taken from different point of views. The shape of the matched curves are similar, however only two matches are consistent with the geometry of the scene.

al. exploited the properties of the *Poisson equation* to characterize shapes and to derive a set of features that can serve as descriptors [46].

We are interested in a descriptor that can be used in the context of image registration where we will focus on image features defined by Jordan curves (i.e. curves that are closed and do not cross themselves, see the example in Figure 4.1 and the Definition (B.1.1)). By using curve features rather than point features one can avoid the problem of detecting neighborhoods that transform consistently with the images (see [87, 89] for extensive surveys regarding the neighborhood adaptation problem and [68, 72] for image registration systems based on curve features). However this comes at the price of developing a matching strategy that is at least affine invariant, so that perspective distortions can be handled robustly.

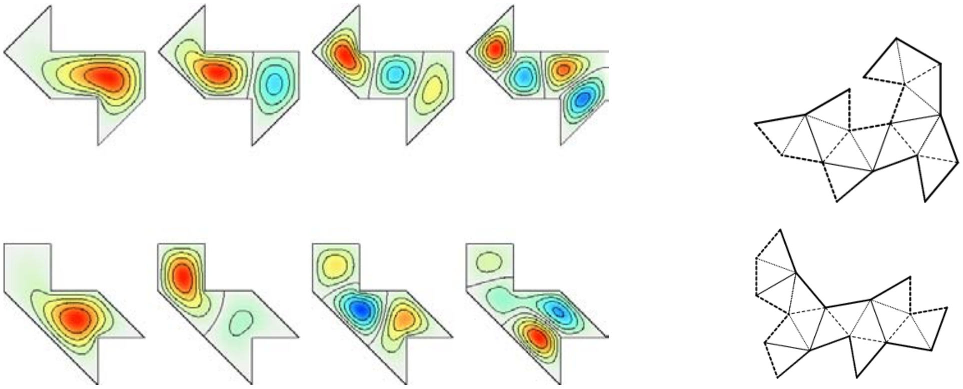
The approaches mentioned before either do not always satisfy the MPEG-7 requirements or they are not completely suitable to establish affine invariant matches between curves. The computation of the CSS descriptors is quite demanding, the algorithm converges slowly if the curve is very complex (i.e. the curve presents many points with large curvature) and depends on some empirical parameters that need to be fine tuned. The comparison of CSS descriptors is not simple. Shape context matrices provide local curve descriptors that are not very compact (since they consist of the coefficients of a matrix) and their comparison is not very fast. Moment invariants of higher orders do not have a clear physical interpretation and the matching procedure requires a normalization process to compensate for the different dynamic range of the moments of different orders. However recently Zhang et al. [133] experimentally showed that Fourier descriptors and Zernike moment descriptors perform better than the CSS descriptors. Shock graphs are very suitable in scenarios where the similarity between curves is defined in terms of structure, but are not the ideal solution if the notion of equivalence is defined within the class of some specific geometric transformation. Moreover the computational complexity for extracting these descriptors and matching them is quite high. Finally note that, with the exception of moments, it is not straightforward to extend the descriptors listed above to represent also the intensity pattern inside the curve. For an extensive quantitative comparison of region descriptors that explicitly take into account the intensity pattern within the region, the interested reader should refer to the survey by Mikolajczyk et al. [88].

In this chapter we will present a curve descriptor that partly satisfies the six principles set by MPEG-7 (which are good retrieval accuracy, compact features, general application, low computation complexity, robust retrieval performance, and hierarchical coarse to fine representation) and a few other important requirements, such as being Rotation Scaling and Translation (RST) invariant, having a clear physical interpretation and being capable of taking into consideration the intensity content of a closed contour (when the curve identifies an image region). The descriptor we propose is novel in the sense that it combines intimately both the information regarding the shape of a region and its intensity content.

This chapter is structured as follows. Section 4.2 introduces the Helmholtz descriptor, it discusses its analytical properties and presents the numerical scheme used to compute the descriptor. Section 4.3 will describe a preprocessing step that aims at extracting the *shape* of a curve to obtain an affine invariant matching algorithm. In Section 4.4 we will show some experimental results and we will evaluate the performance of the descriptors. Finally the conclusions are presented in Section 4.5.

4.2 The Descriptor

In 1966, the mathematician M. Kac published his famous paper entitled “Can One Hear the Shape of a Drum?” [58]. Kac was interested in understanding whether the knowledge of the modes of vibration of a drum was sufficient for univocally inferring the geometric structure of the drum itself. The problem posed by Kac can be related to the problem of constructing curve or region descriptors. In fact, if we imagine that the curve we want to label defines the contour of a drum, it is reasonable to think that the spectrum of such a curve (in terms of modes of vibration) could be an appealing descriptor, given the fact that it can be easily made RST invariant and has a strong physical characterization. Moreover the intensity inside the image region defined by the curve can be used to model the



(a) (b)

Figure 4.2: Figure (a) (courtesy of Dr. T. Driscoll) shows the first four eigenmodes of an isospectral domain. Figure (b) (courtesy of Dr. Buser, Dr. Conway, Dr. Doyle and Dr. Semmler) shows another example of an isospectral domain.

physical properties of the membrane so that the modes of vibration are related not only to the structure of the boundary but also to the region content. With this in mind, the answer to Kac's question becomes crucial, i.e. we would like to have the normal modes of vibration of a drum to identify univocally its geometry (so that we can establish a bijection between the space of the Jordan curves modulo a given transformation and the curve descriptors).

The problem posed by Kac remained unsolved until 1992 when the mathematicians C. S. Gordon, D. L. Webb and S. Wolpert proposed a pair of isospectral drums having the same area and perimeter but different contours. In other words "One Cannot Hear the Shape of a Drum" [45, 15, 30] (see Figure 4.2 for some examples of isospectral drums). Even though for our purposes this fact is unfortunate, since it implies that there may exist curves that are not related by an RST transformation and nonetheless have the same spectrum (i.e. possibly the same descriptor), the experiments presented in Section 4.4 will show how this problem has a limited impact in real life scenarios. Note that an application of the Laplace operator (deeply connected with the spectral properties of planar domains) has also been explored by Saito [111] for image analysis applications and that Sclaroff and Pentland introduced the idea of describing objects in terms of generalized symmetries that remain defined by the object's eigenmodes [114].

In the following subsections we will describe in detail the proposed curve descriptor and the numerical scheme used to compute it.

4.2.1 The Helmholtz Equation

Let Γ be a Jordan curve corresponding to the boundary of Ω, an open subset of $\mathbb{R}^2$. The vibration of the membrane of a drum whose contour is defined by Γ is expressed by the function $w(\boldsymbol{x}, t) : \bar{\Omega} \times \mathbb{R} \to \mathbb{R}$ which solves the wave equation:

$$\triangle w(\boldsymbol{x}, t) - \frac{1}{v(\boldsymbol{x})^2} \frac{\partial^2 w}{\partial t^2}(\boldsymbol{x}, t) = 0$$

where $\triangle$ denotes the Laplacian operator, t indicates time and $v(\boldsymbol{x}) > 0$ indicates the phase velocity of the membrane.[1] This equation can be solved via separation of variables, assuming that w can be decomposed into a spatial part and into a temporal part according to $w(\boldsymbol{x}, t) = u(\boldsymbol{x})q(t)$. It can be shown that the spatial part solves the Helmholtz equation, i.e. the elliptic partial differential equation:

$$\triangle u(\boldsymbol{x}) + \lambda \frac{1}{v(\boldsymbol{x})^2} u(\boldsymbol{x}) = 0 \tag{4.1}$$

[1] For a real membrane the phase velocity is proportional to $\sqrt{\frac{T}{\sigma(\boldsymbol{x})}}$, where T denotes the membrane tension (expressed in Newtons over meters) and σ the membrane density (expressed in kilograms per square meter, and function of the spatial position $\boldsymbol{x}$).

where λ is a suitable scalar. The corresponding boundary problem with Dirichlet boundary conditions is:

$$-\triangle u(\boldsymbol{x}) = \lambda \frac{1}{v(\boldsymbol{x})^2} u(\boldsymbol{x}) \qquad \text{for } \boldsymbol{x} \in \Omega \tag{4.2a}$$

$$u(\boldsymbol{x}) = 0 \qquad \text{for } \boldsymbol{x} \in \Gamma \tag{4.2b}$$

4.2.2 The Descriptor

Our idea is to use the first $N_\lambda + 1$ eigenvalues associated with the Helmholtz equation (4.2) to build an RST invariant descriptor for the curve Γ (in the case where $v(\boldsymbol{x}) = v = \text{const}$) or for the image patch contained in the region Ω (if we set[2] $v(\boldsymbol{x})^2 = \frac{1}{I_s(\boldsymbol{x})}$, where $I_s(\boldsymbol{x})$ denotes the smoothed version of the image intensity at point $\boldsymbol{x}$). As explained in more detail in Appendix C all the eigenvalues associated with (4.2) are real and positive and they can be sorted in order of increasing value: $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$ with $\lambda_k \to \infty$ as $k \to \infty$. These observations justify the following definition:

Definition 4.2.1 *Let Γ be a Jordan curve and let $\lambda_1, \ldots, \lambda_{N_\lambda+1}$ be the first $N_\lambda+1$ eigenvalues that solve (4.2). The corresponding Helmholtz Descriptor (HD) is defined as:*

$$\boldsymbol{F}(\Omega) \overset{\text{def}}{=} \begin{bmatrix} \frac{\lambda_1}{\lambda_2} & \frac{\lambda_2}{\lambda_3} & \ldots & \frac{\lambda_{N_\lambda}}{\lambda_{N_\lambda+1}} \end{bmatrix}^T \in \mathbb{R}^{N_\lambda} \tag{4.3}$$

The invariance of the descriptor with respect to an RST transformation can be understood by observing that a vibrating membrane will produce the same tones when it is rotated and translated, and that a scaling will only affect their amplitude. This intuition is formalized in the following lemma:

[2] The physical intuition behind this choice is that the membrane density at $\boldsymbol{x}$ is directly proportional to the image intensity at the point $\boldsymbol{x}$.

Lemma 4.2.2 *Consider the two Jordan curves Γ_1 and Γ_2 related by an RST transformation:*

$$\Gamma_2 = \{\boldsymbol{x}_2 \in \mathbb{R}^2 : \textit{there exists } \boldsymbol{x}_1 \in \Gamma_1 \textit{ such that } \boldsymbol{x}_2 = sR\boldsymbol{x}_1 + \boldsymbol{t}\}$$

where $s \in \mathbb{R}$ is the scaling factor, $R \in SO(2)$ is a rotation matrix and $\boldsymbol{t} \in \mathbb{R}^2$ is a translation vector. Let also $v_2(\boldsymbol{x}_2) = v_1(\boldsymbol{x}_1)$. Then $\boldsymbol{F}(\Gamma_1) = \boldsymbol{F}(\Gamma_2)$.

Proof: The proof of this lemma follows from the definition of the Laplacian in an orthogonal coordinate system, which is:

$$\triangle = \frac{1}{h_1 h_2}\left[\frac{\partial}{\partial x_1}\left(\frac{h_2}{h_1}\frac{\partial}{\partial x_1}\right) + \frac{\partial}{\partial x_2}\left(\frac{h_1}{h_2}\frac{\partial}{\partial x_2}\right)\right]$$

where h_1 and h_2 are the scale factors of the first fundamental form. It can be easily verified that for a scaling and an arbitrary rotation we have $h_1 = h_2 = s$. Therefore we can write $\frac{1}{s^2}\triangle u_2(\boldsymbol{x}_2) = \triangle u_1(\boldsymbol{x}_1)$. Thus the eigenpairs that solve:

$$\begin{aligned} -\triangle u_1(\boldsymbol{x}_1) &= \lambda \frac{1}{v_1(\boldsymbol{x}_1)^2} u_1(\boldsymbol{x}_1) && \text{for } \boldsymbol{x}_1 \in \Omega_1 \\ u_1(\boldsymbol{x}_1) &= 0 && \text{for } \boldsymbol{x}_1 \in \Gamma_1 \end{aligned}$$

can be used to construct the solutions for:

$$\begin{aligned} -\triangle u_2(\boldsymbol{x}_2) &= (s^2\lambda) \frac{1}{v_2(\boldsymbol{x}_2)^2} u_2(\boldsymbol{x}_2) && \text{for } \boldsymbol{x}_2 \in \Omega_2 \\ u_2(\boldsymbol{x}_2) &= 0 && \text{for } \boldsymbol{x}_2 \in \Gamma_2 \end{aligned}$$

by scaling the eigenvalues by s^2 and by letting $u_2(sR\boldsymbol{x}_1 + \boldsymbol{t}) = u_1(\boldsymbol{x}_1)$. Since the components of the descriptors are ratios of eigenvalues, the scaling factor vanishes and the assertion holds true. ■

4.2.3 Numerical Scheme

The second order finite difference scheme we used to solve (4.2) is a reasonable compromise between accuracy and computational complexity. The step size of the $N \times N$ discretization mesh is calculated according to:

$$h = \frac{\max_{\boldsymbol{x}\in\Gamma} \|\boldsymbol{x} - \boldsymbol{m}(\Omega)\|}{\Delta} \tag{4.6}$$

where $\boldsymbol{m}(\Omega)$ is the center of gravity of the region Ω (which will be defined formally in Section 4.3) and Δ is a parameter that defines the mesh resolution. The spatial derivatives are approximated by the second order central difference formulae:

$$\frac{\partial^2 u}{\partial x^2}(\boldsymbol{x}) \approx \frac{u(x+h,y) - 2u(x,y) + u(x-h,y)}{h^2}$$
$$\frac{\partial^2 u}{\partial y^2}(\boldsymbol{x}) \approx \frac{u(x,y+h) - 2u(x,y) + u(x,y-h)}{h^2}$$

which provide the discretized version of (4.2a) that reads as:

$$-\frac{u_{p+1,q} + u_{p-1,q} + u_{p,q+1} + u_{p,q-1} - 4u_{p,q}}{h^2} = \lambda \frac{1}{v_{p,q}^2} u_{p,q}$$

where $0 \leq p \leq N-1$ and $0 \leq q \leq N-1$ are the indices of the mesh points. Under these assumptions, the solution for (4.2) is obtained by solving a generalized eigenvalue problem:

$$L\boldsymbol{u} = \lambda V \boldsymbol{u}$$

where the linear operator L is given by the sparse symmetric matrix:

$$L = -\frac{1}{h^2}\begin{bmatrix} A & I_N & 0 & \dots & 0 \\ I_N & A & I_N & \dots & 0 \\ 0 & I_N & A & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & A \end{bmatrix} \in \mathbb{R}^{N^2 \times N^2} \quad A = \begin{bmatrix} -4 & 1 & 0 & \dots & 0 \\ 1 & -4 & 1 & \dots & 0 \\ 0 & 1 & -4 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -4 \end{bmatrix} \in \mathbb{R}^{N \times N}$$

Note that I_N is the $N \times N$ identity matrix, the vector $\boldsymbol{u} \in \mathbb{R}^{N^2}$ is obtained by row scanning the discretization grid so that: $\boldsymbol{u}_{Np+q} = u(x_p, y_q)$ and V is a diagonal matrix such that $V_{Np+q} = \frac{1}{v(x_p,y_q)^2}$. As we anticipated before, $v(x_p, y_q)^2$ is chosen to be inversely proportional to the smoothed version of the image $I_s(x_p, y_q)$, which is obtained via a convolution with an isotropic Gaussian kernel with standard deviation σ (in our current implementation the standard deviation is set to be equal to 2.5 pixels). This is done to ensure that v satisfies the required analytic smoothness properties. The size of the problem can be reduced by removing the entries of the vector $\boldsymbol{u}$ that correspond to the points outside of the domain Ω or to the points on the boundary. This is equivalent to removing the corresponding rows and columns in the matrix L and V: after the reduction the matrix L is no more block tridiagonal (as shown in the sparsity pattern of Figure 4.3(b)) but it still is diagonally dominant. In our implementation the eigenvalues/eigenvectors are computed using the Fortran library ARPACK [66] (accessed through Matlab) that takes advantage of the sparse and symmetric structure of L. To improve the numerical stability of the algorithm we balance the matrices by scaling them, so that $\|L\|_\infty = \|V\|_\infty = 1$ and we solve the modified sparse eigenvalue problem:

$$\left(V^{-\frac{1}{2}} L V^{-\frac{1}{2}}\right) \boldsymbol{w} = \mu \boldsymbol{w}$$

where $\boldsymbol{u} = V^{-\frac{1}{2}} \boldsymbol{w}$ and μ is the scaled eigenvalue. Figure 4.3(b) shows an example of the matrix L associated with the region outlined by the green boundary in Figure 4.3(a). Figure 4.3(c) displays the third eigenmode that solves (4.2). The bumps on the eigenmode surface follow from the fact that the membrane density is proportional to the image intensity. The computation of the Helmholtz Descriptor (HD) in the non uniform case takes on average 1.5 seconds on a 2.8Ghz Pentium

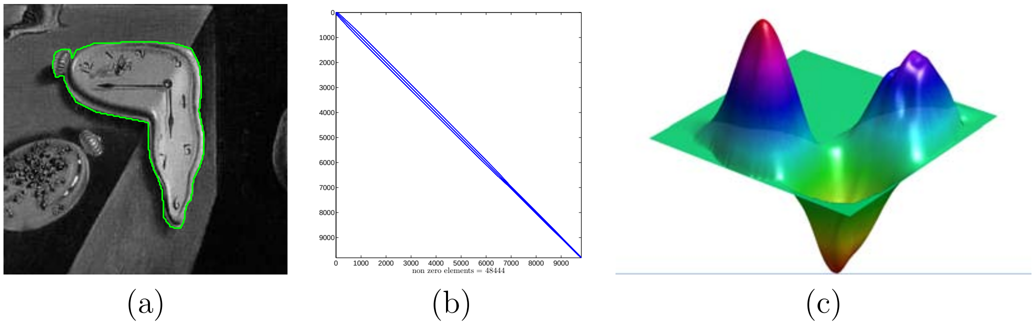


Figure 4.3: Figure (b) displays the sparsity pattern of the matrix L, i.e. it shows a dot for each non zero entry of L. The matrix L is associated with the image region defined by the green boundary in (a) (cropped from the painting *Persistence of Memory* by Salvador Dalí). The size of the matrix L is about $10^4 \times 10^4$ but only $4.8{\cdot}10^4$ elements are non zero. Figure (c) shows the third eigenmode of the Helmholtz equation (4.2).

4 for $\Delta = 30$.

4.2.4 Comparing the Descriptors

As mentioned before, it has been theoretically proven that there exist different curves that have the same spectrum. However this event is quite rare (where the notion of "rare" can be formalized more precisely, see [45]) as the experiments presented in Section 4.4 will confirm. Because of this, the similarity between the descriptors is defined in terms of the weighted Euclidean:

$$d(\boldsymbol{F}(\Omega_1), \boldsymbol{F}(\Omega_2)) = \|\boldsymbol{F}(\Omega_1) - \boldsymbol{F}(\Omega_2)\|_W = \sqrt{\sum_{k=1}^{N} w_k \left[\boldsymbol{F}_k(\Omega_1) - \boldsymbol{F}_k(\Omega_2)\right]^2} \tag{4.7}$$

where the weights are defined according to $w_k \overset{\text{def}}{=} \exp\left(\frac{k-1}{N_\lambda - 1} \log \rho\right)$. The parameter ρ defines the ratio of the weight of the last component of the descriptor with respect to the first one. Experimentally we found that $\rho = 0.75$ is a sensible choice.

The rationale behind the introduction of a weighted distance is related to the physical interpretation of the components of the descriptor: the coefficients with larger indices are associated with the fast modes of vibration of the membrane. These modes are more sensitive to perturbations of the shape of the curve and therefore it is reasonable to weight them less when comparing two curve/regions (in [141] some numerical simulations confirmed that the eigenvalues with larger indices are those more affected by the morphological noise). On the other hand the smallest eigenvalues of a matrix are those more affected by the finite precision mathematical operations.

Remark 4.2.3 *In general the task of studying analytically how the spectrum of a region is affected by the perturbations of the boundary is a complex problem. Even if this problem goes well beyond the scope of this chapter, we would like to mention the approaches described in the classical book of Kato ([61], ch. 6, p. 423) and in two recent papers by Noll [99] and by Ngo [96] that attempt to relate quantitatively the perturbations of the domain boundary to the value of the eigenvalues. It is also possible to approach the problem after the Helmholtz equation has been discretized, by considering morphological perturbations that correspond to the removal of rows and columns from L and V and evaluating the bounds on the eigenvalues defined by the interlacing theorems thoroughly discussed in [53, 80].*

4.3 Achieving Affine Invariance

The descriptors we have introduced in Section 4.2 are RST invariant. However very often it is necessary to match curves or image regions in an affine invariant

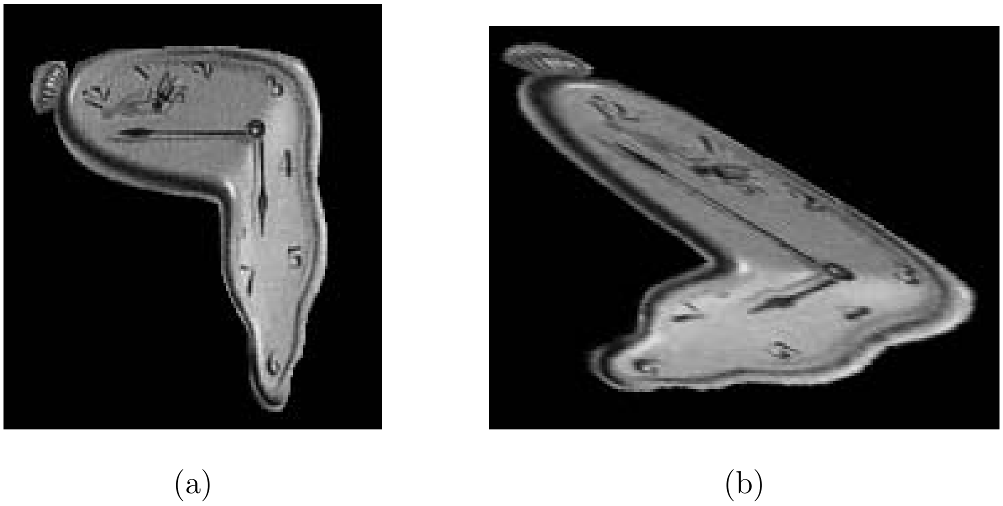


Figure 4.4: An example of two image regions related by an affine transformation (cropped from the painting *Persistence of Memory* by Salvador Dalí).

fashion. As an example, consider planar curves imaged from two different viewpoints using a distant camera, where distant is with respect to the camera focal length. In this case the perspective distortion can be approximated by an affine transformation (see the examples in Figure 4.1 and Figure 4.4). We will describe in detail a procedure that allows us to map a curve (or an image region) in a normalized coordinate system where affine-related objects become congruent modulo a geometric rotation (a discussion of related approaches can be found in [1] Chapter 5, [118] and [139]). First we will consider the case where the content of the region is uniform (uniform case) and then we will generalize the results to cases where we take into consideration the intensity content (non uniform case).

4.3.1 Uniform Case

Let's first introduce the following quantities:

- Let $V(\Omega) \overset{\text{def}}{=} \int_{\Omega} d\boldsymbol{x}$ be the area of Ω, where $d\boldsymbol{x}$ is the infinitesimal area element.

- Let $\boldsymbol{m}(\Omega) \overset{\text{def}}{=} \frac{1}{V(\Omega)} \int_\Omega \boldsymbol{x} \, d\boldsymbol{x}$ be the centroid of Ω.
- Let $\Sigma(\Omega) \overset{\text{def}}{=} \frac{1}{V(\Omega)} \int_\Omega [\boldsymbol{x} - \boldsymbol{m}(\Omega)] [\boldsymbol{x} - \boldsymbol{m}(\Omega)]^T d\boldsymbol{x}$ be the covariance of Ω.

Definition 4.3.1 *Let Γ be a Jordan curve. The shape of Γ is a new Jordan curve such that:*

$$S(\Gamma) \overset{\text{def}}{=} \left\{ \boldsymbol{s} \in \mathbb{R}^2 : \boldsymbol{s} = \Sigma(\Omega)^{-\frac{1}{2}} [\boldsymbol{x} - \boldsymbol{m}(\Omega)] \ \textit{for } \boldsymbol{x} \in \Gamma \right\} \tag{4.8}$$

This definition is important because it allows us to relate affine-transformed curves, as stated in the following theorem and illustrated in Figure 4.5.

Theorem 4.3.2 (Uniform Normalization) *Let Γ_1 and Γ_2 be two Jordan curves related by an affine transformation:*

$$\Gamma_2 = \left\{ \boldsymbol{x}_2 \in \mathbb{R}^2 : \exists \boldsymbol{x}_1 \in \Gamma_1 \ \textit{such that } \boldsymbol{x}_2 = A\boldsymbol{x}_1 + \boldsymbol{b} \right\}$$

where $A \in \mathbb{R}^{2\times 2}$ is a non-singular matrix and $\boldsymbol{b} \in \mathbb{R}^2$. Then the shapes of Γ_1 and Γ_2 are geometrically congruent via a 2-dimensional rotation.

Proof: Before beginning with the proof we want to emphasize the fact that all the steps are not dependent on the dimension n of the space that hosts the curve. Let $\Gamma_1 = \partial\Omega_1$ and $\Gamma_2 = \partial\Omega_2$. We want to show that the matrix:

$$R \overset{\text{def}}{=} \Sigma(\Omega_1)^{\frac{1}{2}} A^T \Sigma(\Omega_2)^{-\frac{1}{2}} \tag{4.9}$$

establishes the congruence relation between $S(\Omega_1)$ and $S(\Omega_2)$. The first step is verifying that (4.9) is a rotation matrix. To achieve this goal we first prove the following identity:

$$\Sigma(\Omega_2) = A\Sigma(\Omega_1)A^T$$

Since the relation between the area of Ω_1 and Ω_2 is:

$$V(\Omega_2) = \int_{\Omega_2} d\boldsymbol{x}_2 = \int_{\Omega_1} |\det(A)|\; d\boldsymbol{x}_1 = |\det(A)| V(\Omega_1)$$

we can write:

$$\begin{aligned} \boldsymbol{m}(\Omega_2) &= \frac{1}{V(\Omega_2)} \int_{\Omega_2} \boldsymbol{x}_2\; d\boldsymbol{x}_2 \\ &= \frac{1}{|\det(A)|V(\Omega_1)} \int_{\Omega_1} (A\boldsymbol{x}_1 + \boldsymbol{b})\, |\det(A)|\; d\boldsymbol{x}_1 \\ &= A\frac{1}{V(\Omega_1)} \int_{\Omega_1} \boldsymbol{x}_1 d\boldsymbol{x}_1 + \boldsymbol{b}\frac{1}{V(\Omega_1)} \int_{\Omega_1} d\boldsymbol{x}_1 \\ &= A\boldsymbol{m}(\Omega_1) + \boldsymbol{b} \end{aligned}$$

and therefore:

$$\begin{aligned} \Sigma(\Omega_2) &= \frac{1}{V(\Omega_2)} \int_{\Omega_2} [\boldsymbol{x}_2 - \boldsymbol{m}(\Omega_2)]\, [\boldsymbol{x}_2 - \boldsymbol{m}(\Omega_2)]^T\, d\boldsymbol{x}_2 \\ &= \frac{1}{|\det(A)|V(\Omega_1)} \int_{\Omega_1} A\, [\boldsymbol{x}_1 - \boldsymbol{m}(\Omega_1)]\, [\boldsymbol{x}_1 - \boldsymbol{m}(\Omega_1)]^T\, A^T |\det(A)|\; d\boldsymbol{x}_1 \\ &= A\Sigma(\Omega_1)A^T \end{aligned}$$

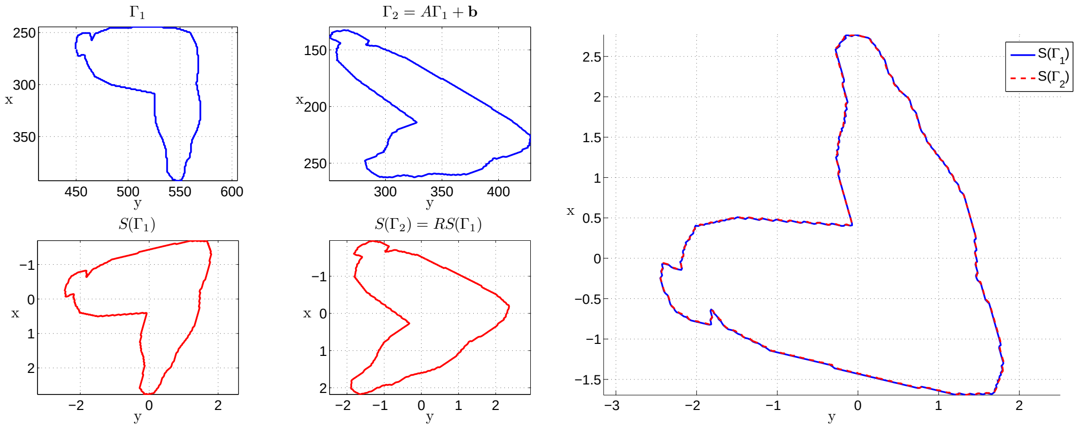
which proves the equality. To show that (4.9) is indeed a rotation matrix it is enough to verify that:

$$R^T R = \Sigma(\Omega_2)^{-\frac{1}{2}}\, \underbrace{A\Sigma(\Omega_1)^{\frac{1}{2}}\Sigma(\Omega_1)^{\frac{1}{2}}A^T}_{\Sigma(\Omega_2)}\, \Sigma(\Omega_2)^{-\frac{1}{2}} = I$$

The proof is concluded observing the following two facts:

- For any $\boldsymbol{s}_1 \in S(\Gamma_1)$ there exits $\boldsymbol{s}_2 \in S(\Gamma_2)$ such that $\boldsymbol{s}_1 = R\boldsymbol{s}_2$.

  To prove this statement note that if $\boldsymbol{s}_1 \in S(\Gamma_1)$, then there exists $\boldsymbol{x}_1 \in \Gamma_1$ such that $\boldsymbol{s}_1 = \Sigma(\Omega_1)^{-\frac{1}{2}}\, [\boldsymbol{x}_1 - \boldsymbol{m}(\Omega_1)]$. Now let $\boldsymbol{x}_2 = A\boldsymbol{x}_1 + \boldsymbol{b}$ and $\boldsymbol{s}_2 =$



(a) (b)

Figure 4.5: The four plots on the left show the curve Γ_1, its affine transformation $\Gamma_2 = A\Gamma_1 + \boldsymbol{b}$ and the corresponding curve shapes $S(\Gamma_1)$ and $S(\Gamma_2)$ in the case where the content of the curve is uniform. The right plot illustrates the congruency between $S(\Gamma_1)$ and $S(\Gamma_2)$. The displayed curves are extracted from Images 4.4(a) and (b).

$\Sigma(\Omega_2)^{-\frac{1}{2}}\left[\boldsymbol{x}_2 - \boldsymbol{m}(\Omega_2)\right] \in S(\Omega_2)$: we want to show that $\boldsymbol{s}_1 = R\boldsymbol{s}_2$. This follows immediately from the chain of equalities:

$$\begin{aligned}
\boldsymbol{s}_1 &= \Sigma(\Omega_1)^{-\frac{1}{2}}\left[\boldsymbol{x}_1 - \boldsymbol{m}(\Omega_1)\right] \\
&= \Sigma(\Omega_1)^{-\frac{1}{2}} A^{-1}\left[\boldsymbol{x}_2 - \boldsymbol{m}(\Omega_2)\right] \\
&= \underbrace{\Sigma(\Omega_1)^{-\frac{1}{2}} A^{-1} \Sigma(\Omega_2)^{\frac{1}{2}}}_{R^{-T}=R} \Sigma(\Omega_2)^{-\frac{1}{2}}\left[\boldsymbol{x}_2 - \boldsymbol{m}(\Omega_2)\right] \\
&= R\Sigma(\Omega_2)^{-\frac{1}{2}}\left[\boldsymbol{x}_2 - \boldsymbol{m}(\Omega_2)\right] = R\boldsymbol{s}_2
\end{aligned}$$

- For any $\boldsymbol{s}_2 \in S(\Gamma_2)$ there exits $\boldsymbol{s}_1 \in S(\Gamma_1)$ such that $\boldsymbol{s}_2 = R^{-1}\boldsymbol{s}_1$

  This claim can be proven similarly to what we just did before.

■

4.3.2 Non Uniform Case

Let $I(\boldsymbol{x})$ be the intensity value of a single channel image at the location $\boldsymbol{x}$; we modify the quantities introduced in Section 4.3.1 as:

- Let $V(\Omega) \overset{\text{def}}{=} \int_\Omega I(\boldsymbol{x}) d\boldsymbol{x}$ be the weighted area of Ω, where $d\boldsymbol{x}$ is the infinitesimal area element.
- Let $\boldsymbol{m}(\Omega) \overset{\text{def}}{=} \frac{1}{V(\Omega)} \int_\Omega I(\boldsymbol{x}) \boldsymbol{x} \, d\boldsymbol{x}$ be the weighted centroid of Ω.
- Let $\Sigma(\Omega) \overset{\text{def}}{=} \frac{1}{V(\Omega)} \int_\Omega I(\boldsymbol{x}) \left[\boldsymbol{x} - \boldsymbol{m}(\Omega)\right] \left[\boldsymbol{x} - \boldsymbol{m}(\Omega)\right]^T d\boldsymbol{x}$ be the weighted covariance of Ω.

In this case Theorem 4.3.2 becomes:

Theorem 4.3.3 (Non Uniform Normalization) *Let Γ_1 and Γ_2 be two Jordan curves related by an affine transformation:*

$$\Gamma_2 = \left\{\boldsymbol{x}_2 \in \mathbb{R}^2 : \exists \boldsymbol{x}_1 \in \Gamma_1 \textit{ such that } \boldsymbol{x}_2 = A\boldsymbol{x}_1 + \boldsymbol{b}\right\}$$

where $A \in \mathbb{R}^{2\times 2}$ is a non-singular matrix and $\boldsymbol{b} \in \mathbb{R}^2$. Moreover suppose that the intensity pattern in Ω_1 and Ω_2 is related according to:

$$I_2(\boldsymbol{x}_2) = I_2(A\boldsymbol{x}_1 + \boldsymbol{b}) = I_1(\boldsymbol{x}_1)$$

Then the shapes of Γ_1 and Γ_2 are geometrically congruent via a 2-dimensional rotation.

Proof: The proof follows exactly the same lines of the proof of Theorem 4.3.2, since it is straightforward to show that:

- $V(\Omega_2) = |\det(A)| V(\Omega_1)$

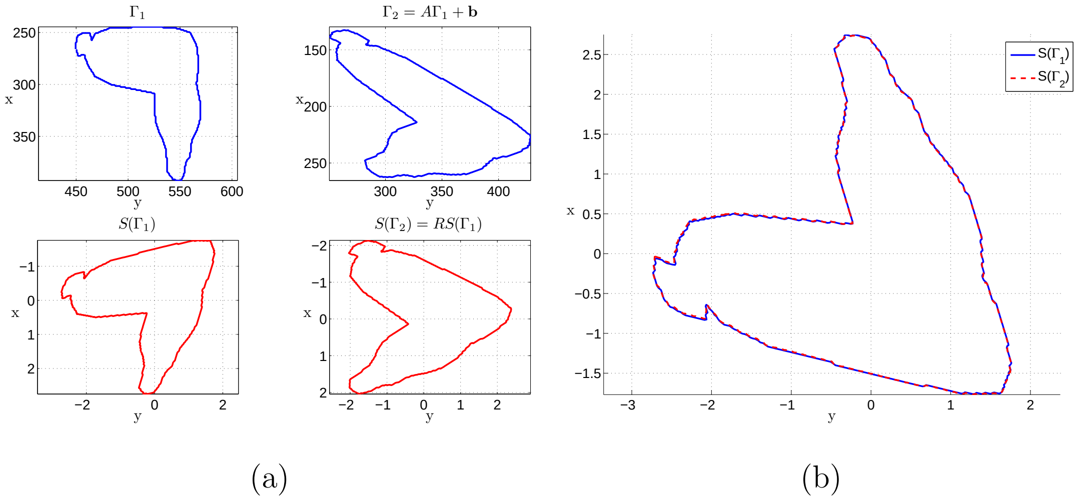


Figure 4.6: The four plots on the left show the curve Γ_1, its affine transformation $\Gamma_2 = A\Gamma_1 + \boldsymbol{b}$ and the corresponding curve shapes $S(\Gamma_1)$ and $S(\Gamma_2)$ in the case where the content of the curve is uniform. The right plot illustrates the congruency between $S(\Gamma_1)$ and $S(\Gamma_2)$. The displayed curves are extracted from Images 4.4(a) and (b).

- $\boldsymbol{m}(\Omega_2) = A\boldsymbol{m}(\Omega_1) + \boldsymbol{b}$

- $\Sigma(\Omega_2) = A\Sigma(\Omega_1)A^T$

also in the presence of the weighting factor related to the image intensity. ■

Figure 4.6 shows an example of the non uniform normalization procedure.

4.3.3 Coupling the Normalization Procedure with the Helmholtz Descriptor

If we want to use the descriptors introduced in Section 4.2 in the context of affine-invariant matching we just have to extract the shape of a curve Γ and then calculate the RST invariant descriptors of $S(\Gamma)$ (using or not the content of the region). This two-step approach can be coupled with any RST invariant curve descriptor (such as the Zernike Moment Descriptors).

4.4 Experimental Results

The experimental results that we will present in this section are divided in two groups. First we will test the performance of the Helmholtz descriptor using a semi-synthetic dataset and then we will use the proposed descriptor to establish matches between images of natural scenes.

4.4.1 Performance Evaluation on a Semi-Synthetic Data Set

The dataset for the experiments described in this section has been extracted from the Amsterdam Library of Object Images (ALOI, see [43]). We considered 250 frontal images of different objects and for each view we synthetically generated 9 other images by applying an homographic transformation that simulates a change in the position of the camera. Each homography is generated following the procedure explained in Section 3.3.2. The images are generated for values of α uniformly distributed in the interval $[0.65, 1.35]$. Further, the image is rotated by a random angle in $[-\pi, \pi]$. Figure 4.7 shows an example of the images generated via this procedure. The objects are segmented using the masking information included in the original ALOI dataset.

In the experiments described in this section, we will compare the performance of the Helmholtz descriptor versus the Zernike Moment Descriptor, which has been shown to perform very well in the context of shape matching and retrieval

[133, 135]. Experimental comparisons of the uniform HD versus the Curvature Scale Space Descriptor can be found in [141]. The performance of the descriptors is evaluated using the *precision-recall curve* calculated (over the dataset described previously) as follows. Each curve Γ (or region Ω) is used in turn as the query. Let $A(\Gamma, N_r)$ denote the set of N_r retrievals (based on the smallest distances (4.7) from Γ in the descriptor space) and $R(\Gamma)$ the set of 10 images in the dataset relevant to Γ. The *precision* is defined as:

$$P(\Gamma, N_r) \overset{\text{def}}{=} \frac{|A(\Gamma, N_r) \cap R(\Gamma)|}{N_r}$$

and measures the proportion of items retrieved that are relevant. Similarly, the *recall* is defined as:

$$C(\Gamma, N_r) \overset{\text{def}}{=} \frac{|A(\Gamma, N_r) \cap R(\Gamma)|}{10}$$

and measures the proportion of relevant items that are retrieved. Note that the same quantities can be defined in the case where the query is a region Ω. The notation $|\cdot|$ denotes cardinality. The precision recall curve is plotted by averaging precision and recall over all Γ, for different values of N_r. On the plots, each marker corresponds to a different value for N_r ranging from 1 to 20. Moreover dashed red lines refer to the Zernike Moment Descriptor (ZMD), whereas continuous blue lines refer to the HD.

Figure 4.8(a) compares the performance of the descriptors after the curves/regions have been normalized using the uniform or non uniform normalization. Both the descriptors have 36 components and are uniformly quantized using 8 bits. For the HD the parameters are $\Delta = 30$, $\sigma = 2.5$ and $\rho = 0.75$. Both for the ZMD and for the HD the performance is better if the regions are normalized using the

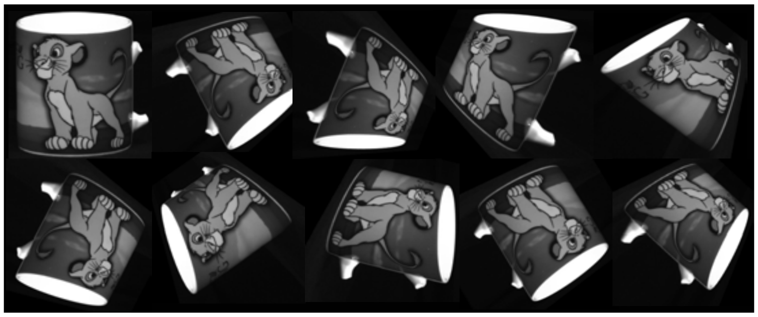
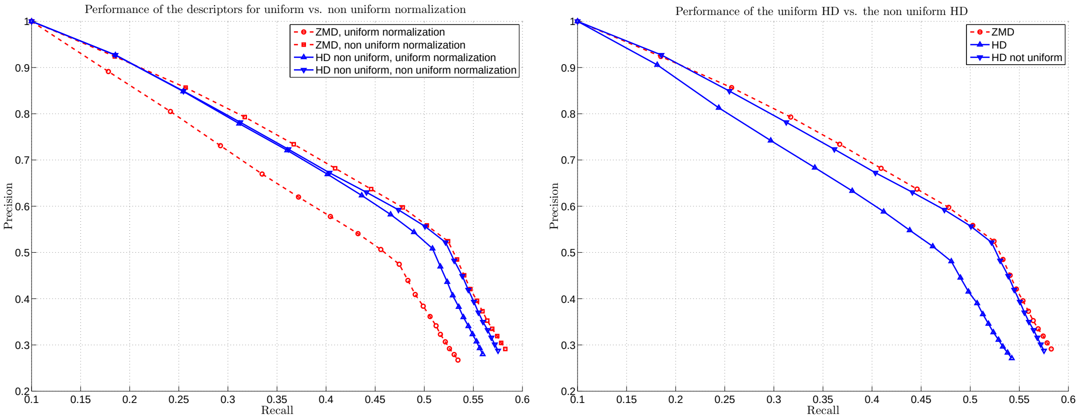
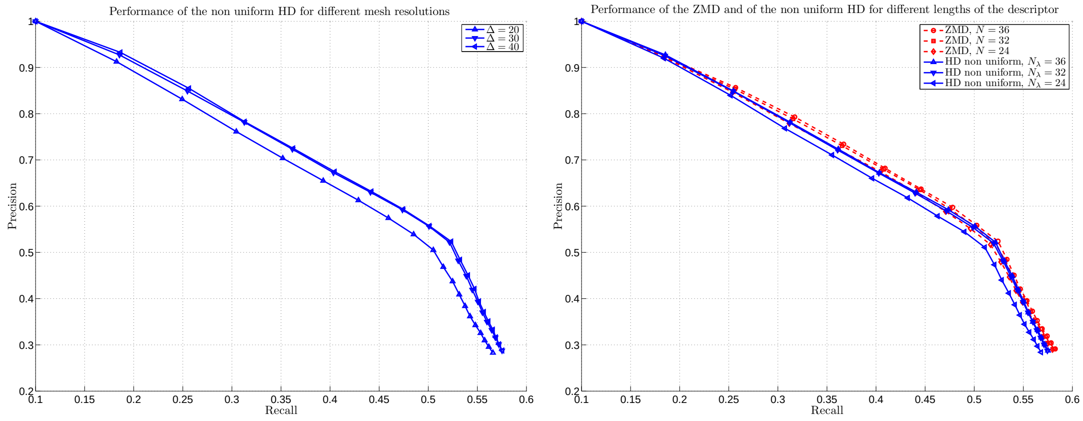


Figure 4.7: A set of images synthesized using the homographies generated using the method described in Section 3.3.2 plus an arbitrary rotation.



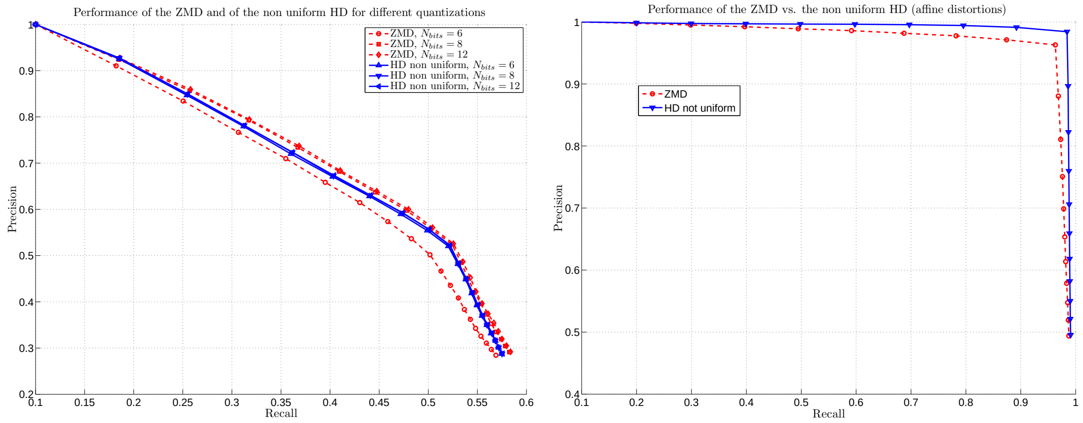
(a) (b)

Figure 4.8: Figure (a) compares the performance of the descriptor in the presence of uniform or non uniform normalization. Figure (b) compares the performance of the uniform HD vs. the non uniform HD. In both experiments the descriptors have 36 components and are uniformly quantized using 8 bits. For the HD the parameters are $\Delta = 30$, $\sigma = 2.5$ and $\rho = 0.75$.



(a) (b)

Figure 4.9: Figure (a) shows the behavior of the non uniform HD for different resolutions of the discretization mesh parameterized by Δ. Figure (b) compares the performance of the ZMD versus the non uniform HD for different lengths of the descriptor. The parameters for the HD are $\sigma = 2.5$ and $\rho = 0.75$.



(a) (b)

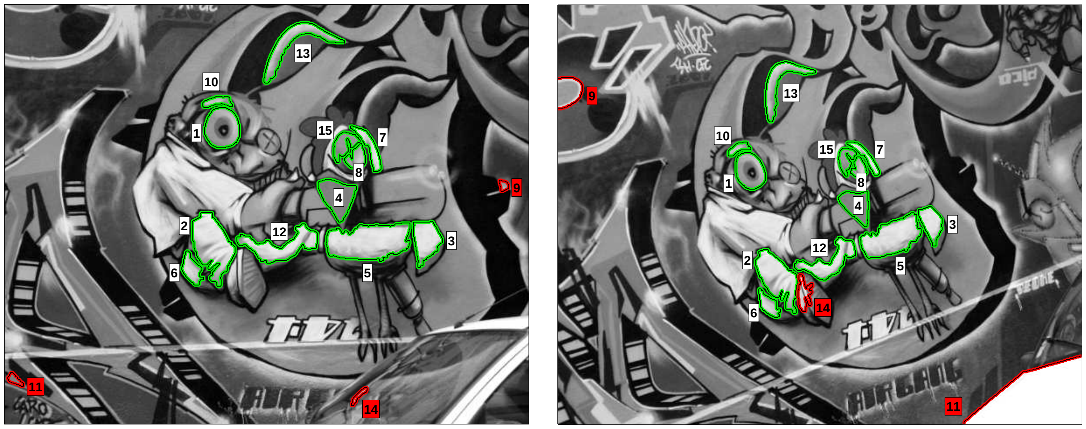
Figure 4.10: Figure (a) displays the precision recall curves for the ZMD and for the non uniform HD while varying the number of bits used to quantize the descriptor components. The parameters for the HD are $\sigma = 2.5$ and $\rho = 0.75$. Figure (b) compares the ZMD and the non uniform HD when the dataset is generated using an affine distortion model for the images.

non uniform procedure described in Section 4.3.2. This can be simply explained observing that the dataset contains several objects that have a similar shape but a different and distinctive image content. The non uniform HD seems to be less affected by the type of normalization used. This can be understood by observing that the descriptor combines the information of the shape with the information of the content of the considered region. Figure 4.8(b) compares the performance of the uniform HD vs. the non uniform HD. The parameters of the descriptors are the same as in the previous experiment. The precision recall curves confirm the intuition that the non uniform Helmholtz descriptor captures the intensity information contained inside the region and that this has a beneficial impact on the overall performance of the approach. Quite surprisingly the performance of the non uniform HD is essentially equivalent to the performance of the ZMD. We believe that this is due to the fact that the numerical scheme used to solve the Helmholtz equation can be refined and improved. We will elaborate more on this claim at the end of this section. Figure 4.9(a) shows the behavior of the non uniform HD for different resolutions of the discretization mesh parameterized by Δ (see Equation (4.6)). As before, the remaining parameters for the HD are $\sigma = 2.5$ and $\rho = 0.75$ with the descriptors coefficients quantized using 8 bits. As it was pointed out in [141], the results indicate that the descriptor is reasonably stable for values of $\Delta \geq 30$. The experiment illustrated in Figure 4.9(b) compares the performance of the ZMD versus the non uniform HD for different lengths of the descriptor. For both of them the performance fluctuations are rather limited. However we can observe a drop in performance for the HD for $N_\lambda = 24$. This might indicate that the components of the Helmholtz descriptor with larger indices bring more information than the corresponding ones for the ZMD. Figure 4.10(a) displays the precision recall curves for the ZMD and for the non uniform HD while varying the number of bits used to quantize the descriptor components. The ZMD presents larger fluctuations than the HD: we hypothesize that this behavior is related to the fact that the coefficients of the Zernike descriptors cover a larger dynamic range than the ratios of the eigenvalues of the Laplacian and hence they are more affected by quantization issues.

From these experiments we conclude that the non uniform Helmholtz descriptor provides a performance (measured in terms of precision recall) that is comparable to the performance obtained by the Zernike Moment Descriptor, which can be considered one of the state of the art descriptors for curve/region description, matching and retrieval [133]. We believe that the numerical method used to compute the solutions of the Helmholtz equation can be greatly improved with an immediate impact on the performances of the HD. This opinion is mainly supported by the observation that the HD provides a joint description of shape and content. The normalization procedure, that can compensate the distortion introduced by an homographic transformation only up to a first order of approximation is also a critical step in the overall procedure. A visual inspection of the eigenmodes of the normalized regions indicates that perturbations due to a non satisfactory normalization may produce completely different Helmoholtz descriptors. Figure 4.10(b) compares the ZMD and the non uniform HD when the dataset is generated using an affine distortion model for the images (and therefore the normalization procedure carries out its task with no approximations): as expected the results obtained using the non uniform Helmholtz descriptors are superior to those obtained using the Zernike descriptors. Regarding the problem of isospectrality introduced in Section 4.2, a manual inspection of a sample set of mismatched curves/regions seems to confirm the intuition that with real imagery the generation of identical spectrums from different curves/regions is an unlikely event. As a concluding remark we want to point out that there exists a tradeoff between the distinctiveness and the robustness provided by the descriptors. In the context of image registration, where we would like to establish matches between images, we would like to avoid situations where the distance between the descriptors of different (but visually similar) curves is small enough to produce a mismatch. On the other hand, a certain degree of robustness is needed to compensate for the approximations introduced by the normalization procedure. Both these two aspects will be experimentally explored in greater detail in the next section.

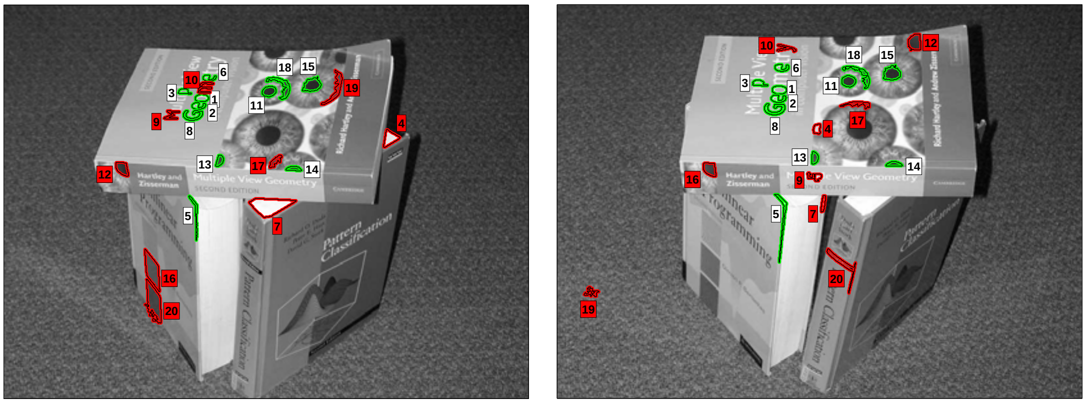
4.4.2 Performance Evaluation on Real Images

The performance of the descriptors has been tested on a set of pairs of real images, representing outdoor and indoor scenes. Every image pair displays the same scene acquired under different points of view. The first stage of the processing consists in finding in each image a set of candidate curves for matching. To accomplish this task we used the level set decomposition of the intensity values of the image, which are known to enjoy several important invariance properties [73]. We observed that for our purposes a good strategy is to slice the intensity profile at two levels, one large and one small. This way it is possible to identify dark regions as well as bright regions. Moreover the size of the extracted regions is large enough so that the intensity content is non uniform. As pointed out in the previous section, this is crucial for the generation of distinctive descriptors in the presence of curves that have very similar shapes. After the Helmholtz descriptors are calculated for each curve/region, the matching is performed using the weighted Euclidean distance (4.7) (with $\rho = 0.75$). In Figures 4.11 to 4.14 we show the results of the curve matching procedure. Figure 4.11 (a) and (b) show that the matching using the HD is facilitated if the detected regions are large enough. More specifically, Region 1 has a round shape that is present elsewhere in the image (e.g. the other eye of the puppet or Region 9 in image (b)). However its distinctive intensity content is likely to be captured by the Helmholtz descriptor. Similar considerations can be extended to the Regions 1, 11 and 15 displayed in Figure 4.12. In particular Region 1 includes the edges of the inside border of the letter "o", making such a region distinguishable from the remaining two. In Figure 4.13, Regions 1, 4 and 5 have similar shapes but we may still argue that the information contained in the intensity pattern can be a relevant to the final matching result. Similar considerations hold for the harbor scene in Figure 4.14. As a final remark, we would like to suggest how this approach could be used to bootstrap the estimation of the mappings (such as homographies or fundamental matrices) that relate the geometry of 3D scenes acquired from points of view separated by a wide base line.



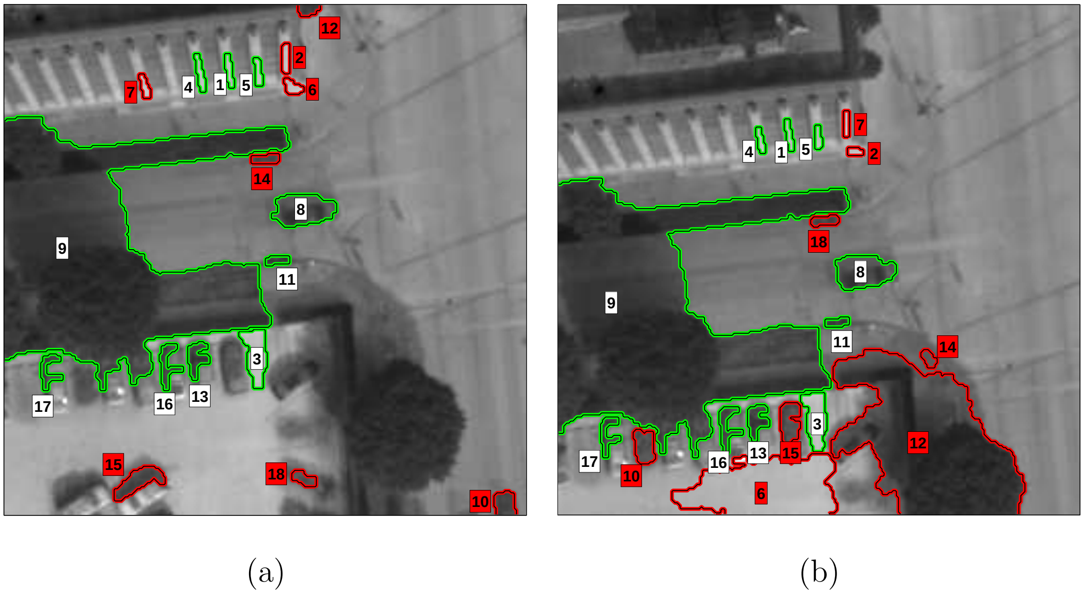
(a) (b)

Figure 4.11: Results of the matching procedure for the Graffiti scene. In all the examples the HD descriptor is composed of 32 components quantized using 8 bits and $\Delta = 30$. The numbers with white background (green region boundaries) identify curve/regions correctly matched, while those with red background (red region boundaries) correspond to mismatches.



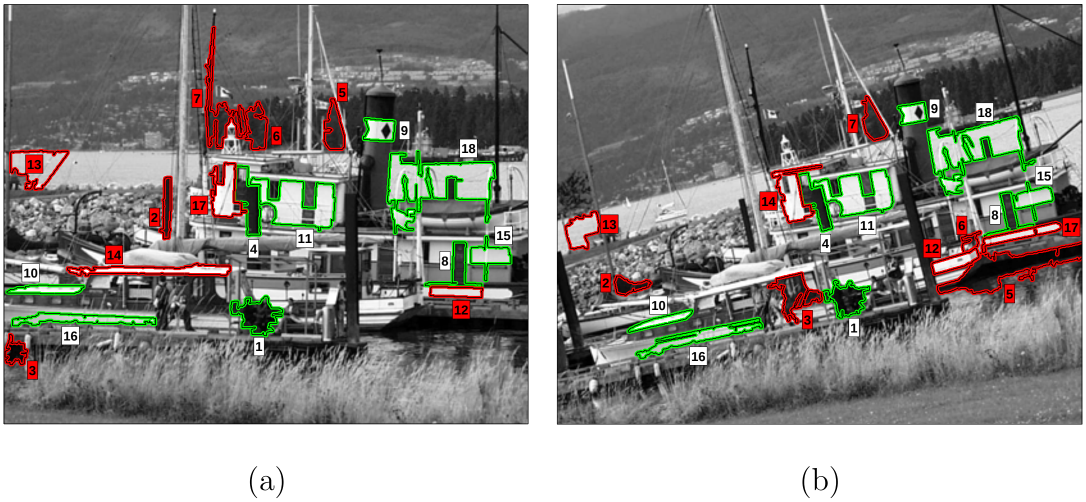
(a) (b)

Figure 4.12: Results of the matching procedure for the Books scene. See the caption of Figure 4.11 for the experimental conditions and the typographical conventions.



(a) (b)

Figure 4.13: Results of the matching procedure for the LA street scene. See the caption of Figure 4.11 for the experimental conditions and the typographical conventions.



(a) (b)

Figure 4.14: Results of the matching procedure for the Harbor scene. See the caption of Figure 4.11 for the experimental conditions and the typographical conventions.

4.5 Conclusions and Future Work

In this chapter we presented a curve/region descriptor that is based on the solution of the Helmholtz equation. This descriptor has a strong physical characterization, since it is related to the modes of vibration of a membrane shaped as the considered region and with a density that is proportional to the region intensity. Together with the descriptor we presented a normalization procedure that is capable of extracting the shape of a curve/region. More precisely, curves (or image regions) are mapped to a normalized coordinate system where affine-related objects become congruent modulo a geometric rotation. The performance of the descriptors has been tested both on a semi-synthetic dataset and on real images and it has been compared with one of the state of the art descriptors, the Zernike moment descriptor. The results of the experiments show that the HD performs well in the context of similarity based curve/region retrieval and curve/region matching. Moreover the normalization procedure proved to be an important tool to compensate for the geometric distortions present in images acquired from different points of view. Both the descriptor and the normalization procedure combine intimately and in an original way the information regarding the shape of the object with the information carried by its visual appearance.

The initial studies that have been presented in this chapter open a number of interesting research perspectives. First of all the calculation of the descriptor would greatly benefit from advanced numerical methods [41, 30, 48, 52] that could solve the Helmholtz equation with an higher degree of accuracy and possibly faster (it is known that finite difference schemes may introduce spurious modes and that there exists a dependence between the grid resolution and the largest index of the eigenpair that can be computed). We also believe in the importance of quantitatively characterizing the influence that the perturbations on the boundaries of the curves have on the coefficients of the descriptors or equivalently the sensitivity of the HD with respect to morphological perturbations. Another interesting research perspective consists understanding the semantics of the descriptors, i.e. how they relate to the visual properties of the curve/regions [32]. We would also like to emphasize the fact that the theory that supports both the normalization procedure and the calculation of the modes of vibration of a membrane is independent of the dimensionality of the considered objects and could be generalized to deal with regions extracted from three dimensional imagery (such as CAT images). Finally we are interested in region detectors that are able to identify image portions that have a rich intensity content and that present a high degree of repeatability in the presence of perspective distortions. We believe that the curves obtained starting from the level set decomposition of the intensity surface of an image [73] could be a good input for the Helmholtz descriptor.

Chapter 5

RANSAC Stabilization

"A probability space is a triple $(\Omega, \mathcal{F}, P)$*."*

A. Kolmogorov

"Probability is the ratio of the number of favorable outcomes
to the number of possible outcomes."

P. Zuliani

Given the need to estimate the parameters of (multiple) geometric or photometric models in the presence of a large number of outliers, we develop a robustification framework that improves the results obtained using RANSAC. The novel contributions of this chapter are:

- The introduction of a *stabilization framework* that improves the quality of estimates obtained using RANSAC in the presence of large uncertainties of the noise scale and multiple instances of the model (see Section 5.3).

- The introduction of a *pseudo-distance* to quantify the dissimilarity between geometric transformations (see Section 5.3).

- The reduction of the problem of *grouping similar models* to the problem of identifying the largest maximal clique in a graph (see Section 5.3).

- The validation of the stabilization framework by means of extensive experiments using both synthetic and real data (see Section 5.4).

5.1 Introduction

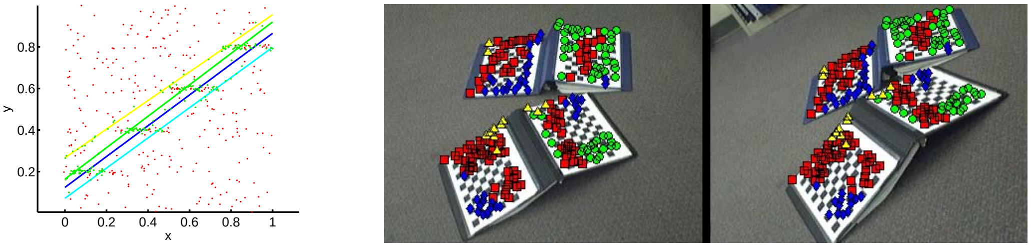
The RANSAC algorithm (RANdom Sample And Consensus) was first introduced by Fischler and Bolles [35] as a method to estimate the parameters of a certain model in the presence of large amounts of outliers (the percentage of outliers can be larger then 50%, which is commonly assumed to be a practical limit in many other statistical methods [55, 110, 79, 120]).

RANSAC has been widely used in the computer vision and image processing community for many different purposes and several modifications have been proposed [128, 20, 97, 131, 126, 21, 142] to improve the speed of the algorithm, the robustness and accuracy of the solution and to decrease the dependency from user defined constants. However, despite these various modifications, the RANSAC algorithm is basically composed of two steps that are repeated in an iterative fashion (hypothesize-and-test framework). First Minimal Sample Sets (MSS) are randomly selected from the input dataset and the model parameters are computed using only the elements of the MSS. The cardinality of the MSS is the smallest sufficient to determine the model parameters (as opposed to other approaches, such as for example least squares, where the parameters are estimated using all the data available with appropriate weights). In the second step RANSAC checks which elements of the full dataset are consistent with the model instantiated with the parameters estimated in the first step.[1] The set of such elements is called Consensus Set (CS) . RANSAC terminates when the probability of finding a better consensus set drops below a certain threshold.

Much of the attention that has been devoted to RANSAC aimed at improving its performance in many different ways. In MLESAC [128] Torr et al. evaluate the quality of the CSs calculating their likelihood. Chum et al. proposed a randomized version of RANSAC [20] to reduce the computational burden to identify a good CS. Chum also proposed to guide the sampling procedure if some a priori information regarding the data distribution is known (PROSAC [21]). A similar attempt was pursued by Tordoff et al. (Guided-MLESAC [126]). Other researchers tried to cope with difficult situations where multiple model instances are present and the noise scale is not known (see the work by Wang and Suter [131] and the multiRANSAC extension described in [142]). In [97] Nistér proposed a paradigm called Preemptive RANSAC that allows real time robust estimation of the structure of a scene and of the motion of the camera.

In our work we are interested in reducing the bias and improving the accuracy of the RANSAC estimate when we do not have precise information regarding the noise scale and when multiple model instances are present in the original dataset. This scenario arises frequently when we want to register the planar structures that are present in an image pair.

[1] If an existing model is supported by few data points the algorithm can miss it, especially when the noise scale is overestimated.



(a) (b)

Figure 5.1: Figure (a) shows an uncorrect fitting of four lines. Figure (b) shows an uncorrect fitting of planar homographies: the shape and color of the markers identifies set of points that fit the same homography. Note that markers belonging to the same group fit homographies that are not consistent with the structure of the scene.

5.1.1 The Problem of the Noise Scale

As mentioned before, the notion of "goodness" for a CS depends on how well its elements fit the instantiated model (even though departures from this paradigm can be found in the MINPRAN estimator [119]). But to identify a CS we first need to define a threshold that distinguishes inliers from outliers, (or, in other words, we need to know the *noise scale*). If this threshold is too small we risk to select only some of the true inliers. On the other hand, if such threshold is too large we may include data points that actually are outliers or pseudo-outliers (i.e. data points which are inliers for a different model) . In both cases the estimate of the model parameters are likely to be biased. Things may become even worse when multiple instances of a model are present in the data: in this case we have to cope both with true outliers as well as with pseudo outliers. As noted in [142] the inaccurate inlier detection for the initial (or subsequent) parameter estimation contributes heavily to the instability of the estimates of the parameters for the remaining models. Figure 5.1 shows two examples that support the previous claim.

5.2 Preliminaries

To facilitate the discussion that follows, it is convenient to introduce a suitable formalism to describe the steps for the estimation of the model parameters and for the construction of the Consensus Set (CS). As usual we will denote vectors with boldface letters and the superscript $^{(h)}$ will indicate the h^{th} iteration. The symbol $\hat{x}$ indicates the estimated value of quantity x. The input dataset which is composed of N elements is indicated by $D = \{\boldsymbol{d}_1, \dots, \boldsymbol{d}_N\}$ and we will indicate a Minimal Sample Set (MSS) with the letter s. Let $\boldsymbol{\theta}\left(\{\boldsymbol{d}_1, \dots, \boldsymbol{d}_h\}\right)$ be the parameter vector estimated using the set of data $\{\boldsymbol{d}_1, \dots, \boldsymbol{d}_h\}$, where $h \geq k$ and k is the cardinality of the MSS.[2] The manifold $\mathcal{M}$ is defined as:

$$\mathcal{M}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \left\{\boldsymbol{d} \in \mathbb{R}^d : f_{\mathcal{M}}(\boldsymbol{d}; \boldsymbol{\theta}) = 0\right\}$$

where $\boldsymbol{\theta}$ is a parameter vector and $f_{\mathcal{M}}$ is a function whose zero level set contains all the points that fit the model instantiated with a given parameter vector. We define the error associated with the datum $\boldsymbol{d}$ with respect to a manifold $\mathcal{M}(\boldsymbol{\theta})$ as the distance from $\boldsymbol{d}$ to $\mathcal{M}(\boldsymbol{\theta})$:

$$e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta})) \stackrel{\text{def}}{=} \min_{\boldsymbol{d}' \in \mathcal{M}(\boldsymbol{\theta})} \operatorname{dist}(\boldsymbol{d}, \boldsymbol{d}')$$

where $\operatorname{dist}(\cdot,\cdot)$ is an appropriate distance function. Using this error metric we define the CS as:

$$S\left(\boldsymbol{\theta}\right) \stackrel{\text{def}}{=} \{\boldsymbol{d} \in D : e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta})) \leq \delta\} \tag{5.1}$$

where δ is a threshold which can either be fixed by the user or estimated automatically [131]. In practical applications we are interested in relating the value

[2] Suppose we want to estimate a line: in this case the cardinality of the MSS is 2, since at least two distinct points are needed to uniquely define a line.

of δ to the statistics of the noise that affects the data. Consider the special case where the distance function is the Euclidean norm so that we can write:

$$e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta})) = \min_{\boldsymbol{d}' \in \mathcal{M}(\boldsymbol{\theta})} \sqrt{\sum_{i=1}^{n} (d_i - d_i')^2} = \sqrt{\sum_{i=1}^{n} (d_i - d_i^*)^2}$$

and suppose the datum $\boldsymbol{d}$ is affected by Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_\eta I)$ so that $\boldsymbol{\eta} = \boldsymbol{d} - \boldsymbol{d}^*$. We want to calculate the value of δ that bounds, with a given probability P_{inlier}, the error generated by a true inlier contaminated with Gaussian noise. More formally we want to find the value δ such that:

$$P[e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta})) \leq \delta] = P_{inlier} \tag{5.2}$$

Following [51], p. 118, we can write the following chain of equations:

$$P[e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta})) \leq \delta] = P\left[\sum_{i=1}^{n} \eta_i^2 \leq \delta^2\right] = P\left[\sum_{i=1}^{n} \left(\frac{\eta_i}{\sigma_\eta}\right)^2 \leq \frac{\delta^2}{\sigma_\eta^2}\right]$$

and since $\eta_i/\sigma_\eta \sim \mathcal{N}(0,1)$, the random variable $\sum_{i=1}^{n} \left(\frac{\eta_i}{\sigma_\eta}\right)^2$ has a χ_n distribution. Hence:

$$\delta = \sqrt{\sigma_\eta^2 F_{\chi_n}^{-1}(P_{inlier})} \tag{5.3}$$

where $F_{\chi_n}^{-1}$ is the inverse cumulative distribution function associated with a χ_n random variable. At this point we want to emphasize how critical is the choice of the noise threshold. First we may have some uncertainties regarding the noise statistics (often we don't know the probability distribution of the error, and even when it is reasonable to assume a Gaussian distribution we may lack a reliable estimate of the standard deviation). Secondly the error threshold depends on the probability P_{inlier}. Too large values of P_{inlier} will produce a large threshold with the risk of including some outliers as well. On the other hand, too small values of P_{inlier} will generate a value for δ which is too small, and possibly some inliers will be discarded.

5.2.1 RANSAC Overview

A pictorial representation of the RANSAC fundamental iteration together with the notation just introduced is shown in Figure 5.2. As mentioned before, the RANSAC algorithm is composed of two steps that are repeated in an iterative fashion (hypothesize-and-test framework). First a MSS $s^{(h)}$ is selected from the input dataset and the model parameters $\boldsymbol{\theta}^{(h)}$ are computed using only the elements of the selected MSS. Then, in the second step, RANSAC checks which elements in the dataset D are consistent with the model instantiated with the estimated parameters and, if it is the case, it updates the current best CS S^* (usually the CS with the largest cardinality). The algorithm terminates when the probability of finding a better CS drops below a certain threshold. In the next paragraphs we will discuss how to estimate the number of iterations that RANSAC is supposed to perform.

How many iterations?

Let q be the probability of sampling from the dataset D a MSS s that produces an accurate estimate of the model parameters (henceforth such an s will be called *stable* MSS). Consequently the probability of picking a non stable MSS (i.e. a MSS that produces a biased estimate of the true model parameter vector) is $1-q$. If we construct h different MSSs, then the probability that all of them are non stable is $(1-q)^h$ (this quantity tends to zero for h going to infinity). We would

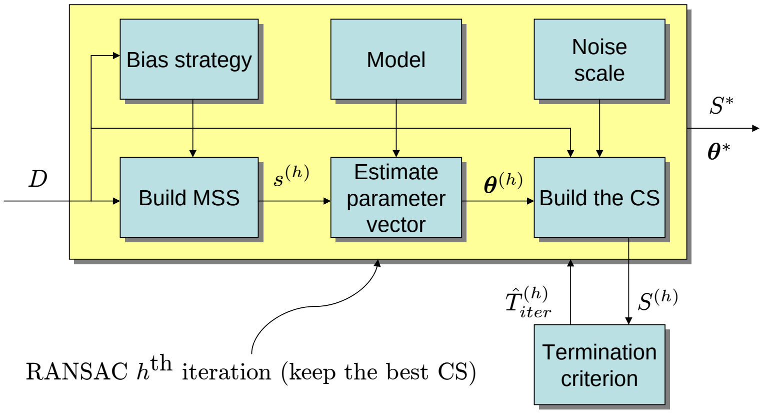


Figure 5.2: Pictorial representation of the fundamental RANSAC iteration.

like to pick h (i.e. the number of iterations) large enough so that the probability $(1-q)^h$ is smaller or equal than a certain probability threshold ε (often called *alarm rate*), i.e. $(1-q)^h \leq \varepsilon$. The previous relation can be inverted so that we write:

$$h \geq \left\lceil \frac{\log \varepsilon}{\log(1-q)} \right\rceil \tag{5.4}$$

where $\lceil x \rceil$ denotes the smallest integer larger than x. Therefore we can set:

$$\hat{T}_{iter} = \left\lceil \frac{\log \varepsilon}{\log(1-q)} \right\rceil \tag{5.5}$$

Constructing the MSSs and Calculating q

If we imagine that the inliers inside the dataset D are noise free, then any MSS entirely composed of inliers will generate the true value of the parameter vector.[3] If all the elements in the dataset have the same probability of being selected, then

[3] As long as we disregard numerical approximations and we pick the noise threshold δ to be an arbitrarily small positive number.

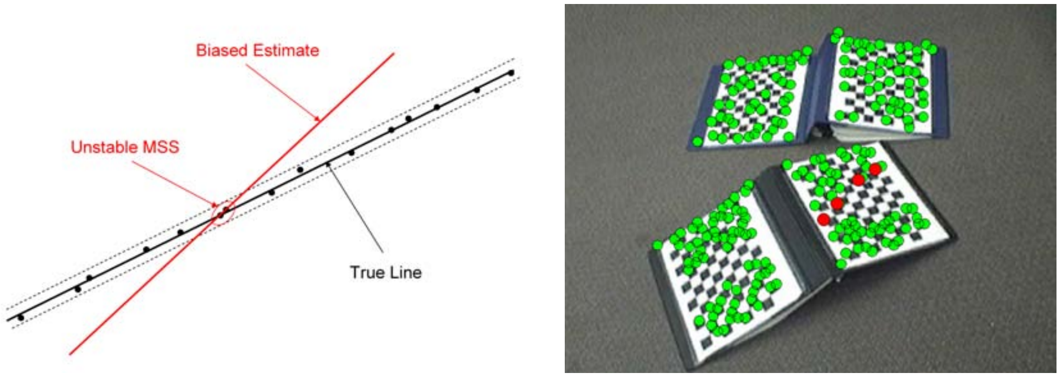
the probability of obtaining a MSS composed only of inliers is:

$$q = \frac{\binom{N_I}{k}}{\binom{N}{k}} = \frac{N_I!(N-k)!}{N!(N_I-k)!} = \prod_{i=0}^{k-1} \frac{N_I - i}{N - i} \tag{5.6}$$

where N_I is the total number of inliers. Unfortunately, to compute q we should know N_I which is generally not known a priori. However it is easy to verify that for any $\hat{N}_I \leq N_I$ we have $q(\hat{N}_I) \leq q(N_I)$ and consequently $(1 - q(N_I))^h \geq \left(1 - q(\hat{N}_I)\right)^h$ (where we made explicit the dependency of q on the number of inliers). Therefore we can estimate the maximum number of iterations using the cardinality of the largest set of inliers found so far (call this $\hat{N}_I$), which can be regarded as a conservative estimate of N_I. Hence, the iteration threshold can be fixed to:

$$\hat{T}_{iter} = \left\lceil \frac{\log \varepsilon}{\log\left(1 - q(\hat{N}_I)\right)} \right\rceil \tag{5.7}$$

We began this paragraph under the assumption that the inliers are noise free: clearly this is not a realistic situation in real life applications: certain MSSs, even when composed only by inliers, can produce biased estimates of the model parameters, as shown in the examples in Figure 5.3. To cope with this problem we need to construct MSSs that satisfy some specific constraints, for example regarding the spatial distribution of their elements [60] or the numerical stability of the estimates they produce. As an example, consider the problem of estimating an homography via the DLT algorithm ([51], p. 88). If we pick four points that approximately belong to a line, then the linear system whose solution defines the homography estimate is poorly conditioned: small amounts of noise can drastically bias the estimate (see Figure 5.3(b)).



(a) (b)

Figure 5.3: Examples of unstable MSSs. (a) Even though the points composing the unstable MSS are within the noise threshold δ (which defines the plane band contained within the dashed lines parallel to the black thick line) they produce a parameter estimate that is biased (red thick line). (b) The red markers identify a MSS of corresponding features (green markers) that belong to a planar surface (the second image is omitted) but are almost collinear: small noise perturbations will cause the homography estimate to be biased.

5.2.2 The Distance Between Two Models

Before presenting the robustification procedure, we want to define a non negative scalar value that measures the “distance” between two models (in other words we want to answer quantitatively questions such as: “How similar are two lines?” or “How different are two homographies?”). Suppose we have two sets of data $D_1 \subseteq D$ and $D_2 \subseteq D$ and let $\boldsymbol{\theta}(D_1)$ and $\boldsymbol{\theta}(D_2)$ be the parameters of the models estimated using these two data sets. We define the pseudo distance between the model instantiated with the parameter vector $\boldsymbol{\theta}(D_1)$ and the model instantiated with the parameter vector $\boldsymbol{\theta}(D_1)$ to be:

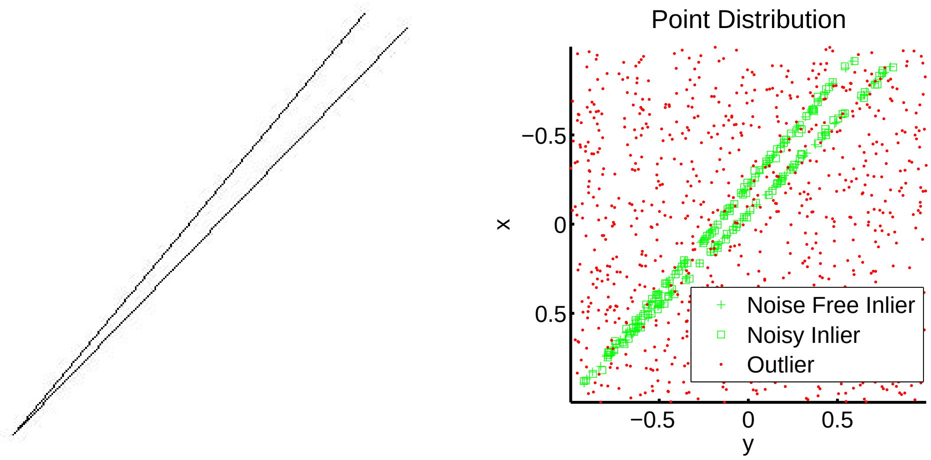
$$\text{pseudo-dist}(\boldsymbol{\theta}(D_1), \boldsymbol{\theta}(D_2)) \overset{\text{def}}{=} \frac{1}{2}\left[\frac{1}{|D_2|}\sum_{\boldsymbol{d}\in D_2} e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}(D_1))) + \frac{1}{|D_1|}\sum_{\boldsymbol{d}\in D_1} e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}(D_2)))\right] \quad (5.8)$$

(where $|D|$ indicates the cardinality of the set D). Even though (5.8) is non negative and symmetric it is not a distance function: in fact (5.8) can be equal to zero only if $D_1 \equiv D_2$ is a MSS. Moreover the concept of triangle inequality does not have a clear interpretation and in general it does not seem to hold.

5.3 The Robustification Procedure

The robustification procedure operates on the output of RANSAC and is composed of three steps: the *Minimal Sample Set (*MSS*) voting procedure*, the construction and processing of the *relationship matrix* and finally the parameter estimation via *robust statistics methods* [55, 110, 79, 120]. The first step is used to identify which Minimal Sample Sets (MSS) formed from the initial set of inliers selected by RANSAC instantiate models that are not likely to happen by chance. The second step is used to increase the robustness of the method in the presence of multiple models (since we want to identify *only one* model) and finally the last step will provide a robust estimate of the model parameters together with an estimate of the noise scale. In the following subsections we will describe in greater detail each of the previous steps.

To make our description clearer we will focus on the problem of estimating a line parameterized as $\theta_1 x + \theta_2 y + \theta_3 = 0$ under the constraint that $\|\boldsymbol{\theta}\| = 1$. More



(a) (b)

Figure 5.4: Image (a) shows the sketch of two lines and image (b) shows the dataset that will be used as our toy problem. Out of 1000 points 150 belong to the true lines (crosses) and they are corrupted with Gaussian noise with zero mean and standard deviation $\sigma = 10^{-2}$ (squares).

specifically we will consider the dataset shown in Figure 5.4(b), obtained from the sketch represented in Figure 5.4(a). Out of 1000 points uniformly distributed in the plane region $[-1, 1] \times [-1, 1]$, the inlier set is composed of 150 points, and each inlier is perturbed by zero mean Gaussian noise with standard deviation $\sigma_\eta = 10^{-2}$.

5.3.1 Step 1: The MSS Voting Procedure

To motivate the MSS voting procedure consider the set $\hat{D}_I$ formed by the elements classified as inliers by RANSAC (which are shown as red circles in Figure 5.6(a)) using a noise scale five times larger than the true one (i.e. $\hat{\sigma}_\eta = 5 \cdot 10^{-2}$ with $P_{inlier} = 0.9$). It can be seen that the data points belonging to different models and several outliers are lumped together: this undesirable effect is likely to bias the estimate of the model parameters. This observation is valid in general, despite the considered model and the outlier distribution. The procedure that we will describe in the following is likely to mitigate this problem, *by identifying the dominant models in a given set of data and by discarding the outliers* (or in other words retaining the inliers and the pseudo-outliers and discarding the true outliers).

Consider N_s distinct MSSs $s_1, \dots, s_{N_s}$ constructed within the set $\hat{D}_I$ and the corresponding parameter vectors $\boldsymbol{\theta}(s_1), \dots, \boldsymbol{\theta}(s_{N_s})$. With a little abuse of notation we define $e_i^n = e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}(s_i)))$ to be the n^{th} smallest error produced by an element $\boldsymbol{d} \in \hat{D}_I$ and by the model instantiated with the parameter vector $\boldsymbol{\theta}(s_i)$. A detail of the histogram distribution of the 10^{th} smallest error for our example is shown in Figure 5.6(b). The width h of the histogram bins is calculated following the rule by Freedman and Diaconis (which is summarized in [56]):

$$h = 2 \ \text{IQR}(\{e_1^n, \dots, e_{N_s}^n\}) \ N^{\frac{1}{3}}$$

where IQR returns the interquartile range (the 75^{th} percentile minus the 25^{th} percentile) and N denotes the number of data.

Thresholding the Histogram

The MSSs that contribute to the initial portion of the histogram can be regarded as *representatives of structures that are unlikely to happen by chance.* To make this statement more precise consider the following situation. Let ρ be the percentage of contamination of the set $\hat{D}_I$, so that the number of outliers amounts to $\rho|\hat{D}_I|$ (and consequently the number of inliers is $(1-\rho)|\hat{D}_I|$). Moreover let's assume that the manifold errors produced by the outliers are *uniformly distributed* in the interval $[-\delta, \delta]$ (where δ is given by (5.3)). As usual, the inlier errors are assumed to be *normally distributed* with standard deviation σ_η. Therefore the expected number of inliers and outliers that will produce an error in the interval $[-\bar{\delta}, \bar{\delta}]$ (where $\bar{\delta} \leq \delta$) is given by:

$$N_I(\bar{\delta}) = P[e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta})) \leq \bar{\delta}]\ (1-\rho)|\hat{D}_I|$$

$$N_O(\bar{\delta}) = \frac{\bar{\delta}}{\delta}\ \rho|\hat{D}_I|$$

In the case of lines, where the cardinality of a MSS is 2, the number of MSSs that are generated only by the inliers is:

$$N_{s,I}(\bar{\delta}) = \frac{1}{2} N_I(\bar{\delta})(N_I(\bar{\delta}) - 1)$$

and the number of the MSSs contaminated by the presence of at least one outlier is:

$$N_{s,O}(\bar{\delta}) = \underbrace{\frac{1}{2} N_O(\bar{\delta})(N_O(\bar{\delta}) - 1)}_{\text{two outliers}} + \underbrace{N_I(\bar{\delta}) N_O(\bar{\delta})}_{\text{one inlier and one outlier}}$$

Figure 5.5(a) shows the curves parameterized by $\bar{\delta}$ that displays the ratio:

$$\frac{N_s(\bar{\delta})}{N_s(\delta)} = \frac{N_{s,I}(\bar{\delta}) + N_{s,O}(\bar{\delta})}{\binom{|\hat{D}_I|}{2}}$$

versus the ratio:

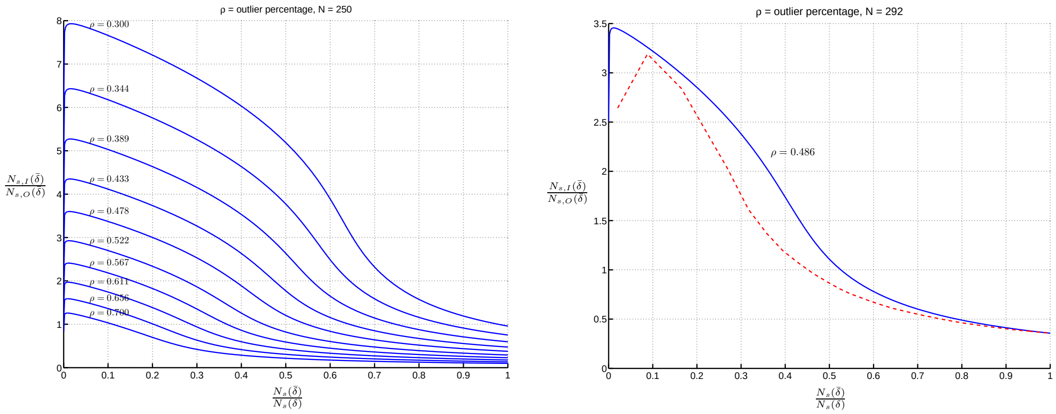
$$\frac{N_{s,I}(\bar{\delta})}{N_{s,O}(\bar{\delta})}$$

for different values of ρ. This curve can be used to better understand the structure of the cumulative histogram of the n^{th} sorted error to select a meaningful threshold. The MSSs that compose the initial part of the cumulative histogram are likely to be mostly generated by inliers. Hence, by selecting these MSSs we ensure that the ratio between the number of outlier free MSSs and the number of outlier contaminated MSSs is sufficiently large. A comparison of the theoretical curve versus the curve obtained from a real cumulative histogram is shown in Figure 5.5(b). Note that the real curve tends to stay below the theoretical curve: we believe that this happens because our initial hypothesis is not fully satisfied (the inlier error distribution is Gaussian only if the initial parameter estimate used to compute the errors is very close to the true parameter value). We select the threshold T_e to be the value for which the cumulative sum of the histogram of the n^{th} error contains at least 5% of the total number of MSSs. From Figure 5.5(a) it can be seen that for this value the ratio between outlier free and outlier contaminated MSSs is larger than one for contamination percentages up to 70%. On the other hand we do not want to consider a percentage of the cumulative histogram that is too small: in this case certain configurations of outliers could hide the true model.

All the MSSs for which:

$$\min_{1 \leq j \leq N_s} e_j^n \leq e_i^n \leq T_e$$

are selected as representative of "interesting" models. In our example the red lines in Figure 5.6(b) show these thresholds superimposed on the error histogram. Figure 5.7(a) shows the selected MSSs by means of a green segment connecting the two points that form the MSS: the two models that have been lumped together by RANSAC have been identified and the real outliers have been discarded. Finally a comment regarding the number N_s of MSSs that are generated. The total number of possible MSSs with cardinality k that can be obtained from $\hat{D}_I$ is $N_s = \binom{|\hat{D}_I|}{k}$. Of course this number can soon become extremely large: to cope



(a)

(b)

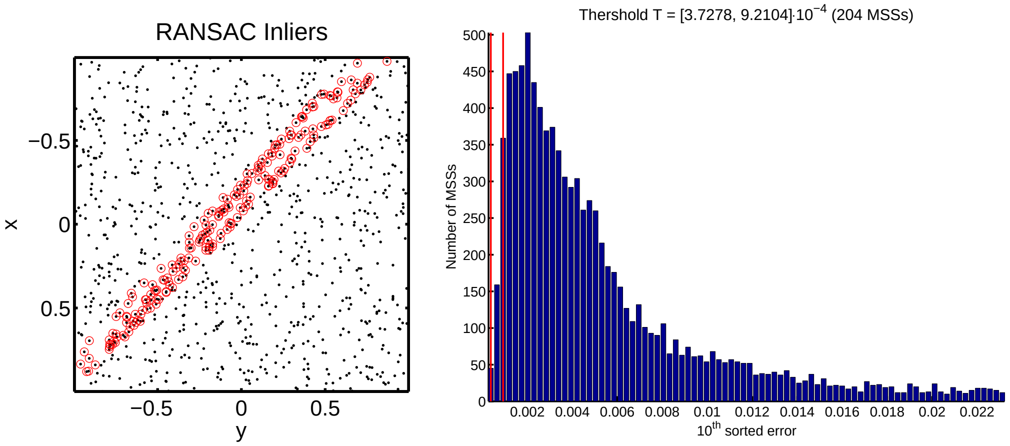
Figure 5.5: Plot (a) displays the ratio between outlier free MSSs and outlier contaminated MSSs (i.e. $\frac{N_s(\bar{\delta})}{N_s(\delta)}$) versus the ratio $\frac{N_{s,I}(\bar{\delta})}{N_{s,O}(\bar{\delta})}$ for different values of the contamination ρ. Plot (b) compares one of the theoretical curves (solid blue line) with the curve (red dashed line) obtained experimentally under the same conditions (percentage of contamination and number of points).

with this problem we uniformly subsample the set of all possible combinations of the MSSs so that $N_s \leq 10^4$.

5.3.2 Step 2: The Relationship Matrix

At this stage we have a set of MSSs that instantiate the dominant model (or models) that are present in the initial dataset $\hat{D}_I$. We would like to group together the models that are "similar" to each other and for this purpose we use the pseudo distance (5.8). More specifically we define the relationship matrix to be a symmetric matrix whose entries are given by:

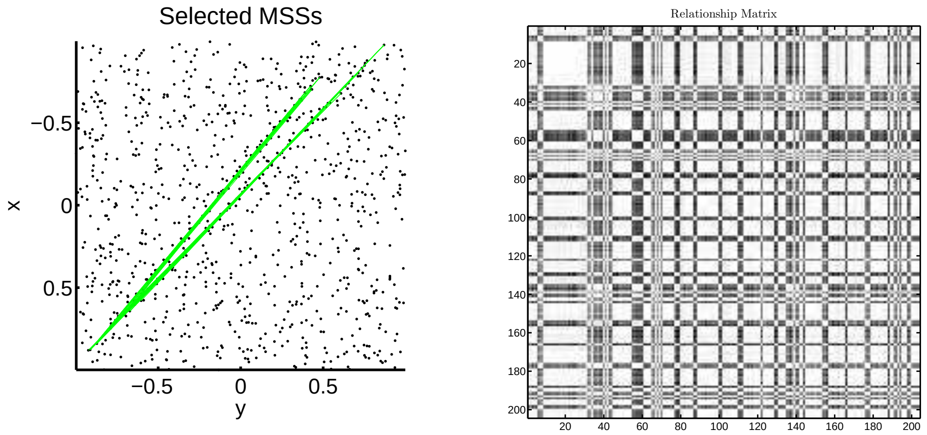
$$\mathcal{R}_{h,k} = \begin{cases} \text{pseudo-dist}(\boldsymbol{\theta}(s_h), \boldsymbol{\theta}(s_k)) & \text{if } h \neq k, \\ 0 & \text{otherwise.} \end{cases} \tag{5.10}$$



(a) (b)

Figure 5.6: In (a) the red circles represent the inliers that have been identified by RANSAC. The error threshold was calculated using (5.3) and assuming $\hat{\sigma}_\eta = 5 \cdot 10^{-2}$ and $P_{inlier} = 0.90$. Figure (b) shows the histogram of the 10^{th} error associated with the parameter vectors estimated from 10^4 MSSs formed within the set of inliers found by RANSAC. The red lines indicate the portion of the histogram that corresponds to "interesting" models.

where s_h and s_k are MSSs selected by the voting procedure described before. The relationship matrix for our example is shown in Figure 5.7(b), where brighter colors indicate smaller values. To select a threshold value to distinguish between models that can be considered equivalent from models that are distinct we resort to the histogram of the values of the upper triangular part of $\mathcal{R}$ (diagonal excluded). The value of the pseudo-distance $T_\mathcal{R}$ which corresponds to the first "relevant" valley after the first "relevant" peak is used to group together the equivalent models. The rationale behind this idea is that in the presence of multiple models the histogram will exhibit a multimodal structure: the first mode (the one that peaks for smaller values of the pseudo distance) accounts for the pseudo distances between equivalent models, all the other modes (if any) account for the pseudo



(a)

(b)

Figure 5.7: In (a) the green segments correspond to the point pairs (i.e. the MSSs) that were selected in the MSS voting procedure. Note that they identify the dominant models that were lumped together by RANSAC including the inliers present between the two lines. Figure (b) shows the relationship matrix $\mathcal{R}$. Brighter colors indicate model pairs that are similar to each other.

distances between different models. This behavior is clearly shown in Figure 5.8(a).

Identifying the Histogram Valley

A histogram can be thought as an estimate of the probability density function that generates a set of data. More generally a probability density function can be conveniently estimated using kernel based methods (also known in the pattern recognition community as Parzen windows techniques [31]). In the one dimensional case, the kernel density estimator has the following expression:

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \tag{5.11}$$

where x_i are the data samples, h is the bandwidth and K is a bounded function with compact support that satisfies the following constraints:

$$\int_{\mathbb{R}} K(x)\ dx = 1 \qquad \lim_{x \to \infty} xK(x) = 0$$

$$\int_{\mathbb{R}} x\ K(x)\ dx = 0 \qquad \int_{\mathbb{R}} x^2\ K(x)\ dx = c$$

In general the quality of a kernel is measured by the integral of the mean of the squared error between the true density and its estimate. The kernel that minimizes an asymptotic approximation of this quantity is the Epanechnikov kernel:

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{if } x^2 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Another possible choice is the Gaussian kernel, that does not have differentiability problems (such as the Epanechnikov kernel at the boundaries of its support). Its expression is:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

However note that such kernel has infinite support (on the contrary the Epanechnikov kernel has finite support).

To estimate the histogram threshold we first perform a gradient ascent to identify the histogram peak, followed by a gradient descent to identify the next valley. During this process we check that the ratio between the height of the valley and of the peak of the histogram is small enough; this is to avoid spurious valleys that are not produced by the presence of different models. Our approach can be regarded as a simplified version of the mean shift algorithm [24]. In more detail, at the l^{th} step, the distribution $\hat{f}$ can be approximated via a Taylor expansion about the point $x^{(l)}$ as:

$$\hat{f}(x^{(l+1)}) \approx \hat{f}(x^{(l)}) + \frac{\partial \hat{f}}{\partial x}(x^{(l)})(x^{(l+1)} - x^{(l)})$$

Since:

$$\hat{f}(x^{(l+1)}) - \hat{f}(x^{(l)}) \approx \frac{\partial \hat{f}}{\partial x}(x^{(l)})(x^{(l+1)} - x^{(l)})$$

in order to decrease or increase the value of $\hat{f}$ we update the current position according to:

$$x^{(l+1)} = x^{(l)} + \alpha \text{ sign}\left(\frac{\partial \hat{f}}{\partial x}(x^{(l)})\right)$$

where α defines the size of the step (negative for a descent or positive for an ascent). The value of α is reduced whenever the sign of the function variation changes. The derivative of the $\hat{f}$ can be easily calculated from (5.11); for the

Gaussian kernel its expression is:

$$\frac{\partial \hat{f}}{\partial x}(x) = -\frac{1}{Nh}\sum_{i=1}^{N}\frac{x-x_i}{h}K\left(\frac{x-x_i}{h}\right)$$

If the number of points is very large we can speed up the computation by dropping the terms for which $K\left(\frac{x-x_i}{h}\right) \approx 0$ (i.e. the points for which $\frac{x-x_i}{h}$ is large[4]). The procedure to identify a stationary point (that includes either a local maximum or a local minimum) of the probability density function $\hat{f}$ is formally described in Algorithm 1. The function TERMINATION is used to stop the loop when the variation of the position of the stationary point or the value of the probability density function or the step size drop under a certain threshold.

Grouping Equivalent Models

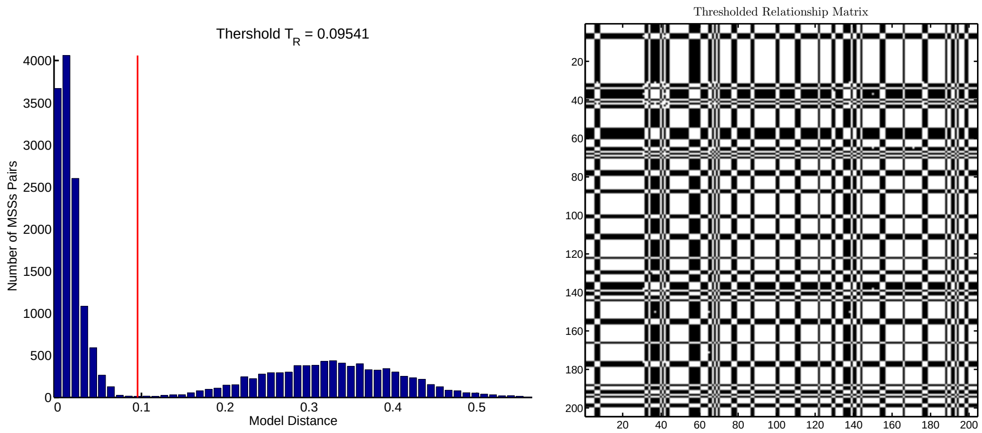
The new matrix $\mathcal{R}_T$ obtained by thresholding the relationship matrix $\mathcal{R}$ is shown in Figure 5.8(b). This new binary matrix can be interpreted as the adjacency matrix associated with a simple undirected graph $G(V, E)$ whose vertices correspond to the MSSs selected by the MSS voting procedure and the edges represent connections between MSSs who instantiate models that can be considered equivalent. It follows immediately that identifying the largest subset of MSSs that originate equivalent models corresponds to finding the maximal clique[5] of the graph G with the largest cardinality. This in an NP-hard problem [10, 25] that can be solved using trust region heuristics in $O(|V|^3)$ [16]: this method has been experimentally shown to be exact on small graphs and very efficient on

[4] As an example, for $x = 7$ we have that $\exp(-\frac{x^2}{2}) \approx 2.2 \cdot 10^{-11}$ which is negligible for most applications.

[5] A clique is a subset Q of V such that any two vertices of Q are adjacent. It is called maximal if there is no other vertex in the graph connected with all the vertices of Q.

Algorithm 1 Identification of a Stationary Point of the Probability Density Function.

```
FindStationaryPoint($x^{(0)}, h, \texttt{mode}$)
  $\hat{f}^{(0)} \leftarrow$ EstimatePDF($K, x^{(0)}$)
  $l \leftarrow 1$
  if $\texttt{mode} = \texttt{ASCEND}$
     then $\alpha \leftarrow h$
     else $\alpha \leftarrow -h$
  repeat
           $\frac{\partial \hat{f}}{\partial x}(x^{(l-1)}) \leftarrow$ EstimatePDFDerivative($K, x^{(l-1)}$)
           $x_{new} \leftarrow x^{(l-1)} + \alpha \ \ \text{sign} \left( \frac{\partial \hat{f}}{\partial x}(x^{(l-1)}) \right)$
           $\hat{f}_{new} \leftarrow$ EstimatePDF($K, x_{new}$)
           if ($\texttt{mode} = \texttt{ASCEND}$ and $\hat{f}_{new} < f^{(l-1)}$) or
              ($\texttt{mode} = \texttt{DESCEND}$ and $\hat{f}_{new} > f^{(l-1)}$)
              then $\alpha \leftarrow 0.5\ \alpha$
              else $x^{(l)} \leftarrow x_{new}$
                   $f^{(l)} \leftarrow \hat{f}_{new}$
           $l \leftarrow l + 1$
     until TERMINATION($x^{(l)}, x^{(l-1)}, \hat{f}^{(l)}, \hat{f}^{(l-1)}, \alpha$)
  return
```

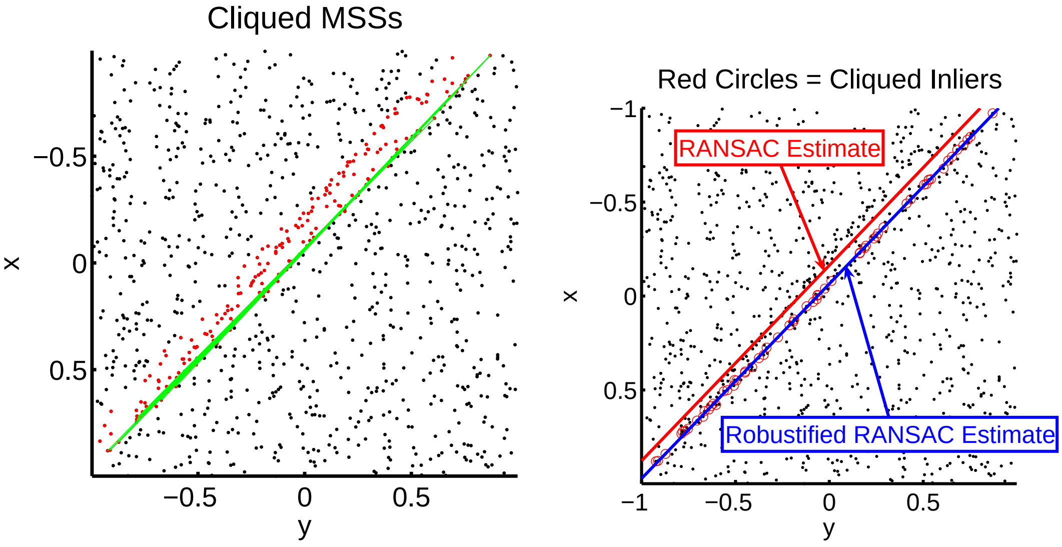


(a) (b)

Figure 5.8: The histogram in Figure (a) shows the model distance distribution (for the upper triangular part of $\mathcal{R}$). Note the bimodal nature of the histogram due to the presence of multiple distinct models. The red line indicates the threshold that is chosen in correspondence of the first valley of the histogram after the first peak. The corresponding thresholded relationship matrix is displayed in Figure (b).

various maximum clique problem instances. Figure 5.9(a) shows the MSSs associated with the largest maximum clique for the considered example: as expected the "cliqued" MSSs contain points that all belong to the same model instance. Henceforth we will indicate the data belonging to the largest maximal clique as $D_C = \left\{ \boldsymbol{d}_1^C, \ldots, \boldsymbol{d}_{N_C}^C \right\}$.

The approach we just described bears a certain resemblance to the spectral factorization methods introduced for clustering and grouping purposes [102, 95]. The main idea behind these methods is to calculate the eigenvector corresponding to the largest eigenvalue of the relationship matrix. Its dominant non zero components can be used to directly identify the elements that belong to the largest group. One of the problems of this approach is the selection of a threshold for



(a) (b)

Figure 5.9: In (a) the green segments correspond to the point pairs (i.e. the MSSs) that were selected by identifying the largest maximal clique of the relationship matrix. Figure (b) compares the initial RANSAC estimate versus the robustified estimate obtained using the Lorentzian M-estimator.

the components of the eigenvector when the groups are not well separated. This problem worsens when multiple clusters of similar dimensions are present in the data. The second step of the robustification directly addresses these problems (which are likely to affect the data obtained by the voting procedure) and reduces the risk of lumping together distinct models. Note also that Pavan et al. [101, 100] to cope with similar difficulties introduced a graph theoretic framework where the grouping process is solved identifying the dominant set in a graph (which can be though as an extension of the maximal clique when the edges connecting the graph nodes are weighted).

5.3.3 Step 3: Parameter Estimation via Robust Statistics Methods

At this point we have a set of data that are considered inliers for a certain model. However we want to increase the robustness of our method further so that we can cope with situations where at the end of the second step we still have some data that could bias our estimate. We achieve this goal first by robustly estimating the noise scale, and then by re-estimating the model parameters via an M-estimator.

Consider an initial estimate $\boldsymbol{\theta}_C$ of the parameter vector obtained from the data forming the largest maximum clique detected in Step 2 and let $e_i = e(\boldsymbol{d}_i^C, \mathcal{M}(\boldsymbol{\theta}_C))$ be the corresponding errors. A popular robust noise scale estimator [110, 120] is the sample median:

$$\sigma \overset{\text{def}}{=} 1.4826 \left(1 + \frac{5}{|D_C| - p}\right) \sqrt{\operatorname{med}_i e_i}$$

where p is the dimension of the parameter space (in our example $p = 2$ because the parameter vector is constrained to have unit norm) and med is the median function. Such an estimator is bounded when the data include less than 50% of outliers. In our method we will use a variant called MAD which takes into consideration the fact that the data points may not be centered [109, 120]:

$$\sigma^* \overset{\text{def}}{=} 1.4826 \ \text{med}_i \ |e_i - \text{med}_j \, e_j| \tag{5.13}$$

The median and MAD estimators have breakdown points of 50% and both methods are biased for multiple-mode cases even when the data contains less than 50% outliers. However this event is unlikely to happen thanks to the analysis of the relationship matrix carried out in Step 2. Therefore the robustified set of inliers is defined as:

$$S^* \overset{\text{def}}{=} \{\boldsymbol{d} \in D_I : e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}_C)) \leq 3\sigma^*\} \tag{5.14}$$

The coefficient 3 is selected based on the assumption that the error distribution of the inliers is Gaussian.

Finally, using the elements in S^*, we will re-estimate the model parameters using the robust estimator [55, 120] that is introduced in the reminder of this section. Suppose that the error $e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))$ is a random variable described by the probability density function p. We define the likelihood of the estimated set of inliers to be:

$$\mathcal{L}_{\mathcal{M}}(S^*, \boldsymbol{\theta}) = \prod_{\boldsymbol{d} \in S^*} p(e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta})))$$

and consequently the log-likelihood as:

$$L_{\mathcal{M}}(S^*, \boldsymbol{\theta}) = \sum_{\boldsymbol{d} \in S^*} \log p(e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))) = -\sum_{\boldsymbol{d} \in S^*} \rho\left(e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))\right)$$

where $\rho(\cdot) = -\log(\cdot)$. The larger is the likelihood (or the log-likelihood) the more likely are the errors to be described by the probability density function $p(\cdot)$. The M-estimate of $\boldsymbol{\theta}$ is defined as:

$$\hat{\boldsymbol{\theta}} \overset{\text{def}}{=} \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, L_{\mathcal{M}}(S^*, \boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{\boldsymbol{d} \in S^*} \rho\left(e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))\right) \tag{5.15}$$

where ρ (the negative logarithm of the error probability distribution function) is called *estimator function*. If the error distribution is Gaussian, with zero mean and standard deviation σ, then:

$$p_{\mathcal{M}}(\boldsymbol{d}, \boldsymbol{\theta}) = \frac{1}{Z} \exp\left(-\frac{1}{2} \frac{e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))^2}{\sigma^2}\right)$$

where Z is an appropriate normalization factor. Consequently we can write:

$$\rho\left(e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))\right) = -\log Z - \frac{1}{2} \frac{e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))^2}{\sigma^2}$$

If we drop the terms that do not depend on $\boldsymbol{\theta}$ we have:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{\boldsymbol{d} \in S^*} e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))^2$$

which shows the well known fact that classical least squares produces the maximum likelihood estimate of a vector in the presence of Gaussian noise. But what happens if the error distribution is not Gaussian? To answer this question we recall the concept of breakdown point for an estimator. This quantity indicates the fraction of the data that can be arbitrarily far from the model manifold without compromising the parameter estimate. Least square estimators $\rho(x) = x^2$ have a 0% breakdown point, since a single outlier can cause the parameter estimate to deviate arbitrarily from its real value. This can be understood by observing the behavior of the *influence function* ψ, which is the derivative of the estimator function and, as the name suggests, determines the influence of a datum on the value of the parameter estimate [136]. In the least square case it is equal to $\psi(x) = 2x$. Such a function increases linearly with no bounds and so does the influence of the outliers. To gain robustness an estimator should give "less importance" to errors whose magnitude is much larger than expected. This happens when the influence function either diverges at a slow rate or saturates or redescends. Interestingly enough such a behavior is deeply connected with the choice of an error probability distribution function that has heavier tails than a Gaussian. As an example, consider the Cauchy or Lorentzian distribution:

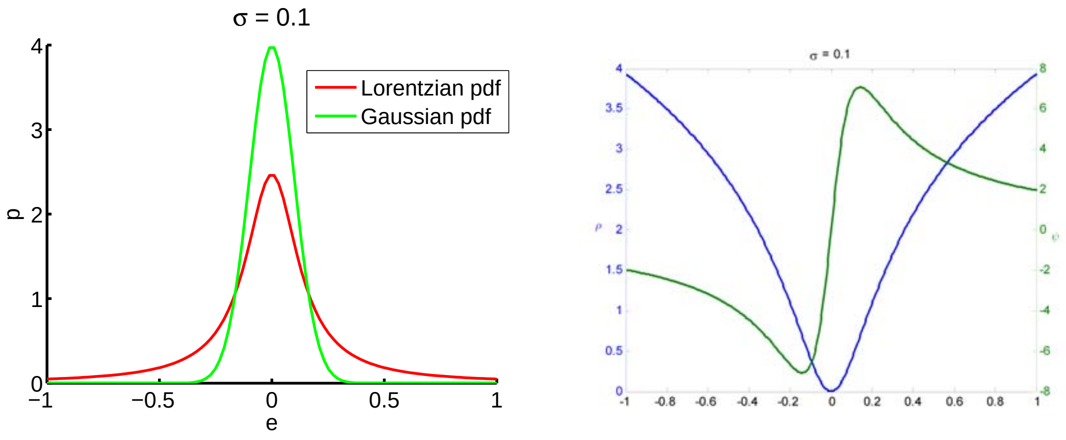
$$p(e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))) = \frac{1}{Z} \cdot \frac{1}{1 + \frac{1}{2}\left(\frac{e(\boldsymbol{d}, \mathcal{M}(\boldsymbol{\theta}))}{\sigma}\right)^2}$$

(where Z is a normalization factor) which is compared with the Gaussian distribution in Figure 5.10(a): its tails contribute to the probability distribution more that those of a Gaussian. The corresponding M-estimator and its influence function have the following expressions:

$$\rho(x) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \tag{5.16}$$

$$\psi(x) = \frac{2x}{2\sigma^2 + x^2} \tag{5.17}$$

The influence function $\psi(x)$ redescends after reaching an extremum point at $x = \pm\sqrt{2}\sigma$ and asymptotically tends to zero (see Figure 5.10(b)). To carry out the re-estimation of the parameters we use a Lorentzian M-estimator. Note that we are not claiming that the correct modeling of the error is obtained assuming a Cauchy distribution: it is an heuristic choice that, in the light of the previous observations



(a) (b)

Figure 5.10: The plots in (a) compare a Gaussian distribution to a Cauchy distribution. Note how (for the same value of the parameter σ) the tails of the Gaussian distribution decay much faster that the tails of the Cauchy distribution. The plots in (b) display the Lorentzian estimator (blue) and its influence function (green). Note the redescending behavior of ψ.

regarding the tails of the distributions, has been shown to be appropriate in many situations.

To minimize the negative log-likelihood (5.15) we utilize a subspace trust region algorithm based on the interior-reflective Newton method described in [18, 9, 11]. The scaling for the estimator is chosen to be equal to $2.3849\sigma^*$ that, as explained in [136], allows one to achieve 95% of asymptotic efficiency for a Gaussian distribution with standard deviation σ^* (see [136]). The final line estimate for our example is shown in figure 5.9(b).

5.4 The Robustification Procedure for Generic Models

In this paragraph we will discuss how the three steps of the robustification procedure should to be tailored when we want to estimate the parameters of models that are more complicated than a 2D line (in particular, we will consider the task of estimating planar homographies ([51], p. 87). It is worth remarking that the robustification procedure is similar in spirit to the refinement step that is carried out to improve the initial numerical solution of a (possibly ill conditioned) linear system. In our case RANSAC produces the initial estimate of the parameters (or, equivalently, of the set of inliers) and the robustification procedure refines such estimate.

5.4.1 Robustification for Complex Models

The complexity of a model is related to the number of its parameters or, almost equivalently, to the minimum number of data necessary to estimate the parameters (i.e. the cardinality of the Minimal Sample Set (MSS)). As we mentioned before, the total number of possible MSSs with cardinality k that can be obtained from $\hat{D}_I$ is $N_s = \binom{|\hat{D}_I|}{k}$. This number can soon become extremely large (the cardinality of a MSS to estimate an homography is 4). One way to cope with this problem is to subsample the set of all possible combinations of the MSSs and to avoid those that are unstable (see Section 5.2.1). In the case of homographies, this can be accomplished by randomly generating 4 reference points $R_1, \dots, R_4$ and by forming a MSS that contains the elements in D that are the closest to the reference

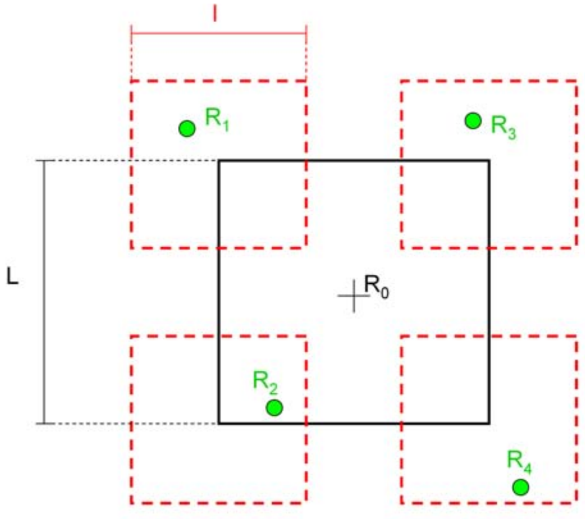


Figure 5.11: The sampling rule to construct stable MSSs for the estimation of planar homographies. The black square position is defined by the point R_0 and the position of the reference points $R_1, \dots, R_4$ is obtained by perturbing uniformly the corners of the black square (allowed locations are within the red dashed squares). The geometry of the sampling scheme is specified by the two parameters L and l that in general are functions of the image resolution and of the size of the planar structures we are interested to detect.

points (see Figure 5.11). The parameters L and l should be set according to the resolution of the image and to the size of the planar structures we are interested to detect. Experimental results showed that the error magnitude variation between the set of inliers and the set of outliers tends to be smoother when we are dealing with models that are more complicated than just a line, hence it is appropriate to consider the histogram distribution of the n^{th} smallest error with $n > 10$ (in the case of homographies we choose $n = 20$). Finally, we still select the threshold T_e to be the value for which the cumulative sum of the histogram of the n^{th} error contains at least 5% of the total number of MSSs. Even though the analysis carried out in Section 5.3.1 was specific for a line model, we believe that the constraint to obtain stable MSSs compensates for the growth of the space of the possible MSSs configurations.

The second and third step of the robustification procedure remain almost the same; however some extra care is required for the estimation of the histogram bandwidth (it is important to check that the Freedman and Diaconis rule produces meaningful values and is not affected by a few samples that produce very large values of the n^{th} error).

5.4.2 Handling Multiple Models

To handle multiple models (or different instances of the same model) it has been suggested to sequentially apply RANSAC and to remove the inliers from the dataset as each model instance is detected [130, 60] (sequential RANSAC). However, as mentioned in Section 5.1.1, inaccurate inlier detection for the initial (or subsequent) parameter estimation contributes heavily to the instability of the estimates of the parameters for the remaining models [142]. As we will show experimentally in Section 6.1.2, a sequential approach will greatly benefit by letting the robustifiaction procedure follow RANSAC after a model has been detected. However a problem arises: should we remove the data identified as inliers after Step 2 (cliqued inliers) or after Step 3 (i.e. after the noise standard deviation has been estimated via (5.13))? We believe that the answer is application dependent: if we want to reduce the risk of neglecting points that actually belong to a given model we should remove the inliers that are identified at the end of Step 3. On the other hand, if we are concerned about false positives (i.e. points that actually do not fit accurately the model but whose influence is still relevant in the robust estimation framework) we should consider for removal the “cliqued” inliers at the end of Step 2.

5.5 Experimental Results

5.5.1 Line Detection Experiment

The goal of this experiment is to study the effect of the stabilization procedure on the estimation of the parameters *in the presence of a large uncertainty on the noise scale*. To this purpose we will evaluate the results obtained from RANSAC and from the stabilization procedure applied to RANSAC's results. We will consider a line on the plane parameterized as $\theta_1 x + \theta_2 y + \theta_3 = 0$ under the constraint that $\|\boldsymbol{\theta}\| = 1$. We will consider four situations where the number of inliers (points belonging to the line whose coordinates are affected by Gaussian noise with standard deviation σ_η) is respectively $N_I = 200, 150, 100, 50$ points. The final dataset is composed of $N = 1000$ points and is obtained by adding to the inliers $N_O = N - N_I$ outliers uniformly distributed in the square $[-1, 1] \times [-1, 1]$. For every value of N_I we will generate 10 different problem instances and for each problem instance we will run the algorithm 50 times for 5 different values of the initial estimate of the noise scale: $\hat{\sigma}_\eta = 9\sigma_\eta, 6\sigma_\eta, 3\sigma_\eta, \sigma_\eta, 0.5\sigma_\eta$. The results will be averaged over the entire set of experiments for each value of N_I and of $\hat{\sigma}_\eta$. For the sake of visualization of the experimental results, we will consider an equivalent parameter representation for the line, given by the spherical coordinates of the vector $\boldsymbol{\theta}$ (recalling that $\|\boldsymbol{\theta}\| = 1$):

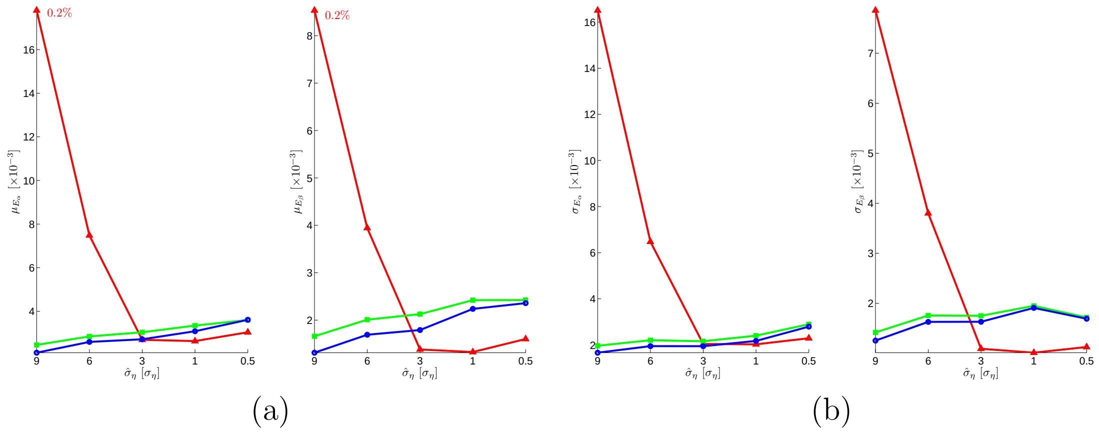
$$\alpha = \arctan \frac{\theta_2}{\theta_1}$$

$$\beta = \arccos \theta_3$$

The estimation error E is defined with respect to the ground truth value of each parameter (e.g. $E_\alpha = |\hat{\alpha} - \alpha|$, where the hat indicates the estimated quantity). To avoid cases where RANSAC in the first place missed completely the line, we consider only the experiments that produced values for E_α and E_β that are less than 0.1 radians.

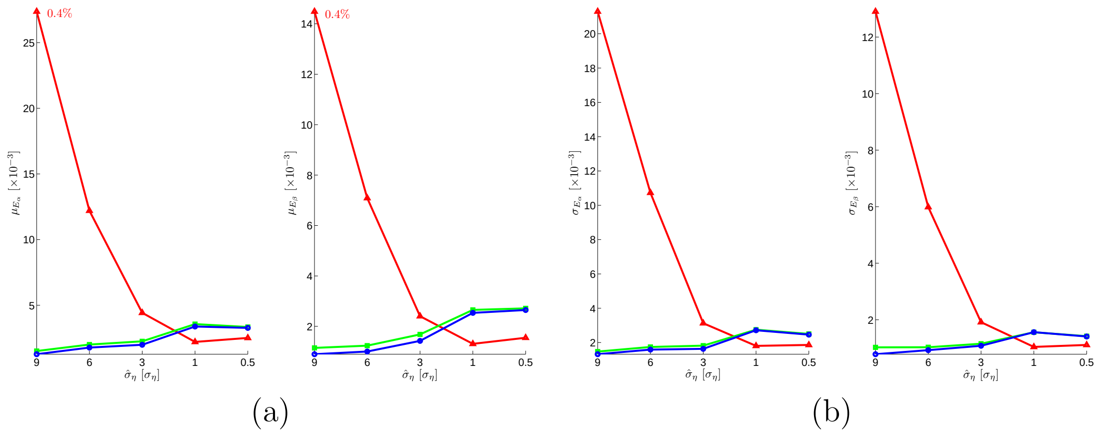
The results of the experiments are shown in Figures 5.12, 5.13, 5.14 and 5.15. The plots in (a) display the mean (μ_{E_α} and μ_{E_β}) of the estimation error of the parameters, whereas the plots in (b) show the standard deviation of the estimation error (σ_{E_α} and σ_{E_β}). Both the mean and the standard deviation are plotted versus the initial estimate of the noise scale (which is expressed in terms of multiples of the true noise standard deviation σ_η). The red triangle-marked line represents the results obtained from RANSAC (without robustification), the green square-marked line the results obtained by estimating the parameters using least squares applied to the Minimal Sample Sets (MSS) identified at the end of the second step of the stabilization procedure and the blue circle-marked line the results at the end of the whole robustification procedure. The percentage that is associated with the markers in plots (a) indicates how many experiments have been discarded because the error E was greater than 0.1 radians (each point is calculated averaging 50 runs of the algorithms for 10 different problem instances, for a total of 500 experiments). When the percentage is omitted none of the experiments has been discarded.

The experiments show that the robustification procedure greatly reduces the bias of the parameter estimates and the bias standard deviation especially when the noise scale is considerably overestimated. Moreover, for a given noise scale,



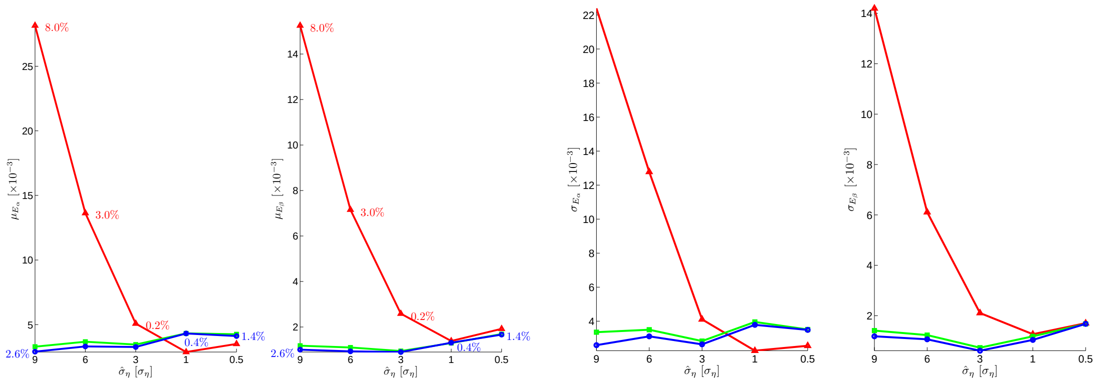
(a) (b)

Figure 5.12: The plots in (a) display the mean (μ_{E_α} and μ_{E_β}) of the estimation error of the parameters, whereas the plots in (b) show the standard deviation of the estimation error (σ_{E_α} and σ_{E_β}). Both the mean and the standard deviation are plotted versus the initial estimate of the noise scale (which is expressed in terms of multiples of the true noise standard deviation σ_η). The red triangle-marked line represents the results obtained from RANSAC (without robustification), the green square-marked line the results obtained by estimating the parameters using least squares applied to the MSSs identified at the end of the second step of the stabilization procedure and the blue circle-marked line the results at the end of the whole robustification procedure. In this experiment the number of the inliers was $N_I = 200$.



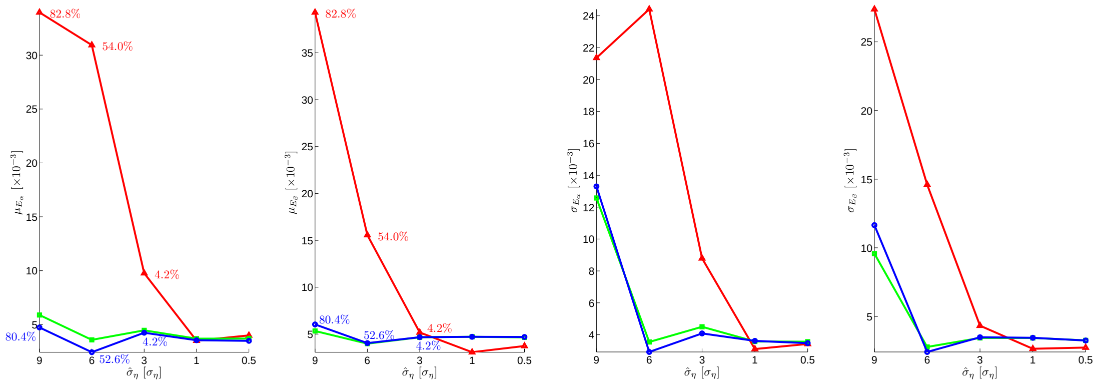
(a) (b)

Figure 5.13: See Figure 5.12 for a description of the plots. In this experiment the number of the inliers was $N_I = 150$.



(a) (b)

Figure 5.14: See Figure 5.12 for a description of the plots. The percentage that is associated with the markers in plots (a) indicates how many experiments have been discarded because the error E was greater than 0.1 radians (each point is calculated averaging 50 runs of the algorithms for 10 different problem instances, for a total of 500 experiments). When the percentage is omitted none of the experiments has been discarded. In this experiment the number of the inliers was $N_I = 100$.



(a) (b)

Figure 5.15: See Figure 5.12 for a description of the plots. In this experiment the number of the inliers was $N_I = 50$.

the benefits of the robustification procedure are more relevant when the number of the inliers is smaller. The robustification procedure performance degrades for values of the noise scale that are close to the ground truth. From the graphs it follows that it is actually convenient to overestimate the noise scale provided that the initial set of inliers contains a sufficiently large subset of elements that actually belong to the true model.

5.5.2 Line Intersection Experiment

The goal of this experiment is to study the effect of the stabilization procedure on the discrimination of "close" models in the presence of a large uncertainty on the noise scale. To this purpose we will compare two sets of results. The first ones are obtained from the recursive application of RANSAC to the initial dataset followed by the removal of the inliers after a model has been identified. The second results are obtained using the same approach with the only relevant difference that RANSAC is followed by the robustification procedure. The dataset is composed of two lines that intersect forming an angle ϕ that in our experiments is either 9° or 6° (the smaller is ϕ the closer are the models). We considered a number of inliers for each line which is either 100 or 75 points. The final dataset is composed of $N = 1000$ points adding to the inliers $N_O = N - N_I$ outliers uniformly distributed in the square $[-1, 1] \times [-1, 1]$. An example of the dataset we consider in the experiments is given in Figure 5.16.

The results of the experiments are shown in Figures 5.17, 5.18, 5.19 and 5.18. The plots in (a) display the mean (μ_{E_α} and μ_{E_β}) of the estimation error of the parameters averaged over the two models, whereas the plots in (b) show the stan-

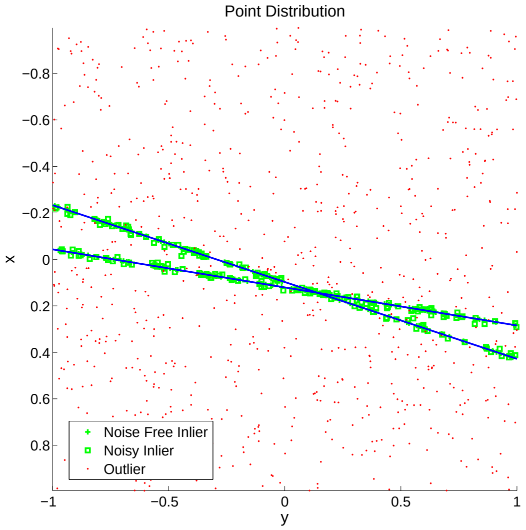
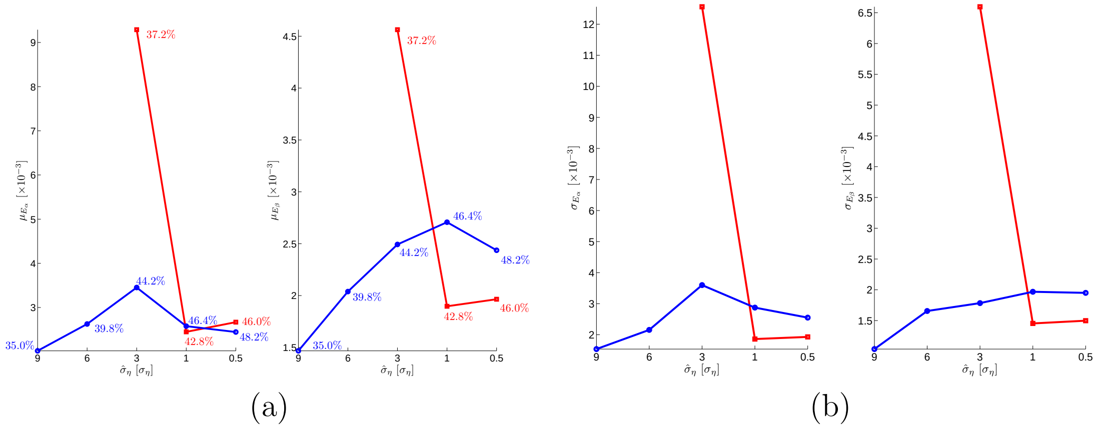


Figure 5.16: An example of the dataset used for the line intersection experiment. Each line contains 100 points and they intersect forming an angle of 9°.

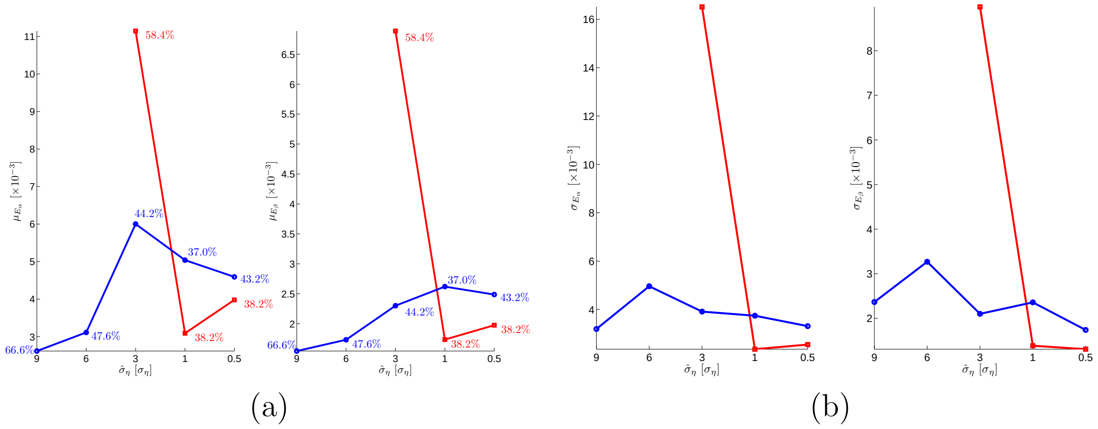
dard deviation of the estimation error (σ_{E_α} and σ_{E_β}, again averaged over the two models). Both the mean and the standard deviation are plotted versus the initial estimate of the noise scale (which is expressed in terms of multiples of the true noise standard deviation σ_η). The red square-marked line represents the results obtained from the sequential application of RANSAC (without robustification), the blue circle-marked line the results obtained from the sequential application of RANSAC (with robustification). Also in this experiment the percentage that is associated with the markers in plots (a) indicates how many experiments have been discarded because the error E was greater than 0.1 radians (each point is calculated averaging 50 runs of the algorithms for 10 different problem instances, for a total of 500 experiments). When the percentage is omitted none of the experiments has been discarded.

The experiments confirm the benefits that are obtained by applying the robus-



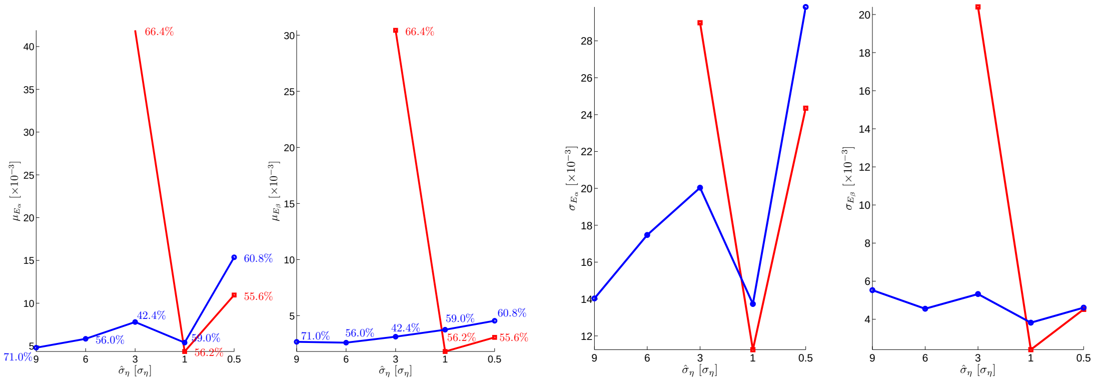
(a) (b)

Figure 5.17: The plots in (a) display the mean (μ_{E_α} and μ_{E_β}) of the estimation error of the parameters averaged over the two models, whereas the plots in (b) show the standard deviation of the estimation error (σ_{E_α} and σ_{E_β}, again averaged over the two models). Both the mean and the standard deviation are plotted versus the initial estimate of the noise scale (which is expressed in terms of multiples of the true noise standard deviation σ_η). The red square–marked line represents the results obtained from the sequential application of RANSAC (without robustification), the blue circle-marked line the results obtained from the sequential application of RANSAC (with robustification). In this experiment the number of the inliers for each line was $N_I = 100$ and the angle between the lines was $\phi = 9°$.



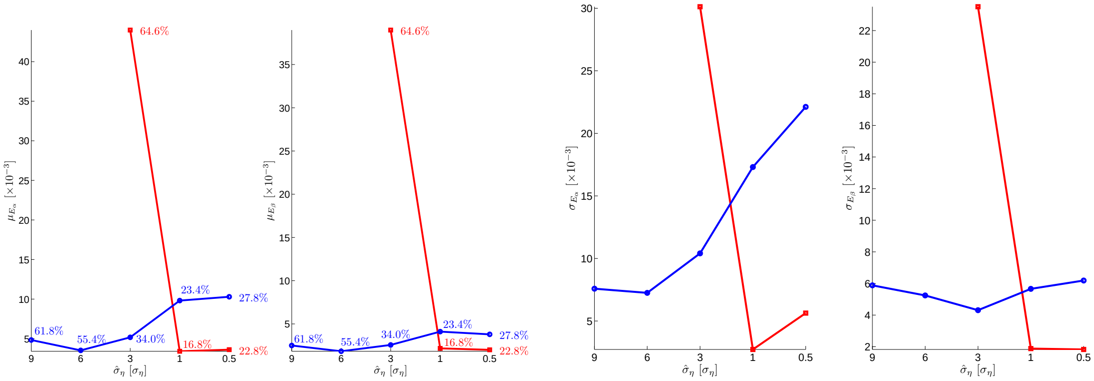
(a) (b)

Figure 5.18: See Figure 5.17 for a description of the plots. In this experiment the number of the inliers for each line was $N_I = 75$ and the angle between the lines was $\phi = 9°$.



(a) (b)

Figure 5.19: See Figure 5.17 for a description of the plots. In this experiment the number of the inliers for each line was $N_I = 100$ and the angle between the lines was $\phi = 6°$.



(a) (b)

Figure 5.20: See Figure 5.17 for a description of the plots. In this experiment the number of the inliers for each line was $N_I = 75$ and the angle between the lines was $\phi = 6°$.

tification procedure to the estimates obtained from RANSAC: similarly to what happened in the line detection experiments, as long as the noise scale is overestimated, the robustification procedure greatly reduces the bias of the parameter estimates and the bias standard deviation.

5.5.3 Multiple Homographies Experiment

This experiment is similar in spirit to the experiment described in Section 5.5.2. In this case we want to study how the robustification procedure performs in the presence of multiple instances of more complicated models. To this purpose, we will consider the planar homographies ([51], p. 87) induced by the checkerboard patterns displayed in Figure 5.21. The use of checkerboard patterns was meant to facilitate the task of the feature tracker, so that we could obtain a reasonably dense and uniform distribution of point features over the planar portions of the image. On the other hand, the checkerboards were deliberately arranged in a challenging way, so that the angle formed by the plane normals is relatively small and the structures are spatially close to one another.

Since we do not know the ground truth for the parameters, we will evaluate the performance of the algorithm in terms of its capability to group together point features that undergo the same motion (motion segmentation). Let's indicate with $D_{I,i}$ the set of inliers relative to the i^{th} homography and with $\hat{D}_{I,i}$ the set of inliers determined by the sequential application of RANSAC at the i^{th} step (followed or not by the robustification procedure). First we will identify the index pair that corresponds to the two sets that have the largest number of elements in common.

More precisely, $\hat{D}_{I,i^*(j)}$ is associated with $D_{I,j}$ if:

$$i^*(j) = \underset{1 \leq i \leq W}{\operatorname{argmax}} \frac{|\hat{D}_{I,i} \cap D_{I,j}|}{|D_{I,j}|}$$

where W is the total number of models. Then for $1 \leq j \leq W$ we will define the percentage of correctly detected inliers as:

$$C_j = \frac{|\hat{D}_{I,i^*(j)} \cap D_{I,j}|}{|D_{I,j}|} 100 \tag{5.19}$$

and the percentage of wrong inliers as:

$$B_j = \left(1 - \frac{|\hat{D}_{I,i^*(j)} \cap D_{I,j}|}{|\hat{D}_{I,i^*(j)}|}\right) 100 \tag{5.20}$$

Note that the denominators of the two expressions are different: this is because in expression (5.19) we want to measure what percentage of the "true" inliers are correctly identified, whereas in (5.20) we measure the percentage of points that have been associated with the wrong planar region.[6] When $\hat{D}_{I,i^*(j)} \equiv D_{I,j}$ we have that $C_j = 100\%$ and $B_j = 0\%$.

The experiments have been carried out averaging the results of 50 runs of the algorithm on two checkerboard sequences and the results are displayed in Figures 5.22 and 5.23. In both experiments the MSS sampling was guided by the procedure illustrated in Figure 5.11 (we set $L = 20$ pixels and $l = 10$ pixels). For the sequential RANSAC without robustification procedure we studied the performance of the algorithm for $\hat{\sigma}_\eta = 0.25, 0.5, 1, 1.5$ pixels. When the robustification procedure was enabled we considered the following values for the noise standard deviation: $\hat{\sigma}_\eta = 1, 2, 3, 4, 5$ pixels. We choose different intervals for $\hat{\sigma}_\eta$ because

[6] It may happen that some points fit a certain homography even if they do not belong to the portion of the image that induced such homography. However we consider this situation to have a negligible impact on the evaluation of the overall performance of the algorithm.

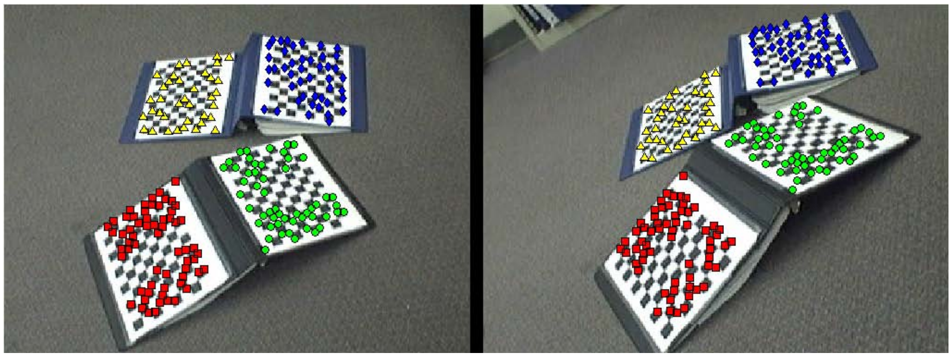
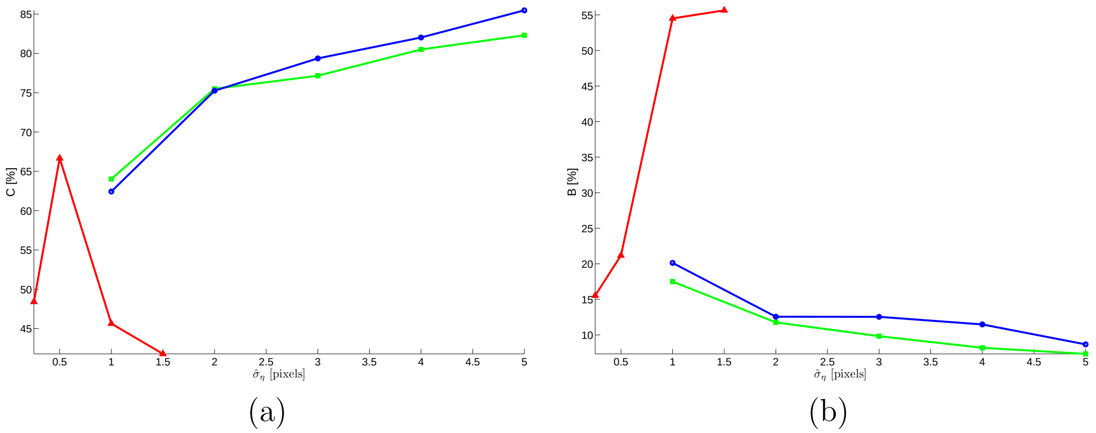


Figure 5.21: An example of the dataset used for the homography detection experiment. There are four homographies induced by the feature points generated by the checkerboard patterns. Features belonging to the same planar region are displayed with markers with the same shape and color.

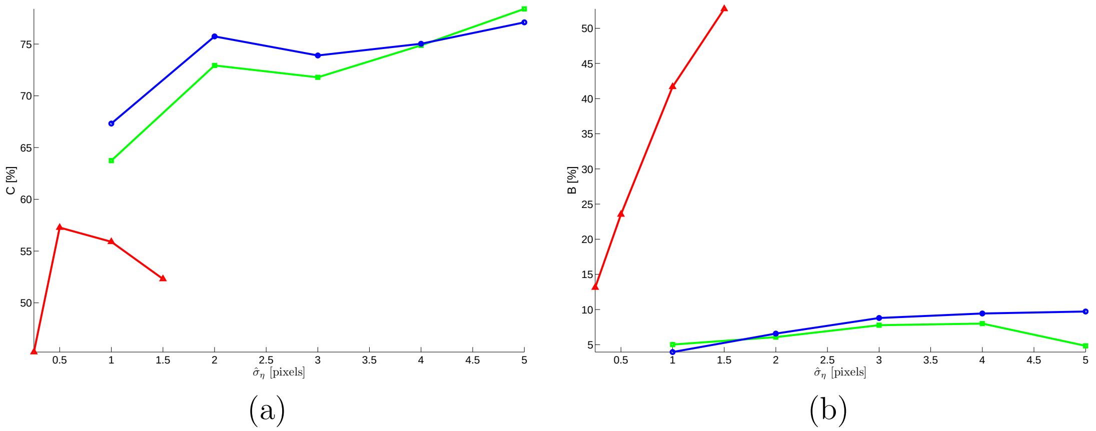
large values of the noise standard deviation will not produce meaningful results if the robustification procedure is not enabled (as shown by the plots in Figures 5.22(b) and 5.23(b), RANSAC performs very poorly when $\hat{\sigma}_\eta$ becomes larger or equal than 1).

From the experimental results we observe that sequential RANSAC with no robustification is very sensitive to the selection of the noise threshold. On the other hand, the robustification procedure makes the results more stable across a larger set of values for the noise standard deviation. This behavior was anticipated by the results obtained in the synthetic experiments, and also the fact that overestimating the noise scale produces better results. Moreover the experiments support the claim made at the end of Section 5.4: to obtain larger values of C_j it is advisable to remove the inliers identified at the end of Step 3 and to reduce the value of B_j it is better to remove the inliers identified at the end of Step 2.



(a) (b)

Figure 5.22: Experimental results for the segmentation of the coplanar point features in the first Checkerboards image pair. Plot (a) shows the mean value of C_j and plot (b) shows the mean value of B_j for $1 \leq j \leq 4$ for different values of the noise scale. The red triangle-marked line displays the results obtained applying sequentially RANSAC not followed by the robustification procedure. The green square-marked line displays the results obtained applying sequentially RANSAC followed by the robustification procedure and with the removal of the "cliqued" inliers, whereas the blue circle-marked line is obtained removing the inliers identified using the noise scale estimate obtained in the third step of the robustification procedure (see (5.13)).



(a) (b)

Figure 5.23: Experimental results for the segmentation of the coplanar point features in the second Checkerboards image pair. See Figure 5.22 for the explanation of the plots.

5.6 Conclusions and Future Work

In this chapter we presented an algorithm that is able to improve the unbiasedness of the parameter estimates obtained from RANSAC, especially in cases where there is a large uncertainty regarding the standard deviation of the noise that contaminates the data and multiple model instances are present.

As described in Section 5.3, the robustification procedure is composed of three steps. First a Minimal Sample Set (MSS) voting procedure aims at identifying the MSSs that produce parameter estimates that are unlikely to happen by chance. Then, in the second step, the MSSs that instantiate different models are grouped together. This is done by introducing a new pseudo distance measure for model similarity (5.8) and by reducing the model clustering problem to the problem of identifying the maximum clique of a graph. Finally, in the third step, robust estimators are used to both estimate the noise standard deviation and the parameters of the models.

As discussed in Section 5.4, the procedure generalizes quite straightforwardly for models of different complexity. However we believe that the selection of the histogram threshold in the first step of the robustification procedure deserves further investigation. In particular we would like to study how the *Helmholtz principle* and the concept of *meaningfulness* (which have been intensively studied by Desolneux, Lisani et al. [73, 27] and have been applied to several image analysis problems) could help in automatically analyzing the structure of the error histogram.

We have shown that the robustification procedure improves the estimates of the homographies that relate planar structures in image pairs. However we are aware that the procedure should be tested on larger sets of images (containing planar structures). This is to verify the performance of the robustification framework when the distribution of the point features is non isotropic and the impact of the camera distortions is not negligible. More generally, we advocate a thorough study of the identifiability and distinguishability of the models that describe the geometric transformations that are induced by camera motions in image registration scenarios.

Chapter 6

Applications

"What is the tangible output?"

B. S. Manjunath

This chapter contains an overview of the algorithms developed in the previous chapters integrated into a registration and mosaicking system. Using the framework developed in Chapter 2, we introduce the concept of *characteristic structure* of a point neighborhood and show how it can be used to improve the detection of matching points between image pairs related by large scale variations. We then devote our attention to the development of a set of techniques to obtain a seamless mosaic of the registered images. The contributions contained in this chapter can be summarized as follows:

- We apply the framework based on condition theory to identify the *characteristic structure* of a point neighborhood and show how this can be used to establish matches between images related by large scale variations (see Section 6.1).

- We explore the possibility of using *indexing* and *dimensionality reduction techniques* to speed the computation of tentative image correspondences (see Section 6.2.1).

- We introduce a *novel robust equalization procedure* to correct the photometric appearance of two images that are to be fused together (see Section 6.2.2).

- We present a *physically motivated* algorithm to calculate the best stitching line between registered images (see Section 6.2.3).

6.1 Point Neighborhood Characteristic Structure Detection

Many early vision tasks are performed by processing the intensity information in the neighborhood of an image point. Examples of these tasks include the detection of reference or tie points, the construction of robust descriptors [92, 39, 50, 98, 106, 116, 112, 5, 59, 87, 63, 129, 74] for applications such as tracking and correspondence, image identification and retrieval. To develop algorithms that behave consistently despite changes in the viewing conditions such as translations, rotations, scalings or perspective distortions, it is crucial to *automatically identify the characteristic structure of the neighborhood* of an image point. For example, given a corresponding point pair in two images related by rotation and scaling, the corresponding point neighborhoods should be rotated and scaled by the same amount. Lowe [74] recently proposed a point detector robust with respect to relevant image projective transformations. Kadir et al. introduced in [59] the concept of image saliency and used it to identify the characteristic scale about an image point. Baumberg [5] and Mikolajczyk et al. [87] presented an iterative procedure to detect point neighborhoods despite the affine distortions of the image. Both approaches largely draw from the image scale space theory (mainly developed by Florack [36] and Lindeberg [70, 71]) to identify the characteristic structure of the neighborhood of a point. One of the most important ideas proposed by Lindeberg is the principle for automatic scale selection (see [71], p. 83):

> In the absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data.

In the literature several derivative based functions have been proposed to identify the characteristic scale of an image [85], such as the gradient magnitude, the image Laplacian, the difference of Gaussians and the Harris function. In this section, we will propose as evidence of the characteristic structure of a point neighborhood the local minima of a function that measures the computational stability of the neighborhood itself (or, more intuitively, how much the intensity pattern in the neighborhood preserves its structure under the effect of noise). As foreseen by Lindeberg, such a function is expressed in terms of a nonlinear combination of the derivatives of the image intensities. One interesting observation about our formulation is that it supports the principle proposed by Lindeberg within a mathematical framework and with no need for any heuristics. In Section 6.1.1 we shall show how suitable combinations of image derivatives are indeed *fundamental evidences* that reveal the characteristic structure of a (circular) point neighborhood.

This section is organized as follows: Section 6.1.1 will specialize and extend the results described in Chapter 2 to identify the characteristic radius of a circular neighborhood. The experimental results and the applications of this method will be presented in Section 6.1.2 and some concluding remarks will be presented in Section 6.3.

6.1.1 Detecting the Characteristic Structure

The definition of the condition number in Section 2.4.1 is the fundamental building block for the principle we propose to reveal the characteristic structure of a point neighborhood. This can be stated as follows:

> The intensity pattern of a generalized image in a neighborhood $\Omega(\boldsymbol{x})$ reflects the characteristic $\boldsymbol{T}$-structure of the image about $\boldsymbol{x}$ if the condition number $K_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}(\Omega(\boldsymbol{x}))$ is minimized for local perturbations of the neighborhood itself.

Note that the notion of characteristic neighborhood is intimately related to the geometric transformation $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}$ that is chosen to model the effects of the noise. We can give a formal definition of the $\boldsymbol{T}$-characteristic neighborhood as follows:

Definition 6.1.1 *The $\boldsymbol{T}$-characteristic neighborhood of a generalized image $\boldsymbol{I}$ about a point $\boldsymbol{x}$ is defined as:*

$$\hat{\Omega}(\boldsymbol{x}) \overset{\text{def}}{=} \underset{\delta\Omega}{\operatorname{argmin}}\, \hat{K}_{\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}}(\Omega(\boldsymbol{x}) + \delta\Omega) \tag{6.1}$$

where $\delta\Omega$ represents a local perturbation of the neighborhood.

To translate the previous principle into a computational algorithm, we need to specify how the neighborhood is parameterized and the functional form of the geometric transformation that models the effect of the noise. These two choices represent a tradeoff between complexity and accuracy: neighborhoods with many degrees of freedom and transformations described by many parameters can model precisely the effects of the noise but they are difficult to handle and do not lead to efficient implementations. For the time being we will focus our attention on circular neighborhoods that can be simply parameterized as $\Omega_r(\boldsymbol{x}) = \{\boldsymbol{y} : \|\boldsymbol{x} - \boldsymbol{y}\| \leq r\}$ where r represents the radius of the neighborhood. The transformation we will consider is intended to model the perturbations produced by the noise in the scaling, rotation and translation of the intensity pattern of an ordinary image ($n = 2$):

$$\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{x} + \begin{bmatrix} \theta_3 & -\theta_4 \\ \theta_4 & \theta_3 \end{bmatrix} (\boldsymbol{y} - \boldsymbol{x}) + \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{6.2}$$

Intuitively one would expect the quantity that measures the stability of a given neighborhood to be invariant with respect to a certain class of geometric transformations, such as those that are used to model the influence of the noise. In general this fact does not follow from the algebraic properties of the Generalized Gradient Matrix (GGM), as extensively discussed in 2.5.1 and in Example 2.5.3, but this difficulty can be overcome by mapping the neighborhood $\Omega(\boldsymbol{x})$ onto a normalized neighborhood.

Some Numerical and Computational Considerations

To compute the condition number (B.1) we need to decide which norm has to be used. Some of the consequences of this choice have been explored (for slightly different purposes) in [140, 64] and discussed in Chapter 2 in relation to the Spectral Generalized Corner Detector Functions (SGCDF). Here we will analyze two alternatives, outlining their advantages and disadvantages.

The first option is to let $q \to \infty$. It follows immediately that $\|M\|_{\infty,Schatten} = \lim_{q\to\infty} \|M\|_{q,Schatten}$ is equal to the largest singular value of M, i.e. $\sigma_1(M)$. Consequently, since the first $\min\{h,k\}$ singular values of the pseudo inverse of a matrix M are the reciprocal of the singular values of M, the condition number computed via the ∞-Schatten norm is given by:

$$\hat{K}_{\boldsymbol{T_\theta},\boldsymbol{x}} = \max_{1\le i\le p} \sigma_i(A^\dagger) = \max_{1\le i\le p} \frac{1}{\sigma_i(A)} = \frac{1}{\min_{1\le i\le p} \sigma_i(A)} = \frac{1}{\sigma_p(A)}$$

(where A is the GGM and $p = 4$ for the transformation (6.2)). This norm measures the maximum distortion produced by the noise in the intensity pattern in $\Omega(\boldsymbol{x})$. The worst scenario happens when the noise "tweaks" the singular vector corresponding to σ_p, i.e. when $\boldsymbol{u}_i^T\boldsymbol{\eta} = 0$ for every $i \neq p$ and $\boldsymbol{u}_p^T\boldsymbol{\eta} = \|\boldsymbol{\eta}\|$ (where $\boldsymbol{u}_i$ denotes the i^{th} left singular vector of A and $\boldsymbol{v}_i$ denotes the i^{th} right singular vector). In fact, as it follows from equation (2.20) in the proof of Theorem 2.4.2, we can write that $\Delta\boldsymbol{\theta} = -A^\dagger\boldsymbol{\eta} = -\left(\sum_{i=1}^{p} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T\right)\boldsymbol{\eta} \propto \frac{1}{\sigma_p}\boldsymbol{v}_p$, i.e. the effect of the noise is maximally amplified along the affine space $\overline{\boldsymbol{\theta}} + \langle\boldsymbol{v}_p\rangle$ in the parameter space. If we make the assumption that the components of the vector $\boldsymbol{\eta}$ are i. i. d. Gaussian variables, then $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2 I)$. Since linear transformations of jointly Gaussian vectors are still jointly Gaussian vectors and $A^\dagger\left(A^\dagger\right)^T = (A^TA)^{-1}$, then from

$\boldsymbol{\Delta\theta} = -A^{\dagger}\boldsymbol{\eta}$ it follows that $\boldsymbol{\Delta\theta} \sim \mathcal{N}\left(\mathbf{0}, \sigma_{\eta}^{2}(A^{T}A)^{-1}\right)$. This result is consistent with the observation that on average the perturbation of the parameters is zero and justifies the choice of defining the condition number (2.14) in terms of the supremum of the ratio between $\|\Delta\boldsymbol{\theta}\|$ and $\|\boldsymbol{\eta}\|$.

The second option is to let $q \to 0$. As shown in Theorem 2.5.8, we can write:

$$\lim_{q\to 0} \frac{1}{\sqrt[q]{p}} \|\hat{K}_{\boldsymbol{T_{\theta,x}}}(\Omega(\boldsymbol{x}))\|_{q,Schatten} = \frac{1}{\det(A^{T}A)^{\frac{1}{p}}} \tag{6.3}$$

This observation has a crucial importance to alleviate the computational burden associated with the calculation of the condition number (B.1) for the geometric transformation (6.2). In fact its value can be simply obtained by raising the determinant of a 4×4 symmetric matrix to the power $-1/p$. Identity (6.3) makes it possible to compute a dense map of the condition number for different values of the neighborhood radius r, in a way that resembles the pyramidal approach described in [74]. The right hand side of equation (6.3) is deeply connected with the corner detector proposed by Rohr (and compared in [107] with other detectors): maximizing the determinant of $A^{T}A$ is equivalent to minimize the condition number and therefore it allows one to identify the characteristic dimension of the neighborhood.

The Algorithm: Design Issues and Practical Implementation

We will now describe some of the practical issues that arise in the implementation of the ideas discussed before. Recall that the ultimate goal is to find points of local minimum for the condition number (B.1).

In our implementation the neighborhood $\Omega_r(\boldsymbol{x})$ (of radius r) is warped on a unit circle sampled with resolution Δ (see Figure 2.6). The image derivatives are computed convolving the image with the derivatives of Gaussian kernels, as suggested in [36, 70]. The weighting function that we used to compute (2.17) is a raised cosine that only smooths the intensities values at the boundaries of $\Omega_r(\boldsymbol{x})$. We opted for this function rather than for a traditional Gaussian to guarantee that the points inside the region $\Omega_r(\boldsymbol{x})$ are equally weighted. The analytical expression of w is given by:

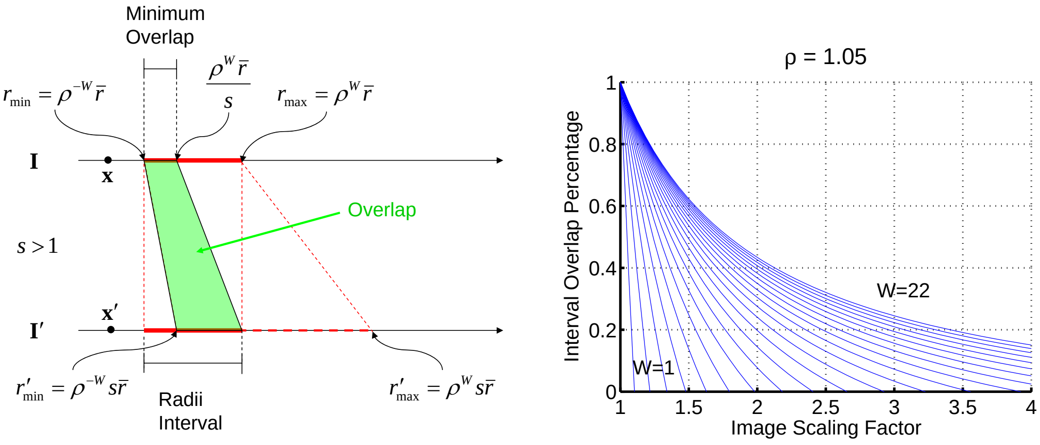
$$w(\boldsymbol{y}_i - \boldsymbol{x}) = \begin{cases} 1, & \text{for } \|\boldsymbol{y}_i - \boldsymbol{x}\| \leq r - \delta \\ \frac{1}{2}\cos\left[\frac{\pi}{2}\left(\frac{\|\boldsymbol{y}_i - \boldsymbol{x}\|}{\delta} + 1 - \frac{r}{\delta}\right)\right] + \frac{1}{2}, & \text{for } r - \delta < \|\boldsymbol{y}_i - \boldsymbol{x}\| < r + \delta \\ 0 & \text{for } \|\boldsymbol{y}_i - \boldsymbol{x}\| \geq r + \delta \end{cases}$$

where δ denotes the transition band.

The condition number is calculated for different values of the neighborhood radius. The radii sequence is constructed in such a way that the ratio between two consecutive values of r is always equal to ρ (i.e. $r^{(w)} = \rho^w \bar{r}$ where $-W \leq w \leq W$ and $\bar{r}$ is a reference radius). The choice of the values for W and ρ is task dependent. For example, suppose we have two images where one is a scaled version of the other and we want to detect the radius of the characteristic neighborhood for two corresponding points. Then W and ρ should be selected to ensure a minimum overlap between the radii intervals for a given scaling factor $s > 1$. From Figure 6.1(a) it can be inferred that the overlap percentage between two radii intervals is:

$$\text{Overlap Percentage} = \frac{\text{Minimum Overlap}}{\text{Radii Interval Length}} = \frac{\rho^W \bar{r} - \rho^{-W} s\bar{r}}{s\bar{r}\left(\rho^W - \rho^{-W}\right)} = \frac{\rho^{2W} - s}{s\left(\rho^{2W} - 1\right)} \tag{6.4}$$

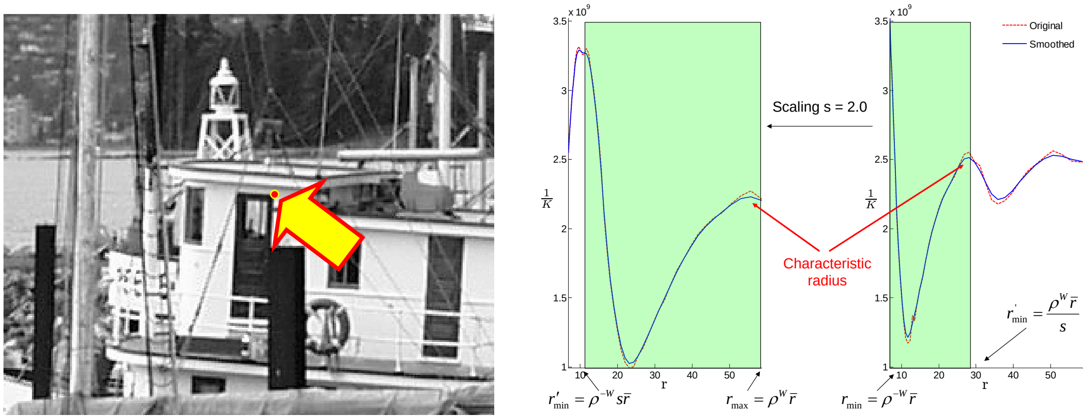
This function is plotted in Figure 6.1(b) for $\rho = 1.05$ and $1 \leq W \leq 22$: if the



(a) (b)

Figure 6.1: (a) This image illustrates pictorially the overlapping (showed in green) between the radii intervals (thick red continuous and dashed lines) for two scaled images $\boldsymbol{I}$ and $\boldsymbol{I}'$. (b) The plot shows how the overlapping percentage varies as a function of the image scaling for a fixed value of ρ and different values of W. As expected, for a fixed W, the smaller the scaling factor the larger the overlap percentage.

scaling factor is $s \approx 2.0$ then for $W = 22$ the overlap percentage is about 45%. Since we can detect the same characteristic dimension for two scaled versions of the neighborhood only if the condition number has a minimum in the overlapping interval, then W and ρ should be chosen according to the expected scaling factor between the images. A common way to tackle this problem is to have an initial estimate of $\bar{r}' = s\bar{r}$ (analogously to what is done in [86]). Figure 6.2 shows an example of the condition number curve for two corresponding points of an image that has been scaled by a factor $s = 2.0$. As far as the speed of the proposed scale detector is concerned, our research-oriented implementation takes about 0.09 seconds to compute 45 values of the condition number for a given image point on a 2.4GHz Pentium 4.



(a) (b)

Figure 6.2: (a) The point $\boldsymbol{x}$ is indicated by the (red) dot pointed by the thick arrow. (b) The inverse of the condition number for the original image and its scaled version ($q = 0$, $\Delta = 40$). The ratio between the points of maximum is $\frac{55.72}{28.14} \approx 2.0$. The fluctuations in the overlapping intervals of $1/\hat{K}_{\boldsymbol{T_\theta},\boldsymbol{x}}$ are consequences of discretization and numerical approximation. The solid curve is the smoothed version of the original dashed curve.

Finally an important remark regarding the numerical stability of $\hat{K}_{\boldsymbol{T_\theta},\boldsymbol{x}}$. As discussed before, the estimate of the condition number is a function of the singular values of the matrix $A(\Omega(\boldsymbol{x}))$, which are affected by the perturbations of the matrix coefficients (see [53], p. 419). In practice we observed that the singular values of A (and hence $\hat{K}_{\boldsymbol{T_\theta},\boldsymbol{x}}$) are noticeably affected by the orientation of the discretization grid used to warp the region $\Omega(\boldsymbol{x})$ on the unit circle (especially when the resolution of the discretization grid is small). To alleviate this problem we align the discretization grid for each value of r to the direction of the average intensity gradient of the image region $\Omega(\boldsymbol{x})$ computed over all the image channels. The condition curves are also smoothed with a Gaussian low pass filter that limits the occurrence of spurious maxima.

6.1.2 Experimental Results

In this section we carry out a set of experiments to validate our approach to detect the characteristic radius of a point neighborhood. Our results will be compared with the scale detection method based on the local maxima of the image Laplacian that in [85] was shown to give the best results. However, before presenting the quantitative results, we would like to give a qualitative proof of concept for our approach by identifying the characteristic radii for a set of corresponding points manually picked for two image pairs related by a scaling (see Figure 6.3). The characteristic radii correspond to the minimum value attained by the 0-Schatten norm condition number for values of r in the interval $[6.8, 58.5]$ pixels (in this experiment $\rho = 1.05$). The neighborhood dimensions appear to be consistent with the structure of the intensity they enclose: the intensity pattern reveals the position, the orientation and the scale despite the possible perturbations introduced by the noise.

Synthetic Experiments

Starting from 7 different natural images whose dimensions are 800×600 pixels, (6 of them are shown in Figure 3.1) we synthesized for each one of them a set of 8 new images related to the original ones via an arbitrary rotation and a scaling (logaritmically spaced in the interval $[1, 4]$). We detected a set of N_p points (about 10^3) in the original image using the Shi-Tomasi detector [116] and we generated the corresponding point pairs using the transformation that relates the original image to each new synthesized version. The characteristic neighborhood of every point was calculated using both the algorithm outlined in Sec-

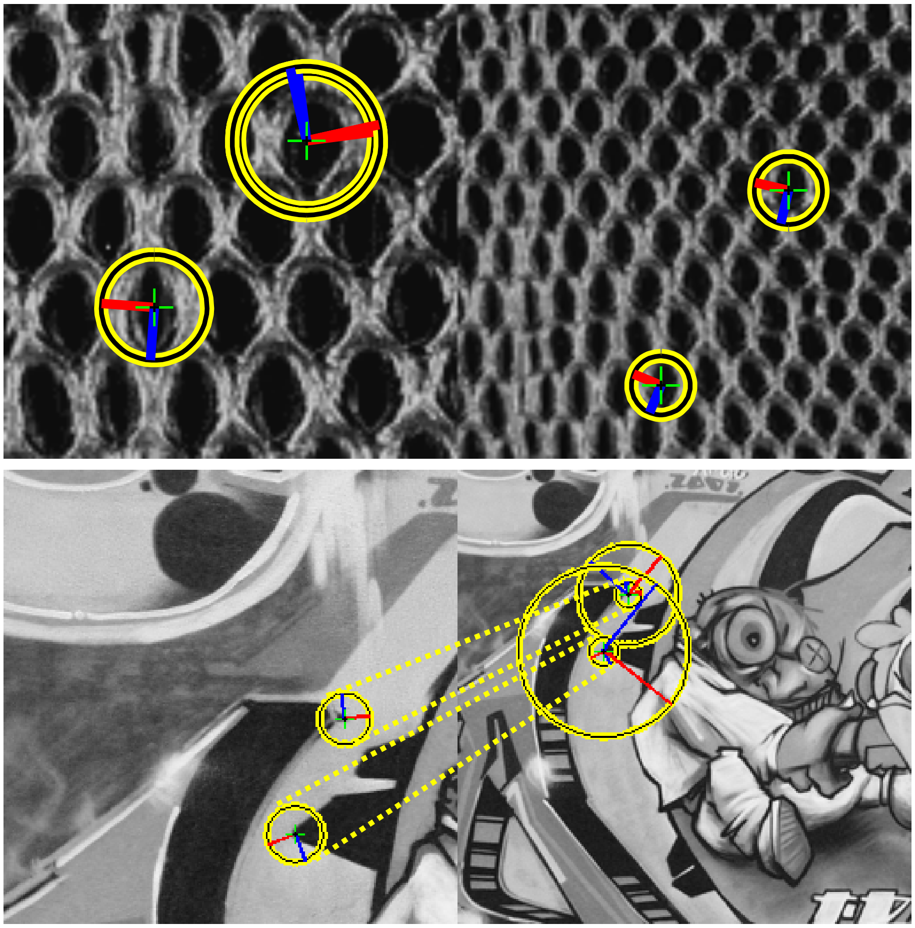


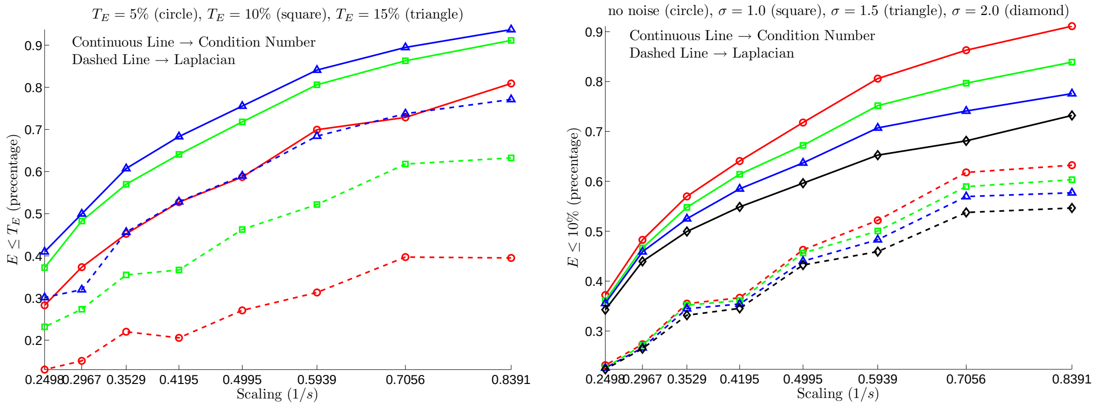
Figure 6.3: These two figures show the characteristic neighborhoods (represented by bright circles) identified by the algorithm described in Section 6.1.1 for a set of points chosen manually ($c = 0$, $\Delta = 40$). Qualitatively, the dimension of the neighborhoods varies consistently with the scale of the image and it defines a distinctive region whose orientation, translation and scale can be easily identified despite the perturbations due to the noise. The low resolution image of the bottom pair generates multiple characteristic radii for each point; however the smaller ones match the radii identified in the higher resolution version of the same image (as shown by the dashed lines).

tion 6.1.1 (Condition Number Detector (CND)) and the algorithm based on the image Laplacian that associates the characteristic radii to the local maxima of $L(\Omega(\boldsymbol{x})) = \sum_{\boldsymbol{y}\in\Omega(\boldsymbol{x})} w(\boldsymbol{y}-\boldsymbol{x}) \left|\sum_{j=1}^{m}\sum_{i=1}^{n} \frac{\partial^2 \boldsymbol{I}_j}{\partial y_i^2}(\boldsymbol{y})\right|$ (Laplacian Detector (LD)). For each corresponding point pair we computed a measure that quantifies the relative discrepancy between the true scaling and the estimated scaling according to:

$$E_i \overset{\text{def}}{=} \min_{\substack{1\le k\le N_r\\ 1\le h\le N'_r}} \frac{1}{s}\left| s - \frac{r_i^{(h)}}{r_i'^{(k)}}\right| \tag{6.5}$$

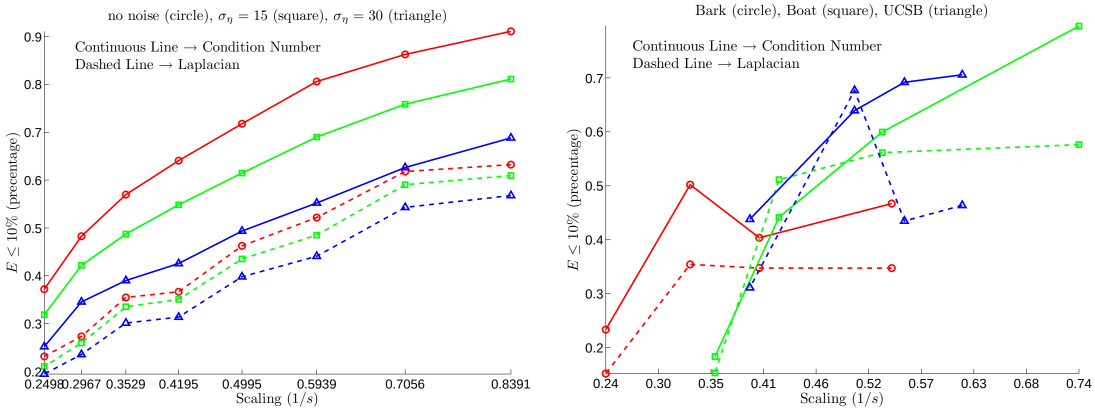
where N_r (respectively N'_r) indicates the number of characteristic radii r_i (respectively r'_i) detected for the point $\boldsymbol{x}_i$ (respectively $\boldsymbol{x}'_i$). In our experiments we choose $N_r = N'_r = 2$ (only the smallest two minima of the CND were considered, or, in the case of the LD, the best two maxima). For each image set we plotted the percentage of points whose error (6.5) was less than 5, 10 and 15 percent versus the scaling. All the experiments were carried out using gray level images and the following parameters: $\rho = 1.05$, $W = 22$ and $\bar{r} = 20$. As far as the CND/LD were concerned, we set $c = 0$ and $\Delta = 20$.

The results of the experiments are summarized in Figures 6.4(a, b) and 6.5(a). Each curve plots the percentage of points with a relative error (6.5) less or equal than the percentage threshold T_E versus the image scaling factor. Over the set of considered images, the CND (continuous line) performs consistently better than the LD (dashed line). This observation remains valid also in the presence of Gaussian perturbations of the position of the point $\boldsymbol{x}$ (see Figure 6.4(b)) and of Gaussian perturbations of the image intensity (see Figure 6.5(a)). In the noise free experiment the CND did not detect a characteristic scale for 12% of the points, whereas the LD for 8% of the points (the percentages are averaged over all the



(a) (b)

Figure 6.4: Figure (a) compares the performance of the CND versus the LD for different values of T_E. Figure (b) compares the performance of the CND versus the LD for $T_E = 10\%$ and different perturbations of the point position $\boldsymbol{x}$.



(a) (b)

Figure 6.5: Figure (a) compares the performance of the CND versus the LD for different perturbations of the image intensity. Figure (b) compares the performance of the CND versus the LD over a set of real images related by transformations that can be approximated by a rotation, a translation and a scaling. All the curves are drawn for $T_E = 10\%$.

scales).

Real Imagery Experiments

Figure 6.5(b) displays the performance of the detectors on three groups of natural images (namely Bark, Boat and UCSB), where each group contains six views related by a known homography that can be approximated with a rotation, a translation and a scaling. The performance of the CND is overall still better than the LD's one, even though a lot of factors that were not modeled in the synthetic experiments (such as projective distortions and non rigid scene changes) tend to reduce the differences between the detectors. We hypothesize that the different performances for the considered image sequences can be explained as follows. The Bark images (for which the difference between the detectors is more evident) are accurately related by a Rotation Scaling and Translation (RST) transformation and closely reflect the experimental conditions of the synthetic experiments. The manmade structures that characterize the UCSB sequence tend to facilitate the job of the detectors and consequently to reduce the differences in their performances (especially for larger scalings). Finally we identify the Boat sequence as the most challenging one, since the RST model is a just crude approximation of the transformation that relates the images. Small changes in the camera position originate projective effects that are not always captured by a simple RST model. Moreover the objects composing the scene do not always satisfy the rigid motion constraints. The effects induced by the RST approximation become non negligible at larger scalings, when the performance of the detectors decreases at the same rate.

Figures 6.6 and 6.7 show some examples of point correspondences that fit an homographic model between images related by large scalings. The point descriptors are calculated over the characteristic neighborhoods. In the top image, the color of the lines connecting the corresponding points indicates the fitting error: red indicates larger errors, green smaller errors. The bottom image displays the quality of the alignment: red or blue areas in the overlapping portion of the images indicate the presence of alignment errors, whereas levels of gray indicates that the alignment is accurate.

6.1.3 Comments

We presented a theoretical framework to detect the characteristic structure of a point neighborhood. Our theory, whose foundations where laid in Chapter 2, is general, in the sense that it applies to multichannel images whose spatial dimension can be greater than 2. During our experiments we noticed that our approach did not benefit much from the use of multiple channels, at least for natural images, where the RGB channels are highly correlated. This is consistent with the experimental results obtained in Chapter 3. However we believe that considering the information of multiple channels could become crucial when dealing with images acquired using multiple sensors (where the channels are highly uncorrelated). The condition number based detector performs consistently better than the state of the art scale detector based on the signature of the image Laplacian. Our future work aims at extending our algorithm within the proposed framework, by considering affine transformations to model the effect of noise and the geometric relations between the images. We believe that such an extension could provide a

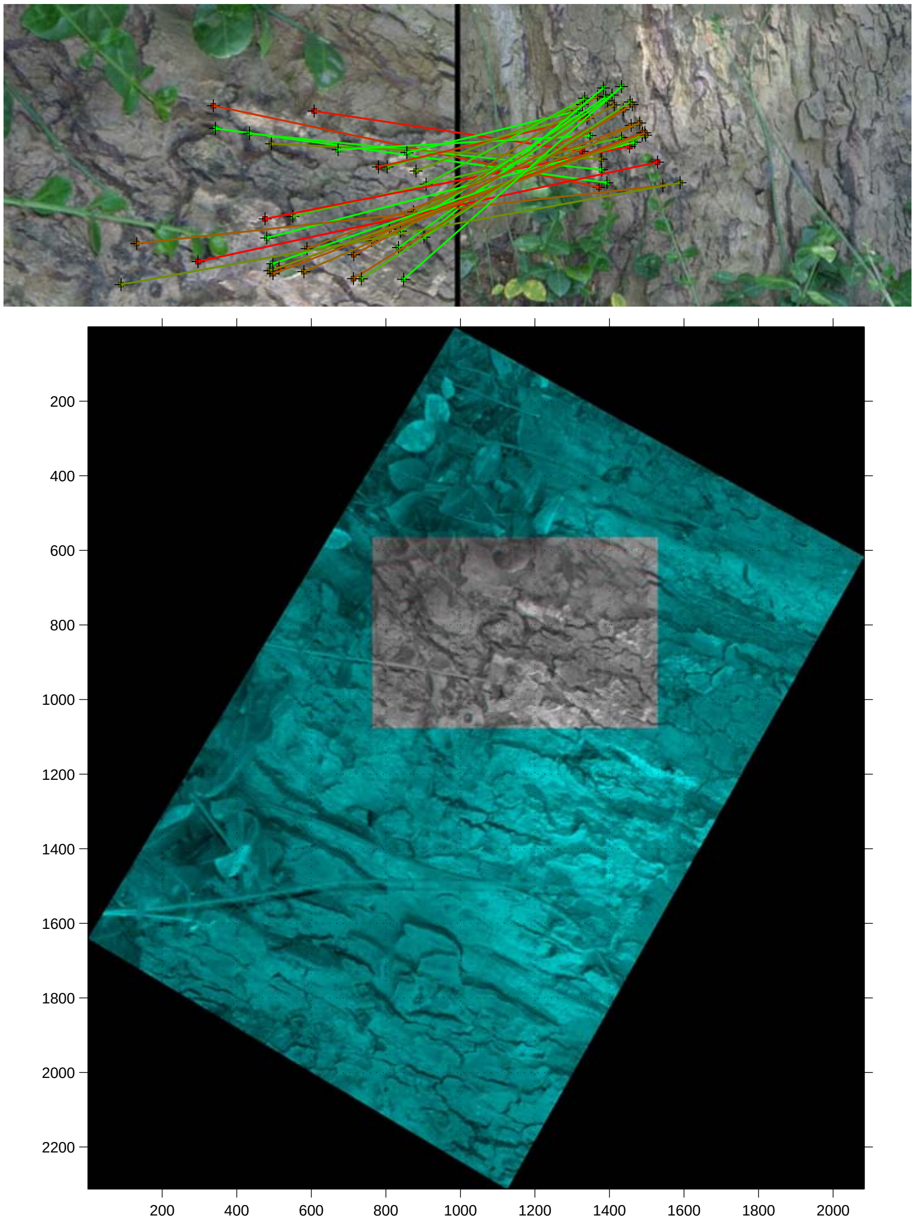


Figure 6.6: The top images display the point correspondences fitting an homographic model between two images of the Bark sequence related by an approximate scaling $s = 2.4$. The color of the lines connecting the corresponding points indicates the fitting error: red indicates larger errors, green smaller errors. The bottom image displays the quality of the alignment: red or blue areas in the overlapping portion of the images indicate the presence of alignment errors, whereas levels of gray indicates that the alignment is accurate.

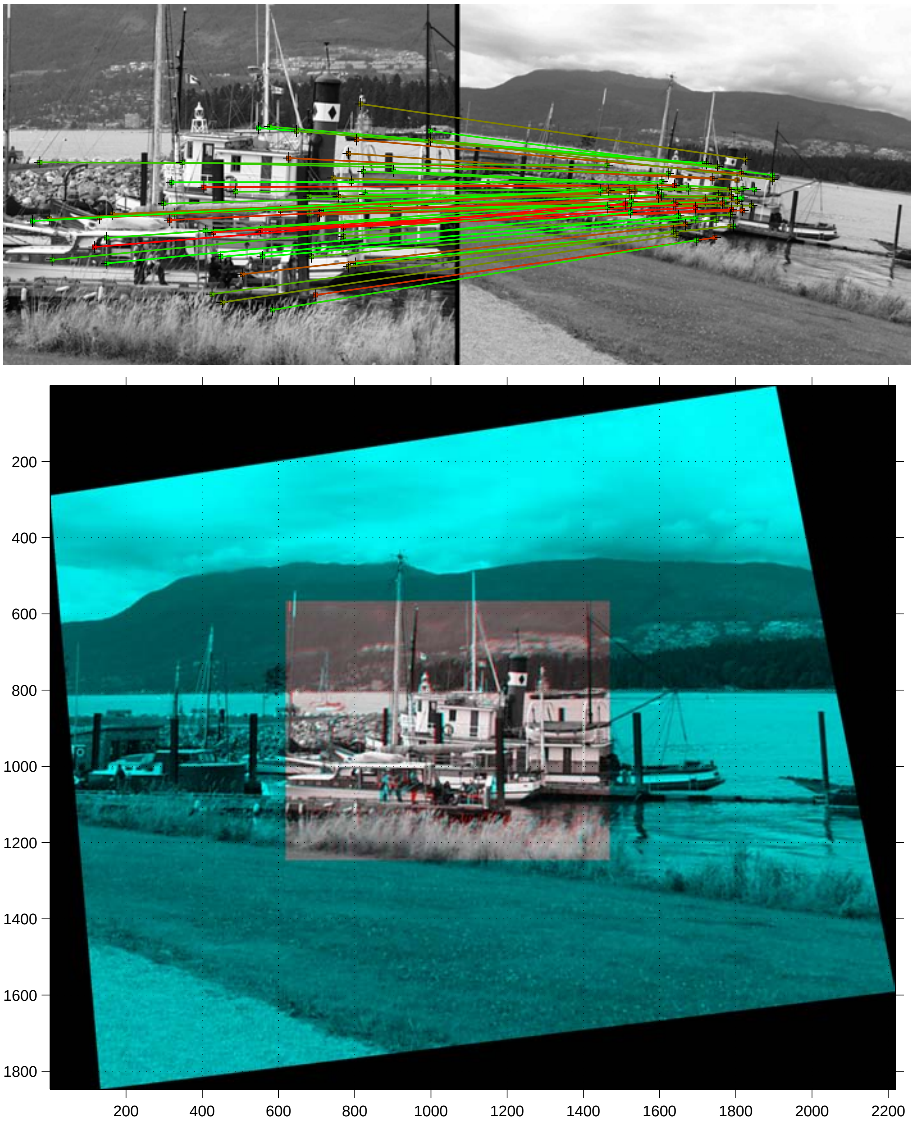


Figure 6.7: The images show the correspondences fitting an homographic model between two images of the Boat sequence related by an approximate scaling $s = 2.2$. Note the registration artifacts due to non rigid motions in the scene. See the caption of Figure 6.6 for more details regarding the meaning of the images.

principled method for region adaptation algorithms that limits the need to resort to empirical and heuristic considerations.

6.2 Image Registration and Mosaicking

In the last few years the quest for robust and efficient algorithms to register several images to the same coordinate system and to create large seamless photomosaics has prompted an intensive research that has been summarized in some remarkable papers [122, 13, 94] and has originated a few commercial applications [82, 84, 83, 28, 17, 104, 34, 22]. In this section we will describe how the frameworks developed in the previous chapters can be integrated in a simple registration system and we will develop a set of tools to obtain seamless mosaicks.

6.2.1 Estimating the Transformation Between Images

As mentioned in the introduction, image registration is the process of aligning two or more images to a coordinate system that is coherent with the three dimensional structure of the scene. There are several transformations that can be used to model the geometric distortion that relates an image pair, such as translations, scalings, rotations, affine and projective mappings. Scene depth discontinuities and self occlusions complicate the task of registering together a set of images, since in general is not possible to obtain a result that is completely coherent with the 3D structure of the scene without explicitly or implicitly estimating its 3D geometry [51, 78]. On the other hand there are several situations where, under relatively mild assumptions, it is still possible to align the images so that the 3D structure of the scene is well approximated. In particular if the camera is far from the imaged objects (where far is relative to the camera focal length) and the scene is essentially planar (so that the depth discontinuities are small with respect to the average distance of the scene from the camera optical center), then an homographic model can describe faithfully the mapping between images. An homography is a non singular linear transformation in the projective space and it can be estimated via the DLT algorithm ([51], p. 88) starting from a set of point correspondences on the camera image plane (the cameras *do not need to be calibrated* beforehand). The problem of estimating homographies is understood very well and the DLT algorithm can be easily embedded in a RANSAC framework that allows one to deal with large quantities of outliers or pseudo outliers, as discussed in Chapter 5. In this section we will focus our attention on how to establish correspondences among the feature points detected using the methods discussed in Chapter 2.

Establishing Tentative Correspondences

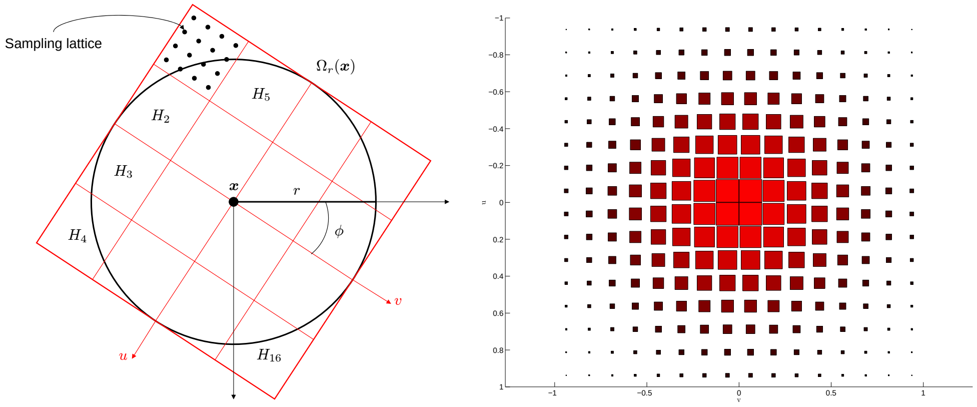
Tentative correspondences between an image pair are established by finding those point pairs whose descriptor distance is minimal. An extensive survey of descriptors constructed processing the intensity information in the neighborhoods of the points can be found in [88].

In our current system we use the same descriptor introduced by Lowe [74] within the SIFT framework. Using the methods described in Chapter 2 and in Section 6.1.1 we can associate with each point $\boldsymbol{x}$ a characteristic circular neighborhood $\Omega_r(\boldsymbol{x})$ (where r denotes the characteristic radius). The dominant gradient direction obtained from equation (2.32) defines the orientation ϕ of the neighborhood. Using this information we create the N_H^2 square patches as shown in Figure 6.8(a). For each patch we construct the histogram H_i that describes the angular distribution (quantized in N_{bins} bins) of the image gradients weighted by their magnitude. Such histograms capture the rough spatial structure of the patch. The gradient values (that are obtained by convolving the image patch with the derivatives of Gaussian filters) are sampled on a lattice composed of N_s^2 points uniformly distributed over the patches. To improve the robustness of the descriptor in presence of misalignments or projective image distortions, the gradient magnitudes are modulated by a Gaussian weighting function (see Figure 6.8(b)). Further, the histograms are smoothed by convolving them with a Gaussian low pass filter and normalized so that the vector formed with their entries has a unitary Euclidean norm. The components whose value is larger than T_H are finally clamped to T_H. The final descriptor is formed concatenating the resulting histograms:

$$\boldsymbol{F}(\Omega_r(\boldsymbol{x})) = \left[\begin{array}{ccc} \bar{H}_1 & \dots & \bar{H}_{N_H} \end{array} \right]^T \tag{6.6}$$

where $\bar{H}_i$ indicates the histogram constructed for the i^{th} patch after weighting, smoothing and clamping. Table 6.1 shows the values of the parameters used in our implementation; it follows that the final descriptor has $N_H^2 N_{bins} = 128$ components.

A correspondence is established when the distance between the descriptors is the minimum among all the possible pairings of points belonging to the image $\boldsymbol{I}$ and to the image $\boldsymbol{I}'$. If not accurately designed, the search of the nearest neighbors can rapidly become computationally very intense, since it requires the evaluation



(a) (b)

Figure 6.8: Figure (a) shows the arrangement of the square patches that are used to compute the histogram that compose the descriptor. In this case $N_H = 4$ and $N_s = 4$. Figure (b) shows how the samples are weighted to improve the descriptor robustness in presence of region misalignments or projective distortions; the larger and brighter is the square marker associated with each sample, the higher is the corresponding weight.

of $N_P N'_P$ Euclidean distances between 128 dimensional vectors, where N_P and N'_P are the number of feature points computed respectively in $\boldsymbol{I}$ and $\boldsymbol{I}'$ (typical values are in the order of thousands). We mitigated this problem by resorting to *dimensionality reduction* and *kd-tree based indexing*.

- **Dimensionality Reduction**. We reduce the dimensionality of the descriptors (6.6) by applying the principal component analysis (PCA). The PCA covariance matrix is computed using 3289 descriptors obtained from natural images. Figure 6.9 shows the first 12 principal components that are used to carry out the projection. In our system we reduced the dimensionality of the descriptors to 32 components.

- ***k*d-tree Based Indexing**. The purpose of *k*d-tree structures [91] is to

Parameter Description	Symbol	Value	Units
Neighborhood radius	r	N. A.	pixels
Number of histograms per edge	N_H	4	N. A.
Number of samples per patch edge	N_s	4	N. A.
Number of histogram bins	N_{bins}	8	N. A.
Standard deviation of the overall weight	σ_W	$\frac{2}{3}$	r
Standard deviation of the differentiation filter	σ_D	$\frac{1}{9}$	r
Standard deviation of the smoothing filter	σ_S	1	bins
Components threshold	T_H	0.2	N. A.

Table 6.1: Summary of the parameters used to implement the descriptors used in the SIFT framework and described in Section 6.2.1.

decompose the descriptor space into a set of relatively small number of cells containing only a few descriptors (or, equivalently, only a few feature points). Hence a kd-tree data structure provides a fast way to identify the nearest neighbor of a query and consequently to establish a point match even when the search space has a large dimensionality. This happens by traversing the tree from the root to the leaves until the cell containing the nearest neighbor is identified. The process is completed by scanning all the descriptors in the cell to identify the one closest to the query.[1] Note that this approach bears a resemblance to the work by Beis et al. [7].

Refining the Correspondences

The tentative set of image correspondences usually contains a large number of mismatches. This may happen for several different reasons such as:

- bad detection of the point neighborhoods,

[1] For our system we used the kd-tree implementation by S. Michael [81].

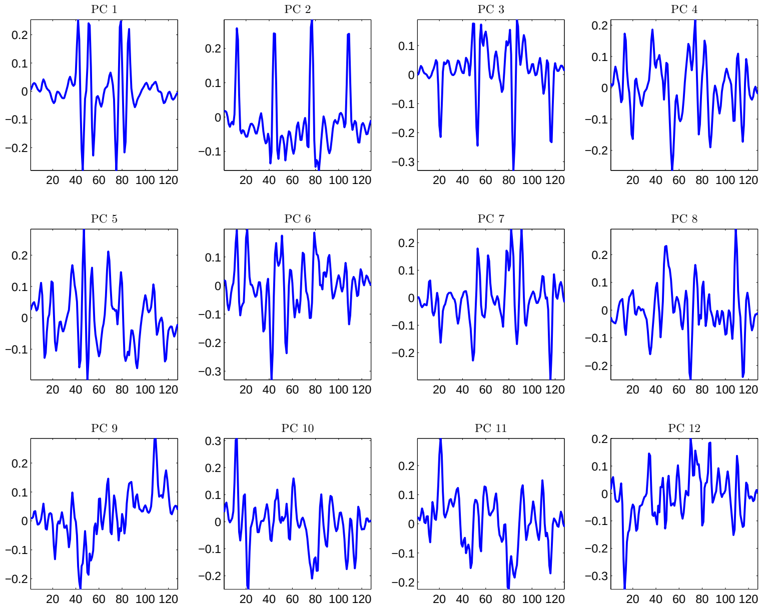


Figure 6.9: The first 12 principal components that define the basis of the low dimensional subspace onto which the descriptors are projected.

- presence of photometric and geometric perturbations that deeply modify the image structure,
- repeating intensity patterns,
- loss of distinctiveness of the descriptors due to the dimensionality reduction
- erroneous detection of the nearest neighbor descriptor using the kd-tree based indexing.

Even if the RANSAC framework is capable of handling large quantities of outliers, the number of iterations (and consequently the time) needed to identify the largest consensus set is roughly inversely proportional to the percentage of inliers. To see this consider equation (5.6), that returns the probability of obtaining a minimal sample set composed only of inliers (rewritten here for sake of convenience):

$$q = \frac{\binom{N_{P,I}}{k}}{\binom{N_P}{k}} \approx \left(\frac{N_{P,I}}{N_P}\right)^k$$

where $N_{P,I}$ indicates the good correspondences out of the total number of tentative matches N_I. Recalling that for $x \to 0$ we can approximate $\log(1+x)$ with x, the number of iterations (5.7) becomes approximately:

$$\left\lceil -\left(\frac{N_P}{N_{P,I}}\right)^k \log \varepsilon \right\rceil$$

One way to reduce the number of iterations is to limit the search space for the minimal sample sets (analogously to what was proposed in [130], [60]). This can be achieved by biasing the probability of selecting the element composing a Minimal Sample Set (MSS). In our implementation the probability of selecting the pair $(\boldsymbol{x}_i, \boldsymbol{x}'_i)$ as an element of the MSS is given by:

$$p_i = \frac{1}{Z} \exp\left(-\frac{1}{2}\frac{d_i^2}{\sigma_{dist}^2}\right) \tag{6.7}$$

where $d_i \overset{\text{def}}{=} \text{dist}(\boldsymbol{F}(\Omega_r(\boldsymbol{x}_i)), \boldsymbol{F}(\Omega'_{r'}(\boldsymbol{x}'_i)))$ is the distance between the descriptors and Z is a normalization factor that ensures that the probabilities sum to one. The scaling factor σ_{dist} is determined so that the ratio between the probability of selecting the pair whose descriptors have the smallest distance and the probability of selecting the pair whose descriptors have the largest distance is equal to ρ_{dist}. In formulæ, if we let $D_{min} \overset{\text{def}}{=} \min_i d_i$ and $D_{max} \overset{\text{def}}{=} \max_i d_i$, we can write:

$$\sigma_{dist}^2 = \frac{D_{min}^2 - D_{max}^2}{2\log\rho}$$

In our implementation we set $\rho_{dist} = 10^{-4}$. Moreover we limit the search for the elements that form the MSSs to those corresponding point whose probability

Parameter Description	Symbol	Value	Units
False Alarm Rate (5.4)	ε	10^{-3}	N. A.
MSS bias probability ratio	ρ_{dist}	10^{-4}	N. A.
Point position noise	σ_η	2	pixels
Inlier probability (5.2)	P_{inlier}	0.9	N. A.
Probability threshold	T_P	0.10	N. A.

Table 6.2: Summary of the RANSAC parameters to identify the point correspondences satisfying an homographic transformation.

(6.7) is larger than $T_P \max_i p_i$. The other parameters used within the RANSAC framework are tabulated in Table 6.2.

The combination of robust descriptors and robust estimation methods enables us to establish image correspondences between pairs that are related by large geometric distortions even if the model used to describe the camera motion is very approximate. Some examples of registration in presence of large scaling between images are shown in Figures 6.6 and 6.7. Figures 6.10 and 6.11 show a couple of examples where the registration of the dominant planar structure of the scene is successfully achieved in presence of large perspective distortions.

6.2.2 Robust Image Equalization

In order to construct photorealistic mosaics it is often necessary to correct the photometric appearance of the images that are to be fused together. In real life scenarios, brightness and contrast changes occur unavoidably because of the image formation process and may be originated by several different and concurrent factors. Some of them are:

- changes of the intensity and of the radiance properties of the light sources,

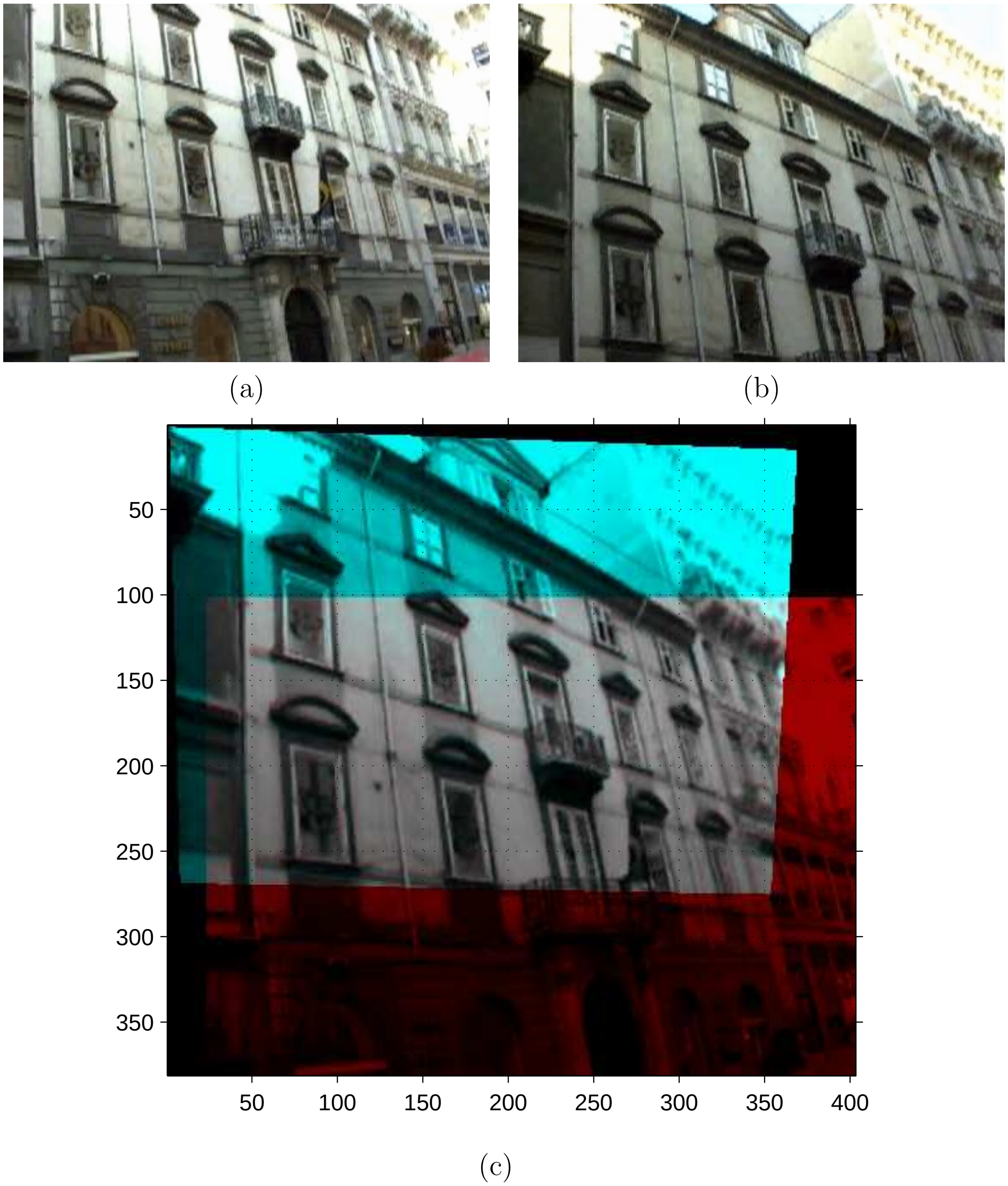
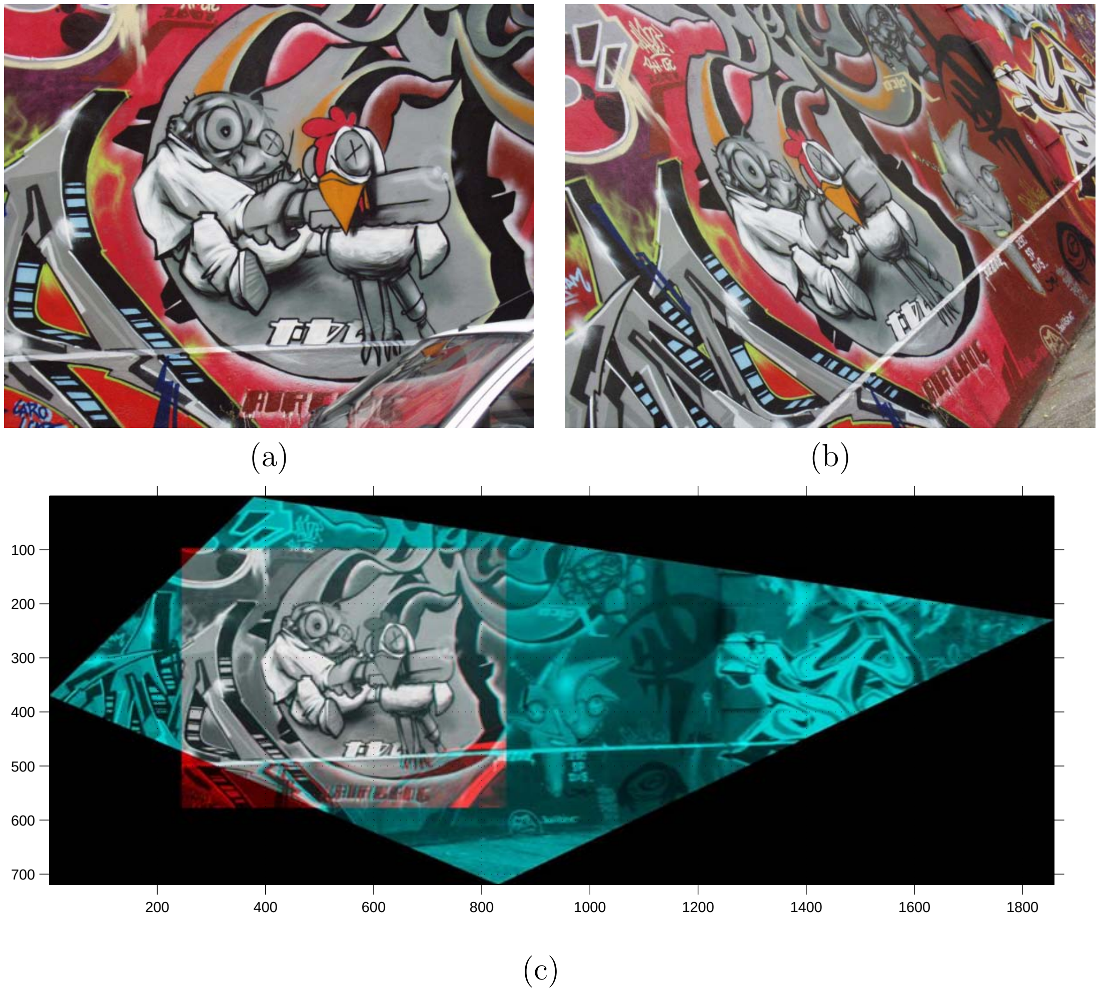


Figure 6.10: The images show the registration of the facade of a building seen from different viewpoints (the house of James Joyce in via S. Nicolò, Sonne, Trieste, Italy). Note that even if the original images are quite noisy and there is strong projective distortion, our registration procedure is able to find 241 point correspondences and the final alignment obtained via a planar homography is very accurate.



(a)

(b)

(c)

Figure 6.11: The images show the registration of the facade of a Graffiti scene seen from different viewpoints. Note that even if there is strong projective distortion, our registration procedure is able to find 78 point correspondences and the final alignment obtained via a planar homography is very accurate.

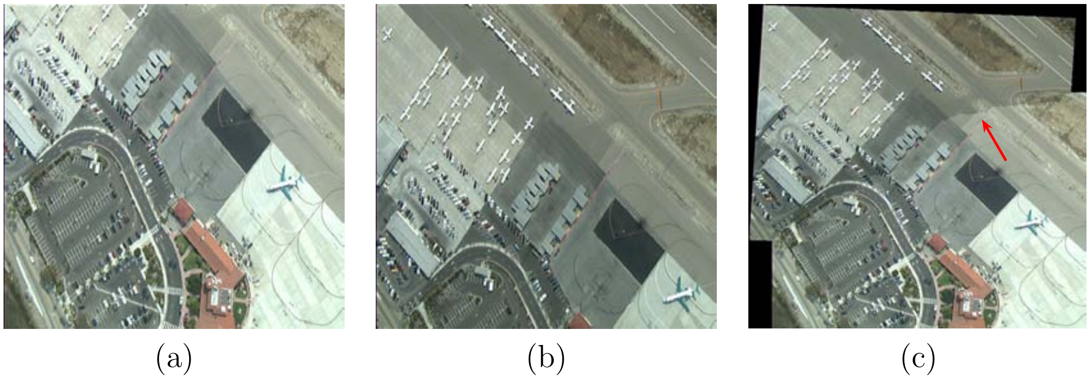


Figure 6.12: Images (a) and (b) show a pair of images to be mosaicked that are non equalized. The resulting mosaic is shown in image (c). The red arrow shows a detail near the stitching line where the different photometric properties of the images become very evident.

- changes of the mutual position between the light source and the surface of the objects composing the scene,
- nonlinearities of the sensor response to the incident light.

An example of a photometric variation is illustrated in Figure 6.12. The images (a) and (b) are acquired modifying the camera sensor properties and this results in a different visual appearance that becomes very noticeable (see the detail pointed by the red arrow in Figure 6.12(c)) when the two images are juxtaposed to create a mosaic[2] even though the scene structure is consistent. To compensate for this undesirable effect we want to estimate a photometric transformation that relates the intensity value of each channel of the images that are to be mosaicked together.

In the literature several approaches have been proposed to model photometric variations in context such as: feature detection [129], feature tracking [116, 125, 57] and area based registration [2, 4]. The main difference between our approach and

[2] The stitching curve is calculated using the method that will be described in Section 6.2.3.

the previous methods is that we want to achieve a *global photometric compensation* whereas the previous methods compensate the photometric distortion *locally*. In some ways our approach resembles the gain compensation procedure introduced by Brown [13]. The intensity value of each point on the camera image plane is a function of the portion of light coming from the source that is reflected by the object surface and is described via the *Bidirectional Reflectance Distribution Function* (BRDF, see [40] p. 60 and [77]). The general form of this function is usually very complex but it can be greatly simplified under the following assumptions:

- surfaces do not fluoresce or emit light,
- the object surface is Lambertian.[3]
- the light source and the cameras are far from the object surface,
- the object surface is approximately planar,

If we also take into consideration the photometric transformations induced by automatic gain control of the CCD amplifier, we can relate the image intensity via the simple affine model:

$$I'(\boldsymbol{x}') = aI(\boldsymbol{x}) + b$$

Since our goal is to create mosaics that are as realistic as possible, we choose to work in the YUV color space, that models the human color perception more faithfully than the usual RGB model. In our algorithm we will also use an affine

[3] A surface is Lambertian if the incident light is scattered so that the apparent brightness of the surface remains the same regardless of the observer's angle of view. For this reason Lambertian surfaces are also called ideal diffuse surfaces.

relation to model the distortion of the chrominance components of the image. The overall model is described by the following set of equations:

$$Y'(\boldsymbol{x}_i') = a_Y Y(\boldsymbol{x}_i) + b_Y \tag{6.8}$$

$$U'(\boldsymbol{x}_i') = a_U U(\boldsymbol{x}_i) + b_U \tag{6.9}$$

$$V'(\boldsymbol{x}_i') = a_V V(\boldsymbol{x}_i) + b_V \tag{6.10}$$

Given the demanding simplifying assumptions used to derive this model and the fact that the scene may be not completely rigid (in the sense that objects may appear or disappear in the overlapping portion of the images), it is necessary to estimate the parameters of the equations (6.8), (6.9) and (6.10) in a robust fashion using a paradigm capable of handling large quantities of outliers. Once again we resort to RANSAC [35].

The implementation of the algorithm is quite straightforward, but there are a few issues that require special care.

- **Aliasing**. Consider the limit situation where we want to equalize two checkerboard images where the intensity alternates from black to white every other pixel. Suppose that the images are related by a simple translation: a registration misalignment of just one pixel will cause the algorithm to map the intensity of black pixels into the intensity of white pixels and viceversa. Even though in real life scenarios such a case would be quite pathological, we guard against analogous situations by performing a low pass filtering of the images before the sampling to construct the MSSs. This prevents abrupt intensity variations to bias the estimates of the transformation parameters, even in the presence of small registration errors.

- **Intensity Saturation**. During the image acquisition process, the camera sensor may saturate, for example in presence of specular reflections. If the reflections are located in corresponding areas of the images, RANSAC may identify the largest consensus set with a group of saturated pixels and consequently the affine transformation becomes simply the identity. To mitigate this problem we discard from the space of the valid MSSs all those sets containing pixels whose luminance is larger than 0.95 and smaller than 0.05 (given that the luminance is normalized between 0 and 1).

- **Computational Complexity**. To reduce the computational burden associated with the identification of the consensus set it is advisable to subsample the overlapping area of the images $\Omega_{overlap}$ (see Figure 6.15(c)).

The examples in Figures 6.18(b), 6.13 and 6.14 illustrate the results obtaining applying the robust equalization technique. The noise threshold to discriminate inliers from outliers was set to $2.5 \cdot 10^{-3}$ (the dynamic range of the YUV channels is $[0, 1]$).

6.2.3 Image Stitching

Consider two images $\boldsymbol{I}$ and $\boldsymbol{I}'$ that are registered to the same coordinate system and let $\Omega_{overlap}$ denote the region where the two images overlap. Our goal is to develop a fast algorithm to identify the best stitching curve in $\Omega_{overlap}$ that will allow us to juxtapose $\boldsymbol{I}$ and $\boldsymbol{I}'$ producing a seamless mosaic. Figure 6.15 illustrates pictorially the problem.

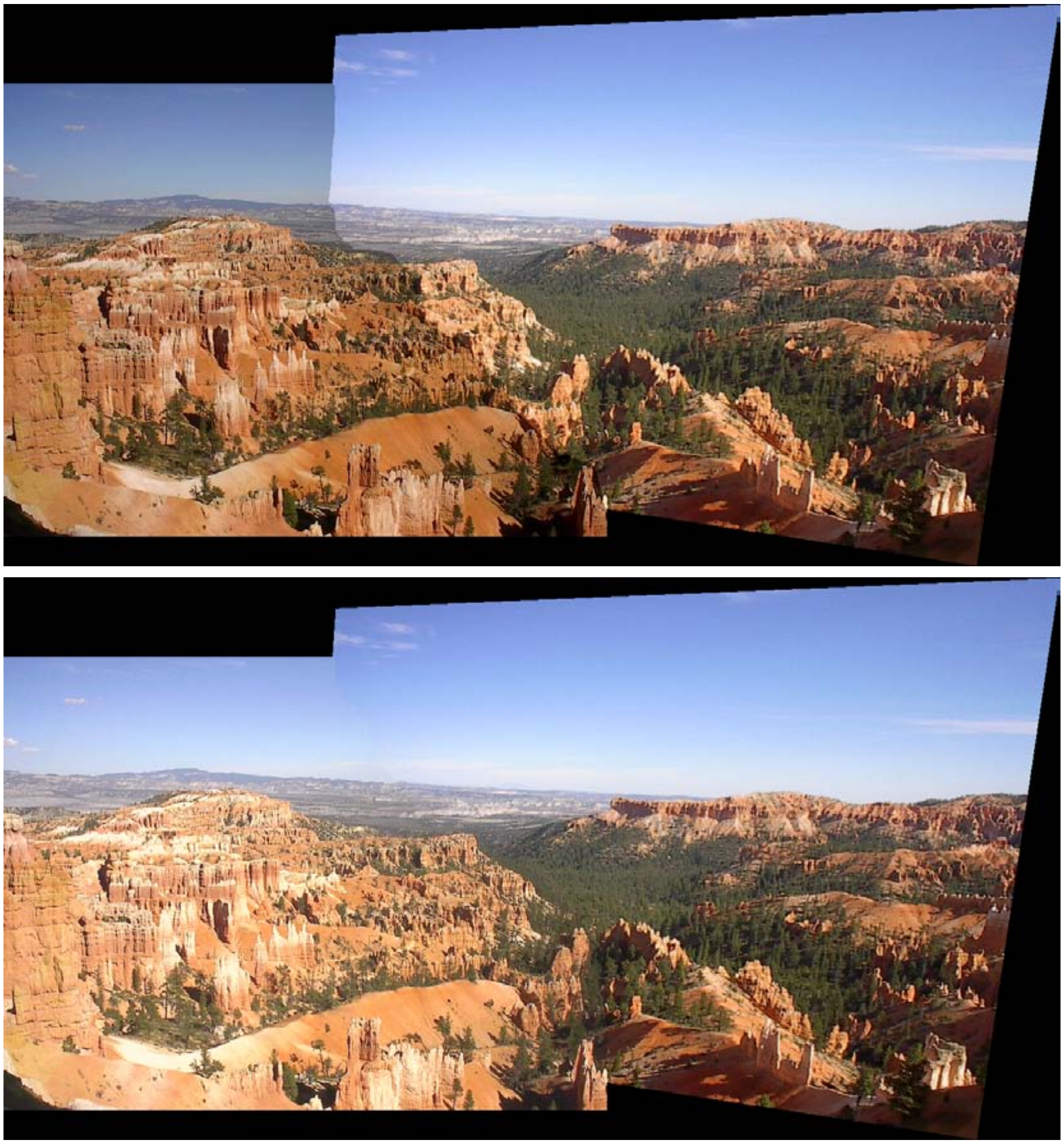


Figure 6.13: The figure displays the mosaic result obtained from two views of Bryce Canyon (image courtesy of G. Pau) without robust equalization (top image) and with robust equalization (bottom image). The stitching curve is obtained using the algorithm described in Section 6.2.3. Note how the photometric stitching artifacts are greatly reduced in the bottom mosaic.

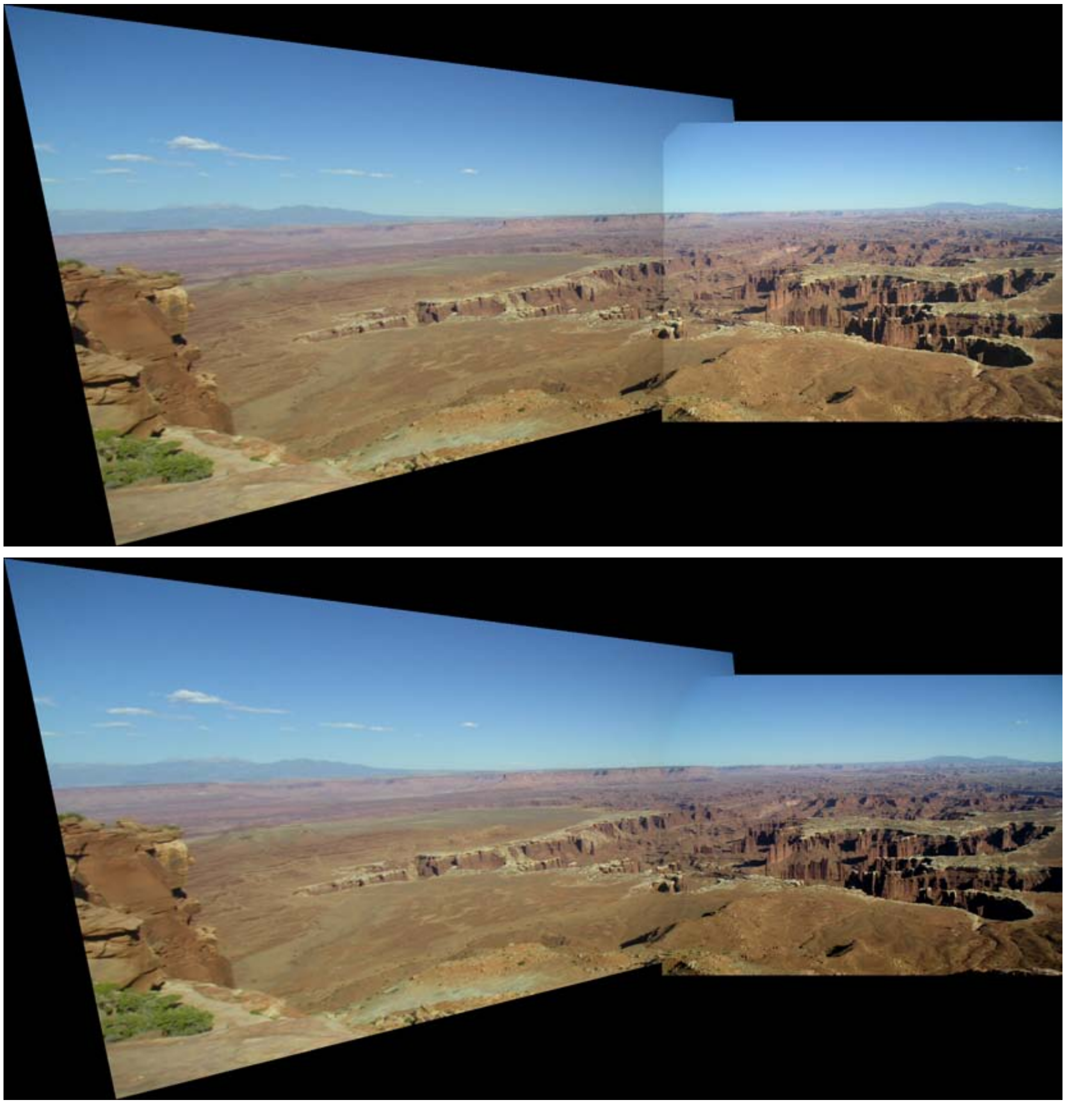
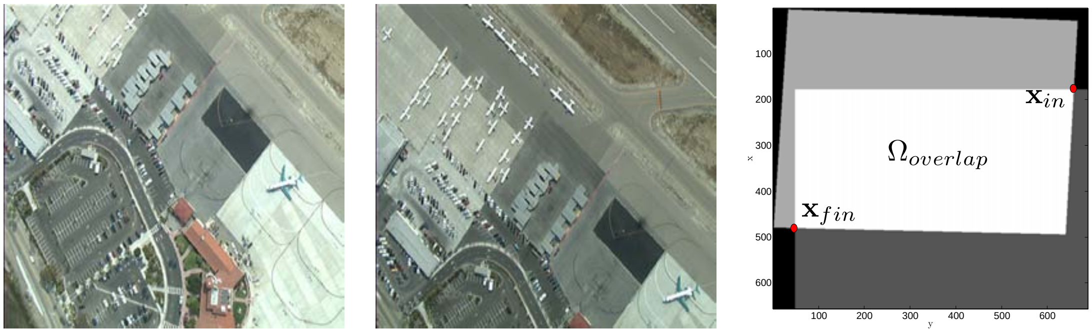


Figure 6.14: The figure displays the mosaic result obtained from two views of the Grand Circle (image courtesy of G. Pau) without robust equalization (top image) and with robust equalization (bottom image). The stitching curve is obtained using the algorithm described in Section 6.2.3. Note how the photometric stitching artifacts are greatly reduced in the bottom mosaic.



(a) (b) (c)

Figure 6.15: The images in (a) and (b) are to be registered with respect to the same coordinate systems. The overlapping region after the registration is displayed in (c). Our goal is to find a simple curve that connects $\boldsymbol{x}_{in}$ to $\boldsymbol{x}_{fin}$ so that the seam between image (a) and image (b) is as little noticeable as possible.

Constructing the Stitching Curves

In order to fix the ideas, in our discussion we will refer to the image pair displayed in Figure 6.15. Our goal is to find a curve that connects the points $\boldsymbol{x}_{in}$ and $\boldsymbol{x}_{fin}$ in such a way that the seam between the images (a) and (b) is as little noticeable as possible.[4] In the next paragraph we will try to formalize this idea.

Suppose that $\boldsymbol{I}$ and $\boldsymbol{I}'$ are aligned with respect to the same coordinate system. We can define the *intensity error function* for each point in the overlapping area as:

$$E(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(G_\sigma * \|\boldsymbol{I} - \boldsymbol{I}'_w\|\right)(\boldsymbol{x})$$

where $\boldsymbol{x} \in \Omega_{overlap}$, G_σ is a Gaussian smoothing kernel and $\boldsymbol{I}'_w$ denotes the warped version of image $\boldsymbol{I}'$ (so that $\boldsymbol{I}'_w(\boldsymbol{x}) = \boldsymbol{I}(\boldsymbol{x})$). Let's also consider a simple curve

[4] In general the overlapping area between two registered images can be as complicated as a convex polygon with six sides if the original images are rectangular and are warped according to a "physically meaningful" planar homography. However such cases can be treated rather straightforwardly as extensions of the basic situation illustrated in Figure 6.15.

that connects the points $\boldsymbol{x}_{in}$ and $\boldsymbol{x}_{fin}$:

$$\begin{aligned}\boldsymbol{\gamma}_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}} : [0,1] &\rightarrow \Omega_{overlap}\\ s &\mapsto \boldsymbol{\gamma}(s)\end{aligned}$$

such that s is its arclength parametrization, $\boldsymbol{\gamma}_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}(0) = \boldsymbol{x}_{in}$, and $\boldsymbol{\gamma}_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}(1) = \boldsymbol{x}_{fin}$. Then the intensity error function integrated along such a curve can be expressed as:

$$C(\gamma_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}) \overset{\text{def}}{=} \int_0^1 f(E(\gamma_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}(s)))\, ds \tag{6.11}$$

where f is an appropriate continuous and monotonically decreasing function (of the intensity error). We can define the minimum cumulative stitching cost at $\boldsymbol{x}$ as the minimum cost (6.11) achievable among all the simple curves that connect $\boldsymbol{x}_{in}$ to $\boldsymbol{x}$:

$$U(\boldsymbol{x}) \overset{\text{def}}{=} \min_{\gamma_{\boldsymbol{x}_{in},\boldsymbol{x}}} C(\gamma_{\boldsymbol{x}_{in},\boldsymbol{x}}) \tag{6.12}$$

Then the optimal stitching curve is the curve that minimizes the minimum cumulative stitching cost at $\boldsymbol{x}_{fin}$, i.e. :

$$\gamma^*_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}} \overset{\text{def}}{=} \operatorname*{argmin}_{\gamma_{\boldsymbol{x}_{in},\boldsymbol{x}}} C(\gamma_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}})$$

As extensively discussed in [23, 29], the solution of (6.12) satisfies the Eikonal equation:

$$\|\nabla U(\boldsymbol{x})\| = f(E(\boldsymbol{x})) \tag{6.13}$$

Equation (6.13) describes the propagation of a wave front with speed $\frac{1}{f(E(\boldsymbol{x}))}$ at each point $\boldsymbol{x}$ neglecting reflections. Hence, the optimal stitching curve corresponds to the path traversed by a wave that propagates starting from the point $\boldsymbol{x}_{in}$ till the point $\boldsymbol{x}_{fin}$ in the least amount of time. From these considerations it follows that

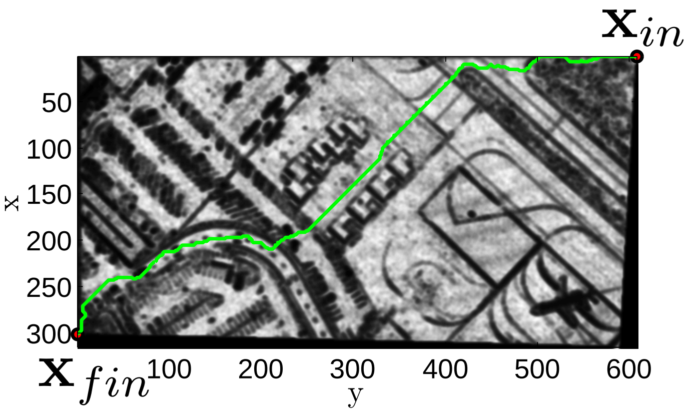


Figure 6.16: Brighter colors indicate the points in $\Omega_{overlap}$ for the images (a) and (b) displayed if Figure 6.15 where the propagation speed of the wave front is faster. The green path corresponds to the optimal stitching curve $\gamma^*_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}$.

we shall design f so that the wave propagates quickly through areas where the error is small and slowly where the error is large. This can be obtained choosing:

$$f(E) = \frac{1}{1 - \frac{2}{\pi}\arctan\left(\alpha\frac{E-E_{min}}{E_{max}-E_{min}}\right)} \tag{6.14}$$

where E_{max} and E_{min} are respectively the maximum and minimum value of E over $\Omega_{overlap}$ and α is a positive scalar that regulates the rate at which the speed goes to zero when the error becomes larger. Such parameter can be set by imposing that the minimum value of the propagation speed is $\rho \in [0, 1]$ times the maximum speed:

$$\alpha = \tan\frac{\pi}{2}(1-\rho)$$

In our implementation we set $\rho = 10^{-3}$. Figure 6.16 shows the speed map in the overlapping area (for the images in Figure 6.15) calculated using the robustly equalized intensities of the images. Note how the function (6.14) rapidly saturates to prevent the stitching curve to traverse locations with large intensity errors.

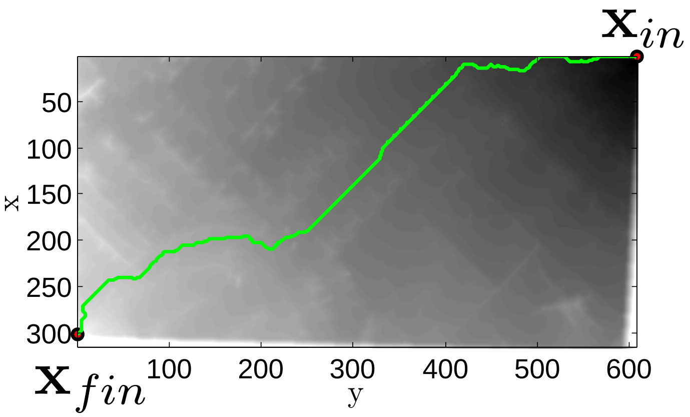


Figure 6.17: The the value of $U(\boldsymbol{x})$ over the overlapping area for the images in Figure 6.15. Brighter colors indicate that the time needed to the wave front to propagate from $\boldsymbol{x}_{in}$ to $\boldsymbol{x}$ is larger. The green path corresponds to the optimal stitching curve $\gamma^*_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}$.

The Algorithm

The stitching algorithm is composed of two steps: first the solution of the Eikonal equation (6.13) is computed for each $\boldsymbol{x}$ in $\Omega_{overlap}$ using the fast marching algorithm [115] and then the stitching curve $\gamma^*_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}$ is computed.

The fast marching method allows us to compute the solution of the Eikonal equation (6.13) in $O(N \log_2 N)$ steps, where N is the number of pixels that compose the overlapping area $\Omega_{overlap}$.[5] Figure 6.17 displays the value of $U(\boldsymbol{x})$: brighter colors indicate that the time needed for the wave front to propagate from $\boldsymbol{x}_{in}$ to $\boldsymbol{x}$ is larger.

In order to determine the best stitching curve between $\boldsymbol{x}_{in}$ and $\boldsymbol{x}_{fin}$, we finally need to descend on the surface $U(\boldsymbol{x})$ from $\boldsymbol{x}_{fin}$ to $\boldsymbol{x}_{in}$. The descent path is unique, since it can be shown that U has a convex structure: starting from any point $\boldsymbol{x}$ and following the gradient descent direction we will always reach $\boldsymbol{x}_{in}$ (as explained in

[5] In our application we used the fast marching implementation by G. Peyré [103].

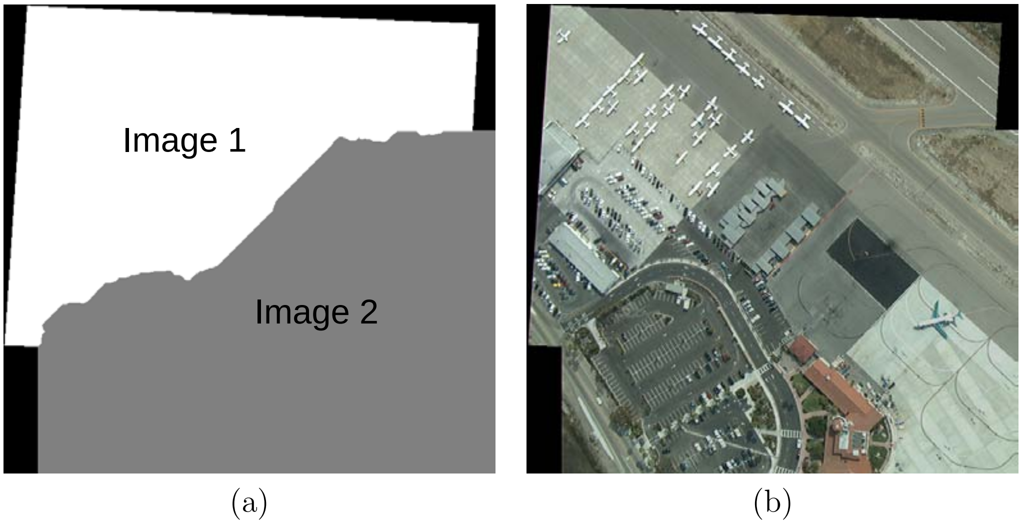


Figure 6.18: Figure (a) shows the partition of the overlapping area induced by the stitching curve. Image (b) shows the final result obtained juxtaposing the robustly equalized source images (see Figure 6.15(a) and (b)) along the optimal stitching curve. Compare Image (b) with Image 6.12(c).

[29], U has only one local minimum that, because of the convexity, coincides with the global minimum $U(\boldsymbol{x}_{in}) = 0$). The stitching curve can be simply constructed via a steepest gradient descent. Such a procedure can be further simplified given the discrete structure of the images: the position of the point $\boldsymbol{x} \in \gamma^*_{\boldsymbol{x}_{in},\boldsymbol{x}_{fin}}$ is updated by looking at its 8 connected neighbors and by moving towards the neighboring point for which U has the smallest value. The best stitching curve for the example shown in Figure 6.15 is the green curve superimposed on Figures 6.16 and 6.17. The final result of the stitching procedure can be seen in Figure 6.18.

Remark 6.2.1 *A final note regarding the possibility of solving the best stitching curve problem using Dijkstra's algorithm [25]. This algorithm solves the minimum cost path problem in a weighted graph and has the same complexity of the* *fast marching method (i.e. $O(N \log_2 N)$, where N is the number of pixels in the overlapping area $\Omega_{overlap}$). In order to use such an algorithm to find the best stitching curve, one should consider the image as an oriented graph where each pixel is a node, and the adjacent vertices are formed by the 4 (or 8) neighboring pixels. However, as pointed out in [29], such an approach would suffer from the so called metrication error: different curves are produced by different choices of the pixel connectivity and such paths are not invariant with respect to simple image transformations such as rotations. An approach based on Dijkstra's algorithm was proposed by Davis [26].*

Improving the Stitching: Wavelet Based Blending

Even if the stitching curve is guaranteed to find the minimum error path to juxtapose two images, it is not given for granted that the seam will be perceptually non noticeable. The first approach to fuse seamlessly two images dates back to 1983, when Burt and Adelson [14] developed a method using splines to blend the subband coefficients obtained via a multiresolution Laplacian decomposition of the images. This seminal work paved the way for other approaches based on more general wavelet decompositions of the images [69, 54, 121, 13]. More recently Zomet, Levin, Peleg, and Weiss proposed a stitching framework called GIST that works in the gradient domain of the image [67, 138]. In general GIST seem to produce consistently good results, even though it is computationally quite demanding. As far as the other stitching algorithms are concerned (including feathering and the approaches based on wavelet decompositions of the images) the authors noted that "*each stitching algorithm works better for some images* *and worse for others*".

In our system we used a wavelet based blending approach scheme the image subbands corresponding to the horizontal (H), vertical (V) and diagonal (D) details are fused about the stitching line using a raised cosine weight. Let d_{max} be the semiwidth of the blending band (note that it remains constant for all the levels of the wavelet decomposition) and let $\boldsymbol{x}^{(l)}$ be closest the point to $\boldsymbol{y}$ on the stitching curve at level l. For the sake of notation we will indicate with the subscript 1 the quantities relative to the first image and with the subscript 2 the quantities relative to the second image after warping. Then we can define the weighting function:

$$W_i^{(l)}(\boldsymbol{y}) = \begin{cases} 1, & \text{for } \|\boldsymbol{y} - \boldsymbol{x}^{(l)}\| > d_{max} \text{ and } \boldsymbol{y} \in \Omega_i^{(l)} \\ -\frac{1}{2}\cos\left[\frac{\pi}{2}\left(\frac{\|\boldsymbol{y}-\boldsymbol{x}^{(l)}\|}{d_{max}} + 1\right)\right] + \frac{1}{2}, & \text{for } \|\boldsymbol{y} - \boldsymbol{x}^{(l)}\| \leq d_{max} \text{ and } \boldsymbol{y} \in \Omega_i^{(l)} \\ \frac{1}{2}\cos\left[\frac{\pi}{2}\left(\frac{\|\boldsymbol{y}-\boldsymbol{x}^{(l)}\|}{d_{max}} + 1\right)\right] + \frac{1}{2}, & \text{for } \|\boldsymbol{y} - \boldsymbol{x}^{(l)}\| \leq d_{max} \text{ and } \boldsymbol{y} \in \Omega_j^{(l)} \\ 0, & \text{for } \|\boldsymbol{y} - \boldsymbol{x}^{(l)}\| > d_{max} \text{ and } \boldsymbol{y} \in \Omega_j^{(l)} \end{cases}$$

where $1 \leq i, j \leq 2$ and $i \neq j$. Note that on the stitching line the resulting signal is given by the average of the original signals. If A indicates the approximation signal and $S \in \{A, H, V, D\}$, then the images are combined at each level according to:

$$\boldsymbol{I}_{S,blended}^{(l)}(\boldsymbol{y}) = W_1^{(l)}(\boldsymbol{y}) I_{1,S}^{(l)}(\boldsymbol{y}) + W_2^{(l)}(\boldsymbol{y}) I_{2,S}^{(l)}(\boldsymbol{y})$$

Figure 6.19 shows an example of the blending procedure. The equalization procedure is not able to compensate for the photometric distortion and the seam between the images is very evident. However the blending procedure is able to produce a seamless mosaic of the two images. On the technical side, we carried out a 3 level decomposition of the images using compactly supported biorthogonal spline wavelets (the same used in JPEG2000) and the blending band has a semiwidth of 16 pixels.

6.2.4 Registration and Mosaic Examples

To give an idea of the practical computational complexity of our system we consider a pair of images whose resolution is 800 × 600 and overlap for about 50% of the area of the images. Our current research-oriented implementation running on Pentium 4 3.40GHz is able to extract and label about 3000 tie points of the images in about 1.5 seconds (`C` and `C++` implementation). The preliminary correspondences are established in about 3 seconds (hybrid Matlab, `C` and `C++` implementation for the dimensionality reduction and kd-tree based nearest neighbor search). Finally 236 matches that fit an homographic model are found in 3 seconds (the RANSAC algorithm is once again an hybrid implementation in Matlab, `C` and `C++`).

The effectiveness of the image registration and mosaicking system is proven by the examples shown in Figure 6.20, Figure 6.21 and 6.22.

6.3 Conclusions

In this chapter we have applied the framework based on condition theory to identify the characteristic structure of a point neighborhood and showed how this can be used to establish matches between images related by large scale variations. We have also studied, designed and implemented the fundamental modules that

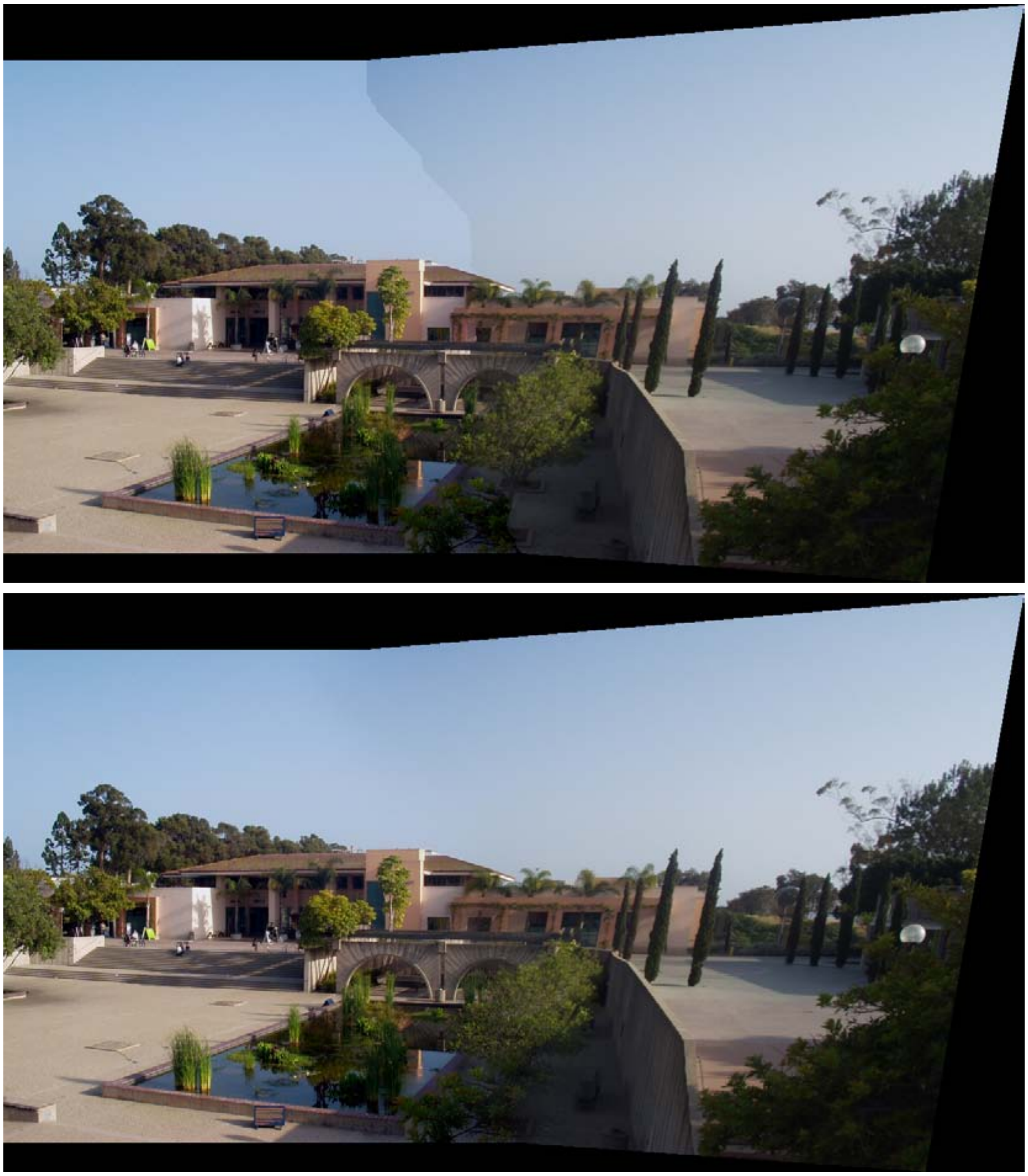


Figure 6.19: The equalization procedure described in Section 6.2.2 is not able to compensate for the photometric distortion and the seam between the images is very evident (top image). However the blending procedure is able to produce a seamless mosaic of the two images (bottom image).

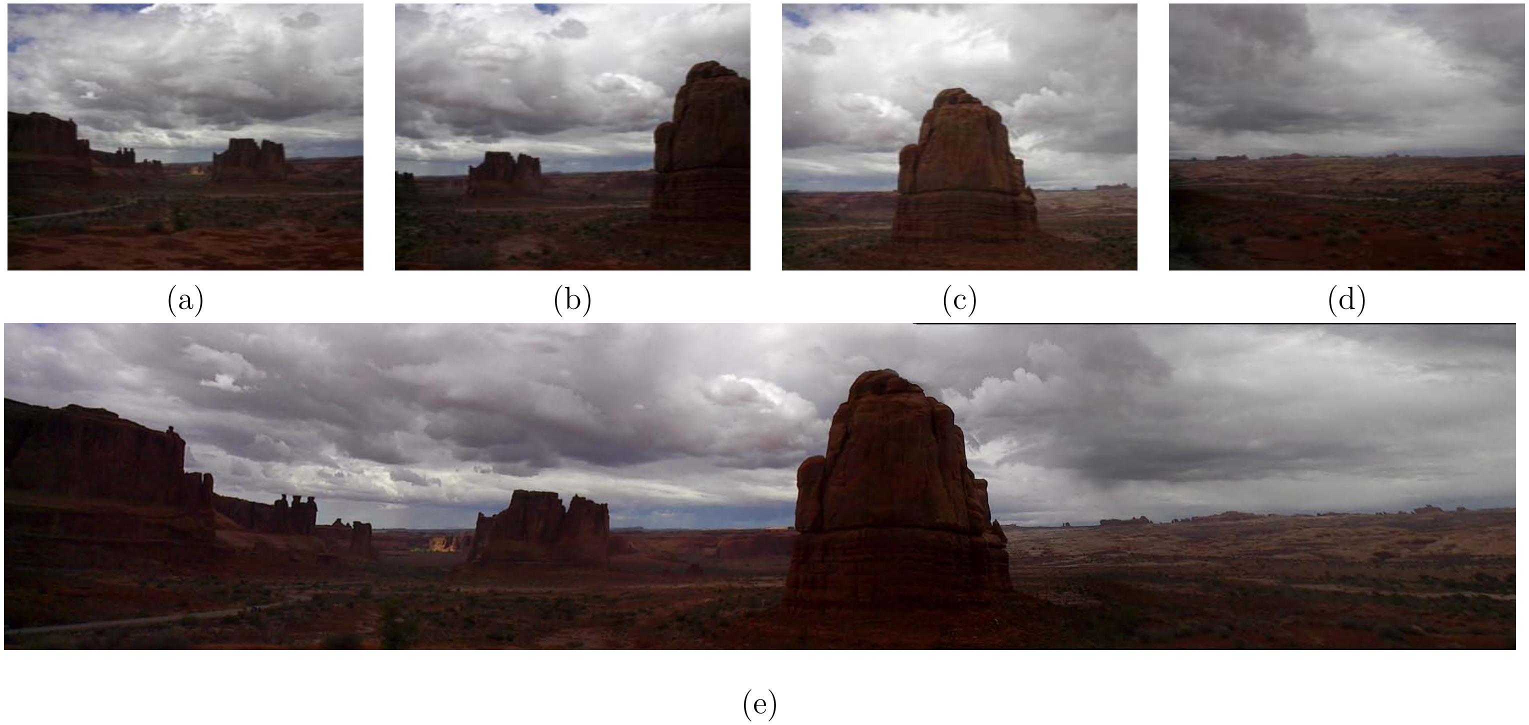
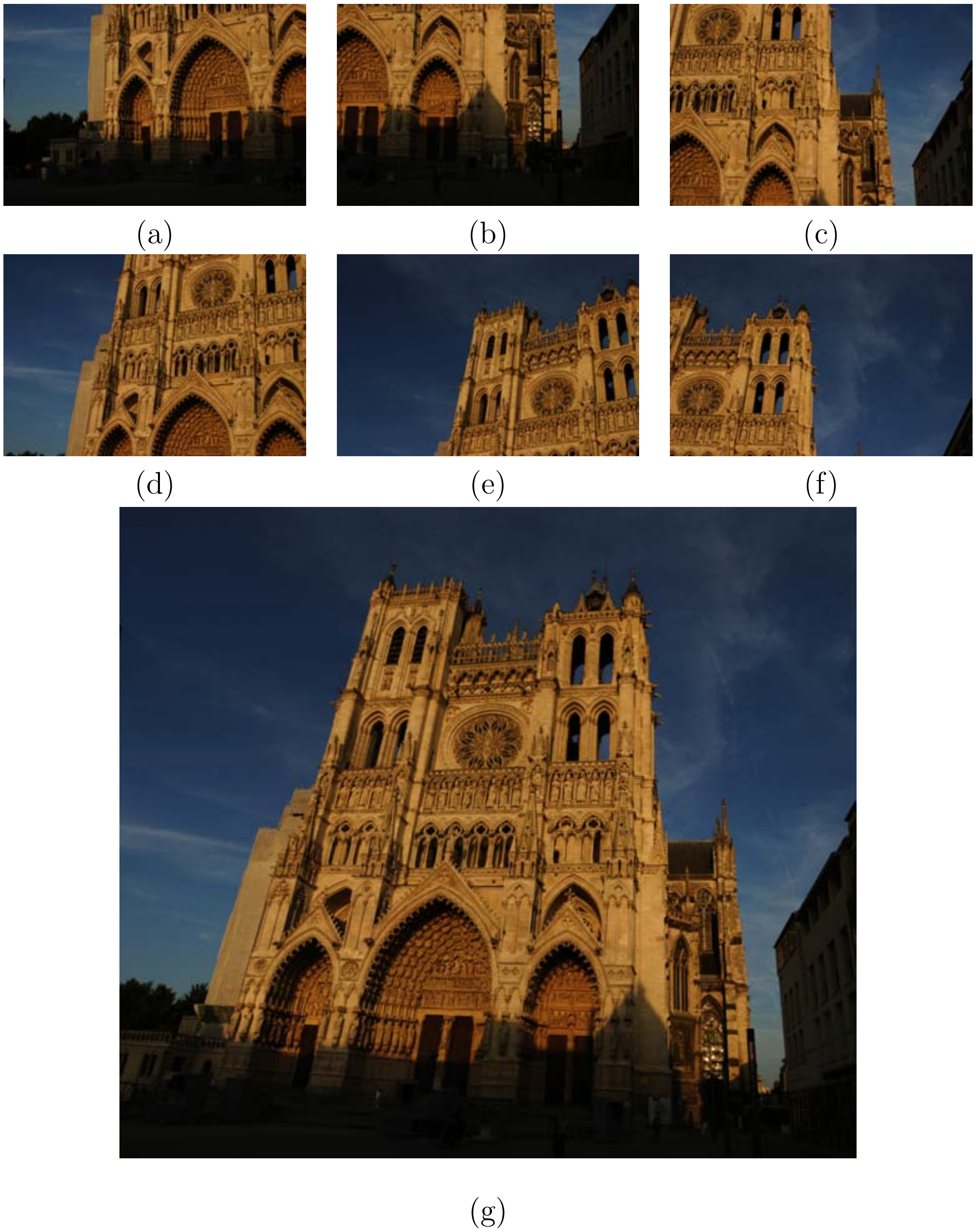


Figure 6.20: The images (a) to (d) of the Grand Circle are registered to form the seamless mosaic in image (e) using the techniques developed in this thesis (image courtesy of G. Pau). Image (e) has been manually cropped to remove the black areas generated during the mosaicking process due to lack of visual information.



(a) (b) (c)

(d) (e) (f)

(g)

Figure 6.21: The images (a) to (f) of the Cathedral of Our Lady of Amiens are registered to form the seamless mosaic in image (g) using the techniques developed in this thesis (image courtesy of J. Nieuwenhuijse, copyright by New House Internet Services BV, www.ptgui.com). Image (g) has been manually cropped to remove the black areas generated during the mosaicking process due to lack of visual information.

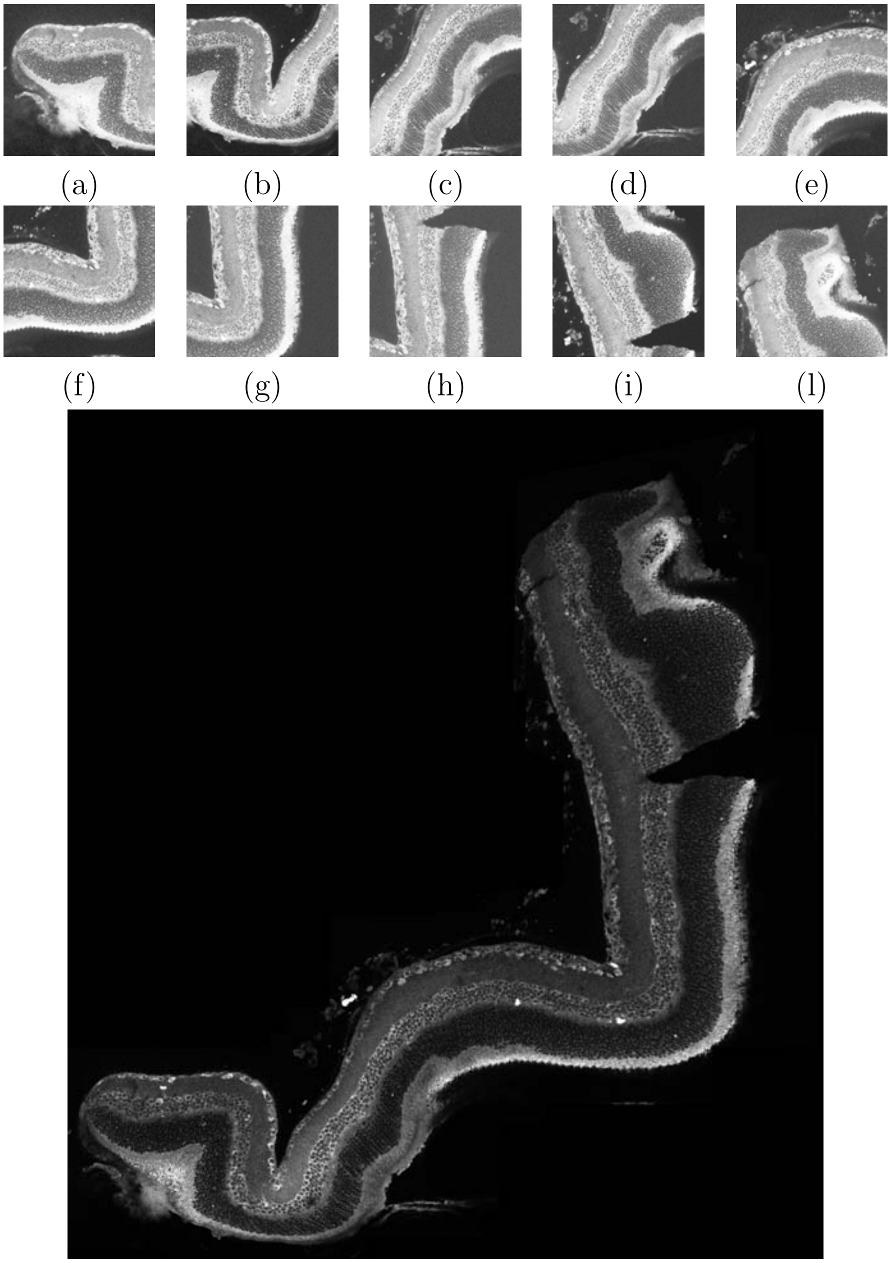


Figure 6.22: The confocal images (a) to (l) of a 3-day detached mouse retina (tissue is stained with bromodeoxyuridine) are registered to form the seamless mosaic in image (m) using the techniques developed in this thesis (image courtesy of Dr. S. K. Fisher, Dr. G. Lewis and Dr. M. Verardo).

register together images in order to produce seamless mosaics (see Figure 1.2). Once again this was done trying to limit the need to resort to empirical consideration by casting the problems in well defined mathematical frameworks. The effectiveness of the resulting modules has been shown by accurately registering and mosaicking images coming from very different domains minimizing the need for parameter tuning and manual intervention.

Chapter 7

Conclusions and Future Work

"Io veggio ben che già mai non si sazia
nostro intelletto, se 'l ver non lo illustra
di fuor dal qual nessun vero si spazia.[1]*"*

Dante

The dominant leitmotif of this dissertation was the development of principled and general mathematical frameworks that allowed us to design a set of modules composing an image registration and mosaicking system. Within these frameworks we were able to tackle a multitude of problems without necessarily resorting to empirical and heuristic considerations. At the same time, our tools enabled us to understand and quantify the tradeoffs between speed, efficiency robustness and accuracy.

[1] Well I perceive that never sated is
Our intellect unless the Truth illume it,
Beyond which nothing true expands itself.

from Paradiso IV, 124–126, translation by H. W. Longfellow

However much remains to be done in order to address the issues that emerged in the course of our investigations. For the sake of discussion we will consider two sets of open problems that we plan to study in the future.

7.1 Low Level Open Problems

In this section we will discuss the issues that stem from the theoretical analysis presented in this dissertation.

Condition Theory for Other Image Analysis Tasks

As mentioned in Chapter 2 and as discussed in the example developed in Appendix B, the results obtained from condition theory are applicable in domains other than point feature detection, when it is necessary to quantify the effect of noise on the estimation of certain quantities. One of the main difficulties is to develop useful and meaningful models that lead to expression of the condition number which are practically computable. Consider for example the notion of $\boldsymbol{T}$-characteristic neighborhood that was introduced in (6.1):

$$\hat{\Omega}(\boldsymbol{x}) \overset{\text{def}}{=} \operatorname*{argmin}_{\delta\Omega} \hat{K}_{\boldsymbol{T}_{\boldsymbol{\theta}},\boldsymbol{x}}(\Omega(\boldsymbol{x}) + \delta\Omega)$$

Setting up a gradient descent method to identify the characteristic neighborhood is not trivial, not only because of the computational and analytical complexity associated with the calculation of the descent direction, but also because the detector response surface is likely to exhibit many local minima. In recent years the numerical analysis community has developed methods to obtain reasonable estimates of the condition number requiring little computational efforts [65] which would allow one to explore more exhaustively the space of the admissible neighborhood configurations. We believe that these contributions can be transferred to several image analysis tasks like the estimation of $\boldsymbol{T}$-characteristic neighborhood mentioned above. In this connection, we note how that the the algorithms proposed by Baumberg in [5] and Mikolajczyk et al. in [86] aim at solving a problem that shares many commonalities with the task mentioned above, the main difference being the way the neighborhood is parameterized and the quantity that is optimized. The extension of our approach to more complex geometric transformation within a gradient descent framework is a challenging research topic that is worth pursuing to tackle the problem of establishing wide baseline correspondences between images in a more principled manner.

Feature Point Localization

Some aspects regarding the localization properties of corner detectors have been studied by Rohr et al. [107], who introduced analytical models of graylevel corners. Even if these models provide some interesting insights regarding the accuracy properties of a corner detector, they could hardly account for all the possible cases that arise when dealing with real images. Lucchese et al. [75] used saddle points obtained from the gradient normal matrix to detect the corners of a checkerboard pattern with high subpixel accuracy. Unfortunately such an approach does not generalize straightforwardly to generic intensity patterns. Torr proposed in [127] to achieve subpixel accuracy by fitting a quadratic surface to the detector response calculated at the nine pixels around a feature point. However this approach introduces a systematic bias, since the surface fitting happens minimizing an algebraic distance (instead of a geometric distance) and therefore is not invariant with respect to common geometric transformations. It is our belief that the discrete representation of an image imposes intrinsic limitations regarding the localization accuracy that can be achieved by a corner detector. Identifying these theoretical bounds is a crucial step in applications where the position of the feature points serves as the input to complex algorithms where accuracy is non negotiable.

Multidimensional Extensions

The theoretical analysis presented in Chapter 2 and Chapter 4 generalize independently from the dimensionality of the signals. This is in general not true as far as the computational issues are concerned.

When analyzing the Spectral Generalized Corner Detector Functions (SGCDF) we introduced some results that can be used to reduce the detector computational complexity in higher dimensional scenarios such as those arising when processing 3D images (like tomographic images or videos where the third dimension is associated with time). We plan to investigate how the results discussed in Chapter 2 can improve the practical design of low complexity detectors. We want also to investigate if there exist image domains where the information contained in different channels could improve the detector performance.

The curve descriptors introduced in Chapter 4 also generalize to spatial domains whose dimension is greater than 2. However the computational complexity grows exponentially and it is necessary to consider alternative methods to solve the Helmholtz equation. Much remains to be done to understand how the descriptors are affected by geometric perturbations of the domain and if they can be used to robustly label both local neighborhoods or 3D domains to establish point correspondences.

Non Rigid Registration

In this dissertation we considered global geometric transformations to register images. This means that the *same* set of parameters is used to model the relation some geometric transformation between the overlapping portions of the images. However there are situations where such an approach is not applicable, because the images that are to be register are subject to deformations that can only be represented locally in closed form. Several techniques have been developed to perform nonrigid registration starting from images that are roughly aligned, especially for medical applications [47]. However the problem of wide baseline non rigid or locally rigid registration it is still open. We are interested in exploring bottom-up approaches where the images that are to be registered are initially decomposed in smaller tiles that are first put in correspondence and then locally registered to one another. Such as approach can benefit from the fast and robust estimation methods that we studied in this dissertation, so that it is possible to cope with the presence of outliers, i.e. portions of images that do not match.

7.2 System Level Open Problems

In this section we will discuss some of the challenges that arise designing an image registration system that is capable to deal with large quantities of images using limited computational resources and minimizing the need for human intervention.

Registration Refinement Procedures

Nowadays it is reasonable to expect that the images that are to be registered and mosaicked have a high resolution (i.e. their dimensions are in the order of thousand square pixels). In this case the refinement process is guided by the desire of exploiting the additional (high resolution) information without compromising the performance of the algorithms. We strongly believe that the methods discussed in the previous chapters can be used to bootstrap the search for image correspondences, possibly within a pyramidal framework where the feature locations are propagated and refined across the pyramid levels. We foresee that robust area based template matching methods [2] could be used to perform such refinement. We also advocate for extensive experiments and for a thorough theoretical analysis to quantify the improved accuracy of the final registration when using high resolution images.

Local Photometric Compensation

In Section 6.2.2 we developed a method to robustly equalize the intensity of the mosaicked images over their overlapping area. To achieve this goal, we assumed that the intensity transformation is the same for all the pixels in the overlapping portion of the images. We believe that such an assumption could be be too simplistic in certain scenarios and therefore more complex models might be required. The first extension that we would like to introduce involves the derivation of an intensity transformation that depends on a more accurate study of the BRDF function so that the position of the pixels is also taken into consideration.

Constructing Minimum Distortion Panoramas

In this dissertation we focused on the registration of image pairs. However in many situations required to register a whole set of images (call this set $\mathcal{I}$) in order to produce large mosaics which are frequently called *panoramic views* or *panoramas*. One immediately recognizes that the complexity of detecting correspondences between all the images that are to be registered is quadratic in the number of the images and therefore any naive approach becomes computationally expensive as soon as the cardinality of $\mathcal{I}$ is large. For a numerical example suppose that we are able to register two images in 10 seconds. If $\mathcal{I}$ contains 10 images then the processing of all the possible pairs takes about 9 minutes. To improve the performance of the system we plan to use two approaches. First we plan to develop a pair rejection criterion based on the results of the preliminary nearest neighbors matchings. We expect that the distances of the descriptors of the preliminary matched points follows distinctive distributions when the images actually overlap and when they do not.

If we consider a graph whose vertices are the images to be registered and whose edges represent the transformation between images, then the *forest of spanning trees* of such a graph describes the set of mosaics that can be obtained from $\mathcal{I}$. Two vertices of the graph are adjacent if and only if it is possible to establish a transformation between the two images. If we associate to each edge a weight that is proportional to the distortion that is necessary to register the connected images (perhaps using the model distance introduced in Section 5.2.2), then the forest of *minimum* spanning trees will return the set of mosaics that are obtained minimizing the total warping distortion.

2.5D Registration

When multiple images are to be registered together, the final mosaic depends on the image that is chosen as reference view. As mentioned before, one way to choose the reference view is to identify which is the minimum distortion panorama. On the other hand, if for example we are interested in creating a mosaic out of a video sequence, varying dynamically the reference view will result in a "spatio-temporal mosaic", i.e. a video where each frame consist of a mosaic produced with a different reference view. Such a temporally varying mosaic could be useful to enhance the perception of depth discontinuities that could otherwise be lost in a static scenario.

Automatic Quality Assessment of Registration

To the best of our knowledge, there is almost no work that aims at assessing automatically the quality of the registration. We believe that this task has a great importance everywhere the accuracy of the registration is an issue. A simple comparison of the image intensities is usually not enough, since the presence of local outliers due to object motions, specularties, saturations, depth discontinuities, occlusions may cause erroneous evaluations of the quality of the registration. We believe that local approaches based on robust comparisons of the intensities in the neighborhoods of the corresponding points are likely to produce more reliable and consistent results.

Appendix A

Some Useful Analytical Results

"Brevissimus quisque dilucidissimus est.[1]*"*

Cicero

A.1 Some Useful Inequalities

Lemma A.1.1 (Norm Inclusion) *Let $\boldsymbol{x}^* = x^* \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T \in \mathbb{R}^p$. Then for any $\boldsymbol{x} \in \mathbb{R}^p$ such that $\|\boldsymbol{x}\|_q = \|\boldsymbol{x}^*\|_q$ we have that $\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}^*\|_1$.*

Proof: The finite form of the Jensen's inequality ([49], p. 83) states that for any real continuous convex function ϕ we have:

$$\phi\left(\frac{\sum_{j=1}^p x_j}{p}\right) \leq \frac{\sum_{j=1}^p \phi(x_j)}{p} \tag{A.1}$$

If we consider the convex function $\phi(x) = x^q$, then (A.1) reads as:

$$\left(\frac{\sum_{j=1}^p x_j}{p}\right)^q \leq \frac{\sum_{j=1}^p x_j^q}{p}$$

or equivalently:

$$\|\boldsymbol{x}\|_1 = \sum_{j=1}^p x_j \leq p\left(\frac{\sum_{j=1}^p x_j^q}{p}\right)^{\frac{1}{q}} = p^{1-\frac{1}{q}}\|\boldsymbol{x}\|_q$$

By assumption $\|\boldsymbol{x}\|_q = \|\boldsymbol{x}^*\|_q = p^{\frac{1}{q}}x^*$ and therefore the lemma is proved once we observe that we can write $\|\boldsymbol{x}\|_1 \leq px^* = \|\boldsymbol{x}^*\|_1$. ■

[1] The briefer, the clearer.

A.2 Some Linear Algebra Facts

The results contained in this section are analyzed in greater details in [44, 53, 80]. Here we will provide a list of some basic results that have been tailored for the proofs and the applications described in the previous chapters.

A.2.1 Matrix Norms

Let's first introduce the vector *q-norm* defined for $q \geq 1$:

$$\|\boldsymbol{x}\|_q \overset{\text{def}}{=} \left(\sum_i |x_i|^q \right)^{\frac{1}{q}} \tag{A.2}$$

When q is omitted it is assumed to be equal to 2 (Euclidean norm). Vector q-norms are convex functions. Definition (A.2) allows one to define the *induced matrix q-norm* as:

$$\|A\|_q \overset{\text{def}}{=} \sup_{\boldsymbol{x} \neq 0} \frac{\|A\boldsymbol{x}\|_q}{\|\boldsymbol{x}\|_q} \tag{A.3}$$

It can be shown that $\|A\|_2 = \sigma_{max}(A)$, where $\sigma_{max}(A)$ is the maximum singular value of the matrix A. The *Schatten matrix q-norm* is defined as:

$$\|A\|_{S,q} \overset{\text{def}}{=} \left(\sum_i \sigma_i(A)^q \right)^{\frac{1}{q}} \tag{A.4}$$

where $\sigma_i(A)$ is the i^{th} singular value of the matrix A. Note that $\|A\|_{S,\infty} = \lim_{q \to \infty} \|A\|_{S,q} = \sigma_{max}(A)$, i.e. the ∞-Schatten norm coincides with the matrix spectral norm. The Schatten q-norm is a unitarily invariant norm.

A.2.2 Spectral Properties of Symmetric Matrices

Lemma A.2.1 *Consider a matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$. The squared singular values of A coincide with the eigenvalues of $A^T A$.*

Proof: Consider the SVD decomposition $A = U \Sigma V^T$. Then, because of the orthogonality of U, the following chain of equations hold:

$$A^T A = \left(U \Sigma V^T \right)^T U \Sigma V^T = V \Sigma^T \Sigma V^T$$

The claim is proved observing that $\Sigma \Sigma^T = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$. ■

Lemma A.2.2 *Consider a full rank matrix $A \in \mathbb{R}^{m\times n}$ with $m \geq n$. Then:*

$$\|A^\dagger\|_2^2 = \frac{1}{\lambda_{min}(A^TA)}$$

where $\lambda_{min}(A^TA)$ indicates the smallest eigenvalue of A^TA.

Proof: For any matrix M we have $\|M\|_2^2 = \lambda_{max}(MM^T)$, where $\lambda_{max}(MM^T)$ is the largest eigenvalue of MM^T. Therefore:

$$\|(A^TA)^{-1}A^T\|_2^2 = \lambda_{max}((A^TA)^{-1}A^TA(A^TA)^{-1}) = \lambda_{max}((A^TA)^{-1}) = \frac{1}{\lambda_{min}(A^TA)}$$

■

Lemma A.2.3 *Let $P \in \mathbb{R}^{n\times n}$ be an orthogonal projection matrix that orthogonally projects the point $\boldsymbol{x} \in \mathbb{R}^n$ onto the $n_P < n$ dimensional space generated by the orthonormal basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_P}$. Then if we extend the basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_P}$ to span the entire space $\mathbb{R}^n$, then singular value decomposition of P is:*

$$P = \begin{bmatrix} V & \boldsymbol{u}_1 & \ldots & \boldsymbol{u}_{n-n_P} \end{bmatrix} \begin{bmatrix} I_{n_P} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^T \\ \boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_{n-n_P}^T \end{bmatrix} \tag{A.5}$$

where $V = \begin{bmatrix} \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_{n_P} \end{bmatrix}$

Proof: The matrix associated with the orthogonal projection onto the column space of V is $P = VV^T$, which can be rewritten as:

$$P = \begin{bmatrix} V & \overline{\boldsymbol{u}}_1 & \ldots & \overline{\boldsymbol{u}}_{n-n_P} \end{bmatrix} \begin{bmatrix} I_{n_P} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^T \\ \overline{\boldsymbol{u}}_1^T \\ \vdots \\ \overline{\boldsymbol{u}}_{n-n_P}^T \end{bmatrix}$$

The SVD decomposition being unique, the assertion holds true. ■

A.2.3 Interlacing Properties of the Singular Values

Theorem A.2.4 *Let $A \in \mathbb{R}^{m\times p}$ be an arbitrary matrix and $A_c^{(l)} \in \mathbb{R}^{m\times p-l}$ be the matrix obtained by deleting any l columns of A. If $m \geq p - l$, then for any $1 \leq i \leq p - l$:*

$$\sigma_i \geq \sigma_{c,i}^{(l)} \geq \sigma_{i+l} \tag{A.6}$$

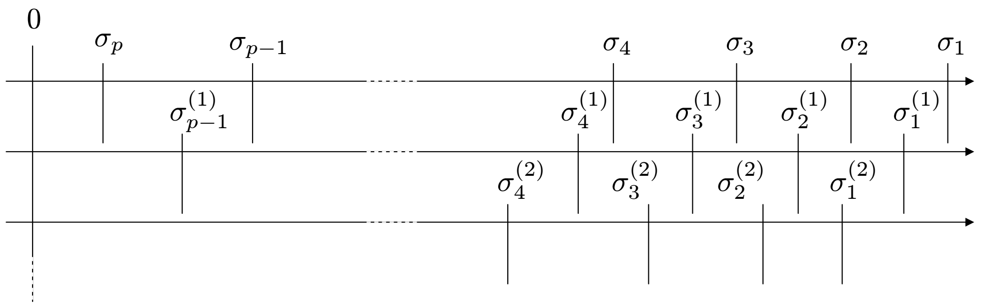


Figure A.1: The interlacing of the singular values of a generic matrix as a consequence of repeated column removal. The number in parenthesis indicates the number of columns that have been removed.

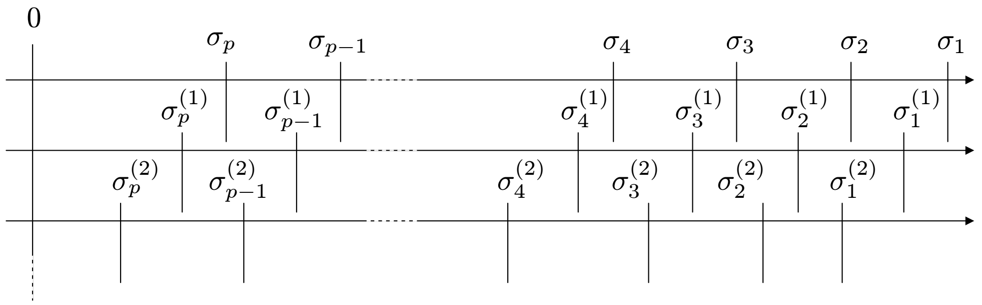


Figure A.2: The interlacing of the singular values of a generic matrix as a consequence of repeated row removal. The number in parenthesis indicates the number of rows that have been removed.

Let $A_r^{(l)} \in \mathbb{R}^{m-l \times p}$ be the matrix obtained by deleting any l rows of A. If $m-l > p$ then for any $1 \leq i \leq p-l$:

$$\sigma_i \geq \sigma_{r,i}^{(l)} \geq \sigma_{i+l} \tag{A.7}$$

Proof: The fundamental form of the interlacing theorem (see [53], p. 419) says that when any of the columns of A are removed to generate the matrix $A_c^{(1)}$, the singular values interlace according to:

$$\sigma_1 \geq \sigma_{c,1} \geq \sigma_2 \geq \sigma_{c,2} \geq \ldots \geq \sigma_{p-1} \geq \sigma_{c,p-1} \geq \sigma_p$$

When the process of removing columns is iterated, the singular values are arranged as shown in Figure A.1. This establishes the first part of the theorem. On the other hand, when any of the rows of A are removed to generate the matrix $A_r^{(1)}$, the singular values interlace according to:

$$\sigma_1 \geq \sigma_{r,1} \geq \sigma_2 \geq \sigma_{r,2} \geq \ldots \geq \sigma_p \geq \sigma_{r,p}$$

and the arrangement depicted in Figure A.1 validates the second claim of the theorem. ■

A.2.4 Fast Diagonalization of Symmetric 2×2 Matrices

Consider a real symmetric matrix $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. We want to diagonalize A using the least number of operations, so that:

$$A = \begin{bmatrix} C & -S \\ S & C \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} C & S \\ -S & C \end{bmatrix}$$

where $C = \cos\theta$ and $S = \sin\theta$. The eigenvalues can be computed first defining:

$$\alpha = a + c$$

$$\beta = (a - c)^2$$

$$\gamma = 4b^2$$

$$\delta = \sqrt{\beta + \gamma}$$

and then letting:

$$\lambda_1 = 0.5(\alpha - \delta)$$

$$\lambda_2 = 0.5(\alpha + \delta)$$

for a total number of 5 additions, 5 multiplications and the extraction of a square root. Finally the angle θ that defines the eigenvectors univocally (modulo a reflection about the origin) can be obtained as:

$$\theta = \arctan \frac{\lambda_1 - a}{b}$$

A.3 Some Optimization Facts

In this section we will present some optimization results that turn out to be useful for the theory developed in the previous chapters. An extensive treatment of optimization theory can be found in [9, 76].

Theorem A.3.1 (Lagrange Multipliers Necessary Conditions) *Let $\boldsymbol{x}^*$ be a local minimum for $f : \mathbb{R}^p \to \mathbb{R}$ subject to the equality constraints:*

$$\boldsymbol{h}(\boldsymbol{x}) = 0$$

where $\boldsymbol{h} : \mathbb{R}^p \to \mathbb{R}^m$. Suppose the Jacobian $J\boldsymbol{h}(\boldsymbol{x}^*)$ is full rank (i.e. the gradients of the constraints at $\boldsymbol{x}^*$ are linearly independent). Let's consider the Lagrangian function:

$$L(\boldsymbol{x}, \boldsymbol{\gamma}) = f(\boldsymbol{x}) + \boldsymbol{\gamma}^T \boldsymbol{h}(\boldsymbol{x})$$

Then there exist a unique Lagrange multiplier vector $\boldsymbol{\gamma}^*$ such that:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\gamma}^*) = 0$$

Theorem A.3.2 (Karush-Kuhn-Tucker Necessary Conditions) *Let $\boldsymbol{x}^*$ be a local minimum for $f : \mathbb{R}^p \to \mathbb{R}$ subject to the equality constraints:*

$$\boldsymbol{h}(\boldsymbol{x}) = 0$$

where $\boldsymbol{h} : \mathbb{R}^p \to \mathbb{R}^m$ and to the inequality constraints:

$$\boldsymbol{g}(\boldsymbol{x}) \leq 0$$

where $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^r$ and the inequality operator is applied component-wise. Suppose f, $\boldsymbol{h}$ and $\boldsymbol{g}$ are continuously differentiable and define the Lagrangian function to be:

$$L(\boldsymbol{x}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = f(\boldsymbol{x}) + \boldsymbol{\gamma}^T \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\delta}^T \boldsymbol{g}(\boldsymbol{x})$$

Then there exist unique Lagrange multipliers vectors $\boldsymbol{\gamma}^*$ and $\boldsymbol{\delta}^*$ such that:

$$\begin{aligned} \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\gamma}^*, \boldsymbol{\delta}^*) &= 0 \\ \boldsymbol{\delta} &\geq 0 \\ \delta_j &= 0 \quad \textit{if the } j^{th} \textit{ constraint is inactive, i.e. } g_j(\boldsymbol{x}^*) < 0 \end{aligned}$$

Lemma A.3.3 *Consider the function $f : \mathbb{R}^p \to \mathbb{R}$:*

$$f(\boldsymbol{x}) = \prod_{j=1}^{p} x_j - \alpha \left(\sum_{j=1}^{p} x_j \right)^p$$

such that $\alpha < \frac{1}{p^p}$. Then $\boldsymbol{x}^* = \frac{c}{p} \left[\begin{array}{ccc} 1 & \ldots & 1 \end{array} \right]^T \in \mathbb{R}^p$ maximizes f under the constraints $\boldsymbol{x} > 0$ (where, as usual, the inequality applies to the components of the vector $\boldsymbol{x}$) and $h(\boldsymbol{x}) = \|\boldsymbol{x}\|_q - c$ for every $q \geq 1$, $c > 0$.

Proof: We start showing that the lemma holds in the case $q = 1$ ignoring the constraint $\boldsymbol{x} > 0$ (which will be imposed a posteriori at the end of the proof). To apply Theorem A.3.2 we need to convert this problem into a minimization problem. Thus we will consider $f' = -f$ and define the Lagrangian associated with this problem as:

$$L(\boldsymbol{x}, \delta) = f'(\boldsymbol{x}) + \delta\left[h((x)) - c\right]$$

Since $\|\boldsymbol{x}\|_1 = \sum_{j=1}^{p} x_j$, the first order necessary condition yields:

$$\frac{\partial L}{\partial x_i}(\boldsymbol{x}, \delta) = S - \frac{P}{x_i} + \delta = 0$$

where $S = \alpha p \left(\sum_{j=1}^{p} x_j\right)^{p-1}$ and $P = \prod_{j=1}^{p} x_j$. Therefore we obtain the following set of equations:

$$\begin{array}{rcl} (S+\delta)x_1 & = & P \\ \vdots & & \vdots \\ (S+\delta)x_p & = & P \end{array}$$

which imply that all the components of $\boldsymbol{x}^*$ are equal, i.e. $x_1^* = \ldots = x_p^* = x^*$.

Let's first consider the case where the constraint is inactive, i.e. $h(\boldsymbol{x}^*) < 0$ in which case $\delta^* = 0$. Then we obtain the equation $x^* = \frac{P}{S} = \frac{x^{*p}}{\alpha p^p x^{*p-1}} = \frac{x^*}{\alpha p^p}$ that is satisfied if and only if $x^* = 0$.

Now let's study the case where the constraint is active. This is equivalent to requiring that $x^* = \frac{c}{p}$ and consequently $S = \alpha p c^{p-1}$ and $P = \frac{c^p}{p^p}$. Hence we have $\delta = \frac{P}{x^*} - S = pc^{p-1}\left(\frac{1}{p^p} - \alpha\right)$. Since the hypothesis of the lemma state that $\frac{1}{p^p} - \alpha \ > 0$, we conclude that $\delta > 0$.

From the previous discussion it follows that the only point that satisfies the necessary conditions and whose components are strictly positive is $\boldsymbol{x}^* = \frac{c}{p}\left[\begin{array}{ccc} 1 & \ldots & 1 \end{array}\right]^T$. Therefore $f'(\boldsymbol{x}) \geq f'(\boldsymbol{x}^*)$ for any point $\boldsymbol{x}$ inside the volume V determined by the constraints $\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}^*\|_1 = c$ and $\boldsymbol{x} > 0$.

The fact that the lemma holds also when $q > 1$ follows directly from Lemma A.1.1. In fact any point $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_q = \|\boldsymbol{x}^*\|_q$ is contained in the volume V and tangent to $\|\boldsymbol{x}\|_1 = c$ at $\boldsymbol{x}^*$. Hence f' is still minimized at $\boldsymbol{x}^*$. ■

Appendix B

Condition Theory for Curve Landmarks Detection

"The same equations have the same solutions."

R. P. Feynman

We believe that the condition theory based framework that we used to study the corner detectors in Chapter 2 has many potential applications in the image analysis domain. In this appendix we will show how it can be used to detect curve landmarks.

B.1 The Model

We first define formally a *planar curve*.

Definition B.1.1 *Let I be an interval of real numbers; then a planar curve is a continuous mapping $\boldsymbol{\gamma} : I \rightarrow \mathbb{R}^2$. The curve $\boldsymbol{\gamma}$ is said to be simple if it is injective. If $I = [s_{in}, s_{fin}]$ is a closed bounded interval and $\boldsymbol{\gamma}(s_{in}) = \boldsymbol{\gamma}(s_{fin})$ then the curve is closed. Curves such that $\boldsymbol{\gamma}(s_1) \neq \boldsymbol{\gamma}(s_2)$ if $s_1 \neq s_2$ for all $s_1, s_2 \in I$ (except possibly the case where $s_1 = s_{in}$ and $s_2 = s_{fin}$) are called simple. A simple closed curve is also called a Jordan curve.*

Hereafter will study the sensitivity of Jordan curves following the same line of thought presented in Section 2.4.1. The expression for a curve perturbed by noise is given by:

$$\widetilde{\boldsymbol{\gamma}}(s) \overset{\text{def}}{=} \boldsymbol{\gamma}(s) + \boldsymbol{\eta} \tag{B.1}$$

We choose to model the effect of the noise by two transformations that operate on the curve neighborhood described by the image of the curve parameter interval

$\omega(s_0) \subseteq I$. The first one is parameterized by the p_T dimensional vector $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}} + \Delta\boldsymbol{\theta}$ and describes the geometric distortion of the curve:

$$\boldsymbol{\gamma}(s) \mapsto \boldsymbol{T}_{\overline{\boldsymbol{\theta}}+\Delta\boldsymbol{\theta},\boldsymbol{\gamma_0}}(\boldsymbol{\gamma}(s)) \tag{B.2}$$

where $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{\gamma_0}} : \boldsymbol{\gamma}(s) \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ and $\overline{\boldsymbol{\theta}}$ represents the identity in the parameter space (i.e. $\boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{\gamma_0}}(\boldsymbol{\gamma}(s)) = \boldsymbol{\gamma}(s)$) and $\boldsymbol{\gamma}(s_0) = \gamma_0$. The second transformation, parameterized by the p_S dimensional vector $\boldsymbol{\psi} = \overline{\boldsymbol{\psi}} + \Delta\boldsymbol{\psi}$, models the effect of the noise on the parameterizations of the curve:

$$\boldsymbol{\gamma}(s) \mapsto \boldsymbol{\gamma}(S_{\overline{\boldsymbol{\psi}}+\Delta\boldsymbol{\psi},s_0}(s)) \tag{B.3}$$

where $S_{\boldsymbol{\psi},s_0} : I \subseteq \mathbb{R} \to \mathbb{R}$ and, as before, we indicate with $\overline{\boldsymbol{\psi}}$ the identity in the parameter space (i.e. $S_{\overline{\boldsymbol{\psi}},s_0}(s) = s$). Hence the perturbed version of the curve can be written as:

$$\widetilde{\boldsymbol{\gamma}}(s) = \boldsymbol{T}_{\overline{\boldsymbol{\theta}}+\Delta\boldsymbol{\theta},\boldsymbol{\gamma_0}}(\boldsymbol{\gamma}(S_{\overline{\boldsymbol{\psi}}+\Delta\boldsymbol{\psi},s_0}(s))) \tag{B.4}$$

Following the definition 2.4.1, we measure the sensitivity of a curve neighborhood $\omega(s_0) \subseteq I$ using the notion of differential condition number introduced in (2.2.2).

Definition B.1.2 *The condition number associated with the curve neighborhood $\omega(s_0)$ with respect to the transformations $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{\gamma_0}}$ and $S_{\boldsymbol{\psi},s_0}$ is defined as:*

$$K_{\boldsymbol{T}_{\boldsymbol{\theta}},s_0}(\omega(s_0)) = \lim_{\delta\to 0} \sup_{\|\boldsymbol{\eta}\|\leq\delta} \frac{\|\Delta\boldsymbol{\zeta}\|}{\|\boldsymbol{\eta}\|} \tag{B.5}$$

where $\boldsymbol{\zeta} = \left[\begin{array}{cc} \boldsymbol{\theta}^T & \boldsymbol{\psi}^T \end{array}\right]^T$ and $\boldsymbol{\gamma}(\boldsymbol{s_0}) = \gamma_0$.

The following theorem is similar to Theorem 2.4.2 since it provides a computable expression to estimate the condition number.

Theorem B.1.3 *If $\boldsymbol{T}_{\boldsymbol{\theta},\boldsymbol{\gamma_0}}$ is affine, then the first order estimate of the condition number (B.5) is given by:*

$$\hat{K}_{\boldsymbol{T}_{\boldsymbol{\theta},\gamma_0},S_{\psi,s_0}}(\omega(s_0)) = \|A^\dagger\left(\omega(s_0)\right)\| \tag{B.6}$$

where † denotes the pseudo inverse of the matrix:

$$A\left(\omega(s_0)\right) \overset{\text{def}}{=} \left[\begin{array}{c} A(\boldsymbol{y}_1) \\ \vdots \\ A(\boldsymbol{y}_N) \end{array}\right] \in \mathbb{R}^{2N\times p_T+p_S} \tag{B.7}$$

which is formed by the N sub-matrices:

$$A(s_i) \overset{\text{def}}{=} w(s_i - s_0) \left[\; J_{\boldsymbol{\theta}} \boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{\gamma}_0}(\boldsymbol{\gamma}(S_{\overline{\boldsymbol{\psi}},s_0}(s_i))) \quad J_s\boldsymbol{\gamma}(S_{\overline{\boldsymbol{\psi}},s_0}(s))\nabla_{\boldsymbol{\psi}} S_{\overline{\boldsymbol{\psi}},s_0}(s_i) \; \right] \tag{B.8}$$

obtained from a set of N points that sample the neighborhood $\omega(s_0)$. The scalar function $w(s_i - s_0)$ denotes the weight associated with the point s_i.

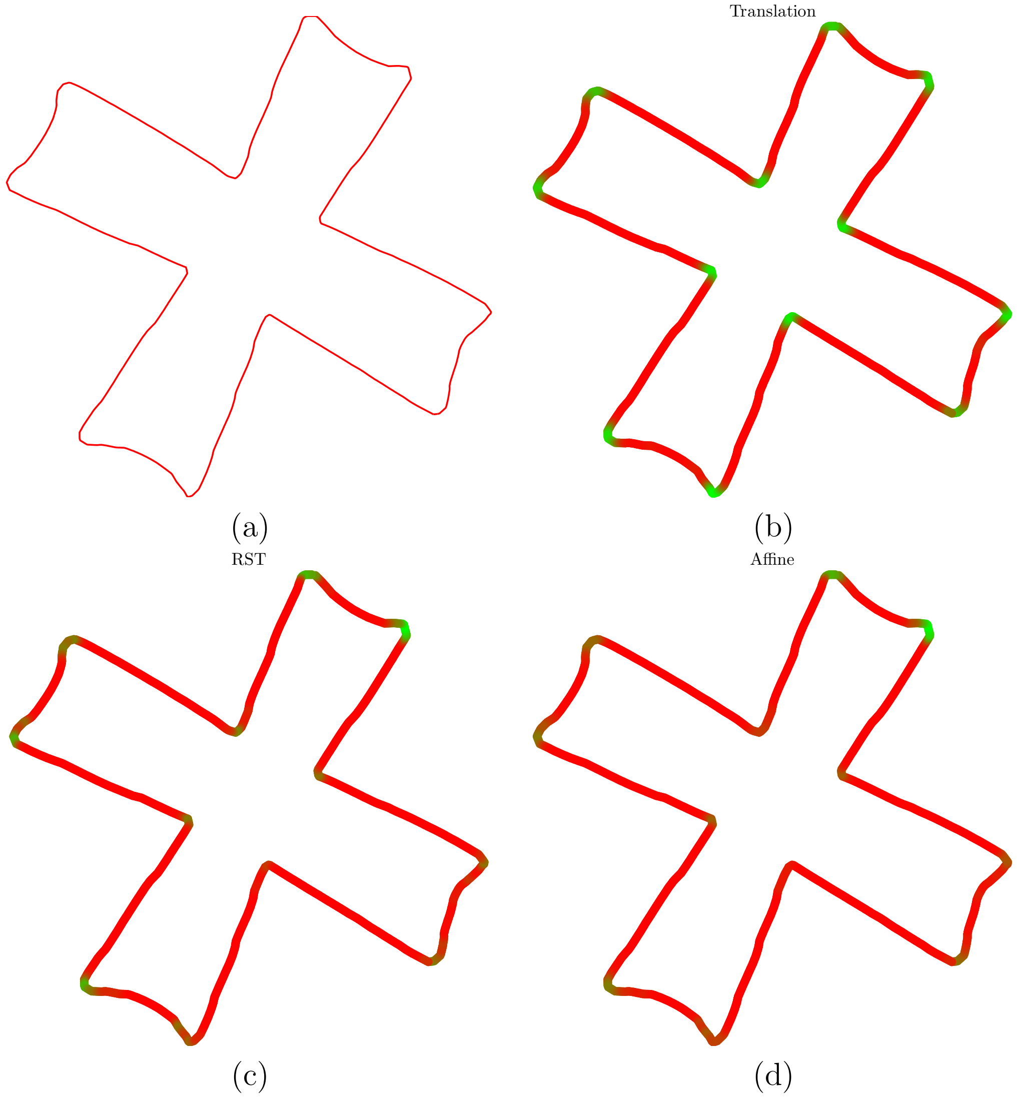
Proof: The proof proceeds along the same lines as the proof of Theorem . Also in this case we use the Taylor expansion to simplify the expression for the curve perturbation (B.4). The result expressed by equation (B.8) is a matter of straightforward differentiation, after noticing that $J_{\boldsymbol{\theta}} \boldsymbol{T}_{\overline{\boldsymbol{\theta}},\boldsymbol{\gamma}_0}(\boldsymbol{\gamma}(S_{\overline{\boldsymbol{\psi}},s_0})) = I_2$. ■
This theorem allows us to extend naturally the definition of Spectral Generalized Corner Detector Functions (SGCDF) introduced in Section 2.5. Note that in this context a *corner* is actually a *landmark* on the curve.

B.2 An Example

Consider the Jordan curve represented in Figure B.1(a). Assume that the curve is parameterized by arc length and that it is represented by $N_P = 10^3$ equally spaced samples $\{\boldsymbol{\gamma}(s_1), \ldots, \boldsymbol{\gamma}(s_{N_P})\}$. Moreover suppose that the neighborhood $\omega(s_0)$ is composed of the points $\{s_0 - r\Delta s, \ldots, s_0 + r\Delta_s\}$, where $\Delta s = s_{i+1} - s_i$. In our numerical example we choose $r = 16$ (which means that we are focusing on curve neighborhoods whose length amounts to 3.3% of the total curve length). In our experiments we considered three types of geometric transformation that, mutatis mutandis, are defined in (2.45), (2.46) and (2.47). We choose to model the effect of noise on the curve parameterizations by a simple affine function: $S_{\boldsymbol{\psi},s_0}(s) = s_0 + \psi_1(s - s_0) + \psi_0$. The curve landmarks are identified calculating the response of the generalized Noble-Förstner (harmonic mean) detector:

$$f_{NF}(\omega(s_0)) \overset{\text{def}}{=} \begin{cases} \frac{1}{\sum_{i=1}^{p_T+p_S} \frac{1}{\sigma_i(A)^2}} & \text{if } \sigma_i(A) \neq 0 \text{ for every } i, \\ 0 & \text{otherwise.} \end{cases} \tag{B.9}$$

Figures B.1(b), B.1(c) and B.1(d) show the response of the detector (B.9) corresponding to a translational, Rotation Scaling and Translation (RST) and affine geometric transformation. Green colors indicate large responses whereas red colors indicate small responses. As one may expected from the discussion in Section 2.4, the detector response is larger where the structure of the curve is less affected by the geometric and parametric distortions discussed above. This is particularly evident for neighborhoods containing the corners formed at the



(a) (b) (c) (d)

Figure B.1: Figures (b), (c) and (d) show the response of the detector (B.9) corresponding to a translational, RST and affine geometric transformation for the curve in (a). Green colors indicate large detector responses whereas red colors indicate small detector responses.

center of the cross: if the effect of the noise is modeled via translations, then the structure of the curve corresponding to $\omega(s_0)$ remains distinguishable. However, when the noise is allowed to affect the scale of the curve (RST transformation) or the local axis skew (affine) the value of the detector becomes much smaller. In a curve matching context the neighborhoods that maintain a distinctive structure in presence of local translations become non distinguishable in presence of local scalings or affine transformations.

Appendix C

Some Analytical Properties of the Helmholtz Equation

"I'm pickin' up good vibrations."

Beach Boys

Consider[1] the boundary problem with Dirichlet conditions (4.2) rewritten here for sake of convenience:

$$\begin{aligned} -\triangle u(\boldsymbol{x}) &= \lambda \frac{1}{v(\boldsymbol{x})^2} u(\boldsymbol{x}) && \text{for } \boldsymbol{x} \in \Omega \\ u(\boldsymbol{x}) &= 0 && \text{for } \boldsymbol{x} \in \Gamma \end{aligned}$$

One trivial solution for this problem is $u(\boldsymbol{x}) \equiv 0$ everywhere. We are interested in the nontrivial solutions of this problem, which can be proven to exist for a wide class of regions Ω for discrete values of the parameter λ and for $v > 0$. Such solution are in the form of a countable set of pairs of eigenvalues/eigenfunction, i.e. (λ_k, u_k). The next lemmas will show that such eigenpairs have the following properties:

- the eigenfunctions are orthogonal
- the eigenvalues are real and the eigenfunctions can be chosen to be real.
- the eigenvalues are positive

[1] For a more thorough treatment of the results presented in this appendix we refer the reader to the texts by Weinberger [132], Carrier et al. [19] and Evans et al. [33].

Lemma C.0.1 *Let (λ_k, u_k) and (λ_l, u_l) be two eigenpairs that solve equation (4.2). Then:*

$$\langle u_k | u_l \rangle_{\frac{1}{v^2}} \propto \delta_{kl}$$

where δ_{kl} is the Kronecker delta and the notation $\langle \cdot | \cdot \rangle$ denotes the weighted inner product on Ω: $\langle f | g \rangle_w = \int_\Omega w(\boldsymbol{x}) f(\boldsymbol{x}) g(\boldsymbol{x})\ d\boldsymbol{x}$.

Proof: Consider the pair of equations:

$$-\triangle u_k = \lambda_k \frac{1}{v^2} u_k \tag{C.2a}$$

$$-\triangle u_l = \lambda_l \frac{1}{v^2} u_l \tag{C.2b}$$

and multiply both sides of (C.2a) by u_l and of (C.2b) by u_k. If we subtract both members and we integrate over Ω we obtain:

$$\int_\Omega \left(u_l \triangle u_k - u_k \triangle u_l \right)\ d\boldsymbol{x} = (\lambda_l - \lambda_k) \int_\Omega \frac{1}{v^2} u_k u_l\ d\boldsymbol{x} \tag{C.3}$$

The left hand side can be rewritten using Green's second identitity as:

$$\int_\Omega \left(u_l \triangle u_k - u_k \triangle u_l \right)\ d\boldsymbol{x} = \int_\Gamma \left(u_l \frac{\partial u_k}{\partial \boldsymbol{n}} - u_k \frac{\partial u_l}{\partial \boldsymbol{n}} \right)\ d\boldsymbol{x}$$

where $\boldsymbol{n}$ denotes the normal at the boundary. Since u_k and u_l are identically equal to zero on the boundary, the left hand side of (C.3) must vanish. Consequently, since $\lambda_l \neq \lambda_k$, the proof is concluded observing that the right hand side of (C.3) yields:

$$\int_\Omega \frac{1}{v^2} u_k u_l\ d\boldsymbol{x} = \langle u_k | u_l \rangle_{\frac{1}{v^2}} = 0$$

■

Lemma C.0.2 *The eigenvalues that satisfy the equation (4.2) are real and the corresponding eigenfunctions can be chosen to be real.*

Proof: We will proof that the eigenvalues are real by contradiction. Let $\lambda \in \mathbb{C}$ and let u be the corresponding eigenfunction (not identically equal to zero) that solves (4.2). It is straightforward to verify that the complex conjugates of the eigenpairs will also satisfy (4.2). Hence, letting $(\lambda_k, u_k) = (\lambda, u)$ and $(\lambda_l, u_l) = (\lambda^*, u^*)$ and following the same steps of the proof of Lemma C.0.1 we conclude that:

$$\int_\Omega \frac{1}{v^2} u u^*\ d\boldsymbol{x} = \int_\Omega \frac{1}{v^2} |u|^2\ d\boldsymbol{x} = 0 \tag{C.4}$$

It follows immediately that (C.4) is satisfied only if $u \equiv 0$, which contradicts the hypothesis. Hence $\lambda \in \mathbb{R}$. Now suppose there exists a complex eigenfunction corresponding to λ: $u = v + jw$. Clearly both u and w satisfy (4.2), hence we can always choose a real eigenfunction. ■

Lemma C.0.3 *The eigenvalues that satisfy the equation* (4.2) *are positive.*

Proof: If we multiply both members of equation $\triangle u_k = -\lambda_k \frac{1}{v^2} u_k$ by u_k and we integrate over the region Ω we obtain:

$$\int_\Omega u_k \triangle u_k \; d\boldsymbol{x} = -\lambda_k \int_\Omega \frac{1}{v^2} u_k^2 \; d\boldsymbol{x}$$

Applying Green's first identity to the left hand side we obtain:

$$\int_\Omega u_k \triangle u_k \; d\boldsymbol{x} = \int_\Gamma u_k \frac{\partial u_k}{\partial \boldsymbol{n}} \; d\boldsymbol{x} - \int_\Omega \|\nabla u_k\|^2 \; d\boldsymbol{x}$$

Since u_k is identically equal to zero on the boundary we can write:

$$\lambda_k \int_\Omega \frac{1}{v^2} u_k^2 \; d\boldsymbol{x} = \int_\Omega \|\nabla u_k\|^2 \; d\boldsymbol{x}$$

The proof is concluded observing that λ_k can be expressed in terms of the ratio of two positive quantities. ■

We conclude this appendix listing a few other important facts.

- The eigenvalues can be sorted in order of increasing value: $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$ with $\lambda_k \to \infty$ as $k \to \infty$

- For a given eigenvalue λ_k there is a finite number of linearly independent eigenfunctions (such number is called the multiplicity of λ_k).

- The first (or principal eigenvalue) has multiplicity 1 and does not change sign over Ω.

- The normalized real eigenfunctions u_k form an orthonormal basis of $L^2(\Omega)$, where the normalization is such that $\int_\Omega \frac{1}{v^2} u_k \; d\boldsymbol{x} = 1$

- If the region Ω is not bounded it may happen that the set of eigenpairs is no longer discrete.

List of Acronyms

Condition Number Based (CNB)

Said of a spectral generalized corner detector function that is related to the condition number of a point neighborhood.

Condition Number Detector (CND)

An algorithm to detect the characteristic scale of a point neighborhood based on the signature of the condition number.

Consensus Set (CS)

The set of data that fit a certain model within a given tolerance.

Generalized Corner Detector Function (GCDF)

A scalar function that returns the quality of the neighborhood of a generalized image for matching/tracking purposes.

Helmholtz Descriptor (HD)

A curve/region descriptor based on the modes of vibration of an elastic membrane.

Generalized Gradient Matrix (GGM)

The matrix that describes the local properties of a point of a generalized image

Laplacian Detector (LD)

An algorithm to detect the characteristic scale of a point neighborhood based on the signature of the image Laplacian.

Minimal Sample Set (MSS)

The smallest set of data necessary to estimate the parameters of a certain model (e.g. the MSS to estimate the parameters of a line has cardinality 2).

Rotation Scaling and Translation (RST)

A geometric transformation that describes a rotation, a scaling and a translation.

Spectral Generalized Corner Detector Function (SGCDF)

A generalized corner detector function that depends solely on the spectral properties of the generalized gradient matrix.

Zernike Moment Descriptor (ZMD)

A curve/region descriptor expanding the domain indicator function on the Zernike moments.

Bibliography

[1] K. Åström. *Invariancy Methods for Points, Curves and Surfaces in Computational Vision.* PhD thesis, Department of mathematics, Lund University, Sweden, 1996.

[2] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, March 2004.

[3] J.L. Barron, D.J. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[4] A. Bartoli. Groupwise geometric and photometric direct image registration. In *British Machine Vision Conference*, Edinburgh, UK, September 2006.

[5] A. Baumberg. Reliable feature matching across widely separated views. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 774–781, Hilton Head Island, South Carolina, June 2000.

[6] S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM Computing Surveys*, 27(3):433–466, 1995.

[7] J. Beis and D. Lowe. Shape indexing using approximate nearest-neighbor search in highdimensional spaces. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1000–1006, 1997.

[8] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.

[9] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, second edition, 1999.

[10] I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo. *Handbook of Combinatorial Optimization*, volume A, chapter The maximum clique problem, pages 1–74. Kluwer Academic Publishers, 1999.

[11] M.A. Branch, T.F. Coleman, and Y. Li. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing*, 21(1):1–23, 1999.

[12] L. Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992.

[13] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 2006. Accepted for publication.

[14] P. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4):217–236, October 1983.

[15] P. Buser, J. Conway, P. Doyle, and K. Semmler. Some planar isospectral domains. *International Mathematics Research Notices*, 9:391 – 400, 1994.

[16] S. Busygin. A new trust region technique for the maximum weight clique problem. *Discrete Applied Mathematics*, 154(15):2080–2096, 2006.

[17] J. Nieuwenhuijse New House Internet Services BV. PTgui. http://www.ptgui.com/. Last visited: September 19, 2006.

[18] R.H. Byrd, R.B. Schnabel, and G.A. Shultz. Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Mathematical Programming*, 40:247–263, 1988.

[19] G. F. Carrier and C. E. Pearson. *Partial Differential Equations, Theory and Technique*. Academic Press Inc., second edition, 1988.

[20] O. Chum and J. Matas. Randomized RANSAC with $T_{d,d}$ test. In *13th British Machine Vision Conference*, September 2002.

[21] O. Chum and J. Matas. Matching with PROSAC - progressive sample consensus. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, volume 1, pages 220–226, San Diego, June 2005.

[22] Cognitech. Video investigator. http://www.linear-systems.com/products/info/cogvi.htm. Last visited: September 22, 2006.

[23] L. D. Cohen and R. Kimmel. Global minimum for active contour models: A minimal path approach. *International Journal of Computer Vision*, 24(1):57–78, August 1997.

[24] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603 – 619, May 2002.

[25] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press and McGraw-Hill, second edition, 2002.

[26] J. Davis. Mosaics of scenes with moving objects. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 354–360, 1998.

[27] J. Delon, A. Desolneux, A. B. Petro, and J.L. Lisani. A non parametric theory for histogram segmentation. Technical Report 05, Centre de Mathématiques et de Leurs Applications, 2005.

[28] H. Dersch. Panorama tools. http://panotools.sourceforge.net/. Last visited: September 19, 2006.

[29] T. Deschamps. *Curve and Shape Extraction with Minimal Path and Level-Sets techniques. Applications to 3D Medical Imaging.* PhD thesis, Université de Paris-Dauphine, December 2001.

[30] T. A. Driscoll. Eigenmodes of isospectral drums. *SIAM Review*, 39(1):1–17, March 1997.

[31] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley Interscience, second edition, 2000.

[32] H. Eidenberger. Statistical analysis of MPEG-7 image descriptions. *ACM Multimedia Systems Journal*, 10(2):84–97, 2004.

[33] L. C. Evans. *Partial Differential Equations (Graduate Studies in Mathematics, 19)*. American Mathematical Society, 1998.

[34] D. Fedorov. imREG/regeemy. http://nayana.ece.ucsb.edu/registration/. Last visited: September 22, 2006.

[35] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.

[36] L. M. J. Florack, B. M. terHaar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, July/August 1992.

[37] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167–174, 1993.

[38] W. Förstner. A feature based correspondence algorithm for image matching. In *International Archives of Photogrammetry and Remote Sensing*, volume 26, pages 150–166, 1986.

[39] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and center of circular features. In *Proc. of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland*, pages 281–305, June 2-4 1987.

[40] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach.* Prentice Hall, August 2002.

[41] L. Fox, P. Henrici, and C. Moler. Approximations and bounds for eigenvalues of elliptic operators. *SIAM Journal of Numerical Analisys*, 4:89–102, 1967.

[42] S. Frantz, K. Rohr, and H.S. Stiehl. Multi-step differential approaches for the localization of 3D point landmarks in medical images. *Journal of Computing and Information Technology*, 6(4):435–447, 1998.

[43] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The Amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.

[44] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The John Hopkins University Press, 1996.

[45] C. Gordon, D. L. Webb, and S. Wolpert. One cannot hear the shape of a drum. *Bulletin of the American Mathematical Society*, 27(1):134–138, July 1992.

[46] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt. Shape representation and classification using the Poisson equation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 61–67, July 2004.

[47] A. Goshtasby, L. Staib, C. Studholme, and D. Terzopoulos, editors. *Computer Vision and Image Understanding: Special Issue on Nonrigid Image Registration*, volume 89. Elsevier Science, February–March 2003.

[48] P. Guidotti and J. V. Lambers. Eigenvalue characterization and computation for the laplacian on general domains. Submitted, October 2005.

[49] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, 1952.

[50] C. Harris and M. Stephens. A combined corner and edge detector. In M. M. Matthews, editor, *Proc. of the 4th ALVEY vision conference*, pages 147–151, University of Manchester, England, Septemeber 1988.

[51] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, second edition, 2003.

[52] K. Höllig, C. Apprich, and A. Streit. Introduction to the web-method and its applications. *Advances in Computational Mathematics*, 23:215–237, 2005.

[53] R. Horn and C. R. Johnson. *Matrix Analisys.* Cambridge University Press, 1999.

[54] C. T. Hsu and J. L. Wu. Multiresolution mosaic. *IEEE Transactions on Consumer Electronics*, 42:981–990, August 1996.

[55] P. J. Huber. *Robust Statistics.* Wiley, 1981.

[56] A. J. Izenman. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, March 1991.

[57] H. Jin, P. Favaro, and S. Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *Proc. of IEEE International Conference on Computer Vision*, volume 1, pages 684–689, Vancouver, BC, July 2001.

[58] M. Kac. Can one hear the shape of a drum? *American Mathematical Monthly*, 73(2):1–23, 1966.

[59] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, November 2001.

[60] Y. Kanazawa and H. Kawakami. Detection of planar regions with uncalibrated stereo using distribution of feature points. In *British Machine Vision Conference*, volume 1, pages 247–256, Kingston upon Thames, London, September 2004.

[61] T. Kato. *Perturbation Theory for Linear Operators*. Springer, February 1995.

[62] J. K. Kearney, W. B. Thompson, and D. L. Boley. Optical flow estimation: An error analysis of gradient based methods with local optimazation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):229–244, March 1990.

[63] C. Kenney, B. Manjunath, M. Zuliani, G. Hewer, and A. Van Nevel. A condition number for point matching with application to registration and post-registration error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1437–1454, November 2003.

[64] C. Kenney, M. Zuliani, and B.S. Manjunath. An axiomatic approach to corner detection. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 191–197, San Diego, California, June 2005.

[65] C. S. Kenney, A. J. Laub, and M. S. Reese. Statistical condition estimation for linear least squares. *SIAM Journal on Matrix Analysis and Applications*, 19(4):906–923, 1998.

[66] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM, 1998.

[67] A. Levin, A. Zomet, S. Peleg, and Y. Weiss. Seamless image stitching in the gradient domain. In *Proc. of the European Conference on Computer Vision*, Prague, May 2004.

[68] H. Li, B. S. Manjunath, and S. K. Mitra. A contour-based approach to multisensor image registration. *IEEE Transactions on Image Processing*, 4(3):320–334, Mar 1995.

[69] H. Li, B.S. Manjunath, and S.K. Mitra. Image fusion using wavelets. *Computer Vision, Graphics & Image Processing: Graphical Models and Image Processing*, 57(3):235–245, May 1995.

[70] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer Academic, Dordrecht, Netherlands, 1994.

[71] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[72] J.L. Lisani. *Shape Based Automatic Images Comparison.* PhD thesis, Univeristé Paris IX-Dauphine, 2001.

[73] J.L. Lisani, L. Moisan, P. Monasse, and J.M. Morel. On the theory of planar shape. *SIAM Journal on Multiscale Modeling and Simulation*, 1(1):1–24, 2003.

[74] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[75] L. Lucchese and S. K. Mitra. Using saddle points for subpixel feature detection in camera calibration targets. In *Proc. of the 2002 Asia Pacific Conference on Circuits and Systems*, Singapore, December 2002.

[76] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, second edition, 2003.

[77] Q.-T. Luong, P. Fua, and Y. G. Leclerc. The radiometry of multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):19 – 33, January 2002.

[78] Y. Ma, S. Soatto, J. Košecká, and S. S. Sastry. *An Invitation to 3-D Vision. From images to Geometric Models.* Springer, 2004.

[79] J. Mendel. *Lessons in Estimation Theory for Signal Processing, Communication and Control.* Prentice-Hall, Englewood-Cliffs, 1995.

[80] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra.* SIAM, 2001.

[81] S. Michael. kdtree. `http://www.mathworks.co.kr/matlabcentral/fileexchange/loadFile.do?objectId=7030&objectType=file`. Last visited: September 8, 2006.

[82] Microsoft. Microsoft digital image pro. `http://www.microsoft.com/products/imaging/default.mspx`. Last visited: September 11, 2006.

[83] Microsoft. Microsoft photo tourism. `http://research.microsoft.com/IVM/PhotoTours`. Last visited: September 11, 2006.

[84] Microsoft. Microsoft photosynth. http://labs.live.com/photosynth. Last visited: September 11, 2006.

[85] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Proc. of IEEE 8th International Conference on Computer Vision*, pages 525–531, Vancouver, Canada, 2001.

[86] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *European Conference on Computer Vision*, pages 128–142, Copenhagen, Denmark, 2002. Springer.

[87] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[88] K. Mikolajczyk and C Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.

[89] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005.

[90] F. Mokhtarian and M. Bober. *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*. Kluwer Academic Publishers, 2003.

[91] A. Moore. A tutorial on *k*d-trees. Extract from PhD Thesis 209, University of Cambridge, Computer Laboratory, 1991.

[92] H. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical Report CMU-RI-TR-3, Carnegie-Mellon University, Robotics Institute, September 1980.

[93] B. S. Morse. *Computation of Object Cores from Grey-level Images*. PhD thesis, University of North Carolina at Chapel Hill, 1994.

[94] R. Szeliski N. Snavely, S. M. Seitz. Photo tourism: Exploring photo collections in 3d. In *ACM Transactions on Graphics (SIGGRAPH Proceedings)*, volume 25, pages 835–846, 2006.

[95] A. Y. Ng, M. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, number 14, 2002.

[96] K. V. Ngo. An approach of eigenvalue perturbation theory. *Applied Numerical Analysis & Computational Mathematics*, 2(1):108 – 125, April 2005.

[97] D. Nistér. Preemptive RANSAC for live structure and motion estimation. In *IEEE International Conference on Computer Vision*, pages 199–206, Nice, France, October 2003.

[98] A. Noble. *Descriptions of Image Surfaces*. PhD thesis, Department of Engineering Science, Oxford University, 1989.

[99] A. Noll. Domain perturbations, capacity and shift of eigenvalues. *Journées équations aux dérivées partielles*, pages 1–10, 1999.

[100] M. Pavan and M. Pelillo. Dominant sets and hierarchical clustering. In *Proc. IEEE International Conference on Computer Vision*, volume 1, pages 362 – 369, Nice, France, 2003.

[101] M. Pavan and M. Pelillo. A new graph-theoretic approach to clustering and segmentation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 145–152, June 2003.

[102] P. Perona and W. Freeman. A factorization approach to grouping. In *Proc. of 5th European Conference of Computer Vision*, pages 655–670, Freiburg, Germany, 1998.

[103] G. Peyré. Toolbox fast marching. http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=6110&objectType=FILE. Last visited: March 23, 2006.

[104] Realviz. Realviz: The stitcher range. http://stitcher.realviz.com. Last visited: September 11, 2006.

[105] J.R. Rice. A theory of condition. *SIAM Journal on Numerical Analysis*, 3:287–310, 1966.

[106] K. Rohr. Modelling and identification of characteristic intensity variations. *Image and Vision Computing*, 10(2):66–76, 1992.

[107] K. Rohr. Localization properties of direct corner detectors. *Journal of Mathematical Imaging and Vision*, 4(2):139–150, 1994.

[108] K. Rohr. Extraction of 3D anatomical point landmarks based on invariance principles. *Pattern Recognition*, 32:3–15, 1999.

[109] P. J. Rousseeuw and C. Croux. Alternatives to the median absolute derivation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993.

[110] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.

[111] N. Saito. Geometric harmonics as a statistical image processing tool for images defined on irregularly-shaped domains. In *Proceedings of IEEE Statistical Signal Processing Workshop*, Boreadux, France, July 2005.

[112] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Proc. of IEEE 6th International Conference on Computer Vision*, pages 230–235, Bombay, India, January 1998.

[113] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[114] S. Sclaroff and A.P. Pentland. Modal matching for correspondence and recognition. *IEEE Transactions on Pattern Analisys and Machine Intelligence*, 17(6):545–561, June 1995.

[115] J. A. Sethian. Fast marching methods. *SIAM Review*, 41(2):199–235, 1999.

[116] J. Shi and C. Tomasi. Good features to track. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, Washington, June 1994.

[117] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. *International Journal of Computer Vision*, 35(1):13–32, November 1999.

[118] D. Sinclair and A. Blake. Isoperimetric normalization of planar curves. *IEEE Transactions on Pattern Analisys and Machine Intelligence*, 16(8):769–777, August 1994.

[119] C. V. Stewart. MINPRAN: A new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):925–938, October 1995.

[120] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, September 1999.

[121] M.-S. Su, W.-L. Hwang, and K.-Y. Cheng. Analysis on multiresolution mosaic images. *IEEE Transactions on Image Processing*, 13(7):952–959, July 2004.

[122] R. Szeliski. Image alignment and stitching: A tutorial. Technical Report MSR-TR-2004-92, Microsoft Research, December 2004.

[123] M.R. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70(8):920–930, 1979.

[124] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography—a factorization method. *International Journal on Computer Vision*, 9(2):137–154, November 1992.

[125] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto. Making good features track better. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 178–183, Santa Barbara, California, June 1998.

[126] B. J. Tordoff and D. W. Murray. Guided-MLESAC: Faster image transform estimation by using matching priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1523–1535, October 2005.

[127] P. H. S. Torr. A structure and motion toolkit in matlab - interactive adventures in S and M. Technical Report MSR-TR-2002-56, Microsoft Research, June 2002.

[128] P.H.S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Journal of Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[129] B. Triggs. Detecting keypoints with stable position, orientation, and scale under illumination changes. In *Proc. of the 8th European Conference on Computer Vision*, volume 4, pages 100–113, 2004.

[130] E. Vincent and R. Laganière. Detecting planar homographies in an image pair. In *2nd International Symposium on Image and Signal Processing and Analysis*, pages 182–187, Pula, Croatia, June 2001.

[131] H. Wang and D. Suter. Robust adaptive-scale parametric model estimation for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1459–1474, 2004.

[132] H. F. Weinberger. *A Firts Course in Partial Differential Equations with Complex Variables and Transform Methods*. Blaidsell Publishing Company, first edition, 1965.

[133] D. Zhang and G. Lu. Evaluation of MPEG-7 shape descriptors against other shape descriptors. *ACM Journal of Multimedia Systems*, 9(1):15 – 30, July 2003.

[134] D. S. Zhang and G. Lu. Generic Fourier descriptors for shape-based image retrieval. In *Proceedings of IEEE International Conference on Multimedia and Expo*, volume 1, pages 425–428, Lausanne, Switzerland, August 2002.

[135] D. S. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.

[136] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing Journal*, 25(1):59–76, 1997.

[137] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.

[138] A. Zomet, A. Levin, S. Peleg, and Y. Weiss. Seamless image stitching by minimizing false edges. *IEEE Transactions on Image Processing*, 15(4):969 – 977, April 2006.

[139] M. Zuliani, S. Bhagavathy, C. S. Kenney, and B. S. Manjunath. Affine-invariant curve matching. In *IEEE International Conference on Image Processing*, October 2004.

[140] M. Zuliani, C. Kenney, and B.S. Manjunath. A mathematical comparison of point detectors. In *Proc. of the 2nd IEEE Workshop on Image and Video Registration*, 2004.

[141] M. Zuliani, C. S. Kenney, S. Bhagavathy, and B. S. Manjunath. Drums and curve descriptors. In *British Machine Vision Conference*, September 2004.

[142] M. Zuliani, C. S. Kenney, and B. S. Manjunath. The MultiRANSAC algorithm and its application to detect planar homographies. In *IEEE International Conference on Image Processing*, September 2005.

Index