

ALMA MATER STUDIORUM
UNIVERSITY OF BOLOGNA

ARCES - ADVANCED RESEARCH CENTER ON ELECTRONIC SYSTEMS FOR
INFORMATION AND COMMUNICATION TECHNOLOGIES "E. DE CASTRO"

**Algorithms for On-Line Image
Registration from Multiple Views**

Ludovico Carozza

Tutor:
Prof. Alessandro BEVILACQUA

Coordinator:
Prof. Claudio FIEGNA

Ph.D. Thesis
January, 2008 - December, 2010

PH.D. PROGRAM IN INFORMATION TECHNOLOGY

CYCLE XXIII-ING-INF/05

Abstract

Images of a scene, static or dynamic, are generally acquired at different epochs from different viewpoints. They potentially gather information about the whole scene and its relative motion with respect to the acquisition device. Data from different (in the spatial or temporal domain) visual sources can be fused together to provide a unique consistent representation of the whole scene, even recovering the third dimension, permitting a more complete understanding of the scene content. Moreover, the pose of the acquisition device can be achieved by estimating the relative motion parameters linking different views, thus providing localization information for automatic guidance purposes. Image registration is based on the use of pattern recognition techniques to match among corresponding parts of different views of the acquired scene. Depending on hypotheses or prior information about the sensor model, the motion model and/or the scene model, this information can be used to estimate global or local geometrical mapping functions between different images or different parts of them. These mapping functions contain relative motion parameters between the scene and the sensor(s) and can be used to integrate accordingly informations coming from the different sources to build a wider or even augmented representation of the scene.

Accordingly, for their scene reconstruction and pose estimation capabilities, nowadays image registration techniques from multiple views are increasingly stirring up the interest of the scientific and industrial community. Depending on the applicative domain, accuracy, robustness, and computational payload of the algorithms represent important issues to be addressed, and generally a trade-off among them has to be reached. Moreover, on-line performance is desirable in order to guarantee the direct interaction of the vision device with human actors or control systems. This thesis follows a general research approach to cope with these issues, almost independently

from the scene content, under the constraint of rigid motions. This approach has been motivated by the portability to very different domains as a very desirable property to achieve. A general image registration approach suitable for on-line applications has been devised and assessed through two challenging case studies in different applicative domains.

The first case study regards scene reconstruction through on-line mosaicing of optical microscopy cell images acquired with non automated equipment, while moving manually the microscope holder. By registering the images the field of view of the microscope can be widened, preserving the resolution while reconstructing the whole cell culture and permitting the microscopist to interactively explore the cell culture. In the second case study, the registration of terrestrial satellite images acquired by a camera integral with the satellite is utilized to estimate its three-dimensional orientation from visual data, for automatic guidance purposes. Critical aspects of these applications are emphasized and the choices adopted are motivated accordingly. Results are discussed in view of promising future developments.

To my family

Acknowledgements

I would like to acknowledge my supervisor Prof. Alessandro Bevilacqua and my colleagues at ARCES, Dr. Alessandro Gherardi and Ing. Filippo Piccinini, who have stimulated my work through useful discussions and observations, suggestions and active collaboration.

Thanks also to the staff of the Bone Regeneration Laboratory, Istituti Ortopedici Rizzoli, Bologna, led by Dr. Enrico Lucarelli, for their partnership in the ADVANCE (Automatic system to Determine Vital AdhereNt CELls) project.

I would like to thank the members of Aerospace Group led by Prof. Paolo Tortora at DIEM, II Faculty of Engineering, University of Bologna, for their partnership in the STARS (Standalone Three-Axis spacecraft oRientation Sensor) project, funded by the University of Bologna as a Strategic Project.

Contents

1	Introduction	1
1.1	Motivations and Goals	2
1.2	Thesis Outline and Contributions	4
2	Image Registration from Multiple Views	9
2.1	Methodological approaches: a survey	9
2.2	General working hypotheses	12
2.3	The registration algorithm	13
2.3.1	Feature detection	14
2.3.1.1	The Shi-Tomasi corner detector	14
2.3.2	Feature matching	15
2.3.2.1	Phase Correlation	15
2.3.2.2	The LKT tracker	16
2.4	Two-view Geometry: model estimation	17
2.4.1	Planar Registration	18
2.4.2	Robust estimation	20
3	On-line Mosaicing of Optical Microscope Imagery	21
3.1	Introduction	22
3.2	Previous Work	25
3.3	The method	26
3.3.1	Flat Field Correction	27
3.3.2	Geometric Registration	29
3.3.2.1	Model Estimation	31
3.3.2.2	Registration Modes	34

CONTENTS

3.4	Experimental Results	36
3.4.1	Metrics for Quality Assessment	37
3.4.2	Results	41
3.4.2.1	Model Selection: Discussion	49
3.4.3	Computational Performance	51
3.5	Conclusions	52
4	Camera Attitude estimation: the STARS project	55
4.1	Introduction	55
4.2	Previous Works	57
4.2.1	Working Hypotheses	58
4.3	A novel vision-based approach	59
4.3.1	Physical and geometrical model	59
4.3.2	Attitude parametrization	62
4.4	Image Registration	63
4.5	Error Analysis	65
4.5.1	Modelling Errors	65
4.5.1.1	Parallax Error	66
4.5.2	Image Registration Error	67
4.5.3	Error dynamic	69
4.6	Simulation Framework	71
4.6.1	Image Acquisition Model	71
4.6.2	Dataset	73
4.7	Experimental results	73
4.7.1	Synthetic imagery	74
4.7.1.1	Theoretical accuracy performance	75
4.7.1.2	Frame-to-frame registration: Results	79
4.7.1.3	Frame-to-frame registration: Discussion	81
4.7.1.4	Dynamic conditions	88
4.7.2	Laboratory Experiments	88
4.8	Conclusions	93

5 Final Conclusions	95
5.1 Automatic On-line Mosaicing of Optical Microscopy Imagery: the AD-VANCE project	96
5.2 Camera pose estimation: the STARS project	97
Bibliography	99
List of Figures	103
List of Tables	107

CONTENTS

Chapter 1

Introduction

Nowadays, vision sensors and systems have gone through a large diffusion in a wide range of intelligent applications, acquiring an increasing importance for the advantages that vision technology can bring. Their relative cheapness and their direct relation with the most meaningful sense of the human perception, the sight, can certainly be considered as the two main aspects that have accelerated their widespread employment in different applicative fields in the last two decades. Moreover, their compactness permits to gather and process a large amount of visual data without requiring a cumbersome set-up and the information retrieved can be used to interact with human actors or other devices, fixed or mobile, to perform intelligent tasks. In general, the exploration of a scene by an agent endowed with a camera can provide a higher level representation of the scene and contribute to its understanding. The images of the scene, taken at different times and from different viewpoints, usually refer to part of the scenes, acquired for a precise moment and in the current field of view (FOV) of the sensor, and this can represent a limit, especially for complex scenes, towards the analysis and the comprehension of the environment explored. In other words, if these images are treated separately, only the visual properties of the objects for a precise *local* (in the spatial and temporal domain) configuration can be analyzed and/or measured. In this way, the relations existing among these different views and due to the preservation of some image property are completely ignored and their potential totally unexploited.

On the other hand, the information shared among the images in different configurations of the vision system and the scene can be utilized to estimate these relations and provide a more complete and consistent representation of the whole scene. Further-

1. INTRODUCTION

more, these relations retain information about the relative motion of the vision sensor and the scene and this property can be employed for pose estimation tasks, in localization and automatic guidance applications, only to cite the most important. In order to recover mutual interactions among different views, pattern recognition algorithms are used to detect invariant features that can be used as references for subsequent *registration*. *Image registration of multiple views* constitutes the main research field that aims at facing these issues and represents an important topic for the scientific and academic research community. Its main goal is to *align* images of the scene in a common and scene-consistent reference frame, by estimating the transformations connecting corresponding patterns of the scene in different views.

In this Chapter the motivations and the potentialities of image registration from multiple views are illustrated, emphasizing the critical aspects to be dealt with in different applicative fields. This better sets up the research approach underlying this thesis. The structure of the thesis is then outlined, summarizing its general methodological approach and the applicative case studies employed for its assessment. Specific contributions of this thesis are explicitly pointed out accordingly as well.

1.1 Motivations and Goals

The general task of registering multiple views under different viewpoints, in different acquisition conditions and depending on scene structure, represents a very complex and challenging research theme. The great effort put in this field by the research community in the last years is motivated by the impact of these research topics in different applicative domains. Patterns related to the same scene feature have to be *recognized* in different views and associated accordingly, taking into account their changes under deformations due to viewpoints motion and possible scene changes (e.g. lighting conditions). Depending on hypotheses or prior information about the sensor model the motion model and/or the scene model, global or local geometrical *mapping functions* between different images or different parts of them can be estimated, and exploited following two main threads.

The first thread concerns *scene reconstruction*, providing a more complete representation of the scene by fusing consistently in a unique reference frame the visual data contained in each view, and/or even recovering the third dimension. In other words,

image analysis of single views can be used to overcome the classical limitation of the restricted field of view to provide a complete scene representation, as in *mosaicing* applications. This representation can be subsequently used in further stages of processing and analysis for complex tasks, like segmentation or object recognition.

Moreover, the geometrical relations between different views retain parameters about the relative motion of the acquisition device with respect to the scene. Accordingly, the relative three dimensional position and orientation parameters, that is the pose, between the device and the scene can be estimated and used to *localize* the acquisition device into the scene environment. This outcome can be used directly to feed a control chain for automatic guidance purposes or indirectly for scene reconstruction (3D reconstruction) or *augmentation* (rendering of virtual objects in Augmented Reality (AR) applications). This second thread proceeds from image analysis to scene synthesis through pose estimation.

Depending on the applicative domains and their related constraints, requirements in terms of accuracy, robustness and computational payload can be more or less compelling. In fact, the algorithms that match among corresponding part of different views can work at different (coarser or finer) detail level. This can rebound on the results in estimating the mapping transform and on the final consistency of the reconstructed scene. The *adaptability* to different acquisition conditions and viewpoint changes is important when repeatability in very different operating conditions is required. Finally, registration algorithms must take into account the computational resources at their disposal, in term of the available hardware technology and the computational performance. The last issue is particularly important in such applications where interactions with other (human or not) agent is required, and a well-timed visual feedback on the properties of the explored scene is needed, accordingly.

This thesis aims at investigating a general approach for on-line image registration from multiple views, more specifically taking into account these issues in two particularly challenging case studies. The first case study regards scene reconstruction through on-line mosaicing of optical microscopy cell images acquired with non automated equipment, while moving manually the microscope holder. By registering the images the field of view of the microscope can be widened, preserving the image resolution while reconstructing the whole cell culture and permitting the microscopist to *interactively* explore the cell culture. In the second application, the registration of

1. INTRODUCTION

terrestrial satellite images acquired by a camera integral with the satellite is utilized to estimate its three-dimensional orientation from visual data, for automatic guidance purposes.

The choice of such different applicative domains is motivated by the desirable goal of providing our approach with *portability*. This goes toward the awkward and still unreached objective of conceiving a general approach working independently from the scene content. Throughout this study, the working hypothesis of rigid motions and absence of deformations are the constraints to be fulfilled.

Moreover, since in both the applications continuous interaction with other agents is required, the corresponding design choices must take into account the computational payload to guarantee suitability for real-time performance. In the next paragraph the structure of the thesis is summarized in more detail, emphasizing the specific contributions while facing these issues.

1.2 Thesis Outline and Contributions

This thesis is structured as follows.

In Chapter 2 the theoretical background beneath image registration from multiple views is summarized. Current state-of-the-art methodological approaches and the related main algorithms are recalled, with some important remarks about their performance in terms of accuracy, robustness and computational payload. This is essential for motivating the design choices that have driven the conception of our registration algorithm. This is described in details in Sect. 2.3, remarking our contribution about the strategies adopted to accomplish an acceptable trade off among accuracy and computational performances.

In Chapter 3 the application of the algorithm we devised to a scene reconstruction case study in the biomedical field is described. In particular, on-line mosaicing of optical microscope imagery with not-automated equipment is investigated. The main motivation behind this application is to widen the nominal field of view of the microscope while fully preserving at the same time the spatial resolution. In fact, a global view of the whole cell culture can be useful to identify special spatial patterns, like

cell colonies, or to build a global reference pattern for subsequent registration or cell tracking in time lapse experiments. The full-resolution image of the whole cell culture can be used for subsequent image analysis steps, like cell segmentation, cell counting, multi-modality image fusion, to cite some of them.

In this context, our approach is capable to overcome some important common limitation in this field. First, it uses only the visual information coming from the acquired images and does not rely at all on motorized stage offsets, thus being suitable for old not automated microscopes whose holder is being moved manually. Second, its on-line performance allows to build the mosaic contemporaneously with the cell culture navigation, permitting to explore the cell culture in its more interesting patterns directly *during* the acquisition. Geometrical accuracy and visual consistency of our method are assessed on large mosaics built from long image sequences using on purpose likelihood metrics. A discussion of the results obtained is provided in order to identify issues still unresolved and propose appropriate future strategies for this application.

This research has been carried out in the context of the ADVANCE (Automatic system to Determine Vital AdhereNt CElls) project, involving the Computer Vision Group (CVG) at ARCES - University of Bologna and the Bone Regeneration Laboratory (BRL), Istituti Ortopedici Rizzoli, Bologna, Italy.

Chapter 4 concerns the application of our registration algorithm on satellite images of the Earth for an automatic guidance application. This research has been carried out in the context of the STARS project (Standalone Three-Axis spacecraft oRientation Sensor) in partnership with the Aerospace Engineering Group at DIEM, II Faculty of Engineering, University of Bologna. A novel autonomous vision-based approach is proposed to estimate the attitude of a remote sensing satellite with high accuracy (up to the order of the arcsec) by registering the Earth images acquired by a camera integral with a satellite along its terrestrial orbit. The estimated three-dimensional orientation can be used for attitude stabilization in view of the automatic guidance and control (AGC) of the satellite. The benefits brought about by this approach are discussed with respect to other state-of-the-art methods currently employed in this field.

The whole methodological chain devised to provide a consistent proof of concept of such an approach is also described in details. Initially, a feasibility study has been carried out to characterize the performance bounds of such an approach, then an

1. INTRODUCTION

on-purpose simulation framework based on synthetic images generated from a georeferenced Earth imagery dataset (LANDSAT 7 ETM+) has been designed. In-depth error source identification and characterization have been carried out based on the theoretical model adopted and the image registration algorithm chosen. This has permitted to identify and isolate, through dedicated simulation configurations and experiments, the critical aspects affecting the overall accuracy. Finally, laboratory tests conducted with a camera mounted on a numerical control unit have been employed to assess the accuracy performance and the influence of the hypotheses derived from the theoretical and simulation frameworks. Discussion of the results obtained encourages the pursue of devising proper compensation strategies.

Finally, Chapter 5 draws conclusions and outlines future developments of this work.

As for my personal research contributions, they have mainly concerned:

- with regard to the ADVANCE project, the geometric registration algorithms, with particular attention to the critical issues (registration model, feature extraction and matching, outlier rejection, etc.) arising in this applicative domain, leading to an automatic and effective method with on-line performance (1, 2). The geometrical algorithms devised here have been implemented in the C++ programming language, also making use of the OpenCV libraries, and utilized by a more complex prototype research for microscope image analysis and data management, integrated with the microscopy system. Microscopy system modeling and system integration issues, as well as photometric registration issues, have been treated taking advantage of the research activity carried out in this field by the other CVG members. The experience of the BRL staff played the main role in the microscopy data acquisition process for the experimental stage.
- with regard to the STARS project, the geometrical modeling of the problem and the image registration issues and algorithms. In particular, I have contributed to design and implement the simulation framework (interacting with an Orbital Simulator provided by the DIEM group) and develop and utilize all the experimental setup. The simulations framework and the registration algorithms have been implemented using Matlab and C++. Specific simulations have been devised

on-purpose to verify the soundness of the model hypotheses and the effectiveness of the registration strategies.

It is worth to notice that this complete analysis methodology constitutes by itself a further novel contribution to the current research state in the related field, permitting comprehensive in-depth error analysis and accuracy performance characterization (3, 4, 5, 6, 7).

1. INTRODUCTION

Chapter 2

Image Registration from Multiple Views

In this Chapter the theoretical background underlying this thesis is outlined. A brief survey of image registration, far from being exhaustive, is initially provided for completeness in Sect. 2.1, emphasizing the common state-of-the-art approaches employed and their applications in different applicative domains. Pros and cons of such approaches are discussed according to critical aspects typically encountered in image registration. In particular, accuracy, robustness and computational complexity are addressed with more emphasis, remarking the applicative requirements (Sect. 2.2). The choices adopted while conceiving our registration algorithm are motivated accordingly, by describing the single stages of the registration algorithm (Sect. 2.3).

2.1 Methodological approaches: a survey

Image registration being a very important research topic in many scientific areas, it has been studied deeply in the last three decades. In fact it is widely employed in Remote Sensing, Biomedical Image Processing, and Computer Vision applications, to cite only some of them. The consistent combination of visual data that are spatially and temporally correlated requires to establish visual correspondences among the different data sources. It implies that a common visual support exists among visual information acquired from different viewpoints (multi-view analysis), from different time instants (multi-temporal analysis), or according to different acquisitions sources (multi-sensors

2. IMAGE REGISTRATION FROM MULTIPLE VIEWS

analysis).

As a consequence, some invariant features must be shared among the images to be registered with respect to a reference pattern. Visual features must be *detected* in different images according to some distinguishing and relevant properties, preserved in spite of different viewing conditions. At the next level, pattern recognition algorithms must have the capability to *match* the visual features among the images. Finally, these correspondences can be used to extract at higher levels spatial and temporal relations among the different views, like their relative pose, for example. Following these steps, the general paradigm of image registration works according to four different stages (8):

- **Feature Detection:** *salient* image properties have to be detected with *repeatability* under different conditions. In other words, the feature chosen must be *preserved* at some level in presence of geometric transformations, photometric changes, noise, etc. Moreover, detected features can be generally localized in the image reference frame at different levels, in general at pixel or even sub-pixel levels. Thus another important property of feature detectors is their spatial *accuracy* (9).

Mainly two approaches are followed for feature detectors, a *featureless dense* or *area-based* approach and a *local sparse* approach. In the first case image properties of a given region, potentially extended to all the image, are used. Generally this approach is based on the image intensity of all the pixels, without actually searching for discriminant features, and delegating the feature matching stage to the registration task.

The second approach relies on the detection of local features, that is localized patterns different in some image property from their surrounding neighborhoods. The properties usually employed to *extract* (and localize) these patterns can fall in the image domain (color, texture, image gradients and their orientation) or in the frequency domain. Since these methods do not work directly on image intensities but rather on image gradients, they are generally more robust to changes in lighting conditions. The resulting structures can be image regions, or contour lines or even local patches (*feature points*), with different level of semantic relevance. A wide literature addresses local structures like *blobs*, *edges*, *corners*, *textons* (10, 11), etc.

- **Feature Matching:** according to this stage, physical patterns of the scene need to be put in correspondence among different views. The most important property of this stage is certainly the *robustness* of the image matching method, since a sufficient number of correct matching should be achieved under different conditions, avoiding *false matchings* that can mislead future stages of image registration.

This stage is strongly correlated with the feature detection stage. Following the main threads used for this stage, two approaches can be outlined. The first, typical of the *featureless area-based* methods, performs exhaustive search in the whole image domain, using likelihood metrics typically employed for template matching (12). The main drawback of such a kind of approach is the computational effort required for the exhaustive search of the image patterns to be matched.

In feature-based approach, *descriptors* with different semantic levels are associated to the detected features. High level descriptors, statistical or morphological, can be associated with a segmented region, for example, as happens for *shape descriptors* (13). In (14), for example, a segmentation algorithm (15) is used to *detect* the luminous profile of headlamp light beams among different views (different beam orientation) for quality assessment purposes. In this case, the segmented contour has been used to compute a *global descriptor* of the profile, that is the *elbow* point showing the maximum change in the profile slope. Methods like this, using global information, can fail when occlusions are present.

Local features descriptors, on the contrary, can cope with this issue. Their sparse nature captures locally image properties that are invariant to several classes of transformations of different nature (geometric, photometric, etc.). Together with robustness, another desirable property of descriptors is their discriminative power, *distinctiveness*, that enforces the probability of correct matchings even in cluttered environments, when “similar” patterns are present. A great number of works have addressed this approach. A recent and extensive survey of local invariant features can be found in (11). However, here it is worth to cite, for the popularity they have currently reached in Computer Vision applications, SIFT (16) and SURF (17) features. Scale invariance is achieved with a multi-scale approach, privileging robustness for SIFT and computational performances for SURF.

2. IMAGE REGISTRATION FROM MULTIPLE VIEWS

- **Warping Model Estimation:** once image correspondences are computed, they can be used to infer the warping transforms linking the different views. Generally, hypotheses about the sensor model, the motion model and the scene structure (rigid, deformable, planar, etc.) are necessary in order to avoid *degenerate configurations* that can cause ambiguity. These mapping transforms can be *global* (valid in the whole image domain) or *local* (referred to local regions). Here we limit our attention to rigid motions for not deformable objects. The research in this field has been characterized by a great effort towards fully automatic approaches (18, 19), working for uncalibrated cameras and in presence of general scene structures, automatically detecting degenerate configurations and estimating accurately the “correct” scene model (20) using *model selection criteria* (21). This goes in the direction of jointly estimating the relative pose parameters between the camera(s) and scene and/or reconstructing the photometric and geometric structure of the scene, preserving scene consistency and accuracy. This is generally achieved by minimizing non linear cost functions on the *global* dataset, that is on the whole image sequence or on a large subset of the acquired images. Accordingly, this often requires the application of iterative minimization algorithms (generally belonging to *bundle adjustment* family) on image *batches*, requiring the sequence to be known in advance. This typically prevents these methods from running on-line.
- **Image Warping:** registration of the images in a common reference frame is finally obtained by warping according to the estimated transform the images in a common reference frame. Generally, first tonal adjustment has to be performed in order to preserve the photometric consistency of the scene taking into account different lighting conditions (22). Once tonally alignment has been performed, image warping can be done using different image interpolation methods (23).

2.2 General working hypotheses

The research approach we have adopted aims at reaching a trade off between performances in terms of accuracy and robustness on the one hand and computational payload on the other hand. The design of this approach has been driven by some working conditions that are commonly encountered in Computer Vision applications:

- the relative motions between the camera and the scene can be modeled as rigid motions, without sensible changes in scale;
- objects of a scene are seen as not deformable;
- image registration must be performed *on-line*, that is *during* the scene exploration, with response timing suitable for interaction with human agents or control electronic devices (typically working at several frames/second). This must be achieved with common (consumer) or limited hardware resources, for which algorithm parallelization on multiple pipelines is not always possible.

2.3 The registration algorithm

Following the requirements addressed in the previous paragraph, we have devised an *hybrid* registration approach. It takes advantage from both the local feature-based and area-based approaches at a reasonable computational cost.

Sparse feature-based approaches are usually capable to provide a high number of features well spread in the image domain. Accordingly, they are more robust to occlusions, in the detection and matching stages, and provides a high number of samples for the model estimation stage. Moreover, to obtain high accuracy in feature localization, a restricted search in a local neighborhood should be preferable to a rough localization based on higher level descriptors. For these reasons, salient features retaining *punctual* image properties have been preferred, even because the low computational cost generally required by their extraction. However, the robustness of the registration algorithms can be affected by such a choice. Area-based methods based on Fourier analysis are, on the other hand, more robust to noise and changes in lighting conditions, but less accurate (generally at pixel level). This robustness is generally obtained since the visual data are treated *globally*. With respect to area-based methods based on image correlation in the spatial domain, their computational cost is lower, due to the fast computation of FFT algorithms (24). Thus the advantageous properties of both these approaches can be merged in order to conceive a *coarse to fine approach*, capable to provide with good robustness a gross initial estimation of the mapping transform, which serves as initialization for a subsequent refinement based on accurate feature tracking.

2. IMAGE REGISTRATION FROM MULTIPLE VIEWS

Interest points are *extracted* from a reference image and then *matched* on the sensed image using the Lukas-Kanade (25) tracking (LKT) algorithm. In order to be robust also to large baselines, feature tracking makes use of a guess location provided by the Phase Correlation algorithm. Once image matchings have been established, *robust* model estimation is performed by using statistical filtering algorithms. In the following paragraphs the choices adopted for each of the stages of our image registration algorithm are discussed.

2.3.1 Feature detection

We have chosen the Shi-Tomasi feature detector (26) as interest point detector essentially for two reasons. The first is that it shows good repeatability with respect to noise, local illumination changes and not too severe geometric transforms. The second reason is that it is the *best* feature for the Lukas-Kanade tracker (25).

2.3.1.1 The Shi-Tomasi corner detector

Shi and Tomasi proposed a stable corner detector analyzing the condition number of the *pseudo-hessian* or auto-correlation matrix: (Eq. 2.1)

$$G(x, y) = \begin{bmatrix} \sum_{i \in W} (I_x(x_i, y_i))^2 & \sum_{i \in W} (I_x(x_i, y_i)I_y(x_i, y_i)) \\ \sum_{i \in W} (I_x(x_i, y_i)I_y(x_i, y_i)) & \sum_{i \in W} (I_y(x_i, y_i))^2 \end{bmatrix} \quad (2.1)$$

on image patches of sizes W centered on points (x, y) of the reference image, being I_x , I_y the local image gradients. It is well known that the image patch can be characterized according to the eigenvalues of G , λ_1 and λ_2 , as:

- a flat region, if both λ_1 and λ_2 are small in value;
- an *edge*, if the eigenvalues are very different and one of them is high, showing image variability in one specific direction;
- a *corner*, if both λ_1 and λ_2 are high in value.

Accordingly, the G matrix is well conditioned if the eigenvalues do not differ too much. At the same time, the eigenvalues must be greater than a certain threshold λ_{min}

in order not to capture noise, so that to fulfill the constraint (Eq. 2.2):

$$\min(\lambda_1, \lambda_2) > \lambda_{th} \quad (2.2)$$

Practically speaking, the minimum eigenvalue is computed for each of the reference image pixels, yielding an *eigenvalues map*. Non maxima suppression is performed on this map on a local neighborhood of the pixel, and the remaining samples are thresholded by a *strength* value λ_{th} . Finally, corner spatial spreadness is ensured rejecting map locations spatially too close to stronger corners.

2.3.2 Feature matching

As discussed above, the feature matching stage proposed here combines in an efficient way two well known algorithms, described in the followed paragraphs.

2.3.2.1 Phase Correlation

The Phase Correlation algorithm in its original formulation is based on the Fourier Shift Theorem, which states that given a reference image $I(x, y)$ and the sensed image $J(x, y)$ related by a global translation (Eq. 2.3):

$$I(x + \Delta x, y + \Delta y) = J(x, y) \quad (2.3)$$

their Discrete Fourier Transforms (DFTs) $F_I(\omega_x, \omega_y)$, $F_J(\omega_x, \omega_y)$ obey Eq. 2.4:

$$F_J(\omega_x, \omega_y) = F_I(\omega_x, \omega_y) \cdot e^{j(\omega_x \Delta x + \omega_y \Delta y)} \quad (2.4)$$

Accordingly, the Normalized Cross Power Spectrum can be expressed through Eq. 2.5:

$$\frac{F_J(\omega_x, \omega_y)}{F_I(\omega_x, \omega_y)} = \frac{F_J(\omega_x, \omega_y) \cdot F_I^*(\omega_x, \omega_y)}{|F_J(\omega_x, \omega_y)| \cdot |F_I(\omega_x, \omega_y)|} = e^{j(\omega_x \Delta x + \omega_y \Delta y)} \quad (2.5)$$

the symbol $(*)$ representing the complex conjugate. The Correlation Surface $CS(x, y)$ can be thus obtained applying the Inverse Fourier Transform (Eq. 2.6):

$$CS(x, y) = F^{-1}(e^{j(\omega_x \Delta x + \omega_y \Delta y)}) = \delta(x - \Delta x, y - \Delta y) \quad (2.6)$$

so that the global translation vector $(\Delta x, \Delta y)$ can be estimated as (Eq. 2.7):

$$(\Delta x, \Delta y) = \underset{(x,y)}{\operatorname{argmax}} CS(x, y) \quad (2.7)$$

2. IMAGE REGISTRATION FROM MULTIPLE VIEWS

This algorithm in his basic version provides global translation components at pixel level. It has been extended also to image rotation (27) and small changes in scale (28), and to sub-pixel level accuracy (29, 30). It is quite robust to additive noise, blurring artifacts and changes in lighting conditions (thanks to normalization in Eq. 2.5) (31). However, the presence of additional geometric distortions, noise and not neglectable modifications in the image content can introduce further spurious *incoherent* peaks in the Correlation Surface, that can even return wrong results in the estimation of the motion parameters. In order to handle the presence of multiple peaks, a threshold equal to the 65 percent of the maximum peak amplitude is applied, and only the $N_{peaks}^{max}=10$ peaks closest to the maximum are kept. For each of these peaks, the correspondent motion parameters are used to determine the corresponding overlapping area among the image pair to be registered. Then a score based on the Normalized Sum of Squared Difference (NSSD) on the overlapping areas is used to choose the *best* peak (the one with the minimum score). Even if quite heuristic, this strategy has proved to be quite robust in our experiments even with images down-sampled with a factor of two, to speed up the algorithm (see Chapter 3 and 4).

2.3.2.2 The LKT tracker

The result of the application of the Phase Correlation provides a guess value for refining the matching stage. We have made use of the Lukas-Kanade tracker in order to estimate residual image transforms with sub-pixel accuracy. This well known method aims at estimating local *optical flow* displacements using a linear approximation of the image intensity spatial and temporal variations. Given a local image patch centered on the pixel (x,y) , under the hypothesis that between the acquisition instants $(t, t + \delta t)$ of the frame couple the *optical flow constraint equation* (32) is fulfilled (Eq. 2.8):

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (2.8)$$

for small motion offsets, the local unknown displacement vector $\bar{v}=(v_x, v_y)$ (assumed locally constant for that image patch) among the image pair $I(x, y), J(x, y)$ can be found by minimizing the likelihood SSD score (Eq. 2.9):

$$\epsilon(v) = \sum_W (I(x, y) - J(x + v_x, y + v_y))^2 \quad (2.9)$$

2.4 Two-view Geometry: model estimation

in a neighborhood W of the feature location. Accordingly, the optimum \bar{v}_{opt} is found for the vector \bar{v} satisfying (Eq. 2.10):

$$\left. \frac{\partial \epsilon(\bar{v})}{\partial \bar{v}} \right|_{\bar{v}_{opt}} = [0 \ 0] \quad (2.10)$$

Expanding Eq.2.10 according to its Taylor series approximation for small displacements, after some passages, the following expression is found for ϵ_{opt} (Eq. 2.11):

$$\epsilon_{opt} = G^{-1} \cdot \bar{b} \quad (2.11)$$

being G the auto-correlation matrix in Eq. 2.1 and \bar{b} a term containing spatial and temporal (that is between the images under registration) derivatives (see (33)). Accordingly, the Shi-Tomasi feature detector ensures that the problem expressed by Eq. 2.11 is well conditioned. The Lukas-Kanade method estimates the vector \bar{v} and uses it iteratively for a subsequent *small signal linearization*, until the algorithm converges or a maximum number of iterations is reached.

This algorithm can reach sub-pixel accuracy but can not handle robustly too large image displacements. Pyramidal implementation of this algorithm (33) aims at coping with this issue, working on more pyramidal levels on which estimating “small” displacement vectors, then back propagated to higher pyramid levels up to the original image. However, this approach can lead even to false matchings when “similar” interest points are spread along the image. For this reason, we have preferred to use a global area-based method like Phase Correlation as a bootstrap step for the tracker initialization.

2.4 Two-view Geometry: model estimation

Given point correspondences on two views, global or local warping transformations can be estimated. In general, they depend on the geometrical configuration linking camera motion and scene, and no assumption about their model can be made if no priors about camera motion and scene structure are given. In this section some features of the *two-view* geometry are briefly recalled, since generally sequential (pairwise) registration approaches are required for on-line applications. Image points are generally represented through their *homogeneous coordinates* (x,y,λ) and accordingly the relations between image correspondences are homogeneous (defined up to a scale factor).

2. IMAGE REGISTRATION FROM MULTIPLE VIEWS

Epipolar Geometry constitutes the general approach to projective geometry between two views. Its most important result is the *epipolar equation* (Eq. 2.12)

$$x'_i \cdot F \cdot x_i = 0 \quad (2.12)$$

which represents a necessary condition once two sets X and X' of N corresponding image points (x_i, x'_i) , $i = 1 \dots N$, are established between the two views. The Fundamental Matrix F is independent from scene structure and retains the camera's relative pose. Accordingly, it can be estimated from image correspondences and exploited to retrieve the camera relative pose (*pose reconstruction*). The knowledge of the mutual camera position and orientation and the image correspondences between the two views can be then jointly employed to reconstruct the three-dimensional structure of the scene (*scene reconstruction*) (34). Estimation of the Fundamental Matrix is subject to ambiguity when *degenerate* configurations are encountered. *Structural degeneracy* is met when the structure of the scene is planar, while *pose degeneracy* is encountered when the two camera centres almost coincide, as for small baselines or pure camera rotations. In these two cases the epipolar geometry approach is not recommended (35), but can luckily be simplified to *planar* projective geometry, referred in the next paragraph. If no priors about the scene or the camera pose are known, model selection (36) can be employed to identify degenerate configurations and adopt the right model accordingly.

2.4.1 Planar Registration

For pure camera rotations (independently from the scene structure) and planar regions (independently from the camera pose) - an approximation that holds also for scenes *distant* from the camera - image correspondence are linked through the *homography* matrix H according to Eq. 2.13:

$$X' = H \cdot X = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot X \quad (2.13)$$

Since it is an homogeneous equation, homography has only 8 degrees of freedom. Homographies are endowed with a group structure so that their composition is still an homography. With these hypotheses, given at least 4 non collinear points matchings, the 9-entries vector \bar{h} of the homography parameters can be estimated, mainly according to two well-known algorithms. The first algorithm, known as the Direct Linear

2.4 Two-view Geometry: model estimation

Transform (DLT) method, aims at estimating the homography parameters rearranging Eq. 2.13 so a linear system in 9 unknowns can be written (Eq. 2.14):

$$A \cdot \bar{h} = 0 \tag{2.14}$$

where the entries of the matrix A are function of the image correspondences coordinates (see (34) for details). A *normalization* algorithm (37) is applied to the image points coordinates in order to perform a preconditioning stage and obtain a numerically stable solution. The vector \bar{h} is estimated solving the (usually overdetermined) constrained problem (Eq. 2.15):

$$\min \|A \cdot \bar{h}\| = 0, \|\bar{h}\| = 1 \tag{2.15}$$

using Singular Value Decomposition (SVD). Accordingly, this algorithm is used to estimate with a low computational effort the homography parameters in a least square sense, that is minimizing the residual in 2.15, also known as *algebraic error*.

The second homography estimation algorithm works by minimizing the *geometric distance* between the *observed* points correspondences (X, X') (i.e. found in the images) and the *estimated* ones (that is computed from the estimated homography) (\hat{X}, \hat{X}') and represents the *gold standard* method. It aims at jointly estimating the homography matrix \hat{H} and the set of image correspondences (\hat{X}, \hat{X}') which minimize the *geometric symmetric distance* (Eq. 2.16):

$$d_H = \sum_i^N d(x_i, \hat{x}_i)^2 + d(x'_i, \hat{x}'_i)^2, \text{ subject to } \hat{x}'_i = \hat{H} \cdot \hat{x}_i, \forall i \tag{2.16}$$

being $d(\cdot, \cdot)$ the Euclidean distance between image points. This algorithm requires the minimization of a non linear cost function and it typically makes use of iterative minimization methods (38), like the Levenberg-Marquardt algorithm. Typically they are initialized with the outcome of the DLT method and are more computationally intensive.

The aforementioned algorithms can be employed for more complex scenes structures and camera motions to estimate local homography mappings on piecewise planar approximations of the scene (39).

2. IMAGE REGISTRATION FROM MULTIPLE VIEWS

2.4.2 Robust estimation

The considerations made above about the model estimation are valid assuming that image correspondences are affected only by *measurement errors*, that is related to image features really corresponding to the *same point* in the scene matched with a limited accuracy. These random errors follow the Central Limit Theorem and are modeled with a Gaussian distribution. However, the matching algorithms can return also *false matchings* that alter this distribution and can dramatically affect the model estimation stage. Accordingly, *robust* model estimation requires the image correspondences to be filtered, so that *outliers* are removed and the model computed on a robust *support* only. To these purposes the RANSAC algorithm (40) has been widely employed in the Computer Vision literature. Its working principle consists in randomly choosing minimal (i.e., with a sufficient number of samples for the model) subsets of the original data, and for each set estimate the model and compute the corresponding *consensus*, that is a function of the number of the other samples of the dataset fitting that model (*inliers*). In order to measure the fitting degree of the generic sample, the model residual is used, and only points with residual under a certain threshold (*distance threshold* d_{th}) are classified as inliers. Finally, the subset with the maximum consensus is chosen and the model refined on the corresponding *consensus* set, containing the inliers for the “winning” model.

The algorithm terminates when a set with consensus greater than a threshold C_{th} is found or after a maximum number of N_{th} attempts. Fixed an empirical value for the probability α that a point is an inlier, and the probability p that at least one of the random sample set is free of outliers, it is possible to compute the value of the parameters d_{th} , C^{th} , and N^{th} , in an adaptive manner, as discussed in (34, 40). In this work the values $\alpha=0.95$ and $p = 0.99$ have been used. As far as homography estimation is concerned, for the matrix H estimated from the generic sample set, the error function employed to compute consensus is the *symmetric transfer error* (Eq. 2.17):

$$\epsilon_t = \sum_i^N (d(x_i, H^{-1} \cdot x')^2 + d(x'_i, H \cdot x_i)^2) \quad (2.17)$$

Chapter 3

On-line Mosaicing of Optical Microscope Imagery

In this Chapter our image registration approach is employed in a *scene reconstruction* application, that is the on-line mosaicing of optical microscope imagery. Through our method, the visual content of single microscope images acquired *while* exploring different areas of biological specimen can be integrated, preserving consistency and level of detail, and thus provide a more complete representation of the whole region under investigation. The application of our approach in this field permits to overcome some existing limitations and provide important additional functionalities. Initially, the motivations behind this application and its potentially strong impact on *high-throughput high-content* image analysis (HTHCA) for the biological research are emphasized. The working context of this application is then described and discussed with respect to the state of the art in this field. Moreover, significant features of our image registration algorithm within this applicative context are analyzed, remarking how they can affect the final outcome. Mosaics resulting from real datasets in different conditions are presented and evaluated. The experimental results obtained in different biological examinations confirm the efficacy of our approach. Besides being visually pleasant, the achieved mosaics exhibit accuracy and consistency with the original images and can be exploited for further global image analysis steps on the whole cell culture.

3.1 Introduction

Microscopic imaging of *in vitro* live cells represents an important tool for researchers to study spatial and temporal evolution patterns of the whole culture. HTHCA (41, 42) in this field is stirring up interest for the advantages that it can bring in speeding up the biologist work, starting from the microscopic imagery acquisition up to its automatic processing and classification of cellular structures. Such image analysis tools can simplify and expedite different stages of biological experiments, since they allow to efficiently combine and process different information taken at different time in different areas of the culture. Besides saving the microscopist from tedious as well as repetitive and time expensive tasks, they can enhance the range of functionalities offered traditionally by the microscope, providing *virtual microscopy* (43) capabilities.

This aspect is particularly important in time lapse microscopy where a global view of the whole cell culture can be useful to identify special spatial patterns, like cell colonies, or to build a global reference pattern for subsequent registration or cell tracking (44). More in general, this is important to provide the microscopist with a more complete comprehension of some particular features of the whole culture, that on the contrary would result awkward to achieve through the user's visual inspection, at not negligible temporal costs.

In (45), for example, these issues are discussed with regard to a neuro-imaging application. In Fig. 3.1, examples of particular spatial patterns of biological interest are shown. As it can be seen, the study of the properties of the cells (number, morphology, etc.) and their configurations can be strongly affected by the spatial extent of the biological structures within the viewed area. To these purposes, *image mosaicing* (also called *montaging* or *tiling*) techniques can be applied to build a wide field-of-view image of the whole cell culture area during the microscopic investigation, while fully preserving spatial resolution of each single image.

The full-resolution image of the whole cell culture can be used for subsequent image analysis steps, like cell segmentation, cell counting, multi-modality image fusion (see Fig. 3.2), efficient image storage (through reduction of the redundancy of the overlapping areas) (46), to cite some of them. Because of the use of the resulting mosaic for subsequent processing and measuring stages, geometric and photometric consistency has to be preserved, with compelling requirements on accuracy and robustness.

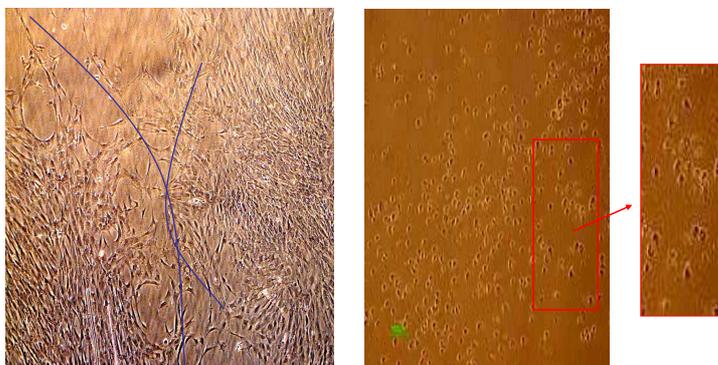


Figure 3.1: Examples of particular spatial cell patterns. **(Left).** Two cell colonies can be identified (blue separation lines) in this cell culture. **(Right).** Typical pattern for cell necrosis (with circular lipid vesicles)

Generally, methods and systems used for microscopic image mosaicing perform in *batch* mode, building the mosaic at a separate stage at the end of the image acquisition on the whole sequence of images. Moreover, the geometric registration of the images makes often use of human intervention and/or relies on known motorized x-y stage offsets of the microscope holder (43, 47) to align the images, then requiring the application of a subsequent *global* refinement stage by minimizing some objective error function. This approach, besides not working for the most common not motorized microscopes, even for current-generation motorized stage microscopes can introduce an additional functionality but however requiring the microscopist to wait the end of the acquisition process to achieve the mosaic. This completely jeopardizes the *interaction* between the microscopic system and the researcher in order to make him know, *during* the acquisition, of the most of the interesting patterns of the cell culture. Our approach permits to *interactively navigate* the cell culture through regions of interest within the acquisition process, providing the microscopist with an immediate visual feedback and contributing to the scene navigation even for cluttered patterns at different cell density. In other words, the explored field of view can be used directly by the microscopist to *localize* structures of interest that would be possibly scanned again, once a better understanding of the whole (often cluttered) area under investigation is reached.

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

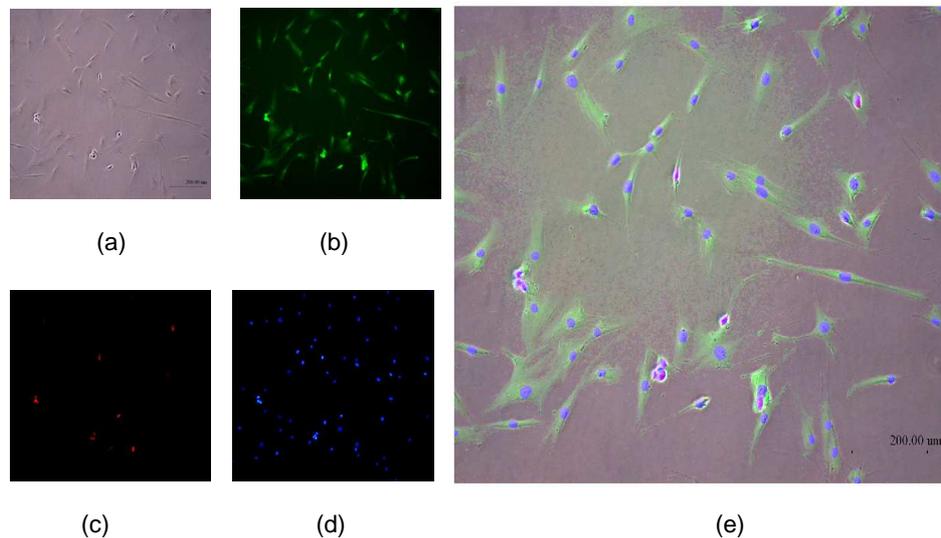


Figure 3.2: (a) Image of a cell culture acquired in phase contrast mode and in fluorescence mode after coloring with different cell markers identifying nuclei and cytoplasm of all the cells (b), nuclei of necrotic cells (c), nuclei of the active cells (d), respectively. (e) Image obtained by multi-modality fusion of the registered images. Functional properties (live, apoptotic cells, etc.) can be studied according to visual features (cell morphology, etc.).

In this Chapter our registration approach is used to design an automatic on-line mosaicing method for optical microscopy imagery based on visual information only, capable to preserve photometric and geometric consistency during the *manual* motion of the microscope holder. In Sect. 3.2 approaches utilized in this research field are illustrated. In Sect. 3.3 the mosaicing algorithm we devised is discussed, remarking the strategies chosen to process light microscopy imagery. First, a photometric-consistent adjusting step based on flat field compensation is used to cope with the uneven spatial distribution of the microscope light illumination field and avoid shading artifacts. Second, accurate image registration is performed in order to geometrically align couples of consecutive frames. Image motion is estimated robustly even in the presence of migrating bodies (e.g. impurities, cell secretions) and condense artifact using statistical filtering of outliers. Global consistency even for looping path sequences is achieved on-

line using a Frame-to-Mosaic registration approach, preventing the registration error to accumulate and amplify. In Sect. 3.4 the experimental setup used to collect and process microscopic images is illustrated, together with useful metrics used for the evaluation of the resulting mosaics. Finally, Sect. 3.5 draws some conclusions about the current achievements and proposes future developments.

3.2 Previous Work

Image Mosaicing represents an important issue in the Computer Vision research community, and accordingly a high number of papers have been published in this field in the last two decades.

Usually, panning motion of the camera and/or planar or general distant scene are assumed, so that parallax effects are not present and planar registration (Sect. 2.4.1) can be applied. Images to be registered can be taken from video shots (48) or can present wide baseline (49), thus affecting the robustness requirements of the matching stage. Both featureless dense correlation-based methods working on pixel intensity (50, 51) or sparse feature-based approaches (49) have been employed in this stage, depending on the computational and accuracy performance required. Then, the warping parameters are estimated, in the majority of the cases, by using sequential pairwise registration at first, and finally by applying *global* registration on the whole image sequence or on a large subset of the acquired images. Global registration is generally performed by minimizing non linear cost functions containing many unknowns, generally depending on the extent of overlapping areas, the number of features, the number of views and the motion model complexity. Works in (48, 51, 52) follow this approach, using iterative optimization in a *bundle adjustment* fashion. Accordingly, the high computational payload associated to this optimization stage requires off-line processing.

More specifically, as far as mosaicing of light microscopy imagery is concerned, it has been often focused on *post processing* the whole sequence of images to provide visually pleasant mosaics. The algorithms employed in this context have different hardware requirements and degrees of automation.

A first class of algorithms follow a *dense* featureless registration approach with likelihood error function based on pixel image intensities and accordingly are computational intensive.

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

The work described in (53) concerns a semiautomated method, based on pairwise registration through image cross correlation, and its implementation on a desktop computer, performing off-line without requiring the holder to be motorized. In (54), a semi-automated method which requires the user to manually align the *tiles* for a subsequent fine registration stage is presented. This work being focused on accuracy performance (up to sub-pixel level), images are pixel-wise registered using a dense featureless approach, thus resulting in a high computational burden that prevents this method to be used on-line. The methods proposed in (55) and (56) are conceived to be used necessarily with high-precision motorized x-y stages. Metadata provided from motorized stage controllers (55) and mosaic initialization through manual alignment (56) are used for a coarse geometric registration, while global tonal and geometric alignments are performed by minimizing a cost function over the pixel intensities of all the images' stack. Accordingly, these methods work in batch mode at the end of the acquisition.

A second class of algorithms relies on sparse feature-based registration approaches, *detecting* and *matching* salient regions in consecutive images. The algorithm described in (57) utilizes Harris detectors (58) to identify salient points and normalized moment of inertia (NMI) as their feature descriptor. This method is applied to single couples of images while general issues regarding the mosaic generation (consistency of photometric and geometric registration) are not addressed in this work. In (59) the software package Autostitch (49) developed for panoramic image generation is tested on microscopic image stacks acquired during manual or motorized motion of the microscopes holder. Autostitch is based on SIFT (16) detection and matching to increase robustness, but it operates off-line. The method proposed in (60) uses wavelet-based edge correlation to detect feature points and normalized cross correlation for their matching. This method is not conceived for on-line mosaicing since it needs global registration to achieve an accurate mosaic.

3.3 The method

The algorithm we devised has been tailored to provide on-line, and hopefully real time, performance, while preserving contemporaneously geometrical and photometric consistency. Inhomogeneous spatial distribution of the microscope light field should be taken into account and compensated in order to avoid *seams*, when compositing

the images into the mosaic, due to shading or vignetting effects. To this purpose, the microscope luminous field has to be modeled and the acquired images *normalized* accordingly. In Sect. 3.3.1 the methods proposed to deal with this issue are described in more detail.

Once the images have been mapped to the same luminous range, geometric registration is needed to represent them in a common geometric reference frame. By registering sequentially couples of consecutive frames through our sparse-features algorithm, the warping transformation between them is estimated and they can be rendered in a common reference frame. This requires that a certain percentage of overlapping areas is shared between the image pair. Considerations on how the acquisition conditions and the scene model affect this warping transform estimation stage are discussed in Sect. 3.3.2.

During the biological specimen navigation, the same area can be revisited more times. It is well known that for (long) looping path sequences, errors due to pairwise registration (Frame-to-Frame registration, F2F) can accumulate, resulting in misregistrations (and accordingly *mosaic seams*) of the corresponding region of the scene. To compensate also for this drift effects preserving on-line capabilities, additional Frame-to-Mosaic (F2M) registration (61, 62) is applied. The geometric registration framework is illustrated in detail in Sect. 3.3.2.

3.3.1 Flat Field Correction

In this paragraph two possible strategies, employed in our framework to estimate light field and to correct for its uneven spatial distribution, are addressed. With respect to natural images, for which it is difficult to model the camera response to light distribution, in microscopy laboratories we can assume that the illumination conditions are relatively well controlled. We have seen from our experiments that the limited extent of the area analyzed during the specimen investigation is more sensitive to the optical layout of the microscopy system rather than to external light variation. Accordingly, the main factor affecting the photometric domain of the acquired images is the different spatial response of the microscope lenses to light. This results in evident vignetting and shading effects that must be compensated before warping the frames into the common geometrical reference frame.

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

Considering image intensities as linearly correlated to light intensity, the illumination field can be estimated by analyzing the spatial distribution of the image gray values. Light distribution within the field of view of the microscope can be estimated from images of an *empty field* or directly from several initial images also containing objects of interest and acquired on-line during the specimen investigation.

In the first case, a stack of empty field images is created during a bootstrap acquisition stage, performed before positioning the specimen on the holder (i.e., in *empty field* conditions). Temporal averaging is performed for each pixel gray level over the stack to obtain a smoothed distribution pattern of the empty field. Then the *illumination field* $IF(x, y)$ is computed normalizing the empty field to its minimum non-zero value. In Fig. 3.3 an example of the estimated illumination field surface is shown.

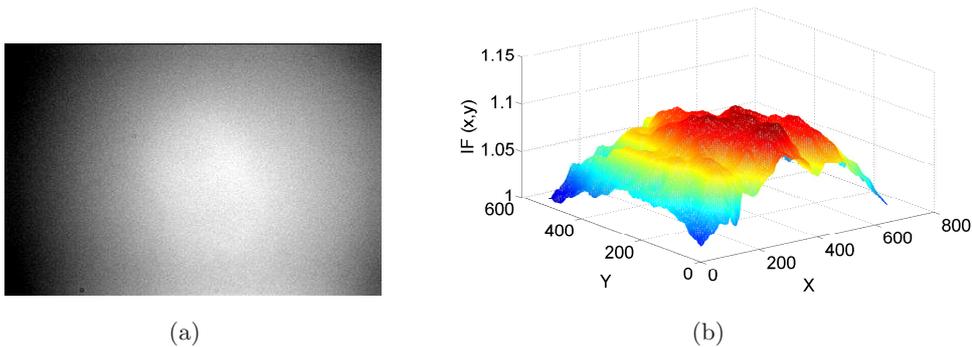


Figure 3.3: (a) Image resulting from averaging on the empty field images stack. (b) Mesh of the estimated empty field, assuming a Gaussian-like shape.

Illumination field can be alternatively estimated from a first set of images of the biological specimen, by measuring image intensities from regions with uniform reflectance and spatially distributed over the specimen area. For these reasons, for their different densities (and reflectance), candidate areas must not contain cells, and are chosen among the regions of *background*, that is belonging to the cell culture medium devoid from cells. To this purpose, an algorithm for background detection has been devised within our research group, and only image intensities extracted from the background areas are used to estimate the illumination field according to a fitting procedure. More details about this algorithm are given in (63).

Once the illumination field is estimated, a *flat field correction* is applied to each acquired image $I(x, y)$ before its geometric registration and warping into the mosaic.

It consists in weighting (through pixel-wise division) image intensities according to the estimated illumination field. Accordingly, the resulting *normalized* images $I^*(x, y)$ are defined by Eq. 3.1:

$$I^*(x, y) = \frac{I(x, y)}{IF(x, y)} \quad (3.1)$$

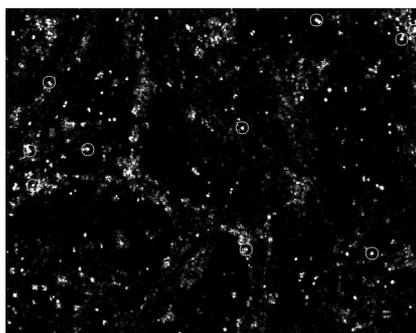
3.3.2 Geometric Registration

We have chosen to adopt a sequential registration approach between consecutive frames based on sparse visual features, rather than using *dense optical flow methods*, essentially for computational reasons. During acquisition, each (normalized) current frame is registered with the previous one in a common reference frame using the algorithms described in Chapter 2.

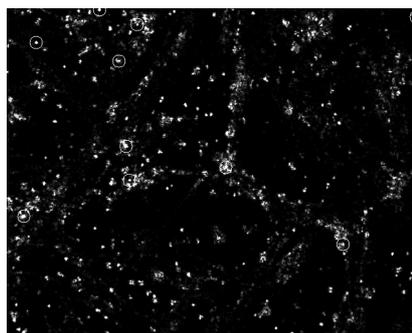
The *coarse-to-fine* strategy proposed represents in fact a good trade-off between accuracy and computational payload. The Shi-Tomasi corner detector is able to extract salient points even from poorly contrasted image, as it happens for phase contrast (64) acquisitions. The corresponding *strength threshold* used in our experiments has been empirically set to the higher 35th percentile of the pseudo-hessian eigenvalues of the image, so that only values greater than this threshold are kept. Although being heuristic, this value has proved to be valid in real world cases. An example of the *eigenvalues map* and the positions of the extracted corner points is shown in Fig.3.4 (a),(b). This example is related to a live stem cell culture and it can be noticed how, in spite of the poor contrast of the cytoplasmic membrane with respect to the cell culture medium, a considerable number of corners can be extracted (and matched).

The Phase Correlation stage can cope with “large” translations of the microscope holder while preserving accuracy. To reduce the computational burden due to the Phase Correlation, this algorithm has been applied on down-sampled images (with a factor of 2), and then the estimated translation at this level is rescaled in the original domain. This gives a coarse approximation for the image translative field that can be refined using the LKT. For each image pair, the frame-to-frame (F2F) warping transformation has to be estimated from the sets of corresponding features. In Fig.3.4 (c),(d) corresponding features in the two views are referred through a common numerical identifier.

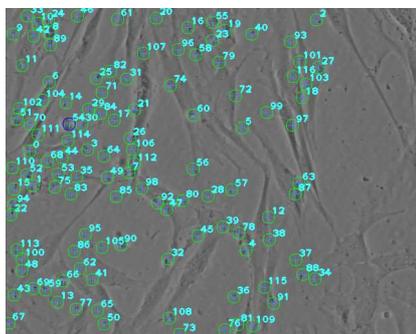
3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY



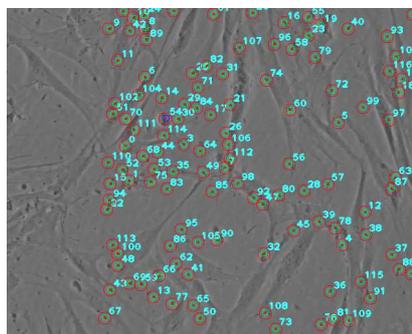
a



b



c



d

Figure 3.4: (a), (b) Eigenvalues maps for a sample image pair of a histological sample. As it can be seen, higher eigenvalues are often located near the borders of cellular membranes. (c), (d) Corresponding features resulting from the registration algorithm in the two views are associated through a common numerical identifier.

3.3.2.1 Model Estimation

The acquisitions conditions, the scene model and the camera model should be taken into account for the selection of the warping model to be estimated. The system can be modeled as composed of a fixed (projective) camera which observes the manual rigid movement of the microscope holder. The thickness of the specimen (some microns) being negligible with respect to the objective working distance (some centimeters), the scene can be considered planar. In these conditions the *parallax effect* can be neglected and corresponding features X_i, X_j on two consecutive views $I_i(x, y), I_j(x, y)$ are related by a planar homography H_i^j according to the homogeneous Eq. 2.13, here recalled for completeness (Eq. 3.2):

$$X_j \cong H_i^j \cdot X_i = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot X_i \quad (3.2)$$

with $\|\bar{h}\|=1$. Thus, proceeding by a set of at least 4 non collinear feature matchings in the system of Eq. 3.2, the estimation of this eight-parameter transform would be required in general.

However, the warping transform actually depends on the real acquisition conditions and can belong even to sub-groups nested in the group of the homographies. In our case study, two additional approximations can be considered as satisfied. The depth extent of the scene (the biological sample) is small if compared with the average distance from the camera principal point, and because of the “small” field of view of the camera, the imaged points can be considered *close* to the optical axis. Under these hypotheses, the perspective camera model can be relaxed to the *affine* model (34) expressed by Eq. 3.3:

$$X_j = A_i^j \cdot X_i = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \cdot X_i \quad (3.3)$$

and the parameters to be estimated decrease to six. Moreover, by proceeding with this complexity reduction approach, we can suppose to neglect the mechanical play of the holder (that can be affected by drift effects due to its continuous use) and the relative deviation of the camera optical axis from the normal to the holder, conceiving accordingly a translative model, with the number of parameters to be estimated reduced

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

Model	k	N_{min}	H
Projective	8	4	$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}, \ \bar{h}\ =1$
Affine	6	3	$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}$
Translative	2	1	$T = \begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix}$

Table 3.1: Properties of the warping models employed in this context. k is the number of parameters (degrees of freedom) of the model, N_{min} is the minimum number of matchings needed to estimate the model.

to two (the x and y translation components)(Eq. 3.4):

$$X_j = T_i^j \cdot X_i = \begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \cdot X_i \quad (3.4)$$

In Table 3.1 the main features of these models are summarized. The estimation of these models is carried on solving the (overdetermined) system in Eq. 3.2, minimizing the algebraic re-projection error, that in these last two cases has the same expression of the geometric distance.

It is worth to notice that the estimation procedure of the warping transform is quite sensitive to the presence of outliers (*false matchings*) and to the percentage of the overlapping area between two consecutive frames. Outliers can be due to the registration algorithm, as it happens when (false) matchings are established between points related to different parts of the scene, or to the presence of migrating impurities corpuscles and slight modifications of not stable structures. Furthermore, the depth of field is not spatially uniform, so that corresponding points in different views can present out-of-focus blurring effect. Rather than characterize each of these error components and compensate for them, the statistical filtering stage based on RANSAC (40) algorithm is employed to cope with these factors.

In Fig. 3.5 corresponding features for a couple of consecutive frames according to this procedure are shown, with a numerical identifier associated to each matched pair

of features. Inliers are identified by corresponding circled cross ticks, while rejected outliers are identified by simple not circled crosses. Boxes with magnified details show how structural modification of cells, among the other factors, can affect the feature matching stage and how the RANSAC stage can cope with this effect.

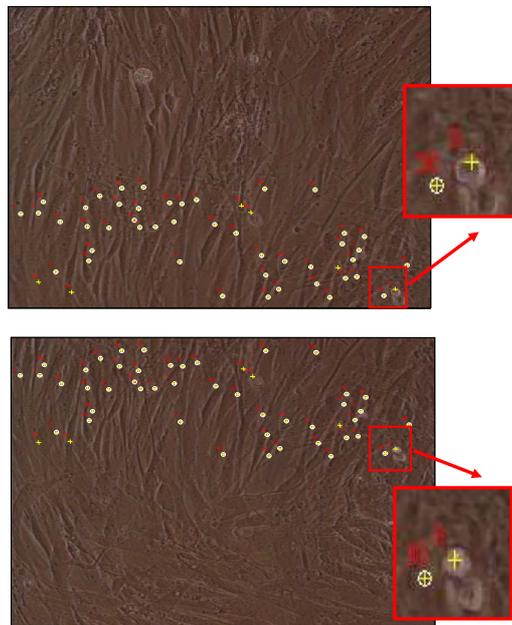


Figure 3.5: Couple of consecutive frames with corresponding feature points superimposed (with common numerical identifier). Circled crosses identify inliers of the RANSAC stage, isolated crosses indicates outliers. Images with magnified details show how the RANSAC stage is able to cope with outliers even due to local morphological changes.

However, outliers are correlated also with the warping model chosen to represent the transform between two views. As discussed in general (21), and more specifically for mosaicing applications in (65), the *model selection* issue for the geometrical transform to be estimated can affect the resulting mosaic. The estimation process for the warping parameters can result unstable if noisy correspondences are established in a small overlapping region. In the presence of outliers, estimating more complex models can

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

lead to ill-conditioned problems, while less complex models can be more *robust* to the risk of *overfitting* the model parameters to the noisy data. In order to avoid *degenerate configurations*, that can lead to large image distortion, the warping model has to be selected *properly*. A very popular criterion for model selection is the GIC (Geometric Information Criterion) and its variants (21), in particular for on-line application, due to the efficiency of its computation. It takes into account different features of the set of models and chooses the model with the least GIC score as the most adequate. GIC is defined as (Eq. 3.5):

$$GIC = \sum_{i=1}^n \frac{e_i^2}{\sigma^2} + \lambda_1 \cdot d \cdot n + \lambda_2 \cdot k \quad (3.5)$$

The first term on the right-hand side (RHS) derives from the Maximum Likelihood Estimation (MLE) of the model parameters, maximizing the probability of observing a set of given features correspondences for a certain set of model parameters. In particular, this term is the log likelihood of the correspondences for a Gaussian distribution of the noise perturbed data, being e_i the re-projection error (between the measured data and the data returned from the estimated model) with variance σ^2 . The third term of RHS is proportional to the number of parameters of the model and penalizes more complex models through the weight λ_2 . For $\lambda_2=2$ the first and the third terms of RHS form the well known Akaike Information Criterion (AIC) (66). GIC is obtained by integrating AIC with the second term of the RHS, which takes into account the number of the samples n used to estimate the model and the dimension d of the model.

This geometric fitting approach aims at selecting the correct model using the observed data from the current couple of frames provided that the noise distribution is known. By choosing $\lambda_1=\lambda_2=2$, as it happens for the Geometric Akaike Information Criterion (GAIC) (67), the influence of the model complexity can be masked from the high number of samples. For this reason, the choice of these weights is important in order to obtain a robust criterion, as discussed in Sect. 3.4.

3.3.2.2 Registration Modes

Once the pair-wise matrix H_i^j has been estimated, obtaining the matrix \widehat{H}_i^j , it is used to perform a *guided search* using the *estimated* features $\widehat{X}_j = \widehat{H}_i^j \cdot X_i$ as guess values for a further LKT tracking. The pairwise warping model is the estimated by using RANSAC

on this last set of image correspondences, and the global transform that warps each image into the *mosaic reference frame* can be computed. We have chosen the mosaic reference frame X_M to be coincident with the first frame X_0 , with H_0^1 being equal to the identity matrix. The *mosaic warping matrix* for the n^{th} frame, M^n , defined by the relation (Eq. 3.6):

$$X_M = X_0 = M^n \cdot X_n \quad (3.6)$$

can be obtained by incrementally chaining, through the matrix left-productory operator (Eq. 3.7),

$$\prod_{i=0}^{n-1} H_i^{i+1} = H_{n-1}^n \cdots H_0^1 \quad (3.7)$$

the transform matrices estimated for the previous frames, according to Eq. 3.8:

$$M^n = \left(\prod_{i=0}^{n-1} H_i^{i+1} \right)^{-1} \quad (3.8)$$

still belonging to the homography group.

This incremental approach is prone to *dead reckoning effects* due to accumulation of the estimation errors, that can turn to increasing geometric deformations and misalignments. This effect is more evident for long looping path sequences, where frames of revisited areas of the scene can present a visible misalignment with the frames already mapped into the mosaic. As discussed above, generally a *global* registration of the collected frame is required to compensate for these effects. However, this would require to know all the image sequence in advance and would result in a computational intensive optimization stage. Alternatively, we have chosen to adopt a Frame-to-Mosaic approach (61, 62) that permits to reach a good trade off between the accuracy of the resulting mosaic and computational performances. According to this approach, the current n^{th} frame is registered not only with the previous $(n-1)^{\text{th}}$ frame, but also with the mosaic built up to that point, M^{n-1} . In other words, a corrective contribution is estimated by registering the current frame with the corresponding area, according to the F2F registration, in the mosaic reference frame. Being H_n^C the matrix taking into account this corrective term, Eq. 3.6 becomes Eq. 3.9:

$$X_M = H_n^C \cdot M^n \cdot X_n \quad (3.9)$$

This further registration has the advantage of compensating mis-registrations at a reduced computational cost, since it requires only one more LKT tracking stage, fed by

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

the estimated model resulting from the F2F registration. Accordingly, the application of the Phase Correlation stage, more computationally expensive, is not needed at this point.

Once the mosaic warping matrix has been computed, the n^{th} image is then warped into the mosaic reference frame, accordingly, using bilinear interpolation and merged into the mosaic using a *stitching* approach, that is replacing the underlying pixel value. This choice has been preferred to traditional *blending* approaches in order to avoid *ghosting* effects due to the motion of particles in adjacent frames. Moreover, as mentioned before, for microscopic imagery mosaic seams are almost due to vignetting and shading effects. Our flat field correction stage, compensating for these artifacts, permits to create a visually pleasant mosaic while keeping the high geometric accuracy achieved by our registration method, as shown in the next paragraph.

3.4 Experimental Results

In order to test our algorithm, image sequences of biological samples have been acquired in phase contrast mode, using standard non-motorized optical microscopy instrumentation, and processed on-line. In particular, an inverted microscope Nikon Eclipse TE2000-U, widely used in research labs, has been equipped with a digital camera (Nikon DXM1200) able to perform live acquisition at 640×512 pixel resolution. The system has been connected to a consumer PC (Intel Pentium 2GHz, 4GB RAM). Fig. 3.6 summarizes the framework used for our experiments. During our experiments, a magnification factor of $100\times$ has been applied, this resulting in a spatial resolution of $1.0152 \mu m/pix$. Before positioning the specimen on the holder, empty-field images have been acquired for several seconds (yielding a stack of 68 elements) in order to estimate the illumination field, as described in Sect. 3.3.1. Then, during acquisition, the holder is moved manually by a microscopist to his liking and the mosaic built on-line accordingly.

To evaluate the results of our algorithm in different working conditions, two different test configurations have been set up. The first consists of images of a histological sample (HS, hereinafter) of altered bone tissue, characterized by the presence of background regions among connected structures. The second refers to a live stem cells culture (SC, afterwards). Both are acquired in phase contrast mode.

Nikon Eclipse
TE-2000UDXM1200 Digital
Eclipse Hardware

PC

Figure 3.6: Experimental setup used in our tests. The inverted microscope Nikon Eclipse TE2000-U has been equipped with a digital camera (Nikon DXM1200). The system has been connected to a consumer PC.

3.4.1 Metrics for Quality Assessment

The quality of the resulting mosaic can be evaluated quantitatively using proper quality indexes. Usually, a consistent metric must be adopted for the comparison of the results with the ground truth data. However, in our case study, since the experimental equipment is not automated, ground-truth data are not available, even as far as the holder motion is concerned. The only ground-truth data available are the single frames acquired during the holder motion, that is the *reference images*. Accordingly, the information contained in the single frame can be compared with its corresponding area in the mosaic according to some metric.

In fact, given the input image sequence and the mosaic $M(x, y)$ built accordingly, it is possible to achieve for each original image, by means of the mosaic warping matrices, the corresponding overlapping area in the mosaic. Given the sequence of N reference images I^i to be mosaicked, let $R_I^i(x, y)$ the i^{th} warped reference image, i.e. achieved by *warping* $I^i(x, y)$ according to the matrix M^i , and $R_M^i(x, y)$ the mosaic region having the same support T^i . The area $R_M^i(x, y)$ of the mosaic is generally the result of the contributions of the $R_I^i(x, y)$ and also of other different warped reference images. Accordingly, $R_M^i(x, y)$ is generally partitioned into $m_i + 1$ regions $A_i^i(x, y) \dots A_{i+m_i}^i(x, y)$, each partition containing only the mosaic pixel derived from one of the reference images that are warped into $R_M^i(x, y)$. This formalism is illustrated through Fig. 3.7.

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

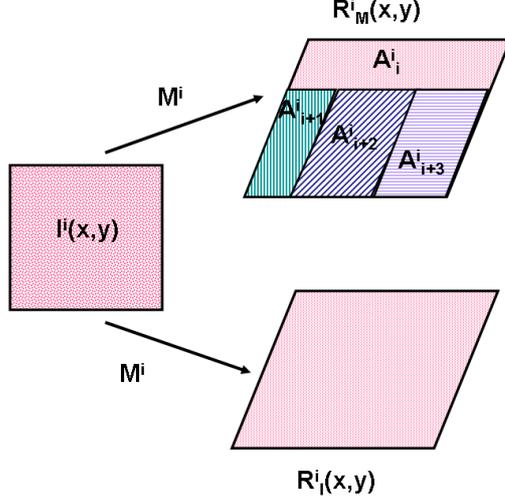


Figure 3.7: Explicative figure for the symbols used to indicate the regions defined to compute quality metric indexes. In this case $m_i=3$. Dimensions are not preserved purposely to improve visualization.

For the sake of clarity, in the following we consider the generic $R_M^i(x, y)$ and omit the index i in the subscript of the expression m_i and in the superscript of the expressions A_j^i . For the generic i^{th} reference image, the support S^i used to compute metric indexes is defined as follows (Eq.3.10):

$$S^i = \bigcup_{j=(i+1)\dots(i+m)} A_j \quad (3.10)$$

that is the union of the partition subsets of $R_M^i(x, y)$, except for the A_i^i region which contains information related to the i^{th} reference frame only. Since $S^i(x, y)$ is a subset of the support T^i , common to both $R_M^i(x, y)$ and $R_I^i(x, y)$, also $R_I^i(x, y)$ can be referred to $S^i(x, y)$. We denote as $P(S^i)$ the number of pixels of S_i .

The Root Mean Squared Error ($RMSE^i$) and the Signal to Noise Ratio (SNR^i) can be defined by considering the image intensities on the support $S^i(x, y)$, as follows (Eq. 3.11, Eq. 3.12):

$$RMSE^i = \sqrt{\frac{\sum_{(x,y) \in S^i} (R_I^i(x, y) - R_M^i(x, y))^2}{P(S^i)}} \quad (3.11)$$

$$SNR^i = \left[\frac{\sum_{(x,y) \in S^i} (R_M^i(x,y))^2}{\sum_{(x,y) \in S^i} (R_I^i(x,y) - R_M^i(x,y))^2} \right] \quad (3.12)$$

usually expressing SNR in [dB].

Accordingly, a sequence of N images yields N samples for each of the two metric indexes. The i^{th} sample of the metric index sequence aims at measuring the likelihood of the i^{th} reference image with the corresponding area $R_M^i(x,y)$ in the resulting mosaic, *provided that* in $R_M^i(x,y)$ only the contributions coming from images different from the reference one are taken into account. In fact, if a region of the mosaic arises from the stitching of a single reference image, the comparison of $R_M^i(x,y)$ with its warped reference image, i.e. $R_I^i(x,y)$, would not be consistent since the two regions are related, in this case, simply by the estimated warping, and their likelihood would be obviously close to the maximum. On the other hand, if errors in image registration occur and propagate, it is likely that the reference images are not registered consistently into the mosaic reference frame, thus even with respect to the other images bringing contribution for *building* (or *hitting*) the corresponding area of the mosaic. In this case, a misalignment in the mosaic should rebound on the metric indexes. Accordingly, only the pixels of the mosaic *not hit* by the reference frame are taken into account in Eq. 3.11 and Eq. 3.12. It is worth to notice that the support of the N^{th} frame is the empty set ($S^N = \emptyset$), since this last frame is integrally stitched on the mosaic, so that the mosaic region $R_M^N(x,y) = R_I^N(x,y)$ is simply obtained warping the reference frame $I^N(x,y)$ according to the mosaic warping matrix. As a consequence, the computation of the metric indexes can be skipped for the last frame, yielding $N - 1$ samples in total.

To better explain how this evaluation methodology works, in Fig. 3.8 the masks related to the $R_M^i(x,y)$ regions are shown for a sample mosaic, each with a corresponding label. For example, observing the mask corresponding to the eighth $R_M^8(x,y)$ region (identified by blue borders), it is hit, besides by itself, also by the ninth, the tenth and the eleventh reference images. The *support* to be considered to compute the metric indexes is the union set of the pixels hit by only these three last reference images.

The expressions in Eq. 3.11 and Eq. 3.12 can be extended taking into account all the $S^i(x,y)$ areas of the first $N - 1$ reference images, and referred to the whole mosaic as global quality metrics. More formally, we define the *global support* S_M as follows

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

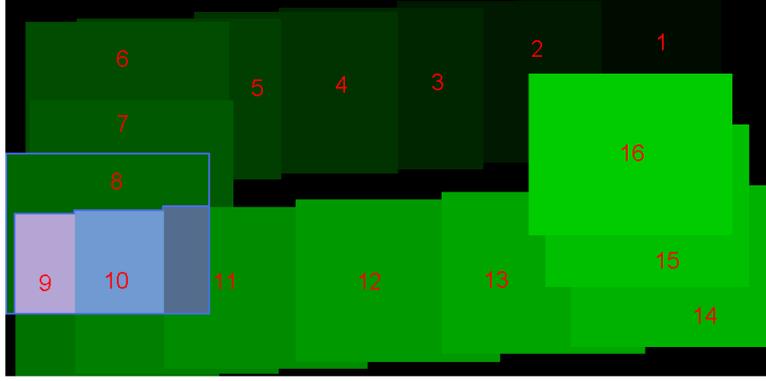


Figure 3.8: Labeled masks of the *hit regions* $R_M^i(x, y)$, being each numerical identifier associated with the corresponding i^{th} warped reference frame ($R_I^i(x, y)$). Blue borders surround $R_M^8(x, y)$. The support used to compute metric indexes for the eighth frame is formed only by the (not green) mosaic areas hit from the ninth, the tenth and the eleventh reference images.

(Eq. 3.13):

$$S_M = \bigcup_{h=1 \dots (N-1)} S^h \quad (3.13)$$

and accordingly the mosaic Root Mean Squared Error $RMSE_M$ and the mosaic Signal to Noise Ratio SNR_M are defined as follows: (Eq. 3.14, Eq. 3.15):

$$RMSE_M = \sqrt{\frac{\sum_{i=1}^{N-1} \left[\sum_{(x,y) \in S^i} (R_I^i(x, y) - R_M^i(x, y))^2 \right]}{\sum_{i=1}^{N-1} P(S^i)}} \quad (3.14)$$

$$SNR_M = \frac{\sum_{i=1}^{N-1} \left[\sum_{(x,y) \in S^i} (R_M^i(x, y))^2 \right]}{\sum_{i=1}^{N-1} \left[\sum_{(x,y) \in S^i} (R_I^i(x, y) - R_M^i(x, y))^2 \right]} \quad (3.15)$$

These metrics have been defined in order to evaluate our results according to three different aspects:

- **flat field normalization:** the effect of the application of the flat field compensation can be discussed with respect to the results obtained processing raw data;
- **choice of the warping model:** once flat field correction has been applied, being all the reference images in the same photometric domain, results obtained choosing different warping models (projective, affine, translative) can be compared;
- **registration mode:** especially for long looping path, the impact of the F2M compensation is evaluated with respect to the incremental F2F registration.

3.4.2 Results

Results related to three image sequences are here presented and analyzed. They have been chosen as representative of the behavior of our algorithms among sample sequences collected in extensive acquisition campaigns. Other results, consistent with the ones reported here, can be found in (1).

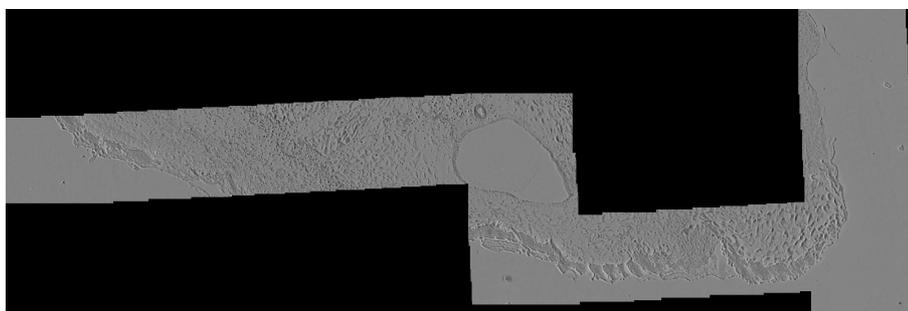
The first sequence (HS) is relative to a set of 60 images (59 image pairs) of a histological sample, registered incrementally (F2F) with the three different warping models. In Fig. 3.9 (a),(b), the resulting mosaics for the translative and affine models are shown, with neglectable differences. Fig. 3.9 refers to a detail of the mosaic obtained with translative model without (c) and with (d) the application of the flat field correction stage. In the first case, it can be clearly noticed the presence of seams due to shading and vignetting effects. These artifacts disappear when the flat field correction is applied.

In Fig. 3.10 (a) the resulting tonally-aligned mosaic for the projective model is shown. It can be noticed how, for the projective model, error propagation leads to large, unnatural, image deformations.

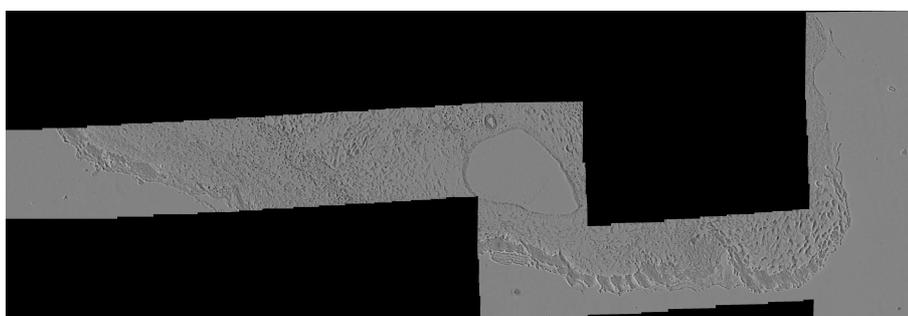
Fig. 3.10 (b), (c) confirm that it is due to the ill-conditioned nature of the estimation of the projective model, being the data used for model estimation small in number and concentrated in a small region.

As far as the quality metric indexes are concerned, Table 3.2 reports the results of the global mosaic indexes $RMSE_M$ and SNR_M (expressed in [dB]) relative to the sequence HS, without (HS-RAW) and with (HS-FF) the application of the flat field compensation.

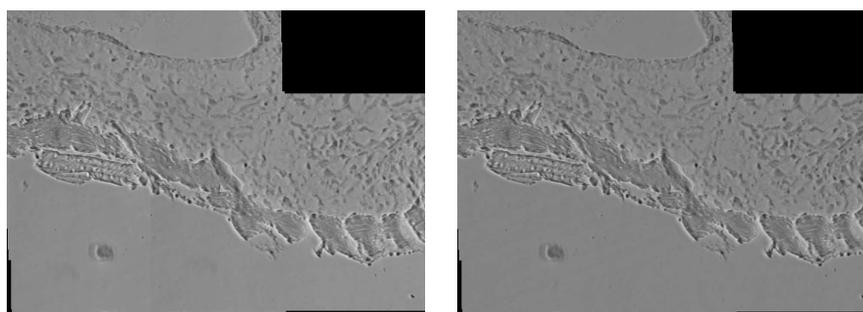
3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY



(a)



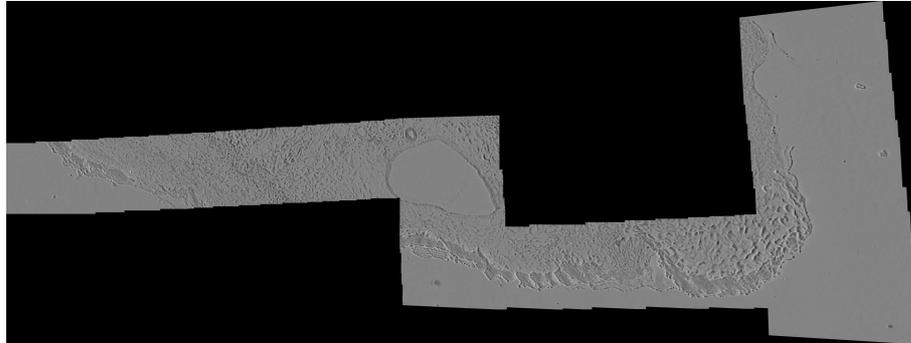
(b)



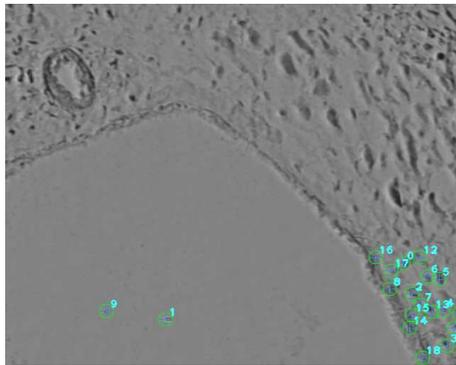
(c)

(d)

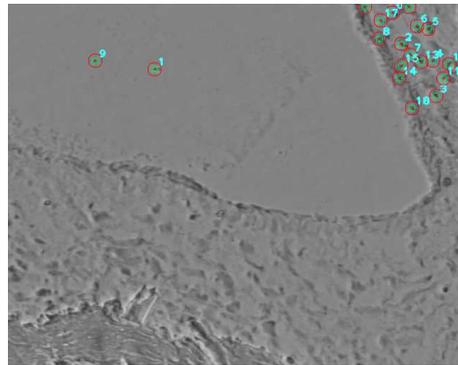
Figure 3.9: Mosaics obtained from the sequence HS using F2F registration with flat field correction for the translative (a) and affine models (b). Detail of the mosaic achieved without applying the flat field compensation ((c)) compared with the correspondent tonally-aligned region (d). Seams due to vignetting and shading effects can be seen (better on the electronic version) in (c), while disappear in (d) thanks to the flat field correction.



(a)



(b)



(c)

Figure 3.10: (a) Mosaic obtained from the tonally aligned sequence HS using the projective model. (b), (c) An image pair with a small number of features concentrated in region. This configuration can make the model estimation an ill-conditioned problem for complex models.

To evaluate consistently the effect of the flat field compensation only, the metric indexes can be compared by row once the model (column) has been selected. In other words, it is not consistent to compare directly indexes before and after the flat field correction for different models, since in this case both the geometric and photometric aspect would affect the final results. Accordingly, first we select a model, and evaluate the effect of the illumination field correction on that model. As in can be seen, fixed a model (column), by applying the flat field correction the RMSE value decreases, for all the three models. Conversely, as expected, the SNR increases when the flat field correction in applied. This is consistent with the improved visual quality of the mosaics

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

Sequence	$RMSE_M$			SNR_M [dB]		
	Trasl	Affine	Proj	Trasl	Affine	Proj
HS-RAW	4.77	4.36	10.21	29.13	29.91	20.51
HS-FF	2.72	1.99	8.62	33.52	36.22	24.78

Table 3.2: Quality metric indexes for the 60-frame HS sequence. The first row concerns results obtained working on raw (not tonally compensated) images, while values in the second row refer to flat field normalized images.

due to the flat field correction stage, for all the adopted models. Considering the models separately, it can be seen that the RMSE *gain* (i.e. its decrease) due to the application of the flat field normalization assumes comparable values for the translative (about 43%) and the affine models (about 54%), while the RMSE value for the projective model is less sensitive (with a gain of about 16%) to the flat field correction. In fact for the projective model the main contribution to the error is reasonably due to the geometric distortion introduced by this model. The SNR gain values (i.e. its increase) are about 15%, 21%, 21%, for the translative, affine and projective models, respectively. Thus, these values are still comparable not only for the translative and affine models, but even for the projective model. For this model, with respect the other two models, to a smaller RMSE gain corresponds a comparable value of the SNR gain, so that necessarily an higher signal intensity gain (related to the numerator in Eq.3.15) must compensate for that. Accordingly, the SNR gain is less sensitive to the model, that is the appearance of the mosaic built according to each of the models is “improved” by the flat field correction in the same way for all the models, each of them introducing its own geometrical distortion independently from the tonal compensation.

In order to evaluate consistently the *soundness* of the selected model with respect to their geometric distortions, the values of the metric indexes for the three models must be compared *once* the flat field correction has been applied, that is comparing values on the HS-FF row only. Considering the RMSE in the HS-FF row, for both the translative and affine models the corresponding values are quite small, both in absolute terms and in comparison with the gray level ranges of the images (0-255). For the projective model, the RMSE values are bigger, due to the large image distortion introduced by using this model. The SNR values show, as expected, an opposite trend,

Sequence	$RMSE_M$			SNR_M [dB]		
	Trasl	Affine	Proj	Trasl	Affine	Proj
SC-1-RAW	4.03	3.78	5.32	29.44	30.01	27.05
SC-1-FF	2.68	2.31	3.42	32.33	33.60	30.21

Table 3.3: Quality metric indexes for the 26-frame SC-1 sequence. The first row concern results obtained working on raw (not tonally compensated) images, while values in the second row refer to flat field normalized images.

decreasing for the projective model. Accordingly, in absolute terms the translative and the affine model performances are very similar (the latter being slightly better), and worsen, as hypothesized, for the projective model.

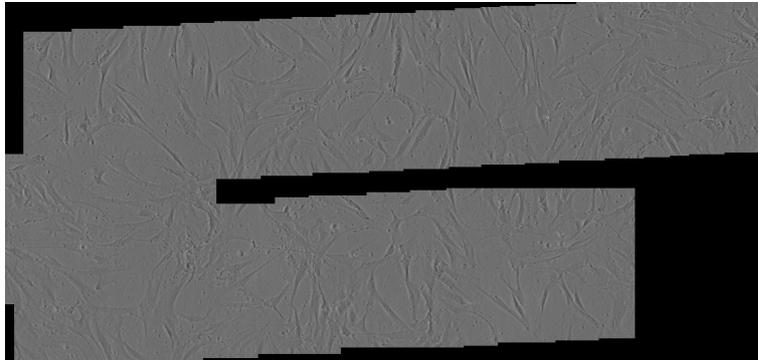
The second sequence is related to 26 images of a live stem cell culture (SC-1). It is characterized by different image content and different degree of overlapping of the frames. In Fig. 3.11 the mosaics obtained by registering these frames with the three models in the F2F mode are shown.

In Table 3.3 the global quality indexes $RMSE_M$ and SNR_M for the mosaics obtained in the F2F mode using the three geometric models, without (SC-1-RAW) and with (SC-1-FF) the application of the flat field compensation, are reported, in order to propose a comparison with the ones obtained with the first sequence.

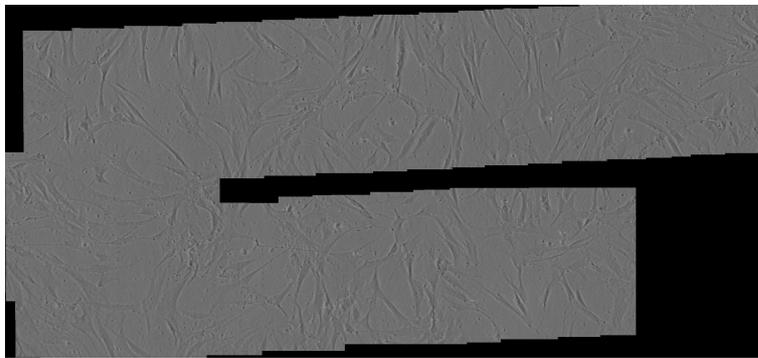
Considerations similar to the ones made for the HS sequence can be made. Chosen a model, the flat field correction yields, as expected, a decrease of the RMSE value and an increase of the SNR value also for the SC-1 sequence. It can be noticed that corresponding values of both the metric indexes in Table 3.3 and Table 3.2 are very similar, except for the projective model. In fact, for the sequence SC-1 the model estimation is more numerically stable, since the overlapping area percentage is at least about the 30% of the image size. In this case a slighter perspective distortion can accumulate during the pairwise registration, yielding, accordingly, “better” values of the metric indexes for the projective model with respect to the sequence HS.

Similar observations can be made also for the RMSE and SNR gain due to the illumination field correction. As far as the RMSE gain is concerned, the values for the translative, the affine and the projective models are equal to about 33%, 38% and 36%,

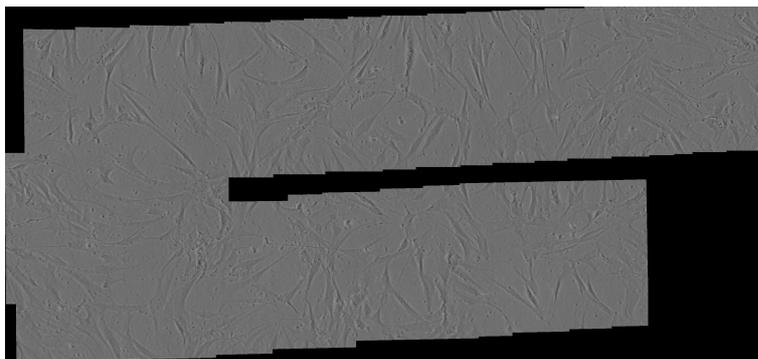
3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY



(a)



(b)



(c)

Figure 3.11: Mosaics obtained registering the tonally compensated SC-1 sequence in the F2F mode using the translative (a), affine (b) and projective model (c), respectively.

respectively. The corresponding values for the SNR gain are equal to about 10%, 12% and 12%, respectively. Accordingly, the gain values for both the metric indexes are quite independent from the model. This behavior has been discussed previously for the *SNR* gain. For this sequence, the reduction of the RMSE brought by the illumination field compensation for the projective model is comparable to the values obtained for the other two models, since the geometric distortion in this case is not too severe with respect to the other two models (see Fig. 3.11). In other words, the geometric registration performance of the different warping models being similar, the flat field compensation affects the mosaic quality in the same way for all the warping models. On the other hand, when high geometrical misalignments are present, like occurs for the projective model in the HS sequence, the corresponding poor quality performance can not be compensated by the tonal alignment.

By comparing the quality indexes models on the row SC-1-FF, that is after the illumination field correction, the affine model is the “winner”, as happens also for the sequence HS, having the lower value for RMSE and the higher for SNR.

The third sequence is built adding 7 frames to the sequence SC-1, with the purpose to form a looping path sequence of 33 images (SC-2). For this sequence, the fifth and the last frames coincide. This sequence has been designed since we want to compare quality indexes for different warping models with Frame-to-Frame and Frame-to-Mosaic registration approaches. Since this comparison concerns the geometric registration, it can be achieved by analyzing the results related to the tonally-aligned configuration, still being valid the considerations assessed above for the flat field compensation. In Fig. 3.12 (a) the resulting mosaic for the affine model is shown. Details obtained from the mosaic achieved by using F2F registration only and by applying the F2M registration are illustrated in Fig. 3.12 (b) and (c), respectively. As it can be seen, seams due to mis-registration of images of a revisited scene are corrected by *re-synchronizing* them with the mosaic built at the current epoch.

Table 3.4 summarizes the quality indexes $RMSE_M$ and SNR_M values achieved for sequence SC. It can be observed that, chosen a model, the F2M registration performs in general better than F2F registration, in terms of a lower RMSE and a higher SNR. Regarding the choice of the model, the translative and affine models show values of the quality indexes that are comparable, and better quality when F2M is applied. This is

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

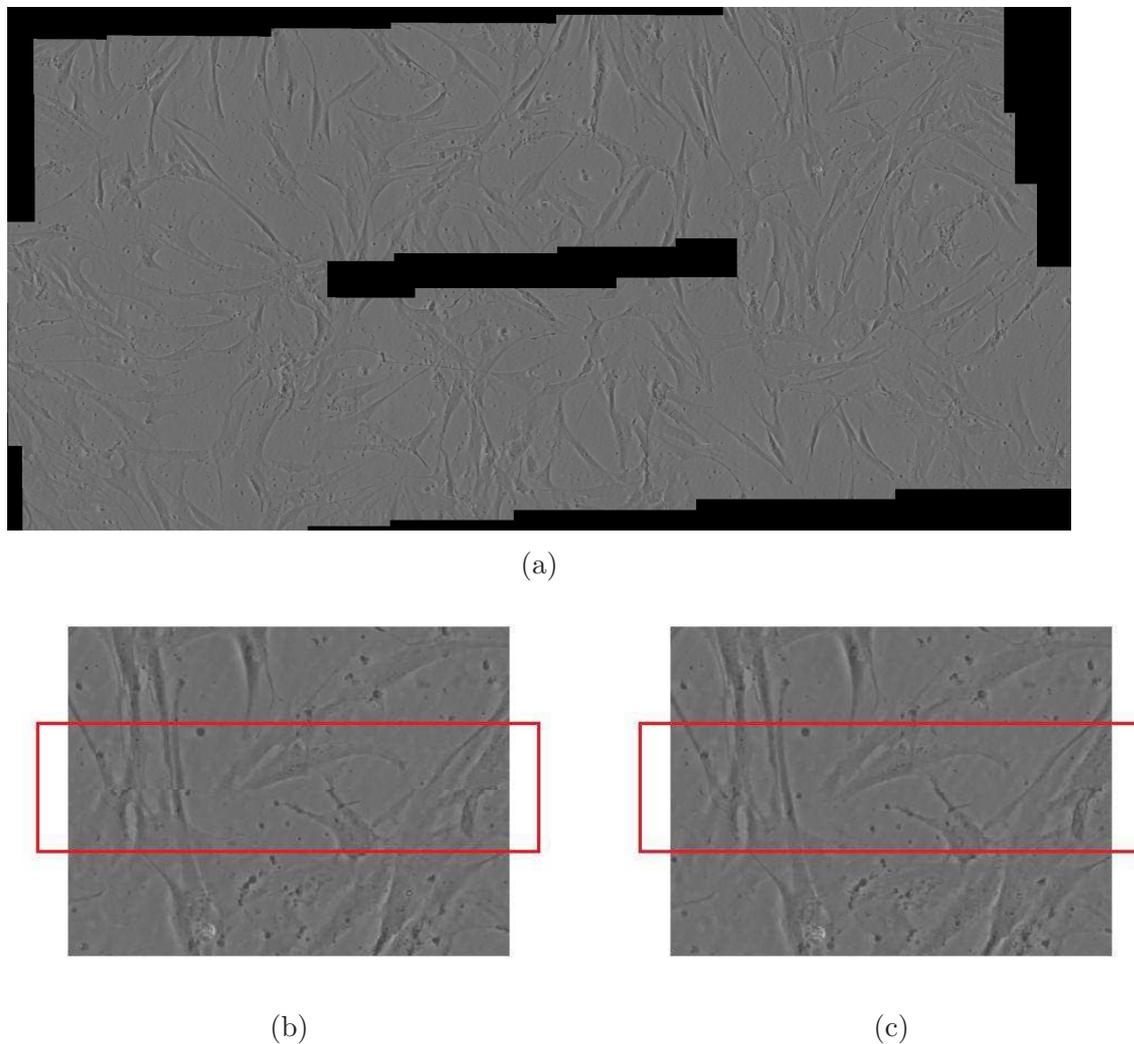


Figure 3.12: (a) Resulting mosaic for the affine model, employing flat field normalization and F2M registration on the sequence SC-2. (b), (c) Details of the mosaics obtained with F2F registration only and with the application of F2M registration. At the center of the red boxes seams due to the geometric drift (b) and their compensation due to F2M registration (c) are clearly visible.

particularly evident with the affine model, for which the RMSE decrease and the SNR increase are due to the compensation of the geometric misalignments, as it can be seen also comparing by visual inspection Fig. 3.12 (b) and (c). For the projective model, in general the overfitting effect can affect even the F2M registration, compensating

Method	$RMSE_M$			SNR_M		
	Trasl	Affine	Proj	Trasl	Affine	Proj
SC-2:F2F	2.63	3.96	9.24	32.51	29.02	21.61
SC-2:F2M	2.59	2.56	8.90	32.63	32.74	21.93

Table 3.4: Values of the global quality metric indexes of the SC-2 sequence for the three different models (columns), without (first row) and with (second row) application of F2M registration.

only partially or even worsening the metric performances. For this reason the next considerations are referred only to the first two models, capable to “fulfill” the physical constraints of the acquisition system and thus less prone to model misclassification, referring to Sect.3.4.2.1 for further observations.

Being the SC-2 composed of 33 images, 32 values of RMSE and SNR can be computed for the first 32 reference images, the last not giving contribution, as discussed in the previous paragraph. In Fig. 3.13 these values are plotted. It can be noticed that, for both the models and the metric indexes, two main behaviors can be identified, observing the values before and after the tenth frame. In fact, the first tenth frames are revisited, with different overlapping area percentages, when closing the loop. For these frames, for both the models, the application of the F2M registration yields a reduction of the RMSE value and an increase of the SNR value, accordingly, with respect to the F2F registration mode. This is especially evident in the fifth frame, that closes the loop. This *measures* and thus assesses the geometric compensation supplied by the F2M mode. On the other hand, after the tenth frame the quality indexes have comparable values between the two registration modes, as expected, since in this interval the F2M registration is done on not-revisited areas and accordingly it does not take advantage of the information contained in the mosaic.

3.4.2.1 Model Selection: Discussion

To discuss the model selection issue, GIC values for the three models have been computed using in our tests $\lambda_1=\ln(r)$, $\lambda_2=\ln(rn)$, being r the dimension of the data. Since features correspondences are expressed by couples of bi-dimensional coordinates, $r=4$. The dimension d of the model has the same value $d=2$ for the three models. The resid-

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

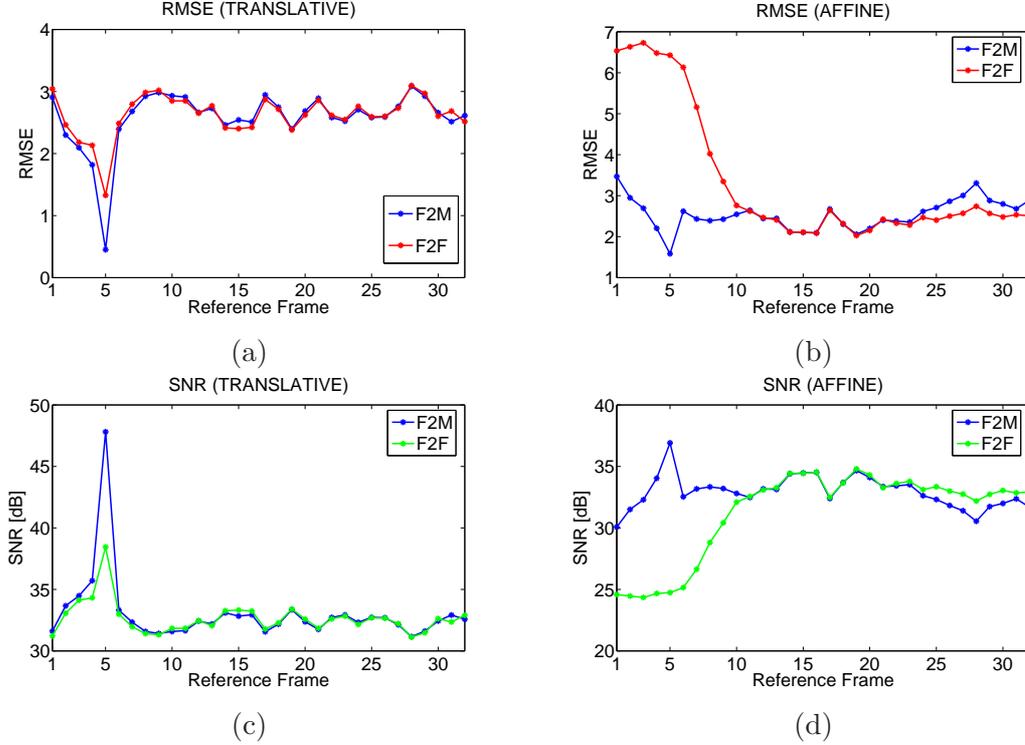


Figure 3.13: (a), (b) Values of the RMSE index for each reference frame for the translative and affine models, respectively, for F2F e F2M registration of the sequence SC-2. (c), (d) Values of the SNR index for each reference frame for the translative and affine models, respectively, for F2F e F2M registration of the sequence SC-2.

uals e_i are computed for simplicity in a Least Square mode (see (21)) over the set of inliers returned by RANSAC. Using these values for the weights in Eq. 3.5, GIC values for the sequence HS and SC-2, respectively, registered according to the F2F mode, are shown in Fig. 3.14.

According to these results, model selection would privilege in almost all the cases the translative model for both the sequences, since in both the cases, even with a better fitting for more complex models (that is, a lower value of the residuals), the corresponding increase of the number of parameters would make the GIC score increase. For the SC-2 sequence (Fig. 3.14 (b)) this is consistent with the values of the metric indexes (Table 3.4) obtained with the translative model, that can be considered globally the best. Moreover, the score distances for the three models show a quite robust margin, being the scores sufficiently “separable” for the three models in this case. On the other

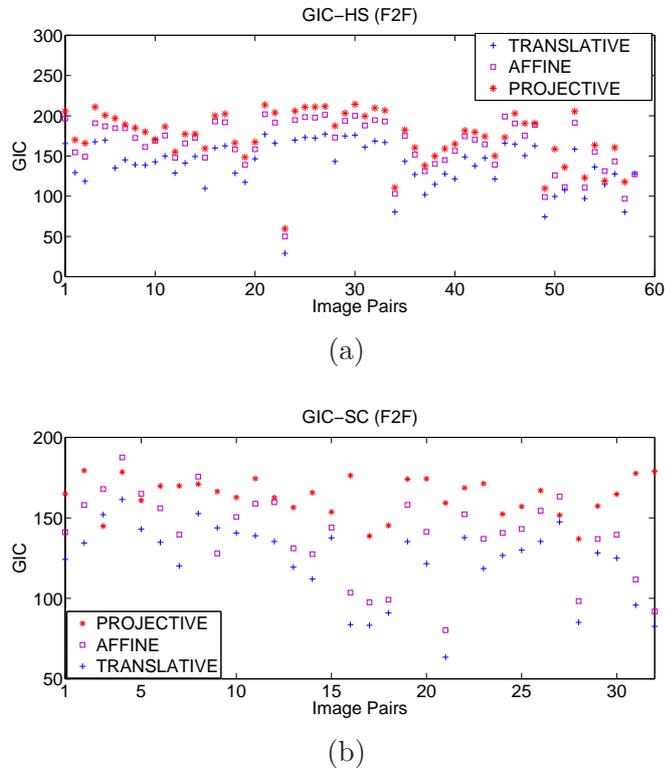


Figure 3.14: GIC values for the HS (a) and SC-2 (b) sequences registered according to F2F mode.

hand, for the HS sequence, according to the quality indexes (Table 3.4) the affine model would be the best, with a noticeable worsening for the projective model. Yet the corresponding GIC scores show (Fig. 3.14 (b)) a quite similar behavior between the affine (the “best”) and the projective model (the “worst”), which make this model selection criterion not robust in this case. For this reason, we are currently proceed toward researching a more robust score function that could take into account also the spatial distribution of the features and weight with a different law the model complexity.

3.4.3 Computational Performance

As far as the computational performances are concerned, the algorithms works on a consumer PC (Intel Pentium 2GHz, 4GB RAM), processing 640×512 images stored on the hard disk at about 2 – 3 frames/second with gray level images. In particular, the main computational payload is due to the Phase Correlation stage, taking approxima-

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

tively 150 – 250 ms, while the remaining 150 – 250 ms are employed for the matching and warping stages. For color images, the warping stage can be heavier, slowing down the temporal performance to approximately 1 frame/second. Since at the moment the algorithm is implemented on a research prototype, we are quite confident that a further optimization stage can be applied to improve on the computational performances, towards *real time* performance with respect to the user reaction time. Moreover, for color images at the moment we are working with the HSV (Hue-Saturation-Value) space, performing flat field normalization on the V channel. This approach works well for “flat images”, that is characterized by similar values of the RGB channels, for which the hue does not change too much. In presence of high variations of the hue and in saturated areas, “false colors” can be generated. We are still working on the flat field normalization and warping stages to try to compensate for these undesired effects.

3.5 Conclusions

In this Chapter we have described the application of our registration algorithms to a scene reconstruction application, the mosaicing of optical microscopy images. By widening the field of view of the microscopy system, the life sciences researchers are provided with a global view of the whole cell culture, and can better identify and localize patterns of interest *during* their inspecting working session. The strategies adopted to preserve the tonal and geometrical consistency of the scene have been presented, discussing the problems encountered. Compensation of uneven illumination in quite controlled condition has been faced by studying the illumination field at system level, without analyzing the single dynamics responsible for this artifact. While working well for images of uniform hue, “false colors” artifacts are still present and proper compensating strategies are currently under study. Geometrical registration approaches have been discussed in order to prevent drift effect that can accumulate in severe distortion, while preserving on-line performance, using an incremental registration approach. To assess the soundness of the methods devised to cope with these issues, on-purpose quality metric indexes has been devised.

The effectiveness of our approach has been proved through experiments on different biological samples, a histological sample and a middle density cell culture. In fact, results about the mosaic quality metric indexes confirm the positive effect of the flat field

normalization and of the Frame-to-Mosaic registration mode. However, the robustness of the geometrical model selection must be improved.

These promising results are driving us to conduct experiments on a wider range of different biological samples, with different cell density, appearance, etc. The robustness of the matching stage in very high-density cultures must be assessed, being likely that in these cases more robust descriptors might be necessary to avoid a high number of mis-matched features.

It is worth to remark that this application has been carried out in the context of the ADVANCE (Automatic system to Determine Vital AdhereNt CELls) project, involving the Computer Vision Group at ARCES - University of Bologna and the Bone Regeneration Laboratory (BRL), Istituti Ortopedici Rizzoli, Bologna, Italy. The algorithms described here have been integrated in a more complex prototype research for microscope image analysis and data management, that is currently under validation by the researchers at BRL.

3. ON-LINE MOSAICING OF OPTICAL MICROSCOPE IMAGERY

Chapter 4

Camera Attitude estimation: the STARS project

In this Chapter our registration algorithm is applied to estimate the attitude of a remote sensing satellite from the Earth images acquired during its orbit. This application have been carried out in the context of the STARS project (Standalone Three-Axis spacecraft oRientation Sensor) in partnership with the Aerospace Engineering Group at DIEM, II Faculty of Engineering, University of Bologna.

The Earth images are exploited as a “native target” in order to recover the three dimensional orientation of the camera with high accuracy. Initially, this novel approach is discussed within the ambit of the state of the art in this field. The working hypotheses of this application are outlined and accordingly a geometric model is formulated. Then properties of the image registration algorithm have been discussed based on the theoretical model, building accordingly a simulation framework. Finally, simulations and experiments designed on purpose provide useful hints about the error sources, performance and some compensation strategy of our approach.

4.1 Introduction

The automatic guidance and control of remote sensing devices in different fields (marine, aerospace, biomedical, etc.) represents a complex and challenging task, often involving the intelligent integration of a great variety of information. Applications of computer vision to these fields are nowadays standing out for the advantages they can carry. The employment of a CCD camera together with a processing unit can replace

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

ensembles of different sensors which would require a subsequent data integration. For example, navigation systems based on *visual odometry* have been conceived to achieve measurements of the motion parameters of the controlled object. This problem has been addressed using ego-motion estimation, where the sensor's pose in an inertial reference frame is recovered by processing images of the surrounding environment acquired from one or more cameras mounted on-board. Usually motion parameters are required to be estimated with a certain order of accuracy, strongly depending on the application's purposes. Accordingly, due to these requirements, an in-depth analysis of error sources and their dynamics can represent a fundamental step to devise some compensation strategy.

In the following paragraphs, we describe a novel vision based approach to cope with a mission critical application, the three-axial attitude estimation of a Low Earth Orbit (LEO) satellite for autonomous guidance purposes. It exploits visual information, extracted from terrestrial images acquired by a camera integral with the satellite, to estimate satellite orientation from their registration. The attitude parameters are recovered from the geometrical transformation that maps views of the Earth acquired at different epochs along the orbit. For such an application, where high accuracy is required, the methodological approach adopted is crucial. In general, the theoretical model employed to describe the system, as well as the algorithms chosen to estimate the quantities of interest, can affect the global application performances. For that reason, it is not less important to identify the different sources of error and evaluate how they can affect the resulting overall accuracy, so to devise proper corrective strategies. Starting from the hypothesis introduced by the theoretical model, a simulation framework can be designed to detect, at a first stage, the most important critical aspects. To this purpose, specific simulations configurations can be designed to identify the "weight" of each error components on the overall performance. Finally experiments are used to reject or confirm the hypothesis previously arising, and consequently to assess the soundness of the design strategies devised.

The structure of this chapter traces out this approach. Sect. 4.2 deals with other approaches previously employed in this field. In Sect. 4.2.1 the working hypotheses and the requirements of the problem are illustrated. The novelty of our approach and the advantages brought are then discussed in Sect. 4.3. This paragraph outlines the methodological stages we have gone through for assessing the proof of concept of this application. In particular, physical and geometrical models adopted for the sensor and for the scene (the Earth) are described. From the overall model the equations to recover the satellite attitude from image correspondences are also derived. In Sect. 4.4

the peculiar features of our image registration algorithm in this context are discussed in details with respect to other approaches. In Sect. 4.5 an in-depth error analysis is carried on to describe the influence of different sources of error. Sect. 4.7 reports the results regarding simulations of realistic orbits on synthetic satellite images and experiments related to laboratory images. The discussion of the results in Sect. 4.7 introduces to the futures perspectives summarized in Sect. 4.8.

4.2 Previous Works

Attitude estimation for the automatic guidance of remote unmanned systems has been generally faced by researchers by using specific devices and sensors in terrestrial or aerial applications and space missions. Most of the currently available navigation systems based on attitude estimation follow two different approaches.

In the first approach, a combination of sensors is employed. In terrestrial and aerial applications, GPS, accelerometers and gyroscopes are widely employed. For satellite missions, where estimation accuracy is a key requirement, orbital gyrocompass and horizon sensors are used to estimating pitch, roll and yaw attitude angles. However, accuracy is limited to tenths of degree (68). Furthermore, in Low Earth Orbit (LEO) system, Earth's horizon appears as not being perfectly circular and infrared radiation deflected by the atmosphere can lead to false detections, that worsen accuracy as low as few degrees (68).

The second approach exploits vision based methods. In particular, the Simultaneous Localization and Mapping (SLAM) approach has been conceived to cope with dead reckoning effects for long looping image sequences by reducing drift errors. The research presented in (69) well resumes the state of the art, where a combined approach of SLAM and visual odometry is proposed to reduce the impact of cumulative errors. Although attitude estimation is also provided, it shows a limited accuracy (some degrees). In (70), sequences of monocular aerial images acquired in real time are compared with a geo-referenced Earth image database to estimate position and velocity of the aircraft. This also helps to reduce drift errors in position estimation. However, no results are reported for the attitude, although authors state to recover it. In addition, processing large database images could be not compliant with the limited resources available on board of a satellite. However, these approaches aim mostly at guaranteeing robustness in looping path sequences, while accuracy is left in the background, since it does not represent a key issue for these applications. Accordingly, error analysis is not carried out in depth in order to increase the accuracy.

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

As far as it concerns specifically satellite systems, *star trackers* constitutes the state-of-the-art sensors to be employed for accurate attitude estimation. They aim at estimating the three-dimensional orientation of the aircraft by matching stars from acquired pictures of the celestial sphere with a referenced database of stars (*star field*) stored on board. They generally suffer from Sun and Moon light blinding, so that they are generally employed in couples in a counter-phased set-up and require specific cooling plants. This can bear on the global power consumption, which constitutes a compelling feature for such power budget-limited systems. In particular, the authors in (71) report a fully autonomous star tracker capable of achieving an angular estimation accuracy below 1 arcsec. The system is able to operate in different modes depending on the environmental conditions. However, only indoor testing has been reported and the accuracy strongly depends on the quality and number of matched stars. Also in (72), sub-pixel centroid estimation is achieved by matching the camera Point Spread Function (PSF) with the star's pixel values. The overall accuracy is influenced by the centroid estimation as well as the probability of correct star identification. The authors report an accuracy for attitude estimation of about 2 arcsec for both synthetic and real (urban scene) images. Nevertheless, stars trackers are not immune to problems, as reported in (73), where the radiation impact is analyzed. The testing of stars trackers in real working condition has been recently reported in (74). The 3-sigma accuracy along the three axis is less than, or equal to, [7 7 70] arcsec, respectively. Here, also the effects of the satellite motion on performance, stray light and direct Sun blinding are discussed, analyzing in which conditions the recognition will fail.

4.2.1 Working Hypotheses

Satellites are usually equipped with remote sensing devices, like imaging sensors, and they are oriented in *nadir* pointing attitude, that is one axis always looks at the Earth during the orbit. Our approach aims at estimating the three-axial orientation (attitude) from Earth images acquired by a calibrated CCD camera integral with the satellite, its position supposed to be known (generally provided by other devices). In such an application, strict requirements must be fulfilled in terms of accuracy and computational performances. What follows is a brief resume of the main constraints related to our target application.

- LEO satellites have a typical orbital radius between 650 – 700 km and a ground velocity which varies from 5.5 to 7.5 *km/s*, covering about 15 circular non-geostationary orbits per day, with a period of about 94 *min/orbit*, about 30

of which in eclipse condition. These factors can affect the camera frame rate, the exposure time to be chosen and the design of the optical layout of the sensor;

- limited computational resources are available on board, generally one-older generation hardware architectures (486 μ P family);
- limited available power: this can play an important role on the choice of sensors that can be mounted on board, their sensing elements and the need of on-purpose cooling devices;
- limited available volume, with direct consequences especially on the choice of the optics layout.

The motion velocity has to be taken into account since it can affect the image quality (e.g., because of blurring). The working frame rate is also important for the feedback frequency in the spacecraft control, since iterative algorithms for refining the estimation can be too “heavy” to be supported from the HW available on mini satellites. Therefore, in the choice of the algorithms for attitude estimation, the computational complexity must be kept under control.

4.3 A novel vision-based approach

We have proposed a novel vision-based approach that had not ever been explored before to determine the satellite attitude. According to this approach, the satellite three-dimensional orientation is derived by analyzing the geometrical relation between couple of subsequent, and partly overlapping, images of the Earth taken by a camera integral with the orbiting satellite. The aim of our algorithm is to estimate, from the set of correspondences established by the image registration algorithm on these image pairs, the satellite orientation parameters in an appropriate reference frame, and in particular their deviation with respect to nominal (i.e. *stabilized*) values. Our approach requires that several aspects are investigated, stemming from the model used to describe the problem, up to the mathematical methods employed to solve it and their algorithmic implementation.

4.3.1 Physical and geometrical model

The Earth is modeled as a sphere of known radius ($R_E = 6371$ km), rotating at the constant velocity ω , taken equal to the Earth mean angular velocity $\Omega_E = 7.272 \cdot 10^{-5}$ rad/s. The camera is modeled as a full projective pinhole camera, supposed to be calibrated

perpendicular to the satellite orbital plane in the direction that sees the satellite moving in the counter-clock-wise way (\vec{H}_{LORF} in Fig. 4.1) and the x-axis, such that the system forms a right-handed triad – in case of circular orbit the x-axis is directed like the satellite velocity vector ($\vec{\Theta}_{LORF}$ in Fig. 4.1). This reference frame depends on the *orbital parameters* of satellite trajectory, that is the right ascension O and the inclination I (see Fig. 4.1). The satellite attitude should be stabilized towards the target *nominal* configuration aligned with this reference frame;

- the Satellite (body) reference frame (SRF): SRF is integral with the satellite, with its origin in the satellite centre of mass and axes nominally aligned with LORF. The displacement in orientation from LORF (which represents the nominal attitude) is measured as roll, pitch and yaw angles;
- the Camera reference frame (CRF): CRF is integral with SRF, with the optical centre of the camera coinciding with the satellite centre of mass (or being at a known displacement, without loss of generality);
- the Image reference frame (IRF) is related to CRF through the matrix of the camera intrinsic parameters (known after camera calibration).

As it can be seen from Fig. 4.1, a roll perturbation corresponds to camera panning, while a pitch perturbation leads to camera tilt. The first four reference frames are commonly used in the aerospace field of satellite navigation (75) and here reported for completeness. From the motion displacement estimation in the IRF the values of deviation from the nominal attitude should be expressed in the LORF (or SRF). However, generally in satellite navigation the current attitude is referred to an inertial reference system, the ECI, thus requiring the chaining the geometrical transformations linking the intermediate reference frames.

Under these assumptions, given a point on the scene (i.e., the Earth), with assigned geographic coordinates (latitude and longitude in the ECEF), it *corresponds* to different coordinates in the IRF in two consecutive image frames, according to the motion field of the satellite. In general, two views of the same scene under general camera rigid motions are linked through the well-known relation (Eq. 4.1 (34)):

$$\begin{bmatrix} x' \\ y' \\ 1' \end{bmatrix} = KR'R^{-1}K^{-1} \left(\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} - \frac{KR(C' - C)}{Z_{cam}} \right) \quad (4.1)$$

where:

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

- $[x \ y \ 1]^T$ and $[x' \ y' \ 1]^T$ represent the homogeneous coordinates of the *corresponding points* (i.e. related to the same point on the scene) in the previous epoch t and in the current epoch t' , respectively;
- K is the matrix of the intrinsic parameters of the camera. In our application, K is supposed to be estimated once and for all through an off-line calibration process during an earlier sensor characterization procedure;
- the rotation matrices of CRF with respect to the *world* (scene) reference frame are represented by R and R' in the previous and the current epoch, respectively;
- the symbols C and C' stand for the three dimensional coordinates of the camera optical centre expressed in the *world* reference frame in the first and in the second epoch, while Z_{cam} is the component along the z-axis of CRF of each of the projection rays of the image points in the first epoch. The ratio in Eq. 4.1 forms the *parallax shift* term which must be subtracted from the coordinates of the first image reference frame. The difference in camera optical centre due to the translation of the satellite is responsible for the *parallax effect*, whereas the term Z_{cam} takes into account the three-dimensional structure of the scene. Given satellite position and orientation and being known the model of the Earth, it is possible to compute Z_{cam} for every pixel of the camera by intersecting the visual rays with the terrestrial sphere.

Accordingly, after the *parallax shift* term compensation, corresponding points in the two image reference frames are related through a plane homography H_∞ (the *infinite homography*) retaining the attitude variation $R'R^{-1}$ (Eq. 4.2):

$$H_\infty = \lambda KR'R^{-1}K^{-1}, \quad \lambda \neq 0 \quad (4.2)$$

The parameter λ can be retrieved from Eq. 4.2 once H_∞ has been estimated, yielding Eq. 4.3:

$$\begin{aligned} \det(H_\infty) &= \lambda^3 \det(KR'R^{-1}K^{-1}) = \lambda^3 \det(K) \det(K^{-1}) \\ &= \lambda^3 \end{aligned} \quad (4.3)$$

being the rotation matrices orthogonal.

4.3.2 Attitude parametrization

We have adopted a convenient parametrization for representing the aircraft attitude. Let the rigid-body motion group denoted as $SE(3)$. We represent the group of three

dimensional rotation $SO(3)$ using unit quaternions (76). This choice has been expressly carried out because of the specific advantages it introduces. In fact, unit quaternions have a group structure (S^3 from now onward), they are closely related to the intuitive axis-angle representation of rotations and the composition of rotations is performed by 4-components vector products, resulting in a more computational advantageous and numerically stable solution. Attitude estimation from noisy data usually yields non-orthonormal rotation matrices, which must be somehow re-orthogonalized, while quaternions need only to be normalized to the unit vector. For these properties, quaternions are usually employed in satellite dynamics (75).

Unit quaternions q are related to the rotation matrix R through an homomorphism $R(q) : S^3 \rightarrow SO(3)$. Usually, in satellite dynamics the rotation matrix associated with the satellite attitude quaternion q satisfies the relation (Eq. 4.4):

$$\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} \right)_{ECI} = R(q) \begin{bmatrix} X_{Sat} \\ Y_{Sat} \\ Z_{Sat} \end{bmatrix} \quad (4.4)$$

thus referring the satellite orientation to ECI. By using the Direction Cosine Matrix corresponding to the quaternion q (77), $DCM(q) = R(q)^T$, and taking into account the Earth rotation velocity ω , after some passages Eq. 4.2 can be rewritten as Eq. 4.5:

$$H = \lambda K DCM(q(t')) R_3(\omega(t' - t))^T DCM(q(t))^T K^{-1} \quad (4.5)$$

Therefore, if in Eq. 4.6 we set:

$$R(\Delta q) (= DCM(\Delta q)^T) \doteq \frac{1}{\lambda} (K^{-1} H_\infty K)^T \quad (4.6)$$

then the *relative orientation quaternion* Δq can be computed by the estimated homography H_∞ (77). Accordingly, starting from the estimation of the previous attitude quaternion $q(t)$, the current *absolute* attitude quaternion can be estimated using the relation expressed in Eq. 4.7:

$$q(t') = q_\omega(\omega(t' - t)) \circ q(t) \circ \Delta q \quad (4.7)$$

where (\circ) represents the canonical product between quaternions.

4.4 Image Registration

Our approach provides that image correspondences are established between image consecutive pairs and fed into the geometrical model described in Sect. 4.3.1.

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

In this applicative context, the image registration algorithm must be capable to *detect* with high accuracy *features* on satellite imagery yet being preserved scene generality, because of the variety assumed by terrestrial landscape. Once *extracted*, the features corresponding to the same scene point (i.e., the Earth surface) on images on two subsequent epochs has to be *matched* reliably. A common approach to the problem used in airborne image analysis (78) is to exploit native structured patterns, like mountains, roads, coastal lines, etc. However these high-level patterns are not always present on the scene and their detection is usually based of voting algorithms (e.g generalized Hough transform) that can be computationally intensive. Moreover, since the attitude perturbations between consecutive epochs involved in realistic orbits are not too severe, strong invariance to perspective distortions is not a strong requirement. Accordingly, robust local region descriptors, like SURF (17) for example, which privileges robustness in cluttered and “distorted” scenes over accuracy, have not been considered. Moreover, such *region* descriptors requires a further layer of “refinement” to provide a *punctual* sub-pixel accuracy.

On the other hand, sparse punctual features can be extracted directly from the natural textured patterns on Earth imagery with a low computational payload. It is possible to use high gradient points, that generally present good robustness to variation in image geometry and illumination, since they do not rely directly on image intensity, and generally works at sub-pixel accuracy. Their sparseness within the scene image permit to deal with partial occlusions of the scene (for example due to clouds). The algorithm described in Chapter 2 well suits these motivations. For each couple of frames to be registered, Shi and Tomasi feature points (26) are extracted each time from the first image. The prevailing motion component in our context is due to the motion of the satellite, and consequently features are no longer visible after a couple of frames. Accordingly, it is preferable to re-extract *stronger* features each time and track them only on the subsequent frame, temporally closer and then less prone to scene changes, thus better preserving the optical flow conservation constraint. Moreover, choosing small patches can retain more punctual information, but can be then less discriminative in very textured patters like Earth’s satellite images. In this context, the application of our *coarse-to-fine* approach permits to estimate the global 2-D translational components of the image motion field $(\Delta x, \Delta y)$ and then measure the residual *local* motion field vector at a sub-pixel accuracy for each feature.

After the correction with the parallax term (see Eq. 4.1), using the DLT algorithm (34) jointly with the RANSAC (40) outlier rejection method to remove *false matchings*, a robust estimation of H is achieved. Once the matrix H has been esti-

mated, it is used to perform a *guided search* using the *estimated* features estimated accordingly (that is *projected* according to H) as guess values for a further LKT tracking on the current image. Model is estimated again on the inliers (after RANSAC filtering) of this new, more robust set of correspondences. It is worth remarking that this registration algorithm does not need any model of terrestrial landmarks to be learned in advance.

4.5 Error Analysis

In order to assess our method, it is important to identify the main sources of error that can stem from such a kind of approach. In particular, approximations due to chosen models have to be at first investigated, considering the resulting performances as the best-case. Then, errors intrinsic to the image registration algorithm have to be evaluated with respect to this best case. Finally, the dynamic model underlying the satellite attitude estimation, based on sequential registration, is analyzed in order to characterize error propagation and devise proper compensating strategies.

4.5.1 Modelling Errors

At a first level, the mathematical models chosen to represent the acquisition system and the (world) scene introduce approximations that can affect the whole accuracy in attitude estimation. In fact, in the ideal condition of image *continuous* domain (i.e. sensors with infinite resolution) the attitude should be estimated *without errors* by computing the matching between point sets simply through ray projection between the couple of frames (and so generating the corresponding set of “*computed matchings*”).

The procedure used to find out these set of computed matchings can be summarized as follows:

- a sparse set $S(t)$ of n points p uniformly distributed is chosen in the first image plane (in case of discrete domains, they represent *pixels* and have integer coordinates);
- the *geo-referenced* projections of the image planes on Earth are found, taking into account the camera and the Earth model chosen. Practically speaking, points p are projected onto the spherical terrestrial surface through visual rays passing by the camera centre at the epoch t , so as to find on Earth the set $S(t)_E$ expressed in latitude and longitude;

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

- starting from each point of the set $S(t)_E$, the matching set $S(t')$ in the second frame is found *back-projecting* visual rays from the set $S(t)_E$ towards the camera centre at the epoch t , until they intersect the plane of the second image, so as to find directly the *matching coordinates*;

In the continuous domain, the errors in computing the homography are just the numerical ones, so that this situation represents the best case yielding an upper bound for the registration accuracy. However, the model described is only theoretical since in real working conditions the image domain is *discrete*, due to the nature of the sensor matrix, and necessarily the image formation process introduces quantization errors that must be taken into account.

To evaluate the effect of this error on the attitude estimation, simulations with geometrical model of the Earth and the camera sensor have been performed in both the cases. Related results are discussed in Sect. 4.7.1.

4.5.1.1 Parallax Error

The parallax term is due to the motion of the satellite along the orbit, with its position supposed to be known. With current technology, satellite position can be provided in real time using GPS/DGPS, with an accuracy within one meter. Its knowledge makes the estimation of the attitude parameters more robust and accurate (79).

As it can be seen in Eq. 4.1, the parallax term depends through a non linear relation on the previous attitude estimation and affects the homography estimation independently from the algorithm chosen for feature matching. Accordingly, by defining the parallax term as in Eq. 4.8:

$$f(q(t)) = \frac{KDCM(q(t))(C' - C)}{Z_{cam}(q(t))} \quad (4.8)$$

Eq. 4.7 becomes:

$$q(t') = q_\omega(\omega(t' - t)) \circ q(t) \circ \Delta q(S(t, t'), f(q(t))) \quad (4.9)$$

where $S(t, t')$ represents the two sets of corresponding points in the two views at consecutive epochs, these depending on the registration algorithm. The registration contribution can be decoupled performing a simulation by annulling the motion of the satellite, but leaving the attitude free to vary. This procedure is employed in conjunction with Eq. 4.9 to study the error propagation, in Section 4.5.3. The impact of this contribution on the overall accuracy is discussed in Section 4.7.

4.5.2 Image Registration Error

Combining the *matching* resolution among satellite image pairs of our image registration algorithm with the ground resolution achieved by the camera sensor permits to evaluate the accuracy reachable in estimating the attitude. The registration algorithm described in Sect 4.4, working with discrete images, shows necessarily a finite precision in matching corresponding features between two epochs. Since it aims at *tracking local gradients* with high accuracy, the image formation process can play a key role on the resulting performance. Building *a priori* a model of the contribution of the image registration error is not a feasible task. A method to overcome this limitation is to assess the registration accuracy experimentally. When evaluating the accuracy of the method, the experimental setup introduces further sources of error which cannot be left out of consideration to achieve a reliable measure of the system accuracy. In particular, due to the lack of raw satellite image sequences acquired in real conditions, the evaluation of our attitude determination system has been carried out using two sets of experiments, using respectively:

- **Synthetic images.** Synthetic images are *generated* from a compatible database of satellite images of the Earth. They present artifacts but work on the *real operating* scene and environment necessary to test the capabilities of feature extractor and tracker. Tracking errors can be caused by image *gradient artifacts*, with tracking resolution being limited by the quantization error intrinsic in the discrete nature of the sensor. To study the impact of this contribution, earlier simulations have been carried out over “synthetic” images *generated by sampling* a geo-referenced Earth imagery dataset. More in details, geo-referenced views of the Earth acquired at different attitude along the orbit are sampled from the database by *projecting* the camera’s Field Of View (FOV), computed at known position and spatial orientation, over the corresponding *geographic* area in the database. Realistic ground truth data for the satellite pose along its orbit are provided by the orbital simulator described in Sect 4.6. The state of the satellite imaging sensor is *geo-referenced* at each epoch, that is the corresponding *maps* of latitude and longitude on Earth are computed for each pixel of the sensor, taking into account the camera model (projective) and the Earth model (spherical and rotating at a known velocity) chosen. In this way, for each point in the current image reference frame, the corresponding latitude and longitude *viewed* on Earth by each sensor location can be achieved. To generate the corresponding terrestrial image, a magnitude value (i.e. gray level) must be assigned to each sensor

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

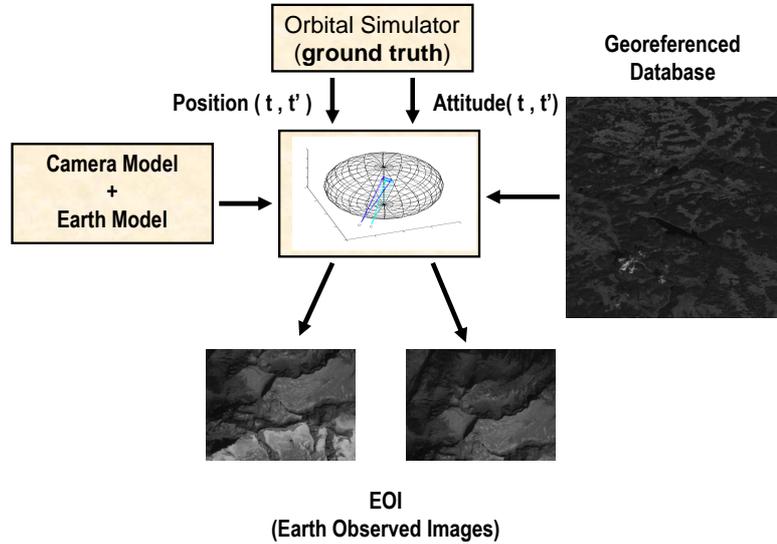


Figure 4.2: Diagram of the procedure used to generate the EOIs sequences. The ground-truth pose data are used to geo-reference the camera field of view, assigning to each pixel the geographical coordinates of the points seen on Earth. Image interpolation on a geo-referenced satellite database is finally used to generate the EOIs couples to be registered.

pixel according to the values of the image database *tile* that covers the sensor geographic map built in this way (at least, the camera geo-referenced FOV). This procedure is summarized in Fig. 4.2 and generate, for each orbit, a corresponding sequence of Earth observed images (EOI). Since the image model provided by the satellite image database is *discrete* by definition, image interpolation is used to sample the geo-referenced tile in correspondence of the sensor geographic maps. The interpolation technique chosen can affect the image generation process and the accuracy of the tracking algorithm, accordingly, due to the artifacts introduced by the sampling process (aliasing, edge halos, etc.).

However, it should be noticed that EOIs are generated from an already available database, obtained after the acquired raw images have undergone a certain number of processing stage (e.g., radiometric normalization, etc.). For this reason, the related simulations occur in quite optimistic conditions as far as it concerns the number of factors affecting the real image acquisition conditions in such a kind of application. The purpose of the test with the EOIs is to evaluate the registration algorithm provided that the acquisition conditions are properly compensated, for example through a complete characterization of the sensor, or through other

strategies however out of the scope of this work.

- **Laboratory testbed.** Acquiring sequences of images in real time from a real world environment permits to evaluate the accuracy (and therefore the error) in real operating conditions. At the current stage of the project, it has not been possible to assess our algorithm directly on the raw data really acquired by an orbiting satellite. We have tested our approach with images of similar scene content but acquired with a laboratory testbed. With respect to the test with EOIs, these tests occur in more realistic *acquisition condition*, being necessary to take into account also camera calibration, lens distortion, not fully controlled lighting conditions, etc. Furthermore, contrarily to what happens for synthetic images, although the image domain is discrete the object domain is represented by a *continuous* model. Therefore, sampling and interpolation errors due to synthetic images are left out of scope, although these kinds of errors still remain as intrinsic to the digital sensor acquisition system. However, a great error contribution in these cases can be due to the instrumentation required to capture the ground-truth. Whereas on the one hand with real time images the quantization error is kept as low as possible, on the other hand the ground-truth uncertainty is much higher. As we will see, simulations with laboratory images present a sort of “uncertainty principle” in which ground-truth accuracy and quantization noise play the main role.

4.5.3 Error dynamic

Our sequential approach is intrinsically affected by the propagation of all the errors typologies aforementioned, since the attitude computed at each epoch is fed into the attitude determination chain for estimation at the next epoch (according to Eq. 4.9). To this purpose, three different configurations for simulations have been devised to understand the impact of the different error sources to error propagation:

- in *NoUPD* configuration, the attitude estimation is not propagated, that is the attitude in the initial epoch coincides with the ground-truth attitude ($q^{true}(t)$), so that the error is due to the F2F image registration only (Eq. 4.10):

$$q^{est}(t') = q_{\omega} \circ q^{true}(t) \circ \Delta q(S(t, t'), f(q^{true}(t))) \quad (4.10)$$

- in *UPD* configuration, the attitude estimation is propagated, thus following this relation (Eq. 4.11):

$$q^{est}(t') = q_{\omega} \circ q^{est}(t) \circ \Delta q(S(t, t'), f(q^{est}(t))) \quad (4.11)$$

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

- in *spurious SP* configuration, the attitude estimated at the previous epoch is not propagated in the parallax term $f(q(t))$, so that the following relation (Eq. 4.12) holds:

$$q^{est}(t') = q_\omega \circ q^{est}(t) \circ \Delta q(S(t, t'), f(q^{true}(t))) \quad (4.12)$$

In the NoUPD configuration, the estimated attitude $q^{est}(t)$ is affected by the image registration error only. By analyzing this error it is possible to predict the attitude errors trends in dynamic working conditions, that is the UPD configuration. In particular, even considering a zero-mean normal distribution for the error of the attitude determination process from image pairs, this could determine a drift in the subsequent system evolution. The worst case condition occurs when the F2F error is modeled as a skewed distribution, that is a non-zero mean process. This systematic error can heavily affect the following attitude determination, since there is always a bias that is propagated to the subsequent attitude evaluations.

In other words, a bias on this error that moves its distribution away from null mean can cumulate into drift effects in the UPD configuration. This is evident when the attitude variation concerns almost one of the three angles, so that for rotation composition in this case commutative property basically holds and the error trend in the UPD configuration approximates the cumulative sum of the errors in the NoUPD configuration. However, even a non-biased temporal distribution of NoUPD pairwise errors can yield a cumulative error that at a certain epoch can exceed the accuracy required. This is a well-known problem of “dead reckoning” systems. Trying to reduce the error between couples of subsequent estimations, in order to minimize the overall cumulative error, does not solve the problem. A typical solution is to periodically re-sync the attitude along the trajectory with a known fixed reference (it could come from an external instrumentation, with compatible accuracy). However, in our approach a feasible solution is to have a geo-referenced map of just some patches of the Earth along the orbit. In this way, the absolute attitude can be determined through image registration when necessary. Different strategies can be adopted to cope with this problem, but their analysis is out of the scope of this work.

Finally, through the SP configuration the impact of the parallax error on error propagation can be evaluated with respect to the contribution of the registration error. In fact, similar results related to the UPD and SP configurations means that the main contribution to drift is given by the registration errors, being the influence on parallax term negligible.

Sect. 4.7.1 goes deep into discussion of the results related to these configurations using synthetic images.

4.6 Simulation Framework

The simulation framework devised in our application integrates three main blocks:

- an *orbital simulator*, developed by the DIEM group, to provide ground-truth attitude data. It consists of an *Orbit Propagator* (which integrates the Two-Body problem equations) and an Attitude Propagator (which describes the attitude dynamics and kinematics). The outcome of this block is a sequence of *state vectors* $[t, Alt, Lat, Lon, q1, q2, q3, q4]$, containing for each time instant t the position coordinates (altitude, latitude, longitude, expressed with respect to ECEF) of the satellite, and the attitude unitary quaternion (expressed with respect to the ECI). For nominal attitudes, the camera centre “looks” for each epoch at the point on the Earth at the corresponding latitude and longitude present in the state vector.
- an *image sequence generator*, capable to geo-reference the camera field of view according to the state vector. The corresponding area under the camera view is firstly computed from the geo-referenced database, following the procedure described in Sect.4.5.2. Then, different kinds of image interpolation can be used in order to generate the EOIs sequence corresponding to the orbit. Bilinear and bicubic interpolation have been used since they are sensitive to low-frequency and high-frequency components of the image, respectively.
- an *image registration framework*, which processes the EOIs sequence and the ground-truth state vector and returns for each couple of consecutive frames several useful outputs, including the estimated attitudes and the errors with respect to the ground-truth ones.

All these blocks, though being general, have to take into account the application requirements. In the following subsections the specific choices adopted for this application are discussed.

4.6.1 Image Acquisition Model

Based on the image registration algorithm described in 4.4, we have studied a proper model for the image acquisition process that is compliant with problem requirements.

As a first preliminary stage, the relation between sensor size (ground area, accordingly), overlapping area between subsequent frames (for feature tracking), frame rate and ground velocity has been analyzed. The overlapping area between two consecutive frames should be reasonably kept at least at about 25% – 30%, in order not to yield

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

Acquisition frame rate [fps]		1	2	5	10
Ground velocity [pixel/s]		460	230	92	46
Sensor size [pixel]	Ground area [km ²]	Overlapping area (%)			
320 × 240	4.8 × 3.6	0	28.1	71.3	85.6
640 × 480	9.6 × 7.2	28.1	64.1	85.6	92.8

Table 4.1: Relation among resolution, acquisition frame rate and overlapping area between consecutive frames.

Model Simulation Parameters	
Earth Model	$R_E = 6371 \text{ km}$ $\Omega_E = 7.272 \cdot 10^{-5} \text{ rad/s}$
Sensor Model	Dimensions = 320 × 240 pixel Pixel Size = 8 μm Focal Length = 336.7 mm
Orbit Model	Altitude = 650 km Ground Velocity = 6.9 km/s

Table 4.2: Parameters for the Earth model, the sensor matrix and the orbit used in our simulations.

an ill-conditioned problem when estimating the homographic transform from successfully matched features. Table 4.1 shows how the sensor resolution (in pixel) and the acquisition frame rate affect the overlapping region. Therefore, we can see how for a small size image (320 × 240 pixel) the minimum frame rate can be of 2 fps only, while when doubling the resolution, the image processing algorithm can run at 1 fps. However, low frame rates have to be avoided since the attitude control feedback would be responsiveness and would expose the satellite to longer perturbations. According to these considerations, the parameters used in our simulations, compliant with the application requirements, are summarized in Table 4.2.

Given these parameters, the instantaneous field of view (IFOV) subtended by each pixel is of 4.9 arcsec. The required *tracking accuracy*, that is the minimum accuracy (in terms of sub-pixels) that must be reliably achieved between two views to achieve the required attitude accuracy, can be computed according to the procedure described in Sect 4.5.1. To give an idea about the orders of magnitude of the values involved, given

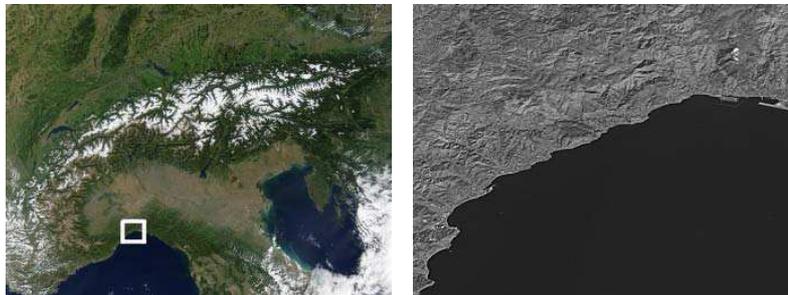


Figure 4.3: Image from NASA Blue Marble (**left**) and Landsat7 databases (**right**).

the parameters in Table 4.2, a separate perturbation for each of the attitude angles of 1 arcsec leads to a shift of about 0.2 pixel to be appreciated, for roll and pitch, while results in a shift of the order of 10^{-3} arcsec for yaw perturbation. Being the system less sensitive to this latter perturbation, we should expect worse estimates for this angle, accordingly.

The working frequency of the orbital simulator is $10Hz$ and $10fps$ for our algorithm, accordingly.

4.6.2 Dataset

Regarding the Earth satellite imagery chosen for our simulations, different georeferenced databases publicly available have been examined, also considering that the resolution on ground should be compatible with the performance of the tracker algorithm in terms of accuracy. The choice has fallen on Landsat7 ETM+, since it better fits the resolution required. In particular, Landsat7 images have a proper geolocation tag and cover 4×4 arcdeg of the Earth, with a spatial resolution of 15 m/pixel (covering 0.5 arcsec in latitude/longitude per pixel) compatible with the sensor's parameters reported above. Just as a reference, Fig. 4.3 (left) shows an image from NASA Blue Marble (250 m/pixel) and the corresponding highlighted image from the NASA Landsat7 databases (right).

4.7 Experimental results

In this Section, the experiments designed to analyze the different kinds of errors are described and the related results are discussed, according to the classification made in the previous sections. In general, for each of the test described, the attitude estimated

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

at each epoch t' by our system, $q_{est}(t')$, is compared to the ground-truth quaternion $q_{GT}(t')$, provided in different modalities according to the nature of the experiment.

The *attitude error quaternion* is defined in Eq. 4.13 as:

$$err_q(t') = (q_{GT}(t'))^{-1} \circ q_{est}(t'); \quad (4.13)$$

and converted in roll-pitch-yaw offsets, expressed in arcsec, to compute displacements from ground-truth data. It retains the rotation that should be applied to the satellite to bring it from the ground-truth attitude to the estimated one.

4.7.1 Synthetic imagery

In this paragraph the results for tests performed on synthetic image sequences are reported. A first set of simulations (S1, hereinafter) of 301 frames (300 image couples) has been performed, covering a portion of a slightly perturbed low Earth *near polar* circular orbit with an altitude of 650 km. The satellite ground velocity is about 6.9 km/s, spanning on the region bounded by the geographical coordinates $[44N, 8E] \div [48N, 12E]$. The corresponding ground truth attitude data, expressed in the LORF (and so with respect to the nominal attitude), are illustrated in Fig. 4.4, left column. In Fig. 4.4, (right column), the frame-by-frame angular offsets are shown.

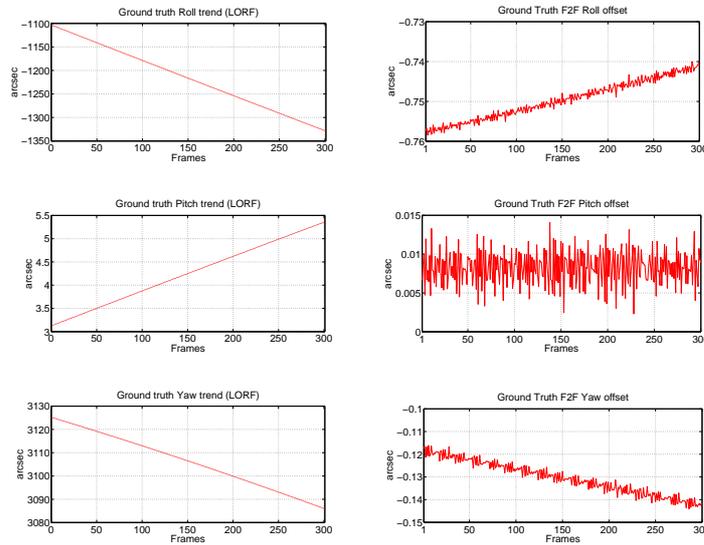


Figure 4.4: (Left) Ground-truth attitude trends and (right) the corresponding F2F ground-truth angular offsets for the simulation S1, expressed with respect to the LORF.

As it can be seen, the roll angle is negative and it increases in absolute value, with a slightly decreasing velocity, while the pitch value is increasing with an almost constant velocity. The yaw trend is decreasing as well, but with a slightly increasing velocity. These ground-truth attitude variations lead to corresponding image optical flows that can be *theoretically* estimated from the models adopted, that is using the computed matchings. By analyzing the results of the related simulations, and comparing them with the ones obtained on the synthetic imagery, some interesting error dynamics can be identified and studied.

4.7.1.1 Theoretical accuracy performance

By using in the simulations the *computed matchings* from the procedure described in Sect. 4.5.1, the accuracy upper bounds of the adopted model can be characterized. In Fig. 4.5 the results regarding the accuracy and the statistical distribution for the three attitude errors for pairwise registration in the NoUPD configuration are reported, considering the *computed matchings* for a continuous model of the sensor.

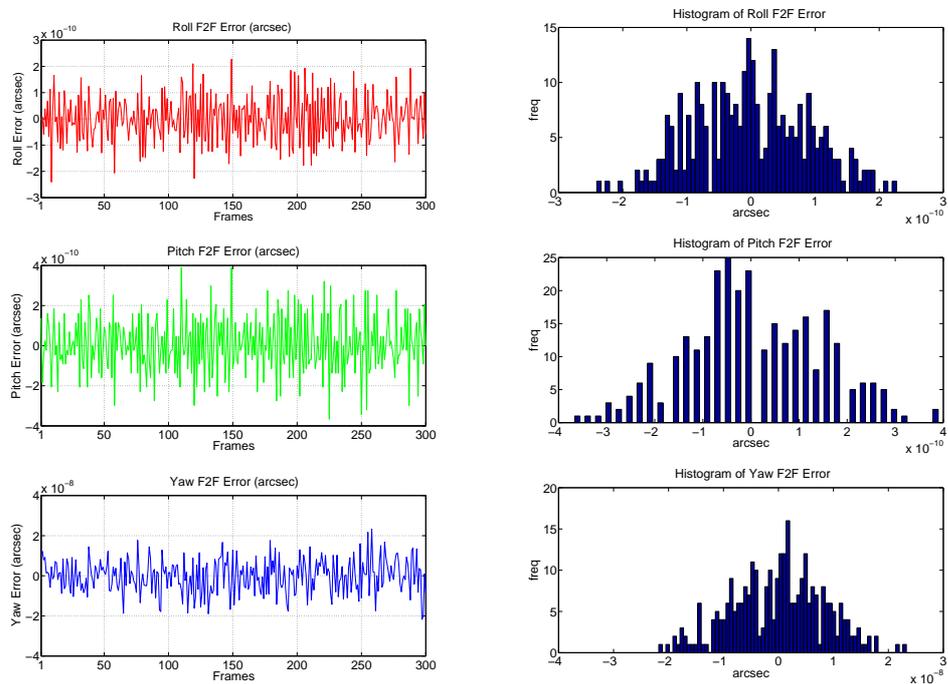


Figure 4.5: (Left Column) F2F attitude errors values for the S1-NOUPD using computed matchings. (Right Column) Histograms of the F2F attitude errors for S1-NOUPD using computed matchings.

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

In this case only numerical errors affect the accuracy, which reaches the order of $[10^{-10}, 10^{-10}, 10^{-8}]$ arcsec, for roll, pitch and yaw attitude angles, respectively. It can be noticed that error distributions are not biased, as expected from numerical errors randomness. For these reasons, the computed matchings can be considered, up to negligible numeric approximations, as the *ground-truth corners* that ideally a feature tracking system should strive for in order to reach the highest accuracy (without systematic errors).

On the other hand, other factors, like sensor discretization, can lead to a limited performance of the tracker, with interesting effects. Initially, the NoUPD configuration for this sequence has been analyzed to evaluate the F2F error accuracy and distribution (S1-NOUPD). Simulations for S1-NOUPD have been performed assuming that the accuracy of the tracker is limited to a certain decimal digit, with a rounding effect, while estimating the motions in X and Y directions in the image plane. Roll (camera pan) or pitch (camera tilt) perturbations lead to predominant horizontal or vertical motion, respectively, in the image plane. Even though the X and Y components are both correlated with the perturbation of any of the camera angles, results for roll and pitch errors can be inferred analyzing the X and Y motion components. Fixed in the sensor matrix a sample pixel to be tracked, the X and Y coordinates of the corresponding computed matching have the trends shown in the first row of Fig. 4.6 (green circles). The corresponding rounded values at the second decimal digit have been superimposed (blue stars), and the differences between the computed value and their rounded version (that is the *tracking error* in this case) are reported in the second row. These periodic patterns are due to the rounding effect applied on the monotonic trends in the first row, caused, in turn, by the monotonic attitude variations of the ground-truth input data (Fig. 4.4 (left column)). In this case, the hypothetical tracker *appreciate* the second decimal digit while *sensing* the third decimal digit.

As far as attitude errors are concerned, results obtained rounding the computed matchings at the second and the first decimal digit are reported in Fig. 4.7, left and right column, respectively. Comparing these results with the ones obtained with the computed matchings (Fig. 4.5), we can observe that the attitude estimation accuracy decays as the rounding decimal digit is more significant, as expected. In particular, for roll, pitch and yaw angles, respectively, the order of magnitude of the error is $[10^{-2}, 10^{-2}, 10^0]$ arcsec when rounding is applied at the second decimal digit. It reaches the order of magnitude of $[10^{-1}, 10^{-1}, 10^0]$ arcsec when the rounding is applied on the first decimal digit.

Moreover, the periodic oscillating patterns of the tracking error components induce

4.7 Experimental results

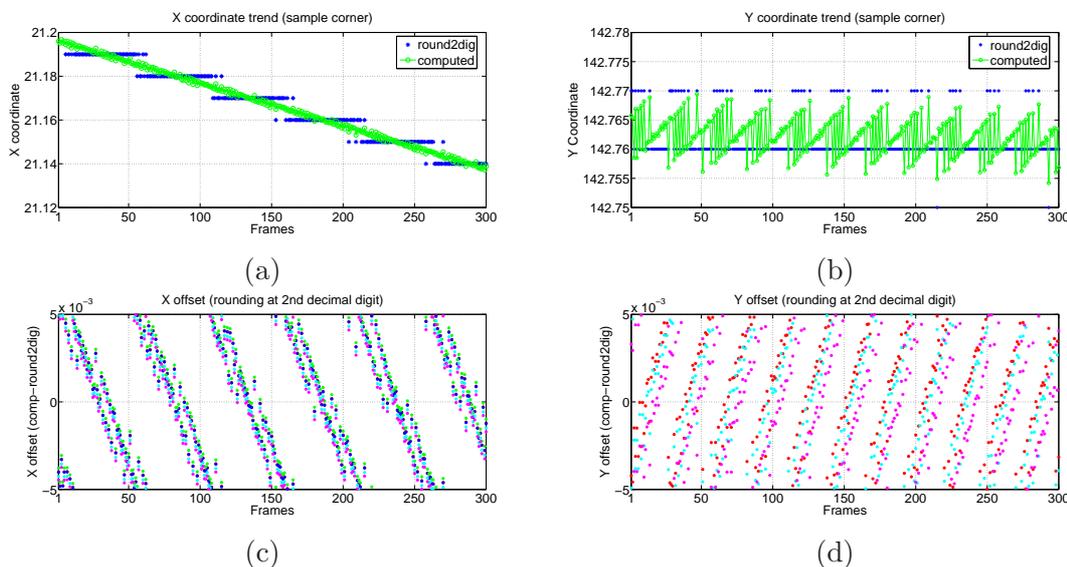


Figure 4.6: (a) Trend of the X coordinate of a sample computed corner (green circles) and its value rounded at the second decimal digit (blue stars), and the X tracking error due to rounding (c), for sample pixels. (b) Trend of the Y coordinate of the sample computed corner (green circles) and its value rounded at the second decimal digit (blue stars), and the Y tracking error due to rounding, for sample pixels (d).

similar patterns in the attitude errors. Observing the left and the right column in Fig. 4.7, it can be seen that this period depends on the trend of the ground-truth input data and on the chosen rounding decimal digit. As far the roll error is concerned, its trend is ruled by the slowly decreasing trend of the X coordinate of the computed corner (Fig. 4.6, (a)). In fact, rounding at the second digit is sensitive to the third digit of the computed value, as well as rounding at the first digit is sensitive to the second digit of the computed value, and so on. Accordingly, since the the second digit of the computed value varies more slowly than the third digit, the period increases as the rounding digit passes from the second (Fig. 4.7, (a)) to the first one (Fig. 4.7, (b)). In general, the period will be much shorter as the chosen rounding digit is less significant.

The pitch error has a different behavior, being ruled by the trend of the Y coordinate of the computed corner (Fig. 4.6, (b)). The latter does not affect, in the observed frame interval, the rounding at the first digit, which yields the constant value 142.8 for all the frame in this interval. Accordingly, the corresponding pitch error (Fig. 4.7, (d)) does not change sign and is ruled by the variations of the green curve in Fig. 4.6, (b), from this constant value. On the other hand, if the rounding is applied to the second

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

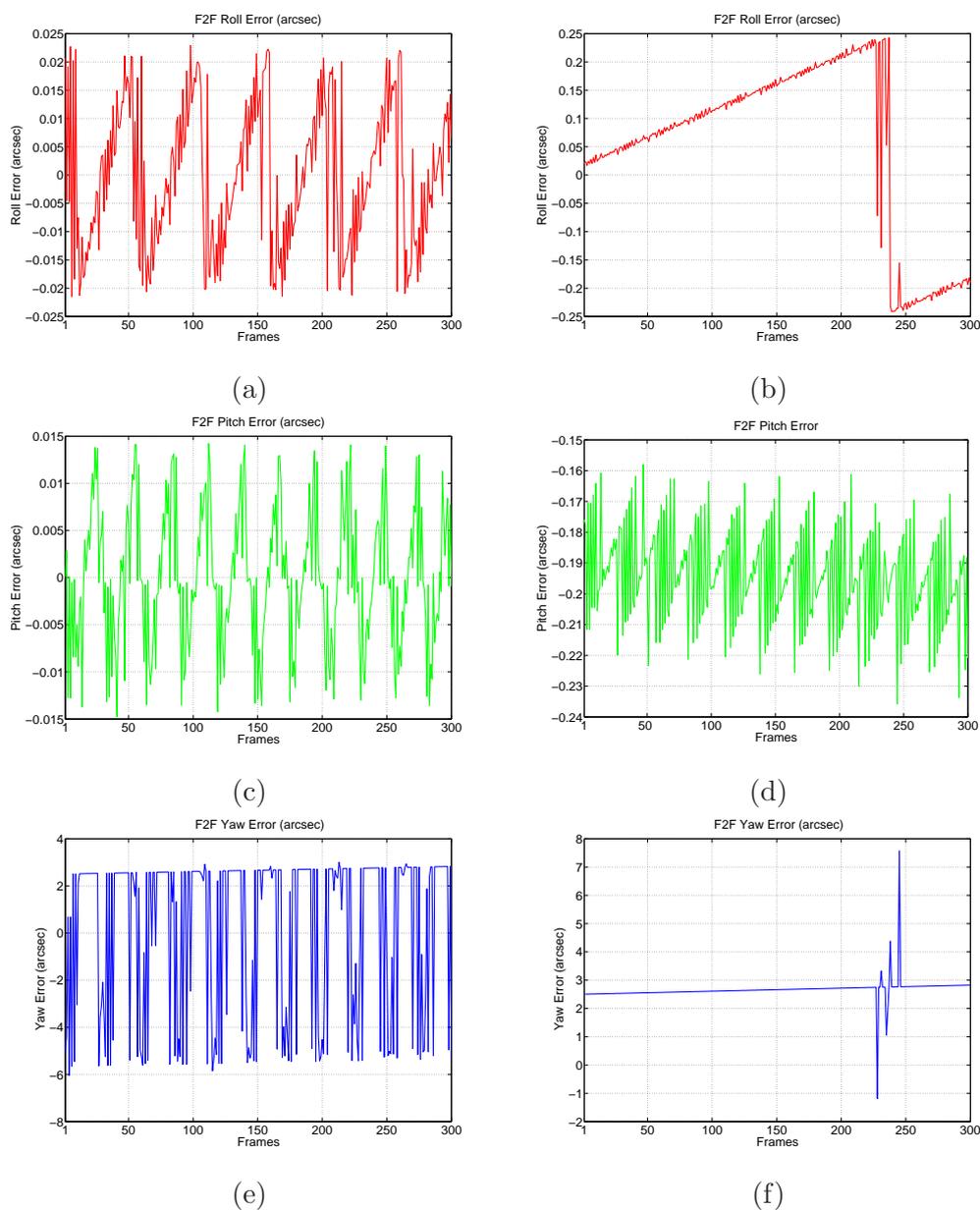


Figure 4.7: Attitude error trends obtained rounding the computed corners at the second (left column) and first (right column) digit.

decimal digit, the corresponding error trend (Fig. 4.7, (c)) is ruled by the periodic pattern in Fig. 4.6, (d). The yaw error is theoretically affected by a composition of the dynamics observed for the roll and pitch error. As it can be seen from Fig. 4.7, (e),

(f), its behavior predominantly follows the roll angle one, and so it is mainly induced by the variation in the X component of the computed corner, which presents a greater magnitude in the observed simulation.

Furthermore it is worth to notice that, considering an interval included in a semi-period, the error is generally biased and its accumulation can yield to drift effects when used to update the subsequent epoch. These observations are useful in order to better explain the results obtained by the our registration algorithm on synthetic imagery and described in the next section.

4.7.1.2 Frame-to-frame registration: Results

The image sequence corresponding to simulation S1 is generated according to the procedure summarized in Fig. 4.2. The image frames are thus the result of the image interpolation process applied to the georeferenced image database. In Fig. 4.8 a couple of the frames with the features extracted and tracked is shown.

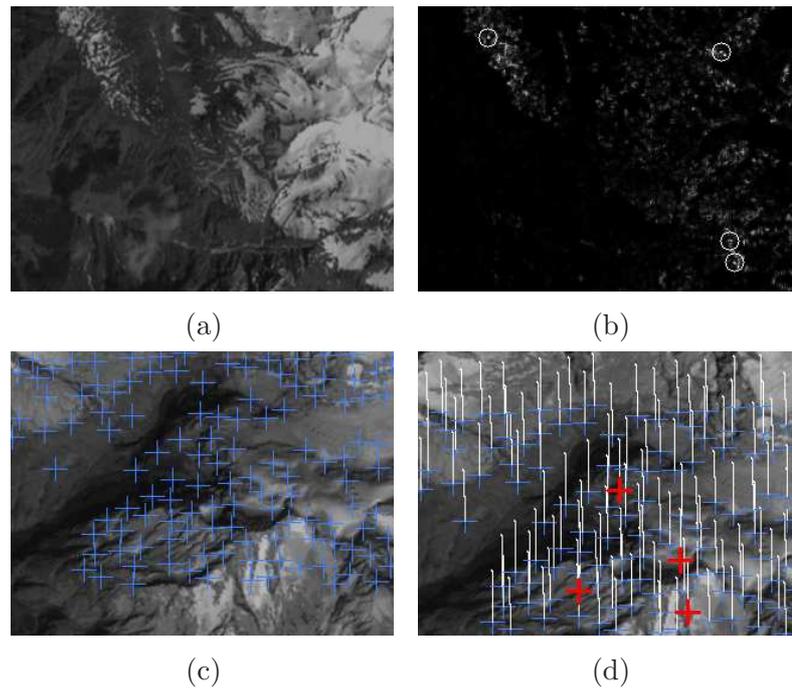


Figure 4.8: (a) A sample image and the corresponding normalized map (b) of the eigenvalues of the pseudo-hessian for the extraction of the corner points. Couple of consecutive frames with the extracted (c) and tracked (d) corner points.

Similarly to what has been done with the computed matchings, initially the image se-

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

quence has been processed in the NOUPD configuration in order to evaluate the F2F error accuracy and distribution. In Fig. 4.9 the related results for the attitude error are shown.

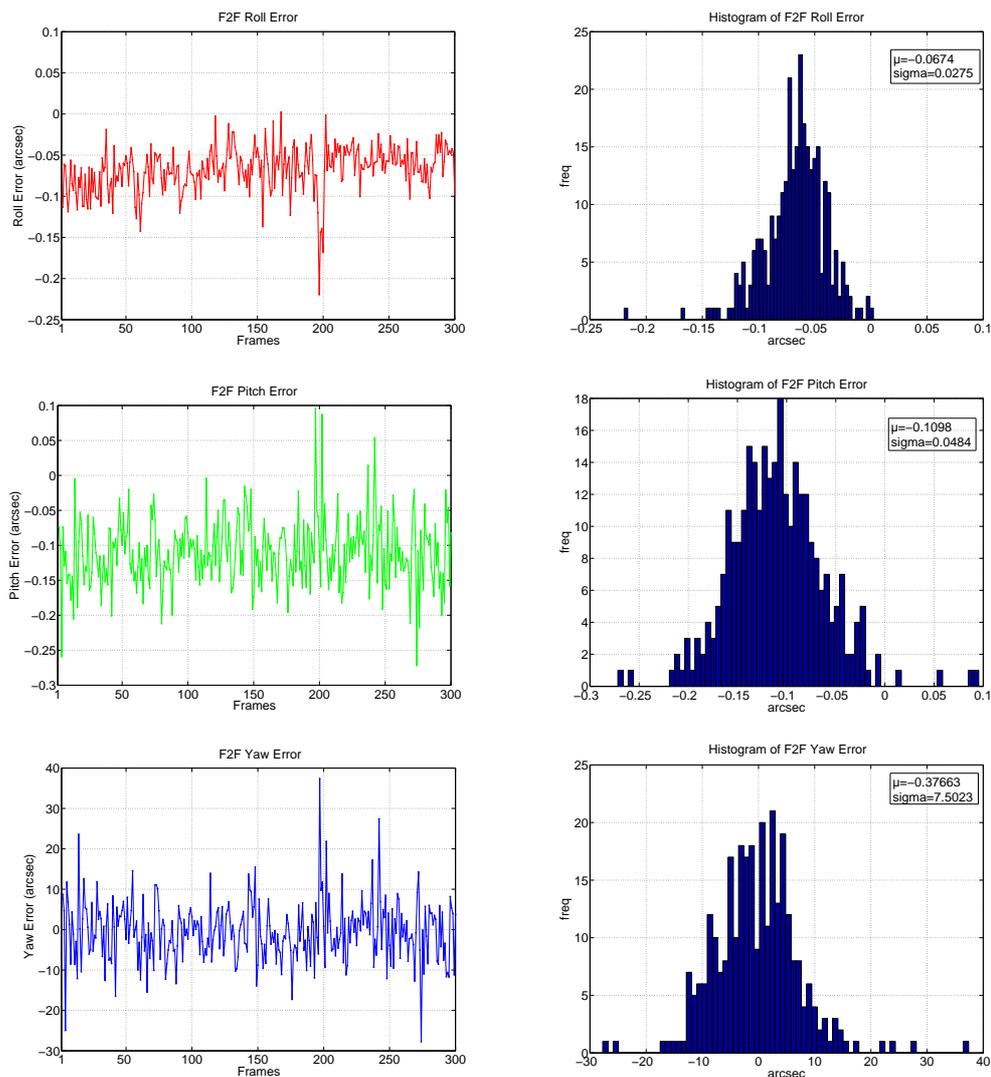


Figure 4.9: Left Column. F2F attitude errors values for the S1-NOUPD. **Right column.** Histograms of the F2F attitude errors for S1-NOUPD.

Regarding accuracy, it is of the order of decimals of arcsec for roll and pitch angles while it is of tens of arcsec for the yaw angle, because of the far lower sensitivity of the image motion field to yaw perturbations intrinsic to the system. As far as the error dis-

tribution is concerned, the histograms of the attitude errors, expressed in arcsec, show a very skewed distribution for roll and pitch. This can yield drift when the absolute orientation is computed and used for the subsequent epoch.

4.7.1.3 Frame-to-frame registration: Discussion

These results stem from the different contributions aforementioned, among which the image registration algorithm (and the tracker performance) plays the main role. To better characterize the error due to the image registration error only, different test sequences (S2, afterward) has been generated with the satellite in a *fixed* position, by just varying one angle (S2-angle) at a time with a fixed step (S2-step) for each sequence, while keeping the other two angles unchanged. In this situation the parallax term is null and the error distribution in this NoUPD configuration can be employed to foresee the error trend in the UPD configuration.

A first set of simulations S2, S2-ROLL-GEO1, has been performed with the satellite fixed at an altitude of 650 km, “looking” at the geographical coordinates $[46.5N] \div [9E]$, starting from the nominal attitude and varying for 1000 frames the roll angle with a constant step. A second set of simulations S2 of 1000 frames, S2-ROLL-GEO2, has been performed in the same manner but over a different sequence of images, in particular with the satellite “looking” at the geographical coordinates $[47.5N] \div [9E]$. In all these simulations, steps of different entity are used as ground truth data to modify with a uniform offset the X and Y component of the corners’ motion vectors for each couple of consecutive frames. In this way, the tracking errors, obtained comparing the tracked corners with the computed corners as the ground-truth *target* corners, can be used to explain results in terms of attitude errors.

It is well known that projections of fixed scene point (points on the Earth) onto the image plane during roll (camera pan) or pitch (camera tilt) perturbations move along hyperbolic paths (80), thus with both image X and Y components changing. However, as previously hinted, for roll perturbation the predominant component to be analyzed is the X component and, in the same way, the Y component has to be studied for simulations with only pitch perturbations. Similarly, the set S2-PITCH-GEO2 refers to the same configuration used for S2-ROLL-GEO2, but with the pitch angle changing.

To study the error trends, we have defined some useful *features* as follows. The set $S(t, t')_{tr}$ of the tracked corners for the generic frame couple consists in position in the image plane of the generic i^{th} corners, extracted from the first frame (X_{ex}^i, Y_{ex}^i) and tracked in the second frame (X_{tr}^i, Y_{tr}^i) . We indicate as (X_c^i, Y_c^i) the corresponding computed corner and as $S(t, t')_c$ the resulting set of computed corners. The *estimated*

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

frame-to-frame mean motion vector at the frame t' , $V_{tr}(t')$, is defined as the average motion vector over the set $S(t, t')$, that is (Eq. 4.14):

$$V(t')_{tr} \doteq \begin{bmatrix} V_x \\ V_y \end{bmatrix}_{tr}^{(t')} = \begin{bmatrix} \frac{\sum_{i=1}^N (X_{ex}^i - X_{tr}^i)}{N} \\ \frac{\sum_{i=1}^N (Y_{ex}^i - Y_{tr}^i)}{N} \end{bmatrix}_{tr}^{(t')} \quad (4.14)$$

where N is the cardinality of the set $S(t, t')$. Similarly, the *computed* frame-to-frame mean motion vector $V_c^{t'}$ can be defined using in the Eq. 4.14 the computed corners instead of the tracked ones. The mean frame-to-frame *tracking error*, $\delta V(t')$, is defined as:

$$\delta V(t') \doteq \begin{bmatrix} \delta V_x \\ \delta V_y \end{bmatrix}^{(t')} = \begin{bmatrix} V_x \\ V_y \end{bmatrix}_{tr}^{(t')} - \begin{bmatrix} V_x \\ V_y \end{bmatrix}_c^{(t')} \quad (4.15)$$

that is the average of contributions related to the single i^{th} corner, ($\delta V_x^i(t') = X_c^i - X_{tr}^i$, $\delta V_y^i(t') = Y_c^i - Y_{tr}^i$). The statistical distribution of the contributions ($\delta V_x^i(t')$, $\delta V_y^i(t')$) is related to the probability, for that frame (t'), of rounding up or down the ground-truth value of the motion vector components. For example, considering a sample couple of consecutive frames from the sequence S2-ROLL-GEO1 (with step $\Delta = 1$ arcsec), the histograms of ($\delta V_x^i(t')$, $\delta V_y^i(t')$) are shown in Fig. 4.10. As it can be noticed, the

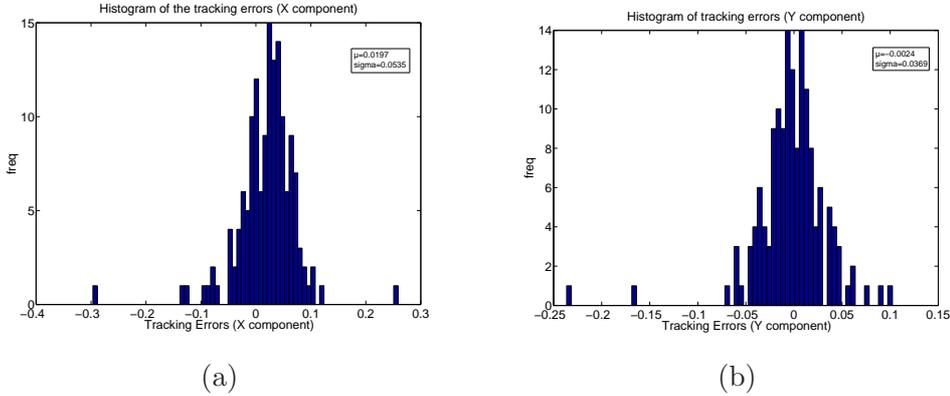


Figure 4.10: Histograms of δV_x^i (a) and δV_y^i (b) for a sample frame couple from the sequence S2-ROLL-GEO1 (with step $\Delta = 1$ arcsec).

distribution of δV_x^i is biased towards positive values, while the distribution of δV_y^i is unbiased, as expected. This holds for all the other frames of the sequence of the same simulation, so that the mean values ($\delta V_x(t')$, $\delta V_y(t')$) have the temporal trends reported in Fig. 4.11. By averaging these values over the sequence of frames (the parameter t'

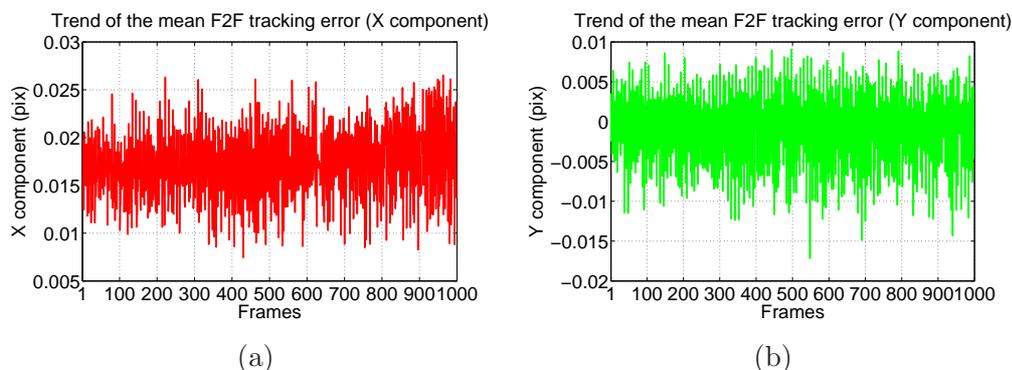


Figure 4.11: Temporal trends of $\delta V_x(t')$ (a) and $\delta V_y(t')$ (b) for the sequence S2-ROLL-GEO1 (with step $\Delta = 1$ arcsec).

disappears), for each image sequence the mean computed motion vector, $\bar{V}^c = (\bar{V}_x^c, \bar{V}_y^c)$, and the mean tracking error, $\delta \bar{V} = (\delta \bar{V}_x, \delta \bar{V}_y)$, can be computed for that sequence. Similarly, the corresponding standard deviations $(\sigma(\delta \bar{V}_x), \sigma(\delta \bar{V}_y))$ can be estimated. All these *features* take into account the variability in the corner structure for each image of the sequence and the variability in the image content along a sequence. In Table 4.3 the ground-truth values of these features for the S2 sets are summarized.

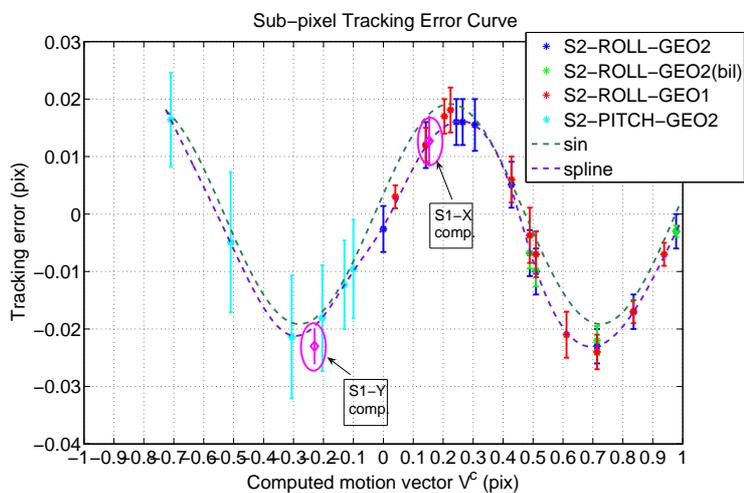
For such simulations, the Phase Correlation stage is able to return the *correct* (expected) X and Y offsets, but being at the same time sensitive to the half pixel, with a rounding effect to the closest integer. Accordingly, we are interested in the decimal part of the \bar{V}^c components, that belongs obviously to the interval $[0 : 1)$. In Fig. 4.12 the values of $\delta \bar{V}_x$, for the simulations S2-ROLL-GEO1 and S2-ROLL-GEO2, and $\delta \bar{V}_y$, for the simulations S2-ROLL-GEO2, are plotted against the values of \bar{V}_x^c and \bar{V}_y^c , respectively. Error bars refers to the corresponding standard deviations $(\sigma(\delta \bar{V}_x), \sigma(\delta \bar{V}_y))$. Since the behavior of the X component of these features for roll perturbations can be likened to the one of the Y component for the pitch perturbations, the X and Y subscripts can be removed and the corresponding samples can be plotted together in the same figure. Natural cubic spline and sinusoidal fitting have been superimposed in order to extrapolate a trend. It is worth remarking that even the samples of the features $\delta \bar{V}_y$ and $\delta \bar{V}_x$ resulting from the simulation S1-NOUPD (with the satellite moving) fall on this curve (see the text boxes in Fig. 4.12 (a) corresponding to the light magenta samples). Moreover, the low standard deviations show the high *precision* of the motion vector estimation process, which indicates stability in the estimation process for different corners and images. The sparse image registration approach using a high number of

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

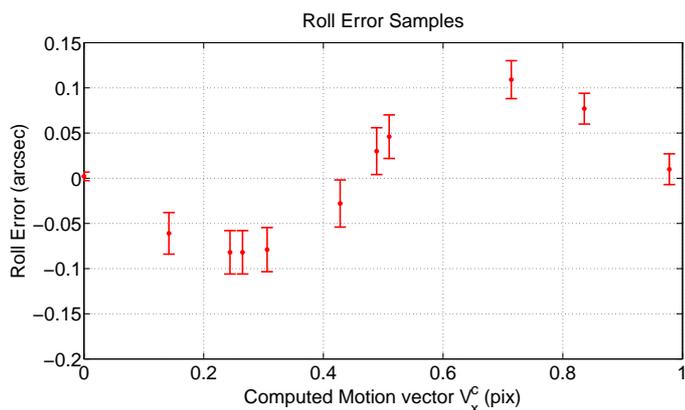
S2 Ground-Truth Simulations Parameters			
	Δ (arcsec)	\bar{V}_x^c	\bar{V}_y^c
S2-ROLL-GEO1	0.700	0.142	0.00
	1.0	0.204	0.00
	1.1	0.224	0.00
	2.4	0.489	0.00
	2.5	0.510	0.00
	3.0	0.612	0.00
	3.5	0.714	0.00
	7.0	1.428	0.00
	9.0	1.836	0.00
	9.5	1.938	0.00
	10.0	2.04	0.00
S2-ROLL-GEO2	0.0	0.0	0.0
	0.7	0.142	0.00
	1.2	0.244	0.00
	1.3	0.265	0.00
	1.5	0.306	0.00
	2.4	0.489	0.00
	2.5	0.510	0.00
	3.5	0.714	0.00
	7.0	1.428	0.00
	9.0	1.836	0.00
	9.7	1.978	0.00
S2-PITCH-GEO2	0.50	0.0	-0.102
	0.64	0.0	-0.130
	1.00	0.0	-0.204
	2.50	0.0	-0.51

Table 4.3: Parameters for S2 simulations on synthetic images. Values of \bar{V}_x^c and \bar{V}_y^c induced by the angular perturbation are visualized as rounded to the third decimal digit.

corners permits to increase the measurements and provide a more *accurate* estimation. Furthermore, observing Table 4.3 and Fig. 4.12 (a), we can infer that the tracking error depends only on the sub-pixel value of the ground truth motion field and not on its



(a)



(b)

Figure 4.12: (a) Sub-pixel tracking error curve obtained by fitting tracking error samples from different simulations, against computed motion vectors \bar{V}^c , with $\bar{V}^c = \bar{V}_y^c$ for PITCH perturbations (in cyan) and $\bar{V}^c = \bar{V}_x^c$ for ROLL perturbations (other colors). Usually bicubic interpolation is used for EOIs generation, but similar behavior is encountered also when bilinear interpolation is employed (in green). Light magenta circles surrounds samples obtained from the simulation S1-NOUPD, which still belong to the estimated sub-pixel tracking error curve. (b) Sample values related to the roll error for the simulations S2-ROLL-GEO2 are reported against the computed motion vectors $\bar{V}^c = \bar{V}_x^c$. They follow, as expected, the same trend extrapolated from the samples in (a), up to the sign of the error.

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

complete value. In other words, the tracking errors corresponding to even very different attitude perturbations (0 or 10 arcsec, for example) can be of the same order of attitude (see the last entry of the parameters related to S2-ROLL-GEO1 in Table 4.3). Thus there is an intrinsic accuracy limit, so that measuring perturbation values comparable with this limit does not make sense. However the accuracy provided by our method complies with the accuracy requirements of the application, since our framework is able to measure, for example, roll perturbations of 1 arcsec with an accuracy of the order of 10^{-1} arcsec.

The corresponding angular error trends, averaged on the 1000 frames, yield the samples related to the mean attitude error. For example, in Fig. 4.12 (b) the samples related to the roll error for the simulations S2-ROLL-GEO2 are reported. They follow, as expected, the same trend extrapolated from the Fig. 4.12 (a), up to the sign of the error. In fact, to positive perturbations in the roll angle a positive X component of the computed motion vector is associated. Thus, underestimating (overestimating) it, as it happens for positive (negative) values of $\delta\bar{V}_x$, leads to a rounding down (up) error in angle estimation.

The samples in Fig. 4.12 (a) clearly form a periodic pattern (with period equal to one pixel), that we call Sub-pixel Tracking Error Curve, oscillating between a positive and a negative part. The tracking error in absolute value reaches its minimum in multiples of the semi-pixel, while it is maximum at the halves of the semi-pixel (around 0.25 and 0.75). This pattern can be conveniently observed within the window $[-0.5:0.5]$ in order to better understand it. It stems that for positive components of the ground-truth motion vector (interval $[0:0.5]$), the tracker underestimate the ground truth value. For negative component of the ground-truth motion vector (interval $[-0.5 : 0]$) the tracker overestimate the ground truth value. Considering the absolute values of the entities involved, in both the cases an underestimation occurs, so that it seems that the tracker tend to the theoretical position of the corner without ever reaching it, because of the limited image resolution, due to discretization, and a corresponding limited tracking accuracy.

In Fig. 4.13 a qualitative sketch of the tracker behavior with respect to the computed features is presented. As it can be noticed, the pixel centers (integer coordinates) N and $N + 1$ behave like *basins of attraction*. In fact there is a symmetric behavior of the tracker with respect to the pixel centers, with a rounding-down effect, increasing in absolute value as the distance from the basin of attraction grows. The semi-pixel position ($N + 0.5$) seems to behave like a *separator* between the basins of attraction, also being already sensed by the Phase Correlation stage. So, when the input angular

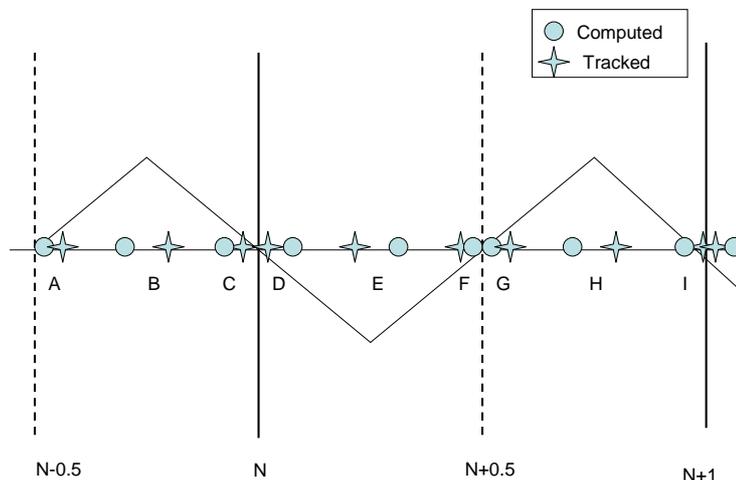


Figure 4.13: Qualitative sketch of the tracker behavior with respect to the computed features is presented. Circles represent *computed* features, while stars represent *tracked* corners. The trend of the tracking error is qualitatively similar to the one of the sub-pixel tracking error curve (Fig. 4.12 (a)).

perturbation make the computed corner fall on the other side of the semi-pixel, the basin of attraction “switches” (inversion F-G in Fig. 4.13) on the other pixel integer coordinate. This symmetry makes necessarily the tracking error reach its maximum (B, E, H in Fig. 4.13) at the halves of the semi-pixels. More formally, the extrapolated curve is continuous and has 2 zeros at integer coordinates, with different signs in the two semi-periods. For the Bolzano’s Theorem, a third zero must exist, and as it can be seen it is located near the half of the pixel. Moreover, since also the concavity of the (twice continuously differentiable) curve changes sign between the two semi-periods, this is also an *inflection point*. Considering the semi-periods, in each of them for the Rolle’s Theorem a stationary point must exist, as in fact happen in 0.25 and 0.75, respectively.

This behavior is coherent with the LKT algorithm, that works on small image patches around the corner guess position (already taking into account the Phase Correlation contribution). In this case, where also image interpolation is present, the sub-pixel motion displacement is sought iteratively. First, the pixel center closest to the computed corners will show the maximum image likelihood, since it affect with a

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

higher weight the image formation process. Starting from this pixel center, the LKT algorithm then tend to converge, but with limited accuracy, to the computed corner.

Potentially the sub-pixel tracking error curve can be used to *compensate* the tracking error, so that even in case of regular variation of the ground truth perturbation data, as it happens for S1, the resulting histogram of the attitude error would have an unbiased distribution.

4.7.1.4 Dynamic conditions

In dynamic conditions, the estimated attitude is then propagated to the next frames. To evaluate the impact of the parallax term on error propagation, the UPD and SP simulation configurations should be analyzed. In Fig. 4.14 the results related to the S1 image sequence are reported for pitch and yaw angles (results for roll angle are comparable to the pitch ones). It can be seen the drift effect expected from the ac-

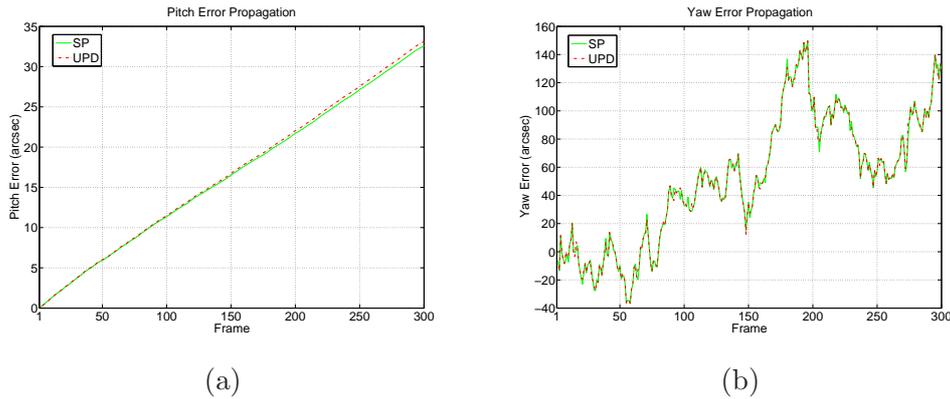


Figure 4.14: Error propagation in the UPD (dashed line) and SP (continuous line) configurations for pitch and yaw angles.

cumulation of the biased F2F error in the UPD configuration. As expected from the considerations made in Sect. 4.5.3, since the trends are similar, the main contribution to drift is retained in the propagation of the registration error. However, looking at the pitch angle, for example, it can be seen that how the two curves tend to diverge, as the impact of the attitude updating even in the parallax term begins to become relevant.

4.7.2 Laboratory Experiments

Our laboratory testbed consists of a CCD commercial camera (MV-BLUEFOX-121G) mounted on a Numerical Control Unit (NCU), capable of performing three-axis rotations, and watching a printed airborne scene.

The NCU allows to measure *ground-truth* attitude angles with respect to the NCU reference frame, with a sensitivity of $\sigma = 0.06^\circ$ ($\equiv 216$ arcsec). This equipment has been assessed following the vision-based procedure explained in (14).

The camera lens distortion and intrinsic parameters have been obtained once and for all after an off-line calibration process (81). Images acquired at a resolution of 1024×768 pixel are processed with our algorithm on an old generation consumer PC (AMD2000+, 1.66GHz, 1GB RAM). The camera optical axis is looking at the planar image of an airborne scene stitched onto a flat panel. Fig. 4.15 shows the experimental

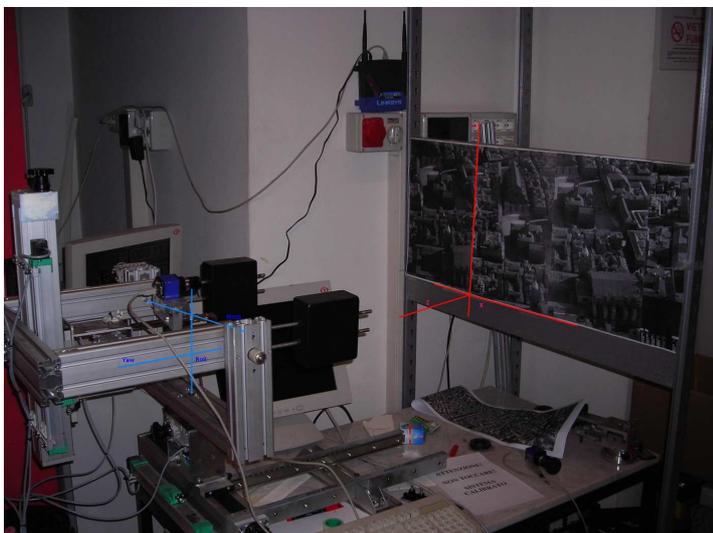


Figure 4.15: A picture of the NCU system together with the reference test image. NCU reference frame (blue) and world reference frame (red) are also sketched.

test bed used.

Several test series have been performed varying one angle along one axis at a time. Thus, we can measure roll, pitch and yaw perturbations one at a time, by leaving the other two components fixed. In order to compare the ground-truth data returned from the NCU with the measures expressed in terms of camera reference frame, it is necessary to initially align the two reference frames to a common zero reference. Two devices have been adopted to compensate for this. Since the present instrumentation does not provide directly the position of the camera center, the camera has been arranged into the NCU so that its principal point belongs to the rotation axis (for the angle of interest) of the NCU reference frame, following the common *parallax adjustment* procedure usually employed in photography. In this way, since the camera center does not move during

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

the rotation, the knowledge of the relative offset with respect to the center of the NCU reference frames is not necessary, and the parallax term is nullified. Moreover, the axis-angle notation (Rodrigues rotation vector (82)) has been adopted. Through this strategy, the (unknown) fixed angular misalignment between the two reference frames does not affect the measurements, since it cancels out with this representation.

In this paragraph the results related to two different test sequences are reported. In the first test sequence (NCU-ROLL) we vary, closing a looping path, roll (camera pan) angles within a range of $[-0.98^\circ \dots + 13.15^\circ]$, with a mean angular step of about 30 hundredths degree, thus acquiring 75 frames. In order to reduce the influence of possible drift due to mechanical play, sometimes the camera has been moved back to its previous position. A couple of consecutive frames processed by our algorithm is shown in Fig. 4.16.

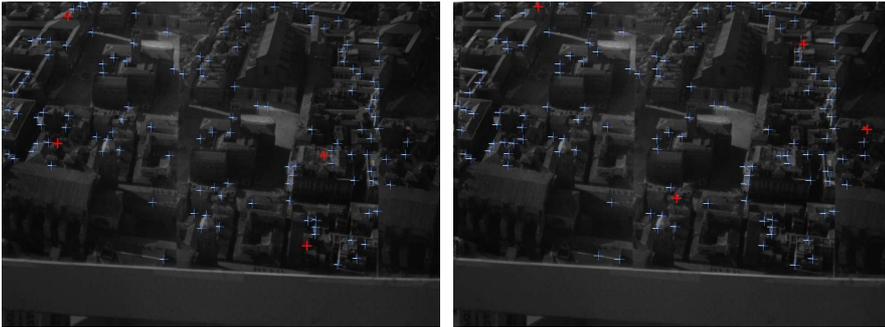


Figure 4.16: A couple of frames from the sequence NCU-ROLL processed by our algorithm.

Fig. 4.17 reports the results related to this test. A similar behavior should characterize sequences with pitch perturbation. In Fig. 4.17 (a) the correlation between the NCU ground-truth data and the estimated attitude angles, for this test, is shown, with the Pearson’s correlation coefficient equal to $p = 0.996$. Accordingly, this coefficient being very close to the unit, a strong linear correlation exists between the ground-truth roll perturbations (expressed in the NCU reference frame) and the ones measured by our framework, expressed in the camera reference frame. This confirms that the Rodrigues rotation vector representation makes the measurements consistent with the ground truth data.

In Fig. 4.17 the experimental pairwise (b) and with attitude updating (c) errors, for the whole test sequence, are shown together with the NCU uncertainty range (dotted lines). The F2F registration angular error almost always lies inside the NCU uncertainty

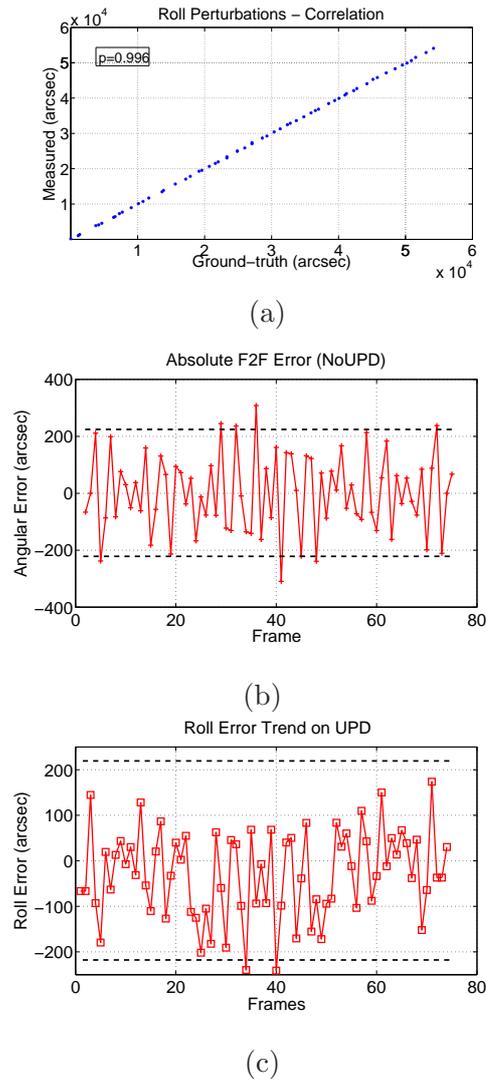


Figure 4.17: (a) Correlation among the ground-truth roll perturbations and the ones measured by our system. Absolute (b) and with attitude updating (c) F2F roll error (expressed in arcsec).

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

range and it spreads across the whole range of accuracy provided by the NCU system. In particular, now the absolute F2F error has a mean value close to zero ($\mu = 0.4$ arcsec) with a standard deviation $\sigma = 136.52$ arcsec. This means that the error distribution over the set of 75 samples is slightly biased, thus not resulting in an evident drift in the UPD configuration. We have tried to interpret these results in the light of the considerations made above for the sub-pixel tracking error curve. However, the accuracy of the equipment available at the moment is not sufficient to provide a so fine analysis. In other words, the ground-truth data are affected themselves by a discretization error. The quasi-unbiased distribution is probably due to the noise added on the ground-truth data, which probably leads to samples falling randomly in the positive or negative part of this curve.

Results concerning a second test sequence of 68 frames for yaw (camera tilt) perturbations (NCU-YAW) are shown in Fig. 4.18. The distances involved in this simulation

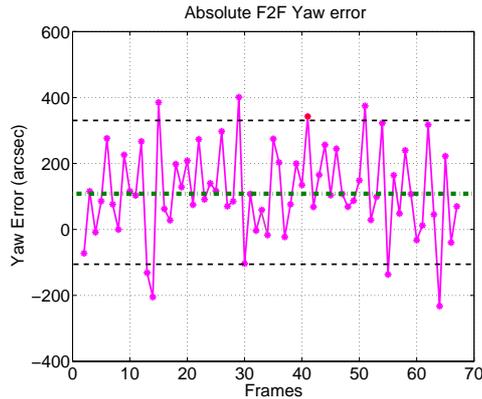


Figure 4.18: Absolute F2F yaw error (expressed in arcsec) resulting from the NCU-YAW test sequence.

(88, 4 cm in *nominal* conditions) does not affect dramatically the sensitivity of the system to the image motion field, as conversely happens for the simulation with synthetic images (see consideration made for Fig. 4.9, yaw angle). Because of the particular set up of the NCU, the parallax adjustment in this case could not be performed. As a consequence, the camera rotates around the NCU yaw axis with a non-zero radius, that we have tried to measure and estimate indirectly by tracking reference points over several circular trajectories. However this rough estimate on the camera center position with respect to the axis of rotation (about 14 cm) leads to a systematic error in the parallax term. This is the reason for the strong bias in the F2F absolute error in Fig. 4.18, with

a mean value of $\mu = 113.4$ arcsec and a standard deviation of $\sigma = 136.87$ arcsec, this being comparable with the value obtained for roll and pitch.

4.8 Conclusions

In this Chapter we have presented the application of our registration algorithms for a camera pose reconstruction application, the estimation of a satellite three-dimensional orientation exploiting Earth images acquired by an on-board camera. In such a kind of application, limited hardware resources are available and they must be sufficient to provide with high accuracy (1 arcsec) and at high rate (several frames/second) the satellite attitude (expressed in Roll-Pitch-Yaw displacements from nominal values), for automatic guidance purposes. We have focused our attention on the methodological stages adopted to assess the performance of our novel registration approach in this particular case study. First, the Earth-Satellite system has been modeled. The achieved model and a realistic orbital simulator have been integrated into a simulation environment, providing attitude ground-truth values, utilized to study the accuracy performance at more levels.

Initially, the accuracy theoretical performances have been derived according to the system model and the numerical error has been characterized, taking the resulting estimations as the best-case reference configuration. In order to assess our image registration method, synthetic image sequences have been generated in the simulation environment referring the camera field of view on a geo-referenced database. Through these simulations, the accuracy of the registration algorithm in the image domain (pixel tracking error) can be associated with the error in attitude estimation. In particular, as far as pairwise registration is concerned, taking as reference the theoretical models, experiments with synthetic images have shown that the tracking accuracy is between 10^{-1} - 10^{-2} pixels. The corresponding attitude estimation error is of the order of decimals of arcsec for roll and pitch angles, while it is of tens of arcsec for the yaw angle, because of the far lower sensitivity of the image motion field to yaw perturbations intrinsic to the system.

To study more in depth the tracker behavior, the effect of the rounding of the tracker precision has been investigated in relation to the ground truth data. This has permitted to characterize and thus predict the tracker behavior in terms of accuracy, both in the image domain and in terms of angular error. Furthermore, different simulation configurations have permitted to isolate different error sources and study the error dynamics. They have confirmed, together with laboratory experiments with real world

4. CAMERA ATTITUDE ESTIMATION: THE STARS PROJECT

images, the correctness of the hypothesis made when characterizing the tracker, and accordingly the soundness of the methodological approach adopted.

These promising results, without claiming to be thorough, have encouraged us to develop our approach towards the estimation of the complete pose (position and attitude) of the satellite. In fact, in our approach the position of the satellite is supposed to be known, being provided in real time by using GPS/DGPS devices. General approaches aiming at jointly estimating position and attitude usually make use of epipolar geometry (Sect. 2.4). However, the depth relief of Earth Surface in the camera FOV is very close to a plane, making the scene configuration *degenerate* for the epipolar geometry (35), so that preliminary tests gave unstable results for both position and orientation. On the other hand, approximating the Earth Surface as a plane can lead to estimate both the position and the orientation, but affecting accuracy. We are currently evaluating the accuracy of some approaches proposed in literature (83, 84, 85).

Chapter 5

Final Conclusions

This dissertation has concerned the application of an hybrid image registration method to two challenging case studies, requiring at the same time high accuracy and on-line performances with limited computational resources. The motivations of this research arise from the *need of generality* that a method should have in different working conditions while preserving, at the same time, a good trade off between accuracy and computational costs. Accordingly, the applications treated by our study belong to very different applicative domains, but have in common compelling requirements about the registration accuracy, their degree of automation and the computational resources available.

The first case study refers to the automatic on-line mosaicing of optical microscope imagery, while the second aims at estimating the three-dimensional orientation of a satellite through image registration of Earth images acquired by a camera mounted on board. Despite of their different applicative domain, both these research projects have been developed following a common methodological approach.

A model of the studied problem has been initially designed, taking into account configurations that could be critical for the application. In particular, simulations with the theoretical model and with synthetic data have been useful to validate hypotheses about the model and/or identify error sources. This has permitted to better design experiments in well controlled configurations, in order to validate, in collaboration with expert partners, the results achieved.

It is worth remarking that, following this methodological approach, for both these two cases, the application of our image registration algorithm has proved additional functionalities and improvements with respect to approaches currently employed in the respective research field. The contributions in terms of originality and performance of

5. FINAL CONCLUSIONS

our approach are summarized more specifically with regard to the two case studies in the following paragraphs.

5.1 Automatic On-line Mosaicing of Optical Microscopy Imagery: the ADVANCE project

Current approaches for mosaicing of optical microscope images aim at building a complete representation of the biological specimen investigated using motorized holders and performing in *batch mode*. With respect to these approaches, our method does not require automated equipment, working even with old-generation not motorized optical microscopes, and works on-line with consumer hardware, widening the microscope field of view and permitting the *interactive* navigation of the whole cell culture. It has to be noticed that this application is motivated by the real demand of microscopists and biologists for high content high throughput image analysis, since it permits to observe more complex spatial patterns referred to the whole extent of the cell culture.

In such an application, several problems had to be faced in order to preserve photometric and geometric consistency. Several components, as the microscope and the camera optical layout, the holder mechanical setup, the typology and the state of the investigated sample, etc., interact in complex ways affecting the final result. Rather than studying separately these components, we have studied their *effects* in term of photometric and geometric distortion at system level.

The inhomogeneous illumination response of the system has been compensated characterizing the system illumination field. In particular, vignetting and shading effects have been identified as the predominant components responsible for possible “tonal seams”, and normalization according to the estimated illumination field has proved to cope with this issue.

Also the different causes that can lead to geometric mis-registration of the images have been analyzed at system level, using robust model estimation to take into account scene content modifications, being imaged live structures. The model estimation problem has been discussed in depth, emphasizing how less complex models can generally give more stable and robust solutions. Moreover, a strategy to compensate also the geometric drift, due to the pairwise registration approach, while preserving at the same time on-line capabilities, has been successfully applied.

This approach has proved to be effective with different biological samples, providing

seamless visually pleasant mosaics. To assess in a quantitative manner the quality of the achieved mosaics, an original evaluation approach has been conceived, using the acquired images directly as ground-truth data. Tests performed on sequences of different content have assessed the consistency of this methodology.

It is worth remarking that the research carried on in this project has led to the development of a research prototype that is currently being validated from expert biologists at the Bone Regeneration Laboratory, Istituti Ortopedici Rizzoli, Bologna, Italy. At the same time, scientific works related to this study have been accepted (1) or are in preparation (2) for peer reviewed conferences proceedings and journals. Our collaboration to this project is still active in order to test and validate our approach with a wider range of cellular patterns, different for morphology, density, appearance, etc.

5.2 Camera pose estimation: the STARS project

This research application has emphasized the need of high accuracy in pose estimation problems. In fact it is fundamental that the satellite attitude is *measured* with high accuracy (at the order of 1 arcsec) in order to perform a correct stabilization during its orbit. We have conceived an original vision-based approach that has the advantage of exploiting natural patterns present on Earth images acquired by an on-board camera, avoiding drawbacks typical of traditional approaches (i.e. *star trackers*). Moreover, by registering terrestrial images, our approach is capable to estimate the satellite attitude with comparable and potentially more accurate performance.

In order to reach this accuracy, several methodological stages have been devised, this representing in itself an original contribution to this research field. In fact, a physical model of the system has been initially conceived and the corresponding theoretical performances analyzed.

At a subsequent level, the design of a realistic simulation environment with orbital ground-truth data, in partnership with experts in the field, has permitted to identify more clearly the error sources, and in particular the ones deriving from the image registration method. In fact, thanks to this approach, realistic image sequences, that are related to the terrestrial scene that would be viewed by a camera integral with the satellite during its orbit, can be generated.

Moreover, this has permitted to characterize the accuracy performance of the registration algorithm, relating tracking errors in the image domain (at sub-pixel level) to angular errors in attitude estimation. This has yielded the estimation of an error curve

5. FINAL CONCLUSIONS

that can be potentially employed for error compensation, even avoiding potential drift effects.

Furthermore, our general simulation environment has permitted to devise different simulation configurations in order to characterize separately the error sources and their dynamics. In fact, in such research projects, it is fundamental to characterize, when possible, the main error sources, being real tests not always possible on the field.

To test the hypotheses formulated in the previous stages, we have conceived a more consistent assessment methodology. With respect to common approaches that make use of indirect image-based metrics to evaluate the attitude estimation error, in our case the ground truth data about camera orientation have been provided directly by a Numerical Control Unit. This has permitted to perform in controlled conditions and with more flexibility particular test configurations that are of interest, and can constitute in general a benchmark for other registration approaches. Experiments with this test-bed have shown high accuracy performance for our registration method.

This research project has led to the publication of several scientific works on peer reviewed conferences proceedings (3, 4, 5, 6), and other works are currently in preparation (7).

I am currently working at the Computer Vision Group - ARCES, University of Bologna, under a research grant in the context of the PERFECT project. It concerns image registration of Computed Tomography imagery for motion artifacts compensation in liver perfusion.

Bibliography

- [1] L. CAROZZA, A. BEVILACQUA, AND F. PICCININI. **An Incremental Method for Mosaicing of Optical Microscope Imagery.** In *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (to appear)*, 2011. [6](#), [41](#), [97](#)
- [2] A. BEVILACQUA, A. GHERARDI, L. CAROZZA, AND F. PICCININI. **On-line Mosaicing of Optical Microscope Imagery.** to be submitted to the *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 11)*, 2011. [6](#), [97](#)
- [3] A. BEVILACQUA, L. CAROZZA, AND A. GHERARDI. **A novel vision-based approach for autonomous space navigation systems.** *International Symposium on Visual Computing (ISVC)*, **5876**:837–846, 2009. [7](#), [98](#)
- [4] A. BEVILACQUA, L. CAROZZA, AND A. GHERARDI. **An image registration approach for accurate satellite attitude estimation.** *International Symposium on Visual Computing (ISVC)*, **5876**:827–836, 2009. [7](#), [98](#)
- [5] A. BEVILACQUA, A. GHERARDI, AND L. CAROZZA. **A vision-based approach for high accuracy assessment of satellite attitude.** *Proceedings of the 5th IEEE International Workshop on Dynamical Vision at ICCV 2009 (oral)*, pages 743–750, 2009. [7](#), [98](#)
- [6] A. BEVILACQUA, C. BIANCHI, L. CAROZZA, A. GHERARDI, N. MELEGA, D. MODENINI, AND P. TORTORA. **Standalone three axis attitude determination from Earth Images.** *Proc. of the 20th Space Flight Mechanics Meetings*, **136**, 2010. [7](#), [98](#)
- [7] L. CAROZZA, A. BEVILACQUA, AND P. TORTORA. **Standalone three axis attitude determination from Earth Images.** to be submitted to *Journal of Guidance, Control and Dynamics*, 2011. [7](#), [98](#)
- [8] B. ZITOVA AND J. FLUSSER. **Image registration methods: a survey.** *Image and Vision Computing*, **21**(11):977–1000, 2003. [10](#)
- [9] A. HAJA, B. JAHNE, AND S. ABRAHAM. **Localization accuracy of region detectors.** In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008. [10](#)
- [10] SONG-CHUN ZHU, CHENG EN GUO, YIZHOU WANG, AND ZIJIAN XU. **What Are Textons?** *International Journal of Computer Vision*, **62**(1):121–143, 2005. [10](#)
- [11] T. TUYTELAARS AND K. MIKOLAJCZYK. **Local Invariant Feature Detectors: A Survey.** *Foundations and Trends in Computer Graphics and Vision*, **3**:177–280, 2008. [10](#), [11](#)
- [12] K. BRIECHLE AND U. D. HANEBECK. **Template matching using fast normalized cross correlation.** *Proceedings of SPIE, Optical Pattern Recognition XII*, **4387**, 2001. [11](#)
- [13] P. TRESADERN AND I. REID. **An evaluation of Shape Descriptors for Image Retrieval in Human Pose Estimation.** In *18th British Machine Vision Conference*, **2**, pages pp. 800–809, 2007. [11](#)
- [14] A. BEVILACQUA, A. GHERARDI, AND L. CAROZZA. **An automatic system for the real time characterization of vehicle headlamp beams exploiting image analysis.** *IEEE Transactions on Instrumentation and Measurement (TIM) 2010*, **59**(10):2630–2638, 2010. [11](#), [89](#)
- [15] A. BEVILACQUA, A. GHERARDI, AND L. CAROZZA. **A visual perception approach for accurate segmentation of light profiles.** *2009 LNCS International Conference on Image Analysis and Pattern Recognition (ICIAR)*, **5627**:168–177, 2009. [11](#)
- [16] D. G. LOWE. **Object recognition from local scale-invariant features.** In *Proceedings of the International Conference on Computer Vision*, pages 1150–1157, 1999. [11](#), [26](#)
- [17] H. BAY, A. ESS, T. TUYTELAARS, AND L. VAN GOOL. **Speeded-Up Robust Features (SURF).** *Computer Vision and Image Understanding archive*, **110**(3):346–359, 2008. [11](#), [64](#)
- [18] M. POLLEFEYS, L. VAN GOOL, M. VERGAUWEN, F. VERBIEST, K. CORNELIS, J. TOPS, AND R. KOCH. **Visual modeling with a Hand-Held Camera.** *International Journal of Computer Vision*, **59**(3):207–232, 2004. [12](#)
- [19] M. VERGAUWEN AND L. VAN GOOL. **Web-based 3D reconstruction service.** *Machine Vision and Applications*, **2**(17):411–426, 2006. [12](#)
- [20] D. BULATOV. **Metric reconstruction from video sequences.** *Pattern Recognition and Image Analysis*, **18**:300–308, 2008. [12](#)
- [21] N. GHEISSARI AND A. BAB-HADIASHAR. **A comparative study of model selection criteria for computer vision applications.** *Image and Vision Computing*, **26**(12):1636–1649, December 2008. [12](#), [33](#), [34](#), [50](#)
- [22] D. HASLER AND S. SSTRUNK. **Mapping colour in image stitching applications.** *Journal of Visual Communication and Image Representation*, **15**:65–90, 2004. [12](#)
- [23] P. THEVENAZ, T. BLU, AND M. UNSER. **Image Interpolation and Resampling.** In *Handbook of Medical Imaging, Processing and Analysis*, pages 393–420. Academic Press, 2000. [12](#)
- [24] M. FRIGO AND S. G. JOHNSON. **The Design and Implementation of FFTW3.** *Proceedings of the IEEE, Special Issue on Program Generation, Optimization, and Platform Adaptation.*, **93**(2):216–231, 2005. [13](#)

BIBLIOGRAPHY

- [25] B. D. LUCAS AND T. KANADE. **An Iterative Image Registration Technique with an Application to Stereo Vision.** In *Proceedings of the 1981 DARPA Image Understanding Workshop*, pages 121–130, April 1981. [14](#)
- [26] J. SHI AND C. TOMASI. **Good Features to Track.** *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 593–600, June 1994. [14](#), [64](#)
- [27] E. DE CASTRO AND C. MORANDI. **Registration of Translated and Rotated Images using Finite Fourier Transforms.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**(5):700–703, 1987. [16](#)
- [28] Y. KELLER, A. AVERBUCH, AND M. ISRAELI. **Pseudo-Polar Based Estimation of Large Translations Rotations and Scalings in Images.** *IEEE Transactions on Image Processing*, **14**(1):12–22, 2005. [16](#)
- [29] Y. KELLER AND A. AVERBUCH. **A projection-based extension to phase correlation image alignment.** *Signal Processing*, **87**(1):124–133, January 2007. [16](#)
- [30] H. FOROOSH, J. B. ZERUBIA, AND M. BERTHOD. **Extension of Phase Correlation to subpixel registration.** *IEEE Transactions on Image Processing*, **14**(1):12–22, 2002. [16](#)
- [31] H. SHEKARFOROUSH, M. BERTHOD, AND J. ZERUBIA. **Subpixel image registration by estimating the polyphase decomposition of cross power spectrum.** In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 532–537, June 1996. [16](#)
- [32] B. K. P. HORN. **Determining Optical Flow.** *Artificial Intelligence*, (1-3):185–203, August 1981. [16](#)
- [33] JEAN-YVE BOUGUET. **Pyramidal Implementation of the Lukas Kanade Feature Tracker: Description of the algorithm.** Technical report, Intel Research Laboratory, 1999. [17](#)
- [34] R. HARTLEY AND A. ZISSERMAN. *Multiple view geometry in computer vision*. Cambridge Academic Press, 2003. [18](#), [19](#), [20](#), [31](#), [61](#), [64](#)
- [35] A. ZISSERMAN AND S. MAYBANK. **A Case Against Epipolar Geometry.** In J. MUNDY, A. ZISSERMAN, AND D. FORSYTH, editors, *Applications of Invariance in Computer Vision*. Springer-Verlag, 1994. [18](#), [94](#)
- [36] P. H. S. TORR. **Model selection for two view geometry: A review.** *Shape Contour and Grouping in Computer Vision*, pages 822–822, 1998. [18](#)
- [37] R. I. HARTLEY. **In Defense of the Eight-Point Algorithm.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(6):580–593, 1997. [19](#)
- [38] K. MADSEN, H. B. NIELSEN, AND O. TINGLEFF. **Methods For Solving Non-linear Least Squares Problems.** Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, 2004. [19](#)
- [39] F. DORNAIKA AND R. CHUNG. **Mosaicking images with parallax.** *Signal Processing: Image Communication*, **19**(8):771–786, 2004. [19](#)
- [40] MARTIN A. FISCHLER AND ROBERT C. BOLLES. **Random Sample and Consensus: A Paradigm for Model Fitting with application to Image Analysis and Automated Cartography.** *Comm. of the ACM*, **24**(6):381–395, 1981. [20](#), [32](#), [64](#)
- [41] T. R. JONES, A. E. CARPENTER, D. M. SABATINI, AND P. GOLAND. **Methods for high-content, high-throughput image-based cell screening.** In *Proceedings of the Workshop on Microscopic Image Analysis with Applications in Biology, held in association with MICCAI06 (Medical Image Computing and Computer-Assisted Intervention)*, pages 65–72, 2006. [22](#)
- [42] R. WOLLMAN AND N. STURMAN. **High throughput microscopy: from raw images to discoveries.** *Medical Image Analysis*, (120):3715–3722, 2007. [22](#)
- [43] D. J. ROMER, K. H. YEARSLEY, AND L. W. AYERS. **Using a modified standard microscope to generate virtual slides.** *Anat. Rec. (Part B: New Anat.)*, **272**(3):91–97, 2003. [22](#), [23](#)
- [44] K. LI, E. D. MILLER, M. CHEN, T. KANADE, L. E. WEISS, AND P. G. CAMPBELL. **Cell population tracking and lineage construction with spatiotemporal context.** *Medical Image Analysis*, **12**(5):546–566, October 2008. [22](#)
- [45] D. L. PRICE, S. K. CHOW, N. A. B. MACLEAN, H. HAKOZAKI, S. T. PELTIER, M. E. MARTONE, AND M. H. ELLISMAN. **High-Resolution Large-Scale Mosaic Imaging Using Multiphoton Microscopy to Characterize Transgenic Mouse Models of Human Neurological Disorders.** *Neuroinformatics*, **1**(4):65–80, 2006. [22](#)
- [46] J. MEYER. *Multimodal Image Registration for Efficient Multi-resolution Visualization*, pages 137–151. Springer, 2007. [22](#)
- [47] M. GAGNON, R. TAYLOR, J. V. LANGE, T. LEE, C. COLLINS, R. DRAUT, AND E. KUJAWSKY. **Automated Microscopic Image Acquisition, Compositing and Display - US Patent 7027628B1**, 2006. [23](#)
- [48] R. MARZOTTO, A. FUSIELLO, AND V. MURINO. **High Resolution Video Mosaicing with Global Alignment.** In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2004*, **1**, pages 692–698, 2004. [25](#)
- [49] M. BROWN AND D. LOWE. **Automatic Panoramic Image Stitching using Invariant Features.** *International Journal of Computer Vision*, **74**(1):59–73, 2007. [25](#), [26](#)
- [50] H. SHUM AND R. SZELISKI. **Construction and refinement of panoramic mosaics with global and local alignment.** In *IEEE International Conference on Computer Vision*, pages 953–958, 1998. [25](#)
- [51] S. COORG AND S. TELLER. **Spherical Mosaics with Quaternions and Dense Correlation.** *International Journal of Computer Vision*, **37**(3):259–273, 2000. [25](#)
- [52] P. F. MCLAUCHLAN AND A. JAENICKE. **Image mosaicing using sequential bundle adjustment.** *Image and Vision Computing*, **20**(1-2):751–759, August 2002. [25](#)
- [53] S. R. OTT. **Acquisition of high resolution digital images in video microscopy: automated image mosaicing on a desktop microcomputer.** *Microscopy research and techniques*, **38**(3):335–339, 1997. [26](#)

- [54] P. THEVENAZ AND M. UNSER. **User-friendly semiautomated assembly of accurate image mosaics in microscopy.** *Microscopy research and techniques*, (70):131–146, 2007. 26
- [55] C. SUN, R. BEARE, V. HILSENSTEIN, AND P. JACKWAY. **Mosaicing of Microscope Images.** In *Proceedings of IEEE Digital Image Computing: Techniques and Applications 2005 (DICTA '05)*, pages 343–348, 2005. 26
- [56] S. K. CHOW, H. HAKOZAKI, D. L. PRICE, N. A. B. MALLEAN, T. J. DEERNICK, J. C. BOUWER, M. E. MARTONE, S. T. PELTIER, AND M. H. ELLISMAN. **Automated microscopy system for mosaic acquisition and processing.** *Journal of Microscopy*, 222(2):76–84, 2006. 26
- [57] F. LU AND S. XIA. **Microscopic Image Mosaicing Algorithm Based on Normalized Moment of Inertia.** *Advanced Intelligent Computing Theories and Applications, LNCS Communications in Computer and Information Science*, 2(21):1010–1017, 2007. 26
- [58] C. HARRIS AND M. STEPHENS. **A combined corner and edge detector.** In *Proceedings of the 4th Alvey Vision Conference, Manchester, UK, 31 August - 2 September, 1*, pages 147–151, 1988. 26
- [59] B. MA, T. ZIMMERMANN, M. ROHDE, S. WINKELBACH, F. HE, W. LINDENMAIER, AND K. E. J. DITTMAR. **Use of Autostitch for automatic stitching of microscope images.** *Micron*, (38):492–499, 2007. 26
- [60] W. J. HSU, W. F. PAUL POON, AND Y. N. SUN. **Automatic seamless mosaicing of microscopic images: enhancing appearance with color degradation compensation and wavelet based blending.** *Journal of Microscopy*, 231(3):408–418, 2008. 26
- [61] F. WINKELMAN AND I. PATRAS. **Online globally consistent mosaicing using an efficient representation.** In *IEEE International Conference on Systems, Man and Cybernetics*, 4, pages 3116–3121, 2004. 27, 35
- [62] A. BEVILACQUA AND P. AZZARI. **A fast and reliable image mosaicing technique with application to wide area motion detection.** In *Proceedings of International Conference on Image Analysis and Recognition (ICIAR'07)*, 4633, pages 501–512, 2007. 27, 35
- [63] A. GHERARDI, A. BEVILACQUA, AND F. PICCININI. **Illumination field estimation through Background Detection in Optical Microscopy.** In *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (to appear)*, 2011. 28
- [64] D. B. MURPHY. **Phase contrast microscopy.** *Fundamentals of Light Microscopy and Electronic Imaging*, pages 97–112, 2001. 29
- [65] Y. KANAZAWA AND K. KANATANI. **Stabilizing image mosaicing by model selection.** *Lecture Notes in Computer Science*, 2018:35–51, 2000. 33
- [66] A. AKAIKE. **A new look in the statistical model identification.** *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. 34
- [67] K. KANATANI. **Model selection for geometric inference.** In *Proceedings of the Fifth Asian Conference on Computer Vision*, 1, pages xxi–xxxii, 2002. 34
- [68] V.L.PISACANE AND R.C.MOORE. *Fundamentals of Space Systems, 2nd ed.*, chapter 5. Oxford University Press, 2005. 57
- [69] F. CABALLERO, L. MERINO, J. FERRUZ, AND A. OLLERO. **Vision-Based Odometry and SLAM for Medium and High Altitude Flying UAVs.** *Journal of Intelligent and Robotic Systems*, 54:137–161, 2009. 57
- [70] G.CONTE AND P.DOHERTY. **An integrated UAV navigation system based on aerial image matching.** *Proceedings of the IEEE Aerospace Conference*, pages 1–10, 2008. 57
- [71] C. RUOCCHIO, D. ACCARDO, G. RUFINO, S. MATTEI, AND A. MOCCIA. **Development and testing of a Fully Autonomous Star Tracker.** *2nd IAA Symposium on Small Satellites for Earth Observation, April 12-16, Berlin, Germany*, April 1999. 58
- [72] M. KOLOMENKIN, S. POLLAK, I. SHIMSHONI, AND M. LINDENBAUM. **Geometric voting algorithm for star trackers.** *Aerospace and Electronic Systems, IEEE Transactions on*, 44(2):441–456, April 2008. 58
- [73] J.L. JORGENSEN, G.G. THUESEN, M.BETTO, AND T.RIIS. **Radiation impacts on star-tracker performance and vision systems in space.** *Acta Astronautica*, 46(2-6):415–422, 2000. 2nd IAA International Symposium on Small Satellites for Earth Observation. 58
- [74] G.D. ROGERS, M.R. SCHWINGER, J.T. KAIDY, T.E. STRIKWERDA, R. CASINI, A.LANDI, R. BETTARINI, AND S. LORENZINI. **Autonomous star tracker performance.** *Acta Astronautica*, 65(1-2):61–74, 2009. 58
- [75] C. D. HALL. *Spacecraft Attitude Dynamics and Control*. February 2003. 61, 63
- [76] B. K. P. HORN. **Closed Form solution of absolute orientation using unit quaternions.** *Journal of the Optical Society of America A*, 4(4):629–642, 1987. 63
- [77] L.VICCI. **Quaternions and Rotations in 3-Space: The Algebra and its Geometric Interpretation.** *Technical Report TR01-014, Dept. of Computer Science, University of North Carolina at Chapel Hill*, 2001. 63
- [78] K.ROHR AND C.DREWNIOK. **Model-Based Landmark Extraction and Correspondence Finding for Aerial Image Registration.** *Video Registration, Chapter 5*, 2003. 64
- [79] R.CARCERONI, A.KUMAR, AND K.DANILIDIS. **Structure from motion with known camera positions.** *IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, 1:477–484, 2006. 66
- [80] W. BURGER AND B. BHANU. **Estimating 3D egomotion from perspective image sequences.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1040–1058, November 1990. 81
- [81] J. Y. BOUGUET. **Camera Calibration Toolbox for Matlab.** *Supplied by Computer Vision Research Group, Dept. of Electrical Engineering, California Institute of Technology*. 89
- [82] AMERICAN SOCIETY OF PHOTOGRAMMETRY AND REMOTE SENSING. *Manual of Photogrammetry. Fifth Edition*. 2004. 90

BIBLIOGRAPHY

- [83] C.-P. LU, G. D. HAGER, AND E. MJOLSNES. **Fast and Globally Convergent Pose Estimation from Video Images**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(6):610–622, 2000. [94](#)
- [84] G. SCHWEIGHOFER AND A. PINZ. **Fast and globally convergent structure and motion estimation for general camera models**. In *17th British Machine Vision Conference*, pages 147–156, 2006. [94](#)
- [85] E. MALIS AND V. MANUEL. **Deeper understanding of the homography decomposition for vision-based control**. Technical report, INRIA, 2007. [94](#)

List of Figures

3.1	Examples of particular spatial cell patterns. (Left) . Two cell colonies can be identified (blue separation lines) in this cell culture. (Right) . Typical pattern for cell necrosis (with circular lipid vesicles)	23
3.2	(a) Image of a cell culture acquired in phase contrast mode and in fluorescence mode after coloring with different cell markers identifying nuclei and cytoplasm of all the cells (b) , nuclei of necrotic cells (c) , nuclei of the active cells (d) , respectively. (e) Image obtained by multi-modality fusion of the registered images. Functional properties (live, apoptotic cells, etc.) can be studied according to visual features (cell morphology, etc.).	24
3.3	(a) Image resulting from averaging on the empty field images stack. (b) Mesh of the estimated empty field, assuming a Gaussian-like shape. . . .	28
3.4	(a) , (b) Eigenvalues maps for a sample image pair of a histological sample. As it can be seen, higher eigenvalues are often located near the borders of cellular membranes. (c) , (d) Corresponding features resulting from the registration algorithm in the two views are associated through a common numerical identifier.	30
3.5	Couple of consecutive frames with corresponding feature points superimposed (with common numerical identifier). Circled crosses identify inliers of the RANSAC stage, isolated crosses indicates outliers. Images with magnified details show how the RANSAC stage is able to cope with outliers even due to local morphological changes.	33
3.6	Experimental setup used in our tests. The inverted microscope Nikon Eclipse TE2000-U has been equipped with a digital camera (Nikon DXM1200). The system has been connected to a consumer PC.	37

LIST OF FIGURES

3.7	Explicative figure for the symbols used to indicate the regions defined to compute quality metric indexes. In this case $m_i=3$. Dimensions are not preserved purposely to improve visualization.	38
3.8	Labeled masks of the <i>hit regions</i> $R_M^i(x, y)$, being each numerical identifier associated with the corresponding i^{th} warped reference frame ($R_I^i(x, y)$). Blue borders surround $R_M^8(x, y)$. The support used to compute metric indexes for the eighth frame is formed only by the (not green) mosaic areas hit from the ninth, the tenth and the eleventh reference images. .	40
3.9	Mosaics obtained from the sequence HS using F2F registration with flat field correction for the translative (a) and affine models (b) . Detail of the mosaic achieved without applying the flat field compensation ((c)) compared with the correspondent tonally-aligned region (d) . Seams due to vignetting and shading effects can be seen (better on the electronic version) in (c) , while disappear in (d) thanks to the flat field correction.	42
3.10	(a) Mosaic obtained from the tonally aligned sequence HS using the projective model. (b) , (c) An image pair with a small number of features concentrated in region. This configuration can make the model estimation an ill-conditioned problem for complex models.	43
3.11	Mosaics obtained registering the tonally compensated SC-1 sequence in the F2F mode using the translative (a) , affine (b) and projective model (c) , respectively.	46
3.12	(a) Resulting mosaic for the affine model, employing flat field normalization and F2M registration on the sequence SC-2. (b) , (c) Details of the mosaics obtained with F2F registration only and with the application of F2M registration. At the center of the red boxes seams due to the geometric drift (b) and their compensation due to F2M registration (c) are clearly visible.	48
3.13	(a) , (b) Values of the RMSE index for each reference frame for the translative and affine models, respectively, for F2F e F2M registration of the sequence SC-2. (c) , (d) Values of the SNR index for each reference frame for the translative and affine models, respectively, for F2F e F2M registration of the sequence SC-2.	50
3.14	GIC values for the HS (a) and SC-2 (b) sequences registered according to F2F mode.	51
4.1	(a) Reference frames adopted to model the framework. (b) Satellite reference frame with attitude perturbations.	60

4.2	Diagram of the procedure used to generate the EOIs sequences. The ground-truth pose data are used to geo-reference the camera field of view, assigning to each pixel the geographical coordinates of the points seen on Earth. Image interpolation on a geo-referenced satellite database is finally used to generate the EOIs couples to be registered.	68
4.3	Image from NASA Blue Marble (left) and Landsat7 databases (right).	73
4.4	(Left) Ground-truth attitude trends and (right) the corresponding F2F ground-truth angular offsets for the simulation S1, expressed with respect to the LORF.	74
4.5	(Left Column) F2F attitude errors values for the S1-NOUPD using computed matchings. (Right Column) Histograms of the F2F attitude errors for S1-NOUPD using computed matchings.	75
4.6	(a) Trend of the X coordinate of a sample computed corner (green circles) and its value rounded at the second decimal digit (blue stars), and the X tracking error due to rounding (c) , for sample pixels. (b) Trend of the Y coordinate of the sample computed corner (green circles) and its value rounded at the second decimal digit (blue stars), and the Y tracking error due to rounding, for sample pixels (d)	77
4.7	Attitude error trends obtained rounding the computed corners at the second (left column) and first (right column) digit.	78
4.8	(a) A sample image and the corresponding normalized map (b) of the eigenvalues of the pseudo-hessian for the extraction of the corner points. Couple of consecutive frames with the extracted (c) and tracked (d) corner points.	79
4.9	Left Column. F2F attitude errors values for the S1-NOUPD. Right column. Histograms of the F2F attitude errors for S1-NOUPD.	80
4.10	Histograms of δV_x^i (a) and δV_y^i (b) for a sample frame couple from the sequence S2-ROLL-GEO1 (with step $\Delta = 1$ arcsec).	82
4.11	Temporal trends of $\delta V_x(t')$ (a) and $\delta V_y(t')$ (b) for the sequence S2-ROLL-GEO1 (with step $\Delta = 1$ arcsec).	83

LIST OF FIGURES

4.12	(a) Sub-pixel tracking error curve obtained by fitting tracking error samples from different simulations, against computed motion vectors \bar{V}^c , with $\bar{V}^c = \bar{V}_y^c$ for PITCH perturbations (in cyan) and $\bar{V}^c = \bar{V}_x^c$ for ROLL perturbations (other colors). Usually bicubic interpolation is used for EOIs generation, but similar behavior is encountered also when bilinear interpolation is employed (in green). Light magenta circles surrounds samples obtained from the simulation S1-NOUPD, which still belong to the estimated sub-pixel tracking error curve. (b) Sample values related to the roll error for the simulations S2-ROLL-GEO2 are reported against the computed motion vectors $\bar{V}^c = \bar{V}_x^c$. They follow, as expected, the same trend extrapolated from the samples in (a), up to the sign of the error.	85
4.13	Qualitative sketch of the tracker behavior with respect to the computed features is presented. Circles represent <i>computed</i> features, while stars represent <i>tracked</i> corners. The trend of the tracking error is qualitatively similar to the one of the sub-pixel tracking error curve (Fig. 4.12 (a)). . .	87
4.14	Error propagation in the UPD (dashed line) and SP (continuous line) configurations for pitch and yaw angles.	88
4.15	A picture of the NCU system together with the reference test image. NCU reference frame (blue) and world reference frame (red) are also sketched.	89
4.16	A couple of frames from the sequence NCU-ROLL processed by our algorithm.	90
4.17	(a) Correlation among the ground-truth roll perturbations and the ones measured by our system. Absolute (b) and with attitude updating (c) F2F roll error (expressed in arcsec).	91
4.18	Absolute F2F yaw error (expressed in arcsec) resulting from the NCU-YAW test sequence.	92

List of Tables

3.1	Properties of the warping models employed in this context. k is the number of parameters (degrees of freedom) of the model, N_{min} is the minimum number of matchings needed to estimate the model.	32
3.2	Quality metric indexes for the 60-frame HS sequence. The first row concerns results obtained working on raw (not tonally compensated) images, while values in the second row refer to flat field normalized images.	44
3.3	Quality metric indexes for the 26-frame SC-1 sequence. The first row concern results obtained working on raw (not tonally compensated) images, while values in the second row refer to flat field normalized images.	45
3.4	Values of the global quality metric indexes of the SC-2 sequence for the three different models (columns), without (first row) and with (second row) application of F2M registration.	49
4.1	Relation among resolution, acquisition frame rate and overlapping area between consecutive frames.	72
4.2	Parameters for the Earth model, the sensor matrix and the orbit used in our simulations.	72
4.3	Parameters for S2 simulations on synthetic images. Values of \bar{V}_x^c and \bar{V}_y^c induced by the angular perturbation are visualized as rounded to the third decimal digit.	84