Universidade de Aveiro
2003

Paulo Miguel de Jesus Dias

Reconstrução Tridimensional de Ambientes Reais Usando dados Laser e de Intensidade

Three dimensional Reconstruction of Real World Scenes Using Laser and Intensity data

Universidade de Aveiro
2003

Paulo Miguel de Jesus Dias

Reconstrução Tridimensional de Ambientes Reais Usando dados Laser e de Intensidade

Three dimensional Reconstruction of Real World Scenes Using Laser and Intensity data

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica do Dr. Francisco Vaz, Professor Catedrático do Departamento de Engenharia Electrónica e Telecomunicações da Universidade de Aveiro e do Dr. Vítor Sequeira, Investigador do Centro Comum de Investigação da Comissão Europeia em Ispra, Itália.

o júri

presidente — Reitor da Universidade de Aveiro

Doutor Francisco António Cardoso Vaz
Professor Catedrático da Universidade de Aveiro

Doutor João Guilherme Miranda Gonçalves
Professor Associado da Universidade de Aveiro

Doutor Joaquim João Estrela Ribeiro Silvestre Madeira
Professor Auxiliar da Universidade de Aveiro

Doutor Robert Fisher
Professor da School of Informatics, Universidade de Edinburgh

Doutor Vítor Sequeira (Co-Orientador)
Investigador do Joint Research Centre, Ispra (Itália)

acknowledgements

First of all, I want to thank Dr. Vítor Sequeira who supervised my work during the 4 years I worked at the Joint Research Centre in Italy. His knowledge about 3D reconstruction, his scientific competence, his dedication, and the many discussions we had, have been of extreme importance for this work and this thesis.

My thanks also to Prof. Francisco Vaz who trusted me five years ago and offered me the possibility to start my PhD in a new and exciting area: 3D reconstruction. He has always been available each time I needed to discuss (most of time through internet) or solve the many scientific and administrative problems that occurred along these last years.

I am also grateful to Prof. João Gonçalves who gave me, with Prof. Vaz, the possibility to start this adventure in November 1998 in his working group. He was always paying attention to the evolution of my work and the many discussions and ideas we exchanged have been of great importance for this work. Working within his group has been a beautiful experience for me.

To Fernand Sorel and André Poucet who have been successive heads of the unit where my work was integrated within the Institute for the Protection and Security of the Citizen. They provided all the support I needed and the facilities that made possible this work.

I also want to thank André Poucet and Sergio Contini for giving me access to the satellite images and maps that are presented in Appendix B. My thanks to Matteo Sgrenzaroli who provided the digital photographs of the San Stefano church presented in chapter 6.

To EC-Camera Research and Training Network (contract number ERB FMRX-CT97-0127) who supported my first year in Italy. In addition, this network also gave me the opportunity to exchange ideas with people working in the same field and with many other young scientists. This cooperation and these meetings were most valuable to share concepts and develop new ideas that appear along this work.

To Institute of Electronics and Telematics Engineering of Aveiro (IEETA) who gave me the time and means to achieve the redaction of my thesis in 2003.

My gratitude also to the Portuguese Foundation of Science and Technology (FCT) that made possible the major part of this work through the PhD grant PRAXIS XXI/BD/19555/99.

Thanks to all the colleagues that have worked with me at the Surveillance and Information Retrieval sector during my years in Italy. Mauro Ruggeri, who provided the data structure used for the segmentation of triangulated models. Marco Fiocco who processed many 3D models for me. And all the others: Calixto Calderon, Emanuele Bovisio, Josep Mouriño, Marco Franzé, Angelo Rossini. A special thank to Elena Stringa who shared the same office for almost three years and to her husband Paul Smits. They have been true friends always available.

acknowledgements

To all the people who introduced me to the beauty of the Alps by giving me the opportunity to hike with them during week ends, in particular, Michael kleith, Roland Hiederer and Gabriele Reina.
Thanks also to the Scouts of Somma Lombardo who gave me the possibility to be a scout leader during my years in Italy.
To Céline, her support has been very important during the redaction of this thesis, despite the distance between us.
Finally, I want to express my deepest gratitude to my family: my parents, my three brothers, Jorge, José, Fernando and my sister Maria. You have always been by my side despite the distance, and had always supported my decisions. I would not have been so far without your help and support. I am fortunate to have you.

resumo

O objectivo do trabalho apresentado nesta tese é a criação de modelos tridimensionais completos e de alta resolução de ambientes reais (informação geométrica e de textura) a partir de imagens passivas de intensidade e de sensores de distância activos. A maior parte dos sistemas de reconstrução 3D são baseados em sensores laser de distância ou em câmaras fotográficas, mas muito pouco trabalho tem tentado combinar estes dois tipos de sensores. A extracção de profundidade a partir de imagens de intensidade é complicada. Por outro lado, as fotografias fornecem informação adicional sobre os ambientes que pode ser usada durante o processo de modelação, em particular para definir, de uma forma precisa, as fronteiras das superfícies. Isto torna os sensores activos e passivos complementares em varios modos e é a ideia de base que motivou o trabalho apresentado nesta tese.

Na primeira parte da tese, concentramo-nos no registro entre dados oriundos de sensores activos de distância e de câmaras digitais passivas e no desenvolvimento de ferramentas para tornar este passo mais fácil, independente do utilizador e mais preciso. No fim, com esta técnica, obtém-se um mapa de textura para os modelos baseado em várias fotografias digitais. O modelo 3D assim obtido é realizado baseado nos dados de distância para a geometria e nas fotografias digitais para a textura. Com estes modelos, obtém-se uma qualidade fotográfica: uma espécie de fotografia de alta resolução em 3D dum ambiente real.

Na segunda parte da tese, vai-se mais longe na combinação dos dados. As fotografias digitais são usadas como uma fonte adicional de informação tridimensional que pode ser valiosa para definir com precisão as fronteiras das superfícies (onde a informação de distância é menos fiável) ou então preencher falhas nos dados ou aumentar a densidade de pontos 3D em áreas de interesse.

abstract

The objective of the work presented in this thesis is to generate complete, high-resolution three-dimensional models of real world scenes (3D geometric and texture information) from passive intensity images and active range sensors. Most 3D reconstruction systems are based either in range finders or in digital cameras but little work tries to combine these two sensors.
Depth extraction from intensity images is complex. On the other hand digital photographs provide additional information about the scenes that can be used to help the modelling process, in particular to define accurate surface boundary conditions. This makes active and passive sensors complementary in many ways and is the base idea that motivates the work in this thesis.
In the first part of the thesis, we concentrate in the registration between data coming from active range sensors and passive digital cameras and the development of tools to make this step easier, more user-independent and more precise. In the end, with this technique, a texture map for the models is computed based on several digital photographs. This will lead to 3D models where 3D geometry is extracted from range data, whereas texture information comes from digital photographs. With these models, photo realistic quality is achieved: a kind of high-resolution 3D photograph of a real scene.
In the second part of the thesis, we go further in the combination between the datasets. The digital photographs are used as an additional source of three-dimensional information that can be valuable to define accurate surface boundary conditions (where range data is less reliable) or even to fill holes in the data or increase 3D point density in areas of interest.

Table of Contents

List of Figures

List of Tables

1 Introduction

3D modelling of real world scenes is an important topic of research with applications in many areas such as virtual museums, game and entertainment, architecture description and modelling, virtual reality, robotic navigation, archaeology, inspection, cultural heritage and many industrial applications like reverse engineering. Modelling is the creation of a computer representation of real world environments that merges together data coming from one or more sensors. The representation of the geometric and texture information of a real scene is a very challenging task due to the acquisition of large-scale data, complexity of the geometry and difficulties to cope with reflectance properties of the objects and variations in the lighting in the scenes.

Two approaches, depending on the type of sensor (laser range scanners or digital cameras), are typically used to face the 3D reconstruction problem.

Laser range scanners provide arrays of distances to several points in a scene: the range images. This representation links directly the acquired data points to the depth in a $2^{1/2}$D sketch: a viewer centred representation with explicit information about surfaces and discontinuities [Marr82].

Digital cameras are used to acquire photographs of the scenes. These photographs are 2D arrays of reflected light from the objects but do not contain any explicit information about depth. Further processing is necessary, to calibrate cameras and compute 3D models.

This thesis aims to demonstrate how “active range sensing” and “passive intensity images” can be registered and combined into a long-range 3D system. The techniques developed enable the combination of both range and intensity data to compute 3D models of large scenes with photo-realistic quality.

1.1 State of the art

In the last decades, 3D reconstruction has become an important research field. As a matter of fact, many techniques and methods have been developed to model real scenes using all three dimensions.

Depending on the sensor, these techniques can be divided in two groups. The first one is composed by active techniques based on range images to measure the environment in three dimensions. Techniques of the second group are passive and get the 3D information from several 2D digital photographs or video sequences. A comparative evaluation of the performance of active and passive 3D vision systems can be found in [El Hakim95].

The following sections will give an overview of these two groups of techniques pointing out their main advantages and limitations.

1.1.1 Active range sensing

There are two main types of active range sensors. Both are characterized by the fact that they emit a signal to the environment and therefore are called active sensors. The first type is based on triangulation (such as structured light systems). It works by projecting laser stripes or a grid of lines into a given surface. 3D position of the object is then recovered from the analysis of the deformation of the image projected by the sensor. The accuracy of the measurement in this case will depend on the configurations of the triangulation system and thus such techniques are well suit for short distance applications (typically 3m).

The second widely used type of sensors is Laser Range Finders (LRF). These sensors emit a pulse into the environment and will measure the distance to the target by analysing the reflected signal. Measurements can be between the emitted and reflected signal (pulsed wave Time-of-Flight LRF), or the difference of phase between emitted and reflected signal (Phase-shift LRF). With this kind of sensors, the accuracy will not depend directly on the distance to the target, making Laser Range Finders suitable for measurements that can go up to kilometres. Given their larger range, LRF are the natural choice for real world modelling. Most of LRF provide a range image (an array of distances to the measured point) and also a reflectance image. This reflectance image is an array containing reflected light for each measured point. Visually, the reflectance image is similar to a 2D infrared photography of the scene that is fully registered with the range data (most of laser range scanners work with infrared lighting).

The first experiments with range sensing appeared in the late seventies [Oshima79] and used active range sensing to get information about the depth and the shape of visible surfaces. Since then, quality, resolution and speed of the sensors have been improved regularly and nowadays many companies are specialized in the commercialisation of such devices. A list of the principal 3D scanners is available in [Boehler02]. The most diffused mid-range laser scanners are those from [iQsun, Callidus, Optech, Riegl, Fröhlich, Cyra, Mensi]. Additional information about the laser scanners used to acquire the data of this thesis is presented in Appendix A.

Some of these scanners are provided with 3D modelling software, but most of them only produce an array of measurements that can be transformed into a cloud of points that contains the spatial position of each measurement. The next paragraphs present several methods that use range imagery to build 3D models of real world environments.

A system was developed in [Sequeira96] that reconstructs the 3D geometric model of a real world scene from multiple range images. The system includes range data acquisition, geometrical feature extraction, registration, integration of multiple views and perception planning to select the next capture-points. This work was extended within the RESOLV project [Sequeira99] to construct textured 3D models from real scenes. It resulted in the Autonomous Environmental Sensor for Telepresence (AEST): a mobile platform that merges together several functions, including range and intensity data acquisition, triangulation, registration of data acquired from different locations, registration of video texture, as well as perception planning and navigation tools.

The VIT group [El-Hakim98] combines in a mobile platform a laser range sensor and several CCD cameras. The global position of the sensors is fixed and well known. Information from the sensors is used to register the different views acquired and compute the global 3D model of the environment.

Another technique is proposed by [Stamos00] who uses a set of 3D range scans and 2D photographs of a scene. From range scans, a 3D model describing the geometry of the scene is created. The position of the 2D cameras is then recovered with respect to the extracted geometry. Finally, the digital photographs are used to render 3D models. The system is optimised for urban scenes with strong geometrical contents.

Range scanners are also used in cultural heritage. The most famous example is the Michelangelo project [Levoy00]. Within this project, a triangulation-based sensor [Cyberware] provides registered range and colour-image data and extract the 3D geometric information of the statue. The main challenge of this project is related to the resolution (the objective was to capture the geometry of the chisel marks which needs a resolution of ¼ of mm) and the size of the final models since the scanned statues can be up to 5m tall. The project results in hundreds of scans and a final model of Michelangelo's David. More recent work [Bernardini02] uses similar methods to produce a 3D model of Michelangelo's Florentine Pietá.

The great Buddha project [Miyazaki00] is another example of the use of range imaging in cultural heritage applications. Its aim is to model the great Buddha of Kamakura. The project considers three modelling aspects. The geometric modelling uses a commercial laser scanner to acquire the 3D geometry of the statue. Image rendering methods are used in the photometric modelling to interpolate new images to texture the model. Finally, environmental modelling consists in displaying the object inside a given environment (a temple around the Buddha in the project).

Another technique uses a range scanner and a CCD camera at a fixed position to generate textured CAD models of outdoors urban environments [Zhao01]. The vehicle is also equipped with a GPS/Odometer based navigation system that provides additional information for the registration of the laser range data during the generation of the CAD model.

All these systems use range data to create a geometric description of the environment. A full model of real world scenes can hardly be produced from a single view. Thus, several range images have to be acquired and registered into the same co-ordinate frame to achieve complete models. The most widely used algorithm to perform this range registration task is the Iterative Closest Point [Besl92].

Most of the systems presented above also acquire digital photographs. These images are not used to compute the geometry but are important in the last phase of the modelling process, to texture the models whose geometry was extracted from range imagery. Some attempts to register automatically range and intensity information exist but normally are limited to constrained situations. In [Smith99], pose estimation is evaluated by matching control points over similar intensity and reflectance images. Range to image registration is

also achieved by [Neugebauer99], but his method only refines an initial solution achieved through interactive selection of correspondences between the two datasets. In [Stamos00], a method based on vanishing lines and geometry constraints (orthogonality, windows) is used to perform the registration. The method is only reliable in scenes with strong geometric contents. Finally, a system was developed by [Kurazume02] to map texture on 3D geometric models using the reflectance image. Pose estimation is computed by establishing correspondences between 3D points and 2D colour edges extracted from the colour images.

Despite these attempts to align automatically image and range information, most of the systems still rely on hard constraints and solve the registration problem by fixing the position and orientation of the camera with respect to the laser scanner or with user interaction.

The possibility to link directly acquired data with its depth constitutes the principal quality of active range sensing as it makes 3D modelling processes more easy. Furthermore, given their working principle, these sensors do not need any texture in the acquired data to perform well and give a uniform cloud of points.

The main disadvantage of laser range scanners are their cost, dimensions and weight. Other problems are related to the limited spatial resolution (the laser's footprint is not an ideal point but an ellipse. Since its surface increases with the distance to the sensor, laser scanners have some difficulties in sharp edges) or the influence of surface properties in the quality of the data. Finally, due to their dimensions and mechanics they are often difficult to operate and calibrate.

1.1.2 3D reconstruction from intensity images

3D reconstruction based on intensity images has been investigated along the last decades probably because it uses a process similar to human vision. These techniques, as the human eyes, acquire several images from the environment and extract depth by analysing the disparity between them. The process is passive since the sensors only measure the reflected natural light without emitting any signal.

Many techniques have been developed to reconstruct 3D environments from intensity images. Photogrammetry [Atkinson96] produces 3D models of high accuracy from digital photographs, but it requires heavy user intervention to compute the models (application of

markers in the environment, manual selection of features, etc.). Other techniques based on stereo or motion, produce 3D models in a more automatic and easy way but are less accurate [Faugeras93]. Since these techniques rely on automatic point matching, the quality of the results is highly influenced by the type of scene, lighting conditions, texture, etc.

In [Faugeras98], the authors use control points in a sequence of video images to recover epipolar geometry and compute the projection matrices for all the camera positions. Information about the scenes (such as known angles or parallel lines) or self-calibration is used to reconstruct the scene up to an unknown affine transformation. Virtual artefacts (pedestrians or trees) are removed from the final texture by combining several images of the sequence with "mosaic" techniques.

Closed image sequences are also the input of the system presented by [Fitzgibbon98b]. It automatically recovers the 3D geometry of a scene by extracting and matching 2D point features over triplets of images. The fundamental matrix and the trifocal tensors between triples of images are then recovered, based on robust techniques. This information provides the 3D projective geometry of the system, necessary to perform 3D reconstruction.

Another technique that combines 2D images to extract 3D information can be found in [Liebowitz99]. This method creates graphical models of a scene from a limited number of images. The main novelty comes from the use of geometric constraints, which are common in architectural scenes (parallelism or orthogonality), to recover the geometry of the scene.

Most of the image based reconstruction methods results in a model of the system in the projective space. Self-calibration techniques are necessary to achieve metric representation from one or several images. These techniques are largely discussed and presented in [Pollefeys99].

The MIT city scanning project [Coorg99] uses hundreds of digital photographs to compute a geometric model of large urban scenes. The position of the acquired images is annotated and several images taken from the same point are grouped into wide mosaics called nodes. The transformation between different nodes is refined based on vanishing points for rotation [Antone00] and a Hough transform for translation [Antone02]. From this information, significant vertical façades are detected and a global model is computed.

Another approach using interactive methods has been developed at the Berkeley University. An interface ("Façade") gives the user the possibility to guide the matching of

lines or other geometrical primitives over several images to get a geometrical description of the scanned scenes [Debevec01]. Recent works are based on the same approach as "Façade" but within a fully automatic system. In these methods, the system uses particular scene constraints (several principal directions) [Werner02] or parameterised blocks (such as windows, doors, buttresses…) [Dick01] as geometrical primitives.

An advantage of these techniques is the high resolution of the digital photographs that permits very accurate results on well-defined targets. The acquisition step is normally easier when compared with laser range scanners. Finally, these techniques provide texture-maps based on real colour-photographs, which makes final results much more realistic and appealing.

Nevertheless, the lack of direct relation between images and depth of the measured object is a major problem for intensity-based techniques. Thus, acquired images need to be rich in texture to achieve precise matching so that interesting points or edges can be easily identified. As a result, 3D datasets will have varying density depending on the texture in the images. Another limitation is linked to the fact that, being passive, these sensors are affected by changes in illumination.

1.2 Proposed approach

3D reconstruction of real world scenes is a very challenging task due to the variety of possible scenes to scan and difficulties to work with real data. Passive and active techniques used in 3D reconstruction have their limitations and, separately, none of these techniques can solve all the problems inherent to the modelling of real environments. Nevertheless, the techniques are complementary in many ways as shown in Table 1-1.

	Range	Intensity
Cost	Laser Range Finders are expensive sensors	Low cost since any digital camera can be used
Acquisition	Often difficult with large sensors	Easy, with a digital camera
Resolution	Limited spatial resolution	High-resolution digital photographs
Texture map	No colour texture map, or black and white reflectance	Possibility to provide a realistic colour texture map
Lighting	Independent from external lighting	Highly dependent on lighting conditions
Texture relevance	No need of texture in scene	Texture is crucial for good results
3D processing	Provide directly 3D measurements	Difficult to extract 3D depth from images

Table 1-1: Advantages (shaded) and limitations of range and intensity based tecnhiques

The above list makes clear how range and intensity images can complement each other. Thus, our aim is to combine them to compensate the limitations of each sensor with the advantages of the other one. Some works already combine range and intensity for 3D reconstruction. An example is the work proposed by [El-Hakim02], who acquires both range and intensity to achieve 3D reconstruction of real scenes. Each data is treated separately and the registration is done interactively in a final step of the modelling process. In the Michelangelo's Florentine Pietá project [Bernardini02], a range sensor is combined with a fixed custom-made lighting system. The resulting photometric data is used to compute consistent normals and colours, and increase the resolution of the final model [Rushmeier99].

In our technique, the data is combined in an earlier step. Range data is used to guide and correct the camera calibration and many tools have been developed to automate this step. Registered intensity data can then be used to help the range segmentation. The main innovation is the use of intensity images not only as texture for the models, but also as an additional source of 3D information that can be used to refine or complete an initial model, computed from range imagery.

1.3 Original contributions

The main contributions of this thesis can be summarized as follows:

- A complete set of tools has been developed for calibration of range and intensity data. Depending on the images, many different techniques can be used to guide the camera calibration process.
- A re-projection procedure has been created to associate intensity image pixels with reflectance images according to the computed camera models. The resulting images are used as texture-maps and permit to check rapidly the quality of the registration.
- An iterative calibration technique is used to tune the initial calibrations by computing new correspondences in the re-projected images.
- Complete texture maps for the models are obtained from several registered intensity images. Z-buffering and blending are considered when computing the final texture maps.

- Geometric features, extracted from a single registered image, are used to guide and improve the segmentation of lines and quadratics in a triangulated model.
- A procedure has also been developed to extract 3D information from two or more registered intensity images, using passive triangulation. This passive triangulation process is useful to improve the initial camera models, passively triangulate primary features in intensity image to guide feature extraction in range images and, finally, to add 3D information into range data.

1.4 Thesis outline

This thesis is divided in three parts:

- Chapters 2 and 3 deal mainly with problems of calibration of the cameras to bring both datasets, intensity and range, into the same reference frame.
- Chapters 4 and 5 demonstrate how registered intensity data can be used to improve the geometry of the models computed from range imagery.
- Chapter 6 concludes the thesis by presenting an example with two different laser scanners. Chapter 7 summarizes the main contributions of the work and presents possible future work.

Chapter 2 presents the methods used to match the data from both sensors. This step is necessary to compute correspondences between range and intensity data. These correspondences will be used afterwards to calibrate the cameras.

Chapter 3 briefly remembers the Tsai camera model [Tsai86] used to calibrate intensity images. It presents also the methods developed to refine and use the camera calibration. At this stage, intensity images are only used as a texture for the range model.

Chapter 4 shows the first experiments in which a single intensity image is used to correct and guide the segmentation of basic features (first and second order curves). This chapter concentrates on 2D edges from the photographs and uses these features to guide the segmentation in triangulated models.

In chapter 5, 3D information coming from intensity is combined with range data. Here, two or more intensity images are used to triangulate 2D features based on the initial

camera calibration, but without any need of range data. Using this extra information allows, first, to increase the quality of the camera calibration, and secondly to add 3D points in the original range cloud. The result is a 3D dataset that merges together 3D information from both sensors.

Chapter 6 presents some experimental results of the technique with two complex indoor and outdoor environments.

Chapter 7 summarises the main contributions of this work and presents some possible future work.

Two Appendices are also included in this thesis. Appendix A describes the experimental set up and gives technical information about the sensors used to acquire the data presented in the thesis. Appendix B shows another possible application of the resizing algorithm presented in Chapter 2, used in this case to register satellite images.

2 Matching Laser Reflectance with Intensity Image

This chapter presents the methods used to match the intensity and range data. This step is necessary to compute correspondences, between the datasets, that will be used afterwards to calibrate the cameras.

2.1 Introduction

Most commercial laser scanners return for each measurement both the distance to points and the reflected light intensity (Figure 2-1). This leads to two different sets of data: 3D models can be computed from the **distance measurements**, whereas an image can be generated from the **reflectance information** (Figure 2-2).

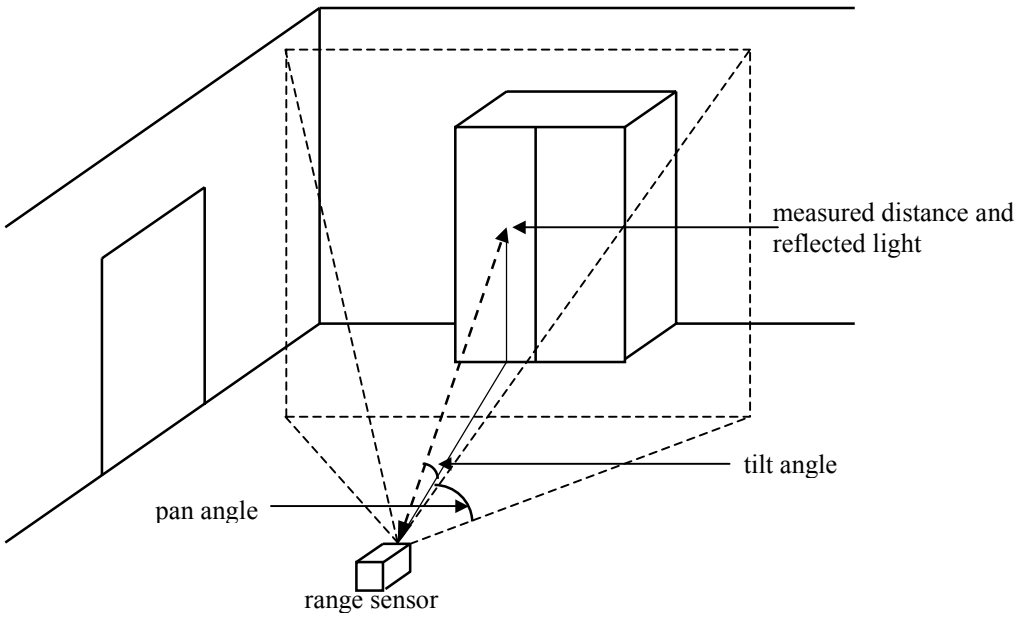


Figure 2-1: Range measurement

(a) (b)

Figure 2-2: Range cloud (a) and reflectance image (b) acquired with a Riegl LMS-Z210 laser scanner (1000x240 pixels)

This chapter presents some techniques developed to perform an automatic matching of points between reflectance (acquired with a laser scanner) and intensity images (acquired with digital cameras). We consider an un-calibrated situation: no previous knowledge is available about the position or internal parameters of the sensors. The choice of using reflectance images rather than three-dimensional information for the matching was motivated by the fact that reflectance images are 2D images such as photographs. The matching can then be performed in 2D using techniques already existing for matching intensity images [Heipke96, Lang95, Brown92].

In our case, an additional difficulty appears compared to traditional matching techniques, since the images to match are not coming from the same sensor. The camera is a high-resolution passive sensor whereas the laser scanner is an active sensor of lower resolution. This leads to images with different field of view, resolution, and properties (illumination and colour for instance).

Some existing works try to register automatically range and intensity data. Elstrom developed an automatic system for registering intensity and range data by computing correspondences between reflectance and intensity images [Elstrom98]. In this case, no automatic resizing of the images is done to ensure that main features overlap in the images. This has to be ensured during the acquisition step or by precise user-selection of a reduced part of the intensity image. Another example of automatic range-intensity registration is presented in [Stamos01a]. In this system, the registration needs the detection of three major vanishing lines in the images to extract pose information. This reduces the field of application of the algorithm to urban scenes with strong geometry contents. Kurazume uses edge detection in reflectance and intensity images to align both datasets by using an iterative process [Kurazume02]. None of these methods considers directly the problem of different resolutions between reflectance and intensity images and thus input data must be

already formatted by the user or carefully acquired. All these methods also consider calibrated cameras with known internal parameters and thus only pose estimation is computed.

Usually, a matching procedure uses either features or area-based techniques. Feature-based techniques use pre-processed features to match the images (edges for example). Area-based techniques work directly on the grey values of the images to extract matching information. Our system uses a mixed feature and area based matching technique to get corresponding points between the images to be registered. The technique can be divided in two steps:

- Resizing of the reflectance image based on the intensity image. This step uses edges to solve the problem of the difference of resolution and field of views between the images to match. The output is a resized reflectance image with the same size and with the main features located in the same area than the intensity image.
- Matching of the resulting images to find some corresponding points using area based techniques.

2.2 Resizing algorithm

The goal of the resizing algorithm is to apply a planar affine transformation to the reflectance images to make them "fit" as well as possible to the intensity images. The resulting images will be used in the matching process. The whole algorithm uses the red channel of the intensity image because of its higher similarity with the infrared reflectance image.

The transformation between the images is a 3D transformation characterised by a rotation, a translation and a distortion due to the sensor used (camera and scanning device parameters). Depending on the devices used, this transformation can be characterised by up to 15 parameters [Luong94]. In this phase, a first approximation is going to be performed using only five parameters: the planar translation co-ordinates in both axis, the X and Y scaling factors and the planar rotation angle between both images. These parameters are modelled with a 3x2 affine matrix.

Obviously, the result will not be exact but an approximation of the transformation between the two images. Still, the main features of both images will be overlapping enough to allow the use of normal area based matching algorithms to extract matching points.

The whole affine resizing process is illustrated in Figure 2-3.

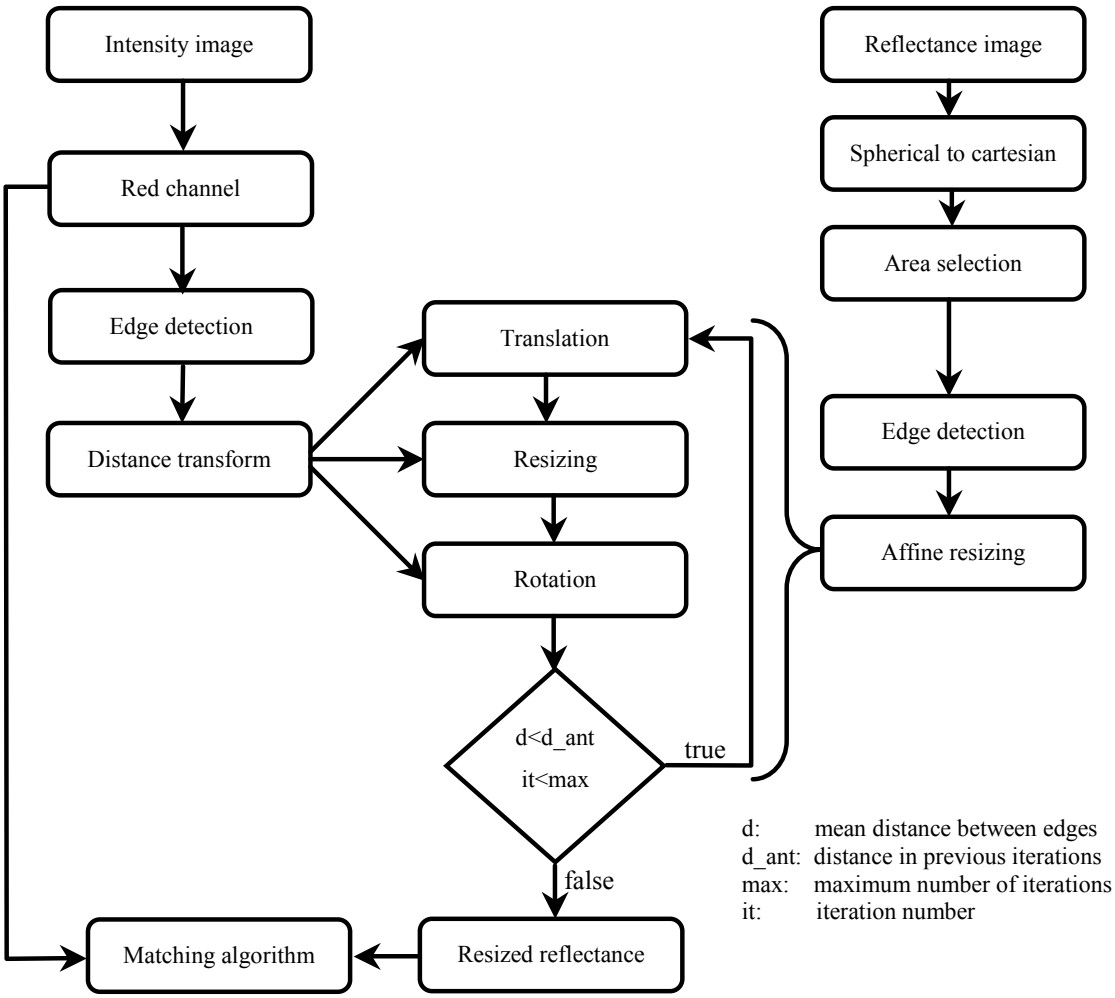


Figure 2-3: Algorithm used to perform the affine planar transformation

2.2.1 Conversion from spherical to planar co-ordinates

Since the images to register come from different acquisition systems, they also are expressed in different co-ordinates. Intensity images come from a projection using a pinhole camera and thus are in Cartesian co-ordinates. On the other hand, when laser scanners acquire range images, each pixel position corresponds to a Pan and Tilt angle. In the resulting image, the points are expressed in spherical co-ordinates.

In our matching process, the first transformation applied to the reflectance image is a change from spherical to Cartesian co-ordinates.

Spherical co-ordinates are different from Cartesian co-ordinates: the Euclidean distance between adjacent pixels depends on the angular distance and not on the real Cartesian distance between points. Due to this particularity, the image in spherical co-ordinates presents distortions mainly in areas far from its centre.

Figure 2-4 illustrates the geometry associated to the spherical co-ordinates in reflectance images.

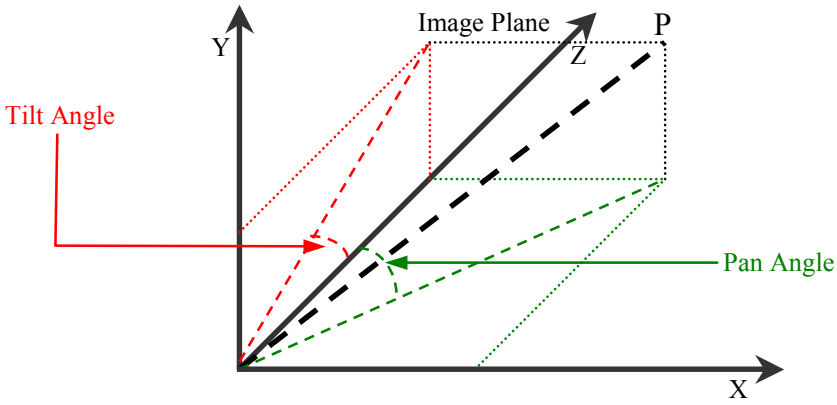


Figure 2-4: The spherical co-ordinate system of reflectance images

The conversion starts by computing the following value:

$$\text{base} = \min\left(\left\lceil\frac{\text{width}}{2\times\tan(\text{pwidth})}\right\rceil;\left\lceil\frac{\text{height}}{2\times\tan(\text{twidth})}\right\rceil\right) \qquad (2.1)$$

where (height ,width) are the intensity image dimensions and (pwidth, twidth) are the maximum angles of the laser scanner in pan and tilt directions.

The base value is used for normalization purposes. Its use guarantees that the size of the image in Cartesian co-ordinates will not be larger than the original reflectance image. The

equations that relate Cartesian co-ordinates (r,c – for row and column) and spherical co-ordinates (PanAngle and TiltAngle) are:

$$\text{PanAngle} = \text{atan}\left(\frac{\text{c} - \frac{\text{width}}{2}}{\text{base}}\right) \tag{2.2}$$

$$\text{Tilt Angle} = \text{atan}\left(\frac{\text{r} - \frac{\text{height}}{2}}{\text{d}}\right) \tag{2.3}$$

$$\text{where d} = \frac{\text{base}}{\cos(\text{PanAngle})}$$

Using these formulas, it is possible to associate to each (r,c) pair in the Cartesian image its corresponding pan and tilt angles in the spherical representation. Since each reflectance point is localised by its Pan and Tilt angle, it is possible to associate directly the grey value of the reflectance image to the point (r,c) in Cartesian co-ordinates.

A bilinear interpolation (see Figure 2-5) between the four closest neighbours is used when the computed angle does not correspond to an integer value. The grey value is computed as a combination of the values of the four neighbouring pixels:

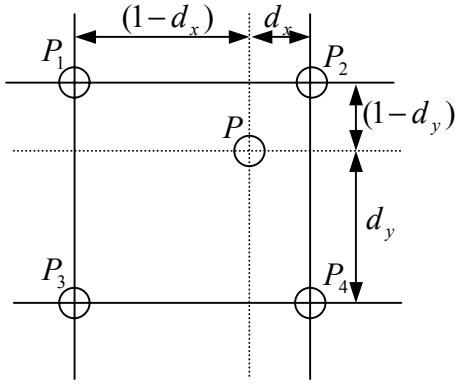


Figure 2-5: Bilinear interpolation

$$P = d_x d_y P_1 + (1-d_x) d_y P_2 + d_x (1-d_y) P_3 + (1-d_x)(1-d_y) P_4 \tag{2.4}$$

Figure 2-6 present the results of the conversion from spherical (a) to Cartesian (b) co-ordinates for a reflectance image of a church in Barza, Italy.

(a) (b)

Figure 2-6: Conversion from spherical to Cartesian co-ordinates of a reflectance image of the Barza church

The transformation is reversible: once the correspondences are computed, it is easy to associate the Cartesian co-ordinates of the points and their co-ordinates in the original reflectance image.

2.2.2 Edge detection

The evaluation of the affine transformation is based on edge detection. It uses the distance transform performed on a binary image obtained from edge detection. The evaluation of the quality of the estimated transform will be based on the median distance between edges in both images.

Some of the most used edge detectors are presented in [Heath97] and [Brejl98]. In this work, edges are detected using a Canny edge detector and nonmaxima suppression is applied to ensure that only strong edges are detected. The Canny method has been selected for its simplicity and capability to provide continuous edges, although it requires two thresholds for the edge detection operation. The Canny edge detector applies low and high thresholds so that edge hysteresis can be considered and thus, it optimises edge continuity [Canny86]. Low and high thresholds are computed automatically for intensity and reflectance images based on statistical characteristics of the gradient over the images. Since the resolution of intensity images is normally higher, the formulas to compute the thresholds are different for intensity and reflectance images:

Intensity Image:
$$\text{LowI} = \text{MeanI} \tag{2.5}$$
$$\text{HighI} = \text{MeanI} + \frac{\text{StdI}}{2} \tag{2.6}$$

where, LowI and HighI are the Canny thresholds for edge detection in the intensity image MeanI and StdI are mean and standard deviation of gradient in the intensity image.

Reflectance images:
$$\text{LowR} = \text{MeanR} - \frac{\text{StdR}}{2} \tag{2.7}$$
$$\text{HighR} = \text{MeanR} \tag{2.8}$$

where, LowR and HighR are the Canny thresholds for edge detection in the reflectance image.
MeanR and StdR are mean and standard deviation of the gradient in the reflectance image.

2.2.3 Distance Transform

The resizing algorithm needs to compute the distance transform of the intensity image. This image will be the "fixed" image, whereas the reflectance image will be the "resized" image, transformed according to the estimated planar transformation.

We chose to keep the intensity image as the fixed image because photographs have normally more edges than reflectance images and thus, computing the affine transformation of the intensity edges would be more time consuming. In addition, the distance transform applied on the intensity images will be more precise since more edges are normally detected.

In most of the cases, the reflectance image is larger than the digital photograph (since laser scanners have a larger field of view). This makes necessary the selection of a reduced part of the reflectance image corresponding to the area covered by the intensity image (see Figure 2-9). To optimise the speed of the algorithm, both images are scaled to the resolution of the smaller one to make the computation faster.

The distance transform is an image in which each pixel value corresponds to the distance to the closest edge. Different techniques exist to perform this transformation and usually, there is a compromise between speed and accuracy. The Euclidean distance transform gives the exact distance to the closest edge in each pixel but is time consuming.

Techniques that try to approximate global distance using local distance propagation (using local masks in the image) are much faster but less accurate [Borgefors84, Borgefors86b, Qin-Zhong84].

The selected implementation is the one proposed by [Borgefors86a]. It consists in the application of a forward and a backward mask (Figure 2-7) on the image. It has the advantage to be computationally efficient and gives a good approximation of the Euclidean distance. On the laboratory images presented in Figure 2-8, the implementation from Borgefors performed in 16ms. In the same test images, our implementation of the Euclidean transform (that computes for each pixel the exact distance to the closest edge) needed 28.8s.

The algorithm begins with a zero/infinity image (zero if a point is an edge, infinity if not). The two masks are then passed over the image once each. The forward mask is applied from top to bottom and left to right and the backward mask from bottom to top and from right to left. The new value of the pixel under the 0 in the mask is the minimum of the five sums of the image pixel's values with the local distance in the mask. For all pixels (*k,l*) in mask *C*, we have:

$$P(i,j) = \min\left[I(i+k, j+l) + C(k,l)\right] \tag{2.9}$$

with, $P(i,j)$ the value of pixel (i,j) in distance transform, $I(i,j)$ the value of pixel (i,j) in image and $C(k,l)$ the value of mask C at position (k,l).

The coefficient values proposed by [Borgefors86a] to minimise the error to the Euclidean distance are presented in Figure 2-7. Figure 2-8 presents the results of the distance transform in a laboratory at the Joint Research Centre (JRC) in Italy. Figure 2-8(a) is the original digital photograph from the scene, Figure 2-8(b) is the resulting distance transform on this image and Figure 2-8(c) presents in the same image the detected edges and the distance transform.

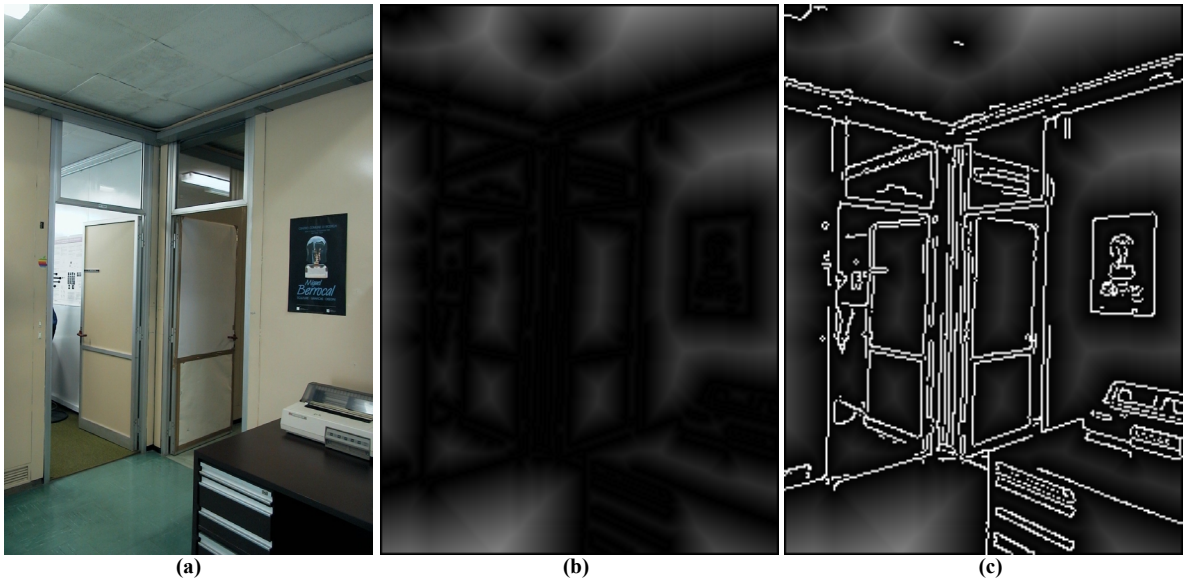
4	3	4
3	0	

(a)

(b)

	0	3
4	3	4

Figure 2-7: The forward (a) and backward (b) masks used for the distance transformation



(a) (b) (c)

Figure 2-8: The distance transform in an image of the laboratory 030 at JRC

2.2.4 Affine transform evaluation

In this section we are seeking for the best set of values to characterise the planar transformation between both images. The five values evaluated within the affine transformation are the scaling factors in X and Y directions (S_x and S_y), the translation values (Δ_x and Δ_y) and the rotation angle (θ) [Halcon02]. Given (x,y,1) the homogeneous co-ordinates of a pixel in the original image, and (x',y',1) the co-ordinates of the pixel after the affine transform, we have:

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} = T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.10}$$

where T, is a 3x3 matrix composed by the following three transformations:

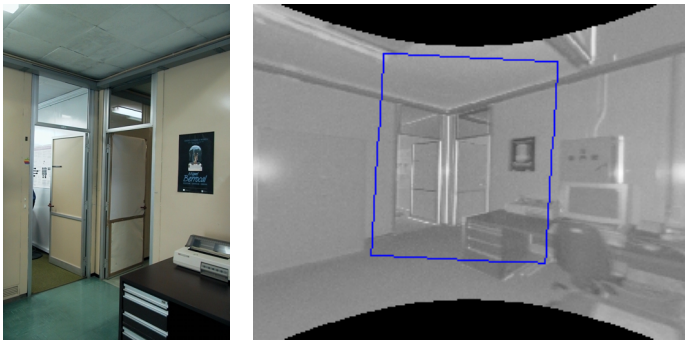
T = Rotation x Scaling x Translation,

with Rotation:
$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.11}$$

Scaling:
$$\begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.12}$$

Translation:
$$\begin{bmatrix} 1 & 0 & \Delta_x \\ 0 & 1 & \Delta_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.13}$$

The algorithm starts by transforming the reflectance image into Cartesian co-ordinates (see section 2.2.1). Once the transformation is computed, the user must select a part of the Cartesian reflectance image that corresponds approximately to the same area of the intensity image. The position, size and orientation of this window will be used to compute the first affine transformation (see Figure 2-9).



(a) (b)

Figure 2-9: Original photograph (1024x1536) and Cartesian reflectance image with user defined window

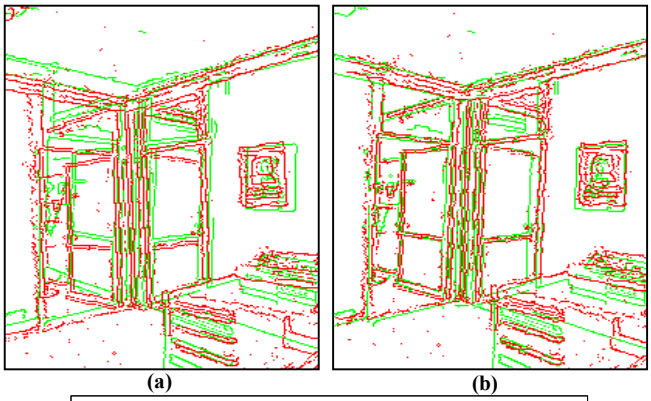
An iterative algorithm will improve the initial approximation. In each step, several affine transforms are computed within an interval around the actual estimates. The transform that minimizes the average distance between edges is selected as the new transformation for the next iteration. The average distance between edges is computed by applying the affine transformed edges as a mask over the distance transform image. The distance to the closest edge is determined for each edge point (as the pixel value in the distance transform image) and used to compute the average distance between edges in the two images.

The algorithm converges when the average distance does not change between two iterations. In addition a maximum number of ten iterations is set to avoid situations where the algorithm can oscillate between two values.

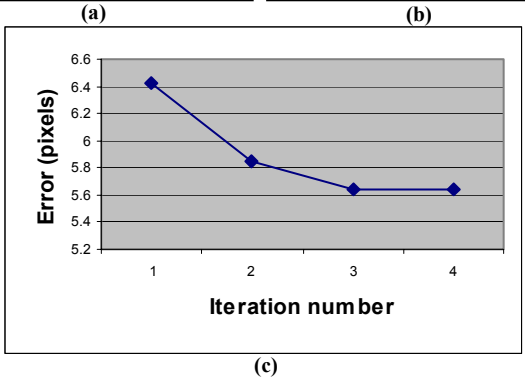
Figure 2-10 presents some results of the resizing process with several images. For each image, the figure presents the superposition of edges (in green, from the fixed intensity images and in red from the resized reflectance image) at the beginning (Figure 2-10a) and at the end (Figure 2-10b) of the resizing algorithm, as well as the evolution of the average distance between the edges in the images (Figure 2-10c).

The first example is an indoor environment: a laboratory at the Joint Research Centre in Italy. The second example is an outdoor environment: a church in Barza, Italy. An example

acquired with a different scanner and camera is presented in the Bornholm church (DK) example. The final example is from a Farmhouse in Laveno, Italy.

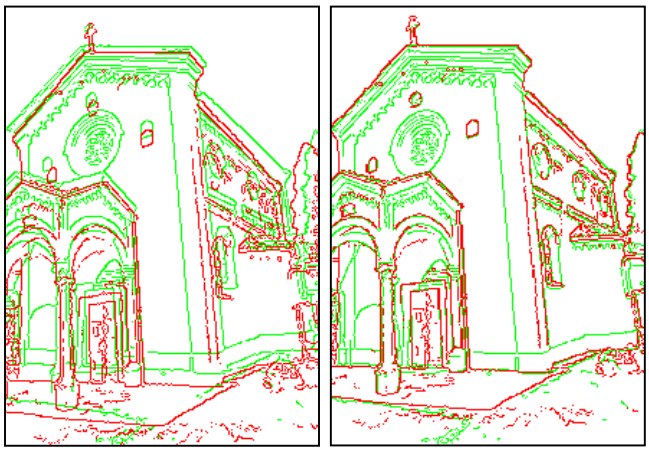


(a) (b)

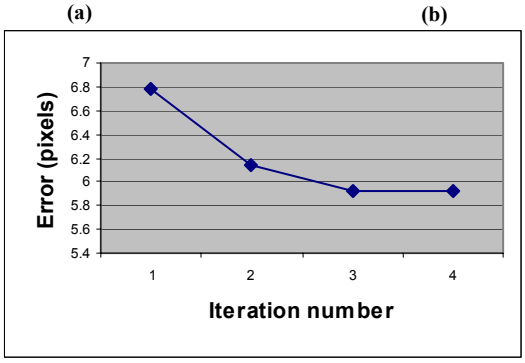


(c)

Laboratory 030 at JRC

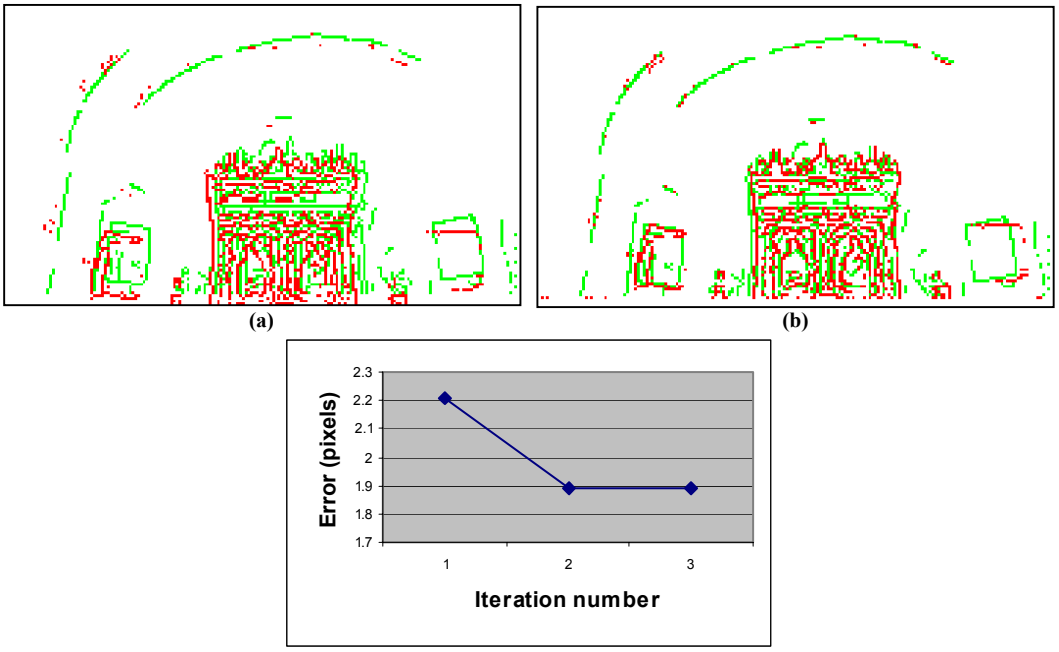


(a) (b)

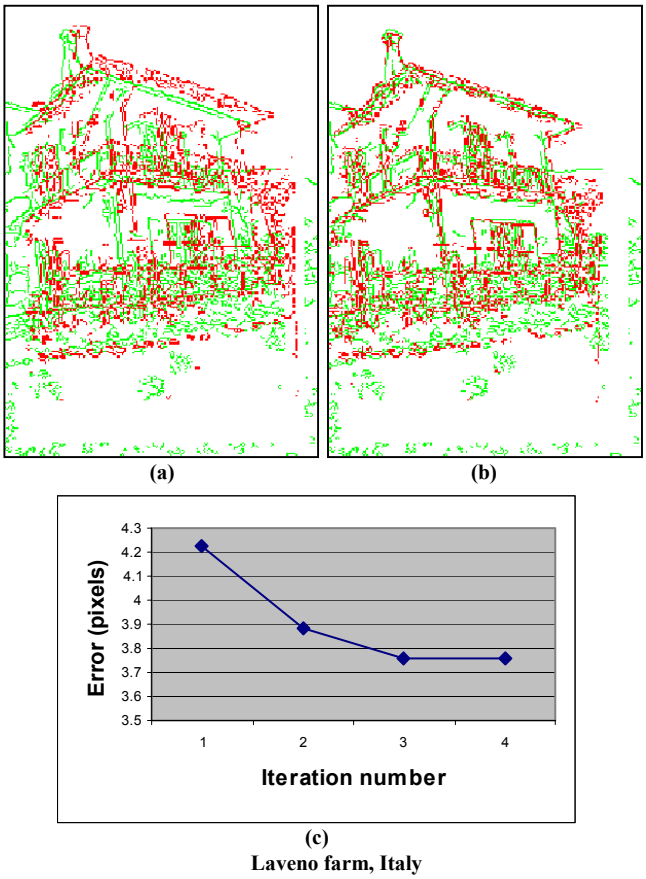


(c)

Church in Barza, Italy



Round Church in Bornholm , Danemark



Laveno farm, Italy

Figure 2-10: Results of resizing algorithm with four different examples

(a- original edges, b- edges after resizing, c- average distance between edges along the process)

In most of the cases, the algorithm does not need to be very precise because a global approximation is usually enough for the matching algorithm. In the laboratory image, despite the final error of 5.63 pixels between the edges, the result of the resizing is sufficient for the subsequent matching step. In this example 4 iterations have been performed and the algorithm took approximately 1.1 seconds in a Pentium IV at 2.0 GigaHertz. Similar results have been obtained with the Barza (final error of 5.92 pixels in 4 iterations and 1.2 s) and the Bornholm examples (final error of 1.89 pixels in 3 iterations and 0.3 s). Results with the Laveno Farmhouse examples give a final error of 3.76 pixels in 4 iterations and 0.83 seconds.

Once the algorithm has converged, the reflectance image is resized using the computed affine transformation. In this resizing step, no interpolation is used even though the visual results are worse (the resized image is not as smooth as when using interpolation). This choice has been made because smoothing the image would decrease the performance of the corner detector used in the matching algorithm. The algorithm is also faster if no interpolation is done. Finally, visual quality is not a very important issue in this part of the process as we are only seeking for correspondences between both images and the output will be pixel co-ordinates in both images and not the images themselves.

A limitation of the algorithm is the importance of the first approximation depending on the area of interest defined by the user. A large error in this approximation can influence the algorithm and make it perform poorly. Still, this user selection is necessary for the system to work with un-calibrated images taken from unknown viewpoints. If some restrictions are imposed in the acquisition system (for example a fixed camera on the top of the laser scanner), additional information about the relative position of the sensors can be used to make this selection automatic.

2.3 Additional Tools

The resizing procedure of section 2.2 is based on a planar transformation with five degrees of freedom. The real transformation between the reflectance and the intensity images is a much more complex three-dimensional transform. Thus, the quality of the approximation will depend on the characteristics of the images and the conditions of acquisition. For example, these approximations will fail when images are taken from very different viewpoints.

To enlarge the field of application of our algorithm, some additional tools have been developed and are presented in the following sections.

2.3.1 Rotation Evaluation based on vanishing points

During the resizing step, it is important to ensure that intensity images and reflectance images are as similar as possible. Some additional transformations can be introduced in the reflectance image to make it more similar to the intensity image, and increase the performance of the resizing algorithm. The result will be a more accurate affine transformation.

After the spherical to Cartesian transform (see section 2.2.1), another transformation that can be evaluated is the 3D rotation between the datasets. This one is based on the use of vanishing points and, thus, its use is reduced to images with strong geometrical contents, where the three principal vanishing directions can be easily extracted.

The automatic computation of the rotation between two different images is not an easy task, but it can be quite simplified with the imposition of some constraints. A well-known technique to compute the camera 3D rotation between two photographs is based on vanishing points. In this case, it is possible to extrapolate the 3D main directions from the vanishing points. The rotation is then computed by matching two main 3D directions in the range image with two 3D direction extracted from the vanishing point analysis [Stamos00].

2.3.1.1 Vanishing point detection

Vanishing lines are commonly used to detect main directions in images and extract 3D direction information. This information can be helpful to register several images by evaluating the rotation existing between the different data acquired. Works using vanishing points to extract camera pose information can be found in [Caprile90, Tuytelaars98, Heuvel98, Rother00, Partington01].

2.3.1.1.1 Line Detection

The line detection process is used to extract linear features from 2D images. It analyses each 2D connected edge (detected with a Canny edge detector [Canny86]). For each one, a RANSAC [Fischler81] technique selects randomly and iteratively two points in the edge and computes the 2D line with more support (grey point in Figure 2-11).

The algorithm determines the intersection between this line (grey points in Figure 2-11) and the edge (in thin dash). Finally, the largest connected region of the intersection is computed and considered as a line if longer than a threshold (red line in Figure 2-11).

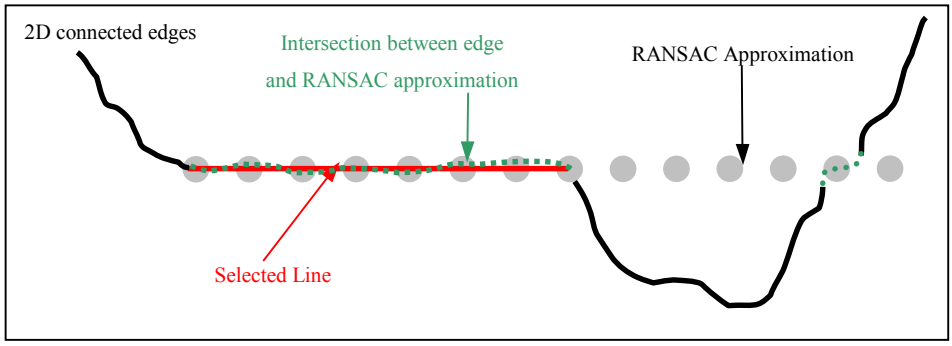


Figure 2-11: The line extraction process

2.3.1.1.2 Gaussian representation

Several methods exist for detection of vanishing points [Shufelt99]. Most of them consist in a mapping from the image plane to another space, which can deal with infinite points. Infinite points are particularly important since vanishing lines are very often parallel and lead to vanishing points located at infinity.

In our case, the Gaussian representation was chosen [Barnard83]. In this representation, a point in the plane is projected into a sphere: a pixel in the image plane corresponds to a

point in the Gaussian sphere and is referenced by two angles (θ, ϕ). The possibility to treat points at infinity as any other points constitutes the main advantage of this representation. Lines in the image plane correspond to great circles in the Gaussian sphere as illustrated in Figure 2-12.

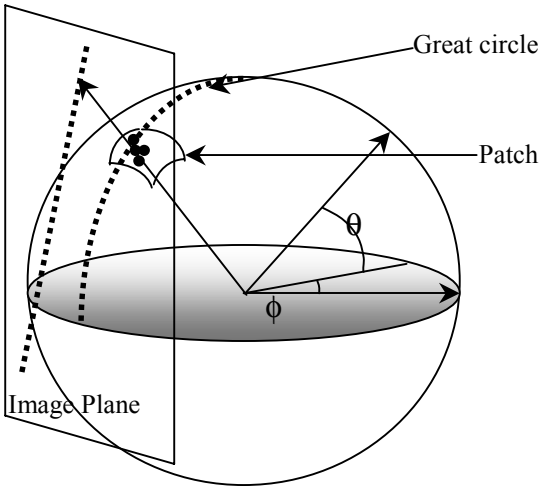
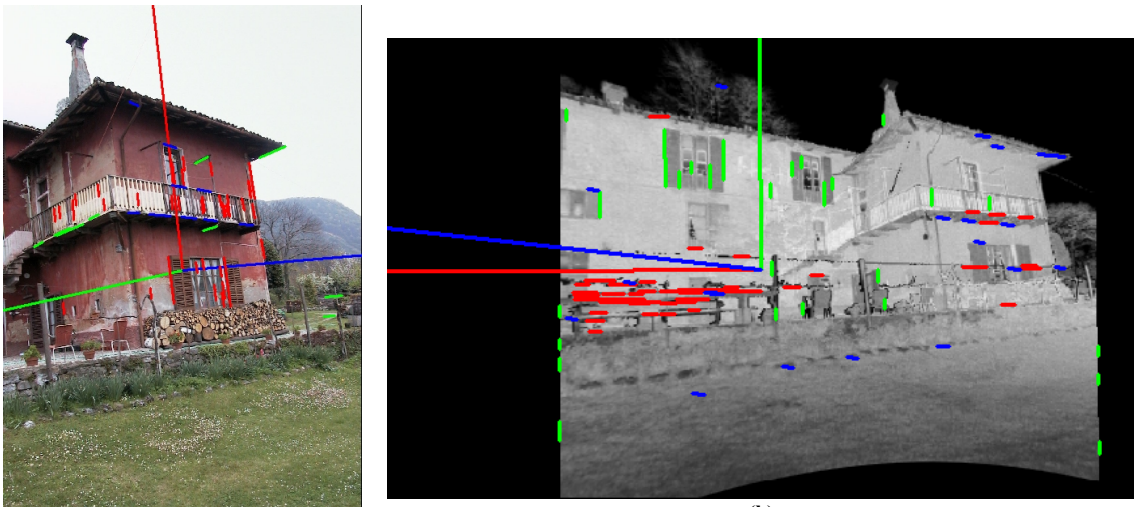


Figure 2-12: The Gaussian sphere representation used for vanishing point detection

The vanishing line detection algorithm starts by dividing the Gaussian sphere into several patches corresponding to surfaces of the sphere with the same area. For each patch, a buffer is updated according to the number of lines whose great circle intercepts the patch (see Figure 2-12).

For rotation evaluation we are interested in detecting only three orthogonal directions in the image. Thus, to avoid detection of vanishing points in the same area of the Gaussian circle, a filtering is introduced. When a vanishing point is selected, the filtering consists in checking the percentage of lines in each patch that converge to an already selected vanishing point. If this percentage is higher than a threshold, the patch is rejected since its probability to be closer to another already detected vanishing point is high.

Figure 2-13 shows the results of vanishing point detection for a digital photograph and the Cartesian reflectance image of a farmhouse in Laveno/Italy.



(a) (b)

Figure 2-13: Vanishing point detection results

2.3.1.2 Estimation of the major 3D directions from Intensity images

When a parallelipipedic solid object is represented under 3D perspective projection, it produces three vanishing points corresponding to three orthonormal directions. The three vanishing points and the centre of projection form a right tetrahedron [Becker95]. This property can be used to compute the centre of projection and thus extract the 3D direction corresponding to the three vanishing points. Figure 2-14 represents the right tetrahedron where *A*,*B* and *C* are the detected vanishing points in the image, *P* the orthocentre of the triangle *ABC* and *Q* is the centre of projection of the image.

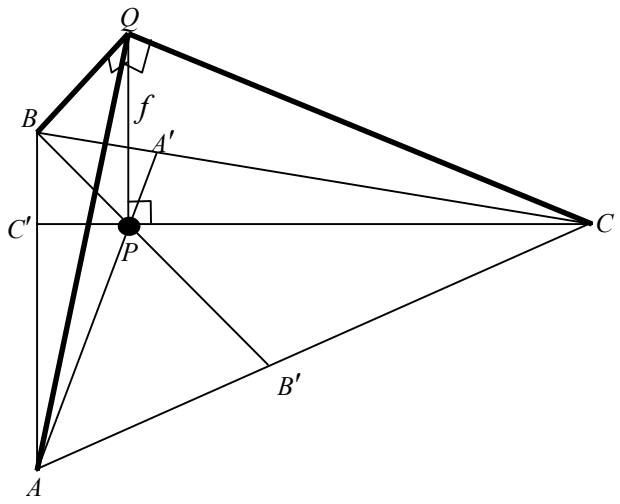


Figure 2-14: The right tetrahedron representation

The 2D co-ordinates of the principal point P in the plane ABC, can be computed as the intersection of lines BB' and CC'. The projection C' of C into AB is given by:

$$C' = \frac{(B-A)^T (C-A)(B-A)}{(B-A)^T (B-A)} + A \tag{2.14}$$

and the projection B' of B onto AC is:

$$B' = \frac{(C-A)^T (B-A)(C-A)}{(C-A)^T (C-A)} + A \tag{2.15}$$

thus, the 2D co-ordinates of P (s,t) are the solution of the linear system:

$$B + s(B'-B) = C + t(C'-C) \tag{2.16}$$

and, the principal distance f, from the principal point P to the centre of projection Q is:

$$f = \left[(P-A)^T (B-P)\right]^{1/2} \tag{2.17}$$

Using this information, it is possible to find the 3D directions corresponding to the vanishing points.

2.3.1.3 Estimation of the major 3D directions from range images

The extraction of the main 3D directions in range images does not need the tetrahedron representation, since range images already contain three-dimensional information about the scene. To find out the 3D directions, the direction of each line supporting a given vanishing point is computed (from the segment co-ordinates in 3D). The angle between pairs of lines is computed and lines with an angle smaller than a threshold are grouped into the same cluster. At the end, the cluster regrouping more lines is selected as the main direction for the given vanishing point.

2.3.1.4 Rotation evaluation

The evaluation of a rotation in 3D needs only the matching of two 3D directions. Since sections 2.3.1.2 and 2.3.1.3 provide three directions, it is possible to select only two of them for the rotation estimation. The selected two pairs of directions are the ones that initially have the smallest angle between them. This choice is justified by the fact that range and intensity images are usually acquired from close viewpoints and, thus, the 3D rotation angle should not be too large.

Once two matching directions are selected, the algorithm proposed by [Faugeras93] is used to evaluate the 3D rotation between these two directions. Given $n_i(x_i, y_i, z_i)$ and $n_i'(x_i', y_i', z_i')$ the two pairs of corresponding 3D directions, the rotation matrix R between them is the one that minimizes the expression:

$$C = \sum_{i=1}^{2} \left\| n_i' - R n_i \right\|^2 \tag{2.18}$$

Using quaternions [Walker91] to represent the rotation R, the expression can be written under the following form, where q is the quaternion that represents the rotation.

$$C = \sum_{i=1}^{2} \left\| n_i' \times q - q \times n_i \right\|^2 = \sum_{i=1}^{2} q^T A_i^T A_i q \tag{2.19}$$

The desired rotation axis is the eigenvector corresponding to the smallest eigenvalue of matrix A. This eigenvalue is the angle of the rotation R and the matrix A is defined as follows:

$$A = \sum_{i=1}^{2} A_i^T A_i \quad \text{with} \quad A_i = \begin{bmatrix} 0 & x_i - x_i' & y_i - y_i' & z_i - z_i' \\ x_i' - x_i & 0 & -z_i - z_i' & y_i + y_i' \\ y_i' - y_i & z_i + z_i' & 0 & -x_i - x_i' \\ z_i' - z_i & -y_i - y_i' & x_i + x_i' & 0 \end{bmatrix} \tag{2.20}$$

2.3.1.5 Re-projection of rotated range

The rotation R is used to create a new reflectance image that considers the rotation existing between the two sensors during the acquisition. This is done by rotating the range cloud of points using rotation R, as illustrated in Figure 2-15(a).

The second step consists in creating a new reflectance image by projecting the 3D point cloud according to a Cartesian projection (see section 2.2.1). The result is a new 2D reflectance image, which seems to be taken with the same rotation as the photograph, see Figure 2-15(b). Bilinear interpolation is used to interpolate the 3D co-ordinates and get a final reflectance image without holes.

In this example, the 3D co-ordinates of the extracted matching directions are:

Intensity directions: (0.16,0.97,0.18) and (-0.57,-0.05,0.82)

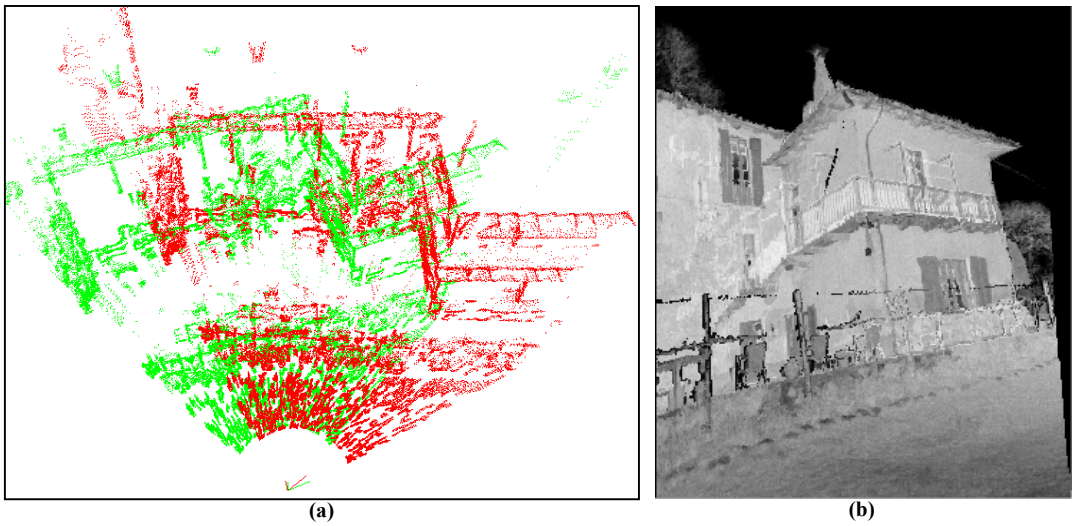
Range directions: (0.12,0.99,0.08) and (-0.79,-0.18,0.59)

The final Euler angles of the estimated rotation between the images are:

Angle in Z direction: -3.94º

Angle in Y direction: 17.34º

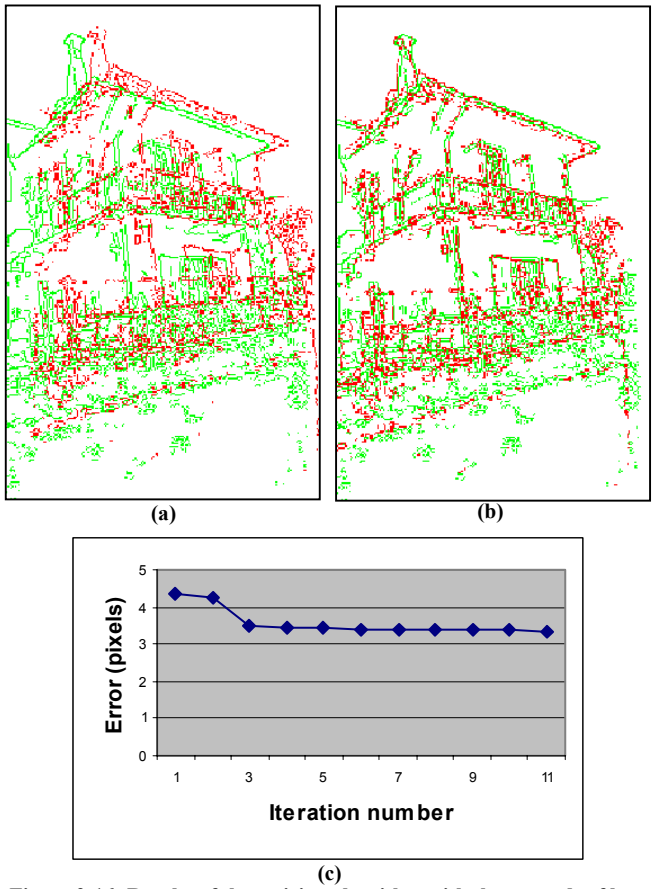
Angle in X direction: 2.72º



(a) (b)

Figure 2-15: (a) the range cloud of points in its original position (red) and after the rotation (green) (b) The rotated reflectance image

The rotated image obtained with this process is used as input in the resizing algorithm presented in section 2.2. Figure 2-16 presents the results of the resizing algorithm with the rotated reflectance. In this case, the final error of the resizing is of 3.33 pixels, against 3.76 when using the original reflectance with no transformation (see Figure 2-10d)



(a) (b)

(c)

Figure 2-16: Results of the resizing algorithm with the rotated ref lectance

(a- original edges, b- edges after resizing, c- evolution of average distance between edges during the process)

2.3.2 Fast Texture mapping using panorama images

The basic idea of the process is to use commercial software (e.g. Photovista [Ring97]) to combine several photographs in a panoramic view. This panoramic view is similar to a reflectance image because it is in spherical co-ordinates and lens distortion is also considered. Figure 2-17 presents the result of the panorama fusion of three images into a panoramic view with Photovista software.

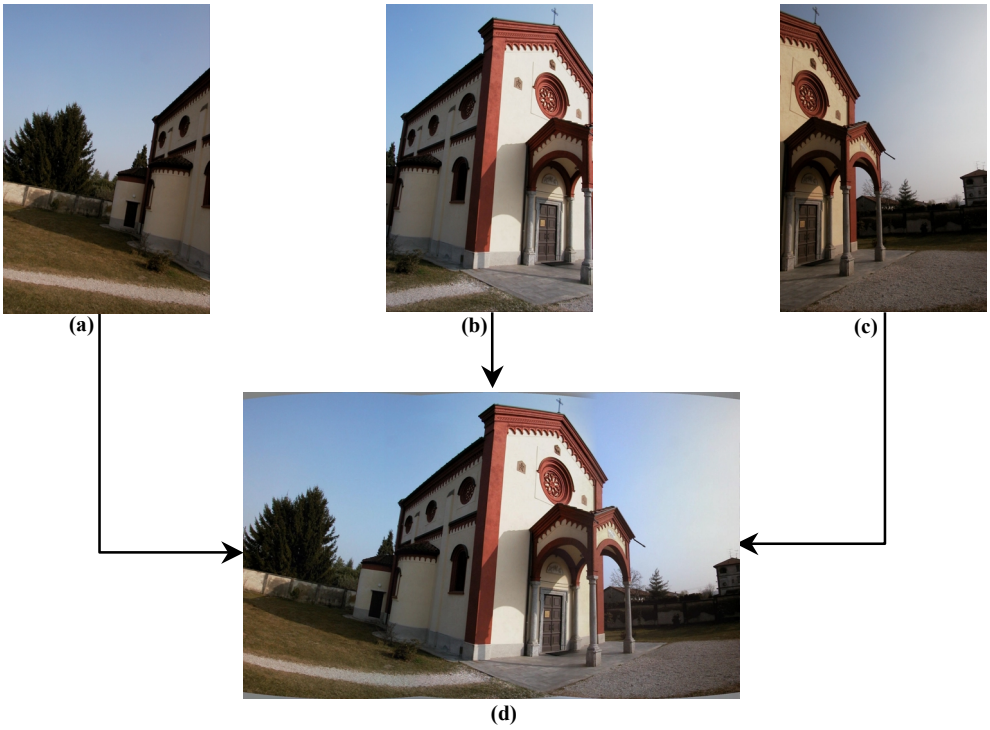
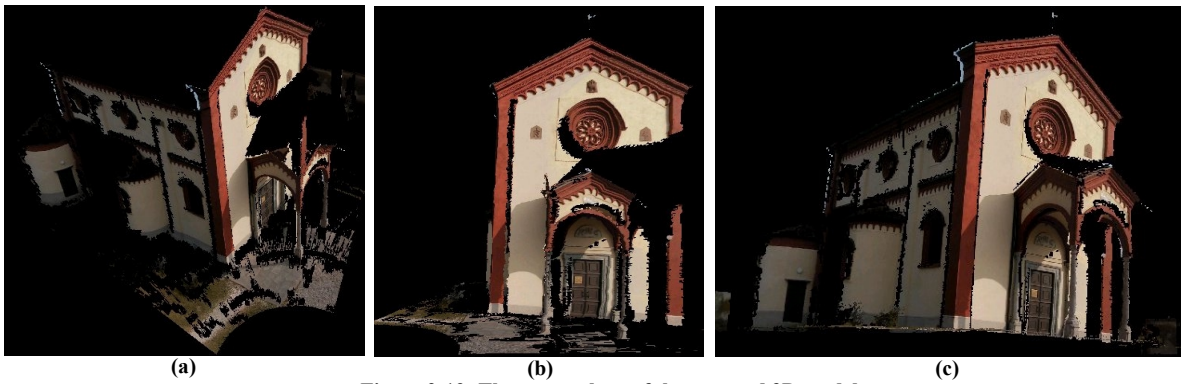


Figure 2-17: Panorama computation from three images with Photovista commercial software

Our resizing algorithm is then used to match the edges of the reflectance and the panorama image. The output of the algorithm is a cropped panoramic image corresponding to the best fitting between the edges (see Figure 2-18). This cropped image can be directly used as a texture map and, in most of cases, already gives a very good texture for the final 3D model (see Figure 2-19).

(a) (b) (c)

Figure 2-18: Reflectance image, resizing results and texture map (sub-part of a panorama of 3 photographs)



(a) (b) (c)

Figure 2-19: Three snapshots of the textured 3D model

2.4 Matching

The resizing algorithm presented in the previous section provides two images with the same resolution and with overlapping global features (edges). The following step of our matching procedure uses classical area based techniques to extract matching points from these images. Matching between reflectance and intensity images is a difficult task because of the different origins of the data. First of all, the light in reflectance images comes only from the laser instead of the natural light for photographs. Secondly, the wavelength of the laser will interfere with the colours on the scene and will affect the final colours of the object in the reflectance image. Finally, orientation and reflecting properties of the scene can prevent the laser beam from returning to the laser scanner and thus holes can appear in the reflectance image.

The implemented matching algorithm starts by selecting points of interest in both reflectance and intensity images. We decide to use corners since they correspond to areas of the images with a large gradient variation and thus, easier to identify using area based techniques. They also are more robust to errors coming from the resizing algorithm. Once corners are identified in the images, the displaced frame difference is computed between these points to select corresponding corners.

2.4.1 Corner detection

Since the application of the Moravec interest operator in 1979 [Moravec79], a lot of research has been done in corner and interest point detection [Kitchen82, Harris88, Deriche90, Mehrotra90, Schmid98, Smith98]. The corner detector implemented is the one presented in [Haralick93]. This corner detector first selects windows of interest and afterwards computes more precisely the position of the points inside the selected windows of interest.

The different steps leading to the windows of interest detection are:

- Calculation of the derivative images using the gradient operator for each row and column (*r,c*):

$$f_r(r,c): \text{horizontal gradient}$$

$$f_c(r,c): \text{vertical gradient}$$

- Determination of the normal matrix for each window of the size chosen for the operator:

$$N = \begin{bmatrix} \sum_{window} f_r^2(r,c) & \sum_{window} f_r(r,c)\cdot f_c(r,c) \\ \sum_{window} f_r(r,c)\cdot f_c(r,c) & \sum_{window} f_c^2(r,c) \end{bmatrix} \tag{2.21}$$

- Calculation of the weight (*w*) and circularity (*q*) of each window

$$w(r,c) = \det(N) \tag{2.22}$$

The weight *w* is also known as the "Beaudet measure for cornerness" and is proportional to the variation of gradient inside the window. It is high in windows where the gradient variation is important (potential points of interest).

$$q(r,c) = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 \tag{2.23}$$

where, λ_1 λ_2 are the eigenvalues of the normal matrix *N*,

q represents the circularity of the ellipse associated to the normal matrix *N*. This value is proportional to the ratio minor axis/major axis of the ellipse; it is high if the two axis have the same size (1 if there are equal) and less in other cases.

The circularity is used to reject points located on edges. In this situation, the ellipse has an important major axis in the direction of the edge and q is less than one.

- Thresholding

  The thresholding is done using the following rule:

$$w^*(r,c) = \begin{cases} w(r,c) & \text{if } w(r,c) > w_{\min} \text{ and } q(r,c) > q_{\min} \\ 0 & \text{else} \end{cases} \tag{2.24}$$

- Suppression of nonmaxima of weight (w^*) within a specified window defined by the user.

To select the window of interest, the algorithm needs the following parameters:

- A gradient operator, here the Sobel gradient operator is used:

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

- The window size for the operator, we use a simple 2 by 2 window.
- The threshold q_{min} for circularity (we used typically 0.9).
- The maximum number of corners to detect. This value is used to compute the neighbourhood for finding local extremas. The minimum weight w_{min} is computed dynamically, using the standard deviation of all the weights. The corners are ordered using their weight and a maximum number of corners is returned according to this ordering.

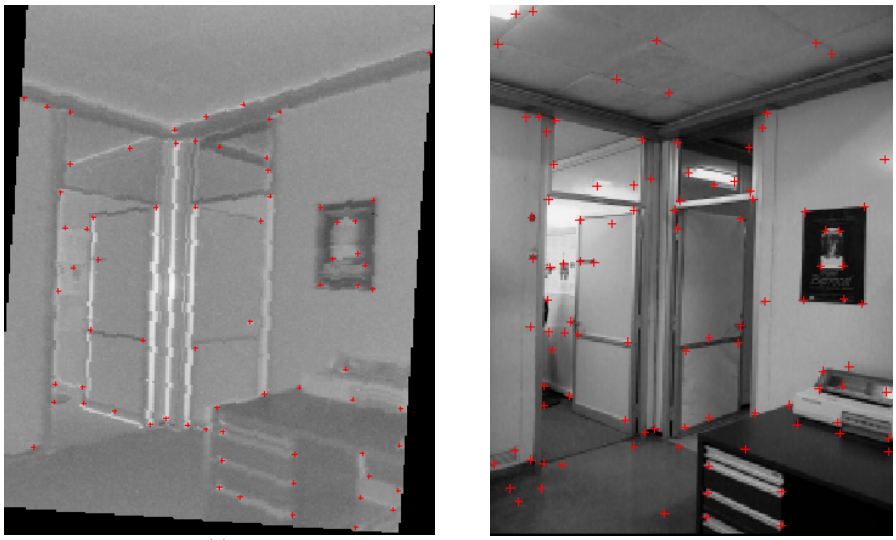
In a second step, the algorithm finds the most interesting point inside the window of interest. This is relatively simple: the point is determined as the weighted centre of gravity of all points inside the window with the product of row and column gradient as weight.

The co-ordinates (x,y) of the corner inside the window of interest are then given by:

$$x = \frac{\sum_{window} (f_r(r,c) \times f_c(r,c) \times c)}{\sum_{Window} f_r(r,c) \times f_c(r,c)} \tag{2.25}$$

$$y = \frac{\sum_{window} (f_r(r,c) \times f_c(r,c) \times r)}{\sum_{Window} f_r(r,c) \times f_c(r,c)} \tag{2.26}$$

Results of corner detection for the laboratory images are presented in Figure 2-20.



(a) (b)

Figure 2-20: Corner detection in the laboratory images, (a) resized reflectance, (b) intensity image

2.4.2 Displaced frame difference

The similarity between intensity images can be evaluated using different measures. The most popular one is the cross correlation but other measures have been developed. In our case, we use the displaced frame difference. This measure gives approximately the same results as cross correlation but is computationally more efficient, making the matching process faster. The displaced frame difference (often abbreviated as *dfd*) is defined as follows:

$$dfd(i,j) = \sum_{u,v} \frac{abs\left|\text{Image}(i-u, j-v) - \text{Template}(l-u, c-v)\right|}{area(\text{Template})} \qquad (2.27)$$

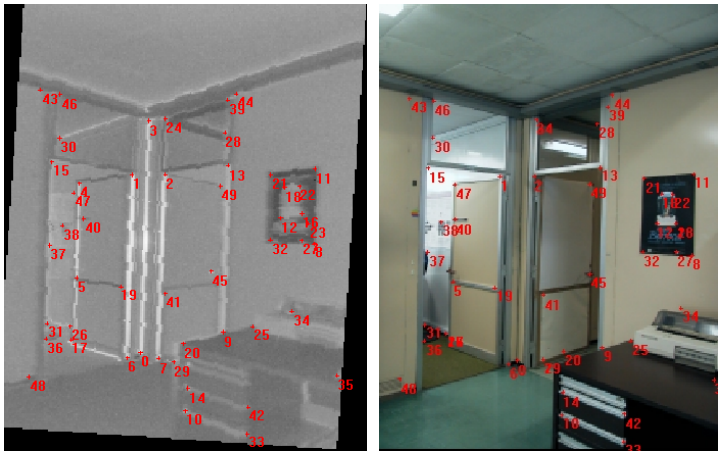
where (i,j) are the co-ordinates of the point of interest in the intensity image (the search image); (l,c) are the co-ordinates in the reflectance image (used as Template) and (u,v) are the co-ordinates within the interval [-TemplateSize/2 ; TemplateSize/2].

Two thresholds must be set by the user for the *dfd* computation:

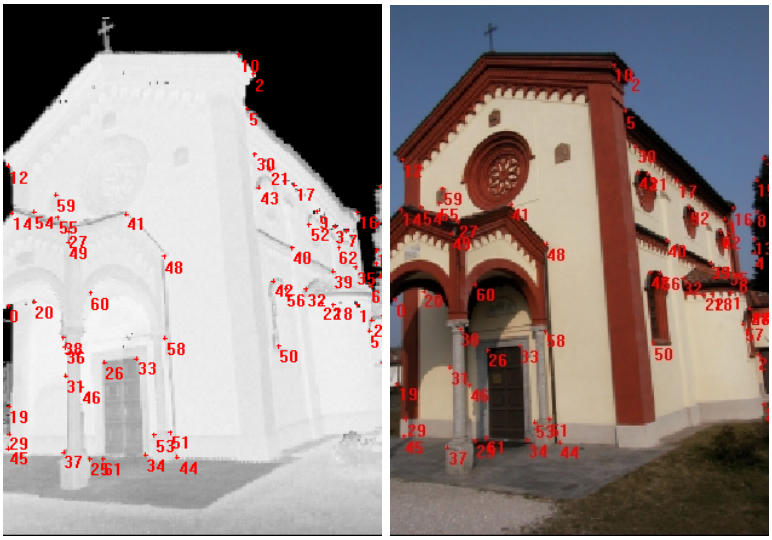
- The search area inside which the *dfd* is going to be computed for each pixel.
- The size template for the *dfd* computation.

The algorithm performs as follows: for each corner in the reflectance image, the *dfd* is computed with the corners in the intensity image located inside the search space. The matching pixel is selected as the intensity corner, within the given window, with the lowest *dfd*. Different thresholds are used in reflectance and intensity corner detection to ensure that more corners will be detected in intensity images.

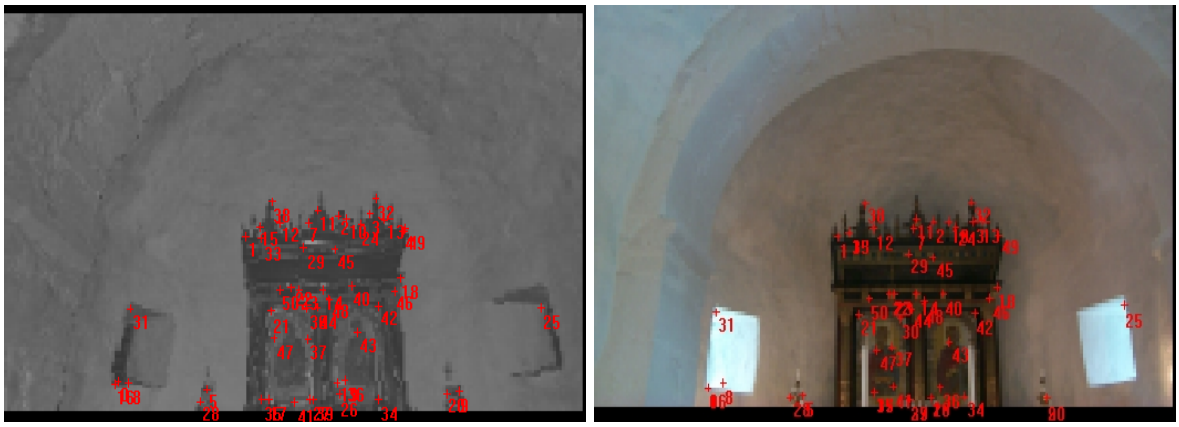
Figure 2-21 presents the final correspondences computed over the test images used on this chapter (see Figure 2-10 and Figure 2-16). In these four examples, the corner detection and matching algorithm performed in less than 2 s on a Pentium IV at 2.0 GigaHertz.



Laboratory: 50 correspondences (average distance between correspondences: 3.9 pixels)



Church of Barza: 63 correspondences (average distance between correspondences: 4.3 pixels)



Bornholm: 51 correspondences (average distance between correspondences: 2.5 pixels)

Laveno farmhouse: 54 correspondences (average distance between correspondences: 4.3pixels)

Figure 2-21: results of the matching algorithm between resized reflectance (left) and intensity image (right)

2.5 User interaction

The system presented in this chapter is almost fully automatic. The only step where user-interaction is necessary is the selection of the area of interest in the reflectance image at the beginning of the resizing process (see Figure 2-9).

The process still presents limitations and needs some rules during data acquisition to perform well. The resizing algorithm performance is related to the initial similitarity between the images. It will fail with images taken from very different viewpoints. The *dfd* matching also has some limitations due to the different nature of the images. Its performance will be reduced in images with little texture information.

For this reason, a user interface gives the possibility to interact with the process in several points to check and/or guide the resizing and matching steps.

The user can guide the process mainly in following operations:

- In the resizing algorithm, the user can tune the edge detection over the images. It can also refine the first affine transform approximation and check its quality by looking at the superposition of the edges. In addition, it is also possible for the user to remove edges that initially are far away from any other edge in the other image.
- The user can also set the size of the intervals in which the translations, scalings and rotations are computed, as well as the number of computed transformations in each step. A trade off between speed and accuracy appears: large intervals and small resolutions increase the accuracy but are very time consuming; smaller intervals and larger resolutions increase speed but result in a worse approximation.
- In the matching algorithm, the user can modify the thresholds for corner detection and *dfd* calculation. It can also visualize and interact with the computed matching in order to remove or add control points over the images.

2.6 Conclusions

This chapter presents many tools that can be used to help and guide the matching between reflectance and photographs. The range of application of these tools is still limited, and no automatic full calibration can be computed for every images. The resizing algorithm needs the reflectance and digital images to be similar and taken from close viewpoints to perform well. Another weak point of this process is the difficulty to get automatically fully reliable correspondences over the reflectance and intensity images. Thus, the images also need to be rich in texture and details to simplify the matching step.

Some tools can be used to improve the performance of the system. This is the case for rotation evaluation. But the domain of application of this technique is reduced and the rotation evaluation will bring some improvements only in images with strong structural contents and where the detection of the three principal vanishing points is possible.

The main innovation is the introduction of an adaptive resizing algorithm to locate reliable registration points. This algorithm can be used in other applications as shown in appendix B, where satellite images are registered using the same technique.

3 Camera calibration

Chapter two presented the tools developed to find corresponding points between reflectance and colour intensity images. This chapter focuses on how to use this information to compute the camera model and re-project intensity images into reflectance images. The resulting images are used as texture-maps for models. They also give a fast and easy way to check the quality of the registration between range and intensity data.

3.1 Calibration

Camera calibration and pose estimation are major issues in computer vision since they are related to many vision problems such as stereovision, structure from motion, robot navigation, change detection [Tsai86, Faugeras93, Fitzgibboon98a, Kumar94, Wilson94, Heikkila97, Pollefeys00, Zhang00, Debevec01, Kurazume02].

Camera calibration consists in the estimation of a model for an un-calibrated camera. The objective is to find the external parameters (position and orientation relatively to a world co-ordinate system), and the internal parameters of the camera (principal point or image centre, focal length and distortion coefficients). One of the most used camera calibration techniques is the one proposed by Tsai [Tsai86]. Its implementation needs corresponding 3D point co-ordinates and 2D pixels in the image. It uses a two-stage technique to compute: first, the position and orientation and, second, the internal parameters of the camera

In many computer vision applications a camera already calibrated is considered. This means that a model of the internal camera parameters is already available. This model can

be provided by the manufacturer or computed using a known target (usually a chessboard plane) [Heikkila97, Zhang00]. This situation, called pose estimation, just needs to recover the six parameters relative to the position and orientation of the camera. Some methods for pose estimation as well as a sensitivity analysis can be found in [Kumar94].

In our implementation [Dias02], we decided to use the well-known Tsai camera calibration method. This method has been chosen for its wide area of application since it can deal with coplanar and non-coplanar points. It also offers the possibility to calibrate internal and external parameters separately. This option is particularly useful since it gives the possibility to fix the internal parameters of the camera, when known, and carry out only pose estimation.

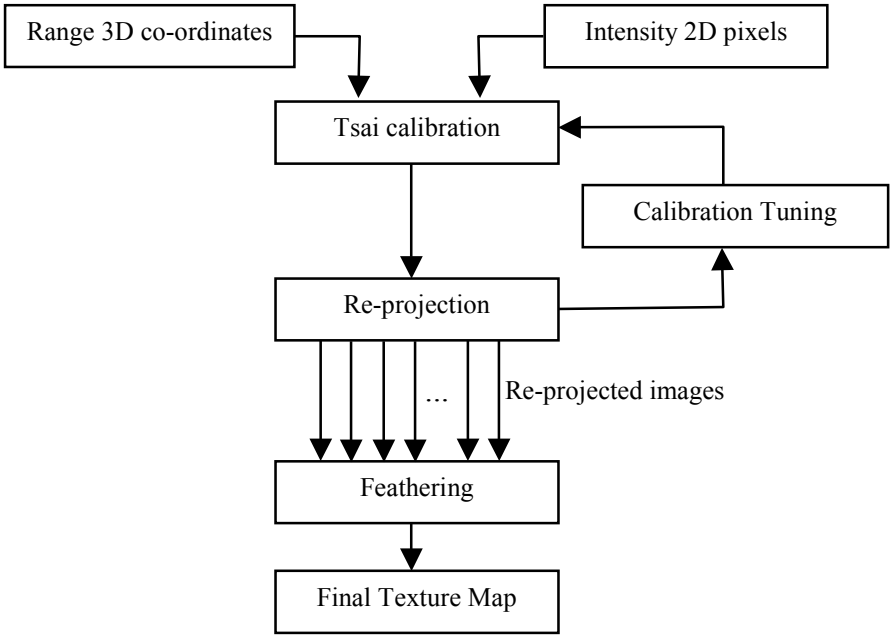


Figure 3-1: The camera calibration and re-projection process

Figure 3-1 illustrates the whole camera calibration process. The 3D-2D correspondences are used to compute the Tsai camera model. Based on this model, the intensity images are re-projected into the reflectance images to produce a re-projected image. This image can be used to refine the initial camera calibration or blended with other re-projected images to produce the final texture map for the model.

3.1.1 Tsai camera model

Given the matching between reflectance and intensity images, and since the reflectance image is fully registered with range data, it is easy to associate the 3D co-ordinates in range data with the corresponding pixel positions in the digital photograph. A Tsai camera calibration technique [Tsai86, Tsai87] is used to find the extrinsic and intrinsic parameters of the camera. The Tsai model is based on a pinhole perspective projection and the following eleven parameters are to estimate:

- *f* - focal length of camera,
- *k* - radial lens distortion coefficient,
- *Cx*, *Cy* - co-ordinates of centre of radial lens distortion,
- *Sx* - scale factor to account for any uncertainty due to imperfections in hardware timing for scanning and digitisation,
- *Rx*, *Ry*, *Rz* - rotation angles for the transformation between the world and camera co-ordinates,
- *Tx*, *Ty*, *Tz* - translation components for the transformation between the world and camera co-ordinates.

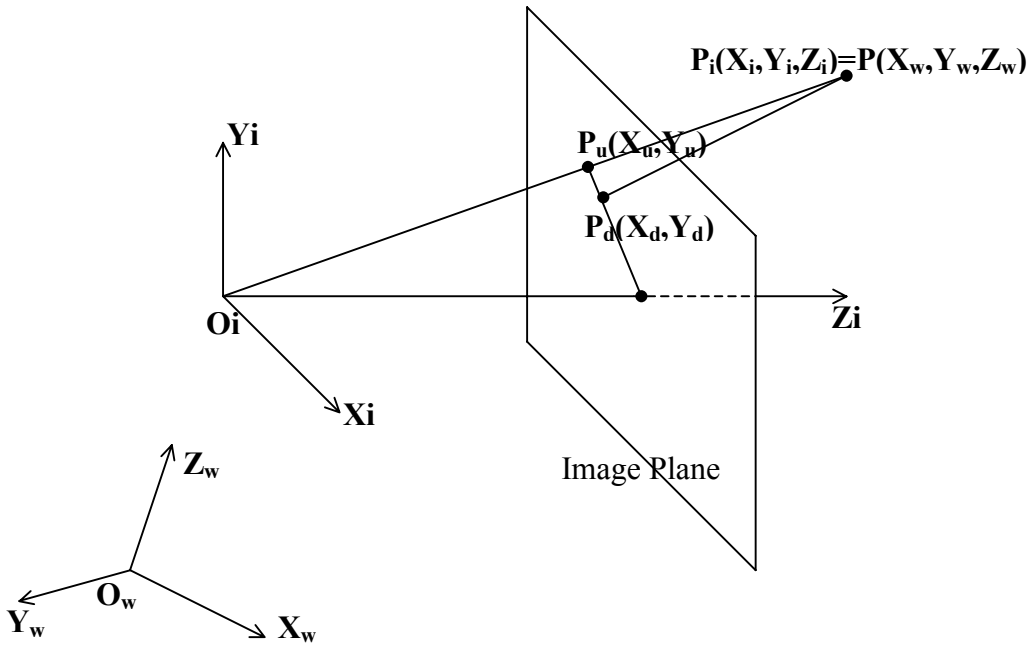


Figure 3-2: Tsai camera re-projection model with perspective projection and radial lens distortion

The transformation from world (X_w, Y_w, Z_w) to image (X_i, Y_i, Z_i) co-ordinates considers the extrinsic parameters of the camera (translation T and rotation R) within the equation:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T \tag{3.1}$$

where R and T characterize the 3D transformation from the world to the camera co-ordinate system and are defined as follows:

$$R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \text{ and } T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \tag{3.2}$$

with:

$$\begin{aligned}
r_1 &= \cos(R_y)\cos(R_z) \\
r_2 &= \cos(R_z)\sin(R_x)\sin(R_y) - \cos(R_x)\sin(R_z) \\
r_3 &= \sin(R_x)\sin(R_z) + \cos(R_x)\cos(R_z)\sin(R_y) \\
r_4 &= \cos(R_y)\sin(R_z) \\
r_5 &= \sin(R_x)\sin(R_y)\sin(R_z) + \cos(R_x)\cos(R_z) \\
r_6 &= \cos(R_x)\sin(R_y)\sin(R_z) - \cos(R_z)\sin(R_x) \\
r_7 &= -\sin(R_y) \\
r_8 &= \cos(R_y)\sin(R_x) \\
r_9 &= \cos(R_x)\cos(R_y)
\end{aligned}$$

(Rx, Ry, Rz) are the Euler angles of the rotation around the three axes.
(Tx, Ty, Tz) are the 3D translation parameters from world to image co-ordinates.

The transformation from 3D position (in the image co-ordinate frame) to the image plane is then computed through the following steps (see Figure 3-2):

- Transformation from 3D world co-ordinates (X_i, Y_i) to undistorted image plane (X_u, Y_u) co-ordinates

$$X_u = f\frac{X_i}{Z_i} \tag{3.3}$$

$$Y_u = f\frac{Y_i}{Z_i} \tag{3.4}$$

- Transformation from undistorted (X_u, Y_u) to distorted (X_d, Y_d) image co-ordinates

$$X_u = X_d(1 + kr^2) \tag{3.5}$$

$$Y_u = Y_d(1 + kr^2) \tag{3.6}$$

where $r = \sqrt{X_d^2 + Y_d^2}$, and k is the lens distortion coefficient.

- Transformation from distorted co-ordinates in image plane (X_d, Y_d) to the final image co-ordinates (X_f, Y_f)

$$X_f = \frac{S_x X_d}{d_x} + C_x \tag{3.7}$$

$$Y_f = \frac{Y_d}{d_y} + C_y \tag{3.8}$$

with d_x, d_y : distance between adjacent sensor elements in the X and Y direction. d_x and d_y are fixed parameters of the camera. They depend only on the CCD size and the image resolution,

(X_f, Y_f) the final pixel position in the image.

3.1.2 Robust estimation

Automatic matching systems return normally a significant number of wrong correspondences. Sometimes a few correspondences are completely wrong and can prevent any precise estimation if considered: these points are known as outliers. Outlier handling is widely studied in the literature and robust regression methods used to deal with this effect can be found, for example, in [Rousseeuw87, Torr97, Gracias97]. These works show that random techniques can be a valid alternative to achieve good estimation of the camera's parameters, for datasets containing a large percentage of outliers.

Given their different natures, reflectance and intensity images are difficult to match. As a result, the output of the matching algorithm (chapter 2) will be corrupted with many outliers. To ensure good calibration estimation even in these conditions, a well-known random method has been selected: the RANdom SAmple Consensus estimation technique (RANSAC) [Fischler81].

The RANSAC method computes several calibrations based on a minimal sub-set of correspondences selected randomly. For each of these "random calibrations", the technique searches for all supporting correspondences (correspondences with a re-projection error below a given level). All correspondences are then used to compute a new camera calibration. The process is repeated and the estimate having the larger set of points and the minimum error is selected as the final calibration. The algorithm needs three parameters:

i) The maximum error (in pixels) to consider a correspondence pair as supporting a given calibration;
ii) The sub-set of points used in each trial for the first evaluation;
iii) The number of trials.

For parameter i) we select a minimal error of re-projection of 10 pixels. The implementation we used for the Tsai camera calibration is the one in [Wilson94]. In this package, the intensity images are always normalized to a 512x512 pixels image. For this reason the re-projection error is always fixed to 10 pixels and does not need to be updated according to the size of the images.

Regarding parameter ii) we used the minimal number of points needed to perform the calibration as initial subset. As an example, for the full non-coplanar calibration, a subset of 11 points was used since it is the minimal number of correspondences the Tsai algorithm requires to perform the calibration in these conditions. The Wilson implementation permits to distinguish four different cases for the calibration, based on the co-planarity of the initial correspondences and the knowledge of the internal parameters of the camera [Wilson94]. Table 3-1 presents the four situations and the minimum correspondences needed for the algorithm. These values will correspond to the parameter ii of the RANSAC algorithm in each case.

Calibration type	Fixed parameters	Minimum Correspondences
1 - Coplanar with intrinsic	Internal parameters	5
2 - Non-coplanar with intrinsic	Internal parameters	7
3 - Full coplanar	None	7
4 - Full non-coplanar	None	11

Table 3-1: The different camera calibration cases considered

At the beginning of the calibration process, the 3D data is analysed for co-planarity. The median plane which best fits the 3D co-ordinates (in the sense that the sum of the squared error is minimized [Eberly00]) is computed. The maximum orthogonal distance from the 3D data to the median plane is evaluated and if this distance is lower than a user-defined threshold, the data is considered as coplanar and co-planarity functions are called (cases 1 and 3 in Table 3-1). In addition, if the internal parameters of the camera are already known (from previous calibration or through the manufacturer), the user can also provide this data and proceed only with pose estimation (cases 1 and 2 in Table 3-1).

To estimate correctly parameter iii) (number of attempts of the RANSAC algorithm) we need to know the number of outliers existing in the data. In this case, given a set of N points with b outliers and an initial selection of m points in the dataset, the number n of iterations to ensure that we select a group with no outliers (with a probability of p between 0 and 1) is given by [Lacey00]:

$$n = \frac{\log(1-p)}{\log(1-\left[1-\frac{b}{N}\right]^m)} \tag{3.9}$$

In our case, we do not know the percentage of outliers in the initial data; furthermore, it will depend on the type of image (the error will increase in poorly textured images where matching algorithms performance is worse). If we assume a full non-coplanar camera calibration (selection of 11 points) and 25% of outliers, we get with a probability of 75%:

$$n = 32.13$$

To validate these assumptions, 100 calibrations have been computed based in the 50 correspondences obtained on the laboratory image in chapter 2. The tests have been done with the following parameters:

i: error of 10 pixels.

ii: 11 points used for the initial sub-set (full non-coplanar calibration)

iii: 30 Iterations.

Figure 3-3 presents the results for the Laboratory scene. The histogram bars indicate the iteration in which the best estimate was found. The blue line represents the number of correspondences used to compute the best estimation. The red line is the final average calibration error. It is obtained by re-projecting the points used in the calibration (blue line) into the image and computing the distance between the original correspondence and the re-projected point.

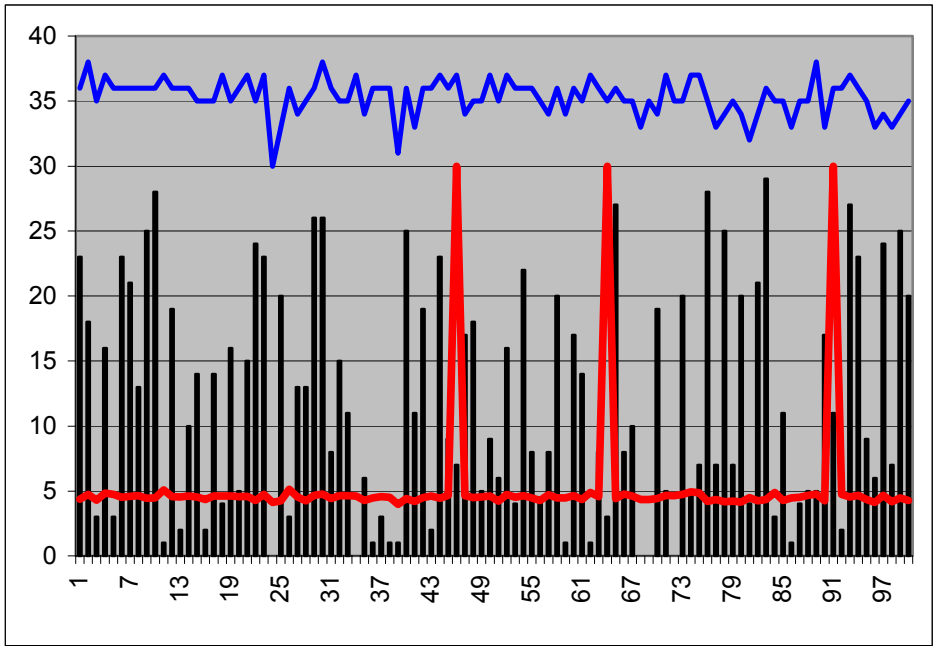


Figure 3-3: Results of 100 RANSAC calibrations using the 50 correspondences of the laboratory image

In the laboratory example, only three of the hundred calibrations failed with these parameters (we consider a failed calibration when the calibration error is larger than 30 pixels). All the other attempts achieved a good calibration with an average error of 4.78 pixels. The average number of points used was 35.34 and the maximum number of correspondences used in a calibration was 38.

The same experiment has been done with the other three sets of data presented in sections 2.2.4.and 2.3.1. The results are summarized in Table 3-2.

Example	Laboratory030	Barza church	Bornholm	Laveno farm
Number of matching	50	63	51	54
Missed calibrations	3	4	18	0
Successful calibrations	97	96	82	100
Average error (pixels)	4.78	4.27	6.98	4.73
Maximum number of points used	38	49	41	52
Average number of points used	35.34	44.00	32.63	46.4

Table 3-2: Results of 100 RANSAC calibration in the four examples used in chapter 2

Table 3-2 demonstrates that the parameters chosen for the RANSAC estimation give good results. In all the cases, less than 25% of the calibrations failed. In the Bornholm example, the bad performance is due to some particular features of the image: all the correspondences are located in a reduced part of the image. This situation is very difficult to calibrate since the Tsai camera calibration performs better when the correspondences are well distributed all over the image.

In addition, with these parameters the processing time is reasonable (around 12.5s in a Pentium IV at 2GHz for the laboratory example). Increasing the number of RANSAC iterations would lead to larger processing times with a minimal gain.

3.2 Re-projection into the reflectance image

After the calibration procedure, a full model for the camera is available. Using this information, each 3D co-ordinate in the range image can be re-projected into the intensity colour image according to the camera model. Since range and reflectance are directly registered, it is possible to establish an association between pixels in reflectance and intensity images, and compute a new reflectance image based on the intensity colour values.

The final image is useful to evaluate the quality of the registration in an easy and fast way. It can also be used directly to texture map the 3D models, giving a much more realistic impression than for a model only textured with the reflectance image. The complete procedure is summarized in Figure 3-4 with the example acquired at the round church in Borhnolm (Denmark).

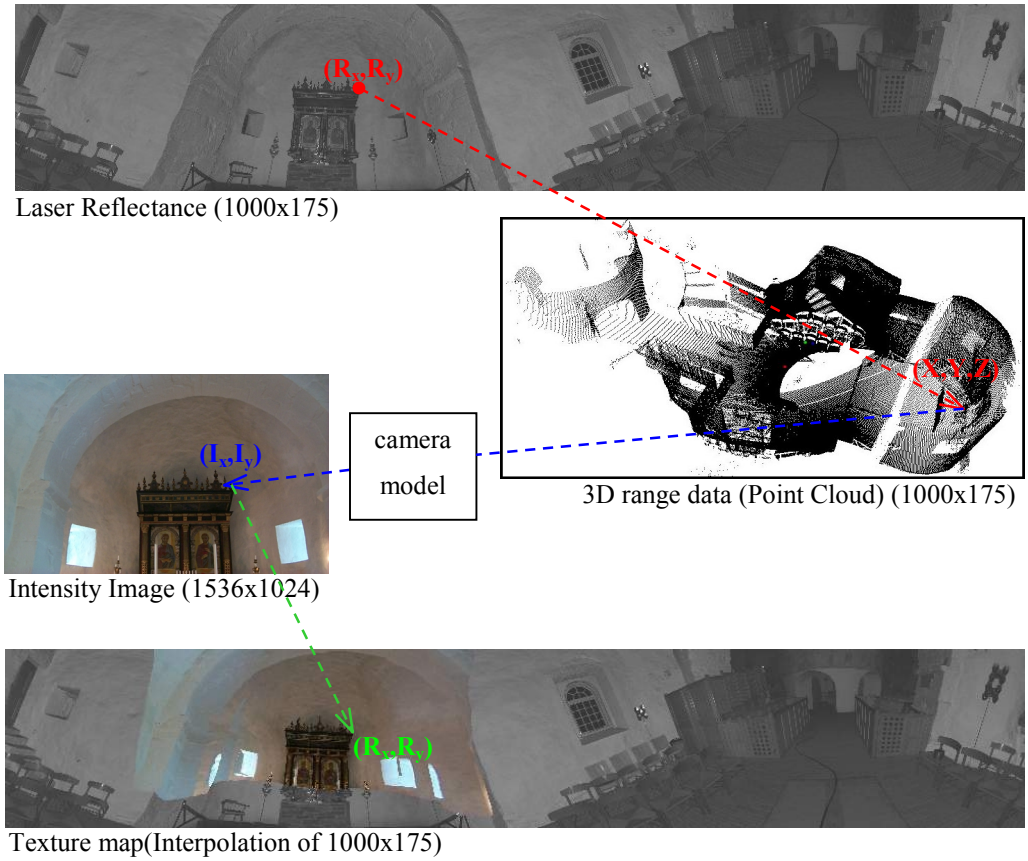


Figure 3-4: The re-projection procedure with an image of the Bornholm church

3.2.1 Texture map

The 3D co-ordinates, when projected in the intensity image, will not correspond normally to an integer pixel value. To avoid distortion in the colours, a bilinear interpolation (see section 2.2.1) is used to compute the resulting RGB value for the re-projected image.

Another issue is related to the different resolutions between reflectance and intensity images. The laser reflectance is usually of lower resolution than the digital photographs, and thus, back-projecting only the known 3D position will produce a texture map with the same resolution as the reflectance image. To avoid this limitation, the reflectance can be resampled to a higher resolution, and a bilinear interpolation is used to compute the "extra" 3D positions. Using this interpolation process makes possible to take advantage of the full resolution of the intensity image for the texture maps.

3.2.2 Z-buffering

Projecting the pixels of the intensity image into the reflectance image can lead to errors due to occlusions in the scene. To avoid this problem, a Z-buffer is filled with the Z co-ordinate of the closest 3D points for each pixel of the intensity image. When computing the texture map, only nearest points according to the Z-buffer are actually re-projected to ensure that occluded areas will not be projected in the texture map.

This problem is illustrated in Figure 3-5. Figure 3-5(a) shows the re-projection results without any Z-buffering. In this situation, some pixels of the intensity images appear more than once in the texture map and in wrong locations (arrows in the figures point at areas with occlusion errors). Using the Z-buffer technique solves the problem as shown in Figure 3-5 (b).

(a) (b)

Figure 3-5: Re-projection without Z-buffering (a) and with Z-buffering (b)

3.3 Calibration tuning

Due to the different nature of the images, a single run of the algorithm can lead to a calibration that is not fully satisfactory. To improve the reliability of the calibration process, an iterative procedure has been introduced. Figure 3-6 gives an overview of the tuning algorithm.

The matching process (see chapter 2) ensures the computation of a first approximation for the camera model. Based on this first calibration, the intensity image is re-projected into the reflectance co-ordinate frame, providing a texture map. At this point, the red channel of the re-projected image is used as an input for a new matching algorithm.

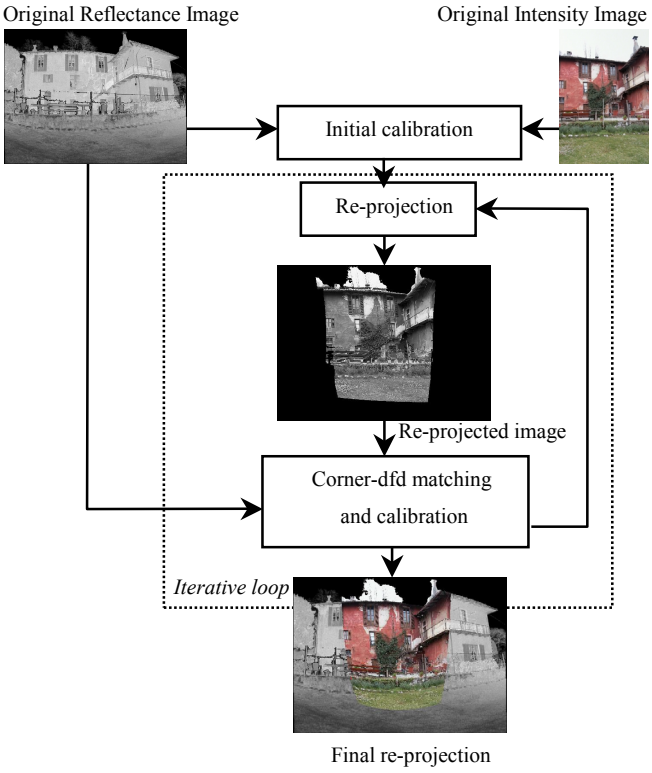


Figure 3-6: The calibration refinement process

At this stage of the process, reflectance and re-projected images are very similar (since the distortion introduced by the camera internal and external parameters is already considered by the re-projection) and a matching procedure is used. The matching algorithm used is the one presented in section 2.4. It uses the reflectance corners selected by the robust estimation (see 3.1.2) and detects corners in the re-projected image. Displaced frame difference (*dfd*, see section 2.4.2) is computed between neighbour corners and the corner of lower *dfd* is selected as the new matching point. The 3D co-ordinates of the range image and the corresponding pixels in the intensity image constitute the input to compute a new calibration.

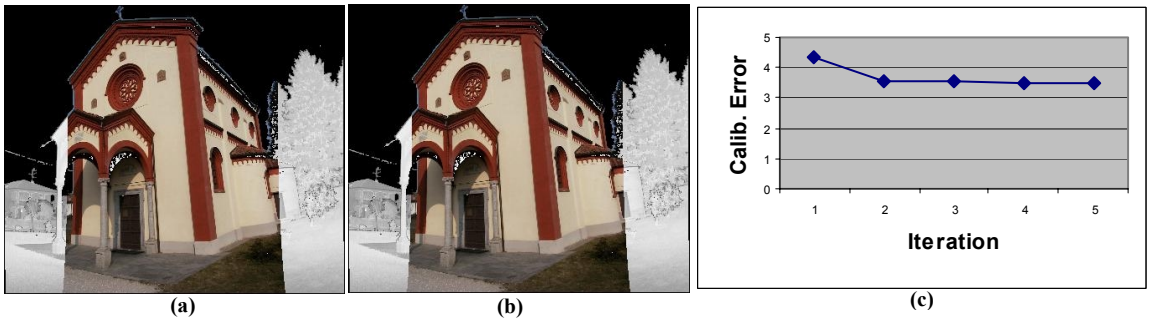
The process is repeated as long as the variation of the calibration error is larger than 5% between 2 iterations. This error is calculated by re-projecting the 3D co-ordinates into the intensity image and by computing the Euclidean distance between the re-projected point and the original correspondence used in the calibration. In each iteration, a new re-projected image is constructed from the last camera model and new correspondences are estimated.

To ensure the convergence of the algorithm and the improvement of the initial camera model estimation, the calibration error of each correspondence pair is computed and recorded. In every new iteration, only matching pairs for which the calibration error decreases are updated, and the other are kept unchanged. In this situation, no robust estimation is used since the technique ensures that no outliers are present within the correspondences.

Figure 3-7 (a) presents the re-projection with the initial calibration obtained using the correspondences computed in chapter 2. Figure 3-7(b) is the re-projection after calibration tuning. Finally, Figure 3-7(c) presents the evolution of the calibration error along the whole process. Table 3-3 gives an overview of the whole calibration process before and after the calibration tuning.

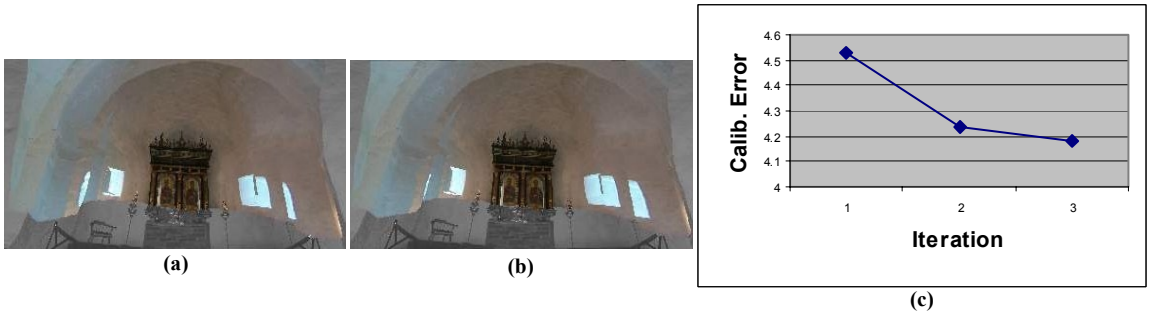
(a) (b) (c)

Laboratory 030 before (a) and after (b) calibration tuning and evolution of calibration error (c)



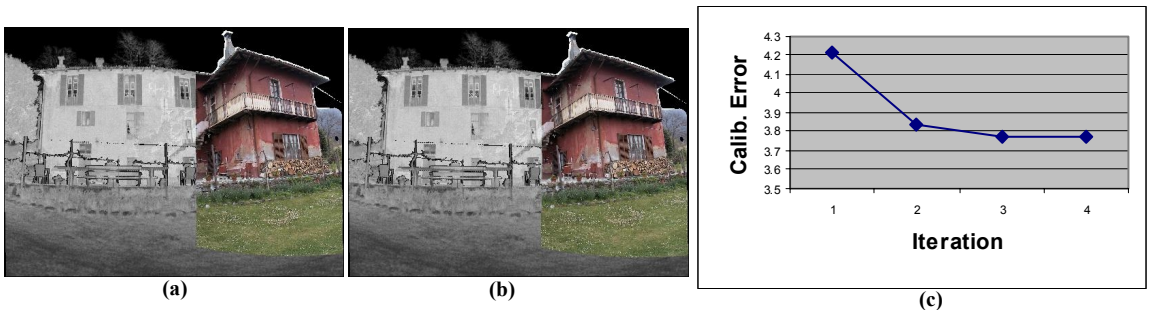
(a) (b) (c)

Church Barza before (a) and after (b) calibration tuning and evolution of calibration error (c)



(a) (b) (c)

Bornholm round church before (a) and after (b) calibration tuning and evolution of calibration error (c)



(a) (b) (c)

Laveno farm house before (a) and after (b) calibration tuning and evolution of calibration error (c)

Figure 3-7: results of the calibration tuning process

	Example	Laboratory 030	Barza Church	Bornholm Church	Laveno Farm
Initial calibration	Initial number of correspondences	50	63	51	54
	Correspondences used in RANSAC	34	46	36	44
	Initial calibration error	4.11 pixels	4.36 pixels	4.53 pixels	4.21 pixels
	Initial calibration processing time	9.92s	11.11s	16.39s	11.67s
Tuning	Final calibration error after tuning	3.83 pixels	3.46 pixels	4.18 pixels	3.77 pixels
	Number of iterations of tuning	4	4	2	3
	Tuning processing time	5.6s	11.21s	9.16s	8.74s

Table 3-3: Results of the calibration procedure before and after tuning (times in a Pentium IV at 2.0GHz)

3.4 Feathering

Typically, the field of view of a range sensor is much larger than the one from a normal camera: It is usual nowadays to find laser scanners capable of acquiring 270° or even 360° images. In these cases, more than one photograph is necessary to cover the whole reflectance image. To solve this problem, after each image is re-projected into the laser co-ordinate frame, a feathering operation is applied. This process consists in blending the data values in the zone where they overlap. The operation will result in a gradual transition from one image to another and will hide small misalignments between images.

The feathering process computes, for each pixel, the number of images that have a pixel projected in this position and the distance to the closest contour of each image. The final value of the pixel will be a weighted combination of the intensity values in each image according to the distance to the closest border. The idea is to decrease the influence of a given re-projected image as we move to its edges, in order to have a smooth transition to the next image. The formula used to compute the final intensity value in each RGB image is:

$$\text{Intensity}(x,y) = \frac{\sum_{i=1}^{n} \text{Intensity}_i(x,y) \times dist_i(x,y)}{DISTANCE\_TOTAL}, \tag{3.10}$$

with $DISTANCE\_TOTAL = \sum_{i=1}^{n} dist_i(x,y)$

$\text{Intensity}_i(x,y)$: intensity of pixel (x,y) in image i.

$dist_i(x,y)$: Euclidean distance of pixel (x,y) to the closest border in image i.

n: number of images with a pixel projected in (x,y)

Different methods for computing the value *dist(x,y)* have been tried since the texture images have normally a high resolution and the processing time starts to be a non-negligible issue. As in most engineering problems, there is a trade off between time and quality: the faster is the method, the worse is the quality. Three different experiments have been done with average, rectangular and weighted blending and are presented in the next sections.

3.4.1 Average blending

Average Blending is the simplest blending process implemented. It consists, for each pixel of the texture map, in adding the RGB values of all the images that contribute to this pixel and divide the result by the number of images.

$$\text{Intensity}(x,y) = \frac{\sum_{i=1}^{n} \text{Intensity}_i(x,y))}{n} \tag{3.11}$$

with

$\text{Intensity}_i(x,y)$: intensity of pixel (x,y) in image i.

n: number of images with a pixel projected in (x,y)

Figure 3-8 presents the results of the average blending of twelve re-projected images in the laboratory scene. Areas for which no colour information is available are filled with the reflectance image. The process took approximately 1.5s in a Pentium IV at 2.0GHz.

Figure 3-8: Average blending of 12 re-projected images of the laboratory scene

3.4.2 Rectangular blending

Rectangular blending approximates the area in which the image is re-projected by a rectangular area (see Figure 3-9) and uses this approximation to compute the distance to the closest border. This process uses the blending formula (3.10), but in this case, the value dist(x,y) is computed as the distance to the closest border of the smallest rectangle that contains the re-projected image i:

$$Dist_i(x,y) = \min(d_{i1}(x,y), d_{i2}(x,y), d_{i3}(x,y), d_{i4}(x,y)) \tag{3.12}$$

where d_1, d_2, d_3 and d_4 are defined as shown in Figure 3-9.

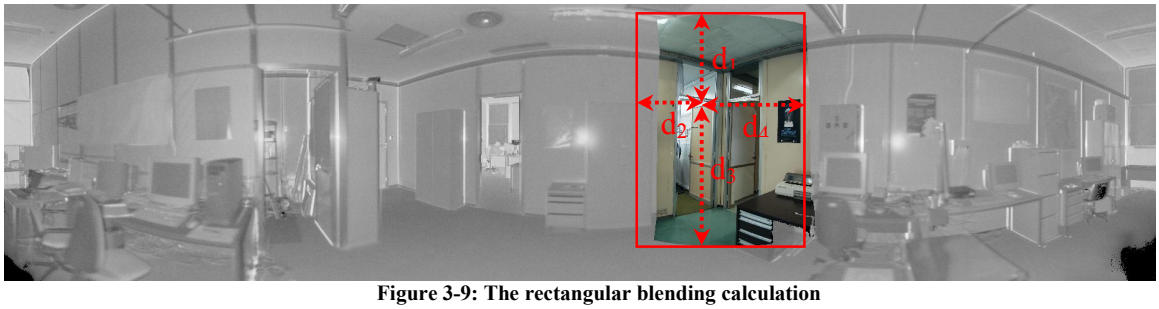


Figure 3-9: The rectangular blending calculation

Results of the rectangular blending are presented in Figure 3-10. In this case, the algorithm performed in approximately 1.6s in a Pentium IV at 2.0GHz.

Figure 3-10: Rectangular blending of 12 re-projected images of the laboratory scene

Rectangular blending offers the advantage to be almost as fast as average blending but with better feathering results. It is particularly interesting when the acquisition leads to re-projected images that can be easily approximated by a rectangular area (this is the case of the laboratory scene). In such conditions, rectangular blending is a good trade off, since it gives good results with a minimal time-cost.

3.4.3 Weighted blending

The last blending technique implemented uses the Euclidean distance to the closest border of the re-projected area. In this case, the computation cost is heavier since the Euclidean distance to each border has to be computed and no approximation is assumed.

Although this method is more time-consuming (the algorithm performed in 20s in a Pentium IV at 2.0GHz for the laboratory images), it leads to the best feathering results as presented in Figure 3-11: the transition between images is smoother than with the other methods.

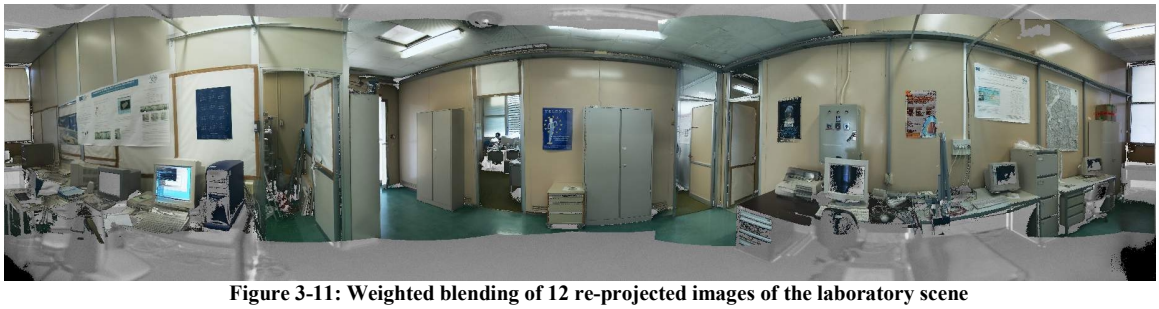


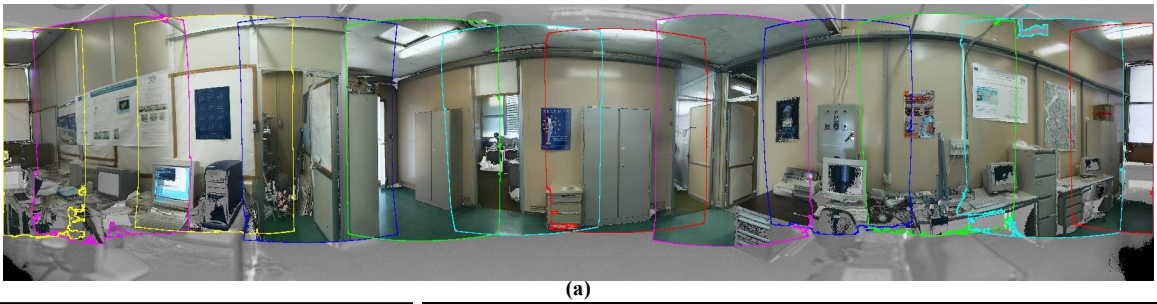
Figure 3-11: Weighted blending of 12 re-projected images of the laboratory scene

3.5 Results

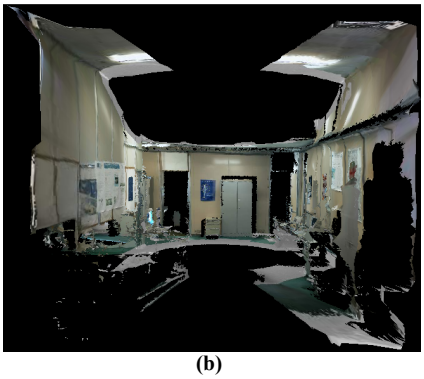
Figure 3-12, Figure 3-13, Figure 3-14 and Figure 3-15 present the different steps of the re-projection and feathering process for the four scenes that have been used to illustrate our algorithms. The figures present two indoor scenes (the laboratory at JRC and the Bornholm round church models) and two outdoor scenes of different kinds of buildings (a church in Barza and a Farm house in Laveno). The laboratory, the Barza church and the Laveno farm were acquired with a RIEGL LMS-Z210 laser scanner and a Canon Proshot 70 digital camera. The church in the Bornholm Island was acquired with a Zoller&Fröhlich LARA laser range scanner [Fröhlich98] and a Minolta digital camera.

For each example we present: the final texture map obtained from the re-projection and feathering of the intensity images (a) and snapshots of the model after applying the texture map (b,c). Table 3-4, Table 3-5, Table 3-6 and Table 3-7 present for each example the results of the calibration of the images used to compute the texture map. These tables show, for each calibrated image, the resizing error (average distance between edges after the resizing), the number of matchings obtained after the resizing algorithm, and the number of matchings actually used in the robust estimation of the camera model. Finally, we also present the initial and final calibration error (before and after the calibration tuning operation).

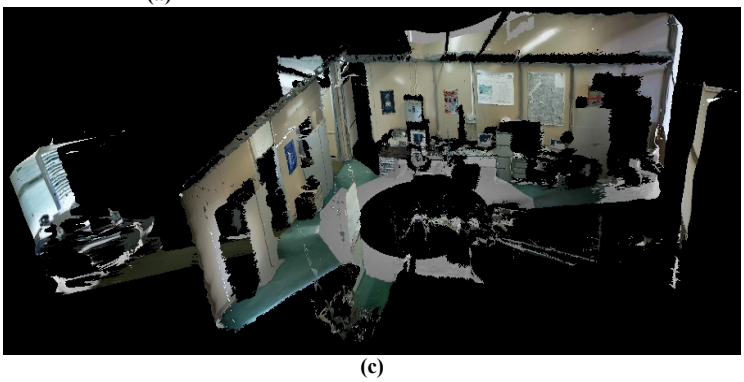
In the laboratory, the Barza church and the Laveno examples, the only intervention of the user was the selection, in the reflectance image, of a rectangle area corresponding approximately to the same field of view of the intensity images. This selection was then used as input for the resizing algorithm. The rest of the process was fully automatic.



(a)



(b)

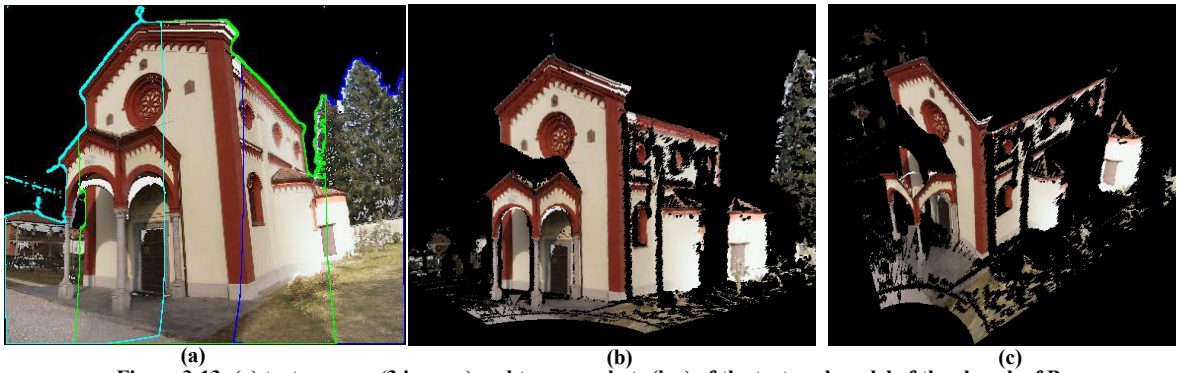


(c)

Figure 3-12: (a) texture map (12 images) and two snapshots (b,c) of the textured model of the laboratory

Image	1	2	3	4	5	6	7	8	9	10	11	12
Resizing error	9.05	7.01	7.11	5.83	5.47	5.44	6.45	5.63	7.18	5.17	7.09	11.12
N. of matching	44	43	62	66	31	51	16	50	40	55	60	45
RANSAC matching	22	24	37	31	20	26	13	34	27	31	33	24
Initial calib. error	3.97	4.97	4.59	5.48	5.16	4.17	3.10	4.11	6.50	5.45	4.49	4.91
Final calib. error	3.61	4.01	3.89	3.96	4.59	3.32	1.85	3.83	4.74	4.23	3.64	4.37

Table 3-4: Calibration results of the 12 intensity images used in the laboratory 030 example

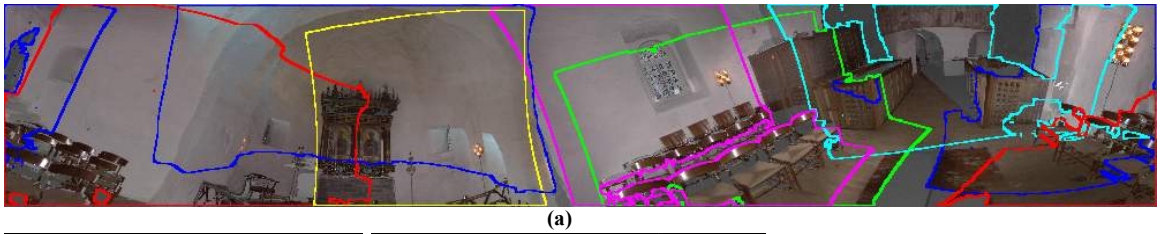


(a) (b) (c)

Figure 3-13: (a) texture map (3 images) and two snapshots (b,c) of the textured model of the church of Barza

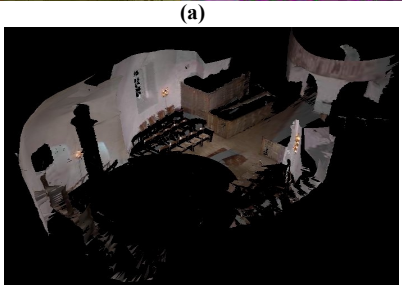
Image	1	2	3
Resizing error	4.60	5.92	3.48
n. of matching	58	63	45
RANSAC matching	25	46	29
Initial calib. error	3.41	4.36	4.98
Final calib. error	2.90	3.46	4.52

Table 3-5:Calibration results of the 3 intensity images used in the church Barza example



(a)

(b)



(c)

(d)

Figure 3-14: (a) texture map (7 images) and three snapshots (b,c,d) of the textured 3D model of the Bornholm church

Image	1	2	3	4	5	6	7
Resizing error	1.89	no	no	no	no	2.71	no
N. of matching	51	17*	14*	15*	13*	55	14*
RANSAC matching	36	15	13	14	12	21	13
Initial calib. error	4.53	3.39	3.97	2.68	2.12	3.98	3.05
Final calib. error	4.18	3.39	3.75	2.68	3.42	3.46	2.68

Table 3-6: Calibration results of the 7 intensity images used in the Bornholm example

In 5 images (*) points were chosen manually since intensity and range data were acquired from different viewpoints

(a)

(b)

(c)

Figure 3-15: (a) texture map (3 images) and two snapshots (b,c) of the textured model of the Laveno farmhouse

Image	1	2	3
Resizing error	2.93	3.41	3.33
N. of matching	60	56	54
RANSAC matching	26	35	44
Initial calib. error	5.27	4.84	4.21
Final calib. error	3.60	3.85	3.77

Table 3-7: Calibration results of the 3 intensity images used in the Laveno farm example

In the Bornholm example, digital photograph and range data were acquired from different viewpoints, making difficult the application of the resizing algorithm. In this situation, control points were manually chosen for five of the seven images (corresponding to “no” resizing and “*” for the control points in Table 3-6).

3.6 Discussion

The main characteristics of the whole registration algorithm presented in this chapter are:

- Independence from the acquisition sensors. The technique has been successfully used with different experimental set ups (laser scanners and cameras from different manufacturers with different resolutions and options);
- No need for previous calibration;
- Simple evaluation of the quality of the registration with the computed re-projected images;
- Possibility for the user to interact with the system and, if necessary, guide the process;
- Polyvalent system that allows the use of several configurations to optimise the registration step. For example, if internal parameters of the camera are known, these parameters can be directly fixed and only external camera parameters will be evaluated.

The algorithm performance depends on the quantity and quality of the 2D features and on the similitude between the images (better results are obtained with images taken from close viewpoints).

The main innovations introduced are the colour feathering of the merged digital photographs, the occlusion handling for selecting the pixels to be used in the final texture map and a calibration-tuning algorithm to refine the first calibration computed based on robust estimation.

A weak point of the calibration procedure is still the difficulty to get automatically fully reliable correspondences over the reflectance and intensity images.

4 3D Edge Correction based on a single intensity image

In the previous chapters, an innovative technique for the semi-automatic registration of laser range data and intensity images was presented. The registered intensity images are used as texture maps to enhance the 3D models computed from range imagery, but all 3D geometry is still extracted only from range data analysis.

The idea developed in this and the next chapters, is based on the use of registered intensity images not only to texture-map but also to improve 3D range segmentation and model geometry (Figure 4-1). This will give the possibility to improve the quality of the models, without acquisition of extra range data, but using the already available 2D intensity information as an additional source of 3D geometry.

The method presented in this chapter does not extract directly 3D information from an intensity image, it uses features detected in the intensity image to improve segmentation in range data [Dias01b]. The process focuses on a simple situation and considers the already triangulated model, and not the cloud of points, to make easier the detection and matching of 3D edges.

The chapter starts by pointing out some of the limitations of range sensors that can be compensated using a single registered intensity image. Our process detects edges in the 3D triangulated model and straight lines in the intensity images. The model of the camera is used to compute correspondences between straight lines in 2D and edges in 3D. The final step of the correction process consists in the alignment of the edges. At the end of the chapter, the process is extended to more complex features (quadratics) and a Z-buffer technique is proposed to handle fully-triangulated models and deal with occlusion problems.

Triangulated model from range data

Range registered digital photograph

Improved 3D model

Figure 4-1: Illustration of model correction based on a single registered image

4.1 Error sources in range acquisition

Laser range finders are subjected to several sources of errors. One of their limitations is the dependence of the measurement accuracy on the type of laser scanner and the integration time. With the Riegl LMS-Z210 scanner used for many acquisitions, the accuracy is typically of 25mm. These devices also have limited spatial resolution depending on the spot size and the angle readout accuracy (0.018º in the Riegl LMS-Z210).

Given their working principle, LRF efficiency depends also on the reflectance properties of the measured objects. A bright object can appear one or two centimetre closer since the reflected beam energy affects the detection thresholds. Errors will also appear with glasses: in this case, no light is reflected and the laser scanner will measure the distance to the closest object behind the glass. A similar situation occurs with mirrors: the distance to the next target reflected in the mirror will be returned [Sequeira96].

One of the main sources of error of Laser Range Finders is that the footprint of the laser beam is not an ideal point but rather a circular area (with the Riegl Laser, the divergence of the beam is approximately 3mrad - meaning a beam width of 30cm at 100m). Effects of this error appear clearly when the footprint lies in a region where depth changes rapidly. In this case the distance measured does not reflect anymore the true distance to the object but a combination of the two areas of the footprint. This effect, known as the “mixed point problem”, is illustrated in Figure 4-2(a). Figure 4-2(b) presents a close view of the resulting edges in the triangulated model of the laboratory scene.

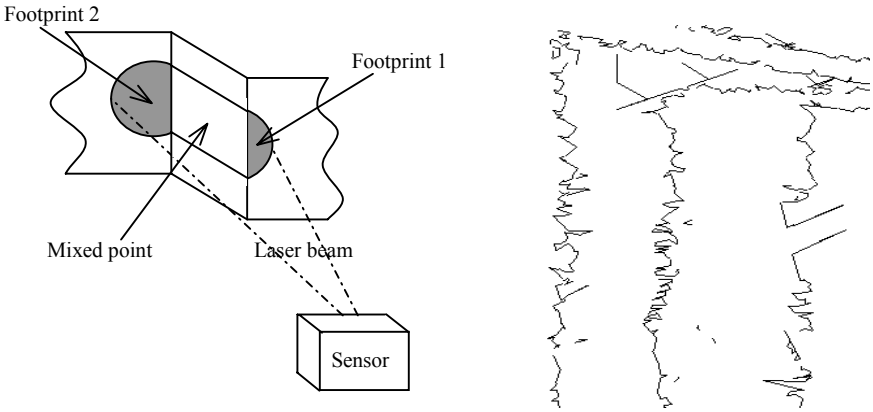
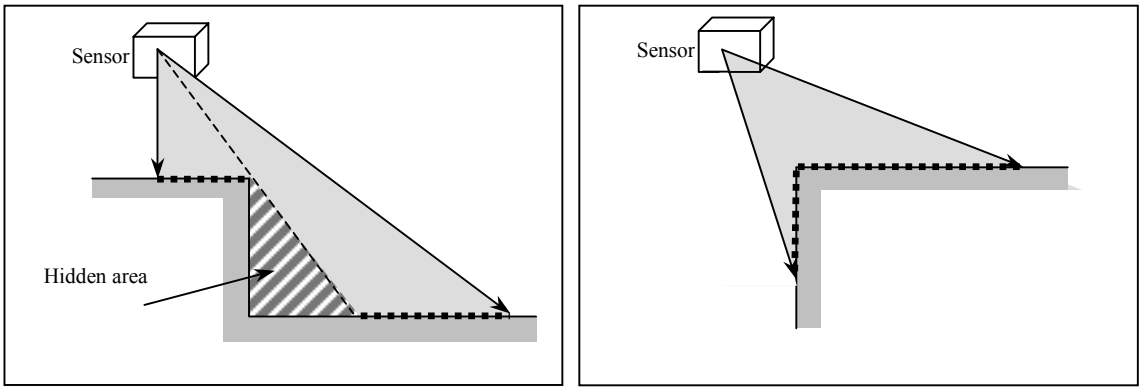


Figure 4-2: The mixed point problem (a) and its effects in the edges of a triangulated model (b)

4.2 Detection of edges in 3D triangulated models

Range images present three different types of geometric edges: jump edges (corresponding to a discontinuity due to an occlusion), roof edges (when the normal angle changes rapidly) and smooth edges (due to a reduced variation in the normal). The process concentrates on jump edges (Figure 4-3 (a)) and roof edges (Figure 4-3 (b)) for several reasons:

- Effects of the mixed point problem appear mainly in these two edges since they both are characterized by a strong discontinuity;
- They are easily detectable in a triangulated 3D model as they correspond to the end of a surface or to rapid changes in the normal of the triangles;
- Jump and roof edges correspond usually to well-marked edges in intensity images (changes in colour, position, etc.), whereas smooth edges often do not appear clearly and are difficult to trace in an intensity image;
- There is no need of re-triangulation of the whole model to correct jump and roof edges since they are located at borders of surfaces. Just "pushing" them to a correct position brings significant improvements and does not implicate the re-triangulation of the whole surface.

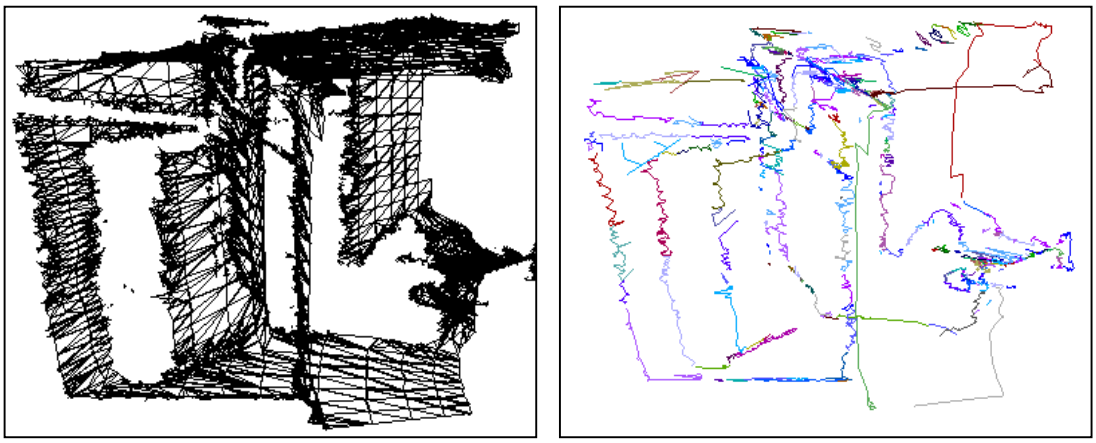


(a) (b)

Figure 4-3: (a) Jump edge and (b) roof edge

4.2.1 Jump Edges

Detection of jump edges in a triangulated model is an easy task since jump edges correspond to the border of surfaces with no contact with any other triangles. Thus, the algorithm simply checks which segments belong only to a single triangle and marks the two vertices of these segments as belonging to a jump edge. Figure 4-4(a) presents the triangulated mesh of a part of the laboratory scene, and Figure 4-4(b), the detected jump edges in this model.



(a) (b)

Figure 4-4: Triangulated model (a) and detected jump edges (b) of the laboratory scene

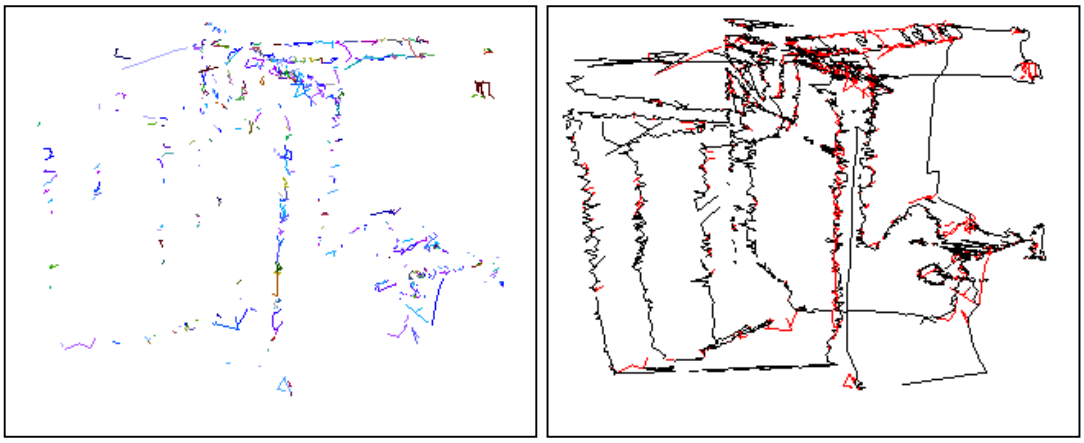
Once jump edges are detected in a model, an ordering algorithm traces all the connected vertices inside each edge. As a result, each detected jump edge is defined as a list of ordered points corresponding to all its vertices. In Figure 4-4(b) each connected jump edge is displayed with a different colour.

An appropriate data structure (that optimises the navigation through the topology of the triangulated model) is used to optimise the detection and ordering processes making the whole procedure fast and efficient. The detection and ordering of 297 jump edges (see Figure 4-4(b)) took approximately 0.28s in a Pentium IV at 2GHz.

4.2.2 Roof Edges

Detection of roof edges is based on the normal of each triangle of the triangulated model. For each triangle, two sides are selected ant the normal is computed as the external product between these two vectors. The value of the angle between the normals of adjacent triangles is used as test condition: if this angle is higher than a given threshold, the edge separating the two triangles is marked as a roof edge.

The same ordering algorithm mentioned in section 4.2.1 is used to order the vertexes inside each roof edge. The roof edges detected in the test scene are presented in Figure 4-5(a) for a threshold of 40 degrees between two adjacent triangles. The detection of these 437 roof edges was performed in 0.3 s in a Pentium IV at 2GHz. Figure 4-5(b) presents in the same snapshot roof (in red) and jump edges (in black).



(a) (b)

Figure 4-5: (a) Roof edges detected with a threshold of 40°, (b) Roof (red) and jump (black) edges in the laboratory scene

4.3 Line detection in calibrated Intensity Image

It is now possible to look for edges of interest in 2D digital photographs: edges that can be used to guide the correction of 3D edges and minimize the errors and visual effects due to the mixed point effect (Figure 4-2).

On a first stage, the techniques concentrate in straight lines since they are easy to detect and correct and appear very often in man-made environments. Furthermore, straight lines are particularly sensitive to the mixed point problem and appear often as a saw tooth, producing a very negative visual effect.

4.3.1 Automatic line detection

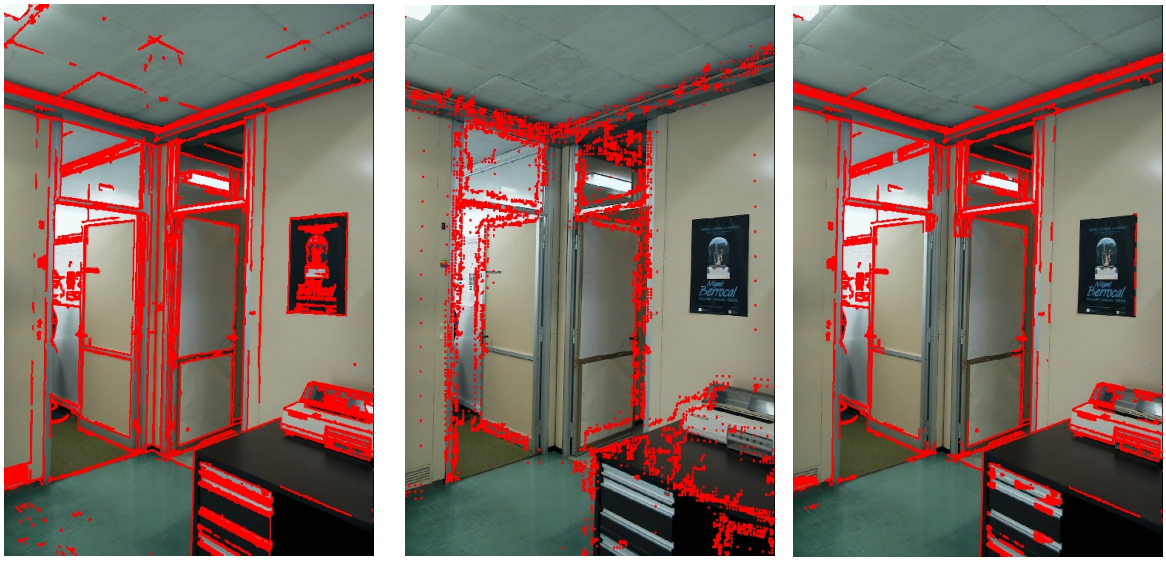
Linear features are often used in 3D reconstruction and matching processes [Clarke96, Tuytelaars98, Baillard99]. In a first attempt, we used already available software to detect straight lines in intensity images, based on a Sobel gradient operator, and followed by grouping and straightening operations. We came to the conclusion that using only image-based techniques was not very efficient since we obtained lots of undesirable edges due to changes in colours, posters, etc. Which correspond to variations in the photograph but not to any 3D geometrical information.

An automatic procedure was then developed to combine 2D visual information (from the intensity image), and 3D spatial information about the 3D edges, to optimise the line detection in the areas of interest and reject most of the undesirable edges.

The base of the line detection procedure is the well-known Canny edge detector [Canny86]. The vertices of the 3D edges detected in the triangulated model (see section 4.2) are re-projected into the intensity image using the available camera model. Edges extracted by the Canny detector but located further than a given distance from any re-projected edge point will be removed. This ensures that only 2D features located in the neighbourhood of 3D edges will be considered. The procedure is illustrated in Figure 4-6. Figure 4-6(a) presents the results of the initial Canny edge detection. The re-projected edge points (from both jump and roof edges) into the photograph image appear in Figure 4-6(b). The resulting edges, Figure 4-6(c), come from the filtering of the original edges with the re-projected edge points.

The main advantage of the filtering operation is the removal of 2D intensity edges that do not correspond to any 3D geometric feature. A typical example is the poster present in

the right wall of Figure 4-6. The Canny edge detector detects the edges of the poster (Figure 4-6(a)) but no 3D edge is present in this area since the poster just lay on a surface and does not bring any additional geometric information (Figure 4-6(b)). In the final result (Figure 4-6(c)), no edges associated to the poster remain, ensuring that the algorithm will concentrate in areas with significant geometrical information.



(a) Original Canny edge detection **(b) Re-projection of the 3D edges** **(b) Edges after the filtering**

Figure 4-6: The filtering based on 3D edge information

The last step of the line detection process will use a RANSAC technique [Fischler81] to extract linear segments from the analysis of the filtered edges. This line detection process is similar to the one already presented in section 2.3.1.1.1.

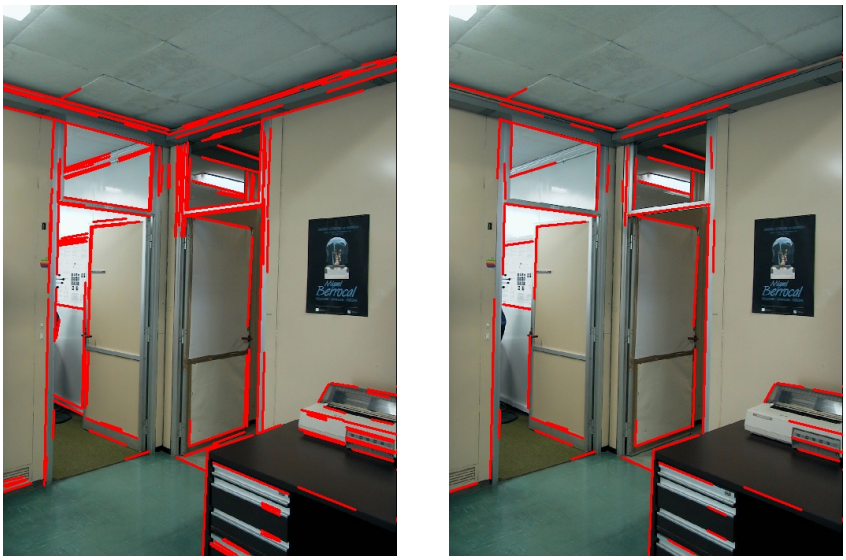
4.3.2 Filtering of redundant straight lines

The line detection process presented in the previous section can lead to many lines concentrated in the same area of the image.

To avoid redundant straight lines in the same area of the intensity image, a filter based on orientation and distance between segments is applied. Given two user-defined thresholds (Th1 and Th2) to set the minimum distance and orientation between segments, a segment s is valid only if for all the other segments i:

$$\begin{aligned} &Dist(i,s) > Th1 \\ &\text{or } Dist(i,s) < Th1 \text{ and } Angle(i,s) > Th2 \\ &\text{or } Dist(i,s) < Th1 \text{ and } Angle(i,s) < Th2 \text{ and } n(s) > n(i) \end{aligned} \qquad (4.1)$$

where, Dist(i,s) is the minimum Euclidean distance between segments i and s, Angle(i,s) is the angle between segment i and s and n(i) is the number of points defining segment i.



(a) (b)

Figure 4-7: The results of the line detection procedure before (a) and after the filtering operation (b)

Figure 4-7 presents the results of line detection, 159 straight lines were detected originally (a), and 59 straight lines remain after the filtering operation (b). The whole procedure (edge and line detection and filtering) was performed in 5.8 s in a Pentium IV at 2GHz.

The filtering operation will remove a significant number of segments. Since the 3D analysis will be done along the lines defined by the segments and not only within the segments, this strong filtering will not result in a reduction of the number of corrected edges.

4.4 Correction of the 3D linear edges

At this point, the process has already selected the 3D edges in the model and the straight lines in the intensity images that could lead to eventual corrections. The next step is to compare and define corresponding lines and correct them in the triangulated model.

4.4.1 Selection of initial 3D line estimation using a RANSAC approach

For each detected segment in the intensity image, the algorithm selects the 3D edge points that are projected close to the line defined by the segment (the measure used is the orthogonal distance to the line).

Figure 4-8, presents a segment in the intensity image (black) and the re-projection of the selected 3D points (white). A RANSAC [Fischler81] technique is applied to pairs of 3D points to determine the 3D line **L** with more support among the 3D points selected.

Figure 4-8: The original detected segment (black) and the re-projected 3D points in its neighbourhood (white)

4.4.2 Determination of the jump edges to correct

Once the 3D line with more support is selected, the algorithm follows the range discontinuities (according to the ordered list defined in section 4.2.1) to select the 3D edges that might correspond to the given 3D line.

Two main conditions have to be fulfilled by an edge to be selected:

- All the points of the edge have to be spatially close to the line **L** in three dimensions. This is verified in 3D using the orthogonal distance; given two points A and B, the orthogonal distance **Dist** of a point C to the line defined by A and B is defined as (where AC, AB and BC are the size of the segments and • is the scalar product of two vectors):

$$\mathbf{Dist} = \frac{\sqrt{AC^2 AB^2 - (AC \bullet AB)^2}}{AB} \qquad (4.2)$$

- The linear regression of the selected 3D edge points re-projected in the digital photographs must have an orientation similar to the original segment. Given θ_s the orientation angle of the initial segment, and θ_l, the orientation angle of the regression line defined by the re-projected 3D points, we have:

$$\theta = \text{abs}\,(\theta_l - \theta_s) < \text{threshold} \tag{4.3}$$

4.4.3 Correction of the 3D co-ordinates

Finally, our algorithm will correct the 3D co-ordinates of points close to straight lines in the triangulated models. This is done in two steps. First, the parameterised equation of the 3D line is computed, and then 3D co-ordinates are modified so that the final points will lie in the computed line. The algorithm will be used iteratively with adaptive thresholds to correct as many edge points as possible.

4.4.3.1 Refining of the 3D line parameters

Given all the points to be corrected according to the 3D line **L,** it is possible to refine the 3D line parameters by computing a linear fitting based on orthogonal regression [Eberly00], and define the 3D line **L'** that best fits all the points. The Origin point (**O**) and Direction vector (**D**) that define the optimal line **L'** between all points to correct $\mathbf{X}_i(x_i, y_i, z_i)$ are defined as follows. O is the centre of gravity of all the points:

$$\mathbf{O} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i\,; \tag{4.4}$$

given **O**(a,b,c), the direction vector **Dir**(D_x,D_y,D_z) is defined as a unit length eigenvector corresponding to the smallest eigenvalue of the matrix [Eberly00, Schneider03]:

$$\delta\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} - \begin{bmatrix}\sum_{i=1}^{m}(x_i-a)^2 & \sum_{i=1}^{m}(x_i-a)(y_i-b) & \sum_{i=1}^{m}(x_i-a)(z_i-c)\\ \sum_{i=1}^{m}(x_i-a)(y_i-b) & \sum_{i=1}^{m}(y_i-b)^2 & \sum_{i=1}^{m}(y_i-b)(z_i-c)\\ \sum_{i=1}^{m}(x_i-a)(z_i-c) & \sum_{i=1}^{m}(y_i-b)(z_i-c) & \sum_{i=1}^{m}(z_i-c)^2\end{bmatrix} \tag{4.5}$$

with $\delta = \sum_{i=1}^{m} (x_i - a)^2 + (y_i - b)^2 + (z_i - c)^2$

4.4.3.2 Correction of single point co-ordinates

After the optimal Line **L'** is found, the three dimensional co-ordinates of the orthogonal projection of each point to correct are computed. Given the original 3D point $X(x, y, z)$, the corrected point $X'(x', y', z')$, $O(a,b,c)$ and $Dir(D_x, D_y, D_z)$ the origin and direction vectors of **L'**, we have:

$$X' = O + t \times Dir \tag{4.6}$$

with

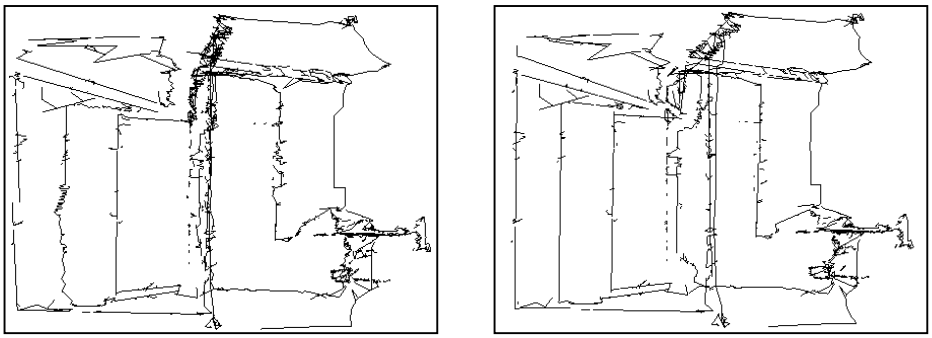
$$t = \frac{xD_x + yD_y + zD_z - aD_x - bD_y - cD_z}{D_x^2 + D_y^2 + D_z^2} \tag{4.7}$$

4.4.4 Iteration with adaptive thresholds

To make the algorithm more effective and less dependent on the user defined thresholds (depending on the images, the results after one iteration can be fairly good or not) the procedure is ran iteratively, each time with new thresholds. The algorithm stops when no correction is computed or when a user-defined number of iterations is reached (three iterations are normally enough). The modified thresholds from iteration to iteration are:

- The distance between a 2D line in the intensity image and a re-projected point, to be considered, is increased from iteration to iteration (Increase of 3 pixels per iteration). (see Figure 4-8)
- The distance **Dist** between the points and the 3D line **L** is increased. This allows a larger search space around the line in successive iterations. (Increase of 2cm per iteration).
- The difference θ in the orientation between the segment in the intensity image and the re-projected edge is also increased. (Increase of 5º per iteration).

Figure 4-9 presents the results for the laboratory scene after one (a) and three iterations (b)



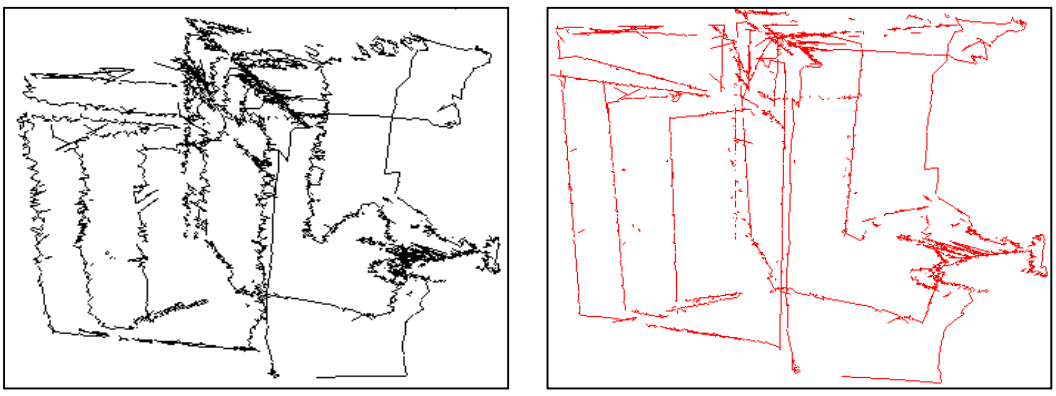
(a) (b)

Figure 4-9: Results of straight-line edge correction after one (a) and three iterations (b)

4.5 Linear correction results

The main test scene (laboratory at Joint Research Centre) is composed of a room with two doors and several windows. It was chosen because it is rich in discontinuities, making it an ideal scene for testing the improvement our technique can bring. In effect, the acquisition of reliable range data is particularly difficult in such a scene since range measurements are prone to errors in discontinuities. At the same time, the scene is rich in geometric contents that can be recovered from the photograph.

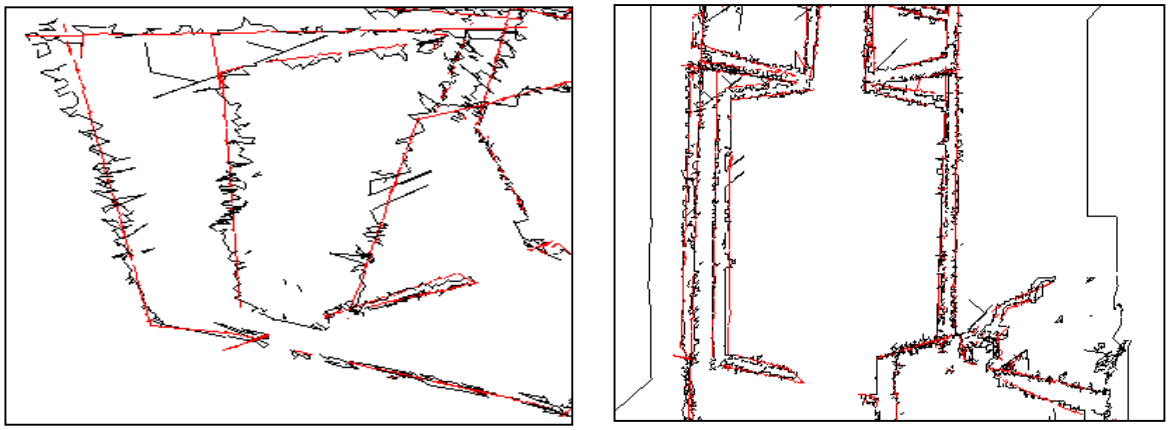
Figure 4-10 presents a snapshot of the original edges (a) and the results of the correction (b) after three iterations of the algorithm. During the whole process, 3821 points co-ordinates were updated, in 4.8s (Pentium IV, 2.0GHz).



(a) (b)

Figure 4-10: Original (a) and Corrected (b) edges

In Figure 4-11, the same snapshots present the original (black) and the corrected (red) edges from two different viewpoints. These snapshots show clearly how the mixed point effects are minimised.



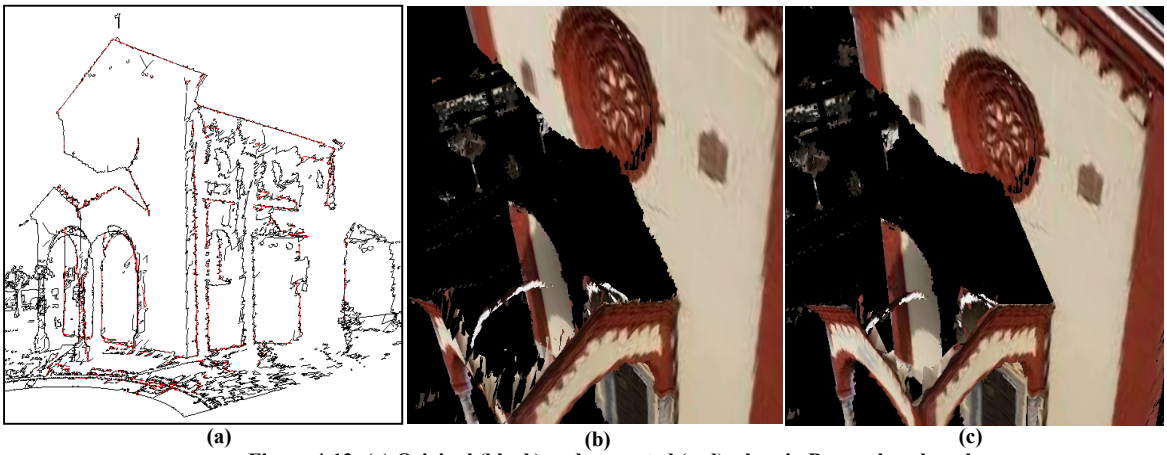
(a) (b)

Figure 4-11: Snapshots of the original (black) and corrected edges (red)

Finally, Figure 4-12(a) presents a snapshot of the textured laboratory model before any correction. The corrected model, after three iterations of the algorithm, appears in Figure 4-12(b).

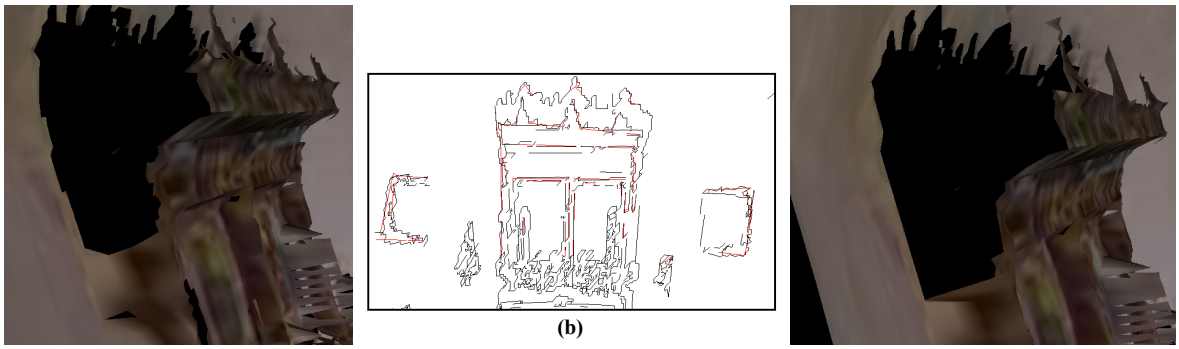
(a) (b)

Figure 4-12: Snapshot of the original (a) and the corrected (b) laboratory model



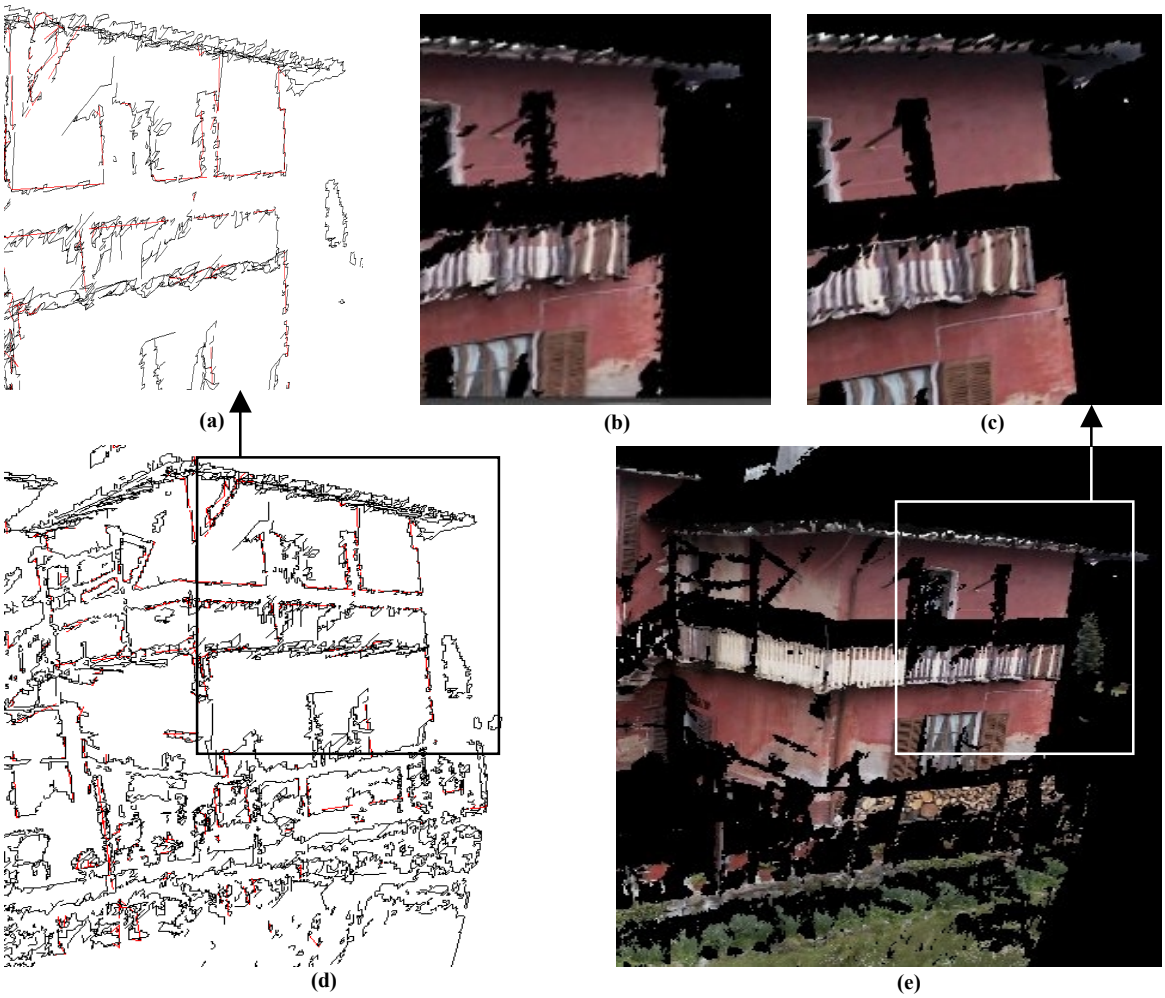
(a) (b) (c)

Figure 4-13: (a) Original (black) and corrected (red) edges in Barza church and snapshot of original (b) and corrected (c) model



(a) (b) (c)

Figure 4-14: (b) Original (black) and corrected (red) edges in Bornholm and snapshots of original (a) and corrected (c) model



(a) (b) (c) (d) (e)

Figure 4-15: Results of the algorithm with the Laveno farmhouse scene: (a)-Details of the edges (black: original edges, red: corrections); (b)-Details of the original textured model; (c)- Details of the corrected textured model; (d)-Global view of the edges; (e)-Global view of the textured corrected model

The algorithm (with the same parameters and number of iterations) was run in the other three test scenes (Figure 4-13, Figure 4-14 and Figure 4-15). The global results are presented in Table 4-1.The average correction error is the average distance between the original and corrected 3D points.

	Lab030	Barza	Bornholm	Laveno
Detected Jump edges	297	1066	566	1164
Detected roof Edges	437	1458	811	2244
Detected lines	159	139	51	59
Filtered lines	59	66	28	38
Number of points corrected	3821	3807	830	2529
Average correction (meters)	0.030	0.053	0.038	0.048

Table 4-1: Results of the straight-line correction in 4 test examples

4.6 Correction of quadratic edges

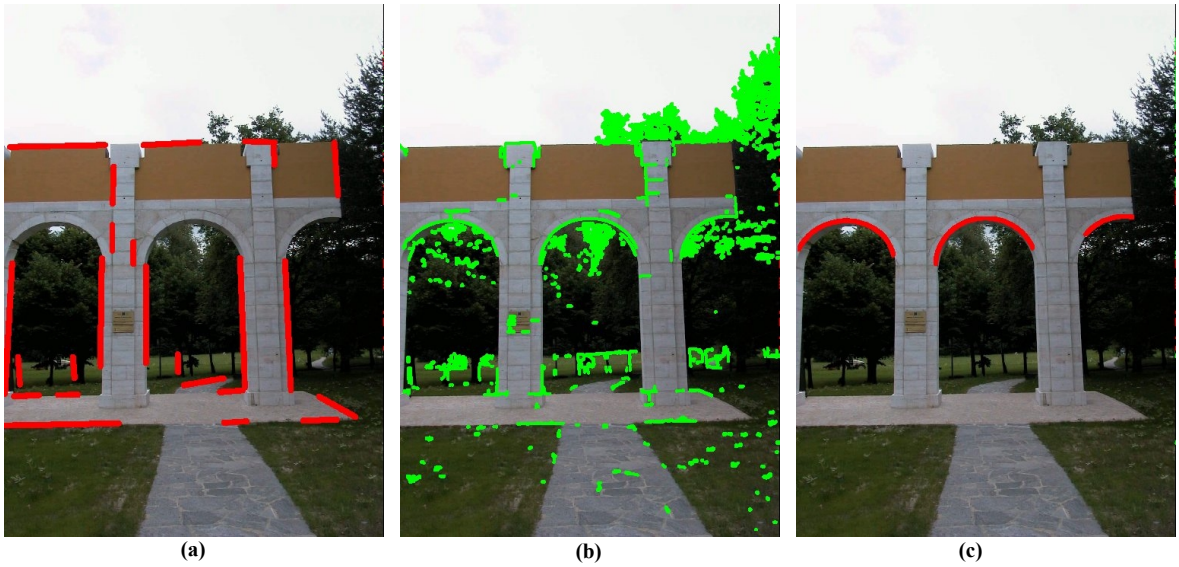
The previous sections concentrate in correcting 3D straight lines in the models using, as a guide, the 2D lines detected in a single registered intensity image. This technique has a reduced range of application, but it can be easily extended to more complex curves. We illustrate this possibility using quadratics. The goal is to demonstrate that the technique developed for straight lines can be easily adapted to any kind of curves with few modifications.

The procedure used to correct quadratics is very similar to the one used for straight lines. The first step is the detection of quadratics in the intensity image. The main difference comes from the fact that the RANSAC approach is used to characterize the 3D "plane of the quadratic" since quadratics curves are defined within a plane. At this point, the 3D points are projected into the "plane of the quadratic" and a RANSAC technique is used to find the quadratic in the plane with more support. The last step consists of re-projecting the 3D points selected as belonging to the quadratic into the curve in the plane.

To illustrate this part of the work, another test scene was selected, since the laboratory is rich in linear elements but does not present many quadratic curves. The selected model is the "arches of São Vicente de Fora" present at the JRC Campus [this building is a copy of the arches present at Lisbon and was used at the European Laboratory for Structural Assessment (ELSA - Joint Research Centre) for seismic experiments. After the tests, the building was moved into the JRC Campus and became part of the JRC Landscape].

4.6.1 Detection of quadratics in intensity images

The quadratic correction process will come naturally after straight line correction. For this reasons instead of applying the detection of quadratics to all the Canny edges, the edges that are in the neighbourhood of straight lines are removed ensuring that the remaining edges will be all of second or higher order. The filtering is done by dilating the original straight lines and removing all the edge pixels that are within the obtained dilated region. Figure 4-16 illustrates this process. Figure 4-16(a) presents the detected straight lines after the dilation (in the arch example, the dilation coefficient used was 8 pixels). The remaining edges after the filtering operation appear in Figure 4-16(b) and finally, Figure 4-16(c) presents the detected quadratics for the arch scene.



(a) (b) (c)

Figure 4-16: Line and Quadratic detection in Arch model: (a) dilated lines, (b) filtered edges, (c) detected quadratics

The detection of the 2nd order curve in the intensity image is done using a similar technique to the one used for line detection (section 4.3.1). A function is used to compute a 2nd order approximation using orthogonal regression and models the quadratic with the following equation [Eberly00]:

$$a_1 + a_2 x + a_3 y + a_4 xy + a_5 x^2 + a_6 y^2 = 0 \qquad (4.8)$$

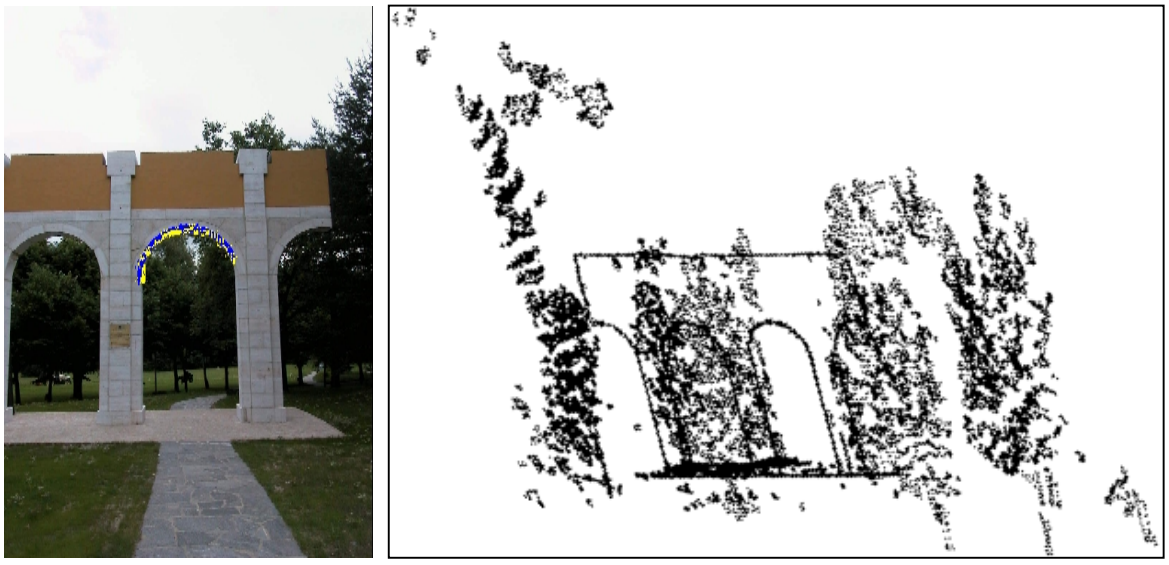
where a_i are the coefficients of the estimated quadratic.

For each of the edges, a RANSAC technique is used to find the quadratic with more support using a subset of 5 points. A threshold is used for the minimum number of points in a quadratic to avoid considering small edges and ensure that the detected quadratics have a reasonable size. In the Arch example, the detection of the quadratics (Figure 4-6(c)) was performed in 5s seconds in a Pentium IV at 2.0GHz.

4.6.2 Computation of the plane of the quadratic

In 3D, quadratic curves cannot be easily modelled with a simple parametric equation, as for straight lines. In this case the surfaces are paraboloids, hyperboloids, ellipsoids, etc. Second order edges (that we are interested in) are defined as 2D curves in 3D planes. At this stage, the 3D edge points re-projected in the neighbourhood of the 2D quadratics of the photograph are going to be used to define the plane with more support, that we call "quadratic plane" so that we can afterwards define the 2D quadratic inside this plane.

Figure 4-17(a) presents one of the detected quadratics and the re-projected edge points. The plane that best fits the re-projected points (in blue are the points that support the given plane) is then computed with a RANSAC technique and, in Figure 4-17(b), all the 3D edge points are re-projected into the "quadratic plane".



(a) (b)

Figure 4-17: The re-projected points (a) and the quadratic plane (b)

4.6.3 3D Points Correction

A problem that we already met before arises again at this point: how to fit a quadratic to a given set of points in a plane. The same technique used for detecting a quadratic in the intensity image is used to find the quadratic that best fits the 3D points using as initial guess, the points re-projected initially close to the quadratic.

The 3D points close to the quadratic in the plane are extracted and the edge ordering is used to avoid breaking connected edges when few points with higher errors are met. Figure 4-18 present some results for two of the detected quadratics. The quadratic that best fits the points appears in blue and the edge points that will be corrected are in red.

The last step of the process consists, for each selected 3D point, to modify its co-ordinates so that the new co-ordinate will be positioned in the quadratic and in the quadratic plane.

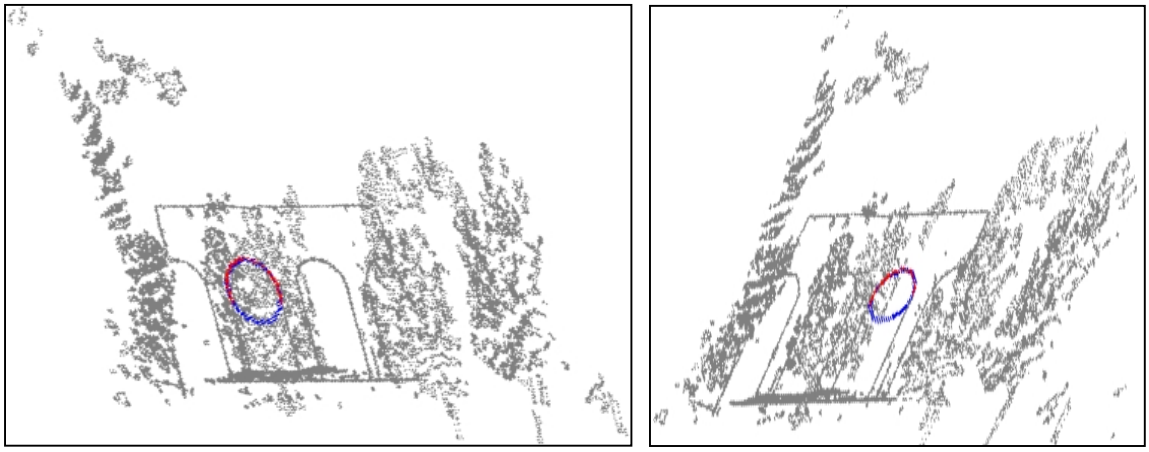
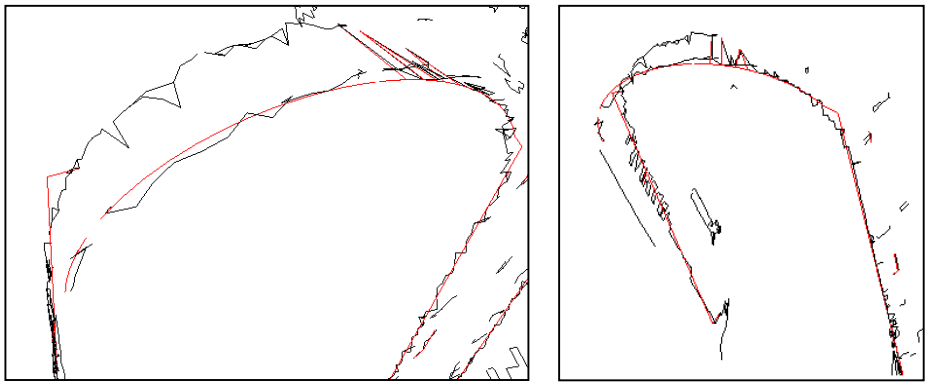


Figure 4-18: Quadratic plane (grey), selected quadratic (blue) and points to correct (red) for two detected curves

4.6.4 Experimental results

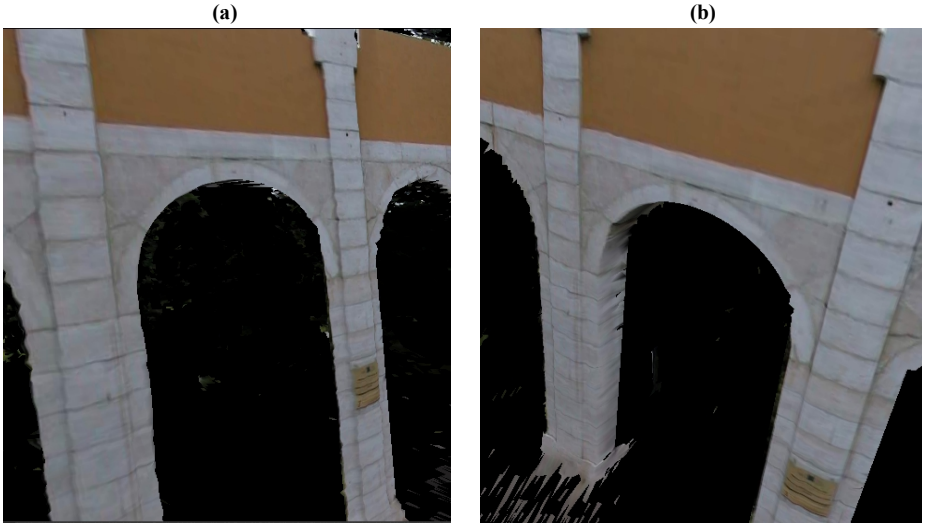
Figure 4-19 and Figure 4-20 present the result of the algorithm in the model of the arches. The three iterations of the quadratic correction algorithm took 42s s in a Pentium IV at 2.0GHz. The whole process (line and quadratic correction) resulted in the update of 1898 point-co-ordinates in the model. Table 4-2 present results of different steps of the process.



(a) (b)

Figure 4-19: The original (black) and Corrected (red) edges of the corrected curve

(a) (b)



(c) (d)

Figure 4-20: Original (a,b) and corrected (c,d) quadratic in the arches model

Detected Jump edges	1179
Detected roof Edges	2516
Detected lines	32
Filtered lines	24
Detected Quadratics	3
Number of points corrected	1898
Average correction (meters)	0.071

Table 4-2: Results of straight line and quadratic correction in Arch example

4.7 Z-buffering for Full model handling

The correction of 3D edges using a single registered intensity image was developed using triangulated models obtained from single range views, since, in these models, the jump edges are particularly numerous making it easier to evaluate and validate the quality of the algorithms.

Obviously, the algorithm also must be able to deal with full 3D triangulated models coming from the fusion of several range views. In this case, the main difference is the necessity to implement a Z-buffer to check which edges are visible from the camera position and ensure that occluded or non-visible edges will not be corrected.

It is then necessary to find the closest 3D points for each intensity image and consider only these points. A similar technique was already used with the range cloud of points in the texture mapping to avoid occlusions errors when re-projecting intensity images into the reflectance (see section 3.2.2).

In the Z-buffering process, the triangulated model is first moved into the camera co-ordinate system. Each triangle is then projected into the photograph (by projecting its 3 vertices) according to the camera model. The equations of the 3 lines defining the triangles are computed and the Z-buffer for all the points belonging to the triangle is set to the maximum Z value of the 3 vertices.

Figure 4-21 presents a photograph of the arches and a re-projected triangle ($\mathcal{V}_1,\mathcal{V}_2,\mathcal{V}_3$) from the model. For each of the triangle segments, the equation of the three borderlines is computed.

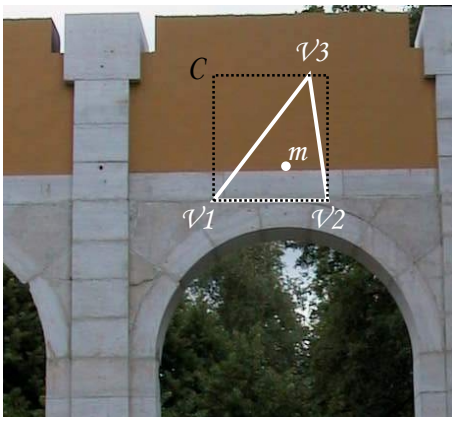
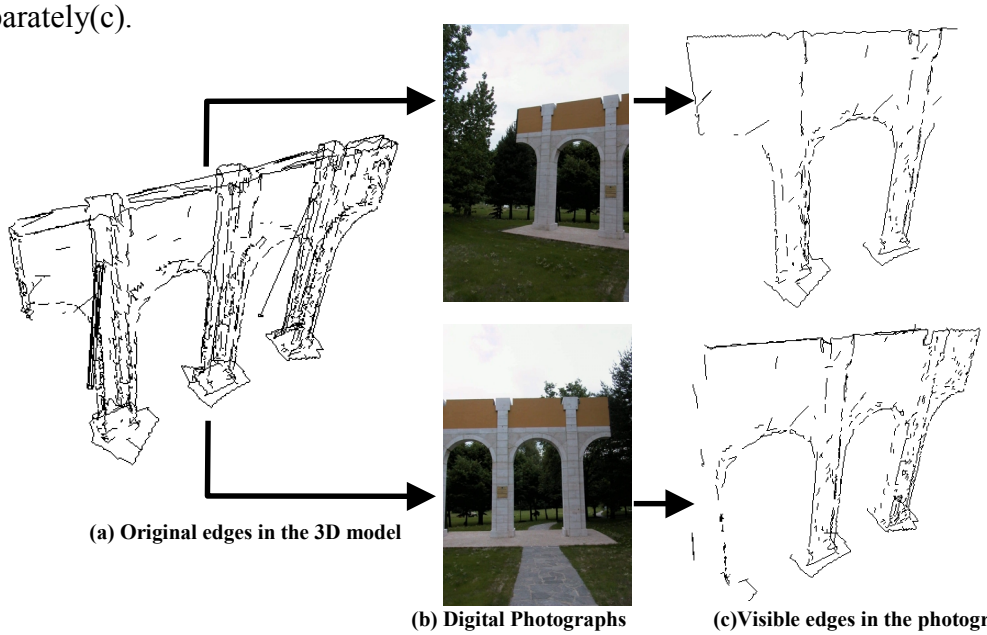


Figure 4-21: Z-Buffer filling based on the re-projected triangles

The sign of each one of the line equation is computed for the median point of the triangle: $sign_1(m)$, $sign_2(m)$, $sign_3(m)$, as well as the maximum Z value for the vertices $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{V}_3$: Z_{max}.

For each point P of the square C, we fill the Z-buffer of the pixels located inside the triangle (pixels for which the line equation signs are equal to $sign_1(m)$, $sign_2(m)$ and $sign_3(m)$) with the maximum value of the Z_buffer already computed.

The result of this operation is an array with for each pixel in the intensity image, the Z-value of the closest 3D point. This Z-buffer is used in the 3D jump and roof edge detection step (see section 4.2) to find out which edges are visible in the intensity image and which are not. Figure 4-22 presents the results for an image of the arch model: all the jump/roof edges detected in the model (a) and the sub-set of edges visible in the two images separately(c).



(a) Original edges in the 3D model **(b) Digital Photographs** **(c)Visible edges in the photograph**

Figure 4-22: Filtering of visible edges in the photographs using Z-buffering

Once the edges visible in each photograph are identified, the line and quadratic correction algorithms presented in the previous sections are used normally but considering only the edges visible in the current photograph. This allows the use of several photographs with a full-triangulated model.

4.8 Discussion

The method presented in this chapter is used to correct the main geometric curves in 3D models with minimal user interaction and with no need of additional data acquisition. The mixed point problem and other effects due to the limited resolution of the laser scanner are significantly minimized.

As it is now, the process can deal with straight lines and second order curves such as quadratics. But the process can be easily extended to other curves. In effect, both the straight line and quadratic correction process works in a similar way. The only difference is the curve fitting function, and the introduction of an additional plane-fitting step for quadratics. To summarize, the algorithm can be used with any kind of geometric 2D/3D edges as long as the following functions are provided for the given curve:

- Fitting of 2D points to the curve: this function has to be able, given a set of 2D points, to return the parameters of the curve that best fits the given points. This function is used when detecting the curve, either in the photograph or in the plane of the curve.
- Computing distance from a 2D point in the plane to the curve: used for selecting the curve that best fits the points based on a RANSAC technique. The best curve will be the one with more points within a given threshold.
- Computing the projection of a point into the curve.

For the fitting functions, we used the ones available in the Magic Software [Eberly00], these functions are time consuming and can be optimised. An improvement could be to use independent and optimised functions for each different quadratics: circles, ellipses, hyperboles, etc. In this case, the curve fitting speed and quality may be improved and the function could be applied successively to the images selecting, for each edge, the curve that best fits.

The thresholds for the quadratic detection are quite severe when the procedure is fully automatic to avoid wrong corrections. These thresholds can be relaxed, but then the user

would probably need to control the edge detection and guide it by indicating wrong fittings (for example, removing small segments or part of other curves)

The whole process is of particular interest in scenes with high geometric contents such as laboratories, buildings, and its performance is reduced in scenes with few geometric entities such as landscapes, trees. The procedure presented is also independent from the scanning devices and can be used with any range scanner or camera.

The technique is a starting point to get a description of the model by straight lines or other curves. A possible extension is to directly create CAD models of buildings from real data.

Another possible extension of the work presented here is to use a similar technique to do surface correction in 3D models. In this case, colour segmentation could be used to isolate areas of interest in the photographs. These areas would be afterward used to detect surfaces in the model (planes, spheres, quadratic surfaces…) and force the 3D points to lie in the exact surface.

5 Combining passive triangulation and range techniques for 3D modelling

In chapter 4, a single registered intensity image was used to improve the segmentation in range images. The method does not really extract 3D information from the intensity image: this one is just used as a guide to isolate and correct 3D features present in range data.

A natural extension of the correction based on a single registered image consists in using two or more registered images to improve the geometry and segmentation of 3D models. In this case, 3D information can be triangulated directly from the already registered intensity images providing a second source of 3D information.

In the following sections, we will present the passive triangulation method implemented to extract 3D information from two or more images. This information will be used for several purposes (see Figure 5-1):

- Calibration tuning: to improve the camera model by comparing passive triangulated points with range points;
- Range segmentation: linear and quadratic features are extracted in range models using intensity data as a guide. This is an extension of the work presented in chapter 4, but using more than a single registered image;
- Dense mapping: used to introduce additional information/points into the original range cloud of points by fusing the information of laser scanners and digital cameras into a single cloud of points.

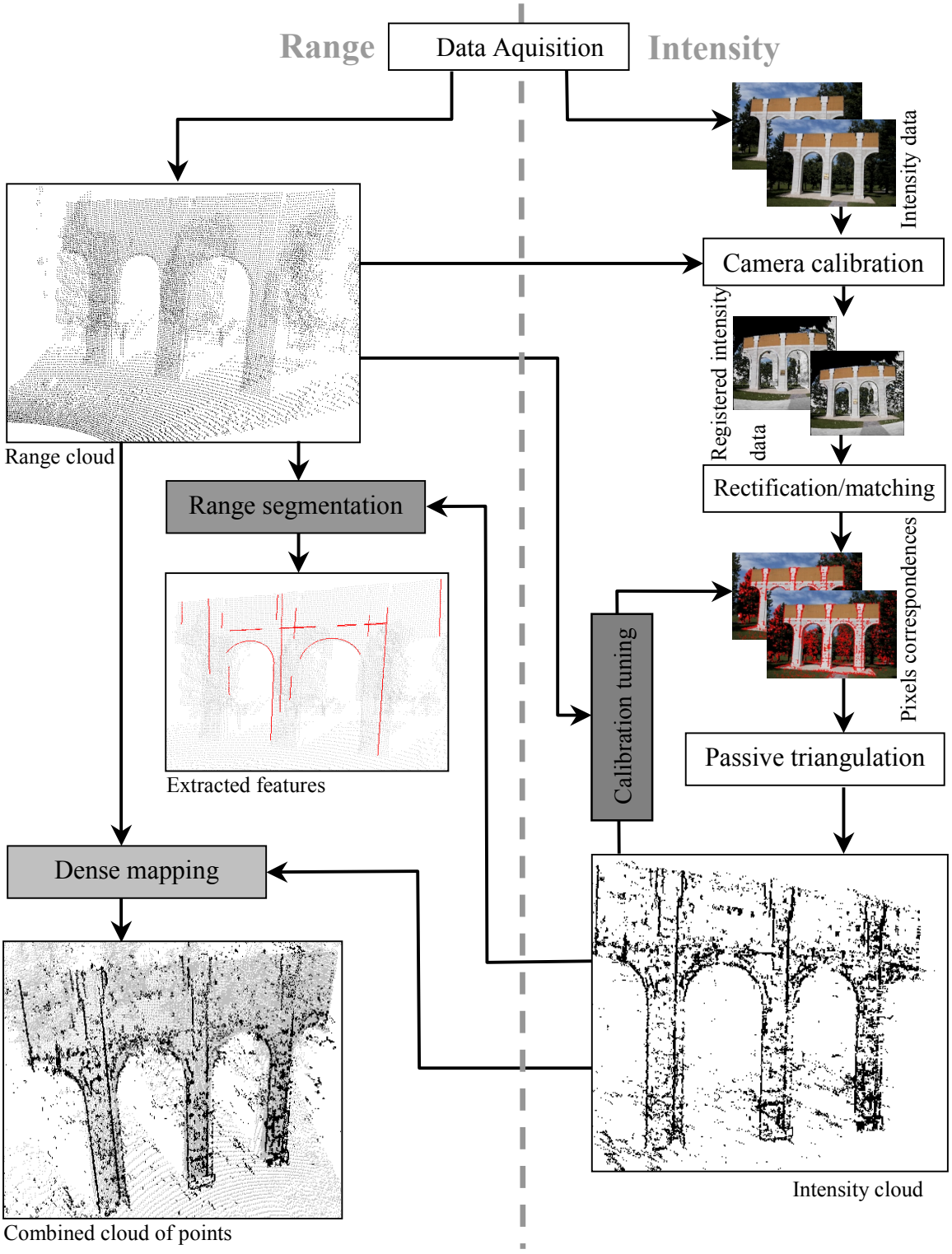


Figure 5-1: The passive triangulation process and its applications

5.1 Matching intensity images

Depth extraction from intensity images is not a new topic of investigation. Many techniques such as stereoscopic vision or photogrammetry already deal extensively with this problem. The algorithms used by our system to compute correspondences over intensity images are based in typical tools such as epipolar constraints, fundamental matrix, image rectification and cross correlation. All these techniques are already well known and well documented [Faugeras93, Pollefeys00, Hartley00].

5.1.1 Fundamental matrix estimation

A powerful tool to help computing correspondences over intensity images taken from close viewpoints is the fundamental matrix. This matrix links two intensity images through epipolar geometry, according to the following equation:

$$X'^{T}_{i} F X_{i} = 0 \qquad (5.1)$$

where X $(x,y,1)$ is a homogeneous co-ordinate in the first image, $X'(x',y',1)$ is the corresponding co-ordinate in the second image and F is the 3×3 Fundamental matrix that links the two images through epipolar geometry.

Figure 5-2 illustrates the epipolar geometry. Given a 3D point P, the projection of P into the image planes I_1 is m_1 and into I_2 is m_2. The line defined by the two centres of projection C_1 and C_2 of the cameras intersects the image planes at the epipoles e_1 and e_2. The epipolar lines ep_1 and ep_2 are the lines defined by the epipoles (e_1 and e_2) and the projections of P in the image planes (m_1 and m_2). Once the fundamental matrix is known, the epipolar line ep_2 corresponding to a point m_1 in image 2 can be computed and the matching problem reduced from two to one dimension.

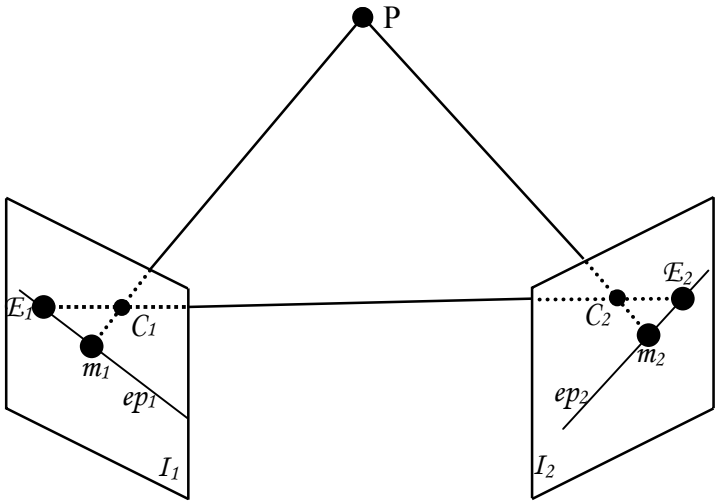


Figure 5-2: The epipolar geometry

Many robust methods have been developed to compute the fundamental matrix [Luong94, Torr97, Salvi01]. They all need at least 7 corresponding pixels over the images to perform well since the fundamental matrix can be determined up to a scaling factor and its determinant is null [Torr97].

The matching algorithm developed selects points of interest in the first image. We used the corner detector implemented in the Open Source Computer Vision Library [OpenCV] based on the algorithm from Haralick and Shapiro [Haralick93]. Sobel first derivative operators are used to compute the derivatives in x and y direction in the image (D_x and D_y). A 2x2 matrix of the sums of the derivatives is then created within a region of interest:

$$C = \begin{bmatrix} \sum D_x^2 & \sum D_x D_y \\ \sum D_x D_y & \sum D_y^2 \end{bmatrix} \tag{5.2}$$

The eigenvalues of C are found to select the pixels with high probability to be corners. Non-maxima suppression is also applied and, finally, the implementation removes corners that are too close to each other to avoid accumulation in reduced areas.

Compared with traditional stereo techniques, it is possible to use additional information coming from the existing calibrations (computed with the semi automatic calibration system presented in chapters 2 and 3). This extra information guides the matching process: a point of interest is selected in the first image, then the algorithm searches for the range point that is re-projected closest to this point according to the camera model. The 3D point is finally re-projected in the second image, being the centre of the search area for the

template matching (see Figure 5-3). The last step consists of using the displaced frame difference to compute the corresponding pixels within the search window in a similar manner to the one presented in section 2.4.2.

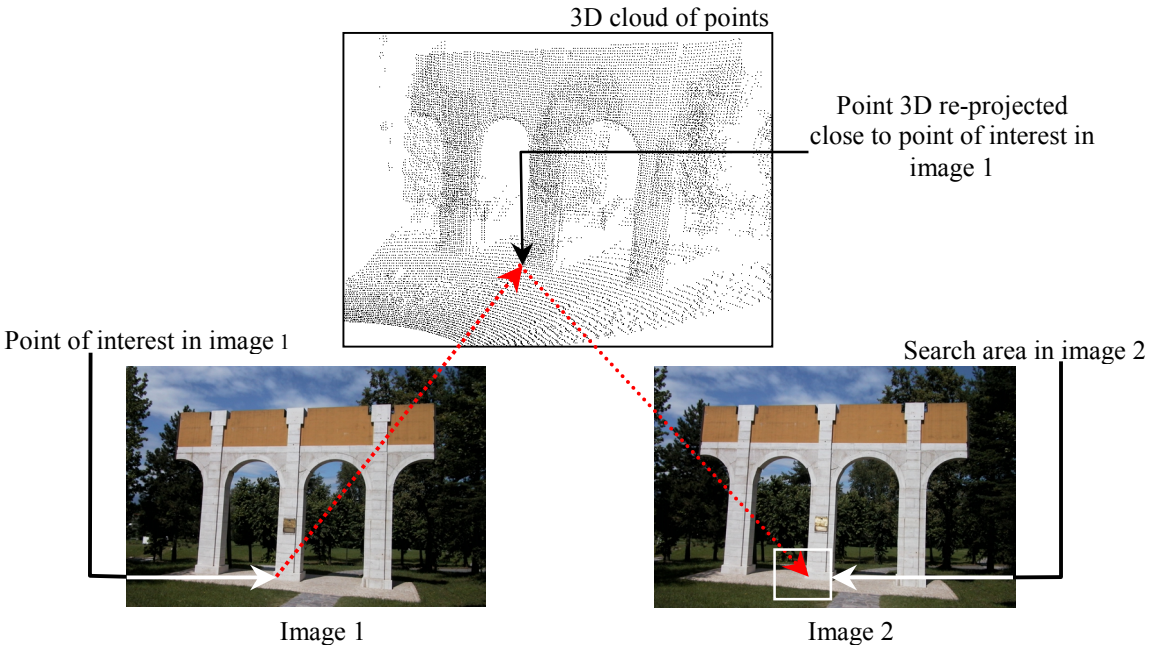
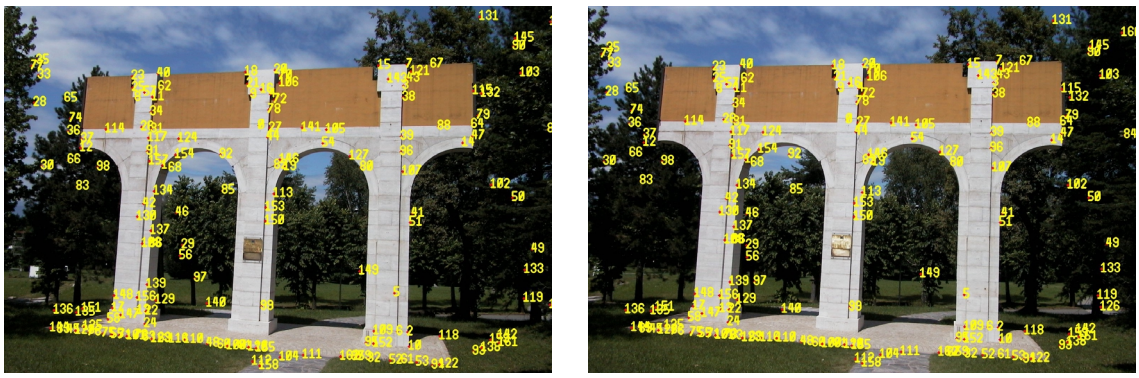


Figure 5-3: Point matching between the intensity images guided by the existing calibrations

Figure 5-4 presents the correspondences computed for a pair of images of the arch model. 163 correspondences were found in 3.0 seconds on a Pentium IV at 2.0GHz.



(a) **(b)**

Figure 5-4: Results of the matching for fundamental matrix estimation

The fundamental matrix is usually estimated using the seven or eight points algorithm. Both methods require the resolution of a system of equations built with the available correspondences. In our case, given the large number of corresponding points, the eight-

point algorithm was chosen for its simplicity and possibility to use robust techniques when more than eight correspondences are avalaible.

The eight points algorithm uses the following equation, derived from (5.1):

$$[xx' \quad yx' \quad x' \quad xy' \quad yy' \quad y' \quad x \quad y \quad 1]F = 0 \qquad (5.3)$$

With,

$X\,(x,y,1)$ and $X'\,(x',y',1)$ two corresponding pixels in the images;

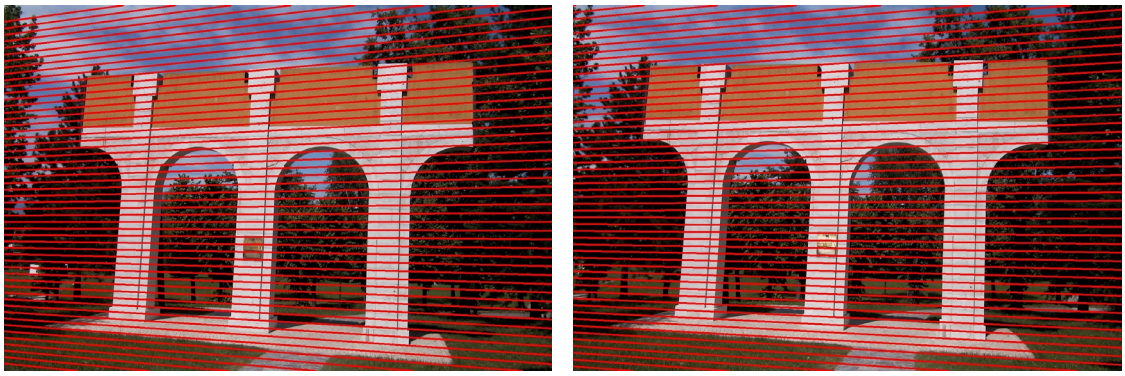
$F\,[f_{11}\ f_{12}\ f_{13}\ f_{21}\ f_{22}\ f_{23}\ f_{31}\ f_{32}\ f_{33}]^T$ a vector composed by the elements of the 3x3 fundamental matrix.

Given eight corresponding pixels, equation (5.3) can be organized into a system of equations that can be solved to extract the elements of the fundamental matrix *F*.

The estimation of the fundamental matrix is an ill-posed problem: small errors in the matching pairs can lead to large errors in the matrix estimation. Given the large number of correspondences available (in Figure 5-4, more than 150 pairs were extracted), it seems logical to apply robust estimation techniques. A RANSAC process similar to the one used for the Tsai camera calibration (see section 3.1.2) is used to increase the reliability of fundamental matrix estimation process.

5.1.2 Rectification

From equation (5.1), it is possible to compute corresponding epipolar lines for a given fundamental matrix *F*. Figure 5-5 presents some of these corresponding epipolar lines computed with the fundamental matrix estimated for the arch images.



(a) (b)

Figure 5-5: Epipolar lines in two intensity images of the arches

The fundamental matrix is used to compute the rectified images (see Figure 5-6). In these images corresponding epipolar lines are forced into the same row. The matching problem is significantly simplified since it is possible to seek for corresponding points only along the epipolar lines and reduce the search space from two to one dimension. The implementation for image rectification used is the one proposed by Intel in its Open Source Computer Vision Library [OpenCV]. More details about image rectification can also be found in [Pollefeys00, Hartley99].

(a) (b)

Figure 5-6: The two rectified images of the arch

5.1.3 Matching

Cross correlation is used with the rectified images to find corresponding points along the epipolar lines. Once more, the seed for the search space comes from the existing calibrations (see Figure 5-3). The points of interest selected in the first image are a sub-sampling of the edges detected using a Canny edge detector.

An additional filtering is used to ensure that only points closed to range discontinuities are considered. Discontinuities in range are detected by analysing each 3D range point neighbourhood. 3D Points where high variations in the depth occur are marked as discontinuities [Sequeira96]. Within this filtering process, all discontinuities of the 3D cloud are re-projected into the images and only correspondences that are spatially closer than a given distance to the discontinuities in both images are considered.

Sub-pixel accuracy is obtained by fitting a second-degree curve to the cross correlation coefficients in the neighbourhood of the disparity. The maximum of the function occurs where the gradient of the function is null. We use the cross correlation values at the points

d-1, *d* and *d*+1 to compute the position of the maximum of the function as follows [Sun02, Anadan89]:

$$x = d + \frac{1}{2} \times \frac{C(d-1) - C(d+1)}{C(d-1) - 2C(d) + C(d+1)} \tag{5.4}$$

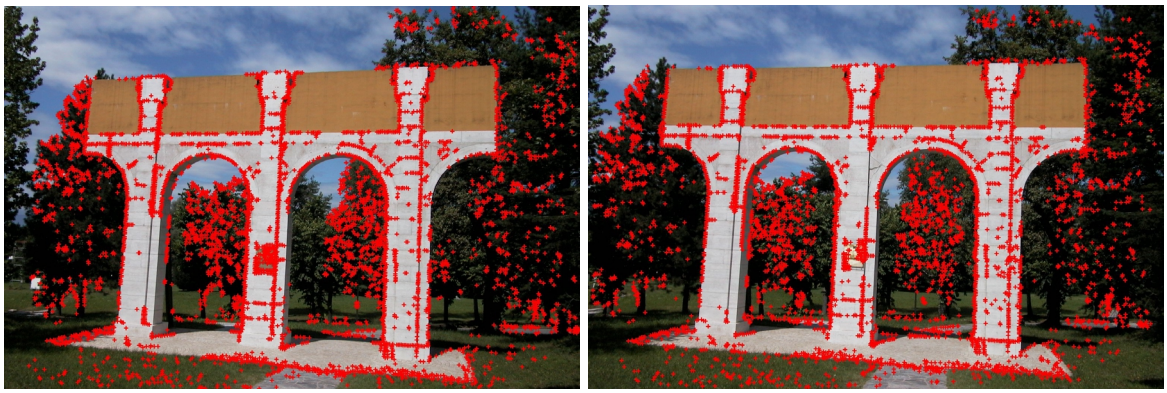
Where $C(d)$ is the cross correlation coefficient at position *d*.

x is the sub-pixel position of the maximum.

Two additional conditions are considered when seeking for correspondences over the intensity images. These conditions are used in most of dynamic matching methods [Cox96, Birchfield99]:

- A threshold is considered for the cross correlation to avoid considering points for which the correlation is too low.
- The matching must be symmetric. If a point p_1 has the highest cross correlation with point p_2, then p_2 must also have its maximum cross correlation with p_1. A maximum disparity between the two symmetric matches is considered.

The results of the matching process using the rectified images, and after the discontinuity filtering is presented in Figure 5-7. The whole process (points of interest detection, filtering and matching) detected 10133 matches in 33.6 Seconds in a Pentium IV at 2GHz. Many of these matchings will be reject during the calibration tuning process since they are near discontinuities in the range images



(a) **(b)**

Figure 5-7: Results of the matching process

5.2 Tuning of the calibration using passive triangulation

This section presents a method developed to improve the initial camera calibrations by comparing 3D points triangulated from passive intensity images with range data.

5.2.1 Triangulation of the 2D correspondences

The correspondences over the intensity images are used with the initial camera calibrations to triangulate 2D pixels in images into 3D points in the space. From the corresponding points in the images, the camera projective rays are computed by inverting the perspective projection equations from the Tsai model and obtaining the parametric equation (the parameters are *s* and *t*) of the 2 rays:

$$L_0(s) = B_0 + sM_0 \tag{5.5}$$

$$L_1(t) = B_1 + tM_1 \tag{5.6}$$

where B_0, B_1 are the origins and M_0, M_1 are the direction vectors of the rays 1 and 2 in 3 dimensions.

The squared-distance between the rays is then [Eberly00]:

$$Q(s,t) = \left|L_0(s) - L_1(t)\right|^2 \tag{5.7}$$

$$Q(s,t) = as^2 + 2bst + ct^2 + 2ds + 2et + f \tag{5.8}$$

With *a*, *b*, *c*, *d*, *e* and *f* defined as ($\bullet$ is the scalar product):

$$a = M_0 \bullet M_0, b = -M_0 \bullet M_1, c = M_1 \bullet M_1$$

$$d = M_0 \bullet (B_0 - B_1), e = -M_1 \bullet (B_0 - B_1)$$

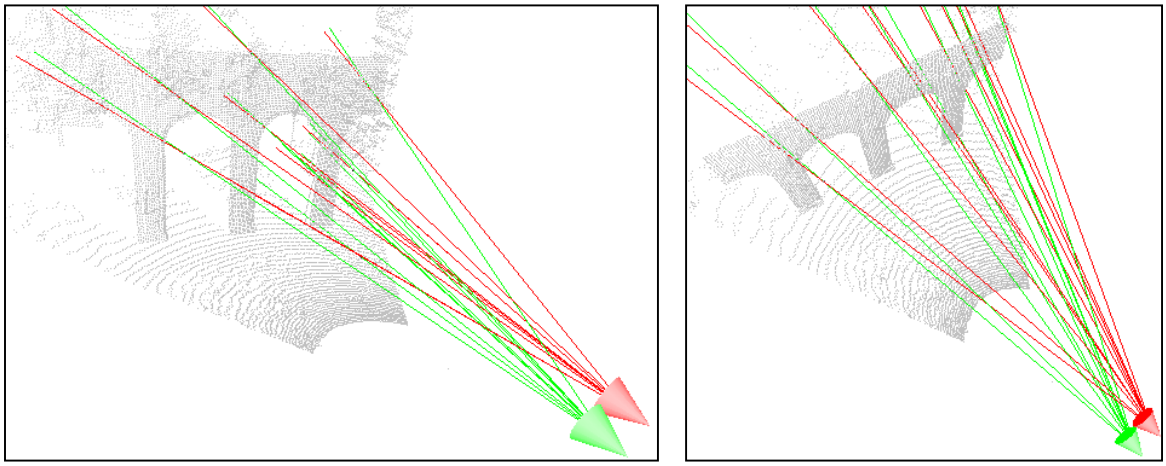
$$f = (B_0 - B_1) \bullet (B_0 - B_1)$$

The minimum distance occurs at the point where the gradient of the function *Q* is null

$$\nabla Q = 2(as + bt + d, bs + ct + e) = (0,0) \tag{5.9}$$

A user threshold defines the maximum distance between two rays above which triangulated points are not considered (in most of the examples, a 5cm threshold was used). Otherwise, the triangulated point is computed as the centre of the segment between the two closest points in the rays.

Figure 5-8 illustrates the triangulation process by presenting two snapshots of the range cloud of points, the position of the camera for the two images considered (represented as two cones) and the projective rays for some points.

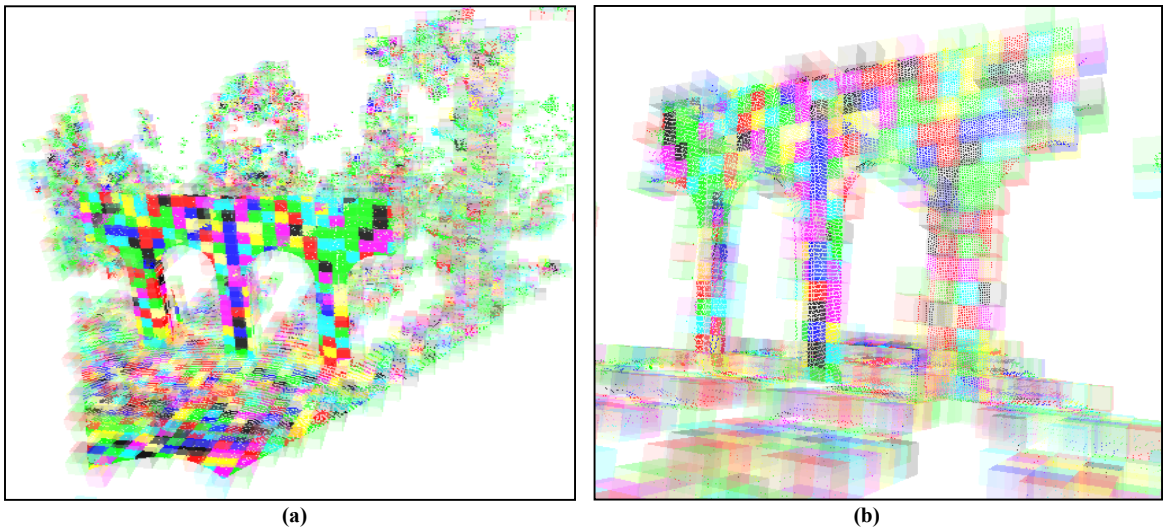


(a) (b)

Figure 5-8: The passive triangulation of 2D pixels into 3D co-ordinates. Range data appear in light grey

5.2.2 Association between triangulated and range 3D points

The process presented in 5.2.1 results, for each matching pair, in the 3D co-ordinates of the triangulated point. It is possible to use this information to evaluate the quality of the camera calibration by measuring the distance from triangulated points to the 3D cloud acquired with the laser scanner. To optimise this computation, a bucketing algorithm sub-divides the whole range cloud of points into indexed cubes to allow fast navigation and fast computation of the closest point inside the range data (Figure 5-9).

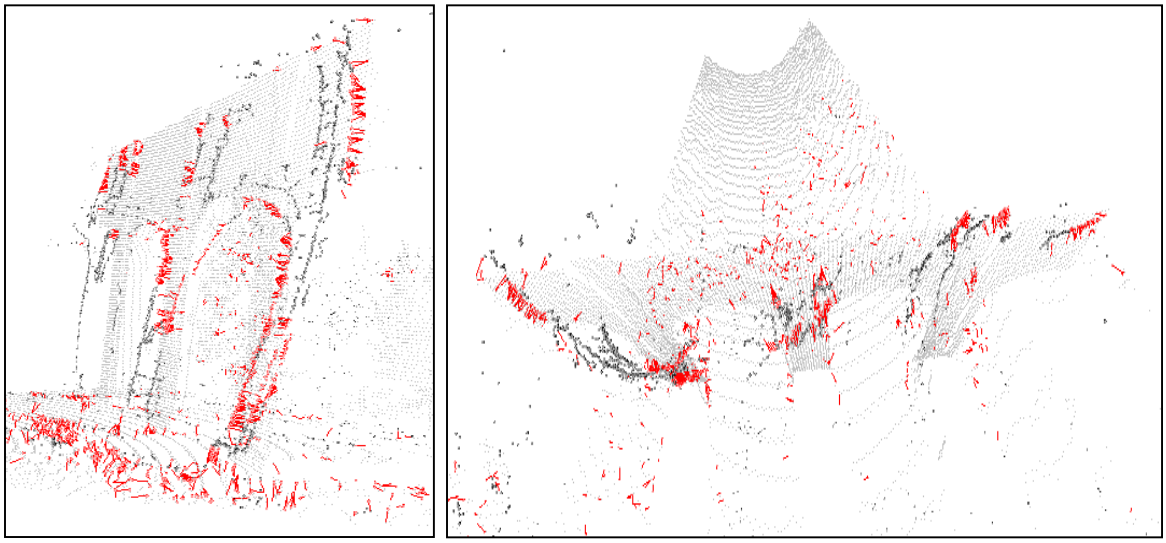


(a) (b)

Figure 5-9: The bucketing of the range data into indexed cubes

At this stage only the range points in the neighbourhood of discontinuities are used, since these are the areas rich in geometrical content and thus valuable for matching. This makes the 3D edge detection a critical step in the process.

A maximum distance, between triangulated and range points, is considered during the process to reject points for which the matching is probably wrong. Two snapshots illustrate the results of the triangulation for the arch example in Figure 5-10. The figure presents the range cloud of points (in light-grey); in red the points considered (for which the distance to the closest range discontinuity is inferior to the threshold (0.3m in this example)) and in black triangulated points that are not considered.



(a) (b)

Figure 5-10: Results of the triangulation using the original camera parameters

5.2.3 Iterative calibration process

To improve the calibration and force triangulated points to converge into the 3D range cloud, the closest discontinuity point in the range data is used with the pixels in intensity images to perform new Tsai camera calibrations.

This process is iterative: the new camera parameters are used to triangulate the 3D points in each iteration and find new 3D triangulated/3D Range correspondences.

This process continues as long as the average distance between triangulated and range points decreases. At this point, an additional optimisation is introduced in the loop to correct matching errors in the intensity images. In these new cycles of the process, pixel positions of the correspondences are updated using the current camera model. The closest 3D range points (used to calibrate the cameras) are re-projected into the images, and the

new pixel co-ordinates in every intensity image are computed as the median between the re-projected point and the original position of the matching.

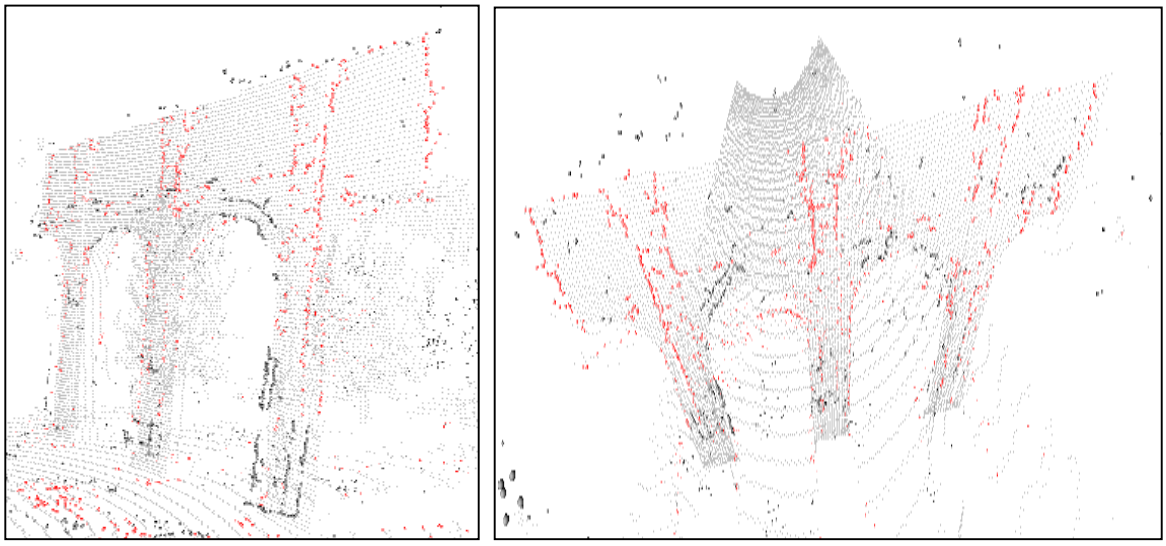
The process was run with two images of the arches. 15 iterations were performed in 2 minutes and 13.2 seconds in a Pentium IV at 2.0GHz. Figure 5-11 and Figure 5-12 present the re-projection images using the initial (a) and the final (b) camera parameters. Figure 5-13 presents two snapshots of the range cloud of points and the final triangulated points using the camera parameters estimated during the process. Finally, a graph with the evolution of the error (average distance, in metres, between triangulated and range points) appears in Figure 5-14. Even if the re-projected images look similar, the 3D error distance between triangulated and range points is significantly reduced.

(a) **(b)**

Figure 5-11: Initial and final re-projection of first image

(a) **(b)**

Figure 5-12: Initial and final re-projection of second image



(a) (b)

Figure 5-13: Results of the triangulation using the final camera parameters

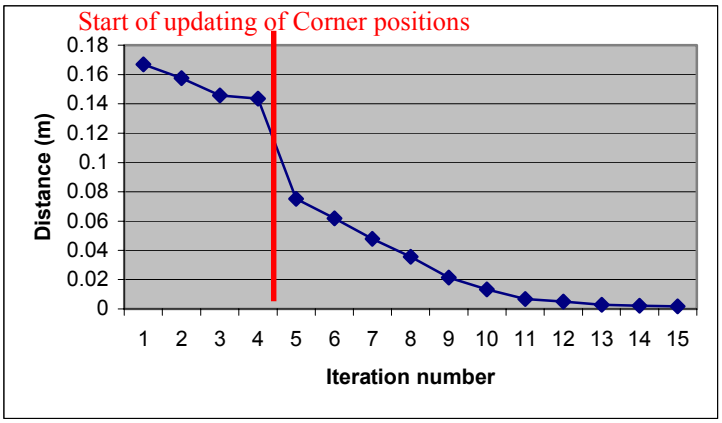


Figure 5-14: The evolution of the average distance between the triangulated and range points along the process

5.2.4 Passive triangulation with more than two intensity images

The technique presented in the previous sections uses only two intensity images to triangulate 3D points. If more than two images are available, the process can be easily extended to use all the available images in the calibration-tuning loop.

The matching process (section 5.1) is used between adjacent images, successively. Points of interest are extracted in the first image, and then correspondences are computed with the second image. The correspondences found in the second image are used as points of interest to find new correspondences in the third image and the process can be repeated as long as other intensity images are available. At the end, we obtain pixel correspondences over the entire set of images.

The main difference when more than two intensity images are used appears in the triangulation step. In this case, the triangulation (section 5.2.1) is done with all possible image pairs and leads to a different triangulated point for each pair of images. The final position is chosen as the centre of gravity between all triangulated points coming from each image pair. The process is illustrated, with three images, in Figure 5-15.

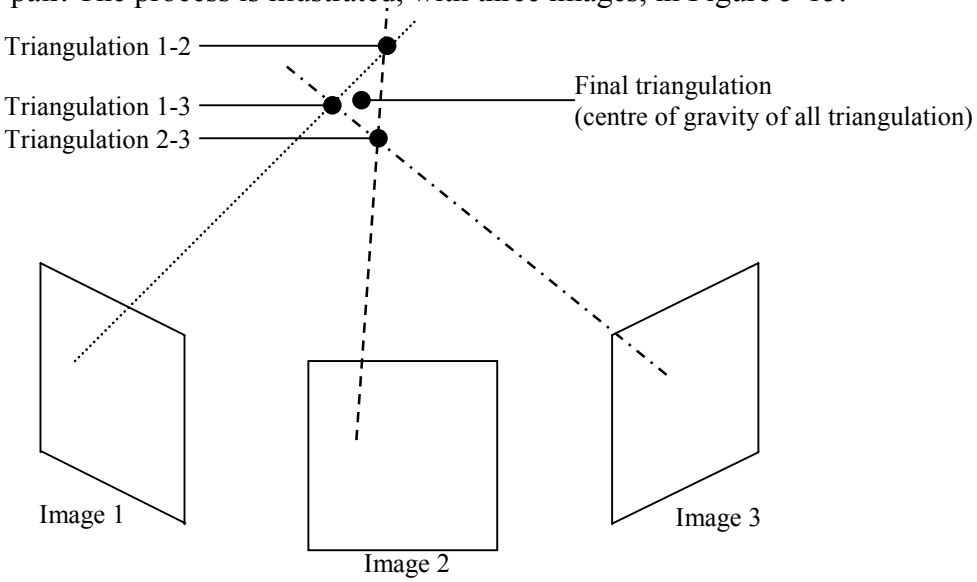
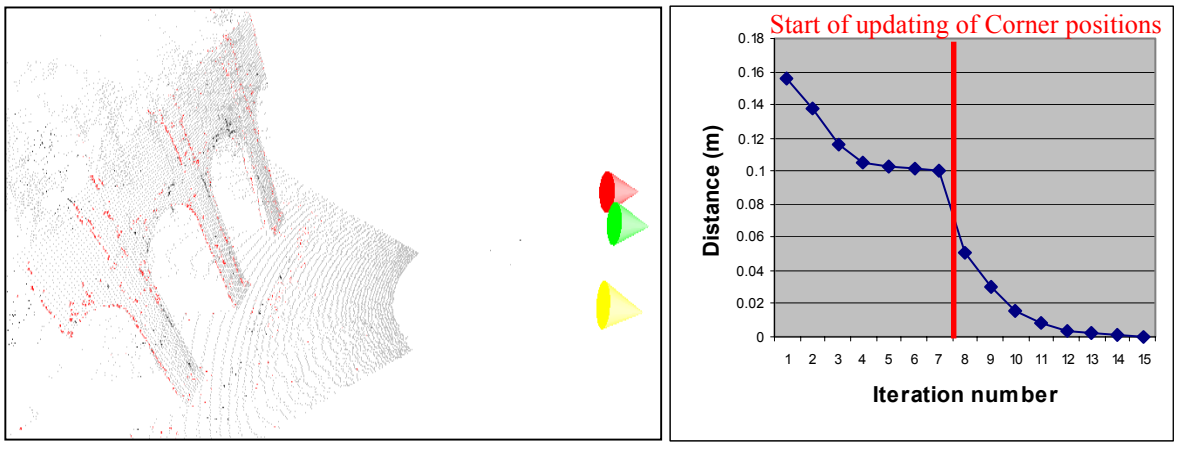


Figure 5-15: Passive triangulation with multiple images

The rest of the tuning process is exactly the same, as presented in sections 5.2.2 and 5.2.3. Figure 5-16(a) presents a snapshot of the cloud of points (light-grey) and the triangulated points after the tuning process (red) as well as the position of the three cameras (the three cones). Figure 5-16(b) illustrates the evolution of the average distance between triangulated and range points. 15 iterations were performed in 3 minutes and 8.2 seconds in a Pentium IV at 2.0GHz.



(a) **(b)**

Figure 5-16: Results of the tuning process for the arch model with passive triangulation of three intensity images

5.3 Improving range segmentation

The following sections present a method that uses calibrated photographs to improve the segmentation of range data by isolating and correcting 3D lines or quadratics curves.

The method is an extension of the process presented in chapter 4 and, thus, a major part of the procedure is similar to the one used in the correction based on a single registered image. In particular, most of the feature detection methods in 2D images are based on the same principles. An interface was also developed to give the user the possibility to correct and add additional 2D features in the 2D detection process.

The method detects features in one image and computes correspondences with other intensity images (using rectification and cross correlation, section 5.1). Then the 3D feature points are triangulated. Finally, the triangulated points are used to isolate the discontinuities in the 3D model and extract corresponding 3D features in range data.

In this situation, we decided to deal directly with the range cloud of points and not with triangulated models as in chapter 4. This choice is justified because 3D information is really extracted from the intensity images (whereas in chapter 4, the photograph is used only as a guide through the camera projective model). Given the possibility of getting 3D information from photographs, it is possible to handle more complex cases and work directly with the range data without any additional processing.

5.3.1 Segmentation in Intensity images

This section will present the methods used to detect 2D features in the registered photographs. Most of these tools have already been used either in the reflectance/intensity matching (chapter 2) or in the correction of range segmentation based on a single intensity image (chapter 4). We will describe briefly the processes pointing out only the main differences and improvements brought.

The line detection algorithm used is similar to the one presented in section 4.3. Since lines are not going to be used as a guide as in chapter 4, but are directly triangulated, it is important to detect the correct extremities of the segments. This is why the filtering operation (see section 4.3.2) is replaced by a grouping algorithm developed to group lines with same orientation and spatially close to each other.

The procedure, illustrated in Figure 5-17, computes between all pairs of segments, the minimum distance between the extremities of the segments (MinDist); and the maximal orthogonal distance between the segments (Dilation). If these two values are lower than the thresholds, the lines are grouped together by selecting the more distant extremities. At the end a new filter is applied to ensure that only long lines are considered.

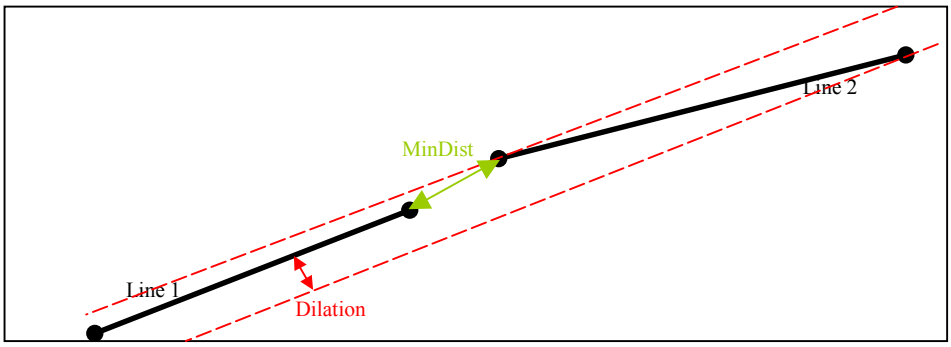


Figure 5-17: The line grouping process

In Figure 5-18, the result of the line detection in the arch example is presented. 20 lines were selected in 1.9 Seconds.

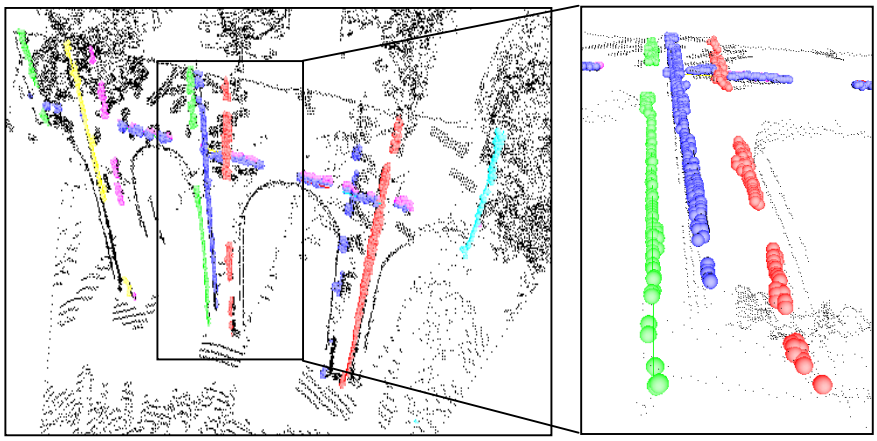
Figure 5-18: Results of the line detection process

For 2D quadratic detection, the method used is the same already presented in section 4.6.1.

5.3.2 3D Line extraction

The 2D features detected in the intensity images will be used to isolate the range discontinuities corresponding to the 2D curves. The main advantage of this technique compared with the method proposed in chapter 4 (using a single image) comes from the fact that the intensity image already gives the approximate 3D location of the features of interest, whereas in the single image method, only the projection of the points is used and no actual depth comes from the intensity images. The situation does not involve any Z-buffer problems since the 3D location is already given from the intensity images and there is no risk of correcting occluded areas.

The 2D features will be passively triangulated pixel-by-pixel using the same rectification/matching method presented in the calibration-tuning step (see 5.1 and 5.2). Figure 5-19 presents two snapshots of the triangulated pixels associated to the lines detected and matched in the intensity images of the arches. In light grey are the range discontinuities and the small coloured spheres correspond to the several triangulated line-points.



(a) (b)

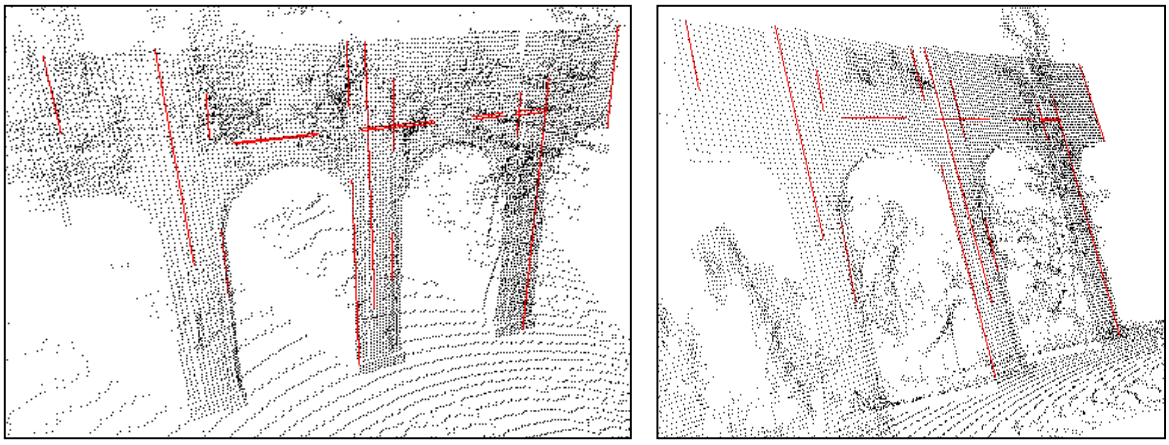
Figure 5-19: Two snapshots of the lines pixels after passive triangulation

The 3D line correction will follow by selecting the 3D range discontinuities located in the neighbourhood (Euclidean distance inferior to a given threshold) of the triangulated points (small spheres in Figure 5-19). The bucketing method for computing distances between triangulated and range points is the same already presented in section 5.2.2.

The last step of the process consists in extracting the line that best fits the selected range discontinuities using an orthogonal fitting (see 4.4.3.1).

The distance between adjacent range points is also considered when selecting the line, in order to select groups of points without any large gap, before associating them to a segment.

The 3D line correction process corrected 20 Lines in the arch model in 11.4 Seconds (Pentium IV at 2.0GHz). Figure 5-20 presents two snapshots with the original range data, and the lines extracted using our process.



(a) (b)

Figure 5-20: The triangulated points and the final lines computed

5.3.3 3D Quadratics extraction

The detection of quadratics in 3D is an extension of the line detection algorithm. It combines techniques used for 3D line extraction (section 5.3.2) and for the correction of quadratics based on a single registered intensity image (section 4.6).

The process starts by triangulating the pixels of the 2D detected quadratics. Each pixel is matched over the two intensity images using rectification and cross correlation matching (see sections 5.1.2 and 5.1.3). The triangulated points are presented in Figure 5-21(a). The small spheres correspond to the triangulated point (a colour is used for each quadratic); in light grey are the discontinuities in the range image.

The neighbourhood of the triangulated points is considered, and range discontinuities close to these points are selected, the Quadratic plane is then computed from these range points. Finally, the range points are projected into the Quadratic plane and an orthogonal fit function is used to select the quadratic with more support. All the process is similar to the one used in section 4.6. The two-extracted quadratics and the range cloud of points are presented in Figure 5-21(b-c), the process took 2.5 seconds in a Pentium IV at 2.0GHz.

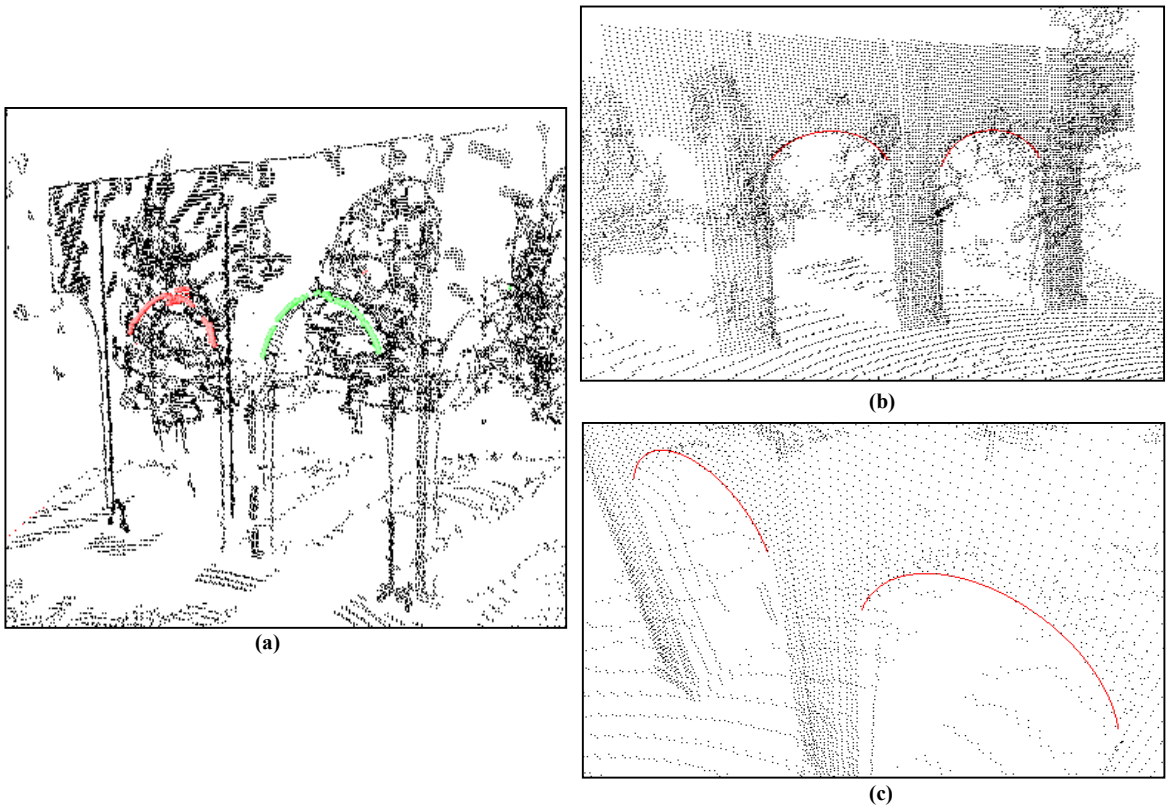


Figure 5-21: The triangulated points (a) and final quadratics computed (c,d)

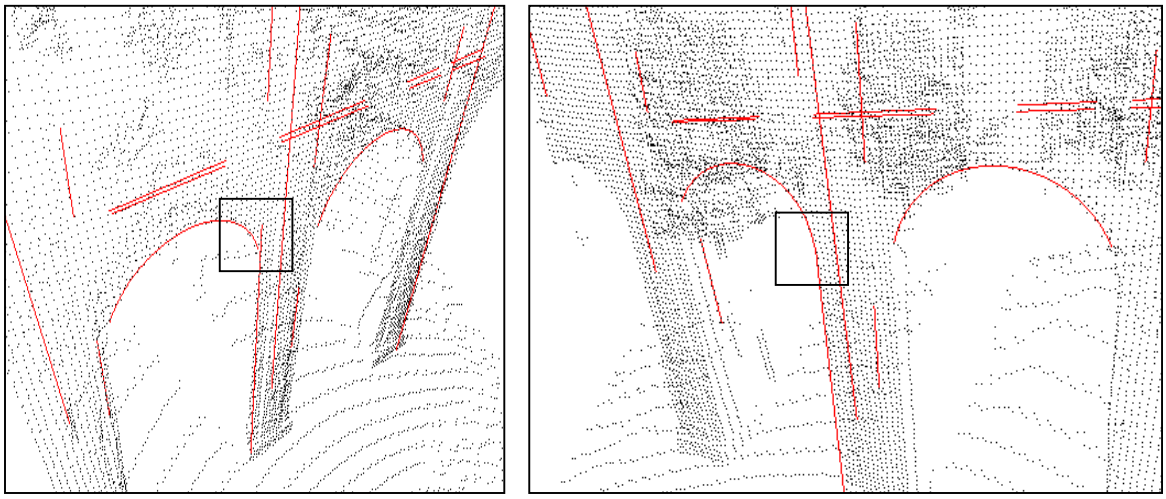
5.3.4 Connection between lines and quadratics

An algorithm was also developed to connect curves with neighbouring extremities. The algorithm searches for curves, with extremities that are spatially close to each other, and perform some correction between both edges, to connect them, and avoid small gaps between neighbouring features.

The algorithm starts by isolating the extremities of all the features (extremities of the lines and extremities of the longest arch for the quadratics). It will then compute the Euclidean distance between all the quadratic and line extremities to isolate neighbouring edges. Once a neighbouring edge (with two extremities closer than a given distance) is found, the algorithm will perform the following actions:

- Compute a new Quadratic plane using the quadratic and line pixels for the orthogonal regression;
- Project all the 3D points of the edges into the new plane;
- Compute the 3D point in the line closest to the quadratic extremity;
- Push the two extremities into the intersection point ensuring the connection of the features.

Figure 5-22(a) presents the lines and quadratics detected using the method described in the previous sections. Figure 5-22(b) presents the same edges after the connection process, a small rectangle points out an area that was corrected. The process took 0.7 seconds in a Pentium IV at 2.0GHz.



(a) (b)

Figure 5-22: The initial edges (a) and after the connection process (b) in the arch example

5.4 Dense mapping to add pixels to the model

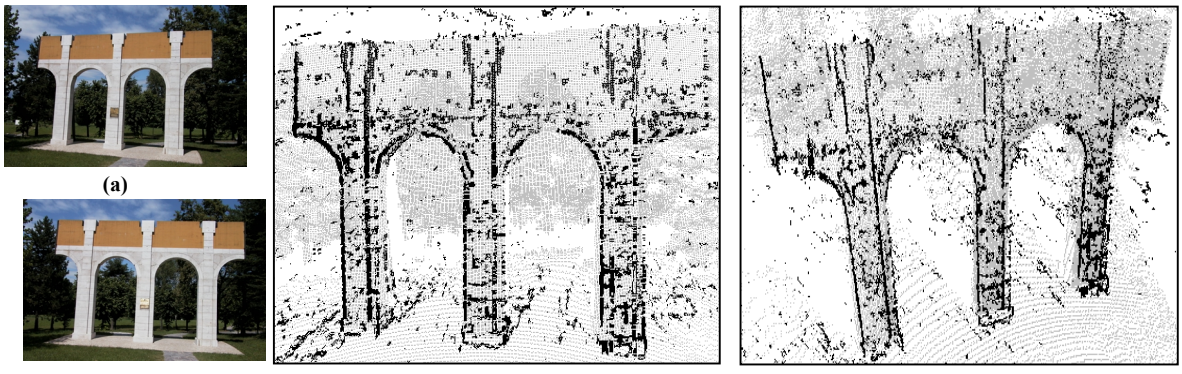
The tools developed along this chapter make possible the extraction of 3D information directly from two or more intensity images. These tools offer a new possibility when computing our 3D models: to add new 3D points to the original range cloud. This can be particularly useful in areas where range data is missing (holes, missing scans, etc.) or to increase the density of points in areas of high interest and 3D content. In these situations, additional 3D data can be computed from pairs of high-resolution intensity images to compensate for lack of range data or to increase the 3D point density.

The technique tries to triangulate all the pixels of the rectified images (a sub-sampling factor is defined to triangulate only part of the intensity image pixels and speed up the process, but results in a reduced number of triangulated points). The triangulation is done with the same tools developed in sections 5.1 and 5.2. The main difference is that the range is not used to guide the matching anymore, since the process is supposed to add data in regions where no range is available. The matching is done along epipolar lines considering that the images were taken from close viewpoints (such as in stereo techniques) meaning that matching points are spatially close in both the images.

To increase the reliability of the matching, the following conditions are considered; these are widely used in dynamic matching techniques [Cox96]:

- Points where the variation of the gradient is lower than a given threshold are not considered, since the matching is difficult in points with few texture information.
- A threshold for the cross correlation is used to reject correspondences that are not reliable.
- The matching results from image 1 to image 2 and from image 2 to image 1 must be symmetric. A maximum disparity value is considered above which the matching is rejected.
- An ordering condition is considered to guarantee that the order of detected correspondences in the two images is the same.

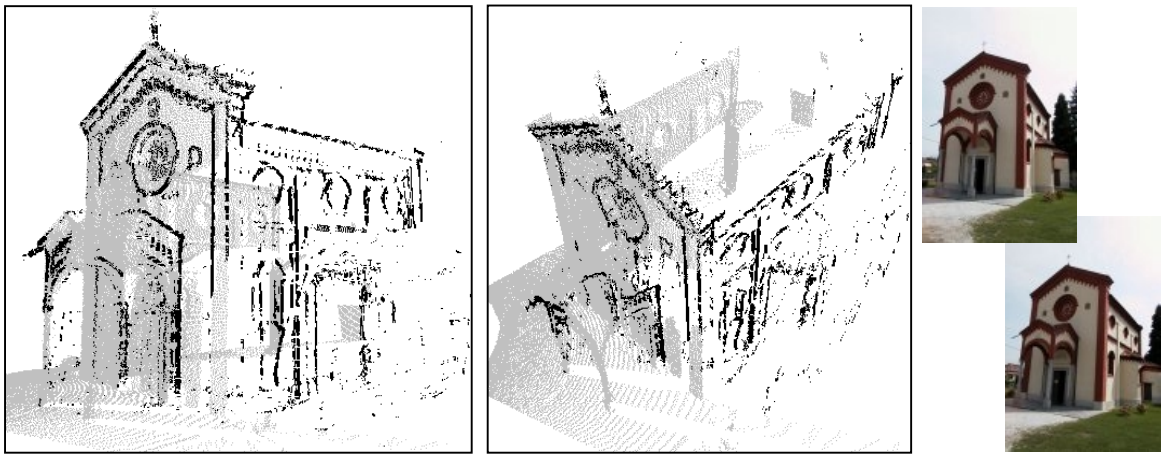
Figure 5-23 presents an example with the arch images. This example illustrates the registration between the data coming from both sources. The two intensity images (Figure 5-23(a-b)) are used for passive triangulation. Figure 5-23(c-d) present in light grey the original range 3D points and in black the 3D points coming from the triangulation of the intensity images. 37851 triangulated points were introduced in 2 minutes and 16.1s using a Pentium IV at 2.0GHz.



(a) (b) (c) (d)

Figure 5-23: Introduction of triangulated points into arch model

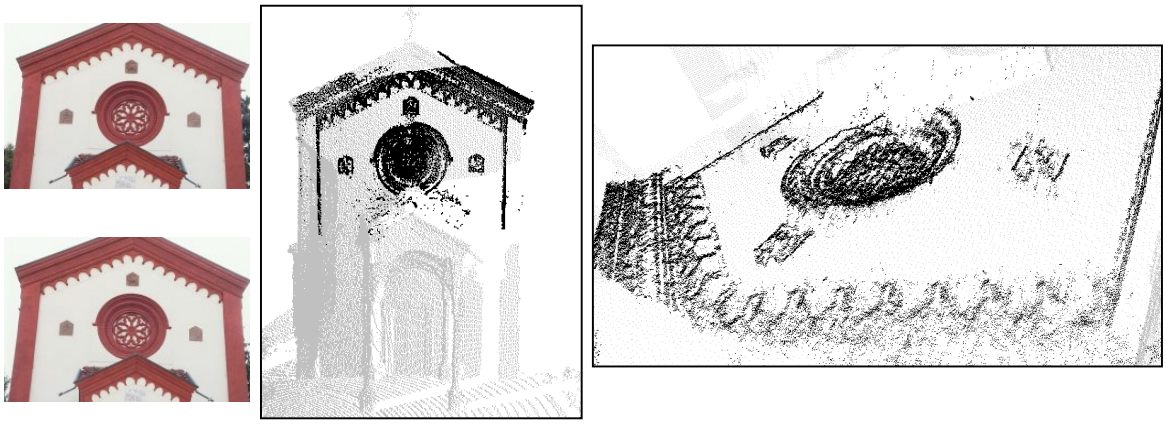
Figure 5-24 presents similar results with images from the church in Barza (Italy). In this case, the range image and the two digital photographs do not cover exactly the same area. The two images were registered and the calibration tuned using the process in section 5.2. This example demonstrates how extra data can be added in areas where no range is available. In this example, the front wall is common to both datasets and was used to register the data, but all the information from the right wall (black points) is extracted from the passive triangulation of the intensity images. 35107 triangulated points (in black) were introduced in 1 minute and 33.5 seconds (Pentium IV at 2.0 GHz) in the range data (light-grey). Figure 5-24(a) and (b) present two snapshots of the 3D points, and Figure 5-24(c) the two images used for triangulation.



(a) (b) (c)

Figure 5-24: Introduction of additional triangulated points into church cloud of points

In Figure 5-25, another example with the Barza dataset illustrates how digital photographs can provide additional high-resolution data to increase the density of 3D points in areas of interest or rich of 3D contents. The two images used for the triangulation (Figure 5-25(a)) cover only a reduced part of the area of the range image. The passive triangulation of these images will result in a large density of points in the "Rosace", for example. Two snapshots of the final cloud of points (in black the triangulated points from intensity data and in light-grey, the original range cloud of points) are presented in Figure 5-25(b-c). 76198 points were added in the initial range cloud of points in 1 minutes and 22.4 s.



(a) (b) (c)

Figure 5-25: Increase of 3D points density with passive triangulation in Barza church dataset

5.5 Conclusions

This section demonstrates how passive triangulation techniques of intensity images can bring improvements in 3D models computed from range data. The combination of both range and passive triangulation techniques allows getting more precise and reliable models since it enables the use of the best data (range or intensity), depending on the case, as they are fully registered.

The calibration is guaranteed of good quality through the calibration-tuning process presented in 5.2. The range segmentation can be improved using the technique presented in 5.3 to extract 3D lines and quadratics, guided by the intensity data. Finally, intensity data can also provide additional 3D data to increase the 3D point density in areas of interest or to complete the range data in areas where no data is available (occlusions, reflections, etc.).

The quality of the process depends mainly on two factors. From one side, the quality of the original calibration of the cameras can influence the process used to tune these calibrations since they are used as initial estimates. This means that if the images are badly registered initially, the algorithm cannot compensate and the result is of poor quality.

The other important factor of the process is the quality of the 3D edge detection in range data. Discontinuities in range data are used in the calibration-tuning process to select the closest 3D point for the next calibration. They are also used in the extraction of 3D lines and quadratics. In the examples used (arches and Barza church), the density of the range data is quite low making difficult the detection of precise discontinuities, which influences the global results of the process.

Fortunately new laser scanners have better resolution, making the detection of discontinuities easier and much more reliable. The use of such data will probably lead to better results than the ones presented in this section.

The acquisition of the intensity data is also an important matter. Mainly in the dense mapping steps, it is important to acquire data from close viewpoints (to increase the reliability of the matching algorithm). Data may also be rich in texture information since otherwise no matching is possible (this is the example of the white wall in the Barza church). On the other side, if images are taken from almost the same position, the triangulation angle is reduced and this can increase the errors in the triangulation step.

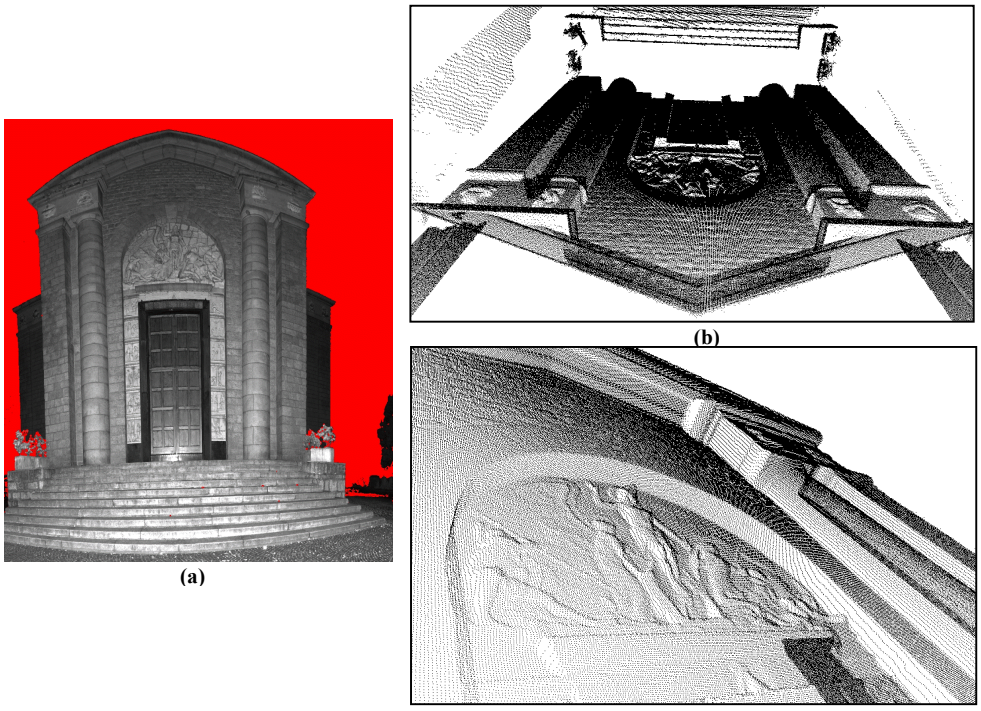
6 Experimental Results

In the previous chapters we used several datasets to illustrate the algorithms developed to register and combine range and intensity data. The goal of this chapter is to unify our work by applying all the algorithms to the same dataset.

The software is applied to an outdoor and an indoor scene in this chapter. The outdoor scene is from a church located in the city of Laveno, Italy, on the shores of the "lago Maggiore". It was selected because of the presence of many details in the carvings located above the main doors. The indoor scene is from the "San Stefano" Church in Verona, Italy.

Both scenes were acquired with different laser scanners, different also from the ones that have been used in the previous chapters. The church of Laveno was acquired with a high resolution Zoller and Fröhlich IMAGER 5003. The reflectance image and two snapshots of the 3D range cloud of points of the model are presented in Figure 6-1. The resolution of the reflectance image is 906x1036 pixels and red points indicate pixels for which no value was measured by the laser scanner. The digital photographs were acquired with Canon PowerShot Pro70 (1024 x 1536 pixels) and Pro90 (1856 x 1932 pixels) digital cameras. Three of these images are presented in Figure 6-2.

The San Stefano church was acquired with a Cyrax 2500 laser scanner providing a reflectance and range image of 338 x 338 pixels (Figure 6-3). The photographs were acquired with a Canon PowerShot G2 digital camera with a resolution of 2272 by 1704 pixels. The resulting data is presented in (Figure 6-4).

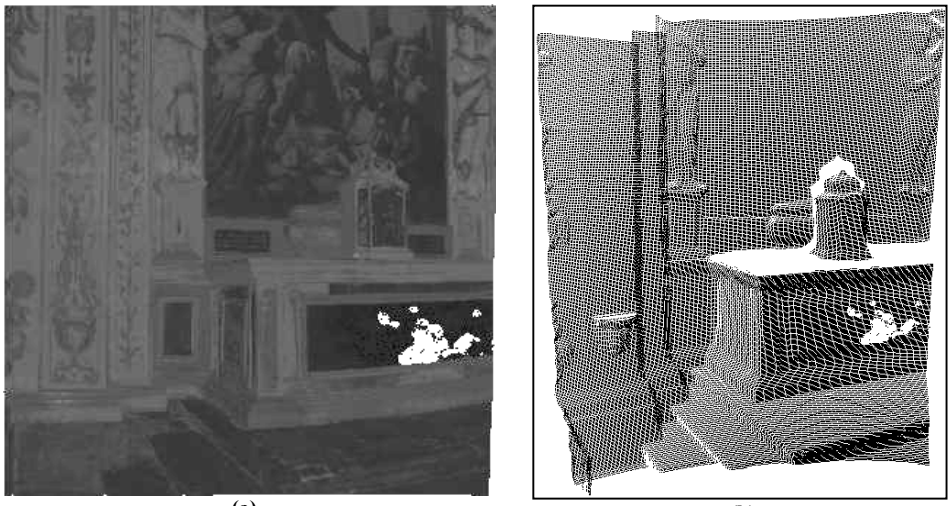


(a) (b) (c)

Figure 6-1: The reflectance (a) and two snapshots (b,c) of the cloud of points of the Laveno church scene (906 x 1036)

(a) (b) (c)

Figure 6-2: Three digital photographs of the Laveno church acquired with a Canon PowerShot70 camera (1024 x 1536)



(a) (b)

Figure 6-3: Reflectance (a) and cloud of points (b) of the San Stefano church scene (338 x 338)

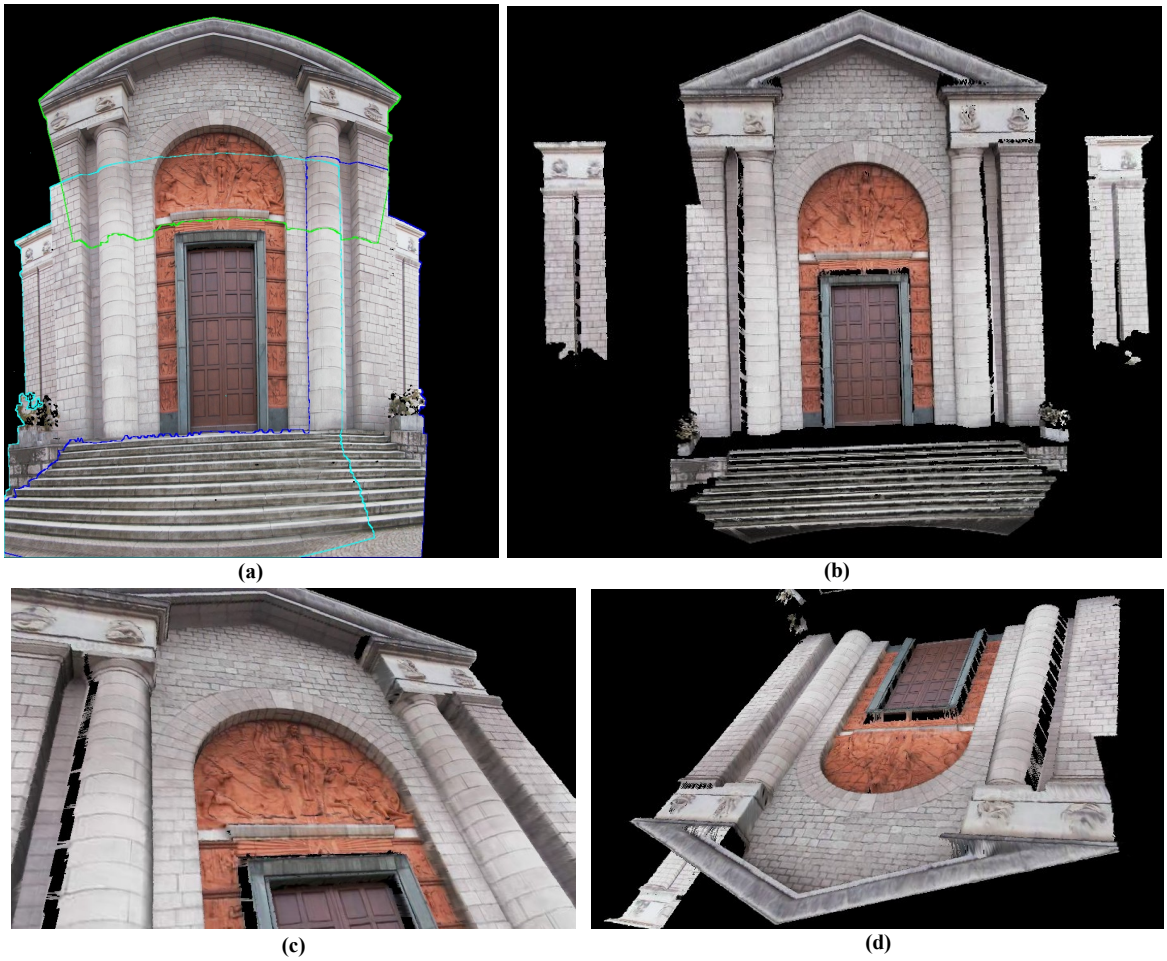
Figure 6-4: Digital photograph of the San Stefano church acquired with a Canon PowerShot G2 camera (2272 x 1704)

6.1 Camera calibration

The techniques presented in chapter 2 and chapter 3 for registering intensity and range data were applied to three digital photographs (presented in Figure 6-2) of the Laveno church scene. Since the acquisition of the digital photographs and range data was done in two different sessions, the digital pictures were not acquired exactly from the same viewpoint as the range data, thus, to get better results, some correspondences were manually added by the user to the ones selected automatically, mainly in the borders of the images to help a good calibration of the lens distortion. The resulting blended texture map and three snapshots of the final textured model are presented in Figure 6-5.

Table 6-1(a) gives the results of the calibration of the three images used to compute the texture map for the triangulated model.

For the San Stefano model, since the field of view of the Cyrax laser scanner is reduced, a single digital photograph is enough to get a texture map for a view of the laser scanner. The results of the calibration are presented in Figure 6-6 and Table 6-1(b).



(a) (b) (c) (d)

Figure 6-5: Texture map (a) and three snapshots of the Laveno church model (b,c,d)

(a) (b) (c)

Figure 6-6: Texture map (a) and two snapshots of the San Stefano model (b,c)

Laveno Church			
Image	1	2	3
Resizing error	5.56	7.50	5.52
n. of matchings	28	25	25
RANSAC matchings	26	25	24
Initial calib. Error	2.24	2.14	2.62
Final calib. Error	2.01	2.14	2.62

(a)

San Stefano Church
1
6.39
42
40
3.11
3.11

(b)

Table 6-1: Results of the calibration of the two test scenes - (a) Laveno church - (b) San Stefano

6.2 3D correction based on a single intensity image

The software developed in chapter 4 was used in the triangulated models of the test scenes to guide the correction of lines and quadratics based on a single registered image. Results of line and quadratic detection in the digital photograph of the Laveno church are presented in Figure 6-7.

The final results of the correction for both models are presented in Figure 6-8 and Figure 6-9. Red edges are the corrected edges whereas black edges are the original edges detected in the triangulated model. Table 6-2 presents some details of the algorithm for both scenes, such as the number of points corrected and the average error for the two models.

(a) **(b)**

Figure 6-7: Results of line (a) and quadratic (b) detection in Laveno church image

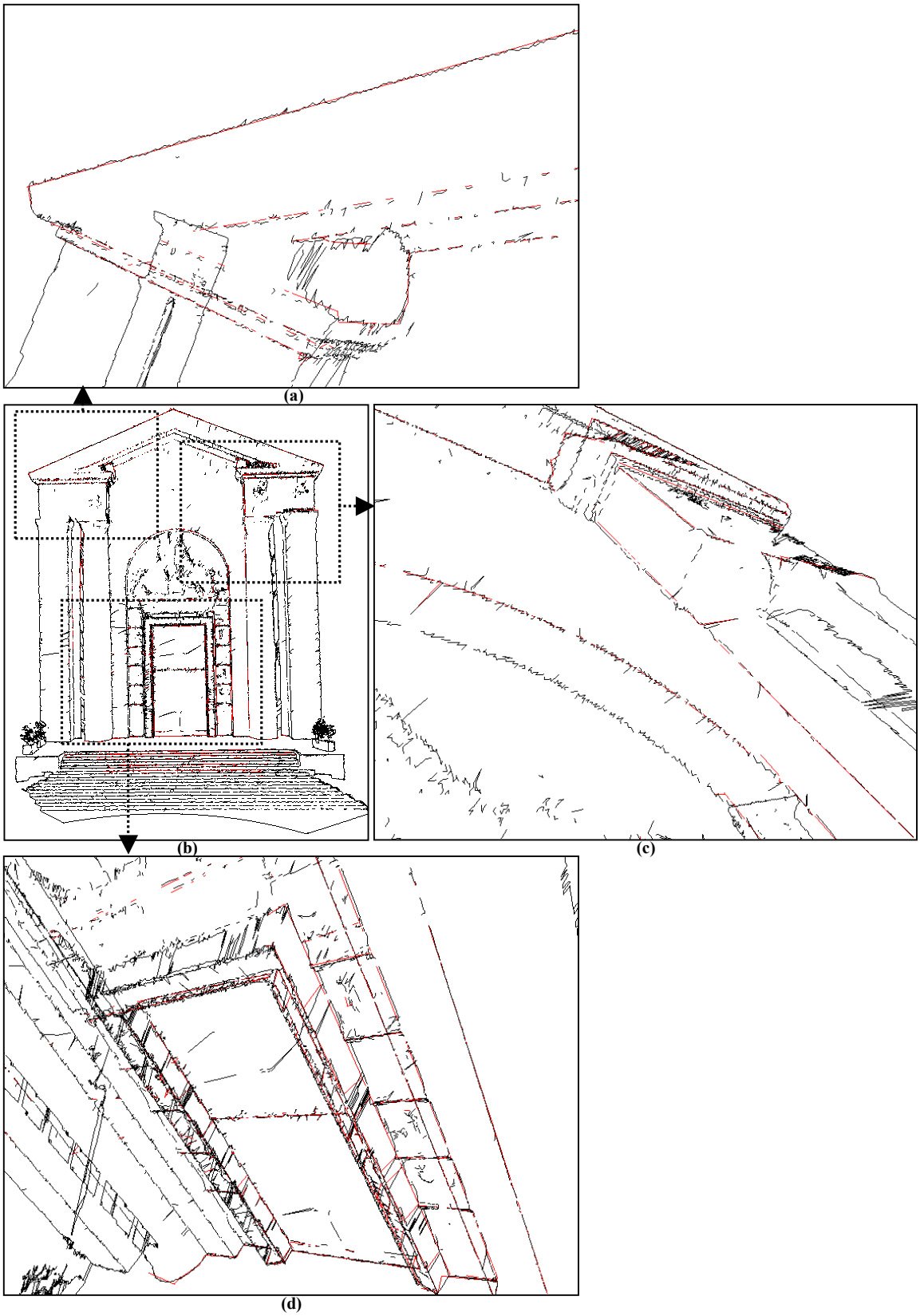


Figure 6-8: Four views of the original (black) and corrected (red) edges in the Laveno church model

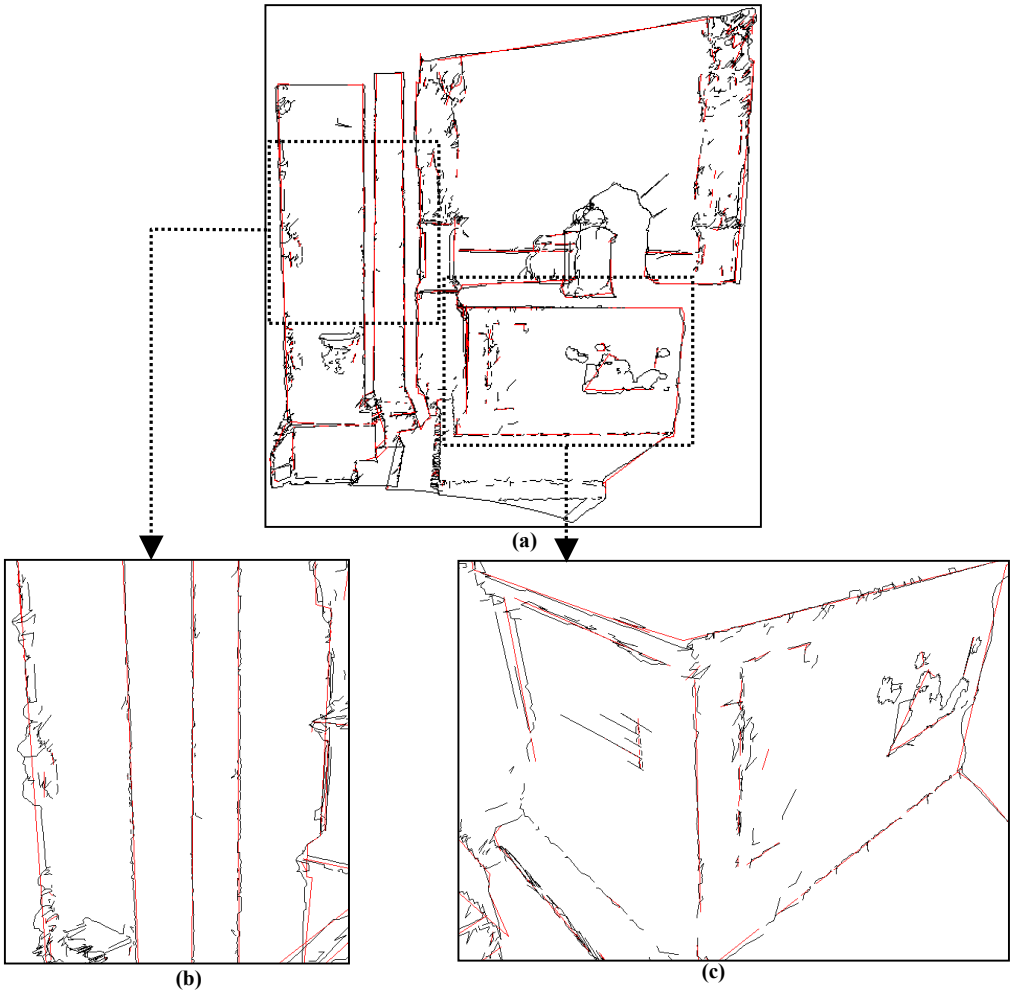


Figure 6-9: the original (black) and corrected (red) edges in the San Stefano model

	Laveno Church
Detected Jump edges	412
Detected roof Edges	6783
Detected lines	159
Filtered lines	51
Detected Quadratics	1
Number of points corrected	17705
Average correction (meters)	0.0246

(a)

San Stefano
71
766
184
88
1
5121
0.0232

(b)

Table 6-2: Results of straight line and quadratic correction in the test scenes - (a) Laveno church - (b) San Stefano

6.3 Combining passive triangulation and range data

In this section, the algorithms presented in chapter 5 were applied to our test data. Two digital photographs of each dataset were selected to illustrate the calibration tuning and the improvement of the segmentation of the initial range data. Finally, pairs of images were used to add 3D pixels to the original 3D range cloud of points.

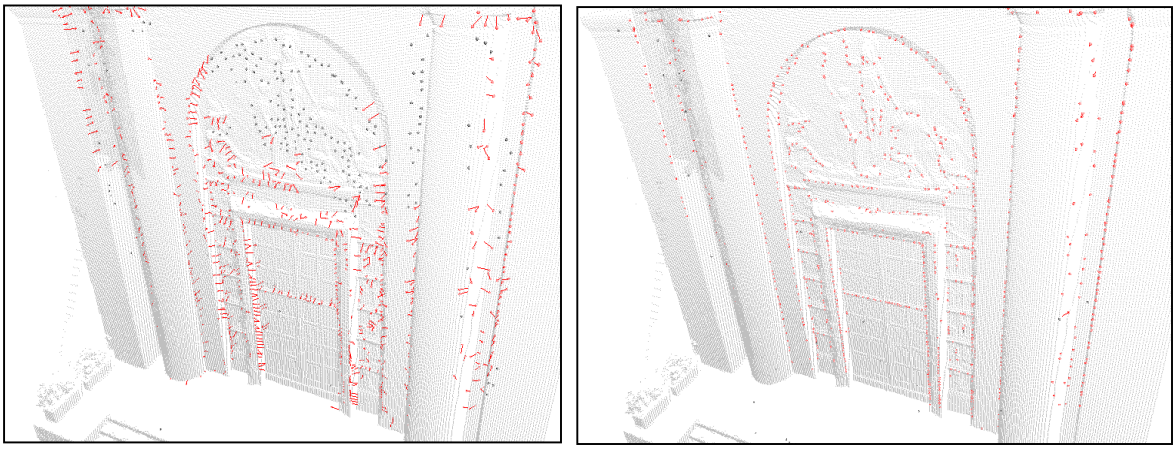
6.3.1 Tuning of the calibration

The two images used to illustrate the calibration tuning for the Laveno church are presented in Figure 6-10. In this figure, red lines correspond to the epipolar lines resulting from the fundamental matrix estimation.

Figure 6-10: Two images of the Laveno Church and the estimated epipolar lines

Figure 6-11 shows the results of the passive triangulation for the Laveno church before (a) and after (b) the calibration tuning process. Figure 6-12 describes the evolution of the average distance between triangulated and range points along the tuning.

Similar results are presented in Figure 6-13 and Figure 6-14 for the San Stefano example. Since the area covered by the photograph is larger than a single range view, the process was ran after registering two range views to use all the area of the photograph. The views were registered using the implementation of the iterative closest point [Besl92] in [Sequeira96].



(a) **(b)**

Figure 6-11: Triangulated points (red) and range points before (a) and after (b) the calibration tuning process (Laveno)

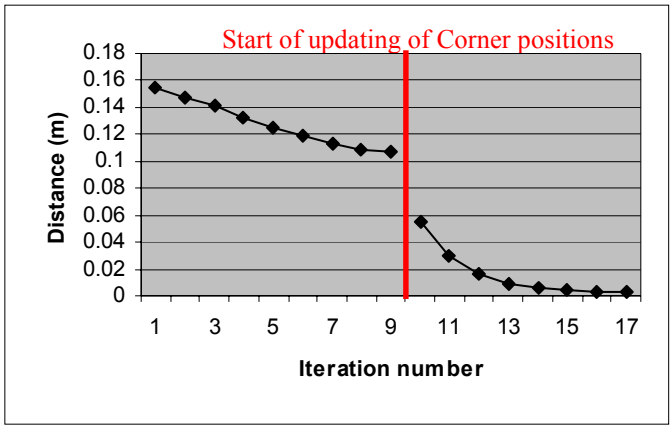


Figure 6-12: Evolution of distance between triangulated and range points along the calibration tuning process (Laveno)

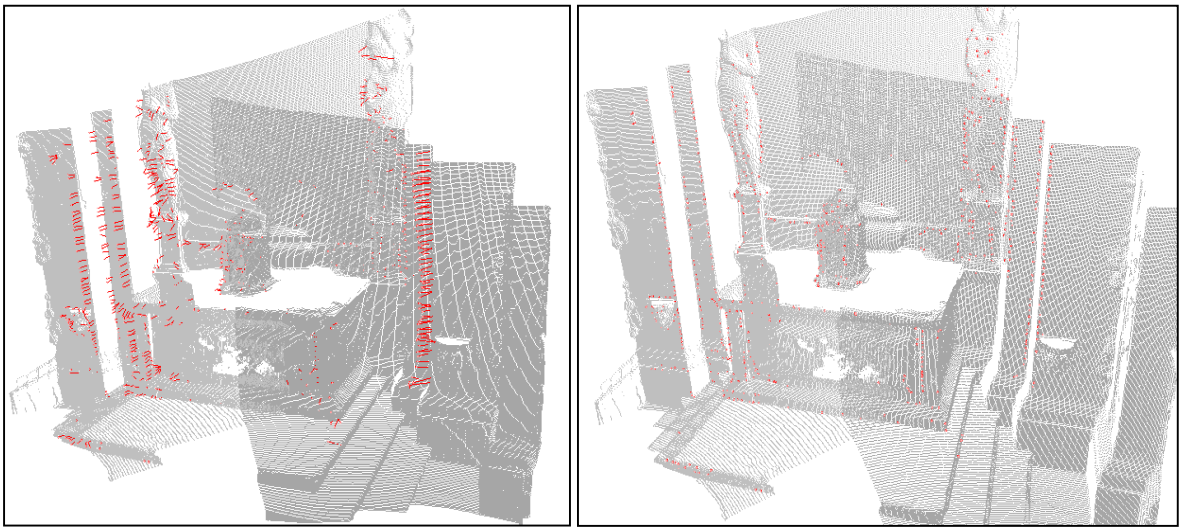


Figure 6-13: Triangulated points (red) and range points before (a) and after (b) the calibration tuning process (San Stefano)

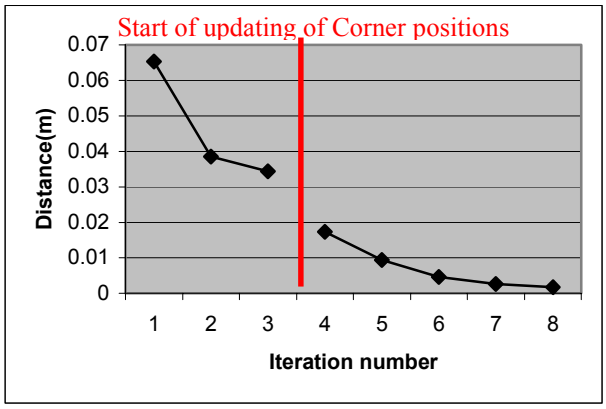


Figure 6-14: Evolution of distance between triangulated and range points along the calibration tuning process (San Stefano)

6.3.2 Improving range segmentation

The algorithm described in section 5.3 is applied to the pair of images of Figure 6-10. The results of the whole process are presented in Figure 6-15 for the Laveno church model: 36 lines and 1 quadratic were corrected along the process.

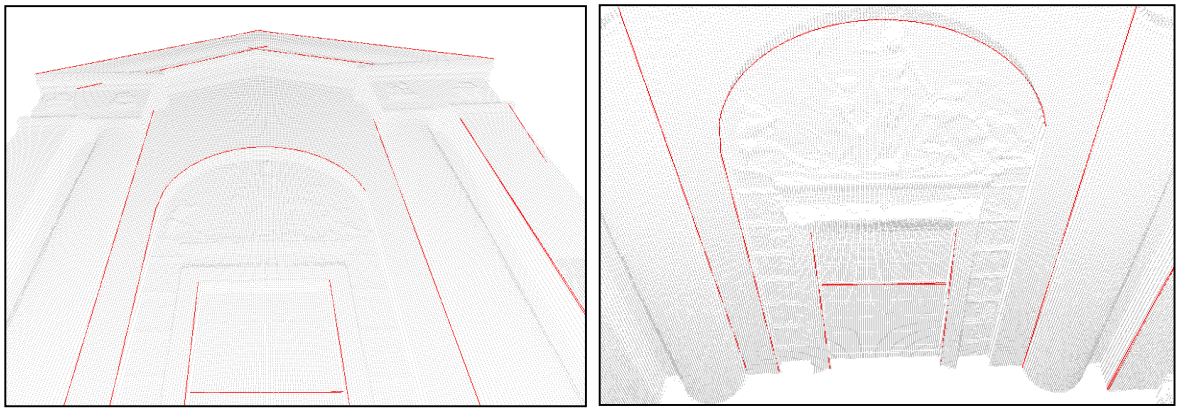


Figure 6-15: Detected features (red) and initial range cloud of points (grey) in Laveno church model

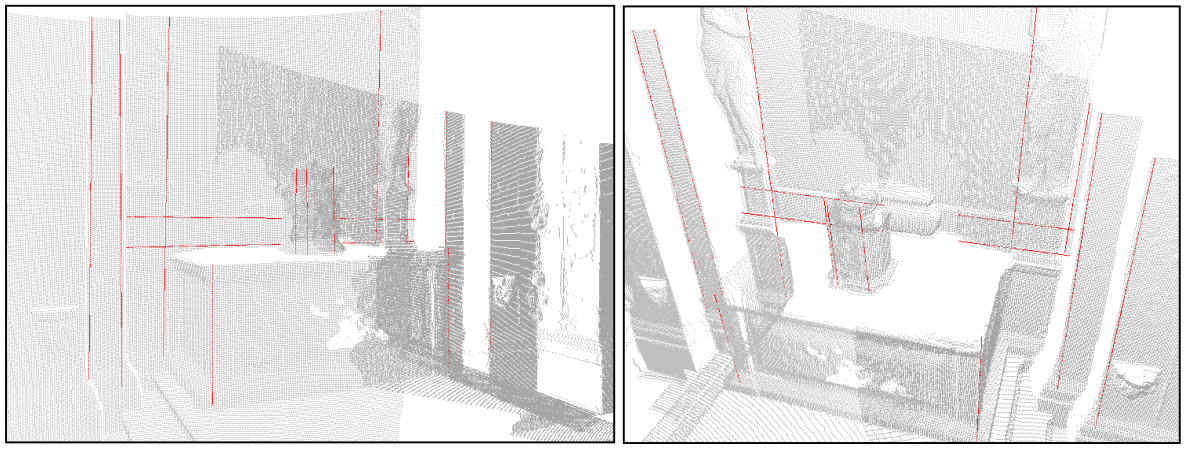


Figure 6-16: Detected features (red) and initial range cloud of points (grey) in San Stefano model

Figure 6-16 presents the results of the correction in the San Stefano model. In this case, 13 lines were corrected.

Given the complexity of the scenes and the large number of edges detected by the Canny detector, some user interaction was necessary to filter the large number of detected edges and guide the detection of lines and quadratics in the 2D digital photograph.

6.3.3 Adding pixels to the model

The tests in this section are related to the addition of extra 3D points into the original range cloud of points using passive triangulation (section 5.4). We selected two pairs of images from the acquired digital photograph to illustrate the process with the data from the Laveno church and two more examples for the San Stefano dataset.

Figure 6-17 presents the results of the process with the two images of Figure 6-10. These photographs were acquired with a Canon Powershot pro70 camera with a resolution of 1024 x 1536 pixels. Our algorithm has introduced 574123 additional points coming from passive triangulation (black points in Figure 6-17) into the original cloud of points (represented with blue points).

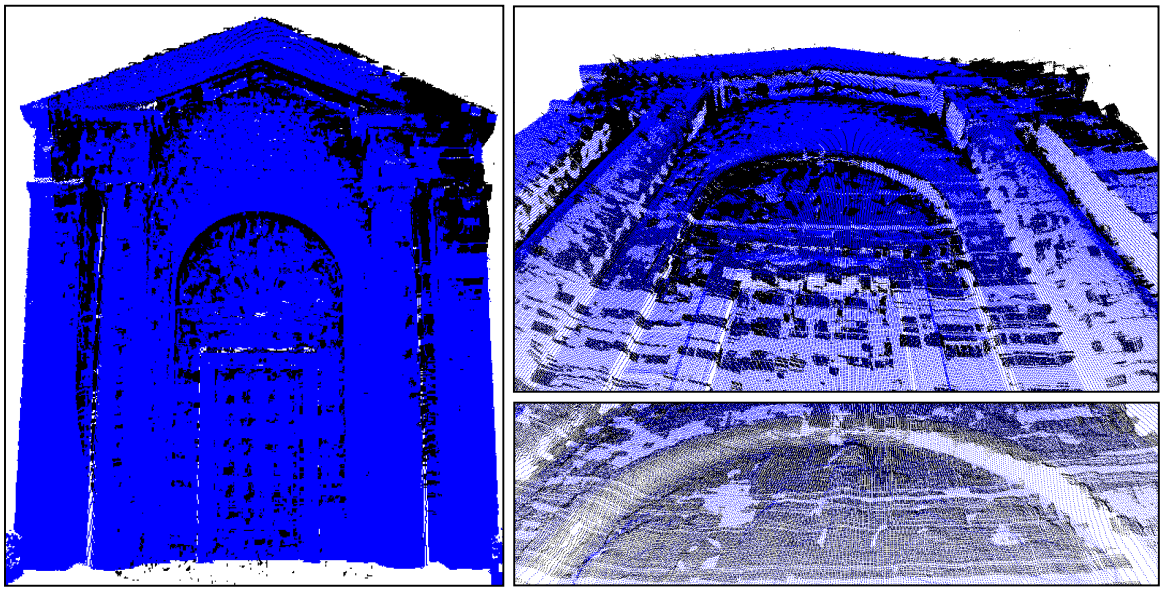
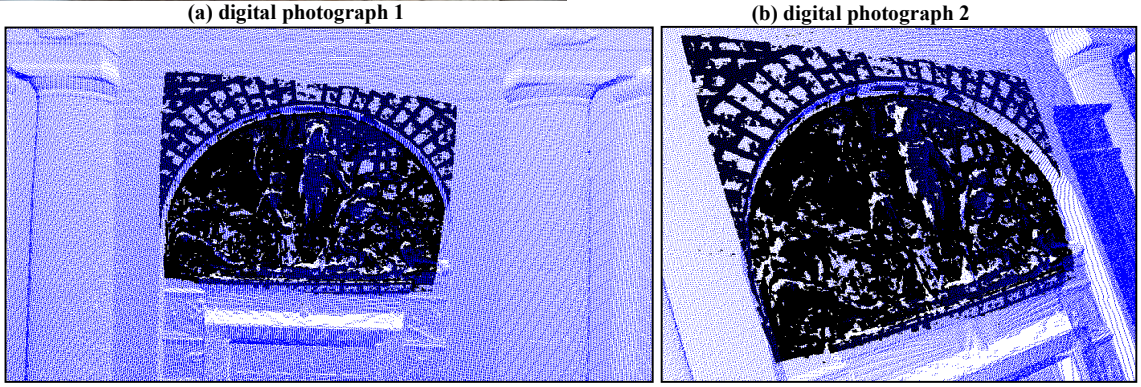


Figure 6-17: Addition of 3D points in the Laveno church based on the two intensity images

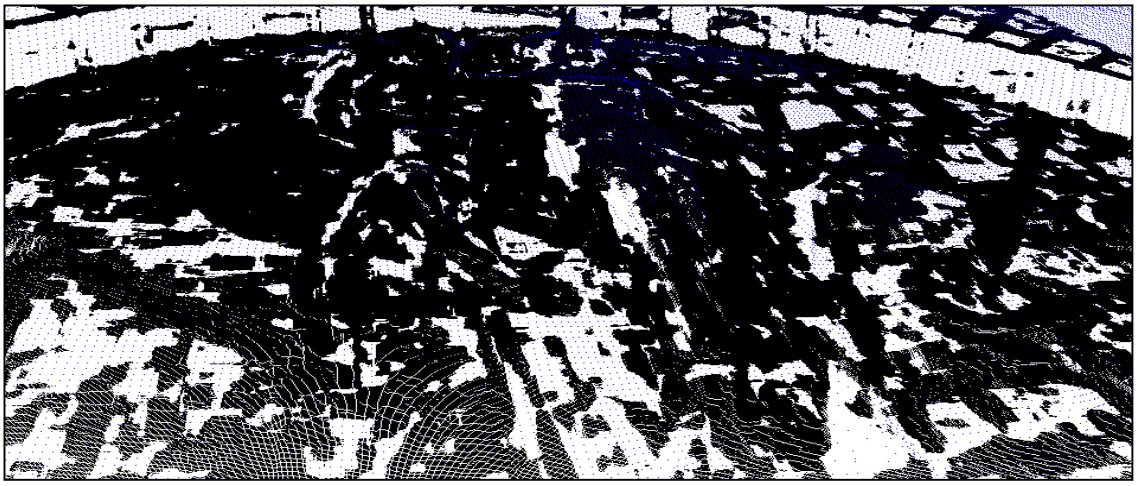
Blue - original cloud of points; Black - 3D points coming from passive triangulation of intensity images

Another example is presented in Figure 6-18. In this case, the two digital photographs (Figure 6-18-a,b) were acquired with a Canon Powershot Pro90 digital camera (with a resolution of 1856 x 1932 pixels). The two selected images are from the carving above the central door to illustrate how our method can be used to increase 3D point density in areas rich in 3D content. In this case, 1620726 triangulated points were added to the cloud of points (see Figure 6-18-c,d and e).

(a) digital photograph 1 (b) digital photograph 2



(c) Original and triangulated points (d) Original and triangulated points



(e) Original and triangulated points

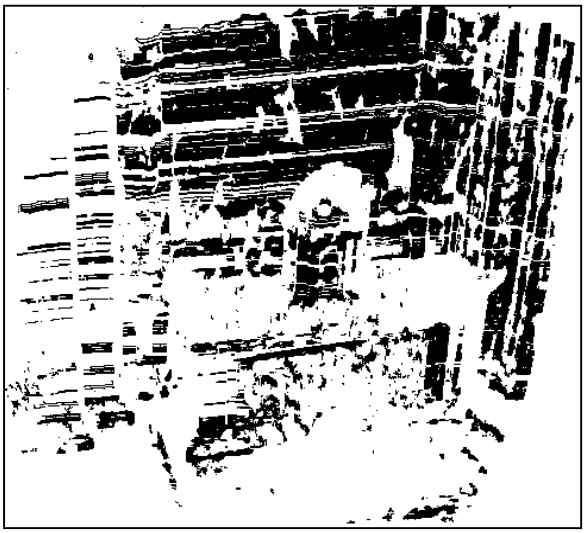
Figure 6-18: Addition of 3D points in carving above the main door of the Laveno church

Blue - original cloud of points; Black - 3D points coming from passive triangulation of intensity images

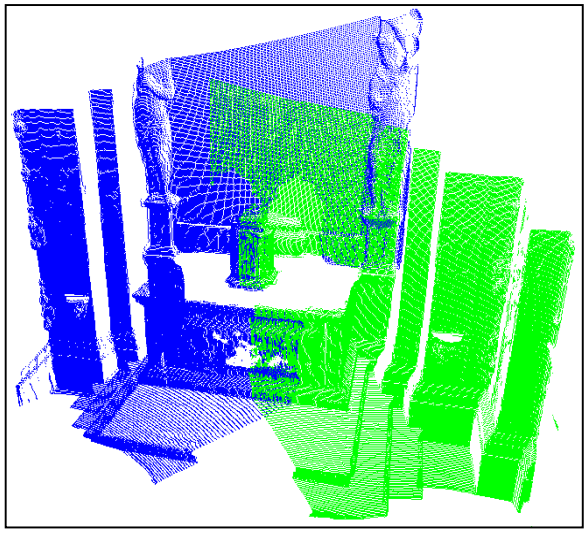
Figure 6-19 is an example from the San Stefano dataset. The intensity images (Figure 6-19-a,b) used for the passive triangulation were acquired with a Canon Powershot proG2 camera (2272 x 1704 pixels). Figure 6-19-c presents the 3D points triangulated from the intensity images and Figure 6-19-d shows, in blue and in green, the two registered range images corresponding to the area of interest. In this case, 798298 triangulated points were added into the original range images (see Figure 6-19-e,f).

(a) digital photograph 1

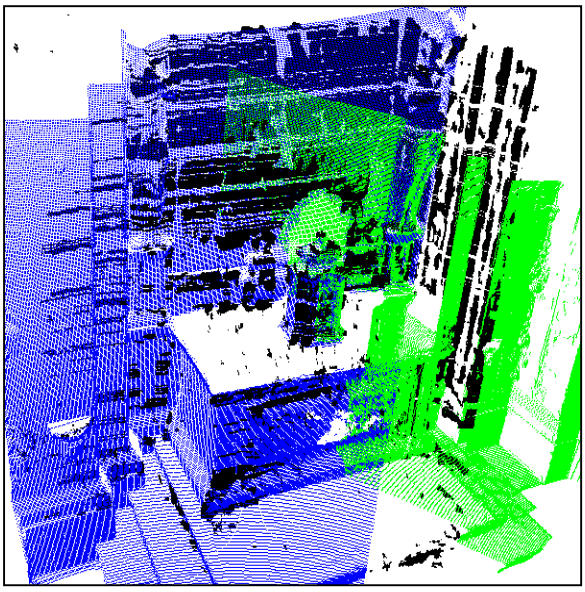
(b) digital photograph 2



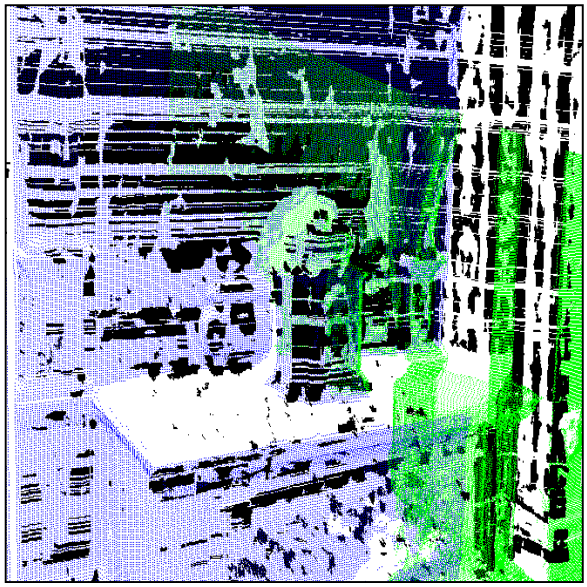
(c) Triangulated points



(d) two registered original range images



(e) Original and triangulated points

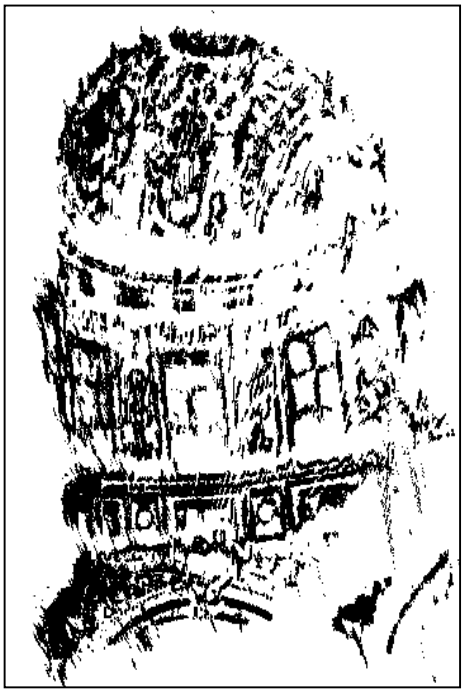


(f) Original and triangulated points

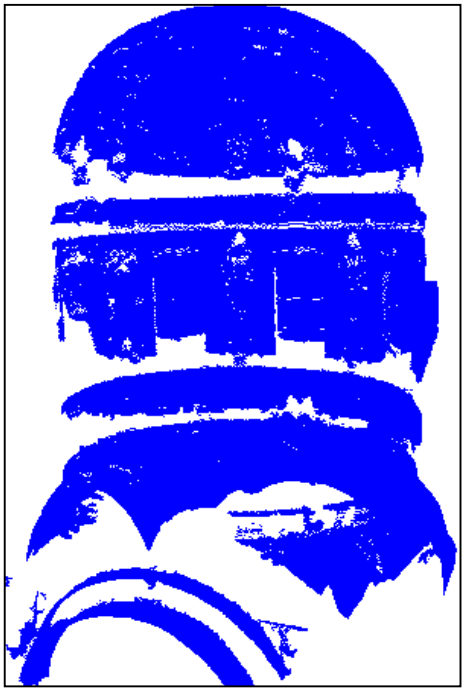
Figure 6-19: Addition of 3D points in San Stefano dataset

(a) digital photograph 1

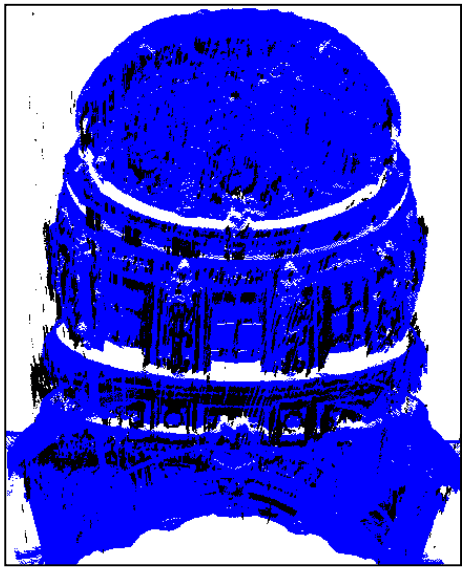
(b) digital photograph 2



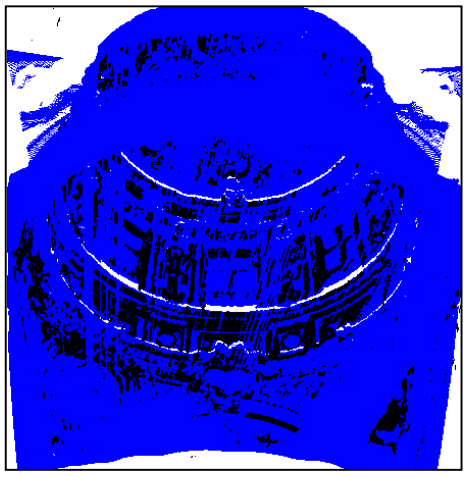
(c) Triangulated points



(d) Original range data



(e) Original and triangulated points



(f) Original and triangulated points

Figure 6-20: Second example of addition of 3D points in San Stefano dataset

Similar results are presented in Figure 6-20 with another range image acquired in the San Stefano church. The two intensity images (Figure 6-20-a,b) were used to triangulate new 3D points (Figure 6-20-c) that can be added to the original range image (Figure 6-20-d). In this example 743661 points were added to the original range image (Figure 6-20-e,f).

6.4 Discussion

In this chapter, all the algorithms developed in the thesis were applied to an outdoor and an indoor dataset acquired with new laser range finders. All the processing was done with a minimum of user interaction. User intervention was necessary only in two steps:

- In the matching phase, where some correspondences were manually added to ensure a good initial calibration.
- In the detection of quadratics and lines in the intensity images. The output of the Canny edge detector is very important in this step since connected edges are the base of all the detecting process. This makes necessary either to adapt the Canny thresholds to the given image or to guide the process by removing additional edges, that are not of interest and can introduce errors in the process.

The results are similar to the ones obtained with the other datasets presented in the thesis proving that our algorithms are scalable (the range data acquired with the Zoller and Fröhlich laser scanner is much larger than the previous datasets used) and can be easily adapted to any type of scanner or digital camera since no knowledge of the acquisition system is necessary.

7 Conclusions and future work

This thesis presents several tools developed to use intensity images, combined with range data, to build 3D realistic models of real world scenes and solve some of the problems inherent to these techniques, when used separately.

Chapter 2 and 3 deal with the problem of registering the datasets and bring range and intensity data into the same reference frame. A complete set of tools has been developed to minimize user intervention during the registration stage. At the same time, many configurations were considered, giving the possibility to complete a good camera calibration with any experimental set up (combination of laser range finder and digital camera).

In chapters 4 and 5, we illustrate how intensity images can be an additional source of 3D information, useful to complete and correct range data. The main innovation is the use of digital photographs not only for texture mapping, but also as an additional 3D sensor that can provide valuable 3D geometric information of the scene. One possible use is the extraction and analysis of straight lines and quadratic features, to minimise the mixed point problem inherent to any range scanner (accurate surface boundary detection is a problem in range data, but can benefit from the use of intensity images). Finally, we achieve a real fusion of the datasets into a single cloud of 3D points that combines both range and intensity data in a new manner.

With these techniques, we build a bridge between two worlds that are traditionally independent when they can be complementary in many ways.

In the next paragraphs, the different tools developed within the thesis are summarised and the key features are outlined. At the end, a global discussion about the main achievements is done and possible future work is presented.

7.1 Summary

1. Affine registration of reflectance and intensity images

Affine registration consists in evaluating the planar affine-transformation between the reflectance and the intensity images. This planar transform is the starting point for all the matching process since it ensures that main features overlap in the two datasets, despite the different resolution of the images.

Additional transformations can be considered before the affine resizing operation, to increase similarity between edges. In this thesis, we consider the spherical to Cartesian correction of the reflectance image, since laser scanners normally return an image in spherical co-ordinates. In some scenes, with strong geometrical contents allowing for the detection of the three main vanishing points, the evaluation of the 3D rotation between the sensors is also considered.

Special attention is necessary during data acquisition, since the different nature of the sensors (laser range finder and digital camera) makes difficult the registration of the data. Both datasets must be acquired from close viewpoints to ensure similar edges in the images. In addition, better results will be achieved in scenes rich in texture information well distributed all over the image.

The affine registration can also provide an initial approximate texture map for the models, based on panoramic images in the “fast texture mapping” (Section 2.3.2).

The system has an integrated interface that allows the user to select areas of interest and refine the initial planar transform estimation in a fast and natural way.

Any kind of 2D images can be treated with our technique since depth information is not used in this step. An example is the use of the planar affine estimation software to register satellite images (see Appendix B).

2. Camera Calibration

The camera parameters are estimated from correspondences computed over the resized images (output of the planar affine registration). We use the camera model from Tsai [Tsai86].

To increase reliability of the system, even in presence of outliers (matching errors), a RANSAC estimation technique is applied during the camera calibration process.

A calibration tuning process has also been developed. Its purpose is to improve an initial calibration estimate by comparing iteratively reflectance and re-projected image.

The camera model is used to compute the "re-projected image" that is directly registered with range data and provides a colour texture map for the model.

Re-projected images are computed considering a Z-buffer technique to avoid occlusion errors in the final texture map image.

A feathering process is applied to merge the re-projection of several images and ensure smooth transitions in the final texture map.

The process will perform better in images rich in texture well distributed all over the three dimensions of the scenes. This is of particular importance during the estimation of the internal parameters of the camera, since lens distortion, for example, needs well-distributed correspondences to be evaluated correctly.

Full calibration of a camera is the default option, but the software also permits the use of additional knowledge to improve and simplify the calibration. For example, internal parameters of the camera can be provided to the system that will, in this case, perform only extrinsic calibration.

A user interface has been developed to give to the user the possibility to interact with the system if necessary. It makes possible to remove or add correspondences and, thus, ensures a good calibration even for difficult images, where automatic calibration performs poorly.

3. Improvement of range segmentation based on a single registered intensity image

The limited spatial resolution of range data induces errors that make difficult, for example, accurate surface boundary detection. In these conditions, registered digital photographs can be valuable tools to correct edges in triangulated models.

The system uses a single registered image to correct straight lines and quadratic features in an existing triangulated model. It detects jump and roof edges in the triangulated 3D model and compares them with 2D features detected in the intensity images.

Good results are obtained in man made environments, which contain a high density of straight lines and quadratics. In this case the effects of the mixed point problem can be significantly minimized. In outdoor environments that are not so structured (due to the presence of trees, clouds, etc.), the correction is more difficult. It can sometimes be necessary to guide the 2D detection of features in the digital photographs in order to avoid wrong corrections.

The system can be easily updated to deal with other features. The key points are to ensure a good and reliable detection of the features in the 2D photographs and to provide the necessary fitting functions.

3D edge points are connected and ordered according to the triangulation information, this allows the correction of complete 3D edges even if only subsets of the 2D features are detected in the intensity image.

Z-buffering is considered to treat complete 3D triangulated models by isolating, for each photograph, the 3D geometry that is actually seen from the camera viewpoint.

4. Use of registered intensity images and photogrammetry techniques to tune existing camera calibration

We use conventional intensity based techniques (fundamental matrix estimation, image rectification, passive triangulation) to improve the initial camera calibration estimation. 3D information is extracted from two or more intensity images (based on the initial estimation of the camera calibration) and compared with range data to refine the initial camera calibrations used for passive triangulation.

Discontinuity detection in range image is a key issue since the comparison is done between triangulated points and range discontinuities. The discontinuity detection process used is based on an analysis of the variations in the depth of range points [Sequeira96]. The use of more precise techniques to detect discontinuities (using for example curve fitting) would probably lead to better results.

5. Improvement of range segmentation based on pairs of digital photographs

This is an extension of the work presented in point 3 (improvement of range segmentation based on a single registered image). The difference is the use of passive triangulation to compute 3D locations from pairs of intensity images. In this case, the correction is guided with intensity images directly in 3D and not just in 2D as previously.

Tests have been done with the range data instead of the triangulated models since, in this situation, 3D information is extracted directly form 2D images.

6. Addition of 3D information from registered intensity images

A real combination of data coming from both sources is done. Points from range are completed with 3D points coming from passive triangulation of intensity images.

There is also the possibility to fill holes or complete range data based on intensity images acquired on the site. The technique also gives the possibility to increase data density in areas of interest.

This technique is especially useful when using range scanners with a reduced resolution, or to get high resolution in some areas by acquiring high resolution photographs.

7.2 Conclusions

In this thesis, we have fulfilled two main objectives:

i) We have developed processes and algorithms to register range and intensity data based on the Tsai camera model. Many processes have been developed with the objective of reducing and helping the user interaction. The calibration produces texture map images that can be used to render triangulated models coming from range analysis. These texture maps allow a fast evaluation of the quality of the registration. They also give the possibility to have 3D photo realistic models of real world scenes. These techniques have been tested both for indoor and outdoor environments and with many different acquisition systems (cameras as well as laser range scanners).

ii) Registered intensity images are used as an additional aid to extract 3D geometric information from the scanned environments. This is one of the main innovations of the thesis, since range and intensity data are traditionally used separately in 3D reconstruction processes. We demonstrate how intensity information can be used to complete and reduce

the effects of some problems inherent to range data acquisition, such as the mixed point problem. It has been shown that one or several registered intensity images can be used to optimise the segmentation of straight lines or quadratics in range data. Combining both data can also be useful to refine the initial camera calibration (based on the selection of 2D correspondences) by using the three dimensions in the calibration operation. Finally, we illustrate how registered intensity data can be a valuable source of additional 3D information, to complete or increase the density of 3D information present in the range data.

The most significant original contributions of the work presented in this thesis are:

- The introduction of an adaptive resizing algorithm to locate reliable registration points between reflectance and intensity images. This algorithm allows automating significantly the registration procedure between range and intensity data.
- The re-projection procedure, with occlusion handling, which allows the creation of a texture map that can be used directly to "paint" the 3D models.
- The colour feathering process used to merge several re-projected intensity images in a single texture map.
- The calibration tuning process used to refine the first calibration coming from robust estimation using the re-projected images.
- The techniques developed to improve the segmentation in range images from a single registered intensity image. This tool gives the possibility to use intensity images to minimize problems that are inherent to range data such as the mixed point problem.
- The calibration tuning process that triangulates 3D points from two or more intensity images and compares them with range data. This process is particularly useful since it gives the possibility to refine the calibrations using 3D information, whereas the initial calibrations come only from 2D image analysis.
- The use of passive triangulation in registered intensity images to improve the segmentation in range image and introduce additional 3D points in the range cloud of points. Here both range and photogrammetry techniques are really combined since the final 3D information comes from the two datasets.

7.3 Future Research

Along the work developed in this thesis, improvements and ideas about possible extensions of the work came out. Among them, the most interesting ideas are following:

Improvements

The system as it is can still be improved in many ways. The fully automatic calibration procedure only works with images containing a lot of structure information distributed in all the 3D space, such as buildings, for example. In addition, its performance also depends on a good acquisition discipline. For example, acquisition of range and intensity data must be done from the same viewpoint and not too close to the target (because of lens distortion). Still, the possibility for the user to interact easily with the system (by correcting the correspondences) and the many calibration options (such as the possibility to fix the internal camera parameters) make this system a useful tool to compute camera calibration even in difficult conditions.

The introduction of additional camera models into the system could help to reduce the final calibration errors since the Tsai camera model only considers one radial term in the lens distortion. Using other camera models with a more precise description would probably reduce the final calibration errors.

We also tried to keep our system as polyvalent as possible so that it can process any kind of data. If we concentrate on a particular case, with known sensors and dealing with a specific type of environment, it would be possible to optimise easily the several parameters of the camera calibration and feature detection. Doing so, we would increase the reliability of the system and the quality of the results for the particular case.

Extension of feature correction to surface correction

One of the natural improvements of the work presented here is to extend the feature correction (based on a single registered image or on passive triangulation of intensity images) to surface correction. This thesis focuses in solving the mixed point problem that is mainly located in surface boundaries, but no 3D improvement is brought in surfaces. It would be possible to extend our work to do some surface estimation and correction. In this case, the 2D feature detection in intensity image would be replaced by segmentation based

on colour for example. The final correction would consist in fitting planes or quadratic surfaces to the data using the intensity segmentation as an input.

Direct Range to CAD conversion

Another possible extension of this work is the development of methods to convert directly 3D models into primitives describing the environment. From the acquired data, features and curves could be extracted and the several areas of the model could be interpreted by the system. A simple example would be the extraction of boundaries and surfaces corresponding to walls, and replace in the model the wall by an ideal plane. This kind of data interpretation would lead to a large simplification of the representation of the models.

Re-triangulation of models

The work presented here stops in an early stage of the combination of intensity and range data. At the end we get either additional points, or some features. It would be of interest to study how to integrate and use this data during the triangulation of the 3D points. We could develop, for example, tools for the re-triangulation of parts of the models using pre-defined features (such as lines or quadratics), or to select areas of interest where the high-density data provided by intensity image should be used. Some early work in this direction has already been done in the last years at JRC and some of the results can be found in [Ruggeri01].

Adaptation of the technique to accept other 2D information

Another interesting extension of this work would consist in testing and improving the software so that it could deal with other data than digital photographs. Thermal images or X-rays are some of the 2D images that could be re-projected into 3D models using our techniques. It would be interesting to see how our software could treat this data and which adaptation would be necessary to be able to use them to texture the models. In some cases, the software would probably be effective, but the introduction of other kinds of camera models may be needed to deal with specific data.

A Appendix A - Experimental Set up

This appendix presents the various systems that have been used to acquire the range and intensity data used along this thesis.

The first section of the appendix presents, with some detail, the several laser range scanners used in this work. At the end, a table gives an overview of the main 3D laser scanners actually available in the market as well as their main characteristics.

The second section of the appendix presents the digital cameras used to acquire the photographs of the different models.

In Table A-1 the different datasets are presented, as well as the acquisition systems used for each model.

3D Model	Laser Range Scanner	Digital Camera
Laboratory 030	RIEGL LMS-Z210	Canon Powershot Pro70
Barza church	RIEGL LMS-Z210	Canon Powershot Pro70
Bornholm church	ZF LARA laser	Minolta digital Camera
Laveno Farmhouse	RIEGL LMS-Z210	Canon Powershot Pro70
JRC Arches	RIEGL LMS-Z210	Canon Powershot Pro70
Laveno church	Zoller&Fröhlich IMAGER 5003	Canon Powershot Pro70 and Pro90
San Stefano church	Cyrax 2500	Canon Powershot Pro G2

Table A-1: Acquisition systems used in this thesis

A.1 Laser Range Finder

Time of flight laser scanners used in this thesis are either pulsed wave or phase shift laser scanners. They are based on a laser that sends a signal to the scene. In pulsed wave scanners, the distance to the measured points is computed based on the time the reflected beam needs to return to the scanning device. In phase shift scanners, the distance is recovered from the phase difference between the emitted and received signals. In order to acquire data along the horizontal and vertical axes, these lasers are mounted inside a rotating head and are deflected by a rotating mirror.

In the thesis four of these laser scanners were used. Many of the models were acquired with a Riegl LMS-Z210 laser scanner property of the laboratory of the Joint Research Centre (Italy), where most of the work was developed. The data from the Bornhom church was acquired within the CAMERA TMR research network (CAd Modelling of Built Environments from Range Analysis [Camera]) by one of the partners of the project. In this case a Zöller and Fröhlich LARA laser scanner was used [Hancock98]. For the data presented in the experimental results, two other devices were used. The church of Laveno was acquired with a laser scanner acquired by the Joint Research Centre in 2003. This scanner is a high resolution IMAGER5003 from Zöller and Fröhlich. Finally, a Cyrax 2500 Laser scanner was used in the San Stefano church.

(a) (b) (c)

Figure A-1: (a) The Riegl LMS-Z210 Laser scanner of the JRC with a Canon Powershot Pro 70 Camera on top
(b) Zöller and Fröhlich imager 5003 laser scanner
(c) Cyrax 2500 Laser scanner

Two photographs of the laser scanners existing at the Joint Research Centre are presented in Figure A-1. Figure A-1(a) is the Riegl LMS-Z210 with on top a digital camera, whereas Figure A-1(b) is a photograph of the Zoller & Fröhlich IMAGER5003. Finally, the Cyrax 2500 is presented in Figure A-1(c).

Table A-2 presents the main specifications of the Riegl, the Zoller & Fröhlich and the Cyra laser scanners. The laser safety is given according to the European Safety Regulation EC 60825-1.

Main specifications	Riegl LMS Z210	Zoller and Fröhlich Imager 5003 Laser - LARA 53500	Cyrax 2500
Measurement Technique	Pulsed Wave	Phase Shift	Pulsed Wave
Laser safety class	Class 1	Class 3R	Class 3R
Resolution	5 mm	0.82 mm	2 mm
Min Range	2 m	0.4 m	1.5 m
Max Range	120 m	53.5 m	100
Range accuracy	15 mm	≤ 5mm	<6 mm
Field of view Vertical scan Horizontal Scan	80 deg 333 deg	310 deg 360 deg	40 deg 40 deg
Angular resolution	0.072deg	0.01deg	0.25 mm at 50 m
Measurement per seconds	9333	125000	1000
Laser wavelength	0.9μm	0.78 μm	0.532 μm
Beam divergence	3 mrad*	0.1 to 0.3 mrad*	≤6mm for 0-50 m
Dimensions	490x210 mm	300x180x350 mm	158x132.5x169mm
Weight	13 Kg	13 Kg	20.5 Kg

*3 mrad correspond to 30 cm beam width per 100 m of range.

Table A-2: Main specifications of Riegl LMS-Z210, Z&F IMAGER 5003 and Cyrax 2500 Laser scanners

Laser range scanner are still a relatively new technology, but there were many improvements in this kind of devices along the last years. The devices are becoming faster, more precise and with a very large resolution, as the demand increases. Nowadays, they are already used in many applications such as architecture, industrial applications, cultural heritage, modelling, robotics, surveying, virtual reality, preservation, reverse engineering… The main problem of these devices is still their cost, mainly due to the relative complexity of the optical hardware.

If many range scanners exist for low range applications, middle range scanners that can be used for modelling of real scenes are still rare: only few manufacturers provide such tools. Table A-3 presents the principal laser scanners that are commercially available and can be used for modelling of real world scenes (laser safety according to EC 60825-1).

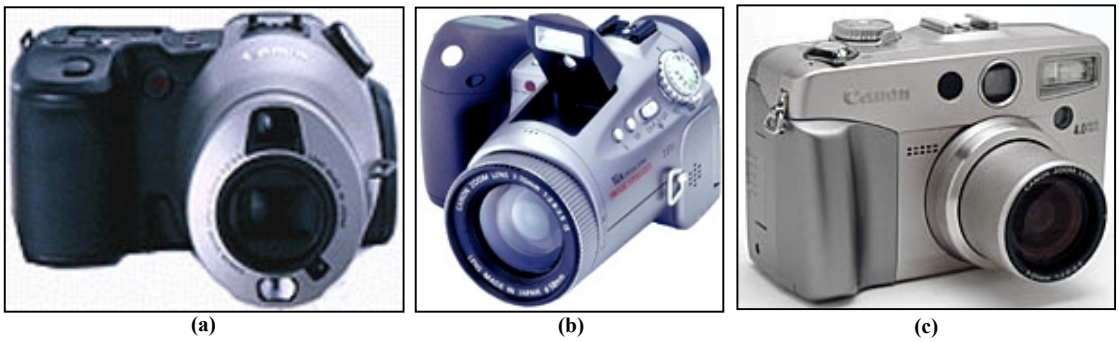
Manufacturer	Scanner	Measurement Technique	Max Range (m)	Eye safety Class	Scan rate (pts/sec)	Field of view (degrees)
3rdTech	Deltasphere	Pulsed Wave	12	IIIa*	19200	360x145
Callidus Precision Systems	Callidus	Pulsed Wave	80	1	1750	140x360
Cyra Technologies	Cyrax 2500	Pulsed Wave	100	3R	1000	40x40
iQsun GmbH, iQvolution AG	IQsun 880	Phase Shift	76	3R	240000	360x320
Mensi	GS 100	Pulsed Wave	100	3R	1000	360 x 60
MetricVision	MV 200	Phase Shift	60	1	1000	360 x90
Optech	ILIRIS_3D	Pulsed Wave	350	1	2000	40x40
Riegl Laser Measurement Systems	LMS-Z210	Pulsed Wave	120	1	9333	333x80
	LPM-800HA	Pulsed Wave	250	1M	1000	360x280
Zoller+Froehlich GmbH	IMAGER 5003	Phase Shift	53.5	3R	125000	310x360

***: Given using the American National Standards Institute (ANSI) safety classification.**

Table A-3: List of principal laser scanners suitable for 3D reconstruction of real world scenes

A.2 Digital Cameras

Most of the intensity data was acquired with the digital cameras available at the Joint Research Centre. Three commercial digital cameras were used: a Canon Powershot pro 70, and its evolution the Canon Powershot pro 90. The photographs of the model of San Stefano were acquired with a Canon Powershot G2 camera. In Figure A-2 pictures of these cameras are presented. For the Bornholm example, a Minolta digital camera was used.



(a) (b) (c)

Figure A-2: (a) Canon Powershot Pro 70, (b) Pro 90 and (c) G2 digital Cameras

Table A-4 present the main specifications of the three cameras presented in Figure A-2.

Main specifications	Canon PowerShot Pro 70	Canon PowerShot Pro 90IS	Canon PowerShot Pro G2
Pixels	1.68 Mpixels	3.34 Mpixels	3.87 Mpixels
Objective	6-15 mm (Equivalent 28-70 in 35mm)	7-70mm (Equivalent 37-370 in 35mm)	7-21mm (Equivalent 34-102 in 35mm)
Resolution	1536 x 1024	1856 x 1392	2272 x 1704
Dimension	145 x 85 x 132 mm	126.5 x 83.9 x 139.1	120.9 x 76.6 x 63.8 mm
Weight (without battery)	690 g	680 g	425 g
Compression	JPEG	JPEG, RAW, AVI (video)	JPEG, RAW, AVI (video)

Table A-4: Main Specification of Canon Powershot Pro 70, Pro 90 and G2 digital Cameras

A.3 Discussion

This chapter presents the experimental set ups used to acquire the models presented in the thesis. All the software was run with four different laser scanners and several digital cameras showing that the software is completely independent from the acquisition device. Since the applications use directly the data, no information about the specifications of the acquisition devices is actually needed. For range data, the software only needs the reflectance image and the co-ordinates of the range points. The spherical to Cartesian transformation needs in addition the angles of the field of view. For intensity data, only the digital photographs are needed. If additional information about the cameras (such as internal parameters) is known, this information can be used, but this is not necessary.

B Appendix B - Use of the resizing algorithm on satellite images

Registration of images is a relevant task in a wide variety of applications. In this appendix the methodologies discussed in the previous chapters are applied to the field of geospatial information (remote sensing satellite images and maps).

At the Joint Research Centre, we had the opportunity to discuss about the registration of images with several people. In particular, people working with geospatial information are confronted everyday with the problem of matching and registering satellite images and maps of the same area, but acquired at different times and with different systems.

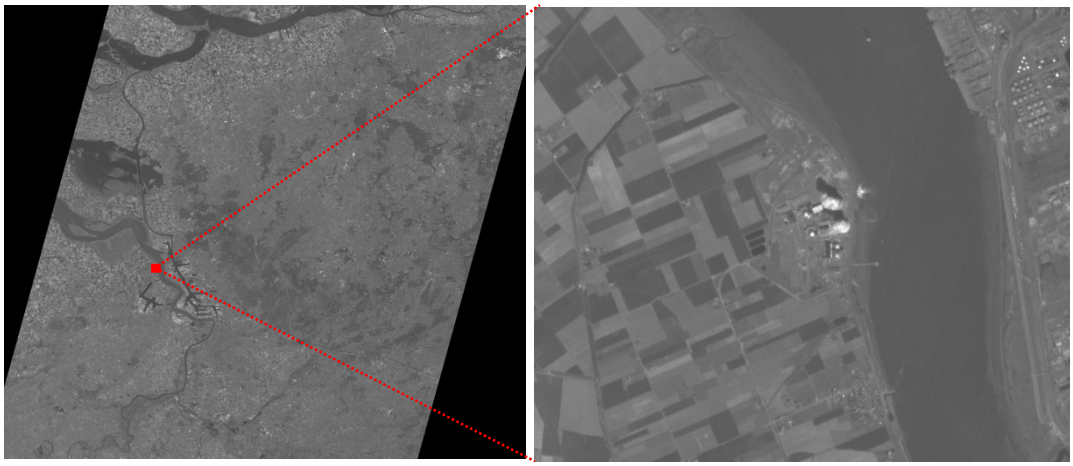
The problem is similar to the one faced in the second chapter of this thesis, when registering reflectance and intensity images. Also in this situation, the data comes from different sensors with different resolution and properties.

Tests were done to check the possibility to use the affine resizing algorithm, presented in chapter 2, to register satellite images. Some results of these tests are presented in this appendix

B.1 Satellite images used

Tests were done with two different sets of images. In the first test, two satellites acquired images of a nuclear facility in different times and with different resolutions. The laboratory that asked to realize those tests is involved in nuclear safety matters. Their goal is to take aerial views from nuclear facilities at different times and compare the new images with old ones to detect changes, like for instance the construction of new non-authorized buildings.

The two satellite images are presented in Figure B-1 and Figure B-2. Figure B-1(a) presents the full Spot panchromatic image (17592 by 15892 pixels), acquired at a resolution of 10m. Figure B-1(b) present a reduced area of the full image corresponding to the area of interest that will be used in our algorithm.



(a) (b)

Figure B-1: the first satellite image and a zoom of the area of interest (Spot panchromatic with 10 m resolution)

The second image, acquired a few months later, is an IRS-1C panchromatic image. It has a resolution of 5.8m. The satellite image used for the experiments is presented in Figure B-2; its resolution is 1923 by 2163 pixels.

Figure B-2: The second satellite image

The algorithm was tested with a second set of images. In this case, the objective was to register a satellite image of the Joint Research Centre (Figure B-3) with a 2D map of the site (Figure B-4). The rectangular area in Figure B-3 corresponds to the area of interest.

Figure B-3: The ikonos image of the Joint Research Centre, 4214 by 3744 pixels. The rectangle indicates the area of interest

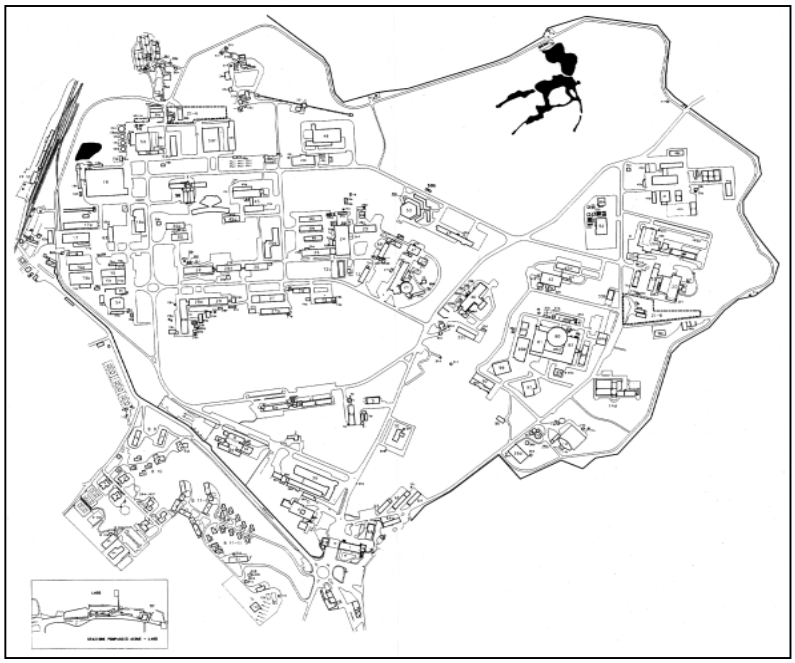


Figure B-4: Map of the Joint Research Centre

B.2 Results

The software developed for the affine registration of the reflectance and intensity images described in chapter 2 was used, without any modification, on these images.

For the first test (images in Figure B-1 and Figure B-2), the final superposition of the edges after the algorithm is presented in Figure B-5. Figure B-6 is the final result of the registration between the two satellite images.

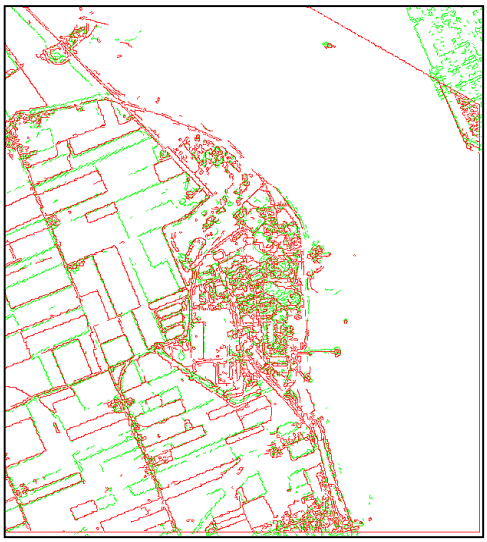


Figure B-5: The edge superposition after the affine resizing process with the first set of images

Figure B-6: Result of the registration

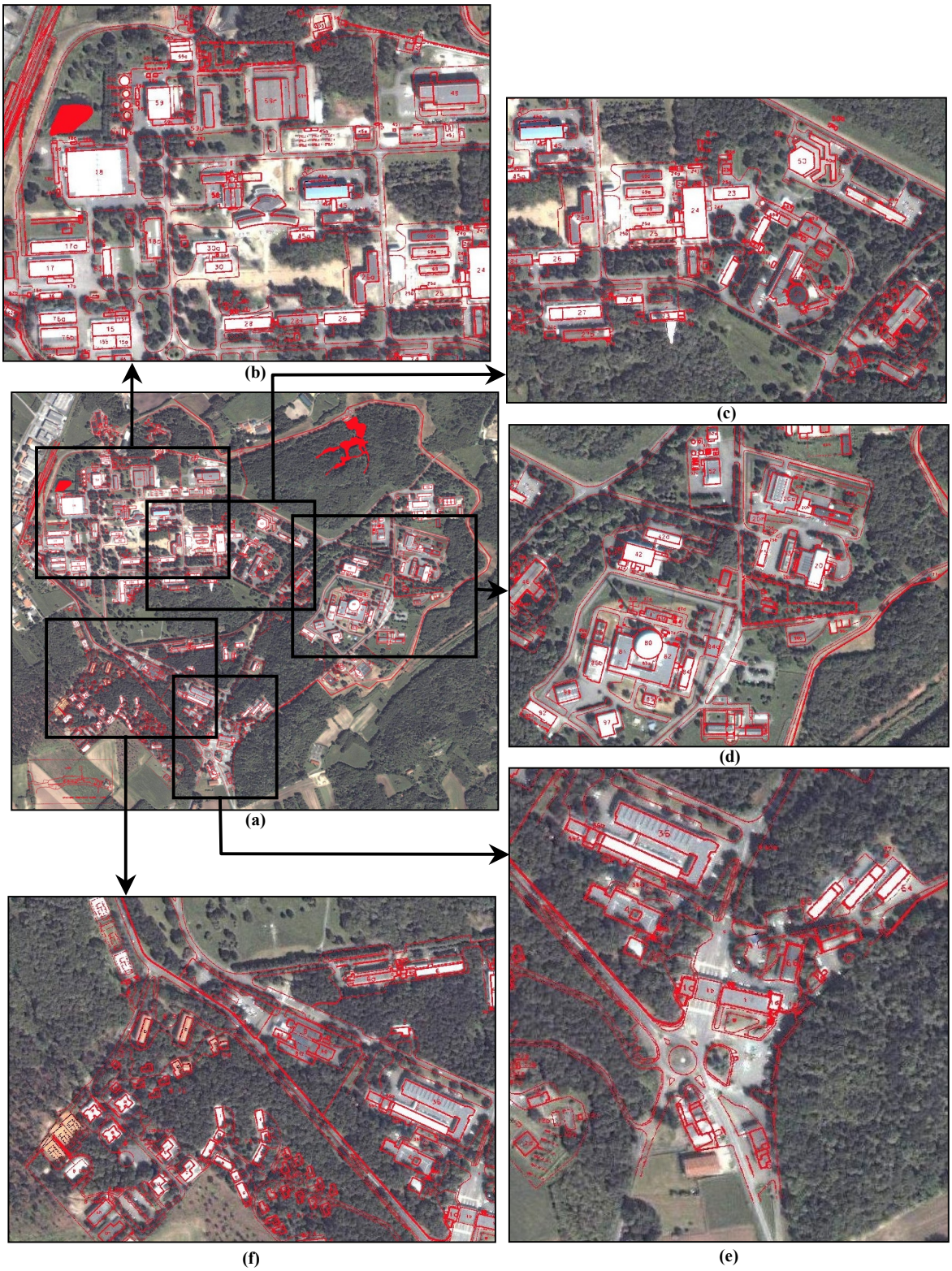


Figure B-7: Results with the second set of images

Figure B-7 presents the results of the registration of the JRC satellite image and map. Figure B-7(a) presents in the same image the superposition of the 2D map and the satellite image after the affine resizing process. Figure B-7(b to f) give the possibility to check the quality of the registration by zooming different areas of the image.

B.3 Discussion

These tests in a different area of application illustrate that the developed algorithm for affine resizing is valid for a variety of applications, since it does not depend on the sensors but is just based on the images. It is, then, of interest for other applications like the one presented here, and can help the automation of tasks that are often depending on a human operator choosing corresponding points manually. The main limitation is the fact that only the planar affine transformation is considered, meaning that additional deformations between the images are not considered. Yet the software is useful when other transformations are negligible compared to the planar transformation. In these situations our algorithm gives a fast and relatively good quality registration.

Although the software was not built to deal with satellite images and can be optimised to perform better on this particular type of images, still the results were comparable with the ones obtained with other methods typically used in these satellite applications (manual selection of correspondences, geo-referenced images, etc.).

Another valuable advantage of our software is the user-interface that allows correcting and/or guiding the automatic process and getting satisfying results even with difficult images (with different edges detected in both the images).

References

[Anadan89] P. Anadan,
A Computational Framework and an Algorithm for the Measurement of Visual Motion,
International Journal of Computer Vision, Vol. 2, No. 3, pp. 283-310, January 1989.

[Antone00] M. Antone, S. Teller,
Automatic Recovery of Relative Camera Rotations for Urban Scenes.
Proceedings of IEEE Conf on Computer Vision and Pattern Recognition 2000, Vol. II, pp. 282-289, June 2000.

[Antone02] M. Antone, S. Teller,
Scalable, Extrinsic Calibration of Omni-Directional Image Networks.
International Journal of Computer Vision, Vol. 49, No.2, pp. 143-174, Sept./Oct. 2002

[Atkinson96] K.B. Atkinson, editor,
Close Range photogrammetry and Machine vision.
Whittles publishing, 1996.

[Baillard99] C. Baillard, C. Schmid, A. Zisserman, A.W. Fitzgibbon,
Automatic Line Matching and 3D Reconstruction of Buildings from Multiple Views.
Proceedings of Conference on Automatic Extraction of GIS Objects from Digital Imagery, Technische Universittät München, Germany, pp. 69-80, September 1999

[Barnard83] S.T. Barnard
Interpreting perspective images.
Artificial intelligence 21, Vol. 21, No. 4, pp. 435-462, 1983.

[Becker95] S. Becker,V. M. Bove Jr,
Semiautomatic 3-D Model Extraction From Uncalibrated 2-D Camera Views.
SPIE Symposium on Electronic Imaging : Science & Technology, San Jose, February pp. 5-10, 1995.

[Becker96] S. Becker,
Vision-assisted modelling for model-based video representations,.
Ph.D. Thesis, MIT, March 1996.

[Bernardini02] F. Bernardini, H. Rushmeier, I.M. Martin, J. Mittleman, G Taubin,
Building a Digital Model of Michelangelos's Florentine Pietá.
IEEE Computer Graphics and Applications, Vol. 22, No. 1, pp. 59-67, Jan/Feb. 2002.

[Besl92] P.J. Besl, N.D. McKay,
A method for registration of 3D shapes.
IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 14, No 2, pp. 239-256 , February 1992.

[Birchfield99] S. Birchfield, C. Tomasi,
Depth discontinuities by pixel-to-pixel stereo,
International Journal on Computer Vision, No.3, pp. 1-25, December 1999.

[Boehler02] W. Boehler, A. Marbs,
3D Scanning Instruments.
CIPA, Heritage Documentation - International Workshop on Scanning for Cultural Heritage Recording, pp. 9-12, Corfu, Greece, September 2002.

[Borgefors84] G. Borgefors,
An improved Version of the Chamfer Matching Algorithm.
7th international conference on pattern recognition, pp. 1175-1177, 1984.

[Borgefors86a] G. Borgefors,
A new distance transformation aproximating the Euclidean distance.
8th international conference on pattern recognition, pp. 336-338, 1986.

[Borgefors86b] G Borgefors,
Distance Transformations in Digital.
Computer Vision, Graphics and Image Processing Vol 34, No. 3, pp. 344-371, June 1986.

[Brejl98] M. Brejl,
Comprehensive exam.
Department of electrical and Computer Engineering, University of Iowa, March 1998

[Brown92] L. G. Brown,
A Survey of Image Registration Techniques.
Surveys Vol. 24, No. 4, pp. 325-376, December 1992,

[Callidus] *Callidus Precision Systems,*
www.callidus.de.

[Camera] *CAd Modelling of Built Environments from Range Analysis*
EC funded TMR (Trans-Mobility of Researchers) network (ERB FMRX-CT97-0127)
http://www.dai.ed.ac.uk/homes/rbf/CAMERA/camera.html

[Canny86] J.F. Canny,
A Computacional Approach toEdge Detection.
IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 6
pp. 679-698, 1986

[Caprile90] B. Caprile, V. Torre,
Using Vanishing Points for Camera Calibration.
International Journal on Computer Vision, Vol. 4, No. 2, pp. 127-140, March 1990.

[Clarke96] J.C. Clarke, S. Carlsson, A. Zisserman,
Detecting and tracking linear features efficiently.
Proceedings of 7th British Machine Vision Conf., Edinburgh, pp. 415-424,1996.

[Coorg99] S. Coorg, S. Teller,
Extracting Textured Vertical Facades from Controlled Close-Range Imagery.
Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 1999, pp. 625-632, 1999.

[Cox96] I. Cox, S. Hingorani, and S. Rao,
A Maximum Likelihood Stereo Algorithm,
Computer Vision and Image Understanding, Vol. 63, No.3, pp. 542-567, May 1996

[Cyberware] *Cyberware*
http://www.cyberware.com.

[Cyra] *Cyra Technologies,*
www.cyra.com.

[Debevec01] P. Debevec,
Reconstructing and Augmenting Architecture with Image-Based Modelling, Rendering and Lighthing.
Proceedings of the International Symposium on Virtual Architecture (VAA'01), pp. 1-10, Dublin 21-22 June 2001.

[Deriche90] R. Deriche, G. Giraudon,
Accurate corner detection: An analytical study.
In Proc. 3rd Int. Conf. on Computer Vision, pp 66-70, 1990.

[Dias00] P. Dias, V. Sequeira , J.G.M. Gonçalves, F. Vaz
Automatic Registration of Laser Reflectance and Colour Intensity Images for 3D Reconstruction.
Proccedings of the 8th Symposium on Intelligent Robotic Systems, SIRS 2000. The University of Reading, England, pp. 191-197, July 2000.

[Dias01a] P. Dias, V. Sequeira , J.G.M. Gonçalves, F. Vaz,
Combining Intensity and Range Images for 3D Architectural modelling,
Proccedings of the Virtual and Augmented Architecture, VAA'01. Trinity College, Dublin, pp. 139-145, June 2001.

[Dias01b] P. Dias, V. Sequeira , J.G.M. Gonçalves, F. Vaz,
Fusion of Intensity and Range Data for Improved 3D Models.
Proccedings of the 8th International Conference on Image Processing, IEEE ICIP'01. Thessaloniki Greece, Vol III, pp. 1107-1110, October 2001.

[Dias02] P. Dias, V. Sequeira , J.G.M. Gonçalves, F. Vaz,
Automatic Registration of Laser Reflectance and Colour Intensity Images for 3D Reconstruction
Journal of Robotics and Autonomous Systems-Elsevier, Vol 39 (3-4), pp. 157-168, June 2002.

[Dick01] A.R. Dick, P.H. Torr, S.J. Ruffle, R. Cipolla,
Combining single view recognition and multiple view stereo for architectural scene.
8th IEEE International Conference on Computer Vision, pp. 268-274, July 2001.

[Eberly00] D.H. Eberly,
3D Game Engine Design: A Practical Approach to Real-Time Computer Graphic.
Morgan Kaufmann Publishers, September 2000.

[El Hakim95] S.F. El-Hakim, J.A. Beraldin, F. Blais,
A Comparative Evaluation of the Performance of Passive and Active 3-D Vision Systems.
SPIE Proc. 2646, St.Petersburg Conf. on Digital Photogrammetry, pp. 14-25, June 1995.

[El-Hakim98] S. F. El-Hakim, C. Brenner, G. Roth,
A multi-sensor approach to creating accurate virtual environments.
ISPRS Journal for Photogrammetry and Remote Sensing, Vol. 53, No. 6, pp. 379-391, December 1998.

[El-Hakim02] S. F. El-Hakim, J.A. Beraldin, M. Picard,
Detailed 3D reconstruction of monuments using multiple techniques.
Proceedings of the Intern. Workshop on Scanning for Cultural Heritage Recording - Complementing or Replacing Photogrammetry, pp.13-18, Greece. September 2002.

[Elstrom98] M.D. Elstrom,
A stereo-based technique for the registration of color and ladar images.
Master Degree thesis, University of Tennessee, Knoxville, August 1998.

[Falkenhagen94] L. Falkenhagen
Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuous Surfaces.
Paker, Y. and Wilbur, S. (Ed.), Image Processing for Broadcast and Video Production, Springer series on Workshops in Computing, pp. 115-127, Hamburg 1994.

[Faugeras93] O. Faugeras,
Three-Dimensional Computer Vision.
MIT Press, 1993.

[Faugeras98] O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Seller, C. Gauclin, I. Zoghlami,
3D reconstruction of urban scenes from image sequences.
Computer Vision and Image Understanding, Vol. 69, No. 3, pp. 292-309, 1998.

[Fischler81] M.A. Fischler, R.C Bolles,
Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.
Communications of the ACM, Vol. 24, No. 6, pp. 381-395, 1981

[Fitzgibbon98a] A.W. Fitzgibbon, A. Zisserman,
Automatic 3D Model Acquisition and Generation of New Images from Video Sequences.
In Proceedings of European Signal Processing conference (EUSIPCO '98), Rhodes, Greece, pages 1261-1269, 1998.

[Fitzgibbon98b] A.W. Fitzgibbon, A. Zisserman,
Automatic camera recovery for closed or open image sequences.
European Conference on Computer Vision (ECCV'98), Vol. 1, pp. 311-326, 1998.

[Fröhlich] *Zoller and Fröhlich GmbH*
http://www.zofre.de.

[Fröhlich98] C. Fröhlich C., M. Mettenleiter, F. Haertl,
Imaging Laser Radar for High-Speed Monitoring of the Environment.
Zoller&Fröhlich technical report 1998.

[Gonçalves00] J.G.M Gonçalves, V. Sequeira, E. Wolfart, P. Dias,
3D Reconstruction of Monuments 'as-built',
Proccedings of the 5th International Conference "Cultural Heritage Networks Hypermedia and MEDICI Framework Day", Milan, Italy, September 2000

[Gracias97] N. Gracias,
Application of Robust Estimation to Computer Vision: Video Mosaics and 3D-Reconstruction
Master Degree thesis, IST-Technical University of Lisbon, Portugal, 1997.

[Halcon02] Machine Vision technologies (MVTec)
Halcon User's Manual, December 2002
http://www.mvtec.com/halcon

[Hancock98] J. Hancock, D. Langer, M. Hebert, R. Sullivan, D. Ingimarson, E. Hoffman, M. Mettenleiter, C. Froehlich.
Active Laser Radar for High-Performance Measurements.
Proceedings of International Conference on Robotics and Automation (ICRA '98), Vol. 2, pp. 1465-1470, Belgium, May 1998.

[Haralick93] R.M. Haralick, L.G. Shapiro,
Computer and Robot Vision.
Addison-Wesley,1992 and 1993.

[Harris88] C.G. Harris, M. Stephens,
A combined corner and edge detector.
In 4th Alvey Vision Conference, pp 147-151, 1988.

[Hartley99] R. Hartley,
Theory and Practice of Projective Rectification,
International Journal of Computer Vision No. 35, Vol. 2, pp, 115-127, 1999.

[Hartley00] R.I. Hartley, A. Zisserman,
Multiple View Geometry in Computer Vision,
Cambridge University Press, June 2000

[Heath97] M. Heath, S. Sarkar, T. Sanocki, K.W. Bowyer,
A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms.
IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 12, pp. 1338-1359, December 1997.

[Heikkila97] J. Heikkila, O. Silven,
A Four-Step Camera Calibration Procedure with Implicit Image Correction.
In Proc. of IEEE Computer Vision and Pattern Recognition, pp. 1106-1112, 1997.

[Heipke96] C. Heipke,
Overview of image matching techniques.
OEEPE Workshop on the application of Digital Photogrammetric Workstations, 1996.

[Heuvel98] F.A. Van Den Heuvel,
Vanishing point detection for architectural photogrammetry.
International Archives of Photogrammetry and Remote Sensing, Vol. 22 part 5, pp. 652-659, 1998.

[iQsun] *iQsun GmbH, iQvolution AG,*
www.iqsun.com - www.iqvolution.com.

[Kitchen82] L. Kitchen, A. Rosenfeld,
Gray level corner detection.
Pattern recognition letters, 1, pp. 95-102, 1982

[Koch95] R. Koch
3D Surface Reconstruction from Stereoscopic Image Sequences,
International Conference on Computer Vison ICCV'95, Cambridge, MA, USA, pp. 109-114, 1995

[Kumar94] R. Kumar and A. Hanson.
Robust Methods for Estimating Pose and a Sensitivity Analysis.
CVGIP-Image Understanding, Vol. 60, No. 3, pp. 313-342, 1994.

[Kurazume02] R. Kurazume, K. Nishino, Z. Zhang, and K. Ikeuchi,
Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute.
Proc. of Fifth Asian Conference on Computer Vision (ACCV), Vol. I, pp. 99-106, January 2002.

[Lacey00] A.J. Lacey, N. Pinitkarn, N.A. Thacker
An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibration,
British Machine Vision Conference 2000, Univ. of Bristol, pp. 646-655, Sept. 2000.

[Lang95] F. Lang, W. Förstner,
Matching Techniques.
Proc. 2nd Course in Digital Photogrammetry, Bonn, Feb. 1995

[Levoy00] M. Levoy et S. Rusinkiewicz, M. Ginzton, J. Ginsberg, K. Pulli, D. Koller, S. Anderson, J. Shade, B. Curless, L. Pereira, J. Davis, D. Fulk,
The Digital Michelangelo Project: 3D scanning of Large Statues.
pp. 131-144, SIGGRAPH 00.

[Liebowitz99] D. Liebowitz, A. Criminisi, A. Zisserman
Creating Architectural Models from Images.
Eurographics, Vol. 18, No. 3, pp. 39-50, September1999.

[Luong94] Q.T. Luong, O.D. Faugeras,
The Fundamental Matrix: Theory, Algorithms, and Stability Analysis,
International Journal of Computer Vision, 17, pp. 43-75, 1996.

[Marr82] D. Marr,
Vision.
W.H. Freeman, 1982.

[Mehrotra90] R. Mehrotra, S. Nichani, N. Ranganathan,
Corner Detection.
Pattern Recognition, Vol. 23, No. 11, pp 1223-1233,1990

[Mensi] *Mensi,*
www.metricvision.com.

[Miyazaki00] D. Miyazaki, T. Ooishi, T. Nishikawa, R. Sagawa, K. Nishino, T. Tomomatsu, Y. Takase, K. Ikeuchi,
The Great Buddha Project: Modelling Cultural Heritage through Observation.
In Proc. of the Sixth International Conference on Virtual Systems and MultiMedia (VSMM 2000), pp.138-145, October 2000.

[Moravec79] H. P. Moravec,
Visual Mapping by a Robot Rover.
International Joint Conference on Artificial Intelligence, pp. 598-600, 1979.

[Neugebauer99] P. Neugebauer, K. Klein,
Texturing 3D Models of Real World objects from multiple unregistered photographic views.
Eurographics, Vol. 18, No. 3, pp. 245-256, September 1999.

[OpenCV] Intel Corporation,
Open Source Computer Vision library.
http://www.intel.com/research/mrl/research/opencv.

[Optech] *Optech,*
www.optech.on.ca.

[Oshima79] M. Oshima Y. Sarai,
A Scene Description Method Using Three-Dimensional Information.
Pattern recognition, Vol. 11, pp. 9-17,1979.

[Partington01] M. Partington, C. Jaynes
A Hybrid Image and Model Based Approach to Photorealistic Building Reconstruction.
Proceedings of Virtual ang Augmented Architecture (VAA'01), pp.243-254, 2001.

[Pollefeys99] M. Pollefeys,
Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences.
Ph.D. Thesis, Dept. of Electrical, Katholieke Universiteit Leuven, 1999.

[Pollefeys00] M. Pollefeys,
3D Modelling from Images,
Tutorial notes, in conjunction with ECCV 2000, Dublin, Ireland, June 2000.

[Qin-Zhong84] Q.Z. Ye,
The signed Euclidean distance Transform and its applications.
9th international conference on pattern recognition, pp. 495-499, Rome, 1988.

[Riegl] *Riegl Laser Measurement Systems*
http://www.riegl.com.

[Riegl99] RIEGL GmbH,
Laser Mirror Scanner LMS-Z210- Technical document & user's instruction manual.
November 1999.

[Ring97] J. Ring, B. Chargin, N. Machotka,
Photovista User´s Guide.
1996-1997 Live pictures, Inc.

[Rother00] C. Rother,
A New Approach for Vanishing Point Detection in Architectural Environments.
British Machine Vision Conference, pp. 382-391, 2000.

[Rousseeuw87] P.J. Rousseeuw, A.M. Leroy,
Robust Regression and outlier detection.
John Wiley & Sons New York, 1987.

[Ruggeri01] M. Ruggeri, P. Dias, V. Sequeira, J.G.M. Gonçalves,
Interactive tools for quality enhancement in 3D modelling from reality.
In Proceedings of the 9th Symposium on Intelligent Robotics Systems, SIRS 2001, pp. 157-165, LAAS-CNRS, Toulouse France, 18-20 July 2001.

[Rushmeier99] H. Rushmeier, F. Bernardini.
Computing consistent normals and colors from photometric data.
Second Intl. Conf. on 3D Digital Imaging and Modelling - 3DIM'99, pp. 99-108, Ottawa, Canada, October 4-8, 1999.

[Salvi01] J. Salvi, X. Armangué, J. Pagès,
A Survey Addressing the Fundamental Matrix Estimation Problem.
Proceedings of IEEE International Conference an Image processing, Vol II, pp,.209-212, October 2001.

[Schmid97] C. Schmid, A. Zisserman,
Automatic Line Matching Across Views.
Proceedings of IEEE Conference on Computer Vision and pattern recognition, pp. 666-671, 1997.

[Schmid98] C. Schmid, R. Mohr, C. Bauckhage,
Comparing and evaluating interesting points.
International Conference on Computer Vision, pp 230-235, 1998.

[Schneider03] P.J. Schneider,D.H. Eberly,
Geometric Tools for Computer Graphics.
Morgan Kaufmann Publishers, September 2003.

[Sequeira96] V. Sequeira,
Active Range Sensing for Three-Dimensional Environment Reconstruction.
PhD Thesis, Dept. of Electrical and Computer Engineering, IST-Technical University of Lisbon, Portugal, 1996.

[Sequeira99] V. Sequeira, K. Ng, E. Wolfart, J.G.M. Gonçalves, D. Hogg,
Automated Reconstruction of 3D Models from Real Environments.
ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier), Vol. 54, pp. 1-22, 1999.

[Sequeira00] V. Sequeira, E. Wolfart, J.G.M. Gonçalves, P. Dias,
The Third Dimension, BBC-1 Tomorrow's World Special Programme on Venice broadcast in the UK on 15.Nov.2000.
The 4'30" programme presents the 3D Reconstructed model of the "Sala dello Scrutinio" at Venice Doges' Palace, November 2000.

[Shufelt99] J.A. Shufelt,
Performance Evaluation and Analysis of Vanishing Point Detection Techniques,
IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 21, No. 3, pp. 282-288, March 99.

[Smith98] P. Smith, D. Sinclair, R, Cipolla, K. Wood,
Effective Corner Matching.
British Machine Vision Conference, 1998.

[Smith99] P. W. Smith, M. D. Elstrom,
Stereo Based Registration of Multi-Sensor Imagery for Enhanced Visualization of Remote Environments.
Proceedings of the IEEE International Conference on Robotics and Automation, pp. 1948-1953, May 1999.

[Stamos00] I. Stamos, P. K. Allen,
3-D Model Construction Using Range and Image Data.
IEEE International Conference on Computer Vision and Pattern Recognition, Vol. I, pp. 531-536, South Carolina, June 2000.

[Stamos01a] I. Stamos, P. K. Allen,
Automatic Registration of 2-D with 3-D Imagery in Urban Environments.
International Conference on Computer Vision, Vol. II, pp 731-736, 2001.

[Stamos01b] I. Stamos,
Geometry and Texture Recovery of Scenes of Large Scale: Integration of Range and Intensity Sensing.
PhD Thesis, Graduate School of Arts and Science, Columbia University, 2001.

[Sun97] C.Sun,
A Fast Stereo Matching Method.
Digital Image Computing: Techniques and Applications, Massey University, Auckland, New Zealand, pp.95-100, 10-12 December 1997.

[Sun99] C. Sun,
Multi-Resolution Stereo Matching Using Maximum-Surface Techniques,
Digital Image Computing: Techniques and Applications, Perth, Australia, pp.195-200, 7-8 December 1999.

[Sun02] C. Sun
Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques.
International Journal of Computer Vision, Vol 47, No.1-3, pp. 99-117, May 2002

[Torr93] Torr, P.H.S., Murray, D.W.,
Outlier Detection and Motion Segmentation,
Proceedings of SPIE Sensor Fusion Conference VI, Vol. 2059, pp. 432-443, September 1993.

[Torr95] P. Torr,
Motion segmentation and outlier detection,
PhD thesis, Dept. of Engineering Science, University of Oxford, 1995.

[Torr97] *Torr, P.H.S.*, *Murray, D.W.*,
The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix.
International Journal of Computer Vision, Vol. 24, No. 3, pp. 271-300, September October 1997.

[Torr98] Torr, P.H.S., Zisserman, A. and Maybank, S.
Robust Detection of Degenerate Configurations for the Fundamental Matrix
In Computer Vision and Image Understanding, vol. 71, no. 3, pp. 312-333, September, 1998.

[Tsai86] R.Y. Tsai,
An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision.
Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 364-374, 1986.

[Tsai87] R.Y. Tsai,
Metrology Using Off-the-Shelf TV Cameras and Lenses.
IEEE Journal of Robotics and Automation, Vol. 3, No. 4, pp. 323-344, August 1987.

[Tuytelaars98] T. Tuytelaars, L. Van Gool, M. Proesmans, T. Moons,
The cascaded Hough Transform as an Aid in Aerial Image Interpretation.
Proc. of 6th IEEE Conference on Computer Vision, pp. 67-72, India, January 1998.

[Walker91] M. W. Walker, L. Shao, R. A. Volz,
Estimating 3-D Location Parameters Using Dual Number Quaternions.
Computer Vision Graphics and Image Processing, Vol. 54, No. 3, pp. 358-367, November 1991.

[Werner02] T. Werner, A. Zisserman,
New technique for automated architectural reconstruction from Photographs.
Proc. 7th European Conference Computer Vision, Vol II, pp. 541-555, Copenhagen, May 2002.

[Wilson94] Reg G. Willson,
Modeling and Calibration of Automated Zoom Lenses.
Ph.D. thesis, Department of Electrical and Computer Engineering,
Carnegie Mellon University, January 1994.

[Zhang00] Z. Zhang.
A flexible new technique for camera calibration.
IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, pp. 1330-1334, 2000.

[Zhao01] H. Zhao, R. Shibasaki,
Reconstructing Textured CAD Model of Urban Environment using Vehicle-borne Laser Range Scanners and Line Cameras.
Proceedings of second international workshop on Computer Vision Systems (ICVS2001), pp. 284-297, Vancouver, Canada, July 7-8, 2001