Chapter 1
Introduction

1.1 Objectives

The aim of this book is to provide a complete framework for efficient 3D modeling. More specifically, given an image sequence of a scene, the objective is to provide a high-precision textured 3D reconstruction of the scene with virtually no human intervention.

The main focus of this book is on efficient modeling of 3D underwater scenes for scientific studies. Nevertheless, as shown in this work, we have successfully applied the technique in other areas of interest such as the reconstruction of small scale objects, outdoor natural scenes, urban environments, etc.

Although some successful 3D reconstruction algorithms have been reported in literature, they are limited to specific applications. Most techniques assume controlled or structured environments, where illumination, camera motion and scene geometry priors can be used. More importantly, these techniques can be applied to very limited scenes only, due to the complexity of the 3D reconstruction problem.

In contrast, we aim to develop an online generic framework for 3D scene reconstruction that can cope with wide areas of complex and highly unstructured environments. In order to achieve this, we focused on the following aspects:

Online process. The entire framework has been designed to process the data sequentially, enabling its use on online applications such as robot navigation and mapping.

Flexibility of acquisition. The 3D reconstruction algorithm uses image sequences that can be acquired by using any type of video/still cameras, with no constraints on the acquisition process. Moreover, the framework can readily cope with camera occlusions and temporary failures.

Stand-alone framework. While additional information can be integrated into the 3D reconstruction process, the framework does not require any

T. Nicosevici & R. Garcia: *Efficient 3D Scene Modeling and Mosaicing*, STAR 87, pp. 1–11.
DOI: 10.1007/978-3-642-36418-1_1

additional sensor information. This increases the flexibility of the reconstruction process while decreasing the acquisition costs. In this way, underwater sequences can, for example, be acquired by using cameras mounted on inexpensive Remotely Operated Vehicles (ROVs) or even by divers using hand-held cameras. On the other hand, without absolute positioning sensors, vision systems are prone to drifting. We address this shortcoming by proposing a novel online cross-over detection system, that allows the detection of loops in camera trajectory along with camera pose[1] correction.

Efficient 3D modeling. The framework employs a novel online 3D model simplification algorithm, that allows mapping of larger, more complex scenes.

1.2 Motivation

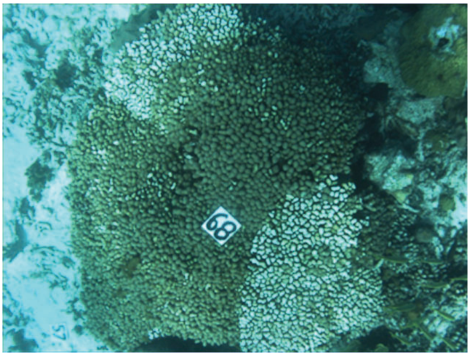
Throughout the history of the Earth, the most determinant element, that shaped it as we know it today, are the oceans. They are the origin of life on Earth and home of the widest biodiversity. Moreover, the oceans are the major factor in our climate, literally affecting almost every aspect of our daily lives.

Apparently paradoxical, the oceans represent the least studied region of the Earth's surface. The main reason behind this is the inaccessibility and hostility of this environment. This, however, is changing at a rapid pace. Our urge to find alternative food and energy sources, to understand climate changes and geological phenomena have determined the scientists to multiply their efforts into understanding this complex environment. Moreover, the latest technological advances provide the scientists with the basis for more efficient means to explore the underwater environment.

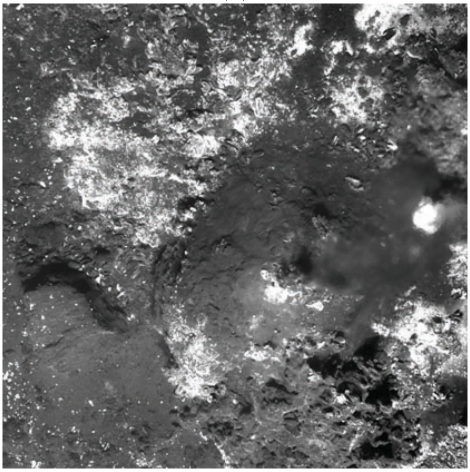
In this context, this work proposes a valuable tool for remote underwater studies. Images acquired by scuba divers using hand-held cameras can be used to obtain high detail textured 3D models of the seafloor. Using cameras mounted on Unmanned Underwater Vehicle (UUV) we can obtain 3D maps of high depth underwater regions that otherwise would be inaccessible to humans. Just to name a few, this proposal has applicability in (see Figure 1.1 for some examples of scientific oriented underwater imagery):

- **Biology.** Visual 3D maps of marine habitats provide important clues in studying the marine species and their interaction with the environment.
- **Ecology.** The impact of human activities on our environment has become a matter of great concern nowadays. Climate changes, intensive fishing and the destruction of habitats greatly affect the underwater biodiversity. In this context, 3D models can be used to observe and monitor the changes that take place in the underwater habitats, such as coral reefs.

[1] Camera position and attitude (orientation).



(a)



(b)

(c)

Fig. 1.1 Motivation – Scientific underwater imagery. (a) a coral reef head near Bahamas; (b) underwater lava formations captured during MoMARETO'06 cruise – courtesy of IFREMER and (c) amphoras near Pianosa Island in the Mediterranean Sea – courtesy of Venus Project.

- **Geology.** Shape and texture of the regions with increased geological activity greatly aid geologists in order to understand the complex geological phenomena that take place underwater.
- **Archeology.** In our pursuit to understand history, we continuously search the depths of the oceans for new clues about our past. Unfortunately, in most cases the artifacts are too fragile or too inaccessible to be recovered for studying. In this case, 3D models can provide a viable solution for remote archeological studies.

Nowadays, an increasing number of underwater studies employ ROVs as an alternative to scuba divers. This eliminates the risks the divers are exposed to, especially in deep waters, while allowing more efficient studies. However, the use of ROVs poses a series of drawbacks: their operation requires specialized personnel and their range and depth is limited by the length of their umbilical cable, which connects the ROV with the ship. Various research groups have focused their efforts on developing underwater vehicles that would carry out missions autonomously. This requires that the vehicles be able to model the environment in order to navigate through it. The images acquired by cameras mounted on these vehicles can be processed and 3D maps of the environment can be obtained. Furthermore, the obtained maps can be used for navigation in subsequent missions where, for example, successive surveys of the same area are needed.

With the wide accessibility of high computational power, the use of wide area 3D modeling has become an area of interest in fields much closer to the end-user:

- **Urban 3D modeling.** Applications such as *Google Earth* [58] (see Figure 1.2) or certain navigation applications such as *iGo* [75] offer 3D models of urban landmarks. The process of constructing the 3D models could be highly simplified by using automated 3D modeling techniques. Additionally, one could imagine applications where tourists would able to obtain 3D models along with information about the landmarks they are visiting.
- **Architecture.** Indoor / outdoor 3D models of buildings can be obtained for virtual marketing purposes. Also, using augmented reality, one could visualize beforehand the results of the restoration of a historical building for example, etc.
- **Virtual reality.** Computer games and virtual community applications such as *Second Life* [158] could be enriched with 3D models of real-life objects and buildings.

1.3 Challenges

The human brain interprets visual information provided by the eyes by generating 3D images of our surroundings. We use this information in order to

Fig. 1.2 Motivation – Urban architecture. Google Earth view of downtown Miami depicting 3D models of the most iconic buildings.

orient ourselves and move through the environment. The flexibility and power of abstraction of the brain allows it to easily cope with constant challenges in our environment such as moving objects, lighting artifacts and so on. When it comes to computer vision however, things become ever more difficult. For this, in order to achieve flexibility and robustness, we need to address a series of challenges:

- Unstructured, natural 3D scenes consist of large amounts of objects with diverse shapes and textures. As the camera moves through the scene, objects constantly occlude each other (see Figure 1.3).
- Light changes (*e.g.* motion of the light source), moving shadows, altering of light reflections in specular surfaces due to point of view changes drastically modify the photometric properties of the scene. This effect is particularly emphasized in underwater scenes, where sun flicker (changes in light pattern due to sunlight being refracted on moving sea surface) dramatically changes the illumination field (refer to Figure 1.4 for details).
- Moving objects such as cars and pedestrians in urban environments or fishes and algaes in underwater environments (see Figure 1.5) violate the rigid scene assumption, inducing errors in scene geometry estimation.

All these are common challenges faced by computer vision systems. However, the underwater environment poses specific challenges that make underwater imagery a particularly difficult task (Figure 1.6):

- In water, light suffers a much higher rate of attenuation than in atmospheric conditions. This limits the maximum distance between the camera and the scene, resulting in a narrow coverage of the camera. In order to

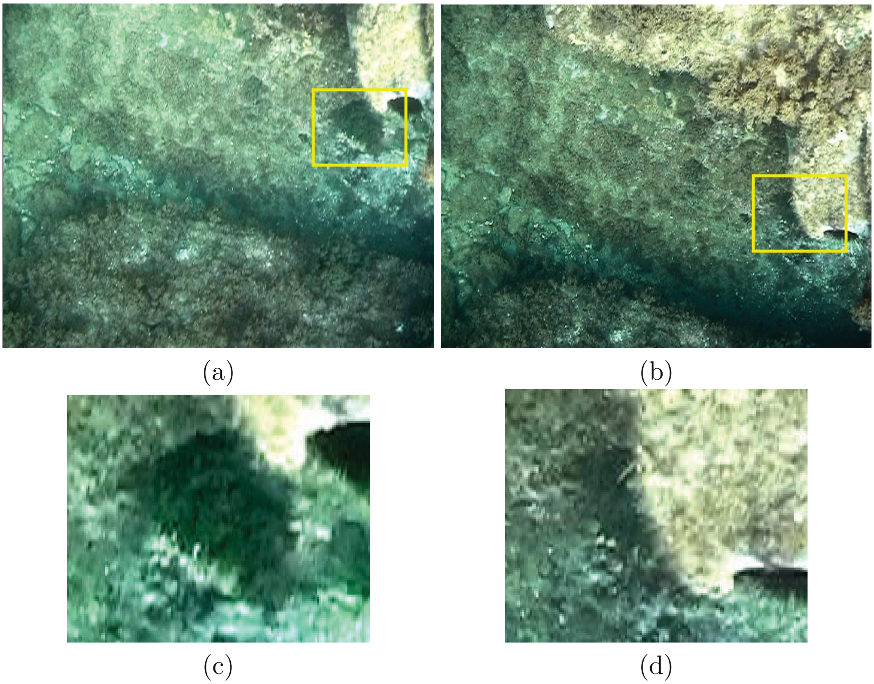
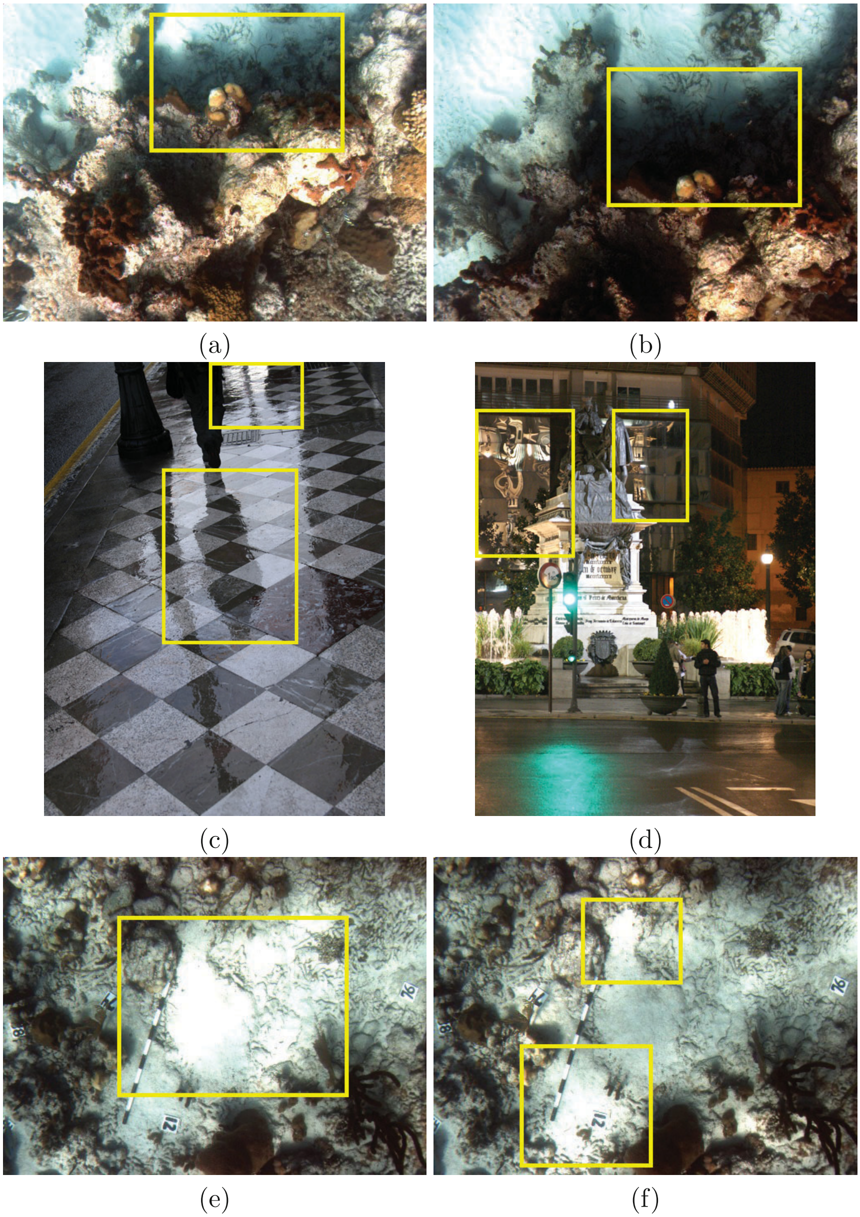


Fig. 1.3 Challenges – Scene occlusions. (a) and (b) show the same underwater scene from two different camera view points; The region marked by the yellow rectangles is shown in detail in (c) and (d). Note the rock is clearly visible in (a) and (c) but almost occluded in (b) and (d).

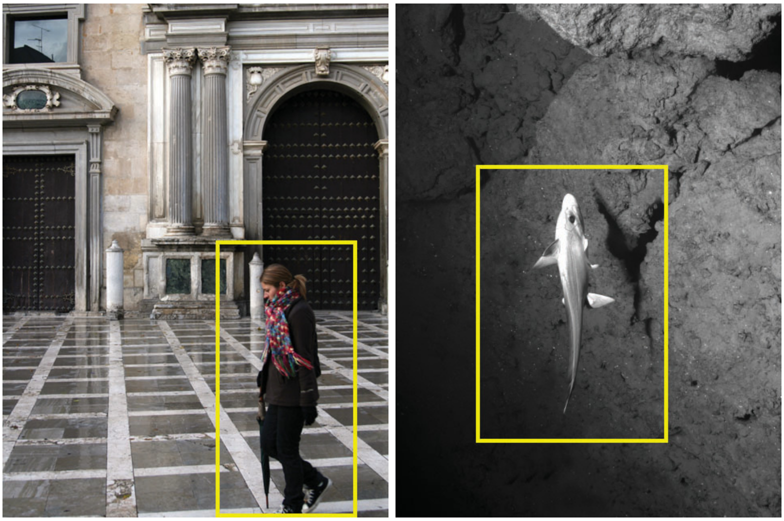
cover wide scene areas, large amounts of images have to be merged. Furthermore, due to light attenuation, at great depths, additional illumination sources have to be used. Generally, underwater vision systems employ focus lights as the latter can illuminate the scene at greater ranges. The drawback of the focus lights, however, is that they induce highly non-uniform lighting fields [50] (refer to Figure 1.6a for details).

- The contrast of underwater images is reduced due to light absorbtion, decreasing the signal-to-noise ratio (see Figure 1.6b).
- Small suspended particles present in the sea water such as plankton and sediments generate the so called *scattering effect*. Practically, the scattering effect takes place due to the light changing direction when it enters in contact with the particles. The forward scattering bends the light beams traveling from the scene towards the camera, resulting in a blurring effect that reduces the level of detail of the images (Figure 1.6c). On the other hand, the backward scattering refracts the light from the light source towards the camera decreasing the contrast and inducing noise in the images (Figure 1.6d).



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

Fig. 1.4 Challenges – Light artifacts. (a) and (b) illustrate an underwater scene a few frames apart. The details of the outlined region are visible in (a) while not distinguishable in (b) due to shadowing. Specular surfaces such as wet pavement (c) and windows (d) are highly reflective, inducing lighting artifacts. (e) and (f) show two frames of an underwater scene taken only $150ms$ apart. The light pattern changes drastically due to sun flickering.



(a) (b)

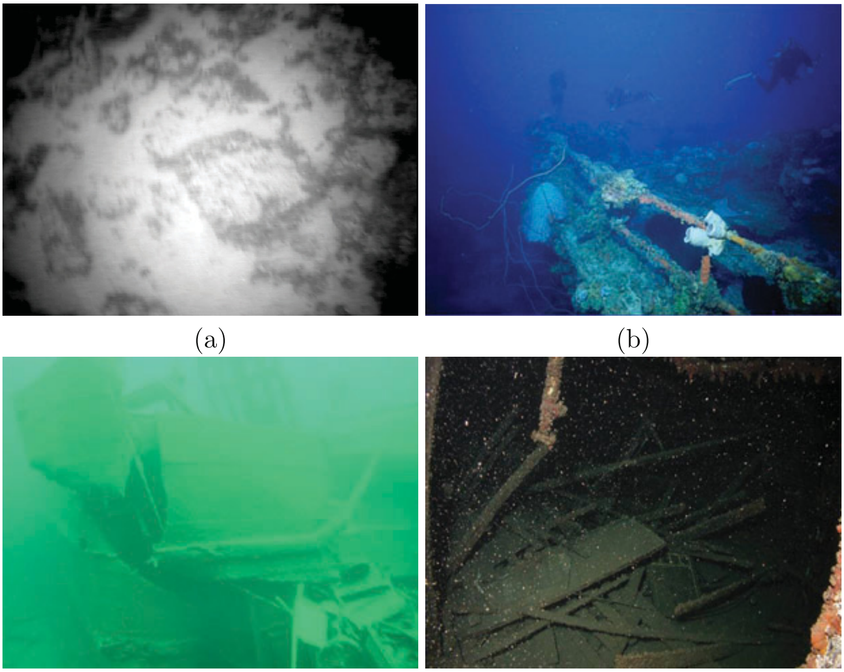
Fig. 1.5 Challenges – Moving objects. Moving objects such as pedestrians in urban scenes (a) and fishes in underwater scenes (b) violate the rigid scene assumption.

In addition to environment challenges, vision systems pose specific fundamental problems:

- Our aim is to develop a 3D modeling system where the camera can move freely through the scene. In order to maintain the generic character of the proposal, we assume that no external sensor information is used. In this case, scene geometry and camera motion are computed incrementally. Over wide scenes, small errors in the camera pose and scene geometry estimation build up over time. This error build-up can lead to estimations that drift away from the reality.
- Scene models are composed of 3D vertices. These vertices can be seen as discrete samples of the surfaces that are present in the scene. Over wide scene areas, millions of such vertices can be generated. Such large number of scene samples can prove too large to be effectively managed by vision systems.

1.4 Contributions

This book has contributions at different levels, briefly described hereafter. A more extensive account of the contributions is presented in Chapter 6.



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

Fig. 1.6 Challenges – Underwater environment. (a) the use of additional illumination sources induce non-uniform illumination fields; (b) light attenuation limits the range of the camera: the two divers in the background are hardly visible; (c) forward scattering blurs and decreases the contrast of underwater images and (d) backward scattering induces image noise.

1.4.1 Structure from Motion

We develop a novel Structure from Motion (SfM) algorithm, where the scene model is generated using a two-step approach: (*i*) camera pose is directly obtained from the scene model and (*ii*) using the camera pose, the scene model is updated and extended. This approach reduces the accumulation of error and results in more accurate scene models.

Also, we propose a novel dual camera pose recovery method, which allows SfM algorithm to successfully cope with both planar and non-planar scenes.

1.4.2 Ortho-mosaic and Rendering

We propose a novel approach to generate synthetic 2D visual maps – *ortho-mosaics*. By exploiting the geometry of the scene, the approach takes into account surface normals and camera poses in order to assure maximum resolution and minimum distortions during the ortho-mosaic rendering. This method results in accurate and visually pleasant scene maps.

1.4.3 Loop Closure Detection

Loop closures are situations where the camera revisits an already surveyed region. These regions allow us to impose additional constraints in the geometry of the scene, hence reducing the 3D estimation errors. This work proposes an online loop closing detection algorithm that uses Bag of Words (BoW) to measure visual similarities among camera frames. There are three main novelties that we propose here: (*i*) the visual vocabularies are built incrementally, enabling the use of the algorithm for online applications; (*ii*) the algorithm requires no training stage and no user intervention, and (*iii*) the feature clustering process uses a global data distribution criteria, resulting in more efficient visual vocabularies.

1.4.4 Vertex Selection

Scene models are formed from thousands to millions of 3D vertices. Most of these vertices are geometrically redundant (4 or more vertices laying on the same plane, 3 or more vertices laying on the same straight edge, etc.). We propose an online approach which analyzes the geometry of the scene and selects only those vertices that are geometrically representative for the scene. The method uses plane-parallax techniques that allow us to approximate the shape of the scene without explicitly recovering its geometry. In this way, feature selection can be carried out sequentially, as the scene model is being built.

The result is a 3D scene model with drastically reduced complexity that, at the same time, maintains the accuracy of the original model.

1.5 Book Outline

In Chapter 2 we review previous work on image registration techniques, visual feature extraction and matching, mosaicing and 3D reconstruction techniques. The review details those aspects of literature that are relevant to this book. Modern visual feature extractors and descriptors are described thoroughly as they constitute the basis for the proposed 3D reconstruction framework. Also, other 3D reconstruction algorithms are discussed along with their limitations, illustrating the motivation behind this work.

Chapter 3 presents the proposed 3D reconstruction algorithm. The first part of the chapter provides a detailed description of each step of the algorithm. The algorithm is validated through a series of experiments presented in the last part of the chapter. In here, we discuss various experiments that we have carried out in underwater, natural and structured environments.

In Chapter 4, we present the cross-over detection algorithm. The algorithm is built on top of the 3D reconstruction process and allows online detection of loops in the camera trajectory. In the first part of the chapter, we provide a review of the literature regrading the cross-over detection problem. Next, a detailed description of the proposed algorithm is provided. Finally, we present a series of experiments along with a comparison with a state of the art loop closure detection algorithm.

Chapter 5 discusses the online model simplification algorithm, that works in parallel with the 3D reconstruction algorithm. First, we discuss the existing work related to 3D model simplification, followed by a detailed description of our simplification algorithm. The chapter concludes with a series of experimental results and comparison with a widely used model simplification algorithm.

Finally, Chapter 6 summarizes the contributions of the book and discusses ongoing and future work. This chapter also presents the publications of the author, that are most significant to the development of this book.