We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

1

Image-Laser Fusion for *In Situ* 3D Modeling of Complex Environments: A 4D Panoramic-Driven Approach

Daniela Craciun[1,2], Nicolas Paparoditis[2] and Francis Schmitt[1]
[1]*Telecom ParisTech CNRS URA 820 - TSI Dept.*
[2]*Institut Geographique National - Laboratoire MATIS*
France

1. Introduction

One might wonder what can be gained from the image-laser fusion and in which measure such a hybrid system can generate automatically complete and photorealist 3D models of difficult to access and unstructured underground environments.

Our research work is focused on developing a vision-based system aimed at automatically generating in-situ photorealist 3D models in previously unknown and unstructured underground environments from image and laser data. In particular we are interested in modeling underground prehistoric caves. In such environments, special attention must be given to the main issue standing behind the automation of the 3D modeling pipeline which is represented by the capacity to match reliably image and laser data in GPS-denied and feature-less areas. In addition, time and in-situ access constraints require fast and automatic procedures for in-situ data acquisition, processing and interpretation in order to allow for in-situ verification of the 3D scene model completeness. Finally, the currently generated 3D model represents the only available information providing situational awareness based on which autonomous behavior must be built in order to enable the system to act intelligently on-the-fly and explore the environment to ensure the 3D scene model completeness.

This chapter evaluates the potential of a hybrid image-laser system for generating in-situ complete and photorealist 3D models of challenging environments, while minimizing human operator intervention. The presented research focuses on two main aspects: (i) the automation of the 3D modeling pipeline, targeting the automatic data matching in feature-less and GPS-denied areas for in-situ world modeling and (ii) the exploitation of the generated 3D models along with visual servoing procedures to ensure automatically the 3D scene model completeness.

We start this chapter by motivating the jointly use of laser and image data and by listing the main key issues which need to be addressed when aiming to supply automatic photorealist 3D modeling tasks while coping with time and in-situ access constraints. The next four sections are dedicated to a gradual description of the 3D modeling system in which we project the proposed image-laser solutions designed to be embedded onboard mobile plateforms, providing them with world modeling capabilities and thus visual perception. This is an

important aspect for the in-situ modeling process, allowing the system to be aware and to act intelligently on-the-fly in order to explore and digitize the entire site. For this reason, we introduce it as ARTVISYS, the acronym for ARTificial VIsion-based SYStem.

2. The in-situ 3D modeling problem

The *in-situ* 3D modeling problem is concerned with the automatic environment sensing through the use of active (laser) and/or passive (cameras) 3D vision and aims at generating in-situ the *complete* 3D scene model in a step by step fashion. At each step, the currently generated 3D scene model must be exploited along with visual servoing procedures in order to guide the system to act intelligently on-the-fly to ensure in-situ the 3D scene model completeness.

Systems embedding active 3D vision are suitable for generating in-situ *complete* 3D models of previously unknown and high-risk environments. Such systems rely on visual-based environment perception provided by a sequentially generated 3D scene representation. Onboard 3D scene representation for navigation purposes was pioneered by Moravec's back in the 1980s (Moravec, 1980). Since then, Computer Vision and Robotics research communities have intensively focused their efforts to provide vision-based autonomous behavior to unmanned systems, special attention being given to the vision-based autonomous navigation problem. In (Nister et al., 2004), Nister demonstrated the feasibility of a purely vision-based odometry system, showing that an alternative for localization in GPS-denied areas can rely on artificial vision basis. Several research works introduced either 2D and 3D Simultaneous Localization and Mapping (SLAM) algorithms using single-camera or stereo vision frameworks (Durrant-White & Bailey, 2006), (Bailey & Durrant-White, 2006). While gaining in maturity, these techniques rely on radiometric and geometric features' existence or exploit initial guess provided by navigation sensors (GPS, IMUs, magnetic compasses) employed along with dead-reckoning procedures.

Scientists from Robotics, Computer Vision and Graphics research communities were introducing the 3D modeling pipeline (Beraldin & Cournoyer, 1997) aiming to obtain photorealist digital 3D models through the use of 3D laser scanners and/or cameras. Various 3D modeling systems have been developed promoting a wide range of applications: cultural heritage (Levoy et al., 2000), (Ikeuchi et al., 2007),(Banno et al., 2008), 3D modeling of urban scenes (Stamos et al., 2008), modeling from real world scenes (Huber, 2002), natural terrain mapping and underground mine mapping (Huber & Vandapel, 2003), (Nuchter et al., 2004), (Thrun et al., 2006).

Without loss of generality, the 3D modeling pipeline requires automatic procedures for data acquisition, processing and 3D scene model rendering. Due to the sensors' limited field of view and occlusions, multiple data from various viewpoints need to be acquired, aligned and merged in a global coordinate system in order to provide a complete and photorealist 3D scene model rendering. As for SLAM techniques, the main drawback standing behind the automation of the entire 3D modeling process is the *data alignment* step for which several methods have been introduced.

For systems focusing on 3D modeling of large-scale objects or monuments (Levoy et al., 2000), (Ikeuchi et al., 2007),(Banno et al., 2008) a crude alignment is performed by an operator off-line. Then the coarse alignment is refined via iterative techniques (Besl & McKay, 1992). However, during the post-processing step it is often observed that the 3D scene model is

incomplete. Although data alignment using artificial markers produces accurate results, it cannot be applied to high-risk environments due to time and in-situ access constraints. In addition, for cultural heritage applications, placing artificial landmarks within the scene causes damages to the heritage hosted by the site. The critical need for an in-situ 3D modeling procedure is emphasized by the operator's difficulty to access too small and too dangerous areas for placing artificial landmarks and by the need to validate in-situ the 3D scene model completeness in order to avoid to return on site to complete data collection.

Existing automatic data alignment methods perform *coarse alignment* by exploiting prior knowledge over the scene's content (Stamos et al., 2008) (i.e. radiometric or geometric features' existence, regular terrain to navigate with minimal perception) or the possibility to rely on navigation sensors (GPS, INS, odometry, etc.). In a second step, a *fine alignment* is performed via iterative methods.

Since in our research work the environment is previously unknown, features' existence cannot be guaranteed. In addition, in underground environments and uneven terrain navigation sensors are not reliable and dead-reckoning techniques lead to unbounded error growth for large-scale sceneries. A notable approach reported by Johnson (Johnson, 1997) and improved by Huber (Huber, 2002) overcomes the need of odometry using shape descriptors for 3D point matching. However, shape descriptors' computation requires dense 3D scans, leading to time consuming acquisition and processing, which does not cope with time and in-situ access constraints.

A main part of this chapter focuses on providing image-laser solutions for addressing the automation of the 3D modeling pipeline by solving the data alignment problem in feature-less and GPS-denied areas. In a second phase, we propose to exploit the world modeling capability along with visual servoing procedures in order to ensure in-situ the 3D scene model completeness.

3. Proposed solution: Automatic 3D modeling through 4D mosaic views

This section resumes how we solve for the automation of the 3D modeling pipeline through the use of 4D mosaics. We start by introducing the hardware design and by summarizing the 4D-mosaicing process. In order to ensure in-situ the 3D scene model completeness, Section 3.3 proposes a 4D-mosaic driven acquisition scenario having as main scope the automatic digitization and exploration of the site.

3.1 Testbed. Hard- and soft-ware architecture

We designed a *dual* system for performing in-situ 3D modeling tasks in large-scale, complex and difficult to access underground environments. Since in such environments navigation sensors are not reliable, the proposed system embeds only 2D and 3D vision sensors, unifying photorealism and high resolution geometry into *4D mosaic views*. Figure 1 illustrates the ARTVISYS's hardware along with the proposed *4D-mosaicing* process. We describe hereafter several ARTVISYS's features and justify the proposed design.

RACL[1] dual-system (Craciun, 2010). The proposed hardware architecture falls in the category of the RACL dual sensing devices, embedding a high-resolution color camera mounted on a motorized pan-tilt unit and a 3D laser-range-finder, which are depicted in Figures 1 a) and b), respectively. There are several reasons for choosing a RACL design:

[1] RACL system: Rigidly Attached Camera Laser system

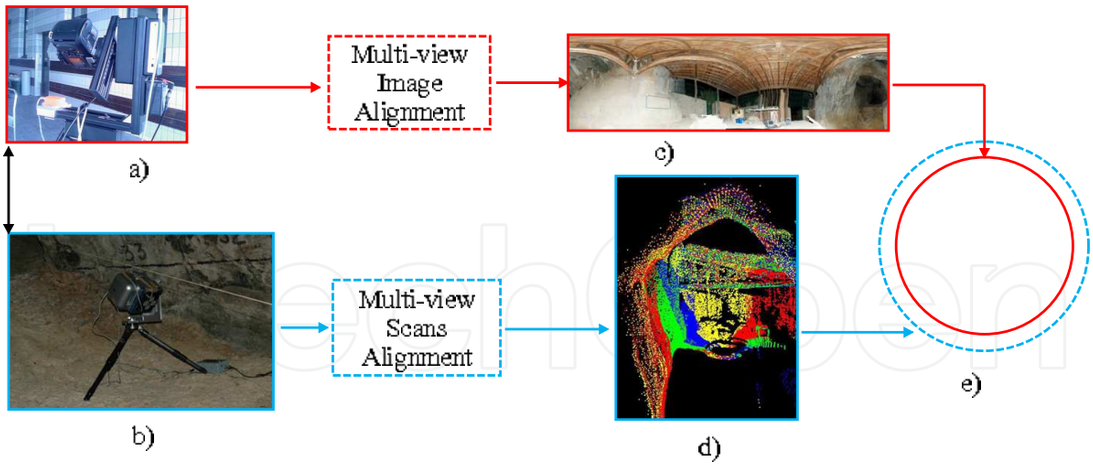


Fig. 1. The 4D-mosaicing process proposed for integration onboard ARTVISYS. a) NIKON D70 ® digital camera mounted on Rodeon ® motorized pan-tilt unit, b) Trimble ® 3D laser-range-finder during a data acquisition campaign undertaken in the Tautavel prehistoric cave (France) by the French Mapping Agency in October 2007, c) a Gigapixel color mosaic resulted from an image sequence acquired in the Tautavel prehistoric cave using an automatic image stitching algorithm which we introduce in Section 5 of this chapter, d) a 3D mosaic resulted from several overlapped scans acquired in the Tautavel prehistoric cave, matched by an automatic multi-view scan-matcher proposed in Section 4, e) alignment the 3D mosaic onto the Gigapixel one to produce the 4D mosaic, process described in Section 6 of this chapter.

- image-laser complementarity has been widely emphasized and investigated by several research works (Dias et al., 2003), (Stamos et al., 2008), (Zhao et al., 2005), (Cole & Newman, 2006), (Newman et al., 2006). There is no doubt that employing the two sensors separately, none can solve for the 3D modeling problem reliably.
- RACL systems overcomes several shortcomings raised by FMCL[2] ones. In particular, image-laser alignment and texture mapping procedures are difficult due to occluded areas in either image or laser data.

Addressing time and in-situ access constraints. An in-situ 3D modeling system must be able to supply fast data acquisition and processing while assuring the 3D scene model completeness in order to avoid to return on site to collect new data.
To this end, we design a *complementary* and *cooperative* image-laser fusion which lead to a *4D mosaicing sensor* prototype. The *complementary* aspect is related to the data acquisition process: in order to deal with time and in-situ access constraints, the proposed acquisition protocol consists in acquiring low-resolution 3D point clouds and high-resolution color images to generate in-situ photorealist 3D models. The use of both sensors rigidly attached leads to a *cooperative* fusion, producing a dual sensing device capable to generate in-situ omnidirectional and photorealist 3D models encoded as 4D mosaic views, which to our knowledge are no achievable using each sensor separately.

[2] Freely Moving Camera Laser system

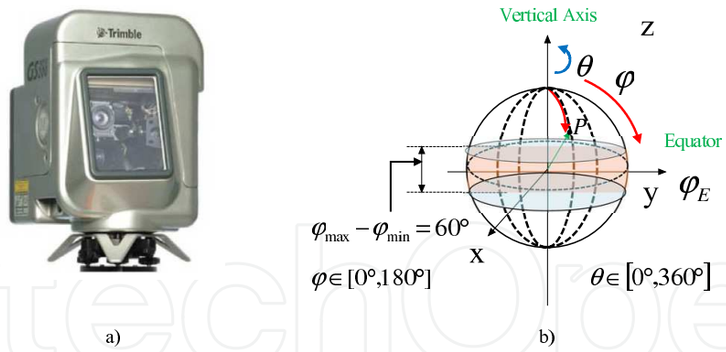


Fig. 2. a) Trimble ® laser range finder delivering 5000 points per second with an accuracy of 3mm at 100m. The dimensions of the laser range finders are: 340mm diameter, 270mm width and 420mm height. The weight of the capturing device is 13.6kg. b) the field of view covered by the sensor.

3.2 Introducing 4D mosaic views: omnidirectional photorealist 3D models

In this chapter we solve for the automation of the 3D modeling pipeline by introducing the *4D mosaic views* as fully spherical panoramic data structure encoding surface geometry (depth) and 3-channel color information (red, green and blue). A *4D mosaic* is generated within three steps (process illustrated in Figure 1), each of which being described in Sections 4, 5 and 6 of this chapter and for which we provide a brief description hereafter.

3D Mosaics from laser-range-finders (LRFs). First, a 3D laser scanner acquires several partially overlapped scans which are aligned and merged into a fully 3D spherical mosaic view via a multi-view scan matching algorithm for which a detailed description is provided in Section 4. Figure 1 d) illustrates an example of a 3D mosaic obtained from real data acquired in the Tautavel prehistoric cave. Since our work is concerned with the 3D modeling in unstructured and underground environments, we introduce an automatic scan matcher which replaces the two-post processing steps usually performed by the currently existing scans alignment techniques (coarse alignment via manual or GPS pose and ICP-like methods for fine registration). The proposed method does not rely on feature extraction and matching, providing thus an environment-independent method.

Gigapixel panoramic head. Second, the motorized panoramic head illustrated in Figure 1 a)acquires a sequence of high-resolution images which are further automatically stitched into a Gigapixel color mosaic via a multi-view image matching algorithm for which a description is given in Section 5. Figure 1 c) depicts an example of the obtained optical mosaic. Since the nowadays image stitching algorithms present several limitations when dealing with unstructured environments, one of our main concern in this chapter is the ability to match images in feature-less areas.

4D-Mosaicing. Third, the 3D mosaic and the 2D optical Gigapixel one are aligned and fused into a photorealist and geometrically accurate 4D mosaic. This is the last step of the 4D-mosaicing process corresponding to Figure 1 e) and for which a mosaic-based approach for image-laser data alignment is proposed in Section 6. The estimated pose is exploited to generate in-situ a 4D mosaic (4-channel: red, green, blue and depth) which to our knowledge has not been reported until now.

The proposed 3D modeling pipeline leads to a vision-based system capable to generate in-situ photorealist and highly accurate 3D models encoded as 4D mosaics for each ARTVISYS's

spatial position, called *station*. The next section introduces the 4D-mosaic-driven in-situ 3D modeling process performed by ARTVISYS aiming to ensure in-situ the 3D scene model completeness.

3.3 4D Mosaic-driven in situ 3D modeling

When dealing with the in-situ 3D modeling problem in large scale complex environments, one has to generate dynamically 3D scene models and to deal with occluded areas on-the-fly, in order to ensure automatically the 3D scene model completeness. This calls for an intelligent 3D modeling system, which implies the computation of the Next Best View (NBV) position (Dias et al., 2002) from which the new 4D mosaic must be acquired in order to sense the occluded areas. In addition, the system must be able to navigate from it's current position to the next best estimated 3D pose from which the next 4D mosaic must be acquired. This implies path planning, autonomous navigation and fast decision making capabilities. A detailed description on this process can be found in (Craciun, 2010).

4D-mosaic-driven acquisition scenario. Due to occlusions, several 4D mosaics must be autonomously acquired from different 3D spatial positions of the system in order to maximize the visible volume, while minimizing data redundancy. To this end, the 4D mosaicing sensor prototype comes together with a a *4D-mosaic-driven acquisition scenario* performed in a stop-and-go fashion, as illustrated in Figure 3.

The acquisition scenario starts by acquiring a 4D-mosaic which is further exploited to detect the occluded areas. In Figure 3, they corresponds to the blue segments representing depth discontinuities associated to each station. In a second step, the system must estimate the 3D pose from which the next 4D-mosaic must be acquired in order to maximize the visible volume. In a third step, the 4D mosaics are matched and integrated within a global 3D scene model which is further exploited to iterate the two aforementioned steps until the 3D scene model completeness is achieved.

Unambigous wide-baseline data alignment. The main advantage of 4D-mosaic views is represented by the fact that they encode explicit color information as 3-channel components (i.e. red, green and blue) and implicit shape description as depth for a fully spherical view of the system's surroundings. The four dimensional components are required in order to ensure reliably further processing, such as unambiguous data matching under wide viewpoint variation.

4. Multi-view rigid scans alignment for in-situ 3D mosaicing

This section presents the first step of the *4D mosaicing process* introduced in Figure1. We describe a multi-view scans alignment technique for generating in-situ 3D mosaics from several partially overlapped scans acquired from the same 3D pose of the system. We first describe the data acquisition scenario, followed by the global approach and an overview of experimental results obtained on real data gathered in two prehistoric caves are presented.

4.1 3D mosaicing acquisition scenario

In Section 3 we presented the hardware design of the proposed system which includes a Trimble® scanning device illustrated in Figure 2 a) providing a cloud of 3D points and their associated light intensity backscattering, within a field of view of 360° horizontally x 60° vertically, as shown in Figure 2 b). When mounted on a tripod, due to the vertical narrow

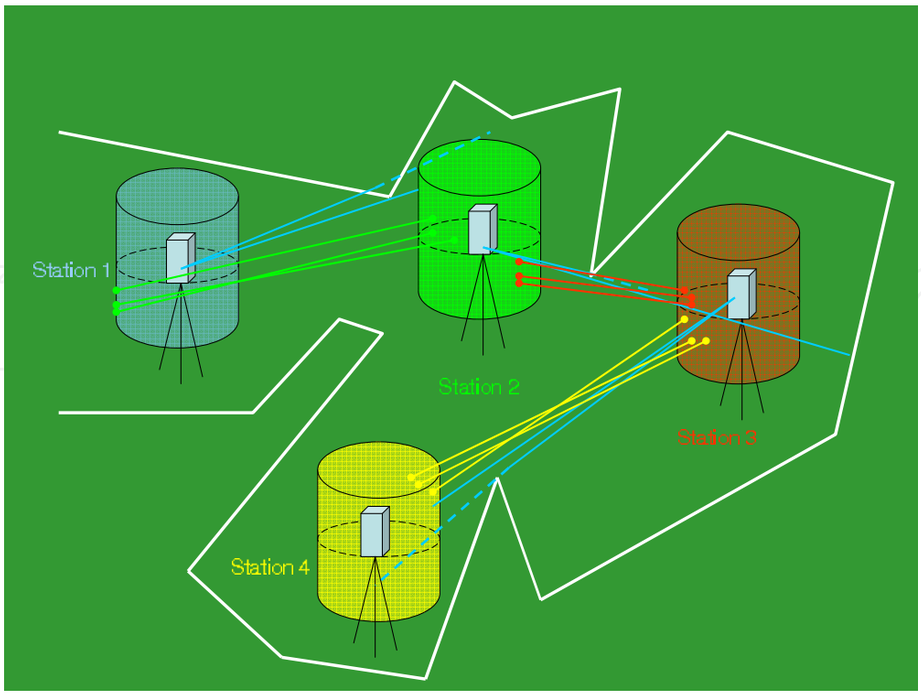


Fig. 3. The 4D-mosaic-driven acquisition scenario performed by ARTVISYS.

field of view, the scanning device is not suitable for the acquisition coverage of ceiling and ground. Therefore, we manufactured in our laboratory a L-form angle-iron shown in Figure 4 a).

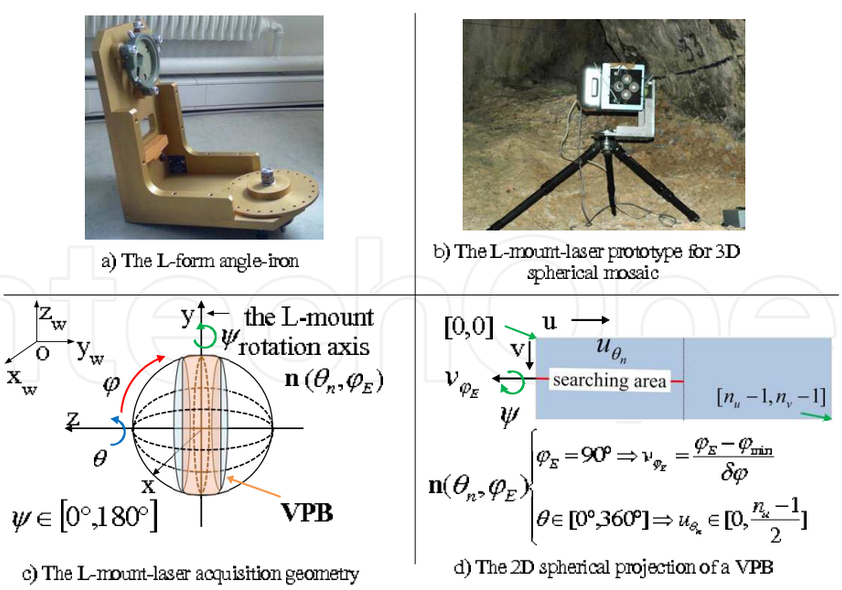


Fig. 4. The 3D mosaicing acquisition geometry.

The L-mount-laser prototype illustrated in Figure 4 b) captures all the area around its optical center within 360° vertically and 60° horizontally as shown in Figure 4 c), which we call a *vertical panoramic band* (VPB). Given a spatial position of the tripod, which we call a *station*, the

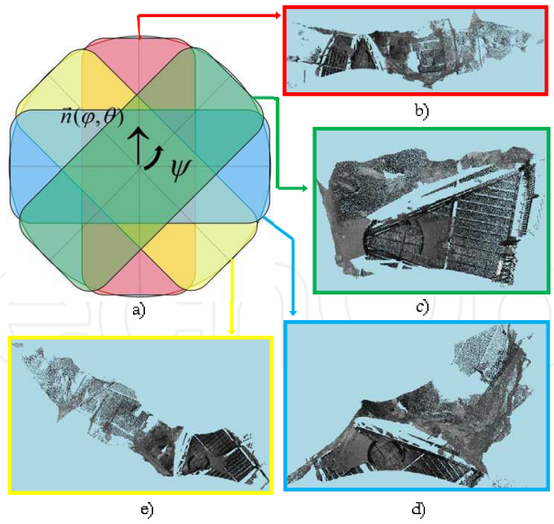


Fig. 5. Example of the 3D mosaicing acquisition scenario performed in the Tautavel prehistoric cave - France. (a)Top view of the acquisition: the laser acquires 4 VPBs as it rotates around its vertical axis $\mathbf{n}(\theta\mathbf{n}, \varphi\mathbf{n})$ with different values of ψ: (b) VBP 1 corresponding to $\psi \approx 0°$, (c) VPB 2 for $\psi \approx 45°$ (d) VPB 3 for $\psi \approx 90°$, (e) VPB 4 for $\psi \approx 135°$.

scenario consists in acquiring multiple overlapping VPBs in order to provide a fully $360° \times 180°$ 3D spherical view. For this purpose, the L-mount-laser is turned around its vertical axis $\mathbf{n}$ (superposed with the scan equator axis, Oy) with different imprecisely known orientations ψ, acquiring one VPB for each orientation, as shown in Figure 5. The L-mount-laser rotation angle ψ may vary within the range of $[0°, 180°]$. For this experiment the L-mount-laser was turned manually, but using a non-calibrated turning device it is straight forward. Generally, $N_{scenario} = 4$ VPBs are acquired to provide a fully 3D spherical view, separated by a rotation $\psi_{max} \simeq 45°$ providing an overlap of $\simeq 33\%$ which our algorithm can handle (to be compared to the state of the art (Makadia et al., 2006), for which a minimum overlap of 45% is required).
Minimum overlap guaranteed. The proposed acquisition scenario facilitates considerably the scan matching task providing a constant and minimal overlapping area situated at the bottom (ground) and top (ceiling) areas of the 3D spherical view. This is an important key issue when performing 3D modeling tasks in large-scale environments, where the amount of the acquired and processed data must be minimized.
The following section introduces an automatic scan alignment procedure which aligns 4 VPBs (composing a fully spherical view) wrt a global coordinate system and integrates them into a single 3D entity, providing thus *in situ* a fully 3D spherical view of the system's surrounding.

4.2 Automatic multi-view rigid scans alignment

Let $S_0, ..., S_{N-1}$ be N partially overlapping scans acquired from different viewpoints. Since each scan is represented in the sensor's local coordinate system, the multi-view scan matching problem consists in recovering each sensors' viewpoints with respect to a global coordinate system, thereby aligning all scans in a common reference system. Generally, the first scan in a sequence can be chosen as the origin, so that the global coordinate system is locked to the coordinate frame of that scan. An absolute pose $\mathbf{T}_i, i = \{0, .., N-1\}$ is the 3D linear operator which rigidly transforms the 3D coordinates of a point $\mathbf{p} \in S_i, \mathbf{p} = (p_x, p_y, p_z, 1)^t$ from the

local coordinate system of scan S_i to the global (or world) coordinate system: $\mathbf{p}_w = \mathbf{T}_i \mathbf{p}_i$. In order to estimate the absolute poses $\mathbf{T}_i$, it is necessary to compute the relative poses $\mathbf{T}_{ij}, j = \{0, .., N-1\}$ and the corresponding overlaps $\mathbf{O}_{ij}$ for each pair of scans via a pair-wise scan matching procedure. Due to the mutual dependency which lies between the overlaps $\mathbf{O}_{ij}$ and the relative poses $\mathbf{T}_{ij}$, the multi-view scan matching is a difficult task.

Pair-wise rigid poses. We developed a pair-wise scan matcher algorithm by matching 2D panoramic views, solving *simultaneously* the above interrelated problems using a pyramidal dense correlation framework via quaternions. The pair-wise scan matching procedure exploits either intensity or depth 2D panoramic views, which encode spatial and appearance constraints increasing therefore the robustness of the pair-wise scan matching process. We solve for the pose estimation in two steps, within a hybrid framework: the rotation $\mathbf{R}$ is first computed by matching either intensity or depth data in the 2D panoramic image space, while the residual translation is computed a posteriori by projecting back in the 3D space the rotationally aligned panoramic images.

The proposed method employs an adaptable pyramidal framework which is the key issue for modeling in occluded environments, providing robustness to large-scale sparse data sets and cutting down the combinatory. In addition, the pyramidal structure emphasizes the tradeoff between the two key aspects of any scan matcher, the accuracy and the robustness. In this work, the accuracy is related to the subpixel precision attached to the dense correlation step, while the robustness component is related to the capability of the scan matcher to handle large motions, performing pose estimation in a coarse to fine fashion.

The global *multi-view fine alignment* is built upon a topological criterion introduced by (Sawhney & Ayer, 1996) for image mosaicing and employed by (Huber, 2002) for matching partially overlapped 3D point clouds. We extend this criterion in order to detect scans which do not correspond to the currently processed sequence (introduced in (Craciun et al., 2010) as *alien* scans). Next, the global multi-view fine alignment refines the pair-wise estimates by computing the best reference view which optimally registers all views into a global 3D scene model.

A detailed description of our method and a quality assement using several experiments performed in two prehistoric underground prehistoric caves may be found in (Craciun et al., 2008), (Craciun et al., 2010), (Craciun, 2010).

4.3 Multi-view scans alignment experiments

Data input. We applied the 3D mosaicing scenario described in Section 4.1 in two prehistoric caves from France: Moulin de Languenay - trial 1 and Tautavel - trials 2, 3 and 4. Each trial is composed by sequence of 4-VPBs acquired nearly from the same 3D position. In order to evaluate the robustness of the proposed method wrt different scanning devices and different scans resolutions, we performed several tests on data acquired with different acquisition setups.

Moulin de Languenay - trial 1: time and in-situ access constraints were not noticed and therefore the Trimble® GS100 laser was set to deliver multi-shot and high resolution scans.

Tautavel - trials 2, 3, 4: the experiments were run in a large-scale and "difficult-to-access" underground site. Therefore, the acquisition setup was designed to handle large-scale scenes while dealing with time and in-situ constraints. In particular, Trimble® GS200 was employed to supply accurate measurements at long ranges. In addition, during experiments we focused to limit as much as possible the acquisition time by setting the sensing device to acquire

one-shot and low resolution scans, emphasizing the robustness of our algorithm with respect to sparse large scale data sets caused by depth discontinuities. Figures 6 illustrates the rendering results for trial 2, obtained by passing each 4-VPBs sequence to the automatic intensity-based multi-view scan matcher.

Trial	Mode	$(\bar{\mathbf{r}} \pm \sigma_{\bar{\mathbf{r}}}) \times 10^{-2}$(m)	$(\Delta\bar{\mathbf{r}} \pm \Delta\sigma_{\bar{\mathbf{r}}}) \times 10^{-2}$(m)	♯points, CPU time (min)
Trial 1	Intensity	3.913 ± 15.86	0.793 ± 2.22	1.508×10^6
GS100	Depth	3.12 ± 13.64		16.44
Trial 2	Intensity	1.18 ± 16.14	1.94 ± 0.84	2.5829×10^6
GS200	Depth	3.12 ± 16.98		27.39
Trial 3	Intensity	0.332 ± 4.15	0.021 ± 0.154	2.6079×10^6
GS200	Depth	0.353 ± 4.304		27.66
Trial 4	Intensity	0.184 ± 1.249	0.007 ± 0.884	2.5321×10^6
GS200	Depth	0.191 ± 0.365		26.28

Table 1. Results of the global 3D scene models. The fourth column illustrates that the accuracy may vary following the mode used with an order of 10^{-2} of the pose estimates wrt the mode used. The last column illustrates the number of points and runtime obtained for each trial.

Table 1 provides the global residual errors obtained for all trials. When analyzing the residual mean errors, we observe the inter-dependency between the alignment accuracy and the number of points provided by the capturing device for pose calculation. The experiments demonstrates the robustness and the reliability of our algorithm in complex environments where depth discontinuities lead to large scale sparse data sets. The fourth column of Table 1 illustrates that following the scan matcher mode, the results' accuracy may vary between $[10^{-2}, 10^{-3}]$.

Runtime. The experiments were run on a 1.66 GHz Linux machine using a standard CPU implementation. The last column of Table 1 shows that the proposed approach exhibits robustness to registration errors with a reasonable computation time. Nevertheless, since the algorithm was originally designed in a multi-tasking fashion, it allows for both sequential and parallel processing on embedded platforms. In (Craciun, 2010) we provide the embedded design for parallel implementation on a multi-core embedded platform.

5. Automatic gigapixel optical mosaicing

The second step in the 4D mosaicing process illustrated in Figure 1 is represented by multi-view image alignment for generating in-situ a fully spherical panoramic view of the system's surroundings. We describe hereafter the Gigapixel mosaicing system designed to be integrated within the proposed in-situ 3D modeling process driven by 4D mosaic views. Further reading on the research work presented in this section can be found in (Craciun et al., 2009), (Craciun, 2010).

5.1 Gigapixel mosaicing system

The inputs of our algorithm are several hundreds of ordered high resolution images acquired from a common optical center. The capturing device illustrated in Figure 7 is previously parameterized with the field of view to be cover and the desired overlap between adjacent images. The method proposed in this paper uses the complementarity of the existing image

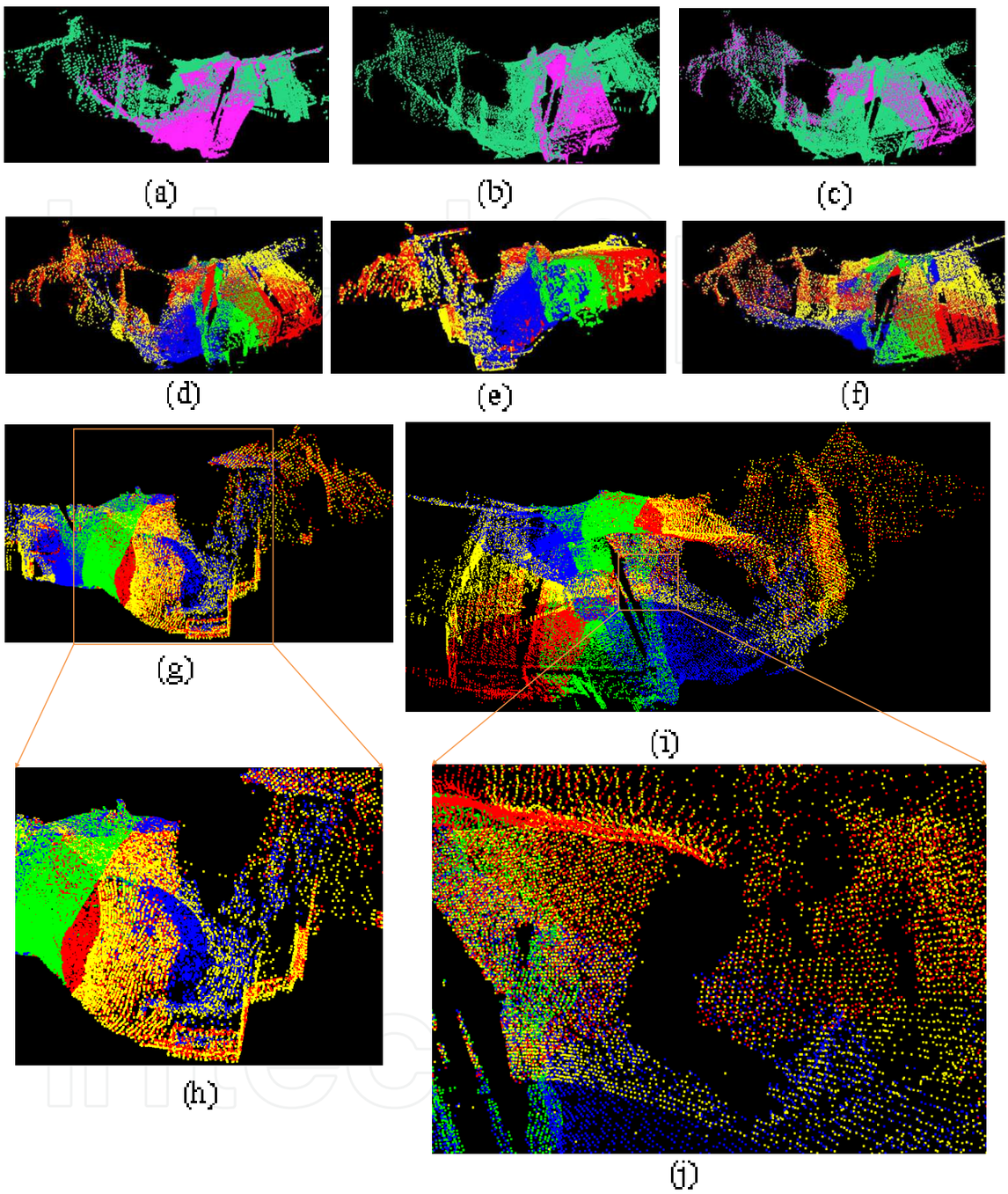


Fig. 6. Multiview Scan Matching results on data sets acquired in Tautavel prehistoric cave, France - Trial 2. (a) S_1 - green, S_2 - magenta, (b) S_{12} - green, S_3 - magenta, (c) S_{123} - green, S_4 - magenta, (d) Multiview scan alignment - Top-down view, S_1 - yellow, S_2 - blue, S_3 - green, S_4 - red, (e) Front-left view, (f) Top view, (g) Front-right view, (h) Zoom-in outdoor front-right view, (i) Bottom-up view, (j) Zoom-in cave's ceiling.

alignment techniques (Snavely et al., 2006), (direct vs. feature-based) and fuses their main advantages in an efficient fashion.
First, a global-to-local pairwise motion estimation is performed which refines the initial estimates provided by the pan-tilt head. We solve for rotation using a pyramidal patch-based correlation procedure via quaternions. The pyramidal framework allows to handle very noisy initial guess and big amounts of parallax.
In order to provide robustness to deviations from pure parallax-free motion[3], the global rotation initializes a patch-based local motion estimation procedure. The pairwise procedure outputs a list of locally matched image points via a translational motion model. Since the matched points do not correspond to any corner-like features, we introduce them as *anonymous features* (AF).
Second, the multi-view fine alignment is achieved by injecting the AF matches in a bundle adjustment engine (BA) (Triggs et al., 1999).Comparing to Lowe's method(Brown & Lowe, 2007), the proposed algorithm can deal with feature-less areas, providing therefore an environment-independent method for the image alignment task.
The following sections describe the overall flow of processing. First, we briefly introduce the camera motion parametrization. Second, we introduce the global-to-local pairwise motion estimation, followed by the multi-view fine alignment description.

(a) (b)

Fig. 7. Mosaicing acquisition System: a NIKON® D70 digital camera (a) with its optical center fixed on a motorized pan-tilt head (Rodeon manufactured by Clauss®) attached to a tripod base (b).

5.2 Camera motion parametrization

Assuming that the camera undergoes purely rotations around it's optical center the camera motion can be parameterized by a 3×3 rotation matrix $\mathbf{R}$ and the camera calibration matrix $\mathbf{K}$. Under the pinhole camera model, a point in space $\mathbf{p} = (p_x, p_y, p_z)^T$ gets mapped to a 2D point $\mathbf{u} = (u_x, u_y)^T$ through the central projection process, which can be written using the

[3] In practice we may notice visible seams due to images' misalignment. One of the main reason is that the motorization of the capturing device yields some vibration noise which is further amplified by the tripod platform. Moreover, unmodeled distortions or failure to rotate the camera around the optical center, may result small amounts of parallax.

homogenous coordinates $(u_x, u_y, 1)^T$ as following:

$$\begin{pmatrix} u_x \\ u_y \\ 1 \end{pmatrix} \cong \mathbf{KR} \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} \tag{1}$$

where, $\mathbf{K} = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix}$ contains the intrinsic parameters, i.e. the focal f and the principal point offset (x_0, y_0). The inversion of Equation 1 yields a method to convert pixel position to 3D-ray. Therefore, using pixels from an image (I_2) we can obtain pixel coordinates in another image (I_1) by applying the corresponding 3D transform and by projecting the transformed points into the I_1's space using equation 1. This principle can be summarized by the warping equation which is expressed as:

$$\hat{\mathbf{u}}_1 \cong \mathbf{K}_1 \mathbf{R}_1 \mathbf{R}_2^{-1} \mathbf{K}_2^{-1} \mathbf{u}_2 \tag{2}$$

Assuming that all the intrinsic parameters are known and fixed for all n images composing the mosaic, i.e. $\mathbf{K}_i = \mathbf{K}, i = 1, .., n$, this simplifies the 8-parameter homography relating a pair of images to a 3-parameter 3D rotation

$$\hat{\mathbf{u}}_1 \cong \mathbf{K} \mathbf{R}_{12} \mathbf{K}^{-1} \mathbf{u}_2 \tag{3}$$

Rotation parametrization. We employ unit quaternions $\mathbf{q}_\theta$, $\mathbf{q}_\varphi$, $\mathbf{q}_\psi$ for representing rotations around the tilt, pan and yaw axis which are denoted by their corresponding vectors $\mathbf{n}_\theta = (1, 0, 0)$, $\mathbf{n}_\varphi = (0, 1, 0)$, $\mathbf{n}_\psi = (0, 0, 1)$. The 4 components of an unit quaternion representing a rotation of angle θ around the n_θ axis are given by $q_\theta = (q_\theta^w, \mathbf{n}_\theta) = (q_\theta^w, q_\theta^x, q_\theta^y, q_\theta^z)^T$.
The orthogonal matrix $\mathbf{R}(\dot{\mathbf{q}})$ corresponding to a rotation given by the unit quaternion $\dot{\mathbf{q}}$ is expressed by:

$$\mathbf{R}[\dot{\mathbf{q}}] = \begin{pmatrix} q_0^2 + q_x^2 - q_y^2 - q_z^2 & 2(q_x q_y - q_0 q_z & 2(q_0 q_y + q_x q_z) \\ 2(q_0 q_z + q_x q_y) & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2(q_y q_z - q_0 q_x) \\ 2(q_x q_z - q_0 q_y) & 2(q_0 q_x + q_y q_z) & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{pmatrix} \tag{4}$$

Capture deviations from parallax-free or ideal pinhole camera model. In order to handle deviations from pure parallax-free motion of ideal pinhole camera model we improve the camera motion model by estimating a local motion estimation provided by a patch-based local matching procedure.

5.3 Global-to-local pair-wise motion estimation

The proposed framework starts with the global rotation estimation followed by the parallax compensation which is performed via a patch-based local motion estimation.

5.3.1 Rigid rotation computation

The motion estimation process follows four steps: (i) pyramid construction, (ii) patch extraction, (iii) motion estimation and (iv) coarse-to-fine refinement. At every level of the pyramid $l = 0, ..., L_{max}$ the goal is to find the 3D rotation $\mathbf{R}^l$. Since the same type of operation is performed at each level l, let us drop the superscript l through the following description.
Let $\mathbf{R}(\mathbf{q}_\theta, \mathbf{q}_\varphi, \mathbf{q}_\psi)^{init}$ be the initial guess provided by the pan-tilt head, where $(\theta, \varphi, \psi)_{hard}$ denote the pitch, roll and yaw angles, respectively expressed in the camera coordinate system.

The optimal rotation is computed by varying the rotation parameters (θ, φ, ψ) within an homogeneous *pyramidal searching space*, $\mathcal{P}_{\mathcal{SS}}$, which is recursively updated at each pyramidal level. $\mathcal{P}_{\mathcal{SS}}$ is defined by the following parameters: θ range $\Delta\theta$, φ range $\Delta\varphi$, ψ range $\Delta\psi$ and their associated searching steps, $\delta\theta, \delta\phi, \delta\psi$.
The rotation angles are computed by applying rotations $\mathbf{R}_{(\theta,\varphi,\psi)}, (\theta, \varphi, \psi) \in \mathcal{P}_{\mathcal{SS}}$ to the 3D rays of recovered from pixels belonging to I_2 and matching the corresponding transformed pixels with pixels from I_1. For a given rotation $\mathbf{R}_{(\theta,\varphi,\psi)}, (\theta, \varphi, \psi) \in \mathcal{P}_{\mathcal{SS}}$ we can map pixels $\mathbf{u}_2$ from I_2 in the I_1's space using the warping equation expressed in Equation 3.

$$\hat{\mathbf{u}}_1 \cong \mathbf{K}\mathbf{R}_{(\theta,\varphi,\psi)\in\mathcal{P}_{SS}}\mathbf{K}^{-1}\mathbf{u}_2 \tag{5}$$

We obtain the rotated pixel from I_2 warped in the I_1's space which yields an estimate of I_1, noted $\hat{I}_1$. The goal is to find the optimal rotation which applied to pixels from I_2 and warped in the I_1's space minimizes the difference in brightness between the template image I_1 and its estimate, $\hat{I}_1(\mathbf{u}_2; \mathbf{R}_{(\theta,\varphi,\psi)})$.
Since images belonging to the same mosaic node are subject to different flash values, we employ the Zero Normalized Cross Correlation score [4] to measure the similarity robustly wrt illumination changes. The similarity score $\mathcal{Z}$ is given in Equation (6), being defined on the $[-1, 1]$ domain and for high correlated pixels is close to the unit value.

$$-1 \leq \mathcal{Z}(I_1(\mathbf{u}), I_2(\hat{\mathbf{u}})) = \frac{\sum_{\mathbf{d}\in\mathcal{W}}[I_1(\mathbf{u}+\mathbf{d}) - \bar{I}_1(\mathbf{u})][I_2(\hat{\mathbf{u}}+\mathbf{d}) - \bar{I}_2(\hat{\mathbf{u}})]}{\sqrt{\sum_{\mathbf{d}\in\mathcal{W}}[I_1(\mathbf{u}+\mathbf{d}) - \bar{I}_1(\mathbf{u})]^2 \sum_{\mathbf{d}\in\mathcal{W}}[I_2(\hat{\mathbf{u}}+\mathbf{d}) - \bar{I}_2(\hat{\mathbf{u}})]^2}} \leq 1 \tag{6}$$

The global similarity measure is given by the mean of all the similarity scores computed for all the patches belonging to the overlapping region. For rapidity reasons, we correlate only border patches extracted in the overlapping regions.

$$\mathbf{E}[\mathbf{R}_{(\theta,\varphi,\psi)}] = \frac{1}{N_w}\sum_{j=0}^{N_w-1}\Phi_j\mathcal{Z}(I_1(\mathbf{u}^j), I_2(\hat{\mathbf{u}}^j_{\mathbf{R}_{(\theta,\varphi,\psi)}})) \tag{7}$$

Φ_j defines a characteristic function which takes care of "lost"[5] and "zero"[6] pixels and N_w denotes the number of valid matches belonging to the overlapping area.
The global dissimilarity score $\mathbf{E}(\mathbf{R}_{(\theta,\varphi,\psi)})$ is defined on the interval $[0, 1]$. The optimal rotation $\hat{\mathbf{R}}_{(\theta,\varphi,\psi)}$ is obtained by maximizing the global similarity score $\mathbf{E}[\mathbf{R}_{(\theta,\varphi,\psi)}]$ over the entire searching area $\mathcal{P}_{SS}$.

$$\hat{\mathbf{R}}_{(\theta,\varphi,\psi)} = \arg\max_{(\theta,\varphi,\psi)\in\mathcal{P}_{SS}} \mathbf{E}[\mathbf{R}_{(\theta,\varphi,\psi)}] \tag{8}$$

5.3.2 Non-rigid motion estimation

In order to handle deviations from pure-parallax motions or from ideal pinhole camera, we use the rotationally aligned images to initialize the local patch matchingprocedure. Let $\mathbf{P}_1 =$

[4] For each pixel, the score is computed over each pixel's neighborhood defined as $\mathcal{W} = [-w_x, w_x] \times [-w_y, w_y]$ centered around $\mathbf{u}_2$ and $\hat{\mathbf{u}}_1$ respectively, of size $(2w_x + 1) \times (2w_y + 1)$, where $w = w_x = w_y$ denote the neighborhood ray.
[5] the pixel falls outside of the rectangular support of I_2
[6] missing data either in $I_1(\hat{\mathbf{u}}^j_{\mathbf{R}})$ or $I_2(\hat{\mathbf{u}}^j_{\mathbf{R}})$, which may occur when mapping pixels $\hat{\mathbf{u}}^j_{\mathbf{R}}$ in the I_2's space

$\{\mathcal{P}(\mathbf{u}_1^k)|\mathbf{u}_1^k \in I_1, k = 1, ..., N_1\}$ and $\mathbf{P}_2 = \{\mathcal{P}(\mathbf{u}_2^k)|\mathbf{u}_2^k \in I_2, k = 1, ..., N_2\}$ be the patches extracted in image I_1 and I_2 respectively, which are defined by a neighborhood $\mathcal{W}$ centered around $\mathbf{u}_1^k$ and $\mathbf{u}_2^k$ respectively. For each patch $\mathcal{P}(\mathbf{u}_1^k) \in \mathbf{P}_1$ we search for its optimal match in I_2 by exploring a windowed area $\mathbf{W}^{\mathbf{SA}}(\mathbf{u}_2^k; \hat{\mathbf{R}})$ centered around $(\mathbf{u}_2^k; \hat{\mathbf{R}})$, where $\mathbf{SA}$ denotes the searching area ray.
Let $\mathbf{P}_2^{k,\mathbf{SA}} = \{\mathcal{P}(\mathbf{u}_2^m)|\mathbf{u}_2^m \in \mathbf{W}^{\mathbf{SA}}(\mathbf{u}_2^k; \hat{\mathbf{R}}) \subset I_2, m = 1, .., M\}$ be M patches extracted from the warped image's searching area centered around $(\mathbf{u}_2^k; \hat{\mathbf{R}})$, with 1-pixel steps. For each patch $\mathcal{P}(\mathbf{u}_2^m)$ we compute the similarity score $\mathcal{Z}(I_1(\mathbf{u}^k), I_2(\mathbf{u}^m))$ and we perform a bicubic fitting in order to produce the best match with a subpixel accuracy and real time performances. The best match is obtained by maximizing the similarity score $\mathcal{Z}$ over the entire searching area $\mathbf{W}^{\mathbf{SA}}$.

$$\mathcal{P}(\hat{\mathbf{u}}_2^k) = \arg \max_{\mathbf{u}_2^m \in \mathbf{W}^{\mathbf{SA}}(\mathbf{u}_2^k; \hat{\mathbf{R}})} \mathcal{Z}(I_1(\mathbf{u}^k), I_2(\mathbf{u}_2^m)) \tag{9}$$

In order to handle "lost" or "zero" pixels, patch matches corresponding to uncomplete warped patches are discarded. This yields a list of matched patches $\mathcal{P}(\mathbf{u}_1^k)$ and $\mathcal{P}(\hat{\mathbf{u}}_2^k)$ which gives the possibility to compute a local translational model for each patch: $\mathbf{t}^k = \|\mathbf{u}_1^k - \hat{\mathbf{u}}_2^k\|$ and compensates eventual parallax motions or deviations from the ideal pinhole camera model. Moreover, the local motion allows the possibility to establish a mean translational motion model over the entire image space, noted $\bar{\mathbf{t}}$. The list of the patch matches are further injected into a bundle adjustment engine for multi-view fine alignment and gap closure.

5.3.3 Experimental results

Figures 8 and 9 illustrate the results obtained by running the global-to-local image motion estimation procedure on an image pair gathered in the Tautavel prehistoric cave, France. The capturing device was set to acquire high resolution images of size 3008×2000 with an overlap of $\simeq 33\%$. In order to evaluate our technique with respect to a feature-based method, we show the results obtained on an image pair for which the SIFT detection and matching failed. The rotation computation starts at the lowest resolution level, $L_{max} = 5$ where a fast searching is performed by exploring a searching space $P_{SS}^{L_{max}} = 5°$ with 1-pixel steps in order to localize the global maximum (Fig. 8c). The coarse estimation is refined at higher resolution levels $l = L_{max} - 1, .., 0$ by taking a P_{SS} of 4 pixels explored with 1-pixel steps. Since deviations from parallax-pure motion are negligible we speed up the process by computing the local motion directly at the highest resolution level, $l = 0$ (Fig. 9). The residual mean square error $(\bar{\mathbf{r}})$ and the standard deviation $(\sigma_{\mathbf{r}})$ of the pairwise camera motion estimation $[\hat{\mathbf{R}}, \mathbf{t}^k]$ are computed using the reprojection error in the 2D space given by:

$$\bar{\mathbf{r}}_{2D} = \frac{1}{N} \sum_{k=1}^{k=N} \|\mathbf{u}_i^k - \mathbf{K}\hat{\mathbf{R}}_{ij}^T \mathbf{K}^{-1}(\hat{\mathbf{u}}_j^k - \mathbf{t}^k)\| \tag{10}$$

5.4 Multi-view fine alignment using the existent BA solutions

Given the pairwise motion estimates $\hat{\mathbf{R}}_{ij}$ and the associated set of AF matches $\mathbf{P}(i,j) = \{(\mathbf{u}_i^k \in I_i; \hat{\mathbf{u}}_j^k \in I_j)|i \neq j, j > i\}$, we refine the pose parameters jointly within a bundle adjustment process (Triggs et al., 1999). This step is a critical need, since the simple concatenation of pairwise poses will disregard multiple constraints resulting in mis-registration and gap. In order to analyze the behavior of the existent BA schemes when consistent matches are injected into

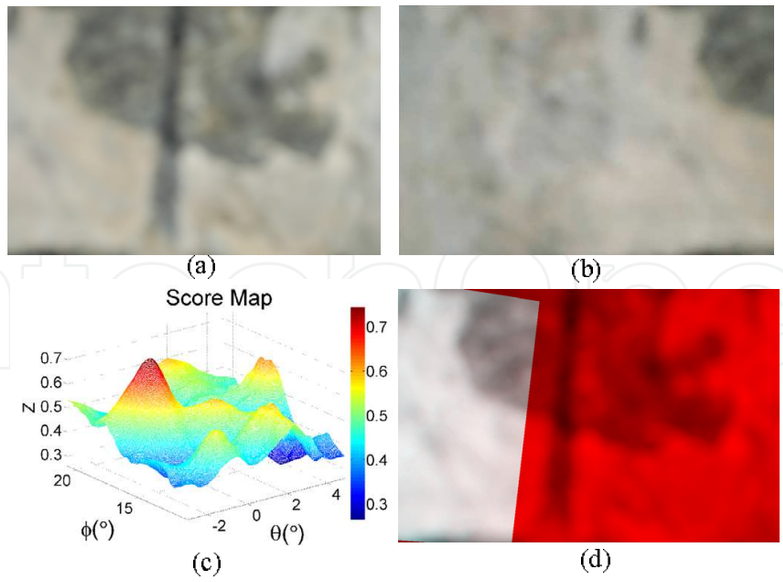


Fig. 8. Rigid Rotation Estimation.(a)origin I_1, (b)image to align I_2, (c)global maximum localization at level $L_{max} = 5$, (d)rotationally aligned images at level $l = 0$: I_1-red channel, the warped image $I_2(\mathbf{u}; \hat{\mathbf{R}})$-green channel, $\hat{\mathbf{R}}(\theta, \varphi, \psi) = (17.005°, 0.083°, 0.006°)$.

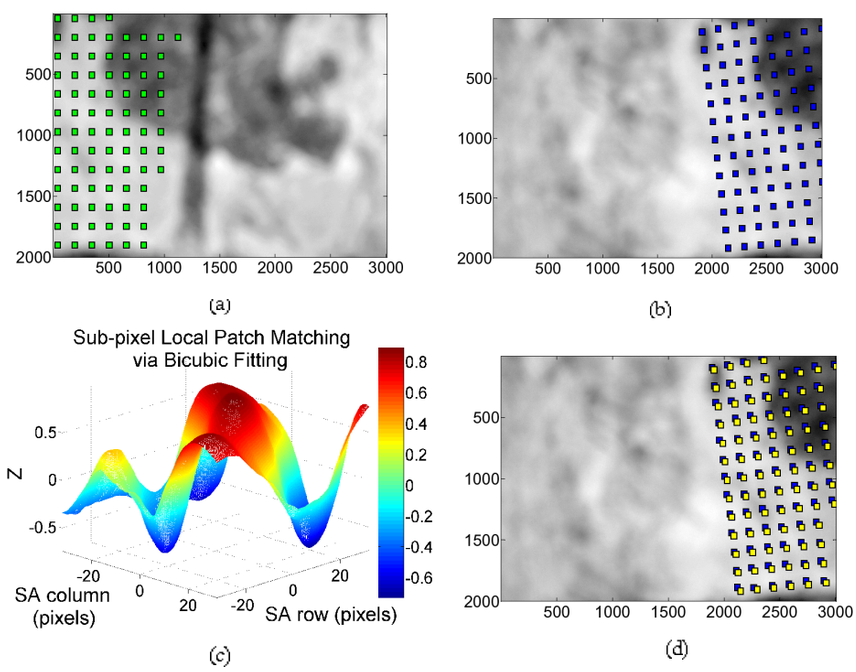


Fig. 9. Anonymous Features Matching Procedure. $\mathcal{W} = 15$ pixels, 85 AF matches. (a)$\mathcal{P}(\mathbf{u}_1^k)$, (b)$\mathcal{P}(\mathbf{u}_2^k)$ extraction in I_2 using the rotation initialization, (c)Bicubic fitting for an arbitrary patch: $\mathbf{SA} = 32$ pixels, matching accuracy: 0.005 pixels, (d)AF-based optical flow: $\mathcal{P}(\mathbf{u}_2^k)$ blue, $\mathcal{P}(\hat{\mathbf{u}}_2^k)$ yellow, $\bar{\mathbf{t}} = [1.6141, 1.0621]$ pixels. $\bar{\mathbf{r}} \pm \sigma_{\mathbf{r}} = 0.08 \pm 0.01$

it, we run the BA step integrated within the Autopano Pro v1.4.2 (Kolor, 2005) by injecting AF pairings pre-computed by the proposed global-to-local pair-wise image alignment step described in Section 5.3.

As in (Brown & Lowe, 2007), the objective function is a robust sum squared projection error. Given a set of N AF correspondences $\mathbf{u}_i^k \longleftrightarrow \hat{\mathbf{u}}_j^k, k = 0, .., N-1$ the error function is obtained by summing the robust residual errors over all images:

$$e = \sum_{i=1}^{n} \sum_{j \in I(i)} \sum_{k \in \mathbf{P}(i,j)} h(\mathbf{u}_i^k - \mathbf{K}\hat{\mathbf{R}}_{ij}^T \mathbf{K}^{-1} \hat{\mathbf{u}}_j^k) \tag{11}$$

where n is the number of images, $I(i)$ is the set of adjacent images to image I_i and $h(\mathbf{x})$ denotes the Huber robust error function (Huber, 1981) which is used for outliers' rejection. This yields a non-linear least squares problem which is solved using the Levenberg-Marquardt algorithm. A detailed description of this approach may be found in (Brown & Lowe, 2007).

Trial Tautavel Prehistoric Cave. Since our research work is focused on generating *in situ* complete and photorealistic 3D models of complex and unstructured large-scale environments, the Gigapixel mosaicing system was placed in different positions in order to generate mosaics covering the entire site. We illustrate in this section two examples of high-resolution mosaic views acquired from different spatial poses of the system corresponding to the cave's entrance and center.

Autopano Pro and AF matches. Figures 10 (a), (b) and Table 2 illustrate the mosaicing results obtained by injecting the AF pairings into the BA procedure integrated within the AutopanoPro v1.4.2 which took in charge the rendering process using a spherical projection and a multi-band blending technique. The mosaic's high photorealist level is emphasized by a high-performance viewer which allows for mosaic visualization using 4-level of detail (LOD), as shown in Figures 10 (c)-(f).

Residual errors. The BA scheme includes a self-calibration step and minimizes an error measured in the 2D image space, causing the rejection of correct AF matches and leading to relatively high mis-registration errors, as shown by the fourth row of Table 2. In practice we observed that this shortcoming can be overcome by injecting a high number of AF matches. However, this may be costly and when a low number of matches are used, there is a high probability that all of them to be rejected, producing the BA's failure. Since we can not afford this risk, our first concern is to improve the multi-view fine alignment process by simultaneously computing the optimal quaternions using a criterion computed in the 3D space in order to reduce the residual error when using a minimum number of AF correspondences. To this end, we propose an analytical solution for the multi-view fine alignment step (Craciun, 2010) .

Runtime. For this experiment we employed the original Rodeon® platform, i.e. without the improvements. Therefore, the searching range for the rotation refinement was considerably high, i.e. $\pm 5°$, leading to a computationally expensive rotation estimation stage. The upgraded-Rodeon® (Craciun, 2010) reduces the computational time by a factor of 5.83 for an experimental version of the implementation, i.e. without code optimization. Moreover, the number of images to be acquired is reduced to $N_{im} = 32$ which decreases by a factor of 4 the acquisition time.

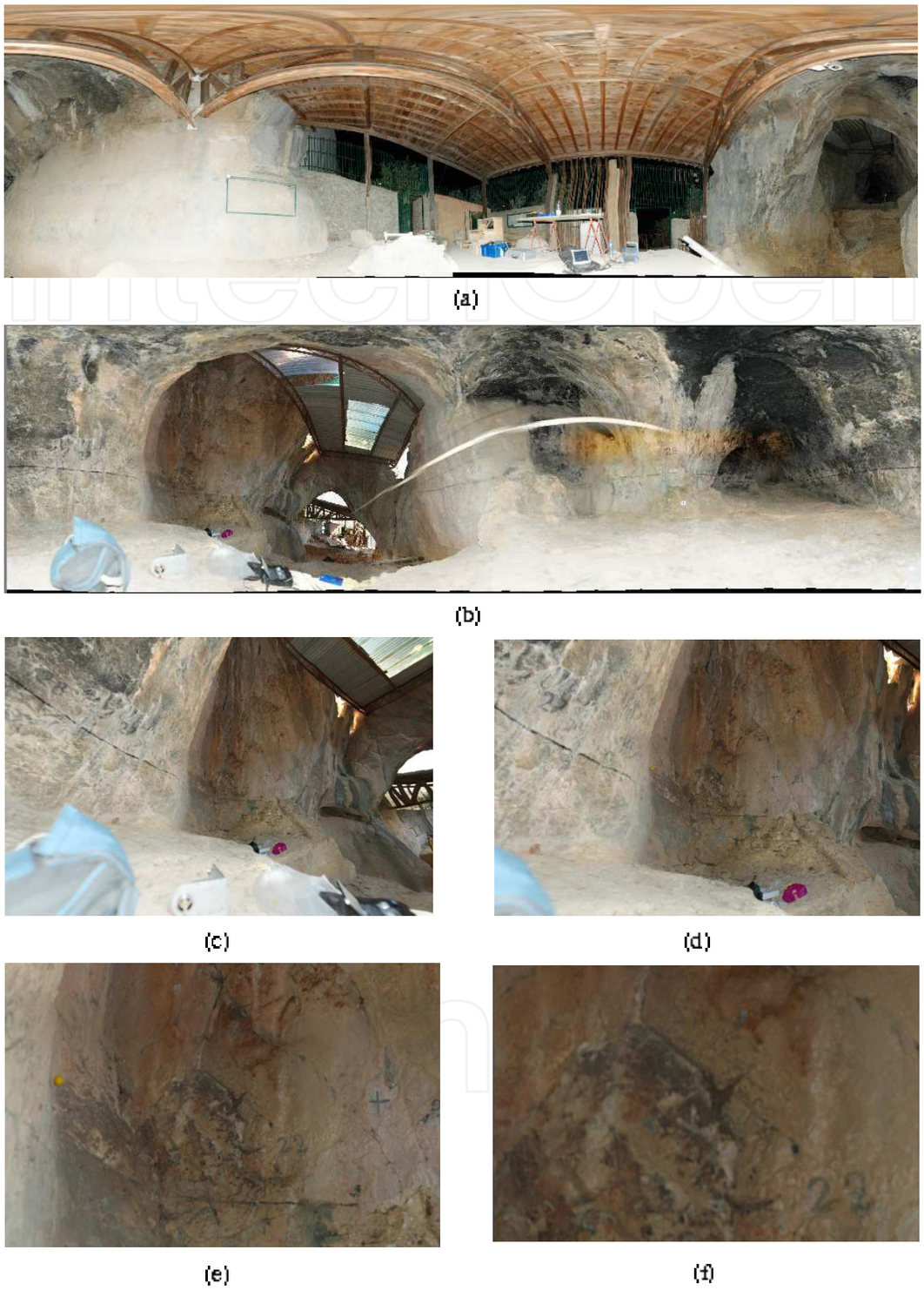


Fig. 10. Mosaicing tests on data sets acquired in Tautavel prehistoric cave using the Rodeon® platform. The mosaics were generated by injecting the AF matches into the BA process integrated within Autopano Pro v1.4.2. (a) - cave's entrance, (b) - cave's center, (c)-(f) 4-LODs corresponding to the right part of mosaic (b).

Mosaic	Figure 10 (a)	Figure 10 (b)
$\#N_{station}$	272	168
FOV(°)	360×108.4	360×105.37
Size(pixels)	43365×13057(567 Mp)	43206×12646 (546 Mp)
e(pixels)	1.93	1.76
# AF matches	21840	13440
CPU (time)	8h 12min	5h 33min

Table 2. Qualitative results corresponding to mosaics generated using Autopano Pro and AF matches when running on a 1.66 GHz Linux machine equipped with 2Gb of RAM memory. The mosaics illustrated in Figures 10 (a) and 10 (b) correspond to the cave's entrance, center respectively.

6. Generating 4D dual mosaic views from image and laser data

The last stage of the 4D mosaicing process illustrated in Figure 1 consists in aligning the 3D mosaic onto the 2D color one, unifying them in a photorealist and geometrically accurate 3D model. This section describes a mosaic-based approach for image-laser data alignment. The reconstruction of the 3D scene model is performed within two steps: (i) an integration step exploits the 3D mosaic to generate 2D meshes and (ii) a texture mapping procedure enables the photorealist component of the 3D scene model.

6.1 Data input and problem statement

Figure 11 illustrates the two inputs of the image-laser alignment procedure. In order to facilitate the visualization of the FOV[7] imaged by each sensor, Figure 11 depicts both the 3D spherical and the 2D image projections associated to each input, i.e. the 3D mosaic generated by the laser and the 2D mosaic obtained from the Gigapixel camera which was down-sampled to meet the 3D mosaic resolution. It can be observed that both sensors are capturing the same FOV, having their optical centers separated by a 3D rotation and a small inter-sensor parallax. In order to build photorealistically textured panoramic 3D models, one must register the 3D spherical mosaic M_{BR-3D} and the color Giga-mosaic M_{HR-RGB} in a common reference coordinate system in order to perform the texture mapping stage.

Pose estimation under calibration constraints. Since the two capturing devices (laser scanner and camera) are supposing to acquire the same FOV, they can be either rigidly attached or used successively, one after another. However, in both cases, it is difficult to calibrate the system such that the parallax is completely eliminated. Consequently, it is possible to model the transformation between the two sensors through a 3D Euclidian transformation with 6-DOF (i.e. three for rotation and three for translation) as illustrated in Figure 11. The following section is dedicated to the description of the image-alignment algorithm allowing to compute transformation relating their corresponding optical centers.

6.2 Automatic pyramidal global-to-local image-laser alignment

We employ a direct correlation-based technique within a feature-less framework. In order to cope with time and in-situ access constraints, we cut down the pose estimation combinatory using a pyramidal framework.

[7] Field of View

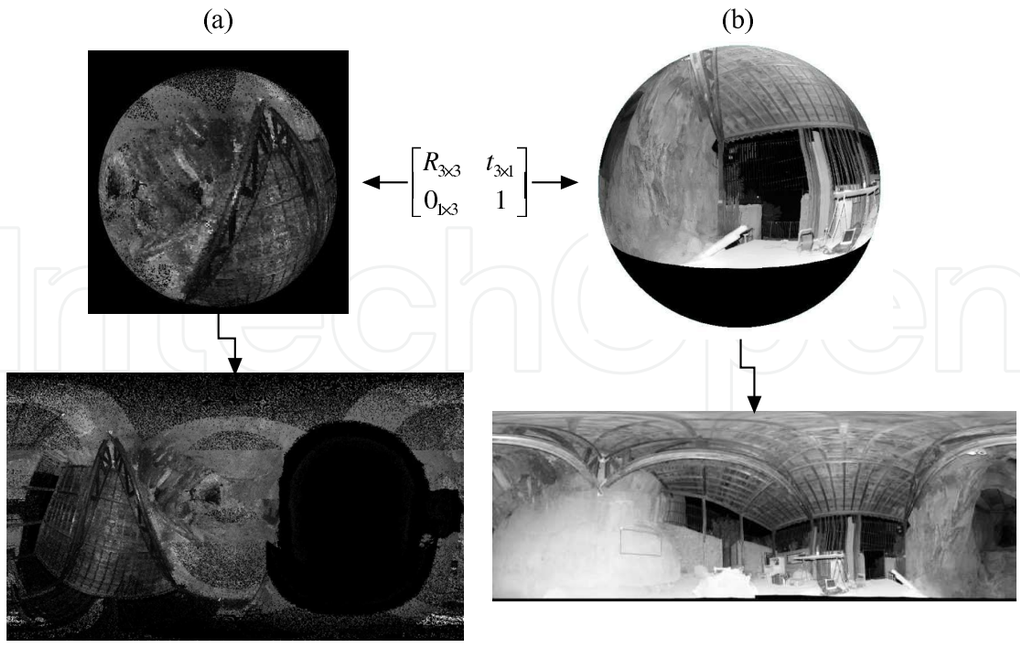


Fig. 11. The two inputs of the panoramic-based image-laser alignment procedure exemplified on a data set acquired in Tautavel prehistoric cave. We illustrate the spherical and image plane projections associated to each input. (a) M_{BR-3D} - the scan matcher output by the 3D mosaicing process described in Section 4. FOV $360° \times 180°$, size: 2161×1276, angular steps $[\delta\theta, \delta\varphi]_{BR-3D} = [0.002906°, 0.00246°]$, (b) the optical mosaic obtained using the algorithm described in Section 5. FOV: $360° \times 108.4°$

Figure 12 illustrates the image-laser fusion pipeline which can be split in two main processes, each of which being detailed through the following description. Since the entire pose estimation method is very similar to the pair-wise global-to-local alignment described in Section 5.3, the following subsections resume several specifications related to its appliance on fully spherical mosaic views.

6.2.1 Pre-processing

The proposed image-laser alignment method correlates the reflectance acquired by the LRF with the green channel of the optical mosaic M_{HR-G}. To do so, we first recover automatically the parameters of the spherical acquisition through a 2D triangulation procedure in order to compute the 2D projection of the 3D mosaic. This stage of the algorithm is very important as it provides the topology between the 3D points and allows fast interpolation.

Generating pyramidal structures for each input: M_{BR-G} and M_{BR-3D}. We generate pyramidal structures of $L_{max} = 3$ levels for both inputs $M_{BR-3D} = \{M^l_{BR-3D} | l = 0, .., L_{max-1}\}$ and $M_{BR-G} = \{M^l_{BR-G} | l = 0, .., L_{max-1}\}$, where the mosaic size ranges from $[2162 \times 1278]$ up to $[270 \times 159]$ corresponding to levels $l = 0, .., L_{max}$.

6.2.2 Pose estimation

The pose estimation procedure employs a hybrid scheme, the 3D rotation is computed by minimizing a radiometric criterion in the 2D mosaic space, while the translation is computed by back-projecting the rotationally aligned mosaics in the 2D space via a local patch matching

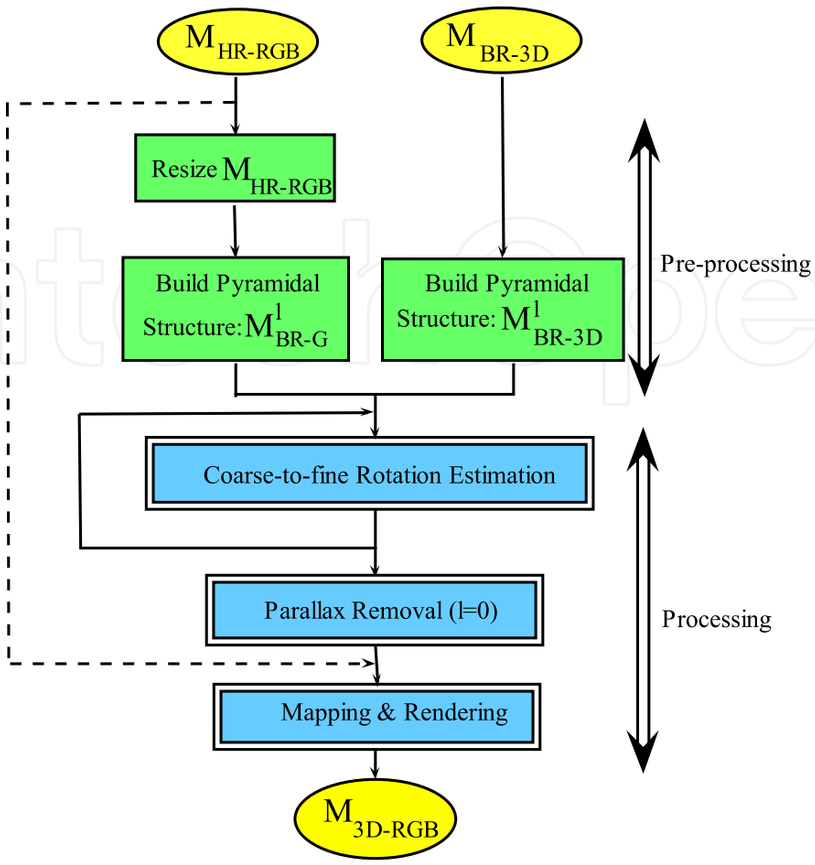


Fig. 12. Image-laser fusion pipeline. Inputs: 3D mosaic M_{HR-RGB} and 2D Giga-pixel color mosaic M_{BR-3D} illustrated in Figures 11 (a) and (b), respectively. The pre-processing and processing steps are highlighted in green and blue, respectively.

procedure. The proposed approach lead to a two-steps rigid transformation computation process: first, the 3D global rotation $\mathbf{R}_{(\theta,\varphi,\psi)}$ is computed in a pyramidal fashion, while the second step is dedicated to the inter-sensor parallax compensation being performed only at the highest resolution level.

Correction of 3D mosaic distortions. As mentioned in Section 4, the 3D mosaic acquisition combines several bands acquired through laser's rotations which may introduce wavy effects within the 3D mosaic geometry. These effects are captured within the inter-sensor parallax computation step which is performed through a non-rigid motion estimation procedure. Consequently, in order to correct the 3D mosaic's geometry, the alignment procedure is performed by aligning the 3D mosaic onto the 2D optical one, M_{BR-G}.

Figure 13 (a) shows that the superposition of the two images does not result in grey-level due to the different responses given by the sensing devices. Figure 13 (b) illustrates a close-up view of the superposed mosaics showing that the global rotation does not model completely the motion separating the camera and the laser, and consequently the inter-sensor parallax must be introduced within the estimated motion model.

Parallax removal. As for the local patch matching procedure described in Section 5, this stage of the algorithm uses the rotationally aligned mosaics. We recover the parallax between the

laser's and the optical mosaicing platform by performing a local patch matching procedure at the highest resolution of the pyramidal structure.
The patch matching procedure outputs a 2D translational motion for each patch, estimating a non-rigid motion over the entire mosaic space. This vector field is used for the parallax removal stage. In addition, the non-rigid motion allows to compute a mean translation motion model defined over the entire mosaic space $\bar{\mathbf{t}}_{2D}$. The parallax is removed in the 2D image space by compensating each $\bar{\mathbf{t}}_{2D}$, obtaining therefore the warped 3D mosaic $\hat{M}_{BR-3D}$ aligned onto the 2D mosaic. Figure 13 (c) depicts the result of the laser-camera alignment procedure.
Accuracy. Although the Giga-pixel mosaic produced using the Autopano Pro software (details are presented in Section 5) has a residual error of 3.74 pixels, it becomes negligible in the down-sampled mosaic M_{BR-G} used for the registration process. A sub-pixel accuracy can be achieved by using a bicubic fitting, as described in Section 5.

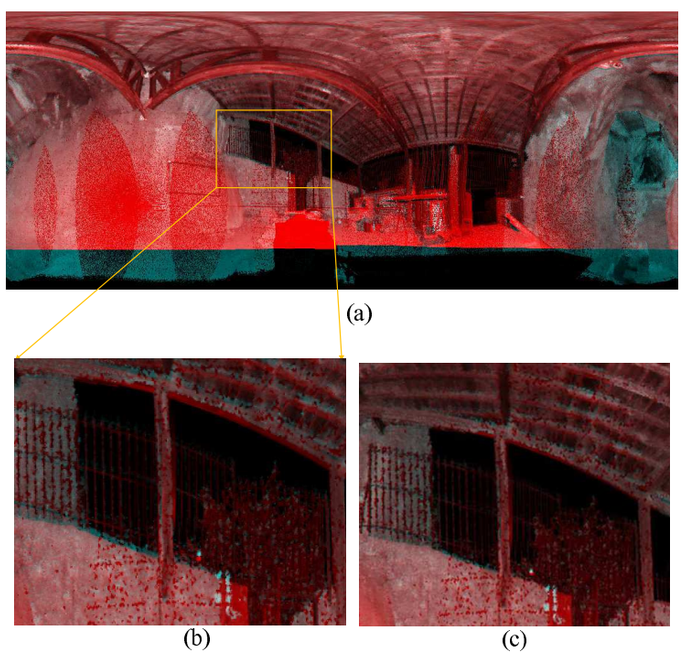
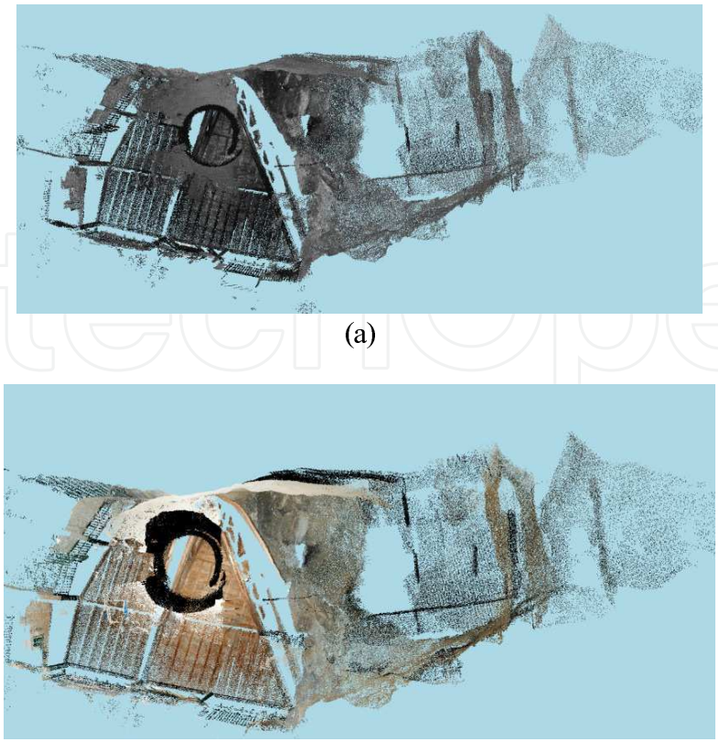


Fig. 13. Experimental results of the parallax removal procedure obtained on data sets acquired in Tautavel prehistoric cave: (a) Superposed aligned mosaics: M_{BR-G} - red channel, $\hat{M}_{BR-3D}$ - greed channel. (b) zoom in - before parallax removal, (c) zoom in - after parallax removal. The compensated parallax amount: $\bar{\mathbf{t}}_{2D} = [-1.775, -0.8275]^T$ pixels.

6.3 Texture mapping and rendering

Since the main goal of our research work is concerned with the in-situ 3D modeling problem, we are mainly interested in producing a fast rendering technique for visualization purposes in order to validate in-situ the data acquisition correctness. To this end, a simple point-based rendering procedure may suffice. Nevertheless, off-line a more artistic rendering can be performed by sending data to a host wirelessly connected to the target.
In-situ point-based visualization. The employed method simply associates the RGB-color to its corresponding 3D coordinate. In order to emphasize the photorealist rendering results obtained when using high-resolution texture maps, Figure 14 compares the rendering results



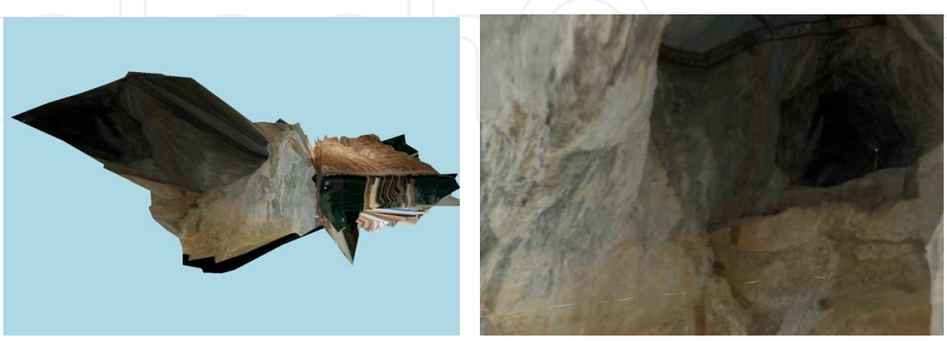
(a)

(b)

Fig. 14. Texture mapping results. (a) The 3D point cloud displayed using the intensity acquired by the LRF. (b) The colored 3D point cloud using the down-sampled optical mosaic M_{BR-RGB}.

obtained by first using the intensity acquired by the 3D scanning device illustrated in Figure 14 (a), while the rendering using the texture maps obtained from the color mosaic is shown in Figure 14 (b).

Off-line mesh-based rendering. We apply an existing 2D meshing algorithm developed in our laboratory by Mathieu Brèdif which assigns to each polygon the RGB-color corresponding to its 3D coordinates. Figures 15 illustrates the rendering results showing that the complex



(a) (b)

Fig. 15. Mesh-based rendering of the Tautavel prehistoric cave. (a) Outdoor view. (b) Indoor view of the 3D model.

surface geometry of the environment lead to depth discontinuities, requiring for a meshing algorithm robust to missing data.

7. Conclusions and future research directions

This chapter aimed at providing solutions for in-situ 3D modeling in complex and difficult to access environments, targeting the automation of the 3D modeling pipeline, and in particular the data alignment problem in feature-less areas. We proposed an image-laser strategy which lead to a *4D mosaicing sensor prototype* able to acquire and process image and laser data to generate in-situ photorealist omnidirectional 3D models of the system's surroundings.

2D, 3D and 4D mosaic views. We propose hardware and software solutions for generating in-situ 2D, 3D and 4D mosaic views in feature-less and GPS-denied areas, making them suitable for map-building and localization tasks. In addition, they provide long-term features tracking, ensuring reliable data matching in feature-less environments. The aforementioned advantages are exploited within a *4D-mosaic-driven acquisition scenario* aiming to ensure the 3D scene model completeness.

Automatic data alignment in feature-less areas. This leads to a two-steps strategy which addresses the automation of the 3D modeling pipeline by solving for its main data alignment issues through the image-laser fusion. We first address a simple problem, i.e. *same viewpoint and small-parallax data alignment*, resulting in automatic 2D and 3D mosaicing algorithms, to provide in a second step image-laser solutions, i.e. the *4D mosaic views*, to solve for *wide-baseline* 3D model alignment using a joint 2D-3D criterion to disambiguate feature matching in feature-less areas.

In our research work, we integrate the 4D mosaicing sensor within a vision-based system designed to supply site surveys and exploration missions in unstructured and difficult to access environments.

8. References

Bailey, T. & Durrant-White, H. (2006). Simultaneous localization and mapping: Part II, *In Proceeding of IEEE Robotics and Automation Magazine* 13(2): 99–110.

Banno, A., Masuda, T., Oishi, T. & Ikeuchi, K. (2008). Flying Laser Range Sensor for Large-Scale Site-Modeling and Its Applications in Bayon Digital Archival Project, *In International Journal of Computer Vision* 78(2-3): 207–222.

Beraldin, J.-A. & Cournoyer, L. (1997). Object modeling creation from multiple range images: Acquisition, calibration, model building and verification, *In Proceedings of International on Recent Advances on 3-D Digital Imaging and Modeling* pp. 326–333.

Besl, P. J. & McKay, N. D. (1992). A method for registration of 3d-shapes, *In IEEE Transactions on Pattern Recognition and Machine Intelligence* 14(2): 239–256.

Brown, M. & Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features, *In International Journal on Computer Vision* 74: 59–73.

Cole, D. M. & Newman, P. M. (2006). Using laser range data for 3d SLAM in outdoor environments, *In Proceedings of IEEE International Conference on Robotics and Automation (ICRA'06)* .

Craciun, D. (2010). Image-laser fusion for 3d modeling in complex environments, *Ph D Thesis Telecom ParisTech* .

Craciun, D., Paparoditis, N. & Schmitt, F. (2008). Automatic pyramidal intensity-based laser scan matcher for 3d modeling of large scale unstructured environments, *In Proceedings of the Fifth IEEE Canadian Conference on Computer and Robots Vision* pp. 18–25.

Craciun, D., Paparoditis, N. & Schmitt, F. (2009). Automatic Gigapixel mosaicing in large scale unstructured underground environments, *In Tenth IAPR Conference on Machine Vision Application* pp. 13–16.

Craciun, D., Paparoditis, N. & Schmitt, F. (2010). Multi-view scans alignment for 3d spherical mosaicing in large scale unstructured environments, *In Journal Computer Vision and Image Understanding* pp. 1248–1263.

Dias, P., Sequeira, V., Gonï£¡laves, J. G. M. & Vaz, F. (2002). Automaic registration of laser reflectance and colour intensity images for 3d reconstruction, *Robotics and Autonomous Systems* 39(3-4): 157–168.

Dias, P., Sequeira, V., Vaz, F. & Goncalves, J. (2003). Underwater 3D SLAM through entropy minimization, *In Proceedings of the 3D Digital Imaging and Modeling (3DIM03)* pp. 418–425.

Durrant-White, H. & Bailey, T. (2006). Simultaneous localization and mapping: Part I, *In Proceeding of IEEE Robotics and Automation Magazine* 13(2): 99–110.

Huber, D. (2002). Automatic Three-dimensional Modeling from Reality, *Ph. D. thesis, Robotics Institute, Carnegie Mellon University,Pittsburgh, PA* .

Huber, D. & Vandapel, N. (2003). Automatic 3d underground mine mapping, *The 4th International Conference on Field and Service Robotics* .

Huber, P. J. (1981). *Robust Statistics*, John Wiley & Sons, New York.

Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M. & Okamoto, Y. (2007). The Great Buddha Project: Digitally Archiving, Restoring, and Analyzing Cultural Heritage Objects, *In International Journal of Computer Vision* 75(1): 189–208.

Johnson, A. (1997). Spin-images: A representation for 3-d surface matching, *PhD thesis, Robotics Institute, Carnegie Mellon University* .

Kolor (2005). Autopano pro, *http://www.autopano.net/en/* .

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J. & Fulk, D. (2000). The Digital Michelangelo Project: 3D Scanning of Large Statues, *In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* pp. 131–144.

Makadia, A., Patterson, A., & Daniilidis, K. (2006). Fully automatic registration of 3d point clouds., *In Proceedings of Compute Vision and Pattern Recognition CVPR'06* pp. 1297–1304.

Moravec, H. P. (1980). Obstacle avoidance and navigation in the real world by a seeing robot rover, *Ph. D. thesis, Stanford University, Stanford, California* .

Newman, P., Cole, D. & Ho, K. (2006). Outdoor SLAM using visual appearance and laser ranging, *In Proceedings of International Conference on Robotics and Automation* .

Nister, D., Naroditsky, O. & Bergen, J. (2004). Visual odometry, *In Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)* pp. 652–659.

Nuchter, A., Surmann, H. & Thrun, S. (2004). 6D SLAM with an application in autonomous mine mapping, *In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'04)* .

Sawhney, H. S. & Ayer, S. (1996). Compact representation of video thourgh dominant multiple motion estimation, *In IEEE Transactions on Pattern Recognition and Machine Intelligence* 18(8): 814–830.

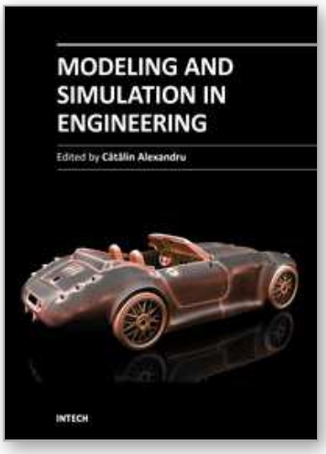
Snavely, N., Seitz, S. M. & Szeliski, R. (2006). Photo tourism: exploring photo collections in 3d, *In Proceedings of ACM SIGGRAPH'06* .

Stamos, I., Liu, L., Chen, C., Wolberg, G., Yu, G. & Zokai, S. (2008). Integrating Automated Range Registration with Multiview Geometry for the Photorealistic Modeling of Large-Scale Scenes, *In International Journal of Computer Vision* 78(2-3): 237–260.

Thrun, S., Montemerlo, M. & Aron, A. (2006). Probabilistic terrain analysis for high-speed desert driving, *In Proceedings of Robotics: Science and Systems* .

Triggs, B., McLauchlan, P., Hartley, R. & Fitzgibbon, A. (1999). Bundle adjustment - a modern synthesis, *In Proceedings of the of the International Workshop on Vision Algorithms: Theory and Practice* pp. 298–372.

Zhao, W., Nister, D. & Hsu, S. (2005). Alignment of Continuous Video onto 3D Point Clouds., *In IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(8): 1308–1318.



Modeling and Simulation in Engineering

Edited by Prof. Catalin Alexandru

ISBN 978-953-51-0012-6
Hard cover, 298 pages
Publisher InTech
Published online 07, March, 2012
Published in print edition March, 2012

This book provides an open platform to establish and share knowledge developed by scholars, scientists, and engineers from all over the world, about various applications of the modeling and simulation in the design process of products, in various engineering fields. The book consists of 12 chapters arranged in two sections (3D Modeling and Virtual Prototyping), reflecting the multidimensionality of applications related to modeling and simulation. Some of the most recent modeling and simulation techniques, as well as some of the most accurate and sophisticated software in treating complex systems, are applied. All the original contributions in this book are jointed by the basic principle of a successful modeling and simulation process: as complex as necessary, and as simple as possible. The idea is to manipulate the simplifying assumptions in a way that reduces the complexity of the model (in order to make a real-time simulation), but without altering the precision of the results.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Daniela Craciun, Nicolas Paparoditis and Francis Schmitt (2012). Image-Laser Fusion for In Situ 3D Modeling of Complex Environments: A 4D Panoramic-Driven Approach, Modeling and Simulation in Engineering, Prof. Catalin Alexandru (Ed.), ISBN: 978-953-51-0012-6, InTech, Available from: http://www.intechopen.com/books/modeling-and-simulation-in-engineering/image-laser-fusion-for-in-situ-3d-modeling-of-complex-environments-a-4d-panoramic-driven-approach

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.