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

4,800
Open access books available

122,000
International authors and editors

135M
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

11

A Stereo Vision Framework for 3-D Underwater Mosaicking

A. Leone, G. Diraco and C. Distante
Institute for Microelectronics and Microsystems, National Research Council
Lecce (Italy)

1. Introduction

Research on automatic mosaic creation for underwater applications has been investigated in the last fifteen years. The reconstruction of complex 3-D structures is useful in several underwater applications; in particular, 3-D mosaicking constitutes an important tool for seabed exploration, improving visualization and navigation in the underwater medium. Moreover, underwater measurement systems have been used extensively in marine research to estimate the size of interesting objects such as organisms and structures. In addition, visual sensing can be an enabling technology for Autonomous Underwater Vehicles (AUVs), which have the critical requirement to maintain an ongoing representation of its relative position with respect to an environmental representation. In these contexts, a better perception of the underwater environment can be achieved by using image processing algorithms for a suitable representation of the seabed. In optic sensing field and underwater environment, shape acquisition concerns with sensing activity including Shape from Stereopsis, Shape from Photometric Stereo, Shape from Motion and Active Stereo. Main aspects about Shape from Stereopsis will be presented focusing the attention on synchronism in the acquisition at hardware/software levels. The usage of triggered expensive equipments in synchronized stereo sequence acquisitions could present some technical limitations, such as low frame rate and poor images resolution, demoting the quality of the 3-D mosaic. For the previous reasons, an ad-hoc algorithmic solution will be detailed to limit asynchronism problems due to time gap (delay) in stereo frames acquisition, when low-cost non-professional (non-triggered) hardware is used. To achieve a metric reconstruction, the presented framework requires a calibration phase whereas a new stereo frame-selection scheme based on the Epipolar Gap Evaluation (EGE) will be discussed to overcome asynchronisms. Normally, dense depth maps obtained by evaluating dense disparity maps allow the construction of high quality 3-D mosaics. A detailed discussion about methodologies for dense stereo matching will be addressed to handle the ill-posed stereo correspondence problem. A joined use of local (respectively global) matching methods and imaging enhancement algorithms will be considered to identify an appropriate amount of right correspondences in severe underwater conditions (backscattering, brightness constancy violation condition). The chapter treats an important aspect of the 3-D mosaic reconstruction as the registration by ego-motion and camera pose estimation. Classical solutions (i.e. Iterative Closest Point algorithm) and new trend in the

Source: Stereo Vision, Book edited by: Dr. Asim Bhatti,
ISBN 978-953-7619-22-0, pp. 372, November 2008, I-Tech, Vienna, Austria

registration activity (Zhang's Epiflow algorithm) will be discussed emphasizing application limits and how the computational cost grows in size with the amount of considered features. A new registration approach will be presented conjugating Epiflow and ICP, under a simplified motion model.

2. Related works for underwater 3-D reconstruction

In the field of 3-D reconstruction, terrestrial applications have encouraged extensive work over the last three decades; on the other hand a limited amount of underwater applications have been explored primarily for mapping and positioning. The first step of the structure reconstruction is the shape acquisition that can be addressed in different ways, as taxonomically shown in Fig.1. In underwater research the shape acquisition is mainly performed with expensive optical sensing methods (Negahdaripour et al., 2002; Khamene et al., 2001) and non-optical sensing methods, working often in conjunction (these methods are underlined in Fig.1). Shape from X is a generic name for techniques that extract shape from images. Normally, in underwater environment the optical sensing techniques include Shape from Stereopsis (Zhang, 2005), Shape from Photometric Stereo (Negahdaripour et al., 2002), Shape from Motion (Khamene et al., 2001) and Active Stereo (Narasimhan et al., 2005). In the field of non-optic underwater sensing, acoustic cameras are employed for 3-D mosaic reconstruction (Castellani et al., 2004), whereas both acoustic and optic cameras are often used providing scene information that cannot be recovered from each sensor alone. Three-dimensional scene structures captured by a camera may be detected and acquired observing the apparent motion of brightness patterns from images. The primary visual motion cue useful for shape acquisition is the perceived movement of brightness patterns, known as optical flow (Horn, 1986) which is an approximation of the 3-D world motion field. The 3-D reconstruction from differential motion cues requires accurate optical flow computation. In

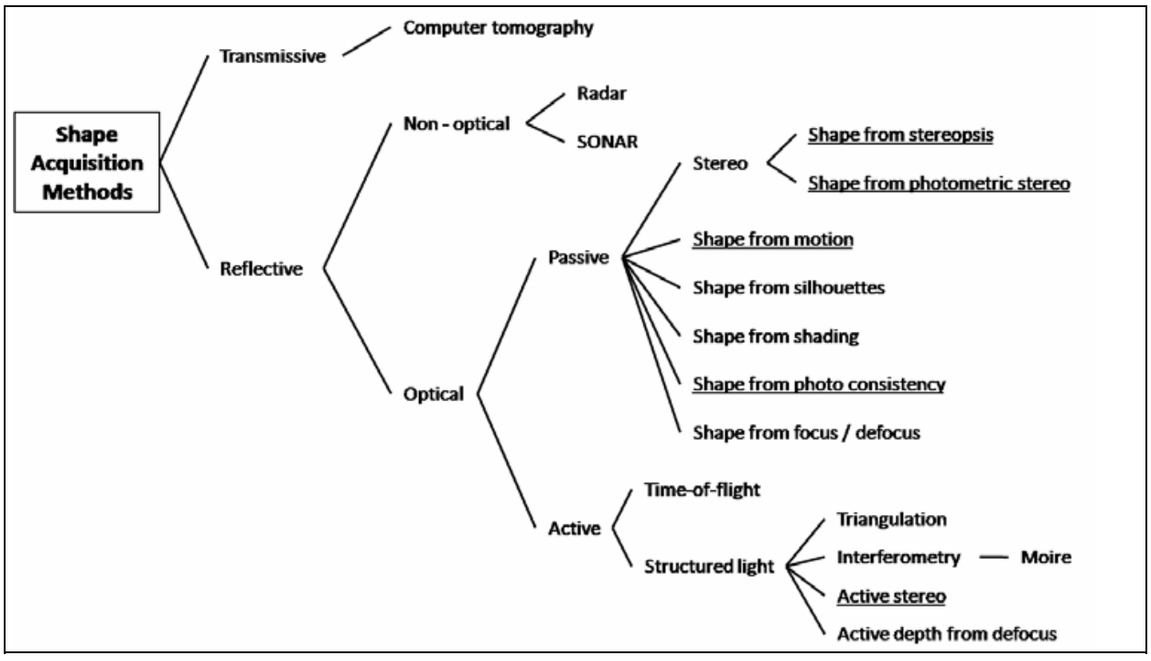


Fig. 1. Short taxonomy of shape acquisition methods.

the last fifteen years, theoretical developments in visual motion studies have established a unified framework for the treatment of the Structure from Motion (SFM) and Structure form Stereo (SFS) problems (Faugeras, 1992), also known as *3-D Reconstruction from Multiple Views*. 3-D reconstruction from multiple views involves extracting target features from one image, matching and tracking these features across two or more images, and using triangulation to determine the position of the 3-D target points relative to the camera. Visual motion methods have been well-studied, requiring densely-sampled image sequences. Instead, the trade-off in stereo vision is between the stereo correspondence problem and a more accurate and robust 3-D reconstruction (Scharstein & Szeliski, 2002). A large amount of works has addressed the correspondence problem, attempting to overcome the various difficulties of the large-displacement correspondence problem: occlusions, large rotations and disparities, photometric and projective distortions (Lucas & Kanade, 1981; Tomasi & Kanade, 1991). The problem of outliers has been solved by the deployment of robust estimation methods (Zhang et al., 1995). The motion and structure may be estimate by using several recursive schemes. Extended Kalman Filter is the most popular approach for jointly estimation of motion and structure with satisfactory results. However, the computation cost of this approach grows cubically with the amount of features, causing a great bottleneck for real-time performance. Stereo matching is one of the most active research areas in computer vision. Stereo matching is a hard problem due to ambiguity in un-textured and occluded areas. Only dominant features, such as points of interest, can be matched reliably. This motivates the development of progressive approaches (Szeliski & Scharstein, 2002). The reduced local disparity in search range makes progressive approaches very efficient in computation and robust. However, the seed initialization remains a computational bottleneck, although, robustness can be improved by enforcing the left-right symmetry constraint. Recently, Graph Cuts and Belief Propagation furnish combinatorial optimization frameworks in which the performance for global stereo algorithms are considerably improved according to an evaluation framework applied to a standard reference stereo data set (Scharstein & Szeliski, 2002). However, the underlying brightness constancy assumption of combinatorial optimization methods severely limits the range of their applications. The earliest attempts for 3-D reconstruction use methods based on volume intersection, as Shape from Silhouette (Laurentini, 1994). Traditional methods, such as stereo, handle large visibility changes between images by solving the correspondence problem between images. The most prominent approaches in 3-D reconstruction are the Voxel Coloring (Slabaugh et al., 2001) and the Space Carving (Kutulakos & Seitz, 1998). These approaches use the color consistency to distinguish surface points from other points in a scene. Cameras with an un-occluded view of a non-surface point see surfaces beyond the point, and hence inconsistent colors, in the direction of the point. Initially, the environment is represented as a discretized set of voxels, and then the algorithm is applied to color the voxels that are part of a surface in the scene. Another promising approach in 3-D reconstruction is the Marching Cubes algorithm for rendering iso-surfaces from volumetric scan data (Lorenson, 1987). The algorithm produces a triangle mesh surface representation by connecting the patches from all cubes on the iso-surface boundary. In underwater scenario, research is directed at exploring the use of vision, potentially in conjunction of other sensors, to automatically control unmanned submersibles, including positioning and navigation by utilizing a photo-mosaic as a two-dimensional visual map. Recent activities combine video imagery taken from multiple views of a scene to derive size and depth measurements and 3-D

reconstructions. These activities support (semi-) autonomous or operator-supervised missions pertaining to automatic vision guided station keeping, location finding and navigation, survey and mapping, trajectory following and online reconstruction of a composite image, search and inspection of subsea structures. These tasks require an accurate estimation of camera position, together with fast, accurate correspondence determination, particularly for real-time registration. Common sources of error include non-planar seafloor, moving objects, illumination variations, transect superposition, positioning drift. A number of studies over the last several years have also addressed the 3-D reconstruction for various applications. Khamene and Negahdaripour incorporate cues from stereo, motion and shading flow for 3-D reconstruction in underwater (Khamene & Negahdaripour, 2003); Majidi and Negahdaripour suggest the use of 3-D reconstruction for global alignment of 3-D sensor positions (Madjidi & Negahdaripour, 2005); Nicosevici et al introduce 3-D reconstruction from motion video and representation of the surface topography by piecewise planar surfaces for the construction of orthomosaics (Nicosevici et al., 2005). Hogue et al have developed a stereo vision-inertial sensing device deployed to reconstruct complex 3-D structures in both the aquatic and terrestrial domains (Hogue et al., 2007). The sensor temporally combines 3D information, obtained using stereo vision algorithms with a 3 DOF inertial sensor. The resulting point cloud model is then converted to a volumetric representation and a textured polygonal mesh is extracted using the Marching Cubes algorithm (Lorenson, 1987).

3. Algorithmic framework

3.1 Overview

This section discusses about the framework for 3-D mosaic reconstruction of a seabed based on the optic sensing technique known as *Shape from Stereopsis*. In Shape from Stereopsis two stereo frames are acquired at the same time, so that normally triggered frame grabber (or other techniques that guarantee synchronism in stereo sequence) is required. However, synchronized stereo sequences use expensive equipments and present some technical limitations, such as low frame rate and poor time resolution. For the previous reasons the attention is focused in stereo reconstruction and 3-D structures estimation by using unsynchronized cameras. Svedman proposes to acquire three images sequentially from the left, right, and again from the left camera. A virtual image from the left camera synchronized with the right image is created by interpolating matching interesting points in the two left images (Svedman et al., 2005). Others approaches estimate the time difference between views, synthesizing synchronous image pairs for dense depth information estimation. The depth estimation is based on stereo correspondence evaluation. Depth and ego-motion estimation leads to the ill-posed stereo correspondence problem. Although for egomotion estimation a sparse set of correspondence points is sufficient, depth estimation requires dense correspondences. In underwater environment normalized cross-correlation is generally employed for the robustness to brightness gain. In dense stereo matching, the best performing algorithms use either the Belief Propagation (Sun et al., 2002) or Graph Cuts (Boykov et al., 2001). However, these algorithms are tested on standard data sets under restricted conditions and/or controlled environments (small disparity range, image sequences that satisfy the brightness constancy assumption). The brightness constancy model is often violated, e.g. for most images in the JISCT collection (Bolles, 1993). Zhang proposes a revised data cost to improve global stereo matching algorithms (Graph Cuts,

Belief Propagation) in underwater environment (Zhang, 2005). 3-D mosaic construction requires two-frames registration and then ego-motion estimation. Iterative Closest Point (ICP) algorithm is used for 3-D registration (Besl & McKay, 1992). The epiflow framework proposed by Zhang is based on the integration of motion and stereo epipolar geometries for ego-motion tracking (Zhang, 2005). Summarizing, an inexpensive asynchronous stereo framework for 3-D seabed reconstruction and orthomosaic is presented. In order to achieve a metric reconstruction, a calibration phase is necessary, so that the asynchronism in stereo sequence can be an issue for proper calibration. To overcome this problem a new stereo frame selection scheme based on the Epipolar Gap Evaluation (EGE) is used (Section 5). In order to handle the stereo correspondence problem, local and global matching methods are evaluated in Section 6, with adoption of a suitable similarity measure (ZNCC) and image enhancement technique (CLAHE). Finally, in Section 7 the registration approach is presented conjugating Epiflow and ICP schemes under a simplified motion model.

3.2 Framework architecture

The main building blocks of 3-D seabed mosaic reconstruction system are the following: (a) shape acquisition, (b) depth estimation and (c) mosaic registration and rendering. In Fig. 2 the overall system architecture is shown. The block 1 deals with shape acquisition, asynchronous stereo sequence calibration by using the Epipolar Gap Evaluation (EGE)

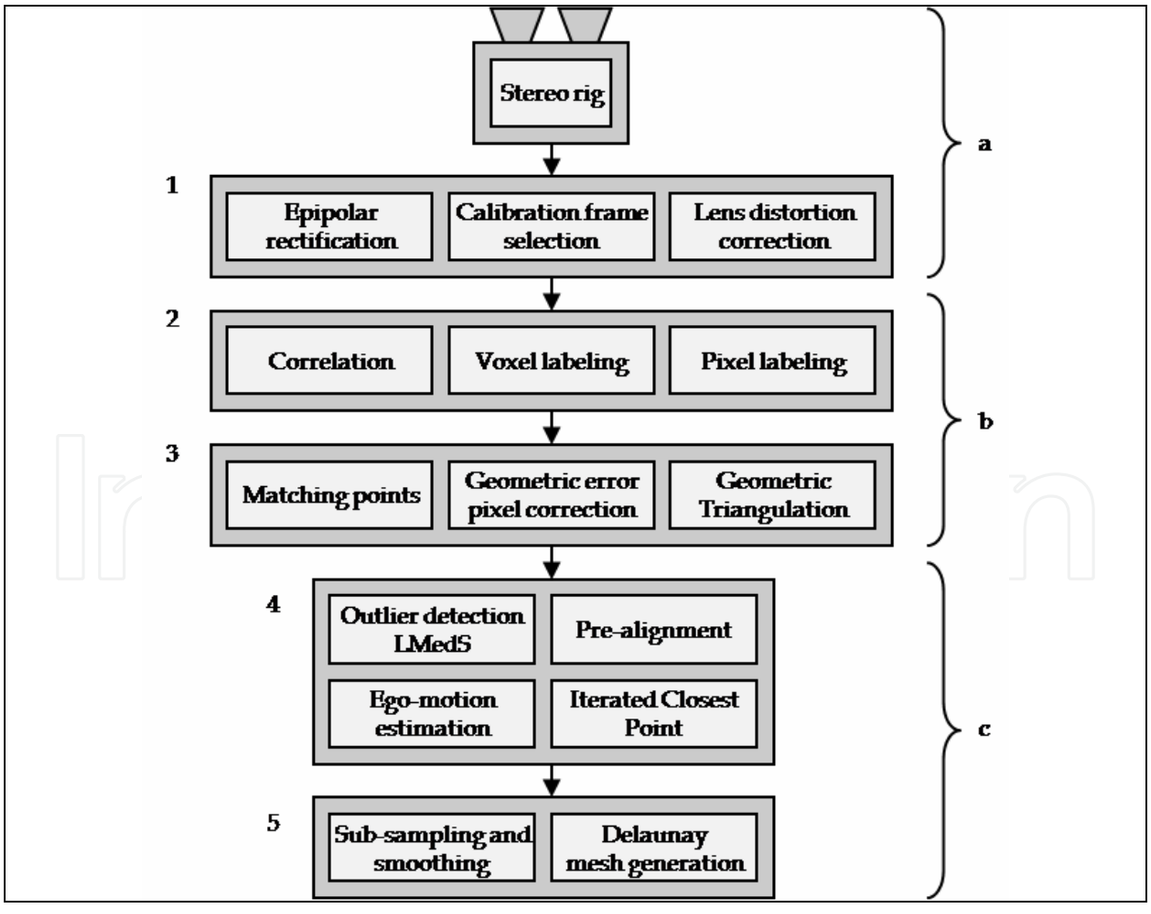


Fig. 2. Overall framework architecture.

scheme, lens distortion correction and histogram equalization. Depth estimation is accomplished in blocks 2 and 3: the former regards the disparity map estimation and the latter computes the subsequent depth map by using geometric triangulation. As described in the following, local and global methods are used in stereo disparity calculation. The block 4 involves fusion and registration of 3-D reconstructions, through system ego-motion estimation, pre-alignment and ICP registration refinement. Finally, triangular Delaunay interpolation and rendering take place in block 5.

4. Structure from stereopsis in underwater environment

4.1 Overview

Starting from two different views of the same scene, it is possible to obtain a 3-D structure reconstruction. For simplicity, the discussion is restricted to scenes consisting of points only; further information about multi-view reconstruction methods can be found in (Hartley & Zisserman, 2003). The input for 3-D reconstruction phase is a set of correspondences in left and right camera images. Let the camera matrices $\mathbf{P}^L$ and $\mathbf{P}^R$, that describe the correspondences $x_i^L \leftrightarrow x_i^R$ in terms of $\mathbf{P}^L X_i = x_i^L$ and $\mathbf{P}^R X_i = x_i^R$, where the point X_i projects to the two given data points x_i^L and x_i^R (see Fig. 3). Unfortunately, neither the projection matrices $\mathbf{P}^L$ and $\mathbf{P}^R$, nor the points X_i and $x_i^L \leftrightarrow x_i^R$ are known *a priori*. The camera matrices estimation is part of the calibration activity (Subsection 5.2), whereas the correspondences estimation $x_i^L \leftrightarrow x_i^R$ deals with the stereo correspondence problem (Subsection 6.2 and 6.3), and the estimation of point X_i is the first step for 3-D reconstruction by using geometric triangulation (Subsection 6.4).

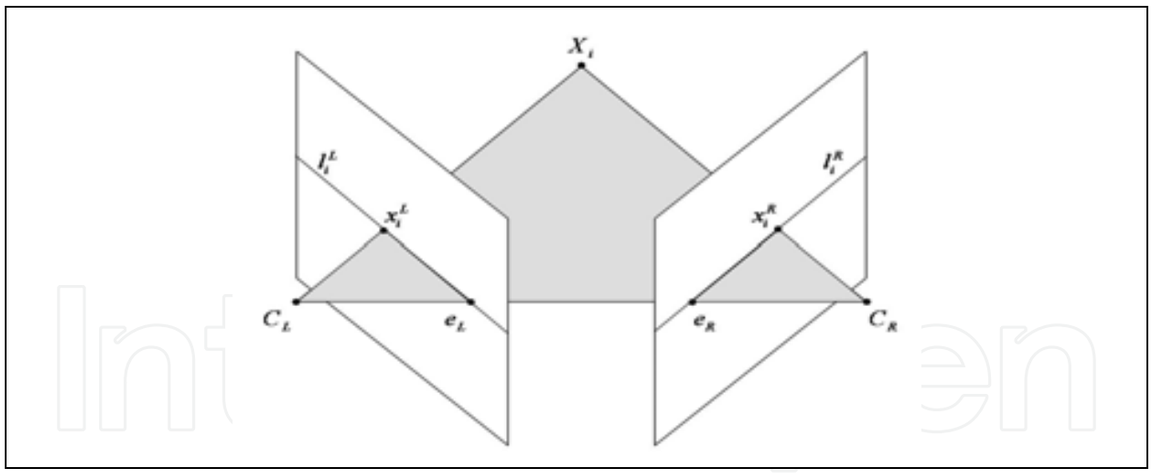


Fig. 3. Epipolar geometry: the optic rays passing through the X_i point and C_L , C_R optical centers intersect the left and right image planes in (x_i^L , x_i^R) image points, respectively.

The main tool in 3-D reconstruction from two views is the *Fundamental Matrix* (Faugeras, 1992), which represents geometrically the constraint for the projected image points (x_i^L , x_i^R) of the same 3-D point X_i. This constraint, also called *Epipolar Constraint*, arises from the coplanarity of the camera centers, the images points (x_i^L , x_i^R) and the space point X_i. Given the Fundamental Matrix $\boldsymbol{F}$, each pair of matching points (x_i^L , x_i^R) must satisfy the following relation:

$$\left(x_i^R\right)^T \mathbf{F} x_i^L = 0 \tag{1}$$

with *F* a 3×3 matrix of rank 2. Furthermore, Eq.1 can be rewritten as:

$$\left(x_i^R\right)^T l_i^R = 0 \tag{2}$$

where

$$l_i^R = \mathbf{F} x_i^L \tag{3}$$

is the *Epipolar Line* associated with the left point x_i^L. In other words, the Eq.2 states that the point x_i^R belongs to the epipolar line l_i^R. The discussed concepts are summarized in Fig. 3. Points e_L and e_R are the *epipoles*, i.e. the points of intersection of the *baseline* (the line joining the camera centers) with the image planes. A 3-D projective reconstruction can be obtained from the knowledge of the Fundamental Matrix alone, so that the pair of camera matrices can be determined up to a 3-D projective ambiguity and the Fundamental Matrix can be estimated directly from a set of point correspondences. Instead, if the camera matrices are known, by means of calibration activity, a metric reconstruction is possible. Therefore, a reliable calibration activity is needed since a 3-D metric reconstruction of the seabed surface is desired. The calibration activity can be affected by three kind of error: synchronization error, algorithm error and matching error. Algorithm error and marker matching error deal with the specific calibration procedure, whereas synchronization error relates the synchronism in stereo frames. The synchronization error can be minimized by computing calibration matrices for each stereo pair in which calibration pattern is viewed and, then, by evaluating the *Epipolar Gap* (EG). Let a pair of matching points x_i^L and x_i^R, the Epipolar Gap is defined as the Euclidean distance between one of these points (i.e. x_i^R) and the epipolar line associated with the other one, as follow:

$$\mathrm{EG_i} = \mathrm{dist}\left(x_i^R, l_i^R\right) \tag{4}$$

If x_i^L and x_i^R matches exactly, Eq.2 is satisfied and then EG_i=0; otherwise EG_i provides a measure (the Euclidean distance) for the epipolar gap between x_i^L and x_i^R. The calibration activity is realized by a careful selection of a stereo pair from the video sequence in which the calibration pattern is acquired. Since there are many stereo pairs in which the calibration pattern is acquired, the Epipolar Gap Evaluation (EGE) is used as selection criterion. The selection of calibration stereo pairs is motivated by the asynchronism of the acquisition system based on a non-triggered frame grabber. Aim of this step is to minimize the acquisition time distance between the left and the right frame in the stereo pair, selecting a stereo pair affected by minimum time difference. The calibration frame selection requires a large amount of calibration parameters (camera matrices), therefore a calibration pattern (an object with known metric measures) must be employed and detected in the acquired video sequence. In the next section, a calibration object (its marker points imprinted on pattern) detection from video sequence is discussed.

4.2 Metric calibration
4.2.1 Pattern detection

A calibration object can be acquired to provide metric correspondences between points in the image coordinates and points in real world coordinates. Generally, an accurate camera coordinate system-based positioning of the calibration object is hard, especially in underwater environment. Hence, the relationship between target and camera coordinate systems needs to be recovered from reliable correspondence estimation. Several camera calibration methods employing 2-D and 3-D calibration targets were proposed in (Tsai, 1987; Heikkia & Solven, 1996). Calibration algorithms use an object on which a square checker pattern with a known size is printed. In this work a pair of still images of 3-D checkerboard is required, according to the method proposed in (Tsai, 1987). Normally, in underwater applications the checkerboard is arranged on the sea floor and the stereo acquisition system moves around during calibration activity (see Fig. 4.b), in contrast with the ideal calibration conditions in which both checkerboard and stereo acquisition system are more likely to be fixed (see Fig. 4.a). The corners of the checker pattern are used as control points for the camera calibration algorithm (see Fig. 5.a). Bakstein examines several pattern geometries to

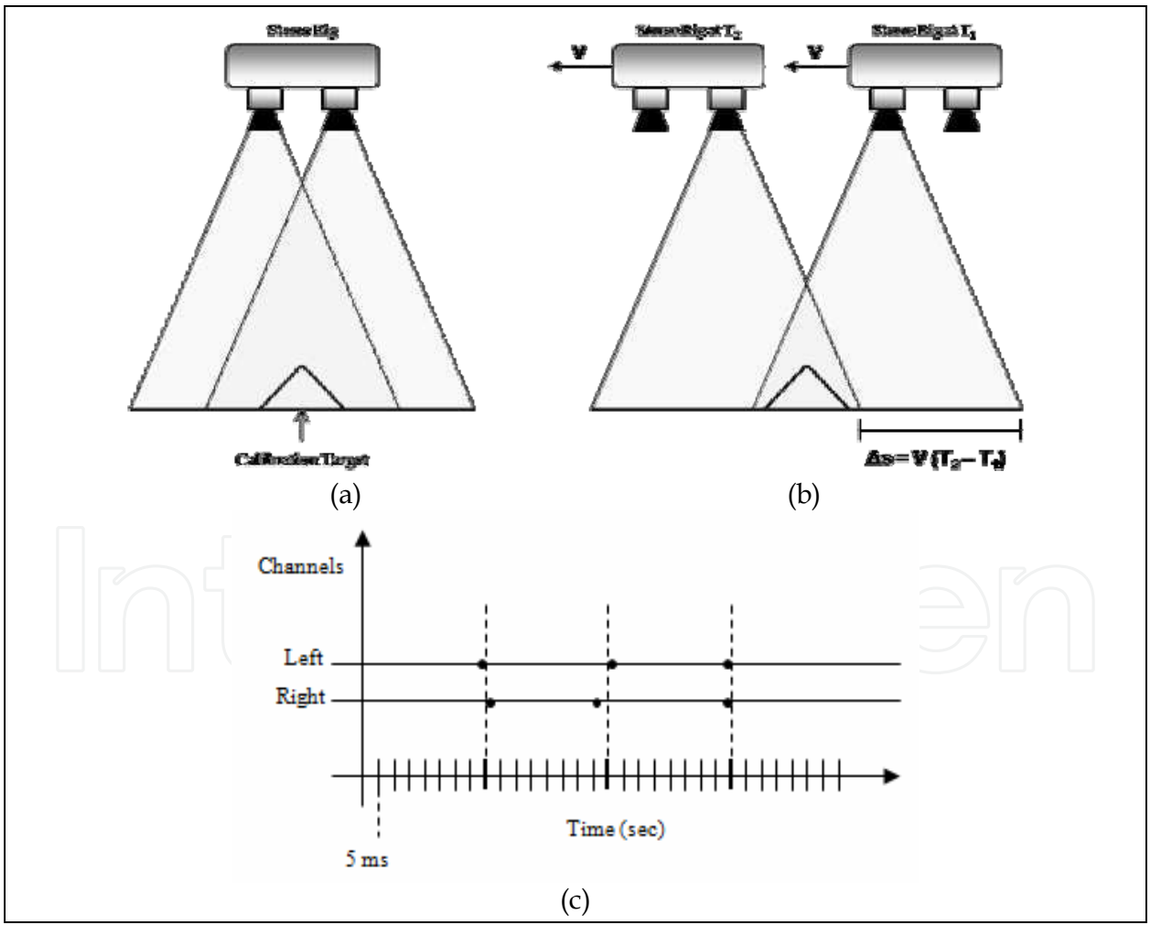


Fig. 4. *a*) Ideal acquisition conditions for proper calibration; *b*) Real calibration conditions: the system moves with velocity **v**; *c*) Stereo acquisition timing at 25 fps. Dashed lines denote ideal acquisition time instants, while dots denote effective left/right acquisition time instants.

derive the control points for camera calibration by using an automated thresholding strategy, so that the checker pattern is chosen for the very low sensitivity by thresholding errors (Bakstein, 1999). Furthermore, a black-and-white checker pattern provides very good contrast and size of the squares could be customized to the object distance for better perception. Asynchronism in stereo acquisition is a problem in the calibration activity, since the epipolar geometry constraint isn't satisfied. This aspect is further explained in Fig.4.c. A 25 fps acquisition system acquires two calibration checkerboard images, but due to asynchronism the effective acquisition time instants (represented by dots in figure) lay around the ideal ones (represented by dashed lines). Unlike others techniques based on view synthesis or time estimation between views, the presented approach tries to solve this problem making only use of calibration parameters encoded in camera matrices. For a generic acquisition system, the calibration error is defined as:

$$\varepsilon = \varepsilon_{syn} + \varepsilon_{cal} + \varepsilon_{match}\,, \tag{5}$$

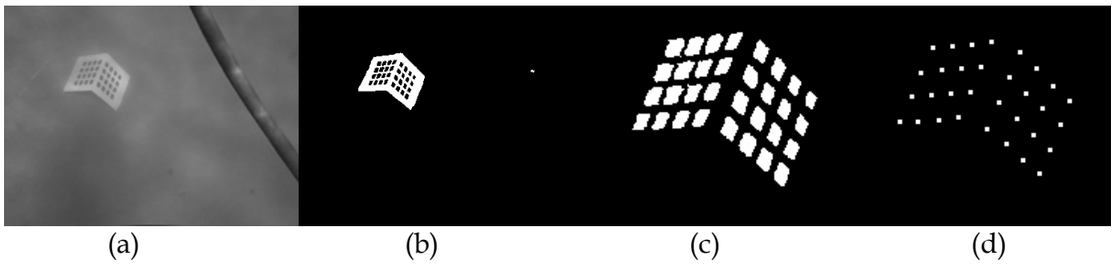


Fig. 5. Calibration object: *a*) Original image; *b*) Segmented image; *c*) Pattern's rectangles; *d*) Centroids.

where ε_{syn} is the synchronization error, ε_{cal} is the calibration algorithm error and ε_{match} is the error affected by calibration pattern points selection. The calibration error minimization takes place by evaluating the epipolar gap for each stereo pairs in which calibration pattern is acquired. The epipolar gap estimation is more accurate as greater amount of calibration pattern images is available. This calibration method detects the calibration object in a rotation independent way and it localizes checkerboard square corners corresponding to calibration target ones by using line-fitting and label-sorting strategies. In order to estimate intrinsic and extrinsic cameras parameters, checkerboard square vertexes must be detected in acquired images and matched with corresponding checkerboard square vertexes in real calibration target. Standard corner detection approaches (Harris & Stephens, 1988) does not give satisfactory results by detecting checkerboard square vertexes, since underwater images present poor quality. To overcome the problem, a projective-geometric approach is investigated, according to the following discussion. Preliminarily, the calibration target is segmented-out (see Fig.5) from undistorted image (Fig.5.a) by using well-known morphological operations (such as connected components for small object removing and thresholding binaryzation). Once the pattern is detected in the image, the checkerboard square vertexes are identified through a line-fitting approach as explained in the following. As well-known, image plane projection preserves the intersection property. Hence in each image the pattern target appears to have squares aligned with the eight non-intersecting lines L_q (see Fig.6.a) and with four non-intersecting lines L_t for each pattern side (see Fig.6.b) due to prospective projection. The previous observation provides a scheme for labeling and

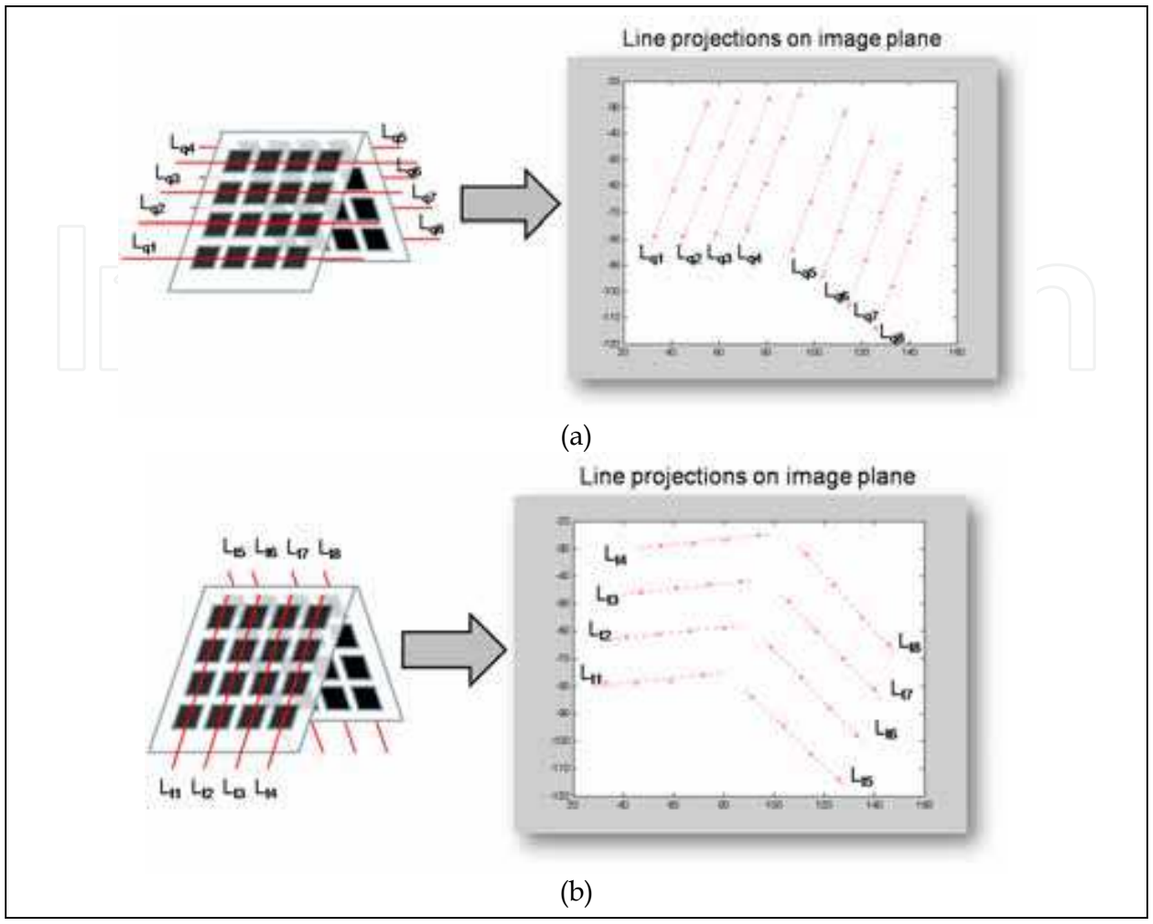


Fig. 6. *a*) projection of 8 non intersecting lines on image plane; *b*) projection of 4 non intersecting lines for each side on image plane.

sorting the centroids as shown in Fig. 7.a. At each iteration the line passing through centroids P_1 and P_i is considered as the reference line r, where the P_1 is the outer centroid with the lowest abscissa. Let L be a list containing $\left(j, d_j^r, d_j^o\right)$ triples where j is the centroid label, d_j^r is the Euclidean distance of P_j from line r, and d_j^o is the Euclidean distance of P_j from axes origin O. Firstly, centroids are sorted by distance d_j^r producing a new list L' in which collinear centroids are grouped. Each group of centroids in L' is newly sorted by distance from the axes origin. The loop exit test is based on the evaluation of errors E_θ and E_d. As shown in Fig.7.c, E_θ measures the collinearity along L_q as a maximum shifting from average angular coefficient:

$$E_\theta = \max_{i=1,\ldots,32} \left| \frac{\theta_i - \bar{\theta}}{\theta_{\max} - \theta_{\min}} \right|, \tag{6}$$

Experimentally, E_θ <0.7 guarantees a satisfactory collinarity along Lq lines. Instead, E_d measures the maximum distance of middle centroids from line passing through outers centroids. A maximum value of 2 pixel for E_d guarantees a good collinearity along L_t lines.

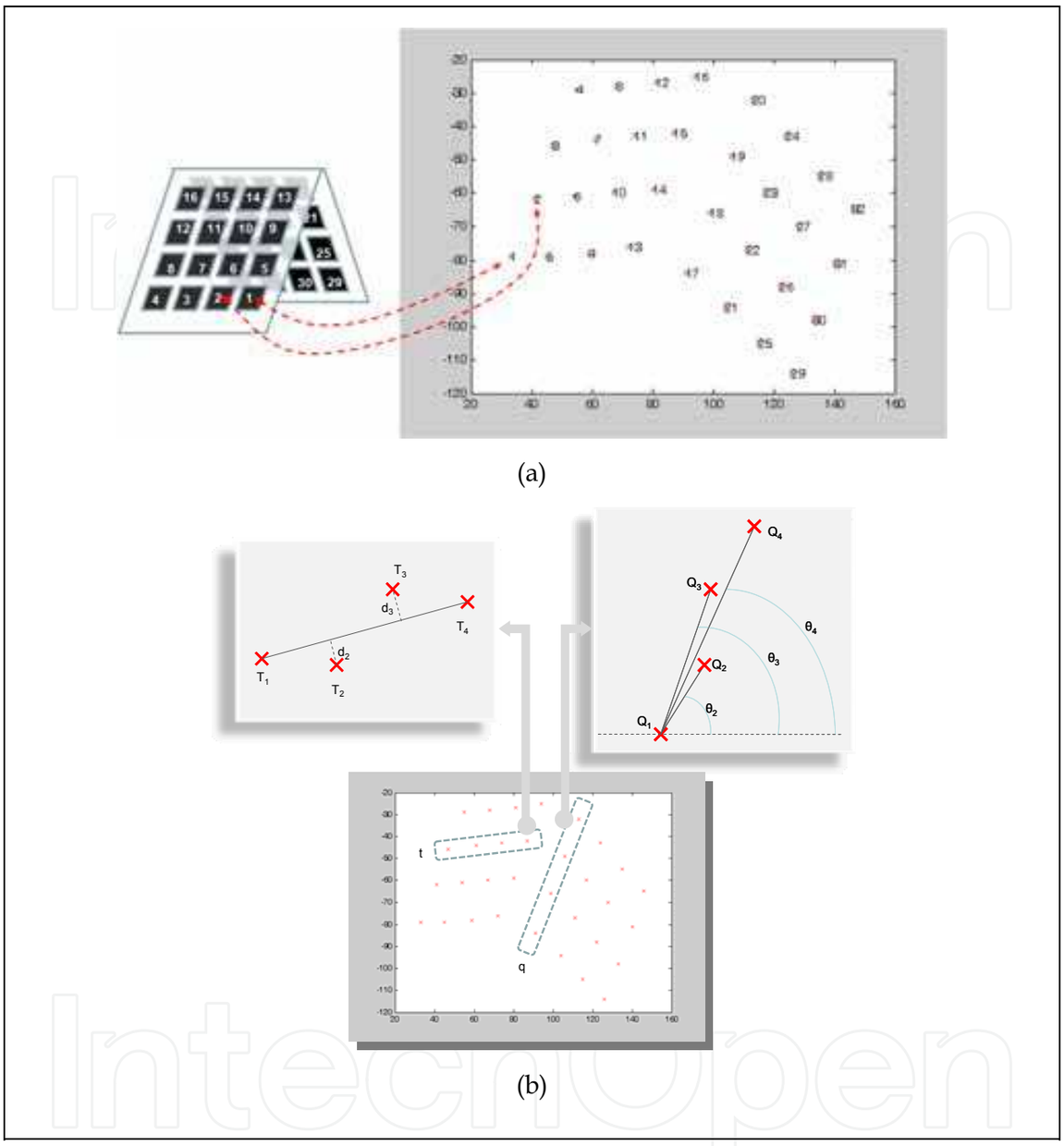


Fig. 7. *a*) Centroid labels mapping and sorting criterion; *b*) Test to evaluate centroid collinearity with L_T and L_Q lines.

Centroid labels are sorted out by evaluating collinearity and distance from the reference line, according to the pseudocode provided in Tab. 1. The goal of the next step is to locate the four vertexes for each labelled and sorted checker square of the calibration pattern. Therefore, for each checkerboard square Northern, Southern, Eastern and Western points are located as intersection of square borders with the two estimated sets of lines L_q and L_t estimated in the previous step (Fig.8.a,b).

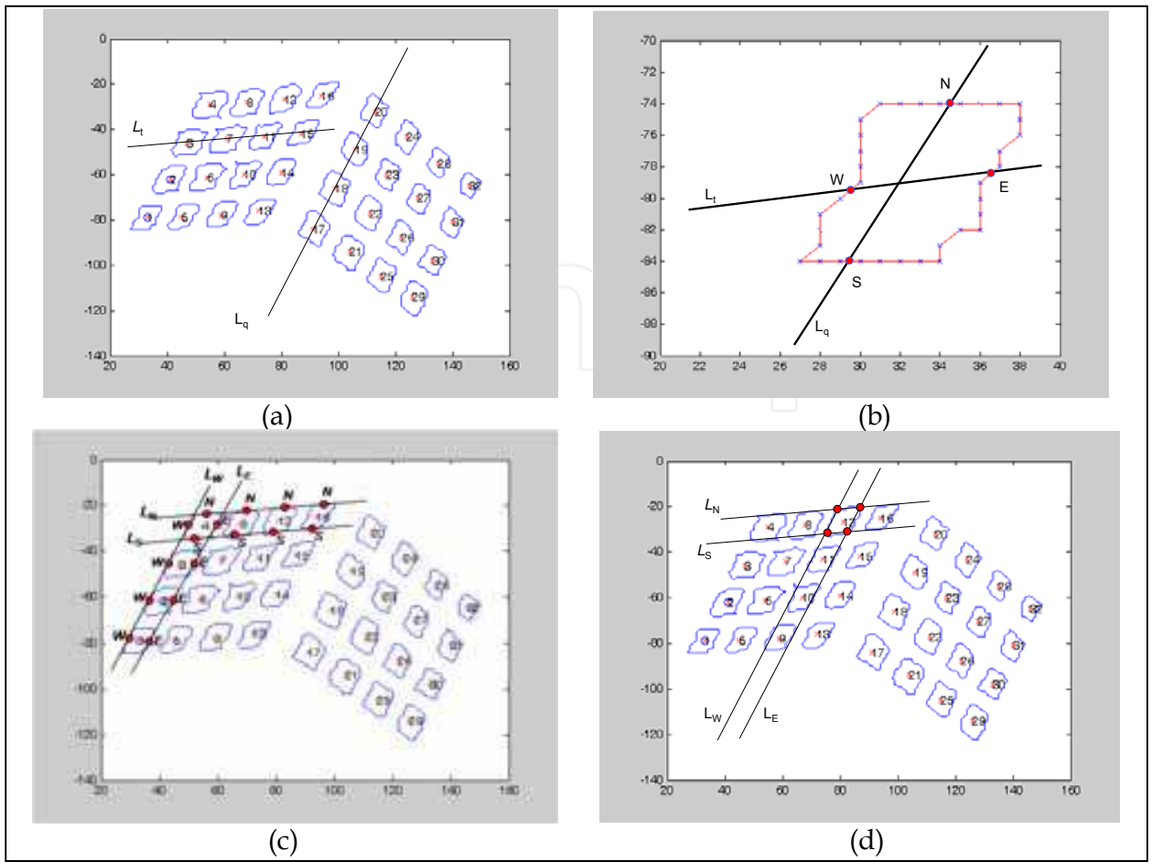


Fig. 8. *a*) Overall viewing of intersections; *b*) Single checker square intersections; *c*) L_N, L_S, L_E, L_W lines are estimated by fitting of N, S, E, W points respectively in each pattern side; *d*) Square vertexes obtained by L_N, L_S, L_E, L_W lines intersection.

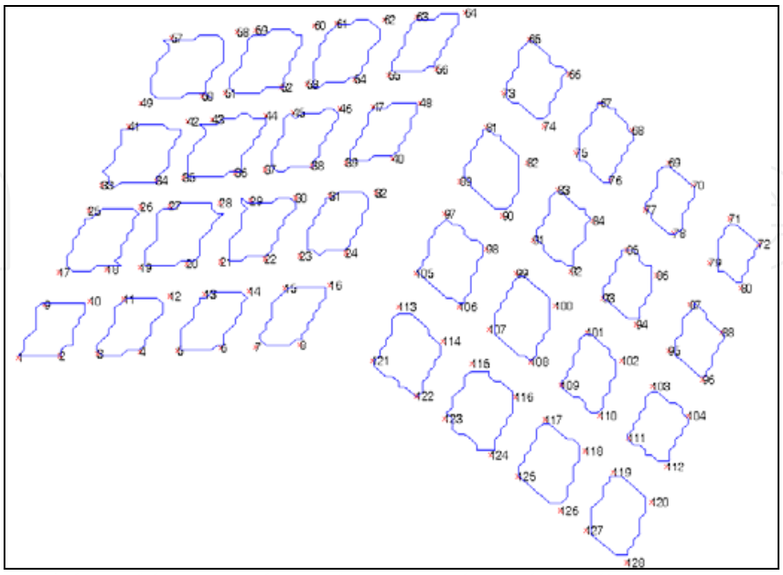


Fig. 9. The result of the calibration object detection algorithm: all square vertexes are labelled and ordered coherently with calibration object marker points.

For i = 1 **to** 32 (amount of checkerboard squares)
r is the line passing through points P_1 and P_i
$\forall j = 1,...,32 : d_j^r = dist\left(P_j, r\right), d_j^O = dist\left(P_j, O\right)$
$L = \left\{ j = 1,...,32 \mid \left(j, d_j^r, d_j^O\right) \right\}$
L' = Sort (L) with respect d_j^r
L'' = Sort (L') with respect d_j^O
if (E_θ < 0.7 and E_d < 2) **exit loop**
next i

Table 1. Pseudocode for centroids labeling algorithm.

Once N, E, S, W points are located, the lines L_N, L_E, L_S, L_W can be estimated through line fitting as represented in Fig.8.c. L_N, L_S, L_E, L_W lines are estimated by fitting of N, S, E, W points respectively in each pattern side. Finally, for each checker square the four vertexes are obtained by lines intersection as $L_N \cap L_W$, $L_N \cap L_E$, $L_S \cap L_W$, $L_S \cap L_E$ (see Fig.8.d). The result of the calibration object detection algorithm is shown in Fig.9, where all square vertexes are labelled and ordered in image coherently with the object marker points of the calibration pattern in real world coordinates.

4.2.2 Epipolar gap evaluation

The calibration activity receives a careful selection of a stereo pair from the video sequence in which the calibration pattern is acquired. The selection of calibration stereo pairs is motivated by the asynchronism of the acquisition system based on a non-triggered frame grabber. Since there are many stereo pairs showing the calibration pattern, the *Epipolar Gap Evaluation* (EGE) is used as a selection criterion. Aim of this step is to minimize the acquisition time distance between the left and the right frame in the stereo pair; the purpose is to select a stereo pair affected by minimum time difference (Fig. 4.c). The selection criterion can be schematized with two sequential activities such as: the *Calibration Matrices Estimation* and the *Epipolar Gap Evaluation*. The scheme is depicted in Fig.10.a. The first step receives as input the h-th stereo frame pair, in which the calibration pattern is acquired, providing the corresponding calibration matrices as output, according to Tsai's calibration method. The second block receives as input the calibration matrices from the previous block and the stereo frame pair in which the seabed surface under reconstruction is acquired. The EGE provides the calibration matrices that minimize the Epipolar Gap on the overall reconstructing frames set. In EGE, a sparse set of stereo matches

$$\mathbf{M}_k^{\mathrm{L}} \times \mathbf{M}_k^{\mathrm{R}} = \left\{ x_{k,1}^{\mathrm{L}}, \ldots, x_{k,n_k}^{\mathrm{L}} \right\} \times \left\{ x_{k,1}^{\mathrm{R}}, \ldots, x_{k,n_k}^{\mathrm{R}} \right\}, \quad n_k = \left|\mathbf{M}_k^{\mathrm{L}}\right| = \left|\mathbf{M}_k^{\mathrm{R}}\right| \tag{7}$$

is computed for the k-th stereo pair $\left(I_k^L, I_k^R\right)$. Hence, each set $\mathbf{M}_k^{\mathrm{L}} \times \mathbf{M}_k^{\mathrm{R}}$ is compared with the Fundamental Matrix $\mathbf{F}_h$ for frame rejection purpose. The frame rejection is based on the Maximum Epipolar Gap (MEG) evaluated as:

$$\mathrm{MEG}_{k,h} = \max_{j=1,\ldots,n_k} \left\{ \mathrm{EG}_{k,j}^h \right\} \tag{8}$$

where $\mathrm{EG}_{k,j}^h$ is the epipolar gap defined in Eq.4 and here generalized as follow:

$$\begin{aligned} \mathrm{EG}_{k,j}^{h} &= \mathrm{dist}\left(x_k^R, l_{k,j,h}^R\right) \\ l_{k,j,h}^R &= \mathbf{F}_h x_{k,j}^L \,,\ \forall k = 1,\ldots,r \,,\ \forall j = 1,\ldots,n_k \,,\ \forall h = 1,\ldots,s \,. \end{aligned} \tag{9}$$

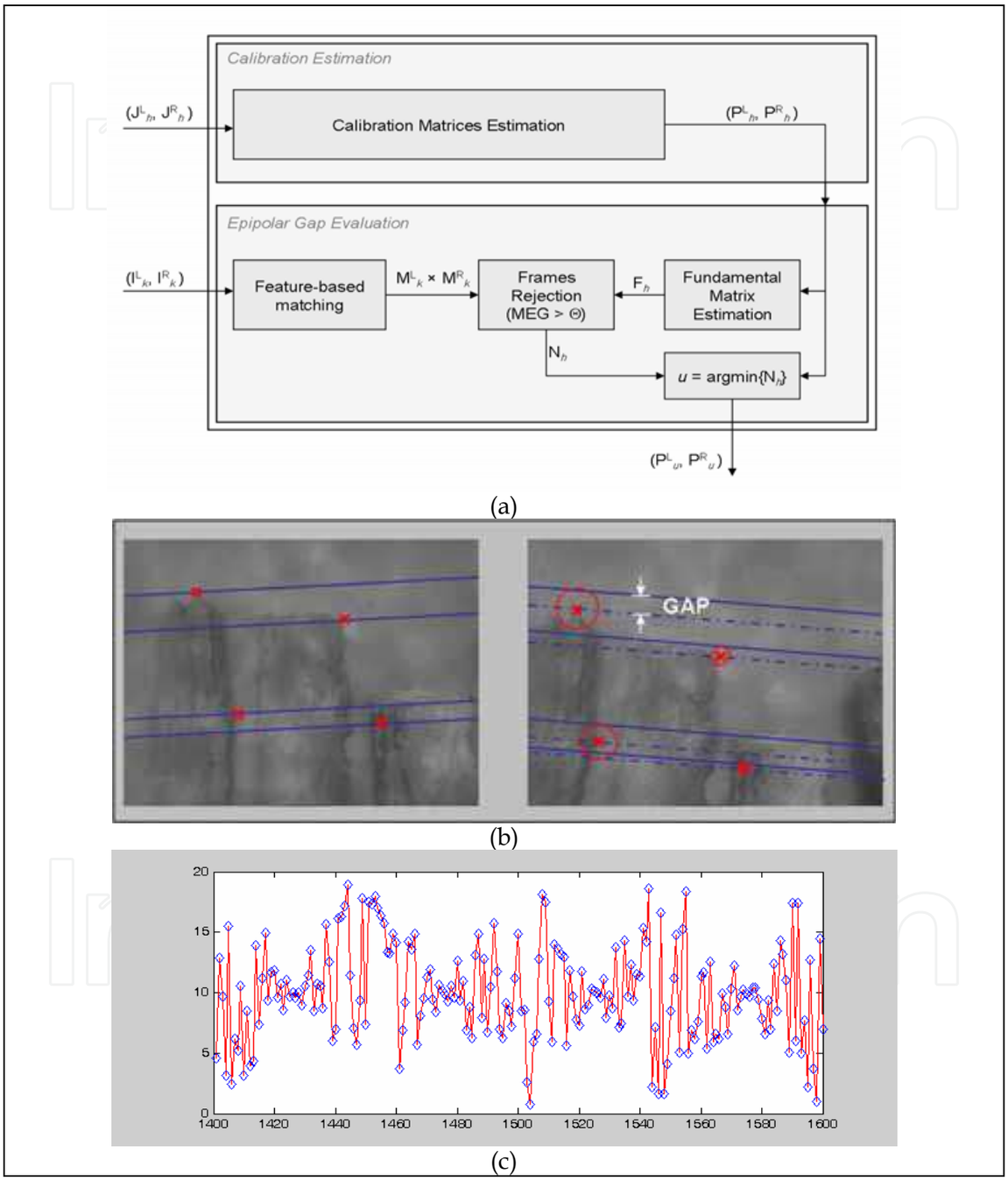


Fig. 10. *a*) Block scheme for the calibration frames selection criterion based on the Epipolar Gap Evaluation. *b*) Epipolar gap evaluated in 4 points ; *c*) Experimental result. The MEG is evaluated within calibration sequence. The optimal value u in Eq.11 is achieved for frame h = 1504 with $\theta \approx 1$ pixel.

Thus, MEG is used as a filtering rejection stereo pairs for which the corresponding value exceeds a prefixed threshold. If the amount of rejected couples is given by:

$$N_h = \left| \left\{ k = 1, \ldots, r \mid \mathrm{MEG}_{h,k} > \theta \right\} \right| \tag{10}$$

then, the chosen calibration pairs will be $\left(\mathbf{P}_u^{\mathrm{L}}, \mathbf{P}_u^{\mathrm{R}}\right)$ with

$$u = \arg \min_{h=1,\ldots,s} \left\{ N_h \right\}. \tag{11}$$

5. Seabed reconstruction

5.1 Overview

Stereo reconstruction is based on the epipolar geometry as discussed in Section 5.1 and illustrated in Fig. 3. Let the camera matrices $\mathbf{P}^{\mathrm{L}}$ and $\mathbf{P}^{\mathrm{R}}$ (two 4×4 projection matrices in homogeneous coordinates) and (x_i^L , x_i^R) two corresponding points in the left-right images, the epipolar constraint in Eq.1 must be satisfied. The constraint may be interpreted geometrically in terms of the rays in space corresponding to the two image points x_i^L and x_i^R ; in particular each point lies on the epipolar line of the other point according to Eq.2 and Eq.3. Referring to Fig. 3 this means that the two rays $\overrightarrow{X_i C_L}$ and $\overrightarrow{X_i C_R}$ back-projected from image points x_i^L and x_i^R lie in a common *epipolar plane* passing through the two camera centres C_L and C_R . Since the two rays lie in a plane, they will intersect in the 3-D point X_i related to the points x_i^L and x_i^R through cameras projections. X_i can be determined as intersection of the two rays back-projected from corresponding points x_i^L and x_i^R by using a stable triangulation method (Hartley & Zisserman, 2003). Unfortunately, these corresponding points are not known a-priori leading to the *Stereo Correspondence Problem*, so that, given the point x_i^L , the problem is to determine another point x_i^R such as these two points are the projections of the same 3-D point onto the left and the right image plane, respectively. Epipolar constraint allows to simplify the matching algorithmic complexity by reducing the searching area close to the epipolar lines. Further algorithmic simplifications are possible by means of *Epipolar Rectification* process, that is, the calculation of an appropriate projective transformation producing as output a stereo image pair in which the epipolar lines are transformed to lines parallel with the *x*-axis in the image plane. Rectification allows to perform stereo matching along horizontal scan lines. In this situation, the horizontal offset between corresponding image points $d_i = x_i^L - x_i^R$ is referred as *disparity*. Stereo correspondences are estimated by using matching algorithms that received rectified stereo image pairs in input and provide a dense disparity map as output. The disparity map is a range image in which each pixel describes the disparity value estimated respect to a reference image. Given a left disparity map represented with a 256 grey levels, each right matching point is:

$$x_i^R = x_i^L - \frac{p_i}{256}\Delta - d_{\min} \tag{12}$$

where Δ is the disparity range, $d_{\min}$ is the minimum disparity value, and p_i is the disparity map value corresponding to the x_i^R point (in 256 gray levels). In this framework, for stereo

disparity maps calculation local and global methods are explored. In both cases, the major difficulties are related to the underlying assumption of brightness constancy in the stereo algorithms, systematically violated in underwater environment (Bolles et al., 1993).

5.2 Stereo correspondences: local methods

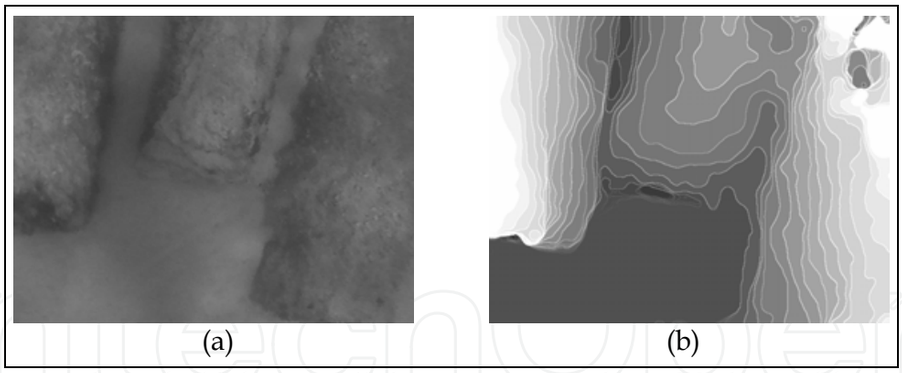
Local approaches are based on a search windows horizontally displaced in one view with respect to the other view for each allowed disparity. Matching measures generally used in local methods are *Sum of Absolute Differences* (SAD), *Sum of Squared Differenced* (SSD), *Normalized Cross Correlation* (NNC) and *Sampling Insensitive Measurement*. In this context the best choice is the *Normalized Zero-Mean Cross-Correlation* (ZNCC), that is quite insensitive at brightness variations typical of underwater environment. ZNCC measure is defined as:

$$\text{ZNCC}\quad(x,y,d)=\frac{\sum_{w_x,w_y\in W}\left(\mathrm{I_L}\left(x+w_x,y+w_y\right)-\bar{\mathrm{I}}_{\mathrm{L}\,x,y}\right)-\left(\mathrm{I_R}\left(x+w_x+d,y+w_y\right)-\bar{\mathrm{I}}_{\mathrm{R}\,x,y}\right)}{\sqrt{\left(\sum_{w_x,w_y\in W}\left(\mathrm{I_L}\left(x+w_x,y+w_y\right)-\bar{\mathrm{I}}_{\mathrm{L}\,x,y}\right)^2\right)\cdot\left(\sum_{w_x,w_y\in W}\left(\mathrm{I_R}\left(x+w_x+d,y+w_y\right)-\bar{\mathrm{I}}_{\mathrm{R}\,x,y}\right)^2\right)}} \tag{13}$$

from which the disparity is calculated as

$$\text{Disparity(x, y)} = \max_{d_{\mathrm{MIN}}\le d\le d_{\mathrm{MAX}}} \text{ZNCC}(x,y,d)\,. \tag{14}$$

In Fig.11 is reported en example of ZNCC stereo matching applied on a rectified seabed image.



(a) (b)

Fig. 11. *a*) original rectified left image; *b*) disparity map with ZNCC obtained from (a).

5.3 Stereo correspondences: global methods

Unlike local approaches, in global approaches the correspondence problem is stated in term of a cost function subject to minimization. The most critical choice is the optimization technique, that essentially falls into two main categories: continuous or discrete optimization. In discrete minimization, the best stereo matching estimation use combinatorial optimization via Belief Propagation (Sun et al., 2002) or Graph Cuts (Boykov et al., 2001). Two Graph Cuts formulation based on the expansion moves scheme are now investigated: Voxel Coloring (Kolmogorov & Zabih, 2002) and Pixel Labeling (Boykov et al., 1998). Graph Cuts based methods address cost function optimization by an energy function that can be represented, in general, as:

$$E(L) = D(L) + V(L), \tag{15}$$

where L is a disparity labeling (i.e. a label set in which each label represents a disparity value), $D(\cdot)$penalizes the variation from observed intensities (data penalties term), and $V(\cdot)$ is the smoothness term that encourages spatial coherence by penalizing discontinuities between neighboring pixels. When cost functions involve convex smoothness term a global minimum is indeed reachable via Graph Cut in polynomial time. Unfortunately, convex smoothness terms do not represent an optimal choice for the stereo problem, in which it is preferable to use a non-convex smoothness function to avoid the over-penalizing of large jumps in disparity.
One simple non-convex function commonly used in stereo vision is the Potts model (Potts, 1952). Despite the simplicity of the Potts model, the resulting problem formulation is proven to be NP-hard (Kolmogorov & Zabih, 2004). However, a strong local optimum can be estimated for non-convex smoothness terms by application of expansion move methods (Boykov et al., 2001). Image enhancement with *Contrast Limited Adaptive Histogram Equalization* (CLAHE) (Pizer et al., 1987) is often employed in underwater imaging application to mitigate brightness variations effects. Both voxel coloring and pixel labeling methods suffer for underwater violation of brightness constancy assumption. However, CLAHE enhancement permits to mitigate brightness variation effects, as explained in Fig.12.

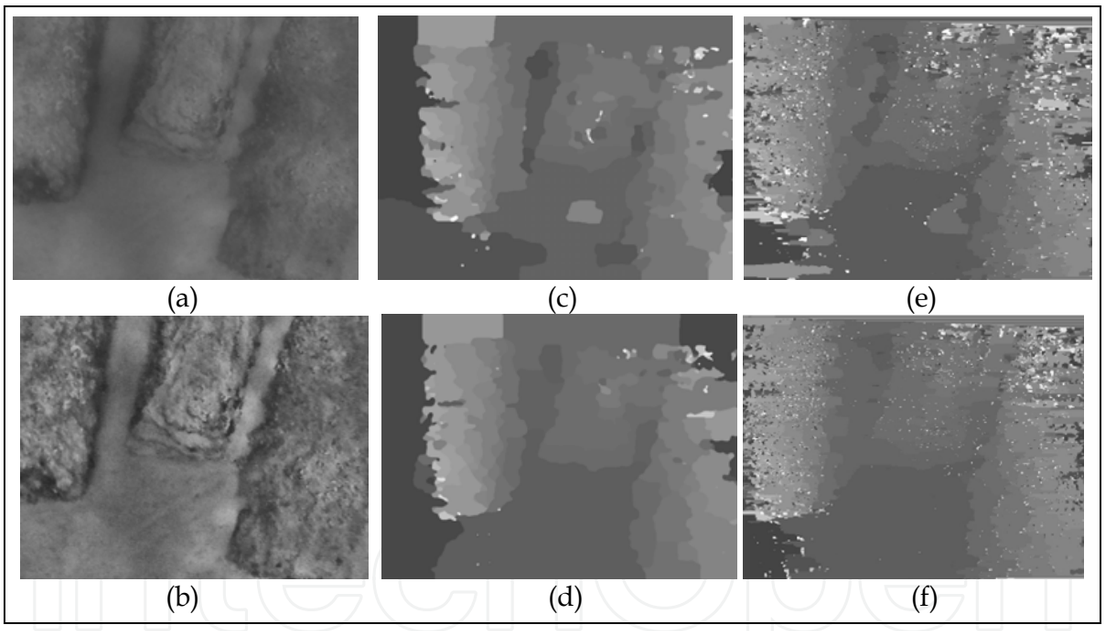


Fig. 12. *a*) original rectified left image; *b*) CLAHE enhanced rectified left image; *c*) disparity map with voxel obtained from (a), *d*) voxel based disparity map obtained from enhanced image (b); *e*) pixel labeling without enhancement; *f*) pixel labeling with CLAHE.

5.4 3-D Reconstruction from stereo correspondences

As prefaced in Section 6.1, given a set of stereo correspondences it is possible to recover the location of the 3-D points by using geometric triangulation. Tsai's calibration method furnishes the camera matrices $\mathbf{P}^L$ and $\mathbf{P}^R$ that relate the image correspondences (x_i^L, x_i^R) with the 3-D point X_i through the relations $\mathbf{P}^L X_i = x_i^L$, $\mathbf{P}^R X_i = x_i^R$, and the epipolar constraint in Eq.1. It follows that x_i^R lies on the epipolar line $l_i^R = \mathbf{F} x_i^L$ and the two rays back-

projected from image points x_i^L and x_i^R lie in a common epipolar plane. Since they lie in the same plane, they will intersect at some point. This point is the reconstructed 3-D scene point X_i. While it is possible to recover the 3-D scene point given only the two imaged points (x_i^L, x_i^R), the accuracy is highly dependent upon the exact matching $x_i^L \leftrightarrow x_i^R$. Generally errors are involved, hence a set of points is usually used leading to an over-determined linear system. In particular, camera parameters and image locations are known only approximately. The back-projected rays therefore do not actually intersect in space. It can be shown that intersection equations can be solved in a least squares sense (Kanatani, 1993). Triangulation is addressed in more details in (Hartley & Zisserman, 2003). Fig. 13 shows the 3-D points calculated by triangulation starting from disparity maps obtained by using methods explained in above sections 6.2 and 6.3. The following sections treat how can be obtained a whole mosaic reconstruction starting from single reconstructions like as shown in Fig. 13.

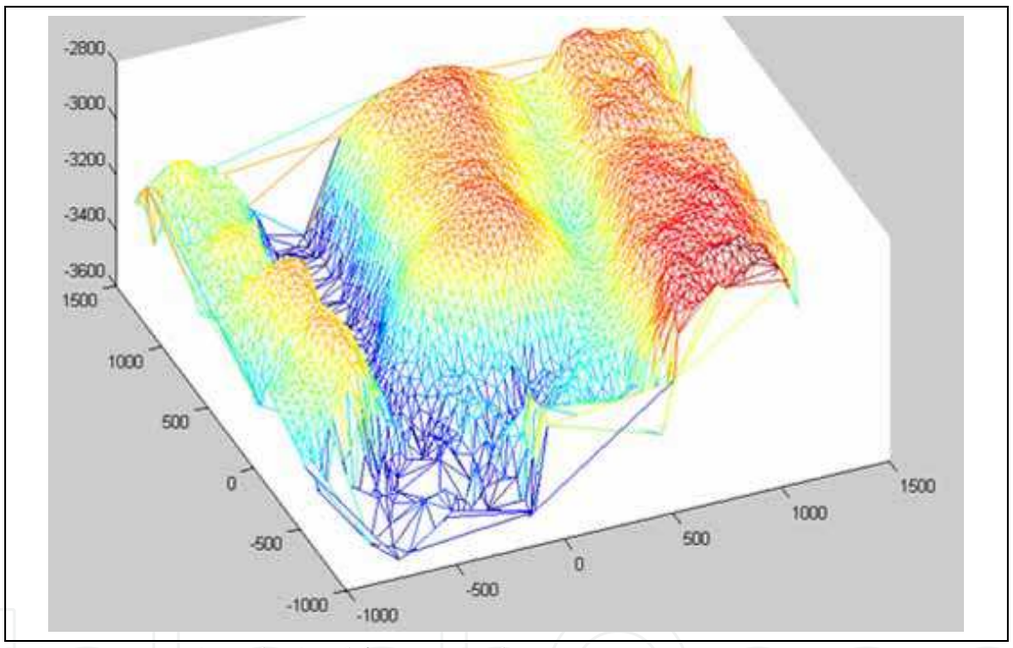


Fig. 13. A reconstruction obtained from a single stereo scan.

6. Seabed mosaic

6.1 Overview

To construct a 3-D mosaic multiple scans must be registered. The registration activity is referred to the alignment of each two-frames reconstruction with the others. A pre-alignment phase can simplify the registration activity, since pitch and roll are not present in motion model (Castellani et al., 2004). For this purpose a pre-alignment activity based on ego-motion estimation through four-frames feature matching (i.e. two following stereo pairs) is used, by evaluating the displacement of left and right matches in two corresponding images. Dimensional effects are normalized by scaling factor according to (Kim & Chung, 2006). Once right correspondences are defined in the four-frame, ego-motion tracking is performed considering displacements through consecutive 3-D point sets.

6.2 Alignment

The correspondences between adjacent frames are estimated with a reliable feature-based matching as explained in the next subsection, while the correspondences between stereo pairs are constrained by the epipolar geometry. Given four frames composed by two following stereo pairs (I_1^L, I_1^R) and (I_2^L, I_2^R), four-frame correspondences are determined according with the scheme shown in Fig. 14.a. Firstly, the correspondences q_1 and q_2

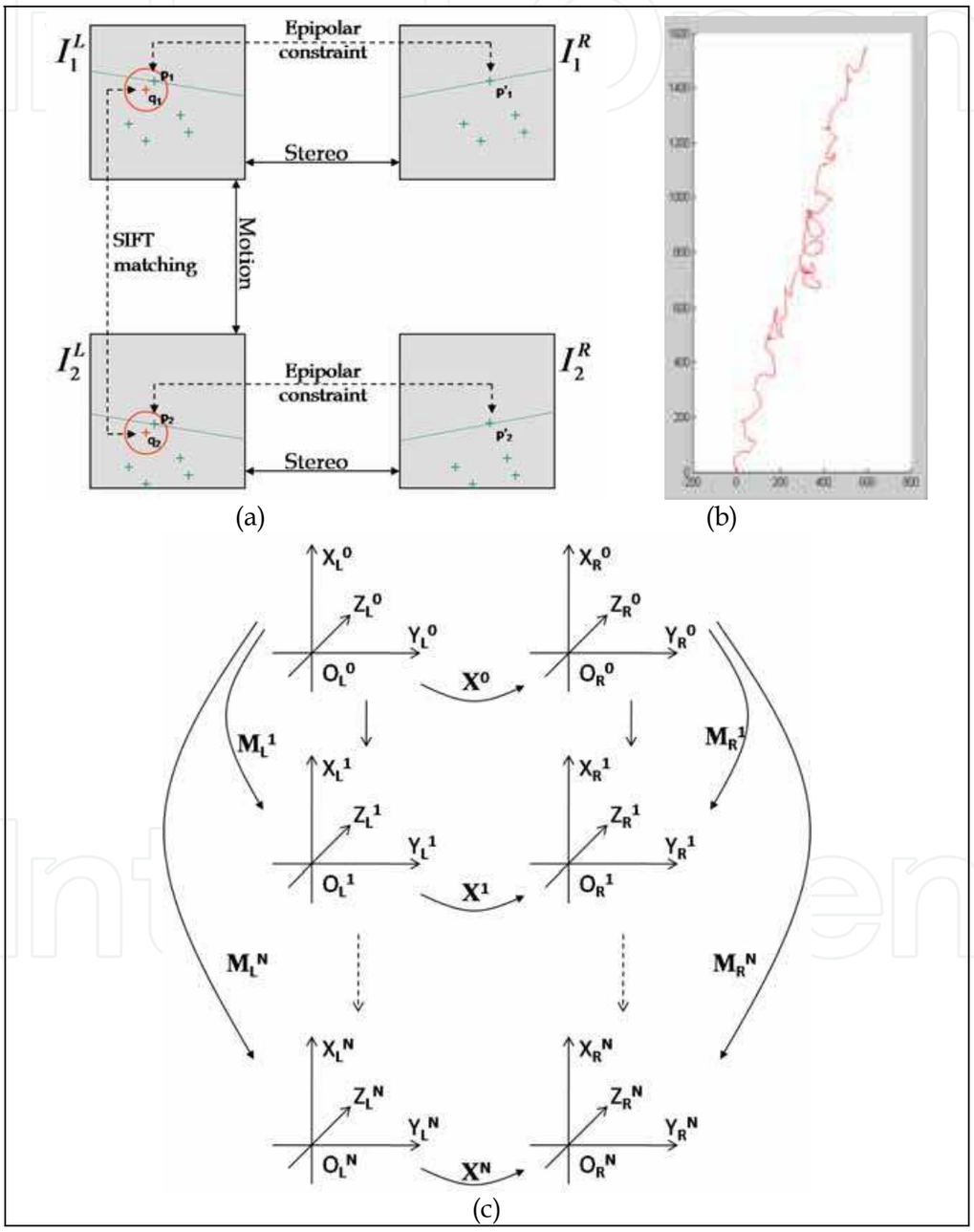


Fig. 14. *a*) Four-frame correspondences estimation scheme; b) Estimated ego-motion trajectory; *c*) The schematic diagram of a moving stereo camera.

between consecutive frames I_1^L and I_2^L are estimated by using the SIFT feature-based matching (see the next subsection). Afterwards, the nearest neighbors p_1 and p_2 to q_1 and q_2 respectively, are selected inside the epipolar constraint. As final result, four matching points p_1, p'_1 and p_2, p'_2 are defined. The scale factor can be determined by using only motion correspondence and the epipolar constrain as explained in the following discussion. Let $\mathbf{R}_x^i$, $\mathbf{R}_L^j$, $\mathbf{R}_R^j$ be rotation matrices and let $\mathbf{t}_x^i$, $\mathbf{t}_L^j$, $\mathbf{t}_R^j$ be translational not zero vectors (for i = 0, …, N and j = 1, …, N). Let be defined the following matrices:

$$\mathbf{X}^i = \begin{pmatrix} \mathbf{R}_x^i & \mathbf{t}_x^i \\ 0 & 1 \end{pmatrix}, \quad i = 0, \ldots, \mathrm{N} \tag{16}$$

$$\mathbf{M}_L^j = \begin{pmatrix} \mathbf{R}_L^j & \mathbf{t}_L^j \\ 0 & 1 \end{pmatrix}, \ \mathbf{M}_R^j = \begin{pmatrix} \mathbf{R}_R^j & \mathbf{t}_R^j \\ 0 & 1 \end{pmatrix}, \quad j = 1, \ldots, \mathrm{N} \tag{17}$$

Let $\bar{\mathbf{t}}_L^j = \mathbf{t}_L^j / \left\| \mathbf{t}_L^j \right\|$ and $\bar{\mathbf{t}}_R^j = \mathbf{t}_R^j / \left\| \mathbf{t}_R^j \right\|$ for j = 1, …, N. Under these assumptions $\mathbf{R}_L^j$, $\bar{\mathbf{t}}_L^j$ and $\mathbf{R}_R^j$, $\bar{\mathbf{t}}_R^j$ can be considered as the motion parameters obtained up to scale using motion correspondence for the left and right cameras respectively. Let the left/right coordinate systems of a moving stereo camera (Fig.14.c). The following relations are satisfied:

$$\mathbf{M}_L^i \mathbf{X}^i = \mathbf{X}^0 \mathbf{M}_R^i \ , \quad i = 1, \ldots, \mathrm{N} \tag{18}$$

Let $s_L = \left\| \mathbf{t}_L^1 \right\|$ and $s_R = \left\| \mathbf{t}_R^1 \right\|$, the Eq.18 can be rewritten as:

$$\begin{cases} \mathbf{R}_L^i \mathbf{R}_x^i = \mathbf{R}_x^0 \mathbf{R}_R^i \\ \mathbf{R}_L^i \mathbf{t}_x^i + s_L \bar{\mathbf{t}}_L^i = s_R \mathbf{R}_x^0 \bar{\mathbf{t}}_R^i + \mathbf{t}_x^0 \end{cases} \quad i = 1, \ldots, \mathrm{N} \tag{19}$$

The Eq.19 can be rewritten in matricial form as in the following Eq.20:

$$\begin{pmatrix} \bar{\mathbf{t}}_L^1 & -\mathbf{R}_x^0 \bar{\mathbf{t}}_R^1 \\ \vdots & \vdots \\ \bar{\mathbf{t}}_L^{\mathrm{N}} & -\mathbf{R}_x^0 \bar{\mathbf{t}}_R^{\mathrm{N}} \end{pmatrix} \begin{pmatrix} s_L \\ s_R \end{pmatrix} = \begin{pmatrix} \mathbf{t}_x^0 - \mathbf{R}_L^1 \mathbf{t}_x^1 \\ \vdots \\ \mathbf{t}_x^0 - \mathbf{R}_L^{\mathrm{N}} \mathbf{t}_x^{\mathrm{N}} \end{pmatrix} \tag{20}$$

The system in Eq.20 gives a unique solution except in a degenerate motion set as proven in (Kim & Chung, 2006). Given the 4-frame matching points and the scale factors estimated as mentioned above, two corresponding 3-D point sets, X_1 and X_2, are calculated between the two following stereo correspondences. Hence, ego-motion is estimated by evaluating 3-D point displacements from X_1 to X_2. Fig.14.b shows the estimated navigation trajectory.

6.3 Correspondences between adjacent frames

Scale Invariant Feature Transform (SIFT) feature (Lowe, 1999) is used as feature matching method, reducing the effect of outlier by using the Least Median of Square (LMedS) method (Zhang et al., 1995). Generally, feature detection methods, such as the Harris detector

(Harris & Stephens, 1988), are sensitive to the affine distortion of image. Therefore, they are not suitable to build feature sets in image acquired in uncontrolled environments. SIFT feature matching (Lowe, 1999) is widely used because it is invariant to affine transforms. These characteristics are suitable to be employed with imagery obtained at different camera angles by using ROV/AUV. SIFT feature algorithm is based upon finding locations within the scale space of an image. Features are identified by detecting maxima and minima in the *Difference of Gaussian* (DOG) pyramidal space. A subpixel location, scale and orientation are associated with each SIFT feature. In order to achieve high specificity, a local feature is formed by measuring the local image gradients at many orientations in coordinates relative to the location, scale and orientation of the feature. Although the SIFT feature matching algorithm has low bad matching error rate, if the outliers are presented in estimation of transformation matrix, the recovered camera motion is incorrect and it is impossible to register the model correctly. Therefore, the LMedS approach is employed and the amount of sample is estimated by using the following relation:

$$P = 1 - (1 - (1 - \varepsilon)^{p})^{m}, \tag{21}$$

where P is the probability of a good sample for LMedS, and ε, p, m denote the ratio of false matched, the sample size and the number of sample required. Choosing m = 108 with $\varepsilon = 0.5$ and p=4 in Eq.21, the probability of a good sample is 99.9%.

6.4 Registration refinement

After a pre-alignment phase based on ego-motion estimation, Iterated Closest Point (ICP) algorithm (Besl & McKay, 1992) is employed for 3-D registration refinement. The registration activity maps each single scan reconstruction into the same coordinate system, solving the absolute orientation problem when correspondences between each single scan reconstruction and the others are unknown. For each iteration ICP algorithm alternates the following two step: 1) calculation of the closest points between reconstructions and assuming this points as correspondent, 2) overlap of the reconstructions by determining the right transformation using absolute orientation. Given two reconstructed points set X_1 and X_2 subject to registration, the above absolute orientation referred in step (2) is resolved through the following minimization using the Kanatani's method (Kanatani, 1993):

$$\min_{\mathrm{R},\mathrm{t}} \sum_{i=1}^{N} \left\| \mathrm{x}_{1,\mathrm{i}} - \left(\mathrm{R}\, x_{2,i} + \mathrm{t} \right) \right\|^2, \tag{22}$$

where (R, t) is the sought transformation (rotation and translation), and (x_{1i}, x_{2i}) is the closest point pair determined in step (1). ICP iterations proceed until the overlapping error doesn't go down a given threshold. The final 3-D mosaic is obtained processing thousands of single scan reconstruction. The final mosaic reported in Fig.15 concerns the reconstruction of a wide underwater area that contains seven columns (only five are visible) in cipolin marble dating to the Roman age.

8. Conclusion and future work

A framework for seabed 3-D mosaic reconstruction has been presented. The three mainly troublesome aspects discussed are asynchronous stereo acquisition, depth estimation and the 3-D mosaic registration. The use of an inexpensive asynchronous stereo sequence is explained,

suggesting a new scheme for accurate calibration frames selection. Moreover, disparity map calculation in both points of view, local and global, is considered. Brightness constancy violation is treated adopting cross-correlation and histogram equalization. Results are evaluated by using ground truth data. Moreover, a four-frame features tracking scheme for ego-motion estimation has been suggested, combining epiflow advantages and ICP registration refinements. Current ongoing and future work involve improvements on ICP module, more accurate ground truth results evaluation and the implementation of the Extended Kalman Filter-based ego-motion and structure recovery for off-line 3-D mosaicking.

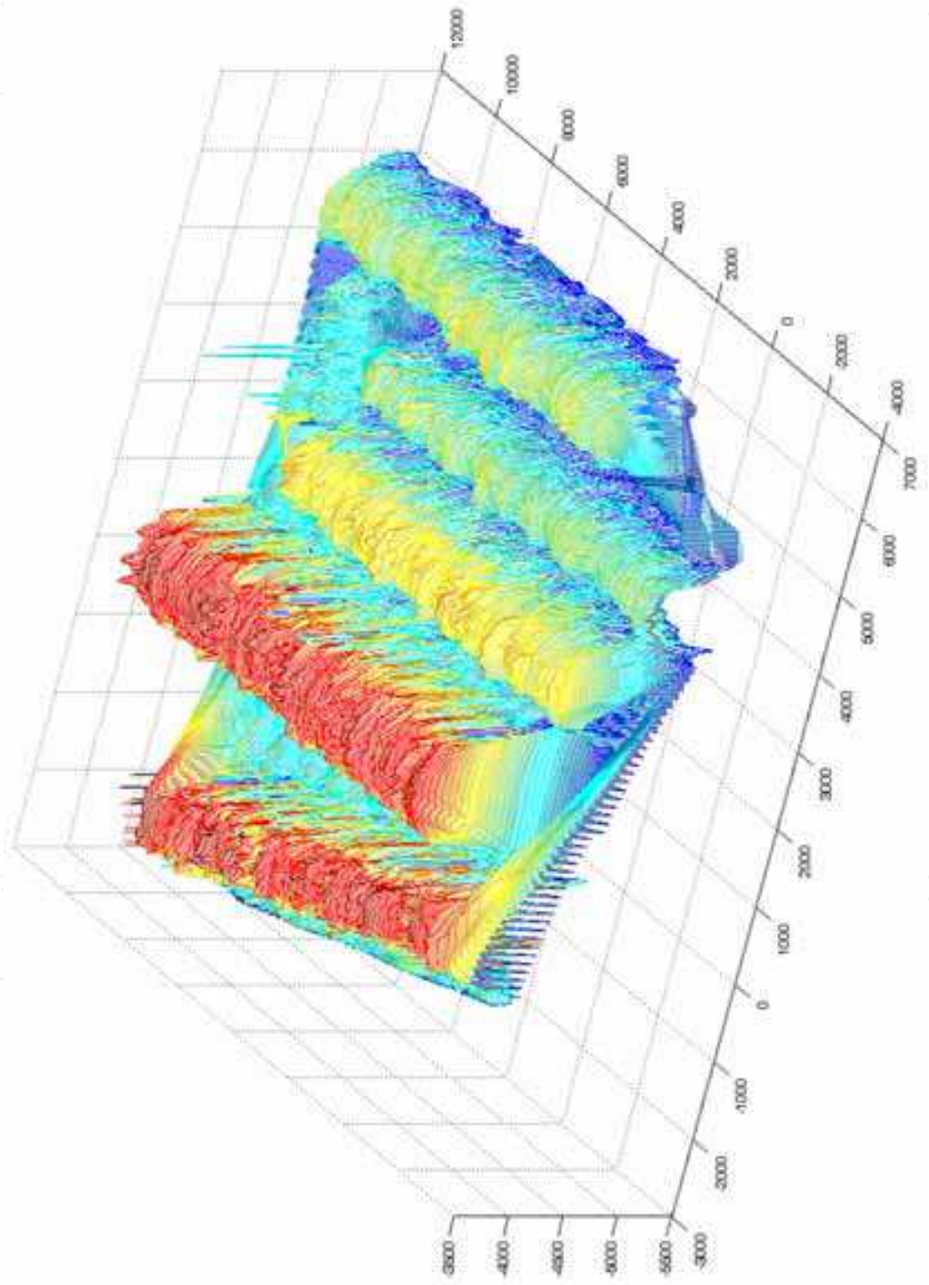


Fig. 15. The final 3-D mosaic reconstruction. Units are expressed in millimetres.

9. References

Bakstein, H. (1999). A Complete DLT-Based Camera Calibration with a Virtual 3D Calibration Object, Diploma Thesis, Charles University, Prague

Besl, P. & McKay, N. (1992). A method for registration of 3D shapes, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 14, pp. 239-256, 0162-8828

Bolles, R. ; Baker, H. & Hannah, M. (1993). The JISCT stereo evaluation, In Proc. DARPA Image Understanding Workshop, pp. 263-274, 1993

Boykov, Y.; Veksler, O. & Zabih, R. (1998). Markov Random Fields with Efficient Approximations. IEEE CVPR, p. 648, 1998

Boykov, Y.; Veksler, O. & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, pp. 1222-1239

Castellani, U.; A. Fusiello, Murino, V., Papaleo, L., Puppo, E., Repetto, S. & Pittore, M. (2004). Efficient On-line Mosaicing from 3D Acoustical Images. Oceans, MTS/IEEE TECHNO-OCEAN, Vol. 2, pp. 670–677, 9-12 Nov 2004

Faugeras, O. (1992). What can be seen in three dimensions with an uncalibrated stereo rig? In Proc. European Conference on Computer Vision, pp. 563-578, Santa Margherita, Italy, May 1992

Harris, C. & Stephens, M.J. (1988). A combined corner and edge detector, In Alvey Vision Conference, pages 147–152, 1988

Hartley, R. & Zisserman, A. (2003). *Multiple View Geometry In Computer Vision*, Cambridge University Press, Second Edition, 2003

Heikkia, J. & Solven, O. (1996). Calibration Procedure for Short Focal Length Off-the-shelf CCD Cameras, In Proc. Of the 13th International Conference on Pattern Recognition, pp. 166 – 170, Vienna, Austria, 1996

Hogue, A.; German, A. & M. Jenkin. (2007). Underwater environment reconstruction using stereo and inertial data, Systems Man and Cybernetics ISIC IEEE International Conference on, pp. 2372-2377, 7-10 Oct. 2007

Horn, B. (1986). *Robot Vision*, MIT Press, 1986

Kanatani, K. (1993). *Geometric Computation for Machine Vision*. Oxford University Press, 1993

Khamene, A.; Madjidi, H. & Negahdaripour, S. (2001). 3-D Mapping of Sea Floor Scenes by Stereo Imaging, Oceans, MTS/IEEE Conference and Exhibition, Vol. 4, pp. 2576-2583, 5-8 Nov. 2001

Khamene, A. & Negahdaripour, S. (2003). Motion and structure from multiple cues; image motion, shading flow, and stereo disparity, *Computer Vision and Image Understanding*, Vol. 90, pp. 99-127, May 2003

Kim, J.H. & Chung M.J. (2006). Absolute motion and structure from stereo image sequences without stereo correspondence and analysis of degenerate cases, *Pattern Recognition*, Vol. 39, No. 9, pp. 1649-1661, 0031-3203

Kutulakos, K. & Seitz, S. (1998). What do N photographs tell us about 3D shape? Technical Report TR680, Computer Science Department, University of Rochester, January 1998

Kolmogorov, V. & Zabih, R. (2002). Multi-camera scene Reconstruction via Graph Cuts, European Conference on Computer Vision, 2002

Kolmogorov, V. & Zabih, R. (2004). What energy functions can be minimized via graph cuts?, *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, pp. 147-159

Laurentini, A. (1994). The visual hull concept for silhouette-based image understanding, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 2, pp. 150-162

Lorenson, W. (1987). Marching Cubes: A High Resolution 3D Surface Construction Algorithm, *Computer Graphics*, Vol. 21, No. 4, pp. 163-169

Lowe, D.G. (1999). Object recognition from local scale-invariant features, IEEE Proc. ICCV, pp. 1150-1157, 1999

Lucas, B. & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision, In Proc. International Joint Conference on Artificial Intelligence, pp. 674-679, Vancouver, Canada, August 1981

Madjidi, H. & Negahdaripour, S. (2005). Global alignment of sensor positions with noisy motion estimates, *IEEE Trans. Robotics and Automation*, Vol. 21, December 2005

Narasimhan, S.G.; Nayar, S.K., Sun, B. & Koppal, S.J. (2005). Structured Light in Scattering Media, Proceedings of the Tenth IEEE International Conference on Computer Vision, Vol. 1, pp. 420-427, 17-21 Oct. 2005

Negahdaripour, S.; Zhang, H. & Han, X. (2002). Investigation of Photometric Stereo Method for 3-D Shape Recovery from Underwater Imagery, Oceans MTS/IEEE, Vol. 2, pp. 1010-1017, 39-31 Oct. 2002

Nicosevici, T.; Negahdaripour, S. & Garcia, R. (2005). Monocular-based 3D seafloor reconstruction and ortho-mosaicing by piecewise planar representation, In Proc. OCEANS MTS/IEEE Conference, Washington, DC, September 2005

Pizer, S.M.; Ambum, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Romeny, BM.H., Zimmerman, J.B., & Zuiderveld, K. (1987). Adaptive Histogram Equalization and Its Variations, Comp. Vis. Graphics & Im. Proc., pp. 1355-368

Potts, R. (1952). Some generalized order-disorder transformation, In Proceedings of the Cambridge Philosophical Society, Vol. 48, pp. 106-109, 1952

Scharstein, D. & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Stereo and Multi-Baseline Vision, IEEE Workshop, pp. 131-140, 9-10 Dec 2002

Slabaugh, G.; Culbertson, W., Malzbender, T. & Schafer, R. (2001). A survey of volumetric scene reconstruction methods from photographs, In Proc. International Workshop on Volume Graphics, pp. 81-100, Stony Brook, NY, June 2001

Sun, J.; Shum, H. & Zheng, N. (2002). Stereo matching using belief propagation, In Proc. European Conference on Computer Vision, pp. 510-524, Copenhagen, Denmark, May/June 2002

Svedman, M.; Goncalves, L., Karlsson, N., Munich, M. & Pirjanian, P. (2005). Structure from Stereo Vision using Unsynchronized Cameras for Simultaneous Localization and Mapping, International Conference on Intelligent Robots and Systems, Intelligent Robots and Systems, IEEE/RSJ, pp. 3069-3074, 2-6 Aug. 2005

Tomasi, C. & Kanade, T. (1991). *Shape and Motion from Image Streams: a Factorization Method - Part 3 Detection and Tracking of Point Features*, Computer Science Department, Carnegie Mellon University, April, 1991

Tsai, R.Y. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Journal of Robotics and Automation RA*, Vol. 3, No. 4, pp. 323 – 344

Zhang, Z.; Deriche, R., Faugeras, O. & Luong, Q. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence*, Vol. 78, pp. 87-119, 1995

Zhang, H. (2005). *Automatic sensor platform positioning and 3-d target modelling from underwater stereo sequences*, PhD Thesis, Coral Gables, Florida, Dec 2005

Stereo Vision

Edited by Asim Bhatti

ISBN 978-953-7619-22-0

Hard cover, 372 pages

Publisher InTech

Published online 01, November, 2008

Published in print edition November, 2008

The book comprehensively covers almost all aspects of stereo vision. In addition reader can find topics from defining knowledge gaps to the state of the art algorithms as well as current application trends of stereo vision to the development of intelligent hardware modules and smart cameras. It would not be an exaggeration if this book is considered to be one of the most comprehensive books published in reference to the current research in the field of stereo vision. Research topics covered in this book makes it equally essential and important for students and early career researchers as well as senior academics linked with computer vision.

How to reference

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

A. Leone, G. Diraco and C. Distante (2008). A Stereo Vision Framework for 3-D Underwater Mosaicking, Stereo Vision, Asim Bhatti (Ed.), ISBN: 978-953-7619-22-0, InTech, Available from: http://www.intechopen.com/books/stereo_vision/a_stereo_vision_framework_for_3-d_underwater_mosaicking

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

IntechOpen