Ocean Engineering 38 (2011) 1207–1219

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

A new global alignment approach for underwater optical mapping ☆

Armagan Elibol *, Rafael Garcia, Nuno Gracias

Computer Vision and Robotics Group, Underwater Vision Lab, University of Girona, Spain

ARTICLE INFO

Article history:
Received 22 January 2010
Accepted 15 May 2011
Editor-in-Chief: A.I. Incecik
Available online 14 June 2011

Keywords:
Mapping
Mosaicing
Underwater robotics
Environmental monitoring

ABSTRACT

Lately, underwater vehicles have become important tools for exploration, monitoring and creation of maps of the seabed. Within mapping applications, the maps obtained from optical data are becoming essential in different study areas such as biological, geological and archaeological surveys, or in detection of benthic temporal changes. However, the underwater medium is very challenging for optical sensors and does not allow the area of interest to be imaged in a single image. Therefore, image mosaicing methods are necessary. Although recent advances in detection of correspondences between images have resulted in highly effective image registration methods, global alignment methods are still needed to obtain a globally coherent mosaic. In this paper, we propose a new global alignment method which works on the mosaic frame and does not require non-linear optimisation. Additionally, a simple image rectifying method is presented to reduce the down-scaling effect which might occur when minimising errors defined in the mosaic frame. Moreover, this rectifying method can also be seen as an alternative and straightforward way of incorporating different sensor information if available. The proposed framework has been tested with underwater image sequences. The resulting method is faster than its counterparts while providing the same level of registration quality.

1. Introduction

Deep water surveys and seabed imagery have become available to the science community thanks to rapid developments in underwater robotics (Escartin et al., 2008). Consequently, optical underwater images are becoming an important tool to study the structure and characteristics of the seafloor, enabling studies at the level of biological communities.

The characteristics of the underwater environment often present difficult challenges for the processing of optical images: significant attenuation and scattering of visible light, lack of image contrast and presence of non-rigid objects. As an example, considering the case of benthic mapping below the photic zone, the need of artificial illumination causes shadows that move in opposite direction with respect to the motion of the camera. Furthermore, light attenuation does not allow images to be taken from a large distance (Pegau et al., 1997; Loisel and Stramski, 2000) and make image registration difficult, potentially leading to inaccurate estimates of registration parameters and image misalignments. Given the need for close range image acquisition, image mosaicing techniques are required to compose high-resolution maps of the surveyed area to obtain a global perspective of the underwater terrain (Gracias and Santos-Victor, 2000; Pizarro and Singh, 2003; Vincent et al., 2003; Leone et al., 2006; Richmond and Rock, 2006; Rzhanov et al., 2006).

This paper considers mission scenarios such as low cost diver cameras for coastal surveying, low cost underwater robots and swarms of small AUVs for distributed mapping relying on cameras for positioning and navigation (e.g., see Fig. 1). We focus on the case of when no any additional information is available apart from the images. The underlying motivation is that it is possible to develop efficient mosaicing methods even in this case. Although most underwater survey platforms have multiple forms of navigation sensing, there are scenarios where vision might be the only sensor providing XY positioning with respect to the bottom (for example when there are no Ultra Short Base Line (USBL), Doppler Velocity Log (DVL) or Inertial Navigation System (INS)). The least expensive sensors such as altitude and heading are omnipresent, but provide very limited information to constrain the number of potential matches, which mainly requires a XY position sensor. In this context, the quality constraints for building image mosaics can be very strict, especially when the mosaic is later used for the localisation of interest areas and the detection of temporal changes (Delaunoy et al., 2008; Escartin et al., 2008). Hence, highly accurate image registration methods are necessary.

Most image mosaicing approaches assume that time-consecutive images can be properly registered (Pizarro and Singh, 2003;

☆ Preliminary work behind the method in this paper was presented at the 4th International Symposium on Visual Computing (Elibol et al., 2008).

* Corresponding author at: Department of Mathematical Engineering, Yildiz Technical University, Istanbul, Turkey.
E-mail addresses: aelibol@yildiz.edu.tr (A. Elibol), rafa@eia.udg.edu (R. Garcia), ngracias@eia.udg.edu (N. Gracias).

0029-8018/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.oceaneng.2011.05.013

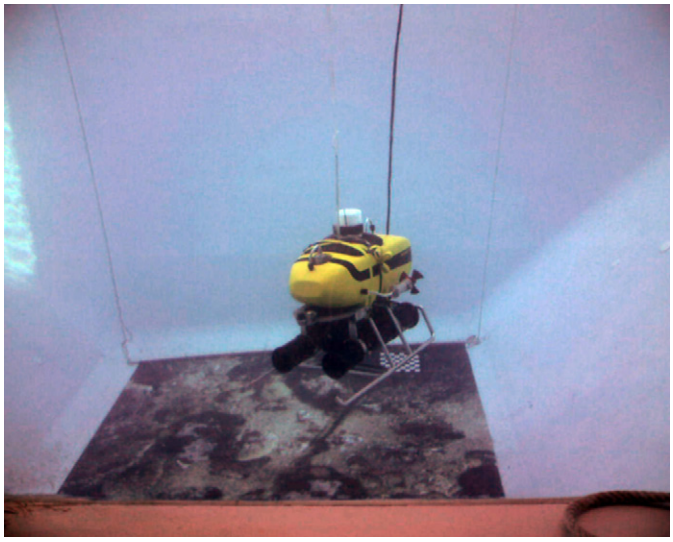

Fig. 1. Technical tests of Unmanned Underwater Vehicle developed by our group, operating in the test pool of the Underwater Vision Lab (University of Girona). The floor of the pool is covered by a poster of real sea floor as to simulate realistic sea floor images for assessing algorithms.

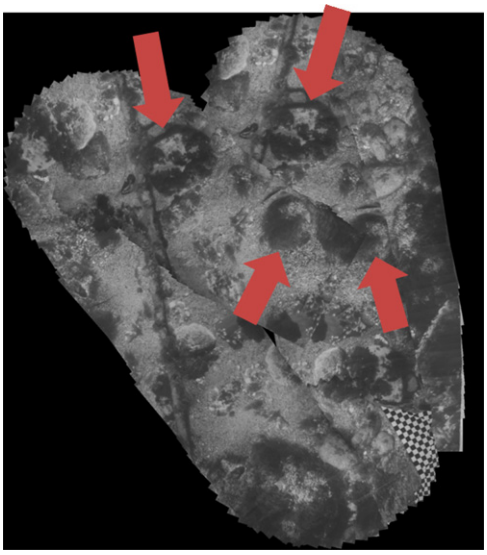

Fig. 2. Example of error accumulation from registration of sequential images. The same benthic structures appear in the different locations in the lower part of the mosaic due to the error accumulation (trajectory drift).

Negahdaripour and Xu, 2002; Rzhanov et al., 2000). Under this assumption, the relative displacement among pairs of images can be obtained by cascading the motion parameters that relate time-consecutive images (Garcia et al., 2001; Elibol et al., 2008). However, in the absence of other sensor information capable of providing absolute orientation and/or position information, the cascading of time-consecutive motion estimates will result in error growth (e.g., see Fig. 2), typical of dead reckoning positioning (Garcia et al., 2001). When the trajectory of the camera provides an overlap between non-consecutive images (closed-loop trajectory), then global alignment techniques can be applied to reduce error accumulation. In the context of this paper, we refer to the global alignment as the problem of finding the image registration parameters that best comply with constraints imposed by image matches. These matches can be time-consecutive or not.

Several methods have been proposed in the literature to solve the *global alignment* problem (Szeliski and Shum, 1997; Davis, 1998; Capel, 2004; Kang et al., 2000; Marzotto et al., 2004; Sawhney et al., 1998; Pizarro and Singh, 2003; Gracias et al., 2004; Can et al., 2002; Ferrer et al., 2007; Cervantes and Kang, 2006). Global alignment usually requires the minimisation of an error term, which is defined from image correspondences. Global alignment methods can be classified according to where the error term is defined. Commonly the error term is defined either in the image frame (Szeliski and Shum, 1997; Capel, 2004; Marzotto et al., 2004; Ferrer et al., 2007) or in the mosaic frame (Davis, 1998; Kang et al., 2000; Sawhney et al., 1998; Pizarro and Singh, 2003; Gracias et al., 2004; Can et al., 2002). Szeliski and Shum (1997) proposed to minimise distances between correspondences, defining an error function on the image coordinate system. Minimising this error term using non-linear least squares has the disadvantage that the gradients have to be computed with respect to the motion parameters. Davis (1998) proposed a method based on solving a linear equation system obtained from motion parameters between images. However, that formulation is limited to the restrictive case of pure camera rotation, where translation is not allowed. Sawhney et al. (1998) first and Pizarro and Singh (2003) later defined an error function on the mosaic frame and minimised the distances between correspondences when they are mapped to the mosaic frame. Unfortunately, when the minimisation is carried out on the mosaic frame, the solution tends to reduce the size of the mosaiced images, since reducing its size also decreases the error term. This bias towards the reduction of the image size is referred as the *scaling* problem. In order to avoid this problem, they introduced an additional term that penalises changes on the image size when the images are mapped to a mosaic frame. However, when the distance of the camera to the scene is not constant, this penalty promotes misalignments between images. Gracias et al. (2004) minimised the same error term by linear recursive and batch formulation for the similarity type of homographies by using all correspondences. However, the proposed linear method cannot cope with projective homographies which are commonly required to build a mosaic from an arbitrary moving camera. Similarly, Can et al. (2002) proposed the linear joint estimation of two combined error terms. The first term is to measure errors against feature locations in the image frame which is chosen as a global frame and the second term minimises the distance between correspondences when they are mapped onto the mosaic frame. However, minimising those error terms for projective homographies cannot be done linearly. Therefore, the proposed method cannot be used for projective homographies. Capel (2004) formulated the global alignment problem using, as unknowns, both the absolute homographies (i.e., the transformations between each image and the common mosaic frame) and the location of point features in the mosaic frame. As the error term is defined in image frame, it does not suffer from the scaling problem. However, as the dataset grows, the total number of unknowns dramatically increases. Kang et al. (2000) and Marzotto et al. (2004) showed that using a grid of points on the mosaic frame produces good results. Although this strategy has the advantage of distributing the errors, it has some disadvantages, such as: (1) the point location must be defined very carefully so that every image and overlapping areas have enough grid points to calculate the homography and (2) since the points are distributed arbitrarily, they may fall in a textureless area, making it difficult to match them in another image. More recently, Ferrer et al. (2007) proposed a global alignment method for creating large-scale underwater photo-mosaics that combines image registration information and 3D position estimates provided by navigation sensors available in deep water surveys, but also requiring non-linear optimisation. Cervantes and

Kang (2006) have presented a technique very similar to our proposal in this paper. Their method modifies the accumulated homographies according to the reappearance of feature positions. They formulate the problem as a image-to-mosaic registration with the aim of real-time mosaicing (known as "online mosaicing"). In this case, if an error occurs while mapping one image onto the mosaic, future image-to-mosaic registrations are most likely to fail.

Homographies that encode motion models that allow scale changes are prone to introduce distortions on the image size (e.g., see Fig. 3). Several methods have been proposed in the literature to deal with this distortion problem (Vercauteren et al., 2006; Wang et al., 2005; Mohan et al., 2008). In Vercauteren et al. (2006), a model of the relationship between the motion and the distortion was proposed to reduce the distortions when using a laser scanning device, namely in fibred confocal microscopy. Wang et al. (2005) proposed a geometric correction of images of rectangular planar patterns (e.g., documents, maps). In this case, a planar homography is used to rectify images and reduce the geometric distortions. Their method assumes that the image contains rectangular objects and requires four corners of the object to be extracted manually which precludes the application in natural or unstructured environments. Recently, Mohan et al. (2008) proposed a similar approach that uses a planar homography to rectify the selected part of the image. However, a surface or an object to be rectified needs to be selected manually and the camera intrinsic parameters have to be known. These requirements make the method difficult to apply to images that do not contain man-made structures.

In summary, a common bottleneck of the methods described above is the computational cost required to provide a consistent globally aligned mosaic. In this paper, we present a combined method for global alignment and distortion reduction capable of creating large-area image mosaics with low computational cost. The method makes use of features tracked along the image sequence and the minimisation is performed on the mosaic frame. It avoids non-linear optimisation by alternating two iterative linear steps. An initial estimation is obtained by accumulating time-consecutive homographies. The error in the initial estimation is modelled by propagating the uncertainty using the methods described in Ochoa and Belongie (2006) and Haralick (1998). Since covariances provide some information about error distribution of the initial estimation, this uncertainty information is used to reduce the total number of iterations required by the global alignment method. In addition, regarding the distortion problem, we propose a method to reduce the distortions associated with the construction of mosaics which uses a projective model.

The rest of the paper is organised as follows: Section 2 summarises some image mosaicing and global alignment aspects. Next, Section 3 is devoted to detail the proposed method to solve the global alignment problem. Then, distortion reduction based on four-point warping is described in Section 4. Some results are illustrated in Section 5 and, finally, we present our conclusions in the last section.

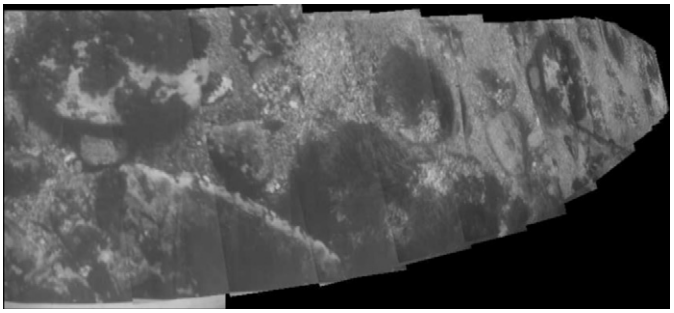

Fig. 3. First 15 images of an underwater dataset. Distortions on the image size are introduced by accumulating projective homographies. The sequence starts on the left and moves to the right. It can be observed that from left to right distortion on the images grows.

2. Feature-based image mosaicing

Feature-based image mosaicing is accomplished through two main steps: *spatial alignment* of the images, also known in the literature as image registration and image *intensity blending* for building the final mosaic. Spatial alignment of the images is normally solved by first finding the motion between pairs of images in the sequence (pairwise alignment). Next, global alignment is applied to deal with the errors occur while representing pairwise motions onto common frame obtain globally coherent mosaics. Finally, image intensity blending methods are needed to deal with intensity differences between images and obtain a seamless mosaic.

2.1. Feature-based image registration

2D image registration is the process of overlaying two or more views of the same scene taken from different viewpoints. Several approaches exist to register images (Zitová and Flusser, 2003). In this paper, we follow a feature-based approach, which is detailed below. The processing pipeline is illustrated in Fig. 4. Feature-based registration methods rely on the detection of salient features using Harris (Harris and Stephens, 1988, Hessian Beaudet, 1978) or Laplacian (Lindeberg, 1998) feature detectors. These features are detected in the two images to be registered, and then a correlation or SSD (*sum of squared distances*) measure is computed around each feature for each assumed geometric transformation of the image. This had been the trend for several years, until recent proposal of the SIFT (*scale invariant feature transform*) algorithm by Lowe (2004). The satisfactory results of this method have greatly speeded up the development of salient point detectors and descriptors, taking feature-based matching techniques to the forefront. Compared to all formerly proposed schemes, SIFT and subsequent developed methods such as SURF (Bay et al., 2006) demonstrate considerably greater invariance to image scaling, and rotation, and robustness under change in illumination and 3D camera viewpoint. These methods solve the correspondence problem through a pipeline that involves (1) feature detection, (2) feature description and (3) descriptor matching. Feature detection is based on either Hessian or Laplacian detectors (the "Difference of Gaussians" of SIFT is an approximation to the Laplacian, and SURF uses an approximation to the Hessian). Feature description exploits gradient information at a particular orientation and spatial frequencies (see Mikolajczyk and

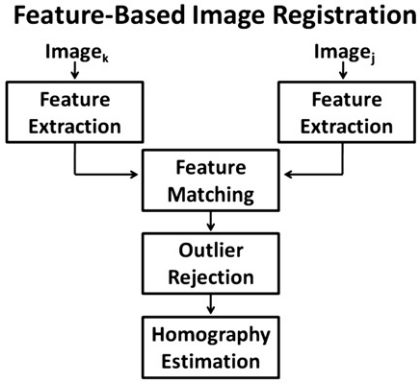

Fig. 4. Pipeline of feature-based image registration between an overlapping image pair.

Schmid, 2005 for a detailed survey on descriptors). Finally, the matching of features is normally based on the Euclidean distance between their descriptors. In this way corresponding points are detected in each pair of overlapping images.

The initial matching frequently produces some incorrect correspondences, which are called *outliers*. Outliers need to be identified and removed, typically with a robust estimation algorithm (e.g., RANSAC Fischler and Bolles, 1981 or LMedS Meer et al., 1992). These algorithms are used to estimate the dominant image motion, which agrees with the largest number of points. Outliers are identified as the points that do not follow that dominant motion. After outlier rejection, a homography can be computed from the inliers through orthogonal regression.

2.2. Global alignment

The aim of global alignment is to overcome cumulative error and build a seamless mosaic obtained from correctly aligned images. Let ${}^{t-1}\mathbf{H}_t$ denote the relative homography between tth and $(t-1)$th image in a sequence. If the first image of the sequence is chosen as the global frame, the global projection of image t into the mosaic frame is denoted as ${}^{1}\mathbf{H}_t$. ${}^{1}\mathbf{H}_t$ is known as the *absolute homography*, and it can be calculated by composing (or cascading) the relative homographies ${}^{1}\mathbf{H}_t = {}^{1}\mathbf{H}_2 \cdot {}^{2}\mathbf{H}_3 \cdot \ldots \cdot {}^{t-1}\mathbf{H}_t$. Unfortunately, the estimated correspondences between image pairs are subject to localisation errors due to noise, and the accuracy of the resulting homography is affected. Therefore, computing absolute homographies by cascading noisy relative homographies results in cumulative error. Thus, global alignment reestimates the absolute homographies by best satisfying geometric constraints arising from the matches between overlapping image pairs. These matches result from sequential and non-sequential (closed loop) image pairs and form an over-constrained set of equations.

3. Iterative global alignment

Our approach is inspired by the method proposed by Capel (2004) that tries to simultaneously minimise both the homographies and the position of features on the mosaic image. In this method, the same feature point correspondences need to be identified over all views, requiring feature tracking. Let ${}^{t}x_i$ denote the coordinates of the ith interest point defined on the coordinate system of image t. ${}^{t}x_i$ is the image projection of point ${}^{m}x_j$ which is called the pre-image point and is also usually projected in different views. In Capel (2004), all the image points that correspond to the projection of the same pre-image point are called N-view matches. The cost function to be minimised is defined as

$$\varepsilon_1 = \sum_{j=1}^{M} \sum_{{}^{t}x_i \in \eta_j} \| {}^{t}x_i - {}^{t}\mathbf{H}_m \cdot {}^{m}x_j \|_2 \tag{1}$$

where M is the total number of pre-image points, η_j is the set of N-view matches and ${}^{t}\mathbf{H}_m$ is a mosaic-to-image homography,[1] and $\| \cdot \|_2$ corresponds to the Euclidean norm. In Eq. (1), both the homographies and the pre-image points are unknowns. The total number of unknowns is (DOFs[2] of homography) × (number of views) + 2 × (number of pre-image points). Eq. (1) can be minimised by applying non-linear least square methods. The residues inside the error term ε_1 are measured in the image frame, but parameterized with points defined on the mosaic frame (see Fig. 5(a)). This formulation avoids the image scaling bias that occurs when the residues are measured on the mosaic frame.

In this paper, we propose an alternative approach that avoids non-linear optimisation. For this, we first transfer ε_1 to the mosaic frame by means of the following equation:

$$\varepsilon_2 = \sum_{j=1}^{M} \sum_{{}^{t}x_i \in \eta_j} \| {}^{m}x_j - {}^{m}\mathbf{H}_t \cdot {}^{t}x_i \|_2 \tag{2}$$

where ${}^{m}\mathbf{H}_t$ is equal to $({}^{t}\mathbf{H}_m)^{-1}$. Direct minimisation of the error term in Eq. (2) biases the estimation towards small image sizes since smaller images lead to smaller differences between ${}^{m}x_j$ and ${}^{m}\mathbf{H}_t \cdot {}^{t}x_i$. If we analyse the error term in Eq. (2), we can observe that the minimisation can be divided into two linear sub-problems (sub-steps): the first step is to minimise the error by considering the homography values to be constant. Therefore, they are not taken into account as unknowns. The problem is then reduced to a special case (*one free point*) of the *quadratic placement problem* (Boyd and Vandenberghe, 2004). This special case has an analytic solution which is the average of the coordinates of all image points after being reprojected onto the mosaic frame under the Euclidean norm (see Fig. 5(b)). The coordinates of the pre-image points (${}^{m}x_j$) in the mosaic frame can be found as the mean of the position of each point multiplied by the corresponding absolute homography. In the first step, as homographies are constant, the ${}^{m}\mathbf{H}_t \cdot {}^{t}x_i$ term in Eq. (2) is known and the equation can be rewritten as follows:

$$\varepsilon_2 = \sum_{j=1}^{M} \sum_{{}^{t}x_i \in \eta_j} \| {}^{m}x_j - {}^{m}x_i^t \|_2 \tag{3}$$

where ${}^{m}x_i^t = {}^{m}\mathbf{H}_t \cdot {}^{t}x_i$. An estimate of ${}^{m}x_j$ is given by minimising (3), which leads to

$${}^{m}\hat{x}_j = \frac{1}{n_j} \sum_{{}^{t}x_i \in \eta_j} ({}^{m}x_i^t) \tag{4}$$

where n_j is the total number of images in which feature point ${}^{m}\hat{x}_j$ appears. The second step is to recalculate new absolute homographies using the new point set $({}^{t}x_i, {}^{m}\hat{x}_j)$ which is computed linearly and independently for each homography.

The error accumulates as the sequence gets longer. This means that tracked feature positions are getting farther from their real positions as they are getting farther away from the chosen global image frame. This knowledge can be introduced into the minimisation process as weights, while calculating the position of features on the mosaic frame during the first step of the first iteration. In order to choose the weights adequately, we propose to propagate the uncertainty of the initial estimation and use it as weights.

As initial estimation, the absolute homography of image i, ${}^{1}\mathbf{H}_i$, is calculated by cascading the relative homographies, given in the following equation:

$${}^{1}\mathbf{H}_i = {}^{1}\mathbf{H}_{i-1} \cdot {}^{i-1}\mathbf{H}_i \tag{5}$$

where $i=2,\ldots,N$. The uncertainties of relative homographies, ${}^{i-1}\boldsymbol{\Sigma}_i$, are calculated from matched points by using the method described in Haralick (1998). Covariance matrices of initial absolute homographies, ${}^{1}\boldsymbol{\Sigma}_i$ for $i=2,\ldots,N$ are propagated by using the first order approximation of Eq. (5), assuming that covariances of time-consecutive homographies are not correlated (Ochoa and Belongie, 2006; Eustice, 2005):

$${}^{1}\boldsymbol{\Sigma}_i = \mathbf{J}_{{}^{1}h_{i-1}} \cdot {}^{1}\boldsymbol{\Sigma}_{i-1} \cdot \mathbf{J}^{T}_{{}^{1}h_{i-1}} + \mathbf{J}_{{}^{i-1}h_i} \cdot {}^{i-1}\boldsymbol{\Sigma}_i \cdot \mathbf{J}^{T}_{{}^{i-1}h_i} \tag{6}$$

where $i=2,\ldots,N$, $\mathbf{J}_{{}^{1}h_{i-1}}$ and $\mathbf{J}_{{}^{i-1}h_i}$ are the Jacobian matrices of Eq. (5) with respect to the parameters of ${}^{1}\mathbf{H}_{i-1}$ and ${}^{i-1}\mathbf{H}_i$. As the first image is chosen as a global frame, its covariance matrix, ${}^{1}\boldsymbol{\Sigma}_1$, is set

[1] m stands for the mosaic frame. This frame can be one of the image frames or a different arbitrary coordinate frame. In this work, we have chosen the first image frame as a mosaic frame therefore, m is equal to 1. In order to keep the generality, we have used m in the notation.

[2] DOF stands for degrees of freedom.

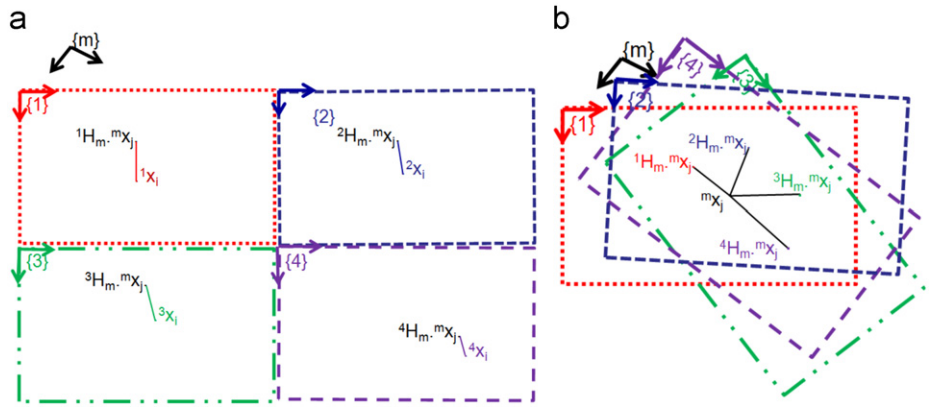

Fig. 5. Capel's method vs our proposal (iterative method). Comparative example. (a) Capel's method: Consider the scene point ${}^{m}x_j$ which has been matched in four different images. Capel's method tries to minimise the sum of distances between the projection of the scene point onto image frames and its identified position on the image frame by simultaneously estimating the position of the scene point and mosaic-to-image homography parameters. (b) Iterative method: The position of the scene point, ${}^{m}x_j$, is unknown but its projections onto the images are known. Once these points are mapped onto the mosaic frame then problem reduces to *quadratic placement problem*. The position is the one where sum of distances to the other points is minimum.

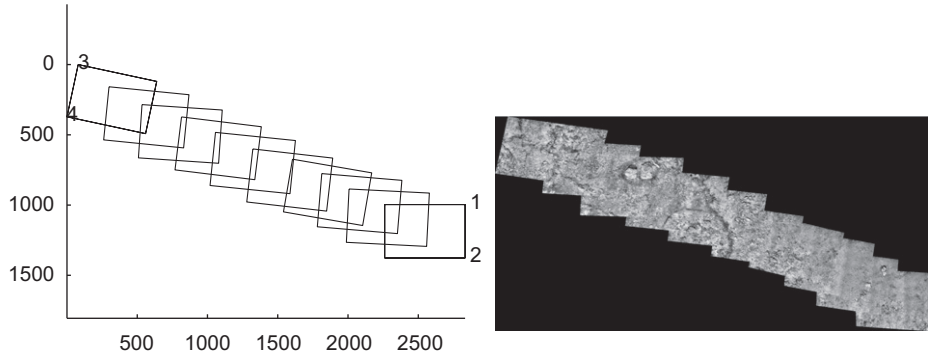

Fig. 6. Selected corner points of the mosaic computed based on Euclidean transformation (steps 1 and 2).

to zero. Then, the uncertainty of the initial estimation is used as a weight in Eq. (4) while calculating the position of features on the mosaic frame during the first step of the first iteration:

$$ {}^{m}\hat{x}_j = \frac{1}{p_j} \sum_{{}^{t}x_i \in \eta_j} w_t \cdot ({}^{m}x_i^t) \tag{7} $$

where $w_t = \sqrt{|{}^{1}\boldsymbol{\Sigma}_t^{-1}|}$ and $p_j = \sum_{{}^{t}x_i \in \eta_j} w_t$. The inclusion of the weighting factors allows for reaching the final result faster (see Fig. 11) as the uncertainty estimation provides some information about error in the initial estimation.

These two linear steps can be executed iteratively until a selected stopping criterion is fulfilled. A typical stopping criterion is to set a threshold on the decrease rate of error term ε_2. It should be noted that this approach has two main advantages over existing methods. First, it avoids non-linear optimisation by iterating two linear steps. This is relevant in the case of large-area mosaics. As non-linear optimisation is not used, the computational cost is much lower, and therefore, minimisation is faster. Both Gauss–Newton and Levenberg–Marquardt methods are frequently used for solving the non-linear least square problems. Those methods use the (augmented) normal equations in their central step (Hartley and Zisserman, 2004). The computation cost of solving the normal equations has complexity n^3 in the number of parameters and it is repeated several times until it converges. Minimising a cost function with respect to a large set of unknowns becomes computationally very expensive. Although there are some improvements for sparsely constructed systems (Hartley and Zisserman, 2004; Lourakis and Argyros, 2008), computational cost can still be prohibitive for large problems.

The main computational cost of the proposed method comes from the second step, which involves computing a set of independent homographies. We have used the Direct Linear Transformation (DLT) algorithm (Hartley and Zisserman, 2004) which uses Singular Value Decomposition (SVD) to compute each homography. For a given $p \times r$ matrix $\mathbf{A}$, the computational cost of the SVD to solve the $\mathbf{A} \cdot b = 0$ system is $4pr^2 + 8r^3$ (Hartley and Zisserman, 2004), and is linear with the number of rows. This computational cost is lower than those of non-linear least square minimisation methods. The second advantage of our proposal with respect to the state of the art is that it requires much less memory to process data, when compared to non-linear methods. One of the major inconveniences of non-linear methods is the memory requirement to store the Jacobian matrix at each iteration. Therefore, the proposed method can be easily applied to large datasets without requiring high-end computation platforms.

4. Reducing image size distortions

As mentioned before, the cascading of sequential motion estimates leads to an error accumulation. This cumulative error affects the size of images. We propose a simple method to reduce those scale distortions. If there is no other information on image positions (e.g., from navigation sensors such as USBL, DVL or INS), our approach initially aligns the images with Euclidean homographies which have 3DOF (1D rotation plus 2D translations) so that there are no changes on the image size. This provides a good approximation in the cases where an underwater

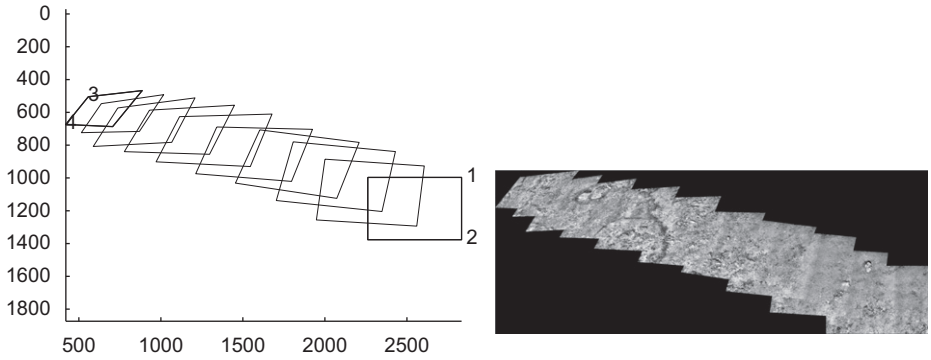

Fig. 7. Corner points of the original mosaic as computed from projective transformation (steps 2 and 3).

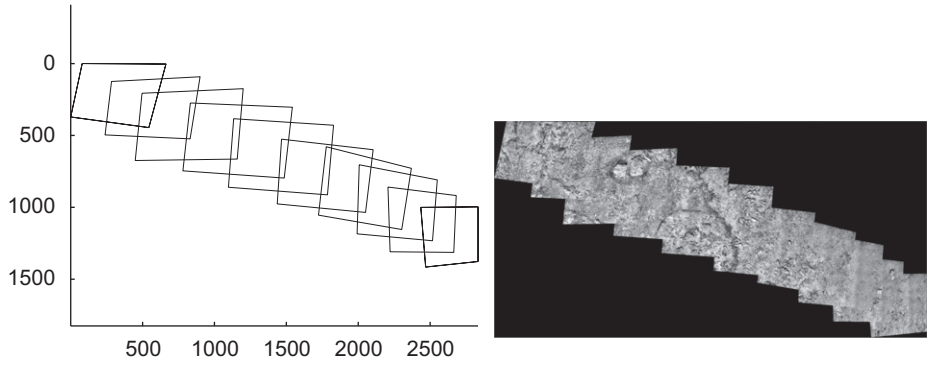

Fig. 8. Final mosaic after applying the four-point homography (step 6).

Table 1
Four-point warping algorithm.

Step 1	Align images using Euclidean homographies
Step 2	Extract the coordinates of corners of the mosaic
Step 3	Align images using projective homographies
Step 4	Repeat step 2 for the mosaic obtained in step 3
Step 5	Compute the projective homography between mosaics obtained in steps 1 and 3
Step 6	Update the absolute homographies and build the final mosaic

Table 2
Characteristics of the datasets.

Data	Total number of images	Image size in pixels	Total number of overlapping pairs	Total number of correspondences	Total number of tracked features	Mapped area (m^2)
Dataset 1	860	384×288	22,116	4,028,557	93,515	400
Dataset 2	263	3008×2000	4746	239,431	18,614	300

robot carries a down-looking camera has small changes in roll and pitch, and keeps approximately constant distance to the seafloor. The coordinates of the four corners of the resulting mosaic aligned through an Euclidean transformation are extracted by using the absolute homographies of the first and last image.[3] Then, images are aligned with projective or affine homographies. The coordinates of the corners of the aligned mosaic are extracted. These corners are used as correspondences of the corners obtained from the Euclidean model. An example can be seen in Figs. 6 and 7. The projective homography between the two mosaic images is calculated. Next, the projective homographies are then multiplied by this four-point homography so that absolute homographies with less distortion are obtained (Fig. 8). The algorithm is summarised in Table 1.

The homography between four corners of two mosaics is comparable to the rectification homography in Liebowitz and

[3] They are not necessarily the first and last image, but can be the images located at the borders of the mosaic, corresponding to the maximum and minimum coordinates along the x and y axes (see Fig. 6).

Zisserman (1998) which is used to reduce the distortions on the image size. In their work this homography is decomposed into three different matrices which account for similarity, affine and pure projective transformations. Each of them is calculated by taking into account the metric properties of the scene such as angles and length ratios, which need to be known, thus reducing the application scenarios. This contrasts with our approach where we have computed this homography without computing the each matrix explicitly as we do not have any information about the properties of the scene. Four correspondences are the minimum number of matched features required to compute the projective homography. This means that the four corners of the mosaic are mapped onto the same position as the ones of the Euclidean mosaic. Therefore, in the final mosaic, the length between mosaic corners and the angles between lines will be same as those of the Euclidean mosaic.

This approach can also be used when information about image positions on the mosaic frame is available from navigation sensors and/or a number of world points with known x and y coordinates may be available and they could be used for rectifying. As the projective homography, rectifying homography, calculated from four correspondences is an exact mapping of points, therefore, after applying this homography, four points in both mosaic have the same coordinates. Furthermore, due to this exact mapping property, multiplying all absolute homographies of

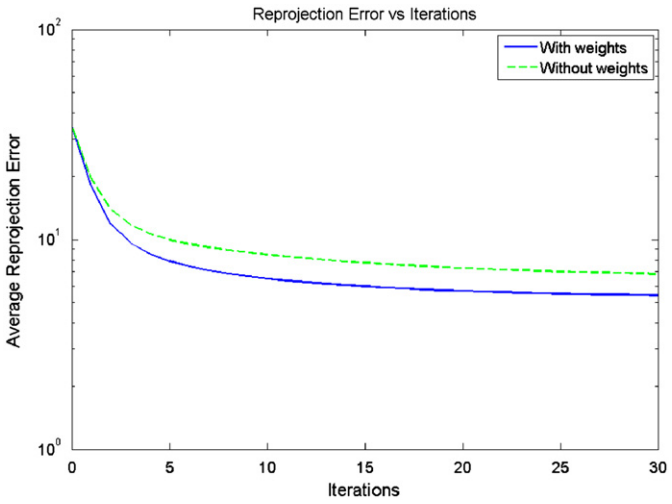

Fig. 11. Change in the average reprojection error with (Eq. (7)) and without (Eq. (4)) using weights for the first dataset. The horizontal axis corresponds to the iterations and the vertical axis shows the average reprojection error in pixels in logarithmic scale.

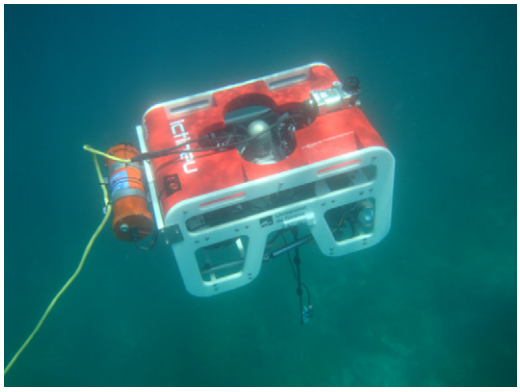

Fig. 9. Snapshot of our Unmanned Underwater Robot *ICTINEU*^{AUV}, operating in the Mediterranean Sea during acquisition of the first dataset. The robot carries a down-looking camera mounted on a bar. The 3D relief of the scene is negligible compared to the altitude of the camera.

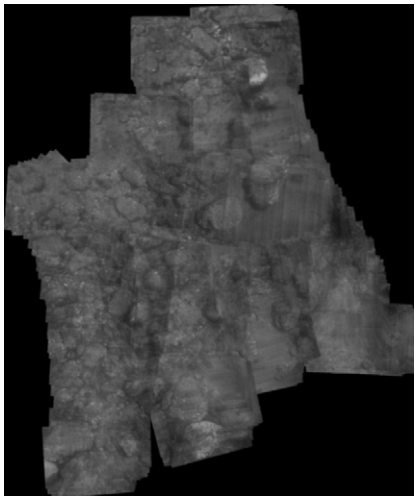

Fig. 12. Initial estimation of the first dataset.

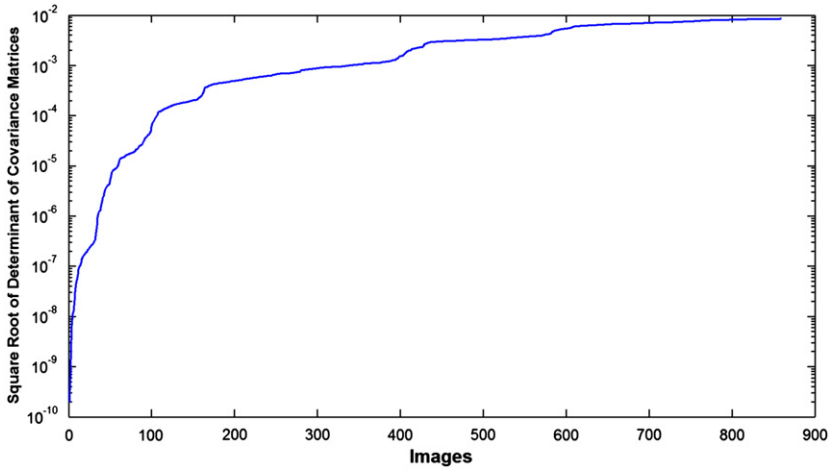

Fig. 10. Uncertainties of the initial estimation. The horizontal axis corresponds to the image index and the vertical axis shows the square root of the determinant of covariance matrices in logarithmic scale.

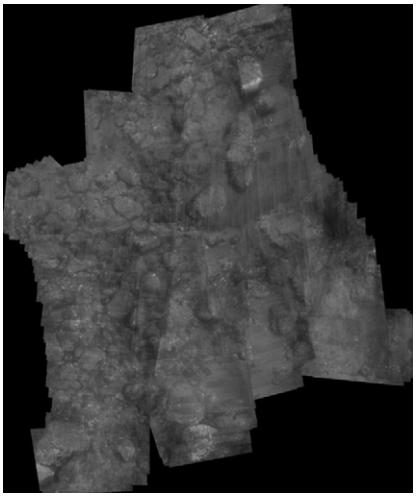

Fig. 13. Resulting mosaic of the proposed method.

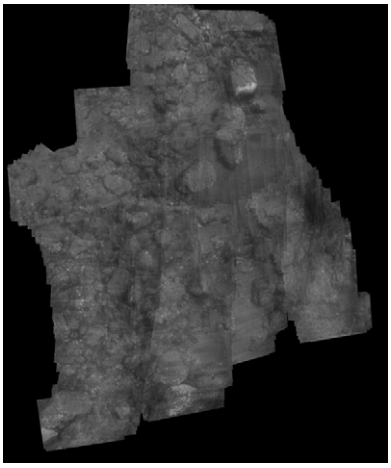

Fig. 14. Resulting mosaic of the Capel's method.

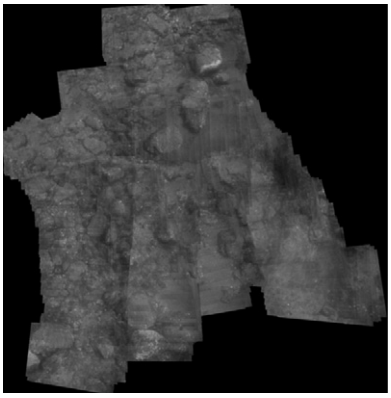

Fig. 15. Mosaic obtained through the bundle adjustment approach described in Ferrer et al. (2007).

images with this rectifying homography does not cause any change on the local alignment between images while globally reducing the distortion on the image sizes.

5. Experimental results

We have tested our method using large-area underwater image sequences. The main characteristics of the datasets are summarised in Table 2. The first dataset covers a large area of the

Fig. 16. Resulting mosaic with Capel's method.

Fig. 17. Resulting mosaic with the proposed method.

Table 3
Results of the tested methods.

Data	Measure	Initial estimation	Capel's method	Proposed method	Bundle adjustment[a]
Dataset 1	Avg. rep. error in pixels	34.79	4.76	5.04	5.15
	Total time in seconds		99,876.70	22,652.10	44,988.30
Dataset 2	Avg. rep. error in pixels	673.56	37.18	35.79	41.04
	Total time in seconds		9080.40	1397.20	8084.90

[a] In this work, n in Eq. (8) is chosen as five (Sawhney et al., 1998; Gracias et al., 2003). This method was not totally implemented in MATLAB^{TM} environment. It makes use of some C code through MEX files. Hence, the time reported here cannot be used to compare with the other two methods accurately but it provides an idea about required time.

Table 4
Distortion measures of the final mosaics for the second dataset.

	Initial estimation	Capel's method, Fig. 16	Iterative method, Fig. 17	Euclidean corners, Fig. 18	Image centres, Fig. 19	Bundle adjustment, Fig. 20
Max/min ratio	1.258	1.490	1.569	1.258	1.312	1.297
Deviation (%) to bundle adjustment	3.053	14.811	20.910	3.053	1.110	0.000

seafloor that was acquired by the *ICTINEU*^{AUV} underwater robot (Ribas et al., 2007) during sea experiments in Colera on the Mediterranean coast of Spain (see Fig. 9). Although the hydrodynamics of open frame vehicles is known to be less efficient than that of closed hull type robots, *ICTINEU*^{AUV} is suitable for applications not requiring movements at high velocities, but requiring travelling close to the seabed. The acquired dataset comprises 860 low-resolution images (384×288 pixels) and covers approximately 400 m^2. Before applying our method, the camera was calibrated and images were compensated for radial distortion. Features are detected and matched between images by using SIFT (Lowe, 2004). Then, RANSAC (Fischler and Bolles, 1981) is used to reject outliers and estimate the motion between images. The total number of overlapping image pairs is 22,116. Features are tracked along the images using the initial estimation of the topology. The total number of tracked features is 93,515 and the number of correspondences among all overlapping image pairs is 4,028,557.

In order to illustrate the advantages of our approach with respect to the closest method in the literature, we implemented Capel's Method. The minimisation of the cost function in Eq. (1) was carried out using the MATLAB^{TM} *lsqnonlin* function for large-scale methods. The optimisation algorithm requires the computation of the Jacobian matrix containing the derivatives of all residuals with respect to all parameters. The Jacobian matrix has a clearly defined block structure, and the sparsity pattern is constant (Capel, 2004). In our implementation, analytic expressions were derived and used for computing the Jacobian matrix. We have compared the performance of our method with Capel's method and the bundle adjustment approach proposed in Ferrer et al. (2007). Our performance criterion corresponds to the average reprojection error over all correspondences, i.e.:

$$\varepsilon_3 = \frac{1}{s}\sum_{k}\sum_{t}\sum_{j=1}^{n} \| {}^{k}\mathbf{x}_j - {}^{1}\mathbf{H}_k^{-1} \cdot {}^{1}\mathbf{H}_t \cdot {}^{t}\mathbf{x}_j \|_2 + \| {}^{t}\mathbf{x}_j - {}^{1}\mathbf{H}_t^{-1} \cdot {}^{1}\mathbf{H}_k \cdot {}^{k}\mathbf{x}_j \|_2 \quad (8)$$

where k and t are a pair of images that were successfully matched, n is the total number of correspondences between the overlapping image pairs, s is the total number of correspondences and $({}^{1}\mathbf{H}_k, {}^{1}\mathbf{H}_t)$ are the absolute homographies for images k and t, respectively.

It should be noted that this error measure does not depend on the selected global frame as it uses absolute homographies to compute a relative homography. Therefore, if all the absolute homographies are mapped to any other arbitrary frame, the reprojection error will remain the same. Hence, the first image frame is chosen as a global frame. The evolution of the uncertainty of the initial estimation is illustrated in Fig. 10. The error accumulates as the sequence gets longer and the uncertainty rapidly grows.

The initial estimation and the resulting mosaics are represented in Figs. 12–14. The average reprojection error of the initial estimation is 34.79 pixels. Capel's method has $93{,}515 \times 2 + 860 \times 8 = 193{,}910$ unknowns and the Jacobian matrix is $1{,}383{,}878 \times 193{,}910$. Since Capel's and the proposed method use different error terms, the selected stopping criterion might not have the same meaning for both approaches. Therefore, we have set a threshold of 6 pixels on the average reprojection error in order to compare the computational time of Capel's and the proposed method with and without the use of weights. Capel's method required 31,525 s to obtain an average reprojection error of 5.72 pixels. The proposed method without using uncertainty weights needed 8443 s to achieve an average error, 5.79 pixels, smaller than the threshold. On the other hand, using the uncertainty weights, the same method required 4087 s to reach an average reprojection error 5.77 pixels.

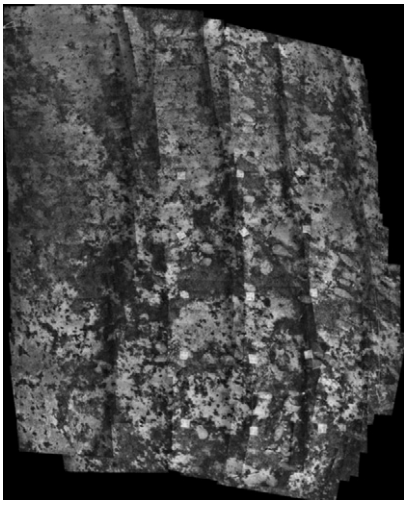

Fig. 18. Resulting mosaic with the proposed method and the distortion reduction approach explained in Section 4.

We have also tested the performance of the methods for the same running time of 19,150 s.[4] Capel's method provided an average reprojection error 8.65 pixels. The proposed method without weights provided an error 5.31 pixels and using weights the average reprojection error was 5.08 pixels. From the results, the use of uncertainties as a weight in the first iteration allowed us to reach the stopping criteria in less iterations, thus reducing the computational cost. Capel's method required 99,876 s to reach a point where the error was not decreasing anymore and the average reprojection error was 4.76 pixels. The proposed method with uncertainty weighting required 22,652 s and the average reprojection error was 5.04 pixels.

The method described in Ferrer et al. (2007) required 44,988 s and the average reprojection error was 5.15 pixels. The resulting mosaic of this last approach can be seen in Fig. 15.

The second dataset is composed of 263 images of size 3008×2000 and covers approximately 300 m^2. The dataset has been acquired in the Mediterranean Sea, close to Pianosa Island. The total number of overlapping image pairs is 4746. The number of correspondences between overlapping pairs is 239,431 and the

[4] This running time was chosen as an approximate mean of the running times of the previous experiment, respectively, 31,525 and 8443 s and was tuned according to Capel's method in order not to stop the method without completing its current iteration.

Fig. 19. Resulting mosaic with the proposed method and incorporating four image centres framework.

Fig. 20. Resulting mosaic with the method in Ferrer et al. (2007).

number of tracked features is 18,614. Therefore, the number of unknowns is $18{,}614 \times 2 + 263 \times 8 = 39{,}332$. The average reprojection error of the initial estimation is 673.56 pixels. Capel's method required 9080 s and the average reprojection error was 37.18 pixels. The proposed method required 1397 s and average error was 35.79 pixels. Table 3 presents the computational time (in seconds) and average reprojection error (in pixels) calculated by using all correspondences for the tested methods over the datasets. All testings were performed in a desktop computer with an Intel Core^{TM}2 2.66 GHz processor, 3 GB RAM and a 64-bit operating system, running MATLAB^{TM}.

The resulting mosaics are illustrated in Figs. 16 and 17. It can be noted that both Capel's and the proposed method caused noticeable distortions to the image size. To quantify the amount of distortion, we have computed the max–min distance ratio (Liu et al., 2008) between the corners of the final mosaics. This criterion requires the knowledge of the true size and/or the ratio of the certain object in the image. However, as our scene is not a man-made environment, it is difficult to define a certain number for this ratio (e.g., if a mosaic has a shape of a square, this ratio must be equal to 1). Therefore, we have used the ratio of the resulting mosaic obtained with bundle adjustment as a comparison baseline. For each mosaic, max–min distance ratios are given in the first row of the Table 4 while the second row shows the deviations of other methods in percentage from the bundle adjustment. From the table and as well as the resulting mosaics, it can be noticed that both Capel's method and the proposed method have caused some distortions on the image size. This is mainly due to the initial estimation. Moreover, in Fig. 17, it can be seen that images which are in the outer transects suffer from higher distortion than the ones located in inner transects in order to get better aligned. This effect is due to the fact that inner images have more overlapping area and contain more tracked features, and during the execution of the first step, the mean position of the tracked features is somewhere closer to the inner images. Therefore, outer images tend to move the most during the optimisation process in order to get better alignment. Alignment in Fig. 17 is better than that of Fig. 16 as the reprojection error is smaller. However, the distortion on the image size is bigger. The proposed method is working on the mosaic frame. Therefore, the resulting mosaic depends on the initial positioning of the images. This is also true for the other methods that require non-linear

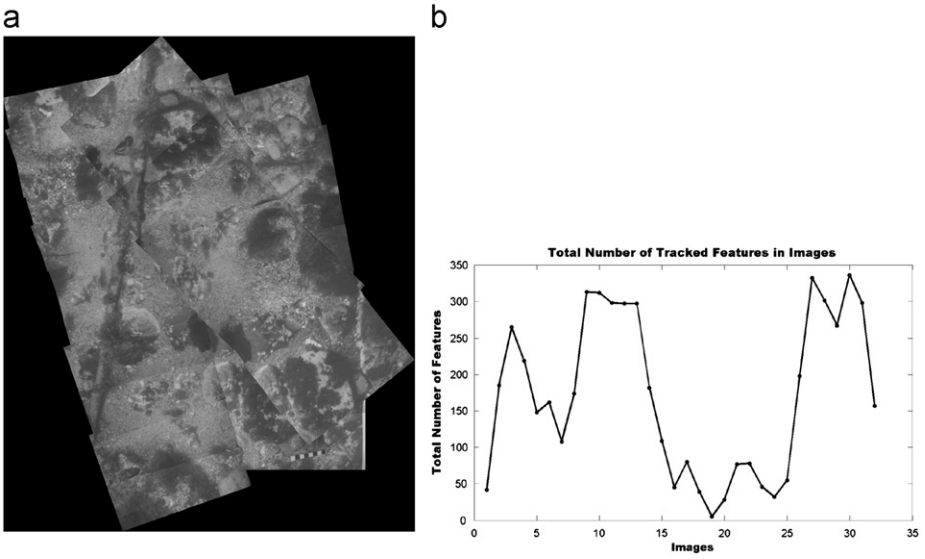

Fig. 21. Initial estimation and number of tracked features of the underwater sequence. (a) Mosaic with the accumulated homographies and (b) distribution of tracked features along images.

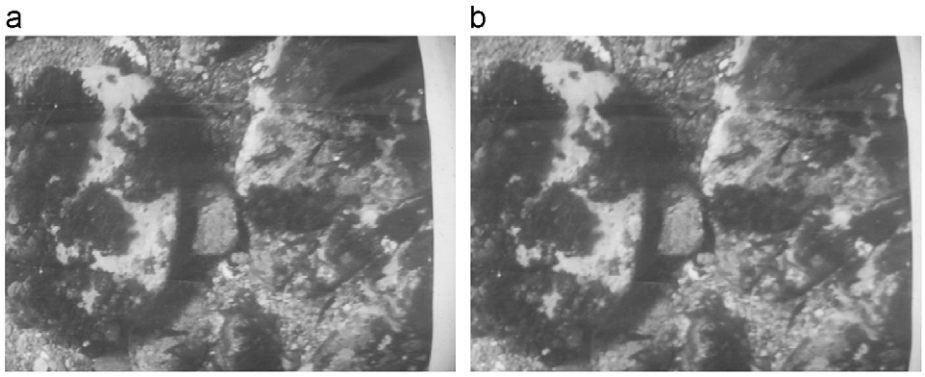

Fig. 22. Radial distortion was partially compensated. (a) Original image acquired by a robot and (b) rectified image.

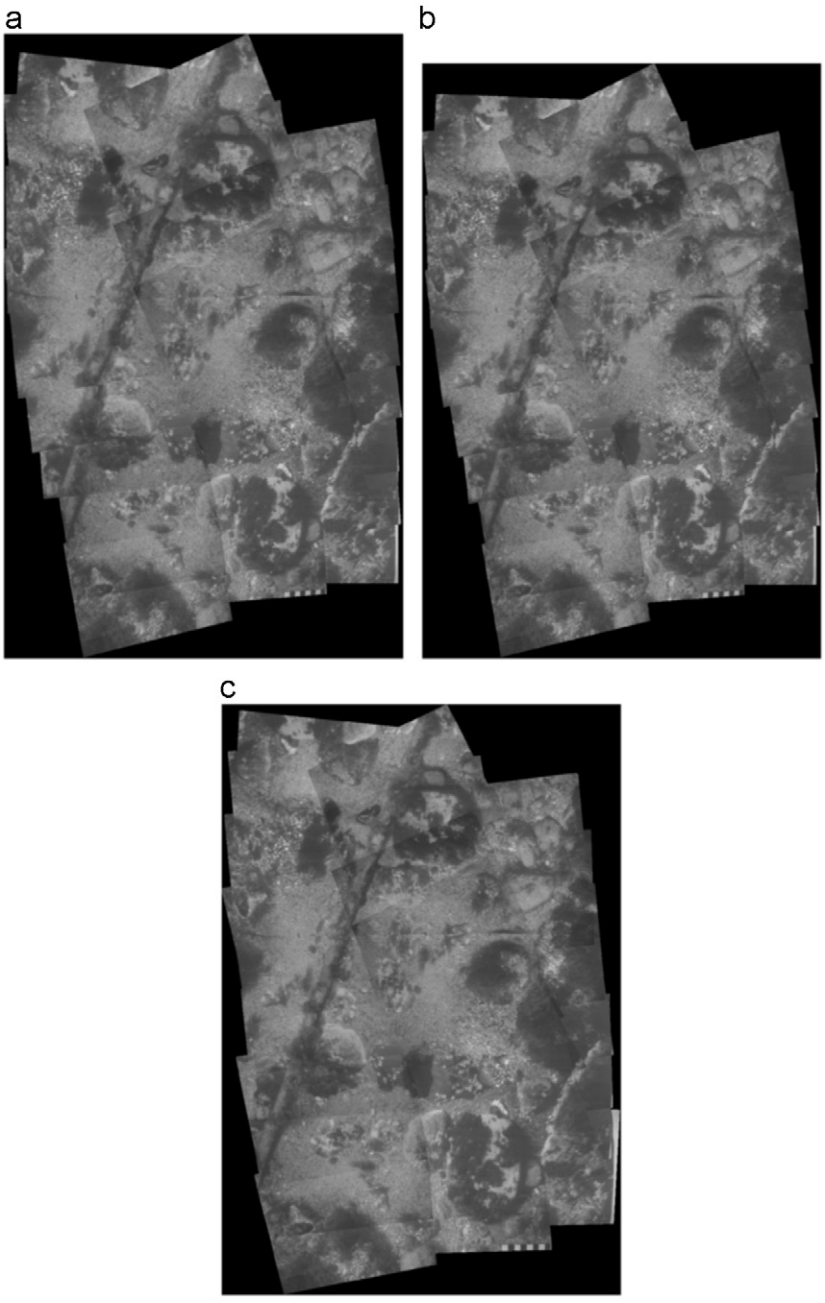

Fig. 23. Resulting mosaics of the underwater image sequence and ground-truth mosaic obtained by registering each image to the poster. (a) Capel's method, (b) proposed method and (c) ground-truth.

optimisation as they need an initial estimation and they might fall into a local minimum which does not provide a correctly aligned mosaic.

It should be noted that the distortion on the image size can be reduced by using the approach detailed in Section 4, which provides the result illustrated in Fig. 18. If some fiducial points and/or navigation data of the robot are known, the available information can be easily incorporated in the proposed approach in Section 4 as well. Fig. 19 shows the resulting mosaic of the proposed framework applied under the knowledge of the real coordinates of the four image centres that are at the top-left, top-right, bottom-left and bottom-right of the mosaic. We have also applied the method proposed in Ferrer et al. (2007) to the dataset. This method takes into account not only image correspondences but also other sensor information when available. Fig. 20 shows the resulting mosaic by using image correspondences and additional sensor information about four image centres that are in the corners of mosaic. From Figs. 19 and 20, one can verify that the resulting mosaics are very similar. Our proposed image rectifying method allows to incorporate the fiducial point information in an easy way, without requiring a non-linear optimisation.

The last image sequence was acquired with remotely operated vehicle develop by our group under controlled conditions, to allow for obtaining positioning ground-truth. For this, a large poster with seafloor texture was placed at the bottom of a test pool. In particular, since the floor of the pool is planar and the robot performs movements in 3D space, camera motion can be adequately described by 2D planar transformations. This image set consists of 159 images of size 376×280 and covers approximately 18 m^2. Additional images of a checkered board were acquired for camera calibration.[5] Before applying our method, the images were compensated for radial distortion. An example of original and rectified images are given in Fig. 22. Thirty-two key-frames were selected by calculating at least 50% overlap. Then, non-time-consecutive overlapping image pairs were found. The total number of overlapping image pairs between key-frames is 150. This number later refined by choosing image pairs that have at least 20% of overlap. Final total number of overlapping image pairs is 85. Fig. 21(a) shows the initial estimation calculated by accumulation. Average reprojection error is 56.50 pixels computed over 32 100 correspondences. If a feature has appeared in three or more images, it is added to the list of tracked features. The total number of tracked features is 1116 and their distribution with the images is given in Fig. 21(b).

The resulting mosaics are depicted in Fig. 23(a) and (b). In both mosaics, some misalignments can be seen because the distribution of tracked features is not close to being uniform. Some images contain very few tracked features, e.g., the 19th image has only five features. The average reprojection error calculated over 32 100 correspondences is 6.79 pixels for Capel's method and 6.83 for the proposed method. The running time for 20 iterations is 34.08 s for Capel's method and 6.95 s for the proposed method.

In order to compare the trajectories obtained by the tested methods, we have registered individual images to the image of the poster and the resulting trajectory was used as a ground-truth. Resulting trajectories can be seen in Fig. 24. Maximum drift between the ground-truth trajectory with the one obtained by the proposed method is 31.01 pixels while it is 61.60 pixels for the trajectory obtained by Capel's method.

[5] The accuracy of the calibration was limited by the fact that only fronto parallel images of the grid were possible to obtain.

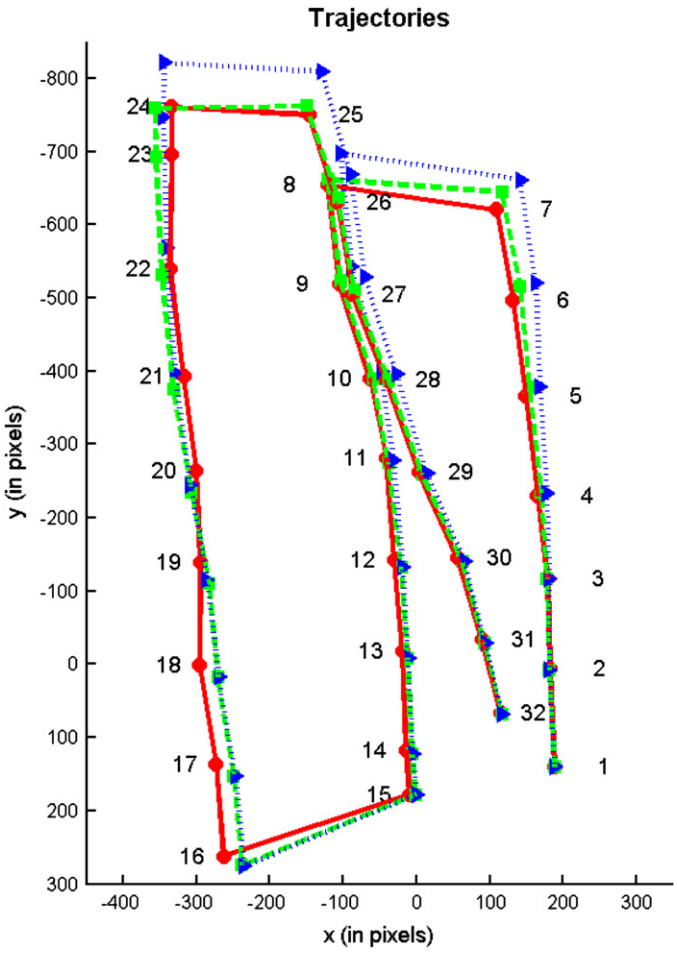

Fig. 24. Solid (red) line shows the ground-truth trajectory obtained by registering individual images to the image of the poster. Dashed (green) line denotes the trajectory obtained by the proposed method while the dotted (blue) line shows the trajectory of Capel's method. Top left corner of the first image is chosen as an origin of the mosaic frame. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

Over the past few years, underwater imagery has become one of the major data sources in different studies of the seabed owing to the impressive development in underwater robotics. However, the underwater medium does not allow images to be taken from a high altitude with respect to the seabed. As a result, a surveyed area cannot be imaged by a single high-resolution image as opposed to the situation in land cartography. This increases the need for image mosaicing methods specially suited to deal with large image sets, as the area of interest in underwater mapping is often quite large. In this work, an iterative global alignment method has been proposed to overcome some of the limitations of current state of the art techniques in image mosaicing. Normally, global alignment requires the minimisation of an error term, which is defined from image correspondences. This error term can be defined either in the image frame or in the mosaic coordinate system. In both cases, non-linear minimisation is required. This new approach provides similar results without the need of non-linear optimisation. We propagate the uncertainties of the initial estimation and use them as a weight during the first iteration. As this models the error of the initial estimation, using the uncertainties as a weight significantly reduces the running time.

The proposed method has been tested with different underwater image sequences covering a large area and comparative

results have been presented to illustrate its performance. As the proposed method is not demanding in terms of computational effort and memory, it makes an important contribution opening up the possibility to deal with very large mosaicing problems using off-the-shelf computing hardware. The results in this paper illustrate that our proposal is faster than its counterparts, while attaining the same level of registration quality. We also introduce a simple rectifying method to overcome the possible distortions that might appear on the image size due to error accumulation and lens distortions. This new method can also be used to quickly fuse any additional navigation data of the robot if available.

Acknowledgements

This work was partially funded through the Spanish Ministry of Education and Science (MCINN) under grant CTM2010-15216, EU Project FP7-ICT-2009-248497. Armagan Elibol was funded by Generalitat de Catalunya under grant 2004FI-IQUC1/00130. Nuno Gracias was supported by MCINN under the *Ramon y Cajal* program.

References

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: speeded up robust features. In: European Conference on Computer Vision, Graz, Austria, pp. 404–417.

Beaudet, P.R., 1978. Rotational invariant image operators. In: IAPR International Conference on Pattern Recognition, pp. 579–583.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.

Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L., 2002. A feature-based technique for joint linear estimation of high-order image-to-mosaic transformations: mosaicing the curved human retina. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (3), 412–419.

Capel, D., 2004. Image Mosaicing and Super-resolution. Springer-Verlag, London.

Cervantes, A., Kang, E.Y., 2006. Progressive multi-image registration based on feature tracking. In: International Conference on Image Processing, Computer Vision, & Pattern Recognition, vol. 2, Las Vegas, USA, pp. 633–639.

Davis, J., 1998. Mosaics of scenes with moving objects. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. I, Santa Barbara, CA, USA, pp. 354–360.

Delaunoy, O., Gracias, N., Garcia, R., 2008. Towards detecting changes in underwater image sequences. In: OCEANS 2008-MTS/IEEE Techno-Ocean, Kobe, Japan, pp. 1–8.

Elibol, A., Garcia, R., Delaunoy, O., Gracias, N., 2008. A new global alignment method for feature based image mosaicing. Advances in Visual Computing Lecture Notes in Computer Science, vol. 5359; 2008, pp. 257–266.

Escartin, J., Garcia, R., Delaunoy, O., Ferrer, J., Gracias, N., Elibol, A., Cufi, X., Neumann, L., Fornari, D.J., Humpris, S.E., Renard, J., 2008. Globally aligned photomosaic of the lucky strike hydrothermal vent field (mid-Atlantic ridge, 37°18.5′N): release of georeferenced data, mosaic construction, and viewing software. Geochemistry Geophysics Geosystems 9 (12), Q12009.

Eustice, R., 2005. Large-area Visually Augmented Navigation for Autonomous Underwater Vehicles. Ph.D. Thesis. Massachusetts Institute of Technology and Woods Hole Oceanographic Institution.

Ferrer, J., Elibol, A., Delaunoy, O., Gracias, N., Garcia, R., 2007. Large-area photo-mosaics using global alignment and navigation data. In: MTS/IEEE OCEANS Conference, Vancouver, Canada, pp. 1–9.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24 (6), 381–395.

Garcia, R., Batlle, J., Cufí, X., Amat, J., 2001. Positioning an underwater vehicle through image mosaicking. In: IEEE International Conference on Robotics and Automation, vol. 3, Seoul, Republic of Korea, pp. 2779–2784.

Gracias, N., Costeira, J.P., Victor, J.S., 2004. Linear global mosaics for underwater surveying. In: 5th IFAC Symposium on Intelligent Autonomous Vehicles, vol. I, Lisbon, Portugal.

Gracias, N., Santos-Victor, J., 2000. Underwater video mosaics as visual navigation maps. Computer Vision and Image Understanding 79 (1), 66–91.

Gracias, N., Zwaan, S., Bernardino, A., Santos-Victor, J., 2003. Mosaic based navigation for autonomous underwater vehicles. IEEE Journal of Oceanic Engineering 28 (3), 609–624.

Haralick, R.M., 1998. Propagating covariance in computer vision. In: Theoretical Foundations of Computer Vision, pp. 95–114.

Harris, C.G., Stephens, M.J., 1988. A combined corner and edge detector. In: Alvey Vision Conference, Manchester, UK, pp. 147–151.

Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision. Cambridge University Press, Harlow, UK.

Kang, E., Cohen, I., Medioni, G., 2000. A graph-based global registration for 2d mosaics. In: International Conference on Pattern Recognition, Barcelona, Spain, pp. 1257–1260.

Leone, A., Distante, C., Mastrolia, A., Indiverr, G., 2006. A fully automated approach for underwater mosaicking. In: MTS/IEEE OCEANS Conference, Boston, USA, pp. 1–6.

Liebowitz, D., Zisserman, A., 1998. Metric rectification for perspective images of planes. In: IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, USA, pp. 482–488.

Lindeberg, T., 1998. Feature detection with automatic scale selection. International Journal of Computer Vision 30 (2), 77–116.

Liu, X., Doermann, D., Li, H., Lee, K.C., Ozdemir, H., Liu, L., 2008. A novel 2D marker design and application for object tracking and event detection. In: Proceedings of the 4th International Symposium on Advances in Visual Computing. Springer, Las Vegas, USA, pp. 248–257.

Loisel, H., Stramski, D., 2000. Estimation of the inherent optical properties of natural waters from irradiance attenuation coefficient and reflectance in the presence of Raman scattering. Applied Optics 39, 3001–3011.

Lourakis, M., Argyros, A., 2008. SBA: a software package for generic sparse bundle adjustment. ACM Transactions on Mathematical Software 36 (1).

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60 (2), 91–110.

Marzotto, R., Fusiello, A., Murino, V., 2004. High resolution video mosaicing with global alignment. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. I, Washington, DC, USA, pp. 692–698.

Meer, P., Mintz, D., Rosenfeld, A., 1992. Analysis of the least median of squares estimator for computer vision applications. In: IEEE Conference on Computer Vision and Pattern Recognition, Champaign, USA, pp. 621–623.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10), 1615–1630.

Mohan, S., Avinash, N., Murali, S., 2008. Rectification of perspective distortion using camera parameters—a perspective geometry based approach. ICGST International Journal on Graphics, Vision and Image Processing GVIP 08, 1–7.

Negahdaripour, S., Xu, X., 2002. Mosaic-based positioning and improved motion-estimation methods for automatic navigation of submersible vehicles. IEEE Journal of Oceanic Engineering 27 (1), 79–99.

Ochoa, B., Belongie, S., 2006. Covariance propagation for guided matching. In: 3rd Workshop on Statistical Methods in Multi-Image and Video Processing (SMVP), Graz, Austria.

Pegau, W.S., Gray, D., Zaneveld, J.R.V., 1997. Absorption and attenuation of visible and near-infrared light in water: dependence on temperature and salinity. Applied Optics 36, 6035–6046.

Pizarro, O., Singh, H., 2003. Toward large-area mosaicing for underwater scientific applications. IEEE Journal of Oceanic Engineering 28 (4), 651–672.

Ribas, D., Palomeras, N., Ridao, P., Carreras, M., Hernandez, E., 2007. Ictineu AUV wins the first SAUC-E competition. In: IEEE International Conference on Robotics and Automation, Roma, Italy.

Richmond, K., Rock, S.M., 2006. An operational real-time large-scale visual mosaicing and navigation system. In: MTS/IEEE OCEANS Conference, Boston, USA.

Rzhanov, Y., Linnett, L., Forbes, R., 2000. Underwater video mosaicing for seabed mapping. In: Proceedings of International Conference on Image Processing, vol. 1, Vancouver, Canada, pp. 224–227.

Rzhanov, Y., Mayer, L., Beaulieu, S., Shank, T., Soule, S., Fornari, D., 2006. Deep-sea geo-referenced video mosaics. In: MTS/IEEE OCEANS Conference, Boston, USA, pp. 2319–2324.

Sawhney, H., Hsu, S., Kumar, R., 1998. Robust video mosaicing through topology inference and local to global alignment. In: European Conference on Computer Vision, vol. II, Freiburg, Germany, pp. 103–119.

Szeliski, R., Shum, H., 1997. Creating full view panoramic image mosaics and environment maps. In: SIGGRAPH International Conference on Computer Graphics and Interactive Techniques, vol. I, Los Angeles, CA, USA, pp. 251–258.

Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., Ayache, N., 2006. Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. Medical Image Analysis 10 (5), 673–692.

Vincent, A., Pessel, N., Borgetto, M., Jouffroy, J., Opderbecke, J., Rigaud, V., 2003. Real-time geo-referenced video mosaicing with the matisse system. In: MTS/IEEE OCEANS Conference, vol. 4, San Diego, USA, pp. 2319–2324.

Wang, X., Klette, R., Rosenhahn, B., 2005. Geometric and photometric correction of projected rectangular pictures. In: Proceeding of Image and Vision Computing New Zealand, Dunedin, New Zealand, pp. 223–228.

Zitová, B., Flusser, J., 2003. Image registration methods: a survey. Image and Vision Computing 21 (11), 977–1000.