Chapter 2
Underwater 2D Mosaicing

Abstract The current chapter describes the main steps involved in the photo-mosaic building process. These steps comprehend the geometrical registration and warping of the images into a single common reference frame, along with an estimation of the topology of the trajectory performed by the UV, and a global alignment of the recovered trajectory. A widely extended geometrical registration strategy consists of identifying common image features between the involved images, using different image feature detectors. These image features, once identified, become correspondences that are used to estimate the camera motion between consecutive images, as well as to perform a global alignment of the estimated trajectory. Global alignment of all the involved images allows providing geometrical consistence to the underwater map. At the end of the chapter the problems and issues of the photo-mosaicing process are pointed out, with the aim of demonstrating the relevance of image blending techniques as a final step of the photo-mosaicing process.

Keywords Photo-mosaicing · Image registration · Image alignment · Image warping · Topology estimation · Global alignment · Deep-ocean surveys

Building a photo-mosaic is a task involving two main steps. Firstly, the images should be geometrically registered and warped accordingly into a single common reference frame. Secondly, the rendering of the mosaic should be performed through blending techniques, which allow us to deal with photometric differences and reduce the visibility of registration inaccuracies between the images involved (see Fig. 2.1).

In the context of large-scale underwater photo-mosaicing, deep-ocean surveys are typically composed of hundreds to hundreds of thousands of images. These images are affected by several underwater phenomena, such as the aforementioned scattering and light attenuation, and the sequences may present small or even nonexistent overlaps between consecutive frames. For these reasons, navigation data coming from acoustic positioning sensors (USBL, LBL), velocity sensors (DVL), inclinometers or gyroscopes might be used to estimate the trajectory of the vehicle. This trajectory can

R. Prados et al., *Image Blending Techniques and their Application in Underwater Mosaicing*, SpringerBriefs in Computer Science, DOI: 10.1007/978-3-319-05558-9_2,

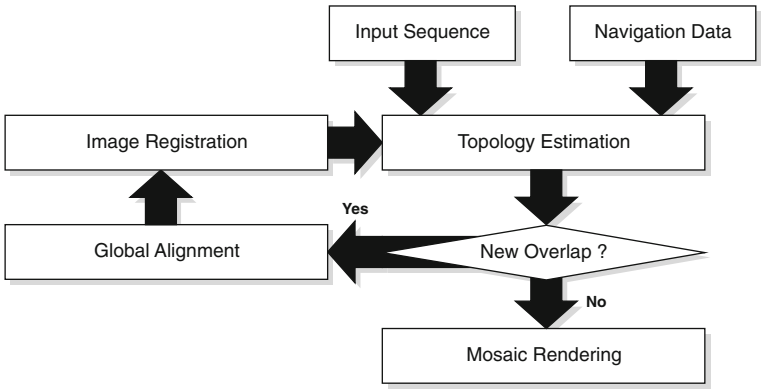

Fig. 2.1 Underwater mosaicing pipeline scheme. The *Topology Estimation*, *Image Registration*, and *Global Alignment* steps can be performed iteratively until no new overlapping images are detected

be later refined by computing *pair-wise alignment* and applying a *global alignment* method [1–7].

2.1 Topology Estimation

When lacking sensor positioning data, such as USBL, LBL or DVL, using time-consecutive image registration, assumed to have an overlapping area, may become the only strategy to estimate the trajectory of the robot and, thus, the motion of the camera. This dead-reckoning estimate suffers from a rapid accumulation of registration errors, leading to drifts from the actual trajectory, but it does provide useful information for non-time-consecutive overlapping images. Matching non-time-consecutive images is a key step in refining the trajectory followed by the robot using global alignment methods. With the refined trajectory, new-non time-consecutive overlapping images can be predicted and attempted to match. This iterative matching and optimization process continues until no new overlapping images are detected. The procedure described is known as *topology estimation* [8, 9] (see Fig. 2.2). If navigation data is available, the topology estimation remains as an indispensable step to obtain globally consistent mosaics and accurate trajectory estimates, specially when dealing with sequences of a large number of images.

Deep-ocean surveys composed of thousands of images make any kind of all-to-all image pair matching strategy to perform a topology estimation unfeasible. Therefore, more sophisticated approaches are needed to perform this task. Elibol et al. [8] proposed an Extended Kalman Filter (EKF) framework, aimed at minimizing the total number of matching attempts and simultaneously obtaining the best possible trajectory. Potential image pairs are predicted by taking into account

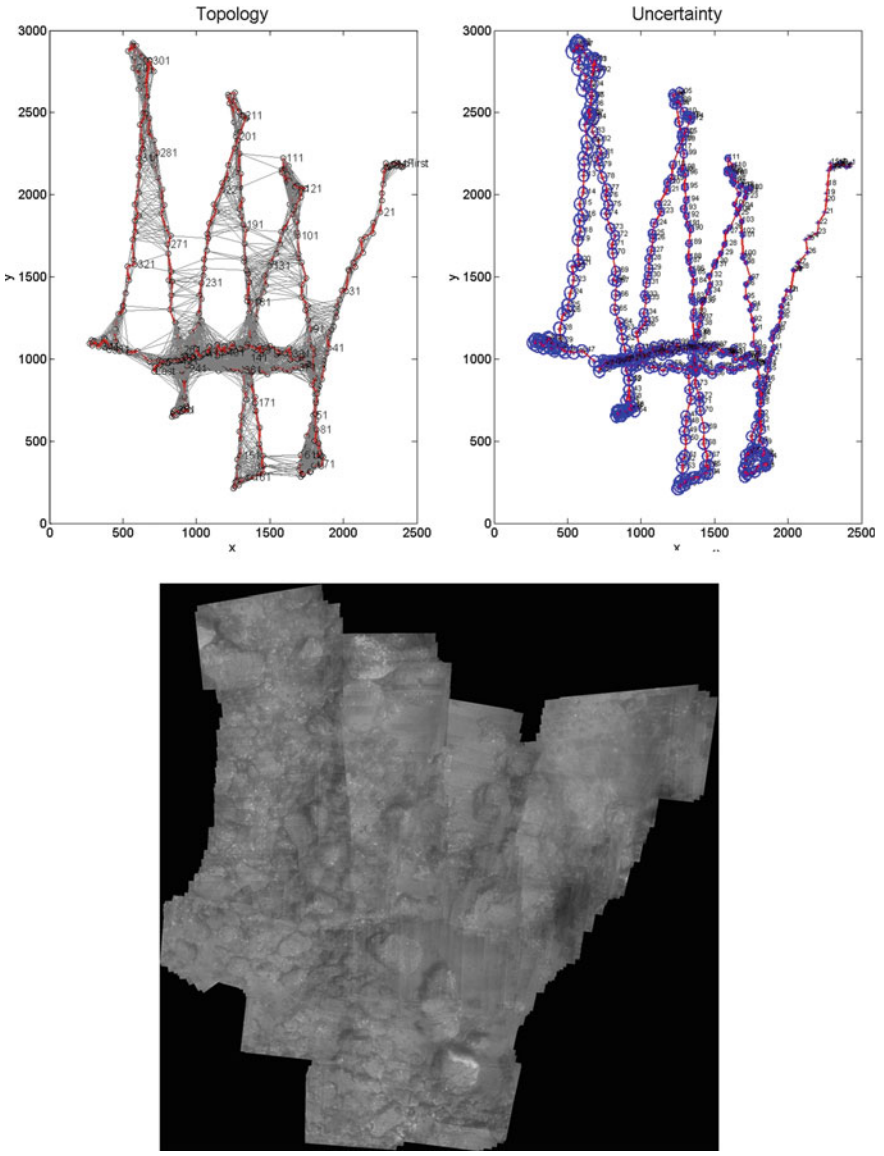

Fig. 2.2 Topology estimation scheme. (*Top-left*) Final trajectory obtained by the scheme proposed in [8]. The first image frame is chosen as a global frame and all images are then translated in order to have positive values in the axes. The x and y axes are in pixels and the scale is approximately 150 pixels per metre. The plot is expressed in pixels instead of metres since the uncertainty of the sensor used to determine the scale (an acoustic altimeter) is not known. The *red lines* join the time-consecutive images while the *black* ones connect non time-consecutive overlapping image pairs. The total number of overlapping pairs is 5,412. (*Top-right*) Uncertainty in the final trajectory. Uncertainty of the image centres is computed from the covariance matrix of the trajectory [5]. The uncertainty ellipses are drawn with a 95 % confidence level. (*Bottom*) Mosaic built from the estimated trajectory

Fig. 2.3 Geometric registration of two different views of the same underwater scene by means of a planar transformation

the uncertainty of the trajectory. Additionally, a different solution to the topology estimation problem in a Bundle Adjustment (BA) framework was proposed in [10]. To obtain a tentative topology, a fast image similarity criterion combined with a Minimum Spanning Tree (MST) solution are used. The topology is improved by attempting image-matching with the pairs of images for which there is the most overlapping evidence.

2.2 Image Registration

Aligning in 2D two or more images taken from different viewpoints consists of finding an appropriate planar transformation which allows overlaying them into a single and common reference frame (see Fig. 2.3). This step, essential in the image mosaicing pipeline, is known as the image registration problem [11] and has been greatly discussed in the literature [12, 13].

The geometrical registration can be performed by means of *direct methods* or *feature-based methods*. Sections 2.2.1 and 2.2.2 present these two main groups of image registration methods.

2.2.1 Direct Methods

This first group of algorithms, also known as feature-less methods, compute the transformation between images by maximizing the photometric consistency over the whole overlapping image regions, and are found to be useful for large overlapping regions as well as small translations and rotations [12, 14, 15]. These methods can be classified in turn into *frequency domain based methods* and *optical flow methods*.

Frequency Domain

Methods based on the frequency domain originally used phase-correlation to estimate the shifts (translations) between an image pair. Later, extensions to account for rotation and scale transformations [16] and affine transformations [17] using log-polar coordinates were also proposed. In practice, the number of authors proposing the use of frequency domain methods for underwater image registration is small [18, 19]. This group of methods are computationally expensive, as they require Fast Fourier Transform (FFT) to be computed over all the images involved.

Optical Flow

Optical flow methods are based on the estimation of the disparity (i.e. apparent motion) of pixels between image pairs. Generally, the optical flow estimates the flow field using the Brightness Constancy Model (BCM), in which it is assumed that the photometric properties of image pixels (luminance and color) remain constant. There are two main groups of algorithms estimating the optical flow. On the one hand, *global methods* such as Horn and Schunck [14] yield dense flow fields, while, on the other hand, *local methods* such as Lucas and Kanade [20, 21] produce non-dense regularized grid flow fields but are less robust to noise. Over the last years, some authors have proposed more robust alternatives to BCM that assume linear changes in illumination, using the Generalized Dynamic Image Model (GDIM) [22, 23] and the color information [24, 25]. Due to the formulation of the problem, optical flow methods are not suited for disparities that exceed 1 pixel. To overcome this issue, multi-resolution approaches such as [26] have been proposed. In this case, the images are gradually decimated and the optical flow is computed from coarse levels towards fine levels. Unfortunately, the method also has some drawbacks. Firstly, it is slow because the optical flow has to be computed at each level. Secondly, the maximum pixel disparity has to be known *a priori* in order to set the number of decimation levels. Furthermore, multi-resolution approaches are very sensitive to noise, since errors in the estimation of optical flow at coarse levels propagate to the fine levels.

2.2.2 Feature-Based Methods

The second group of methods rely on the computation of a transformation between images using a sparse set of points [27–31] and correspondences. Contrarily to direct methods, feature based methods do not require a high frame-rate to ensure a high percentage of overlap between consecutive images. For these reasons, feature-based methods are the most widely used in the literature to perform image registration, and are also used in the work presented, as described in the following sections.

There are two main strategies concerning feature-based pair-wise image alignment (see Fig. 2.4). The first strategy consists of locating the interest points in one image of

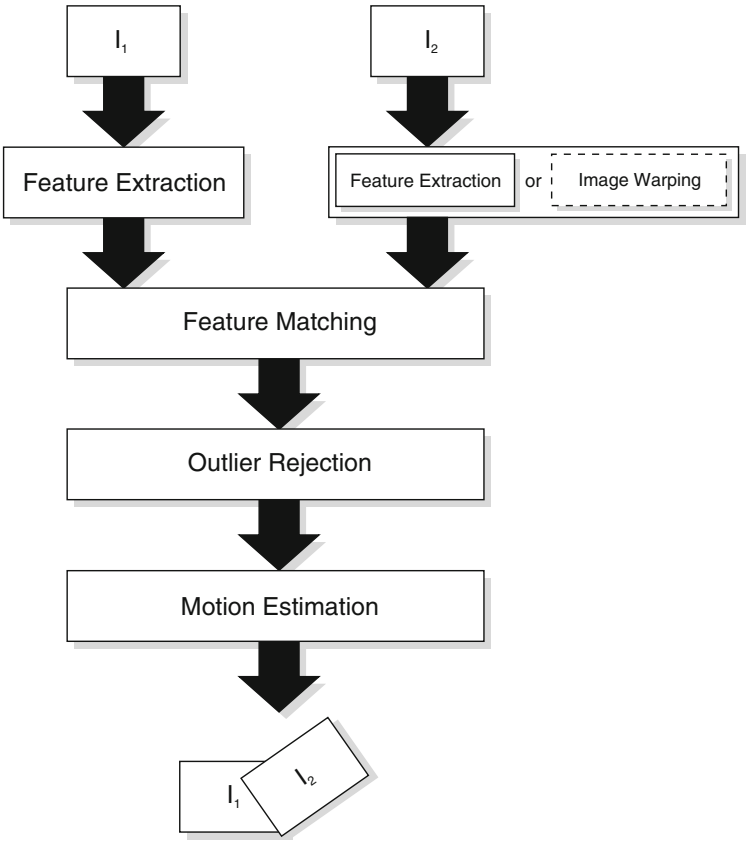

Fig. 2.4 Main steps involved in the pair-wise registration process. The Feature Extraction step can be performed in both images of the pair, or only in one. In this last case, the features are identified in the second image after an optional Image Warping based on a transformation estimation

the pair using some feature detector, such as Harris and Stephens [27], Beaudet [28] or Lindeberg [29], and identifying these in the other. The correspondence problem is solved using cross-correlation or a Sum of Squared Differences (SSD) measure, which is computed using the information of the pixels surrounding the feature, and compared to the value of this measure for a given window of pixels in the other image. The procedure has the advantage of obtaining highly accurate correspondences when changes in rotation and scale are moderate. As a drawback, this strategy requires some prior knowledge to determine the estimated translation between images and the size of the search window, in addition to not being suitable for large changes in rotation and scale. For these reasons, this approach might be used as a refinement step of certain feature-based image alignment strategies [5], after an appropriate warping of the image in which the features found should be identified.

The second strategy is based on the detection of features in both images using invariant feature descriptors, such as SIFT [30], its faster variant SURF [31] (which uses an approximation of the Laplacian and Hessian detectors respectively) or others, and performing the matching, comparing their descriptor vectors. The SIFT descriptor is based on Histograms of Gradient (HOGs) computed in the area surrounding the detected interest points, while SURF describes a distribution of Haar wavelet [32] responses within the neighborhood of the interest point. These feature detectors and descriptors are known to show invariance to a wider range of geometrical and photometrical [33] transformations of the images than the detectors mentioned above. Therefore, these detector and descriptor properties allow us to obtain very robust results, even in the case of strong rotations or scale changes between frames and significant illumination inhomogeneities.

2.3 Motion Estimation

2.3.1 Planar Homography

The planar transformation between two different views of the same flat scene can be described by means of a *planar homography* matrix [34, 35]. This homography is able to describe a motion with up to eight Degrees of Freedom (DOF).

Let us consider a point p, belonging to a 2D plane Π in 3D space. Then, the projections of p into two different images I_1 and I_2 are given in $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^3$ in homogeneous coordinates. Also let the coordinate transformation between the two frames be

$$\mathbf{X}_2 = R\mathbf{X}_1 + T \tag{2.1}$$

where $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^3$ are the 3D coordinates of p relative to camera frames 1 and 2, respectively, taken at times t_1 and t_2. The two projections $\boldsymbol{x}_1, \boldsymbol{x}_2$ of p in images I_1 and I_2 satisfy the epipolar constraint [34]

$$\boldsymbol{x}_2^T E \boldsymbol{x}_1 = \boldsymbol{x}_2^T \widehat{T} R \boldsymbol{x}_1 = 0 \tag{2.2}$$

where E is the essential matrix, containing information about the relative position T and orientation R between the two camera frames 1 and 2, and $\hat{T}$ is the skew-symmetric matrix codifying position T [35].

However, for points on the same plane Π, their images will share an extra constraint that makes the epipolar constraint alone no longer sufficient.

Let $N = [n_1, n_2, n_3]^T \in \mathbb{S}^2$ be the unit normal vector of the plane Π with respect to the first camera frame, and let $d > 0$ denote the distance from the plane Π to the optical center of the first camera. Then we have

$$N^T \mathbf{X}_1 = n_1 X + n_2 Y + n_3 Z = d \quad \Leftrightarrow \quad \frac{1}{d} N^T \mathbf{X}_1 = 1, \quad \nabla \mathbf{X}_1 \in \Pi \tag{2.3}$$

Substituting Eq. (2.3) into Eq. (2.2) gives

$$\mathbf{X}_2 = R\mathbf{X}_1 + T = R\mathbf{X}_1 + T\frac{1}{d} N^T \mathbf{X}_1 = \left(R + \frac{1}{d} T N^T \right) \mathbf{X}_1 \tag{2.4}$$

Then matrix H is defined as follows

$$H = R + \frac{1}{d} T N^T \in \mathbb{R}^{3x3} \tag{2.5}$$

where H is the *(planar) homography matrix*, since it denotes a linear transformation from $\mathbf{X}_1 \in \mathbb{R}^3$ to $\mathbf{X}_2 \in \mathbb{R}^3$ as

$$\mathbf{X}_2 = H\mathbf{X}_1 \tag{2.6}$$

Note that the matrix H depends on the motion parameters R, T as well as the structure parameters N, d of the plane Π. Due to the inherent scale ambiguity in the term $\frac{1}{d} T$ in Eq. (2.5), one can at most recover from H the ratio of the translation T scaled by the distance d.

From

$$\lambda_1 \boldsymbol{x}_1 = \mathbf{X}_1, \quad \lambda_2 \boldsymbol{x}_2 = \mathbf{X}_2, \quad \lambda_2 \boldsymbol{x}_2 = H\mathbf{X}_1 \tag{2.7}$$

we have

$$\lambda_2 \boldsymbol{x}_2 = H\lambda_1 \boldsymbol{x}_1 \quad \Leftrightarrow \quad \boldsymbol{x}_2 \sim H\boldsymbol{x}_1 \tag{2.8}$$

where we recall that $\sim$ indicates equality up to a scale factor. Often, the equation

$$\boldsymbol{x}_2 \sim H\boldsymbol{x}_1 \tag{2.9}$$

itself is referred to as a *(planar) homography* mapping induced by a plane Π.

The homography matrix H encodes information about the camera motion and the scene structure, a fact that facilitates establishing correspondence between points in the first and second images. H can be computed in general from a small number of corresponding image pairs.

2.3.2 Planarity Assumption

The homography matrix allows the description of 2D transformations between images. This motion estimation assumes that the scene is planar (i.e. flat), but this scenario is rare in practice. Nevertheless, it is possible to apply a homography matrix

to register different views of the same scene, even if it is not planar, under certain conditions.

On the one hand, it is possible to use a homography matrix to model the transformation between images when the camera only describes a rotation or change in scale around the same optical center. On the other hand, it can also be assumed that a scene is planar when the camera describes a translation but the magnitude of the scene relief is negligible compared to the distance between the camera and the scene. In any other cases images show the *parallax* effect, i.e. the difference in the apparent position of an object viewed along two different lines of sight, measured by the angle of inclination between those two lines.

The parallax effect impacts both the registration and blending steps. When registering a pair of images showing parallax, the computed homography will try to encode the dominant motion between both views. In that case, if the structures causing the parallax are large enough with respect to the image size, errors in the motion estimation may arise. Furthermore, if two images suffering from parallax are successfully registered, i.e. the dominant motion has been correctly estimated, evident misalignments may appear when overlying both views. This scenario is common in underwater imagery, where the distance between the camera and the scene is not always as important as desired, and consequently image blending techniques have to deal with this problem.

2.3.3 Outlier Rejection

The homography accuracy [36] is strongly tied to the quality of the correspondences used for its calculation. The homography estimation algorithms assume that the only source of error is the measurement of the locations of the points, but this assumption is not always true inasmuch as mismatched points may also be present. There are several factors that can influence the goodness of the correspondences detected. Images can suffer from several artifacts, such as non-uniform illumination, sun flickering (in shallow waters), shadows (specially in the presence of artificial lighting) and digital noise, among others, which can make matching fail. Furthermore, moving objects (including shadows) may induce correspondences which, despite being correct, do not obey the dominant motion between the two images. These correspondences are known as *outliers*. Consequently, it is necessary to use an algorithm able to discern right and wrong correspondences. There are two main strategies to reject outliers widely used in the bibliography [37]: Random Sample Consensus (RANSAC) [38] and Least Median of Squares (LMedS) [39]. LMedS efficiency is very low in presence of Gaussian noise [40, 41]. For this reason, RANSAC has been selected as outlier rejection method in the presented framework.

RANSAC is a robust estimator intended to fit a model to experimental data and is able to smooth data containing a significant percentage of gross errors. This feature makes the approach suitable for image processing applications, where error-prone data is quite frequent. As stated in [38], contrary to other smoothing techniques,

instead of using as much data as possible to obtain an initial solution and then attempting to eliminate the invalid data, RANSAC uses a small set of data as a point of departure and enlarges this set with consistent data when possible. When there is enough data, RANSAC can use a smoothing technique, such as least squares, to compute an improved estimate for the parameters of the model with the mutually consistent data which has been identified. The RANSAC paradigm is tuned up by three parameters: the error tolerance used to determine the compatibility of a given data point to the model, the number N of subsets S_i with size s used to instantiate the model and the threshold T that determines the number of points required to consider that a correct model has been found. RANSAC tries to compute a model candidate based on a set of s data points from S selected randomly. The model is next applied to the rest of the data in order to determine the set of points S_i that are within a distance of a defined threshold. If the size of S_i is greater than any predefined threshold T, the model can be re-estimated with the points in S_i. Otherwise, if the size of S_i is lower than T, a new subset is selected and the process is repeated. After N trials, the largest consensus set S_i is selected and the model is re-estimated. Reliable RANSAC estimates requires that at least one of the candidate models contains the correct parameter values, otherwise the estimator loses its effectiveness.

2.4 Global Alignment

Pair-wise registration of images acquired by an underwater vehicle equipped with a down-looking camera cannot be used as an accurate trajectory estimation strategy. Image noise, illumination issues and the violation of the planar assumption may unavoidably lead to an accumulative drift. Therefore, detecting correspondences between non-consecutive frames becomes an important step in order to close a loop and use this information to correct the estimated trajectory.

The homography matrix ${}^1\mathbf{H}_k$ represents the transformation ot the kth image with respect to the global frame (assuming the 1st frame frame as a global frame) and is known as *absolute homography*. This ${}^1\mathbf{H}_k$ matrix is obtained as a result of the concatenation of the *relative homographies* ${}^{k-1}\mathbf{H}_k$ between the kth and ${}^{k-1}k$th images of a given time-consecutive sequence. As mentioned above, relative homographies have limited accuracy and computing absolute homographies by cascading them results in cumulative error. This drift will cause, in the case of long sequences, the presence of misalignments between neighboring images belonging to different transects (see Fig. 2.5).

The main benefit of *global alignment* techniques is the use of the closing-loop information to correct the pair-wise trajectory estimation by reducing the accumulated drift.

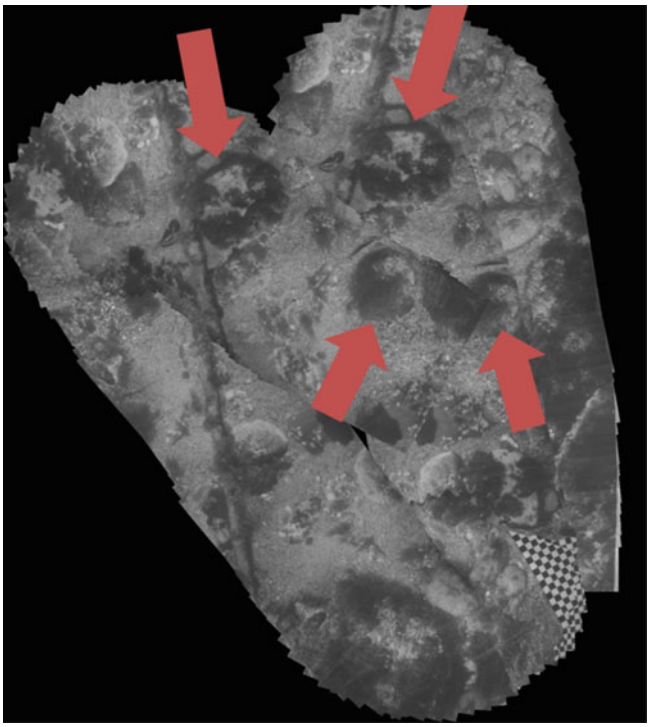

Fig. 2.5 Example of error accumulation from registration of sequential images. The same benthic structures appear in different locations of the mosaic due to error accumulation (trajectory drift)

2.4.1 Global Alignment Methods

There are several methods in the literature intended to solve the global alignment problem [42]. Global alignment methods usually require the minimization of an error term based on the location of the image correspondences. These methods can be classified according to the domain where this error is defined, leading to two main groups: image frame methods [1, 5, 43, 44] and mosaic frame methods [2, 4, 45–48].

Concerning the group of image frame based methods, Davis [45] faced the problem of a camera rotating around its optical axis without translation. The absolute homography was obtained as an accumulation of relative homographies (see Eq. 2.10), and computed solving a sparse linear systems of equations.

$$ {}^{1}\mathbf{H}_i = \prod_{j=2}^{i} {}^{j-1}\mathbf{H}_j \qquad i \geq 2 \tag{2.10} $$

Any image i of a given sequence can be projected to another image space j or to the global frame using the absolute homography of image j, i.e. ${}^{1}\mathbf{H}_i = {}^{1}\mathbf{H}_j \cdot {}^{j}\mathbf{H}_i$,

where ${}^1\mathbf{H}_i$ and ${}^1\mathbf{H}_j$ are unknown and ${}^j\mathbf{H}_i$ is a relative homography. When a closing loop happens, the number of relative homographies becomes greater than the number of images, leading to an over-determined system. Unfortunately, the over parameterization of the system might lead to overfitting if an adequate parametrization of the resolution method is not used.

Another image frame based method was proposed by Shum and Szeliski [49], who defined the error function as:

$$\min_{{}^1\mathbf{H}_2,{}^1\mathbf{H}_3,\ldots,{}^1\mathbf{H}_N} \sum_k \sum_m \sum_{j=1}^{n} \| {}^k\boldsymbol{x}_j - {}^1\mathbf{H}_k^{-1} \cdot {}^1\mathbf{H}_m \cdot {}^m\boldsymbol{x}_j \|_2 \tag{2.11}$$

where ${}^k\boldsymbol{x}_j$ and ${}^m\boldsymbol{x}_j$ are the jth correspondence between images k and m having an overlap area, n the number of correspondences and $\|\cdot\|_2$ the Eculidean norm. Calculating the solution by means of a non-linear least squares minimization has a drawback: the gradients with respect to the motion parameters are quite complicated and have to be provided for the minimization method chosen, e.g. Levenberg-Marquadt.

In the group of mosaic frame based methods, Sawhney et al. [2], proposed a method based on the following error function:

$$E_1 = \min_{{}^1\mathbf{H}_2,{}^1\mathbf{H}_3,\ldots,{}^1\mathbf{H}_N} \sum_k \sum_m \sum_{j=1}^{n} \| {}^1\mathbf{H}_k \cdot {}^k\boldsymbol{x}_j - {}^1\mathbf{H}_m \cdot {}^m\boldsymbol{x}_j \|_2 \tag{2.12}$$

Nevertheless, this solution suffers from what is known as scaling effect of a mosaic-based cost function if no constraints are imposed. This is due to the fact that the cost function has lower values when the image size is smaller, and consequently the function tends to reduce this image size. For that reason, Sawhney et al. [2] extended the method by introducing another term for controlling the scaling effects:

$$E_2 = \sum_{i=1}^{N} \left(\| {}^1\mathbf{H}_i \cdot \boldsymbol{x}_{tr} - {}^1\mathbf{H}_i \cdot \boldsymbol{x}_{bl} - (\boldsymbol{x}_{tr} - \boldsymbol{x}_{bl}) \|_2 + \| {}^1\mathbf{H}_i \cdot \boldsymbol{x}_{tl} - {}^1\mathbf{H}_i \cdot \boldsymbol{x}_{br} - (\boldsymbol{x}_{tl} - \boldsymbol{x}_{br}) \|_2 \right) \tag{2.13}$$

where x_{tr}, x_{bl}, x_{tl} and x_{br} denote the top-right, bottom-left, top-left and bottom-right coordinates of the image corners. E_2 tries to minimize the difference in the diagonal length between the original image size and the image size once projected on the mosaic frame. Nevertheless, this constraint may lead to image misalignments because it violates the distance minimization between correspondences. A weighting factor for this penalization is used, which can be fixed or proportionally grow along the sequence due to error accumulation. The final error function E is the result of the addition of both E_1 and E_2 terms, i.e. $E = E_1 + E_2$. The minimization of this function leads to solutions related by a common translation and rotation that have the same minima [50]. Therefore, Sawhney et al. [2] proposed a new term $\mathbf{H}_1 \cdot (0, 0, 1)^T$ to be added to the error function, in order to fix the problem with the translation of the first image and find only a single solution set. Another solution for this issue has

been proposed by Gracias et al. [4], who fixed one of the image frames as the global mosaic frame and aligned all the images with respect to this one.

Sawhney et al. [2] proposed a graph-based representation of the mosaic for closed loop trajectories. In this case, each node of the graph represents an image whilst each edge represents overlapping areas between the images. Initially, the graph is built only with edges between consecutive images. Edges between non-consecutive images can be added by measuring the distances between the image centers. The goal of this graph is to reduce the total number of products by searching for the optimal path while computing absolute homographies through relative homographies [44, 46], with the aim of reducing the accumulated drift and image distortions.

In the graph representation context, Kang et al. [46] presented an approach to solve the global alignment problem also based on graphs to define the temporal and spatial connectivity between images. Initially, a regular grid of the global frame is defined. Each node of the graph contains a list of corresponding grid points and several lists with the correspondences between these grid points and the points in other images. The correspondences are computed by means of normalized correlation, and the error function is defined as the photometric luminance differences between the points in the mosaic and their projection in other images:

$$E = \sum_i (I_m(\mathbf{p}) - I_i(\mathbf{p}'))^2 \tag{2.14}$$

where $I_m(\mathbf{p})$ is the luminance value of $\mathbf{p}$ in the mosaic and $I_i(\mathbf{p}')$ is the luminance of the projection $\mathbf{p}' =^m \mathbf{H}_i \cdot {}^1\mathbf{p}$ in the ith image. This error function is used to find all the correspondences of each point in the initial grid. The global registration of the different frames is performed by searching for the optimal path connecting each frame to the reference frame. This path, in its turn, is computed by the geometric distance and correlation score between each grid point and its correspondences. Once the images have been registered to the global frame, the location of grid points is adjusted using as a weighting average factor the correlation score between correspondences. Finally, the absolute homographies computed from the accumulation of the relative ones can be recomputed by means of an adjustment transformation, using a linear transformation between the refined grid points and their correspondences.

Marzotto et al. [44] presented a solution close to their of Sawhney et al. [2], which adds another measure to the overlap measure in:

$$d_{ij} = \frac{\max(|\boldsymbol{x}_i - \boldsymbol{x}_j| - |\mathbf{r}_i - \mathbf{r}_j|/2)}{\min(\mathbf{r}_i, \mathbf{r}_j)} \tag{2.15}$$

where x_i and x_j are warped image centers and r_i and r_j are warped image diameters. This additional measure is defined as:

$$\beta_{ij} = \frac{\delta_{ij}}{\Delta_{ij}} \tag{2.16}$$

where δ_{ij} is the overlap measure and Δ_{ij} is the cost of the shortest path between nodes i and j. The optimal path is found by using β values, and the cost is calculated from the weights, d, on the edges. The absolute homographies are obtained as a result of the product of relative homographies through the optimal path. The main advantage of this method to compute the optimal path is that the homographies are less affected by cumulative errors. Similarly to [2], the error function used in the global alignment is defined over a set of grid points, being the error of a given grid point x_k:

$$E_k = \frac{1}{n} \sum_i \sum_j \| \boldsymbol{x}_k - {}^m\mathbf{H}_i \cdot {}^i\mathbf{H}_j \cdot \mathbf{H}_j^{-1} \boldsymbol{x}_k \|_2 \tag{2.17}$$

where n is the number of edges between images containing the grid point x_k and ${}^m\mathbf{H}_i$ and ${}^m\mathbf{H}_j$ denote absolute homographies. The error function is defined as:

$$\min E = \sum_i^m E_i^2 \tag{2.18}$$

where m is the total number of grid points. Unfortunately, there are two main drawbacks to this approach. The first is that point locations have to be carefully selected to ensure enough grid points in both images and overlapping regions in order to compute the homography. The second is that arbitrarily distributed points may fall into textureless areas, making the location of matchings difficult.

With the aim of minimizing both the homography elements and the position of features in the mosaic, Capel [3] proposed a method based on the tracking of features, which requires identifying the same feature in all the different views. Lets consider ${}^t x_i$ as the coordinates of a given i^{th} point defined in the coordinate system of image t and the projection of point ${}^m x_j$ in the mosaic, which is called the pre-image point and is usually projected in different views. All image points corresponding to the projection of the same pre-image point are called N-view matches. This approach proposes the following cost function to be minimized:

$$\varepsilon_1 = \sum_{j=1}^{M} \sum_{{}^t\boldsymbol{x}_i \in \eta_j} \| {}^t\boldsymbol{x}_i - {}^t\mathbf{H}_m \cdot {}^m\boldsymbol{x}_j \|_2 \tag{2.19}$$

where M is the total number of pre-image points, n_j is the set of N-view matches and ${}^t\mathbf{H}_m$ is mosaic-to-image homography. Knowing that the homographies and the pre-image points are unknowns, the total number of unknowns can be be obtained as $n = n_{DOF} \times n_{view} + 2 \times n_{points}$, where n_{DOF} are the number of Degree Of Freedoms (DOFs) of the homography, n_{views} is the total number of views and n_{points} is the total number of pre-image points. The fact of measuring the error term ε_1 in the image frame but being parameterized with points defined in the mosaic frame, allows us to avoid image an scaling bias that appears when measured in the mosaic

frame. As a drawback, the number of unknowns increases significantly as the size of the dataset grows, making it unsuitable for datasets with thousands of images.

BA is a technique to solve the problem of refining visual reconstruction to produce jointly optimal 3D structure and viewing parameter estimates (camera pose and/or calibration) [51, 52]. The solution is intended to be optimal with respect to both structure and camera variations. BA minimizes the reprojection error between the image correspondences. This minimization is defined as the sum of squares of a large number of nonlinear, real-valued functions, and is achieved using nonlinear least squares methods. Concerning image mosaicing, the target of BA is to find optimal camera motion parameters in order to compute absolute homographies [53]. Gracias et al. [54] presented an approach based on the minimization of the following cost function:

$$E = \sum_{i,j} \sum_{k=1}^{n} \left(\| {}^{i}\boldsymbol{x}_k - {}^{i}\mathbf{H}_j \cdot {}^{j}\boldsymbol{x}_k \|_2 + \| {}^{j}\boldsymbol{x}_k - {}^{i}\mathbf{H}_j^{-1} \cdot {}^{i}\boldsymbol{x}_k \|_2 \right) \tag{2.20}$$

where n is the number of matches between images i and j. The total number of unknowns is $6 \times (n_{views} - 1) + 2$. The method requires knowing the intrinsic camera parameters and has high computational requirements due to the use of nonlinear optimization algorithms.

For further details of advantages and disadvantages of the different GA methods the reader is addressed to [55].

2.5 Conclusions

Building photo-mosaics of underwater image surveys is a complex task that faces medium-specific challenges not present in terrestrial or aerial panorama generation. Due to the lack of natural light in deep waters, the UVs should integrate artificial lighting systems. The power of the light sources is limited, specially due to autonomy reasons, and typically does not allow uniform illumination of the whole area covered by a picture. The effects of this lack of power are accentuated by the underwater phenomenon of light attenuation, which leads to a noticeable non-uniform illumination in the images, and constrains the acquisition to a few meters from the seabed. The scattering phenomenon [56], due to suspended particles, is another phenomenon affecting underwater images, and is also affected by artificial lighting inasmuch as light rays collide with the suspended particles. As a result of these phenomena, underwater images suffer from poor and non-uniform illumination and frequently present bright spots due to backward scattering, and lack of sharpness due to forward scattering. The images affected by these problems make the detection of features and consequently the pair-wise registration difficult, giving rise at this point to the importance of the navigation data. The short distance between the camera and the seafloor favours the presence of parallax, which affects the 2D mosaicing approach due to the

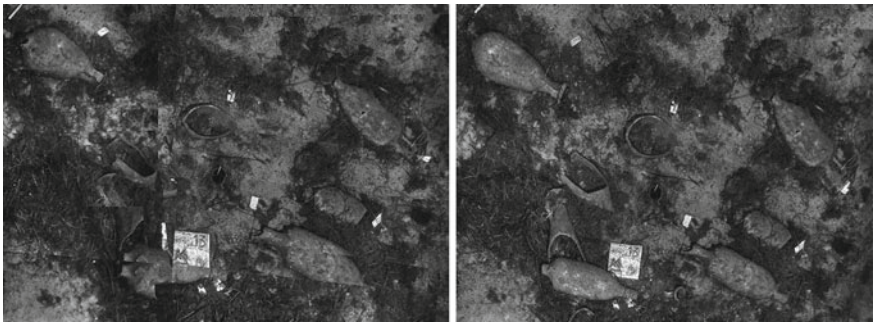

Fig. 2.6 Small area of a mosaic generated from an image set corresponding to a shipwreck in Pianosa (Italy). In the initial mosaic (*left*), before the application of a blending technique, the amphoras and white labels laying on the seafloor appear truncated. In the blended mosaic (*right*), the scene is easily understandable and the discontinuities have disappeared. Images courtesy of Pierre Drap (LSIS, CNRS)

violation of the planar assumption. The parallax effects, in addition to any moving elements in the scene, also impact image registration, and have consequences in the image rendering step. All these factors make the topology estimation and the global alignment [5–8], in conjunction with the use of the available navigation data, very relevant steps to achieve accurate photo-mosaics when dealing with thousands of images.

Given the heterogeneous appearance of the acquired images, and problems such as the planar assumption violation or the presence of moving objects, the use of image blending techniques is required. Apart from the visual appearance, blending techniques are also important for proper interpretation and scientific exploitation of seafloor imagery (e.g. [57, 58]). The structures, objects and areas of interest may cover a wide range of scales, from a few centimeters, i.e. microfauna or rocks, which would appear in individual images, to several hundreds of meters, i.e. topographic scarps or fractures, spanning several frames. The relevance of image blending arises at this point so that the photo-mosaics generated with these techniques present a consistent and uniform appearance (see Fig. 2.6). The blended photo-mosaic, where imaging artifacts have been minimized, allows us to analyze the features of interest, regardless of their size and imaging conditions.

Summarizing, the use of blending techniques in underwater 2D mosaicing is a crucial step when generating high-quality large-scale photomosaics. Preprocessing the images in order to correct non-uniform illumination and enhance their detail also becomes a key step in the mosaicing procedure. Enhanced images are best suited for the feature detection and correspondence finding steps. Providing the images with a good appearance is relevant not only from the aesthetical point of view but also from a functional one.

The problems of image blending and image quality enhancement are treated in the next chapters.

References

1. Szeliski, R., Shum, H.-Y.: Creating full view panoramic image mosaics and environment maps. In: Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), SIGGRAPH'97, pp. 251–258. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1997)
2. Sawhney, H., Hsu, S., Kumar, R.: Robust video mosaicing through topology inference and local to global alignment. In: Proceedings of the European Conference on Computer Vision, Freiburg, Germany, June 1998
3. Capel, D.: Image Mosaicing and Super-Resolution. Springer, Berlin (2004)
4. Gracias, N., Costeira, J.P., Santos-Victor, J.: Linear global mosaics for underwater surveying. In: Proceedings of the IFAC/EURON Symposium on Autonomous Vehicles (IAV), Lisbon, Portugal, July 2004
5. Ferrer, J., Elibol, A., Delaunoy, O., Gracias, N., Garcia, R.: Large-area photo-mosaics using global alignment and navigation data. In: Proceedings of the IEEE OCEANS Conference, pp. 1–9, Oct 2007
6. Elibol, A., Garcia, R., Delaunoy, O., Gracias, N.: A new global alignment method for feature based image mosaicing. In: Proceedings of the International Symposium on Advances in Visual Computing (ISVC), Part II, pp. 257–266. Springer, Berlin, Heidelberg (2008)
7. Elibol, A., Garcia, R., Gracias, N.: A new global alignment approach for underwater optical mapping. Ocean Eng. **38**(10), 1207–1219 (2011)
8. Elibol, A., Gracias, N., Garcia, R.: Augmented state-extended kalman filter combined framework for topology estimation in large-area underwater mapping. J. Field Robot. **27**(5), 656–674 (2010)
9. Elibol, A., Gracias, N., Garcia, R.: Fast topology estimation for image mosaicing using adaptive information thresholding. Robot. Auton. Syst. **61**(2), 125–136 (2013)
10. Elibol, A., Gracias, N., Garcia, R., Gleason, A., Gintert, B., Lirman, D.: Efficient autonomous image mosaicing with applications to coral reef monitoring. In: Proceedings of the Workshop on Robotics for Environmental Monitoring held at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 50–57, San Francisco, USA, Sept 2011
11. Brown, M., Hartley, R.I., Nister, D.: Minimal solutions for panoramic stitching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8, June 2007
12. Szeliski, R.: Image mosaicing for tele-reality applications. In: Proceedings of the IEEE Workshop on Applications of Computer Vision, pp. 44–53, Dec 1994
13. Dani, P., Chaudhuri, S.: Automated assembling of images: image montage preparation. Pattern Recogn. **28**(3), 431–445 (1995)
14. Horn, B., Shunck, B.: Determining optical flow. Artif. Intell. **17**, 185–203 (1981)
15. Shum, H-Y., Szeliski, R.: Construction and refinement of panoramic mosaics with global and local alignment. In: Procedings of the International Conference on Computer Vision (ICCV), p. 953. IEEE Computer Society, Washington, DC, USA, 1998
16. Reddy, B., Chatterji, B.: An FFT-based technique for translation, rotation, and scale-invariant image registration. IEEE Trans. Image Process. **5**(8), 1266–1271 (1996)
17. Wolberg, G., Zokai, S.: Robust image registration using log-polar transform. In: Proceedings of the International Conference on Image Processing (ICIP), vol. 1, pp. 493–496, 2000
18. Rzhanov, Y., Huff, L., Cutter, G.R.: Seafloor video mapping: modeling, algorithms, apparatus. In: Proceedings of the International Conference on Image Processing (ICIP), pp. 868–871, 2002
19. Rzhanov, Y., Mayer, L., Beaulieu, S., Shank, T., Soule, S.A., Fornari, D.J.: Deep-sea georeferenced video mosaics. In: Proceedings of the IEEE OCEANS Conference, pp. 1–6, Sept 2006
20. Lucas, B.D.: Generalized image matching by the method of differences. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA, July 1985 (AAI8601180)

21. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp. 674–679, 1981
22. Negahdaripour, S., Xu, X., Khamene, A., Awan, Z.: 3D motion and depth estimation from sea-floor images for mosaic-based positioning, station keeping and navigation of ROVs/AUVs and high resolution sea-floor mapping. In: Procedings of the IEEE/OES Workshop on AUV Navigation, Cambridge, MA, USA, Aug 1998
23. Negahdaripour, S.: Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. IEEE Trans. Pattern Anal Mach Intell. (PAMI) **20**(9), 961–970 (1998)
24. Madjidi, H., Negahdaripour, S.: On robustness and localization accuracy of optical flow computation from color imagery. In: Proceedings of the 3D Data Processing, Visualization and Transmission (3DPVT), 2nd International Symposium, pp. 317–324, 2004
25. Negahdaripour, S., Madjidi, H.: Robust optical flow estimation using underwater color images. In: Proceedings of MTS/IEEE OCEANS Conference, vol. 4, pp. 2309–2316, Sept 2003
26. Negahdaripour, S., Xu, X., Jin, L.: Direct estimation of motion from sea floor images for automatic station-keeping of submersible platforms. IEEE J. Oceanic Eng. **24**(3), 370–382 (1999)
27. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference, pp. 189–192, Manchester, UK, Aug 1988
28. Beaudet P. R.: Rotationally invariant image operators. In: Proceedings of the International Conference on Pattern Recognition (ICPR), pp. 579–583, Kyoto, Japan, Nov 1978
29. Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comput. Vision **30**, 79–116 (1998)
30. Lowe, D.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision (ICCV), vol. 2, p. 1150. IEEE Computer Society, Washington, DC, USA, 1999
31. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: European Conference on Computer Vision, pp. 404–417, (2006)
32. Haar, A.: Zur theorie der orthogonalen funktionensysteme. Math. Ann. **69**, 331–371 (1910). doi:10.1007/BF01456326
33. Schmid, C., Mohr, R., Bauckhage, C.: Comparing and evaluating interest points. In: Proceedings of the International Conference on Computer Vision (ICCV), pp. 230–235, 1998
34. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2003)
35. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S.: An Invitation to 3-D Vision: From Images to Geometric Models. Springer, Berlin (2003)
36. Negahdaripour, S., Prados, R., Garcia, R.: Planar homography: accuracy analysis and applications. In: IEEE International Conference on Image Processing (ICIP), vol. 1, 1089–1092 (2005)
37. Huang, J-F., Lai, S-H., Cheng, C-M.: Robust fundamental matrix estimation with accurate outlier detection. J. Inf. Sci. Eng. **23**(4), 1213–1225 (2007)
38. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)
39. Rousseeuw, P.J.: Least median of squares regression. J. Am. Stat. Assoc. **79**(388), 871–880 (1984)
40. Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley, New York (1987)
41. Li, X., Hu, Z.: Rejecting mismatches by correspondence function. Int. J. Comput. Vision **89**, 1–17 (2010)
42. Szeliski, R.: Image alignment and stitching: a tutorial. Found. Trends Comput. Graph. Vision **2**(1), 1–104 (2006)
43. Capel, D.P.: Image mosaicing and super-resolution. PhD thesis, University of Oxford, Oxford, UK, 2001

44. Marzotto, R., Fusiello, A., Murino, V.: High resolution video mosaicing with global alignment. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 692–698, June–July 2004
45. Davis, J.: Mosaics of scenes with moving objects. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Santa Barbara, CA, USA, June 1998
46. Kang, E., Cohen, I., Medioni, G.: A graph-based global registration for 2D mosaics. In: Proceedings of the International Conference on Pattern Recognition (ICPR), Barcelona, Spain, Sept 2000
47. Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.: A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicing the curved human retina. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **24**(3), 412–419 (2002)
48. Pizarro, O., Singh, H.: Toward large-area mosaicing for underwater scientific applications. IEEE J. Oceanic Eng. **28**(4), 651–672 (2003)
49. Shum, H.Y., Szeliski, R.: Construction of Panoramic Image Mosaics with Global and Local Alignment, pp. 227–268. Springer, New York (2001)
50. Morris, D.D.: Gauge freedoms and uncertainty modeling for 3D computer vision. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Mar 2001
51. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, ICCV'99, pp. 298–372. Springer, London, UK, 1999
52. Bouguet, J.Y.: Camera Calibration Toolbox. http://www.vision.caltech.edu/bouguetj/calib_doc. June 2008
53. McLauchlan, P.F., Jaenicke, A.: Image mosaicing using sequential bundle adjustment. Image Vision Comput. **20**, 751–759 (2002)
54. Gracias, N., Santos-Victor, J.: Underwater mosaicing and trajectory reconstruction using global alignment. In: Proceedings of the MTS/IEEE OCEANS Conference, pp. 2557–2563, Honolulu, Hawaii, U.S.A., Nov 2001
55. Elibol, A., Gracias, N., Garcia, R.: Efficient Topology Estimation for Large Scale Optical Mapping, Volume 82 of Springer Tracts in Advanced Robotics. Springer, Berlin (2012)
56. Shao, B., Jaffe, J.S., Chachisvilis, M., Esener, S.C.: Angular resolved light scattering for discriminating among marine picoplankton: modeling and experimental measurements. Opt. Expr. **14**(25), 12473–12484 (2006)
57. Barreyre, T., Escartin, J., Garcia, R., Cannat, M., Mittelstaedt, E., Prados, R.: Structure, temporal evolution, and heat flux estimates from a deep-sea hydrothermal field derived from seafloor image mosaics. Geochem. Geophys. Geosyst. **13**(4), 1–29 (2012)
58. Mittelstaedt, E., Escartín, J., Gracias, N., Olive, J.A., Barreyre, T., Davaille, A., Cannat, M., Garcia, R.: Quantifying diffuse and discrete venting at the tour eiffel vent site, lucky strike hydrothermal field. Geochem. Geophys. Geosyst. **13**, (2012)

http://www.springer.com/978-3-319-05557-2

Image Blending Techniques and their Application in Underwater Mosaicing

Prados, R.; Garcia, R.; Neumann, L.

2014, XI, 107 p. 49 illus., 20 illus. in color., Softcover
ISBN: 978-3-319-05557-2