Estimating the Best Reference Homography for Planar Mosaics From Videos

Fabio Bellavia and Carlo Colombo
Computational Vision Group, University of Florence, Florence, Italy
{*fabio.bellavia, carlo.colombo*}*@unifi.it*

Keywords: Hierarchical Mosaicing, Viewpoint Computation, Underwater Vision.

Abstract: This paper proposes a novel strategy to find the best reference homography in mosaics from video sequences. The reference homography globally minimizes the distortions induced on each image frame by the mosaic homography itself. This method is designed for planar mosaics on which a bad choice of the first reference image frame can lead to severe distortions after concatenating several successive homographies. This often happens in the case of underwater mosaics with non-flat seabed and no georeferential information available. Given a video sequence of an almost planar surface, sub-mosaics with low distortions of temporally close image frames are computed and successively merged according to a hierarchical clustering procedure. A robust and effective feature tracker using an approximated global position map between image frames allows us to build the mosaic also between locally close but not temporally consecutive frames. Sub-mosaics are successively merged by concatenating their relative homographies with another reference homography which minimizes the distortion on each frame of the fused image. Experimental results on challenging real underwater videos show the validity of the proposed method.

1 INTRODUCTION

Over the last decade, image mosaicing has received a considerable attention for its wide range of practical applications (Brown and Lowe, 2007). However, despite the recent progress in the field, obtaining good mosaics still remains a challenging and not fully solved task. This is mostly due to the assumption of input data with sufficiently scene distance or image acquired by camera rotations only (Hartley and Zisserman, 2003). These requisites cannot be effectively met in practice, causing image misalignments and ghosting artefacts. In order to avoid or alleviate these issues, several image stitching and post processing techniques have been developed in recent years (Lin et al., 2011; Zaragoza et al., 2014; Zhang and Liu, 2014).

Furthermore, video mosaicing has become quite popular in underwater vision (Pizarro and Singh, 2003; Bellavia et al., 2007; Prados et al., 2012), due to its applications to in situ exploration and autonomous navigation. While common panoramic mosaics assume spherical or cylindrical models, in the case of underwater environments planar surface models are assumed. A problem commonly ignored, yet often present in practice, is the selection of a reference homography reprojection frame on which to attach the various mosaic images. The most common and trivial choice is to use the first frame image or, often supported by geo-referential camera positions, a user predefined one. A bad choice for the reference frame can lead to very distorted mosaics (see Fig. 1 (left)). This can also result after some sequential frame concatenations into degenerate and incorrect configurations. This problem is accentuated in underwater videos due to the unstable trajectory of the acquisition vehicle with roll and pitch shakes and the non-flat truly nature of the seabed in most cases. To the best of our knowledge, methods to solve this issue exist in the literature solely for the case of planar mosaics from rotation-only frames (Capel, 2001).

This paper presents in Sect. 2 a novel multi-step method to estimate the mosaic reference homography in the case of planar mosaics from video sequences. An output example is shown in Fig. 1 (right). This general approach is sided with a robust full mosaic pipeline particularly designed for underwater environments, where the non-planar nature of the scenes make it difficult to match and track the keypoints required to compute inter-frame homographies. Some selected results on real underwater video sequences from different oceanographic campaigns are given in

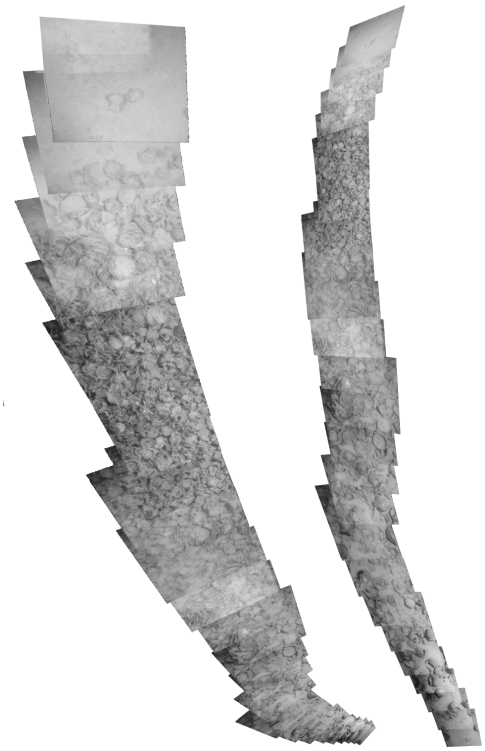


Figure 1: A distorted mosaic due to a wrong reference homography selection (left). Note that in the former case more than half of the frames accumulate on the bottom with large scale variations and distortions. This is not present in the output given by the proposed method (right). In both cases no post-processing color correction or blending have been applied.

Sect. 3, showing the good performance of the overall algorithm. Finally, conclusions and future work are draw out in Sect. 4.

2 METHOD DESCRIPTION

2.1 Overview

The overall pipeline of the method is schematically described in Fig. 2. The approach starts by dividing the video sequence into successive piecewise planar sub-mosaics with low geometric distortions with respect to the original image frames, as described in Sect. 2.2.

Sequential concatenation of temporally consecutive frames is sufficient to produce the piecewise planar sub-mosaics. However, a global strategy is required to match frames locally close but temporally non-consecutive. This is required in order to reduce the propagation of homography estimation errors due to the non-planar real nature of the scene.

An approximated 2D map of the video frame positions is then computed, by considering the average translation between successive frames, which is updated when two close frames are discovered. The best paths on the map are used to robustly track features in spatially close but non-consecutive frames and compute the homography between them. Details of this step are given in Sect. 2.3. Finally, as

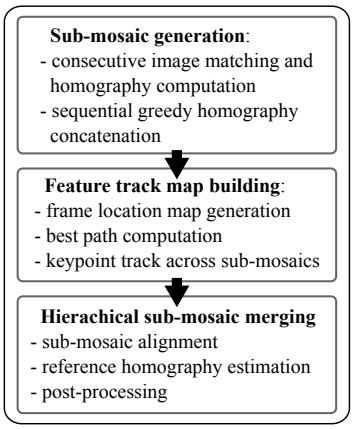


Figure 2: A schematic view of the mosaic pipeline.

described in Sect. 2.4, sub-mosaics are merged using the tracked keypoints one at a time according to a hierarchical clustering, preferring sub-mosaics with high overlap. When two sub-mosaics are merged, the best reference homography between them is found in order to minimize the distortion of the image frames. This is achieved by exploring the search space of the "average" homographies between the two sub-mosaic images. Common blending and photometric post-processing steps are eventually applied to refine the results (Uyttendaele et al., 2001; Kwatra et al., 2003; Brown and Lowe, 2007; Prados et al., 2012).

2.2 Sub-mosaic Generation

Given the video sequence of $n+1$ consecutive frames I_0, I_1, ..., I_n, the first step is to extract the image keypoints to obtain the matches between overlapping frames and their associated homography. For this purpose, the HarrisZ detector (Bellavia et al., 2011) is used, providing robust and reliable corner features.

In the following step, homographies between consecutive image frame pairs (I_k, I_{k+1}) are computed. In particular, feature matches are computed using the sGOr matching selection strategy (Bellavia et al., 2014) under the further assumption that the optical flow between successive frames is bounded. That is, given a generic keypoint pair $(\mathbf{x}_k, \mathbf{x}_{k+1})$ with $\mathbf{x}_i \in I_i$, it must hold

$$\| \mathbf{x}_k - \mathbf{x}_{k+1} \| < \varepsilon_r \quad (1)$$

where the threshold ε_r is set to a third of the diagonal of the rectangular video frame. The homography $\mathrm{H}_{k,k+1} \in \mathbb{R}^{3\times3}$ that maps points from I_k to I_{k+1} is computed on the obtained matches using the noRANSAC method (Bellavia and Tegolo, 2011), a robust extension of the RANSAC 7-point algorithm (Hartley and Zisserman, 2003) using normalized errors.

Finally, a piecewise planar sub-mosaic $S_{i,j}$ going from frame i to frame j (see Fig. 3) is constructed from its keyframe sequence $K_{i,j}$

$$K_{i,j} = I_{k_1}, I_{k_2}, \ldots, I_{k_m} \tag{2}$$

with $i = k_1 < k_2 < \ldots < k_m = j$, as explained hereafter. The first image frame and keyframe I_i of the subsequence is used as reference, so that the homography map $\mathrm{H}_w^{i,j}$ relating the sub-mosaic to any keyframe I_{k_w}, is simply obtained by concatenating the sequential homographies from $\mathbf{H}_{i,i+1}$ to $\mathbf{H}_{k_{w-1},k_w}$:

$$\mathrm{H}_w^{i,j} = \mathbf{H}_{k_{w-1},k_w} \mathbf{H}_{k_{w-1},k_{w-2}} \ldots \mathbf{H}_{i,i+1} \tag{3}$$

The sub-mosaic is then built according to a greedy strategy by sequentially looking for next keyframe in $I_{i+1}, I_{i+2}, \ldots$. The frame I_j is added as keyframe if the overlap between the current sub-mosaic and the projection of frame I_j onto the sub-mosaic is less than a threshold. The sub-mosaic generation process stops when the projection of frame I_{j+1} is too distorted to be included in K_{ij}. In this case I_j is added as the final keyframe with the sub-mosaic $S_{i,j}$ as output and starting from I_{j+1} a new sub-mosaic is grown.

The distortion criterion is defined as follows. Assuming rectangular video frames, their projections into quadrilaterals in the mosaic are considered distorted if one of the following conditions is met: (1) the ratio between the original and projected frames is outside a user-defined range, (2) the area of the bounding box including all common keypoints in the projected mosaic is below a threshold, (3) the ratio between the minimum and maximum semi-axis lengths of the projected quadrilateral exceeds a given value.

2.3 Feature Track Map Building

Sequential sub-mosaics can be merged using only the homography between boundary keyframes, for example using the homography $\mathrm{H}_{j,j+1}$ between the sub-mosaics $S_{i,j}$ and $S_{j+1,w}$, with $i < j < w$. However, this solution is not robust, due to inevitable inaccuracies in the homography which may propagate across the sequence, especially when coming back to an already seen location of the mosaic. According to this observation, adding more robust matches and recognizing loop-closures (Konolige and Agrawal, 2008) may improve the result thanks to a suitable keypoint tracking strategy, implemented as follows.

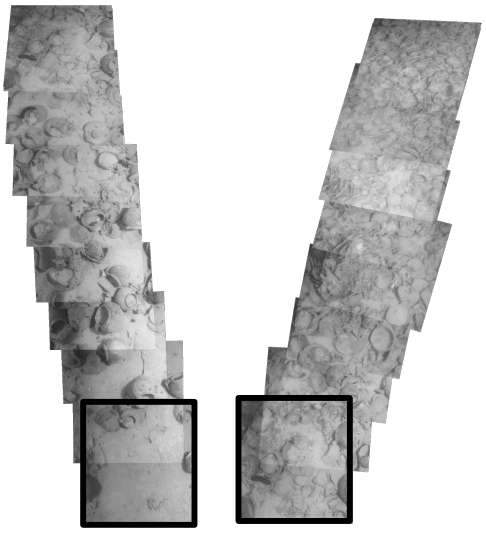


Figure 3: From left to right, two consecutive sub-mosaics from the mosaic of Fig. 1(right). The reference image frames I_{k_1} on which successive frames are concatenated (see text) are highlighted by boxes. No color correction or blending have been applied.

Given a threshold $\varepsilon_t = 1.5\,px$ the enriched set of matches $M_{k-1,k}$ is computed, considering all the keypoint matching pairs and not only the filtered subset given in input to the RANSAC (see Sect. 2.2). With an abuse of notation for indicating homogeneous normalized coordinates, we have

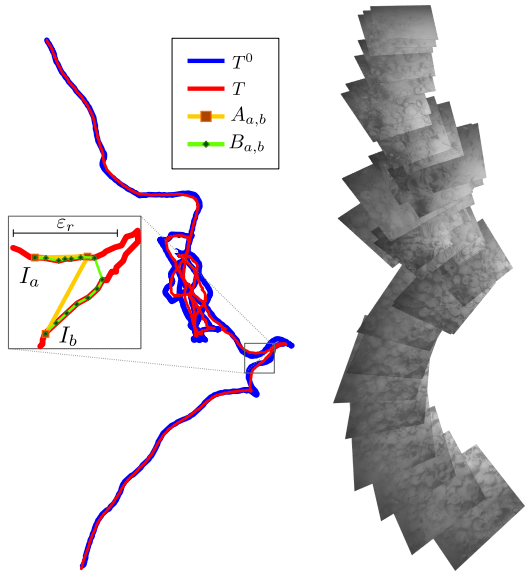
$$\begin{aligned} M_{k-1,k} = \{(\mathbf{x}_{k-1}, \mathbf{x}_k) :\ & \mathbf{x}_{k-1} \in I_{k-1}, \\ & \mathbf{x}_k \in I_k, \| \mathbf{x}_k - \mathrm{H}_{k-1,k}\mathbf{x}_{k-1} \| < \varepsilon_t\} \end{aligned} \tag{4}$$

where matches are selected according to the nearest neighbour approach (Lowe, 2004) on the homography reprojection error. This allows to track more keypoints across non-consecutive frames since longer tracks can be built.

In order to handle loop-closure, a robust global keypoint tracking map is also implemented, see Fig. 4. Under the assumption of an almost planar surface, an initial frame location map $T^0 : \{I_k\} \to \mathbb{R}^2$ for each frame I_k in the video sequence is computed using the average displacement between corresponding matches

$$T^0(I_k) = T^0(I_{k-1}) + \frac{1}{N} \sum_{M_{k-1,k}} (\mathbf{x}_k - \mathbf{x}_{k-1}) \tag{5}$$

The process is started from $T^0(I_0) = \mathbf{0}$, where the summation is on the match pairs $(\mathbf{x}_{k-1}, \mathbf{x}_k) \in M_{k-1,k}$ and $N = |M_{k-1,k}|$ (blue line on Fig. 4(a)). Further iterations i are introduced to progressively update the map T^i, as in the case of the Self-Organizing Map learning method (Haykin, 1998). In detail, given the



(a) (b)

Figure 4: The frame location map (a) corresponding to the mosaic in (b). The first and last iterations T^0 and T are plotted as blue and red lines, respectively, and no blending and color correction have been applied to the mosaic. The zoomed region in (a) shows the minimum path $A_{a,b}$ and the best path $B_{a,b}$ between two frames I_a and I_b. The maximum allowable edge distance is ε_r. (Best viewed in color.)

set of frames I_w in a radius ε_r from I_k

$$J_k^i = \{I_w : \parallel T^i(I_k) - T^i(I_w) \parallel \leq \varepsilon_r\} \tag{6}$$

where ε_r is set as for Eq. 1, T^i is updated to T^{i+1} as follows

$$T^{i+1}(I_k) = T^i(I_k) + \frac{d^i}{L} \sum_{J_k^i} \frac{g_{k,w}}{N} \sum_{M_{k,w}} (\mathbf{x}_k - \mathbf{x}_w) \tag{7}$$

where the outer summation is on the frames $I_w \in J_k^i$, $L = |J_k^i|$, the inner summation as for Eq. 5 and $M_{k,w}$ is defined analogously to $M_{k-1,k}$ (see Eq. 4). The value $0 < d < 1$ is an exponential decay going with the iteration i to limit the iterations and $g_{k,w}$ is a Gaussian weight on the distance $\parallel T^i(I_k) - T^i(I_w) \parallel$. When $d^i \simeq 0$ no further updates are needed, and the final map T is obtained (from blue to red lines as iterations proceed in Fig. 4(a)). Note that the planar displacement approximation of the map cannot handle scale variations, so that there is no perfect correspondence between Fig. 4(a) and Fig. 4(b). Nevertheless, this does not interfere with the final keypoint tracking. The graph $G = (V, E)$ is associated to the map T, where the set of nodes is $V = \{I_k\}$, i.e. the frames of the sequence, and an edge $E_{k,z}$ between vertexes I_k and I_z with an associated weight $\parallel T(I_k) - T(I_z) \parallel$ exists only if $\exists\ i$ such that $I_z \in J_k^i$.

A keypoint $\mathbf{x}_w$ on frame I_w is tracked to the keypoint $\mathbf{x}_z$ in the frame I_z following the chain of matches $(\mathbf{x}_w, \mathbf{x}_{k_1}), (\mathbf{x}_{k_1}, \mathbf{x}_{k_2}), \ldots, (\mathbf{x}_{k_m}, \mathbf{x}_z)$ according to the best path B between the frame I_w and I_z on the graph G. For any match pair in the chain it must hold that $(\mathbf{x}_{k_i}, \mathbf{x}_{k_j}) \in M_{k_i,k_j}$. The best path $E_{w,k_1}, E_{k_1,k_2}, \ldots E_{k_m,z}$ must concatenates short weight edges, since matches from distant frames are more unstable and error prone. Furthermore, the best path length must be short because very long paths accumulate errors.

We computed the best path B from a frame I_k to a frame I_w by an extended version of the Floyd-Warshall all-shortest-paths algorithm (Cormen et al., 2009). Hereafter, both a path and its length with be referred with the same symbol. In the first step we compute the minimum path length A between all the frames in G, using the standard algorithm updating rule at each iteration $0 \leq i \leq |V|$, i.e.

$$A_{a,b}^i = \begin{cases} A_{a,c}^{i-1} + A_{c,b}^{i-1} & \text{if } A_{a,c}^{i-1} + A_{c,b}^{i-1} < A_{a,b}^{i-1} \\ A_{a,b}^{i-1} & \text{otherwise} \end{cases} \tag{8}$$

where $A_{a,b}^i$ is the shortest path length at iteration i between frames I_a and I_b, and $A_{a,b}^0 = E_{a,b}$. Denoting the shortest path length at the last iteration by $A_{a,b} = A_{a,b}^{|V|}$, this is used as follows to bound the length of the best path $B_{a,b}^i$ using a factor $f = 2$. Defining the auxiliary values $C_{a,b}^i$ representing the maximum edge weight in the path between I_a and I_b, the update rule for the last step are

$$B_{a,b}^i = \begin{cases} B_{a,c}^{i-1} + B_{c,b}^{i-1} & \text{if } B_{a,c}^{i-1} + B_{c,b}^{i-1} < f A_{a,b} \text{ and} \\ & \max(C_{a,c}^{i-1}, C_{c,b}^{i-1}) < C_{a,b}^{i-1} \\ B_{a,b}^{i-1} & \text{otherwise} \end{cases} \tag{9}$$

and

$$C_{a,b}^i = \begin{cases} \max(C_{a,c}^{i-1}, C_{c,b}^{i-1}) & \text{if } B_{a,c}^{i-1} + B_{c,b}^{i-1} < f A_{a,b} \text{ and} \\ & \max(C_{a,c}^{i-1}, C_{c,b}^{i-1}) < C_{a,b}^{i-1} \\ C_{a,b}^{i-1} & \text{otherwise} \end{cases} \tag{10}$$

initialized as for the previous step as $B_{a,b}^0 = C_{a,b}^0 = E_{a,b}$, the edge in the best path are updated accordingly. An example of the best path between two frames is given in Fig. 4(a).

Denoting the best path at the last iteration with $B_{a,b} = B_{a,b}^{|V|}$, referring to Sect. 2.2, the robust matches between two sub-mosaics $S_{i,j}$ and $S_{w,z}$ are obtained by trying to track on the best path $B_{s,t}$ each of the keypoints $\mathbf{x}_{i,j}$ of any keyframe $I_s \in K_{i,j}$ to a keypoint $\mathbf{x}_{z,w}$ of any keyframe $I_t \in K_{w,z}$. The obtained robust matches $M_{i,j}^{z,w} = \{(\mathbf{x}_{i,j}, \mathbf{x}_{z,w})\}$ across the sub-mosaics are used to compute the homography $\mathrm{H}_{i,j}^{z,w}$ and finally merge the sub-mosaics as explained in the next section.

Note that for computing the RANSAC given inlier set $M_{i,j}^{z,w}$ an error threshold 5 times greater than that used in the other RANSAC inlier set $M_{k,w}$ is employed. This is done to partially relax the planar surface assumption, which is unreal in concrete cases, and allow larger surface deformations.

2.4 Hierarchical Sub-mosaic Merging

Sub-mosaics are merged incrementally according to their overlap, following a hierarchical clustering algorithm. In particular, defining by $S^0 = \{S_{i,j}\}$ the initial cluster partition, at each step $0 \leq i < |S^0|$, we try to merge all the possible pairs $(S_{i,j}, S_{w,z})$, with $S_{i,j}, S_{w,z} \in S^i$, using the robust homography $\mathrm{H}_{i,j}^{z,w}$ computed as in Sect. 2.3, to which the reference homography $\tilde{\mathrm{H}}_{i,j}^{z,w}$ described next is applied. Denoting by $S_{i,j}^*$ the area of the sub-mosaic $S_{i,j}$ according to the reference homography $\tilde{\mathrm{H}}_{i,j}^{z,w}$ and in similar way for $S_{w,z}^*$, only the sub-mosaic pair with the minimal overlap error $R_{i,j}^{w,z}$

$$R_{i,j}^{w,z} = 1 - \frac{S_{i,j}^* \cap S_{w,z}^*}{S_{i,j}^* \cup S_{w,z}^*} \tag{11}$$

is effectively merged in the next cluster partition S^{i+1}, until no more merges can be done. The reference homography $\tilde{\mathrm{H}}_{i,j}^{z,w}$ between two sub-mosaics is computed by trying to minimize the distortion of all the frames in $K_{i,j}$ and $K_{w,z}$ in the merged mosaic. In particular, considering the pair $(S_{i,j}, S_{w,z})$, we define the auxiliary merged mosaics S^1 and S^2 obtained respectively using the first frames $I_i \in K_{i,j}$ and $I_w \in K_{w,z}$ as references (see Fig. 5 (top and middle rows)). In the first case the homography $\mathrm{H}_{i,j}^{z,w}$ is used to map points of $S_{z,w}$ onto the reference frame I_i, while in the other case the inverse $(\mathrm{H}_{i,j}^{z,w})^{-1}$ maps points of $S_{i,j}$ onto I_w. Both S_1 and S_2 are aligned according to the robust matches $(\mathbf{x}_{i,j}, \mathbf{x}_{z,w}) \in M_{i,j}^{z,w}$ (see Sect. 2.3). In particular, S_1 and S_2 are translated so that the new origins are in their centroids and rotated according to the rotation R of the best similarity transform obtained by the least-square solution (Zhang and Liu, 2014), i.e.

$$\tilde{\mathbf{x}}_1 = \mathbf{x}_1 - \sum_{M_{i,j}^{w,z}} \mathbf{x}_{i,j} \tag{12}$$

$$\tilde{\mathbf{x}}_2 = \mathrm{C}\left(\mathbf{x}_2 - \sum_{M_{i,j}^{w,z}} \mathbf{x}_{w,z}\right) \tag{13}$$

$$\mathrm{C} = \frac{1}{a^2+b^2}\begin{bmatrix} a & b \\ b & a \end{bmatrix} \tag{14}$$

where $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ are the aligned new point coordinates for points $\mathbf{x}_1 \in S^1$ and $\mathbf{x}_2 \in S^2$ respectively. The values

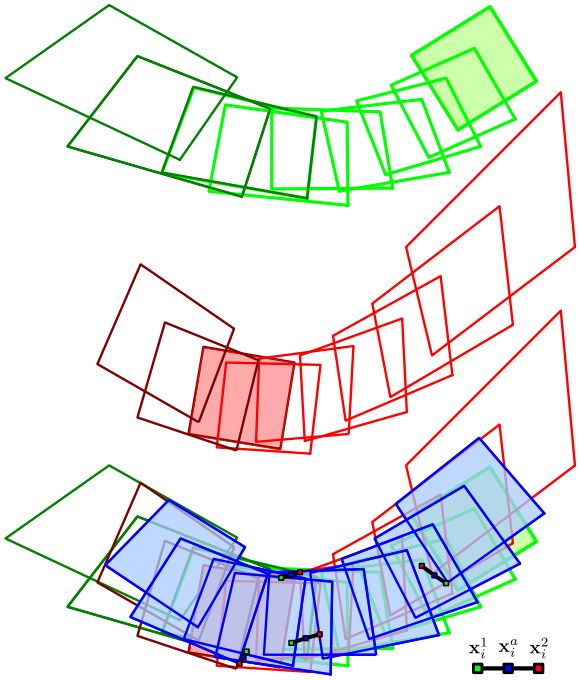


Figure 5: The auxiliary mosaics S^1 (green frames, top row) and S^2 (red frames, middle row) obtained using respectively as reference frames $I_i \in K_{i,j}$ (green filled frame) and $I_w \in K_{w,z}$ (red filled frame) from $S_{i,j}$ (lighter color) and $S_{w,z}$ (darker color) are aligned to find the best reference homography (bottom row). Four pairs of corresponding random sampled points $(\mathbf{x}_i^1, \mathbf{x}_i^2)$ are used to generate the midpoints $\mathbf{x}_i^a$ on which computing the reference homography $\tilde{\mathrm{H}}_{i,j}^{z,w}$ (bottom row, blue frames). The error P is given by accounting for the distortion of each resulting (blue) frame. (Best viewed in color).

a and b are computed on lest-squares according to the similarity transform

$$\mathbf{x}_{i,j} = \begin{bmatrix} a & b & c_1 \\ b & a & c_2 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{w,z} \tag{15}$$

with an abuse of notation for homogeneous coordinates and $a, b, c_1, c_2 \in \mathbb{R}$ (see Fig. 5 (bottom row, green and red frames) for an example).

The reference homography $\tilde{\mathrm{H}}_{i,j}^{z,w}$ is chosen by RANSAC, looking for the "average" homography which minimizes the error P (see Fig. 5 (bottom row, blue frames)). Given four corresponding randomly sampled non-collinear points $\mathbf{x}_1^i$ and $\mathbf{x}_2^i$ on S^1 and S^2 respectively, $1 \leq i \leq 4$, and the associated midpoints $\mathbf{x}_a^i = \frac{1}{2}(\mathbf{x}_1^i + \mathbf{x}_2^i)$, the "average" homography is given by the homography H_a^1 mapping $\mathbf{x}_i^1$ to $\mathbf{x}_i^a$. Under the assumption of $n \times m$ rectangular frames, for each frame $I_k \in K_{i,j} \cup K_{w,z}$, we compute a distortion error P on the quadrilateral $\tilde{I}_k$, obtained by applying

H_a^1 to the image of I_k on S_1

$$P = \max_{I_k \in K_{i,j} \cup K_{w,z}} (P^O + P^N + P^A + P^\alpha) \qquad (16)$$

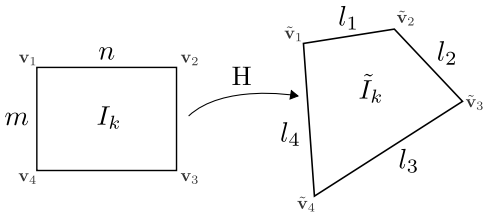


Figure 6: Configuration for computing the error P when projecting the image frame I_k to $\tilde{I}_k$ through the mosaic homography H (see text). Corresponding vertex pairs are $(\mathbf{v}_i, \tilde{\mathbf{v}}_i)$ for $1 \leq i \leq 4$.

Considering the four quadrilateral sides of length l_i of $\tilde{I}_k$ in consecutive order, under the assumption that $n > m$ with l_1, l_3 corresponding to the side of length n on I_k and in similar way l_2, l_4 to m (see Fig. 6), we define the different errors composing P. In particular, P^O measures the error on the ratio between the opposite sides of $\tilde{I}_k$:

$$P^O = 2 - \frac{1}{2}\left(\frac{\min(l_1,l_3)}{\max(l_1,l_3)} + \frac{\min(l_2,l_4)}{\max(l_2,l_4)}\right) \qquad (17)$$

while P^N measures the error on the ratio between consecutive sides of $\tilde{I}_k$:

$$P^N = 1 - \frac{\min(r, \frac{m}{n})}{\max(r, \frac{m}{n})} \qquad (18)$$

where

$$r = \min\left(\frac{l_1}{l_2}, \frac{l_2}{l_3}, \frac{l_3}{l_4}, \frac{l_4}{l_1}\right) \qquad (19)$$

The error P^A gives the error ratio between the frame area nm and the area of its image $A_{\tilde{I}_k}$

$$P^A = 1 - \frac{\min(A_{\tilde{I}_k}, nm)}{\max(A_{\tilde{I}_k}, nm)} \qquad (20)$$

while P^α measures the angular error

$$P^\alpha = (\max(\cos_{12}, \cos_{23}, \cos_{34}, \cos_{41}))^5 \qquad (21)$$

$\cos_{ab}$ being the absolute value of the cosine between the two sides l_a and l_b. An example of resulting mosaic obtained by merging sub-mosaics according to the best reference homography $\tilde{\mathrm{H}}_{i,j}^{z,w}$ of Fig. 5 is shown in Fig. 7. Finally, as post-processing step on the final merged mosaic, multi-band blending (Brown and Lowe, 2007) and color correction using an extension of the Reinhard's method (Reinhard et al., 2001) are applied.

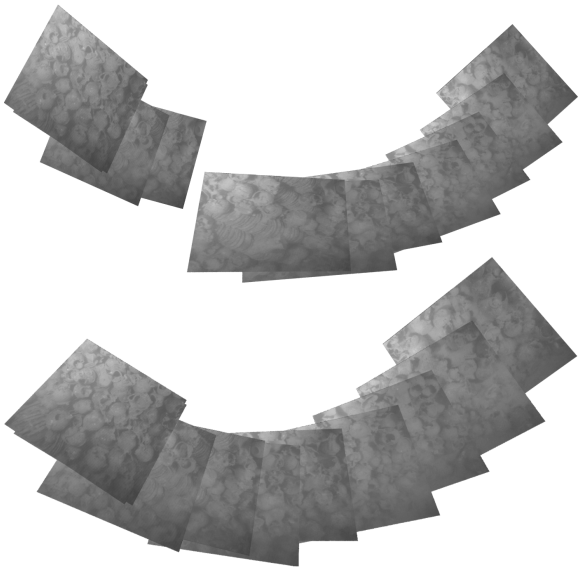


Figure 7: The initial sub-mosaics $S_{i,j}, S_{w,z}$ (top) and the resulting mosaic (bottom) according to the computed reference homography $\tilde{\mathrm{H}}_{i,j}^{z,w}$ of Fig. 5. In both cases no post-processing color correction or blending have been applied.

(a) (b)

Figure 8: Snapshots of the two test video sequences. In the case the video sequence (a) the shaded area containing a fixed robot arm has been cropped. (Best viewed in color.)

3 EXPERIMENTAL RESULTS

We tested the proposed pipeline on two underwater video sequences which together with the software code are freely available online[1]. Snapshots of video frames are shown in Fig. 8, in the case of the first video, the image area of the frame including the robot arm was cropped. To remove redundant data, the original 25 *fps* videos were downsampled to 5 *fps*. The corresponding output mosaics are shown in Fig. 9. As it can be noted, the resulting mosaics are good, with no evident misalignment glitches or strong frame deformations. Underwater scenes are very challenging, due to their high intensity changes and repeated patterns, which make the feature tracking difficult, so

[1] http://www.math.unipa.it/fbellavia/dl/mosaic.zip

(a) (b)

Figure 9: Output mosaics corresponding respectively to the video sequences of Fig. 8(a)-(b). (Best viewed in color).

that the quality of the results strengthen the validity of the feature track map computation of Sect. 2.3.

As it can be seen from the output mosaics, the proposed method is effective in choosing the reference mosaic homography (see Sect. 2.4). Note that, using the first image frame as reference, in the case of Fig. 8(a) would lead to very distorted images, as can be observed even from corresponding initial merged sub-mosaics (see Fig. 5 (top and middle rows, green and red frames)). Indeed, the proposed method allowed us to get in a completely automatic way as good-looking mosaics as those obtained with strong expert user intervention.

4 CONCLUSIONS

This paper proposes a new approach to compute the best reference mosaic homography that minimizes the frame distortions in the case of planar mosaics. For this purpose, a full hierarchical mosaicing pipeline was designed, with particular attention to underwater mosaicing applications that, due to the scene complexity, require robust feature tracking schemes as the one proposed in this paper. Experimental results show the validity of our method, yielding to high quality unsupervised mosaics.

Future work will include more evaluation tests as well incorporating in the pipeline new stitching algorithms (Zaragoza et al., 2014; Zhang and Liu, 2014) to replace the standard 7-point homography computation, with the aim to improve results in the case of strong 3D content.

ACKNOWLEDGEMENT

Thanks to Pamela Gambogi of the "Soprintendenza per i Beni Archeologici della Toscana", Italian Ministry of Culture, for providing the input video sequences.

This work has been carried out during the ARROWS project, supported by the European Commission under the Environment Theme of the "7th Framework Programme for Research and Technological Development".

REFERENCES

Bellavia, F., Gagliano, G., Tegolo, D., and Valenti, C. (2007). Global archaeological mosaicing for underwater scenes. *WSEAS Transactions on Signal Processing*, 2(7):997–1003.

Bellavia, F. and Tegolo, D. (2011). noRANSAC for fundamental matrix estimation. In *British Machine Vision Conference*, pages 1–11.

Bellavia, F., Tegolo, D., and Valenti, C. (2011). Improving Harris corner selection strategy. *IET Computer Vision*, 5(2):87–96.

Bellavia, F., Tegolo, D., and Valenti, C. (2014). Keypoint descriptor matching with context-based orientation estimation. *Image and Vision Computing*, 32(9):559 – 567.

Brown, M. and Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73.

Capel, D. P. (2001). *Image Mosaicing and Super-resolution*. PhD thesis, University of Oxford.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to Algorithms (3. ed.)*. MIT Press.

Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition.

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2 edition.

Konolige, K. and Agrawal, M. (2008). Frameslam: From bundle adjustment to real-time visual mapping. *IEEE Transactions on Robotics*, 24(5):1066–1077.

Kwatra, V., Schödl, A., Essa, I. A., Turk, G., and Bobick, A. F. (2003). Graphcut textures: image and video synthesis using graph cuts. 22(3):277–286.

Lin, W. Y., Liu, S., Matsushita, Y., Ng, T. T., and Cheong, L. F. (2011). Smoothly varying affine stitching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 345–352.

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

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

Prados, R., Garcia, R., Gracias, N., Escartin, J., and Neumann, L. (2012). A novel blending technique for underwater giga-mosaicing. *IEEE Journal of Oceanic Engineering*, 37(4):626–644.

Reinhard, E., Ashikhmin, M., Gooch, B., and Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41.

Uyttendaele, M., Eden, A., and Szeliski, R. (2001). Eliminating ghosting and exposure artifacts in image mosaics. In *Computer Vision and Pattern Recognition*, pages 509–516.

Zaragoza, J., Chin, T. J., Tran, Q. H., Brown, M., and Suter, D. (2014). As-Projective-As-Possible image stitching with moving DLT. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 36(7):1285–1298.

Zhang, F. and Liu, F. (2014). Parallax-tolerant image stitching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3269.