Estimating the Precision of Under-Water Video-Mosaics

E. Michaelsen, N. Scherer-Negenborn
Fraunhofer-IOSB, Gutleuthausstrasse 1, 76275 Ettlingen, Germany
eckart.michaelsen@iosb.fraunhofer.de

Abstract

Under-water video-mosaics are an important tool e.g. for inspection of man-made infrastructure. Cameras may drift in rotation and distance to the surface while the mosaic will often be very much larger than a single frame resulting in long chains of planar homographies. This contribution addresses the problems arising from dead-reckoning drift in such chains. Local patches are rectified using homography decomposition. Experiments with a portal allowing underwater image sequences with mechanical groundtruth are performed. Thus the deviation of the mosaic from a true orthographic map may be estimated.

1. Introduction

Visual inspection of man-made underwater infrastructure from unmanned platforms is of increasing interest. Another application can be mapping of the seafloor. Due to poor visibility conditions the distance between camera and observed surface must be kept short. Therefore, wide-angle lens are used. Large image mosaics are stitched with the camera moving back and forth along the surface. It is of crucial importance to estimate the drift error occurring in such mosaics.

2. Related work

Optimal estimation of homogenous entities, such as planar homography matrices, with more entries than degrees of freedom from image to image correspondences including sound uncertainty propagation is treated in [3]. Robust estimation using RANSAC is state-of-the-art. There are, however, alternatives such as GoodSAC [8], SURSAC [11], or iterative reweighting [13].

Practical state-of-the-art underwater mosaicing is e.g. reported in [5] where the task is environmental or archaeological monitoring. The computer vision and photogrammetric part of the work is more or less standard. But, remarkably, in one of the examples a 100m$^2$ barge is submerged with rails supporting the vehicle such that rotations are impossible. Accordingly the image transformations are often restricted to similarity mappings. Such efforts and restrictions emphasize the difficulties encountered with large mosaics underwater. Also Gracias and Santos-Victor report on mosaicing and trajectory reconstruction from videos taken of a seafloor scene [4]. They again emphasize the difficult lighting and visibility conditions in underwater computer vision. From the constant video stream of a few minutes a set of about hundred pictures is selected with enough overlap and sufficient difference. From these the mosaic and trajectory is estimated assuming a flat scene and using homography decomposition [2]. Underwater mosaic stitching with particular emphasis on the lens distortion induced drift is treated in [1].

3. A method for large mosaics

Stitching mosaics by use of planar homographies brings the risk of producing strong projective distortions including the mapping of pixels into and from infinity and rupture of the image topology. In particular if the mosaic is becoming orders of magnitude larger than the size of the first image with the camera moving further away drift will occur in the projective entries of the accumulated mapping – until division by zero or sign-flip leads to absurd results. Therefore, the method used here uses consecutive homography estimation only for the construction of limited local mosaics. These partial mosaics are then combined using homography decomposition into one large global mosaic avoiding projective distortions. This approach follows [9]. Diligent lens distortion
estimation is crucial here. The automatic procedure used for that is based on regular hole grids [10].

3.1. Chaining local planar homographies

Let \( x' = h(x) \) map from one image into the next, where \( x \) and \( x' \) respectively are the point coordinates. Using homogenous coordinates planar homographies turn out to be linear: \( x' = Hx \) with \( H \) being a 3x3 matrix. Such mapping can be estimated from a set of at least four correspondences of interest points using the linear matrix equation provided that 1) a coordinate system is used that balances the entries into the equation system such that signs have equal frequencies and absolutes are close to unity, and 2) \( H \) is not too far away from the unity matrix (in particular the “projective” entries \( H_{31} \) and \( H_{32} \) should be small).

In a video sequence 1) can be forced and 2) can be assumed. Briefly the procedure is as follows: In each frame of the video a set of interest points \( \{p_i; i=1, ..., k\} \) is extracted using the squared averaged gradient operator. These are tracked in the previous frame using optical flow including image pyramids from open CV base. Among these a consensus set is selected and simultaneously an optimal homography using linear estimation and RANSAC on the correspondences of the \( p_i \) in coordinates transformed accordingly. Thus a long multiplicative chain of image-to-image homographies is estimated.

In the UML activity diagram displayed in Fig. 1 this process is given as upmost loop. If failure occurs – e.g. lack of structure or very repetitive patterns – a default homography is chosen (here unity). Such chain cannot be very long because of drift occurring in the projective entries. For the experiments chaining of about twenty frames was regarded as “long enough” criterion.

3.2. Rectifying and combining partial mosaics by homography decomposition

It is known since [2] that given the internal calibration matrix \( K \) the mapping can be re-stated as \( H' = KH^{-1} \) and decomposed as \( H' = \lambda R + tn^T \) where \( \lambda \) is a scalar factor, \( R \) is the rotation matrix of the camera between the images, \( t \) is a translation vector, and \( n \) is the surface normal of the planar scene. If the mosaic patch was successful the component \( n_3 \) (towards to the camera) cannot be zero. Thus a rotation angle \( \beta = \arctan(n_2/n_3) \) around the x-axis can be calculated eliminating the first entry of \( n \), and following this (if \( n_3 \) is still not zero) a second rotation round the y-axis by \( \alpha = \arctan(n_2/n_3) \) will yield \( n \) to be the z-axis. Applying the accordant rotation as homography to the mosaic patch should then lead to a rectified mapping, i.e. a similarity mapping of the scene keeping e.g. angles invariant. Fig. 1 shows that currently the system will signal “failure” and terminate if the decomposition failed – e.g. if \( t \) turns out too small, and thus the estimation of \( n \) is instable.

3.3. Global Methods and Loop Closure

Today it is recommended to use global methods that maximize precision by considering also correspondences between non-successive images of the video. Most popular is bundle-adjustment [6] and [12]. However, the computational efforts then are rising non-linearly with the video length, and there will be little hope that any correct correspondence can be established between arbitrarily picked distinct
frames if the single frame is much smaller than the final mosaic.

An alternative is occasional loop-closure setting correspondence between non-adjacent frames of the video, whenever chances for correspondence are high [7]. Here the goal was estimating the drift from the pair wise homography chain only, and the error of the surface normal estimation from homography decomposition. However, estimation and minimization of such drift will help identifying auspicious non-adjacent image pairs, where a loop-closure has good potential. Exemplarily, such experiment was appended to the following section.

4. Experiments

Fig. 2 shows the experimental setup in the test-basin of Fraunhofer-IOSB in Ilmenau, Germany. A portal allows computer controlled positioning under water in all three axes with high precision and repeatability. The portal carries a platform with a camera and two lights. At the wall of the basin one calibration target and two example surface targets are mounted. The latter contain standard faults – such as cracks and erosion - which are of interest for the inspectors. One target is made of bricks the other one of concrete.

The only known positioning error of the portal is a catenary of 14mm due to the weight of the portal. This is irrelevant for the example mosaic taken at the center of the pool wall. X, Y and Z axes can be assumed orthogonal with much higher accuracy than the mosaic. Positioning and repeatability are at the order of 1mm.

Fig. 3 shows one of the frames of the video taken of the brick-target using a push-broom trajectory always 400mm off the target travelling 500mm for each x-swath and then 100mm down in z-direction. Traveling speed was 50mm per second. The human observer may notice lens distortions – assuming regular brick structure. Recall that for the machine vision geometry these distortions have been modeled using one parameter and calibrated in the same water using the nearby calibration target [9].

Fig. 4 shows the resulting mosaic. Red dots indicate the position of each camera main inside the mosaic thus giving the estimated trajectory. From this the reader can assess the forward-chaining drift. This video contains about 2000 frames. Homography decomposition was performed and the patch-normal rectified according to the methods given in Sect. 3.2.
For assessing the rectification result the occurring deviations in the scene normal can be taken as rotational error around the two axes parallel to the scene directly because for this experiment the camera was aligned with the surface normal. These results are summarized in Table 1.

**Figure 5: Tilted Underwater Video Frame**

Fig. 5 shows one of the frames of another video taken of the same brick-target using a straight trajectory always 600mm off the target travelling 800mm along the x-axis at a speed of 100mm per second. This time the camera was mounted pointing 12° downwards, so that not only lens distortions but also perspective shortening can be seen.

**Figure 6: Rectified Mosaic of the Tilted Video**

Fig. 6 shows the mosaic resulting from this video. Note that the metal frame of the target appears roughly rectangular in this mosaic indicating that the rectification succeeded. Six partial mosaics are combined into one here. The red trajectory in Figure 6 deviates from a straight line in average 0.85 Pixel. But there is a considerable bias overestimating the camera tilt – as can be seen in Table 1.

**Table 1: Angles around x- and y-axes**

<table>
<thead>
<tr>
<th>Video</th>
<th>#patch</th>
<th>Truth</th>
<th>Estimation</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98</td>
<td>0°, 0°</td>
<td>0.02°, 0.02°</td>
<td>0.8°, 1.4°</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0°, 12°</td>
<td>2.3°, 14.2°</td>
<td>0.2°, 0.7°</td>
</tr>
</tbody>
</table>

**Figure 7: Frame with Pincushion Distortion**

**Figure 8: According Rectified Mosaic**

**Figure 9: Mosaic with Loop-closure**
Additional experiments were made with a different camera displaying pincushion distortion (in contrast to the barrel distortion visible in Figure 5). One frame is displayed in Figure 7. Mosaicing runs quite the same (after setting the internal camera parameters according to the new calibration) and Figure 8 displays the result. Here, the camera path was closed, so that a single loop-closure following Meidow [7] could be applied afterwards. The gain from such global method can be seen in Figure 9.

5. Discussion

Of course we have also conducted experiments with floating platforms and handheld cameras in outdoor waters where the visibility conditions are usually worse. This will also have an effect on the errors. However, for the time being, we have no groundtruth of the trajectory and orientations for such videos. The portal made it possible to conduct the experiments under water. Experiments in air would be less representative for the underwater inspection task because off the influence of the optical density of the medium on the mapping geometry. E.g. the location of the virtual projection center is not easy to be determined, causing error in the calibration of the focal length. It is also important to use the same lighting conditions as will be used on inspection cruises.

We conclude that under such benign visibility conditions and with enough observable structure the reported drift errors are small enough for automatic underwater infrastructure inspection.

References


