Augmented State–Extended Kalman Filter Combined Framework for Topology Estimation in Large-Area Underwater Mapping
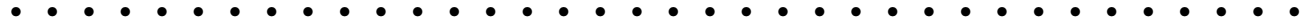
• •

Armagan Elibol, Nuno Gracias, and Rafael Garcia
Computer Vision and Robotics Group, University of Girona, 17071 Girona, Spain
e-mail: aelibol@eia.udg.edu, ngracias@eia.udg.edu, rafa@eia.udg.edu

Over the past few years, underwater vehicles have greatly improved as a tool for undersea exploration and navigation. In particular, autonomous navigation, localization, and mapping through optical imaging have become topics of interest for researchers in both underwater robotics and marine science. Underwater imagery can be used to construct image composites (photomosaics) used in many different application areas such as underwater surveying and navigation. For surveying operations with a low-cost robot limited to a down-looking camera and a sonar altimeter, it is common practice to ensure that there is enough overlap between time-consecutive images as this is the only data source of navigation. When the robot revisits a previously surveyed area, it is essential to detect and match the non-time-consecutive images to close a loop and, thus, improve trajectory estimation. While creating the mosaic, most of the existing algorithms try to match all image pairs to detect the non-time-consecutive overlapping images when there is no additional navigation information. We present a framework to obtain a two-dimensional mosaic with minimum image matching attempts and simultaneously get the best possible trajectory estimation by exploring the contribution of the image pairs matching to the whole system. Different strategies for choosing match candidates have been tested, and the results are given in different challenging underwater image sequences. © 2010 Wiley Periodicals, Inc.

1. INTRODUCTION

When underwater vehicles (UVs) perform missions near the seafloor, optical sensors can be used for navigation, motion planning, localization, and mapping. These sensors are especially useful in the case of low-cost robots, which incorporate a very limited sensor suite. The pose (position and orientation) of a low-cost underwater robot can be calculated by integrating the apparent motion between consecutive images acquired by a down-looking camera carried by the vehicle. Knowledge of the pose at image acquisition instances can also be used to align consecutive frames to form a mosaic, i.e., a composite image that covers the entire scene imaged by the submersible. Several strategies in the literature have attempted to recover the vehicle motion using visual mosaics (Andersen & Taylor, 2007; Gracias & Santos-Victor, 2000; Unnikrishnan & Kelly, 2002). These visual positioning systems allow the vehicle to localize itself on the mosaic map as it is being built (known as simultaneous localization and mapping, SLAM). Once the map has been constructed, the mosaic serves several purposes, such as the following:

1. To carry out map-based navigation, planning the path of the vehicle during the execution of the mission.
2. To use as a high-resolution image to perform some further processing such as localizing interest areas,

enabling the detection of temporal changes in the marine habitats.

Underwater images are also crucial for studying the ocean, especially in understanding biological and geological processes happening on the seafloor. The characteristics of the underwater environment are very challenging for optical imaging, mainly due to the significant attenuation and scattering of visible light (Loisel & Stramski, 2000; Pegau, Gray, & Zaneveld, 1997). Commonly, underwater images suffer from lack of contrast, blurring, and variable illumination due to refracted sunlight or artificial illumination. Moreover, light attenuation does not allow images to be taken from a long distance. Therefore, mosaicking techniques are needed to create high-resolution maps of the surveyed area using a large number of acquired images and to get a global perspective of the underwater terrain (Gracias & Santos-Victor, 2000; Leone, Distante, Mastrolia, & Indiverr, 2006; Pizarro & Singh, 2003; Richmond & Rock, 2006; Rzhanov, Mayer, Beaulieu, Shank, Soule, et al., 2006; Vincent, Pessel, Borgetto, Jouffroy, Opderbecke, et al., 2003). Thus, robotic exploration to construct photomosaics is becoming a common requirement in geological surveys, mapping, and temporal change detection (Delaunoy, Gracias, & Garcia, 2008; Escartin, Garcia, Delaunoy, Ferrer, Gracias, et al., 2008; Eustice, Singh, Leonard, & Walter, 2006).

Consider a scenario in which a low-cost underwater robot surveys an area of interest. Commonly, such a vehicle is equipped with only a video camera, to provide feedback to the pilot while being teleoperated from a mother vessel. Although autonomous underwater vehicles (AUVs) are normally equipped with different sensors such as DVL (Doppler velocity log), INS (inertial navigation system), USBL (ultra short base line) and ring laser gyroscopes, most commercially available low-cost remotely operated vehicles (ROVs) are limited to a video camera, lights, a pressure sensor (AC-ROV, 2009; GNOM Standard ROV, 2009) and in some cases a compass (Proteus 500 ROV, 2009; SeaBotix LBV150BE MiniROV, 2009; SM 1000 Low Cost ROV System, 2009). Generally, when lacking other sensor data (e.g., USBL, DVL, gyrocompass), time-consecutive images are assumed to have an overlapping area. This overlap allows the images to be registered and an initial estimate of the camera trajectory over time to be obtained. This initial dead-reckoning estimation suffers from rapid accumulation of registration errors, leading to drifts from the real trajectory. The initial estimate, however, provides useful information for the detection of non-time-consecutive overlapping images. Matching non-time-consecutive images is a key step to refine the trajectory followed by the robot using global alignment methods (Capel, 2004; Elibol, Garcia, Delaunoy, & Gracias, 2008; Ferrer, Elibol, Delaunoy, Gracias, & Garcia, 2007; Gracias, Costeira, & Victor, 2004; Sawhney, Hsu, & Kumar, 1998; Szeliski & Shum, 1997). With the refined trajectory, new non-time-consecutive overlapping images can be predicted and attempted to match. The iterative matching and optimization process continues until no new overlapping images are detected. This process is known as a *topology estimation*. In the context of this paper, we refer to the topology estimation as the problem of finding overlapping image pairs among different transect(s) of the surveyed area.

Finding matches among non-time-consecutive image pairs is usually referred to as loop closing, i.e., detecting that the area being explored has been visited before. Closing loops is essential to reduce the uncertainty of the trajectory estimation (Beevers & Huang, 2005; Cheng & Han, 2006; Clemente, Davison, Reid, Neira, & Tardós, 2007; Fengda, Lingfu, & Li, 2007; Ho & Newman, 2007; Ila, Andrade-Cetto, Valencia, & Sanfeliu, 2007; Jungho & In-So, 2007). Impressive progress has recently been achieved in the field of SLAM for underwater platforms equipped with cameras (Eustice et al., 2006; Mahon, Williams, Pizarro, & Johnson-Roberson, 2008; Singh, Roman, Pizarro, Eustice, & Can, 2007) and sonar (Newman, Leonard, & Rikoski, 2003; Ribas, Neira, Ridao, & Tardos, 2006; Ruiz, de Raucourt, Petillot, & Lane, 2004). SLAM approaches are well suited for navigation applications, namely in real-time control and localization of vehicles, and have been successfully used for online image mosaicing in medium-size data sets (Garcia, Puig, Ridao, & Cufí, 2002; Richmond &

Rock, 2006). This contrasts with offline batch approaches in which the data are processed a posteriori. By avoiding real-time constraints, large-scale optimization methods can be used on considerably larger data sets, with significantly higher accuracy in the final results (Escartin et al., 2008).

Recent advances in image matching techniques (Bay, Tuytelaars, & Van Gool, 2006; Lowe, 2004; Mikolajczyk & Schmid, 2005; Zitová & Flusser, 2003), such as scale invariant feature transform (SIFT), allow pairs of images to be registered in the absence of prior information on orientation, scale, or motion between images. Such techniques are behind the recent widespread availability of photo stitching softwares in the computer vision community, because they allow panoramas to be created with minimal user input (Brown & Lowe, 2007; Yao & Chamb, 2007). In most cases, these approaches attempt to match all images against all others. Although this is feasible for small data sets, it becomes impractical for the creation of large-area mosaics in which useful surveys may comprise several hundred images (Ferrer et al., 2007). The number of possible image pairs is a quadratic function of the total number of images in the sequence, and the processing becomes impractical for large image sets (e.g., for surveys containing 100, 250, and 500 images, the total numbers of all possible image pairs are 4,950, 31,125, and 124,750, respectively). The image matching process requires the execution of several steps: feature detection, feature description, feature matching, and outlier rejection. Therefore, its computational cost is inherently high. Hence, it is very important to reduce the number of image matching attempts.

Although extensive research has been carried out on control and estimation problems,[1] almost none of the existing methods have been studied in a batch mosaicking framework, to decide which image pairs should be matched and in which order. This paper proposes a solution to the problem of topology estimation in large-scale batch mosaicking using a combined framework of augmented state Kalman filter (ASKF) and extended Kalman filter (EKF). We aim to minimize the number of image matching attempts and simultaneously obtain an accurate trajectory estimation. Our method explores the contributions of image matchings and chooses which images should be matched first. In this study, we have extended our previous work (Elibol, Gracias, & Garcia, 2009) by removing the assumption that all time-consecutive images have an overlapping area. As input, we assume a set of images without any additional information about their alignment. We initialize our framework by using ASKF with the time-consecutive images as if they had an overlapping area. At the end, we also add a step; once all the overlapping images have been found and matched, all of them can be incorporated into an iterated extended Kalman filter (IEKF)

---

[1]Especially in localization, navigation, SLAM, and target tracking.

(Bar-Shalom, Li, & Kirubarajan, 2001), reducing the uncertainty and improving the trajectory estimation. We also present a new derivation that allows the observation mutual information (OMI) (Grocholsky, 2002) to be computed in an efficient way. OMI is used to measure the amount of information one observation can provide to the whole topology within a Kalman filter estimator.

The paper is organized as follows. In the next section, relevant related work is provided on image mosaicking, estimation, and control problems. Section 2.1 is dedicated to some basic definitions of information and Kalman filters. A computationally efficient formula for calculating the OMI is derived in Section 3. The ASKF–EKF combined framework for topology estimation is introduced in Section 4, and a detailed explanation of the framework is given in Sections 4.1 and 4.2. Experimental results are illustrated and discussed in Section 5. Finally, the last section is dedicated to conclusions.

2. RELATED WORK

Several image mosaicking methods have been proposed over the past decade. One of the first mosaicking systems for an underwater robot was proposed by Marks, Rock, and Lee (1995). This system achieved real-time performance due to the use of special-purpose hardware for image filtering and correlation, allowing the creation of "single column" mosaics in real time. Because the robot could control its heading, a very restrictive (translation-only) motion model was assumed, without taking into account rotation, scaling, or perspective distortion. For the system to work, the robot was programmed to keep constant altitude. This work was extended by Fleischer, Rock, and Burton (1997) to deal with loop trajectories, by detecting crossovers in previous mosaicked areas. The same translation-only motion model was used, which would degrade the mosaic quality whenever the robot performed changes in altitude, roll, or pitch. Sawhney et al. (1998) proposed an end-to-end solution for image mosaicking in which the image topology (i.e., the spatial relations between overlapping frames) was iteratively estimated. Spatial consistency was improved by identifying and registering images with large superposition. A simpler approach was followed by Davis (1998) to register images captured by a rotating camera. With small rotations and some assumptions about the camera intrinsic parameters, phase-correlation methods were used for pairwise registration. A system of linear equations was defined for the elements of all the homographies, relating each image with a reference image for which a least-squares solution is obtained. However, no adequate parameterization was used on these elements to take advantage of the special structure of the rotation-induced homography. Duffin and Barrett (1998) used a homography parameterization for global registration that imposed constant camera skew and aspect ratio. Other constraints on the camera and scene

geometry were not taken into account. Unnikrishnan and Kelly (2002) addressed the problem of efficiently distorting strip mosaics to close loops in a smooth way. The proposed solution had low computational complexity and was best suited for cases in which the number of temporally distant overlaps was small compared to the adjacent overlaps. The problem of finding correspondences at the extremities of mosaic segments, required for imposing endpose constraints, was not addressed. Capel (2004) presented a complete mosaicking system with topology estimation and global registration by extending a maximum likelihood estimation for the two-view homographies to the case of multiple N views. However, the total number of unknowns increased drastically. For long sequences, this issue might cause some convergency problems during minimization. Also, this method has not been tested with low overlapping images.

Kalman filter–based image mosaicking has been studied, especially for mosaic-based navigation purposes (Caballero, Merino, Ferruz, & Ollero, 2007; Garcia et al., 2002; Richmond & Rock, 2006). Garcia et al. (2002) presented an ASKF for position estimation of AUVs. It uses image matching to provide incremental one-dimensional (1D) rotation and two-dimensional (2D) translation information (in X and Y) and an altimeter for translation in Z. Results were presented for a simulation using a constant velocity model of an AUV to generate the observation data. In Richmond and Rock (2006), a system combining vision and DVL sensors was proposed to estimate the vehicle position in real time along the mosaic of the visited area. This system combined vision and DVL odometry to get the current state of the vehicle, and image registration is used to bound the odometry drift. Richmond and Rock (2006) acknowledged that mosaics from this real-time system were not as accurate as those generated by offline methods. Caballero et al. (2007) proposed EKF-based image mosaicking to estimate unmanned aerial vehicle (UAV) position. In their model, the state vector was composed of absolute homographies and the state was updated when a loop closure was detected and images were processed sequentially. Eustice, Pizarro, and Singh (2004) proposed a system based on the ASKF with measurements provided by inertial sensors and monocular video. Mahon et al. (2008) presented mapping results of large areas; however, they used an extensive and expensive sensor suite, including DVL to compute the dead reckoning as well as a compass and a tilt sensor to obtain the vehicle's orientation and a pressure sensor to measure depth. Moreover, a stereovision rig is used to provide loop-closure observations. Because of the accuracy of the DVL over short distances, the vision system is not used to provide odometry information. All these sensors allow navigating over large areas. Ila et al. (2007) proposed loop-closure detection through a test composed of two passes. First, the Mahalanobis distance was used to detect the closure, and then the Bhattacharyya distance

was employed to choose the most informative pose pairs. Recently, Ila, Porta, and Andrade-Cetto (2010) proposed a method to keep the most informative links between robot poses using mutual information within a SLAM context, which relates to our paper. However, we consider all potentially possible matching pairs among all images, which is a problem different from that of performing matches from the most recent image to all previous images as the robot moves. Because we are considering all possible matching pairs, it is necessary to have a reasonable ranking criterion in terms of their contribution while obtaining the real topology and the trajectory. All these methods have been developed in the context of position estimation and SLAM; they have not been addressed for fast topology estimation in batch mosaicking.

Exploiting the data provided by more than one sensor has been extensively researched as multisensor data fusion (Gan & Harris, 2001; Jin & Sun, 2003; Yukun, Xicai, & Zhigang, 2007), sensor selection (Mattikalli, Fresnedo, Frank, Locke, & Thunemann, 2007; Oshman, 1994; Pahalawatta, Pappas, & Katsaggelos, 2004; Sun, 2004), sensor management, and scheduling (Kangsheng & Guangxi, 2006; Karan & Krishnamurthy, 2003; Krishnamurthy, 2002; Logothetis & Isaksson, 1999; Ng & Ng, 2000). Information theory, operational research, and optimization theory have been the most important fields to help select data from sensors, fuse the data, and make decisions (Hall & McMullen, 2004). Several researchers have studied methods for improving estimation accuracy using information provided by several sensors (Denzler & Brown, 2002; Grocholsky, 2002; Manyika & Durrant-Whyte, 1994; Whaite & Ferrie, 1997; Xiong & Svensson, 2002).

We formulate our problem as a sensor fusion and management problem within the Kalman filter estimation framework. In our work, image matching between overlapping image pairs is treated as an observation or measurement that comes from a sensor. Then, a predicted gain is calculated as the amount of information the observation provides to the information matrix of the whole system. This is obtained by computing the OMI. If sensor noises are independent, there are basically two sensor fusion methods. The first method consists of stacking (augmenting) the observation vector. Thus, the observation vector increases in size as new observations are available. Stacked covariance and observation transition matrices are used. The second method combines the observations based on the minimum mean-square-error estimation and keeps the observation vector dimension unchanged. Each observation is weighted by the covariance of its noise. If observation noises are dependent, the stacking method is employed, but this time the stacked noise covariance matrix is no longer diagonal. However, the dependent sensor noises can be formulated in such a way that they can be treated as independent noise (Duan, Han, & Tao, 2004; Jin & Sun, 2003).

Sensor management consists of selecting the subset of sensors that will be used to obtain a measurement at time epoch k and is formulated as a combinatorial optimization and/or search problem (Ng & Ng, 2000; Xiong & Svensson, 2002). The aim is to find the best subset to fulfill the sensor constraints (such as energy, computational cost, and bandwidth) and maximize or minimize the selected cost function (i.e., information gain, uncertainty).

We now summarize the concepts of the Kalman filter, the information filter, and some information measures (Mutambura, 1998) used in this paper.

Given a state vector $\mathbf{x}$ and its covariance $\mathbf{P}$, the Kalman filter update equations are as follows (Anderson & Moore, 1979):

$$\mathbf{S}(k) = \mathbf{H}(k)\mathbf{P}(k \mid k - 1)\mathbf{H}(k)^T + \mathbf{R}(k),$$

$$\mathbf{K}(k) = \mathbf{P}(k \mid k - 1)\mathbf{H}(k)^T \mathbf{S}(k)^{-1},$$

$$\mathbf{P}(k \mid k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]\mathbf{P}(k \mid k - 1),$$

$$\mathbf{x}(k \mid k) = \mathbf{x}(k \mid k - 1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\mathbf{x}(k \mid k - 1)], \quad (1)$$

where $\mathbf{S}(k)$ is an innovation covariance matrix and $\mathbf{K}(k)$ is the Kalman gain. $\mathbf{z}(k)$ is the observation value provided by a sensor. The observation noise is a zero-mean Gaussian noise with covariance $\mathbf{R}(k)$. The observation prediction, which can be computed from the state vector, is denoted as $h(x(k \mid k - 1))$, where h is the function that maps the state vector to the observations. As this function is usually nonlinear, the transition matrix from state to observation $\mathbf{H}(k)$, is calculated as follows:

$$\mathbf{H}(k) = \left.\frac{\partial h}{\partial \mathbf{x}}\right|_{\mathbf{x} = \mathbf{x}(k \mid k - 1)}. \quad (2)$$

The notation $(\cdot)(k \mid t)$ refers to a value at epoch k given t. An information filter is the dual form of a Kalman filter and can be found by changing the state vector $\mathbf{x}$ and covariance $\mathbf{P}$ with the information state $\mathbf{y}$ and *Fisher information* matrix $\mathbf{Y}$ (Anderson & Moore, 1979):

$$\mathbf{y}(k \mid t) = \mathbf{P}^{-1}(k \mid t) \cdot \mathbf{x}(k \mid t), \quad (3)$$

$$\mathbf{Y}(k \mid t) = \mathbf{P}^{-1}(k \mid t). \quad (4)$$

An observation $\mathbf{z}(k)$ at time epoch k contributes $\mathbf{i}(k)$ to the information state and $\mathbf{I}(k)$ to the Fisher information matrix by means of sufficient statistics as given in Manyika and Durrant-Whyte (1994):

$$\mathbf{i}(k) = \mathbf{H}(k)^T \mathbf{R}(k)^{-1}[\mathbf{z}(k) - h(\mathbf{x}(k \mid k - 1))$$

$$+ \mathbf{H}(k)\mathbf{x}(k \mid k - 1)], \quad (5)$$

$$\mathbf{I}(k) = \mathbf{H}(k)^T \mathbf{R}(k)^{-1}\mathbf{H}(k). \quad (6)$$

The update equations of the information filter have the following compact form:

$$\mathbf{y}(k \mid k) = \mathbf{y}(k \mid k - 1) + \mathbf{i}(k), \quad (7)$$

$$\mathbf{Y}(k \mid k) = \mathbf{Y}(k \mid k - 1) + \mathbf{I}(k). \quad (8)$$

For N independent sensors, Eqs. (7) and (8) become

$$\mathbf{y}(k \mid k) = \mathbf{y}(k \mid k-1) + \sum_{j=1}^{N} \mathbf{i}(k)_j, \qquad (9)$$

$$\mathbf{Y}(k \mid k) = \mathbf{Y}(k \mid k-1) + \sum_{j=1}^{N} \mathbf{I}(k)_j. \qquad (10)$$

The entropic information about the system can be modeled before and after making an observation, respectively, given in the following equations:

$$\iota(k \mid k-1) = \frac{1}{2} \ln[(2\pi e)^{-n} |\mathbf{Y}(k \mid k-1)|], \qquad (11)$$

$$\iota(k \mid k) = \frac{1}{2} \ln[(2\pi e)^{-n} |\mathbf{Y}(k \mid k)|]. \qquad (12)$$

The mutual information between the discrete random variables **A** and **B** provides an answer for the question "How much information does the random variable **B** give about the random variable **A**?" Following the definition of the mutual information, OMI provides an answer for the question "How much information does the observation provide to the system?" In other words, OMI tells how much the uncertainty of the state will reduce when the observation is realized. Therefore, OMI (Grocholsky, 2002) is an important measure, which can be easily calculated from the information matrices as the change in information, yielding

$$I(k, \mathbf{z}(k)) = \frac{1}{2} \ln \left[\frac{|\mathbf{Y}(k \mid k)|}{|\mathbf{Y}(k \mid k-1)|} \right]. \qquad (13)$$

3. EFFICIENT CLOSED-FORM SOLUTION FOR CALCULATING OBSERVATION MUTUAL INFORMATION

The calculation of OMI has a high computational cost whether a Kalman or an information filter is used. It consists of calculating the determinant of either the covariance or the information matrices. For an $n \times n$ matrix, the time complexity of computing the determinant from its naive definition is of the order $O(n!)$. Using *LU decomposition*,[2] the computational complexity is $O(n^3)$. Therefore, the computational cost of the OMI is $O(n^3)$, where n is the size of the state vector. However, rearranging the equations for OMI and using the structure of the Kalman filter, the computational cost can be further reduced to $O(m^3)$, where m is the size of the observation vector. We now introduce a new derivation that allows this significant reduction. First, we convert the OMI formulation given in Eq. (13) from information form to covariance form:

$$I(k, \mathbf{z}(k)) = \frac{1}{2} \ln \left[\frac{|\mathbf{P}(k \mid k)^{-1}|}{|\mathbf{P}(k \mid k-1)^{-1}|} \right],$$

[2]Used by the det() function of MATLAB™.

$$= \frac{1}{2} \ln \left[\frac{|\mathbf{P}(k \mid (k-1)|}{|\mathbf{P}(k \mid k)|} \right]. \qquad (14)$$

Equation (14) can be reformulated as follows by using the Kalman filter equations from Eq. (1):

$$I[k, \mathbf{z}(k)] = \frac{1}{2} \ln \left[\frac{1}{|\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)|} \right]$$

$$= \frac{1}{2} \ln[|[\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]^{-1}|]. \qquad (15)$$

However, the formula in Eq. (15) is still the same size as the covariance matrix of the system. From the determinant properties (Golub & van Loan, 1996), given two $p \times q$ matrices, **A** and **B**, it holds that

$$|(\mathbf{I}_p + \mathbf{A}\mathbf{B}^T)| = |(\mathbf{I}_q + \mathbf{B}^T\mathbf{A})|. \qquad (16)$$

Therefore, $|[\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]|$ can be rewritten, and Eq. (15) becomes

$$I[k, z(k)] = \frac{1}{2} \ln[|[\mathbf{I} - \mathbf{H}(k)\mathbf{K}(k)]^{-1}|]. \qquad (17)$$

If we premultiply Eq. (1) with $\mathbf{H}(k)$, we obtain

$$\mathbf{H}(k)\mathbf{K}(k) = \underbrace{\mathbf{H}(k)\mathbf{P}(k \mid k-1)\mathbf{H}(k)^T} \mathbf{S}(k)^{-1}$$

$$= [\mathbf{S}(k) - \mathbf{R}(k)]\mathbf{S}(k)^{-1}$$

$$= \mathbf{I} - \mathbf{R}(k)\mathbf{S}(k)^{-1}. \qquad (18)$$

We can now replace $\mathbf{H}(k)\mathbf{K}(k)$ in Eq. (17) with its equivalent in Eq. (18):

$$\mathbf{I} - \mathbf{H}(k)\mathbf{K}(k) = \mathbf{R}(k)\mathbf{S}(k)^{-1},$$

$$[\mathbf{I} - \mathbf{H}(k)\mathbf{K}(k)]^{-1} = \mathbf{S}(k)\mathbf{R}(k)^{-1},$$

$$|[\mathbf{I} - \mathbf{H}(k)\mathbf{K}(k)]^{-1}| = |\mathbf{S}(k)||\mathbf{R}(k)^{-1}|. \qquad (19)$$

Finally, we rewrite Eq. (17) by using the last line in Eq. (19):

$$I[k, \mathbf{z}(k)] = \frac{1}{2} \ln[|\mathbf{S}(k)||\mathbf{R}(k)^{-1}|] \qquad (20)$$

The OMI calculation in Eq. (20) consists of calculating two determinants of matrices sized $m \times m$, instead of calculating $n \times n$ determinants where usually $m \ll n$. Using a different reasoning, we have reached the same solution as those presented in Ertin, Fisher, and Potter (2003) and Ila et al. (2010), which used a Bayesian formulation. Our novel derivation is simpler and easier to follow.

4. ASKF–EKF COMBINED FRAMEWORK FOR TOPOLOGY ESTIMATION

This section discusses how some tools from control and estimation theory can be applied to our problem. Our model is inspired by Kalman filter–based image mosaicking strategies (Caballero et al., 2007; Garcia et al., 2002; Richmond & Rock, 2006). As our interest is batch mosaicking, we do not need any control input and, therefore, our model does not have any state prediction equations. Only update equations are used.

As matching non-time-consecutive image pairs provides additional information about the topology and improves the trajectory estimation, it is indispensable to detect them to get better trajectory estimation. In this context, it is important to measure the impact of matching one image pair in terms of how much information it will provide about the topology. Uncertainty of observations arises from the detected feature points in images. To keep observations uncorrelated, the same feature point should not be used in calculations of the uncertainty of two or more different observations. While computing the uncertainty of the observations, we have used a small subset of detected feature points. This ensures the independence among the observation elements. Moreover, as a design option, each image is used once (at most) in each iteration of the algorithm. Modeling the problem in this way allows us to use the standard for sensor fusion, selection, and management.

4.1. Definitions

1. The *state vector* is created in the initialization step using the ASKF algorithm in Table I and is composed of the homography values that relate every image with the mosaic frame:

$$x_i = \text{vec}(^m\mathbf{H}_i), \quad i = 1, 2, 3, \ldots, N, \tag{21}$$

where N is the total number of images and $\text{vec}(\cdot)$ is the function that converts the homography matrix input into a vector. $\mathbf{P}$ denotes the covariance matrix of the state vector $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N]^T$. We have used similarity homographies that have four degrees of freedom (DOF) (scaling, rotation, and translation in both x and y axes):

$$^m\mathbf{H}_i = \begin{bmatrix} a_i & -b_i & c_i \\ b_i & a_i & d_i \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{x}_i = [a_i \quad b_i \quad c_i \quad d_i]^T.$$

Similarity-type homographies represent an adequate trade-off between (1) encoding the trajectory of a down-looking camera facing an approximately flat surface from a typical surveying altitude above the seafloor and (2) resilience to fast error accumulation, which results from cascading these transformations, in the absence of other sensors (Negahdaripour & Firoozfam, 2001).

2. A new *observation (measurement)* is obtained when two images, i and j, are successfully matched. The observation is represented by the homography

between corresponding images at time epoch k:

$$\begin{aligned} \mathbf{z}(k) &= \text{vec}\left(^i\mathbf{H}_j\right) + \mathbf{v}(k) \\ &= {}^i\mathbf{H}_m \cdot {}^m\mathbf{H}_j + \mathbf{v}(k) \\ &= \text{mat}(\mathbf{x}_i)^{-1} \cdot \text{mat}(\mathbf{x}_j) + \mathbf{v}(k), \end{aligned} \tag{22}$$

where $\text{mat}(\cdot)$ is the function that converts the state vector into homography matrices and $\mathbf{v}(k)$ is the observation noise vector. It is assumed that the observation noise is Gaussian and is not correlated with state noise and its covariance matrix is $\mathbf{R}(k)$.

3. A potential observation is an image pair that has a potential overlapping area as predicted by the state and its uncertainty.
4. One time epoch is defined as a full cycle of the following steps, which are detailed in the next section: generation of possible observation list, selection, image matching, and filter update (see Figure 1).

We have tested five different ranking strategies to select the observations we attempted to match. All these strategies are used while selecting which observations to carry out.

Expected Overlap The expected overlap criterion ranks the pairs according to their overlap probability and thus a greater chance of being successfully matched. The robot trajectory and its uncertainty are used to compute the overlap probability. Computational details are explained in Section 4.2.

OMI The OMI score is calculated for each observation in the potential observation list. To compute this score, a generic observation noise covariance matrix $\mathbf{R}(k)$ is used. This calculated OMI is the *predicted information gain* of the observation.

Expected Overlap Weighted OMI The expected overlap weighted OMI combines the first two ranking criteria using the OMI score as a weight to overlap probability.

Combined Because a loop-closure event can considerably reduce the uncertainty and trajectory drift, it is important to be able to detect such an event as soon as possible. Preliminary results (Elibol et al., 2009) have shown that after a certain number of iterations, the matching of previously unmatched image pairs does not provide significant information. In other words, when we order the potential image pairs according to their OMI score, there is no significant difference between the highest OMI score and the lowest one. At this step, instead of using OMI-based ordering, one could consider using one of the other strategies with less computational cost than OMI. Therefore, a combined strategy is devised: the expected overlap weighted OMI strategy for a small number of initial iterations (manually
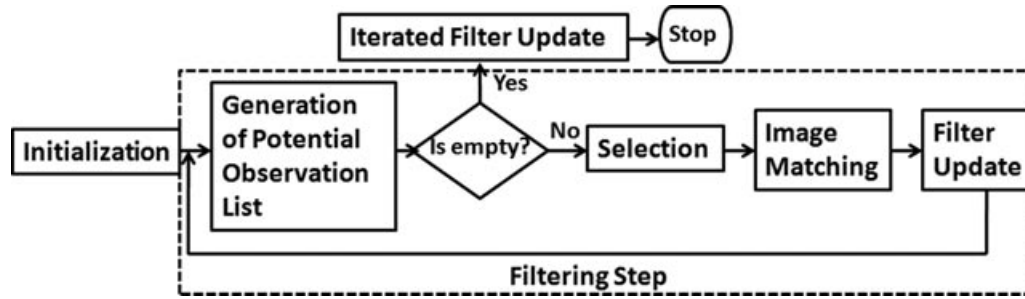
Figure 1. Pipeline of the proposed framework for topology estimation.

chosen between two and five) and then using the expected overlap strategy to rank the potential observations.

Random The random criterion orders image pairs randomly. It is included as a baseline to compare the performances of the other criteria.

4.2. Implementation

Our framework is composed of three main steps: initialization, filtering, and iterated update. The filtering step is divided into four subparts: potential observation list generation, selection, image matching, and filter update. The pipeline is illustrated in Figure 1.

Initialization This step instantiates the state vector and its covariance matrix using an ASKF formulation.[3] The algorithm is outlined in Table I. The first image frame is chosen as the global (mosaic) frame. Time-consecutive images are added to the system one by one as if they had an overlapping area with the previous image. For each new image in the sequence, the state is augmented by adding $\mathbf{x}_g = [1, 0, 0, 0]^T$ to the state vector and a 4×4 diagonal ma-

---

[3]The initialization step is also referred to the ASKF step in the rest of the paper because it is the only step at which the ASKF is employed.

---

trix $\mathbf{P}_g$ to the state covariance. Then an observation between the new image and the previous image is added in the form of an identity mapping $\mathbf{z}_g = [1, 0, 0, 0]^T$ with very high covariance matrix $\mathbf{R}_g$. The purpose of including this observation is to impose the soft prior that time-consecutive images have a greater chance of overlapping than non-time-consecutive images. The filter is then updated by using the Kalman filter update equations (1). Once the state augmentation is finalized, the resulting state vector is composed of identity mappings, and the covariance matrix grows from first image to the last image. The resulting state and covariance are the inputs for the later steps.

Once the state augmentation is finalized, the resulting state vector is composed of identity mappings, and the covariance matrix grows from the first image to the last image. The resulting state and covariance are the inputs for the later steps. The purpose of the initialization is to impose the soft prior that time-consecutive images have a greater chance of overlapping than non-time-consecutive ones.

Potential Overlapping Image List Once the initial state and covariance matrix are computed, a potential observation list is generated. This step requires computing an approximation of the probability that the two given images have an overlap. To compute this, we propose a method related to that of Mahon et al. (2008). In Mahon et al. (2008), the loop-closure hypotheses were computed by first

---

Table I. ASKF step.

| Input	Number of images, generic state vector $\mathbf{x}_g$, generic covariance matrix $\mathbf{P}_g$, generic observation $\mathbf{z}_g$, and observation noise covariance $\mathbf{R}_g$	
Output	State vector and its covariance matrix	
Step 1	Expand the state vector with $\mathbf{x}_g$.	$\mathbf{x}(k \mid k-1) = [\mathbf{x}_g, \mathbf{x}(k-1 \mid k-1)]^T$
Step 2	Expand the covariance matrix with $\mathbf{P}_g$.	$\mathbf{P}(k \mid k-1) = \begin{bmatrix} \mathbf{P}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{P}(k-1 \mid k-1) \end{bmatrix}$
Step 3	Filter update by using $\mathbf{x}(k \mid k-1)$, $\mathbf{R}_g$, $\mathbf{P}(k \mid k-1)$, and $\mathbf{z}_g$	
Step 4	Check whether all images in the sequence are added to the system. If not, go to Step 1	
Step 5	END.	
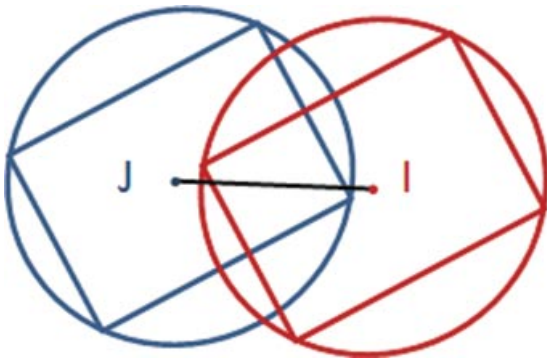
Figure 2. Images are assumed as circles with radius of half of their diagonal.

finding the overlapping pairs with a distance criterion between image centers. This criterion was based on the intersection between two circles, i.e., if the distance between the centers of the circles was smaller than the sum of their radii. Then it computes the likelihood of this overlap by using a displacement distribution sampled on a grid. Cells within the overlap bounds were integrated to estimate the likelihood of overlapping images. A 20×20 grid was used, requiring the calculation of 400 samples. Conversely, our method operates on the discretized distance between image centers including their uncertainties. We also assume that images are circular with a radius of half of the image diagonal (see Figure 2). The covariances of the image centers are propagated from the covariance matrix of the state

vector. The distance between image centers is discretized. In this discrete area, image vectors are generated. These vectors have value 1 where the point lies inside the image and 0 where the point lies outside the image. Because of the uncertainty of the image position, image vectors are convolved with Gaussian filters by taking into account the uncertainty of their centers. Convolution of the resulting vectors gives an approximation of the probabilities of the points to decide whether they belong to the images (see Figure 3). By counting the total number of nonzero elements in the convolution vector, the percentage of overlapping area can be approximated. If the percentage is bigger than a chosen threshold, the image pair is considered to be overlapping and is added to the potential observation list.

Selection After generating the list, different scores (e.g., information gain, expected overlap as described in Section 4.1) for each observation can be calculated for each strategy being tested. However, it is not possible to attempt to match all the observations in the list as it might include several nonoverlapping pairs due to the uncertainty and the trajectory estimation. Therefore, it is necessary to select a subset of the list.

The selection step aims to choose the subset of potential observations in such a way that maximizes the chosen score. This problem can be formulated as a variation of the linear assignment problem and can be solved with binary integer programming (Nemhauser & Wolsey, 1988). Each potential observation is composed of two images, i and j. Let C_k be a subset of all possibilities $A_k = \{(i, j) | i > j, i = j + 1, \ldots, N \quad j = 1, 2, \ldots, N - 1\}$, denoting the
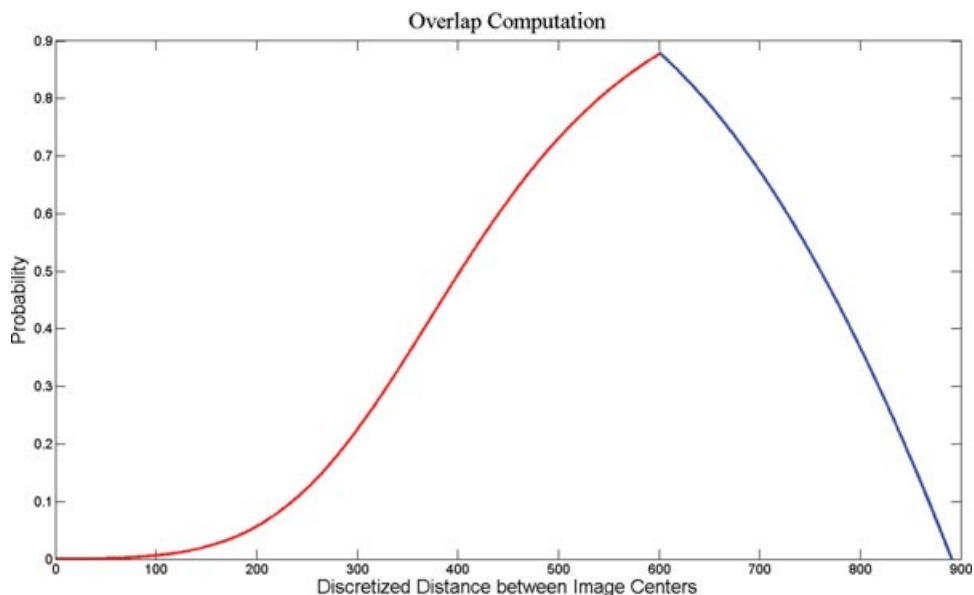

Figure 3. Image vectors convolved with Gaussians in the discretized distance between image centers. The red line denotes the vector for image i, and the blue line is for image j.

potential observation list at epoch k. The selection procedure finds the image indices, i, j that solve the following optimization problem:

$$\max \sum_{(i,j)\in C_k} d_{ij} \cdot \text{Score}(i,j) \quad \text{s.t.}$$
$$\sum_j d_{ij} = 1, \qquad (23)$$
$$\sum_i d_{ij} = 1,$$
$$d_{ij} \in \{0, 1\},$$

where $\text{Score}(i, j)$ is a function that returns the score of matching images i and j and d_{ij} is a decision variable of the observation that is composed of images i and j. The solution of this optimization problem provides the observations that will later be used as an input for the image matching step.

Image Matching After generating and choosing the list of potential observations, image matching starts. The image matching procedure is composed of two substeps: (1) SIFT (Lowe, 2004) is used to detect the features in images and (2) the random sample consensus (RANSAC) (Fischler & Bolles, 1981) algorithm is used to reject outliers and estimate the homography. Only one attempt is made to match each image pair. If the matching is not successful, the pair is marked as a nonmatch and abandoned. If it is successful, the noise covariance of the registration parameters is calculated from the correspondences using covariance propagation (Haralick, 1998), assuming additive Gaussian noise on the point correspondences and performing first-order noise propagation (details are given in the Appendix).

Filter Update The final procedure of the filtering step is to update the state and covariance using EKF formulations in Eq. (1). The filtering step is executed until there are no image pairs left in the potential observation list.

Iterated Filter Update As a final step, an IEKF is applied until the change in the reprojection error is smaller than a chosen threshold.

5. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed framework for ranking the observations using different strategies. Testing is performed on four real data sets, collected from underwater robots on seafloor survey missions. Data sets correspond to planar areas, although the contributions of the paper on reducing image matching attempts are also valid if we use a less generic framework to extend to three dimensions. Possible ways to perform this extension are by modeling the trajectory in three dimensions and assuming planar scenes (Ferrer et al., 2007) or by using the fundamental matrix (Pizarro, Eustice, & Singh, 2009). In the absence of ground truth, we have computed the trajectory

parameters using a bundle adjustment (BA) approach (Ferrer et al., 2007), which uses not only image registration but also navigation sensors (e.g., USBL), if available. This serves as a baseline to compare the results of the different strategies. As time-consecutive image pairs do not necessarily have an overlapping area, we performed all-against-all exhaustive image matching and then we employed BA, which minimizes the reprojection error[4] given in Eq. (24) over homographies. The cost function was minimized, and the parameters were estimated using large-scale methods for nonlinear least squares (Coleman & Li, 1996) as implemented in MATLAB[TM]. The optimization algorithm requires the computation of the Jacobian matrix containing the derivatives of all residuals with respect to all trajectory parameters. Fortunately, this Jacobian matrix is very sparse; each residual depends on only a very small number of parameters (Triggs, McLauchlan, Hartley, & Fitzgibbon, 2000). The result of applying BA is provided in the last row of the tables. The resulting homography set is used as a reference to compare the results of the proposed topology estimation framework. Our comparison criterion is the average reprojection error over all correspondences that were found by all-against-all image matching, defined as

$$\min_{{}^1\mathbf{H}_2, {}^1\mathbf{H}_3, \ldots, {}^1\mathbf{H}_N} \sum_k \sum_m \sum_{j=1}^n \parallel {}^k\mathbf{p}_j - {}^1\mathbf{H}_k^{-1} \cdot {}^1\mathbf{H}_m \cdot {}^m\mathbf{p}_j \parallel_2$$
$$+ \parallel {}^m\mathbf{p}_j - {}^1\mathbf{H}_m^{-1} \cdot {}^1\mathbf{H}_k \cdot {}^k\mathbf{p}_j \parallel_2, \qquad (24)$$

where k and m are the indices of the successfully matched pair of images ${}^k\mathbf{p}_j$ and ${}^m\mathbf{p}_j$ are the coordinates of the jth correspondence in images k and m, respectively; n is the total number of correspondences between the overlapping image pairs. N is the total number of images, and ${}^1\mathbf{H}_k$ and ${}^1\mathbf{H}_m$ are the homographies.

The first data set covers a challenging, large area of the seafloor that was acquired by the ICTINEU underwater robot (Ribas, Palomeras, Ridao, Carreras, & Hernandez, 2007) during experiments in Colera on the Mediterranean coast of Spain (see Figure 4). The trajectory was composed of seven vertical and two horizontal transects that provide several non-time-consecutive image pairs. It consisted of 430 images of 384×288 pixels, and it covered approximately 400 m^2. Table II summarizes the results for this data set. The first column lists the tested strategies. The second column shows the total number of successfully matched image pairs. The third column contains the total number of image pairs that were not successfully matched due to lack of overlapping areas or failure of the registration algorithm. They are called *unsuccessful observations*. The percentage of the total number of image matching attempts with respect to all-against-all image matching attempts is given in the

---

[4]Additional navigation information, when available, is included in the error term. For details refer to Ferrer et al. (2007).
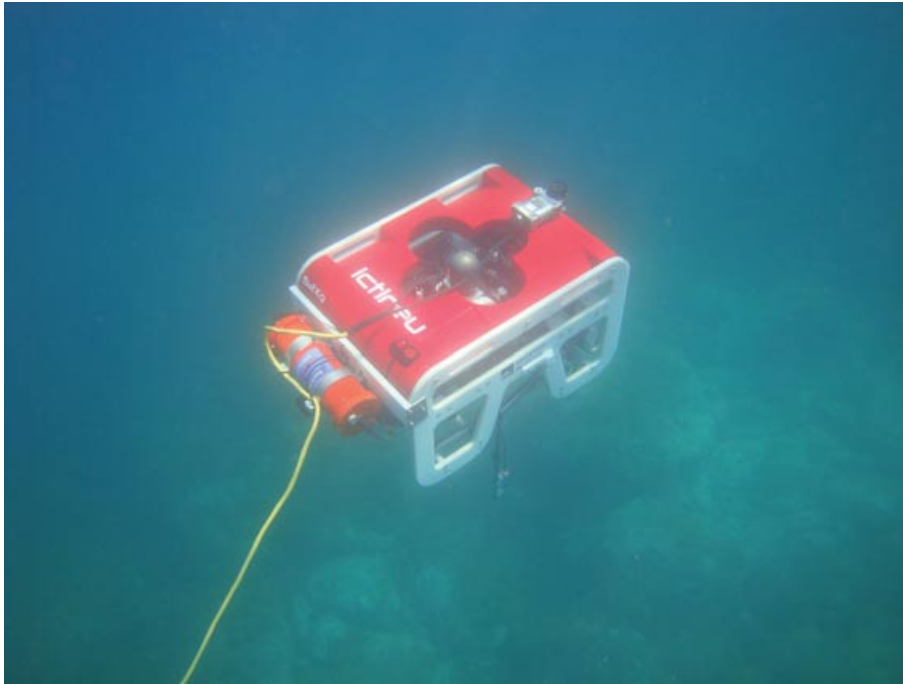
Figure 4. Snapshot of the unmanned underwater robot ICTINEU operating in the Mediterranean Sea during acquisition of the first data set. The robot carries a down-looking camera mounted on a bar.

fourth column. The fifth column denotes how many time epochs were achieved in the filtering step. The last column corresponds to the average reprojection error calculated using all the correspondences with the resulting set of homographies for each tested strategy. Figure 5 shows the final trajectory, and the overlapping image pairs are given in Figure 6. The final mosaic is illustrated in Figure 7.

It can be seen from Table II that, among all strategies, the overlap weighted OMI strategy produces a minimum reprojection error and is also closest to BA. For random ordering strategy, we have executed the proposal framework several times, and the values provided here are average values of the executions. One could conclude that the random strategy has performed well compared to the other

strategies. However, the performance of the random strategy is influenced by the particular trajectory. The trajectory is composed of several overlapping transects, for which the possibility of having overlap between any random image pair is higher than for other cases.

The order of successful observations makes a difference and has a large impact on the resulting trajectory. The OMI selection strategy required the largest total number of image matching attempts. Because of the high uncertainty of the state and the identity state vectors, especially after the initialization step, the generated potential observation list in the first time epoch had the same number of entries as the all-against-all list. Actually, several entries of the list did not have any overlapping areas. Therefore, during the first

Table II. Summary of results for the first data set.

| Strategy	Successful obs.	Unsuccessful obs.	% of attempts with respect to all-against-all	Iterations until stop	Final avg. error in pixels
1. Expected overlap	5,319.00	1,182.00	7.05	65.00	6.07
2. Highest OMI	5,346.00	16,279.00	23.45	245.00	9.14
3. Overlap weighted OMI	5,337.00	3,341.00	9.41	70.00	5.85
4. Random order	5,337.87	2,550.47	8.55	73.26	5.89
5. Combined (1–3)	5,333.00	1,557.00	7.47	67.00	5.91
Bundle adjustment	5,412.00	86,823.00	100.00	—	5.38
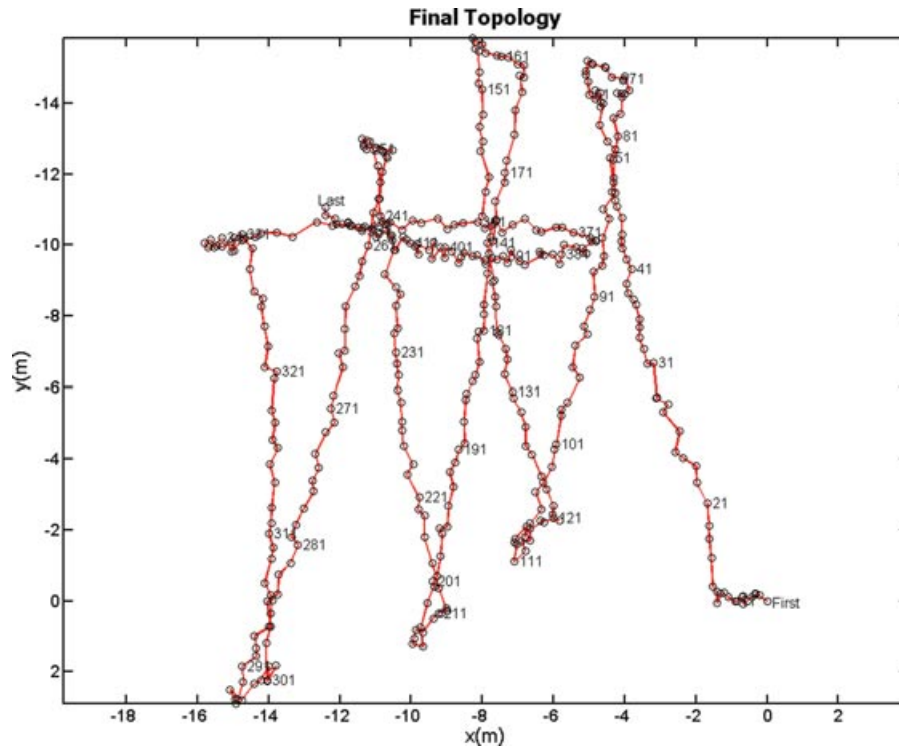
Figure 5. Final trajectory of the first data set. Total number of images was 430. Numbers denote the image centers. For clarity, the links between non-time-consecutive images are not represented.
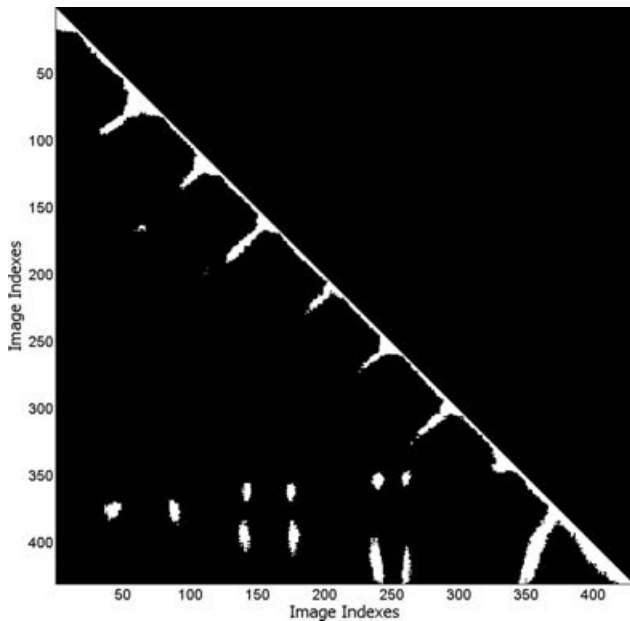

Figure 6. Overlapping image pairs for the first data set. The total number of overlapping pairs was 5,412, and the percentage with respect to all-against-all is 5.87.

iterations, the total number of successful observations in the OMI-based selection strategy was low because OMI selects the observations that provide the most information to the system. After the initialization step, uncertainty grows from the first image to the last image in time order. As all the images are mapped to the same position (identity state vector) in the first iteration, OMI chooses the observations that are composed of images closer to the first and last images. If the trajectory does not have a loop closing around those images, more image matching attempts are required by OMI. For these reasons, OMI attempts to match more image pairs than the other strategies. On the other hand, once the loop-closing image pairs have been detected, there is no need to continue choosing image pairs according to their OMI scores. If the trajectory provides an overlapping area between the first and last images, then one could expect the *combined* ranking criteria to achieve better image matching results. To illustrate this, we tested our framework on a trajectory including only one loop (see Figure 8). This second data set was composed of 30 images extracted from an underwater image sequence acquired by a Phantom 500 ROV during a survey in Andros, the Bahamas (Lirman, Gracias, Gintert, Gleason, Reid, et al., 2007). The results for the second data set are summarized in Table III.
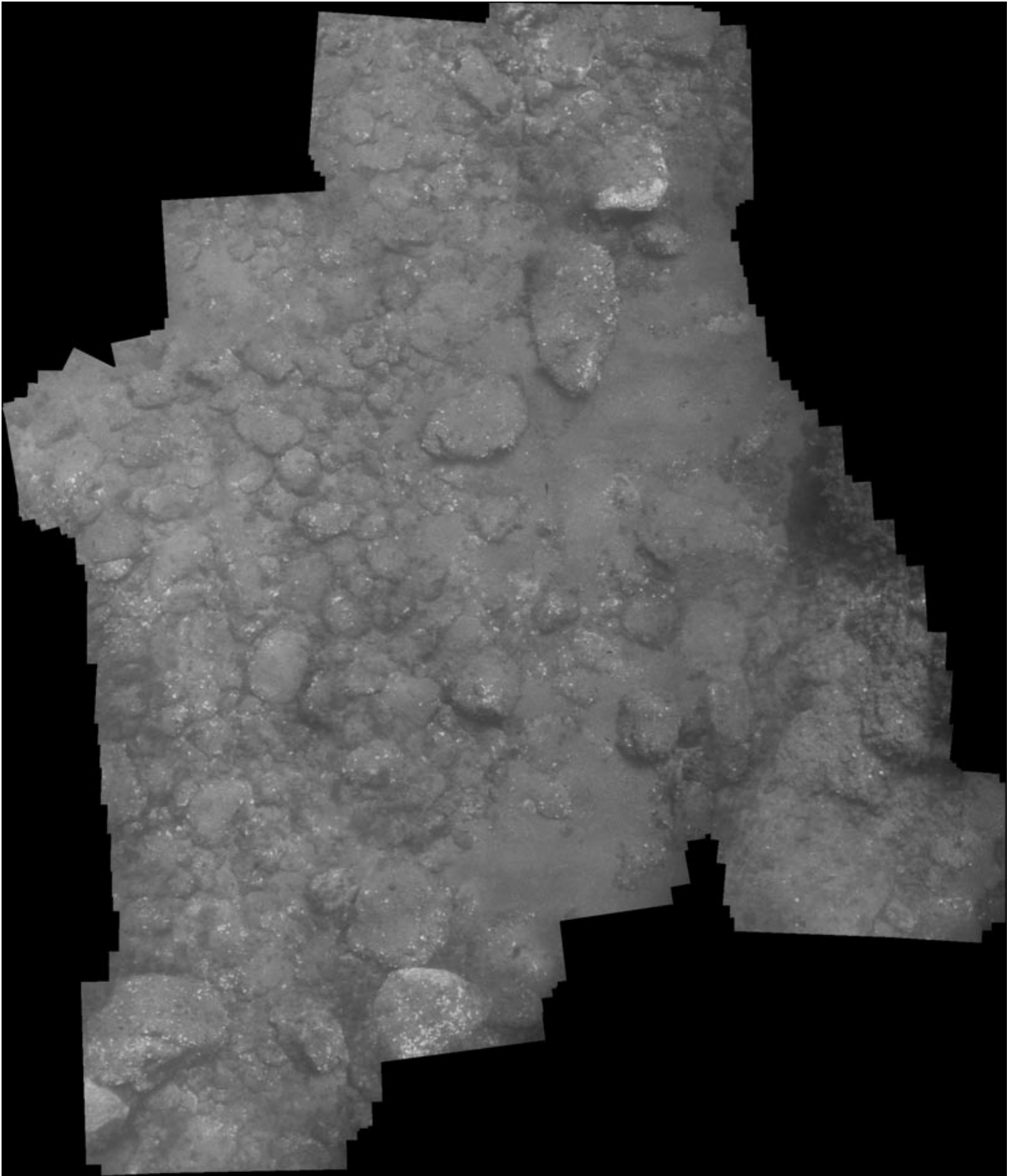
Figure 7. Final mosaic of the first data set, approximately 100 pixels per meter. After global alignment, the final mosaic was blended using a composition of gradient domain imaging and graph cut algorithms (Gracias, Mahoor, Negahdaripour, & Gleason, 2009; Prados, Neumann, Cufí, & Garcia, 2007).
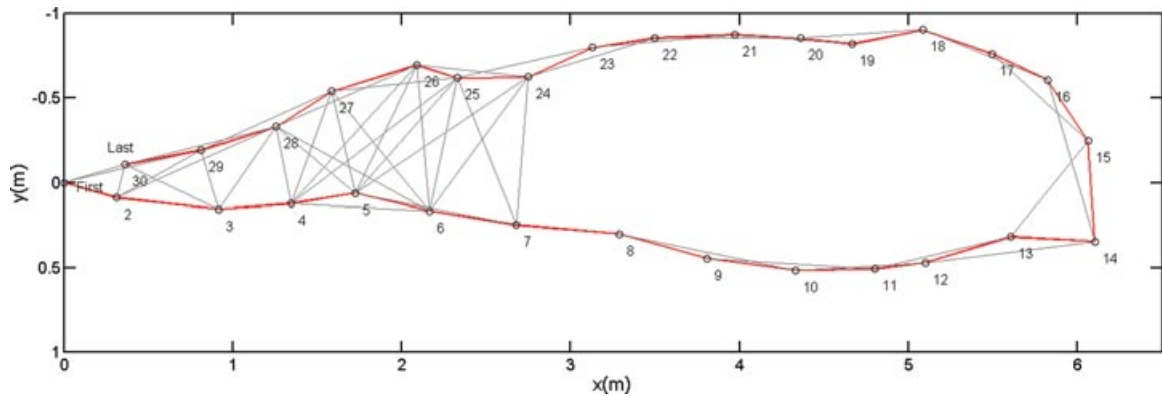
Figure 8. Final topology of the second data set. Numbers denote the image centers, and lines denote the overlapping image pairs. The total number of all overlapping pairs was 75, and the percentage with respect to all-against-all was 17.24.

As expected, the combined strategy performed better than the expected overlap strategy. It was able to obtain more successful image pairs with fewer attempts than the expected overlap strategy. In terms of trajectory accuracy, it also provided better results than the expected overlap.

When OMI-based selection criteria are used, Eq. (20) needs to be computed for each observation in the potential observation list. Therefore, its computational cost is higher than the expected overlap and random selection strategies. However, taking into account that the computational cost of the Kalman filter is much lower than minimizing the reprojection error using nonlinear optimization methods, the overall computational cost is lower than that of BA, which is commonly used in offline batch processing. Moreover, comparing to all-against-all image matching, the total number of image pairs that were attempted to match is much smaller, because our proposal takes into account the uncertainties in the image positions while generating the potential observation list. This also reduces the total time required for the complete topology estimation. To decide whether a pair of images is considered as a potential observation, it is necessary to use a threshold when generating the potential overlapping list. We tested expected overlap and combined ranking strategies with different threshold values to compare their performances and to evaluate the

effect of the threshold. We chose these two strategies as they performed better than all other strategies, according to the previous experiments.

For the third experiment, we extracted a set of images captured by an ROV surveying at approximately 2 m over a coral reef. Images were 512×384 pixels and acquired at 15 frames per second. The subset consisted of 80 images and covers approximately 53 m². The three-dimensional (3D) relief of the scene was negligible compared to the altitude of the robot. The trajectory had the shape of an eight with plenty of non-time-consecutive overlapping image pairs. Figure 9 shows the final topology for the third data set, and the results are summarized in Table IV.

For the higher threshold values, the combined strategy was able to find more overlapping pairs than the expected overlap. After initialization, time-consecutive pairs had high probabilities of having an overlap and the expected overlap ranking criterion selected them first. However, due to the nature of the Kalman filter, matching those images reduced the uncertainty but did not provide a good trajectory estimation. Therefore, if the threshold was high, overlapping image pairs between transects of the trajectory (i.e., loop-closing image pairs) were not detected or considered as a potential observation during the process of generating the potential observation list. Furthermore, as

Table III. Summary of results for the second data set.

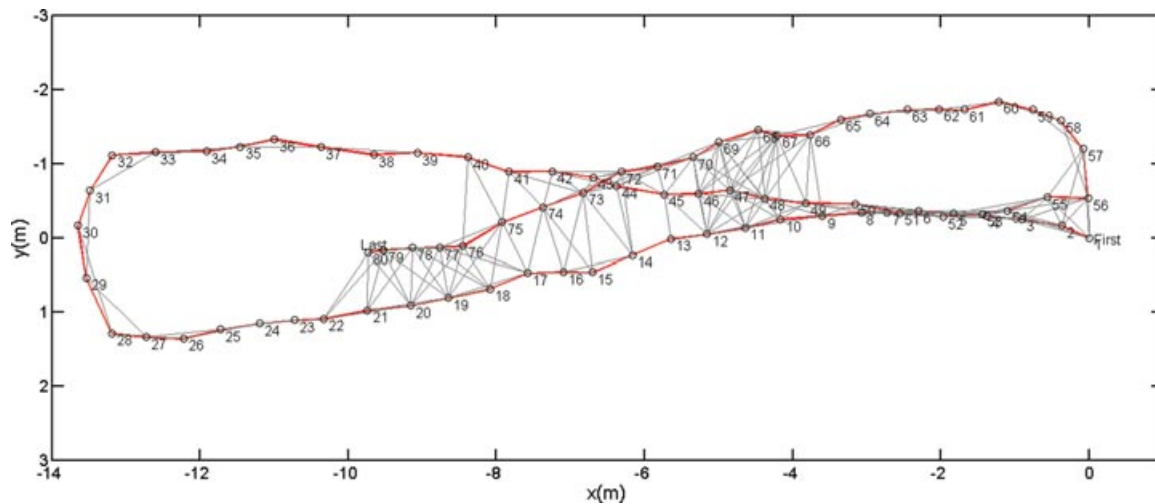| Strategy	Successful obs.	Unsuccessful obs.	% of attempts with respect to all-against-all	Iterations until stop	Final avg. error in pixels
1. Expected overlap	73.0	50.0	28.28	12.0	8.19
2. Highest OMI	75.0	158.0	53.56	20.0	6.90
3. Overlap weighted OMI	74.0	67.0	32.41	14.0	6.99
4. Random order	74.7	110.9	42.67	18.9	7.36
5. Combined (1–3)	74.0	48.0	28.05	12.0	7.13
Bundle adjustment	75.0	360.0	100.00	N.A.	6.78

Figure 9. Final topology of the third data set. Numbers denote the image centers, and lines denote the overlapping image pairs. The total number of all overlapping pairs was 262, and the percentage with respect to all-against-all was 8.29.

Table IV. Comparison of expected overlap and combined strategy for different threshold values.

Strategy	Threshold	Successful obs.	Unsuccessful obs.	Iterations	Avg. error in pixels
Expected overlap	0.5	81	0	3	507.97
Combined		119	109	8	12.13
Expected overlap	0.4	97	2	5	532.72
Combined		170	129	13	8.48
Expected overlap	0.3	132	15	6	517.37
Combined		243	118	15	7.97
Expected overlap	0.2	138	44	7	520.81
Combined		261	244	20	7.88
Expected overlap	0.1	172	194	17	288.99
Combined		262	411	23	7.89
Expected overlap	0.01	262	569	31	8.43
Combined		262	676	31	7.89

the combined strategy uses the expected overlap weighted OMI criterion for the first couple of iterations, it was able to detect some loop-closing image pairs. This resulted in a better trajectory estimation than matching only the time-consecutive ones. The combined strategy was able to get almost the whole topology (only one image pair was missing) for a threshold value of 0.2, and the total number of matching attempts was 505 of 3,160 possibilities. The expected overlap was able to obtain the whole topology successfully for a threshold of 0.01, and the total number of matching attempts was 831.

For the first data set, ranking observations with the expected overlap resulted in a accurate estimate of the topology with a minimum number of image matching attempts. However, the time-consecutive images have overlapping areas (Figure 5). Therefore, applying ASKF at the initializa-

tion step based on the prior of overlapping areas among time-consecutive images yielded a realistic modeling of the uncertainty of the trajectory. We have also tested this approach using a small data set in which there are nonoverlapping time-consecutive images. The data set had 29 images and consisted of two approximately parallel transects, with a few overlapping image pairs between transects (see Figure 10). It covered an area of 20 m^2. Results are summarized in Table V.

The expected overlap criterion failed to find the complete topology for any threshold value apart from zero.[5] This criterion chooses the highest expected overlap. Exploiting the information provided by the time-consecutive

---

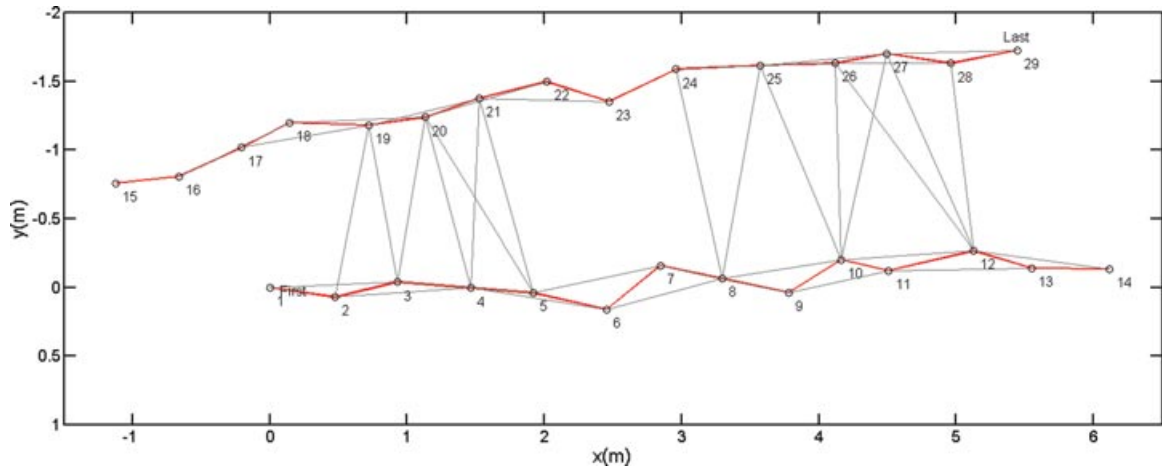[5] A threshold of 0 results in all-against-all matching.

Figure 10. Final topology of the fourth data set. Numbers denote the image centers, and lines denote the overlapping image pairs. The total number of all overlapping pairs is 64, and the percentage with respect to all-against-all is 15.76.

Table V. Summary of results for the fourth data set.

Strategy	Successful obs.	Unsuccessful obs.	% of attempts with respect to all-against-all	Iterations until stop	Final avg. error in pixels
1. Expected overlap	49	38	21.43	9	235.22
2. Highest OMI	64	224	70.94	25	12.27
3. Overlap weighted OMI	64	186	61.58	26	9.58
4. Random order	58.7	143.3	49.75	21	91.28
5. Combined (1–3)	64	185	61.33	26	8.21
Bundle adjustment	64	342	100.00	*N.A.*	6.63

images after the initialization step means that those images have a higher probability of being overlapping image pairs. Therefore, the expected overlap criterion tries to match them first. However, due to the nonoverlapping time-consecutive images between two transects of the trajectory, it failed to find the overlapping pairs between the transects. Such behavior is expected because no additional navigation information was used apart from the image data.

One of the advantages of the proposed approach is computational efficiency related to the naïve but robust approach of matching all images against all others. In the worst-case limit, our framework converges to the all-against-all strategy. However, we showed in the last experiment that our proposal is able to reduce the total number of matching attempts even if the assumption of overlapping time-consecutive images is violated. Our experiments highlight the importance of finding the most informative image pairs at the start of the search to reduce the drift and uncertainty of the trajectory for a low-cost vehicle equipped with only optical sensors. In later iterations, as the trajectory estimation gets closer to the real one and uncertainty on the trajectory reduces, there is no need to look for the

most informative image pairs. In general, we conclude that the combined strategy performs better than the other strategies tested here.

6. CONCLUSIONS

Cost and weight constraints mean that low-cost ROVs usually have a very limited number of sensors. When a low-cost ROV carries out a seafloor survey, it uses a down-looking camera and usually follows a predefined trajectory that provides several non-time-consecutive overlapping image pairs. Finding these pairs is one of the most important requirements for accurate trajectory estimation and quality mapping, which are necessary for the global view of the surveyed area. In this paper, we have presented an ASKF–EKF combined framework to estimate the topology with the minimum number of image matching attempts. The proposed framework allows the use of existing theory for estimation and control problems in batch mosaicking of large areas. All overlapping image pairs that are successfully matched contribute differently in terms of reducing uncertainty and reprojection error. We have proposed a novel and easy derivation to compute the OMI efficiently.

An important conclusion of this study is that it is important to decide which image pair is to be matched and when it is to be matched. In this context, different strategies for ordering image matching have been tested and their performances have been compared. Our future directions will be aimed at obtaining and extracting prior information from images to enable better uncertainty modeling of the trajectory in the initialization step.

7. APPENDIX: OBSERVATION MODEL

We assume that the robot has a down-looking camera and the camera optical axis is orthogonal to the scene, which is approximately flat. We also assume that the camera keeps a constant distance to the plane of the scene. The camera has been intrinsically calibrated, from which the 3×3 intrinsic parameter matrix was obtained.

On the basis of the above, we model the image motion between any pair of images using four parameters, accounting for 1D rotation, 2D translation, and 1D scaling. The image-to-image homography has the general form

$$H = \begin{bmatrix} a & -b & c \\ b & a & d \\ 0 & 0 & 1 \end{bmatrix}.$$

7.1. Noise Model for the Observations

We perform image matching between pairs of images using SIFT (Lowe, 2004) to detect features and RANSAC (Fischler & Bolles, 1981) to reject outliers and calculate the motion. The outcome of this process is two lists of image points that contain the image projections of the same set of world points. We follow the standard assumption that the uncertainty in the matching can be modeled as additive zero-mean Gaussian noise that corrupts one of the lists of point coordinates. Therefore, for a pair of matched points $[u_i^* \ v_i^*]$ and $[u_j \ v_j]$, in image coordinates, we have

$$\begin{bmatrix} u_i^* \\ v_i^* \end{bmatrix} = H_{(1:2,1:2)} \cdot \begin{bmatrix} u_j \\ v_j \end{bmatrix} + H_{(1:2,3)} + \eta,$$

where $\eta \sim N(0, \Sigma_\eta)$.

We assume that $\Sigma_\eta = \sigma^2 \cdot I_2$, where σ is the noise standard deviation in pixels and I_2 is a 2×2 identity matrix. The value of σ was obtained experimentally with underwater data, to account for the inevitable departures from the planarity assumption, because the 3D structure of the scene will induce noise. Experiments with real data showed that this structure for Σ_η is an adequate approximation. Olson (2002) described a method for point matching using correlation where Σ_η has the most general structure.

7.2. Obtaining the Observation Distribution

We model the observations as a Gaussian random variable $z \sim N(\bar{z}, \Sigma_z)$, where $z = [a \ b \ c \ d]^T$ contains the four parameters of the normalized homography H, defined above.

The observation mean $\bar{z}$ is obtained through minimization of a least-squares criterion:

$$\bar{z} = \arg \min \sum_n \left(\begin{bmatrix} u_i^* \\ v_i^* \end{bmatrix} - \begin{bmatrix} a & -b & c \\ b & a & d \end{bmatrix} \cdot \begin{bmatrix} u_j \\ v_j \\ 1 \end{bmatrix} \right)^2,$$

where the lists of point matches were obtained by robust feature matching using the similarity motion model. Details can be found in Gracias and Santos-Victor (2000). The observation covariance Σ_z is obtained by a first-order approximation for the covariance propagation of small noise in implicit functions (Haralick, 1998):

$$\Sigma_z \cong \left(\frac{\partial G}{\partial z} \right)^{-1} \cdot \left(\frac{\partial G}{\partial u} \right)^T \cdot \Sigma_u \cdot \frac{\partial G}{\partial u} \cdot \left(\frac{\partial G}{\partial z} \right)^{-1},$$

where

$G = \dfrac{\partial F}{\partial z}$ (Jacobian of the cost function with respect to parameters),

$F = R^T R$ (cost function – sum of squared residues),

$R = \begin{bmatrix} r_1 \\ \vdots \\ r_N \end{bmatrix}$ (residues vector),

$r_n = \begin{bmatrix} u_i^* \\ v_i^* \end{bmatrix} - \begin{bmatrix} a & -b & c \\ b & a & d \end{bmatrix} \cdot \begin{bmatrix} u_j \\ v_j \\ 1 \end{bmatrix}$ (point residue).

The simple structure of the image motion model allows for compact closed-form expressions for the above entities.

REFERENCES

AC-ROV. (2009). Retrieved September 15, 2009, from http://www.ac-cess.com.

Andersen, E. D., & Taylor, C. N. (2007, November). Improving MAV pose estimation using visual information. In IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA (pp. 3745–3750).

Anderson, B., & Moore, J. (1979). Optimal filtering. Englewood Cliffs, NJ: Prentice-Hall.

Bar-Shalom, Y., Li, R. X., & Kirubarajan, T. (2001). Estimation with applications to tracking and navigation. New York: Wiley-Interscience.

Bay, H., Tuytelaars, T., & Van Gool, L. J. (2006, May). SURF: Speeded up robust features. In European Conference on Computer Vision, Graz, Austria (pp. 404–417).

Beevers, K. R., & Huang, W. H. (2005, April). Loop closing in topological maps. In IEEE International Conference on Robotics and Automation, Barcelona, Spain (pp. 4367–4372).

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

Caballero, F., Merino, L., Ferruz, J., & Ollero, A. (2007, April). Homography based Kalman filter for mosaic building. Applications to UAV position estimation. In IEEE International Conference on Robotics and Automation, Rome (pp. 2004–2009).

Capel, D. (2004). Image mosaicing and super-resolution. London: Springer-Verlag.

Cheng, C., & Han, W. (2006, May). Large-scale loop-closing with pictorial matching. In IEEE International Conference on Robotics and Automation, Orlando, FL (pp. 1194–1199).

Clemente, L., Davison, A., Reid, I., Neira, J., & Tardós, J. (2007, June). Mapping large loops with a single hand-held camera. In Robotics: Science and Systems, Atlanta, GA.

Coleman, T. F., & Li, Y. (1996). An interior trust region approach for nonlinear minimization subject to bounds. SIAM Journal on Optimization, 6(2), 418–445.

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

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

Denzler, J., & Brown, C. (2002). An information theoretic approach to optimal sensor data selection for state estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(2), 145–157.

Duan, Z., Han, C., & Tao, T. (2004, October). Optimal multisensor fusion target tracking with correlated measurement noises. In IEEE International Conference on Systems, Man and Cybernetics (vol. 2, pp. 1272–1278).

Duffin, K., & Barrett, W. (1998, June). Globally optimal image mosaics. In Proceedings of Graphics Interface, Vancouver, Canada (pp. 217–222).

Elibol, A., Garcia, R., Delaunoy, O., & Gracias, N. (2008, December). A new global alignment method for feature based image mosaicing. In Advances in Visual Computing, Las Vegas, NV (pp. 257–266).

Elibol, A., Gracias, N., & Garcia, R. (2009, June). Match selection in batch mosaicing using mutual information. In 4th Iberian Conference on Pattern Recognition and Image Analysis, Povoa de Varzim, Portugal (pp. 104–111).

Ertin, E., Fisher, J., & Potter, L. (2003, April). Maximum mutual information principle for dynamic sensor query problems. In Information Processing in Sensor Networks, Palo Alto, CA (pp. 405–416).

Escartin, J., Garcia, R., Delaunoy, O., Ferrer, J., Gracias, N., Elibol, A., Cufi, X., Neumann, L., Fornari, D. J., Humpris, S. E., & Renard, J. (2008). Globally aligned photomosaic of the Lucky Strike hydrothermal vent field (Mid-Atlantic Ridge, 3718.5′N): Release of georeferenced data, mosaic construction, and viewing software. Geochemistry Geophysics Geosystems, 9(12), Q12009.

Eustice, R., Pizarro, O., & Singh, H. (2004, April). Visually augmented navigation in an unstructured environment using a delayed state history. In Proceedings of the 2004 IEEE International Conference on Robotics and Automation, New Orleans, LA (vol. 1, pp. 25–32).

Eustice, R. M., Singh, H., Leonard, J. J., & Walter, M. R. (2006). Visually mapping the RMS Titanic: Conservative covariance estimates for SLAM information filters. International Journal of Robotics Research, 25(12), 1223–1242.

Fengda, Z., Lingfu, K., & Li, X. (2007, July). Loop-closing by using SIFT features for mobile robots. In Chinese Control Conference, Hunan Province, China (pp. 221–224).

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

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

Fleischer, S., Rock, S., & Burton, R. (1997, October). Global position determination and vehicle path estimation from a vision sensor for real-time video mosaicking and navigation. In MTS/IEEE OCEANS Conference, Halifax, NS, Canada (vol. 1, pp. 641–647).

Gan, Q., & Harris, C. J. (2001). Comparison of two measurement fusion methods for Kalman-filter-based multisensor data fusion. IEEE Transactions on Aerospace and Electronic Systems, 37(1), 273–279.

Garcia, R., Puig, J., Ridao, P., & Cufí, X. (2002, May). Augmented state Kalman filtering for AUV navigation. In IEEE International Conference on Robotics and Automation, Washington, DC (vol. 3, pp. 4010–4015).

GNOM Standard ROV. (2009). Retrieved September 15, 2009, from http://www.gnom-rov.com/products/gnom-standard/.

Golub, G., & van Loan, C. (1996). Matrix computations. Baltimore, MD: Johns Hopkins University Press.

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

Gracias, N., Mahoor, M., Negahdaripour, S., & Gleason, A. (2009). Fast image blending using watersheds and graph cuts. Image and Vision Computing, 27(5), 597–607.

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

Grocholsky, B. (2002). Information-theoretic control of multiple sensor platforms. Ph.D. thesis, University of Sydney.

Hall, D. L., & McMullen, S. A. H. (2004). Mathematical techniques in multisensor data fusion. Norwood, MA: Artech House.

Haralick, R. M. (1998, March). Propagating covariance in computer vision. In 9. Theoretical foundations of computer vision, Dagstuhl, Germany (pp. 95–114).

Ho, K. L., & Newman, P. (2007). Detecting loop closure with scene sequences. International Journal of Computer Vision, 74(3), 261–286.

Ila, V., Andrade-Cetto, J., Valencia, R., & Sanfeliu, A. (2007, November). Vision-based loop closing for delayed state robot mapping. In IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA (pp. 3892–3897).

Ila, V., Porta, J. M., & Andrade-Cetto, J. (2010). Information-based compact pose SLAM. IEEE Transactions on Robotics, 26(1), 78–93.

Jin, X. B., & Sun, Y. X. (2003, June). Optimal centralized state fusion estimation for multi-sensor system with correlated measurement noise. In Proceedings of 2003 IEEE Conference on Control Applications, Istanbul, Turkey (vol. 1, pp. 770–772).

Jungho, K., & In-So, K. (2007, November). Robust feature matching for loop closing and localization. In IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA (pp. 3905–3910).

Kangsheng, T., & Guangxi, Z. (2006). Sensor management based on Fisher information gain. Journal of Systems Engineering and Electronics, 17(3), 531–534.

Karan, M., & Krishnamurthy, V. (2003, July). Sensor scheduling with active and image-based sensors for maneuvering targets. In Proceedings of the Sixth International Conference of Information Fusion, Cairns, Queensland, Australia (vol. 2, pp. 1018–1023).

Krishnamurthy, V. (2002). Algorithms for optimal scheduling and management of hidden Markov model sensors. IEEE Transactions on Signal Processing, 50(6), 1382–1397.

Leone, A., Distante, C., Mastrolia, A., & Indiverr, G. (2006, September). A fully automated approach for underwater mosaicking. In MTS/IEEE OCEANS Conference, Boston, MA (pp. 1–6).

Lirman, D., Gracias, N., Gintert, B., Gleason, A., Reid, R. P., Negahdaripour, S., & Kramer, P. (2007). Development and application of a video–mosaic survey technology to document the status of coral reef communities. Environmental Monitoring and Assessment, 159, 59–73.

Logothetis, A., & Isaksson, A. (1999, June). On sensor scheduling via information theoretic criteria. In Proceedings of the American Control Conference, San Diego, CA (pp. 2402–2406).

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

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

Mahon, I., Williams, S., Pizarro, O., & Johnson-Roberson, M. (2008). Efficient view-based SLAM using visual loop closures. IEEE Transactions on Robotics, 24(5), 1002–1014.

Manyika, J., & Durrant-Whyte, H. F. (1994). Data fusion and sensor management: An information-theoretic approach. Upper Saddle River, NJ: Prentice Hall.

Marks, R. L., Rock, S. M., & Lee, M. J. (1995). Real-time video mosaicking of the ocean floor. IEEE Journal of Oceanic Engineering, 20, 229–241.

Mattikalli, R., Fresnedo, R., Frank, P., Locke, S., & Thunemann, Z. (2007, September). Optimal sensor selection and placement for perimeter defense. In IEEE International Conference on Automation Science and Engineering, CASE 2007, Scottsdale, AZ.

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

Mutambura, A. G. O. (1998). Decentralised estimation and control for multisensor systems. Boca Raton, FL: CRC Press.

Negahdaripour, S., & Firoozfam, P. (2001, November). Positioning and photo-mosaicking with long image sequences; Comparison of selected methods. In OCEANS, 2001. MTS/IEEE Conference and Exhibition, Honolulu, HI (vol. 4, pp. 2584–2592).

Nemhauser, G., & Wolsey, L. (1988). Integer and combinatorial optimization. New York: Wiley-Interscience.

Newman, P. M., Leonard, J. J., & Rikoski, R. R. (2003, October). Towards constant-time SLAM on an autonomous underwater vehicle using synthetic aperture sonar. In International Symposium on Robotics Research, Sienna, Italy (pp. 409–420).

Ng, G. W., & Ng, K. H. (2000). Sensor management what, why and how. Information Fusion, 1(2), 67–75.

Olson, C. F. (2002). Maximum-likelihood image matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(6), 853–857.

Oshman, Y. (1994). Optimal sensor selection strategy for discrete-time state estimators. IEEE Transactions on Aerospace and Electronic Systems, 30(2), 307–314.

Pahalawatta, P. V., Pappas, T. N., & Katsaggelos, A. K. (2004, October). Optimal sensor selection for video-based target tracking in a wireless sensor network. In IEEE International Conference on Image Processing, Singapore (vol. 5, pp. 3073–3076).

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

Pizarro, O., Eustice, R., & Singh, H. (2009). Large area 3D reconstructions from underwater optical surveys. IEEE Journal of Oceanic Engineering, 34(2), 150–169.

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

Prados, R., Neumann, L., Cufí, X., & Garcia, R. (2007). Visually pleasant blending techniques in underwater mosaicing. Instrumentation Viewpoint, (6), 98–99.

Proteus 500 ROV. (2009). Retrieved September 15, 2009, from http://www.hydroacousticsinc.com/marine_technology.php.

Ribas, D., Neira, J., Ridao, P., & Tardos, J. D. (2006, October). SLAM using an imaging sonar for partially structured environments. In IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China.

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

Richmond, K., & Rock, S. M. (2006, September). An operational real-time large-scale visual mosaicking and navigation system. In MTS/IEEE OCEANS Conference, Boston, MA.

Ruiz, I., de Raucourt, S., Petillot, Y., & Lane, D. (2004). Concurrent mapping and localization using sidescan sonar. IEEE Journal of Oceanic Engineering, 29(2), 442–456.

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

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

SeaBotix LBV150BE MiniROV. (2009). Retrieved September 15, 2009, from http://www.seabotix.com/products/lbv150be_features.htm.

Singh, H., Roman, C., Pizarro, O., Eustice, R., & Can, A. (2007). Towards high-resolution imaging from underwater vehicles. International Journal of Robotics Research, 26(1):55–74.

SM 1000 Low Cost ROV System. (2009). Retrieved September 15, 2009, from http://www.sub-find.com/sm_1000.htm.

Sun, S. L. (2004). Multi-sensor optimal information fusion Kalman filters with applications. Aerospace Science and Technology, 8, 57–62.

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

Triggs, B., McLauchlan, P., Hartley, R., & Fitzgibbon, A. (2000, September). Bundle adjustment—A modern synthesis. In Vision Algorithms: Theory and Practice, Corfu, Greece (pp. 298–375). Springer Verlag.

Unnikrishnan, R., & Kelly, A. (2002, May). Mosaicing large cyclic environments for visual navigation in autonomous vehicles. In IEEE International Conference on Robotics and Automation, Washington, DC (pp. 4299–4306).

Vincent, A., Pessel, N., Borgetto, M., Jouffroy, J., Opderbecke, J., & Rigaud, V. (2003, September). Real-time geo-referenced video mosaicking with the MATISSE system. In MTS/IEEE OCEANS Conference, San Diego, CA (vol. 4, pp. 2319–2324).

Whaite, P., & Ferrie, F. P. (1997). Autonomous exploration: Driven by uncertainty. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(3), 193–205.

Xiong, N., & Svensson, P. (2002). Multi-sensor management for information fusion: Issues and approaches. Information Fusion, 3(2), 163–186.

Yao, J., & Chamb, W. K. (2007). Robust multi-view feature matching from multiple unordered views. Pattern Recognition, 40(11), 3081–3099.

Yukun, C., Xicai, S., & Zhigang, L. (2007). Research on Kalman-filter based multi sensor data fusion. Journal of Systems Engineering and Electronics, 18(3), 497–502.

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