ESTIMATING ATTITUDE AND TRAJECTORY OF FORWARD LOOKING IMAGING SONAR USING INTER-FRAME REGISTRATION

Benjamin Henson[a], Yuriy Zakharov[a]

[a]Department of Electronic Engineering, University of York, York, UK. YO10 5DD.

Contact: Department of Electronic Engineering, University of York, York, UK. YO10 5DD.
e-mail: bth502@york.ac.uk

Abstract: *The aim of this work is to produce an accurate attitude-trajectory estimate for a Forward Looking Sonar (FLS) based on the optical flow between consecutive sonar frames. The attitude-trajectory can be used to locate an underwater platform, such as an Autonomous Underwater Vehicle (AUV), to a degree of accuracy suitable for navigation. The attitude-trajectory estimation is performed in three steps. Firstly, a selection of optical flow estimation techniques are applied to estimate a pixel displacement map between consecutive sonar frames in the native polar (range/bearing) format. From this displacement map, a reduced set of comparative statistics are produced. The second step finds the best match for the displacement map statistics when compared to displacement maps for a set of modelled sonar sensor motions within an expected motion range. Thus, from the model, an estimate of the incremental sensor motion between frames is made. Finally, using a weighted regularised spline technique to integrate the incremental motion, an attitude-trajectory is generated for the sonar sensor. To assess the accuracy of the attitude-trajectory estimate, the FLS frames are registered to a global Cartesian grid, building a mosaic of the underwater scene.*

Keywords: *Forward Looking Sonar, registration, mosaic, navigation.*

1. INTRODUCTION

With high resolution forward looking sonars (FLSs) there has been an interest in using techniques akin to the optical vision systems for navigation underwater [1] [2]. In such systems the navigation is based on the inter-frame registration. The sonar sensor motion estimation can also be based on comparison/registration of pairs of sonar frames [1] [2] [3]

[4]. There are however problems when applying optical registration techniques to FLS images. Such images usually have a low resolution and typically have a low signal to noise ratio (SNR). With a single illumination point and the native polar format of FLS frames, for relatively simple sonar sensor motions the map of pixel movements can become complicated [5]. An efficient approach for optical image registration (pixel displacement estimation) is based on the *optical flow*, which relates to the brightness variation within a scene and in time [6]. This paper proposes an algorithm to estimate the motion of an underwater platform based on the *optical flow* between consecutive FLS frames. The work in this paper proposes a method for estimating the pixel displacement map (DM) between a pair of FLS frames. This DM is then converted into the sonar sensor inter-frame motion, which is based on representing a DM via a small set of statistics. This set is compared to statistics pre-computed for modelled motions within an expected motion range. A method is further proposed that uses weighted regularised splines to integrate the inter-frame motions into an attitude-trajectory estimate for the sonar sensor. To validate the attitude-trajectory estimate, a mosaic of an underwater scene was built from a real in-field FLS dataset [7].

2. ESTIMATION OF INTER-FRAME DM

The estimation of the inter-frame DM is based on a selection of *optical flow* estimation techniques applied to two frames represented in the polar (range/bearing) format. The processing is divided into the coarse (pixel-wide) and fine (sub-pixel) displacement estimation.

Coarse displacement estimation - Signal model and sparse recovery. For inter-frame registration we consider a reference and a target frame from the FLS. We assume that any pixel in the target image is a convolution of the reference image and an unknown pixel-specific kernel. The support of the kernel is defined by the dynamics of the underwater platform and the sonar frame rate. In this simplified model we consider all other distortion artefacts, such as revealed or occluded pixels, as measurement noise. The task is to identify the kernel at any given pixel location. To do this a square patch (13×13 pixels) is taken from the target frame in the vicinity of the pixel location of interest. This patch should be sufficiently large to allow kernel identification at low SNRs, but small enough to consider the kernel invariant. The measurement patch is part of a linear system of equations, where the reference frame is used to build a dictionary of pixel-wide patch motions. This system has a sparse solution with a small number of non-zero coefficients, estimated using the Orthogonal Matching Pursuit (OMP) [8]. The coefficient with the highest magnitude defines the coarse DM.

Coarse displacement estimation - Selection of sample points. To reduce the number of sample points ($^1/_{17}$ of the number of pixels in the frame), and hence the complexity, in our algorithm the reference image intensity is used as a probability density function for generating the sample points as described in [9] [10].

Coarse displacement estimation - Mode filter. A square aperture mode filter is used to propagate the sample values, whilst reducing outliers. This builds a DM for the whole frame.

Coarse displacement estimation - Forward and backward DM comparison. A DM is estimated for the motion in the forward direction (from the reference to target frame), and in the backward direction (from the target to reference frame). A comparison between these two estimated DMs allows validation of the registration because one should be the reverse of the other [11]. The comparison for every pixel position is made, if the magnitude of the sum of the forward/backward displacements is less than a threshold then the forward estimate is retained. The pixel locations where the displacement estimate is not retained are interpolated, taking the value of the nearest (in terms of pixels) accepted estimate. Thus, an estimate of the coarse displacement is associated with each pixel location in the reference frame.

Fine displacement estimation. Adaptive filters have been widely used in signal processing applications for system identification [12], and so are a good candidate technique to apply to the *optical flow* problem [13] [14] [15] [16]. Here an adaptive filter is used to identify the convolution kernel, related to the pixel motion, to a sub-pixel precision. The adaptive filter algorithm used is an exponentially weighted recursive least squares (ERLS) [12]. The adaptive filter works most effectively if the change in the convolution kernel from one adaptive iteration to the next is a slow evolution. To achieve this, the pixel order is altered. A permutation table is created using an order derived from a pseudo-Hilbert space filling curve [10] and the DM. The input to the adaptive filter is a sequence of square pixel apertures (regressor vectors) from the reference frame, in the order determined by the permutation table. For each pixel position, the adaptive filter produces an estimate of the convolution kernel (7×7 coefficients) [17] [12]. The fine pixel displacement is estimated using interpolation, assuming a continuous 2-D convolution kernel [18]. An initial fine DM is then created from the sum of the coarse and fine pixel estimates, to which a median filter is applied to produce the fine DM.

3. ESTIMATION OF THE INTER-FRAME SONAR SENSOR MOTION

The fine displacement estimation has produced the DM $\bar{d}(\xi)$, where $\xi = [\xi^{(\psi)}, \xi^{(r)}]$ is a pixel position (beam angle and range, respectively) in the reference frame. The algorithm described in this section estimates an inter-frame motion vector $\boldsymbol{\alpha}$ using $\bar{d}(\xi)$. Firstly, we describe a preprocessing of $\bar{d}(\xi)$ to select an area X_B of the frame that contains information relevant to the dominant motion of the sensor. We then introduce the motion model and establish the relationship between the motion parameters ($\boldsymbol{\alpha}$) and the DM. The motion estimation is then formulated as a least squares (LS) problem. This LS problem can be solved using standard optimization techniques. However, a significantly more efficient method for real-time processing is firstly to transform the DM into a small vector of auxiliary statistics, and then to apply a dichotomous coordinate descent (DCD) search to match the statistics with a set of precomputed modelled 'statistics'.

Preprocessing of the fine DM. The purpose of the preprocessing is to remove unreliable parts of the DM. The first of two steps is a thresholding of the reference frame, this is based on computing a histogram of intensity, and choosing a predefined quantile, for example 35%. The second step is based on processing the DM only at pixels kept after the first step. In this processing, a histogram for magnitudes of the displacements is computed, to which Tukey's rule [19] is applied to identify outliers. The outlier pixels are also removed from further processing, thus finally identifying a reliable set X_B on the reference frame.

The motion model. There are six possible degrees of motion of the sonar sensor, which are translations in axes x, y, z and rotations around each axis. We assume that the altitude z is constant and predefined. The platform roll (rotation around x) and pitch (rotation around y) are considered negligible. Therefore, we consider only translations Δ_x and Δ_y in x and y and rotation Δ_θ around z for estimation: $\boldsymbol{\alpha} = [\Delta_x, \Delta_y, \Delta_\theta]$. The displacement is described by a model $d_{\text{model}}(\xi, \boldsymbol{\alpha}) = \left[\xi^{(\psi)} - \xi_t(\boldsymbol{\alpha})^{(\psi)}, \xi^{(r)} - \xi_t(\boldsymbol{\alpha})^{(r)}\right]$, where $\xi_t(\boldsymbol{\alpha}) = [\xi_t(\boldsymbol{\alpha})^{(\psi)}, \xi_t(\boldsymbol{\alpha})^{(r)}]$ is the new pixel position after the modelled motion and $\xi \in X_B$. The pixel position is transformed from polar coordinates in the reference frame to Cartesian coordinates on the seabed. The motion $\boldsymbol{\alpha}$ is transformed into a new position on the seabed. The new position is then transformed back to polar coordinates $\xi_t(\boldsymbol{\alpha})$.

Estimation of the motion vector $\boldsymbol{\alpha}$. The motion can be estimated by solving the LS optimization problem: $\hat{\boldsymbol{\alpha}}_{\text{LS}} = \arg\min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha})$, where the LS cost function is given by

$J(\boldsymbol{\alpha}) = \sum_{\xi \in X_B} \left\| \bar{d}(\xi) - d_{\text{model}}(\xi, \boldsymbol{\alpha}) \right\|^2$. Obtaining the solution to the LS problem has high complexity because, for an iterative LS solver, a model DM must be regenerated multiple times using the complicated non-linear transforms for every pixel $\xi_t(\boldsymbol{\alpha}) \in X_B$.

DM dimension reduction and DCD search. To reduce the complexity of the inter-frame motion estimator, the fine DM $\bar{d}(\xi)$, $\xi \in X_B$, is represented by a small number of statistics (12 in our example) as follows. Four histograms are computed from the DM: $\boldsymbol{G}_{\pm}^{(j)} = \text{hist}\{\bar{d}^{(j)}(\xi), \xi \in X_B^{\pm}\}$, j=1,2, where X_B^{+} and X_B^{-} denote parts of X_B for positive and negative beam angles and $\bar{d}^{(j)}(\xi)$ denotes the j^{th} coordinate of vector position $\bar{d}(\xi)$. For each of the histograms, three quartiles are computed, thus twelve parameters in total; we denote all of these quartiles as a vector $\boldsymbol{s}$. The vector $\boldsymbol{s}$ is then compared with vectors $\boldsymbol{s}_{\text{model}}(\boldsymbol{\alpha})$ computed on a grid of motions $\boldsymbol{\alpha} \in T$ to find the best match, where T is a grid of sensor motions bounded by the dynamics of the platform. Specifically, the modelled DMs for all possible sensor motions $\boldsymbol{\alpha} \in T$ are pre-computed, and for each of them we store the twelve 'statistics' as was described above. The match is based on the minimization of the Euclidean distance $J_s(\boldsymbol{\alpha}) = \|\boldsymbol{s} - \boldsymbol{s}_{\text{model}}(\boldsymbol{\alpha})\|^2$: $\widehat{\boldsymbol{\alpha}}_S = \arg\min_{\boldsymbol{\alpha} \in T} J_s(\boldsymbol{\alpha})$. The solution of this minimization problem can be found using an exhaustive search over the grid. In our example, the cardinality of T is $|T| = 97 \times 256 \times 301 \approx 8 \times 10^6$ which results in approximately 10^8 addition and multiplication operations to find $\widehat{\boldsymbol{\alpha}}_S$. To further reduce the complexity, we propose to use the DCD search [17] on the grid T. With the DCD search, the complexity is $12\log_2|T| \approx 275$ multiplications and additions. The result of the processing described above when applied to all frames in a sequence of P frames is a sequence of motion vectors $\{\boldsymbol{\alpha}_k\}_{k=1}^{P-1}$. We also retain the values $l_k = J_s(\boldsymbol{\alpha}_k)$. These will be transformed into a set of weights $\{w_k\}_{k=1}^{P-1}$ that characterize the accuracy of the inter-frame motion estimation.

4. ATTITUDE-TRAJECTORY ESTIMATION

The aim is now to produce an estimate of the sonar sensor attitude and position at any time t within an experiment. The incremental movements $\boldsymbol{\alpha}_k = [\Delta_x(t_k), \Delta_y(t_k), \Delta_\theta(t_k)]$ need to be processed to produce positions and attitudes $\boldsymbol{p}(t) = [x(t), y(t), \theta(t)]$ that are in a coordinate system fixed on the seabed; where $t_k = kT_{\text{frame}}$ and T_{frame} is the frame interval. Firstly, we represent $\theta(t)$ as a smoothed spline found as a trade-off between an error in the fit to the data points and smoothness of the spline. One efficient way is to use P-splines [20]. The spline is given by $\theta(t) = \sum_{m=1}^{N_b} c_m B(t - (m-1)\tau)$, $B(t)$ is a cubic B-spline [21], N_b the number of basis functions, c_m are basis expansion coefficients, and in our example $\tau = T_{\text{frame}}/4$. The angular velocity $\theta'(t)$ is then given by $\theta'^{(t)} = \sum_{m=1}^{N_b} c_m b(t - (m-1)\tau)$, where $b(t) = B'(t)$. To find the spline coefficients c_m, a weighted LS optimization problem with a penalty is formulated as follows. For P-splines, the smoothness penalty is efficiently calculated using the difference in the values of the spline coefficients themselves [20]. The cost function takes the form $S = \sum_{k=1}^{P-1} w_k \left[\frac{1}{T_{\text{frame}}} \Delta_\theta(t_k) - \theta'(t_k)\right]^2 + \mu \sum_{m=n+1}^{N_b} (\Delta^n c_m)^2$, where $\mu > 0$ is a regularization parameter, Δ^n is the n^{th} difference operator. In our example, we use $n = 2$, i.e. $\Delta^2 c_m = c_m - 2c_{m-1} + c_{m-2}$. Having obtained the spline $\theta(t)$ we are able now to rotate the incremental motions $\Delta_x(t_k)$, $\Delta_y(t_k)$ onto the seabed coordinate system. The same spline integration procedure as described above is then applied to find the smoothed splines $x(t)$ and $y(t)$. The dynamics of the system can be incorporated into the smoothing regularisation, for instance the regularising parameter μ can be increased until the attitude-trajectory conforms to a maximum expected acceleration of the platform.

5. DAM INSPECTION EXAMPLE DATASET

The example dataset is from the inspection of a dam wall, [7] [22]. Fig. 1 shows a mosaic of 1596 frames from a single track along the dam wall. Using the sensor's estimated attitude-trajectory, the interpolated pixel from each frame is projected onto the seabed with reference to a fixed Cartesian coordinate system. It can be seen that using the estimated attitude-trajectory for the sonar sensor a coherent mosaic can be produced, with small details in sections of the image showing good registration.

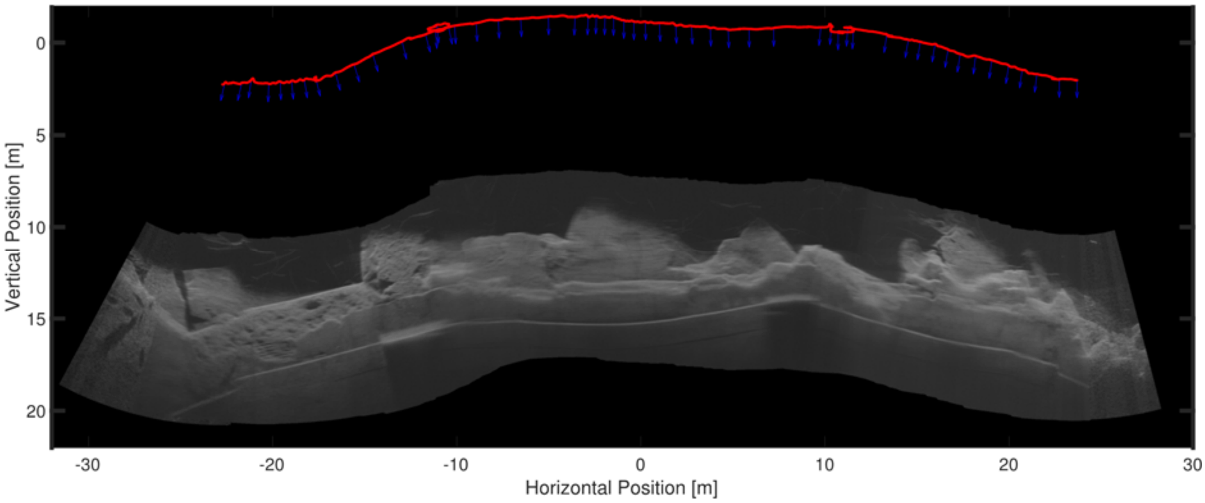

Fig. 1 A mosaic of 1596 frames showing a track along a dam wall (track motion is predominantly in negative y direction). The sonar sensor trajectory is shown in red and the attitude at every 30th frame is shown as a blue arrows

6. CONCLUSION

In this paper, we have proposed a method for estimating the attitude-trajectory of a FLS sonar sensor. The proposed method initially estimates the displacement map (DM) that describes the motion of individual pixels between frames. By comparing the inter-frame DM with those generated analytically from modelled sonar sensor movements, a change in attitude and position for the sensor is estimated. The complexity of the DM comparison is reduced by representing each of the DMs with a set of summary statistics, and the search complexity is further reduced by means of a dichotomous search. The use of smoothing splines (P-splines) is proposed for the integration of the motion and smoothing the result with reference to the physical system dynamics. From this, an attitude-trajectory for the sonar sensor is estimated. In assessing the accuracy of the estimated attitude-trajectory for the sonar sensor, a coherent mosaic has been produced, with small details in sections of the image showing good registration.

7. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Luis A. Conti, University of São Paulo and Acquest Subaquatic Geology and Geophysics [22] for providing the sonar data.

REFERENCES

[1] **H. Johannsson, M. Kaess, B. Englot, F. Hover and J. Leonard**, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010.

[2] **N. Hurtós, D. Ribas, X. Cufí, Y. Petillot and J. Salvi**, "Fourier-based Registration for Robust Forward-looking Sonar Mosaicing in Low-visibility Underwater Environments," *Journal of Field Robotics,* vol. 32, pp. 123-151, 2015.

[3] **N. Hurtós, S. Nagappa, X. Cufí, Y. Petillot and J. Salvi**, "Evaluation of registration methods on two-dimensional forward-looking sonar imagery," in *MTS/IEEE OCEANS-Bergen*, 2013.

[4] **M. D. Aykin and S. Negahdaripour**, "On Feature Matching and Image Registration for Two-dimensional Forward-scan Sonar Imaging," *Journal of Field Robotics,* vol. 30, pp. 602-623, 2013.

[5] **S. Negahdaripour**, "On 3-D motion estimation from feature tracks in 2-D FS sonar video," *IEEE Transactions on Robotics,* vol. 29, pp. 1016-1030, 2013.

[6] **B. K. P. Horn and B. G. Schunck**, ""Determining optical flow": a retrospective," *Artificial Intelligence,* vol. 59, pp. 81-87, 1993.

[7] **L. Conti, M. Rodriques and B. Hanot**, "Hydroelectric Power Plant Inspections," *Hydro International Magazine,* 2016.

[8] **J. A. Tropp and A. C. Gilbert**, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory,* vol. 53, pp. 4655-4666, 2007.

[9] **C. Schretter and H. Niederreiter**, "A direct inversion method for non-uniform quasi-random point sequences," *Monte Carlo Methods and Applications,* vol. 19, pp. 1-9, 2013.

[10] **C.-C. Wu and Y.-I. Chang**, "Approximately even partition algorithm for coding the Hilbert curve of arbitrary-sized image," *IET Image Processing,* vol. 6, pp. 746-755, 2012.

[11] **C. Chailloux, J.-M. Le Caillec, D. Gueriot and B. Zerr**, "Intensity-based block matching algorithm for mosaicing sonar images," *IEEE Journal of Oceanic Engineering,* vol. 36, pp. 627-645, 2011.

[12] **S. Haykin**, Adaptive Filter Theory (2nd Ed.), Prentice-Hall, Inc., 1991.

[13] **M. R. Lynch and P. J. W. Rayner**, "A new approach to image registration utilising multidimensional LMS adaptive filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1988. ICASSP-88.*, 1988.

[14] **P. Elad and A. Feuer**, "Recursive optical flow estimation-adaptive filtering approach," in *IEEE Nineteenth Convention of Electrical and Electronics Engineers in Israel*, 1996.

[15] **G. Caner, A. M. Tekalp, G. Sharma and W. Heinzelman**, "Local image registration by adaptive filtering," *IEEE Transactions on Image Processing,* vol. 15, pp. 3053-3065, 2006.

[16] **B. Henson and Y. Zakharov**, "Local optical-flow estimation for forward looking imaging sonar data," in *MTS/IEEE OCEANS-Monterey*, 2016.

[17] **Y. V. Zakharov, G. P. White and J. Liu**, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Transactions on Signal Processing,* vol. 56, pp. 3150-3161, 2008.

[18] **J. Ren, J. Jiang and T. Vlachos**, "High-accuracy sub-pixel motion estimation from noisy images in Fourier domain," *IEEE Transactions on Image Processing,* vol. 19, pp. 1379-1384, 2010.

[19] **J. W. Tukey**, Exploratory data analysis, Reading, Mass., 1977.

[20] **P. H. C. Eilers and B. D. Marx**, "Flexible smoothing with B-splines and penalties," *Statistical science,* vol. 11, pp. 89-102, 1996.

[21] **M. Unser**, "Splines: A perfect fit for signal and image processing," *IEEE Signal processing magazine,* vol. 16, pp. 22-38, 1999.

[22] "Acquest Subaquatic Geology and Geophysics website," 2017, Accessed 8th, May, 2017.