Robust Global Mosaic Topology Estimation for Real-Time Applications

Nuno Pinho da Silva and João Paulo Costeira

ISR - Instituto Superior Técnico,
Torre Norte, Piso 7
Av.Rovisco Pais, 1 1049-001 Lisboa, Portugal
{nmps, jpc}@isr.ist.utl.pt
http://www.isr.ist.utl.pt

Abstract. This paper proposes an iterative methodology for real-time robust mosaic topology inference. It tackles the problem of optimal feature selection (*optimal sampling*) for global estimation of image transformations. This is called IGLOS: iterative global optimal sampling. IGLOS is a unified framework for robust global image registration (optimum feature selection and model computation are considered within the same methodology). The major novelty is that it does not rely on random sampling procedures. Furthermore, by considering an optimal subset of the total number of correspondences, it naturally avoids trivial solution. IGLOS can cope with any motion parameterization and estimation technique. Applications to underwater linear global mosaics and topology estimation are presented.

1 Introduction

In underwater activities, globally coherent seabed maps are useful tools to a human operator on a survey mission. Also, they have been used as spatial representations to support underwater autonomous navigation [**?**,5,10]. In building image mosaics there are two main tasks: image registration and image rendering. If only pair-wise registration is performed, small levels of noise in the estimation process may lead to large accumulated error, particulary if there are loops in the trajectories where non time consecutive frames overlap [1,8]. Furthermore, underwater applications are particulary prone to outliers due to independent moving objects (e.g., fishes or algae), poor lighting condition and mismatches. Therefore, robust global registration is required. Traditionally, this is accomplished with random sampling based algorithms between overlapping pairs followed by mosaic global topology estimation, in particular, with linear models [3,6].

This paper addresses the problem of robust global mosaic topology inference in real-time operations. Instead of removing outliers by random sampling, our methodology tackles the problem of optimal feature selection by sorting. An iterative approach is propose: from a set of correspondences and a model, we choose a subset of points that minimize the regression error. This methodology

M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 1250–1257, 2005.

is called IGLOS: iterative global optimal sampling. Feature selection process is the same as the one used in least trimmed squares [7]. IGLOS depends upon one single parameter that needs no estimation: *pt*, the required number of features to compute the model (*pt* parameter). Usually, trivial solution is avoided by introducing penalizing terms on the cost function [6,8] or using restrict motion models [2]. In our method, by imposing the choice of *pt* optimal features, spread over the image, we not only increase robustness to outliers but also avoid degeneracy. The major contributions are robustness, optimality and low complexity, which makes it suitable for real-time topology estimation.

2 Iterative Global Optimal Sampling (IGLOS)

2.1 Problem Formulation

Consider the image registration example of Figure 1. For the sake of clarity, assume that camera motion between I_k (image k) and the reference image I_0 is adequately described by an affine transformation $\mathbf{H}_{0k}$[1], and that p correspondences were found between pairs of images (I_i, I_j). Let $h_k = row(\mathbf{H}_{0k})$ be a column vector formed by stacking the first two rows of $\mathbf{H}_{0k}$ (see [4] for details on *row* operator). In matrix form:

$$\mathbf{H}_{0k} = \begin{bmatrix} \alpha_1^k & \alpha_2^k & \alpha_3^k \\ \alpha_4^k & \alpha_5^k & \alpha_6^k \\ 0 & 0 & 1 \end{bmatrix}$$

$$h_k = [\alpha_1^k, \alpha_2^k, \alpha_3^k, \alpha_4^k, \alpha_5^k, \alpha_6^k]^t$$

$$h_0 = [1, 0, 0, 0, 1, 0]^t$$

Using the notation $x = [u\ v\ 1]^t$ and $\mathbf{C}(x) = \begin{bmatrix} u\ v\ 1\ 0\ 0\ 0 \\ 0\ 0\ 0\ u\ v\ 1 \end{bmatrix}$ the registration error between a pair of corresponding points, measured in the reference image I_0 is given by $\varepsilon_{ij}^n = \|\mathbf{C}(x_{ij}^n) \cdot h_i - \mathbf{C}(x_{ji}^n) \cdot h_j\|_2^2$, where $x_{ij}^n \leftrightarrow x_{ji}^n$ is the nth ($n \leq p$) pair of corresponding points between images I_i and I_j, in homogenous coordinates. The global residue is written as:

$$\varepsilon = \left\| \begin{bmatrix} \mathbf{C}(x_{01}^1) & -\mathbf{C}(x_{10}^1) & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{C}(x_{01}^p) & -\mathbf{C}(x_{10}^p) & \mathbf{0} \\ \mathbf{C}(x_{02}^1) & \mathbf{0} & -\mathbf{C}(x_{20}^1) \\ \vdots & \vdots & \vdots \\ \mathbf{C}(x_{02}^p) & \mathbf{0} & -\mathbf{C}(x_{20}^p) \\ \mathbf{0} & \mathbf{C}(x_{12}^1) & -\mathbf{C}(x_{21}^1) \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{C}(x_{12}^p) & -\mathbf{C}(x_{21}^p) \end{bmatrix} \cdot \begin{bmatrix} h_0 \\ h_1 \\ h_2 \end{bmatrix} \right\|_2^2 = \|\mathbf{A} \cdot \bar{h}\|_2^2 = \|\bar{\varepsilon}\|_2^2 \tag{1}$$

[1] This is the most general collineation allowing for the residual vector to be expressed as a linear combination of the motion parameters, for more than 2 images.

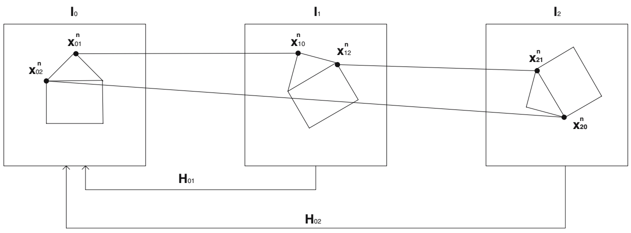

Fig. 1. Image registration example with three images

Note that each row of matrix $\mathbf{A}$ is related to the coordinates of a single matched pair. Given a model estimate $\bar{h}$, finding the optimum set of inliers is tantamount to choosing the entries of $\bar{\varepsilon}$ such that the global error ε is minimum in the least squares sense. This can be done by left multiplying $\mathbf{A}$ by a diagonal matrix $\mathbf{P}$ with $1, 0$ entries respectively. Thus, the problem of global robust registration can be stated through the following optimization problem:

Problem 1.
$$(\mathbf{P}^*, \bar{h}^*) = \underset{\mathbf{P} \in \mathcal{D}_{01}^{pt}\ ,\ \|\bar{h}\| = 1}{\arg\min} \{\|\mathbf{P} \cdot \mathbf{A} \cdot \bar{h}\|_2^2\}$$

where $\mathcal{D}_{01}^{pt}$ is the set of diagonal matrices with $\{0, 1\}$ entries and rank pt (to avoid null solution). The pt parameter is the total number of required inliers (the number of 1's in $\mathbf{P}$). If p_{min} is the minimum number of features to instantiate some model (e.g., $p_{min} = 3$ for affine or $p_{min} = 4$ for the general homography) and p_k correspondences were found between each image pair M_k, in a total of M matched pairs, then pt is the sum of all $pt_k : \ p_{min} \leq pt_k \leq p_k$ correspondences between all matched frames. In the example of Figure 1, $pt_k = p_{min} = 3\ ,\ \forall_k$ then $pt = \sum_{m=1}^{3} pt_k = 9$.

Since ($\mathbf{P}^*$ and $\bar{h}^*$) are unknowns, this problem is a nonlinear optimization problem. Furthermore, it is an integer problem in the $\mathbf{P}$ variable. Its combinatorial nature requires exhaustive search to finding a solution. To avoid this exhaustive search issue, some algorithms randomly sample the search space which is equivalent to randomly assign 1's and 0's in matrix $\mathbf{P}$. Though complex, Problem 1 is separable, in the sense that knowing one variable we can easily compute the other. Decoupling Problem 1, makes possible to avoid combinatorial explosion.

2.2 Iterative Approach for Solving Problem 1

Assume that, at iteration q, one knows an estimate $\mathbf{P}^q$ of $\mathbf{P}$, that is, a subset of correspondences. Knowing h_0 (e.g., for affine transformation $h_0 = [100010]$), the registration error is $\varepsilon = \|(\mathbf{P}^q \cdot \mathbf{A}) \cdot \bar{h} - (\mathbf{P}^q \cdot \mathbf{A}_0) \cdot h_0\|_2^2$, where $\mathbf{A}_0$ collects the columns correspondent to the reference image, $\mathbf{A}$ the remaining columns and $\bar{h}$ is the frame-to-mosaic global model for the remaining frames. Writing $b_0 = (\mathbf{P}^q \cdot \mathbf{A}_0) \cdot h_0$, Problem 1 reduces to

Problem 2. $\bar{h}^* = \arg\min_{\bar{h}} \{\|(\mathbf{P}^k \cdot \mathbf{A}) \cdot \bar{h} - b_0\|_2^2\}$

which solution gives an optimal $\bar{h}$ for the considered set of features. Note that $\mathbf{P}^q$ is idempotent $((\mathbf{P}^q)^t \cdot \mathbf{P}^q = \mathbf{P}^q)$. As long as $\mathbf{A}$ is full rank $(p_k \geq p_{min})$ and assuming affine motion, the solution to Problem 2 is

$$\bar{h}^* = (\mathbf{A}^t \cdot \mathbf{P}^q \cdot \mathbf{A})^{-1} \cdot (\mathbf{P}^q \cdot \mathbf{A})^t \cdot b_0 \tag{2}$$

Given $\bar{h}^*$, an optimal set of correspondences can be found by solving

Problem 3. $\mathbf{P}^* = \arg\min_{\mathbf{P} \in \mathcal{D}_{01}^{pt}} \{\|\mathbf{P} \cdot \bar{\varepsilon}\|_2^2\}$

If an efficient solution for the above problem exists, one may iterate between choosing the optimum set of features with known motion and computing the best global transformation from a set of pt correspondences. Initialization and other implementation issues are discussed in Section 3.

2.3 Optimal Sampling: Inlier Selection

Optimal sampling refers to the selection of the inliers that minimizes ε, that is, efficient solution of Problem 3. Reshape $\bar{h}$ by reintroducing h_0 into the proper entries (considering a sequence of N frames, $\bar{h} = [h_0\ h_1\ , \ldots,\ h_N]^t$ for $I_0 = I_1$). Given the transformation, all pair-wise residues measured in the mosaic frame ε_{ij}^n are stacked into the residual vector $\bar{\varepsilon}$ in ascending image order. Recalling that $\mathbf{P}$ is idempotent, the global registration error can be expressed as

$$\varepsilon = \|\mathbf{P} \cdot \bar{\varepsilon}\|_2^2 = \sum_{j}^{T_p} (\bar{\varepsilon}_j)^2 \cdot \mathbf{P}_{jj} \tag{3}$$

Thus, optimal sampling is accomplished by *sorting the residual vector and choosing the first pt_k entries between each matched pair M_k*. Sorting $\bar{\varepsilon}$ is performed in $\sum_{k=1}^{M} \mathcal{O}(p_k \log p_k)$ complexity, where p_k are the correspondences found in each matched pair M_k , $k = 1, \ldots, M$. pt_k is the number of required inliers, that is, $p_{min} \leq pt_k \leq p_k$. It turns out that this process leads to the criterium of the least trimmed squares [7]. Figure 2 outlines the methodology.

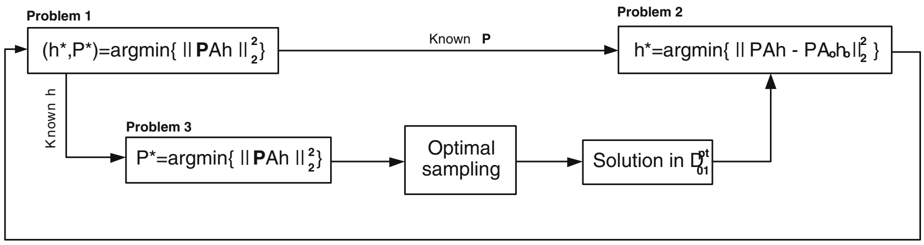

Fig. 2. Outline of the methodology

3 Implementation

Section 2.3 describes how to efficiently solve Problem 3. Given some initial feature subset (or motion estimation), one may iterate between find the optimal subset of features and compute the model with the selected subset until the following is verified

$$| \ \|\mathbf{P}^q \cdot \mathbf{A} \cdot \bar{h}^{q-1}\|_2^2 - \|\mathbf{P}^q \cdot \mathbf{A} \cdot \bar{h}^q\|_2^2 \ | \leq \theta \tag{4}$$

or a $q = K_{max}$ iterations (N is the total of frames in the sequence).

By defining regions on the image, features may be selected by constraining them to these regions in order to avoid degeneracy. Also, in Problem 2 ,b_0 is null everywhere except for the entries correspondent to coordinates in the reference frame. Therefore, reducing the number of selected features introduces less zeros in b_0, thus avoiding a solution close to the trivial solution ($\bar{h} = 0$).

Note that by decoupling Problem 1, inlier selection and motion computation are made independent. Extending IGLOS for accurate global registration is done by replacing Equation 2 with a nonlinear method.

To decrease computational burden, one may consider only the frames that overlap with the last one acquired. This considerably reduces the dimension of $\mathbf{P}$, consequently, the computational cost. Furthermore, instead of using batch least squares, model can be obtained using recursive least squares [3], making IGLOS suitable for real-time applications.

3.1 Initialization

We propose an *iterative initialization*[2]. In case of image mosaicing, we used the assumptions that image motion is smooth. In fact, one reasonable assumption is that the transformation between consecutive image is (picewise) constant. In other words, initial motion estimates between consecutive images are given by $\mathbf{H}_{(l-1)l} = \mathbf{H}_{(l-2)(l-1)}$, where I_l is the last acquired frame. If the transformation between the first 2 frames is known (a global translation which can be easily estimated or computed by other methods), frame-to-mosaic initialization proceeds as follow:

$$\begin{array}{l} \mathbf{H}_{0l} = \mathbf{H}_{0(l-2)} \cdot \mathbf{H}_{(l-2)(l-1)} \cdot \mathbf{H}_{(l-1)l} \\ \mathbf{H}_{(l-2)(l-1)} = \left(\mathbf{H}_{0(l-2)}\right)^{-1} \cdot \mathbf{H}_{0(l-1)} \\ \mathbf{H}_{(l-1)l} = \mathbf{H}_{(l-2)(l-1)} \\ \mathbf{H}_{(l-1)l} = \left(\mathbf{H}_{0(l-2)}\right)^{-1} \cdot \mathbf{H}_{0(l-1)} \end{array} \Rightarrow \mathbf{H}_{0l} = \mathbf{H}_{0(l-1)} \cdot \left(\mathbf{H}_{0(l-2)}\right)^{-1} \cdot \mathbf{H}_{0(l-1)} \tag{5}$$

for $l \geq 3$.

4 Experimental Results

Consider that a planar seabed is filmed by a moving camera pointing downwards. In each selected frame, a set of features were matched with a correlation based

[2] The authors acknowledge Prof. José Santos-Victor for this contribution.

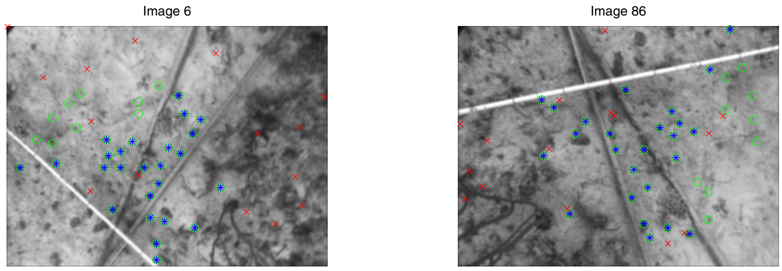

Fig. 3. Two images from the rock sequence, taken at non consecutive time. Superimposed inliers are depicted as circles 'o', outliers as crosses 'x' and IGLOS optimum set as asterisks '*'. No outlier was returned in the optimum.

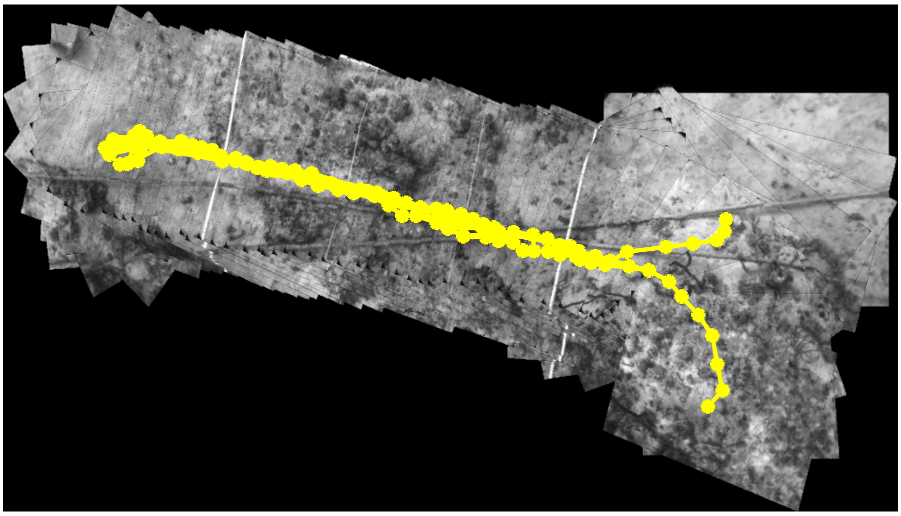

Fig. 4. Global IGLOS mosaic image. Topology is superimposed.

matcher as in [2]. The mosaic in Figure 4, constructed from a N=96 frames sequence[3], contains 610 overlapping pairs (Figure 3). Besides the real outliers, the set was contaminated with 30% of outliers, randomly taken from a uniform distribution over the image plane, to give a benchmark for performance. We require $pt_k = 25$ inliers for every M_k matched pair and affine model was assumed. Setting the first image as the reference frame I_0, we assume h_2 and h_3 (the transform from frames I_2 and I_3 to I_0, respectively) to be a global translation, after which initializations proceed as described in Section 3.1. Images were rendered with the use-last operator.

Figure 4 presents the correctly estimated topology superimposed. The maximum number of iterations per frame was set to $K_{max} = 50$ but the average

[3] The authors acknowledge Nuno Gracias for the image set and rendering procedures.

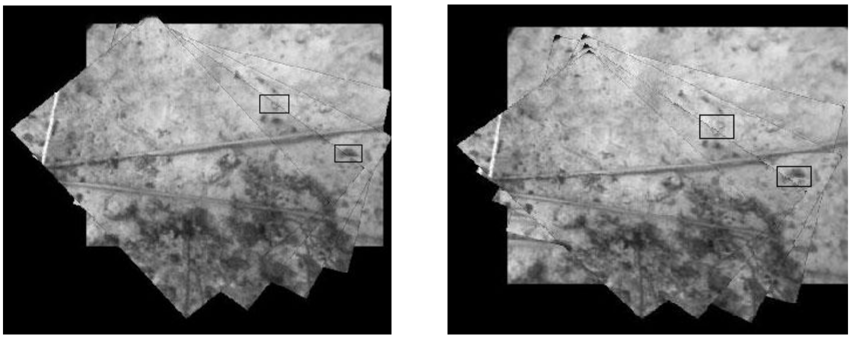

Fig. 5. Details from resulting mosaics. **Left:** from IGLOS mosaic. **Right:** from linear least squares mosaic.

number of iterations per frame was $K = 42.7345$. In the same conditions, a random sampling algorithm would have done $K_p = 24$ iterations per overlapping pair, meaning that $K = \frac{24 \cdot 610}{96} = 152.5$ iterations per frame would be necessary to assure, with 95% of probability (not optimal), that all points sampled i one sample would contain no outliers. The returned optimal set contains no outliers (Figure 3). Besides efficiency, selecting an optimum set of features promotes accuracy on linear mosaic construction, which is important in survey missions. Figure 5 illustrates details of the resulting mosaics with IGLOS and linear least squares. Superimposed boxes highlight the differences. Despite outliers, in the left image it is possible to observe that only one rock and sea weed exist. IGLOS provides a methodology for real-time robust mosaic topology estimation and improves accuracy in the resulting mosaic.

5 Summary and Conclusions

We have formulated the problem of global robust registration as a nonlinear mixed-integer optimization problem. To avoid NP hard problem, an iterative methodology was proposed, IGLOS: *iterative global optimal sampling.* Outliers rejection is performed through pair-wise sorting and model is globally estimated. The applicability of IGLOS to robust global consistent mosaics was discussed.

The methodology presented does not rely on random sampling procedures or on any estimate of the inlier standard deviation to assure robustness. Other motion parameterizations (e.g., similarity or full collineation) and non-linear estimation are straightforwardly introduced in the methodology. Major contributions are robustness with low complexity and optimality in the least squares sense. The tradeoff is dependence of initialization. IGLOS provide a unified framework for robust global registration (optimum feature selection and model computation are performed within the same methodology). In autonomous navigation, IGLOS allows for real-time robust mosaic topology inference.

References

1. Duffin, K., Barret W.: Globally optimal image mosaics. Graphics Interface, pp.217-222, 1998.
2. Gracias, N.: Mosaic-based Visual Navigation for Autonomous Underwater Vehicles. Phd Thesis, Instituto Superior Técnico. Universidade Técnica de Lisboa, 2002.
3. Gracias, N., Costeira, J., Santos-Victor, J.: Linear Global Mosaics for Underwater Surveying. 4th IFAC/EURON Symposium on Autonomous Vehicles, IAV04, Lisbon, Portugal, 2004.
4. Llkepohl, H.: Handbook of Matrices. Springer, 1996.
5. Negadahripour, S., Horn, B.: Direct passive navigation. PAMI, 9(1), January, 1987.
6. Pizarro, O. and Singh, H.: Toward large-area mosaicing for underwater scientific applications. IEEE journal of oceanic engineering, vol. 28(4), October, 2003.
7. Rosseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. John Wiley and Sons, 1987.
8. Sawhney, H., Hsu, S., Kumar, R.: Robust video mosaicing through topology inference and local to global aligment. Proc. European conference on computer vision, Springer-Verlag, June, 1998.
9. Torr, P.H.S., Murray, D.W.: The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. IJCV, 24(3),271-300, 1997.
10. Unnikrishnan, R., Kelly, A.: Mosaicing large cyclic environments for visual navigation in autonomous vehicles. Proc. International conference o robotics and automation (ICRA2002), pp. 4299-4306, Washington DC, USA, May, 2002.