Low overlap image registration based on both entropy and mutual information measures

Cédric de Césare
Ifremer / Laboratoire I3S, CNRS-UNSA
Sophia-Antipolis, France
decesare@i3s.unice.fr

Maria-João Rendas
Laboratoire I3S, CNRS-UNSA
Sophia-Antipolis, France
rendas@i3s.unice.fr

Anne-Gaëlle Allais and Michel Perrier
Ifremer Centre de Méditerranée
La Seyne sur Mer, France
{Anne.Gaelle.Allais,Michel.Perrier}@ifremer.fr

***Abstract*—In this paper, a new approach to image matching based on both entropy and mutual information measures that can correctly perform association under very low overlap conditions is presented. The method is *feature-based*, working on a list of interest points detected on the acquired images. Unlike the classical SSD/RanSaC method, our algorithm is robust to the existence of many similar regions in the images, enabling us to handle situations where the interest points correspond to local details of an habitat that is dispersed with a regular-like structure accross the ocean bottom. Actually, our method requires only that four corresponding points be detected in the two images being aligned. This allows matching under minimal overlap of the images. The improvement in efficiency of our method, when compared to previous techniques, relies on explicitly accounting for ambiguities in the association between the templates of the two images, thus preventing that useful information be discarded at an early association step.**

I. Introduction

Establishing a visual map of the region observed by the onboard camera of a mobile robot is a powerful tool which can be used either for post-mission environment analysis [3] or for on-line navigation [4]. For instance, in [10] we presented a path planning strategy for complete coverage missions, whose aim is to guarantee that the mosaic created after mission completion completely covers a prespecified region, i.e., that it leaves no non-observed holes.

With full generality, visual-based navigation requires image registration both between successive and non successive frames. The first case allows (partial) estimation of the motion of the platform between frame acquisitions (we will refer to it as *image matching*) whereas the second enables the estimation of the relative pose of the platform at distant times (we will refer to it as *image registration*). The main difference between these two problems concerns the expected image overlap, which is assumed large in the former case while no specific assumptions hold in the latter.

For complexity reasons, image matching is often based on template or block matching approaches. Global methods do exist [9], [13], [15], that are able to solve the registration problem, but they are usually too heavy to be implemented on-line, and are more common in medical imaging applications, where run-time is not a major concern. A variety of block-matching approaches has been proposed in the literature, with different specific goals in mind, such as object tracking , motion estimation , creation of underwater visual maps , and underwater navigation with respect to visual maps of the ocean floor [1], [2], [4].

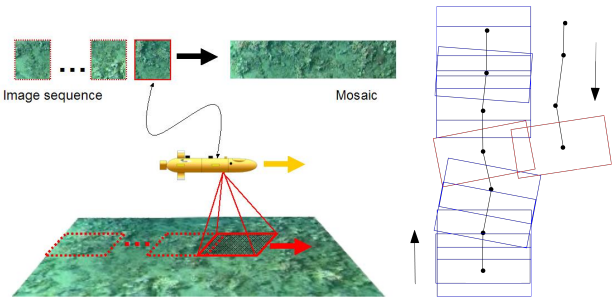

Fig. 1. (Left) Example of a robot creating an image mosaic of an interest region (Right) The robot is returning on the previously observed area.

In this paper, we present a new algorithm that can robustly perform image-to-image registration. The algorithm improves on previous techniques in that it does not require a large overlap between the images being registered, and that it can operate even on quasi-repetitive scenarios, that contain many locally similar textured regions. These increased efficiency and robustness are attained by explicitly acknowledging the plausibility of similarity between distant neighborhoods (i.e., the probability of false associations), and delaying definite block-to-block association to a step that globally evaluates their collective likelihood.

The ability of conditioning the decision about the association of each image block on the associations of the other blocks requires the assumption of a "global" model that jointly constraints them. In our context, this "global" model comes from assumptions on the trajectory of the robot and the world that it observes. We will consider the framework of underwater robotics, and assume that the sensor carrier is traveling at a nearly constant distance from a locally flat surface (the ocean bottom), with a camera placed orthogonally with respect to it, as shown schematically in Figure 1. This conveniently constraints the apparent motion of distinct regions of the images, as we make more precise in the next section. We will not distinguish between image matching and image registration. In fact, our algorithm can be applied in both situations, either between successive images when the platform is travelling along a quasi-linear track (e.g. between the blue rectangles

978-1-4244-2620-1/08/$25.00 ©2008 IEEE

on the right-hand side of fig. 1), and when its image partially overlaps previously observed regions (as for the two red rectangles in the right-hand side of fig. 1).

The paper structure is as follows. Section II presents the image matching problem assessed in the paper. In section III, similarity (or discrepancy) measures usually used in image alignement are reviewed. Section IV presents the classical template-based image registration algorithm RanSac. Section V presents our registration algorithm, explaining how its processing of local ambiguities allows overcoming of some problems associated to previous approaches. In section VI we illustrate the advantages of our algorithm with respect to RanSaC in two sets of examples.

II. Problem formulation

In this section the image matching problem is formulated. The image registration model is first defined, establishing the assumptions underlying our work.

A. Image registration model H

Two perspective images of a *planar* surface are related by an image registration model H known as *homography*. Since we assume that the world observed by the robot is locally planar, the same homography relates the coordinates of *all image points* of two distinct images $\{\tilde{I}_n(s), s \in \mathcal{S}\}$ and $\{\tilde{I}_m(s), s \in \mathcal{S}\}$ – where $\mathcal{S}$ denotes the image support, a subset of the retinal plane – and, neglecting image noise and lightening induced variations,

$$\tilde{I}_n(s) = \tilde{I}_m(Hs), \qquad s \in \mathcal{S} \cap H\left[\mathcal{S}\right] \quad . \tag{1}$$

We will denote by $\mathcal{O}^H$ the *overlap region* corresponding to homography H :

$$\mathcal{O}^H = \mathcal{S} \cap H\left[\mathcal{S}\right] \quad . \tag{2}$$

B. Problem Statement

Ideally, the image registration problem can be formulated as finding the homography H such eq. (1) holds. However, this equality assumes perfect validity of the conditions under which the homography model is valid, in particular, *(i)* that the observed scene is planar, and *(ii)* that there are no image artifacts due to noise or light effects. A more realistic formulation of the problem looks for the homography that minimizes the differences between the two sides of the equality sign, allowing for image-to-image variations of intensity or color due to light changes:

Problem *1:* **Image Registration** *Given two images $\tilde{I}_m$ and $\tilde{I}_n$, find the homography H that minimizes*

$$\mathcal{E}(H) = \mathcal{D}\left(\{\tilde{I}_n(s)\}_{s\in\mathcal{O}^H}, \{f\left(\tilde{I}_m(Hs)\right)\}_{s\in H^{-1}[\mathcal{O}^H]}\right). \tag{3}$$

where $\mathcal{D}(\cdot,\cdot)$ is a given discrepancy measure[1] *between images, for some unknown* distortion *function $f(\cdot) : \mathcal{I}^{|\mathcal{O}^H|} \rightarrow \mathcal{I}^{|\mathcal{O}^H|}$ with $\mathcal{I}$ the set of image values.*

[1]The definition of the discrepancy measure $D(\cdot,\cdot)$ will be discussed in section III.

The problem formulated above considers *global* image registration, requiring that the same homography H and distortion model f hold between all the regions of the image. It is sensitive with respect to sudden changes in the observed scene, for instance if a new object appears between the two acquisitions. A simpler problem, that can be made robust with respect to local variations in the image scene, imposes the homography relation only between local neighborhoods of the image.

Problem *2:* **Template Matching** *Let $\tilde{I}_m(s), s \in \mathcal{S}$ and $\tilde{I}_n(s), s \in \mathcal{S}$ be two images, and $W_n(s), s \in S_n \subset \mathcal{S}$ a template in $\tilde{I}_n$. Find the template $W_m^\star$ that minimizes*

$$\mathcal{E}_{W_n}\left(W_m\right) = \mathcal{D}\left(W_n, f\left(W_m\right)\right), W_m = \{\tilde{I}_m(Hs), s \in S_n\}. \tag{4}$$

In the template matching problem the support S_n is a small neighborhood of the image, increasing the plausibility that the homography model H holds between the two distinct observations. Usually, instead of a generic homography, a simple translation model is assumed: $Hs = s + t_n$, and no distortion operator $f(\cdot)$ is considered, the image being modeled as a noisy version of an unobserved image I, leading to the simple model, such that

$$W_n(s) = \tilde{W}_m^\star(s + t_n) + \epsilon_{n,m}(s), \qquad s \in S_n \quad . \tag{5}$$

where $\epsilon_{n,m}$ is a noise process that combines the noisy components of the two images. In these conditions, template matching is equivalent to estimation of the translation vector t_n. Definition of a statistical model for the noise process allows then application of statistical estimation tools to the problem of estimating the vector t_n that defines the subset of I_m correspondent to W_n.

A slightly different problem consists in finding the correspondences between a *set* of image neighborhoods (or templates).

Problem *3:* **Template Association** *Let $\tilde{L}_i = \{W_i^{(j)}, s \in S_i^{(j)} \subset \mathcal{S}\}_{j=1}^{K_i}, i = m, n$ be two lists of templates. Assume that $K_m < K_n$. Find the injective mapping $\pi : \{1,\ldots,K_m\} \rightarrow \{1,\ldots,K_n\}$ that maps templates in $\tilde{L}_m$ to templates in $\tilde{L}_n$ that minimizes*

$$\mathcal{C}(\pi) = \sum_{j=1}^{K_1} \mathcal{E}_{W_n^{(j)}}\left(W_m^{(\pi(j))}\right) \quad . \tag{6}$$

This is the problem addressed in the paper.

Finally, we also formulate the problem of determining the homography H that maps the pixels coordinates of a series of pairs points detected in two images.

Problem *4:* **Homography from templates** *Consider two images $\tilde{I}_m$ and $\tilde{I}_n$, and let $\tilde{L}_m, \tilde{L}_n$ be lists of corresponding templates. Let $Z_j^{(i)}$ denote the coordinates of the center*

of template $W_j^{(i)}$. Find the homography H that maps the coordinates of the two lists of templates, i.e, that minimizes :

$$d(Z_n^{(i)}, HZ_m^{(i)}), \qquad i = 1, \ldots, \left|\tilde{L}_n\right| = \left|\tilde{L}_m\right| \ , \qquad (7)$$

where $d(.,.)$ is a distance metric.
This problem has a simple linear solution in the elements of the homography matrix H [6].

Finding the solutions of Problems 3 and 4 lead to the solution of Problem 1.

$$\hat{H} = \arg\min_H \ d(Z_n^{(i)}, Z_m^{\hat{\pi}(i)}) \qquad (8)$$

where

$$\hat{pi}(i) = \arg\min_\pi \ \mathcal{C}(\pi). \qquad (9)$$

III. Discrepancy measures

We focus in this section on the definition of the discrepancy measure $\mathcal{D}(\cdot,\cdot)$ that enters in the definition of problems 1 to 3 in the previous section. Let us go back to the structural ideal relation that we want to establish between two images (or image templates):

$$\tilde{I}_n(s) = f\left(\tilde{I}_m(H(\beta)s) : \alpha\right), \qquad s \in \mathcal{O}^H \ . \qquad (10)$$

where we explicity assumed that the admissible distortion functions $f(\cdot)$ are parametrized by a vector of parameters α, and introduced β as the free parameters of the homography H (in the simple translation model, $\beta = t_n$ is a two dimensional vector).

Assuming that

$$\forall\beta, \forall s \in S_n, H(\beta)s \in \mathcal{S} \ , \qquad (11)$$

and that we have noise corrupted versions of the images I_n and I_m with noise[2] $\epsilon \sim p_\epsilon$, the templates of both images can be modelled as follows :

$$\begin{aligned} W_n(s) &= I(s) + \epsilon_n(s), \\ W_m(s) &= f(I(H(\beta)s : \alpha) + \epsilon_m(s). \end{aligned} \qquad (12)$$

which leads to expressing $W_m(s)$ in term of $W_n(H(\beta)s)$:

$$\begin{aligned} W_n'(s)|_{\beta,\alpha} &= f(W_n(H(\beta)s) : \alpha) \\ &\approx f(I(H(\beta)s) : \alpha) + \epsilon_n'(s) \qquad (13) \\ W_m(s) &= f(W_n(H(\beta)s) : \alpha) + \epsilon_m(s) - \epsilon_n'(s) \qquad (14) \end{aligned}$$

We should note that ϵ_n' is dependant on the image and the function f.

Using W_n as a parameter in the model of W_m, we can approximately solve Problem 2 by maximizing the probability over β:

$$p_{\epsilon^\star}\left(W_m(s) - f(W_n(H(\beta)s) : \hat{\alpha}(\beta))\right), \qquad (15)$$

where $\epsilon_m(s) - \epsilon_n'(s) \sim p_{\epsilon^\star}$ is assumed known, and $\hat{\alpha}(\beta) = \arg\max_\alpha \ p_{\epsilon^\star}\left(W_m(s) - f(W_n(H(\beta)s) : \alpha(\beta))\right)|\alpha, \beta)$.

[2]Notation $X \sim p$ indicates that x is distributed according to p.

The corresponding template will be referred as $W_m^\star = \{W_m(s), s \in H(\hat{\beta})S_n\}$. For this criterion, we can thus identify the negative likelihood as the relevant discrepancy function for problem 2:

$$\mathcal{D}_{ML}(\beta) = -\log p_\epsilon^\star\left[W_n - f\left(W_m(H(\beta)s) : \hat{\alpha}(\beta)\right)\right], \qquad (16)$$

If $p_\epsilon^\star$ is assumed to be Gaussian, and $f(\cdot)$ is the identity function (i.e., there is no distortion of image values from one view to the other), this criterion is equivalent to minimizing the SSD (Sum of Squared Differences), classically used for image registration [4] :

$$\mathcal{D}_{\text{SSD}}(\beta) \equiv \sum_{s\in S_n} \rho(s:\beta)^2 \ , \qquad (17)$$

where $\rho(s:\beta)$ are the residues

$$\rho(s:\beta) \equiv W_n(s) - W_m\left(H(\beta)s\right), \ s \in S_n. \qquad (18)$$

This criterion is often used because of its simplicity and good performance under simple noise models. It is, nevertheless, sensitive to lightening variations, unlike the two alternative criteria presented in the next subsections [1], [14], [13].

However we must point out that when we register two images, eq. (11) can never hold [3]. In these conditions, the size of the overlap region $\mathcal{O}^{H(\beta)}$ depends on the value of β considered, and the likelihood functions are defined over *distinct observation spaces*. Maximum Likelihood is thus no longer applicable, since we cannot compare the values of distributions defined over different sets. This difficulty can be solved by using the Minimum Description Length (MDL) [11], and formulating the template matching as a model estimation problem, where different values of β index the individual candidate models being considered. It leads to a penalized version of the Likelihood function, see [9]. If a non-parametric, or semi-parametric approach is used to with respect to the noise distribution p_ϵ, it can be shown that the negative likelihood is penalized version of the *entropy* of the empirical distribution of the residues (18). The penalty term, dependent on the size of the images being matched, automatically sets a threshold that takes into account the variation of the effective size of the observations being matched.

We review below the main criteria that have been used as alternatives to the SSD criterion. We first give two basic definitions required in the rest of the article. See [12] for a detailed presentation and discussion.

A. Basic notions from Information Theory

a) Shannon entropy: Let X be a random variable with distribution p_X. The Shannon entropy[4] $H(X)$ of X is a

[3]Indeed, this would mean that the homography is the identity matrix and that the overlap region has the same size than $\tilde{I}_n$ and $\tilde{I}_m$.

[4]We use the same notation for both entropy and the homography matrices H. In each case, the meaning of H should be clear from the context.

measure of the uncertainty about the values of X. It is, by definition[5]

$$H(X) = \mathrm{E}_{p_X}\left[-\log p_X(x)\right] \quad (19a)$$
$$= -\sum_x p_X(x)\log p_X(x) \ , \quad (19b)$$

where we assumed that X is defined over a discrete set $\mathcal{X}$.

Joint entropy of two random variables $X, Y \sim p_{X,Y}$ is defined in analogous way: $H(X,Y) = \mathrm{E}_{p_{X,Y}}\left[-\log p_{X,Y}(x,y)\right]$. Entropy has a number of remarkable properties, in particular: *(i)* it is always positive, being equal to zero only for the degenerate distributions that put all probability mass in the same point, *(ii)* if $|\mathcal{X}| = K \Rightarrow H \leq \log K$, *(iii)* if $X \sim p_X$ and $Y \sim p_Y$ are statistically independent, $H(X,Y) = H(X) + H(Y)$.

b) Mutual Information: The mutual information $MI(X;Y)$ between two random variables X and Y is a measure of the amount of information one random variable contains about the another. It can be defined by :

$$MI(X;Y) = H(X) + H(Y) - H(X,Y) \quad (20a)$$
$$= H(X) - H(X|Y). \quad (20b)$$

This last equation supports the interpretation of the mutual information as a decrease in the uncertainty about one variable when observations about the other variable are made. It can be shown that $MI(X;Y) \geq 0$, being equal to zero if and only if one of the variables is a function of the other, i.e.

$$MI(X;Y) = 0 \Leftrightarrow X = g(Y) \ , \quad (21)$$

for some bijective function $g(\cdot)$. An important property of mutual information is that it is invariant with respect to bijective functions :

$$MI(X;Y) = MI(g_1(X), g_2(Y)), \quad (22)$$

for any bijective functions g_1 and g_2.

B. Residues Entropy

Use of the *residues entropy*

$$\mathcal{D}_{\mathbf{H}}(\beta) \equiv \widehat{H}\left(\rho(s:\beta)\right), \quad (23)$$

where $\widehat{H}(\rho(s:\beta))$ is an estimator of the entropy of the residues $\{\rho(s:\beta)\}_{s\in\mathcal{O}^{H(\beta)}}$ as a discrepancy metric can be justified based on its *robustness* with respect to several factors : *(i)* lack of knowledge of the image noise distribution p_ϵ; *(ii)* robustness to systematic shifts in the intensity values (that can be a side effect of automatic camera controls). As we discussed above, it is related to the likelihood under non-parametric models of p_ϵ, see [9].

[5] $\mathrm{E}_p[\cdot]$ denotes expected value under distribution p.

C. Mutual information

The use of entropy as a criterion to estimate image correspondences requires the explict determination of $\hat{\alpha}(\beta)$ *for each value*. It assumes thus that we are able to define the set of admissible distortion functions. Mutual information provides a criterion that is invariant with respect to the function $f(\cdot)$[6] relating the pixel intensities in the two images. Since $\mathcal{D}$ must measure discrepancy, we define it as the *negative* of the mutual information between the images being compared:

$$\mathcal{D}_{\mathbf{MI}}(\beta) \equiv -\widehat{MI}\left(W_n(s); W_m\left(H(\beta)s\right)\right), \quad (24)$$

where $\widehat{MI}\left(W_n(s); W_m\left(H(\beta)s\right)\right)$ is an estimator of the mutual information between $W_n(s)$ and $W_m\left(H(\beta)s\right)$. According to this criterion, we should choose β such that the mutual information between correspondent pixels is maximized. As expressed previously, MI is more often estimated by using its relation to entropy, see eq. (20). We point out the increased complexity of the estimation of mutual information with respect to entropy, requiring characterisation of the *joint* distribution of the two data sets.

IV. The RanSaC based algorithm

The classical algorithm for template-based image registration is briefly presented in this section. For a detailed description refer to [6], [4]. It can be decomposed in three main steps (fig. 2):

A. Template detection

The aim of this step is to determine a set of image regions $\tilde{L}_n$ that *(i)* can be well localized inside an image, and *(ii)* are distinctive (in the sense that they resemble no other region in the image). Guaranteeing this last property is computationally demanding, requiring analysis of the pairwise similarity of the set of regions under consideration. For this reason, only property *(i)* is usually taken into consideration. The Harris detector [5] and SIFT points [8], [7] are commonly used, offering good properties in terms of the accuracy of the estimation of their location (for Harris "corners") and robustness with respect to perspective effects (for SIFT points). In the results presented in section VI we consider image templates identified by an Harris detector [7].

B. Template Matching (Problem 2)

The second step solves Problem 2 in section II for all elements of $\tilde{L}_n$ using the SSD criterion. A set of "best matches" $\tilde{L}_m = \{W_m^{\star(j)}, j = 1, \ldots, \left|\tilde{L}_n\right|\}$ is then obtained. We point out that the situation where $W_m^{\star(i)} = W_m^{\star(j)}$ is possible. $\tilde{L}_n$: a binary function $\pi^\star$ which describes the association between the two lists can be defined by :

$$\begin{array}{rccc} \pi^\star : & \tilde{L}_n \times \tilde{L}_m & \rightarrow & [0,1] \\ & \left(W_n^{(i)}, W_m^{(j)}\right) & \rightarrow & \pi^\star(i,j) \end{array} \quad (25)$$

[6] Providing that f is injective.

[7] One should note that any other interest point detector can be used, as long as it is stable: the set of detected points stays constant from image to image.

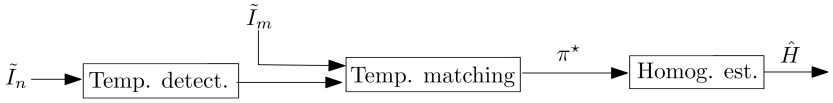

Fig. 2. Classical template-based approach to image registration.

$\pi^\star(i,j) = \delta_{i,j}$ where δ is the Kronecker symbol.

C. Selection and homography estimation

Finally, an homography $\hat{H}$ is estimated by considering a subset of the template correspondences established in the previous step. The homography selection is performed by RanSaC algorithm (fig. 3).

The RanSaC algorithm uses the notion of *inliers* (couples of templates which fit an homography) and *outliers* (couples of templates which do not), and can be briefly summarized as follows.

The algorithme starts by randomly selecting four template couples from $\pi^\star$. An homography H can thus be computed. The next step aims to find how many pairs in $\pi^\star$ fit H within a given tolerance t (26).

$$\sum_i d(Z_n^{(i)}, HZ_m^{(i)})^2 + d(Z_m^{(i)}, H^{-1}Z_n^{(i)})^2 < t. \qquad (26)$$

The set of couples which are below the tolerance is defined as the inlier set. If the number of inliers is large enough, the sum for all the inliers of the error of homography fitting is computed. These steps are repeated N times. Finally the homography for which the fitting homograpy error is the lowest is retained.

V. Methodology

A. General presentation

This section is dedicated to the presentation of our algorithm, a template-based image registration method that can correctly perform association under low overlap conditions, where the RanSaC architecture presented in the previous section will fail. As the RanSaC method, our approach also considers separately the establishment of template-to-template associations and the actual selection and homography estimation. The main distinction consists in the definition of the application $\pi^\star$ that maps the template lists.

In figure 4, the processing chain of our algorithm is presented. The first step performs template detection. The two other blocks address template association and homography estimation. These blocks are slightly different from the standard RanSaC processing chain and are discussed in more detail in the next subsections. However we briefly highlight the main differences :

- Instead of associating a list of templates in one of the images $\tilde{L}_n$ to *the other image* $\tilde{I}_m$, we perform template detection individually in each image, producing two lists $\tilde{L}_n$ and $\tilde{L}_m$.
- Instead of performing a strict association between the members of the two lists, we define a *multi-valued association* between the two lists of templates, that measures the *similarity* of all pairs of templates.

B. Template detection

Template detection step is performed separetely for both images, using the Harris corner detector.

C. Multi-valued association metric

The multi-valued association metric can be defined as follows: Define the *similarity matrix*, denoted by Π^a, as the $|L_n| \times |L_m|$ matrix whose (i,j) element is $\mathcal{D}_a(i,j)$

$$\Pi^a(ij) = \mathcal{D}_a\left(W_n^{(i)}, W_m^{(j)}\right) \ ,$$

where $\mathcal{D}_a$ is one of the discrepancy measures presented in section III: $a \in \{\text{ML}, \mathbf{H}, \text{MI}, \text{SSD}\}$.

The rationale for the method above is the following: using a measure of similarity with small numerical complexity, we start by detecting likely template-to-template associations. These associations are denoted by the mapping function π^0.

The discrepancy measure chosen for this first step is the residue *entropy*. We compute the entropy based similarity matrix Π_H and then associate the template pairs that have a suficiently small discrepancy

$$\pi^0(i,j) = \begin{cases} 1, & \Pi_H(i,j) \leq \gamma_H \\ 0, & \text{otherwise} \end{cases} \ . \qquad (27)$$

$\pi^0(i,j) = 1$ if the two templates are similar relatively to the similarity measure used and otherwise $\pi^0(i,j) = 0$. Note that we impose no constraints on π^0. In particular, it can happen that some templates are associated to no other template, i.e, $\forall j, \pi^0(i',j) = 0$ or $\forall i, \pi^0(i,j') = 0$, or one template of one image, say $\tilde{I}_n$ can be associated to more than one template in the other image: $\exists j_1 \neq j_2 \ : \ \pi^0(i,j_1) \neq 0$ and $\pi^0(i,j_1) \neq 0$. Since π^0 is a binary function, we can represent it by a bipartite graph $\mathcal{G}_{\pi^0}$, whose nodes are the elements of the two lists, and that has an arc between nodes for which $\pi^0 = 1$. Figure 5 illustrates one possible example.

In a second step, we use mutual information to obtain a better discrimination of the selected couples of templates. Finally we obtain the list of possible associations denoted by $\pi^\star$ and a graph $\mathcal{G}_{\pi^\star}$, defined in the same way of $\mathcal{G}_\pi^0$. $\pi^\star$ has the same properties as π^0 and is defined by retaining only the pairs that have large mutual information (small values of $\mathcal{D}_{MI}$) :

$$\pi^\star(i,j) = \begin{cases} 1, & \Pi_H(i,j) \leq \gamma_{MI} \\ 0, & \text{otherwise} \end{cases} \ . \qquad (28)$$

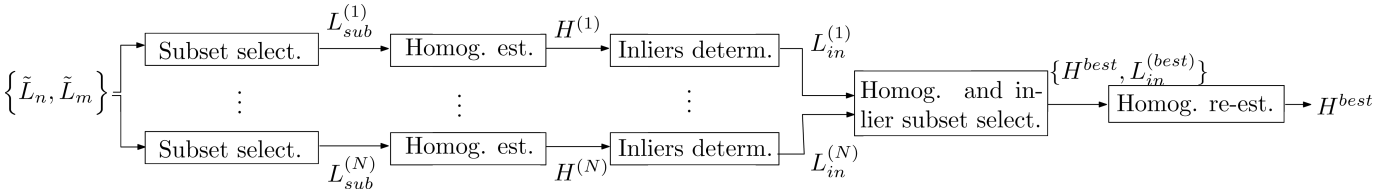

Fig. 3. Last step of the RanSaC method

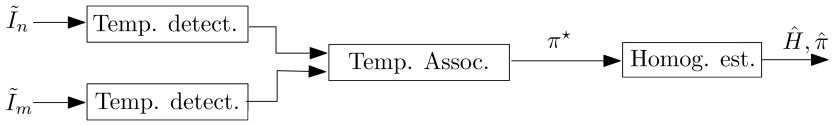

Fig. 4. Our approach to image registration.

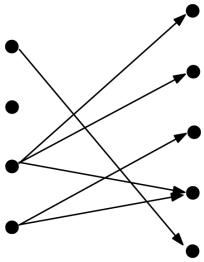

Fig. 5. Graph of an application π^0.

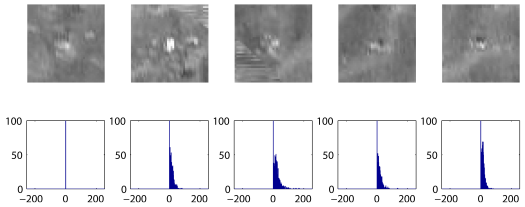

Fig. 7. Templates and the residue histograms representing regions which can be seen as ambiguous. The fourth template is used as the reference to compute the residues.

We point out that no computation is done for the templates which verify $\pi^0(i,j) = 0$. The fact that we do not enforce any special structure in $\pi^\star$ enables us to capture situations where the templates in one image resemble many templates in the other image, as it is frequent in natural terrains, where the interest points detected pick up details of the local habitat, and is dispersed over the ocean flow with a quasi-regular structure, as shown in Figure 6. Figure 7 shows the templates and residue histograms when matching the first template to the other templates. All histograms are concentrated around zero, demonstrating that all templates are similar to each other.

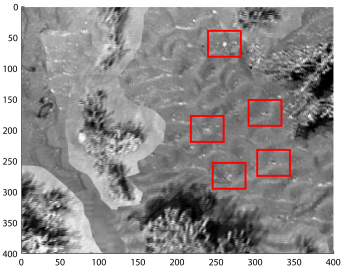

Fig. 6. Regions of the image which can lead to ambiguous associations.

Definition of the thresholds γ_H and γ_{MI} in (27) and (28) is difficult. Our present implementation uses thresholds defined in terms of the maximum values:

$$\gamma_a = \alpha_a \max_{(i,j)} \Pi_a(i,j), \qquad a \in \{H, MI\} \ .$$

In our results, we use a value of $\alpha_H = 0.7$ and $\alpha_{MI} = 0.3$. These threshold have been determined from four series of images. They are voluntarily slighlty lower than what they could be in order to fit several kinds of images.

D. Homography estimation

Using the application $\pi^\star$, we then solve problem 4. Let $\mathcal{F}$ be the set of sub-graphs of $\mathcal{G}_{\pi^\star}$ that have four disjoint arcs, and denote by π a generic element of $\mathcal{F}$. Each π establishes a one-to-one association of four points in each image, and is thus sufficient to determine an homography, that we denote by H^π.

$$\pi \in \mathcal{F} \longrightarrow H^\pi \ .$$

Computing all the possible combinations of four point couples and choosing the best one (for example by determining the highest mutual information measure between the corresponding common part of the images) leads to the best solution in terms of homography estimation. Nevertheless this method is extremely demanding in terms of computation, as all the combinations are examined. Present work studies the possiblibility of defining a preference strategy for considering couples of templates and a stopping criterion.

VI. Experiments

In this section we compare our algorithm with the SSD/RanSac method in two sets of images. The first one considers the registration of images related by a simple translation, with low overlap between them, demonstrating the limitations of other methods even under simple motion models. The second series of results consider real images, testing our approach under general platform motion model, and scene geometries.

A. Low overlap

This first example considers registration of the two images shown in Figure 8. These images are synthetic noise free images, being subsets of a large image created by stitching together real B&W images of the ocean bottom taken with a camera onboard an underwater robot (fig. 16). The selected "images", of size 399×399 are extracted at locations selected along a trajectory manually defined on the large image. These experiments are useful because we have a ground truth to which the result of the algorithm can be compared. Moreover, they perfectly comply with the flat surface model, and thus allow assessment of performance under this ideal condition.

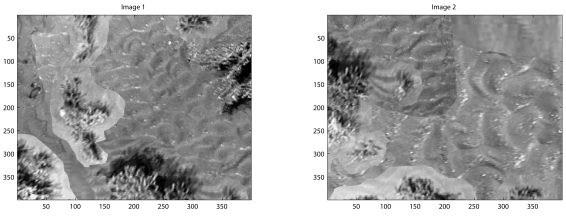

Fig. 8. Images to be matched and their matching.

Even visually, it is difficult to establish the correspondence between the two images. In fact, the image on the left is displaced by 250 pixels on both directions, leading to an overlap of about 14% of the total image size, corresponding to the homography H_0 :

$$H_0 = \begin{bmatrix} 1 & 0 & -250 \\ 0 & 1 & -250 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (29)$$

As shown in fig. 9, 34 and 37 templates have been respectively detected in the first and second images. Among all the possible associations only five are correct. Figure 9 gives a visual representation of the 37×34 discrepancy matrix Π_H. The correct five associations correspond to the black (close to zero) entries. All other values are much larger, indicating the presence of outliers (templates that have no correspondence in the other image).

The homography $\hat{H}$ estimated by our method is $\hat{H} = H_0$,: correctly identifying the true image displacement. Figure 12 show the composition of the two images using the estimated homography.

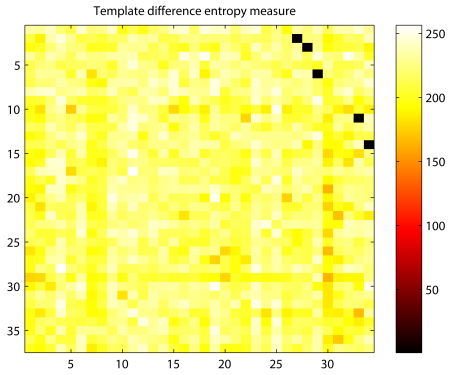

Fig. 9. Template difference entropy measure.

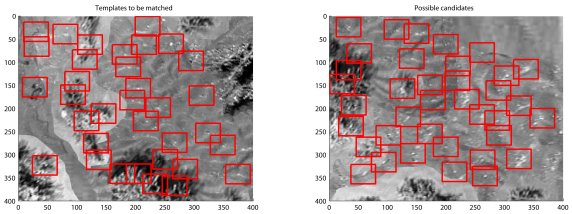

Fig. 10. Templates to be matched and their possible candidates.

The images are noise free. This puts the SSD matching criterion used by the RanSaC method in its best operating condition.

The low overlap generates too many outliers decreasing the probability taht the real homography be selected. Even in this case , the weak percentage of correct associations will prevent it from passing the inlier test. Thus, the RanSaC algorithm cannot produce an estimate of the homography and the classical method of image registration fails.

As we just mentionned, the number of true matchings in this example is five. That means that if the four true correspondences are selected, only one true couple will be left to validate the model. There will be only one inlier whereas false homographies may lead to greater numbers of inliers. In our case, the RanSac based method is thus unable to produce an estimate.

B. Real images

The other experiments presented here use a real underwater video sequence. Figure 13 illustrates the underwater images to be matched. Images are no more noise free and slight changes in lightening can be observed as shown in Figure 15. The final selected templates are presented in figures 14 and 15. The bottom row of templates in Figure 15 plots the template residues between each couples of selected templates, showing that our algorithm leads to good template associations.

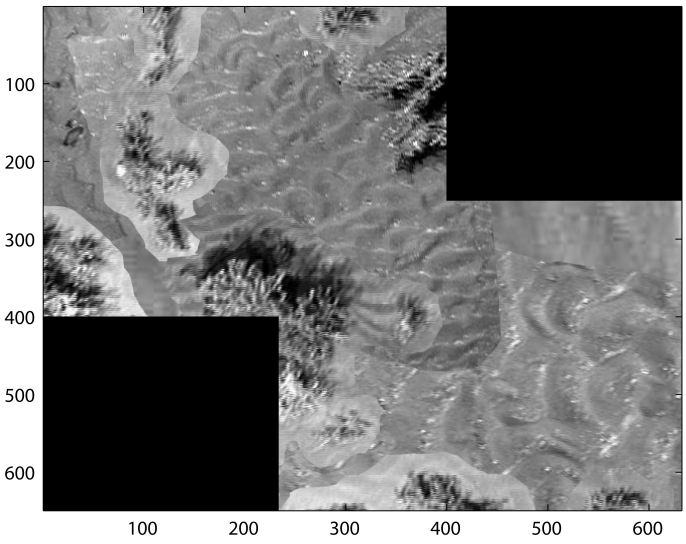

Fig. 12. Matching result.

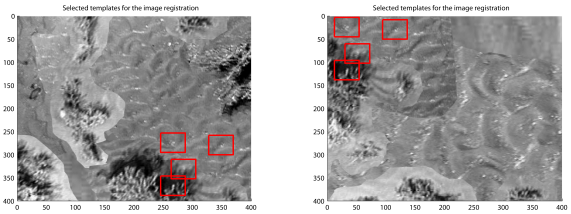

Fig. 11. Selected couples of templates.

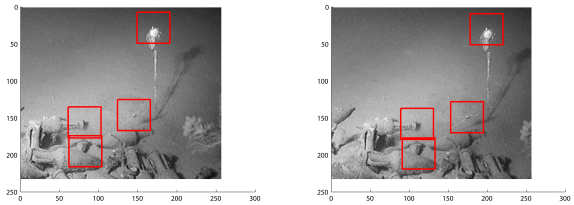

Fig. 14. Selected couples of templates.

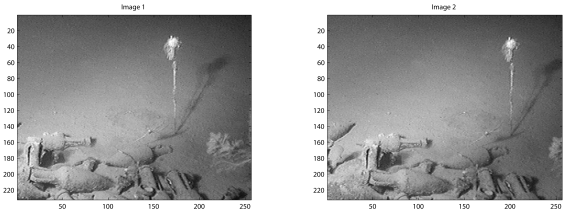

Fig. 13. Real images to be matched.

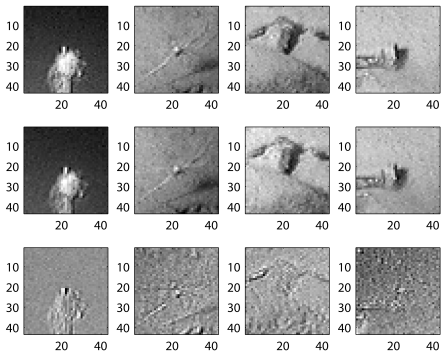

Fig. 15. Selected couples of templates from image 1 and image 2 and their residues.

VII. Conclusion and perspectives

In this document a new template matching approach has been presented in the context of underwater image mosaicing. This approach works for nearly repetitive scenarios, in cases where the classical matching methods may fail. The proposed metod can handle images presenting very low overlap, and takes into account ambiguities between pairs of templates.

References

[1] S. Boltz, E. Wolsztynski, E. Debreuve, E. Thierry, M. Barlaud, and L. Pronzato. A minimum-entropy procedure for robust motion estimation. In *IEEE International Conference on Image Processing*, pages 1249–1252, Atlanta, USA, 2006. ICIP'06.

[2] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *IEEE Conf. on Comp. Vis. and Pat. Rec.*, pages 142–151, Hilton Head Island, South Carolina, 2000.

[3] Lirman D., Gracias N., Gintert B., Gleason A., Reid R., Negahdaripour

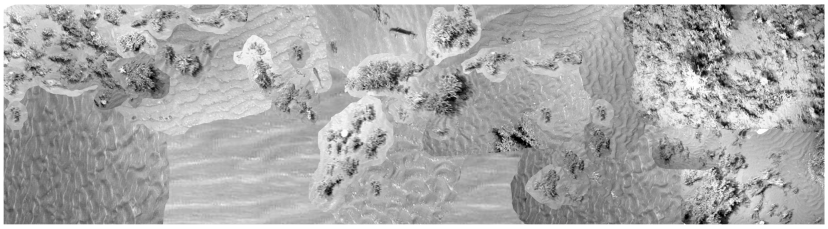

Fig. 16. Large synthetic image where images to be registered are extracted from.

S., and Kramer P. Development and application of a video-mosaic survey technology to document the status of coral reef communities. *Environmental Monitoring and Assessment*, 125, February 2007.

[4] N. Gracias and J. Santos-Victor. Robust estimation of the fundamental matrix and stereo correspondences.

[5] C. Harris and M. Stephens. A combined corner and edge detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[7] D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.

[8] David G. Lowe. Object recognition from local scale-invariant features. In *Proc. of the International Conference on Computer Vision ICCV, Corfu*, pages 1150–1157, 1999.

[9] M.J. Rendas. Construction of video mosaics using the minimum description length. Boston, USA.

[10] M.J. Rendas and C. de Cesare. Controlled acquisition of video mosaics by an autonomous underwater vehicle, 2007. Durham, NH, USA.

[11] J. Rissanen. *Stochastic Complexity in Statistical Inquiry Theory*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1989.

[12] J.A. Thomas T.M. Cover. *Elements of Information Theory*. John Wiley and Sons, 1991.

[13] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *ijcv*, 24(2):137–154, 1997.

[14] E. Wolsztynski, E. Thierry, and L. Pronzato. Minimum-entropy estimation in semi-parametric models. *Signal Processing*, 85(5):937–949, 2005.

[15] L. Zollei. *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images*. Ph.d. thesis, MIT, 2006.