Maximum Likelihood Mosaics

Bernardo Esteves Pires, Pedro M. Q. Aguiar, Member, IEEE

Abstract

The majority of the approaches to the automatic recovery of a panoramic image from a set of partial views are suboptimal in the sense that the input images are aligned, or registered, pair by pair, e.g., consecutive frames of a video clip. These approaches lead to propagation errors that may be very severe, particularly when dealing with videos that show the same region at disjoint time intervals. Although some authors have proposed a post-processing step to reduce the registration errors in these situations, there have not been attempts to compute the optimal solution, i.e., the registrations leading to the panorama that best matches the entire set of partial views. This is our goal. In this paper, we use a generative model for the partial views of the panorama and develop an algorithm to compute in an efficient way the Maximum Likelihood estimate of all the unknowns involved: the parameters describing the alignment of all the images and the panorama itself.

Index Terms

Image alignment/registration, mosaics, panoramic imaging, featureless methods, Maximum Likelihood.

Permission to publish this abstract separately is granted.

I. INTRODUCTION

Very good paper to read. Technically sound, very clear, and very good results. My only minor comment is that the statement that no attempt has been made to find the optimal solution is not true. There are many papers tackling this issue, e.g. work of Pollefeys, zisserman’s group at oxford, some work at INRIA by various groups, the work of Peleg et al., etc.

In this paper, we address the problem of recovering, in an automatic way, a panoramic image, or a mosaic, from a set of uncalibrated partial views, e.g., a set of video frames. Modern digital video systems demand efficient solutions for this problem, e.g., for image stabilization [5], [12] and content-based representations [1]. Other application fields include virtual reality and remote sensing. The key step to the success of the automatic mosaic building is the accurate registration, or alignment, of the input images.

A. Related work

Although some authors have approached the registration problem using classical signal processing techniques, such as Fourier transforms [15], or current image analysis tools, such as integral projections [10], the majority of the papers in the literature are mostly distinguished by either requiring a low-level pre-processing step (feature-based methods) or attempting to register the images directly from their intensity levels (featureless methods).

Feature-based methods, e.g., [6], align the images by first detecting and matching a set of pointwise features. Since reliable feature points must correspond to sharp intensity corners [16], [3], this first step is hard to accomplish in a fully automatic way when processing real videos, particularly when the images are noisy, have low texture, or exhibit a small overlap among them.

In opposition, featureless methods are optimal, in the sense that they estimate the registration parameters by minimizing the difference between the image intensities in a large region, thus leading to more robust solutions to the registration of a pair of views, e.g., [11], [13]. However, when building a panorama from a

Contact author: P. Aguiar, ISR—Institute for Systems and Robotics, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal. E-mail: aguiar@isr.ist.utl.pt. His work was partially supported by FCT grant POSI/SRI/41561/2001.

B. Pires is with The Boston Consulting Group, Lisbon Office, Portugal. E-mail: pires.bernardo@bcg.com.
large set of images, practitioners usually register them sequentially, one at a time. This leads to propagation errors that may become visually noticeable if non-consecutive images cover the same region of the panorama, which is common in applications such as seabed mapping. Although some authors proposed to post-process the registration parameters to deal with this problem [11], [7], there have not been attempts to generalize the highly successful featureless methods to the multi-frame case.

B. Proposed approach: featureless global estimation

The robustness of the featureless approaches to the registration of two views motivated us to develop a featureless method to align a larger set of frames. However, it is not obvious how the two-frame cost function, usually the sum of the image square differences [11], [13], should be generalized to the multi-frame case. We were able to derive the appropriate cost function, which is an original contribution of this paper, by including as unknown, jointly with the registration parameters, the panoramic image itself.

Our approach in this paper is then to formulate the automatic recovery of mosaics from a set partial views, as a classical parameter estimation problem. The input images are modelled as noisy observations of limited regions of the unknown panorama. Naturally, since the images are uncalibrated, the problem includes as unknowns the parameters describing the registration, or alignment, of the entire set of input images. We then use Maximum Likelihood (ML) estimation. To minimize the ML cost with respect to the large set of unknowns, we propose an efficient method. First, we derive the closed-form solution for the estimate of the panorama in terms of the other unknowns (the registration parameters). Then, we plug-in the estimate of the panorama into the ML cost, obtaining an error function that depends on the registration parameters alone. This error function is a weighted sum of the square differences between all possible pairs of input images. We derive a gradient-descent algorithm to minimize this cost.

Like in the current featureless approaches to the registration of two images [11], [13], the derivatives involved in the gradient-descent algorithm to minimize our ML cost, are computed in a simple way in terms of the image gradients.

C. Paper organization

The remaining of the paper is organized as follows. In section II we formulate the registration of multiple images as a classical estimation problem. Section III deals with ML estimation for this problem, i.e., it introduces the Maximum Likelihood Mosaics (MLM) approach. In section IV we develop MLM, using the simpler case of registering a pair of images. We contrast MLM with minimizing the registration error over a fixed window, as usually done in current featureless approaches. Section V generalizes MLM to the multi-frame registration. In section VI we derive the gradient-descent algorithm to minimize the ML cost. Section VII describes experiments and section VIII concludes the paper. Preliminary versions of parts of this work are in [13], [14].

II. Problem Formulation

In this section, we develop a generative model for the partial views of an unknown panorama, and use ML to derive the estimation criterion that will allow us to recover the observed panorama, as well as the registration parameters, i.e., the viewing positions.

A. Generative model

We model each pixel of each image $I_i$, as a noisy sample of the panorama $P$. For simplicity, we consider the image domain to be the entire plane $\mathbb{R}^2$ and, to take care of the limited field of view, we define a window $H$ as $H(x, y) = 1$ in the region observed in the images and $H(x, y) = 0$ in the regions outside the camera field of view. The observation model is then

$$I_i(x_i) = [P(x_0) + R(x_i)]H(x_i),$$ (1)
where R denotes the noise, assumed i.i.d. zero-mean Gaussian, \( x_i \) are the image coordinates \((x, y)\), expressed in the coordinate system of the generic image \( I_i \), and \( x_0 \) are the corresponding coordinates of the panoramic image \( P \), expressed in its own coordinate system (which we will refer to as the reference coordinate system). Image models related to \( (1) \) have been used in the context of segmenting and tracking moving objects in video sequences \([2], [8]\).

The reference coordinate system and the coordinate system of any of the images are related by a generic parametric mapping

\[
x_i = m(\theta_i; x_0).
\]

The parameter vector \( \theta_i \) in \( (2) \) determines thus the mapping between each pixel of the panorama, with coordinates \( x_0 \), expressed in the reference coordinate system, with the corresponding pixel of image \( I_i \), with coordinates \( x_i \). Common parameterizations include translation (2 degrees of freedom (dof)), rotation (1 dof), rigid motion (3 dof), translation+rotation+zoom (4 dof), affine (6 dof), and the projective, or homography (8 dof), see, \( e.g.\), \([11], [9]\). Although our derivations are intentionally left fully generic, in the experiments, we have used the affine mapping.

### B. Estimation criterion

Given a set of \( n \) images, \( \{I_1, \ldots, I_n\} \), our goal is to recover all the unknowns involved: the panorama \( P \) and the set of parameter vectors \( \{\theta_1, \ldots, \theta_n\} \) that define the viewing positions. We use ML. From the observation model \( (1) \), after simple manipulations, we express the symmetric of the log-likelihood function as

\[
L(P, \theta_1, \ldots, \theta_n) = \frac{nN}{2} \ln (2\pi \sigma^2) + (2\sigma^2)^{-1} \sum_{x_0 \in \mathbb{R}^2} \sum_{i=1}^n [I_i(m(\theta_i; x_0)) - P(x_0)]^2 H(m(\theta_i; x_0)),
\]

where \( N \) is the number of pixels in each image and \( \sigma^2 \) is the variance of the observation noise.

### III. Maximum Likelihood Mosaics

To compute the ML estimate of all the unknowns, \( i.e.\), to carry out the minimization of the ML cost, given by the symmetric log-likelihood \( (3) \), with respect to (wrt) \( \{P, \theta_1, \ldots, \theta_n\} \), we start by noticing that the estimate of the panorama \( P \) can be expressed in closed-form as a function of the remaining unknowns.

We derive the expression for the ML estimate \( \hat{P} \) of the panorama by minimizing \( (3) \) wrt a generic pixel value \( P(x_0) \). By making zero the derivative of \( (3) \) wrt \( P(x_0) \), the estimate \( \hat{P} \) at pixel \( x_0 \) is easily obtained as a function of the set of unknown registration parameters, which we will compactly denote by \( \Theta = \{\theta_1, \ldots, \theta_n\} \):

\[
\hat{P}(x_0, \Theta) = \frac{\sum_{i=1}^n I_i(m(\theta_i; x_0)) H(m(\theta_i; x_0))}{\sum_{i=1}^n H(m(\theta_i; x_0))}.
\]

This expression shows that the estimate of the intensity of each pixel \( x_0 \) of \( \hat{P} \) is given by the average of the intensities of the corresponding pixels of all the input images that captured \( x_0 \), \( i.e.\), all the images \( I_i \) for which \( H(m(\theta_i; x_0)) = 1 \).

### IV. Two-Frame Registration

#### A. Cost-function

#### B. Minimization algorithm

It is now clear that it is not possible to evaluate the error \( e(\theta, x) \) in \( (7) \) for every pair of values of \( \theta \) and \( x \). This is the main problem of using a fixed window \( \mathcal{R} \)—registration is only possible when the overlapping region contains \( \mathcal{R} \). This limitation has a particular impact on the behavior of iterative registration algorithms. In fact, to avoid an exhaustive search, \( e.g.\), block matching, the minimization
of $E(\theta)$ in (7) is usually performed by using gradient-based algorithms that iteratively optimize $\theta$. Obviously, at every iteration of the algorithm, the overlapping region (which depends on the current estimate of $\theta$) must contain $R$, since the error $E(\theta)$, as well as its gradient, depend on a sum over $R$.

As illustrated by the first experiment of section VII, the minimum overlap requirement makes hard the automatic registration of arbitrary images. In fact, by specifying a priori a fixed window $R$, we can not cope with all possible situations. If one specifies a small $R$, it may fit into the true overlapping region, but the estimation error will be large due to the smooth minimum of $E(\theta)$. On the other hand, if one specifies a large $R$, it is not possible to register images which have a small overlap. Our goal here is to develop a method to perform the registration in the situations when the overlap between the images is not known a priori.

Instead of using a fixed window $R$, we propose an adaptive window $R_{A}(\theta)$, defined as the largest region for which it is possible to evaluate the error $e(\theta, x)$ as defined in (7). In our iterative optimization, the estimate $\theta$ is computed by refining a previous estimate $\theta_0$, i.e., $\theta = \theta_0 + \delta$. The update $\delta$ is estimated by minimizing the registration error over the adaptive window $R_{A}(\theta_0)$,

$$\delta = \arg \min_{\delta} \sum_{x \in R_{A}(\theta_0)} e^2(\theta_0 + \delta, x), \quad (5)$$

where $e(\theta, x)$ is as defined in (7). The adaptive window $R_{A}(\theta_0)$, whose size and shape depend on the current estimate $\theta_0$ of the motion parameter vector, is the overlapping region between the image $I$ and the image $I'$ registered according to $\theta_0$, i.e., $I'(m(\theta_0, x))$.

To compute the update $\delta$, we develop an adaptive-window-based Gauss-Newton method. Similar methods have been used to minimize (7), i.e., to register images using a fixed window, e.g., [11]. In this method, $e(\theta, x)$ is approximated by its first-order truncated Taylor series expansion, $e(\theta, x) \simeq e(\theta_0, x) + \delta^T \cdot \nabla \theta e(\theta_0, x)$. Using this approximation in (5), and making zero the gradient of the cost function, we get $\delta$ as the solution of the linear system

$$\left( \sum_{x \in R_{A}(\theta_0)} \nabla_{\theta} e \cdot \nabla_{\theta} e \right) \cdot \delta + \sum_{x \in R_{A}(\theta_0)} e \nabla_{\theta} e = 0, \quad (6)$$

where we omit the dependency of $e$ on $x$ and $\theta_0$ for compactness. From the definition of $e$ in (7), we see that $\nabla_{\theta} e$ in (6) is computed from the image gradient, $\nabla_{\theta} e = -\nabla_{\theta} m \cdot \nabla_{x} I'$. The initial guess for $\theta_0$ is such that $m(\theta_0, x)$ is the identity mapping, which corresponds to initializing the algorithm with zero displacement between the images, thus the initial window $R_{A}(\theta_0)$ is the entire image region.

The Gauss-Newton method just described assumes the motion is small. To cope with large displacements, we use a multiresolution scheme. In such scheme, the iterative estimation algorithm is first used in a lower resolution versions of the input images, until a certain stopping criterium is reached. The resulting parameter estimates are then used as initial guesses for the parameters in the the next (higher) resolution and the process is repeated until the original images are used.

Among the number of valid stopping criteria, we combine the two most obvious: i) the maximum number of iterations; and ii) the minimum value of the norm of the update vector $\delta$. Using only ii) is not adequate in the low resolution levels, where it is only necessary to make a coarse estimation of the parameters. In these levels, convergence may be slow and the overall performance of the algorithm is not affected if we simply perform a fixed number of iterations.

C. Impact of the window

The motion of the brightness pattern between two images $I$ and $I'$ is described by the parametric mapping $x' = m(\theta, x)$ that maps each pixel of $I$, with coordinates $x$, into the corresponding pixel $x'$
of $I'$. Featureless approaches to image registration estimate the global motion parameter vector $\theta$ by minimizing the error

$$E(\theta) = \sum_{x \in \mathcal{R}} e^2(\theta, x), \quad e(\theta, x) = I(x) - I'(m(\theta, x)),$$

where the sum is over a fixed, pre-specified, rectangular window $\mathcal{R}$. When the overlap between the images is large, the window $\mathcal{R}$ is simply chosen as a large rectangle in the interior of the image(s). However, when the overlap is small, it is difficult to select a priori an appropriate window $\mathcal{R}$, due to two reasons. First, since it is not known beforehand where the overlapping region is, it’s hard to choose a location for the window $\mathcal{R}$. Second, imposing a priori a small window, leads to less accurate estimates of $\theta$ because not only the minimum of $E(\theta)$ in (7) becomes less sharp but also the local minima phenomena become more severe.

![Graph showing the error $E(\theta)$ for different sizes of the window $\mathcal{R}$](image)

Fig. 1. Error $E(\theta)$ in (7) for different sizes of the window $\mathcal{R}$.

To illustrate the impact of the size of the window, we represent in Fig. 1 the typical evolution of $E$ in (7), as a function of a single motion parameter $\theta$, for several sizes of $\mathcal{R}$. Naturally, as anticipated above, the larger is $\mathcal{R}$, the smaller is the domain $\{\theta\}$ in which $E(\theta)$ can be evaluated. The several local minima and the smoothness of the minimum of $E(\theta)$ at the true value $\theta = 20$ in the top plots, obtained with relatively small windows, contrast with the single sharp minimum of the bottom-right plot, obtained with the largest window (note that the vertical scale is different from plot to plot).
V. Multi-frame Registration

A. Estimate of the registration parameters $\{\theta_1, \ldots, \theta_n\}$

Replacing the ML estimate $\hat{P}$ of the panorama, given by (4), in the symmetric log-likelihood (3), we express this ML cost $L$ as a function of the unknown registration parameters $\Theta$ alone. After algebraic manipulations, we get:

$$ L(\Theta) = \frac{nN}{2} \ln (2\pi\sigma^2) + (4\sigma^2)^{-1} \sum_{x_0 \in \mathbb{R}^2} W^{-1}(x_0, \Theta) \cdots$$

where $E_{ij}$ is the error between the co-registered images $I_i$ and $I_j$,

$$ E_{ij}(x_0, \theta_i, \theta_j) = I_i(m(\theta_i; x_0)) - I_j(m(\theta_j; x_0)),$$

and $W(x_0, \Theta)$ is a weight that counts the number of images that have captured the pixel $x_0$ of the panorama, according to the registration parameters in $\Theta$, i.e.,

$$ W(x_0, \Theta) = \sum_{k=1}^{n} H(m(\theta_k; x_0)).$$

By discarding from (8) the constant terms, i.e., the terms that do not depend on the unknown registration parameters $\Theta$, we conclude that the ML estimate for the problem of global multi-frame registration, is equivalent to the following minimization:

$$ \hat{\Theta} = \arg \min_{\Theta} \sum_{i,j=1}^{n} \sum_{x_0 \in \mathcal{R}_{ij}} \frac{E_{ij}^2(x_0, \theta_i, \theta_j)}{W(x_0, \Theta)}.$$

For simplicity, when deriving (11) from (8), the sums were interchanged and the spatial region of summation was re-defined to take care of the windows $H(\cdot)$ in (8), i.e., $\mathcal{R}_{ij}$ in (11) is the region where the images $I_i$ and $I_j$ overlap,

$$ \mathcal{R}_{ij} = \{x : H(m(\theta_i; x)) \cdot H(m(\theta_j; x)) = 1\}.$$

Expressions (10) and (11) condense one the contributions of this paper—they show that the ML estimate $\hat{\Theta}$ of the registration parameters $\Theta$ is given by the minimum of a particular weighted sum of the square differences between all possible pairs of co-registered input images.

VI. MLM Algorithm

Our algorithm to the minimization of the ML cost (11) uses an iterative scheme inspired in the common approaches to the two-frame problem [11], [13]. In each step, the algorithm updates a current estimate that we denote by $\Theta^0 = \{\theta_1^0, \ldots, \theta_n^0\}$.

A. Iterative minimization of the ML cost

Instead of updating the entire set of parameters $\Theta$ in a single step, which would be computationally complex, we propose a coordinatewise minimization: we update each vector $\theta_q$ at a time, keeping fixed the remaining registration parameters $\{\theta_i = \theta_i^0, i \neq q\}$. The update is $\theta_q = \theta_q^0 + \delta$, where $\delta$ is obtained from (11), after discarding the terms that do not depend on $\theta_q$:

$$ \delta = \arg \min_{\delta} \sum_{i=1}^{n} \sum_{x_0 \in \mathcal{R}_{iq}} \frac{E_{iq}^2(x_0, \theta_i^0, \theta_q^0 + \delta)}{W(x_0, \Theta^0)}.$$
To obtain a closed-form solution for the update $\delta$, we approximate the error $E_{iq}$ by its first-order Taylor series expansion,

$$E_{iq}(x_0, \theta^0_q, \theta^0_i + \delta) \approx E_{iq}(x_0, \theta^0_q, \theta^0_i) + \delta^T \cdot \nabla_{\theta} E_{iq}(x_0, \theta^0_q, \theta^0_i).$$

(14)

From the definition of $E_{iq}$ in [9], the gradient in the Taylor series expansion is easily computed in terms of the spatial gradient of image $I_q$. Furthermore, that gradient does not depend on $\theta^0_i$, thus we will denote it more compactly by $\nabla(x_0, \theta^0_q)$,

$$\nabla(x_0, \theta^0_q) = \nabla_{\theta} E_{iq}(x_0, \theta^0_q, \theta^0_i)$$

(15)

$$= -\nabla_{\theta} m(\theta^0_q, x_0) \cdot \nabla x_i (m(\theta^0_q, x_0)).$$

(16)

By inserting the Taylor series approximation in (13) and making zero the derivative wrt $\delta$, we get the update $\delta$ as the solution of a linear system

$$\Gamma(\Theta^0) \cdot \delta + \gamma(\Theta^0) = 0.$$  

(17)

The matrix $\Gamma(\Theta^0)$ and the vector $\gamma(\Theta^0)$ are obtained as

$$\Gamma(\Theta^0) = \sum_{x_0 \in \mathcal{R}_q} \nabla(x_0, \theta^0_q) \cdot \nabla^T(x_0, \theta^0_q),$$

(18)

$$\gamma(\Theta^0) = \sum_{x_0 \in \mathcal{R}_q} \nabla(x_0, \theta^0_q) \left[ \hat{P}(x_0, \Theta^0) - I_q(m(\theta^0_q, x_0)) \right],$$

(19)

where we used expression (4) for $\hat{P}$. The sums in (18,19) are over the region observed by image $I_q$, $\mathcal{R}_q = \{x : H(m(\theta^0_q, x) = 1\}$.

### B. Interpretation in terms of current algorithms

Since the iterations in standard featureless two-frame alignment algorithms [11], [13] also lead to a system like (17), we now interpret our solution (17,18,19) in terms of those approaches. Define $E_{0q}$ as the difference between image $I_q$ and the previous estimate of the panorama, obtained with the registration parameters $\Theta^0$,

$$E_{0q}(x_0, \Theta^0) = \hat{P}(x_0, \Theta^0) - I_q(m(\theta^0_q, x_0)).$$

(20)

Since the gradient of this error wrt $\theta_q$ is equal to the one defined in (15), we can re-write expressions (18,19) in terms of $E_{0q}$,

$$\Gamma(\Theta^0) = \sum_{x_0 \in \mathcal{R}_q} \nabla_{\theta} E_{0q}(x_0, \Theta^0) \cdot \nabla^T_{\theta} E_{0q}(x_0, \Theta^0),$$

(21)

$$\gamma(\Theta^0) = \sum_{x_0 \in \mathcal{R}_q} \nabla_{\theta} E_{0q}(x_0, \Theta^0) E_{0q}(x_0, \Theta^0).$$

(22)

Expressions (21,22) are equal to the ones that arise from aligning the previous estimate $\hat{P}$ of the panorama with image $I_q$, by using standard featureless methods, see e.g., [11], [13] or [3]. We thus conclude that our global approach lead to an algorithm that refines the estimate of the registration parameters of each image by using the methodology developed to register a single pair of images.
C. Convergence—initialization and multiresolution

Our algorithm starts by aligning the images sequentially, using the standard two-frame approach [11], [13]. Then, we compute an initial estimate of the panorama by using (4). After this, we cyclically refine the registrations parameters of each image. The stopping criterion may either be the error below a small threshold or reaching a maximum number of iterations.

Since the truncated Taylor series is a good approximation only when the vector $\mathbf{\theta}_q$ is close to its initial value $\mathbf{\theta}_0$, estimating the update $\delta$ from (17,18,19) leads to the convergence to the globally optimal ML estimate, only when the initial estimate is close enough to it. However, in practice, e.g., in the first experiment described below, it is common that the initial estimate of the panorama is very rough, due to the propagation of (two-frame based) registration errors. To cope with these situations, we use a coarse-to-fine approach similar to the one proposed in [4], [11]: the parameters are first estimated in the coarsest resolution level, then used as an initialization to the next finer level, until the full image resolution is attained. As illustrated in the following section, this multi-resolution approach succeeds in correcting large miss-registrations.

VII. Experiments

We describe two experiments. The first experiment compares our global approach with the current sequential registration methods. In the second experiment, we illustrate with automatic mosaic building in a seabed mapping context.

A. Adaptive window versus fixed window

To illustrate how our method performs better than the usual fixed window method, we synthesized input images by cropping a real photography and adding noise. This corresponds to a simple translational motion model, which suffices to show the advantage of using our adaptive window method.

In Fig. 2 the overlap between input images is large. The left image of Fig. 2 shows the failure of the algorithm with a fixed window of size 64. Our algorithm and the one with a fixed window of size 128 both lead to good results, see the middle and right images of Fig. 2. Note that, although these two images are visually indistinguishable, the estimate of the global motion provided by our algorithm is more accurate because it minimizes the error over the largest possible window.

![Fig. 2. Registration of a pair of images. Left: using a fixed window of size 64 (registration failure). Middle: fixed window size 128. Right: our algorithm.](image)

In Fig. 3 the overlap between input images is small, thus it is impossible to use a fixed window of a large size. When using a fixed window of size 64, the usual algorithm fails, see the left image of Fig. 3. The right image of Fig. 3 shows that our algorithm succeeds in this challenging situation.

Finally, Fig. 4 shows a mosaic obtained by pairwise registering, sequentially, a set of input images.
B. MLM versus sequential alignment

To have an exact knowledge of the ground truth, we “synthesized” the input images by cropping a real photo and adding noise. In Fig. 5, we represent the evolution of the standard two-frame featureless sequential alignment (e.g., [11], [13]) of those images. Note that the fourth image is miss-aligned and how that error propagates to the alignment of the remaining images. The (highly incorrect) panorama this way obtained, see the bottom right image of Fig. 5 was then used as the initialization for the global method we propose in this paper. After few iterations, our algorithm converged to the panoramic image shown in Fig. 6 which is visually indistinguishable from the ground truth image.

C. Underwater mosaic for seabed mapping

As a final example, we illustrate our method with automatic mosaic construction from video images, captured by an underwater camera in the sea. Although underwater images are particularly difficult to align, due to the absence of salient features, the mosaic recovered by our algorithm is visually correct, see Fig. 8.

As a final example, we use images captured by an underwater camera in the sea. Fig. 7 shows four of those images. The low texture and the almost total absence of salient feature points make these images particularly challenging. Note also that although the overlapping region between images is not very small, its shape is not rectangular. In this situation, the traditional fixed window method would use a small rectangular window inside the overlapping region, thus failing to use all the information available.

In Fig. 8 we represent the seabed mosaic obtained by using our algorithm to sequentially register the images of Fig. 7.
VIII. CONCLUSION

We proposed a new method to build a panoramic image from a set of partial views. Rather than composing the input images in an incremental way, our approach seeks the global solution to the estimation problem, i.e., it computes the panorama that best matches all the partial observations. To minimize the global cost, we derived an efficient gradient descent algorithm that generalizes the current most robust two-frame featureless registration approaches.

REFERENCES

[1] P. Aguiar, R. Jasinschi, J. Moura, and C. Pluempitwiriayawej, “Content-based image sequence representation,” in Digital Video Processing, T. Reed, Ed., CRC Press, 2004.
[2] P. Aguiar and J. Moura, “Detecting and solving template ambiguities in motion segmentation,” in IEEE ICIP, 1997.
[3] ———, “Image motion estimation – convergence and error analysis,” in IEEE ICIP, Greece, 2001.
[4] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani, “Hierarchical model-based motion estimation,” in European Conf. on Computer Vision, Santa Margherita Ligure, Italy, 1992.
[5] F. Dufaux and J. Konrad, “Efficient, robust, and fast global motion estimation for video coding,” IEEE T-IP, 2000.
[6] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
[7] D. Hasler, L. Sbaiz, S. Ayer, and M. Vetterli, “From local to global parameter estimation in panoramic photographic reconstruction,” in IEEE ICIP, Kobe, Japan, 1999.
[8] N. Jojic and B. Frey, “Learning flexible sprites in video layers,” in IEEE Int. Conf. on CVPR, Hawaii, 2001.
[9] D. Kim and K. Hong, “Fast global registration for image mosaicing,” in IEEE ICIP, Barcelona, Spain, 2003.
[10] J. Lee and J. Ra, “Block motion estimation based on selective integral projections,” in IEEE ICIP, Rochester, USA, 2002.
[11] S. Mann and R. Piccard, “Video orbits of the projective group: a simple approach to featureless estimation of parameters,” IEEE Trans. on Image Processing, 1997.
[12] N. Petrovic, N. Jojic, and T. Huang, “Hierarchical video clustering,” in IEEE MMSP, Siena, Italy, 2004.
[13] B. Pires and P. Aguiar, “Registration of images with small overlap,” in Proc. of the IEEE Multimedia Signal Processing Workshop, Siena, Italy, 2004.
[14] ———, “Featureless global alignment of multiple images,” January 2005, submitted to IEEE Int. Conf. on Image Processing.
[15] B. Reddy and B. Chattery, “An FFT-based technique for translation, rotation, and scale-invariant image registration,” IEEE Trans. on Image Processing, 1996.
[16] J. Shi and C. Tomasi, “Good features to track,” in IEEE Int. Conf. on Computer Vision and Pattern Recognition, 1994.
Fig. 5. Sequential registration. Note how the misalignment of the fourth image (middle left) propagates to the remaining ones.
Fig. 6. Proposed approach. Final estimate of the panorama, when our algorithm is initialized with the bottom right image of Fig. 5.

Fig. 7. Sample underwater images.
Fig. 8. Mosaic built by using our algorithm to register the images of Fig. 7.
