Registration of Images with Small Overlap

Bernardo Esteves Pires and Pedro M. Q. Aguiar

Institute for Systems and Robotics—ISR/IST

Av. Rovisco Pais, 1040-001, Lisboa, Portugal

E-mail: {bpires,aguiar}@isr.ist.utl.pt

Abstract— We address the problem of building a mosaic from uncalibrated images. The key issue in this problem is the registration of the images, *i.e.*, the estimation of the parameters describing the global motion of the brightness pattern between them. Current approaches to global motion estimation avoid pre-processing steps, *e.g.*, feature detection and matching, by inferring the motion parameters directly from the image intensities, through the minimization of the sum of the square error over a fixed window. This window is usually chosen as a large rectangle within the region where it is expected that the original images overlap. Although this leads to good results when the images to register have a large overlap, it often fails when the images exhibit only a small overlap, due to the requirement of choosing *a priori* a very small window. To overcome this limitation, we propose in this paper an adaptive window.

I. INTRODUCTION

The problem of registering, or aligning, image data, finds applications in such diverse fields as remote sensing, virtual reality or digital video. While in remote sensing and virtual reality, image registration is necessary to build large images from partial views, in digital video, it is used for stabilization and compression [1].

The key step in image registration is the estimation of the global motion of the brightness pattern. Under common assumptions about the scene geometry or the camera motion, the global motion is described by a small set of parameters, see for example [2], [3] for different parameterizations. Among the approaches to image registration, two very distinct lines of thought are found in the literature. The so-called feature-based methods compute the motion parameters by first detecting and matching features across the images [4]. In opposition, image-based methods avoid pre-processing steps by estimating the motion parameters directly from the image intensity values [2]. Other techniques include the use of Fourier transforms [5] and integral projections [6].

Once feature correspondences are established, global motion estimation is straightforward, regardless of the size of the overlapping region. However, the bottleneck of the feature-based methods is the feature matching step, which becomes a tremendous quagmire in several practical situations, *e.g.*, when processing low textured images. Image-based methods overcome this limitation by estimating the motion parameters through the minimization of the sum of the square error between the images over a pre-specified large region. To avoid

an exhaustive search, like in block matching, the minimizer is usually computed by using a Gauss-Newton method and a multiresolution scheme [2]. Although these methods have been successfully used in applications in which the images have a large overlap, *e.g.*, video stabilization [1], they often fail when the images to register exhibit only a small overlap, due to the need to select *a priori* a very small region.

In this paper, we propose a featureless method that uses an adaptive window. We show that the adaptive window method succeeds in registering uncalibrated images even when the overlapping region is very small.

II. IMPACT OF THE WINDOW

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

$$E(\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathcal{R}} e^2(\boldsymbol{\theta}, \mathbf{x}), \quad e(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{I}(\mathbf{x}) - \mathbf{I}'(\mathbf{m}(\boldsymbol{\theta}, \mathbf{x})), \quad (1)$$

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, its hard to choose a location for the window $\mathcal{R}$. Second, imposing *a priori* a small window, leads to less accurate estimates of $\boldsymbol{\theta}$ because not only the minimum of $E(\boldsymbol{\theta})$ in (1) becomes less sharp but also the local minima phenomena become more severe.

To illustrate the impact of the size of the window, we represent in Fig. 1 the typical evolution of E in (1), as a function of a single motion parameter θ, 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).
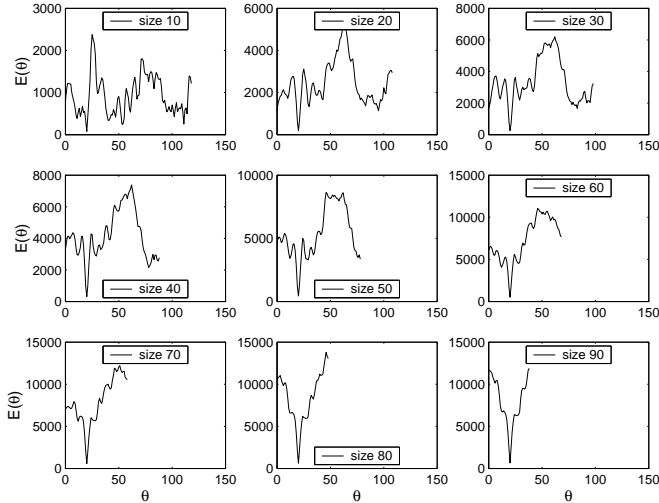
Fig. 1. Error $E(\boldsymbol{\theta})$ in (1) for different sizes of the window $\mathcal{R}$.

III. REGISTRATION ALGORITHM: ADAPTIVE WINDOW

It is now clear that it is not possible to evaluate the error $e(\boldsymbol{\theta}, \mathbf{x})$ in (1) for every pair of values of $\boldsymbol{\theta}$ and $\mathbf{x}$. 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 $E(\boldsymbol{\theta})$ in (1) is usually performed by using gradient-based algorithms that iteratively optimize $\boldsymbol{\theta}$. Obviously, at every iteration of the algorithm, the overlapping region (which depends on the current estimate of $\boldsymbol{\theta}$) must contain $\mathcal{R}$, since the error $E(\boldsymbol{\theta})$, as well as its gradient, depend on a sum over $\mathcal{R}$.

As illustrated by the first experiment of section VI, the minimum overlap requirement makes hard the automatic registration of arbitrary images. In fact, by specifying *a priori* a fixed window $\mathcal{R}$, we can not cope with all possible situations. If one specifies a small $\mathcal{R}$, it may fit into the true overlapping region, but the estimation error will be large due to the smooth minimum of $E(\boldsymbol{\theta})$. On the other hand, if one specifies a large $\mathcal{R}$, 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 $\mathcal{R}$, we propose an adaptive window $\mathcal{R}_A(\boldsymbol{\theta})$, defined as the *largest region* for which it is possible to evaluate the error $e(\boldsymbol{\theta}, \mathbf{x})$ as defined in (1). In our iterative optimization, the estimate $\boldsymbol{\theta}$ is computed by refining a previous estimate $\boldsymbol{\theta}_0$, *i.e.*, $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{\delta}$. The update $\boldsymbol{\delta}$ is estimated by minimizing the registration error over the adaptive window $\mathcal{R}_A(\boldsymbol{\theta}_0)$,

$$\widehat{\boldsymbol{\delta}} = \arg \min_{\boldsymbol{\delta}} \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} e^2(\boldsymbol{\theta}_0 + \boldsymbol{\delta}, \mathbf{x}), \qquad (2)$$

where $e(\boldsymbol{\theta}, \mathbf{x})$ is as defined in (1). The adaptive window $\mathcal{R}_A(\boldsymbol{\theta}_0)$, whose size and shape depend on the current estimate $\boldsymbol{\theta}_0$ of the motion parameter vector, is the overlapping region between the image $\mathbf{I}$ and the image $\mathbf{I}'$ registered according to $\boldsymbol{\theta}_0$, $\mathbf{I}'(\mathbf{m}(\boldsymbol{\theta}_0, \mathbf{x}))$.

To compute the update $\boldsymbol{\delta}$, we develop an adaptive-window-based Gauss-Newton method. Similar methods have been used to minimize (1), *i.e.*, to register images using a fixed window, *e.g.*, [2]. In this method, $e(\boldsymbol{\theta}, \mathbf{x})$ is approximated by its first-order truncated Taylor series expansion, $e(\boldsymbol{\theta}, \mathbf{x}) \simeq e(\boldsymbol{\theta}_0, \mathbf{x}) + \boldsymbol{\delta}^T \cdot \nabla_{\boldsymbol{\theta}} e(\boldsymbol{\theta}_0, \mathbf{x})$. Using this approximation in (2), and making zero the gradient of the cost function, we get $\widehat{\boldsymbol{\delta}}$ as the solution of the linear system

$$\left(\sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} \nabla_{\boldsymbol{\theta}} e \cdot \nabla_{\boldsymbol{\theta}}^T e \right) \cdot \widehat{\boldsymbol{\delta}} + \sum_{\mathbf{x} \in \mathcal{R}_A(\boldsymbol{\theta}_0)} e \, \nabla_{\boldsymbol{\theta}} e = \mathbf{0}, \quad (3)$$

where we omit the dependency of e on $\mathbf{x}$ and $\boldsymbol{\theta}_0$ for compactness. From the definition of e in (1), we see that $\nabla_{\boldsymbol{\theta}} e$ in (3) is computed from the image gradient, $\nabla_{\boldsymbol{\theta}} e = -\nabla_{\boldsymbol{\theta}} \mathbf{m} \cdot \nabla_{\mathbf{x}} \mathbf{I}'$. The initial guess for $\boldsymbol{\theta}_0$ is such that $\mathbf{m}(\boldsymbol{\theta}_0, \mathbf{x})$ is the identity mapping, which corresponds to initializing the algorithm with zero displacement between the images, thus the initial window $\mathcal{R}_A(\boldsymbol{\theta}_0)$ is the entire image region.

IV. MULTIRESOLUTION SCHEME

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 $\widehat{\boldsymbol{\delta}}$. 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.

V. AFFINE MOTION MODEL

We use the affine motion model: $\mathbf{x}' = \mathbf{m}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$. The parameter vector is $\boldsymbol{\theta} = [a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2]^T$ and the image coordinates are linearly related by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \qquad (4)$$

For the affine model (4), the gradient of the error $e(\boldsymbol{\theta}, \mathbf{x})$ with respect to the motion parameter vector $\boldsymbol{\theta}$ is given by

$$\nabla_{\boldsymbol{\theta}} e(\mathbf{x}) = - \left[x\mathbf{I}'_x(\mathbf{x}), y\mathbf{I}'_x(\mathbf{x}), x\mathbf{I}'_y(\mathbf{x}), y\mathbf{I}'_y(\mathbf{x}), \mathbf{I}'_x(\mathbf{x}), \mathbf{I}'_y(\mathbf{x}) \right]^T$$

and the update $\widehat{\boldsymbol{\delta}} = [\widehat{\delta}_{a_{11}}, \widehat{\delta}_{a_{12}}, \widehat{\delta}_{a_{21}}, \widehat{\delta}_{a_{22}}, \widehat{\delta}_{b_1}, \widehat{\delta}_{b_2}]^T$ of the parameter vector $\boldsymbol{\theta}$ is obtained from (3) as

$$\mathbf{D}\widehat{\boldsymbol{\delta}} = \mathbf{d}, \quad \mathbf{d} = \sum_{\mathcal{R}_A} \left[\begin{array}{cccccc} xe\mathbf{I}'_x & ye\mathbf{I}'_x & xe\mathbf{I}'_y & ye\mathbf{I}'_y & e\mathbf{I}'_x & e\mathbf{I}'_y \end{array} \right]^T,$$

$$\mathbf{D} = \sum_{\mathcal{R}_A} \left[\begin{array}{cccccc} x^2\mathbf{I}'^2_x & xy\mathbf{I}'^2_x & x^2\mathbf{I}'_x\mathbf{I}'_y & xy\mathbf{I}'_x\mathbf{I}'_y & x\mathbf{I}'^2_x & x\mathbf{I}'_x\mathbf{I}'_y \\ xy\mathbf{I}'^2_x & y^2\mathbf{I}'^2_x & xy\mathbf{I}'_x\mathbf{I}'_y & y^2\mathbf{I}'_x\mathbf{I}'_y & y\mathbf{I}'^2_x & y\mathbf{I}'_x\mathbf{I}'_y \\ x^2\mathbf{I}'_x\mathbf{I}'_y & xy\mathbf{I}'_x\mathbf{I}'_y & x^2\mathbf{I}'^2_y & xy\mathbf{I}'^2_y & x\mathbf{I}'_x\mathbf{I}'_y & x\mathbf{I}'^2_y \\ xy\mathbf{I}'_x\mathbf{I}'_y & y^2\mathbf{I}'_x\mathbf{I}'_y & xy\mathbf{I}'^2_y & y^2\mathbf{I}'^2_y & y\mathbf{I}'_x\mathbf{I}'_y & y\mathbf{I}'^2_y \\ x\mathbf{I}'^2_x & y\mathbf{I}'^2_x & x\mathbf{I}'_x\mathbf{I}'_y & y\mathbf{I}'_x\mathbf{I}'_y & \mathbf{I}'^2_x & \mathbf{I}'_x\mathbf{I}'_y \\ x\mathbf{I}'_x\mathbf{I}'_y & y\mathbf{I}'_x\mathbf{I}'_y & x\mathbf{I}'^2_y & y\mathbf{I}'^2_y & \mathbf{I}'_x\mathbf{I}'_y & \mathbf{I}'^2_y \end{array} \right].$$

Again, we omit the dependency of $e, \mathbf{I}'_x$ and $\mathbf{I}'_y$ on $\boldsymbol{\theta}$ and $\mathbf{x}$.

Although simply initializing the iterative process with an identity mapping $\mathbf{x}' = \mathbf{x}$, *i.e.*, with $\mathbf{A} = \mathbf{I}_{2\times 2}, \mathbf{b} = \mathbf{0}_{2\times 1}$, may be enough in particular cases, we improve the convergence of the algorithm by first computing a rough estimate of the parameters in $\mathbf{b}$ using a pure translational model, *i.e.*, by keeping fixed $\mathbf{A} = \mathbf{I}_{2\times 2}$ in this first phase.

VI. EXPERIMENTS

We present experiments using three kinds of input images. In subsection VI-A, we synthesize input images from a photography. In subsection VI-B, we use real images obtained with an hand-held camera. Finally, in subsection VI-C, we use a set of very challenging low textured underwater images.

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.

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

Fig. 3. Registration of images with very small overlap. Left: using a fixed window of size 64 (registration failure). Right: our algorithm.

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.
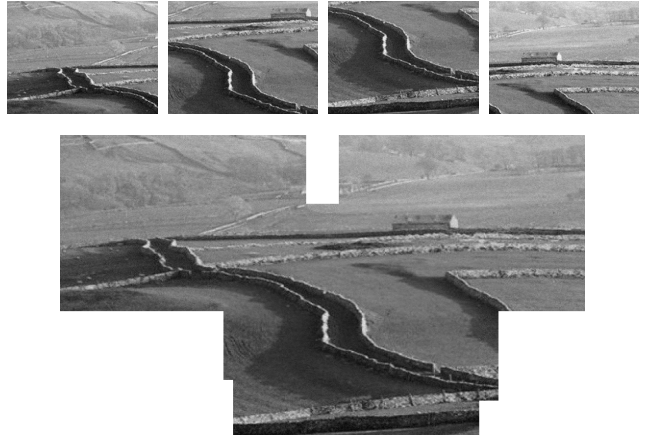

Fig. 4. Mosaic of images with small overlap. Top: original images. Bottom: mosaic built by using our algorithm to register those images.

B. Registration of uncalibrated images with small overlap

We now illustrate the behavior of the multiresolution scheme by using two images obtained with a hand-held camera. These images, shown in the bottom row of Fig. 5, have a small overlapping region. Note also that there are almost no salient feature points inside this overlapping region, making very hard the use of feature-based registration methods.

Each row of images in Fig. 5 corresponds to a different resolution level. We represent both the low resolution versions of the input images and the low resolution mosaic obtained by estimating the motion parameters at the corresponding resolution level. In the top, we show the lowest resolution level, where the images are blurred and the resulting registration is coarse. From top to bottom, the resolution increases. In the bottom, we represent the original input images and the final accurate result of the registration.

Fig. 6. Sample underwater images.

Fig. 5. Three levels of the multiresolution scheme. From top to bottom, the resolution increases. Left column: pairs of images at each resolution level. Right column: registration of the images on the left.

C. Underwater mosaic

As a final example, we use images captured by an underwater camera in the sea. Fig. 6 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. 7, we represent the seabed mosaic obtained by using our algorithm to sequentially register the images of Fig. 6.

VII. CONCLUSION

We proposed an adaptive window to estimate in a featureless way the parameters describing the global motion of the image brightness pattern.

The experiments have shown that our method succeeds in registering uncalibrated images even when they have small overlap and low texture.
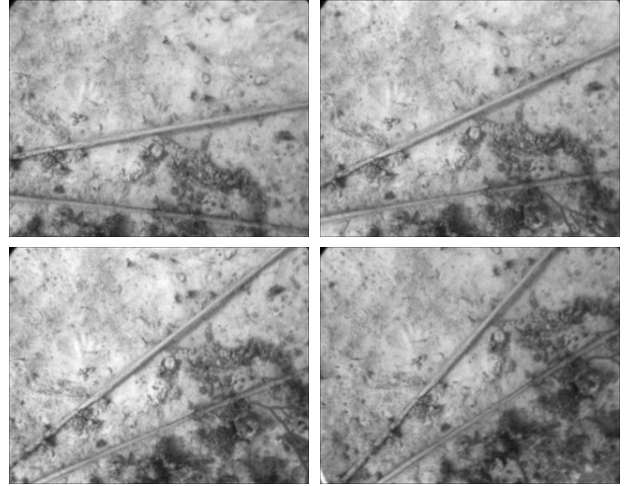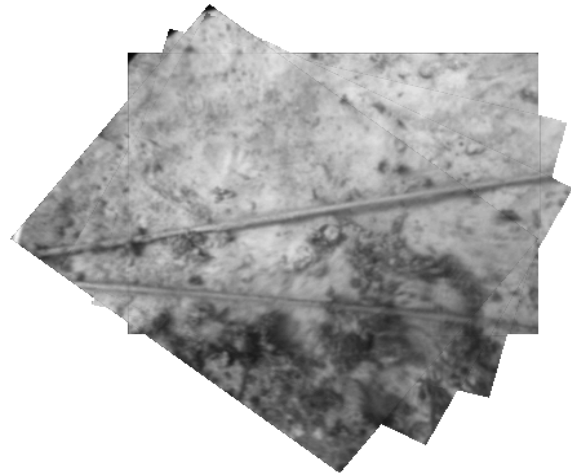

Fig. 7. Mosaic built by using our algorithm to register the images of Fig. 6.

REFERENCES

[1] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE T-IP*, 2000.
[2] S. Mann and R. Piccard, "Video orbits of the projective group: a simple approach to featureless estimation of parameters," *IEEE T-IP*, 1997.
[3] D. Kim and K. Hong, "Fast global registration for image mosaicing," in *Proc. of IEEE ICIP*, 2003.
[4] J. Mundy and A. Zisserman, Eds., *Geometric Invariance in Machine Vision*, MIT Press, 1992.
[5] B. Reddy and B. Chattery, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE T-IP*, 1996.
[6] J. Lee and J. Ra, "Block motion estimation based on selective integral projections," in *Proc. of IEEE ICIP*, 2002.