UNSUPERVISED SIMULTANEOUS REGISTRATION AND EXPOSURE CORRECTION

Pedro M. Q. Aguiar

Institute for Systems and Robotics / IST
Lisboa, Portugal
E-mail: aguiar@isr.ist.utl.pt

ABSTRACT

Early approaches to building mosaics by composing photographic images, assume the input images have similar exposures. Since this is unlikely to happen in practice, it became common to compensate for different exposures in the blending step, after the images have been registered, or aligned [1]. However, registration methods usually assume brightness constancy and fail to align images with different exposures. Recent approaches to this problem lead to computationally complex solutions that require either robust statistics or nonlinear optimization. In this paper we propose a computationally simple method to jointly estimate the registration parameters and the parameters describing the exposure correction, directly from the image intensity values. We obtain closed-form solutions for the estimates of the exposure parameters. This enables the derivation of a simple two-step iterative algorithm to minimize the global cost. Our experiments show that this algorithm succeeds to register real images exhibiting simultaneously very distinct orientations and exposures.

***Index Terms*—** Image registration, Image restoration

1. INTRODUCTION

Two very distinct approaches to image registration, or alignment, have been followed in the past. The so-called feature-based approaches first attempt to find the correspondences between feature points, usually intensity corners, see *e.g.* [2]. Featureless methods avoid this intermediate stage by estimating the registration parameters directly from image intensity values, see *e.g.* [3, 4]. The majority of these methods assume brightness constancy, *i.e.*, that the input images are equally exposed. This does not happen in many practical scenarios, see an illustrative example in Fig. 1. In this paper, we address the problem of automatically registering this kind of photos.

Naturally, in feature-based based approaches, the most difficult task is the feature matching step. When the images have different exposures, feature matching becomes harder, because local brightness patterns differ from one image to the other. To address this difficulty, researchers have proposed time-consuming algorithms that use robust features, such as Zernike moments, and robust statistics (RANSAC) [5], or multi-stage processing schemes [6]. Featureless approaches cope with differently exposed images by capturing the brightness change into the observation model, *i.e.*, by generalizing the optical flow equation. The joint estimation of the parameters describing the geometric registration of the images and the parameters describing the brightness change, leads to the use of computationally expensive nonlinear optimization [7].

Fig. 1. Two photographs of the same scene. Left: "natural" orientation and exposure. Right: tilted and mush darker view.

The approach we propose in this paper is featureless. To minimize the cost function that arises from Maximum Likelihood estimation, we propose an algorithm that exploits the fact that the observation model is linear in the parameters describing the brightness change. We obtain closed-form solutions for the estimates of these parameters and incorporate them into a computationally simple algorithm, where the brightness parameters and the geometric registration parameters are computed in alternate steps. Experiments show that our algorithm is able to automatically align real images, even when they have very distinct orientations and/or exposures.

Paper organization Section 2 briefly overviews the usual approach to the featureless registration of images. In section 3, we describe our method to simultaneously perform exposure correction and registration. Section 4 contains experiments that demonstrate the effectiveness of our approach and section 5 concludes the paper. A MATLAB implementation of the algorithm we propose in this paper is made available in [8].

2. FEATURELESS REGISTRATION OF IMAGES

Two photographs of the same scene are taken from different viewpoints. The model underlying the operation of registering, or aligning, the resulting images $\mathbf{I}_1$ and $\mathbf{I}_2$, simply states that a pixel $\mathbf{x}_1 = [x_1, y_1]^T$ in $\mathbf{I}_1$ and a pixel $\mathbf{x}_2 = [x_2, y_2]^T$ in $\mathbf{I}_2$ that "observe" the same point of the scene, are related by $\mathbf{x}_2 = \mathbf{m}(\boldsymbol{\theta}; \mathbf{x}_1)$, where $\boldsymbol{\theta}$ is vector collecting a (small) set of parameters. Thus, the image intensity levels satisfy

$$\mathbf{I}_1(\mathbf{x}) \simeq \mathbf{I}_2(\mathbf{m}(\boldsymbol{\theta}; \mathbf{x})) \tag{1}$$

and $\mathbf{m}(\boldsymbol{\theta}; \mathbf{x})$ describes the two-dimensional motion of the brightness pattern in the image plane, *i.e.*, $\boldsymbol{\theta}$ characterizes the registration of $\mathbf{I}_1$ and $\mathbf{I}_2$. The dimension of $\boldsymbol{\theta}$ depends on the motion model, which in turn depends on the kind of camera motions and/or scene geometries allowed, see *e.g.* [2]. Common choices are the 2-parameter pure translational, the 4-parameter rigid motion (translation, rotation and zoom), the 6-parameter affine, and the 8-parameter projective models.

In featureless approaches to image registration, the estimate of $\boldsymbol{\theta}$ is computed by minimizing the sum of the square differences between the image intensities, see *e.g.* [3, 4],

$$e(\boldsymbol{\theta}; \mathbf{x}) = \mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2\left(\mathbf{m}(\boldsymbol{\theta}; \mathbf{x})\right), \tag{2}$$

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{\mathbf{x}} e^2(\boldsymbol{\theta}; \mathbf{x}). \tag{3}$$

The estimate $\widehat{\boldsymbol{\theta}}$ in (3) is usually computed by using a Gauss-Newton method where, in each iteration, a previous guess $\boldsymbol{\theta}_0$ is updated, *i.e.*, $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \widehat{\boldsymbol{\delta}}$. In this method, $e(\mathbf{x}, \boldsymbol{\theta})$ 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 (3), and making zero the gradient of the cost function, $\widehat{\boldsymbol{\delta}}$ is obtained as the solution of the linear system

$$\boldsymbol{\Gamma}\left(\boldsymbol{\theta}_0\right) \cdot \widehat{\boldsymbol{\delta}} + \boldsymbol{\gamma}\left(\boldsymbol{\theta}_0\right) = \mathbf{0}, \tag{4}$$

where the matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)$ and the vector $\boldsymbol{\gamma}\left(\boldsymbol{\theta}_0\right)$ are given by

$$\boldsymbol{\Gamma}\left(\boldsymbol{\theta}_0\right) = \sum_{\mathbf{x}} \nabla_{\boldsymbol{\theta}}\, e(\boldsymbol{\theta}_0; \mathbf{x}) \cdot \nabla_{\boldsymbol{\theta}}^T\, e(\boldsymbol{\theta}_0; \mathbf{x}), \tag{5}$$

$$\boldsymbol{\gamma}\left(\boldsymbol{\theta}_0\right) = \sum_{\mathbf{x}} e(\boldsymbol{\theta}_0; \mathbf{x})\, \nabla_{\boldsymbol{\theta}}\, e(\boldsymbol{\theta}_0; \mathbf{x}), \tag{6}$$

see [4]. Note that, from the definition of $e(\boldsymbol{\theta}; \mathbf{x})$ in (2), we see that its gradient $\boldsymbol{\nabla}_{\boldsymbol{\theta}}\, e(\boldsymbol{\theta}_0; \mathbf{x})$ in (5,6) is easily computed from the image spatial gradient as $\boldsymbol{\nabla}_{\boldsymbol{\theta}}\, e = -\boldsymbol{\nabla}_{\boldsymbol{\theta}}\, \mathbf{m} \cdot \boldsymbol{\nabla}_{\mathbf{x}} \mathbf{I}_2$.

3. REGISTRATION AND EXPOSURE CORRECTION: TWO-STEP ITERATIVE OPTIMIZATION

The method outlined in the previous section avoids the quagmire of selecting and matching pointwise features, which made it very popular in the recent past. However, since its robustness is rooted on the brightness constancy imposed by the model (1), it can not be used to register images with distinct exposures. In this case, the image intensity levels are related by a more general model, which is well approximated by

$$\mathbf{I}_1(\mathbf{x}) \simeq \alpha\, \mathbf{I}_2(\mathbf{m}(\boldsymbol{\theta}; \mathbf{x})) + \beta, \tag{7}$$

where the gain α and offset β account for the change of exposure from image $\mathbf{I}_1$ to image $\mathbf{I}_2$, see *e.g.* [9].

To align images $\mathbf{I}_1$ and $\mathbf{I}_2$ under this scenario, we need to jointly estimate both the registration parameters in $\boldsymbol{\theta}$ and the exposure parameters α and β. When the observation noise is white Gaussian, the Maximum Likelihood estimate of these parameters leads to the joint minimization

$$\left\{\widehat{\boldsymbol{\theta}}, \widehat{\alpha}, \widehat{\beta}\right\} = \arg\min_{\boldsymbol{\theta}, \alpha, \beta} E(\boldsymbol{\theta}, \alpha, \beta), \tag{8}$$

where the cost $E(\boldsymbol{\theta}, \alpha, \beta)$ is a generalization of the one in (2,3):

$$e(\boldsymbol{\theta}, \alpha, \beta; \mathbf{x}) = \mathbf{I}_1(\mathbf{x}) - \alpha\, \mathbf{I}_2\left(\mathbf{m}(\boldsymbol{\theta}; \mathbf{x})\right) - \beta, \tag{9}$$

$$E(\boldsymbol{\theta}, \alpha, \beta) = \sum_{\mathbf{x}} e^2(\boldsymbol{\theta}, \alpha, \beta; \mathbf{x}). \tag{10}$$

Two-step iterative minimization We now address the non-linear minimization (8). At a first sight, it might seem attractive to generalize the approach outlined in the previous section, *i.e.*, to develop a Gauss-Newton method to estimate the full set of unknowns $\{\boldsymbol{\theta}, \alpha, \beta\}$. This kind of approach was followed by Y. Altunbasak *et al* [7] in a framework that included as unknowns also the lens distortion parameters and more general illumination changes. Rather than enlarging the vector of unknowns of the Gauss-Newton method, in this paper we propose an approach that takes advantage of the fact that the error (9) is linear in the exposure unknowns $\{\alpha, \beta\}$, leading to a very simple algorithm.

We minimize (8) by using a two-step iterative method, a strategy that has succeed in several optimization problems, *e.g.*, Expectation-Maximization algorithms [10]. In one of the steps, $\boldsymbol{\theta}$ is kept fixed and (8) is minimized with respect to (wrt) α and β, leading to a closed-form solution. In the other step, α and β are kept fixed and (8) is minimized wrt $\boldsymbol{\theta}$, leading to the problem addressed in the previous section.

Estimation of α and β for fixed $\boldsymbol{\theta}$ From the definition of the cost E in (9,10), its derivatives wrt to α and β are

$$\frac{\partial E}{\partial \alpha} = -2 \sum_{\mathbf{x}} \mathbf{I}_2\left(\mathbf{m}(\boldsymbol{\theta}; \mathbf{x})\right) e(\boldsymbol{\theta}, \alpha, \beta; \mathbf{x}), \tag{11}$$

$$\frac{\partial E}{\partial \beta} = -2 \sum_{\mathbf{x}} e(\boldsymbol{\theta}, \alpha, \beta; \mathbf{x}). \tag{12}$$

The estimates $\widehat{\alpha}$ and $\widehat{\beta}$ for fixed $\boldsymbol{\theta}$ are found by minimizing (8) wrt α and β, *i.e.*, by making zero the derivatives (11,12):

$$\frac{\partial E}{\partial \alpha} = 0 \quad \Leftrightarrow \quad \sum \mathbf{I}_1 \mathbf{I}_2 - \alpha \sum \mathbf{I}_2^2 - \beta \sum \mathbf{I}_2 = 0, \tag{13}$$

$$\frac{\partial E}{\partial \beta} = 0 \quad \Leftrightarrow \quad \sum \mathbf{I}_1 - \alpha \sum \mathbf{I}_2 - \beta N = 0, \tag{14}$$

where N is the number of pixels in the region of summation of the cost E and the dependency on $\mathbf{x}$ is omitted for compactness. Since equations (13,14) are linear in α and β, we get closed-form solutions for the estimates $\widehat{\alpha}$ and $\widehat{\beta}$:

$$\widehat{\alpha} = \frac{N \sum \mathbf{I}_1 \mathbf{I}_2 - \sum \mathbf{I}_1 \sum \mathbf{I}_2}{N \sum \mathbf{I}_2^2 - \left(\sum \mathbf{I}_2\right)^2}, \qquad (15)$$

$$\widehat{\beta} = \frac{\sum \mathbf{I}_1 \sum \mathbf{I}_2^2 - \sum \mathbf{I}_2 \sum \mathbf{I}_1 \mathbf{I}_2}{N \sum \mathbf{I}_2^2 - \left(\sum \mathbf{I}_2\right)^2}. \qquad (16)$$

Estimation of $\boldsymbol{\theta}$ for fixed α and β Compensate the exposure of image $\mathbf{I}_2$ according to the current estimates of α and β, *i.e.*, define an image $\mathbf{I}_2'$ related to $\mathbf{I}_2$ by $\mathbf{I}_2'(\mathbf{x}) = \alpha \mathbf{I}_2(\mathbf{x}) + \beta$. Minimizing (8) wrt $\boldsymbol{\theta}$ is exactly the same as registering images $\mathbf{I}_1$ and $\mathbf{I}_2'$ (just note that the error (9) is written in terms of $\mathbf{I}_2'$ as $e(\boldsymbol{\theta}, \alpha, \beta; \mathbf{x}) = \mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2'(\mathbf{m}(\boldsymbol{\theta}; \mathbf{x}))$ and compare this expression to (2)). Thus, we compute $\widehat{\boldsymbol{\theta}}$ by using the method described in the previous section, now with images $\mathbf{I}_1$ and $\mathbf{I}_2'$.

Coarse-to-fine estimation The truncated Taylor series involved in the estimation of $\boldsymbol{\theta}$ is a good approximation only when the initial guess $\boldsymbol{\theta}_0$ is close to $\widehat{\boldsymbol{\theta}}$. To cope with large displacements between the images (positions and orientations), we use coarse-to-fine estimation, a multiresolution approach typical of featureless methods, *e.g.* [3, 4]. Our algorithm starts with very low resolution versions of the input images and gradually increases their detail, until the original full resolution is reached. Besides being crucial to guarantee good convergence, this coarse-to-fine estimation also reduces the computational cost. In fact, since an approximate estimate is achieved by iterating at very low resolution, *i.e.*, at a low computational cost, the subsequent refinement of that estimate requires only very few iterations at larger resolution levels.

Initialization and stopping criteria According to our experience, the two-step algorithm coupled with the coarse-to-fine strategy just described, exhibits good convergence even with the trivial initialization of guessing that the input images are equally exposed and aligned. The initial guess is then $\alpha_0 = 1$, $\beta_0 = 0$, and $\boldsymbol{\theta}_0$ is such that $\mathbf{m}(\boldsymbol{\theta}_0; \mathbf{x}) = \mathbf{x}$. The algorithm stops when the estimates updates are below a small threshold.

4. EXPERIMENTS

We used the method just described to align the images of Fig. 1. To illustrate the behavior of the two-step algorithm, we represent in Fig. 2, from left to right, top to bottom, the evolution of the registration of the right image according to successive estimates of the registration parameters in $\boldsymbol{\theta}$ and exposure parameters α and β. The top left image of of Fig. 2 shows the lower resolution version of the right image of Fig. 1, see how the orientation and exposure of these images is the same. As the registration process evolves, the orientation and the exposure of the successive images in Fig. 2 converges to the ones that best match the left image of Fig. 1. The final result, obtained at the original full resolution, is shown in the bottom right image of Fig. 2.

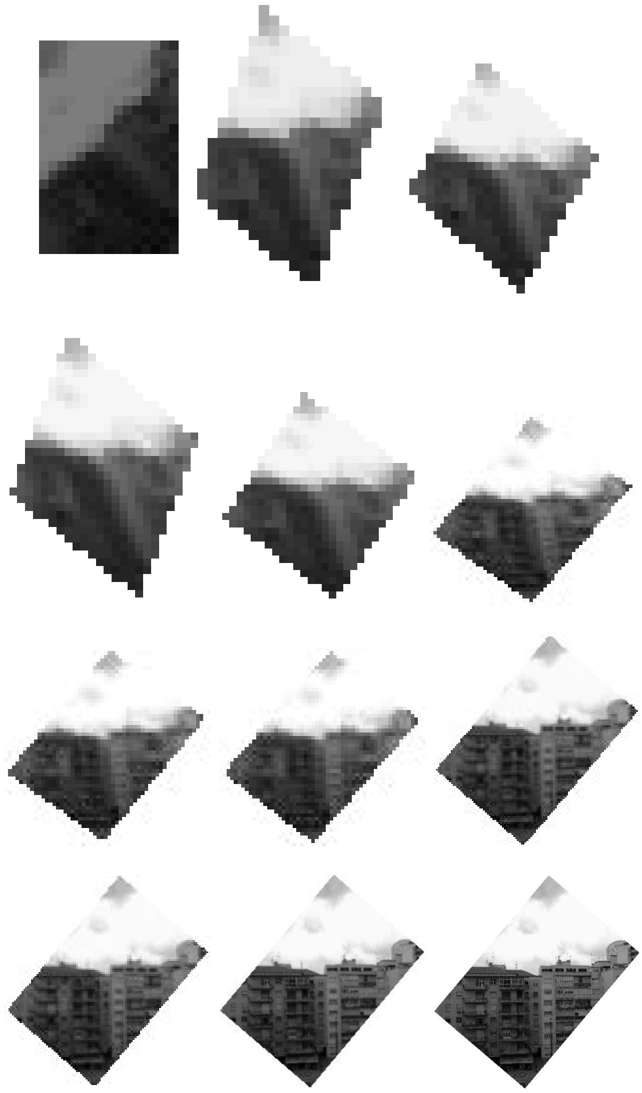

Fig. 2. Evolution of the two-step iterative algorithm when processing the pair of images of Fig. 1. From left to right, top to bottom, we represent the registration of the right image, at increasing resolution levels, according to the estimates of the parameters obtained as the algorithm evolves.

In Fig. 3, we represent the composition of the two images of Fig. 1, after simultaneous registration and exposure correction, *i.e.*, the mosaic obtained by merging the left image of Fig. 1 with the bottom right image of Fig. 2. Two other examples, represented in Fig. 4, use underwater images in a seabed mapping application. Due to the low texture of these images, it would be very hard to compute image-to-image correspondences between pointwise features in an automatic way. Since our method processes directly the intensity values in the entire images, it is robust to the absence of highly textured regions and succeeds to align these underwater images, even when their brightness changes dramatically from image to image, as shown by the mosaics in Fig. 4.

The motion model used in the experiments above is the 6-parameter affine. More experimental results, as well as a MATLAB implementation of the algorithm we propose in this paper, are available from [8]. In all our experiments, the algorithm converged in less than 15 iterations.

Fig. 3. Composition of the images in Fig. 1, using our method for simultaneous registration and exposure correction.

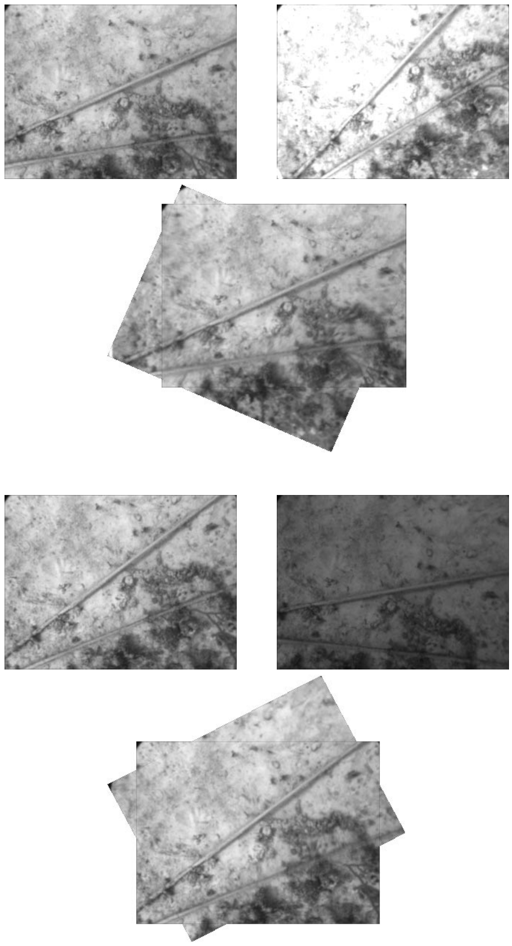

Fig. 4. Mosaics of underwater images for seabed mapping.

5. CONCLUSION

We propose a featureless method to compose, both geometrically and photometrically, a pair of uncalibrated images. This method is robust because it processes directly all the information available, rather than relying on an intermediate feature matching stage. Our method leads to an iterative algorithm that estimates, in alternate steps, the geometric and photometric parameters. This algorithm is computationally simple because one of the steps admits closed-form solution and the other leads to the well studied and already optimized constant brightness featureless motion estimation.

6. REFERENCES

[1] P. Burt and E. Adelson, "Multiresolution spline with application to image mosaics," *ACM Transaction on Graphics*, vol. 2, no. 4, 1983.

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[3] 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.

[4] B. Pires and P. Aguiar, "Featureless global alignment of multiple images," in *IEEE ICIP*, Genova, Italy, 2005.

[5] P. Perez and N. Garcia, "Robust and accurate registration of images with unknown relative orientation and exposure," in *IEEE ICIP*, Genova, Italy, 2005.

[6] D. Xu and T. Kasparis, "Robust image registration under spatially non-uniform brightness changes," in *IEEE ICASSP*, Philadelphia, PA, USA, 2005.

[7] Y. Altunbasak, R. Merserau, and A. Patti, "A fast parametric motion estimation algorithm with illumination and lens distortion correction," *IEEE Trans. on Image Processing*, vol. 12, no. 4, 2003.

[8] `www.isr.ist.utl.pt/~aguiar/mosaics.html`

[9] M. Grossberg and S. Nayar, "Determining the camera response from images: what is knowable?," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, 2003.

[10] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, 1997.