# Groupwise Geometric and Photometric Direct Image Registration

## Adrien Bartoli

**Abstract**—Image registration consists of estimating geometric and photometric transformations that align two images as best possible. The direct approach consists of minimizing the discrepancy in the intensity or color of the pixels. The inverse compositional algorithm has been recently proposed by Baker et al. for the direct estimation of groupwise geometric transformations. It is efficient in that it performs several computationally expensive calculations at a precomputation phase. Photometric transformations act on the value of the pixels. They account for effects such as lighting change. Jointly estimating geometric and photometric transformations is thus important for many tasks. such as image mosaicing. We propose an algorithm to jointly estimate groupwise geometric and photometric transformations while preserving the efficient precomputation-based design of the original inverse compositional algorithm. It is called the dual inverse compositional algorithm. It uses different approximations than the simultaneous inverse compositional algorithm and handles groupwise geometric and global photometric transformations. Its name stems from the fact that it uses an inverse compositional update rule for both the geometric and the photometric transformations. We demonstrate the proposed algorithm and compare it to previous ones on simulated and real data. This shows clear improvements in computational efficiency and in terms of convergence.

**Index Terms**—Image registration, geometric warp, photometric transformation, inverse composition.

✦

## 1 INTRODUCTION

IMAGE registration is the task of applying some transformations to two images so that they match as best possible. This can be seen as the computation of some geometric transformation, for example, a homography, used to deform one of the images to model camera pose, and some photometric transformation, applied to the intensity or color of the pixels, to account, for example, for lighting change.

Image registration has been an important research topic for the past decades. It is central to many tasks in computer vision, medical imaging, augmented reality, and robotics. Applications include image mosaicing [8], object and feature tracking [5], [7], [9], [11], superresolution [6], and visual servoing.

Broadly speaking, two approaches have been proposed: The feature-based and the direct approaches. The feature-based approach, see, e.g., [12], relies on abstracting the input images by the geometric location of a set of carefully chosen, salient features. The direct approach, see, e.g., [8], uses the value, that is, the intensity or color, of the pixels of interest. The *inverse compositional algorithm* of Baker et al. [2] estimates groupwise geometric transformations, such as homographies.[1] It has been shown to be one of the most

1. To be precise, transformations parameterized such that there is a group structure on the parameter vector.

- *The author is with LASMEA, 24, avenue des Landais, 63177 Aubière cedex, France. E-mail: Adrien.Bartoli@gmail.com.*

reliable and computationally efficient registration methods. The efficiency stems from the so-called *inverse compositional trick*, making the Hessian matrix[2] constant (it is the design matrix involved in the linear least squares problem to be solved at each iteration). This makes it possible to precompute its inverse.

This paper is about the registration of two images related by a geometric and a photometric transformation. An example of photometric transformation is "gain and bias," which rescales and offsets the value of the pixels. The *simultaneous inverse compositional algorithm* proposed in [1] by Baker et al. estimates such transformations but at the expense of recomputing and inverting the Hessian matrix at each iteration. An efficient variant called the *project out inverse compositional algorithm* is proposed in [1]. Due to an approximation of the photometric error function, it performs worse than the simultaneous inverse compositional algorithm in terms of convergence frequency and number of iterations.

We propose the *dual inverse compositional algorithm*, which uses the inverse compositional trick for both the geometric and photometric counterparts of the registration, thereby preserving the possibility of precomputing the inverse of the Hessian matrix. We originally proposed this method in [3]. It deals with gray-level and color images and groupwise photometric transformations. The dual inverse compositional algorithm takes different steps to converge compared to the simultaneous inverse compositional algorithm of Baker et al. Thorough experiments show similar convergence properties for these algorithms, with a significantly lower computational cost in favor of the proposed algorithm, and an improved stability compared to the

2. We use the expression "Hessian matrix" for the Gauss-Newton approximation to the "true Hessian matrix."

project out inverse compositional algorithm. The dual inverse compositional algorithm is based on the assumption that the geometric and photometric transformations commute. This assumption prevents its use for estimating nonglobal photometric transformations such as Linear Appearance Variations. It is thus useful mainly for global changes such as lighting and camera setting changes that can mix the different color channels.

*Paper organization.* We formally state the problem and review previous work in Section 2. We present as background material the inverse compositional, the simultaneous inverse compositional, and the project out inverse compositional algorithms of Baker et al. in Section 3. We propose the dual inverse compositional algorithm in Section 4. Some groupwise color photometric transformations are presented in Section 5. We report experimental results on simulated data in Section 6. A conclusion is provided in Section 7. The parameterization of homographic warps is detailed in Appendix A and a proof showing that nonglobal transformations cannot be used with the dual inverse compositional algorithm is reported in Appendix B. We report experimental results on real data in Appendix C. Finally, the proposed dual inverse compositional algorithm is summarized.

*Notation.* Vectors are denoted using bold fonts, for example, $\mathbf{q}$, matrices using roman fonts, for example, E, and scalars in italics, for example, $a$. The entries of a vector or matrix are written, as in $\mathbf{x}^{\mathrm{T}} = (x_1 \; \cdots \; x_n)$, where $\mathbf{x}$ is transposed in this equation. The two norm of a vector $\mathbf{r}$ is written $\|\mathbf{r}\|$. The gradient of a scalar-valued function $f$, in other words, its partial derivative vector with respect to vector $\mathbf{x}$, is denoted $\nabla_{\mathbf{x}} f$. It is evaluated at $\mathbf{0}$, that is, the zero vector, unless specified as in $(\nabla_{\mathbf{x}} f)(\mathbf{x}_0)$:

$$\nabla_{\mathbf{x}} f = (\nabla_{\mathbf{x}} f)(\mathbf{0}) = \left( \left( \frac{\partial f}{\partial x_1} \right)(\mathbf{0}) \cdots \left( \frac{\partial f}{\partial x_n} \right)(\mathbf{0}) \right)^{\mathrm{T}}.$$

Note that, for vector-valued functions, $\nabla$ gives the Jacobian matrix, that is, the matrix containing all of the partial derivatives of the function. Columnwise matrix vectorization is written vect.

The source and target images to be registered are denoted $\mathcal{S}$ and $\mathcal{T}$, respectively. They are seen as functions from $\mathbb{R}^2$ to $\mathbb{R}^c$, where $c$ is the number of channels, that is, $c = 1$ in the gray-level case and $c = 3$ in the color case. For instance, $\mathcal{T}(\mathbf{q})$ is the image value at pixel $\mathbf{q} \in \mathbb{R}^2$. Bilinear interpolation is used for subpixel coordinates. The unit column vector is denoted $\mathbf{1}$ with length given by the context. The geometric and photometric transformations are respectively denoted $\mathcal{G}$ and $\mathcal{P}$, with respective parameter vectors to be estimated denoted $\mathbf{g}$ and $\mathbf{p}$. We make the distinction between the target to source image photometric parameters and the reverse one, respectively denoted $\mathbf{p}$ and $\tilde{\mathbf{p}}$. Note that $\mathcal{G}$ and $\mathcal{P}$ refer to global parameterizations as opposed to local, minimal parameterizations written $\bar{\mathcal{G}}$ and $\bar{\mathcal{P}}$ with parameters $\boldsymbol{\delta}_g$ and $\boldsymbol{\delta}_p$ or $\boldsymbol{\delta}_{\tilde{p}}$, respectively, defined such that the zero vector induces the identity transformation. The geometric transformation is also called the warp.

## 2 PROBLEM STATEMENT AND PREVIOUS WORK

The geometric registration problem is the minimization of a nonlinear least squares error function, given by the discrepancy in the value of the pixels, between the source image $\mathcal{S}$ and the target image $\mathcal{T}$ warped onto the source one by the unknown warp. The warp maps a pixel $\mathbf{q}$ in the region of interest $\mathcal{R}$ defined in the source image to the corresponding pixel $\mathcal{G}(\mathbf{q}; \mathbf{g})$ in the target image. We expect that, given an "appropriate" parameter vector $\mathbf{g}$, $\mathcal{S}(\mathbf{q})$ is "close to" $\mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g}))$, for all $\mathbf{q} \in \mathcal{R}$: This is the brightness constancy assumption. The direct image registration problem is thus formally posed as the minimization of the *photometric error*:

$$\min_{\mathbf{g}} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g}))\|^2. \tag{1}$$

Note that other error functions can be used to deal with, for example, occlusions. Most algorithms linearize each term in the transformation parameters $\mathbf{g}$ and iteratively update an initial guess by solving linear least squares problems. Hardie et al. [6] register several images at once while computing a superresolved one. Baker and Matthews [2] propose the efficient *inverse compositional algorithm* for solving problem (1). More details are given in the next section. It uses a Gauss-Newton, local approximation to the error function. The efficiency stems from the fact that the Hessian matrix involved in the normal equations to be solved at each iteration is constant. Its inverse is thus precomputed.

Problem (1) does not take into account photometric changes, that is, changes in the pixel values. These changes occur, for example, when lighting changes between the acquisition of the two images or when two different cameras are used. They are modeled by a transformation $\mathcal{P}$ with parameter vector $\mathbf{p}$ and give rise to the following minimization problem:

$$\min_{\mathbf{g}, \mathbf{p}} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\mathbf{q}) - \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g})); \mathbf{p})\|^2. \tag{2}$$

A commonly employed photometric model $\mathcal{P}$ is an affine transformation modeling gain and bias (or contrast and brightness). More complex transformations are reviewed in Section 5.

Jin et al. [9] use this model for feature tracking in gray-level images, in contrast to [11], which normalizes the image patches by using the mean and standard deviation of the pixel values. Heigl et al. [7] track points in color images by summing the error over the three channels. Lai and Fang [10] register images with low-order polynomials for modeling spatially varying gain and bias. Vemuri et al. [13] use a forward compositional framework for estimating spline-based free-form deformations. Their modified Newton optimization scheme uses a constant approximation of the Hessian matrix evaluated at the optimum.

Baker et al. extend the inverse compositional algorithm in [1] to deal with linear appearance variations of the source image. In their framework, the photometric transformation is applied to the source image. As will be seen in Section 3.2, this makes the Hessian matrix vary across the iterations, thereby spoiling the computational efficiency of the inverse compositional algorithm: The simultaneous inverse compositional

algorithm estimates and inverts the Hessian matrix at each iteration. Baker et al. propose several approximations to reduce the computational cost, namely, the "project out inverse compositional algorithm" and the "normalization inverse compositional algorithm." They show that these approximations do not behave well for high gain values.

One reason for Baker et al. to apply the photometric transformation to the source image is to handle general linear appearance variations, i.e., based on linear combinations of "eigenimages." This is used in conjunction with 3D Morphable Models of, e.g., faces that do not form a group. In the case of, e.g., homography estimation, there is however no practical reason to handle such transformations.

## 3  BACKGROUND

This section is devoted to the description of the inverse compositional and simultaneous inverse compositional algorithms of Baker et al. proposed in [2] and [1], respectively.

### 3.1  The Inverse Compositional Algorithm

The inverse compositional algorithm, or IC for short, forms the basis for our dual inverse compositional algorithm, presented in Section 3.2. Its advantages are twofold. First, it converges rapidly compared to other optimization schemes. Second, as already mentioned, each iteration is performed efficiently.

The inverse compositional algorithm iteratively updates an initial guess of the sought-after transformation. The key idea is to express the updated transformation as the composition of the current transformation $\mathcal{G}(\cdot; \mathbf{g})$ and the inverse of an incremental transformation $\bar{\mathcal{G}}^{-1}(\cdot; \delta_g)$— this is the inverse compositional update rule— $\mathcal{G}(\cdot; \mathbf{g}) \leftarrow \mathcal{G}(\bar{\mathcal{G}}^{-1}(\cdot; \delta); \mathbf{g})$. Notation $\bar{\mathcal{G}}$ refers to a local parameterization of the warp, as opposed to the global parameterization denoted $\mathcal{G}$. This update rule is written $\mathbf{g} \leftarrow \mathcal{U}_g(\mathbf{g}, \delta_g)$. Details on the global and local parameterizations and the update rule are given in Appendix A for homographic warps.

The optimization is performed over $\delta_g$, the parameter vector of the incremental warp, instead of $\mathbf{g}$. The geometric registration problem (1) is rewritten $\min_{\delta_g} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{G}(\bar{\mathcal{G}}^{-1}(\mathbf{q}; \delta_g); \mathbf{g}))\|^2$. Let $\mathcal{W}$ be the warped target image, that is, $\mathcal{W}(\mathbf{q}) = \mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g}))$. The incremental transformation is then applied to the source image, instead of the target one, leading to $\min_{\delta_g} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q}; \delta_g)) - \mathcal{W}(\mathbf{q})\|^2$. This is the *inverse compositional trick*. Note that this only approximates the original problem (1) since the error function is expressed within the warped and not within the source image. The error function is approximated by first-order Taylor expansion in $\delta_g$, forming a Gauss-Newton approximation: $\min_{\delta_g} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\mathbf{q}) + \mathrm{L}_g^{\mathrm{T}}(\mathbf{q})\delta_g - \mathcal{W}(\mathbf{q})\|^2$. This is a linear least squares problem, which is solved via its normal equations. We define $(\nabla\mathcal{S})(\mathbf{q})$ to be the $(2 \times c)$ Jacobian matrix of the source image at $\mathbf{q}$ and $(\nabla_{\mathbf{g}}\bar{\mathcal{G}})(\mathbf{q}; \mathbf{0})$ to be the Jacobian matrix of the local warp, evaluated at $\mathbf{q}$ and at warp parameters $\mathbf{0}$. It is assumed for simplicity that $\mathbf{0}$ represents the identity warp, as discussed in Appendix A. The Jacobian matrices $\mathrm{L}_g(\mathbf{q})$ are thus obtained using the chain rule as

$\mathrm{L}_g^{\mathrm{T}}(\mathbf{q}) = (\nabla\mathcal{S})(\mathbf{q})^{\mathrm{T}}(\nabla_{\mathbf{g}}\bar{\mathcal{G}})(\mathbf{q}; \mathbf{0})$. They only depend on the source image at the pixels of interest and are thus constant over the iterations. Let $\mathcal{D}(\mathbf{q}) = \mathcal{W}(\mathbf{q}) - \mathcal{S}(\mathbf{q})$ be the error image; the normal equations are $\mathrm{E}_g\delta_g = \mathbf{b}_g$, where the Hessian matrix $\mathrm{E}_g$ and the right-hand side $\mathbf{b}_g$ are $\mathrm{E}_g = \sum_{\mathbf{q} \in \mathcal{R}} \mathrm{L}_g(\mathbf{q})\mathrm{L}_g^{\mathrm{T}}(\mathbf{q})$ and $\mathbf{b}_g = \sum_{\mathbf{q} \in \mathcal{R}} \mathrm{L}_g(\mathbf{q})\mathcal{D}(\mathbf{q})$. The solution $\delta_g = \mathrm{E}_g^{-1}\mathbf{b}_g$ for the local warp parameters is thus computed very efficiently since the Jacobian matrices $\mathrm{L}_g(\mathbf{q})$, as well as the inverse $\mathrm{E}_g^{-1}$ of the Hessian matrix, are precomputed.

Once $\delta_g$ has been computed, parameters $\mathbf{g}$ are updated by composing the current warp with the incremental warp with the update rule $\mathbf{g} \leftarrow \mathcal{U}_g(\mathbf{g}, \delta_g)$. If one uses a homographic warp, for example, then the updated parameters are given by multiplying the current homography by the inverse of the local one. The process is iterated until convergence, determined in our experiments, by thresholding $\|\delta_g\|$ by $\varepsilon = 10\mathrm{e} - 8$.

### 3.2  The Simultaneous Inverse Compositional Algorithm

The simultaneous inverse compositional algorithm, or SIC for short, aims at registering two images by computing both a warp and a parametric photometric transformation $\mathcal{P}$ acting on the pixel values, that is, their intensity of color.

Let $\tilde{\mathbf{p}}$ denote the parameter vector for the photometric transformation from the source to the target image. Baker et al. [1] pose the registration problem as

$$\min_{\mathbf{g}, \tilde{\mathbf{p}}} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{P}(\mathcal{S}(\mathbf{q}); \tilde{\mathbf{p}}) - \mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g}))\|^2. \tag{3}$$

We note that the minimization takes place in the geometric space of the source image but in the photometric space of the target image.

Baker et al. use an inverse compositional update rule for the warp and a forward additive update rule for the photometric transformation. Applying the inverse compositional trick as in the previous section, with $\mathcal{W}(\mathbf{q}) = \mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g}))$ the warped image, yields $\min_{\delta_g, \delta_{\tilde{p}}} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{P}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q}; \delta_g)); \tilde{\mathbf{p}} + \delta_{\tilde{p}}) - \mathcal{W}(\mathbf{q})\|^2$, where we switched the warp and the photometric transformation, i.e., $(\mathcal{P}(\mathcal{S}; \tilde{\mathbf{p}} + \delta_{\tilde{p}}))(\bar{\mathcal{G}}(\mathbf{q}; \delta_g)) = \mathcal{P}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q}; \delta_g)); \tilde{\mathbf{p}} + \delta_{\tilde{p}})$. The first-order Taylor expansion in $\delta_g$ and in $\delta_{\tilde{p}}$ gives

$$\min_{\delta_g, \delta_{(\tilde{p})}} \sum_{\mathbf{q} \in \mathcal{R}} \left\| \mathcal{P}(\mathcal{S}(\mathbf{q}) + \mathrm{L}_g^{\mathrm{T}}(\mathbf{q})\delta_g; \tilde{\mathbf{p}}) \right.$$
$$\left. + (\nabla_{\mathbf{p}}\mathcal{P})^{\mathrm{T}}(\mathcal{S}(\mathbf{q}) + \mathrm{L}_g^{\mathrm{T}}(\mathbf{q})\delta_g; \tilde{\mathbf{p}})\delta_{\tilde{p}} - \mathcal{W}(\mathbf{q}) \right\|^2.$$

Further expansion is achieved using the assumption that $\mathcal{P}$ is an affine transformation for its parameters $\mathbf{p}$, which include many different photometric transformations and linear appearance variations. We define $\check{\mathbf{p}}$ as the linear counterpart of the parameters, i.e., where the intercept vanishes, in other words, without its affine counterpart. For example, if $\mathcal{P}(v; \tilde{\mathbf{p}}) = \tilde{p}_1 v + \tilde{p}_2$, i.e., the gain and bias photometric transformation, then $\check{\mathbf{p}}^{\mathrm{T}} = (\tilde{p}_1\ 0)$. This allows us to simplify the first term as $\mathcal{P}(\mathcal{S}(\mathbf{q}) + \mathrm{L}_g^{\mathrm{T}}(\mathbf{q})\delta_g; \tilde{\mathbf{p}}) = \mathcal{P}(\mathcal{S}(\mathbf{q}); \tilde{\mathbf{p}}) + \mathcal{P}(\mathrm{L}_g^{\mathrm{T}}(\mathbf{q}); \check{\mathbf{p}})\delta_g$. Neglecting the second order terms, we approximate the second term as $(\nabla_{\mathbf{p}}\mathcal{P})^{\mathrm{T}}(\mathcal{S}(\mathbf{q}) + \mathrm{L}_g^{\mathrm{T}}(\mathbf{q})\delta_g; \tilde{\mathbf{p}})\delta_{\tilde{p}} \approx \mathrm{L}_p^{\mathrm{T}}(\mathbf{q})\delta_{\tilde{p}}$, with

$L_p(\mathbf{q}) = (\nabla_{\mathbf{p}}\mathcal{P})(\mathcal{S}(\mathbf{q}); \tilde{\mathbf{p}})$. Note that $L_p(\mathbf{q})$ is independent of $\mathbf{p}$ due to the affine nature of $\mathcal{P}$. It is thus precomputed. Let the error image be $\mathcal{D}(\mathbf{q}) = \mathcal{W}(\mathbf{q}) - \mathcal{P}(\mathcal{S}(\mathbf{q}); \tilde{\mathbf{p}})$; we get $\min_{\delta_g, \delta_{\tilde{p}}} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{P}(L_g^T(\mathbf{q}); \check{\mathbf{p}})\delta_g + L_p^T(\mathbf{q})\delta_{\tilde{p}} - \mathcal{D}(\mathbf{q})\|^2$. We see in this linear least squares problem that the associated Jacobian matrix has a constant part associated to the photometric parameters $\delta_{\tilde{p}}$ and a nonconstant part associated to the warp parameters $\delta_g$. This makes nonconstant the Hessian matrix of the normal equations since the current photometric transformation has to be applied to the Steepest Descent images (the columns of the Jacobian matrix). The optimization problem is rewritten $\min_{\delta_{g\tilde{p}}} \sum_{\mathbf{q} \in \mathcal{R}} \|K_{g\tilde{p}}^T(\mathbf{q}; \tilde{\mathbf{p}})\delta_{g\tilde{p}} - \mathcal{D}(\mathbf{q})\|^2$, with $\delta_{g\tilde{p}}^T = (\delta_g^T \; \delta_{\tilde{p}}^T)$ the joint incremental parameter vector and $K_{g\tilde{p}}$ the joint Jacobian matrices given by $K_{g\tilde{p}}^T(\mathbf{q}; \tilde{\mathbf{p}}) = (\mathcal{P}(L_g^T(\mathbf{q}); \check{\mathbf{p}}) \; L_p^T(\mathbf{q}))$. The normal equations are $E_{g\tilde{p}}\delta_{g\tilde{p}} = \mathbf{b}_{g\tilde{p}}$, with the Hessian matrix $E_{g\tilde{p}}$ and the right-hand side $\mathbf{b}_{g\tilde{p}}$ given by $E_{g\tilde{p}} = \sum_{\mathbf{q} \in \mathcal{R}} K_{g\tilde{p}}(\mathbf{q}; \tilde{\mathbf{p}})K_{g\tilde{p}}^T(\mathbf{q}; \tilde{\mathbf{p}})$ and $\mathbf{b}_{g\tilde{p}} = \sum_{\mathbf{q} \in \mathcal{R}} K_{g\tilde{p}}(\mathbf{q}; \tilde{\mathbf{p}})\mathcal{D}(\mathbf{q})$.

## 3.3 The Project Out Inverse Compositional Algorithm

The project out inverse compositional algorithm, or PO for short, proposed by Baker et al. in [1], aims at reducing the computational cost required by each iteration of the simultaneous inverse computational algorithm. We start its derivation from (3) and rewrite the error function in matrix form using the $\mathcal{L}_2$-norm:

$$\min_{\mathbf{g}, \tilde{\mathbf{p}}} \left\| \begin{pmatrix} \vdots \\ \mathcal{P}(\mathcal{S}(\mathbf{q}); \tilde{\mathbf{p}}) - \mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g})) \\ \vdots \end{pmatrix} \right\|^2.$$

The error vector is projected into a linear subspace $\mathcal{B}$ chosen as a basis for the affine photometric models we use, see Section 5, and its orthogonal complement $\mathcal{B}^\perp$, giving

$$\min_{\mathbf{g}, \tilde{\mathbf{p}}} \left\| \begin{pmatrix} \vdots \\ \mathcal{P}(\mathcal{S}(\mathbf{q}); \tilde{\mathbf{p}}) - \mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g})) \\ \vdots \end{pmatrix} \right\|_{\mathcal{B}}^2$$
$$+ \left\| \begin{pmatrix} \vdots \\ \mathcal{S}(\mathbf{q}) - \mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g})) \\ \vdots \end{pmatrix} \right\|_{\mathcal{B}^\perp}^2.$$

We observe that the second term is independent of the photometric transformation. The project out inverse compositional algorithm consists in minimizing the second term with respect to the geometric parameters $\mathbf{g}$. Minimizing the first term with respect to $\mathbf{p}$ subsequently gives the photometric parameters through a closed-form solution. Minimizing the second term is performed by using a weighted $\mathcal{L}_2$-norm, see [1] for the details, and is implemented with a weighted inverse compositional algorithm.

It is shown that this algorithm has problems with determining the optimal magnitude of the update vector. This arises when a gain is estimated in the photometric transformation, which is the case for all of the transforma-

tions we show in Section 5. In the absence of noise, it is shown in [1] that the computed update vector $\delta_g$ is a gain-weighted version of the ideal one. Two solutions are then suggested. The first one is to use Levenberg-Marquardt instead of Gauss-Newton as a local optimization engine since it dynamically adjusts the step size. The second solution is a step size correction scheme based on dividing $\delta_g$ by the current estimate of the gain. We implemented both solutions. Both work well, but, as reported in [1], they sometimes oscillate around the sought-after solution and may even diverge. Assessing convergence is thus difficult and one has to define a fixed number of iterations and select the best estimate computed so far.

As reported in [1], the project out inverse compositional algorithm has poorer performances than the simultaneous one. The reason is that the algorithm is based on the fact that the error projected in the $\mathcal{B}^\perp$ subspace does not depend on the photometric transformation parameters. This is, however, true only when the alignment is correct. Therefore, in practice, parameter updates are subject to unexpected perturbations that may prevent convergence.

## 4 THE DUAL INVERSE COMPOSITIONAL ALGORITHM

We extend the inverse compositional algorithm to estimate a groupwise photometric transformation along with the warp, as stated in problem (2). The algorithm is summarized in Table 1 and illustrated in Fig. 1. It is dubbed DIC for short.

The main difference compared to the simultaneous inverse compositional algorithm resides in the problem formulation: One of the images, namely, the target one, is taken as a "generator" for the other one, namely, the source image. In other words, while the minimization takes place in the geometric frame of the source image and in the photometric frame of the target image in the formulation by Baker et al., it takes place in the frame of the source image for both photometry and geometry in our formulation. The algorithms also differ in the update rule they employ for the photometric parameters. Baker et al. use a forward additive update rule, while we use an inverse compositional update rule. This allows us to apply the inverse compositional trick for both the geometric and photometric transformations, under the assumption that these transformations commute.

Consider problem (2) and plug in an inverse compositional update rule for both the geometric and photometric transformations, i.e., $\mathcal{G}(\cdot; \mathbf{g}) \leftarrow \mathcal{G}(\bar{\mathcal{G}}^{-1}(\cdot; \delta_g); \mathbf{g})$ and $\mathcal{P}(\cdot; \mathbf{p}) \leftarrow \bar{\mathcal{P}}^{-1}(\mathcal{P}(\cdot; \mathbf{p}); \delta_p)$. Note that the incremental photometric transformation is composed to the left and not to the right of the current transformation, contrarily to the case of the warp. For more details, see Section 5. This gives

$$\min_{\delta_g, \delta_p} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\mathbf{q}) - \bar{\mathcal{P}}^{-1}(\mathcal{P}(\mathcal{T}(\mathcal{G}(\bar{\mathcal{G}}^{-1}(\mathbf{q}; \delta_g); \mathbf{g})); \mathbf{p}); \delta_p)\|^2. \quad (4)$$

The optimization is now to be performed on the incremental parameters $\delta_g$ and $\delta_p$, the latter accounting for the incremental photometric transformation. Using the inverse compositional trick on the photometric transformation, i.e., applying the incremental photometric transformation to the source image instead of the target image, gives

TABLE 1
DIC: The Proposed *Dual Inverse Compositional Algorithm*
for Groupwise Geometric and Photometric Registration
of Gray-Level or Color Images

---

#### OBJECTIVE

Register a target image $\mathcal{T}$ to a source image $\mathcal{S}$ by computing the parameters $\mathbf{g}$ of a geometric registration $\mathcal{G}(\cdot; \mathbf{g})$ and the parameters $\mathbf{p}$ of a photometric registration $\mathcal{P}(\cdot; \mathbf{p})$. Other inputs are the region of interest $\mathcal{R}$ in the source image and an initial value for $\mathbf{g}$ and $\mathbf{p}$. Upon convergence, the photometric error $\sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\mathbf{q}) - \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g})); \mathbf{p})\|^2$ is minimized.

#### ASSUMPTIONS

- Group structure on the warp parameters $\mathbf{g}$
- Group structure on the global photometric parameters $\mathbf{p}$
- Commutativity of the warp and the photometric transformation

#### ALGORITHM

**Pre-computations**

1) Compute the joint Jacobian matrices $\mathrm{L}_{gp}(\mathbf{q})$ for $\mathbf{q} \in \mathcal{R}$ from equation (7):

$$\mathrm{L}_{gp}(\mathbf{q}) = \left( (\nabla \mathcal{S})(\mathbf{q})^\mathsf{T} \left( \nabla_{\mathbf{g}} \bar{\mathcal{G}} \right)(\mathbf{q}; \mathbf{0}) \quad (\nabla_{\mathbf{p}} \mathcal{P})(\mathcal{S}(\mathbf{q}); \mathbf{0}) \right)$$

2) Compute the Hessian matrix $\mathrm{E}_{gp}$ and its inverse from equation (8):

$$\mathrm{E}_{gp} = \sum_{\mathbf{q} \in \mathcal{R}} \mathrm{L}_{gp}(\mathbf{q}) \mathrm{L}_{gp}^\mathsf{T}(\mathbf{q})$$

**Iterations**

1) Warp and photometrically transform the target image $\mathcal{T}$ to $\mathcal{W}$ using the warp $\mathbf{g}$ and the photometric parameters $\mathbf{p}$:

$$\mathcal{W}(\mathbf{q}) = \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g})); \mathbf{p})$$

2) Compute the incremental warp and photometric parameters $\boldsymbol{\delta}_g$ and $\boldsymbol{\delta}_p$:

- Compute the error image $\mathcal{D}$:

$$\mathcal{D}(\mathbf{q}) = \mathcal{W}(\mathbf{q}) - \mathcal{S}(\mathbf{q})$$

- Compute the right hand side of the normal equations $\mathbf{b}_{gp}$ from equation (9):

$$\mathbf{b}_{gp} = \sum_{\mathbf{q} \in \mathcal{R}} \mathrm{L}_{gp}(\mathbf{q}) \mathcal{D}(\mathbf{q})$$

- Solve for the joint incremental parameters $\boldsymbol{\delta}_{gp}$:

$$\boldsymbol{\delta}_{gp} = \mathrm{E}_{gp}^{-1} \mathbf{b}_{gp}$$

3) Update the warp $\mathbf{g}$ and photometric parameters $\mathbf{p}$:

$$\mathbf{g} \leftarrow \mathcal{U}_g(\mathbf{g}, \boldsymbol{\delta}_g) \qquad \mathbf{p} \leftarrow \mathcal{U}_p(\mathbf{p}, \boldsymbol{\delta}_p)$$

---

*The Hessian matrix is constant throughout the iterations.*

$$\min_{\boldsymbol{\delta}_g, \boldsymbol{\delta}_p} \sum_{\mathbf{q} \in \mathcal{R}} \|\bar{\mathcal{P}}(\mathcal{S}(\mathbf{q}); \boldsymbol{\delta}_p) - \mathcal{P}(\mathcal{T}(\mathcal{G}(\bar{\mathcal{G}}^{-1}(\mathbf{q}; \boldsymbol{\delta}_g); \mathbf{g})); \mathbf{p})\|^2. \quad (5)$$

We now use the assumption that the order used to apply the warp and the photometric transformation to an image does not matter, i.e., that $\mathcal{P}(\mathcal{T}(\mathcal{G}(\cdot; \mathbf{g})); \mathbf{p}) = (\mathcal{P}(\mathcal{T}; \mathbf{p}))(\mathcal{G}(\cdot; \mathbf{g}))$. This assumption allows us to switch the photometric transformation and the warps in the second term of (5). Applying the inverse compositional trick once again on the incremental warp gives

$$\min_{\boldsymbol{\delta}_g, \boldsymbol{\delta}_p} \sum_{\mathbf{q} \in \mathcal{R}} \|\bar{\mathcal{P}}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q}; \mathbf{g})); \boldsymbol{\delta}_p) - \mathcal{P}(\mathcal{T}; \mathbf{p})(\mathcal{G}(\mathbf{q}; \mathbf{g}))\|^2.$$

Switching the warp and photometric transformation again, on the second term, and letting $\mathcal{W}$ be the warped and photometrically transformed image, i.e., $\mathcal{W}(\mathbf{q}) = \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q}; \mathbf{g})); \mathbf{p})$, yields

$$\min_{\boldsymbol{\delta}_g, \boldsymbol{\delta}_p} \sum_{\mathbf{q} \in \mathcal{R}} \|\bar{\mathcal{P}}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q}; \boldsymbol{\delta}_g)); \boldsymbol{\delta}_p) - \mathcal{W}(\mathbf{q})\|^2. \quad (6)$$

We show below that the normal equations induced by the Gauss-Newton approximation to problem (6) have a constant Hessian matrix. Similarly to the simultaneous inverse compositional algorithm, we use the first-order Taylor expansion in $\boldsymbol{\delta}_g$ and $\boldsymbol{\delta}_p$, giving

$$\min_{\boldsymbol{\delta}_g, \boldsymbol{\delta}_p} \sum_{\mathbf{q} \in \mathcal{R}} \|\mathcal{S}(\mathbf{q}) + \mathrm{L}_g^\mathsf{T}(\mathbf{q}) \boldsymbol{\delta}_g + (\nabla_{\mathbf{p}} \bar{\mathcal{P}})(\mathcal{S}(\mathbf{q}) + \mathrm{L}_g^\mathsf{T}(\mathbf{q}) \boldsymbol{\delta}_g; \mathbf{0}) \boldsymbol{\delta}_p$$
$$- \mathcal{W}(\mathbf{q})\|^2.$$

Using the assumption that $\mathcal{P}$ is an affine transformation and neglecting the second-order terms yield

$$\min_{\boldsymbol{\delta}_g, \boldsymbol{\delta}_p} \sum_{\mathbf{q} \in \mathcal{R}} \|(\mathrm{L}_g^\mathsf{T}(\mathbf{q}) \ \mathrm{L}_p^\mathsf{T}(\mathbf{q})) \boldsymbol{\delta}_{gp} - \mathcal{D}(\mathbf{q})\|^2,$$

where $\mathcal{D}$ is the error image, i.e., $\mathcal{D}(\mathbf{q}) = \mathcal{W}(\mathbf{q}) - \mathcal{S}(\mathbf{q})$. We denote the joint incremental parameter vector $\boldsymbol{\delta}_{gp}^\mathsf{T} = (\boldsymbol{\delta}_g^\mathsf{T} \ \boldsymbol{\delta}_p^\mathsf{T})$ and define the joint Jacobian matrices $\mathrm{L}_{gp}(\mathbf{q})$ by

$$\mathrm{L}_{gp}^\mathsf{T}(\mathbf{q}) = (\mathrm{L}_g^\mathsf{T}(\mathbf{q}) \ \mathrm{L}_p^\mathsf{T}(\mathbf{q})). \quad (7)$$

The Jacobian matrices $\mathrm{L}_p(\mathbf{q})$ for the photometric parameters are $\mathrm{L}_p(\mathbf{q}) = (\nabla_{\mathbf{p}} \bar{\mathcal{P}})(\mathcal{S}(\mathbf{q}); \mathbf{0})$. As in the original inverse compositional algorithm, the Jacobian matrices only depend on the source image at the pixels of interest. They are thus constant, as well as the Hessian matrix $\mathrm{E}_{gp}$ of the normal equations $\mathrm{E}_{gp} \boldsymbol{\delta}_{gp} = \mathbf{b}_{gp}$, with

$$\mathrm{E}_{gp} = \sum_{\mathbf{q} \in \mathcal{R}} \mathrm{L}_{gp}(\mathbf{q}) \mathrm{L}_{gp}^\mathsf{T}(\mathbf{q}), \quad (8)$$

$$\mathbf{b}_{gp} = \sum_{\mathbf{q} \in \mathcal{R}} \mathrm{L}_{gp}(\mathbf{q}) \mathcal{D}(\mathbf{q}). \quad (9)$$

The warp is updated as in the inverse compositional algorithm,: $\mathbf{g} \leftarrow \mathcal{U}_g(\mathbf{g}, \boldsymbol{\delta}_g)$, and the photometric transformation update rule is written $\mathbf{p} \leftarrow \mathcal{U}_p(\mathbf{p}, \boldsymbol{\delta}_p)$. The proposed dual inverse compositional algorithm handles most groupwise photometric transformation such as those given in Section 5.

## 5 SOME GROUPWISE COLOR PHOTOMETRIC TRANSFORMATIONS

We mention some photometric transformations that can be employed within our framework. The dual inverse compositional algorithm is based on the assumption that the geometric and photometric transformation commute. We show in Appendix B that this prevents the use of nonglobal photometric transformations. Spatially varying transformations such as the polynomial one in [10] are usually nonglobal. However, transformations such as the one in [5], which approximates specular highlights to the first order, can be employed. It is spatially varying but has a constant weight for
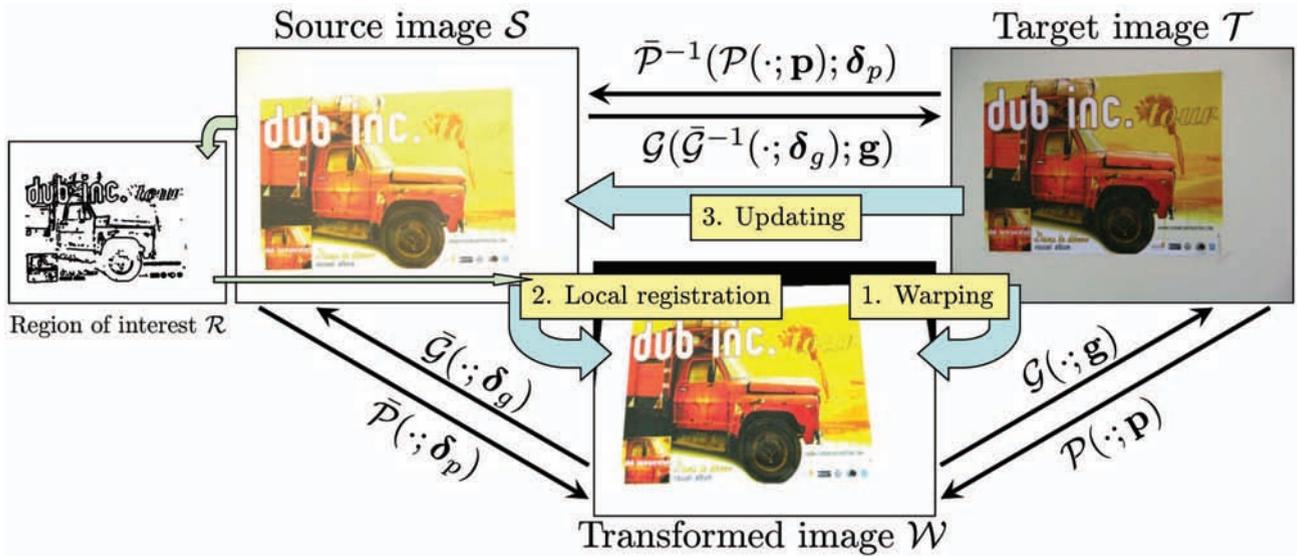
Fig. 1. DIC. The proposed *dual inverse compositional algorithm* extends the inverse compositional algorithm to jointly compute a geometric and a photometric registration, $\mathcal{G}(\cdot; \mathbf{g})$ and $\mathcal{P}(\cdot; \mathbf{p})$, by iterating the three main steps mentioned in the schema. One of the strengths of this approach is that, as in the inverse compositional algorithm, the Hessian matrix involved in the normal equations in Step 2 is constant.

each pixel, depending only on its distance with some point of interest, making its parameters global.

The most common photometric transformation for gray-level images is the aforementioned gain and bias. In the color image case, we use affine transformations, that is, transformations that can be written as $A\mathbf{v} + \mathbf{b}$, where A is a $(3 \times 3)$ matrix combining the three color channels and $\mathbf{b}$ is a 3-vector, modeling a per-channel bias. Finlayson et al. [4] show that linear transformations are well adapted for color constancy in practice. We have tried several variants, summarized below. Note that, similarly to the warp, we use a global and a local parameterization for each transformation. The local parameterizations all guarantee $\bar{\mathcal{P}}(\mathbf{v}; \mathbf{0}) = \mathbf{v}$.

The update rules are obtained by writing the transformation in matrix form by reshaping vector $\delta_p$ to a matrix A and vector $\mathbf{b}$. Inversion and composition are then performed and the resulting set of parameters is vectorized to get the updated vector $\mathbf{p}$.

*Single gain and bias.* This is the direct transposition of the gain and bias transformation of gray-level images to color images. It is due to account for global and uniform lighting change. The global and local transformations are

$$\mathcal{P}(\mathbf{v}; \mathbf{p}) = p_1 \mathbf{v} + p_2 \mathbf{1} \quad \text{and} \quad \bar{\mathcal{P}}(\mathbf{v}; \delta_p) = \mathcal{P}\left(\mathbf{v}; \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \delta_p\right).$$

The $(3 \times 2)$ Jacobian matrix is $L_p^T(\mathbf{q}) = (\mathcal{S}(\mathbf{q})\ \mathbf{1})$ and the inverse compositional update rule is $\mathcal{U}_p(\mathbf{p}, \delta_p) = \frac{1}{1+\delta_{p,1}} (p_1\ p_2 - \delta_{p,2})^T$.

*Multiple gain and bias.* This is a generalization of the gain and bias transformation to color images, with independent gain and bias applied to each color channel. This models global lighting change and the fact that each color channel

may have a different behavior when lighting changes. Finlayson et al. [4] show that this model is effective for color constancy. The global and local transformations are

$$\mathcal{P}(\mathbf{v}; \mathbf{p}) = \begin{pmatrix} p_1 & & \\ & p_2 & \\ & & p_3 \end{pmatrix} \mathbf{v} + \begin{pmatrix} p_4 \\ p_5 \\ p_6 \end{pmatrix},$$

$$\bar{\mathcal{P}}(\mathbf{v}; \delta_p) = \mathcal{P}\left(\mathbf{v}; \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} + \delta_p\right),$$

where $\mathbf{1}$ and $\mathbf{0}$ are $(3 \times 1)$ vectors. The $(3 \times 6)$ Jacobian matrix is $L_p^T(\mathbf{q}) = (\text{diag}(\mathcal{S}(\mathbf{q}))\ I)$ and the inverse compositional update rule is $\mathcal{U}_p(\mathbf{p}, \delta_p) = \left(\frac{p_1}{1+\delta_{p,1}}\ \frac{p_2}{1+\delta_{p,2}}\ \frac{p_3}{1+\delta_{p,3}}\ \frac{p_4-\delta_{p,4}}{1+\delta_{p,1}}\ \frac{p_5-\delta_{p,5}}{1+\delta_{p,2}}\ \frac{p_6-\delta_{p,6}}{1+\delta_{p,3}}\right)^T$.

*Full affine channel mixing.* This generalizes the other photometric transformations by mixing the different color channels and applying a per-channel bias. This is mainly useful for images taken by different cameras or under different lighting colors. The global and local transformations are

$$\mathcal{P}(\mathbf{v}; \mathbf{p}) = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & p_9 \end{pmatrix} \mathbf{v} + \begin{pmatrix} p_{10} \\ p_{11} \\ p_{12} \end{pmatrix},$$

$$\bar{\mathcal{P}}(\mathbf{v}; \delta_p) = \mathcal{P}(\mathbf{v}; \mathbf{u} + \delta_p),$$

with $\mathbf{u}^T = \text{vect}(I_{(3\times3)}) = (1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0)$. The $(3 \times 12)$ Jacobian matrix is

$$L_p^T(\mathbf{q}) = \left(\text{diag}\left(\mathcal{S}(\mathbf{q})^T, \mathcal{S}(\mathbf{q})^T, \mathcal{S}(\mathbf{q})^T\right)\ I\right)$$
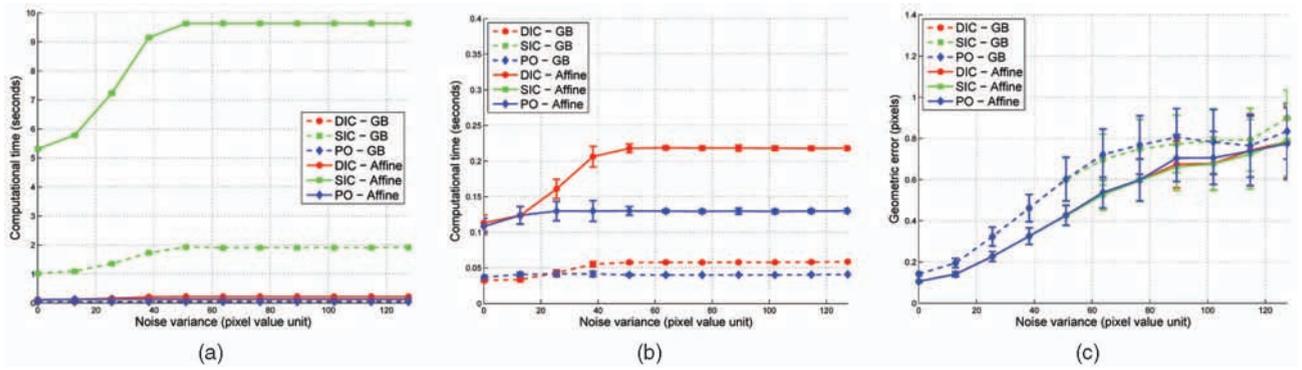
Fig. 2. (a) Computational time versus the variance of pixel noise. (b) A zoom of the graph in (a). (c) Geometric error versus the variance of pixel noise.

and the inverse compositional update rule is

$$\mathcal{U}_p(\mathbf{p}, \delta_p) =$$

$$\left( \begin{array}{c} \text{vect}\left( \left( \begin{array}{ccc} 1 + \delta_{p,1} & \delta_{p,2} & \delta_{p,3} \\ \delta_{p,4} & 1 + \delta_{p,5} & \delta_{p,6} \\ \delta_{p,7} & \delta_{p,8} & 1 + \delta_{p,9} \end{array} \right)^{-1} \left( \begin{array}{ccc} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & p_9 \end{array} \right) \right) \\ \left( \begin{array}{ccc} 1 + \delta_{p,1} & \delta_{p,2} & \delta_{p,3} \\ \delta_{p,4} & 1 + \delta_{p,5} & \delta_{p,6} \\ \delta_{p,7} & \delta_{p,8} & 1 + \delta_{p,9} \end{array} \right)^{-1} \left( \begin{array}{c} p_{10} - \delta_{p,10} \\ p_{11} - \delta_{p,11} \\ p_{12} - \delta_{p,12} \end{array} \right) \end{array} \right).$$

## 6 EXPERIMENTAL RESULTS ON SIMULATED DATA

Our experiments are designed to compare the converge properties and the computational cost of the proposed dual inverse compositional algorithm to other algorithms in various conditions, as well as different photometric transformations, as described in Section 5. All comparisons are done by estimating homographies. Our implementation uses Matlab with a number of routines written in C through Mex files (for example, bilinear image warping and error computation). Timing is measured on a PC equipped with a Pentium M at 1.6 GHz. We report results for SIC (see Section 3.2), PO (see Section 3.3), and the proposed DIC (see Section 4).

*Simulation setup.* Given a texture image, we simulate a 2D homography by displacing four points in random directions by some magnitude $\gamma$ with default value $\gamma = 5$ pixels. This is used to generate the target image in conjunction with a gain $\alpha$ and a bias $\beta$ with default values $\alpha = 1.2$ and $\beta = 15$. For the multiple gains and biases and full affine channel mixing photometric models, we use $\alpha$ with 10 percent Gaussian perturbation for the diagonal entries, some random values for the off-diagonal entries, and $\beta$ with 10 percent Gaussian perturbations for the biases. Finally, centered Gaussian noise with variance $\sigma$ is added to the pixel values in the source and target images, with default value $\sigma = 25.5$, that is, 10 percent of 255 (the maximum value of intensity and of each color channel). Finally, the pixel values are clamped between 0 and 255 in order to simulate sensor saturation. The source image is $600 \times 800$ and 25,392 pixels of interest are used.

We vary some parameters of this setup independently, namely, the noise variance $\sigma$ from 0 percent to 20 percent

(that is, 0 to 51 pixel intensity or color units), the geometric magnitude $\gamma$ from 0 to 20 pixels and the gain from 0.2 to 3. The results are average values over 100 trials.

The algorithms are run for 20 iterations. Results are shown for the gray-level gain and bias model (two parameters) and the full affine channel mixing model (12 parameters).

*Computational time.* We can see in Fig. 2a that the overall computational time needed by DIC and PO is much lower than the one needed by SIC. This also holds against the geometric transformation magnitude and the gain value. Fig. 2b shows that DIC needs slightly more computational time than PO, at worst 1.69 times more. The table below shows detailed timing results in seconds for the parameter update step for a single iteration:

| | G&B | Single G&B | Multiple G&B | Full Affine |
|------|--------|------------|--------------|-------------|
| SIC | 0.0887 | 0.2927 | 0.5070 | 0.8723 |
| PO | 0.0023 | 0.0065 | 0.0065 | 0.0065 |
| DIC | 0.0033 | 0.0098 | 0.0117 | 0.0146 |

The image-warping step respectively takes 0.0270 seconds and 0.0312 seconds in the gray-level and color cases, respectively. We observe that DIC and PO have computational times of the same order of magnitude, while SIC is two orders of magnitude more expensive.

*Geometric error.* This is measured by comparing the estimated transformation at convergence to the true one and reflects the accuracy of the algorithms. As can be seen in Fig. 2c, the geometric error is almost the same for all algorithms. This is also the case for the other experiments we performed and means that all algorithms reach the same accuracy on the estimated transformation when they converge to the sought-after solution.

*Number of iterations.* Fig. 3a shows that the proposed DIC needs about the same number of iterations as SIC, while PO needs more iterations and usually reaches a maximum 20 iterations. Given that the computational cost of an iteration is lower for PO and DIC than for SIC, this explains why the computational time needed by SIC is much larger than the one needed by DIC and PO.

*Photometric error.* The photometric errors (not shown here) reached by all algorithms in all our experiments are identical, meaning that they all are able to minimize the
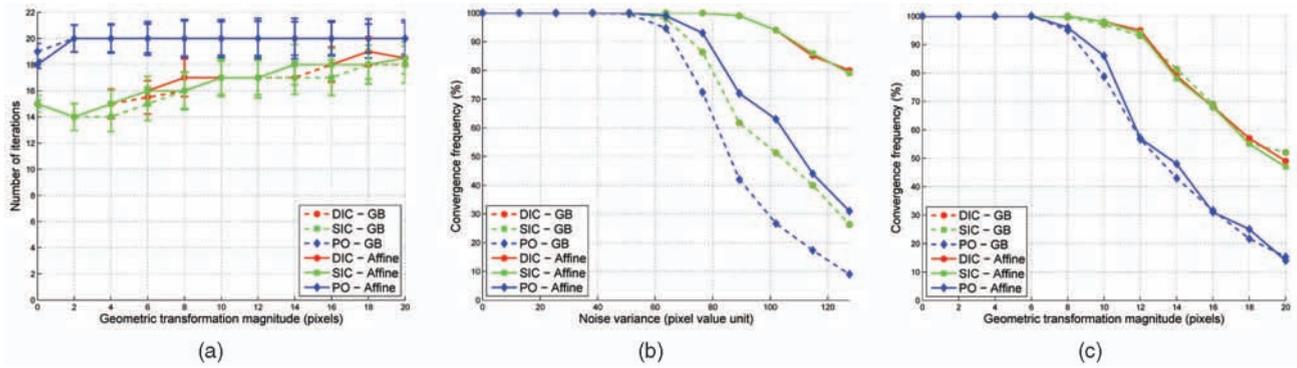
Fig. 3. (a) Number of iterations versus the magnitude of the geometric transformation. (b) Convergence frequency versus the variance of pixel noise. (c) Convergence frequency versus the magnitude of the geometric transformation.

error function to the same extent and, thus, that they all converge to the right solution.

*Convergence frequency.* We observe in Fig. 3b that, beyond a noise variance of 50, the converge frequency drops from 100 percent. Fig. 3c shows that the convergence frequency also decreases when the geometric transformation magnitude $\gamma$ increases. The decrease starts at about 6 pixels for PO and 8 pixels for SIC and DIC and is slightly faster for PO than for the two other methods. Overall, PO has a clearly lower convergence frequency than SIC and DIC. Varying the gain does not affect convergence. The convergence frequencies for DIC and SIC are equivalent.

*Overall.* All three methods are equivalently accurate. PO requires more iterations than SIC and DIC. SIC is much more expensive than PO and DIC and DIC is slightly more expensive than PO. SIC and DIC have equivalent convergence frequencies, significantly higher than the convergence frequency of PO. This means that DIC combines both the stability of SIC with the efficiency of PO and is thus the algorithm we recommend for this kind of groupwise image registration problem.

## 7 CONCLUSION

An algorithm is proposed for the direct registration of two images: the dual inverse compositional algorithm. It is original in that it considers one of the images to be registered as a "generator" for the other one. Its main advantage compared to the simultaneous inverse compositional algorithm is that the Hessian matrix involved in each iteration is constant, making the algorithm very efficient in terms of computational cost. This stems from the fact that, as with the geometric transformation, the photometric one is dealt with using the inverse compositional trick. It is more stable than the project out inverse compositional algorithm. The proposed dual inverse compositional algorithm handles any global groupwise geometric and photometric transformation.

## APPENDIX A

## PARAMETERIZING HOMOGRAPHIES

Groupwise geometric transformations include translations, rotations, affinities, and homographies. We describe the case of homographies. They have 8 degrees of freedom and can be represented by $(3 \times 3)$ homogeneous matrices (i.e., defined up to scale). The representation of the warp by a homography matrix makes it easy to invert a warp or compose two warps, as required by the inverse composition trick, respectively, by inverting the homography matrix and by multiplying the two homography matrices. We use a homography matrix H for the parameterization of the global warp $\mathcal{G}$:

$$\mathcal{G}(\mathbf{q}; \mathrm{H}) = \frac{1}{H_{31}q_1 + H_{32}q_2 + H_{33}} \begin{pmatrix} H_{11}q_1 + H_{12}q_2 + H_{13} \\ H_{21}q_1 + H_{22}q_2 + H_{23} \end{pmatrix}.$$

We constrain H to have a unit two norm. This is enforced each time the update rule is applied by simply dividing H by its two norm.

Following [2], the local warp $\bar{\mathcal{G}}$ is parameterized by an 8-vector $\boldsymbol{\delta}_h$ as

$$\bar{\mathcal{G}}(\mathbf{q}; \boldsymbol{\delta}_h) = \mathcal{G}\left(\mathbf{q}; \mathrm{I} + \begin{pmatrix} \delta_{h,1} & \delta_{h,2} & \delta_{h,3} \\ \delta_{h,4} & \delta_{h,5} & \delta_{h,6} \\ \delta_{h,7} & \delta_{h,8} & 0 \end{pmatrix}\right).$$

This parameterization is such that, as required for deriving the registration algorithms, the identity local warp is obtained for $\boldsymbol{\delta}_h = \mathbf{0}$:

$$\bar{\mathcal{G}}(\mathbf{q}; \mathbf{0}) = \mathcal{G}(\mathbf{q}; \mathrm{I}) = \mathbf{q}.$$

A short calculation shows that the Jacobian of the local warp is

$$(\nabla_{\boldsymbol{\delta}_h} \bar{\mathcal{G}})(\mathbf{q}; \mathbf{0}) = \begin{pmatrix} q_1 & q_2 & 1 & 0 & 0 & 0 & -q_1^2 & -q_1 q_2 \\ 0 & 0 & 0 & q_1 & q_2 & 1 & -q_1 q_2 & -q_2^2 \end{pmatrix}.$$

Inverse composition is performed by multiplying the current homography matrix to the right by the inverse of the incremental one:

$$\mathcal{U}_g(\mathrm{H}, \boldsymbol{\delta}_h) = \mathrm{H}\left(\mathrm{I} + \begin{pmatrix} \delta_{h,1} & \delta_{h,2} & \delta_{h,3} \\ \delta_{h,4} & \delta_{h,5} & \delta_{h,6} \\ \delta_{h,7} & \delta_{h,8} & 0 \end{pmatrix}\right)^{-1}.$$

## APPENDIX B

## THE DUAL INVERSE COMPOSITION ALGORITHMS AND NONGLOBAL PHOTOMETRIC TRANSFORMATIONS

We show that the dual inverse compositional algorithm does not handle nonglobal varying photometric models

such as the Linear Appearance Variations used in Active Appearance Models. We note that the photometric transformation $\mathcal{P}$ now depends on the pixel location $\mathbf{q}$:

$$\mathcal{P}(\mathbf{v};\mathbf{p}) \Rightarrow \mathcal{P}(\mathbf{v};\mathbf{p};\mathbf{q}).$$

We derive the proof for the Linear Appearance Variation model, but the reasoning holds for general nonglobal photometric models. The Linear Appearance Variation model combines basis images $\mathcal{A}_k$ as

$$\mathcal{P}(\mathbf{v};\mathbf{p};\mathbf{q}) = \mathbf{v} + \sum_{k=1}^{l} p_k \mathcal{A}_k(\mathbf{q}).$$

For the case where the basis images are aligned with the source image, the following property holds:

$$\mathcal{P}(\mathcal{S}(\mathbf{q});\mathbf{p};\mathbf{q})) = (\mathcal{P}(\mathcal{S};\mathbf{p}))(\mathbf{q}). \tag{10}$$

When they are aligned with the target image, this transforms as

$$\mathcal{P}(\mathcal{T}(\mathbf{q});\mathbf{p};\mathbf{q})) = (\mathcal{P}(\mathcal{T};\mathbf{p}))(\mathbf{q}). \tag{11}$$

We examine the case where the basis images are aligned with the source image and then when they are aligned with the target image.

## B.1 Basis Images Aligned with the Source Image

*The error function.* Each term of the nonlinear least squares error in (4) is

$$e(\mathbf{q}) = \mathcal{S}(\mathbf{q}) - \bar{\mathcal{P}}^{-1}(\mathcal{P}(\mathcal{T}(\mathcal{G}(\bar{\mathcal{G}}^{-1}(\mathbf{q};\delta_g);\mathbf{g}));\mathbf{p};\mathbf{q});\delta_p;\mathbf{q}).$$

We use the inverse compositional trick on the photometric transformation, giving

$$e(\mathbf{q}) \approx \bar{\mathcal{P}}(\mathcal{S}(\mathbf{q});\delta_p;\mathbf{q}) - \mathcal{P}(\mathcal{T}(\mathcal{G}(\bar{\mathcal{G}}^{-1}(\mathbf{q};\delta_g);\mathbf{g}));\mathbf{p};\mathbf{q}).$$

We then switch the incremental geometric transformation $(\delta_g)$ and the current photometric transformation $(\mathbf{p})$ and, to avoid notational burden, we apply the inverse compositional trick on the geometric transformation at the same time, giving

$$\begin{aligned} e(\mathbf{q}) \approx{}& \bar{\mathcal{P}}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q};\delta_g));\delta_p;\bar{\mathcal{G}}(\mathbf{q};\delta_g)) \\ & - \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q};\mathbf{g}));\mathbf{p};\bar{\mathcal{G}}(\mathbf{q};\delta_g)). \end{aligned}$$

### B.1.2 First Term

By switching the order of the photometric and geometric transformations, the leading term rewrites as

$$e_1(\mathbf{q}) = \bar{\mathcal{P}}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q};\delta_g));\delta_p;\bar{\mathcal{G}}(\mathbf{q};\delta_g)) = (\bar{\mathcal{P}}(\mathcal{S};\delta_p))(\bar{\mathcal{G}}(\mathbf{q};\delta_g))$$

and thus has a constant Jacobian matrix since the photometric transformation does not depend on the pixel location. We rewrite it as

$$e_1(\mathbf{q}) = (\bar{\mathcal{P}}(\mathcal{S};\delta_p))(\bar{\mathcal{G}}(\mathbf{q};\delta_g)) = \left(\mathcal{S} + \sum_{k=1}^{l} \delta_p^k \mathcal{A}_k\right)(\bar{\mathcal{G}}(\mathbf{q};\delta_g)),$$

$$\frac{\partial e_1}{\partial \delta_p^k}(\mathbf{q};\delta_p = \mathbf{0};\delta_g = \mathbf{0}) = \mathcal{A}_k(\mathbf{q}),$$

$$\frac{\partial e_1}{\partial \delta_g}(\mathbf{q};\delta_p = \mathbf{0};\delta_g = \mathbf{0}) = (\nabla\mathcal{S})(\mathbf{q})^{\mathrm{T}}(\nabla_{\mathbf{g}}\bar{\mathcal{G}})(\mathbf{q};\mathbf{0}).$$

### B.1.2 Second Term

The second term has a varying Jacobian matrix. Define

$$e_2(\mathbf{q}) = \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q};\mathbf{g}));\mathbf{p};\bar{\mathcal{G}}(\mathbf{q};\delta_g)).$$

We expand it as

$$e_2(\mathbf{q}) = \mathcal{T}(\mathcal{G}(\mathbf{q};\mathbf{g})) + \sum_{k=1}^{l} p^k \mathcal{A}_k(\bar{\mathcal{G}}(\mathbf{q};\delta_g)).$$

We get

$$\frac{\partial e_2}{\partial \delta_p^k}(\mathbf{q};\delta_p = \mathbf{0};\delta_g = \mathbf{0}) = \mathbf{0},$$

$$\frac{\partial e_2}{\partial \delta_g}(\mathbf{q};\delta_p = \mathbf{0};\delta_g = \mathbf{0}) = \sum_{k=1}^{l} p^k (\nabla\mathcal{A}_k)(\mathbf{q})^{\mathrm{T}}(\nabla_{\mathbf{g}}\bar{\mathcal{G}})(\mathbf{q};\mathbf{0}).$$

These partial derivatives are not constant since they depend on $\mathbf{p}$, which is updated every iteration.

## B.2 Basis Images Aligned with the Target Image

### B.2.1 The Error Function

We replace the argument $\mathbf{q}$ of the photometric transformation by its corresponding point in the target image, which we dub $\gamma = \mathcal{G}(\bar{\mathcal{G}}^{-1}(\mathbf{q};\delta_g);\mathbf{g})$. Each term of the error in (4) is thus

$$f(\mathbf{q}) = \mathcal{S}(\mathbf{q}) - \bar{\mathcal{P}}^{-1}(\mathcal{P}(\mathcal{T}(\gamma);\mathbf{p};\gamma);\delta_p;\gamma).$$

Using the inverse compositional trick on the photometric transformation gives

$$f(\mathbf{q}) \approx \bar{\mathcal{P}}(\mathcal{S}(\mathbf{q});\delta_p;\gamma) - \mathcal{P}(\mathcal{T}(\gamma);\mathbf{p};\gamma).$$

Switching the incremental photometric and geometric transformations and using the inverse compositional trick on the geometric transformation gives

$$f(\mathbf{q}) \approx \bar{\mathcal{P}}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q};\delta_g));\delta_p;\mathcal{G}(\mathbf{q};\mathbf{g})) - \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q};\mathbf{g}));\mathbf{p};\mathcal{G}(\mathbf{q};\mathbf{g})).$$

### B.2.2 Second Term

The second term $f_2 = \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q};\mathbf{g}));\mathbf{p};\mathcal{G}(\mathbf{q};\mathbf{g}))$ is rewritten as

$$f_2(\mathbf{q}) = \mathcal{P}(\mathcal{T}(\mathcal{G}(\mathbf{q};\mathbf{g}));\mathbf{p};\mathcal{G}(\mathbf{q};\mathbf{g})) = (\mathcal{P}(\mathcal{T};\mathbf{p}))(\mathcal{G}(\mathbf{q};\mathbf{g})).$$

This is the warped image $\mathcal{W}$, which does not depend on the unknown incremental parameters $\delta_g$ and $\delta_p$.

### B.2.3 First Term

The first term has a varying Jacobian matrix. Define

$$f_1(\mathbf{q}) = \bar{\mathcal{P}}(\mathcal{S}(\bar{\mathcal{G}}(\mathbf{q};\delta_g));\delta_p;\mathcal{G}(\mathbf{q};\mathbf{g})).$$

We expand it as

$$f_1(\mathbf{q}) = \mathcal{S}(\bar{\mathcal{G}}(\mathbf{q}; \delta_g)) + \sum_{k=1}^{l} \delta_p^k \mathcal{A}_k(\mathcal{G}(\mathbf{q}; \mathbf{g})).$$

We get

$$\frac{\partial f_1}{\partial \delta_g}(\mathbf{q}; \delta_p = \mathbf{0}; \delta_g = \mathbf{0}) = (\nabla \mathcal{S})(\mathbf{q})^{\mathrm{T}}(\nabla_{\mathbf{g}} \bar{\mathcal{G}})(\mathbf{q}; \mathbf{0}),$$

$$\frac{\partial f_1}{\partial \delta_p^k}(\mathbf{q}; \delta_p = \mathbf{0}; \delta_g = \mathbf{0}) = \bar{\mathcal{A}}_k(\mathcal{G}(\mathbf{q}; \mathbf{g})).$$

Therefore, the partial derivatives with respect to $\delta_g$ are constant, but those with respect to the $\delta_p^k$ are varying—they are the warped basis images, depending on the current geometric parameters that are updated every iteration.

## APPENDIX C

## EXPERIMENTAL RESULTS ON REAL DATA

Our goal is to validate our algorithms and compare them to different algorithms on real data sets, to evaluate the photometric transformations presented in Section 5, and to compare the results obtained when using gray-level and color images. We show results for the pair of images shown in Fig. 1. They have been used in Fig. 1 to illustrate the inverse compositional and the dual inverse compositional algorithms. They were acquired by the same camera under different lighting conditions, namely, natural daylight and electric light. The image resolution is $640 \times 480$. The initial warp was chosen as the identity since the images are close enough to enable convergence to the sought-after solution for all methods. The number of pixels of interest chosen near the image edges is 53,889.

We launched the simultaneous inverse compositional, the project out, and the dual inverse compositional algorithms with the images converted to gray level for estimating gain and bias and with the color images and different photometric models (single gain and bias, multiple gains and biases and full affine channel mixing). We show the results in Fig. 4 for the gray-level gain and bias (two parameters) and the color full affine cases (12 parameters). Note that, for PO, we compute the best photometric transformation at each iteration to measure the photometric error, but this is not counted into the measured computational time.

In the gray-level gain and bias case, PO converges to the solution in 53 iterations, while DIC and SIC both require 79 iterations and behave very similarly. All three of them converge to the same solution.

In the color affine case, SIC converges first in 69 iterations, followed by DIC, which requires 95 iterations. Finally, PO uses 204 iterations to converge. All three algorithms converge to the same solution.

We observe that the first iteration increases the photometric error for both SIC and DIC. The magnitude of error variation, both at the increasing and the decreasing phases, is strongly related to the number of parameters in the photometric model. In other words, the more flexible the photometric model is, the steeper the error variation. This behavior is discussed below.
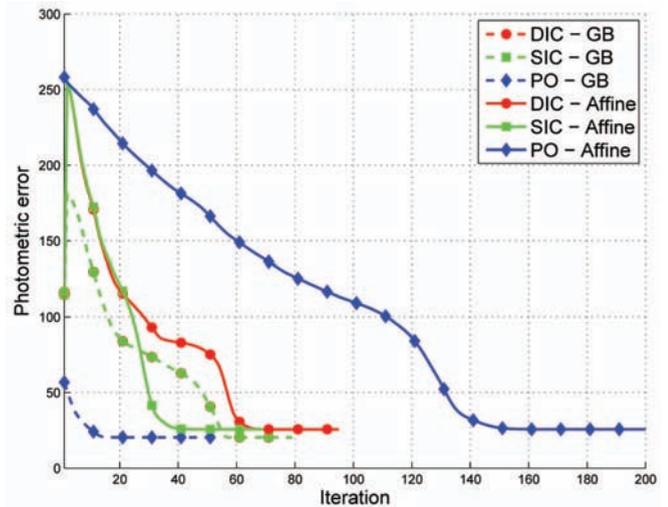


Fig. 4. Color photometric error through the iterations for the image pair shown in Fig. 1 for different photometric transformations and the dual inverse compositional algorithm.

All three algorithms diverge when no photometric model is used. Fig. 5 shows the error image at convergence for the different photometric models.

Finally, we report the total computational time in seconds for each algorithm and each photometric model:

|  | G&B | Single G&B | Multiple G&B | Full Affine |
|---|---|---|---|---|
| SIC | 20.09 | 50.29 | 54.72 | 89.02 |
| PO | 3.88 | 8.57 | 12.81 | 15.24 |
| DIC | 5.95 | 7.03 | 7.45 | 7.81 |

It is clear that SIC is the most expensive algorithm in all cases, being 4 to more than 10 times slower than DIC, while PO is faster in the gray-level gain and bias case but slower in all three color cases.

We observed in our experiments that the photometric error measured throughout the iterations is often increased by the first iteration and, very rarely, by the second iteration for SIC and DIC. This phenomenon is thus not related to the form of the update rule. We have the following hints to understand this behavior:

- The higher the number of parameters in the photometric transformation, the steeper the photometric error curve, both at the increasing and decreasing phases.
- Using edge pixels only makes this behavior stronger compared to using the whole region of interest.

The geometric error reflects the closeness of the estimated warp to the true one. Using simulated data, we assessed the change in the geometric error caused by the first iteration. As already observed by Baker et al. [1], it always decreases. This holds true even if the photometric error increases. In other words, the first iteration brings the warp closer to the sought-after solution, while it takes the photometric transformation away from it since the value of the error function, namely, the photometric error, increases.

The reasons are given as follows: The initial warp causes a geometric misalignment of the images. The intensity or color correspondences from which the photometric transformation

Fig. 5. The error images for different color photometric transformations applied to the image pair in Fig. 1 with the dual inverse compositional algorithm.

is estimated are thus erroneous. All tested algorithms are based on a Gauss-Newton approximation of the error function, which is not fully second order, and thus do not fully preserve the tight coupling between the incremental warp and photometric transformation. The badly estimated incremental photometric transformation, along with the incremental warp, can thus make the photometric error grow, as we observed. This phenomenon is amplified by the fact that the error function is an affine function of the photometric transformation parameters that thus exactly fits the intensity or color correspondences conditioned on the current warp estimate.

## REFERENCES

[1] S. Baker, R. Gross, and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework: Part 3," Technical Report CMU-RI-TR-03-35, Robotics Inst., Carnegie Mellon Univ., Nov. 2003.
[2] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," Int'l J. Computer Vision, vol. 56, no. 3, pp. 221-255, Feb. 2004.
[3] A. Bartoli, "Groupwise Geometric and Photometric Direct Image Registration," Proc. British Machine Vision Conf., 2006.
[4] G.D. Finlayson, M.S. Drew, and B. Funt, "Color Constancy: Generalized Diagonal Transforms Suffice," J. Optical Soc. of Am. A, vol. 11, no. 11, pp. 3011-3019, Nov. 1994.
[5] M. Gouiffes, C. Collewet, C. Fernandez-Maloigne, and A. Trémeau, "Feature Point Tracking: Robustness to Specular Highlights and Lighting Changes," Proc. European Conf. Computer Vision, 2006.
[6] R.C. Hardie, K.J. Barnard, and E.E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images," IEEE Trans. Image Processing, vol. 6, no. 12, pp. 1621-1633, Dec. 1997.
[7] B. Heigl, D. Paulus, and H. Niemann, "Tracking Points in Sequences of Color Images," Pattern Recognition and Image Understanding, 1999.
[8] M. Irani and P. Anandan, "About Direct Methods," Proc. Workshop Vision Algorithms: Theory and Practice, 1999.
[9] H. Jin, P. Favaro, and S. Soatto, "Real-Time Feature Tracking and Outlier Rejection with Changes in Illumination," Proc. Int'l Conf. Computer Vision, 2001.
[10] S.-H. Lai and M. Fang, "Robust and Efficient Image Alignment with Spatially Varying Illumination Models," Proc. Int'l Conf. Computer Vision and Pattern Recognition, 1999.
[11] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto, "Improving Feature Tracking with Robust Statistics," Pattern Analysis and Applications, vol. 2, no. 4, pp. 312-320, 1999.
[12] P.H.S. Torr and A. Zisserman, "Feature Based Methods for Structure and Motion Estimation," Proc. Workshop Vision Algorithms: Theory and Practice, 1999.
[13] B.C. Vemuri, S. Huang, S. Sahni, C.M. Leonard, C. Mohr, R. Gilmore, and J. Fitzsimmons, "An Efficient Motion Estimator with Application to Medical Image Registration," Medical Image Analysis, vol. 2, no. 1, pp. 79-98, 1998.

**Adrien Bartoli** received the PhD degree from INPG/INRIA, Grenoble, France, where he did his research with the Perception Group (ex MOVI) under the supervision of Professors Peter Sturm and Radu Horaud. He was a postdoctoral researcher at the University of Oxford, United Kingdom, in the Visual Geometry Group, under the supervision of Professor Andrew Zisserman. He has been a permanent CNRS research scientist (Chargé de Recherche) at the LASMEA Laboratory, Clermont-Ferrand, France, since October 2004 and a visiting professor at DIKU in Copenhagen, Denmark, from 2006 to 2009. His main research interests are in image registration and structure from motion for rigid and nonrigid scenarios in computer vision. He is the author or coauthor of more than 75 scientific papers. He received the 2004 INPG PhD thesis prize and the Best Student Paper Award at CORESA '07, together with his PhD student Vincent Gay-Bellile and colleague Patrick Sayd. Since September 2006, he has been a co-head of the ComSee research team at LASMEA.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.