

All-pass Parametric Image Registration

Xinxin Zhang, *Student Member, IEEE*, Christopher Gilliam, *Member, IEEE*, and Thierry Blu, *Fellow, IEEE*

Abstract—Image registration is a required step in many practical applications that involve the acquisition of multiple related images. In this paper, we propose a methodology to deal with both the geometric and intensity transformations in the image registration problem. The main idea is to modify an accurate and fast elastic registration algorithm (Local All-Pass—LAP) so that it returns a parametric displacement field, and to estimate the intensity changes by fitting another parametric expression. Although we demonstrate the methodology using a low-order parametric model, our approach is highly flexible and easily allows substantially richer parametrisations, while requiring only limited extra computation cost.

In addition, we propose two novel quantitative criteria to evaluate the accuracy of the alignment of two images (“salience correlation”) and the number of degrees of freedom (“parsimony”) of a displacement field, respectively. Experimental results on both synthetic and real images demonstrate the high accuracy and computational efficiency of our methodology. Furthermore, we demonstrate that the resulting displacement fields are more parsimonious than the ones obtained in other state-of-the-art image registration approaches.

Index Terms—Image registration, parametric fitting, geometric transformation, intensity transformation, registration evaluation, local all-pass filters.

I. INTRODUCTION

IMAGE registration is the process of finding the geometric transformation between two Cartesian coordinate systems so as to align two or more images. Image registration plays a significant role in medical research [1], [2], remote sensing [3], geological prospection, computer vision [4] and many other fields of modern science. Aligning two or more images of the same scene taken by the same sensor is known as monomodal registration, while aligning these images of the same scene taken by different sensors is known as multimodal registration. In this paper, we deal with both monomodal and multimodal registration of two images. In terms of monomodal image registration, a variety of situations should be considered including geometric transformation and intensity transformation in the condition of blur or noise distortions. When it comes to multimodal image registration, image pairs of different modalities [5], [6] have different appearance.

a) *Mathematical setting*: In a pair of images to be registered, the fixed image is referred to as the target image and the other one is the source image, we denote them by I_1 and I_2 , respectively. The two images may be acquired from different uncalibrated cameras, with different resolution, field of view (FOV) and (non-uniform) illumination. However, when these

This work was supported in part by a grant #CUHK14207718 of the Hong Kong Research Grants Council.

C. Gilliam is with the School of Engineering, RMIT University, Australia. X. Zhang and T. Blu are with the Department of Electronic Engineering, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: xxzhang@ee.cuhk.edu.hk; thierry.blu@m4x.org).

images are related only through a geometric transformation, without any change of light intensity the images are related by the brightness constancy equation [7], [8]:

$$I_1(x, y) = I_2(\mathcal{T}(x, y)), \quad (1)$$

where $\mathcal{T}(x, y) \rightarrow (x + u_x(x, y), y + u_y(x, y))$ is a geometric transformation. In the sequel, we will often assimilate the displacement field $(u_x(x, y), u_y(x, y))$ to a complex function: $u(x, y) = u_x(x, y) + iu_y(x, y)$ and so, the key task of image registration is to estimate $u(x, y)$ such that I_2 is aligned to I_1 (see Fig. 1).

Many registration methods rely on the brightness constancy equation (1) or on a Taylor-linearized version. In real applications, however, the brightness constancy assumption is often violated thus these methods are likely to be error prone. Accounting for photometric changes can be done by introducing a functional $\mathcal{F}(\cdot)$ that relates the intensities of the source and the target images:

$$I_1(x, y) = \mathcal{F}(I_2(x + u_x(x, y), y + u_y(x, y))). \quad (2)$$

This functional may be a combination of convolutions, of nonuniform illumination changes, of non-linear pointwise operations etc. Hence, successful image registration should also involve estimating the intensity transformation $\mathcal{F}(\cdot)$.

The inherent assumption in (2) is that there is *only one* displacement value for every pixel. Obviously, when the same pixel overlays several objects (e.g., when the images are 2D projections of 3D scenes) that move differently, this assumption is violated. Faced with this issue, a registration algorithm attempts to restore the displacement that is the *most continuous* (i.e., the most consistent) with the rest of the image, whereas an optical flow algorithm attempts to restore the displacement of the foreground object, resulting in a *possibly discontinuous* displacement field: two different solutions of an ill-posed problem. In this paper, we focus on image registration.

b) *State of the art*: Depending on the geometric transformation model, image alignment algorithms may be either global [9]–[12], or elastic [13]–[19].

Global image registration methods use global parametric models to describe the displacement field, so the registration problem reduces to calculating the parameters of the model which can be fast when very few parameters are considered. In the literature, these methods are usually limited to parametric registration which solves for geometric homographies; i.e., combinations of shifts, rotations, scale changes and shears. In order to cope with more complex transformations, more parameters are needed and the algorithms for retrieving them become unreliable.

Elastic registration methods estimate a displacement vector per pixel and are able to deal with local and complex distor-

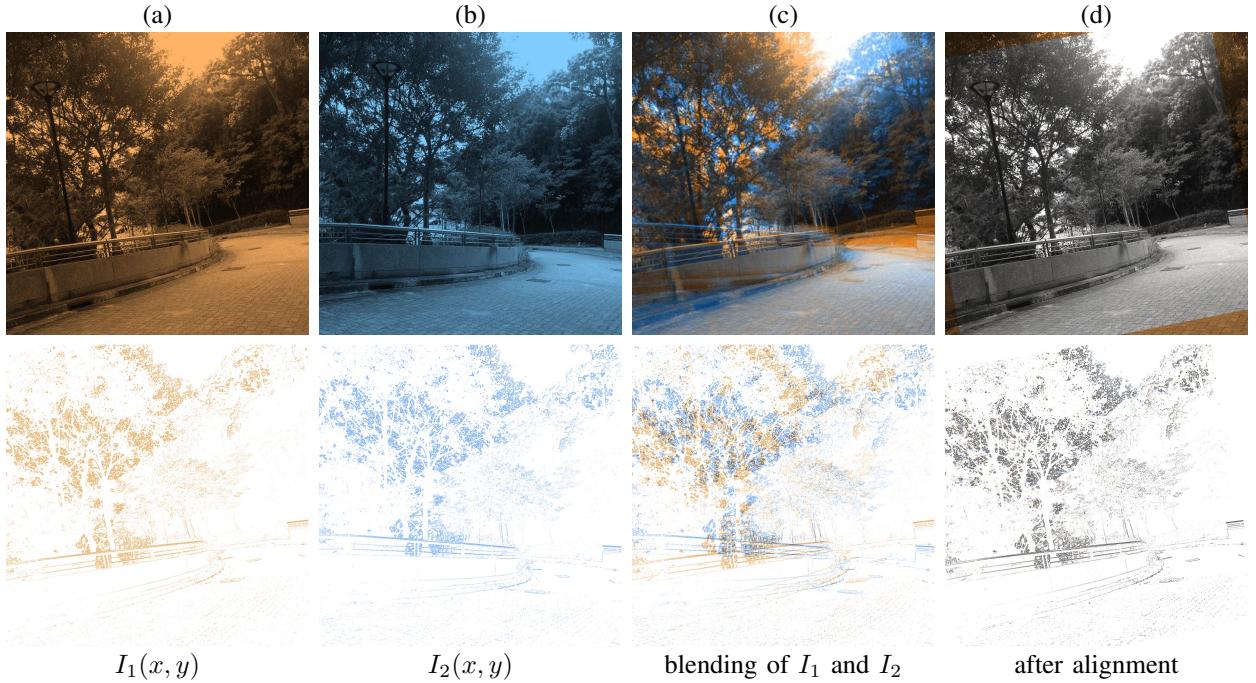


Fig. 1. Top row: the target image (a, orange-toned) is related to the source image (b, blue-toned) through a geometric transformation according to Equation (1) or (2). Blending these images (c) outlines their mis-alignment. Perfect alignment (d) is achieved when the blended image is tone-neutral (i.e., gray). Bottom row: the intensities of the same images at the 10% largest gradient values (visualization better adapted to multimodal images).

tions. But they are usually slow, and quite sensitive to blur, noise and other intensity changes.

A popular registration criterion is the preservation of the brightness between images which, however, leads to algorithms that are very sensitive to intensity inaccuracies. In order to cope with these sensitivity issues, many registration approaches have attempted to remove intensity differences with pre-processing steps. However, these operations usually lead to loss of useful information. Periaswamy and Farid [15], [17] built an intensity model accounting for local and global variation in image intensity. Evangelidis and Psarakis [20] applied a first order Taylor expansion with respect to the transform parameters to deal with the intensity changes. In [21], an iterative method was introduced to estimate both geometric and intensity transformation, in which either monofunctional or bifunctional dependence between intensities of images was assumed. These methods are usually effective for intensity changes but are computationally heavy and are likely to fail when the intensity changes are not deterministic.

Mutual information [22]–[24] is another registration criterion, capable of dealing with random intensity relations between two images, and is one of the most popular strategy for multimodal image registration. Its major drawbacks are the sensitivity to noise [25], to non-uniform illumination changes, and heavy computation.

Using features [26]–[33] leads to yet another approach to image registration: the deformation is estimated by establishing the correspondence between the extracted features or landmarks from the two images. Due to their invariance to translation, rotation and scaling, SIFT descriptors are one of the most popular sets of features used in this type of registration [34]. The SIFT algorithm has undergone various

improvements in recent years, e.g. SURF [35], SIFT flow [33], PCA-SIFT [36], GSIFT [37], CSIFT [38] etc. Although SIFT performs reasonably well on natural images, it fails to produce an acceptable number of features and matches when applied to medical data, because such images usually contain large homogeneous regions with little edge contents [39]. The advantage of feature-based approaches is that they reduce the computational complexity and, in the cases where reliable features are found, can deal with both monomodal and multimodal registration problems. However, the performance of these methods is closely associated with the feature types and the accuracy of the feature detection: a selection of methods [30]–[32] were shown to be relatively robust to intensity changes and modality difference, but they were prone to fail for feature-less images, e.g. images contains large area of sky, sea, grass, very smooth region, etc.

Image registration has also been addressed using deep learning methods [40]–[48] but, in most cases, the problem considered was 3D registration and the available 2D registration algorithms did not perform reliably in our tests. Fortunately, 2D optical flow estimation has also been approached using deep neural networks [49]–[52] and, although such algorithms put a lower emphasis on the continuity of the displacement field as mentioned earlier, we have found comparing with one of them to be meaningful [50].

Assessing the alignment quality is relatively straightforward when the ground-truth displacement field is known. For instance, a number of registration approaches [4], [53] show the quantitative results on synthetic images. However, in real applications this ground-truth is usually not available. To cope with this problem, it is possible to approximate the underlying displacement field from manually labeled landmarks [4], [54],

[55] but this procedure suffers from being non-automatic, non-objective and unlikely to yield subpixel accuracy. Accordingly, alignment quality is generally assessed subjectively by only comparing the images after alignment (e.g. using the PSNR, or normalized cross correlation [56]). A flaw of such subjective assessment is that less attention is paid to the “complexity” of the displacement field thus it is possible to align images perfectly using very unrealistic displacement fields (see Section IV). The realistic nature of the displacement field is of particular importance in areas such as biomedical imaging [57].

c) *Contributions and outline:* The major contribution of this paper is a highly accurate and computationally efficient methodology to solve image registration problems by approximating the displacement field and the intensity changes between the images using parametric representations. These representations can be chosen to be local, global, or mixed, although we will only exemplify our approach using quadratic 2D polynomials (global). We want to stress that the parametric model chosen is but an *approximation* of the displacement field, not its exact representation: our algorithm is able to retrieve that approximation. Of course, we also argue that this approximation is quite accurate for digital cameras.

The framework of our method is based on a highly efficient elastic registration algorithm, the local all-pass filtering algorithm (LAP) [58], which is iterated from large to small filter sizes, in a coarse-to-fine manner. At each iteration, our chosen geometric parametric model is fitted to the LAP displacement field; in turn, our parametric intensity transformation model is fitted to the intensity mismatch after warping. More specifically, the displacement and intensity change are modelled as linear combinations of elementary displacements/intensity changes, respectively, which ensures that the fitting steps can be performed very efficiently by solving linear systems of equations. An exploratory application of this methodology was presented in [59].

Another important contribution of this paper is to propose a set of two indicators that quantify the quality of the alignment procedure: the “salience correlation” (SalC), which evaluates the percentage of features that are common to the two pictures after warping; and the “parsimony” (Pars), which evaluates the complexity of the displacement field retrieved. Hence, between two registration algorithms that achieve a similar salience correlation (typically, above 50%), the “best” one should be the simplest—the most parsimonious.

The rest of the paper is organized as follows. In Section II, we introduce the Local-All Pass (LAP) registration framework which was originally presented in [60]. Then, in Section III, we describe the parametric model for both monomodal registration and multimodal registration in detail including our geometric transformation model and our intensity transformation model. In Section IV, we propose new criteria for evaluating the performance of registration methods. In Section V, we perform extensive comparisons on both synthetic images and real data with several state-of-the-art techniques qualitatively and quantitatively. We also demonstrate (Section V-C) how our algorithm performs in a multispectral image application.

II. THE LOCAL ALL-PASS REGISTRATION ALGORITHM

The main idea of the LAP [58] is that a constant shift is equivalent to filtering with an all-pass filter when the brightness constancy hypothesis is satisfied as illustrated in Fig. 2. Making use of this idea locally implies assuming that the displacement field is slowly changing across space. This contrasts with the use of the optical flow equation (i.e., the first order Taylor approximation of the brightness constancy equation), which assumes that the image itself is slowly changing across space.

In the frequency domain, the shifting relationship between the source image and target image is equivalent to

$$\hat{I}_2(\omega_x, \omega_y) = e^{-iu_x\omega_x - iu_y\omega_y} \hat{I}_1(\omega_x, \omega_y) = \underbrace{\hat{H}(\omega_x, \omega_y)}_{\text{real all-pass filter}} \hat{I}_1(\omega_x, \omega_y)$$

where \hat{I} represents the Fourier transform of the image I and $\omega = (\omega_x, \omega_y)^T$ denotes the 2D frequency coordinates. Now, it turns out that real all-pass filters can always be expressed as

$$\hat{H}(\omega_x, \omega_y) = \frac{P(e^{i\omega_x}, e^{i\omega_y})}{P(e^{-i\omega_x}, e^{-i\omega_y})},$$

where P is a finite impulse response filter. Then, the idea to “localize” the all-pass filter consists in approximating the filter $P(e^{i\omega_x}, e^{i\omega_y})$ as a linear combination of a few (three here) fixed, known real filters with given spatial support. Hence, the approach for determining $\hat{H}(\omega_x, \omega_y)$ amounts to finding the coefficients corresponding to the filter basis. Then, the local displacement is extracted from the impulse response $p_{k,l}$ of the filter P according to

$$u = 2 \frac{\sum_{k,l} (k + il) p_{k,l}}{\sum_{k,l} p_{k,l}}$$

and validated when found to be within the spatial support of P —the resolution of the LAP. In practice, this algorithm is implemented iteratively by changing this spatial support from large to small (poly-filter LAP), in such a way as to deal with both large and small geometric deformations.

Compared with the state-of-the-art deformation estimation methods, the poly-filter LAP algorithm is very fast and highly accurate when the brightness constraint is exactly satisfied. However, the algorithm becomes inaccurate when this constraint is violated, and it is necessary to resort to pre- and post-processing (high-pass filtering, inpainting, smoothing) in order to preserve high-quality results. In addition, as in other elastic registration algorithms, large displacements are not estimated accurately. It is these defects that we wish to remedy with our fitting approach.

III. ALL-PASS PARAMETRIC FRAMEWORK

An overview of our approach to improve the poly-filter LAP registration algorithm is sketched in Fig. 3, and relies on three main steps: given a pair of images I_1 (target), I_2 (source), and a previous estimate, $u^{(j-1)}$, of the displacement field between these images,

- 1) we fit a parametric intensity relation between $I_1(x, y)$ and $I_2(x + u_x^{(j-1)}(x, y), y + u_y^{(j-1)}(x, y))$;

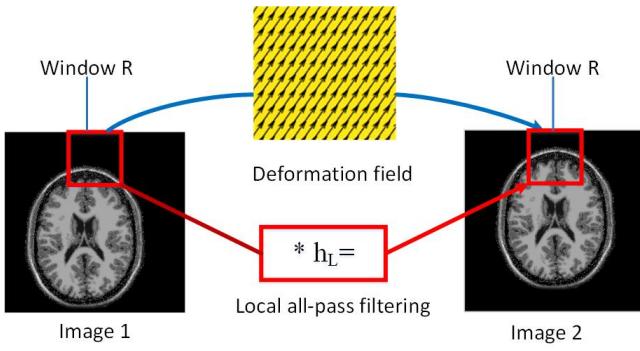


Fig. 2. Illustration of the equivalence between constant displacement field and filtering with an all-pass filter. Note that the locations of the two windows on the right-hand side and left-hand side are exactly the same.

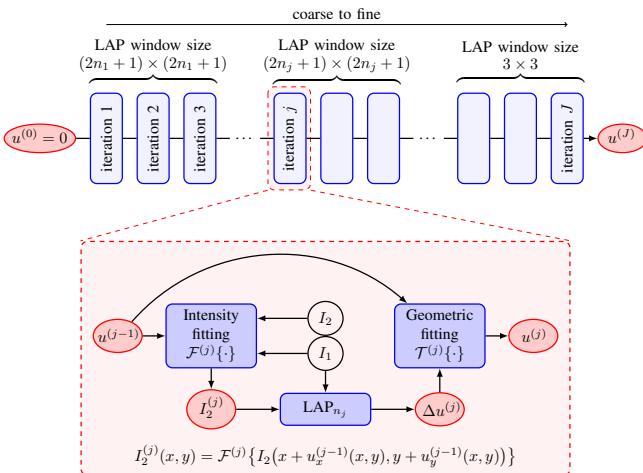


Fig. 3. Registering the source image I_2 to the target image I_1 is performed in a finite number of iterations: the resolution is increased by a factor 2 every three iterations, from a coarse resolution $n_1 = \min(\text{size of } I_1)/4$ to the fine 1-pixel resolution. An enlargement of a single iteration at resolution n_j details the sequence of parametric fittings (displacements, intensities) and LAP.

- 2) we apply one iteration of the LAP algorithm (i.e., at some resolution) to estimate the residual displacement field between $I_1(x, y)$ and the intensity-adjusted warped version of $I_2(x, y)$;
- 3) we fit a parametric expression to the LAP displacement field, and update the full displacement field with this approximation.

Typically, the functional $\mathcal{F}^{(j)}\{\cdot\}$ is a linear combination of elementary functionals including convolutions, non-uniform illumination changes, non-linear point-wise operations, etc. Similarly, the displacement field is expressed as a linear combination of elementary displacement fields, including shifts, rotations, scale changes, shears, but also higher order polynomial distortions etc.

Hence, the basic idea of our algorithm is to find the few parameters that best describe both the intensity transformation and the geometric distortion, and this can be implemented very efficiently because this amounts to solving a linear system of equations for each model. We now give more motivation and details about our parametric models.

A. Parametric fit of the displacement field

a) Methodology: Image registration is inherently ill-posed, hence smoothness or sparse hypotheses [1], [2], [61] are always required so as to constrain the spatial transformation and make the problem well-posed. Our choice is to express these constraints explicitly through a parametric model, contrary to the more standard regularization approach [62], because this will be significantly more computationally efficient—hence, iteration-friendly.

The general setting consists in expressing the displacement field $u(x, y)$ as a linear expansion on a basis of elementary displacements $u_k(x, y)$, i.e.

$$u(x, y) = \sum_{k=1}^K a_k u_k(x, y), \quad (3)$$

where K is the total number of basis functions. For instance, choosing $u_1(x, y) = 1$, $u_2(x, y) = x$ and $u_3(x, y) = y$ allows to represent all linear geometric transformations but, of course, adding other basis elements is likely to increase the accuracy of the approximation. More general polynomial or Fourier series expressions, but also local transformations (e.g., expressed onto a local basis like a uniform B-spline basis [63]) and wavelet decompositions can be considered.

Estimating the displacement on a global level boils down to calculating the K coefficients in (3). This is achieved by minimizing the difference between the displacement that results from the LAP algorithm, u_{LAP} , and our parametric expression for u :

$$\min_{a_k} \sum_{x, y \in \Omega} |u(x, y) - u_{\text{LAP}}(x, y)|^2, \quad (4)$$

where Ω is the fitting region (see Subsection III-C). Note that we use the least-squares error metric without extra regularization. The solution to (4) is equivalent to solving a linear system of K equations with K unknowns, which is quite fast while providing the global minimum of the problem.

b) Specific parametric model: the quadratic expression

$$u(x, y) = a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 y^2 + a_6 x y, \quad (5)$$

has been found to be an accurate representation of the displacement field involved in digital cameras [53], [64], [65]. A partial justification could be that, for objects of limited thickness that are positioned at a well-defined distance from the camera, a reasonable parametric model that accounts for the optical distortions of the resulting photograph is a 2D homography [66]–[68]: $u(x, y) = (\alpha x + \beta y + \gamma)/(ax + by + c)$ where the parameters α, β, γ are complex-valued and a, b, c are real-valued. The denominator $ax + by + c$ is typically proportional to the depth of the object located at (x, y) and so, when this depth z_0 does not change significantly across the object we can assume that $|ax + by + c - z_0|/z_0 \ll 1$, which makes it possible to approximate the denominator $(ax + by + c)^{-1}$ by $z_0^{-1} - z_0^{-2}(ax + by + c - z_0)$ —i.e., a 2D quadratic polynomial.

c) *Warping*: The resulting parametric displacement field is used to warp the source image closer to the target image. Since the estimated displacement is non-integer, it is essential to build a continuous model of the image for image warping. We adopt the shifted linear interpolation [69] for all the LAP windows larger than 5×5 pixels: we have shown in [69] that (optimally) shifted linear interpolation achieves the high quality of cubic interpolation [70], while retaining the low computational cost of linear interpolation. However, at finer filter resolution, i.e. for LAP windows smaller than or equal to 5×5 pixels, we use cubic-OMOMS [71] interpolation which we have shown to reach a quality that is significantly higher than cubic interpolation, at a equivalent computational cost.

Although we mostly use a quadratic polynomial model in the current paper, we want to stress again that our framework allows for much richer “elastic” models based on, e.g., global representations (higher order polynomials, Fourier series) or local representations (B-spline basis, radial basis functions). The linear fitting strategy makes it easy to test empirically bases for the displacement field. We would like to point out, however, that although increasing the model order ensures a potentially more accurate approximation of the displacement field, the ill-posed nature of the registration problem is likely to result in a worse *actual* approximation, when the model order is too high.

B. Estimation of the intensity transformation

As with the representation of spatial transformations, the intensity transformation $\mathcal{F}\{\cdot\}$ in (2) is modelled as a linear combination of elementary intensity transformations, i.e.

$$\mathcal{F}\{\cdot\} = \sum_{l=1}^L \gamma_l \mathcal{F}_l\{\cdot\}, \quad (6)$$

where L is the number of basis functionals $\mathcal{F}_l\{\cdot\}$. Then the intensity transformation problem boils down to finding the L coefficients $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_L]^T$, which is achieved by least-squares minimization of the difference between the target image intensity and the source image intensity after warping, within the fitting region (see Subsection III-C). The parametrization of the intensity transformation allows the algorithm to deal with not only global illumination variations, but also with the possible blurring of either image, or with deterministic correspondences between the intensities of the two images.

This parametrization makes it possible to deal with multimodal image alignment, provided that the expected relation between the intensities of the two modalities is, at least partly, deterministic. That such an explanatory relation exists is actually an underlying assumption in one of the major methods for multimodal image registration, the maximization of mutual information [5], [22]–[24]: it is known that the mutual information of two sources reaches its theoretical maximum when there is a one-to-one relation between them.

1) *Illumination changes*: In a real scenario, due to different illumination, different exposure conditions, different weather, different camera settings or embedded post-processing, the brightness constancy assumption is likely to be violated, an

outcome that may also result from, e.g., shadow, noise, blur or other artifacts. Hence, it is necessary to build a model accounting for the photometric transformation. In terms of natural images, the most common issues that we encounter are different illumination and exposure, e.g. flash/no-flash pairs, images taken under different weather or different times.

In the case of illumination changes, the relationship described in Equation (6) can be chosen as:

$$I_1(x, y) = \alpha(x, y) \times I_2(x + u_x(x, y), y + u_y(x, y)), \quad (7)$$

where the function $\alpha(x, y)$ is used to model the intensity transformation between I_1 and I_2 . As with the geometric transformation, we assume the intensity transformation between two images changes smoothly across the image and use a global quadratic polynomial function to describe the illumination changes:

$$\alpha(x, y) = b_1 + b_2 x + b_3 y + b_4 x^2 + b_5 y^2 + b_6 xy; \quad (8)$$

where b_k ($k = 1, 2, \dots, 6$) are real-valued coefficients. Although the shadow and saturated regions are local problems which cannot be modelled by a quadratic polynomial function, these locations are likely to be excluded automatically from the fitting region during the fitting process (see Subsection III-C). Thus the key task of illumination changes estimation is to calculate the 6 coefficients by solving a linear system of equations. Moreover, the model can be easily changed to a higher order polynomial function or other types. For instance, a linear combination of shifted cubic B-spline functions is able to model local intensity distortions, contrary to our global quadratic polynomial model.

2) *Blurring*: In real applications, blur is a common phenomenon in the process of imaging, including out-of-focus blur, atmospheric blur and motion blur. Registration and motion estimation methods, especially elastic methods are usually sensitive to blur. Different blur levels between images often lead to boundary distortions in the registered results. To deal with this issue, we propose to estimate the blur relation between the two images by approximating it as a linear combination of elementary filters. Here, the type of blur we exemplify is uniform out-of-focus blur and atmospheric blur which are characterized by a Gaussian impulse response.

Assume the target image is more blurry than the source image, the functional relation $\mathcal{F}\{\cdot\}$ in Equation (6) becomes:

$$I_1(x, y) = (h * I_2)(x + u_x(x, y), y + u_y(x, y)), \quad (9)$$

where $h(x, y)$ is a Gaussian filter. If it is the source image instead that is more blurry, then I_1 and I_2 should be swapped: $I_2(x, y) = (h * I_1)(x + u_x(x, y), y + u_y(x, y))$. We use the average gradient of the images to judge which one is more blurry. After the blur compensation, the two images have the same blur level. Finding the adequate out-of-focus blur that relates two images can be done by estimating the variance of the Gaussian function [72], [73], but we choose a more direct approach, based on the observation that a Gaussian function can be approximated very accurately as linear combination of Gaussians with fixed variance

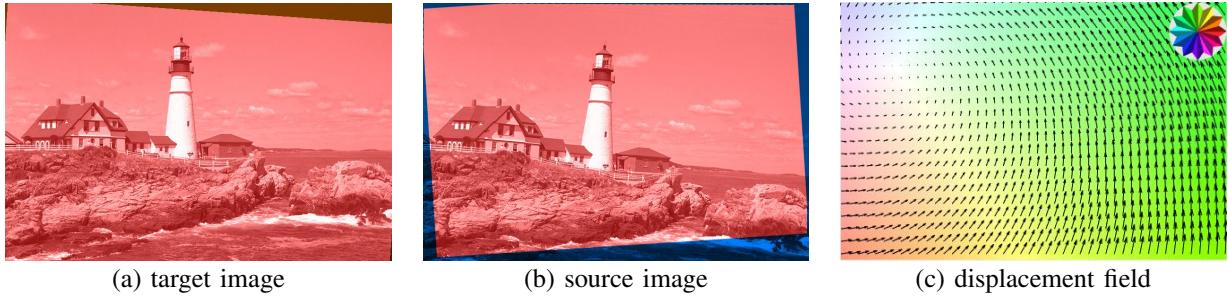


Fig. 4. (a) and (b) illustrate the fitting region \mathcal{R}_1 and its mapping \mathcal{R}_2 (marked in red) within the target image and source image, respectively. The compass rose in (c) indicates our color encoding of the direction of the displacement field (also shown by the arrows).

$$h(x, y) \approx \sum_{n=1}^N c_n h_n(x, y), \quad (10)$$

where $h_n(x, y)$ is one of the N given Gaussian filter basis and c_1, c_2, \dots, c_N are the corresponding coefficients to be estimated. The number N of basis functions depends on our knowledge of the variance σ^2 of the Gaussian function $h(x, y)$: for instance, if that variance is within a range of the form $[s^2, 4s^2]$ for some known s , then we can estimate $h(x, y)$ with an accuracy $\approx 6 \times 10^{-3}$ using only $N = 3$ Gaussian functions of variances $s^2, 2s^2$ and $4s^2$.

As with geometric transformation estimation, estimating the blur transformation is equivalent to calculating N coefficients by solving a small linear system of equations. Hence, it is very fast.

3) *Pixel-wise mapping*: Image pairs of the same subjects, taken by different devices are instances of multimodal images; for example, CT, MRI and PET image pairs in medical imaging, color and infrared image pairs in remote sensing, RGB and depth image pairs, R channel, and G channel and B channel image pairs. Differently from what we have just seen, the corresponding intensities of the same object are unlikely to be modelled adequately by a slowly varying spatial functional.

Instead, we postulate a one-to-one mapping between the intensities in the overlap region of the two images if image pairs are well-registered. As mentioned at the beginning of this section, a possible justification of such a one-to-one relation is that it is the expected outcome of aligning them by maximizing their mutual information. Histogram matching is precisely such a one-to-one mapping [74, p.128]. Accordingly, we suggest to apply a final histogram matching step in (2) after the spatial intensity prediction (6) of the target image from the intensities of the warped source image.

C. Fitting region determination

As with other approaches [75]–[77], our algorithm has to identify the region of overlap (“fitting” region) between the images to register; i.e., the set of pixel locations $(x, y) \in \mathcal{R}_1$ within the target image I_1 where the parametric representation (5) holds. This overlap is characterized by the two conditions

- 1) (x, y) is within the range of the pixels of I_1 ,
- 2) $(x + u_x(x, y), y + u_y(x, y))$ is within the range of the pixels of I_2 .

Within the source image I_2 , this overlap (denoted \mathcal{R}_2) is the set of pixels $(x + u_x(x, y), y + u_y(x, y))$ where (x, y) span \mathcal{R}_1 . An example of fitting region is shown (marked in red) in Fig. 4 (a) and (b).

However, determining a reliable fitting region is a chicken-and-egg problem since the accurate overlap region is only known after successful registration. We solve this problem by updating the fitting region at each iteration of our algorithm based on the following simple exclusion rules: locations (x, y) in the target image I_1 that are

- either invalidated by the LAP algorithm (typically, because the displacement found is larger than the LAP resolution),
- or such that $(x + u_x(x, y), y + u_y(x, y))$ lies outside the range of the pixels of I_2 ,

are excluded, and the pixels that remain form the updated fitting region \mathcal{R}_1 . In particular, shadows, occlusions and more generally whatever is not predictable on the target image is likely going to be invalidated by the LAP—hence, excluded from the fitting region.

During the first coarse iterations, the estimated fitting regions are significantly different from the ground-truth overlap regions in most cases, but as the number of iterations increases, the fitting region becomes increasingly closer to the true overlap region. Here, we benefit from the fact that, in our method, the global parametrization makes it possible to extrapolate the displacement field from a limited region, and this is particularly useful in the case of large deformations where many algorithms usually fail. Using an isotropic diffusion inpainting algorithm to estimate the deformations of erroneous regions [58], [60] would likely be less accurate and more time consuming. Our strategy is not limited to global parametrizations, though: it still works quite well with local parametric representations (3), provided that the number of parameters is small enough [63], [78].

D. Comments

We are improving the poly-filter LAP algorithm [58] from several aspects: First, we are able to cope with intensity changes and do not rely anymore on the intensity constancy assumption—an essential condition for the PF-LAP to succeed. Second, all the pre- and post-processing operations (typically: Gaussian smoothing, median filtering, deformation inpainting) are more accurately and efficiently replaced by the

estimation of the parameters of the displacement model in the fitting region. The gain in accuracy is particularly striking in the case of information loss, shading and occlusion.

IV. EVALUATION METRICS

It is customary to evaluate the quality of image registration algorithms on synthetic data by comparing the displacement fields or registered images with the ground-truth [4], [53]. When a ground-truth is not available, human-labeled landmarks (a non-objective metrics) are usual. However, when such landmarks are not available, it is frequent to rely only on the visual comparison between the registered image and the target image, without consideration for the displacement field. We show in Fig. 5 that this is an issue: it is possible to have a perfect visual alignment, with a terrible displacement field.

In this section, we make a case for two objective metrics: one that measures how two aligned images match (the salience correlation), and one that measures how “simple” the displacement field is (the parsimony).

1) *Salience correlation (SalC)*: Under the condition of strict brightness constancy, the registered image should be exactly the same as the target image. In this case, PSNR is a good measure for similarities. However, in real situations, the PSNR of well-aligned images is often below 20 dB, at a value that can also be reached by ill-aligned images, which shows that it cannot be used as an objective alignment criterion. The normalized cross correlation (NCC) [79] reduces the influence of intensity changes, but is too sensitive to occlusions and noise. Also, despite its success at multimodal image registration as an optimization criterion, using the actual value of the mutual information between two images would also be largely unreliable for the purpose of evaluating alignment quality.

Yet, it is expected that, even in the case of intensity changes or occlusions, the salient features of a correctly registered image $I_2^{(\text{reg})}(x, y) = I_2(x + u_x, y + u_y)$ and the target image $I_1(x, y)$ are well-aligned. This suggests adapting a correlation measure so that it involves only the most salient (i.e., large gradient) pixels—the salience correlation:

$$\text{SalC} = \frac{\left| \int_{\mathcal{R}_1} I_1(x, y) \mu_1(x, y) \cdot I_2^{(\text{reg})}(x, y) \mu_2(x, y) dx dy \right|}{\sqrt{\int_{\mathcal{R}_1} I_1(x, y)^2 dx dy \cdot \int_{\mathcal{R}_1} I_2^{(\text{reg})}(x, y)^2 dx dy}} \times 100\% \quad (11)$$

where \mathcal{R}_1 is the overlap region (defined in Section III-C) and the mask maps μ_1, μ_2 indicate the location of the 10% largest values of $\|\nabla I_1\|$ and $\|\nabla I_2^{(\text{reg})}\|$, respectively. The 10% threshold was chosen based on numerous experiments and, typically, in monomodal registration problems without severe intensity distortion, we observe consistently that the salient features of the image are accurately aligned when the SalC is larger than 50%.

In practice, we implement the gradient in the Fourier domain after mild Gaussian prefiltering (standard deviation 0.75, i.e., a resolution similar to a 2×2 pixel window) so as to focus on the most important features.

2) *Parsimony (Pars)*: Although most image registration algorithms include a smoothness/sparsity constraint (e.g., [62]), we have not found any attempts to evaluate objectively the sparsity of the resulting displacement field in the literature: on real data, the emphasis is mostly laid on the accuracy of the alignment of the features/landmarks. We believe that this is largely insufficient because, it is possible to have a perfect alignment of features with a ridiculously incorrect displacement field, as shown in Fig. 5. We also demonstrate this issue on real images in Fig. 6.

For this reason, we propose an isotropic total variation-based metric that aims to quantify the sparsity/number of degrees of freedom of a displacement field—“parsimony”:

$$\text{Pars} = \frac{\int \|\nabla u(x, y)\| dx dy}{\int |u(x, y)| dx dy} \times C. \quad (12)$$

where the (image size-dependent) normalization constant C is chosen in such a way that $\text{Pars} = 1$ when $u(x, y)$ is the displacement field resulting from a rotation around the center of the image. The smaller the parsimony, the more concise the modelling of the displacement field, which renders it more plausible, according to the heuristic principle known as Ockham’s razor [80].

We provide in Fig. 6 a visual depiction of the salience correlation and the parsimony on a real alignment experiment.

3) *Guidelines*: Obviously, a salience correlation that is as close as possible to 100% and a parsimony as small as possible are desirable features of a registration algorithm. Hence, if algorithm 1 is characterized by, both a larger SalC and a smaller Pars than algorithm 2, then it is reasonable to consider that algorithm 1 is “better” than algorithm 2.

Now, a less decidable situation could happen, where algorithm 1 has a higher SalC (good) and a higher Pars than algorithm 2 (bad): algorithm 1 achieves better alignment of features, at the cost of a more complex displacement field. As we know (see Figs. 5 and 6), the alignment provided by algorithm 1 may well be unrealistic, even more so than algorithm 2. Based on our empirical observation, a partial answer to this dilemma is that, as soon as SalC reaches a minimum of 50%, it is essentially the parsimony that rules the quality of the alignment: achieving a higher percentage of feature alignments than 50% does not significantly increase this quality, whereas a smaller parsimony renders this alignment more likely.

This observation seems to hold true generally for monomodal image registration problems, without significant intensity distortion. In more general situations (e.g., blurring, large intensity variations, multimodal images, optical flow problems), the SalC can be significantly lower than 50%, even after perfect alignment. Then, the application of the principle above (SalC above a threshold, then minimise Pars) should be adapted on a case-by-case basis—typically, identify a valid lower SalC threshold that guarantees feature alignment.



Fig. 5. Unrealistic alignment of “Riemann” to “Lebesgue”. The displacement field is estimated by histogram matching: target and source pixels are paired according to their intensity rank. The salience correlation before alignment is obviously low at 14%, but reaches 64% after registration. This almost surely incorrect alignment demonstrates the need for a measure of the likeliness of the displacement field—the parsimony. Here, the parsimony of the displacement field is particularly large.

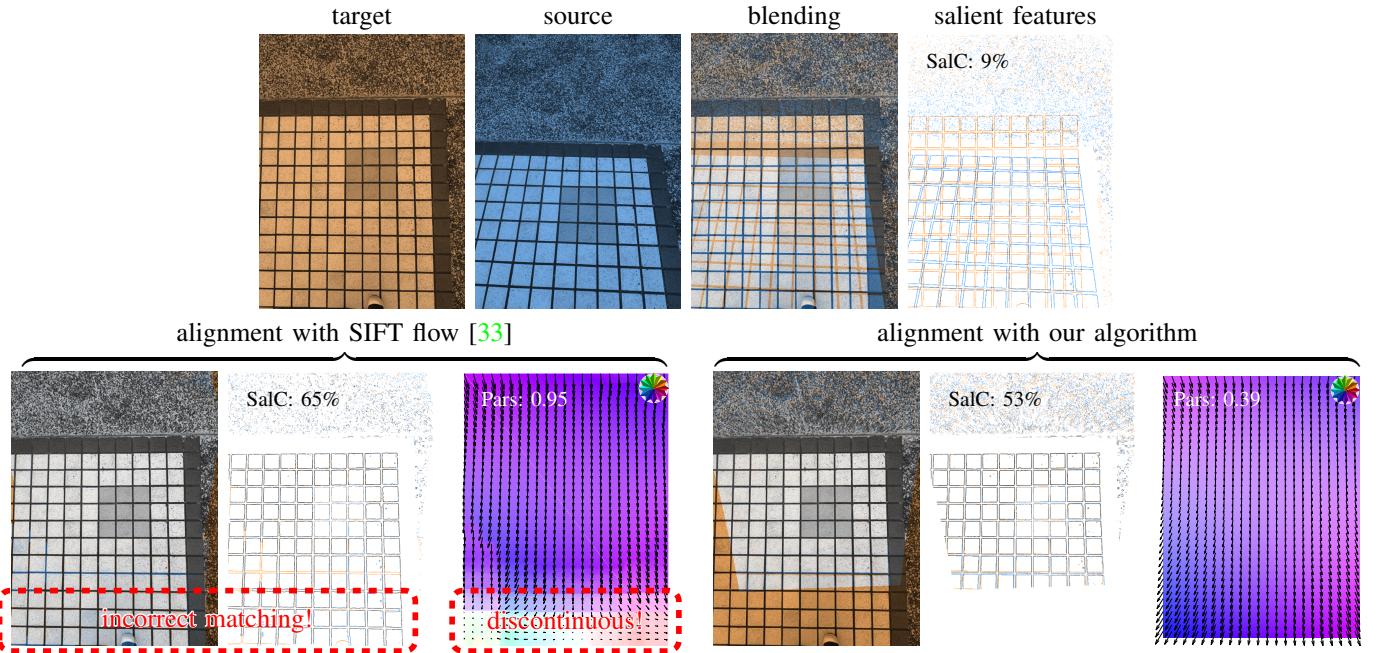


Fig. 6. Visualization of the two registration quality measures proposed in this paper (Section IV): the salience correlation (SalC) quantifies how well salient features of the target and (aligned) source images match—perfect alignment is 100%; the parsimony (Pars) quantifies the “simplicity” of the displacement field—the smaller, the simpler. Notice that, here, although the SIFT flow algorithm yields a seemingly better alignment than our algorithm (larger SalC), this alignment is incorrect (the overlap region found is too large: only 8 rows of tiles are common to the original images, not 11) and this is betrayed by a higher complexity of the displacement field—the parsimony.

V. EXPERIMENTS AND PERFORMANCE EVALUATION

A. Experimental Setting

In this section, extensive experiments are performed on both synthetic and real images with the method that we are proposing. We first demonstrate the accuracy and robustness of our registration algorithm on synthetic tests. Then, on real images, we compare our approach with several state-of-the-art registration methods. We consider three global parametric methods including: a method using enhanced correlation coefficient maximization (ECC) to deal with homography transformations [20]; a parametric method optimizing normalized total gradient (NTG) [81]; and a method using a smoothly varying field to warp and orient features between two images (LAFF) [82]. We also consider seven well-known elastic image

registration methods including: the multiscale LAP algorithm [58]; an intensity-based image registration using residual complexity minimization from the Medical Image Registration Toolbox (MIRT) [83]; the improved Demons algorithm based on the implementation in [84]; an elastic registration using a cubic B-spline free form displacement model implemented in ImageJ (bUnwarpJ) [13]; an algorithm incorporating both geometric and intensity transformation (GIT) [15]; a feature-based method that establishes the correspondence between the SIFT features from each image (SIFT flow [33]); and the implementation freely available in the Elastix toolkit that optimizing mutual information [85]. Furthermore, a deep learning-based optical flow estimation method (SelFlow) [50] is considered. In this paper, we use the public implementation

of these methods. Out of fairness for these methods, not only do we run the algorithms under default configuration, but we also check whether tuning some of their parameters (e.g., number of iterations, window size/resolution depth) results in significant accuracy improvements—and if so, show these results instead.

As shown in Fig. 3, we apply repeatedly the LAP-fitting algorithm three times at each resolution, before refining the resolution of the LAP by a factor 2—and then repeat. The coarsest resolution at which the LAP is applied is determined according to the size of images as $\min(\text{height}, \text{width})/4$. In the parameter fitting part, the displacement field is modelled as a 2D quadratic polynomial (5), whereas the intensity changes, are modelled either as the 2D quadratic polynomial (8), as a sum of three Gaussians (10), or as a one-to-one mapping (typically, using histogram matching, see Section III-B3), depending on the experiment. All algorithms are run on an Intel Core i7-5930K CPU @ 3.50 GHz with 64 GB RAM, using MATLAB R2016b.

B. Monomodal registration

1) *Synthetic experiments*: We generate five random displacement fields, that follow a homography representation $u(x, y) = (\alpha x + \beta y + \gamma)/(ax + by + c)$ chosen from the Oxford dataset [55], and change the intensity of the images under three types of conditions, to demonstrate the robustness of our registration method (see Fig. 7):

- Noise: we add different Gaussian random noise realizations (PSNR = 20 dB) to both the source and the target image;
- Blur: the target image is a Gaussian filtered version (width 1.5 pixels) of the warped source image;
- Intensity change: We generate randomly, slowly changing intensity maps based on the \cos^4 law of vignetting [86] on the target image.

We use the blur model (10) or the intensity model (8) depending on the condition tested.

For these synthetic tests, we have access to the ground-truth deformation. Accordingly we are able to assess the results using direct metrics—the median absolute error (E_{Med}) and the mean absolute error (E_{Mean}) between the estimated displacement field and the ground-truth—and our reference-free metrics—parsimony and salience correlation. It is obvious from the results shown in Table I that, our method is highly accurate and robust to the intensity changes we investigated: the E_{Med} and the E_{Mean} values are of the order of a few percents of a pixel, and can be even smaller in the absence of intensity distortion. Actually, we have observed that even in a very low PSNR setting (e.g., 1 dB), both E_{Med} and E_{Mean} are still less than 1 pixel. Furthermore, the table highlights that this performance is obtained with minimal computation time. In the case of a mixture of noise, blur and intensity distortion, the $E_{\text{Med}}/E_{\text{Mean}}$ values calculated for our algorithm worsens to 0.10/0.14 pixels, still significantly more accurate than for other algorithms.

Finally, the table also highlights the correspondence between small error values and our reference-free metrics proposed in Section IV. Note that the SelFlow method [50] has

been trained with discontinuous displacement fields between target and source images (optical flow). Hence, its larger parsimony values are expected.

2) *Real experiments*: To demonstrate the practicality of the proposed algorithm, we test our method on different types of real images and compare with other state-of-the-art algorithms.

a) *Oxford affine dataset (Mikolajczyk et al. [55])*: This dataset is made of subsets of 6 images, each of which includes accurate estimations of the best homographic transformation between the first image and the other five. We compute the alignment accuracy of various registration algorithms by comparing with this “ground-truth” homography. Note that we considered only three subsets of this dataset, involving blur and illumination changes: the other subsets (apart from “UBC”, which involves intensity distortions) involve displacements that are too large (1/3 to 1/2 of the image size) for any of the registration algorithms tested, including ours, to be successful. Also, to enable some of the comparison algorithms to cope with the large illumination changes in “Leuven” we pre-processed the images by matching the histogram of the source image to the target image. We have checked that this pre-processing is neutral to the other algorithms (no-improvement, no degradation). For the “Bikes” and “Trees” subsets, we model the blur as a sum of three Gaussians (10), and for the “Leuven”, we model the intensity change by the 2D quadratic polynomial models (8).

Table II shows the average performance of all the algorithms tested on these three subsets. A typical alignment result for each subset with the state-of-the-art algorithms can be found in the supplementary material. As can be seen from the results, our algorithm is significantly faster and more accurate than any of the other state-of-the-art algorithms (both E_{Med} and E_{Mean}). In addition, the parsimony of the displacement field found by our algorithm is also smaller or equivalent (i.e., “less complex”) than that found by other algorithms. The salience correlation (SalC) is also larger than or equivalent to that of other algorithms—except for the “Trees” subset. In that subset, the images are characterized by a large depth of field and the change of viewpoint invalidates the warping relation (2): several displacement fields (for different depths) are needed to describe the intensity relation between source and target. As a result, even the “ground-truth” alignment has a very low average SalC (28%), significantly smaller than the SalC of the GIT algorithm (43%)—its large average parsimony (1.89) suggests that the displacements found are incorrect.

b) *Mobile phone images*: We have captured several image pairs using a smartphone camera under various conditions: at different moments of the day, from different viewpoints, and with different types of distortion. Then, we compare the registration algorithms considered here regarding their SalC, Pars, and visualise how well the warped source images align onto the target images. An extract of the results is shown from Fig. 8 to Fig. 10. More extensive comparisons are shown in the supplementary material (also see http://www.ee.cuhk.edu.hk/~xxzhang/welcome_files/TIP_exp.html).

Figure 8 is an example of large scaling and rotation for which most of the other methods are failing (no intensity fitting needed). Fig. 9 demonstrates the performance of our algorithm

TABLE I
EVALUATION COMPARISON FOR THE SYNTHETIC IMAGE REGISTRATION

	Source image	Noiseless Images				Noisy Images (20 dB)				Gaussian Blurry Images				Intensity Distortion (vignetting)							
		E _{Med}	E _{Mean}	Pars	SalC	Time (s)	E _{Med}	E _{Mean}	Pars	SalC	Time (s)	E _{Med}	E _{Mean}	Pars	SalC	Time (s)	E _{Med}	E _{Mean}	Pars	SalC	Time (s)
Global algorithms	Ours	0.002	0.003	0.70	93%	7.7	0.03	0.03	0.70	63%	6.4	0.01	0.01	0.70	57%	13.0	0.02	0.02	0.70	86%	8.4
	ECC [20]	0.30	0.40	0.70	83%	10.2	0.32	0.43	0.70	59%	10.0	0.32	0.41	0.70	54%	10.0	0.29	0.41	0.70	79%	10.2
	NTG [81]	2.30	2.68	0.70	44%	24.3	2.29	2.65	0.70	35%	23.9	2.39	2.66	0.70	38%	24.6	2.35	2.72	0.69	43%	23.8
	LAFFP [82]	0.32	0.48	0.69	78%	2092.7	0.34	0.51	0.69	57%	1693.5	0.32	0.50	0.70	54%	3397.0	0.47	0.69	0.72	72%	1928.0
Elastic algorithms	LAP [58]	0.01	2.45	1.05	90%	7.3	0.60	1.68	1.77	63%	4.9	0.41	1.37	1.62	56%	5.5	0.95	5.45	1.83	72%	10.3
	Demons [84]	0.45	4.71	1.39	73%	14.7	0.73	4.60	1.44	50%	12.9	0.42	3.77	1.37	50%	29.1	5.04	8.87	1.55	49%	14.8
	MIRT [83]	24.63	25.95	4.36	40%	91.1	26.14	27.25	4.96	25%	63.6	22.41	24.90	4.93	30%	72.7	24.67	27.18	4.83	36%	84.2
	bUnwarpJ [13]	1.65	1.67	0.65	53%	103.9	1.80	1.82	0.70	37%	15.5	1.79	1.82	0.71	42%	12.5	1.79	2.72	0.70	45%	14.9
	GIT [15]	0.06	0.11	1.02	89%	959.7	0.53	0.77	1.29	61%	934.6	0.39	0.53	1.22	55%	960.6	0.07	0.20	0.97	85%	942.3
	SIFT flow [33]	0.46	2.54	2.73	77%	49.5	0.74	4.92	2.35	55%	49.5	0.57	2.87	1.90	53%	49.4	0.48	3.50	1.90	74%	49.0
	Elastix [85]	1.86	7.37	1.84	52%	310.0	2.11	7.24	1.44	35%	317.9	2.01	6.80	1.57	39%	320.5	2.03	7.89	1.51	44%	316.4
	SelfFlow [50]	0.47	0.53	0.88	74%	4.2	0.73	1.06	0.86	52%	4.7	0.56	2.20	1.24	51%	4.9	0.50	0.55	0.94	71%	4.5

(1) Bold values indicate the best results. (2) PSNR between the noisy image and original noiseless image is 20 dB. (3) The largest displacement is 80 pixels. (4) The results are calculated in the overlap region and averaged over 5 different displacement fields (homographies). (5) The size of the images is 514×768.

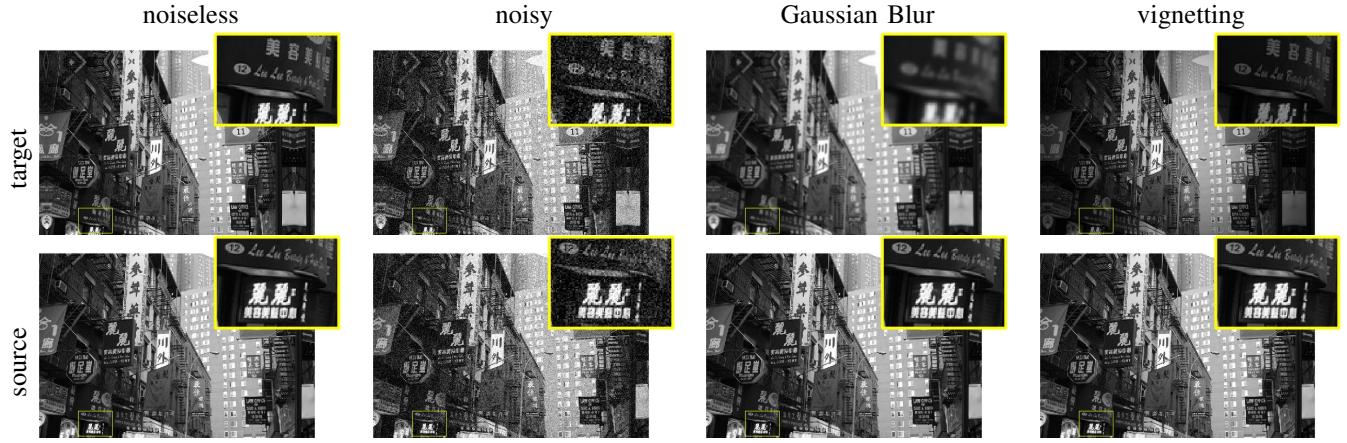


Fig. 7. Typical images and distortions involved in the synthetic experiments summarized by Table I.

when the target image is out of focus. In that example, the intensity changes caused by blurring are parametrized using a sum of three Gaussians (10). In Fig. 10, the target image is taken in the evening (lighting) while the source image is taken during daytime, and its accurate alignment illustrates the robustness of our algorithm in the case of significant intensity changes—parametrized according to (8).

As a general observation, our algorithm always perform the best in terms of visualization, as confirmed by the quantitative alignment metrics that we are proposing (SalC, Pars). Moreover, our algorithm is usually faster than the other algorithms—with the exception of Selfflow. Misalignment in elastic methods is typically characterized by edge and shape distortions, whereas it is characterized by edge and shape displacements in global methods. Misalignment in our algorithm may also, in principle, be affected by these two types of artifacts; however, as demonstrated in these figures, our algorithm is significantly more robust than others to large displacements, significant intensity variations and blurring. We should also point out that the high accuracy and robustness of our algorithm is not obtained at the cost of computational efficiency; on the contrary, despite being implemented in an interpreted language (Matlab, without mex files), our program is frequently as fast or faster than most of the compared

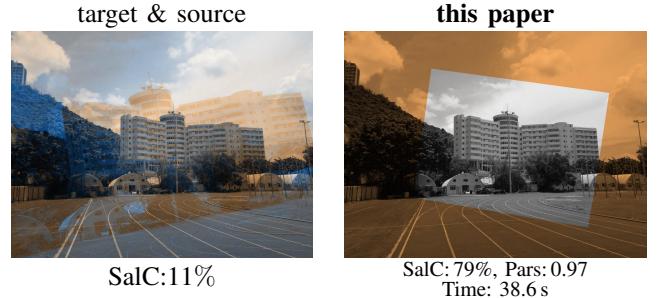


Fig. 8. Registration of real images (960 × 1280 pixels) that have undergone large rotation and scaling, using different algorithms.

algorithms.

C. Multimodal registration

Figure 11 shows a typical multimodal registration in remote sensing: the target image is the near-infrared channel and the source image is the green channel from a multi-spectral dataset (<https://www.sensefly.com/education/datasets/>). We show only the features of these images for a visual evaluation of their alignment: the full images have intensities that are too different, which makes comparisons difficult. In this context, we find that the salient features of the images are accurately aligned when the SalC is larger than

TABLE II
EVALUATION COMPARISON ON THE OXFORD AFFINE DATASET [55]

		Bikes (Blur) (700×1000)					Trees (Blur) (700×1000)					Leuven (Illumination changes) (600×900)				
		E _{Med}	E _{Mean}	Pars	SalC	Time (s)	E _{Med}	E _{Mean}	Pars	SalC	Time (s)	E _{Med}	E _{Mean}	Pars	SalC	Time (s)
	Ground truth	—	—	0.17	52%	—	—	—	0.70	29%	—	—	—	0.17	64%	—
Global algorithms	This paper	0.42	0.53	0.18	53%	10.0	1.36	1.68	0.58	33%	8.9	0.19	0.26	0.15	67%	7.8
	ECC [20]	0.47	0.60	0.18	52%	17.8	2.05	3.32	0.60	28%	14.5	0.21	0.27	0.15	66%	13.0
	NTG [81]	0.87	0.98	0.22	52%	15.5	1.71	1.97	0.58	28%	15.6	0.62	0.89	0.11	61%	32.9
	LAFP [82]	1.79	2.58	0.23	42%	1933.4	4.31	5.24	0.64	26%	2330.9	1.14	1.57	0.29	51%	1446.2
Elastic algorithms	LAP [58]	0.63	0.78	0.21	52%	11.3	1.69	3.22	0.77	35%	9.4	0.33	0.43	0.28	67%	7.9
	Demons [84]	21.64	22.10	1.27	21%	27.4	3.92	7.28	1.28	28%	12.7	0.56	1.58	2.14	61%	25.8
	MIRT [83]	19.14	24.65	3.08	44%	206.9	25.95	26.99	4.52	24%	76.1	2.95	5.34	3.45	59%	106.8
	bUnwarpJ [13]	1.02	1.12	0.22	51%	18.0	2.15	3.00	0.93	30%	24.3	0.34	0.40	0.67	66%	24.7
	GIT [15]	1.38	3.84	1.32	52%	982.9	2.18	4.66	1.89	43%	962.0	0.48	0.96	2.13	64%	948.7
	SIFT flow [33]	0.85	1.94	0.71	52%	84.7	2.50	5.70	3.08	40%	86.2	0.55	0.80	0.92	64%	66.3
	Elastix [85]	6.69	12.38	2.44	36%	299.9	3.78	10.17	2.44	28%	319.5	0.44	1.46	0.85	66%	307.0
	SelFlow [50]	1.08	2.81	0.64	49%	4.3	2.39	7.26	1.41	32%	4.7	0.33	0.53	0.44	64%	4.2

(1) Bold values indicate the best results. (2) Alignment accuracy is computed within the overlap region and averaged over the 5 different image pairs in each subset.

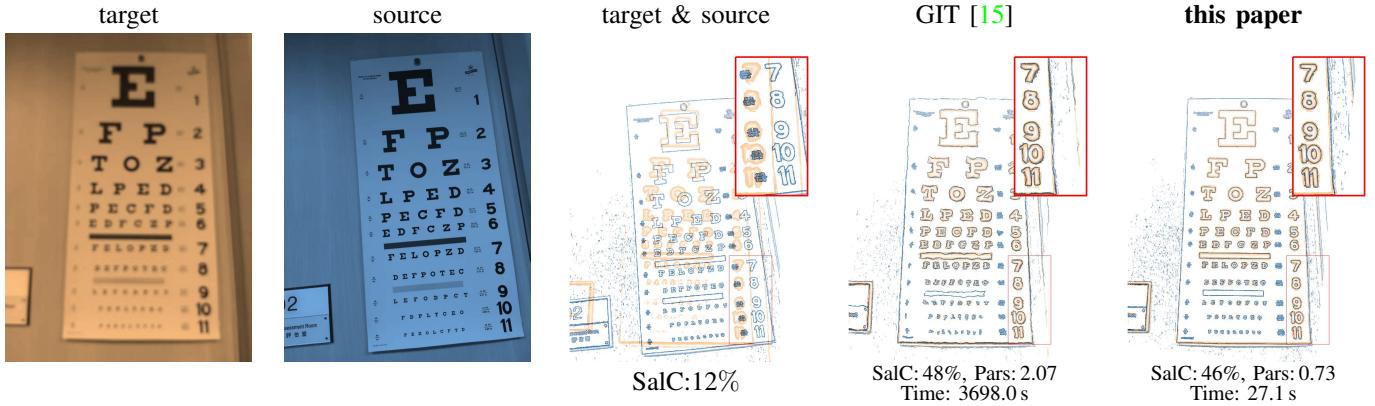


Fig. 9. Registration of real images (1440 × 1080 pixels) corrupted by blurring, using different algorithms. In order to better visualise the misalignments, we outline them on the feature images. The second row shows the pixel intensity at the 10% largest gradient locations.

30%. Most algorithms are unable to reach this threshold, as confirmed by the lack of overlap between the edges of the target (orange) and aligned-source (blue) images. Those who can, achieve this at the cost of a significantly more complex displacement field than our algorithm (Pars = 0.13).

We should point out here that, if we attempt to predict the intensity changes between the source and target images using a parametric expression like (8) only, our algorithm would not be successful. Here, it is essential to use histogram matching (i.e., find a one-to-one intensity mapping) between the two aligned images.

VI. DISCUSSION AND CONCLUSION

We have proposed a novel methodology for image registration that works by repeatedly fitting the dense displacement field obtained from an elastic registration algorithm (LAP), and fitting the intensity between source and target images, according to geometric and intensity parametric models. This methodology essentially requires a choice of bases to represent the displacement field and the intensity changes. Here, we have exemplified our approach with quadratic 2D polynomials, which seem to model quite well the geometric distortions involved in digital cameras. Apart from that, we have empirically

determined that a coarse to fine iterative process whereby the window size involved in the LAP algorithm is reduced by half every three iterations is sufficient to guarantee accurate results consistently.

We have demonstrated the high accuracy, robustness and computational efficiency of this methodology by exemplifying it with a specific 12-parameter quadratic displacement model. Specifically, we showed that such a low order model is sufficient to deal with arbitrary real images taken with a mobile camera, and that the alignment found in various conditions is of higher quality than any of the other state-of-the-art algorithms tested (both parametric and elastic registration algorithms).

There are several reasons why our algorithm works so well: 1. the LAP registration algorithm is very fast, and at the same time very accurate when the intensity of the source and target images match; 2. the (low-order) parametric representation of the displacement ensures that accurate fitting can be achieved even when the overlap region between the target and source images is very limited, which allows to deal with occlusions for instance, and more generally, makes the estimation robust to model mismatch; 3. the intensity prediction part allows to deal with significant intensity changes between the source

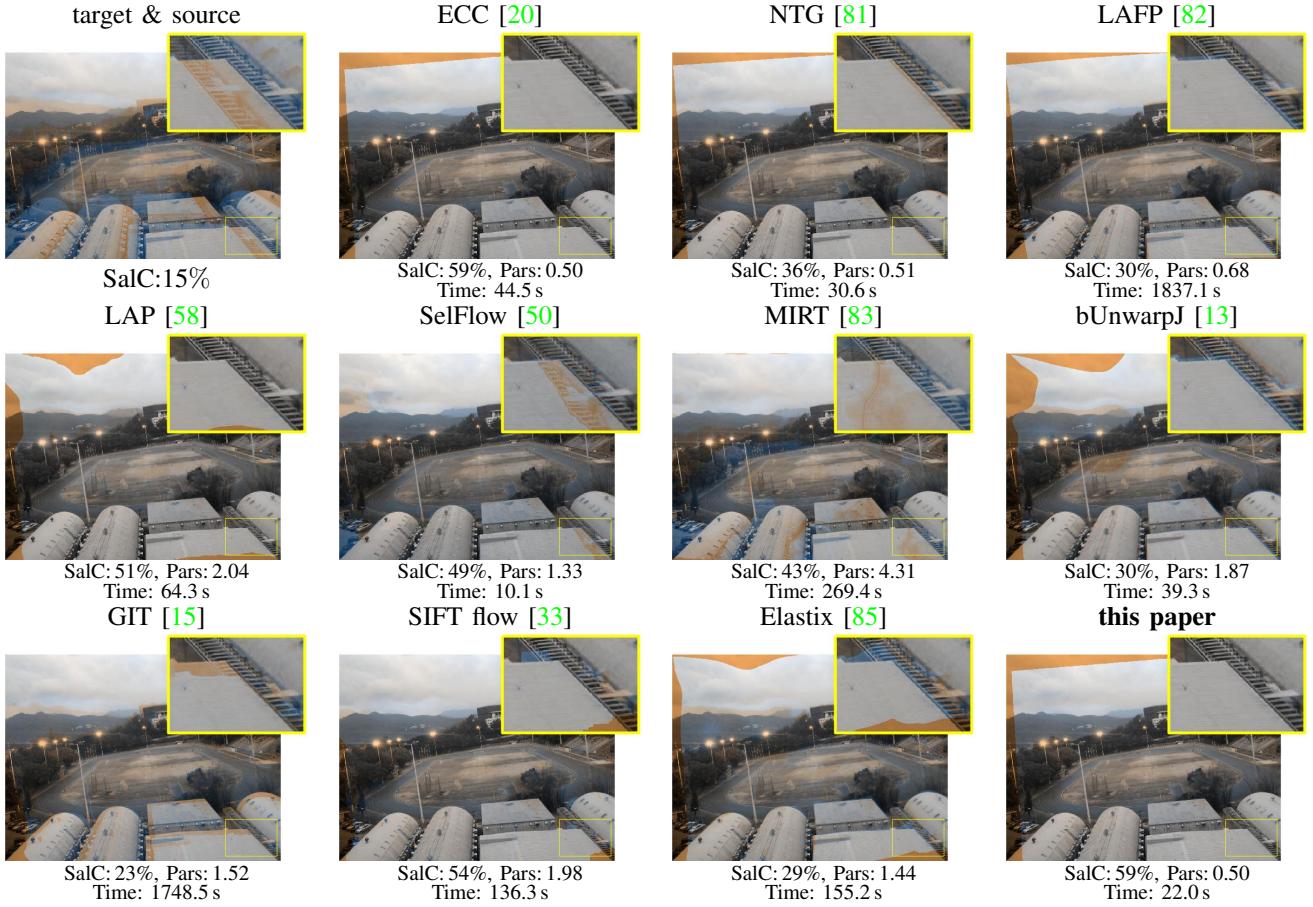


Fig. 10. Registration of real images (900×1200 pixels) that have undergone significant intensity changes (evening vs daylight, changing cloud cover), using different algorithms. Our intensity fitting method uses the model (8). Note that, because of the camera automatic exposure, the two images have similar brightness (on the evening photo, the stadium projectors are turned on).

and target images (e.g., blur, day/night illumination changes, even non-parametric deterministic changes like the ones encountered in multimodal image acquisition). In practice, our algorithm is more likely to fail when the displacement field is so large that the LAP algorithm is not able to estimate it anywhere in the image (e.g., larger than half the size of the image) as we have seen in some subsets of the Oxford dataset.

In order to evaluate the quality of the displacement field retrieved by registration algorithms, we have proposed two “no-reference” metrics: 1. the salience correlation (SalC) which, by focusing on the most salient features of the images, is a sharpened form of signal to noise ratio; 2. the parsimony (Pars), which quantifies the complexity of the displacement field. We showed that only one measure (typically, on the matching of intensities) is insufficient to characterize the quality of a registration algorithm; we believe that our emphasis on the simplicity of the displacement field—low parsimony—is a novel criterion that should be taken into account in the evaluation of registration algorithms, in particular when the ground-truth is unknown.

Our methodology for image registration can easily be applied to more complex models like: higher degree polynomials, Fourier series representation, local spline or wavelet representations. Such an extension would allow our algorithm to deal with more complex, possibly local, geometric transformations

although we expect that this may be more computationally expensive. Too many parameters may also lead to less realistic alignments: to see if this is the case, we have considered a Fourier series parametrisation of the displacement field, for both synthetic (exact brightness consistency) and real images (inconsistent brightness). As shown in Fig. 12, with more parameters the salient features are increasingly better aligned (saturation of SalC at a high value). When brightness consistency is satisfied, Pars saturates at a low value, that characterises the simplicity of the homography used to distort the synthetic images: this demonstrates a form of robustness of our algorithm with respect to over-parametrisation. For real images, where brightness is not fully consistent, we observe a regular increase of the parsimony, even though the SalC saturates at about 80%: minimising the parsimony provided that SalC is large enough (we suggest $\text{SalC} \geq 50\%$) makes the choice that the most realistic solution is the simplest one.

Alternatively, the parametric registration algorithm developed in the current paper may also be completed by the PF-LAP [58] to achieve more complex displacement fields than the quadratic polynomial exemplified here—leveraging on a good estimation of intensity changes between source and target images.

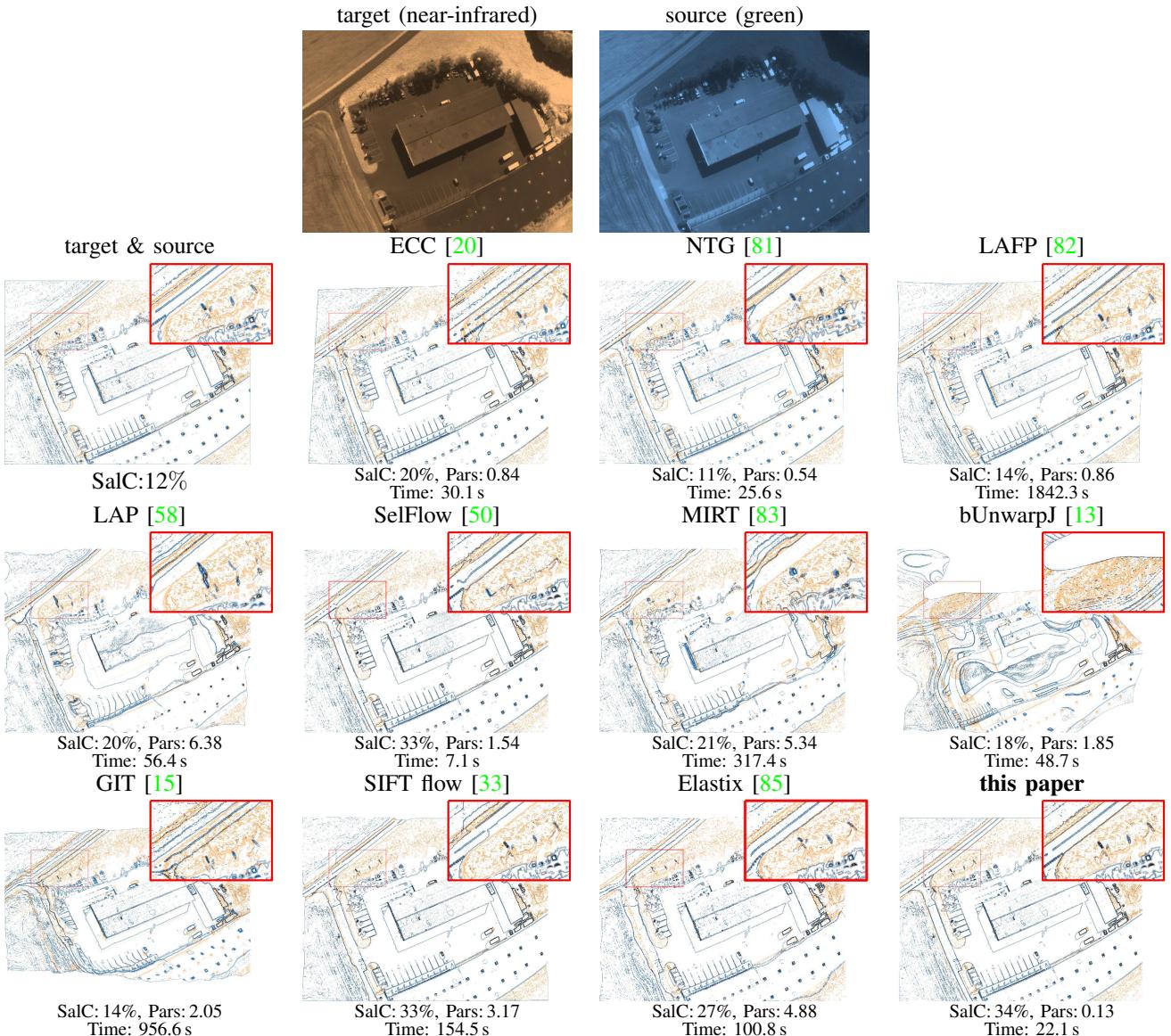


Fig. 11. Registration of multimodal satellite images (960×1280 pixels): overlay of the salient features (lower three rows) resulting from different algorithms. The target and source images (top row) are the near-infrared and green channel of a multispectral image taken from the dataset (<https://www.sensefly.com/education/datasets/>). Our intensity fitting method is histogram matching in this case.

ACKNOWLEDGMENT

This work was supported by a grant #CUHK14207718 of the Hong Kong Research Grants Council. The authors would like to thank Profs. Pierre Moulin and Jan Kybic for useful discussions. We also thank the anonymous reviewers for their comments, which have improved the quality of this paper.

REFERENCES

- [1] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, pp. 1153–1190, 2013.
- [2] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.
- [3] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey," in *International Conference on Image and Signal Processing (ICSIP)*, 2010, pp. 103–112.
- [4] X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and multi-spectral registration for natural images," in *European Conference on Computer Vision*, 2014, pp. 309–324.
- [5] J. P. Pluim, J. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.
- [6] A. Andronache, M. von Siebenthal, G. Szekely, and P. Cattin, "Non-rigid registration of multi-modal images using both mutual information and cross-correlation," *Medical Image Analysis*, vol. 12, no. 1, pp. 3–15, 2008.
- [7] S. Baker, D. Scharstein, and J. P. L. et al., "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, pp. 1–31, 2011.
- [8] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 1, pp. 185–203, 1981.
- [9] R. Woods, S. Cherry, and J. Mazziotta, "Rapid automated algorithm for aligning and reslicing PET images," *Journal of Computer Assisted Tomography*, vol. 16, pp. 620–633, 1992.
- [10] R. Woods, S. Grafton, and C. H. et al., "Automated image registration: I. General methods and intrasubject, intramodality validation," *Journal of Computer Assisted Tomography*, vol. 22, pp. 141–154, 1998.
- [11] F. Albu, "Low complexity image registration techniques based on integral projections," in *International Conference on Systems, Signals and Image Processing (IWSIP)*, 2016, pp. 1–4.
- [12] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Robust and efficient parametric face alignment," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 1847–1854.
- [13] I. Arganda-Carreras, C. O. S. Sorzano, and R. M. et al., "Consistent and elastic registration of histological sections using vector-spline regular-

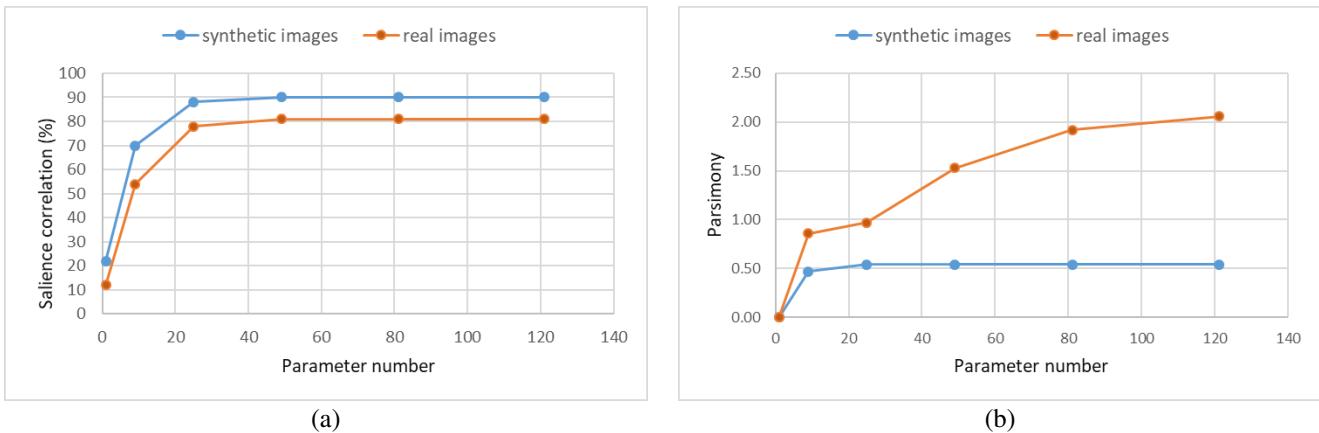


Fig. 12. Increasing the number of Fourier coefficients in the representation of the displacement field, increases the salience correlation (a) and the parsimony (b). The SalC, however, reaches a value larger than 50% quickly and saturates after about 20 parameters. For synthetic images the parsimony saturates quickly at a small value, whereas for real images, it seems to increase without limit, but without significant SalC increase. Images used: same as Fig. 7 (noiseless) and Fig. 8.

- ization,” in *International Workshop on Computer Vision Approaches to Medical Image Analysis (CVAMIA)*, 2006, pp. 85–95.
- [14] R. Bajcsy and S. Kovacic, “Multiresolution elastic matching,” *Computer Vision, Graphics and Image Processing*, vol. 46, pp. 1–21, 1989.
- [15] S. Periaswamy and H. Farid, “Elastic registration in the presence of intensity variations,” *IEEE Transactions on Medical Imaging*, vol. 22, pp. 865–874, 2003.
- [16] S. Klein, M. Staring, and K. M. et al., “Elastix: a toolbox for intensity-based medical image registration,” *IEEE Transactions on Medical Imaging*, vol. 29, pp. 196–205, 2010.
- [17] S. Periaswamy and H. Farid, “Medical image registration with partial data,” *Medical Image Analysis*, vol. 10, pp. 452–464, 2006.
- [18] A. Goshtasby, “Image registration by local approximation methods,” *Image and Vision Computing*, vol. 6, pp. 255–261, 1988.
- [19] J. Kybic and M. Unser, “Fast parametric elastic image registration,” *IEEE Transactions on Image Processing*, vol. 12, pp. 1427–1442, 2003.
- [20] G. D. Evangelidis and E. Z. Psarakis, “Parametric image alignment using enhanced correlation coefficient maximization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1858–1865, 2008.
- [21] A. Roche, A. Guimond, N. Ayache, and J. Meunier, “Multimodal elastic matching of brain images,” *Computer Vision-ECCV 2000*, pp. 511–527, 2000.
- [22] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
- [23] P. Viola and W. M. Wells III, “Alignment by maximization of mutual information,” *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [24] P. Thévenaz and M. Unser, “Optimization of mutual information for multiresolution image registration,” *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2083–2099, 2000.
- [25] J. P. Pluim, J. A. Maintz, and M. A. Viergever, “Interpolation artefacts in mutual information-based image registration,” *Computer vision and image understanding*, vol. 77, no. 2, pp. 211–232, 2000.
- [26] H. S. Alhichri, and M. Kamel, “Virtual circles: a new set of features for fast image registration,” *Pattern Recognition Letters*, vol. 24, pp. 1181–1190, 2003.
- [27] C. Davatzikos, J. Prince, and R. Bryan, “Image registration based on boundary mapping,” *IEEE Transactions on Medical Imaging*, vol. 15, pp. 112–115, 1996.
- [28] Q. Li, G. Wang, and J. L. et al., “Robust scale-invariant feature matching for remote sensing image registration,” *IEEE Geoscience and Remote Sensing Letters*, vol. 6, pp. 287–291, 2009.
- [29] J. Ma, W. Qiu, and J. Z. et al., “Robust L_2 E estimation of transformation for non-rigid registration,” *IEEE Transactions on Signal Processing*, vol. 63, pp. 1115–1129, 2015.
- [30] A. Kelman, M. Sofka, and C. V. Stewart, “Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–7.
- [31] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, “A partial intensity invariant feature descriptor for multimodal retinal image registration,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1707–1718, 2010.
- [32] G. Wang, Z. Wang, Y. Chen, and W. Zhao, “Robust point matching method for multimodal retinal image registration,” *Biomedical Signal Processing and Control*, vol. 19, pp. 68–76, 2015.
- [33] C. Liu, J. Yuen, and A. Torralba, “SIFT flow: Dense correspondence across scenes and its applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011.
- [34] D. G. Lowe, “Object recognition from local scale-invariant features,” in *iccv*. Ieee, 1999, p. 1150.
- [35] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded up robust features,” in *European Conference on Computer Vision*, 2006, pp. 404–417.
- [36] Y. Ke and R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2004, pp. II–II.
- [37] E. N. Mortensen, H. Deng, and L. Shapiro, “A SIFT descriptor with global context,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 184–190.
- [38] A. E. Abdel-Hakim and A. A. Farag, “CSIFT: A SIFT descriptor with color invariant characteristics,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp. 1978–1983.
- [39] D. Sargent, C.-I. Chen, C.-M. Tsai, Y.-F. Wang, and D. Koppel, “Feature detector and descriptor for medical images,” in *Medical Imaging 2009: Image Processing*, vol. 7259, 2009, pp. 72 592Z-1–8.
- [40] Z. Shen, X. Han, Z. Xu, and M. Niethammer, “Networks for joint affine and non-parametric image registration,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4224–4233.
- [41] B. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Işgum, “A deep learning framework for unsupervised affine and deformable image registration,” *Medical image analysis*, vol. 52, pp. 128–143, 2019.
- [42] G. Haskins, U. Kruger, and P. Yan, “Deep learning in medical image registration: A survey,” *arXiv preprint arXiv:1903.02026*, 2019.
- [43] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, “Unsupervised learning for fast probabilistic diffeomorphic registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 729–738.
- [44] H. Li and Y. Fan, “Non-rigid image registration using fully convolutional networks with deep self-supervision,” *arXiv preprint arXiv:1709.00799*, 2017.
- [45] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Işgum, “End-to-end unsupervised deformable image registration with a convolutional neural network,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 204–212.
- [46] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, “Quicksilver: Fast predictive image registration—a deep learning approach,” *NeuroImage*, vol. 158, pp. 378–396, 2017.

- [47] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, "Svnet: Learning deformable image registration using shape matching," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 266–274.
- [48] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, "Robust non-rigid registration through agent-based action learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 344–352.
- [49] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "Deepflow: Large displacement optical flow with deep matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1385–1392.
- [50] P. Liu, M. Lyu, I. King, and J. Xu, "Selfflow: Self-supervised learning of optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4571–4580.
- [51] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2758–2766.
- [52] T.-W. Hui, X. Tang, and C. Change Loy, "Liteflownet: A lightweight convolutional neural network for optical flow estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8981–8989.
- [53] Y. Gavet, M. Fernandes, and J.-C. Pinoli, "Quantitative evaluation of image registration techniques in the case of retinal images," *Journal of Electronic Imaging*, vol. 21, no. 2, pp. 021118–1–7, 2012.
- [54] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, and A. A. Argyros, "FIRE: Fundus image registration dataset," *Journal for Modeling in Ophthalmology*, vol. 1, no. 4, pp. 16–28, 2017.
- [55] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.
- [56] M. M. Fouad, R. M. Dansereau, and A. D. Whitehead, "Image registration under illumination variations using region-based confidence weighted m -estimators," *IEEE Transactions on image processing*, vol. 21, no. 3, pp. 1046–1060, 2012.
- [57] W. R. Crum, T. Hartkens, and D. Hill, "Non-rigid image registration: theory and practice," *The British journal of radiology*, vol. 77, no. suppl_2, pp. S140–S153, 2004.
- [58] C. Gilliam and T. Blu, "Local all-pass geometric deformations," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 1010–1025, 2018.
- [59] X. Zhang, C. Gilliam, and T. Blu, "Iterative fitting after elastic registration: An efficient strategy for accurate estimation of parametric deformations," in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP'17)*, September 17–20, 2017, pp. 1492–1496.
- [60] C. Gilliam and T. Blu, "Local all-pass filters for optical flow estimation," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1533–1537.
- [61] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Non-parametric diffeomorphic image registration with the demons algorithm," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2007, pp. 319–326.
- [62] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, pp. 185–203, 1981.
- [63] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Transactions on image processing*, vol. 12, no. 11, pp. 1427–1442, 2003.
- [64] Y.-j. Wang, G. Farnebäck, and C.-F. Westin, "Multi-affine registration using local polynomial expansion," *Journal of Zhejiang University SCIENCE C*, vol. 11, no. 7, pp. 495–503, 2010.
- [65] G. Farnebäck and C.-F. Westin, "Affine and deformable registration based on polynomial expansion," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 857–864.
- [66] A. Agarwal, C. Jawahar, and P. Narayanan, "A survey of planar homography estimation techniques," *Centre for Visual Information Technology, Tech. Rep. IIIT/TR/2005/12*, 2005.
- [67] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 1. IEEE, 1999, pp. 125–131.
- [68] J. Nemeth, C. Domokos, and Z. Kato, "Recovering planar homographies between 2d shapes," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2170–2176.
- [69] T. Blu, P. Thévenaz, and M. Unser, "Linear interpolation revitalized," *IEEE Transactions on Image Processing*, vol. 13, pp. 710–719, 2004.
- [70] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.
- [71] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-order interpolation of minimal support," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1069–1080, 2001.
- [72] X. Zhang, R. Wang, X. Jiang, W. Wang, and W. Gao, "Spatially variant defocus blur map estimation and deblurring from a single image," *Journal of Visual Communication and Image Representation*, vol. 35, pp. 257–264, 2016.
- [73] C.-T. Shen, W.-L. Hwang, and S.-C. Pei, "Spatially-varying out-of-focus image deblurring with L1-2 optimization and a guided blur map," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1069–1072.
- [74] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Prentice Hall, 2008.
- [75] B. E. Pires and P. M. Aguiar, "Registration of images with small overlap," in *6th IEEE Workshop on Multimedia Signal Processing*, 2004, pp. 255–258.
- [76] K. Wilkie and E. R. Vrscay, "Mutual information-based methods to improve local region-of-interest image registration," in *International Conference Image Analysis and Recognition*, 2005, pp. 63–72.
- [77] G. Roth, "Registering two overlapping range images," in *Second International Conference on 3-D Digital Imaging and Modeling*, 1999, pp. 191–200.
- [78] H.-R. Su and S.-H. Lai, "Non-rigid registration of images with geometric and photometric deformation by using local affine Fourier-moment matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2874–2882.
- [79] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Transactions on Computers*, vol. 100, no. 2, pp. 179–186, 1972.
- [80] E. Sober, *Ockham's razors*. Cambridge University Press, 2015.
- [81] S.-J. Chen, H.-L. Shen, C. Li, and J. H. Xin, "Normalized total gradient: a new measure for multispectral image registration," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1297–1310, 2018.
- [82] Y. Matsushita, "Aligning images in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1–8.
- [83] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 1882–1891, 2010.
- [84] H. Lombaert, L. Grady, and X. P. et al., "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, pp. S61–S72, 2009.
- [85] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.
- [86] S. J. Kim and M. Pollefeys, "Robust radiometric calibration and vignetting correction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 562–576, 2008.



Xinxin Zhang received the M.Sc. degrees from Peking University in computer applied technology in 2016. She is currently pursuing the Ph.D. degree with the Image and Video Processing Group, Department of Electronic Engineering, The Chinese University of Hong Kong. Her research interests include image registration, biomedical and medical imaging, image enhancement and reconstruction, image deblurring, and blur estimation.



Christopher Gilliam received the degree (Hons.) from the Imperial College London, U.K., the MEng. degree in electrical and electronic engineering in 2008, and the Ph.D. degree in electrical and electronic engineering in 2013. He is currently a Research Fellow with the School of Engineering, RMIT University, Australia. He was a Postdoctoral Fellow with the Image and Video Processing Group, Department of Electronic Engineering, The Chinese University of Hong Kong. His research interests include sampling theory and its application in image processing in particular image-based rendering, sparse sampling involving signals with finite rate of innovation, motion estimation and registration for image sequences, in particular biological and medical imaging, and sensor management.



Thierry Blu (M'96-M'06-F'12) was born in Orléans, France, in 1964. He received the Dipl.-Ing degree from the École Polytechnique, France, and Télécom Paris (ENST), Paris, in 1986 and 1988. In 1996, the Ph.D. degree in electrical engineering from ENST in iterated rational filterbanks, and applied to wideband audio coding.

From 1998 and 2007, he was with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland. He is currently a Professor with the Department of Electronic Engineering,

The Chinese University of Hong Kong. His research interests include wavelets, approximation and sampling theory, sparse representations, image denoising, biomedical imaging, optics, and wave propagation.

Dr. Blu was the recipient of two best paper awards from the IEEE Signal Processing Society in 2003 and 2006. He is also the co-author of a paper that received a Young Author Best Paper Award in 2009 from the IEEE Signal Processing Society. He has been a member of the IEEE Signal Processing Theory and Methods Technical Committee from 2008 to 2013, and an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2002 to 2006, THE IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010, and the Elsevier Signal Processing from 2008 to 2011. He has been on the board of Eurasip J. on Image and Video Processing, since 2010, and a member of the IEEE Bio Imaging and Signal Processing Technical Committee.