Geodesic Active Fields – A Geometric Framework for Image Registration

Dominique Zosso*, *Student Member, IEEE,* Xavier Bresson, Jean-Philippe Thiran, *Senior Member, IEEE*

***Abstract*—In this paper we present a novel geometric framework called geodesic active fields for general image registration. In image registration, one looks for the underlying deformation field that best maps one image onto another. This is a classic ill-posed inverse problem, which is usually solved by adding a regularization term. Here, we propose a multiplicative coupling between the registration term and the regularization term, which turns out to be equivalent to embed the deformation field in a weighted minimal surface problem. Then, the deformation field is driven by a minimization flow toward a harmonic map corresponding to the solution of the registration problem. This proposed approach for registration shares close similarities with the well-known geodesic active contours model in image segmentation, where the segmentation term (the edge detector function) is coupled with the regularization term (the length functional) via multiplication as well. As a matter of fact, our proposed geometric model is actually the exact mathematical generalization to vector fields of the weighted length problem for curves and surfaces introduced by Caselles-Kimmel-Sapiro [1]. The energy of the deformation field is measured with the Polyakov energy weighted by a suitable image distance, borrowed from standard registration models. We investigate three different weighting functions, the squared error and the approximated absolute error for monomodal images, and the local joint entropy for multimodal images. As compared to specialized state-of-the-art methods tailored for specific applications, our geometric framework involves important contributions. Firstly, our general formulation for registration works on any parametrizable, smooth and differentiable surface, including non-flat and multi-scale images. In the latter case, multiscale images are registered at all scales simultaneously, and the relations between space and scale are intrinsically being accounted for. Secondly, this method is, to the best of our knowledge, the first re-parametrization invariant registration method introduced in the literature. Thirdly, the multiplicative coupling between the registration term, i.e. local image discrepancy, and the regularization term naturally results in a data-dependent tuning of the regularization strength. Finally, by choosing the metric on the deformation field one can freely interpolate between classic Gaussian and more interesting anisotropic, TV-like regularization.**

***Index Terms*—Biomedical image processing, Computational geometry, Differential geometry, Diffusion equations, Image registration, Scale-spaces, Surfaces.**

I. Introduction

Manuscript received XXXX; revised XXXX. This work is supported by the Swiss National Competence Center in Biomedical Imaging (NCCBI).

D. Zosso and J.-Ph. Thiran are with the Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 11, CH-1015 Lausanne, Switzerland, {dominique.zosso,jp.thiran}@epfl.ch.

X. Bresson is with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, xbresson@cityu.edu.hk

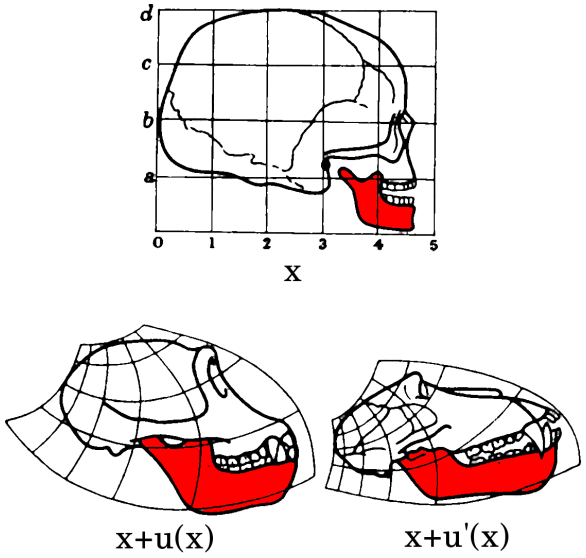


Fig. 1. The skull of a human is registered to chimpanzee and baboon by finding the deformation fields $\mathbf{u}(\mathbf{x})$ and $\mathbf{u}'(\mathbf{x})$, such that human features, e.g. the mandible (shaded), at $\mathbf{x}$ match those of chimpanzee and baboon at $\mathbf{x}+\mathbf{u}(\mathbf{x})$ and $\mathbf{x}+\mathbf{u}'(\mathbf{x})$. Skull sketches reproduced from [2]

IMAGE REGISTRATION is the concept of mapping homologous points of different images, representing a same object. Homology, in turn, is defined as the relation between "organs deriving from the same embryonic blanks"[1]. This fundamental concept is illustrated in Fig. 1.

In practice, however, it is highly difficult to establish homology in images strictly based on this definition. For automatic image registration, it is therefore commonplace to substitute homology by a measurable criterion of image dissimilarity, which is to be minimized. Depending on the nature of the images to be registered, different metrics are used to assess image distances. But also the deformation model and constraints that are applied on that deformation field can vary, as well as the optimization technique that is used to solve the minimization [3]–[6].

Let a deformation field $\mathbf{u}(\mathbf{x}) : \mathbf{x} \in \Omega \subset \mathbb{R}^n \rightarrow (u_1, \ldots, u_p) \in \mathbb{R}^p$ describe the spatial displacement along $p \leq n$ dimensions of an n-dimensional image of support Ω. The determination of this underlying deformation field between two images is an ill-posed inverse problem, requiring additional prior knowledge to make it well-posed. On one hand, parametric deformation models, including rigid and affine transformations, which are defined globally for the

[1] *Homologue:"sont homologues les organes dérivant des mêmes ébauches embryonnaires"*. Le trésor de la langue française informatisé.

whole image space, restrict the degrees of freedom to a small number of parameters. On the other hand, freeform deformations allow for an individual local displacement of each point in the image domain. To restrict such deformation fields to what is believed to be "physically meaningful" deformations, constraints on the field regularity are introduced. Typical regularization constraints reduce the variations of the deformation field $\mathbf{u}(\mathbf{x})$ by defining an additional penalty, e.g., the vectorial total variation functional [7]:

$$\int_\Omega |\nabla \mathbf{u}(\mathbf{x})| \mathrm{d}\mathbf{x} = \int_\Omega \sqrt{\sum_{k=1}^{p} |\nabla u_k|^2} \mathrm{d}\mathbf{x}. \quad (1)$$

Thus, image distance metric and regularization penalty are commonly incorporated into a single energy minimization model, a.k.a. variational model, e.g., [8]. The energy functionals are commonly of the general form

$$E = E_{\text{data}} + r \cdot E_{\text{regularization}}, \quad r > 0. \quad (2)$$

As a typical instance consider an energy functional consisting of the L^2-norm of the difference between the fixed and the moving image, $F, M : \Omega \to \mathbb{R}$, regularized by the aforementioned vectorial total variation:

$$E = \int_\Omega \left(M(\mathbf{x}+\mathbf{u}) - F(\mathbf{x})\right)^2 \mathrm{d}\mathbf{x} + r \cdot \int_\Omega |\nabla \mathbf{u}(\mathbf{x})| \mathrm{d}\mathbf{x}. \quad (3)$$

The balancing parameter r can have a severe impact on the registration result. Its choice is arbitrary and the optimal depends on several conditions.

In their seminal work [9], Sochen, Kimmel and Malladi introduced the powerful Beltrami framework for image denoising and enhancement. This model is based on the Polyakov model [10] introduced in string theory for physics. The Polyakov model represents strings as harmonic maps in high-dimensional and curved spaces defined by Riemannian manifolds. Adopting this pure geometric point of view amounts to seeing objects such as images, shapes, or vector fields as geodesics or harmonic maps. Recently, a new regularization criterion derived from the Beltrami framework was introduced in stereo vision and optical flow modeling [11]–[13]. There, the authors embed the disparity map or the optical flow field, respectively, as harmonic map, and propose to use the Polyakov energy as the regularization term, while keeping the classical additive data terms.

In some registration problems, separate objects (Gestalts) in the images are displaced and deformed independently. This is illustrated by a study on the individual movements of separate parts between slices of histological samples, where regularization has been delimited by explicitly modeled boundaries [14]. Other examples can be found in computer vision, where the optical flow often exhibits piece-wise constant or piece-wise smooth regions, with distinct boundaries [15]. Geometric regularization offers some nice advantages in this respect. The first, flow-driven, TV-like regularizer of [13] intrinsically allows for sharper transitions and isolated regions. Further, there are cases where boundaries in images – in terms of intensities or even texture – are good predictors of deformation field boundaries [16]. In [13], the authors present a second, combined flow-intensity driven regularizer, where image intensity is embedded in the manifold along with the deformation field. Hence, this additional cue increases the geodesic distance between independent homogeneous Gestalts and helps defining sharp deformation field boundaries between them.

Data-dependent regularization has also become important when dealing with outlier pixels. In rigid registration, the influence of mismatching regions can be drastically reduced by cropping the image distance function, e.g., by using Tukey's biweight instead of squared error as an instance of robust statistics [17]. In non-rigid registration, one can estimate a local measure of image data reliability to spatially adapt the strength of regularization [18], while in atlas-based registration this information can equally be derived from atlas statistics.

The goal of this work is to define a novel image registration scheme using a geometric approach. We couple the registration term and the regularization term locally, by multiplication. Hence, we embed the deformation field in a higher dimensional space and define a variational model using the *weighted* Polyakov energy. While the Polyakov energy itself only provides a regularity constraint – harmonic map –, the weighting allows driving the deformation field toward low image dissimilarity. This is in close analogy to geodesic active contours in segmentation [1], where the segmentation term, i.e. the edge detector function, is coupled with the regularizing length function through multiplication as well. Because our model actually represents a mathematical generalization to vector fields of the weighted length problem for curves and surfaces, we call this model *geodesic active fields* (GAF) for image registration.

As will become clearer in the next sections, the GAF framework has several appealing properties. The proposed approach directly generalizes to non-Euclidean images, and thus automatically allows working, e.g., with non-flat or multiscale images. In particular, we will instantiate a model for the simultaneous registration of multiscale images at all scales, where the metric on the deformation field automatically takes care of the specific relations between space and scale. Also, we will show that the geometric GAF energy formulation has the advantage of being invariant with respect to the parametrization of the image domain. To the best of our knowledge, this is the first registration method invariant by re-parametrization. Further, thanks to the Beltrami-like embedding of the deformation field we can benefit of all advantages of geometric regularization, including the freedom to choose the desired degree of anisotropy. Last but not least, the multiplicative link between data and regularization term represents an automatic data-dependent modulation of the local regularization strength by the current alignment quality. The weighting function increases regularization in regions where low matching quality indicates missing confidence, e.g., due to a higher level of noise, whereas lower regularity is required in regions where a good fit can be provided.

The structure of this paper is as follows. The next section will introduce the mathematical tools that build the foundation of the GAF framework. We will recall the Polyakov energy employed in the Beltrami framework and its weighted version

used in GAF, as well as the corresponding minimizing flows. In section III we show how the weighted Polyakov framework can be used to define an abstract geometric image registration model. We then derive from this general image registration model several instances for stereo vision as well as flat and non-flat 2D image registration in section IV. Then, in section V, we instantiate an extension of the framework to multiscale image registration. Section VI studies different weighting functions. Finally, we show some illustrative, preliminary results obtained with our geodesic active fields framework in section VII and we discuss our model in section VIII.

II. Weighted Polyakov Energy

Sochen, Kimmel and Malladi introduced in [19] and [9] a general geometrical framework for low-level vision, based on an energy functional defined by Polyakov in [10]. In this framework which is widely used for image restoration, anisotropic smoothing and scale-spaces, images are seen as surfaces or hypersurfaces embedded in higher dimensional spaces.

A. Beltrami framework

An n-dimensional manifold Σ with coordinates $\sigma^{1\cdots n}$ is embedded in an m-dimensional manifold M with coordinates $X^{1\cdots m}$, with $m > n$. The embedding map $X : \Sigma \mapsto M$ is given by m functions of n variables. For example, a 2D gray-level image can be seen as a surface embedded in 3D: $X : (x, y) \to (x, y, I)$, where $I = I(x, y)$ corresponds to the gray-level intensities of the image. A Riemannian structure can be introduced: the metric $g_{\mu\nu}$ locally measures the distances on Σ, whereas on M distances are measured using h_{ij}.

To measure the weight of the mapping $X : \Sigma \mapsto M$, Sochen *et al.* [9] use the Polyakov energy, known from high energy physics [10], as a natural generalization of the L^2-norm on the embedded image to manifolds:

$$S[X^i, g_{\mu\nu}, h_{ij}] = \int \sqrt{g} g^{\mu\nu} \partial_\mu X^i \partial_\nu X^j h_{ij} \mathrm{d}^n\sigma, \qquad (4)$$

where the Einstein summation convention is used, g is the determinant of the image metric tensor, and $g^{\mu\nu}$ is its inverse, such that $g^{\mu\nu} g_{\nu\gamma} = \delta^\mu_\gamma$ (δ^μ_γ is the Kronecker delta). Naturally, the metric g is chosen as the induced metric, obtained by the *pullback*-relation: $g_{\mu\nu} = h_{ij} \partial_\mu X^i \partial_\nu X^j$. Under such a metric, the Polyakov energy shortens to:

$$S = \int \sqrt{g} \mathrm{d}^n\sigma, \qquad (5)$$

and represents the area of the embedded image surface. Using the Euler-Lagrange equation technique from calculus of variations, the following minimizing flow is obtained:

$$\partial_t X^i = \frac{1}{\sqrt{g}} \partial_\mu \left(\sqrt{g} g^{\mu\nu} \partial_\nu X^i \right) + \Gamma^i_{jk} \partial_\mu X^j \partial_\nu X^k g^{\mu\nu}, \qquad (6)$$

where the Levi-Civita connection Γ^i_{jk}, also called the Christoffel symbol, is defined as

$$\Gamma^i_{jk} = \frac{1}{2} h^{il} \left(\partial_j h_{kl} + \partial_k h_{jl} - \partial_l h_{jk} \right). \qquad (7)$$

Assuming the embedding is in a Euclidean space with Cartesian coordinates, the Christoffel symbols are all equal to zero, and the corresponding gradient descent equation is

$$\partial_t X^i = -\frac{1}{\sqrt{g}} \frac{\delta S}{\delta X^i} = \frac{1}{\sqrt{g}} \partial_\mu (\sqrt{g} g^{\mu\nu} \partial_\nu X^i) \equiv H^i, \qquad (8)$$

known as the Beltrami flow, where H^i denotes the i-th component of the mean curvature vector of the manifold.

A remarkable property of the Beltrami framework is the freedom to choose the metric of the embedding space. For example, let us embed a 2D gray-level image in 3D, using the following metric tensor:

$$h_{ij} = \mathrm{diag}(1, 1, \beta^2), \qquad (9)$$

where $\beta > 0$ is a constant. This allows to set the scale of the feature dimension independently of the spatial dimensions. The pullback relation yields the metric tensor $g_{\mu\nu}$:

$$g_{\mu\nu} = \begin{bmatrix} 1 + \beta^2 I_x^2 & \beta^2 I_x I_y \\ \beta^2 I_x I_y & 1 + \beta^2 I_y^2 \end{bmatrix}. \qquad (10)$$

Its determinant is given by $g = 1 + \beta^2 |\nabla I|^2$. Thus, the Polyakov energy of the embedding reads:

$$S = \int \sqrt{1 + \beta^2 |\nabla I|^2} \mathrm{d}^n\sigma. \qquad (11)$$

If $\beta \to \infty$, the 1 in this energy becomes negligible, and the energy approaches the TV-norm, well-known in image denoising [20], [21]. If, however, $\beta \to 0$, then the minimizing flow approaches the isotropic heat diffusion [9]. The impact of β on the apparent feature amplitudes of an embedded scalar field is illustrated in Fig. 2(a).

Moreover, the features being considered within the Beltrami framework are not restricted to scalar values only, but generalize directly to any vector value. For explicit applications of the framework to denoising of color images and textures, we refer the reader to [22]. For a review of the framework over a variety of manifolds and data structures, see [23].

B. Weighted Polyakov energy

Here, we present a weighted version of the Polyakov energy which will be used to define our registration model. In [24] the Polyakov energy was tuned by a weighting function f:

$$S_f = \int f \sqrt{g} \mathrm{d}^n\sigma, \qquad (12)$$

where $f = f(X^i, g_{\mu\nu}, h_{ij})$. In [24], the weighting function represents an edge-detector, that attracts an evolving contour to the edges in an image so as to segment it. In the GAF framework, the weighting function is an image discrepancy measure, that attracts the deformation field toward well aligned configurations, as will be seen shortly.

Still assuming the embedding is in a Euclidean space with Cartesian coordinates, i.e., h_{ij} is diagonal and constant, the corresponding gradient descent equation is

$$\partial_t X^i = f H^i + \partial_k f g^{\mu\nu} \partial_\mu X^k \partial_\nu X^i - \frac{m \cdot n}{2} \partial_k f h^{ki}, \qquad (13)$$

where fH^i corresponds to a weighted mean curvature flow on

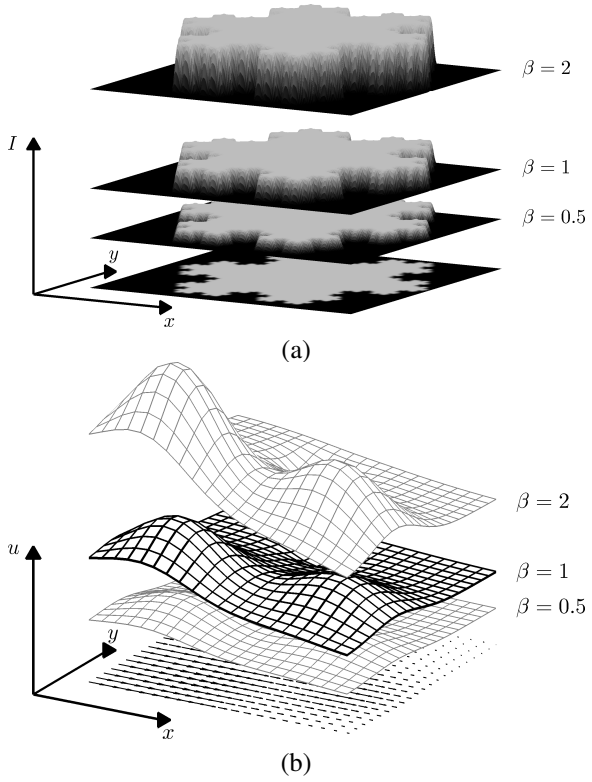


Fig. 2. (a) Beltrami framework: A gray-level image is embedded in 3D according to $X : (x, y) \rightarrow (x, y, I)$. The apparent effect of the aspect ratio β is illustrated as a relative scaling of the surface variations with respect to the spatial dimensions x and y. Minimizing the Polyakov energy smoothens the image, where β interpolates between total variation and Gaussian smoothing. (b) Geodesic active fields framework: The scalar deformation field (lateral shift u along x) of a planar image is embedded in $\mathbb{R}^3$. Formally, the embedding writes $X : (x, y) \rightarrow (x, y, u)$. The arrows illustrate the corresponding deformation field in the image plane. The Polyakov energy measures the area of the embedded surface, and is a measure of regularity of the deformation field. Weighting of the energy allows driving the minimization toward the optimal registration result.

manifolds.

III. Geodesic Active Fields

In this section we define the general evolution equation for the geodesic active fields for image registration. In contrast to the Beltrami framework for image denoising, we do not embed images, but the deformation field that relates the image pair to be registered. The deformation field is embedded as a mapping between the n-dimensional image domain and a m-dimensional space, where $m > n$. This is achieved by letting the components of the deformation field become additional dimensions of the embedding space. A very simple such embedding is illustrated in Fig. 2(b). We will then define metric tensors on the deformation field and the corresponding GAF energy to be minimized. The embedded deformation field manifold then evolves toward a *weighted* minimal surface, where the weighting function attracts it to a deformation field that brings the two images into registration.

In the most general form, we register a pair of n-dimensional images defined on a Riemannian domain Ω with coordinates $\mathbf{x} = (x_1, \ldots, x_n)$. The deformation field acts along $p \leq n$ dimensions, i.e., $\mathbf{u} : \Omega \mapsto \mathbb{R}^p, \mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), \ldots, u_p(\mathbf{x}))$.

The embedding X and the metric tensors h_{ij} and $g_{\mu\nu}$ on the deformation field are chosen as follows:

$$\begin{cases} X : (x_1, \ldots, x_n) \rightarrow (x_1, \ldots, x_n, u_1, \ldots, u_p) \\ h_{ij} \text{ is arbitrary} \\ g_{\mu\nu} = \partial_\mu X^i \partial_\nu X^j h_{ij}, \end{cases} \tag{14}$$

where $x_1, \ldots, x_n$ denote the spatial components of the image and $u_1, \ldots, u_p$ are the components of the dense deformation field. These equations are introduced in the weighted Polyakov functional (12) and its minimization flow (13), leading to the following general registration energy functional and the minimizing evolution flow of the geodesic active fields (GAF):

$$\begin{cases} E_{GAF} = \int f \sqrt{g} \prod_{i=1}^{n} \mathrm{d}x_i \\ \partial_t u_i = f H^{n+i} + \partial_k f g^{\mu\nu} \partial_\mu X^k \partial_\nu u_i - \frac{m \cdot n}{2} \partial_k f h^{k(n+i)}, \quad 1 \leq i \leq p, \end{cases} \tag{15}$$

where the weighting function $f = f(\mathbf{x}, \mathbf{u})$ is arbitrary for now, and will be defined in more detail in section VI.

The main contributions of this framework are:

1) The freedom to register images on any Riemannian manifold, i.e., on any smooth and parametrized surface. This will be developed further in sections IV and V.
2) The invariance under re-parametrization of the proposed energy, like the GAC energy [1] for the segmentation problem.
3) The freedom to choose the metric h_{ij} in the embedding space to obtain different regularizing behavior, as known from the versatility of the Beltrami framework.
4) The intrinsic data-dependent modulation of the local regularization strength thanks to the multiplicative weighting.

In image registration, the property of parametrization invariance is a very rare, but actually highly desirable property. Indeed, there is no reason why the chosen parametrization of the image domain should influence the outcome of the registration process. And yet, many currently used image registration methods lack this important invariance property.

The relevance of the contributions one and two can be clarified with the example of catadioptric images illustrated in Fig. 3. Such images are widely used in omnidirectional vision and robot navigation, for example, where ego-motion and position can be derived from a sequence of images, e.g. [25]–[27]. Because standard image registration methods ignore the paraboloid geometry of the actual image, they agnostically work on either one of the flat parametrized image versions. As can be clearly seen, there are important distortions between the raw, disc representation on the one hand, and the polar panorama view on the other hand, in Fig. 3a) and b), respectively. Obviously, a simple energy like mean squared error, employed in many standard methods, such as the popular Demons algorithm [28], fails to be re-parametrization invariant on those images,

$$\int (M - F)^2 \mathrm{d}x \mathrm{d}y \neq \int (M - F)^2 \mathrm{d}\phi \mathrm{d}r, \tag{16}$$

and the registration result depends on the selected parametriza-

(a) (x, y) (b) (ϕ, r) (c) P^2

Fig. 3. Omnidirectional *bunny* image obtained from a catadioptric system [29]. (a) Flat 2D raw image obtained from the camera, in Cartesian parametrization. (b) Panoramic representation, obtained by polar reparametrization of the raw disc. (c) Mapping of the *bunny* on the parabolic manifold, $P^2 = (r\cos\phi, r\sin\phi, r^2)$, corresponding to the focal projection of the catadioptric system. Classical registration of images of this type with standard methods will produce different results for the disc and panorama parametrizations. The GAF energy uses a metric corresponding to the actual image geometry and is therefore invariant to the chosen parametrization.

tion. In contrast, with the GAF energy, a metric tensor h_{ij} is derived from the actual image geometry, like the paraboloid in Fig. 3c), and the registration result becomes independent of the chosen image representation.

IV. Stereo Vision and Image Registration

In the previous section, we have defined the general, abstract energy of GAF and its corresponding gradient descent flow. In the following paragraphs, we instantiate this general concept for specific applications, namely stereo vision and 2D image registration in the Euclidean case, as well as image registration on non-flat manifolds. In other words, we will define specific image geometries and deformation field embeddings, derive the corresponding metric tensors, and thus concretize the GAF energy and its flow. The weighting function f, however, remains unspecified and will be described in detail only later, in section VI.

A. The general Euclidean case

Let us first consider the case of n-dimensional images defined on well-known Euclidean domains Ω with Cartesian coordinates $\mathbf{x} = (x_1, \dots, x_n)$. We look for a deformation field acting along $p \leq n$ dimensions.

The embedding X of the deformation field, and the corresponding metric tensors h_{ij} and $g_{\mu\nu}$ are chosen as follows:

$$\begin{cases} X : (x_1, \dots, x_n) \to (x_1, \dots, x_n, u_1, \dots, u_p) \\ h_{ij} = \mathrm{diag}(\underbrace{1, \dots, 1}_{n}, \underbrace{\beta^2, \dots, \beta^2}_{p}) \\ g_{\mu\nu} = \partial_\mu X^i \partial_\nu X^j h_{ij} = \delta_{\mu\nu} + \beta^2 \sum_{i=1}^{p} \partial_\mu u_i \partial_\nu u_i, \end{cases} \tag{17}$$

where β is the scaling factor applied to the deformation field components to get the desired aspect ratio. In analogy to the Beltrami framework, this parameter interpolates between isotropic Gaussian, and anisotropic TV-like smoothing of the deformation field. Now, the general Euclidean registration energy functional and the minimizing evolution flow, obtained by plugging the above choice into (15), take the following form:

$$\begin{cases} E_{GAF} = \int f\sqrt{g} \prod_{i=1}^{n} \mathrm{d}x_i \\ \partial_t u_i = fH^{n+i} + \partial_k f g^{\mu\nu} \partial_\mu X^k \partial_\nu u_i - \frac{m \cdot n}{2\beta^2} f_{u_i}, \quad 1 \leq i \leq p. \end{cases} \tag{18}$$

B. Stereo vision

Simply put, in stereo vision the depth information corresponding to a location is encoded as the lateral shift between its representation in two adjacent image acquisitions [30]. The recovered depth information is used in e.g. satellite imaging or robot vision to reconstruct the observed scene. The lateral shift can be determined by registration of the two images, where only lateral deformation is allowed. That is, the deformation field has only one component, along the x-dimension. We choose the following embedding and metric tensors, corresponding to the illustration in Fig. 2b):

$$\begin{cases} X : (x, y) \to (x, y, u) \\ h_{ij} = \mathrm{diag}(1, 1, \beta^2) \\ g_{\mu\nu} = \begin{bmatrix} 1 + \beta^2 u_x^2 & \beta^2 u_x u_y \\ \beta^2 u_x u_y & 1 + \beta^2 u_y^2 \end{bmatrix}, \; g = 1 + \beta^2 |\nabla u|^2. \end{cases} \tag{19}$$

Introducing those equations into (15), we get the following energy functional and evolution equation:

$$\begin{cases} E_{GAF} = \int f\sqrt{1 + \beta^2 |\nabla u|^2} \mathrm{d}x\mathrm{d}y \\ \partial_t u = fH^u + \partial_k f g^{\mu\nu} \partial_\mu X^k \partial_\nu u - \frac{3}{\beta^2} f_u, \end{cases} \tag{20}$$

where H^u is simply the 3rd component of the mean curvature vector:

$$H^u = \frac{g_{11} u_{yy} - 2 \cdot g_{12} u_{xy} + g_{22} u_{xx}}{g^2}. \tag{21}$$

C. 2-D image registration

In the case of registration, involving deformations along all image dimensions, one has $p = n$ and $m = 2n$. Here, as an example without loss of generality, we consider the registration of 2D images. The deformation field is described by u and v, resp. along x and y:

$$(u, v) : (x, y) \in \Omega \mapsto (u, v) = (u(x, y), v(x, y)) \in \mathbb{R}^2. \tag{22}$$

We choose the following embedding and metric tensors:

$$\begin{cases} X : (x, y) \to (x, y, u, v) \\ h_{ij} = \mathrm{diag}(1, 1, \beta^2, \beta^2) \\ g_{\mu\nu} = \begin{bmatrix} 1 + \beta^2(u_x^2 + v_x^2) & \beta^2(u_x u_y + v_x v_y) \\ \beta^2(u_x u_y + v_x v_y) & 1 + \beta^2(u_y^2 + v_y^2) \end{bmatrix} \\ g = 1 + \beta^2(|\nabla u|^2 + |\nabla v|^2) + \beta^4(\nabla u, \nabla v)^2, \end{cases} \tag{23}$$

where $(\nabla u, \nabla v) = u_x v_y - u_y v_x$ is defined as the magnitude of the cross product of the gradient vectors ∇u and ∇v. The expression of the determinant g has become quite cumbersome. The term $\beta^4(\nabla u, \nabla v)^2$ measures the misalignment of the gradients between different deformation field components [22]. All these settings put into the general equations produce

the following energy functional and minimizing flow:

$$\begin{cases} E_{GAF} = \int f\sqrt{1+\beta^2(|\nabla u|^2+|\nabla v|^2)+\beta^4(\nabla u,\nabla v)^2}\mathrm{d}x\mathrm{d}y \\ \partial_t u = fH^u + \partial_k f g^{\mu\nu}\partial_\mu X^k \partial_\nu u - \frac{4}{\beta^2} f_u \\ \partial_t v = fH^v + \partial_k f g^{\mu\nu}\partial_\mu X^k \partial_\nu v - \frac{4}{\beta^2} f_v. \end{cases} \tag{24}$$

D. Registration on non-flat manifolds

One of the main contributions of the proposed framework is that the image domain does not necessarily have to be Euclidean. Indeed, images to be registered can be defined on any Riemannian manifold, i.e., on any smooth and parametrized surface. In the Euclidean case, the spatial coordinates were directly given by the image domain parameters. In the non-Euclidean case, the spatial coordinates of the image are more complicated functions of the domain parameters instead.

To give a basic example, that will be illustrated in section VII, consider a spherical patch S described by two angles, θ and ϕ, on which the images are defined:

$$\begin{cases} S : (\theta,\phi) \in \Omega \subset \mathbb{R}^2 \to (x,y,z) \in \mathbb{R}^3 \\ \Omega = [\theta_{\min},\theta_{\max}] \times [\phi_{\min},\phi_{\max}] \\ 0 < \theta_{\min} < \theta_{\max} < \pi, \quad 0 \le \phi_{\min} < \phi_{\max} < 2\pi \\ x(\theta,\phi) = \sin\theta\cos\phi \\ y(\theta,\phi) = \sin\theta\sin\phi \\ z(\theta,\phi) = \cos\theta \end{cases} \tag{25}$$

The induced metric on S is naturally given by $g_S = \mathrm{diag}(1,\sin^2\theta)$. Further, let the deformation field $(\vartheta(\theta,\phi),\varphi(\theta,\phi))$ act on the two angles describing the patch. This suggests the following embedding:

$$\begin{cases} X : (\theta,\phi) \in \Omega \subset \mathbb{R}^2 \to (\theta,\phi,\vartheta,\varphi) \in \mathbb{R}^4 \\ h_{ij} = \mathrm{diag}(g_S, \beta^2 g_S) = \mathrm{diag}(1,\sin^2\theta,\beta^2,\beta^2\sin^2\theta) \end{cases} \tag{26}$$

where the metric tensor h_{ij} has been set by taking the induced metric g_S of the patch parametrization into account. The pullback relation yields the following metric tensor $g_{\mu\nu}$ in parameter space:

$$g_{\mu\nu} = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{bmatrix} + \beta^2 \begin{bmatrix} \vartheta_\theta^2 + \sin^2\theta\varphi_\theta^2 & \vartheta_\theta\vartheta_\phi + \sin^2\theta\varphi_\theta\varphi_\phi \\ \vartheta_\theta\vartheta_\phi + \sin^2\theta\varphi_\theta\varphi_\phi & \vartheta_\phi^2 + \sin^2\theta\varphi_\phi^2 \end{bmatrix} \tag{27}$$

Given this metric tensor h_{ij}, the embedding space is not Euclidean anymore, and the computation of the mean curvature vector involves the Levi-Civita connection as in (6), to account for the Riemannian part.

For the spherical patch, only two relevant Christoffel symbols computed by (7) differ from zero:

$$\Gamma^{\varphi}_{jk} = \begin{cases} \cot\theta & j \mathrel{\widehat{=}} \varphi, k \mathrel{\widehat{=}} \theta, \\ \cot\theta & j \mathrel{\widehat{=}} \theta, k \mathrel{\widehat{=}} \varphi, \\ 0 & \text{otherwise}, \end{cases} \tag{28}$$

and $\Gamma^{\vartheta}_{jk} = 0 \ \forall j,k$, where, with some abuse of notation, $j \mathrel{\widehat{=}} \varphi$ denotes the one j corresponding to the parameter φ, and consequently $\Gamma^{\varphi}_{jk} = \Gamma^{i}_{jk}|_{i\mathrel{\widehat{=}}\varphi}$. This gives the following evolution equations for the deformation field:

$$\begin{cases} \partial_t\vartheta = fH^{\vartheta} + \partial_k f g^{\mu\nu}\partial_\mu X^k \partial_\nu \vartheta - \frac{4}{\beta^2} f_\vartheta, \\ \partial_t\varphi = fH^{\varphi} + \partial_k f g^{\mu\nu}\partial_\mu X^k \partial_\nu \varphi - \frac{4}{\beta^2\sin^2\theta} f_\varphi, \\ H^{\vartheta} = \frac{1}{\sqrt{g}}\partial_\mu\left(\sqrt{g}g^{\mu\nu}\partial_\nu\vartheta\right), \\ H^{\varphi} = \frac{1}{\sqrt{g}}\partial_\mu\left(\sqrt{g}g^{\mu\nu}\partial_\nu\varphi\right) + 2\cot\theta(g^{\theta\theta}\varphi_\theta + g^{\theta\phi}\varphi_\phi), \end{cases} \tag{29}$$

where $g^{\theta\theta} = g^{\mu\nu}|_{\mu\mathrel{\widehat{=}}\theta,\nu\mathrel{\widehat{=}}\theta}$, and $g^{\theta\phi} = g^{\mu\nu}|_{\mu\mathrel{\widehat{=}}\theta,\nu\mathrel{\widehat{=}}\phi}$.

V. Multiscale Image Registration

A. Motivation

It is today commonly accepted, that the scale at which one measures a certain property becomes an additional dimension of the imaging space. Images are naturally composed of objects which are meaningful only at certain scales of observation [31], [32]. This has given rise to Witkin's patented notion of a *scale-space* [33]. Witkin introduced the concept of artificially generating larger (coarser) scales of an image through low-pass filtering.

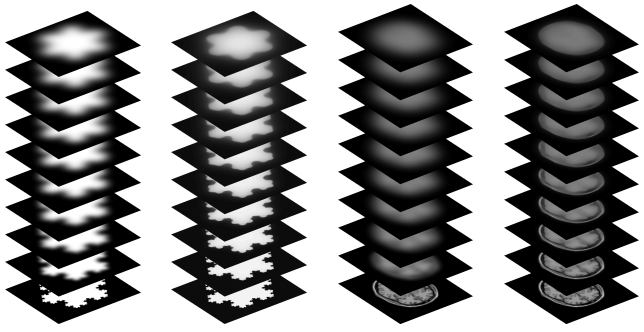
Scale-spaces have particular importance in the context of image registration. As an example, let us consider the human brain. It exhibits a highly convoluted and irregular structure, with high complexity and variability. For example, sulci and gyri vary a lot between subjects. On the other hand, high level structures of the brain – the "big picture" – are highly conserved, such as the two hemispheres, the lobes and main folds. Hence, a hierarchical representation of these structures is important in the context of inter-subject registration: considering the complexity of the cortical surface, directly involving local small-scale features would mislead the registration to be trapped in bad local minima. A robust method needs to rely on large-scale features, describing the main landmarks of the cortex, such as the main gyri or sulci, while small-scale features drive the registration more locally to reach the desired precision [34].

The most intuitive and commonly used approach to multiscale image registration consists of repeated, hierarchical registration at single scales – from coarse to fine. The result of one stage is used as initialization for the next finer scale. This pyramidal approach has reasonable computational load, but the link between scales is relatively weak, however, and unidirectional: information is only relayed from coarse to fine. Here, we propose a method of registering pairs of entire scale-spaces. All scales are registered simultaneously, thus allowing for bidirectional communication between scales.

The geometry of a large class of scale-spaces can be defined by a general metric tensor [35]:

$$h_{ij} = \mathrm{diag}\left(\underbrace{\frac{1}{c^2},\ldots,\frac{1}{c^2}}_{n \text{ times}}, \frac{1}{c^2\rho^2}\right), \tag{30}$$

where the first n elements of the diagonal correspond to the spatial dimensions $x_1,\ldots,x_n$, and the last element refers to the scale σ. c and ρ are two functions that represent the conductance and the density in the general model of heat transfer. The spatial derivative within such a scale-space is



(a) (b) (c) (d)

Fig. 4. (a)–(b) Linear and Beltrami scale-spaces of the Von Koch snowflake. The scale σ increases linearly from bottom to top of the image stack, thus constituting an additional image dimension. (c)–(d) Multiscale representation of a T1-weighted magnetic resonance image of a human brain, in a linear and Beltrami scale-space.

now obtained as $c\nabla$, whereas the scale derivative is given by $\rho c\partial_\sigma$. The natural heat equation, that defines the scale-space, is:

$$\partial_\sigma I = \frac{1}{\rho}\nabla\cdot(c\nabla I). \tag{31}$$

Different choices for c and ρ yield different well-known scale-spaces. The linear scale-space, e.g., corresponds to $c=\sigma$ and $\rho=1$: $\partial_\sigma I = \sigma\Delta I$. The Perona-Malik scale-space is reproduced with $\rho=1$ and $c=\exp(-\alpha|\nabla I|^2), \alpha>0$ [36]. The Beltrami flow of Sochen-Kimmel-Malladi requires $c=\rho=\frac{1}{\sqrt{1+\beta^2|\nabla I|^2}}$ [9], [19]. The linear and the Beltrami scale-space are illustrated at the example of the fractal image of a *Von Koch snowflake*, and a single slice of a T1-weighted brain MR image in Fig. 4.

B. Multiscale active deformation fields

Multiscale images have an additional image dimension: the scale σ. Along this scale-dimension, no deformation takes place. The multiscale deformation field is embedded as follows:

$$\begin{cases} X:(x_1,\dots,x_n,\sigma)\to(x_1,\dots,x_n,\sigma,u_1,\dots,u_n)\\ h_{ij}=\mathrm{diag}(\frac{1}{c^2},\dots,\frac{1}{c^2},\frac{1}{c^2\rho^2},\frac{\beta^2}{c^2},\dots,\frac{\beta^2}{c^2})\\ g_{\mu\nu}=\partial_\mu X^i\partial_\nu X^j h_{ij}, \end{cases} \tag{32}$$

where the structure of the metric tensor h_{ij} is arbitrary, and inspired by (30).

Considering a linear scale-space, i.e., $c=\sigma$ and $\rho=1$, the embedding thus looks like:

$$\begin{cases} X:(x_1,\dots,x_n,\sigma)\to(x_1,\dots,x_n,\sigma,u_1,\dots,u_n)\\ h_{ij}=\frac{1}{\sigma^2}\mathrm{diag}(1,\dots,1,1,\beta^2,\dots,\beta^2)\\ g_{\mu\nu}=\partial_\mu X^i\partial_\nu X^j h_{ij}=\frac{1}{\sigma^2}\left(\delta_{\mu\nu}+\beta^2\sum_{i=1}^{n}\partial_\mu u_i\partial_\nu u_i\right). \end{cases} \tag{33}$$

Again, as for non-flat image domains, the multiscale embedding is not Euclidean, and the Levi-Civita connection (7) is required to compute the complete mean curvature vector according to (6).

Note, that the deformation field $\mathbf{u}=\mathbf{u}(\mathbf{x},\sigma)$ evolves at all scales *simultaneously*. At each scale, the deformation field is attracted by the corresponding data term, while coherence between scales is obtained thanks to the regularizing power of harmonic maps.

C. Multiscale 2-D image registration

In the case of 2D images to be registered, the only relevant non-zero Christoffel symbols computed as (7) are $\Gamma^u_{u\sigma}=\Gamma^u_{\sigma u}=\Gamma^v_{v\sigma}=\Gamma^v_{\sigma v}=-\frac{1}{\sigma}$. The evolution equations for both components (u,v) of the deformation field along (x,y) are

$$\begin{cases} \partial_t u = fH^u+\partial_k f g^{\mu\nu}\partial_\mu X^k\partial_\nu u-\frac{15\sigma^2}{2\beta^2}f_u\\ \partial_t v = fH^v+\partial_k f g^{\mu\nu}\partial_\mu X^k\partial_\nu v-\frac{15\sigma^2}{2\beta^2}f_v\\ H^u=\frac{1}{\sqrt{g}}\partial_\mu\left(\sqrt{g}g^{\mu\nu}\partial_\nu u\right)-\frac{2}{\sigma}\partial_\mu u g^{\mu\sigma}\\ H^v=\frac{1}{\sqrt{g}}\partial_\mu\left(\sqrt{g}g^{\mu\nu}\partial_\nu v\right)-\frac{2}{\sigma}\partial_\mu v g^{\mu\sigma} \end{cases}, \tag{34}$$

where, with some abuse of notation, $g^{\mu\sigma}$ denotes the column ν of the inverse of the metric tensor $g^{\mu\nu}$ corresponding to the scale σ.

VI. What Choice of Weighting Function for the Registration Problem?

The purpose of the weighting function f is to drive the deformation field toward minimal surfaces that bring the two images into registration. As such, the flow should stop when the deformed image perfectly matches the target image. Hence, the weighting function is naturally chosen to be an image distance metric, which approaches zero when the two images match.

A. Deformation model

The weighting function is the place, where the deformation field actually gets to act on the images. Therefore, it is crucial to define the particular deformation model we want to use. First, we work with Euler coordinates. That is, for any pixel in the fixed imaged, the corresponding pixel is looked up in the moving image using a coordinate mapping. The corresponding location in the moving image will almost never fall on an exact pixel location and interpolation will be required.

Here, we use a very simple scheme, where the look-up is based on a shift by addition. The transform operator T is thus defined as:

$$T_0 : T_0\mathbf{x}=\mathbf{x}+\mathbf{u}(\mathbf{x}), \tag{35}$$

where addition is implicitly understood only along the $p\leq n$ dimensions of the image that are deformed. Also, for simplicity we shall ignore any boundary issues and finite support.

This very basic deformation model embodies only a restricted set of properties. By definition, the displacement needs to be at least twice differentiable, otherwise the Riemannian manifold cannot be constructed and mean curvature cannot be computed. Other than that, no further guarantees exist: the deformation is not necessarily invertible as nothing explicitly prevents the Jacobian to become negative. Further, it is not enforced to be surjective (onto), and homeomorphism or even diffeomorphism are not guaranteed properties. It is important to realize, however, that this is a restriction of the employed deformation model and not of the GAF framework as a whole.

More sophisticated deformation models can be used to obtain these properties.

Very recently, Vercauteren *et al.* introduced exponential map diffeomorphisms in the Demons framework [37]. There, at each iteration one looks for an infinitesimally small update $\mathrm{d}\mathbf{s}$ to $\mathbf{s} = 0$, that is applied through composition of its exponential map with the existing diffeomorphic deformation.

We have integrated this more complicated deformation field model into our GAF framework as well. However, diffeomorphisms are beyond the scope of this article and the details of this specific GAF version will be published separately. In the mean time, the reader may refer to [38].

Once the deformation model has been defined, corresponding fixed an moving image locations can be mapped, and the matching quality can be quantified using one of several distance metrics, of which we present some in the following paragraphs.

B. Squared error

If the images have been acquired using similar sensors, one can generally assume that the same entities are pictured at the same feature intensity in both images. An intuitive and simple choice for monomodal image registration subject to additive Gaussian noise is the squared error metric [39]:

$$f^{(1)T_0}(\mathbf{x},\mathbf{u}) = \left(M(T_0\mathbf{x}) - F(\mathbf{x})\right)^2 = \left(M(\mathbf{x}+\mathbf{u}(\mathbf{x})) - F(\mathbf{x})\right)^2, \tag{36}$$

where F and M refer to the fix and moving images respectively. The evolution equation (15) includes the partial derivatives of the weighting function with respect to all components of the embedding. For the function given in (36), these are obtained as follows:

$$\begin{cases} f_{\mathbf{x}}^{(1)T_0} = 2\cdot\left(M(T_0\mathbf{x}) - F(\mathbf{x})\right)\cdot\left(J^T\nabla M(T_0\mathbf{x}) - \nabla F(\mathbf{x})\right) \\ f_{\mathbf{u}}^{(1)T_0} = 2\cdot\left(M(T_0\mathbf{x}) - F(\mathbf{x})\right)\cdot\nabla M(T_0\mathbf{x}) \end{cases} \tag{37}$$

where ∇F and ∇M refer to the gradients of the fix and moving images, respectively, and where J^T denotes the transpose of the Jacobian of the deformed field:

$$J_{ij} = \delta_{ij} + \frac{\partial u_i}{\partial x_j}. \tag{38}$$

C. Local joint entropy

If images of different modality are to be registered, the above squared error metric is not a suitable distance metric anymore. Instead, mutual information is a commonly accepted similarity criterion in this case [40]–[42].

Mutual information is a global measure on the joint (p^{fm}) and marginal (p^f and p^m) histograms of the fixed and moving images:

$$MI = \sum_{i_1,i_2} p^{fm}\ln(p^{fm}) - \sum_{i_1} p^f\ln(p^f) - \sum_{i_2} p^m\ln(p^m), \tag{39}$$

where $p^{fm} = p^{fm}(i_1,i_2)$ etc. Let us assume, that the marginal entropies remain constant throughout the whole registration process, as they only depend on the fix and moving image separately. Maximizing mutual information is thus equal to minimizing the joint entropy.

The same joint entropy, i.e., the expectation of the negative logarithm of the joint probability, can also be computed in the image domain, instead of using the above histograms [43]:

$$H^{fm} = -\sum_{i_1,i_2} p^{fm}\ln(p^{fm}) = \frac{1}{|\Omega|}\sum_{\mathbf{x}} -\ln(p^{fm}), \tag{40}$$

where $p^{fm} = p^{fm}(F(\mathbf{x}), M(\mathbf{x}+\mathbf{u}))$. The negative logarithm denotes the local joint entropy. This local joint entropy has a minimum value of 0 (if the joint probability matches 1), and is unbound positive. This provides us with a local measure that corresponds well to the weighting function criteria stated above.

Consequently, we define the following information theory based weighting function for multi-modal image registration:

$$f^{(2)T_0}(\mathbf{x},\mathbf{u}) = -\ln(p^{fm}(F(\mathbf{x}), M(T_0\mathbf{x}))). \tag{41}$$

Using this weighting function, the goodness of a local alignment is measured by the frequency of similar intensity pairs in the rest of the image.

The partial derivatives along spatial components $f_{\mathbf{x}}^{(2)}$ are easily estimated numerically. The partial derivatives along deformation field components are obtained using the chain rule:

$$f_{\mathbf{u}}^{(2)T_0} = -\frac{p_m^{fm}(F(\mathbf{x}), M(T_0\mathbf{x}))}{p^{fm}(F(\mathbf{x}), M(T_0\mathbf{x}))}\cdot\nabla M(T_0\mathbf{x}), \tag{42}$$

where $p_m^{fm}(i_1,i_2)$ is the partial derivative of the histogram along the dimension corresponding to the moving image.

D. Absolute error

For non-smooth deformation fields, e.g., observed in optical flow-based image registration, the L^1-norm may perform better as data term [44]. The L^1-norm measures the absolute error between the two images,

$$f^{(3)T_0}(\mathbf{x},\mathbf{u}) = \left|M(T_0\mathbf{x}) - F(\mathbf{x})\right|, \tag{43}$$

and it can be approximated by a differentiable function:

$$f^{(3)T_0}(\mathbf{x},\mathbf{u}) = \sqrt{\left(M(T_0\mathbf{x}) - F(\mathbf{x})\right)^2 + \varepsilon^2}, \tag{44}$$

where $1 \gg \varepsilon > 0$. The partial derivatives of the approximated function are obtained easily:

$$\begin{cases} f_{\mathbf{x}}^{(3)T_0} = \frac{(M(T_0\mathbf{x}) - F(\mathbf{x}))}{f^{(3)T_0}(\mathbf{x},\mathbf{u})}\cdot\left(J^T\nabla M(T_0\mathbf{x}) - \nabla F(\mathbf{x})\right), \\ f_{\mathbf{u}}^{(3)T_0} = \frac{(M(T_0\mathbf{x}) - F(\mathbf{x}))}{f^{(3)T_0}(\mathbf{x},\mathbf{u})}\cdot\nabla M(T_0\mathbf{x}). \end{cases} \tag{45}$$

E. Data term and regularization balancing

In practice, we found useful to extend the weighting function by a positive constant, to convey a minimal weight to regularization. This is required in two cases: first, a pixel pair might accidentally fit well and locally produce zero discrepancy. As a consequence without a minimal weight, regularization would not be able to release the trapped pixels from their local minima. On the other hand, minimal regularization

weight is required by the aperture problem, otherwise displacement would not get propagated into matched, homogeneous regions [45], [46]. The general form of the weighting function is thus

$$f = 1 + \alpha f^{(i)}, \tag{46}$$

where $f^{(i)}$ is one of the image distance metrics specified above, and α is the balancing parameter, that scales the image discrepancy w.r.t. the constant minimal weight. This form represents a Polyakov energy functional, where the image distance metric corresponds to an additional penalty weighting. A big α will favor high data-fidelity, whereas a small value limits the modulating impact of the image discrepancy and increases overall regularization. Note that choosing 1 as the minimum weight renders the weighting function in some way symmetric to the square root of the metric tensor determinant, which shares the same lower bound.

VII. Results

We have implemented the geodesic active fields and ran it on several test problems. Here, the results are presented in order of task complexity. As for all forward schemes, the step length, and thus the speed of the registration, is heavily limited by the stability of the integration. The implementation was done using Matlab®(R2009a) on a standard 2.4GHz Intel®Core™2 Duo desktop machine, running a 64bit Fedora Core 11.

A. Mean curvature estimation

In (8), mean curvature is expressed as the anisotropic divergence of the coordinate gradient. Except for the simple stereo case, where an analytical expression of mean curvature was given, explicit expressions are cumbersome. Instead, we propose to estimate the mean curvature vector numerically, by using central differences twice.

In the 2D case of flat and non-flat images, this amounts to the same scheme as was already proposed by [13] and of which numerical properties have been studied and discussed in [47]. The numeric scheme for the mean curvature vector in the multiscale case is obtained in the same manner.

B. A few words on β and regularization

It might be useful to illustrate the influence of the scaling factor β on the deformation field smoothness. The analysis is easiest in the stereo case. For higher co-dimensions the analysis becomes more tedious and is beyond the scope of the present paper. We refer the reader to similar studies in the field of color and vector image denoising [22], and optical flow regularization [13]. To begin with, a pair of images is registered, where the one-directional deformation field u is initialized with a single local impulse. To study the impulse response of the regularization only, we wish $f = 1$ being constant, and set $\alpha = 0$. The so clutched GAF energy now corresponds exactly to the Beltrami framework for image denoising. Without surprise, after a few iterations, the deformation field has diffused, as illustrated in Fig. 5(a). Next, the deformation field is initialized with a unit step, disturbed by uniformly distributed, additive random noise, as shown in Fig. 5(b). In Fig. 5(c)-(e), the role of the parameter β becomes clear: The regularizer changes from Gaussian filtering for low β, to highly anisotropic, feature preserving TV-norm-like filtering at higher values. The actual choice of the parameter value depends on the available *a priori* knowledge on the deformation field regularity for a specific registration task. For computer vision applications such as motion detection and stereo vision, where entire image regions move as individual blocks (Gestalts), a higher β is preferable to allow for sharp deformation boundaries. Also think of the skulls in Fig. 1, where the rigid skull and mandible may be in a different relative pose in the image pair, whereas other applications would require more smooth transitions, thus motivating small β.

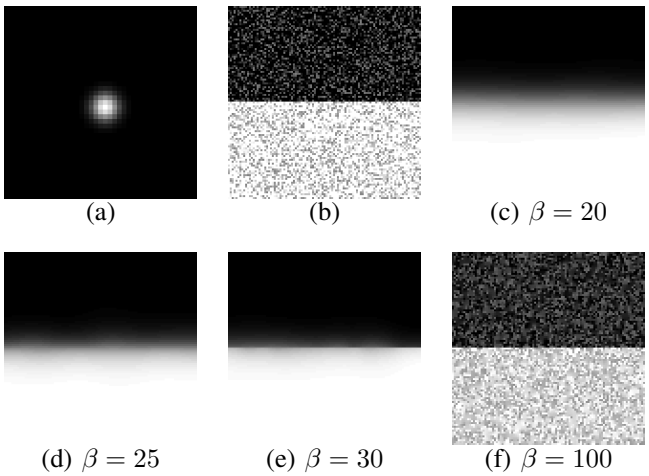


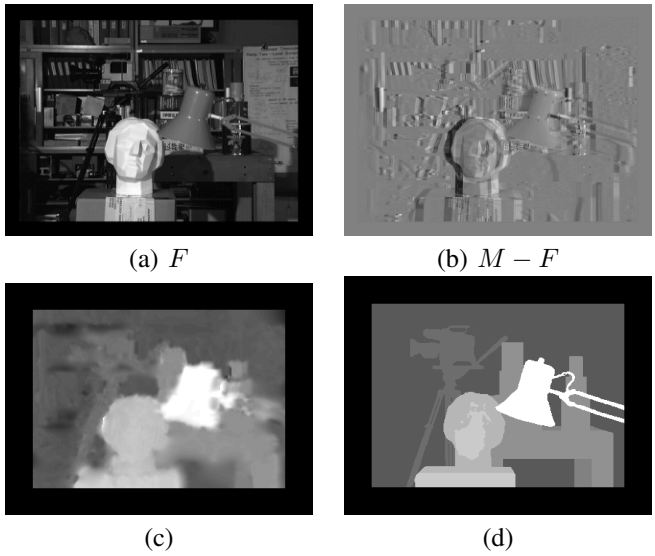
Fig. 5. (a) Mid-time response, i.e. after some 100 iterations, to a single impulse in the deformation field under constant data term $f = 1$. (b) Initial unit-step deformation field with uniform additive noise, and (c)–(f) its smoothing by the regularizer at different β.

C. Application to stereo vision

An example of stereo vision depth recovery was performed as shown in Fig. 6. The image pair `tsukuba` is a well known test image, taken from the middlebury benchmark set for stereo vision. The registration was set up according to the embedding and evolution equation described in section IV-B and using the absolute error weighting function (44). In our current implementation of which the goal is to illustrate the concept, the depth recovery result is fair, but does not yet achieve the quality of specifically tailored state-of-the-art stereo vision tools.

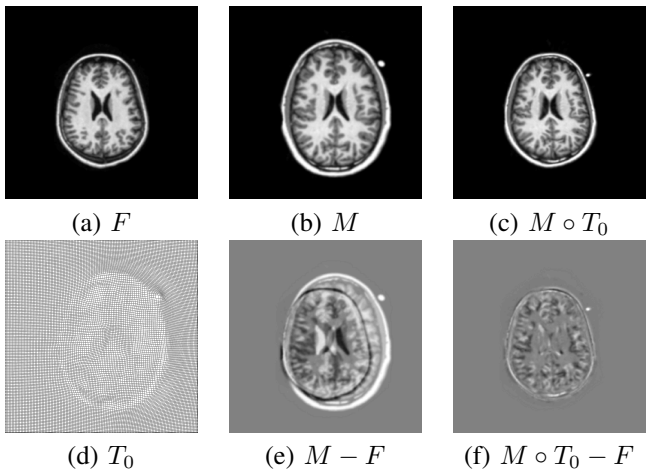
D. Application to medical imaging

The third case deals with registration of a highly misaligned monomodal medical image pair. Two roughly corresponding axial slices through the T1-weighted MRI volume of different subjects are to be registered. The images have a resolution of 256×256 pixels. Registration is set up with the squared error weighting function. The slices are well aligned by registration, as illustrated in Fig. 7. Note, that the subtle differences in the folding pattern cannot effectively be compensated by the



(a) F (b) $M - F$ (c) (d)

Fig. 6. (a)–(b) The `tsukuba` test image for stereo vision and the image pair difference. (c) The recovered disparity map, (d) Ground truth.

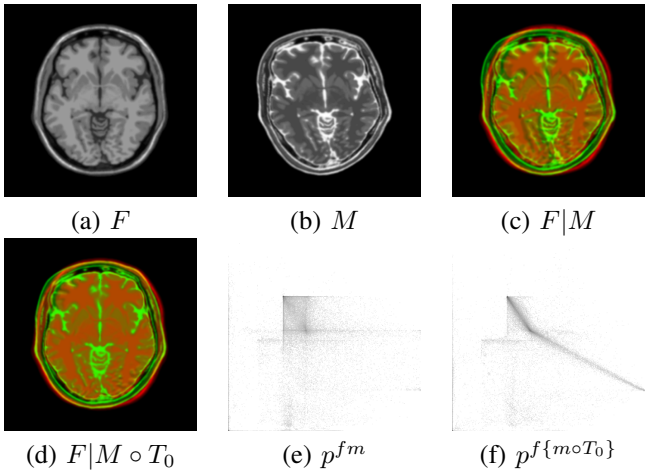


(a) F (b) M (c) $M \circ T_0$ (d) T_0 (e) $M - F$ (f) $M \circ T_0 - F$

Fig. 7. (a)–(b) Fix and moving image. (c) The moving image warped by the recovered deformation field. (d) The estimated deformation field. (e)–(f) Intensity differences before and after registration.

dense deformation field, i.e., the global outline of the skull and brain structures are aligned, but gyri and sulci remain largely individual.

Another case aims at registering a pair of multimodal medical images at resolution 317×317. The first image is the same T1 brain slice as above. The second image now is a deformed slice in T2 weighting. For multimodal registration, we use the information theory based joint local entropy as weighting function (41), again in the diffeomorphic setting using exponential map compositions. At the fine resolution, the resampling of an entire image takes considerable time, as well as the computation of the joint histogram. Accordingly, the whole registration process takes around 3 minutes. Registration is widely successful, as is indicated by the before and after checkerboard and overlay images provided in Fig. 8. Compared to the robust squared error weighting function, the local joint entropy is much more delicate with respect to the initial condition, but allows to register images of different



(a) F (b) M (c) $F|M$ (d) $F|M \circ T_0$ (e) p^{fm} (f) $p^{f\{m\circ T_0\}}$

Fig. 8. (a) The fix T1 image. (b) The artificially deformed T2 weighted image. (c)–(d) Checkerboard overlay of T1 and T2 images prior to and after registration. (e)–(f) The joint intensity histograms prior to and after registration. While initially, the histogram is widely spread, registration results in important histogram focalization.

modalities.

E. Registration on non-flat manifolds

To illustrate the model on a non-flat manifold, we have implemented the spherical patch described in (29). First, the purpose of the pullback relation is nicely illustrated in Fig. 9. We picture the impulse response that corresponds to the diffusion of a local non-zero spot in the deformation field without data term, i.e., α is simply set to 0, thus $f = 1$. On the spherical patch, the impulse response is isotropic and equal both close to the North pole and close to the equator. Isotropy on the spherical manifold requires a high degree of anisotropy in the rectangular parameter domain, as low-θ regions map denser on the sphere. This required anisotropy is directly obtained thanks to the pullback relation between the metrics h_{ij} on the patch and $g_{\mu\nu}$ in the parameter space. Further, the registration has been tested on an artificially deformed pair of topological maps of the Earth, see Fig. 10. The patch spans a good part of the northern hemisphere and some of the southern hemisphere of a globe, hence covering parts of both Americas, entire Europe, Africa, the Atlantic Ocean and of western Asia. Thus, the registration framework is shown to work on non-flat manifolds, such as the sphere.

F. Multiscale image registration

Finally, the multiscale registration case is tested on a pair of artificially deformed T1 brain images. Images are repeatedly lowpass filtered with a Gaussian to generate a linear diffusion scale-space. The multiscale image stacks prior to and after registration are shown in Fig. 11, as well as the corresponding intensity residues. Registration succeeds quite well, as illustrated by the almost entirely removed intensity errors.

VIII. Discussion and Conclusion

In this paper, we have presented a novel, purely geometric method, called geodesic active fields, to register images. The

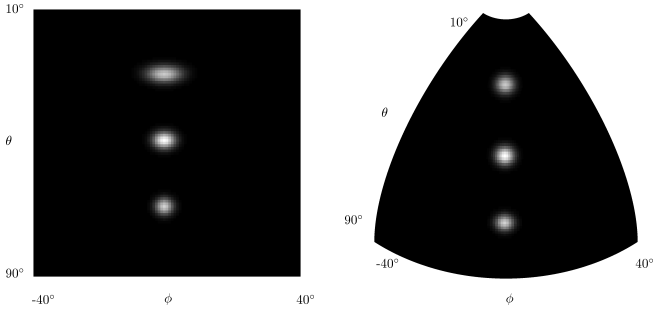


Fig. 9. (a) Diffusion takes place in parameter domain (θ, ϕ), governed by the metric tensor $g_{\mu\nu}$. At low θ, diffusion is highly anisotropic. (b) Diffusion as seen on the embedded spherical patch. Both impulse responses look the same and are isotropic. This is obtained through the pullback relation that links the respective metrics h_{ij} on the patch and $g_{\mu\nu}$ in the parameter space.

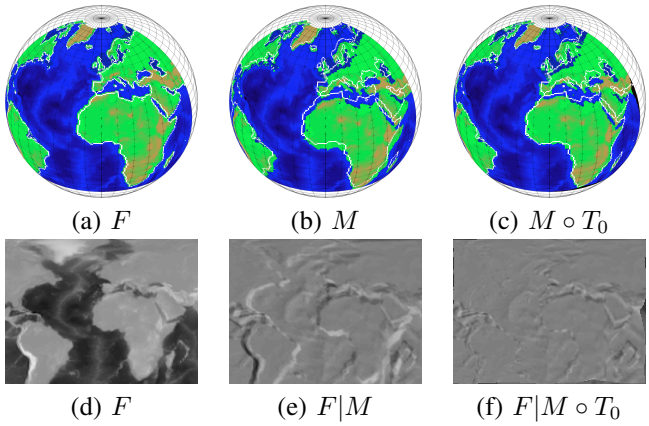


Fig. 10. (a) The fixed map with its approximate coastlines highlighted as white contours. (b) The moving image, with the coastlines of the fixed image superimposed. (c) The warped moving image after registration. The map fits well with the superimposed fixed coastlines, except at a few locations where small misregistration is observed (e.g. Red Sea). The colormap indicates height in meters above (below) sea level. (d) The fixed image in rectangular, flat parameter domain. (e)–(f) The red-green overlay before and after registration.

fundamental idea is to embed deformation fields in a weighted minimal surface energy and evolve the deformation field toward minimal surfaces, while being attracted by configurations that bring the images into registration. The process amounts to looking for an optimal hyper-contour in the space of all possible deformations in terms of image mismatch and deformation field regularity. This point of view reveals the close analogy to geodesic active contours in image segmentation [1], that can be derived from a weighted Polyakov energy as well [24], hence the name geodesic active fields (GAF).

In contrast to classic approaches in variational methods, which make use of purely additive competition between data and regularization term, our method combines the two energy contributions in a multiplicative way. In fact, the data term is represented by a local image distance function, that acts as multiplicative weighting on the geometric regularization term, resulting in a weighted surface energy. We recall the main contributions of the proposed framework:

1. Registration of non-flat and multiscale images. We have derived the minimizing flow of this weighted minimal surface for different image registration configurations. First, the framework applies to standard Euclidean images, defined on Cartesian planes and volumes. Further, our proposed method also directly generalizes to images on Riemannian manifolds, such as non-flat image domains and various scale-spaces, and ultimately the combination of both. In true multiscale registration and in contrast to hierarchic multiresolution approaches, image pairs are registered at all scales simultaneously. Communication between different scales is bidirectionally achieved by the regularization term, smoothing the deformation field across scales. In this context, we contribute a framework which has the advantage over classical approaches of automatically taking the relation between space and scale into account. Useful applications of non-flat image registration can easily be found in computer vision, e.g. motion detection or scene reconstruction from omnidirectional images.

2. Parametrization invariance. The second contribution of the proposed framework is the invariance of the registration result with respect to the parametrization chosen to describe the image domain. This result is also very intuitive, as by construction the employed energy measures the weighted hyperarea of the embedded deformation field, which is inherently independent of the parametrization that is used to describe this manifold.

3. Data-dependent, spatially-adaptive regularization. The multiplicative coupling of data-term and regularization intrinsically produces a data-dependent local modulation of the regularization strength. Naturally, one selects the one image discrepancy measure to be minimized that is the best estimate of alignment quality one can get. It is thus intuitive to let this same reliability estimate tune the local amount of regularization required. In practice, this might be particularly useful in medical image pairs that violate the premier assumption of actual existence of a one-to-one mapping between them, like a pair of images with and without lesions. In these instances, the adaptive regularization might help filling-in "the blanks" with a more regular deformation field than in the surrounding tissues that can be well aligned. It is also useful in images with regions of different noise levels. We thus require a smaller amount of global regularization, compared to classical additive schemes, where the non-adaptive regularization force always causes a bias off the optimal data position in the end result. Also, thanks to the multiplicative coupling, data-term and regularity compete very locally, in contrast to additive methods, where image distance metric and deformation field regularity compete as global measures on the whole image domain. Note, that the data-dependency of the regularization in GAF, based on the current local alignment of images, is different from [18], where regularization strength depends on individual image (gradient) intensities.

4. Geometric regularization. In a similar context, the geometric nature of the regularization, in particular its freedom to choose the amount of anisotropy through the parameter β, can allow for sharper deformation field transitions than classical Gaussian regularization. This is needed in cases, where individual objects move or deform independently, and where deformation cues from separate objects should not overly interact. Also, TV-like regularization reduces the impact

of deformation field outliers, as diffusion of the error is limited. Such outliers can occur at locations of actual image dissent, which can be observed for example with occlusions in stereo vision.

We would like to end this paper with some concluding remarks:

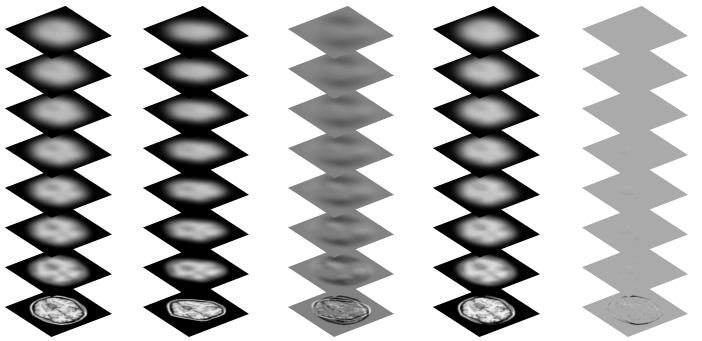
1. Weighting functions. We have provided three instances of weighting function, namely squared error (36), absolute error (44) and local joint entropy (41). On the one hand, the absolute and squared error weighting functions minimize the global L^2 and L^1-norm between the two images, and are suitable for monomodal image registration. The local joint entropy, on the other hand, maximizes the mutual information between images, and lends itself to multimodal image registration.

2. The parameters α and β. It is important to emphasize the role of the parameters α and β. First, β tunes the aspect ratio between the deformation field dimensions and the spatial dimensions in the embedding. In the simplest case of stereo matching, it has been shown that this allows interpolating between L^2 and L^1-norm minimization of the deformation field gradient magnitudes, whereas interpretation is more difficult in the general case. Second, note that β only changes the nature of the regularization, but not its relative weight with respect to the data term, which is precisely the role of the balancing parameter α.

3. Preliminary results, limitations and future work. In this paper, we have only shown preliminary results, based on very simple discretized forward Euler schemes. These are results for illustrative purposes only, that cannot compete with tightly tailored, and specifically tuned state-of-the-art solutions to practical applications. As we focus on the theoretical and methodological aspects of our image registration framework, we did not develop efficient and accurate numerical schemes to challenge established state-of-the-art methods. Consequently we do not compare quantitatively with other registration methods.

The most stringent limitations of the current GAF implementations are numerical stability (mean curvature estimation) and computational complexity (small time steps). Consequently, our next efforts will, therefore, focus on bringing the GAF energy in a suitable form for more efficient numerical implementations, both in terms of speed, accuracy, and stability. On another note, we continue working on the integration of more sophisticated, diffeomorphic deformation models, that have become very popular in medical image registration.

As mentioned, the embedding we propose for GAF corresponds to the flow-driven geometric regularizer proposed in [13]. The second, combined flow-intensity driven regularizer of that article is not exploited in the proposed GAF formulation, but inclusion is straightforward. We propose to go even one step further by embedding textural features rather than intensities, to address cases where Gestalts are defined by regions of homogeneous texture rather than flat intensity.



(a) F (b) M (c) $M - F$ (d) M' (e) $M' - F$

Fig. 11. (a) The multiscale stack of the fixed image of a T1 brain slice. (b) The moving image of a T1 brain slice, obtained by synthetic deformation of the fixed image. (c) The intensity difference illustrates the misalignment at all scales. (d) The registered moving image. (e) The residue after registration is significantly reduced. Some misregistration is observed at the frontal parts of the skull.

REFERENCES

[1] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.

[2] D. W. Thompson, *On Growth and Form*. Cambridge: Cambridge Univ. Press, 1917.

[3] M. A. Audette, F. P. Ferrie, and T. M. Peters, "An algorithmic overview of surface registration techniques for medical imaging," *Med. Image Anal.*, vol. 4, no. 3, pp. 201–17, 2000.

[4] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surveys*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[5] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, Mar. 1998.

[6] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image Vision Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[7] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford, U.K.: Clarendon Press, 2000.

[8] G. Hermosillo, C. Chefd'hotel, and O. Faugeras, "Variational methods for multimodal image matching," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 329–343, 2002.

[9] N. Sochen, R. Kimmel, and R. Malladi, "A general framework for low level vision," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 310–318, Mar. 1998.

[10] A. M. Polyakov, "Quantum geometry of bosonic strings," *Phys. Lett. B*, vol. 103, no. 3, pp. 207–210, 1981.

[11] R. Ben-Ari and N. Sochen, "Non-isotropic regularization of the correspondence space in stereo-vision," in *ICPR 2004*, vol. 4, Aug. 2004, pp. 293–296.

[12] ——, "A geometric approach for regularization of the data term in stereo-vision," *J. Math. Imaging Vis.*, vol. 31, no. 1, pp. 17–33, May 2008.

[13] ——, "A geometric framework and a new criterion in optical flow modeling," *J. Math. Imaging Vis.*, vol. 33, no. 2, pp. 178–194, Feb. 2009.

[14] A. Pitiot and A. Guimond, "Geometrical regularization of displacement fields for histological image registration," *Medical Image Analysis*, vol. 12, no. 1, pp. 16–25, 2008.

[15] S. Sabatini, F. Solari, and G. Bisio, "Lattice models for context-driven regularization in motion perception," in *Neural Nets*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2003, vol. 2859, pp. 35–42.

[16] H.-H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 565–593, Sep. 1986.

[17] M. Reuter, H. D. Rosas, and B. Fischl, "Highly accurate inverse consistent registration: A robust approach," *NeuroImage*, vol. In Press, pp. –, 2010.

[18] L. Tang, G. Hamarneh, and R. Abugharbieh, "Reliability-driven, spatially-adaptive regularization for deformable registration," in *WBIR 2010*, ser. Lecture Notes in Computer Science, B. Fischer, B. Dawant, and C. Lorenz, Eds., vol. 6204, 2010, pp. 173–185.

[19] R. Kimmel, N. Sochen, and R. Malladi, "From high energy physics to low level vision," *Scale-Space Theory in Computer Vision*, vol. 1252, pp. 236–247, 1997.

[20] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation based Noise Removal Algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.
[21] L. Rudin and S. Osher, "Total Variation based Image Restoration with Free Local Constraints," *Proc. IEEE Int. Conf. Image Process.*, vol. 1, pp. 31–35, 1994.
[22] R. Kimmel, R. Malladi, and N. Sochen, "Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images," *Int. J. Comput. Vis.*, vol. 39, no. 2, pp. 111–129, 2000.
[23] N. Sochen, R. Deriche, and L. Lopez-Perez, "The Beltrami flow over manifolds," INRIA, Tech. Rep. 4897, 2003.
[24] X. Bresson, P. Vandergheynst, and J.-P. Thiran, "Multiscale active contours," *Int. J. Comput. Vis.*, vol. 70, no. 3, pp. 197–211, Dec. 2006.
[25] Y. Yagi, S. Kawato, and S. Tsuji, "Real-time omnidirectional image sensor (COPIS) for vision-guided navigation," *IEEE Trans. Robot. Autom.*, vol. 10, no. 1, pp. 11–22, Feb. 1994.
[26] J. Gaspar, N. Winters, and J. Santos-Victor, "Vision-based navigation and environmental representations with an omnidirectional camera," *IEEE Trans. Robot. Autom.*, vol. 16, no. 6, pp. 890–898, Dec. 2000.
[27] R. Bunschoten and B. Krose, "Robust scene reconstruction from an omnidirectional vision system," *IEEE Trans. Robot. Autom.*, vol. 19, no. 2, pp. 351–357, Apr. 2003.
[28] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.
[29] I. Bogdanova, X. Bresson, J.-P. Thiran, and P. Vandergheynst, "Scale Space Analysis and Active Contours for Omnidirectional Images," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1888–1901, Jul. 2007.
[30] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 7–42, 2002.
[31] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. Lon. Series B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.
[32] J. J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, no. 5, pp. 363–370, 1984.
[33] A. P. Witkin, "Scale-space filtering," *Int. Joint Conf. AI*, vol. 2, pp. 1019–1021, 1983.
[34] D. Zosso and J.-P. Thiran, "A scale-space of cortical feature maps," *IEEE Signal Process. Lett.*, vol. 16, no. 10, pp. 873–876, Oct. 2009.
[35] D. H. Eberly, "A differential geometric approach to anisotropic diffusion," in *Geometry-Driven Diffusion in Computer Vision*, B. ter Haar Romeny, Ed. Kluwer, 1994, pp. 371–392.
[36] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
[37] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *Neuroimage*, vol. 45, no. 1, Supplement 1, pp. S61–S72, Mar. 2009.
[38] D. Zosso, X. Bresson, and J.-P. Thiran, "Geodesic active fields - a geometric framework for image registration," Ecole Polytechnique Fédérale de Lausanne (EPFL), Tech. Rep. LTS-REPORT-2010-001, 2010.
[39] A. W. Toga, *Brain Warping*. Academic Press, 1999.
[40] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–52, Mar. 1996.
[41] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.
[42] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
[43] P. Rogelj, S. Kovacic, and J. C. Gee, "Point similarity measures for non-rigid registration of multi-modal data," *Computer Vision and Image Understanding*, vol. 92, no. 1, pp. 112–140, 2003.
[44] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, vol. 4713. Berlin, Heidelberg: Springer, 2007, pp. 214–223.
[45] P. Stumpf, "Über die Abhängigkeit der visuellen Bewegungsempfindung und ihres negativen Nachbildes von den Reizvorgängen auf der Netzhaut," *Zeitschrift fur Psychologie*, 1911.
[46] D. Todorovic, "A gem from the past: Pleikart Stumpf's (1911) anticipation of the aperture problem, Reichardt detectors, and perceived motion loss at equiluminance," *Perception*, vol. 25, no. 10, pp. 1235–1242, 1996.
[47] L. Dascal and N. A. Sochen, "A maximum principle for Beltrami color flow," *SIAM Journal on Applied Mathematics*, vol. 65, no. 5, pp. 1615–1632, 2005.

Dominique Zosso (S'06) was born in Berne, Switzerland, in 1983, and received the MSc. degree in electrical and electronics engineering from Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2006.

He has previously worked as Researcher with the Structural Bioinformatics Group at the Swiss Institute of Bioinformatics and Biozentrum, University of Basel. Since 2007, he is Research and Teaching Assistant at the Signal Processing Laboratory at EPFL, Lausanne, Switzerland, and enrolled in the EPFL Electrical Engineering Doctoral School. His current research interests include PDE and variational models for inverse problems in image processing and computer vision, in particular image registration.

Xavier Bresson received a B.A. degree in theoretical physics in 1998, a M.Sc. degree in electrical engineering from Ecole Supérieure d'Electricité, Paris, and a M.Sc. degree in signal processing from University of Paris XI in 2000. In 2005, he completed a PhD degree in the field of computer vision at Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

In 2006-2010, he was a Postdoc scholar in the Department of Mathematics at University of California, Los Angeles (UCLA) with Tony Chan and Stanley Osher. In 2010, he joined the Department of Computer Science at City University of Hong Kong as assistant professor. His current research works are focused on continuous convex relaxation techniques to find global solutions of non-convex problems in image processing and graph-based problems in machine learning, and a unified geometric framework for energy minimization models in image processing.

Dr. Bresson has published 35 papers in international journals and conferences.

Jean-Philippe Thiran (S'91–M'98–SM'03) received the Elect. Eng. and Ph.D. degrees from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1993 and 1997, respectively.

Since January 2004, he has been an Assistant Professor, responsible for the Image Analysis Group at the Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His current scientific interests include image segmentation, prior knowledge integration in image analysis, partial differential equations and variational methods in image analysis, multimodal signal processing, medical image analysis, including multimodal image registration, segmentation, computer-assisted surgery, and diffusion MRI processing.

Dr. Thiran was Co-Editor-in-Chief of Signal Processing (published by Elsevier Science) from 2001 to 2005. He is currently an Associate Editor of the International Journal of Image and Video Processing (published by Hindawi), and member of the Editorial Board of Signal, Image and Video Processing (published by Springer). He was the General Chairman of the 2008 European Signal Processing Conference (EUSIPCO 2008). He is a senior member of the IEEE, and a member of the MLSP and IVMSP technical committees of the IEEE Signal Processing Society.