SIAM J. IMAGING SCIENCES
Vol. 8, No. 3, pp. 2021–2069

Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics*

Vadim Fedorov†, Pablo Arias†, Rida Sadek†, Gabriele Facciolo‡, and Coloma Ballester†

Abstract. In this paper we study the problem of comparing two patches of images defined on Riemannian manifolds which in turn can be defined by each image domain with a suitable metric depending on the image. For that we single out one particular instance of a set of models defining image similarities that was earlier studied in [C. Ballester et al., *Multiscale Model. Simul.*, 12 (2014), pp. 616–649], using an axiomatic approach that extended the classical Álvarez–Guichard–Lions–Morel work to the nonlocal case. Namely, we study a linear model to compare patches defined on two images in $\mathbb{R}^N$ endowed with some metric. Besides its genericity, this linear model is selected by its computational feasibility since it can be approximated leading to an algorithm that has the complexity of the usual patch comparison using a weighted Euclidean distance. Moreover, we propose and study some intrinsic metrics which we define in terms of affine covariant structure tensors and we discuss their properties. These tensors are defined for any point in the image and are intrinsically endowed with affine covariant neighborhoods. We also discuss the effect of discretization over the affine covariance properties of the tensors. We illustrate our theoretical results with numerical experiments.

Key words. multiscale analysis, similarity measures, degenerate parabolic equations, structure tensors, affine invariance

AMS subject classifications. 68U10, 35Q80, 46N60

DOI. 10.1137/141000002

1. Introduction.

Image comparison is a topic that has received a lot of attention in the image processing and computer vision communities since it is a main ingredient in many applications, such as object recognition, stereo vision, image interpolation, image denoising, and exemplar-based image inpainting, among others. A common way to define a nonlocal similarity measure between two images is to compare the patches (local neighborhoods) around each pair of points formed by taking one point from each image. We consider a general setting in which images are defined on Riemannian manifolds. Such manifolds arise, for instance, for images defined on $\mathbb{R}^N$, endowed with a suitable metric depending on the image.

In [3] it was shown that multiscale analyses of similarities between images on Riemannian manifolds, satisfying a certain set of axioms, are (viscosity) solutions of a family of degenerate PDEs. Our goal in this paper is to study one particular instance of the set of models derived in [3], namely a linear model to compare patches defined on two images in $\mathbb{R}^N$ endowed

*Received by the editors December 12, 2014; accepted for publication (in revised form) July 23, 2015; published electronically September 22, 2015.
http://www.siam.org/journals/siims/8-3/100000.html

†Departament de Tecnologia, Universitat Pompeu Fabra, 08018 Barcelona, Spain (vadim.fedorov@upf.edu, pablo.arias@upf.edu, rida.sadek@upf.edu, coloma.ballester@upf.edu). The research of these authors was partially supported by MICINN project MTM2012-30772, by the ERC Advanced Grant INPAINTING (grant agreement 319899), and by GRC reference 2014 SGR 1301, Generalitat de Catalunya.

‡CMLA, ENS Cachan, 94230 Cachan, France (facciolo@cmla.ens-cachan.fr).

with some metric. Besides its genericity, this linear model is selected by its computational feasibility since the solution of the PDE can be approximated via the convolution with a short-time space-varying kernel, leading to an algorithm that has the complexity of the usual Euclidean patch comparison.

Let us review the fundamentals of the approach in [3]. Given two images u, v defined in their respective image domains (assume $\mathbb{R}^2$ for simplicity), we want to compare their neighborhoods at the points $x, y \in \mathbb{R}^2$, respectively. The simplest way to compare them would be to compare the two neighborhoods of x, y using the Euclidean distance. That is, let us define

$$\mathcal{D}(t, x, y) = \int_{\mathbb{R}^2} g_t(h)(u(x+h) - v(y+h))^2 \, dh, \tag{1.1}$$

where g_t is a given windowing function that we assume to be Gaussian of variance t. This formula gives an explicit comparison and assumes that the image domain is the Euclidean plane. It generalizes approaches to patch comparison applied, for example, in [8, 20, 29, 13].

The purpose in [3] was to define such measures of similarity in the case of images defined on Riemannian manifolds (e.g., the image plane endowed with an anisotropic metric, like the structure tensor). An explicit formula like (1.1) was not possible. As proved in [3] the measure is given by the solution of a degenerate parabolic PDE in the variables (x, y). Let us mention at this point that (1.1) is not an exception; it solves the equation

$$\frac{\partial \mathcal{D}}{\partial t} = \Delta_x \mathcal{D} + 2\mathrm{Tr}(D^2_{xy}\mathcal{D}) + \Delta_y \mathcal{D}, \tag{1.2}$$

which is possibly the simplest case of a linear PDE expressing the multiscale comparison of two image patches. Note that the scale t in (1.1) reflects the size of the patch used for the comparison. In the case of comparing image patches defined on Riemannian manifolds, a large family of possibilities will appear, derived from the axiomatic approach. As in [2, 10] the set of axioms will include architectural axioms and the comparison principle that permit one to define multiscale analyses as solutions of a degenerate parabolic PDE. Further specification can be attained by including linear or morphological assumptions.

In practice in the framework of [3], to fully specify the multiscale analysis of similarities, one needs to provide the manifolds $\mathcal{M}_i$, the corresponding metrics G_i, and an *a priori connection between both manifolds.* The latter is a field of linear maps which for each pair of points $(x, y) \in \mathcal{M}_1 \times \mathcal{M}_2$ maps isometrically the tangent space at x with the tangent space at y. This is a key concept that allows one to actually compare patches in both manifolds. Given two Riemannian manifolds and their respective metrics, there are infinite a priori connections, and the appropriate ones might depend on the application. Think, for instance, about the simple case in which we are comparing two images defined in $\mathbb{R}^2$, one of them being a rotated copy of the other one. In this case, before comparing patches, one might be interested in rotating them. In the present formalism, this rotation is provided by an a priori connection.

In this paper we concentrate on the case in which $(\mathcal{M}_i, G_i) = (\mathbb{R}^N, G_i)$, $i = 1, 2$, and on some of the linear models developed in [3], attending to its representative character and computational feasibility, as will be shown. In this setting, the multiscale similarity measure is given by the solution of the following PDE:

$$\frac{\partial \mathcal{D}}{\partial t} = \mathrm{Tr}(G_1(x)^{-1} D^2_x \mathcal{D}) + 2\mathrm{Tr}(G_1(x)^{-1/2} G_2(y)^{-1/2} D^2_{xy}\mathcal{D}) + \mathrm{Tr}(G_2(y)^{-1} D^2_y \mathcal{D}). \tag{1.3}$$

The computational complexity of solving this equation is determined by the product manifold $\mathcal{M}_1 \times \mathcal{M}_2$, and thus is of order S^4, if each image is determined on a grid of size S^2. For that reason, using a WKB approximation, we develop a practical formula to approximate the solutions of (1.3), making it numerically tractable. WKB theory, named after Wentzel, Kramers, and Brillouin, is well known in quantum mechanics and is used to find approximate solutions to linear partial differential equations with spatially varying coefficients.

We also focus on the problem of defining the metrics G_i such that the resulting multiscale similarity measure is affine invariant. By affine invariance in this setting we mean that the similarity values are invariant to affine warpings of any of the images. For this aim, we propose affine covariant structure tensors as the image metrics. The a priori connection is then determined up to a rotation (for each pair (x, y)), which we determine based on the image contents in each patch.

Additionally, we discuss the effects of camera blur and discretization over the affine invariance properties of the tensors. We illustrate the behavior, results, and properties with numerical experiments in several scenarios.

In summary, the main contributions of this paper are the practical formula for the multiscale similarity measure, the study of the proposed affine covariant structure tensors used to define the image metrics, and the corresponding numerical experiments.

Let us finally say that from the mathematical point of view the basic ingredients of the axiomatic approach in [3] are the papers [2, 12, 11, 10], and the results in [3] are an extension of them. They can be considered as a nonlocal extension (comparing two points) of the multiscale analyses defined using the axiomatic approach in [2].

The similarity measure proposed in this paper operates on intensities within elliptic patches whose size and shape are not fixed a priori. There exist other methods using shape adaptive patches [15, 17]. Some methods for patch or region comparison do not directly consider squared difference of color or intensity values. For example, the normalized cross-correlation between patches can be used [1, 7, 32]. In the object recognition context, it is common to build the scale-space of each image by low-pass filtering and then search through this scale-space to detect scale invariant feature keypoints. Some kind of feature descriptor is calculated at each keypoint, and its own characteristics scale. Finally, these descriptors are used to compare two keypoints. There exist many approaches to keypoint detection [35]. As for description, SIFT descriptors [22] are probably the most widely known among the others. In the object classification context the probabilistic matching using the so-called constellation model [9, 16] is used to find the probability that any two given patches belong to the same class.

Let us summarize the plan of the paper. In section 2.1 we collect some basic notation and definitions about Riemannian manifolds. Section 2.2 is devoted to recalling from [3] the notion of an a priori connection. In section 3 we define the basic set of axioms satisfied by multiscale analyses for image similarity measures defined on Riemannian manifolds, and we express them in terms of solutions of an (eventually degenerate) parabolic equation. In section 4 we consider the case of linear multiscale analyses, naturally obtaining that they are expressed as solutions of a linear equation generalizing the case of (1.2). We study the case of $\mathbb{R}^N$ and develop a WKB approximation formula of the solutions of (1.3) that leads to a feasible algorithm. We study multiscale properties of the proposed similarity measure with some experiments. In section 5

we address the construction of affine covariant tensors and affine covariant neighborhoods associated with them and study their properties both theoretically and experimentally. Then in section 6 we demonstrate more experiments with the proposed similarity measure. Finally, in section 7 we present our conclusions.

2. Preliminaries.

We collect in this section some basic notation and definitions about Riemannian manifolds and also recall from [3] the notion of an a priori connection.

2.1. Notation.

Let $(\mathcal{N}, h)$ be a smooth Riemannian manifold in $\mathbb{R}^{N+1}$. As a particular case we can consider $\mathcal{N} = \mathbb{R}^N$ (or a domain in $\mathbb{R}^N$) endowed with a general metric h. As usual, given a point $\eta \in \mathcal{N}$, we denote by $T_\eta\mathcal{N}$ the tangent space to $\mathcal{N}$ at the point η. By $T_\eta^*\mathcal{N}$ we denote its dual space.

Let η be a point on $\mathcal{N}$, let $U \subseteq \mathbb{R}^N$ be an open set containing 0, and let $\psi : U \to \mathcal{N}$ be any coordinate system such that $\psi(0) = \eta$. Let $h_{ij}(\eta)$ and $\Gamma_{ij}^{\mathcal{N},k}(\eta)$ (indices i, j, k run from 1 to N) denote, respectively, the coefficients of the first fundamental form of $\mathcal{N}$ and the Christoffel symbols computed in the coordinate system ψ around η. For simplicity we shall denote by $H(\eta)$ the (symmetric) matrix $(h_{ij}(\eta))$ and by $\Gamma^{\mathcal{N},k}(\eta)$ the matrix formed by the coefficients $(\Gamma_{ij}^{\mathcal{N},k}(\eta))$, $i, j = 1, \dots, N$, for each $k = 1, \dots, N$.

The scalar product of two vectors $v, w \in T_\eta\mathcal{N}$ will be denoted by $\langle v, w\rangle_\eta$, and the action of a covector $p^* \in T_\eta^*\mathcal{N}$, on a vector $v \in T_\eta\mathcal{N}$, will be denoted by (p^*, v). If $v, w \in T_\eta\mathcal{N}$, we have $\langle v, w\rangle_\eta = h_{ij}(\eta)v^i w^j$, where v^i, w^i are the coordinates of v, w in the basis $\frac{\partial}{\partial x^i}|_\eta$ of $T_\eta\mathcal{N}$. Let us note that we shall use Einstein's convention that repeated indices are summed. Using this basis for $T_\eta\mathcal{N}$ and the dual basis on $T_\eta^*\mathcal{N}$, if $p^* \in T_\eta^*\mathcal{N}$, and $v \in T_\eta\mathcal{N}$, we have $(p^*, v) = p_i v^i$. Notice that we may write $(p^*, v) = h_{ij}(\eta)p^j v^i$, where p^i are the coordinates of the vector p associated with the covector p^*. The relation between both coordinates is given by

$$p_i = h_{ij}(\eta)p^j \quad \text{or} \quad p^i = h^{ij}(\eta)p_j, \tag{2.1}$$

where $h^{ij}(\eta)$ denotes the coefficients of the inverse matrix of $h_{ij}(\eta)$. By a slight abuse of notation, we shall write (2.1) as

$$p^* = Hp \quad \text{or} \quad p = H^{-1}p^*.$$

In this way $H : T_\eta\mathcal{N} \to T_\eta^*\mathcal{N}$. In the case that ψ is a geodesic coordinate system, the matrix H is the identity matrix $I = (\delta_{ij})$, and I maps vectors to covectors, i.e., $I : T_\eta\mathcal{N} \to T_\eta^*\mathcal{N}$ (with the same coordinates in the dual basis). We shall denote by I^{-1} the inverse of I, mapping covectors to vectors.

Maps. Symmetric maps. Quadratic forms. We shall also use this coordinate system to express a bilinear map $\hat{A} : T_\eta\mathcal{N} \times T_\eta\mathcal{N} \to \mathbb{R}$. Indeed, if (A_{ij}) is the matrix of $\hat{A}$ in this basis, and $v, w \in T_\eta\mathcal{N}$, we may write $\hat{A}(v, w) = A_{ij}v^j w^i$. If $A^i_{\ j} = h^{ik}(\eta)A_{kj}$, then $A^i_{\ j}$ determines a map, called $A : T_\eta\mathcal{N} \to T_\eta\mathcal{N}$, such that $\hat{A}(v, w) = \langle Av, w\rangle = (HAv, w)$. Observe that $H(\eta)A : T_\eta\mathcal{N} \to T_\eta^*\mathcal{N}$. Observe also that our notation $A^i_{\ j}$ already indicates that $A = (A^i_{\ j})$ maps vectors to vectors. In our notation, we shall not distinguish between matrices and maps.

As usual, we say that a linear map $L : T_\eta\mathcal{N} \to T_\eta^*\mathcal{N}$ is symmetric if $(Lv, w) = (Lw, v)$ for any $v \in T_\eta\mathcal{N}$, $w \in T_\eta\mathcal{N}$. From now on, we shall use the notation

$$SM_\eta(\mathcal{N}) := \{A : T_\eta\mathcal{N} \to T_\eta^*\mathcal{N},\ A \text{ is symmetric}\}.$$

We shall also write

$$S_\eta(\mathcal{N}) := \{A : T_\eta\mathcal{N} \to T_\eta\mathcal{N},\ H(\eta)A \in SM_\eta(\mathcal{N})\}.$$

Notice that if $A : T_\eta\mathcal{N} \to T_\eta\mathcal{N}$, we define $A^t : T^*_\eta\mathcal{N} \to T^*_\eta\mathcal{N}$ by

$$(A^t p, v) = (p, Av) \qquad \forall v \in T_\eta\mathcal{N},\ p \in T^*_\eta\mathcal{N}.$$

We define $A^{t,h} : T_\eta\mathcal{N} \to T_\eta\mathcal{N}$ by

$$\langle A^{t,h}v, w\rangle = \langle v, Aw\rangle \qquad \forall v, w \in T_\eta\mathcal{N}.$$

From now on, when the point $\eta \in \mathcal{N}$ is understood we write H instead of $H(\eta)$. Notice that $HA^{t,h} = A^t H$.

If $A \in S_\eta(\mathcal{N})$, then $HA \in SM_\eta(\mathcal{N})$ and $(HAv, w) = (v, HAw)$, that is, $\langle Av, w\rangle = \langle v, Aw\rangle$. That is, $A^{t,h} = A$.

Rotations in the tangent space. Let us define a rotation $R : T_\eta\mathcal{N} \to T_\eta\mathcal{N}$ as a linear map that satisfies

$$\langle Rv, Rw\rangle = \langle v, w\rangle \qquad \forall v, w \in T_\eta\mathcal{N}.$$

Notice that rotations satisfy

$$R^t H R = H.$$

Note also that isometries (rotations) satisfy $R^{t,h} = R^{-1}$.

Let $B : T_\eta\mathcal{N} \to T_\eta\mathcal{N}$ be a matrix such that $BI^{-1}B^t = H^{-1}$. Thus $B^tHB = I$ and B is mapping an orthonormal basis of $(T_\eta\mathcal{N}, I)$ to an orthonormal basis of $(T_\eta\mathcal{N}, H(\eta))$.

If $R : T_\eta\mathcal{N} \to T_\eta\mathcal{N}$ is a rotation, then

$$(B^{-1}RB)^t I (B^{-1}RB) = I.$$

That is, $B^{-1}RB$ is a classical rotation.

Gradient and Hessian. Given a function u on $\mathcal{N}$, let us denote by $D_\mathcal{N}u$ and $D^2_\mathcal{N}u$ the gradient and Hessian of u, respectively. In a coordinate system, $D_\mathcal{N}u$ is the covector $\frac{\partial u}{\partial x^i}$, and $D^2_\mathcal{N}u$ is the matrix $\frac{\partial^2 u}{\partial x^i \partial x^j} - \Gamma^{\mathcal{N},k}_{ij}\frac{\partial u}{\partial x^k}$ which acts on tangent vectors. Thus, with this notation $D^2_\mathcal{N}u(\eta) : T_\eta\mathcal{N} \times T_\eta\mathcal{N} \to \mathbb{R}$ is a bilinear map, $\eta \in \mathcal{N}$, and is a symmetric matrix in coordinates. Let us write $\nabla_\mathcal{N}u$ as the vector of coordinates $h^{ij}\frac{\partial u}{\partial x^j}$. Then $|\nabla_\mathcal{N}u(\eta)|^2_\eta = \langle \nabla_\mathcal{N}u(\eta), \nabla_\mathcal{N}u(\eta)\rangle_\eta$. To simplify our notation we shall write Du and ∇u instead of $D_\mathcal{N}u$ and $\nabla_\mathcal{N}u$. The vector field ∇u satisfies $\langle \nabla u, v\rangle_\eta = du(v)$, $v \in T_\eta\mathcal{N}$, du being the differential of u.

The manifold $\mathcal{N} = \mathcal{M}^1 \times \mathcal{M}^2$. Let $(\mathcal{M}^i, g^i)$ be a smooth Riemannian manifold with metric g^i, $i = 1, 2$. Let $\Gamma^{(i)}$ be the connection on $\mathcal{M}^i$. We shall work here with a manifold $\mathcal{N} = \mathcal{M}^1 \times \mathcal{M}^2$ with the metric $h = g^1 \times g^2$ so that $T_\xi\mathcal{N} = T_{\xi_1}\mathcal{M}^1 \times T_{\xi_2}\mathcal{M}^2$, $\xi = (\xi_1, \xi_2) \in \mathcal{M}^1 \times \mathcal{M}^2$. If $(v_i, w_i) \in T_{\xi_1}\mathcal{M}^1 \times T_{\xi_2}\mathcal{M}^2$, $\xi = (\xi_1, \xi_2) \in \mathcal{M}^1 \times \mathcal{M}^2$, then we consider the metric

$$\langle (v_1, w_1), (v_2, w_2)\rangle_\xi = \langle v_1, v_2\rangle_{\xi_1} + \langle w_1, w_2\rangle_{\xi_2} = (G^1(\xi_1)v_1, v_2) + (G^2(\xi_2)w_1, w_2).$$

With a slight abuse of notation, let us write $G(\xi) = \mathrm{diag}(G^1(\xi_1), G^2(\xi_2))$ instead of $H(\xi)$.

Let $\xi = (\xi_1, \xi_2) \in \mathcal{M}^1 \times \mathcal{M}^2$. Let us consider a coordinate system of the form $\psi = (\psi_1, \psi_2) : U_1 \times U_2 \to \mathcal{M}^1 \times \mathcal{M}^2$ with $\psi_i(0) = \xi_i$, U_i being a neighborhood of 0 in $\mathbb{R}^N$. Write $x \in U_1, y \in U_2$. Let us denote the connection on $\mathcal{M}^1 \times \mathcal{M}^2$ as $\Gamma := \Gamma^{(1)} \otimes \Gamma^{(2)}$ with indices $i, j, k \in \{1, \dots, 2N\}$ with $\xi_i = \xi_{1i}$, $i \in \{1, \dots, N\}$, and $\xi_i = \xi_{2(i-N)}$, $i \in \{N+1, \dots, N\}$. Note the coordinates as z^i, $i \in \{1, \dots, 2N\}$, with $z^i = x^i$, $i \in \{1, \dots, N\}$, and $z^i = y^{i-N}$, $i \in \{N+1, \dots, N\}$. Using the formula

$$(\Gamma^{(1)} \otimes \Gamma^{(2)})^k_{ij} = \frac{1}{2} h^{kl} \left(\frac{\partial h_{jl}}{\partial z^i} + \frac{\partial h_{il}}{\partial z^j} - \frac{\partial h_{ij}}{\partial z^l} \right),$$

we obtain

$$(\Gamma^{(1)} \otimes \Gamma^{(2)})^k(x, y) = \begin{pmatrix} \Gamma^{(1)k}(x) & 0 \\ 0 & \Gamma^{(2)k}(y) \end{pmatrix}.$$

We denote by $SM_\xi(\mathcal{N})$ the set of symmetric matrices of size $2N \times 2N$ in $\mathcal{N} = \mathcal{M}^1 \times \mathcal{M}^2$.

2.2. A priori connections on $\mathcal{N} = \mathcal{M}^1 \times \mathcal{M}^2$. This is an important concept in this paper, and we need to clarify it. Suppose that both manifolds $\mathcal{M}^1$ and $\mathcal{M}^2$ coincide with $\mathbb{R}^N$ endowed with the Euclidean metric. Let u, v be two given images in $\mathbb{R}^N$. Then it would be standard to use the L^2 distance to compare the patches centered at x and y,

$$\mathcal{D}(t, x, y) = \int_{\mathbb{R}^N} g_t(h)(u(x+h) - v(y+h))^2 \, dh, \tag{2.2}$$

where g_t is a given window that we assume to be Gaussian of variance t. But if the image v is rotated, we could also use the L^2 distance between u and a rotated version of v (around y), namely

$$\mathcal{D}(t, x, y) = \int_{\mathbb{R}^N} g_t(h)(u(x+h) - v(y+Rh))^2 \, dh. \tag{2.3}$$

We admit that this decision is taken a priori and is done thanks to an operator that connects the tangent plane at both points.

Let $\xi = (\xi_1, \xi_2) \in \mathcal{N} = \mathcal{M}_1 \times \mathcal{M}_2$. Let us consider a coordinate system of the form $\psi = (\psi_1, \psi_2) : U_1 \times U_2 \to \mathcal{M}^1 \times \mathcal{M}^2$ with $\psi_i(0) = \xi_i$, U_i being a neighborhood of 0 in $\mathbb{R}^N$.

Definition 2.1. *We say that $P(\xi)$, $\xi = (\xi_1, \xi_2) \in \mathcal{N}$, is an a priori connection map in $\mathcal{N}$ if $P(\xi) : (T_{\xi_1}\mathcal{M}^1, G^1(\xi_1)) \to (T_{\xi_2}\mathcal{M}^2, G^2(\xi_2))$ is an isometry, i.e.,*

$$\langle P(\xi)v, P(\xi)w \rangle_{G^2(\xi_2)} = \langle v, w \rangle_{G^1(\xi_1)} \quad \forall v, w \in T_{\xi_1}\mathcal{M},$$

and we assume also that the map is differentiable in ξ.

Given an a priori connection $P(\xi) : (T_{\xi_1}\mathcal{M}^1, G^1(\xi_1)) \to (T_{\xi_2}\mathcal{M}^2, G^2(\xi_2))$, we can also define its inverse $P(\xi)^{-1} : (T_{\xi_2}\mathcal{M}^2, G^2(\xi_2)) \to (T_{\xi_1}\mathcal{M}^1, G^1(\xi_1))$. For simplicity, and because the arguments in P clearly specify whether we go from $\mathcal{M}^1$ to $\mathcal{M}^2$ or inversely, we denote $P(\xi_2, \xi_1) = P(\xi_1, \xi_2)^{-1}$ so that we have

$$P(\xi_2, \xi_1)P(\xi_1, \xi_2) = I. \tag{2.4}$$

Let us note that if the (orientable) manifold $\mathcal{M}^1 = \mathcal{M}^2 = \mathcal{M}$ admits an a priori connection (into itself), then there is a section of the frame bundle (bundle of orthonormal frames). This is equivalent to saying that there is a section of the bundle of reference systems. This is the notion of parallelizable manifolds. The manifolds $(\mathbb{R}^N, g(x))$ are parallelizable. If $\mathcal{M}$ has dimension 2, then $\mathcal{M}$ is parallelizable if and only if its Euler–Poincaré characteristic is 0 [33]. Any orientable manifold of dimension 3 is parallelizable [31].

Remark 1. Note that if we have a complete manifold with empty cut locus, we can define the a priori connection in it by parallel transport without ambiguities.

Remark 2. Note that if $P(\xi)$ is an a priori connection and we give two maps $R : \mathcal{M}^1 \to \mathrm{Isom}(T\mathcal{M}^1)$, $\bar{R} : \mathcal{M}^2 \to \mathrm{Isom}(T\mathcal{M}^2)$ (where $\mathrm{Isom}(T\mathcal{M}^i)$ denotes the set of isometry maps in the tangent bundle $T\mathcal{M}^i$), then $\bar{R}(\xi_2)P(\xi)R(\xi_1)$ is also an a priori connection.

In coordinates, $P(\xi)$ expresses the a priori connection in the coordinate system $\psi_1 \to \psi_2$. The isometry property can be written as

$$(P(\xi)^t G^2(\xi_2)P(\xi)v, w) = (G^1(\xi_1)v, w),$$

where $P(\xi)$ is expressed in the basis of $T_{\xi_1}\mathcal{M}^1$ associated with the metric $G^1(\xi_1)$ and the basis of $T_{\xi_2}\mathcal{M}^2$ associated with the metric $G^2(\xi_2)$. Then

$$P(\xi)^t G^2(\xi_2)P(\xi) = G^1(\xi_1). \tag{2.5}$$

Let us recall from [3] how to compute the a priori connection in another coordinate system. Let $\overline{\psi} = (\overline{\psi}_1, \overline{\psi}_2)$ be another coordinate system around ξ. Let $G^i(\xi_i), \overline{G}^i(\xi_i)$, $i = 1, 2$, be the metric matrices represented in the coordinate systems ψ_i, $\overline{\psi}_i$, respectively. Let $B_{G^i,\overline{G}^i}(\xi_i) = D(\psi_i^{-1} \circ \overline{\psi}_i)(0)$, $i = 1, 2$, and $B_{G,\overline{G}}(\xi) = (B_{G^1,\overline{G}^1}(\xi_1), B_{G^2,\overline{G}^2}(\xi_2))$. Note that $B_{G^i,\overline{G}^i}(\xi_i) : (T_{\xi_i}\mathcal{M}^i, \overline{G}^i(\xi_i)) \to (T_{\xi_i}\mathcal{M}^i, G^i(\xi_i))$ is such that $B_{G^i,\overline{G}^i}(\xi_i)^t G^i(\xi_i) B_{G^i,\overline{G}^i}(\xi_i) = \overline{G}^i(\xi_i)$. Note also that all matrices here are uniquely defined. Using the last equality, (2.5) can be expressed as $P^t(\xi)B_{G^2,\overline{G}^2}(\xi_2)^{-t}\overline{G}^2(\xi_2)B_{G^2,\overline{G}^2}(\xi_2)^{-1}P(\xi) = B_{G^1,\overline{G}^1}(\xi_1)^{-t}\overline{G}^1(\xi_1)B_{G^1,\overline{G}^1}(\xi_1)^{-1}$. If we define

$$\overline{P}(\xi) := B_{G^2,\overline{G}^2}(\xi_2)^{-1}P(\xi)B_{G^1,\overline{G}^1}(\xi_1), \tag{2.6}$$

then $\overline{P}(\xi)$ is an a priori connection in the coordinate system $\overline{\psi}_1 \to \overline{\psi}_2$, $\overline{P}(\xi) : (T_{\xi_1}\mathcal{M}^1, \overline{G}^1(\xi_1)) \to (T_{\xi_2}\mathcal{M}^2, \overline{G}^2(\xi_2))$ (see [3]). Then, (2.6) can be rewritten as

$$B_{G^2,\overline{G}^2}(\xi_2)\overline{P}(\xi) = P(\xi)B_{G^1,\overline{G}^1}(\xi_1), \tag{2.7}$$

and we see that both maps $B_{G^1,\overline{G}^1}(\xi_1)$ and $B_{G^2,\overline{G}^2}(\xi_2)$ reflect the same rotation when expressed in the corresponding a priori connections $P(\xi)$ and $\overline{P}(\xi)$, respectively.

Definition 2.2. *We say that the coordinate systems $\psi, \overline{\psi}$ are $P(\xi)$-related if $\overline{P}(\xi)$ is defined by (2.6). We will also say that they are R-related.*

Let us consider the case where $\mathcal{M}^1 = \mathcal{M}^2 = \mathcal{M}$ and $P(\xi)$ is an internal a priori connection given from parallel transport between ξ_1 and ξ_2, which is an isometry. Then one can define $\overline{P}(\xi)$ by parallel transport expressed in the coordinate systems $\overline{\psi}_1$, $\overline{\psi}_2$.

Generation of a priori connections. Fix a geodesic coordinate system around each point of $\mathcal{M}^i$; $I^i(\xi_i)$ is referred to this system for each $\xi_i \in \mathcal{M}^i$. For each $\xi \in \mathcal{N}$, let us consider an isometry map (assuming that it exists)

$$Q(\xi) : (T_{\xi_1}\mathcal{M}^1, I^1(\xi_1)) \to (T_{\xi_2}\mathcal{M}^2, I^2(\xi_2)).$$

Call $\mathrm{Isom}((T\mathcal{M}^1, I^1), (T\mathcal{M}^2, I^2))$ this set of maps. Let us note that this is nothing more than an a priori connection. We just have one concept, and we express it in different coordinate systems. Thus what we are going to do is to give the a priori connection Q in a geodesic coordinate field $\mathcal{GS}$ and derive its expression in another coordinate system field.

Let $B^i(\xi_i) : (T_{\xi_i}\mathcal{M}^i, I^i(\xi_i)) \to (T_{\xi_i}\mathcal{M}^i, G^i(\xi_i))$ be the corresponding canonical maps connecting a geodesic coordinate system $\mathcal{GS}$ around ξ_i to $(T_{\xi_i}\mathcal{M}^i, G^i(\xi_i))$. Thus

$$B^i(\xi_i)^t G^i(\xi_i) B^i(\xi_i) = I^i(\xi_i).$$

The map $B^i(\xi_i)$ is uniquely defined by the coordinate systems. Changing the geodesic coordinate system we get a different matrix.

Let $Q(\xi) \in \mathrm{Isom}((T\mathcal{M}^1, I^1), (T\mathcal{M}^2, I^2))$, where each I is referred to $\mathcal{GS}$, and let us define

$$P(\xi) := B^2(\xi_2) Q(\xi) B^1(\xi_1)^{-1}.$$

Then $P(\xi)$ is an a priori connection map.

Related rotations. Let us consider a coordinate system field and an a priori connection $P(\xi) : (T_{\xi_1}\mathcal{M}^1, G^1(\xi_1)) \to (T_{\xi_2}\mathcal{M}^2, G^2(\xi_2))$ in that system field. Let us consider a second coordinate system field with metric $\overline{G}^i(\xi_i) = G^i(\xi_i)$, $i = 1, 2$, for each $\xi_i \in \mathcal{M}^i$ so that $B_{G^i,\overline{G}^i}(\xi_i)$ is an isometry field. Let $R^i(\xi_i) := B_{G^i,\overline{G}^i}(\xi_i)$. Let $\overline{P}(\xi)$ be the derived connection. Then (2.7) can be written as

$$R^2(\xi_2) = P(\xi) R^1(\xi_1) \overline{P}(\xi)^{-1}. \tag{2.8}$$

We say that $(R^1(\xi_1), R^2(\xi_2))$ are P-related or R-related, and we call $R = (R^1(\xi_1), R^2(\xi_2))$ a diagonally related rotation (or just a diagonal rotation if no confusion arises).

Related germs of functions on $\mathcal{N} = \mathcal{M}^1 \times \mathcal{M}^2$. Let $C_b(\mathcal{N})$ denote the space of bounded continuous functions in $\mathcal{N}$ with the maximum norm. We think of $C_b(\mathcal{N})$ as the space of similarity functions on $\mathcal{N} = \mathcal{M}^1 \times \mathcal{M}^2$. We denote by $C_b^\infty(\mathcal{N})$ the space of infinitely differentiable functions on $\mathcal{N}$.

Let $C \in C_b(\mathcal{N})$. Let us denote

$$(C, \psi)(x, y) = C(\psi_1(x), \psi_2(y)) \qquad \forall (x, y) \in U_1 \times U_2.$$

Thus, we can say that $\psi = (\psi_1, \psi_2)$ and $\overline{\psi} = (\overline{\psi}_1, \overline{\psi}_2)$ are R-related if (2.8) holds. If $\overline{\psi}$ is R-related to ψ, we write $(C, \overline{\psi})$ as $R(C, \psi)$. Note that $R(C, \psi)$ is a linear map for the restriction of functions in $C_b(\mathcal{N})$ to a neighborhood of $(0, 0)$.

Gradient and Hessian. We denote by $SM_\xi(\mathcal{N})$ the set of symmetric matrices of size $2N\times 2N$ in $\mathcal{N}=\mathcal{M}^1\times\mathcal{M}^2$. In coordinates, we denote $D_{\mathcal{N}}C=(D_xC,D_yC)$,

$$D^2_{\mathcal{N}}C=\begin{pmatrix} D_{\mathcal{N},xx}C & D_{\mathcal{N},xy}C\\ D_{\mathcal{N},xy}C & D_{\mathcal{N},yy}C\end{pmatrix}.$$

In coordinates, with $i,j,k\in\{1,\dots,N\}$,

$$D^2_{\mathcal{N}}C=\begin{pmatrix} \frac{\partial^2 C}{\partial x^i\partial x^j} & \frac{\partial^2 C}{\partial x^i\partial y^j}\\ \frac{\partial^2 C}{\partial y^j\partial x^i} & \frac{\partial^2 C}{\partial y^i\partial y^j}\end{pmatrix}-\begin{pmatrix} \Gamma^{(1)k}(x)\frac{\partial C}{\partial x^k} & 0\\ 0 & \Gamma^{(2)k}(y)\frac{\partial C}{\partial y^k}\end{pmatrix}.$$

3. Multiscale analysis of image similarity measures. For simplicity, we shall write $\mathcal{N}=\mathcal{M}^1\times\mathcal{M}^2$. The metric will be denoted by $g=g^1\times g^2$, and $G(\xi),G^1(\xi_1),G^2(\xi_2)$ will be the corresponding matrices, $\xi=(\xi_1,\xi_2)\in\mathcal{N}$. Let $(\kappa):=\kappa_n$ be an increasing sequence of nonnegative constants. We define the following set of functions:

$$\mathcal{Q}((\kappa)):=\{C\in C^\infty_b(\mathcal{N}):\ \|D^\alpha C\|_\infty\le\kappa_n\ \forall n\ge 0,\ \forall|\alpha|\le n\}.$$

As usual, $O(f)$ (resp., $o(f)$) will denote any expression which is bounded by $c|f|$ for some constant $c>0$ (resp., such that $\frac{o(f)}{|f|}\to 0$ as $f\to 0$). Assume that $T_t:C_b(\mathcal{N})\to C_b(\mathcal{N})$ is a nonlinear operator for any $t\ge 0$. We shall denote $C(t,\xi_1,\xi_2)=T_tC(\xi_1,\xi_2)$, $C\in C_b(\mathcal{N})$, $\xi=(\xi_1,\xi_2)\in\mathcal{N}$, $t\ge 0$. By a slight abuse of notation, to highlight the implicit coordinate system $\psi:U_1\times U_2\to\mathcal{N}$, we will denote it as $T_t(C,\psi)$, which can also be written as $C(t,\psi(x,y))=T_t(C,\psi)(x,y)$, $(x,y)\in U_1\times U_2$. Assume that we are given an a priori connection P on $\mathcal{N}$.

In the following we review the axioms for multiscale analyses of similarities.

Architectural axioms:

[Recursivity] $T_0(C)=C$, $T_s(T_tC)=T_{s+t}C\ \forall s,t\ge 0$, $\forall C\in C_b(\mathcal{N})$.

[Infinitesimal generator] $\frac{T_h(C,\psi)-(C,\psi)}{h}\to(\mathcal{A}(C),\psi)$ as $h\to 0+$ for any $C\in C^\infty_b(\mathcal{N})$ and any coordinate system $\psi=(\psi_1,\psi_2)$ around $\xi\in\mathcal{N}$, where $\mathcal{A}$ is the so-called infinitesimal generator for T_t [2]. We assume that

$$T_t(R(C,\psi))(\psi^{-1}(\xi))=R(T_t(C),\psi)(\psi^{-1}(\xi))+o(t)=T_t(C)(\xi)+o(t)\qquad\text{as } t\to 0+ \tag{3.1}$$

for any $C\in C_b(\mathcal{N})$, any coordinate system $\psi=(\psi_1,\psi_2)$, and any R which are P-related rotations. We have denoted by $R(C,\psi)$ the function in the coordinate system $\overline{\psi}$ which is $P(\xi)$-related (or R-related) to ψ.

Writing (3.1) in terms of the generator $\mathcal{A}$ we have

$$R(C,\psi)(0)+t\mathcal{A}(R(C,\psi))(0)+o(t)=R((C,\psi)+t\mathcal{A}(C,\psi))(0)+o(t)$$

$$=C(\xi)+t\mathcal{A}(C,\psi)(0)+o(t).$$

Using the linearity of $R(C,\psi)$, dividing by t, and letting $t\to 0+$ we obtain

$$\mathcal{A}(R(C,\psi))(0)=R\mathcal{A}(C,\psi)(0)=\mathcal{A}(C,\psi)(0) \tag{3.2}$$

for any $C \in C_b(\mathcal{N})$, any coordinate system $\psi = (\psi_1, \psi_2)$, and any R- or P-related rotations.

Remark 3. In $T_t(R(C, \psi))$ the a priori connection is expressed in the coordinate system $\overline{\psi} = (\overline{\psi}_1, \overline{\psi}_2)$. In $R(T_t(C), \psi) = T_t(C)$ the a priori connection is expressed in the coordinate system $\psi = (\psi_1, \psi_2)$. That is, the infinitesimal generator axiom says that both expressions are the same (intrinsic character of T_t) when the coordinate systems are R-related.

Remark 4. The infinitesimal generator axiom contains the invariance with respect to diagonal rotations in the tangent plane of $\mathcal{M}^1 \times \mathcal{M}^2$. When $(\mathcal{M}^i, G^i) = (\mathbb{R}^N, I)$, it amounts to invariance with respect to Euclidean diagonal rotations in $\mathbb{R}^{2N}$. That is, $T_t(RC) = RT_t(C)$ $\forall t \geq 0$, $\forall C \in C_b^\infty(\mathbb{R}^N \times \mathbb{R}^N)$, $\forall R \in O(N)$ (Euclidean rotations in $\mathbb{R}^N$) where $RC(x, y) = C(Rx, Ry)$.

Remark 5. When both manifolds are $\mathcal{M} = \mathbb{R}^N$ with the Euclidean metric, the axioms are just the following:

[Infinitesimal generator] $\frac{T_h C - C}{h} \to \mathcal{A}(C)$ as $h \to 0+$. This holds for any $C \in C_b^\infty(\mathbb{R}^N \times \mathbb{R}^N)$.

In some sense the coordinate system around each point is always the same, the canonical one; they are related by the identity.

[Regularity axiom] $\|T_t(C + h\tilde{C}) - (T_t(C) + h\tilde{C})\|_\infty \leq Mht$ $\forall h, t \in [0, 1]$, $\forall C, \tilde{C} \in \mathcal{Q}((\kappa))$ where the constant M depends on $\mathcal{Q}((\kappa))$.

[Locality] $T_t(C)(x) - T_t(\tilde{C})(x) = o(t)$ as $t \to 0+$, $x \in \mathbb{R}^N$, $\forall C, \tilde{C} \in C_b(\mathcal{N})$ such that $D^\alpha C(x) = D^\alpha \tilde{C}(x)$ for all multi-indices α.

Comparison principle:

[Comparison principle] $T_t C \leq T_t \tilde{C}$ $\forall t \geq 0$ and all $C, \tilde{C} \in C_b^\infty(\mathcal{N})$ such that $C \leq \tilde{C}$.

Since we are not going to consider the morphological comparison of image patches, we will not mention the morphological axiom and its consequences. We refer the reader to [3] for details. We just mention the following:

[Gray level shift invariance] $T_t(0) = 0$, $T_t(C + \kappa) = T_t(C) + \kappa$ $\forall t \geq 0$, $\forall C \in C_b^\infty(\mathcal{N})$, $\forall \kappa \in \mathbb{R}$.

Let us recall the following result from [3] (see also [2, 10]).

Theorem 3.1. *Let T_t be a multiscale analysis satisfying all the architectural axioms and the comparison principle. Then there exists a function $F : SM_\xi(\mathcal{N}) \times T_\xi^*\mathcal{N} \times \mathbb{R} \times \mathcal{N} \to \mathbb{R}$ increasing with respect to its first argument such that*

$$\frac{T_t(C, \psi) - (C, \psi)}{t} \to F(D^2(C \circ \psi)(0), D(C \circ \psi)(0), C(\xi), \xi, G, \Gamma^k) \text{ in } C_b(\mathcal{N}) \text{ as } t \to 0+$$

$\forall C \in C_b^\infty(\mathcal{N})$, ψ being a coordinate system around $\xi \in \mathcal{N}$. The function F is continuous in its first three arguments. If we assume that T_t is gray level shift invariant, then the function F does not depend on C.

The function F is elliptic; i.e., if $A_1, A_2 : T_\xi\mathcal{N} \to T_\xi^*\mathcal{N}$ are two matrices such that A_1, A_2 are symmetric and $A_1 \leq A_2$, $p \in T_\xi^*\mathcal{N}$, $c \in \mathbb{R}$, then

$$F(A_1, p, c, \xi, G, \Gamma^k) \leq F(A_2, p, c, \xi, G, \Gamma^k).$$

Recall that we have denoted $G = (G^1, G^2)$ and $\Gamma = \Gamma^{(1)} \otimes \Gamma^{(2)}$. Notice that we did not denote explicitly the arguments for G, Γ^k. Notice that the first argument in F is a symmetric map from $T_\xi\mathcal{N}$ to $T^*_\xi\mathcal{N}$.

Theorem 3.2. *Let T_t be a multiscale analysis satisfying all the architectural axioms, the comparison principle, and gray level shift invariance. If $C(t,\xi) = T_tC(\xi)$, then C is a viscosity solution of*

$$\frac{\partial C}{\partial t} = F(D^2_{\mathcal{N}}C, DC, \xi, G, \Gamma^k), \tag{3.3}$$

with $C(0,\xi) = C(\xi)$.

The proof that $C(t,\xi) = T_tC(\xi)$ is the viscosity solution of (3.3) follows as in [2, 19].

4. The linear case. Let us particularize the above result when the multiscale analysis on similarity functions satisfies also the linearity axiom (that is, $T_t(aC_1+bC_2) = aT_t(C_1)+bT_t(C_2)$ $\forall a, b \in \mathbb{R}$, $\forall C_1, C_2 \in C_b(\mathcal{N})$).

Theorem 4.1. *Let T_t be a multiscale analysis on similarity functions satisfying all the architectural axioms, the comparison principle, and gray level shift invariance. Assume that T_t is linear. Then*

$$\frac{\partial C}{\partial t} = F(D^2_{\mathcal{N}}C, \xi, G),$$

where

$$\begin{aligned} F(X,\xi,G) = {} & c_{11}(\xi)\mathrm{Tr}((G^1)^{-1}(\xi_1)X_{11}) + 2c_{12}(\xi,G)\mathrm{Tr}(\bar{D}_{12}I^1(\xi_1)^{-1}X_{12}) \\ & + c_{22}(\xi)\mathrm{Tr}((G^2)^{-1}(\xi_2)X_{22}), \end{aligned}$$

where $\bar{D}_{12}$ is an isometry from $(T_{\xi_1}\mathcal{M}^1, G^1(\xi_1)) \to (T_{\xi_2}\mathcal{M}^2, G^2(\xi_2))$. The ellipticity of F implies that $c_{11}, c_{22} \geq 0$.

The previous result was proved in [3], where, moreover, the authors add that

$$2c_{12}(\xi,G)\bar{D}_{12}I^1(\xi_1)^{-1} = B^2(\xi_2)D'B^1(\xi_1)^t, \tag{4.1}$$

and the dependence of $c_{12}(\xi,G)\bar{D}_{12}I^1(\xi_1)^{-1}$ on G is only in $B^2(\xi_2)$ (isometry) and $B^1(\xi_1)^t$ (isometry). D' is a matrix that depends only on ξ (see [3]). We could also write the second term as $2c_{21}(\xi,G)\mathrm{Tr}(\bar{D}_{21}I^2(\xi_2)^{-1}X_{21})$.

Note that if $X = D^2_{\mathcal{N}}C$, then $X_{ii} = D^2_{\mathcal{M}_i}C$ and the operators $c_{ii}(\xi)\mathrm{Tr}((G^i)^{-1}(\xi_i)X_{ii})$ are multiples of the Laplace–Beltrami operator. Notice also that there are no first order terms in these operators. They cannot couple with vectors so that we have the required invariance induced by the rotations of tangent planes.

4.1. The case of $(\mathcal{M}^r, g^r) = (\mathbb{R}^N, g^r)$. To fix ideas we consider $\mathcal{M}^1 = \mathcal{M}^2 = \mathcal{M} = \mathbb{R}^N$ and $g^r_{ij}(x)$ to be general metrics in $\mathbb{R}^N$, $r = 1, 2$. We know that $e_i = G^r(x)^{-1/2}f_i$ is an orthonormal basis of $(T_x\mathcal{M}^r, g^r(x))$ if f_i is a Euclidean orthonormal basis. Let $I^r(x) : (\mathbb{R}^N, g^r(x)) \to (\mathbb{R}^N, (g^r)^{-1}(x))$ be given by $I^r(x)e_i = e^*_i$. Then

$$I^r(x) = G^r(x).$$

If $B^r(x)$ satisfies $B^r(x)I^r(x)^{-1}B^r(x)^t = G^r(x)^{-1}$, then we can take $B^r(x) = I$.

We can define $P(x,y)(v) = G^2(y)^{-1/2}G^1(x)^{1/2}v$, $v \in \mathbb{R}^N$, as the a priori connection of x and y. Then $|P(x,y)v|_{g^2} = |v|_{g^1}\ \forall (x,y) \in \mathbb{R}^{2N}$. Recall that $\bar{D}_{1,2} : (\mathbb{R}^N, g^1(x)) \to (\mathbb{R}^N, g^2(y))$ is an isometry; in this case it is given by $\bar{D}_{1,2} = G^2(y)^{-1/2}G^1(x)^{1/2}$. Then (4.1) is

$$2c_{12}(x,y)\bar{D}_{1,2}I^1(x)^{-1} = 2c_{12}(x,y)G^2(y)^{-1/2}G^1(x)^{-1/2}.$$

The PDE obtained is

$$\frac{\partial C}{\partial t} = a(x,y)\Delta_{\mathcal{M}x}C + 2c_{12}(x,y)\mathrm{Tr}(G^2(y)^{-1/2}G^1(x)^{-1/2}D_{xy}C) + c(x,y)\Delta_{\mathcal{M}y}C, \tag{4.2}$$

where

$$\Delta_{\mathcal{M}x}C = \mathrm{Tr}(G^1(x)^{-1}(D_{xx}C(x) - \Gamma^{(1)}(D_xC)(x))),$$

and a similar expression for the operator $\Delta_{\mathcal{M}y}$ holds.

Remark 6. Note that (first by transposition and then by reordering) we have

$$\begin{aligned}\mathrm{Tr}(G^2(y)^{-1/2}G^1(x)^{-1/2}D_{xy}C) &= \mathrm{Tr}(D_{yx}CG^1(x)^{-1/2}G^2(y)^{-1/2}) \\ &= \mathrm{Tr}(G^1(x)^{-1/2}G^2(y)^{-1/2}D_{yx}C),\end{aligned}$$

which is a symmetric expression in (x,y). If T_t is symmetric in (x,y), then c_{12} is also symmetric.

In the symmetric case, the matrix associated with the operator (4.2) is

$$\begin{pmatrix} a(x,y)G^1(x)^{-1} & c_{12}(x,y)G^2(y)^{-1/2}G^1(x)^{-1/2} \\ c_{12}(x,y)G^1(x)^{-1/2}G^2(y)^{-1/2} & c(x,y)G^2(y)^{-1} \end{pmatrix}.$$

It is positive semidefinite if and only if $a, c \geq 0$ and $ac - c_{12}^2 \geq 0$.

Remark 7. Let A, B be two $N \times N$ matrices, let u, v be two given images, let $C(0,x,y) = (u(x) - v(y))^2$, and let $C(t,x,y) = \int_{\mathbb{R}^N} g_t(h)C(0, x + Ah, y + Bh)\, dh$, where g_t is the Gaussian of scale t. Then $C(t,x,y)$ satisfies the equation

$$\frac{\partial C}{\partial t} = \mathrm{Tr}(AA^tD_x^2C) + 2\mathrm{Tr}(AB^tD_{xy}C) + \mathrm{Tr}(BB^tD_y^2C). \tag{4.3}$$

Note that this equation corresponds to the models described in Theorem 4.1 when the metrics are constant in both images. This will be exploited as a numerical approximation later in this paper, where the construction of the metrics, which is a relevant issue, will be also discussed in detail.

Let us check that $C(t,x,y) = \int_{\mathbb{R}^N} g_t(h)C(0, x + Ah, y + Bh)\, dh$ solves (4.3). Indeed,

$$\begin{aligned}\frac{\partial C}{\partial t} &= \int_{\mathbb{R}^N} (g_t(h))_t C(0, x + Ah, y + Bh)\, dh = \int_{\mathbb{R}^N} \Delta_h g_t(h) C(0, x + Ah, y + Bh)\, dh \\ &= \int_{\mathbb{R}^N} g_t(h)\Delta_h C(0, x + Ah, y + Bh)\, dh = \int_{\mathbb{R}^N} g_t(h)\Delta_h (u(x + Ah) - v(y + Bh))^2\, dh.\end{aligned}$$

Note that

$$\begin{aligned}\Delta_h(u(x + Ah) - v(y + Bh))^2 &= 2\|\nabla_h(u(x + Ah) - v(y + Bh))\|^2 \\ + 2(u(x + Ah) - v(y + Bh))\Delta_h(u(x + Ah) - v(y + Bh)) &=: I + II.\end{aligned}$$

For notation simplicity, let us not denote the arguments of u and v. Then

$$\begin{aligned} II &= 2(u-v)\Delta_h(u-v) = 2(u-v)\left(\mathrm{Tr}(AA^tD_x^2u) - \mathrm{Tr}(BB^tD_y^2v)\right) \\ &= 2(u-v)\left(\mathrm{Tr}(AA^tD_x^2(u-v)) + \mathrm{Tr}(BB^tD_y^2(u-v))\right) \\ &= \mathrm{Tr}(AA^tD_x^2(u-v)^2) - 2\mathrm{Tr}(AA^t\nabla_x(u-v)\otimes\nabla_x(u-v)) \\ &\quad + \mathrm{Tr}(BB^tD_y^2(u-v)^2) - 2\mathrm{Tr}(BB^t\nabla_y(u-v)\otimes\nabla_y(u-v)) \\ &= \mathrm{Tr}(AA^tD_x^2(u-v)^2) - 2\|A^t\nabla_x u\|^2 + \mathrm{Tr}(BB^tD_y^2(u-v)^2) - 2\|B^t\nabla_y v\|^2 \end{aligned}$$

and

$$\begin{aligned} I &= 2\|\nabla_h(u-v)\|^2 = 2\|A^t\nabla_x u - B^t\nabla_y v\|^2 \\ &= 2\|A^t\nabla_x u\|^2 + 2\|B^t\nabla_y v\|^2 - 4\langle A^t\nabla_x u, B^t\nabla_y v\rangle. \end{aligned}$$

Then

$$\begin{aligned} I + II &= \mathrm{Tr}(AA^tD_x^2(u-v)^2) + \mathrm{Tr}(BB^tD_y^2(u-v)^2) - 4\langle A^t\nabla_x u, B^t\nabla_y v\rangle \\ &= \mathrm{Tr}(AA^tD_x^2(u-v)^2) + \mathrm{Tr}(BB^tD_y^2(u-v)^2) + 2\mathrm{Tr}(AB^tD_{xy}(u-v)^2). \end{aligned}$$

Thus

$$\frac{\partial C}{\partial t} = \mathrm{Tr}(AA^tD_x^2C) + \mathrm{Tr}(BB^tD_y^2C) + 2\mathrm{Tr}(AB^tD_{xy}C).$$

4.2. WKB approximations. Let us concentrate on the analysis of the example in Remark 7. We start by writing the operator in (4.3) as

$$\begin{aligned} \frac{\partial C}{\partial t} &= \mathrm{Tr}(AA^tD_x^2C) + 2\mathrm{Tr}(AB^tD_{xy}C) + \mathrm{Tr}(BB^tD_y^2C) \\ &= \mathrm{Tr}\left(\begin{pmatrix} AA^t & AB^t \\ BA^t & BB^t \end{pmatrix}\begin{pmatrix} D_{\mathcal{N},xx}C & D_{\mathcal{N},xy}C \\ D_{\mathcal{N},xy}C & D_{\mathcal{N},yy}C \end{pmatrix}\right). \end{aligned}$$

Note that

$$\begin{pmatrix} AA^t & AB^t \\ BA^t & BB^t \end{pmatrix} = \begin{pmatrix} A & 0 \\ B & 0 \end{pmatrix}\begin{pmatrix} A^t & B^t \\ 0 & 0 \end{pmatrix} =: \Sigma\Sigma^t.$$

Note that neither Σ nor $\Sigma\Sigma^t$ is an invertible operator and we would need to regularize them (for instance by introducing a perturbation ϵI in the $(2,2)$ entry of Σ).

Remark 8. Note that the PDE above can be related to the following stochastic ODE:

$$\begin{aligned} dX_t &= AdW_t, \\ dY_t &= BdW_t, \end{aligned} \tag{4.4}$$

where the Brownian motion dW_t is common in both equations.

The more general linear case derived from (4.2) is

$$\frac{\partial C}{\partial t} = \Delta_{\mathcal{M}x}C + 2\mathrm{Tr}(G^2(y)^{-1/2}G^1(x)^{-1/2}D_{xy}C) + \Delta_{\mathcal{M}y}C, \tag{4.5}$$

where

$$\Delta_{\mathcal{M}x}C = \mathrm{Tr}(G^1(x)^{-1}(D_{xx}C - \Gamma^{(1)}(D_xC)(x,y))),$$

$$\Delta_{\mathcal{M}_y} C = \mathrm{Tr}(G^2(y)^{-1}(D_{yy}C - \Gamma^{(2)}(D_y C)(x,y)))$$

can be subsumed under the previous notation by taking

$$A = G^1(x)^{-1/2}, \qquad B = G^2(x)^{-1/2}.$$

In any case, for Σ to be invertible, let us write (4.3) as

$$\frac{\partial C}{\partial t} = \mathrm{Tr}(\Sigma\Sigma^t D^2_{\mathcal{N}} C) =: \mathrm{Tr}_{(\Sigma\Sigma^t)^{-1}}(D^2_{\mathcal{N}} C), \tag{4.6}$$

where $\mathrm{Tr}_{(\Sigma\Sigma^t)^{-1}}(D^2_{\mathcal{N}} C)$ is a notation for the Laplace–Beltrami operator (eventually degenerated), i.e., the trace of the Hessian with respect to the (eventually degenerated) metric $g := (\Sigma\Sigma^t)^{-1}$. In this case, the result in [37] could give the approximate formula we are looking for. However, we will use another approach.

We would like to obtain an approximation formula of the type

$$C(t+\epsilon, p) = \int K(\epsilon, p, p') C(t, p') dp',$$

where $p = (x, y)$, $p' = (x', y')$.

We proceed using the so-called WKB approximation (from quantum mechanics) as in the paper [30, section 3 and Appendix]. Following the WKB method, we assume the kernel K to be of the form

$$K(t, p, p') = \frac{H(t, p, p')}{\sqrt{t}} e^{-\Psi(p,p')/t}.$$

We are interested in the behavior of the kernel for t small. Without loss of generality, we can assume that $H(t, p, p') = H_0$, a constant. Indeed, as in [30, Appendix] one can check that for t small the leading term corresponds to a constant $H = H_0$. The function Ψ does not depend on t and is positive. The validity of this approximation procedure can be found in [14], for example.

To simplify the notation, let us forget the arguments of the above functions. Since the equation is linear, we may assume that it is satisfied by the kernel K. Then, introducing K into (4.6), some straightforward computations produce an equality with several terms. For short times only the most singular part is dominant. Therefore, by considering the leading terms of order $\frac{1}{t^{5/2}}$ on both sides of the equation, which are the most divergent ones as $t \to 0+$, we have that the term on the left-hand side (obtained from $\frac{\partial K}{\partial t}$) is

$$\frac{H_0}{\sqrt{t}} e^{-\Psi/t} \frac{\Psi}{t^2}$$

and the term obtained from the right-hand side is

$$\frac{H_0}{\sqrt{t}} e^{-\Psi/t} \frac{1}{t^2} \langle \Sigma\Sigma^t \nabla_p \Psi, \nabla_p \Psi \rangle.$$

The equality of both terms gives the following PDE:

$$\langle \Sigma\Sigma^t \nabla_p \Psi, \nabla_p \Psi \rangle = \Psi,$$

i.e.,

$$\|\Sigma^t \nabla_p \Psi\|^2 = \Psi. \tag{4.7}$$

If $\Phi = 2\sqrt{\Psi}$, (4.7) becomes

$$\|\Sigma^t \nabla_p \Phi\|^2 = 1. \tag{4.8}$$

From

$$\Sigma^t = \begin{pmatrix} A^t & B^t \\ 0 & 0 \end{pmatrix},$$

we can write (4.8) as

$$\|A^t D_x \Phi + B^t D_y \Phi\|^2 = 1. \tag{4.9}$$

Let us denote $(P, Q) = (D_x\Phi, D_y\Phi)$, and let $H(P, Q) = \|A^t P + B^t Q\|^2$. Writing the Hamilton–Jacobi equation (4.9) in terms of H, the solution of

$$H(P, Q) = 1$$

is given in terms of the Lagrangian

$$L(\bar{P}, \bar{Q}) = \sup_{(P,Q)} \left\{ \langle (\bar{P}, \bar{Q}), (P, Q) \rangle - H(P, Q) \right\}.$$

Note that, with

$$R = A^t P + B^t Q, \qquad S = A^t P - B^t Q,$$
$$\bar{R} = A^{-1} \bar{P} + B^{-1} \bar{Q}, \qquad \bar{S} = A^{-1} \bar{P} - B^{-1} \bar{Q},$$

we may write

$$L(\bar{P}, \bar{Q}) = \sup_{(R,S)} \left\{ \frac{1}{2} \langle (\bar{R}, \bar{S}), (R, S) \rangle - \|R\|^2 \right\}.$$

Then,

$$\text{if } \bar{S} \neq 0, \text{ then } L(\bar{P}, \bar{Q}) = +\infty;$$
$$\text{if } \bar{S} = 0, \text{ then } L(\bar{P}, \bar{Q}) = \tfrac{1}{16} \|\bar{R}\|^2 = \tfrac{1}{16} \|A^{-1} \bar{P} + B^{-1} \bar{Q}\|^2.$$

Recall that the solution of the Hamilton–Jacobi equation (4.9) is given by

$$\Phi(t, p, p') = \frac{1}{16} \inf_{\mathcal{C}} \int_p^{p'} \|A^{-1} \dot{\gamma}(s) + B^{-1} \dot{\tilde{\gamma}}(s)\|^2 \, ds,$$

where $p = (x, y)$, $p' = (x', y')$, $\mathcal{C}$ is the set of curves $(\gamma, \tilde{\gamma})$ such that γ is a curve joining x to x' and $\tilde{\gamma}$ a curve joining y to y', and

$$A^{-1} \dot{\gamma}(s) = B^{-1} \dot{\tilde{\gamma}}(s),$$

i.e.,

$$\dot{\gamma}(s) = AB^{-1}\dot{\tilde{\gamma}}(s). \tag{4.10}$$

Let us analyze this formula when A, B are constant matrices. Then, by integrating (4.10), we get

$$x' - x = AB^{-1}(y' - y),$$

i.e.,

$$B^{-1}(y' - y) = A^{-1}(x' - x) := h.$$

Thus, we may write

$$x' - x = Ah, \qquad y' - y = Bh.$$

Moreover,

$$\Phi(t, p, p') = \frac{1}{4} \inf_{\gamma} \int_x^{x'} \|A^{-1}\dot{\gamma}(s)\|^2 \, ds \left(= \frac{1}{4} \inf_{\gamma} \int_y^{y'} \|B^{-1}\dot{\tilde{\gamma}}(s)\|^2 \, ds \right). \tag{4.11}$$

Let us solve this equation explicitly in the case of constant matrices A, B. The solution is given by γ being a straight line joining its two endpoints. Indeed, writing $\alpha(s) := A^{-1}\gamma(s)$, then

$$\Phi(t, p, p') = \frac{1}{4} \inf_{\alpha} \int_{A^{-1}x}^{A^{-1}x'} \|\dot{\alpha}(s)\|^2 \, ds. \tag{4.12}$$

The solution is given by

$$\Phi(t, p, p') = \frac{1}{4}\|A^{-1}x - A^{-1}x'\| = \frac{1}{4}\|h\|. \tag{4.13}$$

Recall that $\Phi = 2\sqrt{\Psi}$. Then, for $t > 0$ small enough, we have the approximation

$$C(t, p) = \frac{H_0}{\sqrt{t}} \int e^{-\Psi(t,p,p')/t} C(0, x + Ah, y + Bh) dh$$

for some constant H_0. Then, after adjusting constants,

$$C(t, p) = \int g_t(h) C(0, x + Ah, y + Bh) dh,$$

where $g_t(h)$ denotes the Gaussian of variance t, and we recover the formula given in Remark 7.

In the general case, where A and B are not constant matrices but $A(x) = G^1(x)^{-1/2}$, $B(y) = G^2(y)^{-1/2}$, we have

$$\Phi(t, p, p') = \frac{1}{4} \inf_{\gamma} \int_x^{x'} \|G^1(\gamma(s))^{1/2}\dot{\gamma}(s)\|^2 \, ds = \frac{1}{4} \inf_{\tilde{\gamma}} \int_y^{y'} \|G^2(\tilde{\gamma}(s))^{1/2}\dot{\tilde{\gamma}}(s)\|^2 \, ds. \tag{4.14}$$

By a drastic approximation, writing $A = G^1(x)^{-1/2}$, $B = G^2(y)^{-1/2}$, $h = A^{-t}x' - A^{-t}x$, we obtain the formula (coinciding with the previous one for constant A, B)

$$C(t,p) = \int g_t(h) C(0, x + G^1(x)^{-1/2}h, y + G^2(y)^{-1/2}h) dh. \tag{4.15}$$

We think of the following case:

$$C(0,x,y) = (u(x) - v(y))^2.$$

Remark 9. If we interpret (4.14) intuitively as a geodesic distance $d(x,x')$ in the manifold $\mathcal{M}_1$, and $x' = x + A(x)h$, then

$$C(t,p) = \frac{H_0}{\sqrt{t}} \int e^{-d(x,x+A(x)h)^2/t} C(0, x + A(x)h, y + B(y)h) dh. \tag{4.16}$$

On the other hand, $d(x, x + A(x)h)^2$ could be approximated as in the bilateral filter [34] by $\kappa_{spatial}\|A(x)h\|^2 + \kappa_{color}|u(x) - u(x + A(x)h)|^2$, $\kappa_{spatial}, \kappa_{color} > 0$, where u is the image on $\mathcal{M}_1$.

Formulas (4.15) and (4.16) are both discussed and tested in subsection 4.3 and in section 6.

4.3. Experiments on the multiscale similarity measure. In this section we concentrate on a study of multiscale properties of the proposed patch similarity measure. The following experiments are aimed to demonstrate the behavior of the similarity measure at different scales and motivate the necessity of multiple scales. For this reason, the case of disparity map estimation was selected. Let us note that even though all the experiments in this section were made in the context of disparity map estimation, the complete application of the proposed similarity measure to the depth estimation from stereo images is out of the scope of the current work. Let us assume for now that the computation of an a priori connection map (and hence $A(x) = G^1(x)^{-1/2}$ and $B(y) = G^2(y)^{-1/2}$) for a pair of images is a "black box." It will be defined and studied in detail in section 5.

First experiment. In the first experiment we compute similarity values between a given point on the left image and all the points on the right image of the stereo pair shown in Figure 1. One point was selected in the interior of an object on the left image and another point on a boundary of an object. Let us note that in the case of disparity map estimation we restrict the search for the largest similarity value to the epipolar lines or, in our case, to the corresponding scan lines, because the images are rectified.

Let us specify the values of parameters used in these experiments. For the first experiment, the scale parameter was selected to be $t \in \{16.6, 25, 50, 150, 300, 15000\}$. A big value of t (e.g., $t = 15000$) corresponds to a coarse scale (large window), while a small value (e.g., $t = 16.6$) corresponds to a fine scale (small window). The following section will clarify how the metrics and the a priori connection maps are computed for these experiments. Let us just note for now that there is a parameter related to the a priori connection computation, which we set here to be $r = 150$ (see section 5 for details).

Formulas (4.15) and (4.16) allow us to calculate the patch distance $C(t,x,y)$ between two given points x and y. Given $t > 0$, for simplicity we denote it here as $d(x,y)$. For

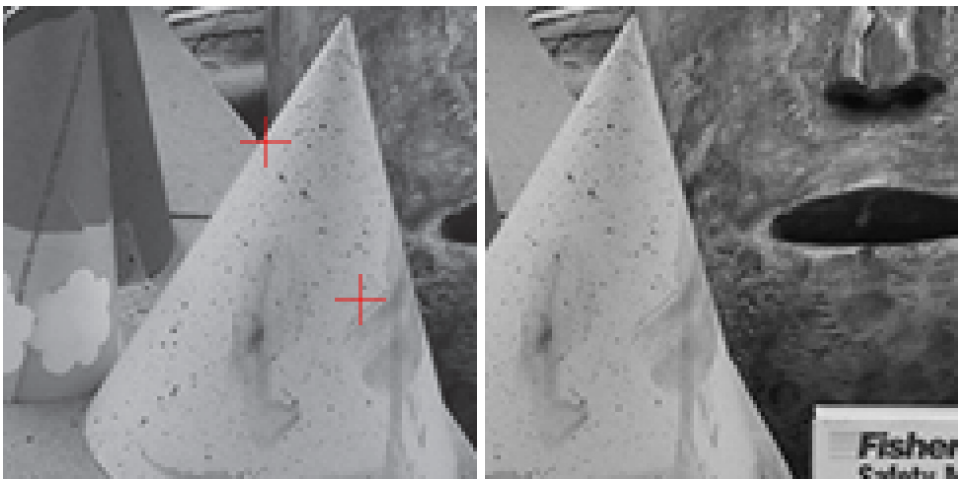

Figure 1. *Stereo pair used in the experiments. Two red crosses show the selected points of interest.*

visualization purposes these distances were color-coded by $c(x,y) = \exp(-\frac{(d(x,y)-d_{min})^2}{2\sigma^2})$, where $\sigma = \frac{d_{max}-d_{min}}{\gamma}$, d_{max} and d_{min} are the maximum and minimum patch distances values, respectively, and $\gamma > 0$ is a parameter. Now, the highest values of $c(x,y)$ correspond to the most similar patches. For all the experiments in this section, $\gamma = 50$, which stretches values close to 1. Note that d_{max} and d_{min} were computed from the whole set of distances across all the scales, which means that equal colors on two different similarity maps correspond to the same similarity values.

Figures 2 and 3 show similarity maps computed using Gaussian weights (see (4.15)). Figure 2 shows maps for a point selected in the interior of the object, while Figure 3 shows maps for a point selected on the boundary. Similarly, Figures 4 and 5 show similarity maps for the same setting but computed with approximated geodesic weights (see (4.16)) as in the bilateral filter (with $\kappa_{spatial} = 1.0$ and $\kappa_{color} = 3.0$). The scan line in which we are interested in the disparity estimation case is marked with small red strokes.

Discussion. The first point was selected inside an object on a region with a smooth texture, far away from the boundary of that object. In this case, one would expect that the better matching can be achieved with a larger window. It is confirmed by the experiments (Figures 2 and 4), showing that big values of t give more distinctive matching, whereas small values (for example, $t = 16.6$) produce equally high similarity values for many adjacent points.

The point on a boundary of the object was selected in such a way that the background of that object undergoes a severe change (Figures 3 and 5). This determines very low similarities for big values of t when the influence of the background is high due to the large window. This is especially noticeable with the geodesic weights which produce sharp edges in the similarity maps.

The underlying premise of geodesic weights is that colors or intensities can be used to distinguish between different objects. That is, the pixels in the neighborhood of x with a color similar to $u(x)$ have more influence on the similarity calculation than pixels with different colors (which probably belong to another object).

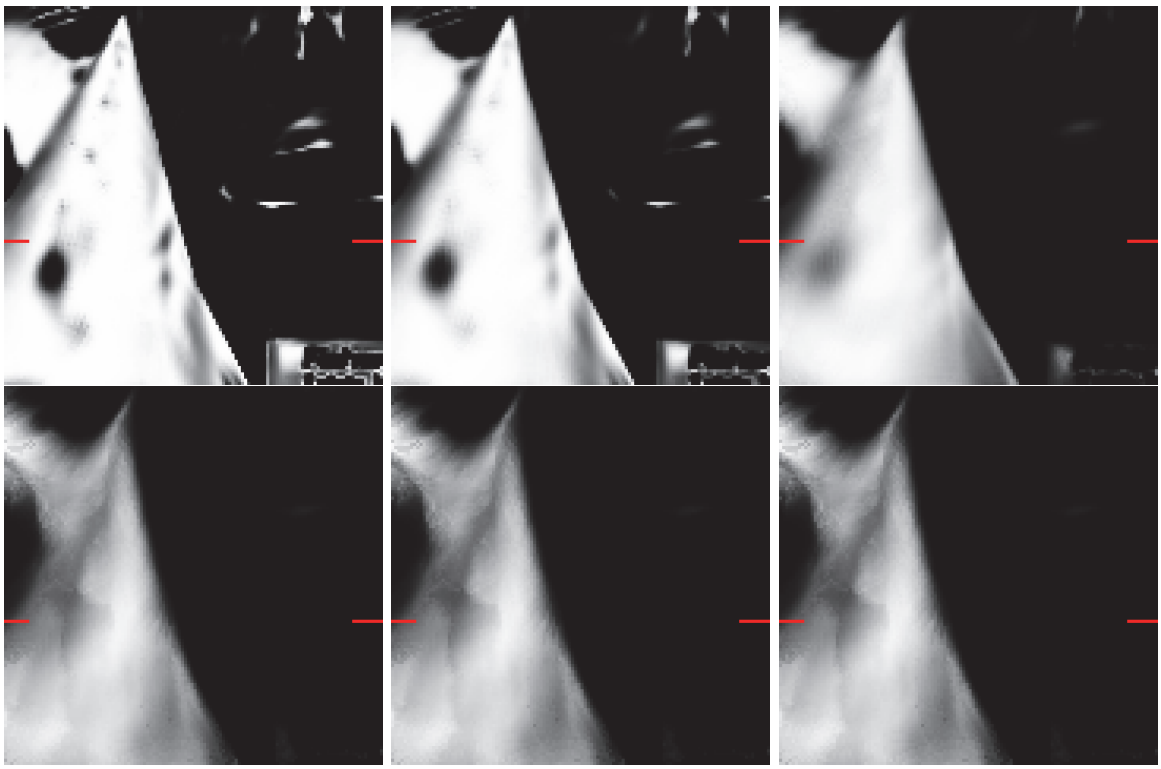

Figure 2. *Similarity maps for the point inside the object calculated using Gaussian weights* (4.15). *From left to right and top to bottom:* $t = 16.6$, $t = 25$, $t = 50$, $t = 150$, $t = 300$, *and* $t = 15000$.

Note that in the approximated geodesic distance (4.16) the influence of the image colors is controlled by κ_{color}. The geodesic weights on a region with similar color are virtually identical to the Gaussian weights. However, if the window contains two objects with different colors, then (for a reasonable choice of κ_{color}) only the pixels with a color similar to the center of the window have a significant contribution in the similarity calculation. Finally, by taking a very large value for κ_{color} one would recover the Gaussian weights (4.15) regardless of the image colors. The choice of κ_{color} is delicate, as it should be small to distinguish objects by their intensity but large enough to capture the variations of intensity within an object. In the context of a denoising application the value of this parameter is usually chosen to be proportional to the noise present in the image [34].

In stereo vision the geodesic weights are particularly important because matching using fixed weights produces the so-called *foreground fattening* effect. This phenomenon occurs when a matching window contains parts of objects with different depths. In this setting background pixels near an occluding edge may get the depth of the occluding edge (which is in the foreground); hence in the estimated depth map the foreground object appears fattened.

The geodesic weights correspond to a well-established stereo-vision technique by Yoon and Kweon [41] that allows one to estimate sharper depth maps near depth discontinuities using weights similar to the bilateral filter [34].

It is worth mentioning that, for the big values of t, points close to the tip of the cone

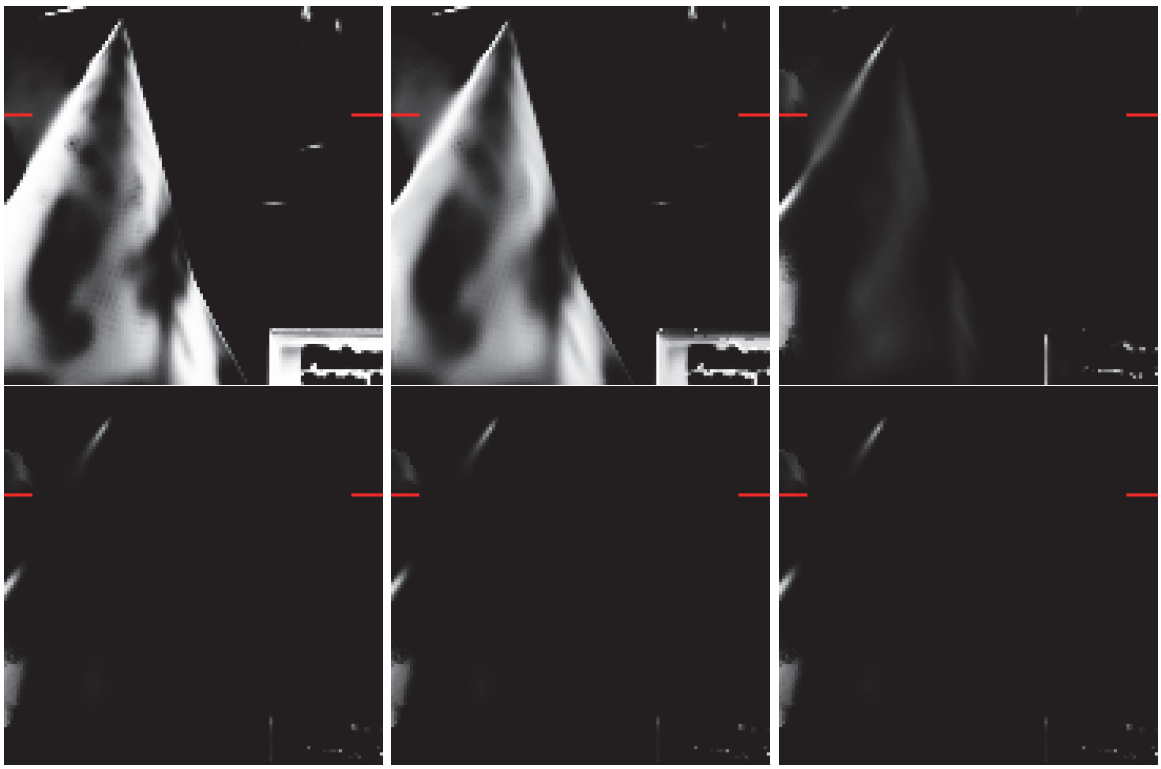

Figure 3. *Similarity maps for the point on the boundary calculated using Gaussian weights* (4.15). *From left to right and top to bottom:* $t = 16.6$, $t = 25$, $t = 50$, $t = 150$, $t = 300$, *and* $t = 15000$.

(Figures 3 and 5) have much higher similarity values than points on the scan line. This happens because in these locations the background has a more similar appearance. In applications other than stereo this might be a desirable best match.

Second experiment. In order to show the effect of changing t on the whole image, we calculate disparity maps for the stereo pair shown in Figure 6. Disparities were obtained by an exhaustive search for the best match in the range of possible offsets $x \in [-55, -5]$. For the second experiment the scale parameter was selected to be $t \in \{16.6, 25, 50, 150, 15000\}$. Parameter r is fixed as before: $r = 150$ (see section 5 for details). For approximated geodesic weights (see (4.16)), the coefficients were set to be $\kappa_{spatial} = 1.0$ and $\kappa_{color} = 3.0$.

Figure 7 shows the ground truth disparity map and the disparity maps computed at different scales using Gaussian weights (see (4.15)). Figure 8 shows the occlusion mask and the errors with respect to the ground truth excluding the occlusion mask. Similarly, Figures 9 and 10 show disparity maps and errors computed using approximated geodesic weights (see (4.16)).

Discussion. As can be seen, calculating similarity values on a coarse scale results in smooth disparity maps which closely match the ground truth within the objects but are oversmoothed at the boundaries. In contrast, fine scale similarities emphasize sharp boundaries of objects but produce a lot of mismatches on flat regions.

Figure 11 shows the closed-up fragment of error maps for the limit cases ($t = 15000$ and

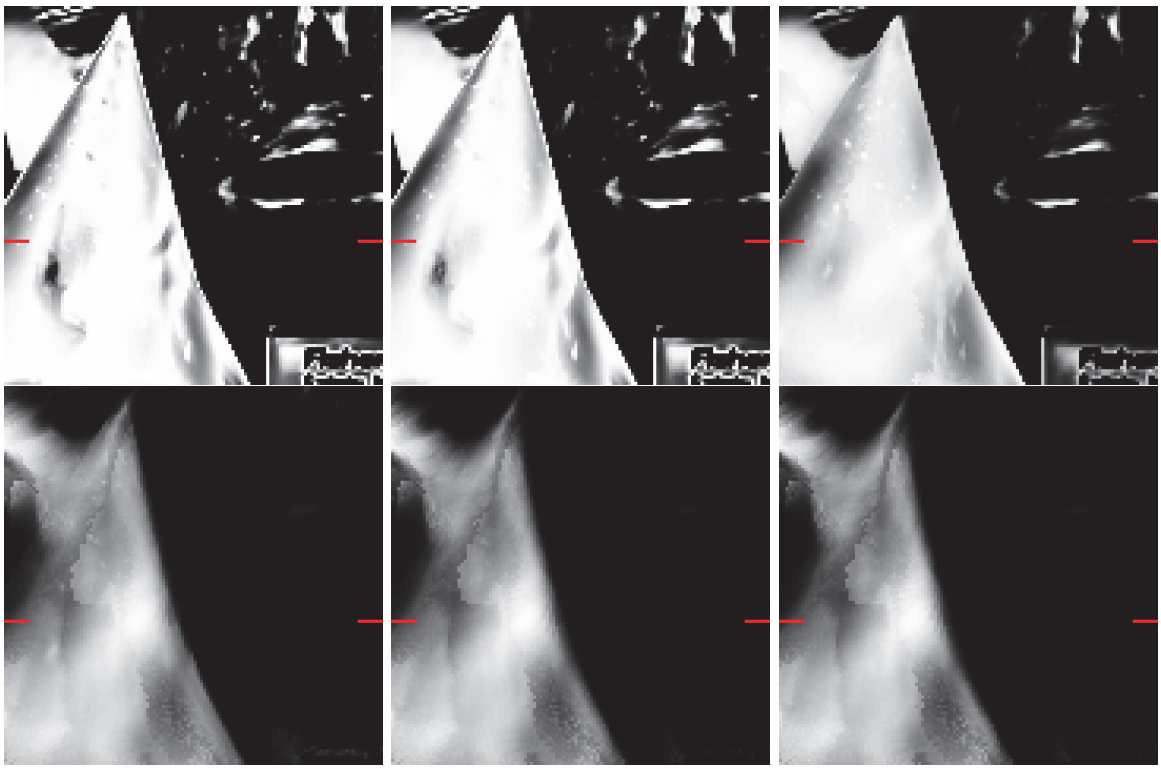

Figure 4. *Similarity maps for the point inside the object calculated using geodesic weights* (4.16). *From left to right and top to bottom:* $t = 16.6$, $t = 25$, $t = 50$, $t = 150$, $t = 300$, *and* $t = 15000$.

$t = 16.6$). Some of the interesting regions are highlighted in red. This confirms that in general, if a foreground object is moving over its background, the similarity measure performs differently at different scales within the object and close to its boundary. The way of "merging" similarity values from different scales in order to produce the best match may vary depending on the application. This issue is out of the scope of the current work and should be studied in detail in the future.

5. Affine covariant structure tensors as metrics on the image domain. In this section we address the construction of affine covariant structure tensors and affine covariant neighborhoods, and we discuss their properties. The affine structure tensor field will be used as an anisotropic metric on the image domain and will allow us to build an a priori connection map to properly compare two images.

As is well known, the structure tensor can be seen as a metric in the image plane [39, 40, 21, 6, 5, 28], and it has been used in image processing and computer vision in fields ranging from nonlinear filtering to motion analysis.

The computation of affine covariant tensors is closely related with the problem of estimating affine covariant regions. The latter has been addressed in the object recognition literature. In [25], the authors compute affine covariant regions on a set of points that are robust to scale changes. Given two matching points, the affine transformation is then estimated up to a ro-

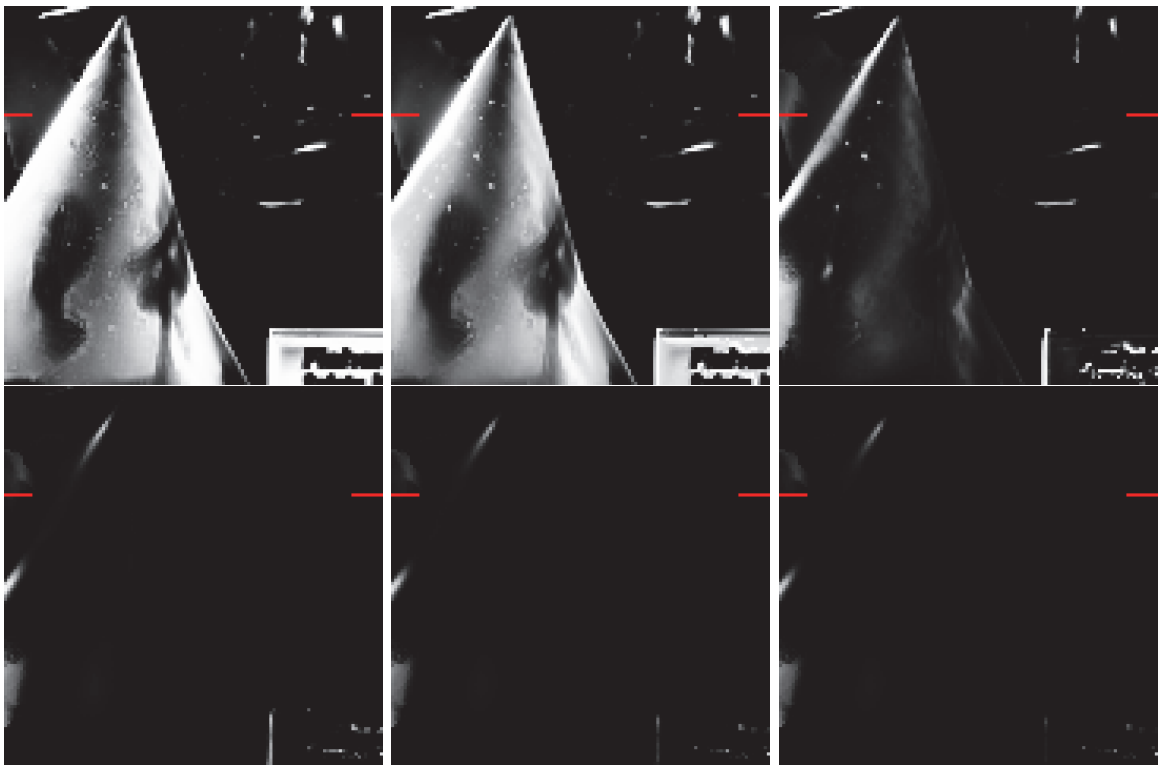

Figure 5. *Similarity maps for the point on the boundary calculated using geodesic weights* (4.16). *From left to right and top to bottom:* $t = 16.6$, $t = 25$, $t = 50$, $t = 150$, $t = 300$, *and* $t = 15000$.

Figure 6. *Stereo pair used for disparity map calculation.*

tation. In [23], the authors build up affine covariant domains referred to as maximally stable extremal regions (MSER). Mainly, MSER are defined as the most contrasted connected components of upper and lower level sets of the image [23]. Though these approaches give very good results in the object recognition context, they do not provide a dense set of regions with

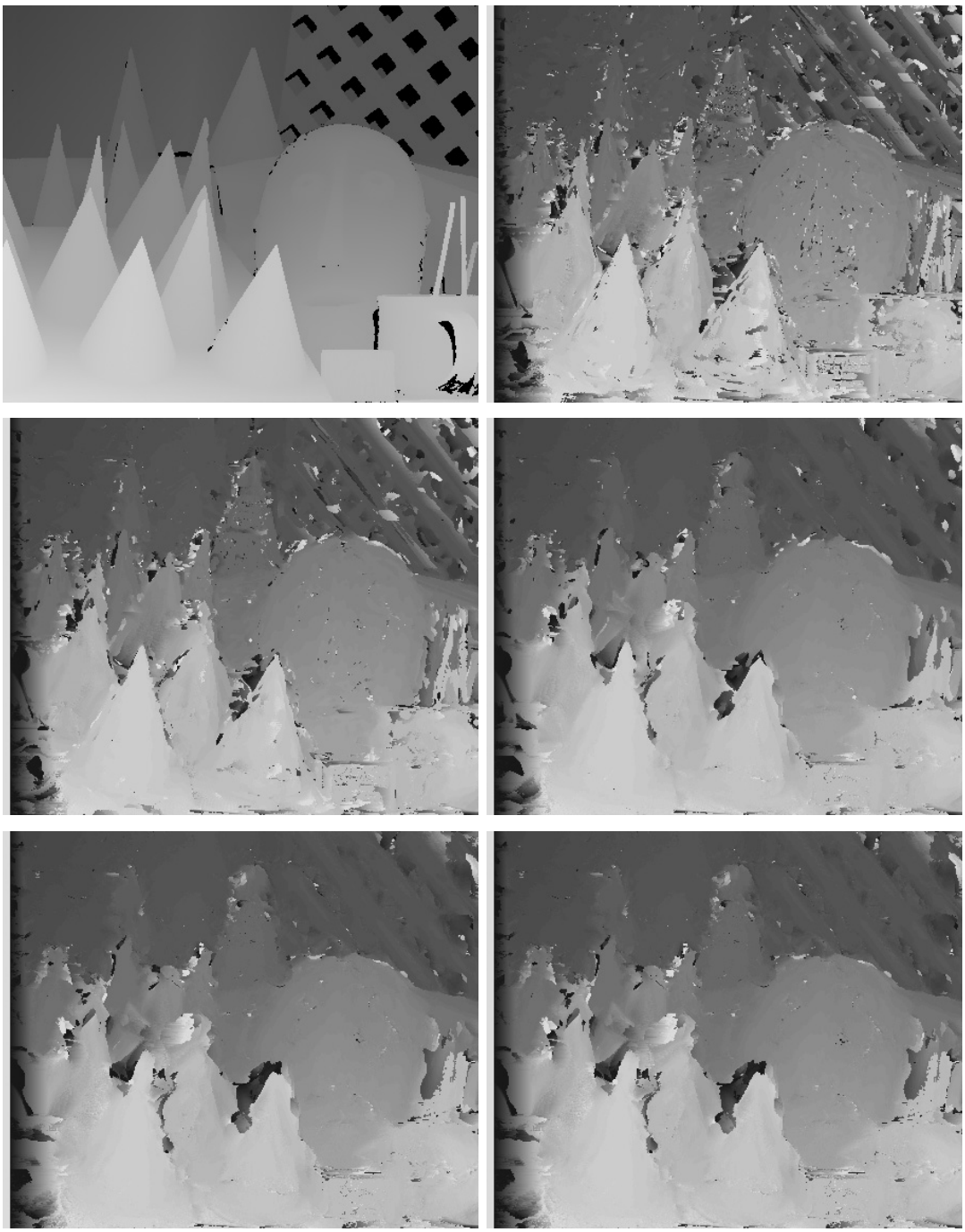

Figure 7. *Disparity maps calculated using Gaussian weights* (4.15). *Brighter colors correspond to bigger offsets. From left to right and top to bottom: ground truth, disparity maps for* $t = 16.6$, $t = 25$, $t = 50$, $t = 150$, *and* $t = 15000$.

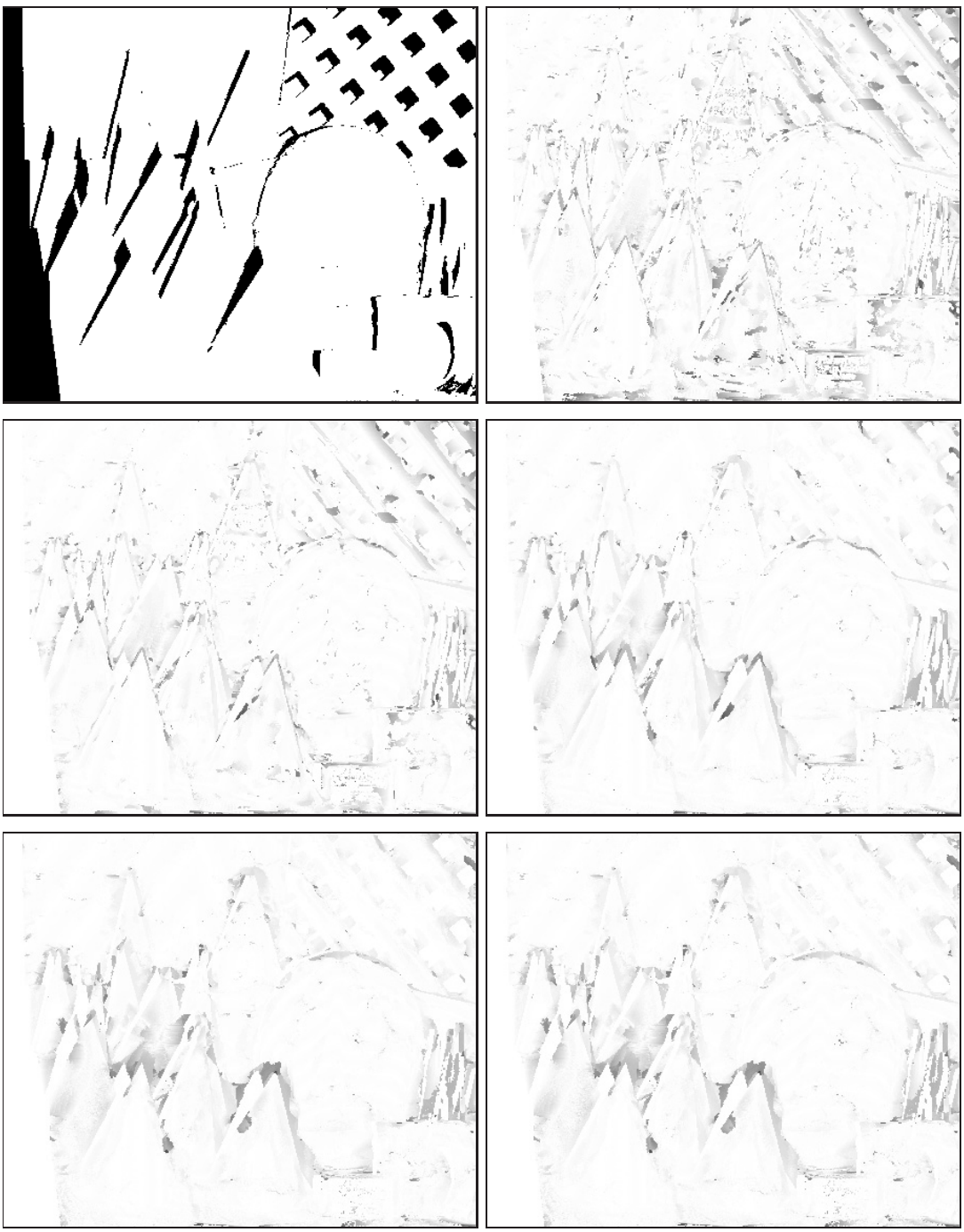

Figure 8. *Errors in disparities with respect to the ground truth excluding the occlusion mask. Disparities were calculated using Gaussian weights* (4.15)*. Darker colors correspond to bigger errors. From left to right and top to bottom: occlusion mask, error maps for* $t = 16.6$*,* $t = 25$*,* $t = 50$*,* $t = 150$*, and* $t = 15000$*.*

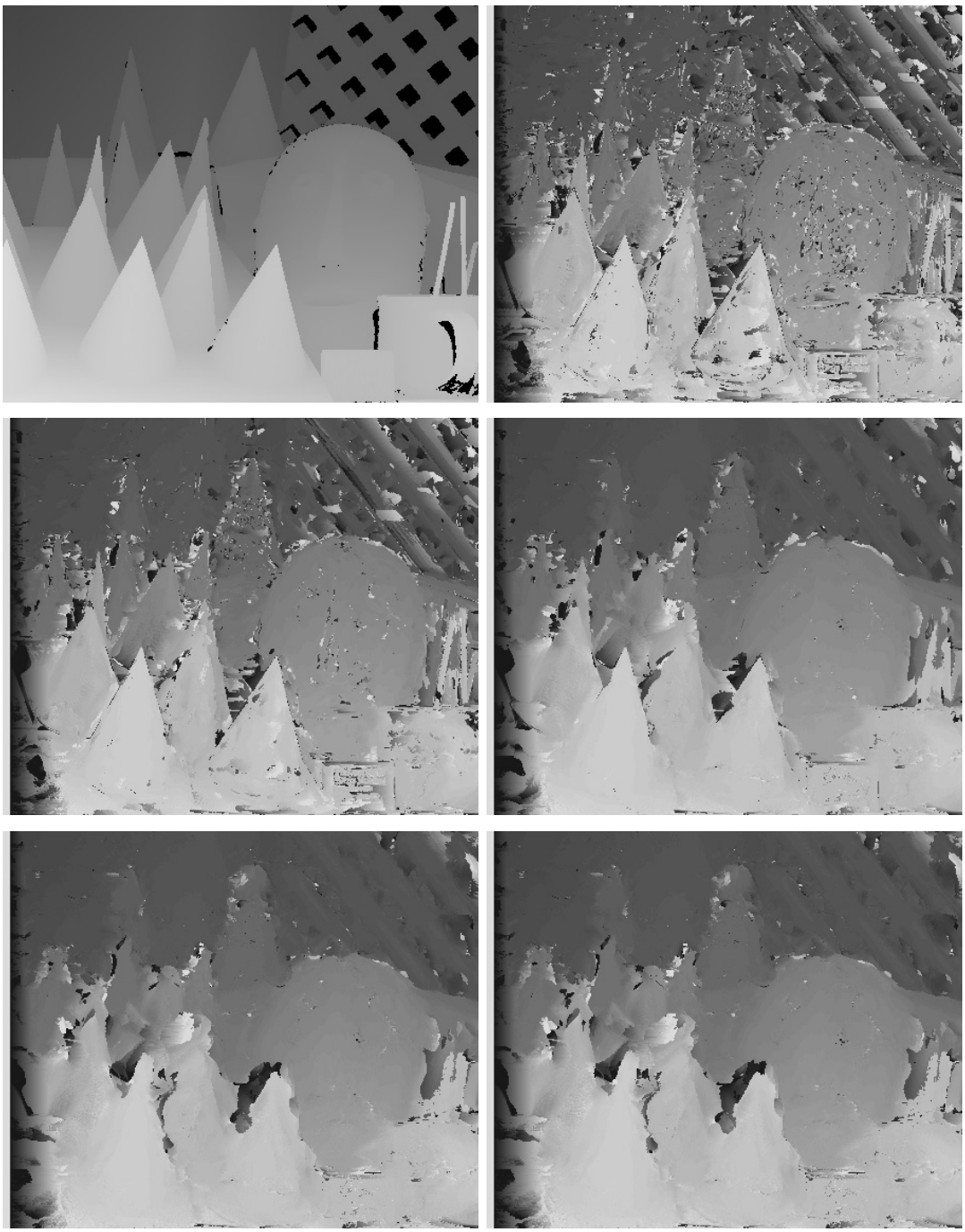

Figure 9. *Disparity maps calculated using geodesic weights* (4.16). *Brighter colors correspond to bigger offsets. From left to right and top to bottom: ground truth, disparity maps for* $t = 16.6$, $t = 25$, $t = 50$, $t = 150$, *and* $t = 15000$.

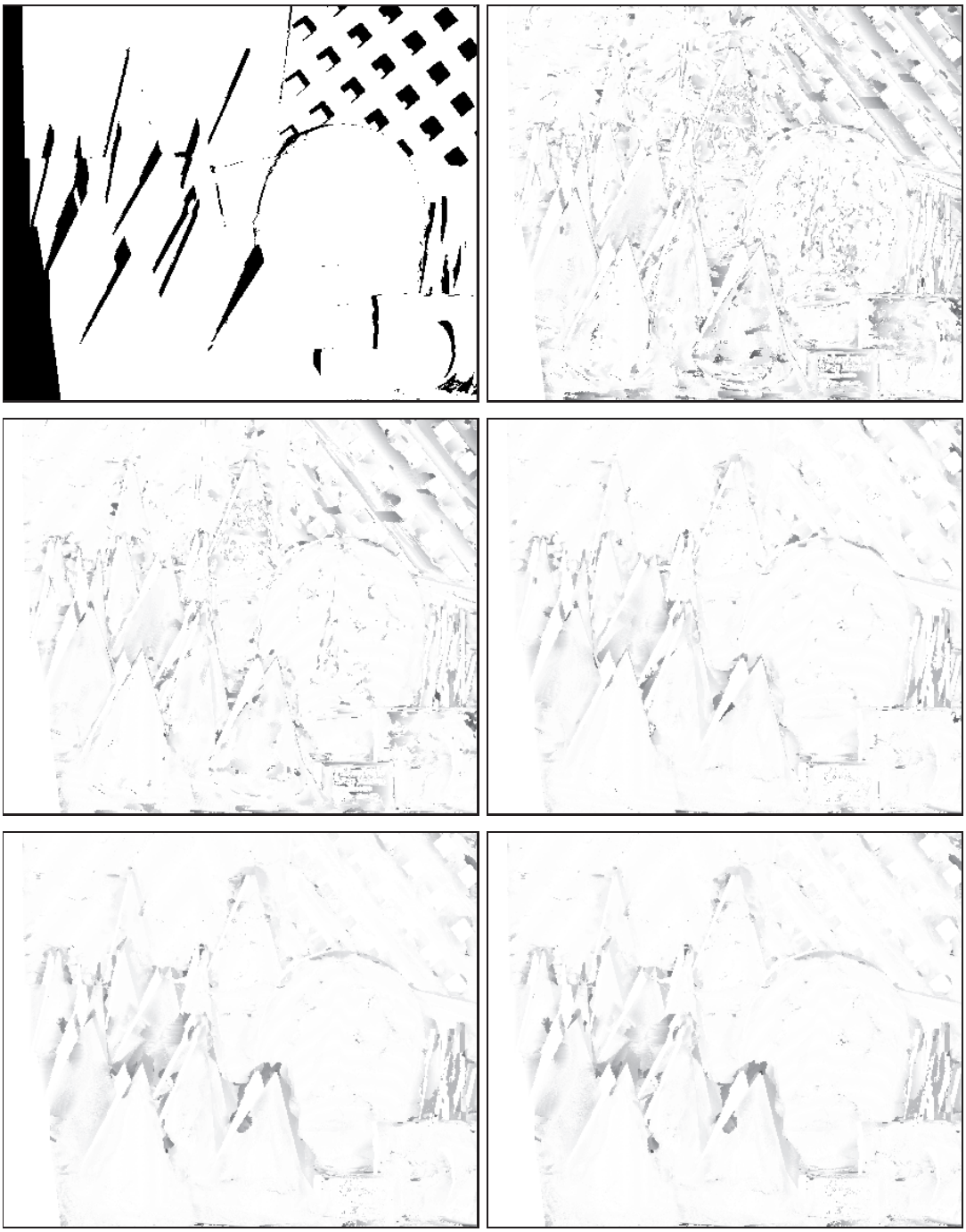

Figure 10. *Errors in disparities with respect to the ground truth excluding the occlusion mask. Disparities were calculated using geodesic weights* (4.16)*. Darker colors correspond to bigger errors. From left to right and top to bottom: occlusion mask, error maps for* $t = 16.6$*,* $t = 25$*,* $t = 50$*,* $t = 150$*, and* $t = 15000$*.*

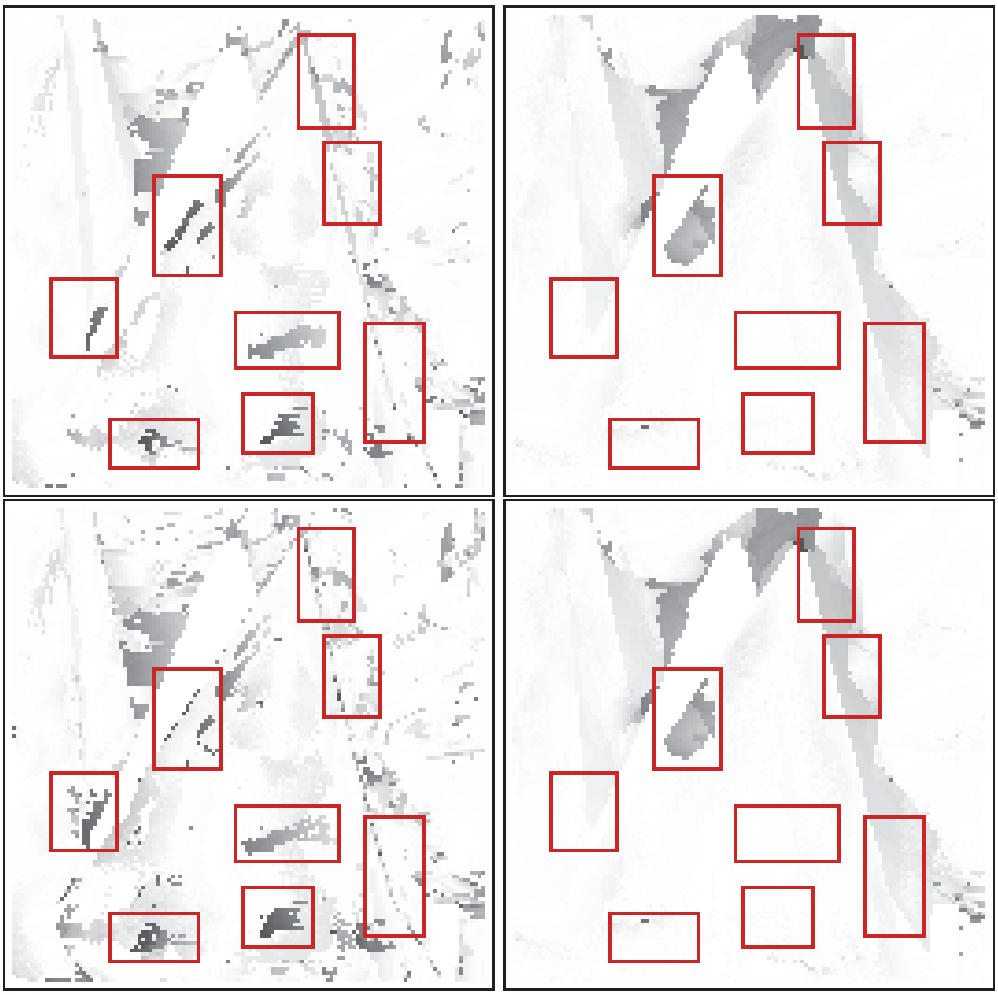

Figure 11. *Closed-up errors with respect to the ground truth excluding the occlusion mask. First row: calculated using Gaussian weights* (4.15) *for* $t = 16.6$, $t = 15000$. *Second row: calculated using geodesic weights* (4.16) *for* $t = 16.6$, $t = 15000$.

a guarantee of their affine covariance.

5.1. Construction of affine covariant structure tensors and neighborhoods. Let us start by describing the construction of structure tensors and their corresponding neighborhoods which are affine covariant.

Let u be a given image, $u : \mathbb{R}^N \to \mathbb{R}$. Let $GL(N)$ be the set of invertible matrices in $\mathbb{R}^N$. Let $A \in GL(N)$. Let us denote $u_A(x) := u(Ax)$.

Let us start by some examples. Assume that we have a metric g on $\mathbb{R}^N$. Let us remark that the map A induces a metric g_A in $\mathbb{R}^N$ such that $G_A(x) = A^t G(Ax)A$, where $G(y)$ denotes the symmetric matrix $(g_{ij}(y))$. This law of transformation holds also with the Hessian of an image u and the Hessian of u_A. Indeed, $D^2 u_A(x) = A^t D^2 u(Ax)A$. Let us remark that it is also the case when we consider the Hessian on u defined on the Riemannian manifold $(\mathcal{M}, g) =$

$(\mathbb{R}^N, g)$ and the Hessian of u_A on $(\mathcal{N}, \bar{g}) = (\mathbb{R}^N, g_A)$. Indeed, $D^2_{\mathcal{N}} u_A(x) = A^t D^2_{\mathcal{M}} u(Ax) A$. The following definition of affine covariant tensors associated with an image generalizes this property for a $(1,1)$ tensor field computed from an image.

Definition 5.1. *Let H_u be a $(1,1)$ tensor defined on $\mathbb{R}^N$ such that, for each $x \in \mathbb{R}^N$, it is represented by an $N \times N$ matrix $H_u(x)$ mapping a vector in $\mathbb{R}^N$ to another vector in $\mathbb{R}^N$. We say that H_u is an affine covariant tensor if it satisfies*

$$H_{u_A}(x) = A^t H_u(Ax) A, \tag{5.1}$$

where $u_A(x) := u(Ax)$ for $A \in GL(N)$.

Let us remark that it includes a slight abuse of terminology, as the term affine covariant tensor refers to our tensor computed from an image.

Another interesting example is the following. Let us consider $F(u) = Du \otimes Du$, where $\otimes$ denotes the tensor product. Then,

$$\begin{aligned} F(u_A)(x) &= Du_A(x) \otimes Du_A(x) = A^t Du(Ax) \otimes A^t Du(Ax) \\ &= A^t Du(Ax) \otimes Du(Ax) A = A^t F(u)(Ax) A. \end{aligned} \tag{5.2}$$

Thus, $F(u)$ is an affine covariant tensor.

This law of transformation is well adapted to define neighborhoods that transform well with respect to affine transformations. This is the object of Lemma 5.2. In Lemma 5.3 we will describe ways to generate affine covariant tensors based on an image by integrating in those affine covariant neighborhoods.

Lemma 5.2. *Let H_u be an affine covariant tensor. Let*

$$B_{H_u}(x,r) = \{y : \langle H_u(x)(y-x), (y-x)\rangle \le r^2\}, \qquad x \in \mathbb{R}^N,\ r > 0. \tag{5.3}$$

Then, $B_{H_{u_A}}(x,r) = A^{-1} B_{H_u}(Ax, r)$.

We will say that $B_{H_{u_A}}(x,r)$ is an *affine covariant neighborhood.*

Proof.

$$B_{H_{u_A}}(x,r) = \{y : \langle H_{u_A}(x)(y-x), (y-x)\rangle \le r^2\} = \{y : \langle A^t H_u(Ax) A(y-x), (y-x)\rangle \le r^2\}. \tag{5.4}$$

Let $\bar{x} = Ax$, $\bar{y} = Ay$. Then

$$B_{H_{u_A}}(x,r) = \{A^{-1}\bar{y} : \langle H_u(\bar{x})(\bar{y}-\bar{x}), (\bar{y}-\bar{x})\rangle \le r^2\} = A^{-1} B_{H_u}(Ax, r). \tag{5.5}$$ ∎

By Lemma 5.2, the set $W_{F_u}(x,r) = \{y : \langle F_u(x)(y-x), (y-x)\rangle \le r^2\}$ also satisfies $W_{F_{u_A}}(x,r) = A^{-1} W_{F_u}(Ax, r)$.

Let us remark that if we define

$$B_u(x,r) = \{y : |Du(x)(y-x)| \le r\}, \tag{5.6}$$

then

$$B_{u_A}(x,r) = \{y : |Du_A(x)(y-x)| \le r\} = A^{-1}\{y : y \in B_u(Ax, r)\} = A^{-1} B_u(Ax, r);$$

i.e., $B_{u_A}(x,r)$ is an affine covariant neighborhood as well.

At this point we have all the ingredients we need to describe the method of construction of affine covariant structure tensors and affine covariant neighborhoods. Let

$$T(u)(x) = \int_{B_u(x,r)} Du(y) \otimes Du(y)\, dy. \tag{5.7}$$

Then

$$T(u_A)(x) = \int_{B_{u_A}(x,r)} Du_A(y) \otimes Du_A(y)\, dy = \int_{A^{-1}B_u(Ax,r)} A^t Du(Ay) \otimes Du(Ay) A\, dy \tag{5.8}$$

and by writing $\bar{y} = Ay$, $y \in A^{-1}B_u(Ax,r)$, we get

$$T(u_A)(x) = A^t \int_{B_u(Ax,r)} Du(\bar{y}) \otimes Du(\bar{y}) |\det A|^{-1}\, d\bar{y} A. \tag{5.9}$$

That is, $T(u_A)(x) = |\det A|^{-1} A^t T(u)(Ax) A$.

Thus, $T(u_A)$ is an affine covariant tensor with a weight expressed by $|\det A|^{-1}$. We still refer to it as an affine covariant tensor density of exponent -1. To cancel the factor $|\det A|^{-1}$, we observe that $\mathrm{Area}(B_{u_A}(x,r)) = |\det A|^{-1}\mathrm{Area}(B_u(Ax,r))$. Therefore, if we define

$$NT(u)(x) = \frac{\int_{B_u(x,r)} Du(y) \otimes Du(y)\, dy}{\mathrm{Area}(B_u(x,r))}, \tag{5.10}$$

we have $NT(u_A)(x) = A^t NT(u)(Ax) A$. In other words, $NT(u)$ is an affine covariant tensor (or an affine covariant tensor density of exponent 0), computed on an affine covariant neighborhood.

Lemma 5.3. *Let H^1 be an affine covariant tensor density of exponent k ($k = 0, -1$), and let H^2 be an affine covariant tensor. Let H^i_A be the tensor after the affine transformation A. Let*

$$T(H^1, H^2)(x) = \int_{B_{H^1}(x,r)} H^2(y)\, dy. \tag{5.11}$$

Then

$$T(H^1_A, H^2_A)(x) = |\det A|^k A^t T(H^1, H^2)(Ax) A. \tag{5.12}$$

That is, $T(H^1, H^2)$ is an affine covariant tensor density of exponent k.

We have taken $k = 0, -1$ because we wanted to cover our examples. Other exponents could be taken as well.

Lemma 5.3 permits us to iterate the above construction (5.10) and redefine for $k \geq 2$

$$NT^{(k)}(u)(x) = \frac{\int_{B_{NT^{(k-1)}(u)}(x,r)} Du(y) \otimes Du(y)\, dy}{\mathrm{Area}(B_{NT^{(k-1)}(u)}(x,r))}, \tag{5.13}$$

where

$$B_{NT^{(k-1)}(u)}(x,r) = \{y : \langle NT^{(k-1)}(u)(x)(y-x),(y-x)\rangle \leq r^2\}. \tag{5.14}$$

Recall that for the initial iteration, $k=0$, the affine covariant neighborhood was given by

$$B_{NT^0(u)}(x,r) = B_u(x,r) = \{y : |Du(x)(y-x)| \leq r\}. \tag{5.15}$$

Let us remark that, when we iterate, the affine covariant neighborhood associated with $NT^k(u)$ does not depend as much as $NT(u)(x)$ on the gradient at the pixel x. The next remark is also motivated by this.

Remark 10. A variant of the above example is

$$\bar{B}_u(x,r) = \{y : |Du(y)(y-x)| \leq r\}. \tag{5.16}$$

Then $\bar{B}_{u_A}(x,r) = A^{-1}\bar{B}_u(Ax,r)$. It can be used to define the $(1,1)$ affine tensor density of exponent -1,

$$\bar{T}(u)(x) = \int_{\bar{B}_u(x,r)} Du(y)\otimes Du(y)\,dy, \tag{5.17}$$

or its normalized version as above that would be an $(1,1)$ affine covariant tensor.

Given an image $u:\mathbb{R}^2\to\mathbb{R}$, consider the affine covariant tensors $NT^{(k)}(u)(x)$ and affine covariant neighborhoods $B_{NT^{(k)}(u)}(x,r)$ for $x\in\mathbb{R}^2$, $k\in\mathbb{N}$, $r>0$. We have empirically observed that after a few iterations this procedure cycles over a finite number of affine covariant tensors (typically 1 to 3). In any case, for x in the image domain, any of these tensors is guaranteed to be an affine covariant tensor. Therefore, the purpose of the iterative algorithm is not to ensure affine covariance, but to diminish dependency on the very first iteration, which depends solely on a single gradient and thus is very sensitive to noise. It can be clarified as follows. Given an image u, we calculate the tensors at every point x with a different number of iterations k. We then calculate the Frobenius norms of the differences $\left\|N_T^{(k)}(u)(x) - N_T^{(k+1)}(u)(x)\right\|_F$ for every two consecutive values of k. Finally, we average the norms over all x. Figure 12 shows these average changes over k for three selected images. The fact that the values go to some nonzero value is explained by the occasional alternation between two affine covariant tensors. This behavior is typical for all tested images.

For the rest of the paper we will denote by $T_u(x)$ the affine covariant structure tensor $NT^{(k)}(u)(x)$ for a fixed value of k (in our experiments, we take $k=30$) and a given value of r. We will say that T_u is the affine covariant structure tensor field associated with u. Similarly we use the notation $B_{T_u}(x)$ to refer to the affine covariant neighborhood $B_{NT^{(k)}(u)}(x,r)$. They are ellipses in the case $N=2$. In Figure 13 we show some of these neighborhoods.

Let us note that r is a free parameter. It controls the size of the affine covariant neighborhood at a given point. On the other hand, the size of the neighborhood is also affected by the texture in the vicinity of that point. In this work r was experimentally chosen for each image and we observed that it may vary over a relatively wide range without any significant impact on the final result.

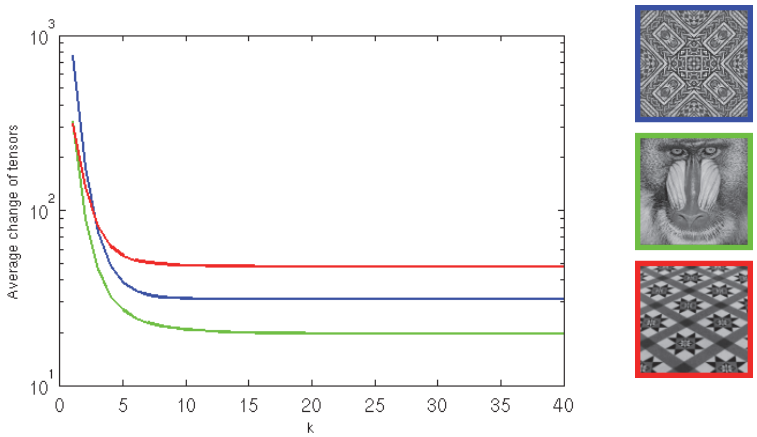

Figure 12. *Average change of tensors from one iteration to another, depending on the number of iterations. Thumbnails of the corresponding images are shown on the right.*

Recall that the initial iteration takes into account nonlocal information due to its infinite band support. In section 5.4 we study the dependency of tensors on the initial iteration (5.15) for different values of r. In section 6.1 we demonstrate a few similarity maps obtained for different values of r. The analysis of the appropriate choice of r, with application to different contexts, will be the object of future work.

Remark 11. Let us make an additional remark on the solutions of equations of the form (4.6). Note that if $A \in GL(N)$ and g is a metric on $\mathbb{R}^N$ defined as above by the structure tensor associated with a given image, $u : \mathbb{R}^N \to \mathbb{R}$, then, as previously noticed, $A : \mathbb{R}^N \to (\mathbb{R}^N, g)$ induces a metric g_A in $\mathbb{R}^N$ such that $G_A(x) = A^t G(Ax)A$. If $\widetilde{\Gamma}$ is the Levi-Civita connection in the metric g_A, then

$$\widetilde{\Gamma}(A^t p) = A^t \Gamma(p) A.$$

Let $(\mathcal{M}, g) = (\mathbb{R}^N, g)$, $(\widetilde{\mathcal{M}}, g) = (\mathbb{R}^N, g_A)$. Therefore, the equation

$$u_t = \mathrm{Trace}_g \left(D^2_{\mathcal{M}} u \right),$$

where $\mathrm{Tr}_g(A) = \mathrm{Tr}(G^{-1}A) = g^{ij}A_{ij}$, is affine invariant. Indeed, since

$$\mathrm{Tr}_{g_A}\left(D^2_{\widetilde{\mathcal{M}}} u_A \right) = \mathrm{Tr}\left(G_A^{-1} D^2_{\widetilde{\mathcal{M}}} u_A \right) = \mathrm{Tr}\left(A^{-1} G^{-1} A^{-t} A^t D^2_{\mathcal{M}} u A \right)$$

$$= \mathrm{Tr}\left(G^{-1} D^2_{\mathcal{M}} u A A^{-1} \right) = \mathrm{Tr}\left(G^{-1} D^2_{\mathcal{M}} u \right) = \mathrm{Tr}_g \left(D^2_{\mathcal{M}} u \right),$$

then

$$\frac{\partial}{\partial t} u_A(t,x) = \frac{\partial}{\partial t} u(t, Ax) = \mathrm{Tr}_g \left(D^2_{\mathcal{M}} u \right)(t, Ax) = \mathrm{Tr}_{g_A}\left(D^2_{\widetilde{\mathcal{M}}} u_A \right)(t,x).$$

That is, u_A solves the same equation as $u(t,x)$ with initial condition $u(Ax)$. This gives an example of a linear multiscale analysis that is affine invariant.

Notice that the metric depends on the initial condition. This may not be obvious from the notation above, but the equation is applied to a given image, and the metric is constructed from the initial condition. This guarantees that whenever for $u(0,x)$ the metric is g, for $u(0,Ax)$ the metric is $g_A(x) = A^t g(Ax) A$. For instance, this is true for the structure tensors.

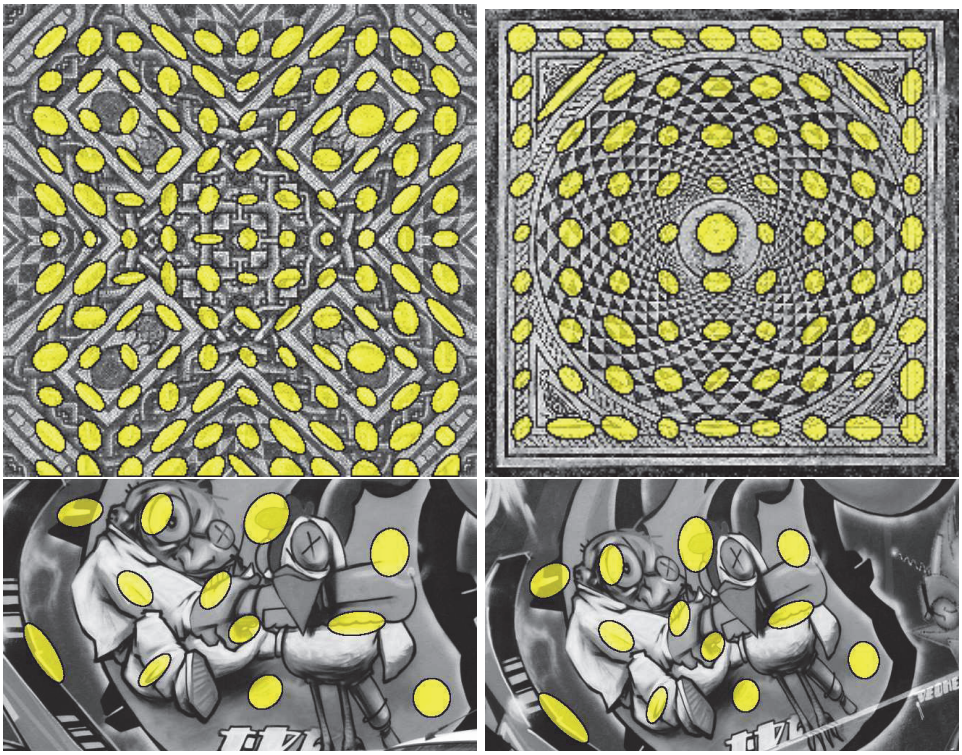

Figure 13. *Images showing (superimposed) some affine neighborhoods. First row: neighborhoods are placed every* 25 *pixels. Second row: neighborhoods are placed at the corresponding points.*

5.2. A priori connection from tensors. Assume that we have two given images $u = I_u : \Omega_u \to \mathbb{R}$, $v = I_v : \Omega_v \to \mathbb{R}$. For each $x \in \Omega_u$, let $T_u(x)$ be the structure tensor of u at x. For each $y \in \Omega_v$, let $T_v(y)$ be the structure tensor of v at y.

We consider (with a previous extension of u and v to $\mathbb{R}^N$ first by an even extension and then by periodicity) the two manifolds $(\mathbb{R}^N, G^1 = T_u(x))$, $(\mathbb{R}^N, G^2 = T_v(y))$ and the PDE

$$\frac{\partial C}{\partial t} = \Delta_{\mathcal{M}x} C + 2\mathrm{Tr}(G^2(y)^{-1/2} G^1(x)^{-1/2} D_{xy} C) + \Delta_{\mathcal{M}y} C. \tag{5.18}$$

Let us clarify for this case the construction of an a priori connection for the approximating formulas (4.15) and (4.16).

For this, let us diagonalize the tensors matrices $T_u(x) = U_u(x) D_u(x) U_u^t(x)$ and $T_v(y) = U_v(y) D_v(y) U_v^t(y)$. Here

$$D_u(x) = \mathrm{diag}(\lambda_{u,1}(x), \lambda_{u,2}(x)), \qquad \lambda_{u,1}(x) \geq \lambda_{u,2}(x),$$

$$D_v(y) = \mathrm{diag}(\lambda_{v,1}(y), \lambda_{v,2}(y)), \qquad \lambda_{v,1}(y) \geq \lambda_{v,2}(y).$$

The matrices $U_u(x), U_v(y)$ are rotation matrices formed with the eigenvectors of $T_u(x), T_v(y)$, respectively. Let $e_{u,i}(x)$ be the eigenvector of $T_u(x)$ associated with the eigenvalue $\lambda_{u,i}(x)$, $i \in \{1, 2\}$. Let $e_{v,i}(y)$ be the eigenvector of $T_v(y)$ associated with the eigenvalue $\lambda_{v,i}(y)$. That is, $e_{u,i}(x)$ is the ith column of $U_u(x)$ and $e_{v,i}(y)$ is the ith column of $U_v(y)$.

Recall that each structure tensor can be described by its corresponding neighborhood, which is an ellipse given by

$$B_{T_u}(x) = \{\bar{x} : \langle T_u(x)(\bar{x} - x), \bar{x} - x\rangle \leq r^2\},$$
$$B_{T_v}(y) = \{\bar{y} : \langle T_v(y)(\bar{y} - y), \bar{y} - y\rangle \leq r^2\}.$$

If we let $A(x) := D_u(x)^{1/2}U_u(x)^t$, then by the change of variables $X = A(x)x'$ we have $A(x)\frac{e_{u,i}(x)}{\sqrt{\lambda_{u,i}(x)}} = f_i$, where f_i is a Euclidean orthonormal basis. This means that we are rotating the ellipse, aligning the minor axis to f_1 and the major to f_2, and changing the length of both axes. Similarly, for the ellipse associated with $T_v(y)$ we let $B(y) := D_v(y)^{1/2}U_v(y)^t$, and by the change of variables $Y = B(y)y'$ we have $B(y)\frac{e_{v,i}(y)}{\sqrt{\lambda_{v,i}(y)}} = f_i$. After these operations both ellipses are transformed to a standard circle of radius r, and hence we can compare them.

Let us remark that

$$P(x, y) = B(y)^{-1}A(x) = U_v(y)D_v(y)^{-1/2}D_u(x)^{1/2}U_u(x)^t \tag{5.19}$$

is an a priori connection.

Similarly to [18] it can be shown that formula (5.19) allows one to obtain exact affinity from the affine covariant structure tensors with an additional constraint of prohibiting vertical disparity as usual, for example, for the stereo case. In the general case, however, the affinity is determined only up to an orthogonal transformation. This result is shown in Lemma 5.4.

We denote by R any orthogonal matrix in $\mathbb{R}^2$ and denote $P_R(x, y) = T_v(y)^{-1/2}\, R\, T_u(x)^{1/2}$. The following result holds.

Lemma 5.4. *Let us assume $u(z) = v(Az)\ \forall z \in \mathbb{R}^N$ for some $A \in GL(N)$. Then, $P_R(x, y) = A$ for $y = Ax$ and some orthogonal matrix R.*

Proof. Consider $y = Ax$. Proving that $A = P_R(x, y) = T_v(y)^{-1/2}RT_u(x)^{1/2}$ is equivalent to proving that $T_v(y)^{1/2}AT_u(x)^{-1/2}$ is an orthogonal matrix.

But T_v is an affine covariant tensor. Therefore, $T_u(x) = A^tT_v(y)A$. This identity is equivalent to $I = \left(T_v(y)^{1/2}AT_u(x)^{-1/2}\right)^t \left(T_v(y)^{1/2}AT_u(x)^{-1/2}\right)$. That is, $T_v(y)^{1/2}AT_u(x)^{-1/2}$ is an orthogonal matrix. ∎

This means that a truly affine invariant a priori connection is defined as

$$P(x, y) = B(y)^{-1}R(x, y)A(x) = U_v(y)D_v(y)^{-1/2}R(x, y)D_u(x)^{1/2}U_u(x)^t, \tag{5.20}$$

where $R(x, y)$ is some additional orthogonal transformation. Due to the fact that $U_u(x)$ and $U_v(y)$ are rotations and $D_u(x)^{1/2}$, $D_v(y)^{1/2}$ are scalings, there is no reflection involved in these transformation and, therefore, $R(x, y)$ is a rotation. For formulas (4.15) and (4.16) we can embed $R(x, y)$ into $A(x)$ or $B(y)$. Let us note that in practical applications this additional rotation $R(x, y)$ can be determined, for example, by estimating dominant orientations within the normalized circles.

5.3. Tensors on real images. Real images acquired with a digital camera are affected by the optical blur (which we assume to be Gaussian) and the sampling. In this section we will analyze to what extent the affine covariance properties of the tensors hold for real discrete

images. For simplicity, we consider planar images defined on $\mathbb{R}^2$. Furthermore, as in [27], we assume an affine camera model; i.e., we disregard perspective effects.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be an infinite resolution image on a plane, seen from a frontal view. We consider an image $u : \mathbb{R}^2 \to \mathbb{R}$, resulting from looking at f from a different viewpoint. Under an affine camera, we can express u as follows:

$$u(x) = f_A(x) = f(Ax),$$

where $A \in GL(N)$ is an affinity matrix associated with the viewpoint (without loss of generality, we omit the translation of the affinity). Throughout this section we will use the notation $f_A = Af$; i.e., we use the same symbol for the affinity matrix A as for the operator that warps an image in accordance with affinity A. Since the normalized structure tensor is a $(1, 1)$ affine tensor, tensors computed at corresponding locations on f and u should be related through formula (5.1).

Both u, f are infinite resolution images, seen from two different viewpoints before acquisition. When acquired by a finite resolution camera the image is modified by a filtering operation G_{Σ_1} and a sampling operator S_1. We assume the G_{Σ_1} to be a Gaussian kernel with covariance matrix Σ_1, such that its width is the smallest one that allows sampling without aliasing, with a sampling step of 1. We assume that $\Sigma_1 = \sigma_1 I$ as defined in [27]. Thus, after acquisition we have $\hat{u}_1 = S_1 G_{\Sigma_1} * Af$ and $\hat{f}_1 = S_1 G_{\Sigma_1} * f$. Let us ignore for the moment the effect of the sampling to focus on the blur. We denote by $u_1 = G_{\Sigma_1} * Af$ and $f_1 = G_{\Sigma_1} * f$ the images before sampling.

In general the warping by A and the Gaussian filter will not commute; thus after acquisition, u_1 and f_1 will no longer be related by an affinity. Indeed, if G_Σ is a Gaussian kernel with covariance matrix Σ and zero mean, then we have that

$$G_\Sigma * Af = AG_{A\Sigma A^T} * f,$$

a property which is sometimes referred to as weak commutativity [27].

Thus, in the affine camera model, a change of viewpoint associated with the affinity A induces an anti-aliasing Gaussian filter $G_{A\Sigma_1 A^T}$. Tensors computed in $u_1 = AG_{A\Sigma_1 A^T} * f_1$ will match tensors in $G_{A\Sigma_1 A^T} * f_1$, but they will differ from the tensors of $G_{\Sigma_1} * f$.

An exception is given by the case in which the Gaussian kernel is isotropic (as G_{Σ_1}) and $A = R$ is a rotation since in that case $R\sigma^2 I R^T = \sigma^2 I$. However, this is not true if the affinities contain scalings and/or tilts. In the following, we describe an experiment to quantify the effect of zooms on the tensors by comparing tensors computed on $G_{A\Sigma_1 A^T} * f$ and $G_{\Sigma_1} * f$ for several scalings.

For a zoom by a factor of $s > 0$, $A = H_s = sI$, we have that the induced Gaussian filtering is $H_s \Sigma_1 H_s^T = s^2 \Sigma_1 = (s\sigma_1)^2 I$. To quantify the effect of zooms in the computation of the tensors, we simulate acquisitions at several scales by creating a Gaussian scale-space $G_{s^2\Sigma_1} * f$ for the classical image *baboon* while varying $s \in [1, 10]$ and computing tensors at each scale. Since we do not have access to an infinite resolution image, we assume that $f_1 = S_1 G_1 * f$ is well sampled and build the scale-space by filtering f_1. This is an approximation to a Gaussian filtering of the continuous f, especially for large kernel widths.

Tensors are computed for each image $G_{s^2\Sigma_1} * f$ for a fixed parameter $r > 0$. Let us denote by $T_s(x)$ the tensors computed at x from image $G_{s^2\Sigma_1} * f$, and let us denote $\lambda_{s,2}(x)$ its

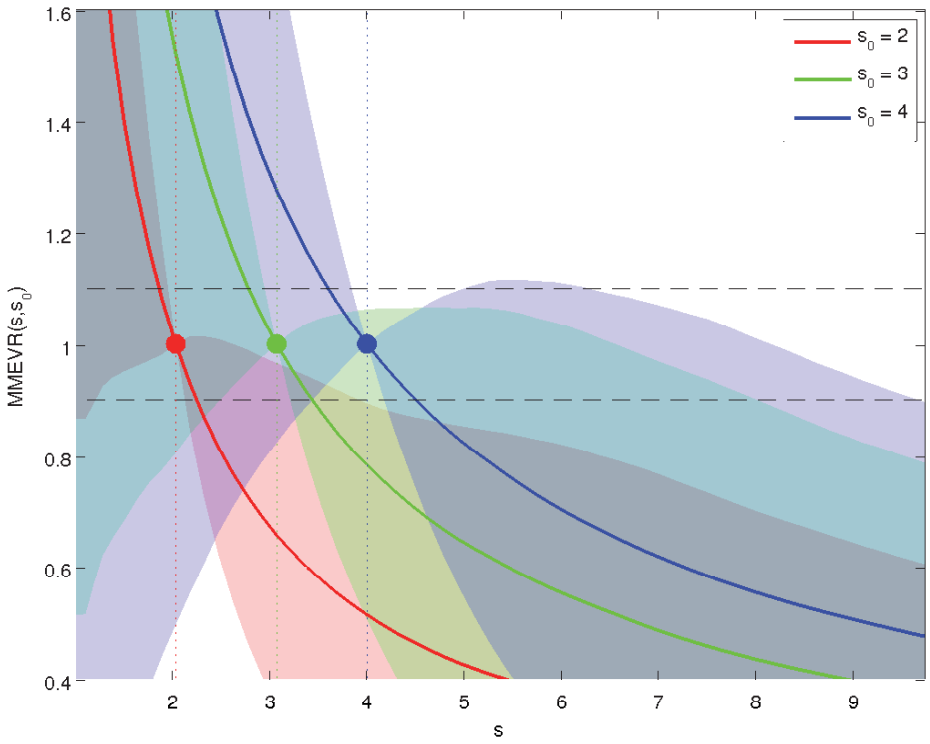

Figure 14. *Median minimum eigenvalue ratio, between tensors computed at scale s and at scale s_0, as a function of s. Plots show the evolution of the median eigenvalue ratio with respect to the scale s for three values of s_0, namely $s_0 = 2, 3$, and 4, indicated by the vertical dotted lines. The horizontal dotted lines show the range 1 ± 0.2, corresponding to a variation of 20% of the median ratio at s_0. The shaded areas show the region around the median of three standard deviations.*

smallest eigenvalue. Figure 14 plots the median in the image of the ratio between the smallest eigenvalue of a tensor computed at scale s and the corresponding tensor at scale s_0 (median minimum eigenvalue ratio):

$$\mathrm{MMEVR}(s, s_0) = \mathrm{median} \left\{ \sqrt{\frac{\lambda_{s,2}(x)}{\lambda_{s_0,2}(x)}} : x \in \Omega \right\}.$$

Three curves are shown, corresponding to $s_0 = 2, 3$, and 4. Together with the median value, we show the region around the median of three standard deviations (due to the presence of outliers, the standard deviation is estimated as 1.4826 MAD, the median absolute deviation).

Note that if we represent a tensor by the elliptical neighborhood $\langle y, T_s(x)y \rangle < 1$, the smallest eigenvalue is inversely proportional to the largest axis of the ellipse. The experiment measures how fast tensors change with the Gaussian filtering associated with a scale change.

The curves decrease monotonically. The reason for this is that as we zoom out (increase s) the image is filtered with a wider kernel, thus decreasing the high frequency content in the image. The gradients will be smaller, and thus the smallest eigenvalue of the tensors will also

be smaller. We show in the figure the range corresponding to a 20% variation with respect to 1, the value of the ratio in s_0. We could consider that, on average, the elliptical regions defined by the tensors change 20% in their size.

Note also that there is a considerable amount of dispersion around the median value. Tensors at different pixels have different behavior with scale, depending on the image in the neighborhood of the pixel.

5.4. Dependency on the initial iteration. The proposed construction method of the affine covariant tensors is an iterative method which starts from some initial region (5.15). The initial region is an infinite band whose orientation depends on the gradient at the central point. Therefore, the resulting tensor depends somehow on all the pixels within the initial band. In general this dependency tends to vanish with a sufficient number of iterations of the proposed scheme. In this section we study how this behavior depends on the image content and the value of r.

In order to study the dependency of the tensors on the initial region we select a highly textured image and create a second image by replacing the peripheral part of the first image with another content (Figure 15). Let $T_u(x)$ and $T_v(x)$ be tensors on the first and second images, respectively. We then calculate the error between two corresponding tensors as $e(x) = \|T_u(x) - T_v(x)\|_F^2$, where $\|\cdot\|_F^2$ is the Frobenius norm. Figure 16 shows the errors, color-coded by $c(x) = \exp(-\frac{e(x)^2}{2\sigma^2})$, where σ is in the order of 10^4.

For rather small values of r (50, 100, 150) many covariant neighborhoods (especially close to strong edges) degenerate into a small set of pixels or even into one pixel (Figure 17). With such small neighborhoods, the tensor estimation tends to converge to an incorrect solution. Obviously, in this case tensors highly depend on the initial iteration, which explains a large number of mismatches in the central part of the images for small r. Moreover, it explains occasional low errors on the peripheral part of the images, where tensors close to strong edges occasionally capture the same direction of these edges, but the corresponding neighborhoods do not have enough information to distinguish between two different images.

With the value of r large enough, dependency of tensors on the initial iteration is negligible in the whole central region. Let us note that tensors near the peripheral region are inevitably influenced by this region in all the iterations and not only during the initial one. Therefore they are irrelevant for this experiment. In section 6.2, we study the proposed similarity measure on a boundary between two objects, undergoing different transformations.

5.5. Related approaches to affine neighborhoods. The proposed iterative method for the computation of affine covariant structure tensors allows us to estimate the a priori connection between two given points and thus a local affinity. Other methods, suitable for the local affinity estimation, can be found in the literature. Our approach for the computation of affine invariant tensors and neighborhoods has much in common with the iterative shape adaptation algorithm of Mikolajczyk and Schmid [25] used in the Harris-Affine feature detector. In this section we compare both approaches in their ability to estimate an affine transformation. We use an implementation of the Harris-Affine feature detector of [38].

The method of [25] was originally proposed in the feature extraction context for object recognition purposes. It starts by building a scale-space for a given image and detecting stable keypoints together with their characteristic scales. Oriented elliptical neighborhoods are then

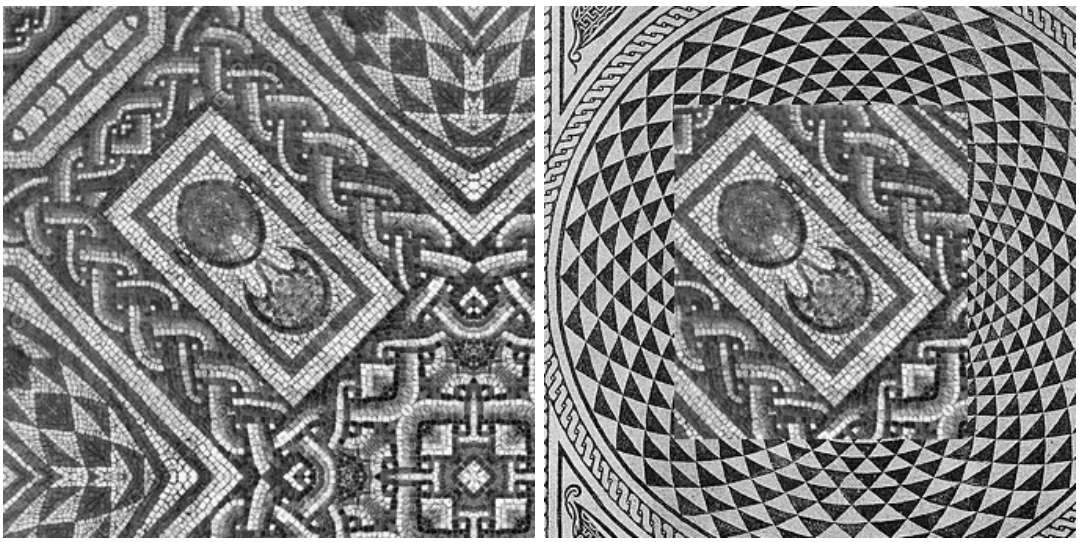

Figure 15. *Two images with identical central parts and different peripheral parts.*

calculated at these keypoints using an iterative procedure. First, a so-called shape adaptation matrix is estimated in the vicinity of a keypoint, and then the image is warped in accordance with this matrix. The amount of information around a keypoint that is considered in the calculation of the shape adaptation matrix is controlled by the integration scale σ_I. Both σ_I and the position of the keypoint are updated in each iteration. The process is repeated until a convergence criterion is met. In the feature detection context, the integration scale σ_I is related to the characteristic scale of a keypoint.

Like the proposed affine covariant structure tensor, the shape adaptation matrix can be used to estimate an affine transformation that aligns the vicinities of two given points. However, for arbitrary points that are not keypoints, one needs to provide σ_I as a free parameter. In this case parameter σ_I can be seen as radius parameter r in our approach. Also, positions of the points should be fixed. Let us note that the method described in [25] is initialized on a circular window given by an isotropic Gaussian which is not affine covariant, whereas the method this paper proposes starts from an affine covariant initial band and guarantees that at any stage of the algorithm the computed tensor is affine covariant.

In order to compare both methods we use an image sequence well known in the feature detection community, the *graffiti* sequence, taken from the test data in [24]. It contains six images showing different views of the same scene. Five ground truth global transformations from the first image to all the others are known. We uniformly sample 2745 points on the first image and use the ground truth transformations to obtain the sets of corresponding points for the other images. We then estimate local affinities for every pair of corresponding points using both methods. Knowing the locations of the corresponding points, we obtain estimated global transformations from local ones. We measure the estimation error by the Frobenius norm of the difference between the estimated and the ground truth transformations. Let us note once again that in both methods the sizes of neighborhoods are controlled by the free parameters: r in our method and σ_I in the Harris-Affine one. However, the values of r and σ_I

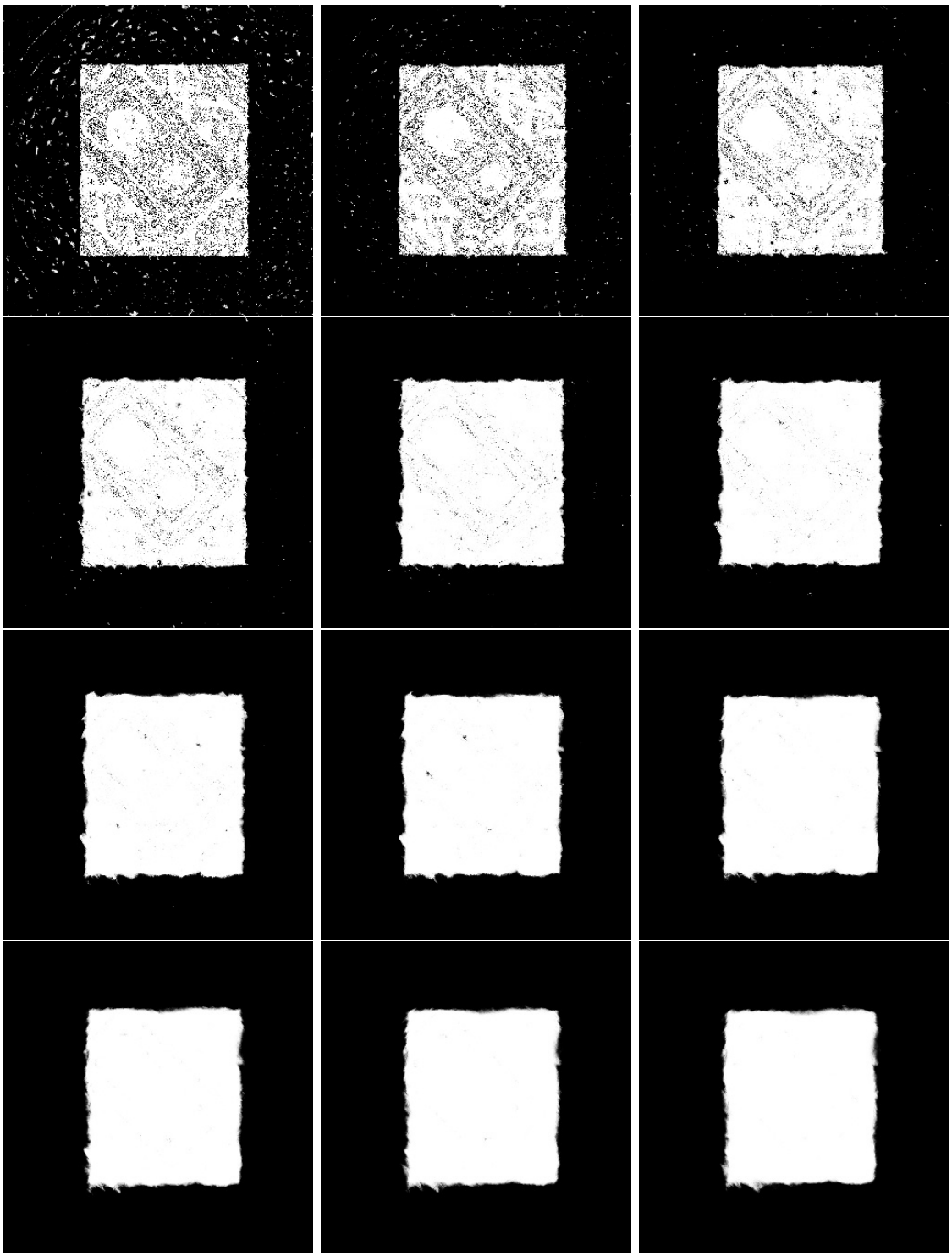

Figure 16. *Errors between corresponding tensors for different values of r. Darker color means bigger error. From left to right and top to bottom: $r = 50$, $r = 100$, $r = 150$, $r = 200$, $r = 250$, $r = 300$, $r = 350$, $r = 400$, $r = 450$. $r = 500$, $r = 550$, and $r = 600$.*

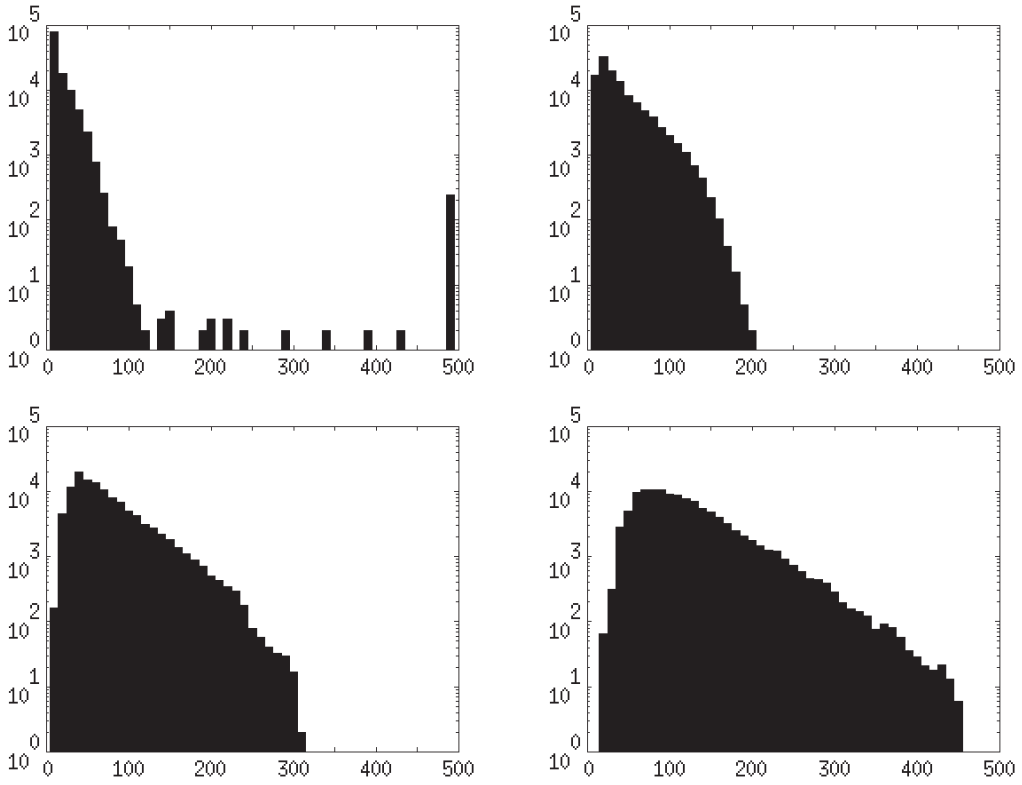

Figure 17. *Histograms of the ellipse sizes for different values of r. Vertical axis is in logarithmic scale; last bar accumulates all larger values. From left to right and top to bottom: $r = 50$, $r = 100$, $r = 150$, $r = 200$.*

could not be easily related. Therefore, for each pair of images we test extensive ranges of r and σ_I and select the ones giving the smallest median value of the error. With this experiment we indirectly evaluate the degree of affine covariance of the proposed covariant structure tensors and affine normalization proposed in [25]. Figure 18 shows statistics for both methods over the five pairs of images. It can be seen that statistically our method performs slightly better.

Let us note that other approaches to affine neighborhoods exist; see, e.g., [4, 36, 23, 26]. However, when applying these methods it is impractical or even impossible to extract affine invariant regions densely. In contrast, our approach by design is capable of producing a dense field of affine invariant neighborhoods.

6. Experimental results. In section 4.3, the proposed affine invariant similarity measure has been studied and experimentally analyzed without taking into account the computation scheme for affine covariant tensors. In this section we consider the whole approach and demonstrate a few more results on the similarity measure. For all the experiments in this section, to compare two points x and y we first calculate the affine covariant tensors and neighborhoods at these points using (5.13), (5.14), and (5.15). We transform the neighborhoods to normalized circles and align these circles by exhaustively searching for a missing rotation. This gives us the a priori connection (5.20) between x and y and allows us to apply the formula (4.15)

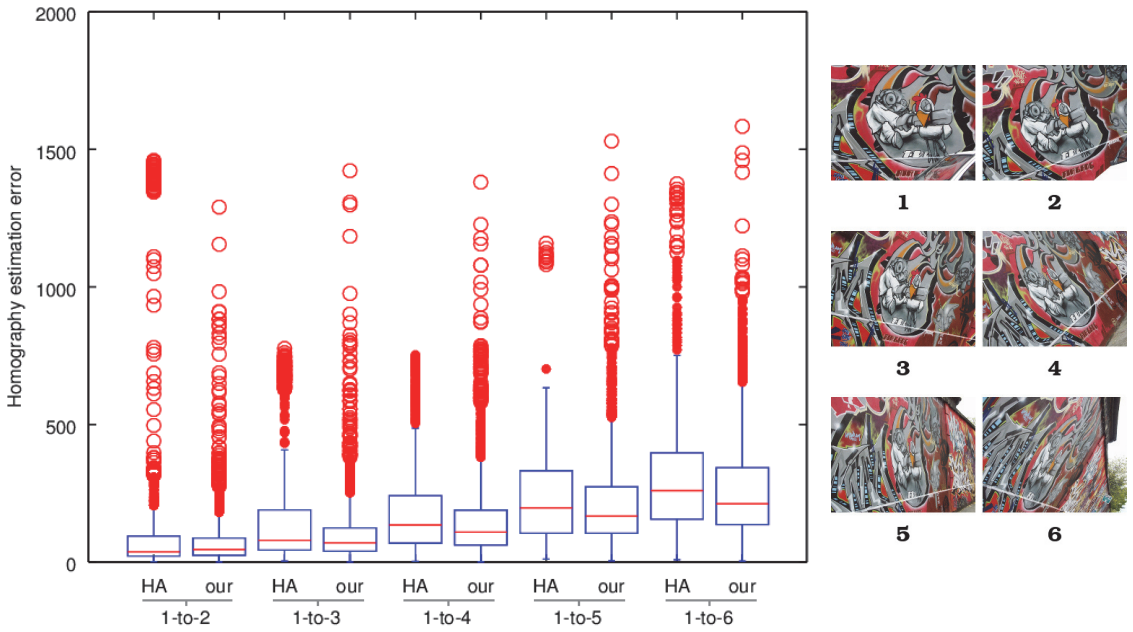

Figure 18. *Boxplot showing statistics of homography estimation error for* 2745 *uniformly sampled points. Homography is estimated for the first image and all the others (labeled as "*1*-to-*2*," "*1*-to-*3*," and so on). Estimation using shape adaptation matrix of Harris-Affine feature detector (labeled as "HA") is compared with our approach using structure tensors (labeled as "our"). As usual, boxes represent the first, second, and third quartiles, the length of the whiskers corresponds to* 1.5 *interquartile range (IQR), and red crosses are outliers. Thumbnails of the images being used are shown on the right.*

to calculate patch similarity (distance) value.

6.1. Similarity measure under different transformations. At first we study the proposed similarity measure between pairs of images, which undergo different kinds of transformations. For this purpose we calculate patch distance values $d(x, y)$ between corresponding points on two images. In each selected pair of images, differing by an affine or planar projective transform, the ground truth correspondences are either given (Figure 19) or easy to estimate (Figures 21 and 23). In the following experiments we demonstrate results for a rotation, an affine transformation, and finally a perspective transformation. In each case we compare the proposed similarity measure with the well-known Euclidean distance given by (1.1). Let us remark that it can be seen as a particular case of (4.15), where $T_u(x) = I$ and $T_v(y) = I$. This similarity measure is modeled by a usual patch distance between square patches of a given size.

In order to visualize similarity maps, we color-code patch distances $d(x, y)$ by $c(x, y) = \exp(-\frac{(d(x,y)-d_{min})^2}{2\sigma^2})$, where $\sigma = \frac{d_{max}-d_{min}}{\gamma}$, d_{max}, and d_{min} are the maximum and minimum patch distance values, respectively, and $\gamma > 0$ is a parameter that controls stretching of values close to 1. Brighter color means higher similarity value. Note that for each pair of images distances were normalized all together using the same d_{min} and d_{max}, which means that equal colors on two different similarity maps correspond to the same similarity values.

In the described setting there is only one object shown on a pair of images under different

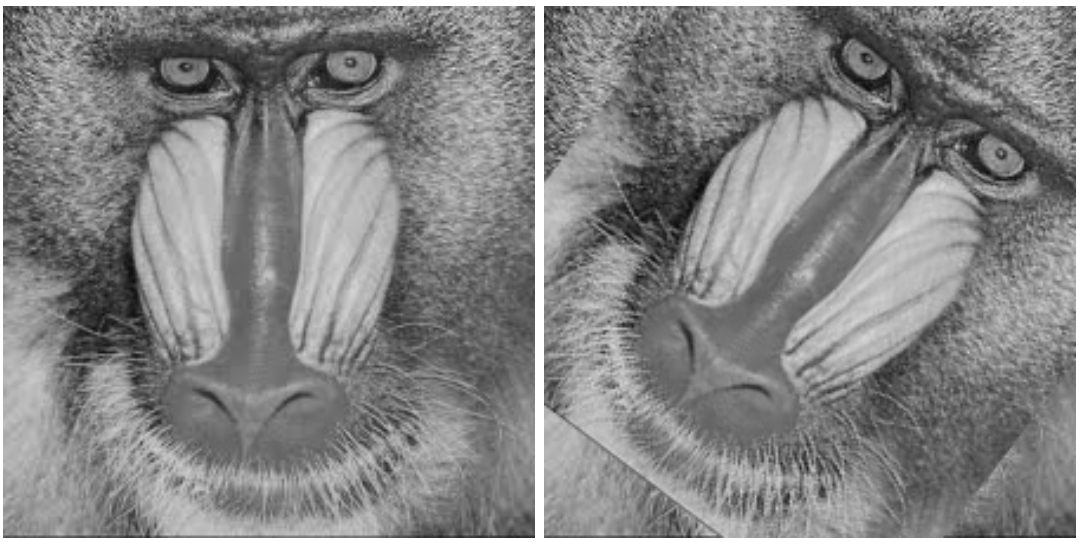

Figure 19. *Original image and rotation of the original image by an angle of* 37 *degrees superimposed on the original.*

transformations; moreover, the ground truth correspondences are provided. Therefore, for the following experiments we fix the scale parameter $t = 150000$ (which corresponds to almost uniform weights within the covariant neighborhoods). We fix $r = 150$ as well.

First experiment. In the first experiment we study rotation on a pair of images shown in the Figure 19. The left image is the original one, and the right image was rotated clockwise by an angle of 37 degrees. Figure 20 shows similarity values calculated from the given correspondences. The distances were color-coded with $\sigma = \frac{d_{max}-d_{min}}{10}$.

Second experiment. For the second experiment we use a pair of images (Figure 21) related by an affine transform. The left image is the fronto-parallel view, and the right image is the affinely rectified view of the same scene. Figure 22 shows similarity values calculated from the given correspondences. The distances were color-coded with $\sigma = \frac{d_{max}-d_{min}}{20}$.

Third experiment. The third pair of images (Figure 23) includes two different views of the same packet of juice. We are interested in the front side of the packet, which is planar. Two given views of that front side are related by perspective transform of a plane. Figure 24 shows similarity values calculated from the given correspondences. The distances were color-coded with $\sigma = \frac{d_{max}-d_{min}}{10}$.

Varying r. Finally, let us present some similarity maps calculated for the same pairs of images but this time varying the r parameter (Figures 25, 26, and 27). These results show that r may vary in a relatively wide range without significant impact on the final result.

The actual choice of r should depend on the application, the sizes of the image features to be captured, and the image noise and texture content themselves. As shown in section 5.4, r should also be big enough to avoid degenerate neighborhoods (reduced to a few pixels). However, together with the previous three experiments, the similarity maps on Figures 25, 26, and 27 verify that for the whole range of r being tested the proposed similarity measure outperforms the usual Euclidean distance (1.1).

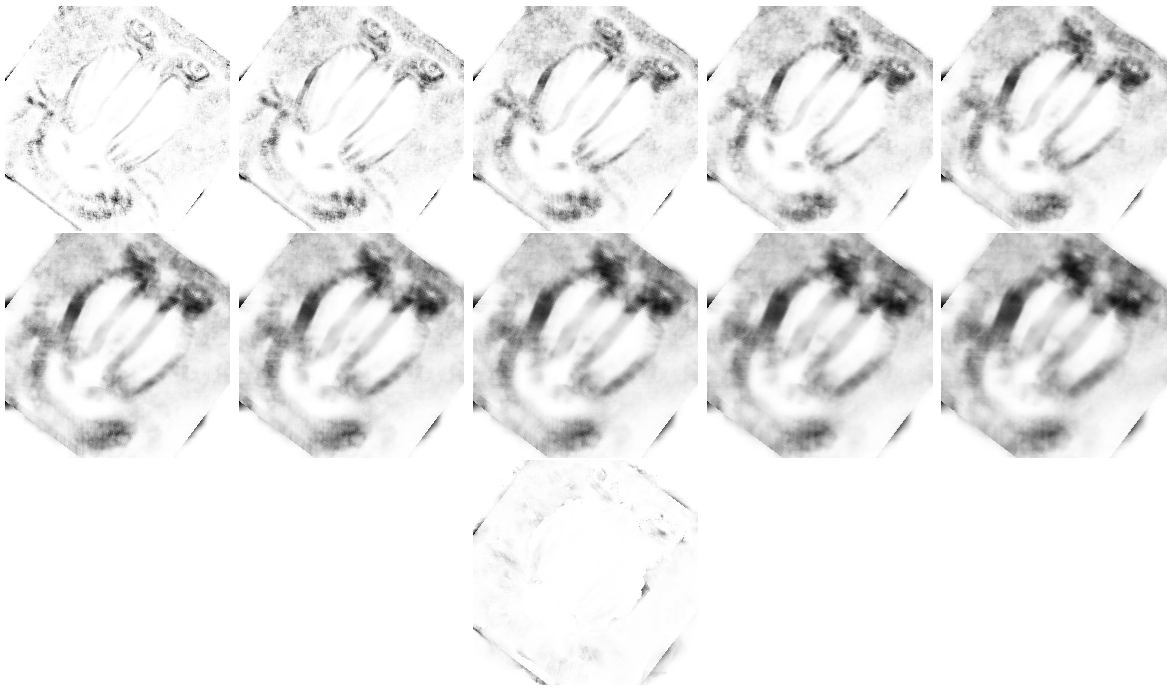

Figure 20. *Similarity maps for given correspondences between the original and rotated images. First row, left to right: similarities calculated using square patches of sizes* 5, 7, 9, 11, 13. *Second row, left to right: similarities calculated using square patches of sizes* 15, 17, 19, 21, 23. *Third row: similarities calculated using the proposed similarity measure.*

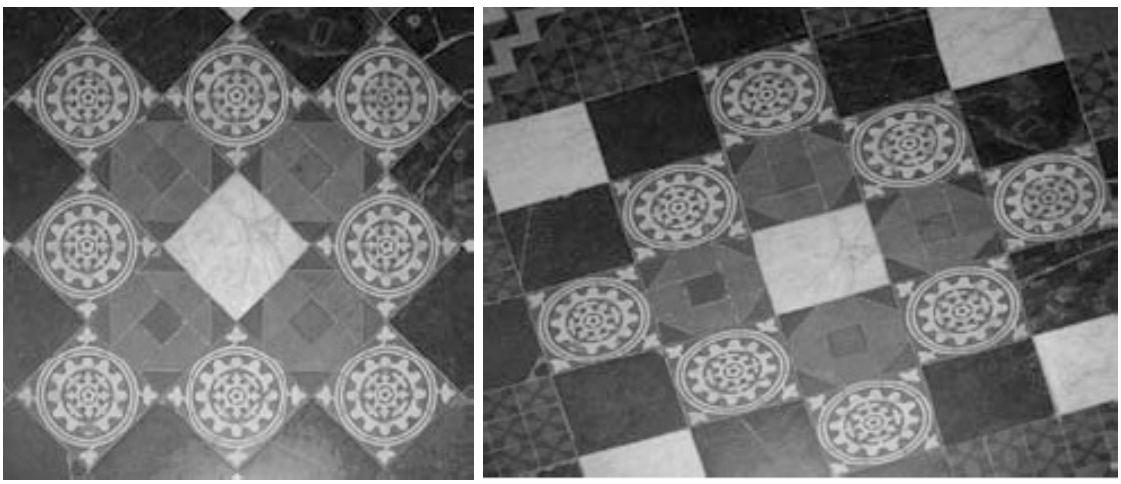

Figure 21. *Fronto-parallel and affinely rectified views of the same scene.*

6.2. Local affinities and discussion on the selection of parameters. In this section we study the behavior of the proposed similarity measure close to the boundary of two different transformations. For doing this we created a pair of synthetic images, shown in Figure 28. On these images two different views of the *graffiti* image are used as backgrounds and two

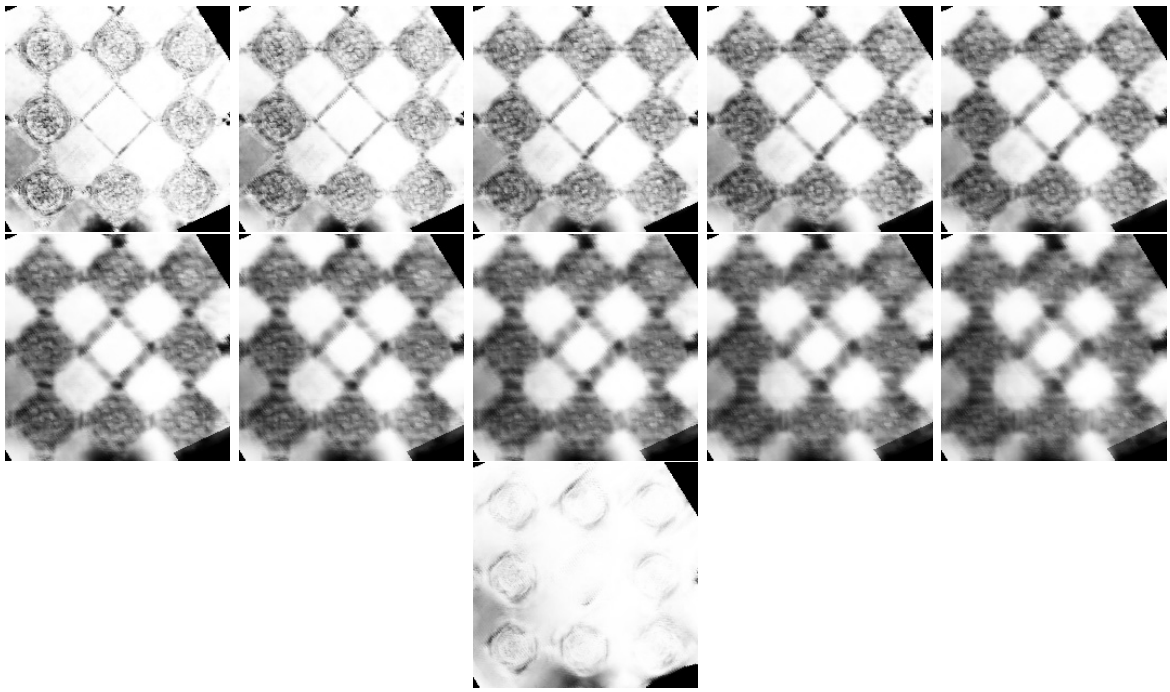

Figure 22. *Similarity maps for given correspondences between the images related by an affinity. First row, left to right: similarities calculated using square patches of sizes* 5, 7, 9, 11, 13. *Second row, left to right: similarities calculated using square patches of sizes* 15, 17, 19, 21, 23. *Third row: similarities calculated using the proposed similarity measure.*

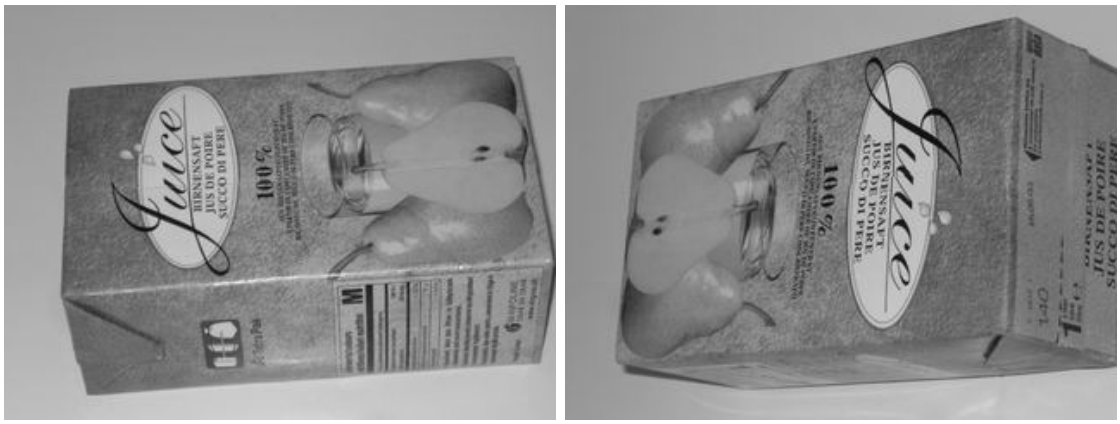

Figure 23. *Two views of the same packet of juice.*

different views of another object are imposed as foreground objects. Backgrounds are related by an affine transformation, and foreground objects are related by another affine transformation. The ground truth correspondences are calculated first for the background, then for the foreground objects, and finally combined using a mask.

There are two main parameters in this work related to the amount of support we allow in

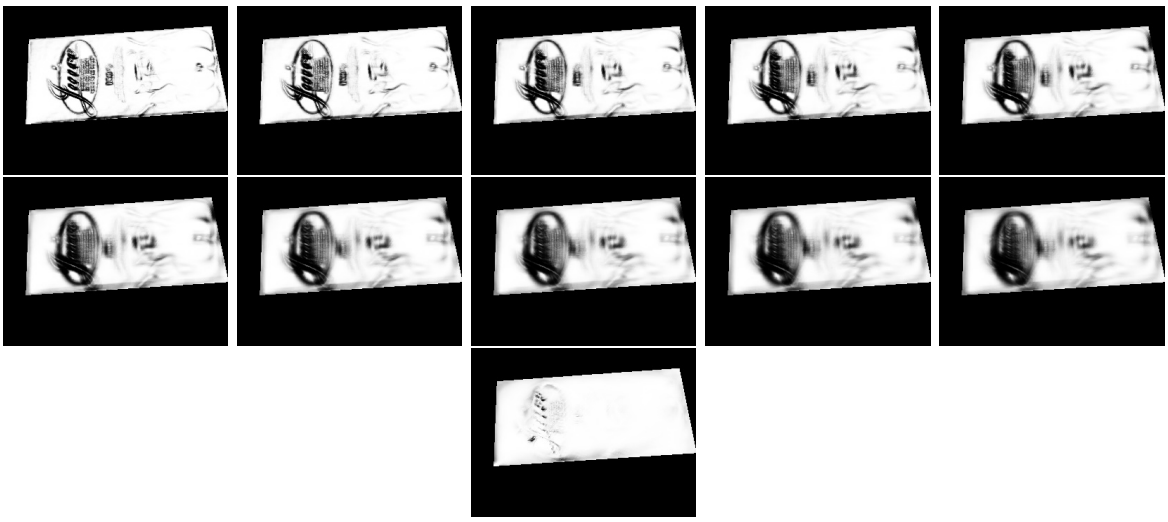

Figure 24. *Similarity maps for given correspondences between two views of the front side of the packet of juice. First row, left to right: similarities calculated using square patches of sizes* 5, 7, 9, 11, 13. *Second row, left to right: similarities calculated using square patches of sizes* 15, 17, 19, 21, 23. *Third row: similarities calculated using the proposed similarity measure.*

Figure 25. *Similarity values for given correspondences calculated using the proposed similarity measure with different values of* r. *First row, left to right:* $r = 50$, $r = 75$, $r = 100$, $r = 150$, $r = 175$. *Second row, left to right:* $r = 200$, $r = 225$, $r = 250$, $r = 275$, $r = 300$.

the comparison of two patches: t and r. In the following experiment both t and r are variable; therefore it is useful to link them together. Let $t = r/\hat{t}$. Small values of $\hat{t}$ correspond to coarse scales (larger window), while big values correspond to fine scales (smaller window). Recall that in (4.15), g_t is a given window that we assume to be Gaussian of variance t. Then $\hat{t}$ can be informally seen as the number of Gaussian sigmas that should fit into the patch, which in our case has elliptical shape and can be normalized to a circle of radius r. By introducing $\hat{t}$ we can specify one set of $\hat{t}$ values for all values of r.

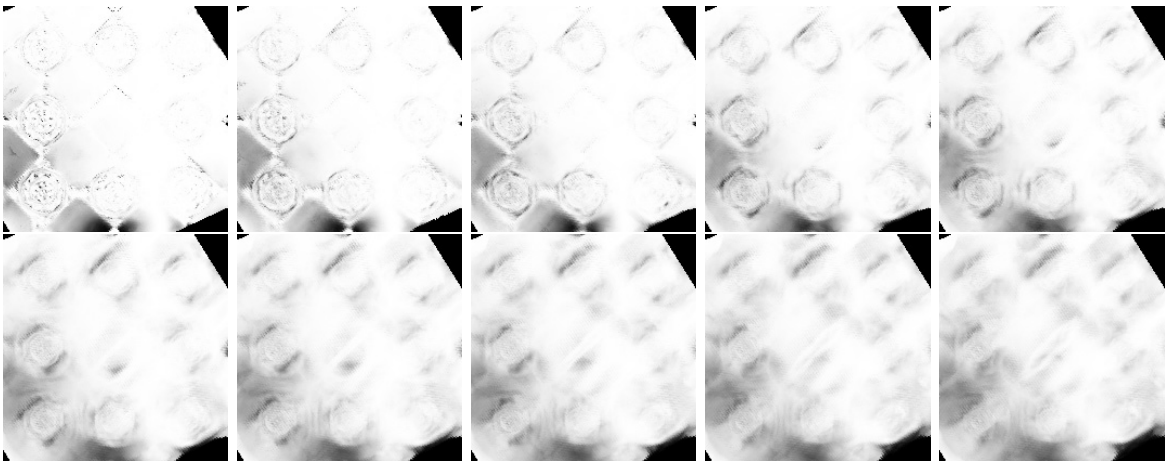

Figure 26. *Similarity values for given correspondences calculated using the proposed similarity measure with different values of r. First row, left to right: $r = 50$, $r = 75$, $r = 100$, $r = 150$, $r = 175$. Second row, left to right: $r = 200$, $r = 225$, $r = 250$, $r = 275$, $r = 300$.*

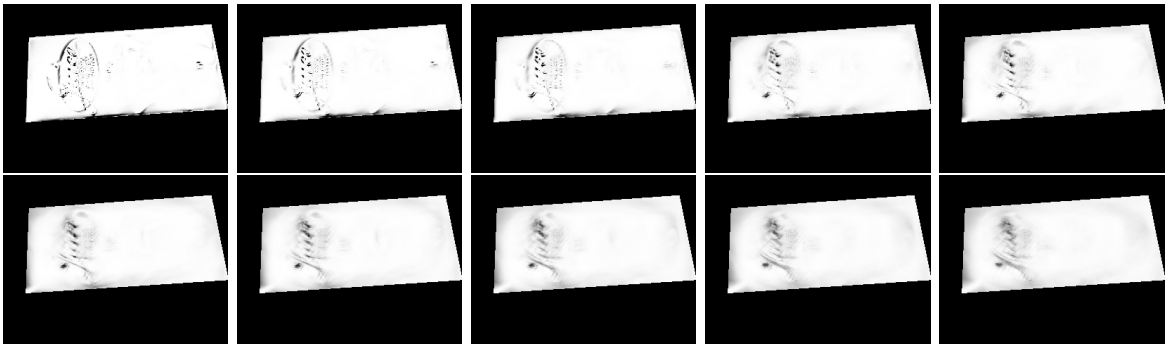

Figure 27. *Similarity values for given correspondences calculated using the proposed similarity measure with different values of r. First row, left to right: $r = 50$, $r = 75$, $r = 100$, $r = 150$, $r = 175$. Second row, left to right: $r = 200$, $r = 225$, $r = 250$, $r = 275$, $r = 300$.*

Figure 28. *Pair of images used in the experiment on the boundary of two different affinities.*

Figure 29 shows similarity maps calculated from the known correspondences for different values of r and scale parameter $\hat{t}$. The set of r values is $r \in \{100, 200, 300\}$, and the set of $\hat{t}$

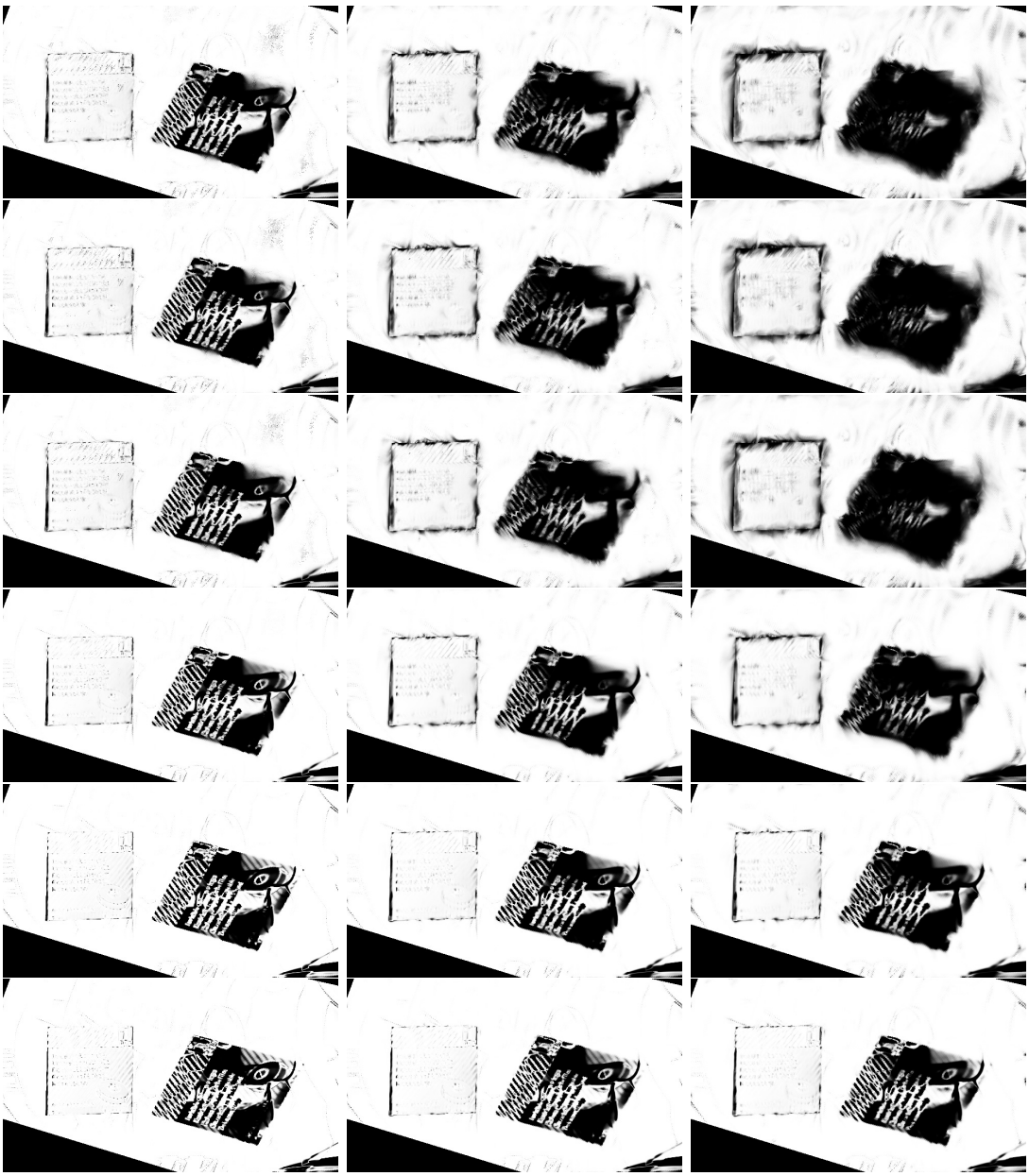

Figure 29. *Similarity values for given correspondences, calculated for different values of* $\hat{t}$ *and* r*. Columns: from left to right* r *is equal to* 50, 100, 150*. Rows: from top to bottom* $\hat{t}$ *is equal to* 0.01, 0.5, 1, 3, 6, 9*.*

values is $\hat{t} \in \{0.01, 0.5, 1, 3, 6, 9\}$. Let us remark that the dark region in the right part of each similarity map is irrelevant to the experiment because it is caused by an occlusion.

This experiment confirms that for all selected combinations of $\hat{t}$ and r the matching is correct far enough away from the boundary of different transformations. Moreover, the size of the problematic region on the boundary may be controlled by adjusting both r and $\hat{t}$ parameters. Keeping in mind that one would like to see r as a parameter that can be selected

once for a concrete pair of images, or even estimated automatically, this emphasizes once again the importance of multiscale similarity computation for the cases where two or more objects undergoing different transformations may meet.

7. Conclusions. In this work we study in detail the linear model to compare patches of images defined on Riemannian manifolds, described first in [3]. Using a WKB approximation we develop two practical formulas (4.15), (4.16) to approximate the solution the corresponding linear PDE (1.3). In order to be able to apply these formulas one needs to provide an a priori connection between any pair of points to be compared. We propose intrinsic metrics allowing us to consider images as Riemannian manifolds. These metrics are defined in terms of affine covariant tensors and affine covariant neighborhoods associated with them. The tensors are then used for obtaining the a priori connection. We accompany theoretical developments with experiments. First we study properties of the proposed similarity measure, assuming that the a priori connection is known. Then we move to the properties of the proposed affine covariant structure tensors in isolation from the similarity measure and also discuss some practical question related to the computation of the a priori connection. Finally, we take a look at the proposed similarity measure as a whole. Many application-specific questions were left aside in order to keep this work general and concise.

Acknowledgment. We would like to acknowledge our mentor, the late Prof. Vicent Caselles, as this work would not have been possible without him.

REFERENCES

[1] S. Agarwal and D. Roth, *Learning a sparse representation for object detection*, in Proceedings of the Seventh European Conference on Computer Vision (ECCV 2002), Lecture Notes in Comput. Sci. 2353, Springer, Berlin, 2002, pp. 113–127.

[2] L. Alvarez, F. Guichard, P.-L. Lions, and J. M. Morel, *Axioms and fundamental equations of image processing*, Arch. Rational Mech. Anal., 123 (1993), pp. 199–257.

[3] C. Ballester, F. Calderero, V. Caselles, and G. Facciolo, *Multiscale analysis of similarities between images on Riemannian manifolds*, Multiscale Model. Simul., 12 (2014), pp. 616–649.

[4] A. Baumberg, *Reliable feature matching across widely separated views*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000), Vol. 1, 2000, pp. 1774–1781.

[5] T. Brox, R. Boomgaard, F. Lauze, J. Weijer, J. Weickert, P. Mrázek, and P. Kornprobst, *Adaptive structure tensors and their applications*, in Visualization and Processing of Tensor Fields, Springer, Berlin, 2006, pp. 17–47, 417–420.

[6] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek, *Nonlinear structure tensors*, Image Vision Comput., 24 (2006), pp. 41–55.

[7] R. Brunelli and T. Poggio, *Face recognition: Features versus templates*, IEEE Trans. Pattern Anal. Mach. Intell., 15 (1993), pp. 1042–1052.

[8] A. Buades, B. Coll, and J.-M. Morel, *A non-local algorithm for image denoising*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005), Vol. 2, 2005, pp. 60–65.

[9] M. C. Burl, M. Weber, and P. Perona, *A probabilistic approach to object recognition using local photometry and global geometry*, in Proceedings of the 10th European Conference on Computer Vision (ECCV' 1998), Lecture Notes in Comput. Sci. 1407, Springer, Berlin, 1998, pp. 628–641.

[10] F. Calderero and V. Caselles, *Multiscale analysis for images on Riemannian manifolds*, SIAM J. Imaging Sci., 7 (2014), pp. 1108–1170.

[11] V. Caselles, L. Igual, and O. Sander, *An axiomatic approach to scalar data interpolation on surfaces*, Numer. Math., 102 (2006), pp. 383–411.

[12] V. Caselles, J. M. Morel, and C. Sbert, *An axiomatic approach to image interpolation*, IEEE Trans. Image Process., 7 (1998), pp. 376–386.
[13] P. Chatterjee and P. Milanfar, *Patch-based near-optimal image denoising*, IEEE Trans. Image Process., 21 (2012), pp. 1635–1649.
[14] J. K. Cohen, F. G. Hagin, and J. B. Keller, *Short time asymptotic expansions of solutions of parabolic equations*, J. Math. Anal. Appl., 38 (1972), pp. 82–91.
[15] C. A. Deledalle, V. Duval, and J. Salmon, *Non-local methods with shape-adaptive patches (NLM-SAP)*, J. Math. Imaging Vision, 43 (2012), pp. 103–120.
[16] R. Fergus, P. Perona, and A. Zisserman, *Object class recognition by unsupervised scale-invariant learning*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003), Vol. 2, 2003, pp. 264–271.
[17] A. Foi, V. Katkovnik, and K. Egiazarian, *Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images*, IEEE Trans. Image Process., 16 (2007), pp. 1395–1411.
[18] J. Garding and T. Lindeberg, *Direct estimation of local surface shape in a fixating binocular vision system*, in Proceedings of the Third European Conference on Computer Vision (ECCV 1994), J.-O. Eklundh, ed., Lecture Notes in Comput. Sci. 800, Springer, Berlin, Heidelberg, 1994, pp. 365–376.
[19] F. Guichard and J.-M. Morel, *Image Analysis and PDEs*, IPAM GBM Tutorials, 2001.
[20] N. Kawai, T. Sato, and N. Yokoya, *Image inpainting considering brightness change and spatial locality of textures and its evaluation*, in Advances in Image and Video Technology, Lecture Notes in Comput. Sci. 5414, Springer, Berlin, 2009, pp. 271–282.
[21] 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., 39 (2000), pp. 111–129.
[22] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, Int. J. Comput. Vis., 60 (2004), pp. 91–110.
[23] J. Matas, O. Chum, M. Urban, and T. Pajdla, *Robust wide-baseline stereo from maximally stable extremal regions*, Image Vision Comput., 22 (2004), pp. 761–767.
[24] K. Mikolajczyk, *Affine Covariant Features: Dataset,* http://www.robots.ox.ac.uk/~vgg/research/affine/ (2007).
[25] K. Mikolajczyk and C. Schmid, *Scale & affine invariant interest point detectors*, Int. J. Comput. Vis., 60 (2004), pp. 63–86.
[26] J.-M. Morel and G. Yu, *ASIFT: A new framework for fully affine invariant image comparison*, SIAM J. Imaging Sci., 2 (2009), pp. 438–469.
[27] J.-M. Morel and G. Yu, *Is sift scale invariant?*, Inverse Probl. Imaging, 5 (2011), pp. 115–136.
[28] G. Peyré, *Manifold models for signals and images*, Comput. Vis. Image Underst., 113 (2009), pp. 249–260.
[29] N. Sabater, J.-M. Morel, and A. Almansa, *How accurate can block matches be in stereo vision?*, SIAM J. Imaging Sci., 4 (2011), pp. 472–500.
[30] N. Sochen, R. Kimmel, and A. M. Bruckstein, *Diffusions and confusions in signal and image processing*, J. Math. Imaging Vision, 14 (2001), pp. 195–209.
[31] E. Stiefel, *Richtungsfelder und fernparallelismus in n-dimensionalen mannigfaltigkeiten*, Comment. Math. Helv., 8 (1935), pp. 305–353.
[32] J. Sullivan, A. Blake, M. Isard, and J. MacCormick, *Bayesian object localisation in images*, Int. J. Comput. Vis., 44 (2001), pp. 111–135.
[33] E. Thomas, *Vector fields on manifolds*, Bull. Amer. Math. Soc., 75 (1969), pp. 643–683.
[34] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, in Proceedings of the Sixth International IEEE Conference on Computer Vision (ICCV 1998), 1998, pp. 839–846.
[35] T. Tuytelaars and K. Mikolajczyk, *Local invariant feature detectors: A survey*, Found. Trends Comp. Graph. Vis., 3 (2008), pp. 177–280.
[36] T. Tuytelaars and L. Van Gool, *Matching widely separated views based on affine invariant regions*, Int. J. Comput. Vis., 59 (2004), pp. 61–85.
[37] S. R. S. Varadhan, *On the behavior of the fundamental solution of the heat equation with variable coefficients*, Comm. Pure Appl. Math., 20 (1967), pp. 431–455.

[38] A. Vedaldi and B. Fulkerson, *VLFeat: An Open and Portable Library of Computer Vision Algorithms*, http://www.vlfeat.org/ (2008).
[39] J. Weickert, *Anisotropic Diffusion in Image Processing*, Vol. 1, Teubner, Stuttgart, 1998.
[40] J. Weickert, *Coherence-enhancing diffusion filtering*, Int. J. Comput. Vis., 31 (1999), pp. 111–127.
[41] K. J. Yoon and I. S. Kweon, *Adaptive support-weight approach for correspondence search*, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006), pp. 650–656.