Optimal Image Alignment With Random Projections of Manifolds: Algorithm and Geometric Analysis

Effrosyni Kokiopoulou, *Member, IEEE*, Daniel Kressner, and Pascal Frossard, *Senior Member, IEEE*

***Abstract*—This paper addresses the problem of image alignment based on random measurements. Image alignment consists of estimating the relative transformation between a query image and a reference image. We consider the specific problem where the query image is provided in compressed form in terms of linear measurements captured by a vision sensor. We cast the alignment problem as a manifold distance minimization problem in the linear subspace defined by the measurements. The transformation manifold that represents synthesis of shift, rotation, and isotropic scaling of the reference image can be given in closed form when the reference pattern is sparsely represented over a parametric dictionary. We show that the objective function can then be decomposed as the difference of two convex functions (DC) in the particular case where the dictionary is built on Gaussian functions. Thus, the optimization problem becomes a DC program, which in turn can be solved globally by a cutting plane method. The quality of the solution is typically affected by the number of random measurements and the condition number of the manifold that describes the transformations of the reference image. We show that the curvature, which is closely related to the condition number, remains bounded in our image alignment problem, which means that the relative transformation between two images can be determined optimally in a reduced subspace.**

***Index Terms*—Manifold condition number, pattern transformations, random projections, sparse representations, transformation manifolds.**

I. Introduction

THE problem of computing the relative geometric transformations between visual patterns is of paramount importance and enjoys numerous applications in various fields, including vision sensor networks, pattern recognition, and medical image analysis, to name just a few [1]. The analysis or comparison of visual patterns is generally only meaningful when images are aligned and placed in a common referential system. The transformed version of a reference pattern can be described as a point of a low-dimensional manifold in a high-dimensional space, which is usually called the *transformation manifold*. The alignment of a query image can then be performed by computing its projection on the transformation manifold. The manifold distance (MD) is defined as the minimum distance between the query image p and the manifold generated by the reference image s (see Fig. 1).

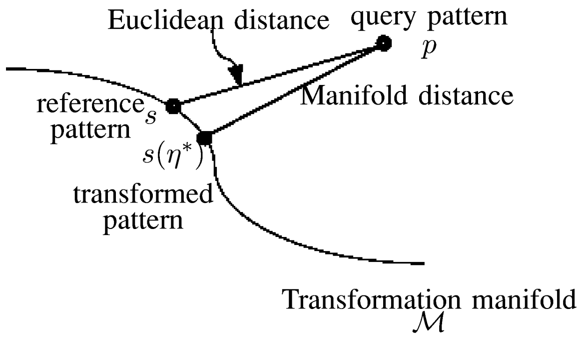


Fig. 1. MD is the minimum distance from a query point p to the transformation manifold $\mathcal{M}$ spanned by the transformed versions of s.

In certain applications, however, we might not have access to the full query image or it might be computationally too expensive to deal with the complete image. These limitations are either due to the design of sensors or due to important constraints in terms of bandwidth or computational resources. In this case, the query image may be given in compressed form by a few linear measurements. The analysis of compressed query images has recently found applications in wide-area persistent surveillance [2] and fast MR imaging [3], for example. The use of linear measurements permits to reduce the complexity of the sensing step and the scale of the image alignment problem; it thus extends to systems with imperfect settings such as low-complexity vision sensors networks (see, e.g., [4]–[6]). When the query image is given by only a few random projections (see Fig. 2) in the image alignment problem, the relative geometric transformation between a query and a reference image is estimated directly from the linear measurements. According to the theory of compressed sensing (CS), a few random projections of a sparse (or nearly sparse) signal are sufficient to preserve its salient information. Simultaneously, it has been shown that random projections of signal manifolds result into approximately isometric embeddings, i.e., pairwise Euclidean distances are nearly preserved in the reduced space [7]–[9]. Equipped with these two properties, it becomes possible to solve the image alignment problem with help of a transformation manifold in the reduced space that is built on the linear measurements performed by the sensors.

Manuscript received December 17, 2009; revised April 26, 2010, September 20, 2010, November 25, 2010; accepted December 17, 2010. Date of publication December 23, 2010; date of current version May 18, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xilin Chen.

E. Kokiopoulou and D. Kressner are with the Seminar for Applied Mathematics, Department of Mathematics, ETH Zurich, CH-8092 Zurich, Switzerland (e-mail: effrosyni.kokiopoulou@sam.math.ethz.ch; daniel.kressner@sam.math.ethz.ch).

P. Frossard is with the Signal Processing Laboratory (LTS4), Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: pascal.frossard@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2010.2102044

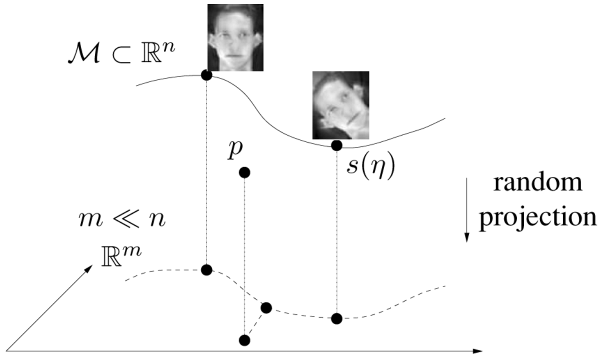


Fig. 2. Image alignment is performed in the reduced space after projections.

We propose a new method for image alignment when the relative transformation consists of a synthesis of translations, rotation, and isotropic scaling. The proposed method estimates the globally optimal transformation between a query image and a reference image s, using a sufficient number of random measurements. For this purpose, we represent s as a sparse linear combination of geometric primitives, called atoms, which are chosen from a parametric, possibly redundant dictionary. This representation permits to build a parametric definition of the transformation manifold, which describes the possible transformations of the reference image. The image alignment problem can then be cast as an MD minimization problem. Building on our previous work [10], [11], we then formulate the pattern alignment problem with random measurements as a DC program by showing that the objective function can be represented as a *difference of convex functions* (DC), in a case where the dictionary is composed of Gaussian atoms. DC programs are nonconvex problems that can be solved with efficient globally optimal algorithms by exploiting their special structure. Our approach therefore provides a constructive way to perform image alignment with random measurements.

At the same time, the results from [9] suggest that the number of measurements necessary for proper alignment depends on the condition number of the pattern transformation manifold, which in turn is closely related to its curvature. We perform a geometric analysis of this manifold, showing that it is well conditioned and providing an explicit upper bound on the curvature. Moreover, we establish an efficient numerical procedure for computing the principal curvature at a certain point on the manifold. These results confirm that the required number of measurements is bounded in our problem, and that image alignment can be solved efficiently in a reduced subspace. In summary, the contribution of this paper consists of providing: 1) a globally optimal approach to image alignment with random projections and 2) theoretical as well as numerical insights into the geometric properties of the pattern transformation manifolds, showing that the image alignment problem can be solved with a finite number of measurements.

The remainder of this paper is organized as follows. In Section II, we formulate the problem of image alignment from random measurements. In Section III, we discuss the representation of transformation manifolds using sparse geometric expansions. We then show in Section IV that the distance between the query pattern and the transformation manifold is a DC function of the transformation parameters. In Section V, we provide a geometric analysis of transformation manifolds that confirms that a limited number of measurements is sufficient for accurate alignment. Finally, experimental results are presented in Section VI.

II. IMAGE ALIGNMENT PROBLEM

A. Problem Formulation

In this paper, we are interested in estimating the relative transformation η^* that best matches two visual patterns. This problem is very common in image processing or computer vision applications that rely on image alignment or matching of views. In particular, we consider the problem where only a compressed version of the query pattern is available in the form of linear measurements captured by the vision sensor. Formally, we consider images that undergo transformations described by a set of parameters η. We further suppose that all of the transformed versions $s(\eta)$ of the reference pattern s can be represented by a transformation manifold $\mathcal{M}$ of low dimension d. Then, we assume that we have m random projections of the query pattern p, obtained by computing inner products with m random signals $z_1, \ldots, z_m$. The image alignment problem is equivalent to an MD minimization problem in the linear subspace defined by the measurement vectors. It can be formulated as a parameter estimation problem as follows.

Transformation estimation problem:

$$\eta^* = \arg\min_{\eta} f(\eta)$$

where

$$f(\eta) = \sum_{i=1}^{m} |\langle s(\eta), z_i\rangle - \langle p, z_i\rangle|. \tag{1}$$

The use of the 1-norm in the objective function is quite popular in image registration and is often referred to as the sum of absolute differences (SAD); see [12]. Given also the fact that the L1-norm is known to be robust to outliers, we have chosen to use this norm in the case of random measurements as well. Moreover, as will be seen below, this norm appears to be most suitable for the DC optimization method proposed in this paper.

The optimization problem (1) for determining the best transformation parameters η^* is typically nonconvex [13]. This makes it hard to solve using traditional methods, such as steepest descent or Newton-type methods due to their local convergence property and the presence of an unknown number of local minima. We show in Section II-B how the transformation manifold could be described in a parametric form when the transformation is a synthesis of shift, rotation and isotropic scaling. The objective function can be written as a difference of two convex (DC) functions, which in turn permits to formulate the optimization problem as a DC program and to solve it globally by a cutting plane method [14, Th. 5.3].

B. Geometric Insights

In the above alignment problem, one still has to choose the number of measurements that leads to efficient transformation estimation. The optimal number of random projections is hard

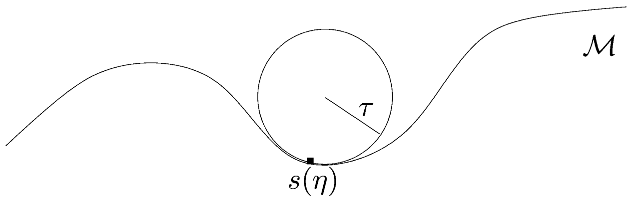


Fig. 3. Large τ corresponds to a well-conditioned manifold.

to define in practice. Suppose that we project the transformation manifolds spanned by two distinct patterns on m random vectors $z_1, \ldots, z_m$. In order to make sure that matching points in the reduced space is close to matching the corresponding points in the original high-dimensional space, the embedding should be nearly isometric, that is, pairwise Euclidean distances should be nearly preserved. Only if this is the case can one reliably perform image alignment in the reduced space and estimate the unknown transformation.

Recently, Baraniuk and Wakin [7] provided an estimate of m that is linear in d and logarithmic in n, which is the number of pixels in the image. We briefly revisit the main result from [7].

Theorem 1: Let $\mathcal{M}$ be a compact d-dimensional manifold in $\mathbb{R}^n$ having condition number $1/\tau$, volume V, and geodesic covering regularity R. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let Z be a random orthoprojector from $\mathbb{R}^n$ to $\mathbb{R}^m$ and

$$m \geq O\left(\frac{d \log(nVR\tau^{-1}\epsilon^{-1}) \log(\rho^{-1})}{\epsilon^2}\right). \tag{2}$$

Suppose $m < n$. Then, with probability exceeding $1 - \rho$, the following statement holds. For every pair of points $x, y \in \mathcal{M}$, we have

$$(1-\epsilon)\sqrt{\frac{m}{n}} \leq \frac{\|Zx - Zy\|_2}{\|x - y\|_2} \leq (1+\epsilon)\sqrt{\frac{m}{n}}. \tag{3}$$

In short, Theorem 1[1] is proved by determining a high-resolution sampling on the manifold and then applying the Johnson–Lindenstrauss lemma [15] to the sampled points. The above theorem implies that, besides d and n, the number m depends logarithmically on other properties of the manifold, such as its condition number $1/\tau$, volume V, and geodesic covering regularity R. Note that R is closely related to the condition number [7], and we will omit its definition. Intuitively, the condition number of $\mathcal{M}$ is defined as $1/\tau$, where τ is the maximum radius of a sphere that, when placed tangent to any point in $\mathcal{M}$, intersects $\mathcal{M}$ only at that point (see [16] for a precise definition). This is illustrated graphically in Fig. 3, showing that a large τ implies a well-conditioned manifold, which has a low curvature and does not, at least locally, intersect itself. Therefore, the condition number is an important property of $\mathcal{M}$ towards characterizing the number m of required random measurements. We show in Section V that our transformation manifold is well conditioned, which means that the number of necessary measurements is clearly bounded in the problem of (1).

III. TRANSFORMATION MANIFOLDS

A. *Visual Pattern Representation*

We show here how one could build a parametric transformation manifold for a reference pattern s. We first explain the representation of the pattern as a linear combination of geometric functions (usually called *atoms*), taken from a structured parametric and possibly redundant dictionary $\mathcal{D} = \{\phi_\gamma, \gamma \in \Gamma\}$ spanning the input space. This representation aims at capturing the most prominent features of the pattern. The atoms in a *parametric* dictionary are constructed by applying geometric transformations to a generating mother function denoted by ϕ. A geometric transformation $\gamma \in \Gamma$ can be represented by the action of an operator $U(\gamma)$, and therefore the parametric dictionary takes the form

$$\mathcal{D} = \{\phi_\gamma = U(\gamma)\phi, \gamma \in \Gamma\}. \tag{4}$$

A transformation γ_i, defining the ith atom, is composed of elementary transformations of the following three types.

- *Translation* by $b_i = [b_{ix}\ b_{iy}]^\top$. $U(b_i)$ moves the generating function across the image, i.e., $U(b_i)\phi(x,y) = \phi(x - b_{ix}, y - b_{iy})$.
- *Rotation* by ω_i. $U(\omega_i)$ rotates the generating function by the angle ω_i, i.e., $U(\omega_i)\phi(x,y) = \phi(\cos(\omega_i)x + \sin(\omega_i)y, \cos(\omega_i)y - \sin(\omega_i)x)$.
- *Anisotropic scaling* by $a_i = [a_{ix}\ a_{iy}]^\top$. $U(a_i)$ scales the generating function anisotropically, i.e., $U(a_i)\phi(x,y) = \phi((x)/(a_{ix}), (y)/(a_{iy}))$.

These elementary transformations yield a transformation $\gamma_i = (b_i, a_i, \omega_i) \in \Gamma$ as a synthesis of translation, anisotropic scaling, and rotation. It can be observed that applying a transformation to the mother function is equivalent to transforming the coordinate system from $\{x, y\}$ to $\{\tilde{x}, \tilde{y}\}$ before applying $\phi(\,\cdot\,)$. In particular, the ith atom $\phi_{\gamma_i} = U(\gamma_i)\phi(x,y)$ with $\gamma_i = (b_i, a_i, \omega_i) \in \Gamma$ can be regarded as the pullback

$$\phi_{\gamma_i}(x,y) = \phi(\Psi_{\gamma_i}(x,y)) \tag{5}$$

where $(\tilde{x}, \tilde{y}) := \Psi_{\gamma_i}(x,y)$ satisfies

$$\begin{aligned}\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} &= \underbrace{\begin{bmatrix}\frac{1}{a_{ix}} & 0\\ 0 & \frac{1}{a_{iy}}\end{bmatrix}}_{A} \underbrace{\begin{bmatrix}\cos\omega_i & \sin\omega_i\\ -\sin\omega_i & \cos\omega_i\end{bmatrix}}_{R(\omega_i)} \underbrace{\begin{bmatrix}x - b_{ix}\\ y - b_{iy}\end{bmatrix}}_{t}\\ &= AR(\omega_i)t.\end{aligned} \tag{6}$$

The approximation of a pattern s with atoms from the dictionary $\mathcal{D}$ can be obtained in different ways. Even if finding the sparsest approximation of s is generally a hard problem, effective suboptimal solutions are usually sufficient to capture the salient and geometric structure of the pattern with only a few atoms. In this work, we have chosen to use Orthogonal Matching Pursuit (OMP) [17, Sec. 9.5.3], which is a simple yet effective algorithm for computing sparse approximations in practice.

Initially, OMP chooses the residual $r_0 = s$ and then proceeds iteratively by selecting in the kth step the atom ϕ_{γ_k} that best

---

[1] Note that, even if $f(\eta)$ in (1) is based on the 1-norm distance in our formulation, Theorem 1 still provides insights on the required number of measurements, by the equivalence of norms in finite-dimensional spaces (i.e., a good match in the 1-norm yields a good match in the 2-norm).

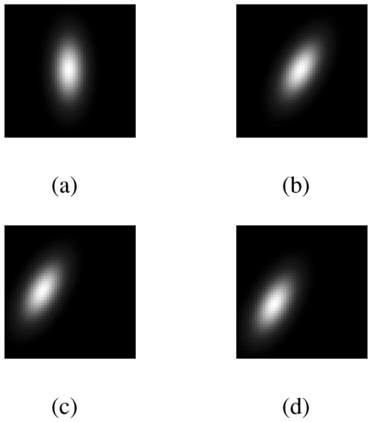


Fig. 4. Sample Gaussian atoms. (a) $b_x = 0, b_y = 0, a_x = 10, a_y = 5, \omega = 0$. (b) $b_x = 0, b_y = 0, a_x = 10, a_y = 5, \omega = -\pi/6$. (c) $b_x = 0, b_y = -10, a_x = 10, a_y = 5, \omega = -\pi/6$. (d) $b_x = 5, b_y = -10, a_x = 10, a_y = 5, \omega = -\pi/6$.

matches the residual r_{k-1}, i.e., $\gamma_k = \arg_{\gamma\in\Gamma} \max |\langle r_{k-1}, \phi_\gamma\rangle|$. Then, γ_k is removed from the residual by projection: $r_k = (I - P_k)r_{k-1}$, where P_k is the orthogonal projector onto $\mathrm{span}\{\phi_{\gamma_k}\}$. After K steps of OMP, the pattern s is approximated by a sparse linear combination of K atoms

$$s \approx \sum_{k=1}^{K} \xi_k \phi_{\gamma_k}. \tag{7}$$

We propose the use of a dictionary of 2-D atoms capturing the geometric information in an image. The generating function ϕ of $\mathcal{D}$ used in this paper is the Gaussian

$$\phi(x, y) = \frac{1}{\rho} \exp(-(x^2 + y^2)). \tag{8}$$

This dictionary has been shown to provide good approximation performances for natural images [18]. Fig. 4 shows a few sample Gaussian atoms corresponding to various geometric transformations γ. In addition, Fig. 5 illustrates the progressive approximation of a human face from a Gaussian dictionary using OMP. Observe that already very few atoms are sufficient to capture the main geometric characteristics of the pattern and that the representation (7) does not need to be very accurate to be useful for alignment purposes.

B. Transformation Manifolds

In the following, we show how all the geometric transformations of the reference image s build a parametric transformation manifold. We restrict scalings to be isotropic, i.e., the geometric transformation η takes the form $\eta = (b, \alpha, \omega)$ consisting of a translation $b = [b_x, b_y]$, an isotropic scaling α, and a rotation ω. The manifold $\mathcal{M}$ of all such transformed images can be expressed mathematically as

$$\mathcal{M} = \{s(\eta) := U(\eta)s, \eta = (b, \alpha, \omega)\}. \tag{9}$$

Although the manifold resides in a high-dimensional space, its intrinsic dimension d is rather small and equal to the number of transformation parameters, which is 4. Fig. 6 shows a few samples from the transformation manifold of a human face, with the transformation restricted to a rotation. The random projections of the resulting manifold are illustrated in Fig. 7.

In general, all possible transformations η forms a group, the so-called similitude group SIM(2) of the plane. As in (6), we denote

$$R(\omega) = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix}, \qquad 0 \le \omega < 2\pi$$

as the rotation matrix for the angle ω. If (b, α, ω) and (b', α', ω') are two elements of the SIM(2) group, then the group law [18] is given by

$$(b, \alpha, \omega) \circ (b', \alpha', \omega') = (b + \alpha R(-\omega)b', \alpha\alpha', \omega' + \omega). \tag{10}$$

In the following, we replace the reference image s by its approximation (7). Using the pullback interpretation (5), it follows that the transformation η applied to s results in

$$s(\eta) = U(\eta)s = \sum_{k=1}^{K} \xi_k U(\eta)\phi_{\gamma_k} = \sum_{k=1}^{K} \xi_k \phi_{\eta\circ\gamma_k} \tag{11}$$

where $\eta\circ\gamma_k$ is a composition of transformations. In other words, the transformation is applied to each constituent atom individually. Furthermore, the group law (10) can be employed to determine the updated parameters of the transformed atoms. Let us emphasize the importance of (11): it permits to express the manifold equation (9) in closed form with respect to the transformation parameters η. The definition of the manifold in (11) is used in the optimization problem (1). We show in Section IV that the resulting manifold distance problem can then be defined as the difference of two convex functions, which enables the use of the DC programming methodology.

IV. DC DECOMPOSITION

A. Properties of Dc Functions

The purpose of this section is to show that the objective function (1) is DC when the dictionary is built from 2-D Gaussian functions. We start with some definitions and basic properties about DC functions [14], [19], [20]. Let $X \subseteq \mathbb{R}^n$ be convex. A function $f : X \to \mathbb{R}$ is called DC on X if there exist two convex functions $g, h : X \to \mathbb{R}$ such that

$$f(x) = g(x) - h(x). \tag{12}$$

A representation of this form is called *DC decomposition* of f. DC decompositions are clearly not unique; for any convex function $c(x)$, the decomposition $f(x) = (g(x) + c(x)) - (h(x) + c(x))$ is also DC. We will make use of the following two properties.

Proposition 1 (Properties of DC Functions [19, Sec. 4.2]): Let $f = g - h$ and $f_i = g_i - h_i, i = 1 \ldots, m$ be DC functions. Then, the following functions are also DC:

a) $\sum_{i=1}^{m} \lambda_i f_i = [\sum_{\{i:\lambda_i\ge 0\}} \lambda_i g_i - \sum_{\{i:\lambda_i<0\}} \lambda_i h_i] - [\sum_{\{i:\lambda_i\ge 0\}} \lambda_i h_i - \sum_{\{i:\lambda_i<0\}} \lambda_i g_i]$.
b) $|f| = 2\max\{g, h\} - (g + h)$.

B. DC Form of the Objective Function

We now combine Proposition 1 with our previous results [11] to prove the main result of this paper.

Theorem 2: The objective function f in (1) is DC.

Fig. 5. Progressive OMP approximation of a human face (leftmost) with 20, 50, 80, 110, and 140 Gaussian atoms (from left to right).

Fig. 6. Samples from the transformation manifold of the human face. From left to right, the samples correspond to rotation angles from 0 to 2π with step $\pi/4$.

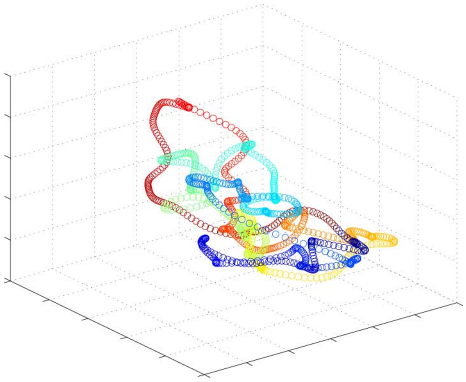


Fig. 7. Random projections in 3-D of the rotation manifold of Fig. 6. The samples correspond to rotation angles from 0 to 2π with step $\pi/500$. The color/shading is a linear map of the rotation angles.

Proof: Recall that

$$\begin{aligned} f(\eta) &= \sum_{i=1}^{m} |\langle s(\eta), z_i\rangle - \langle p, z_i\rangle| \\ &= \sum_{i=1}^{m} \left| \sum_{k=1}^{K} \xi_k \langle \phi_{\eta_k}, z_i\rangle - \langle p, z_i\rangle \right| \end{aligned} \tag{13}$$

where $\eta_k = \eta \circ \gamma_k$. In [11] we have shown that (i) the transformed generating functions ϕ_{η_k} are DC, (ii) the inner products $\langle \phi_{\eta_k}, z\rangle$ between the atoms and a fixed pattern z are DC, and (iii) the inner product $\langle s(\eta), z\rangle = \sum_{k=1}^{K} \xi_k \langle \phi_{\eta_k}, z\rangle$ is also a DC function of η.

In particular, each function $\langle s(\eta), z_i\rangle = \sum_{k=1}^{K} \xi_k \langle \phi_{\eta_k}, z_i\rangle$ corresponding to a measurement vector z_i, with $1 \leq i \leq m$, is DC. Note that $f_i(\eta) := \langle s(\eta), z_i\rangle - \langle p, z_i\rangle$ remains DC since the second term is constant and does not depend on η. Assume now that the DC decomposition of each function f_i is given by $f_i(\eta) = g_i(\eta) - h_i(\eta)$.

By Proposition 1b), the absolute value of a DC function is DC and hence

$$|f_i(\eta)| = 2\max\{g_i, h_i\} - (g_i + h_i) = \tilde{g}_i(\eta) - \tilde{h}_i(\eta)$$

is also DC. Finally, the objective function in (13) is DC since it is simply a sum of M DC functions

$$\begin{aligned} f(\eta) &= \sum_{i=1}^{m} |f_i(\eta)| = \sum_{i=1}^{m} (\tilde{g}_i(\eta) - \tilde{h}_i(\eta)) \\ &= \underbrace{\sum_{i=1}^{m} \tilde{g}_i(\eta)}_{g(\eta)} - \underbrace{\sum_{i=1}^{m} \tilde{h}_i(\eta)}_{h(\eta)}. \end{aligned}$$

□

C. DC Programs

An optimization problem is called a DC program if it takes the form

$$\begin{aligned} \min_x \quad & f(x) = g(x) - h(x), \\ \text{s.t.} \quad & x \in X = \{x \in \mathbb{R}^n : \delta(x) \leq 0\} \end{aligned} \tag{14}$$

where $g, h : X \to \mathbb{R}$ are convex functions and $\delta : \mathbb{R}^n \to \mathbb{R}$ is a convex function. The next proposition provides an optimality condition for (14).

Proposition 2 [14]: The point $x^* \in X$ is an optimal solution to the DC problem (14) if and only if there exists $t^* \in \mathbb{R}$ such that

$$0 = \inf\{-h(x) + t : x \in X, t \in \mathbb{R}, g(x) - t \leq g(x^*) - t^*\}. \tag{15}$$

In this work, we have chosen to solve the DC Program (14) by the outer approximation cutting plane algorithm proposed in [14, Sec. 5.3], for its simplicity and also due to the fact that the parameter space in our problem is four-dimensional. However, we should mention that our framework could also be combined with other DC solvers such as Branch-and-Bound schemes [14, Sec. 5.1 and 5.2] and DCA [21].

The efficiency of the DC solver mainly depends on the overall cost for evaluating DC decompositions. The cost for evaluating one DC decomposition grows proportionally with $K \cdot n_1 \cdot n_2$, where K is the number of atoms and $n_1 \times n_2$ is the resolution of the image. Hence, sparsity directly affects the efficiency of our method. This is where the choice of the dictionary becomes important, since it has to capture the main features of natural images with a few atoms only. This is the case for the dictionary of 2-D Gaussian functions used in this paper [18].

The cutting plane method quickly converges to the vicinity of the global minimizer but then typically saturates. Hence, to obtain high accuracy it is more beneficial to switch to a local optimization method, such as Newton, once the iterate is sufficiently close to the global minimizer, see, e.g., [11].

V. GEOMETRIC ANALYSIS OF TRANSFORMATION MANIFOLDS

We have seen in Section II that the condition number of the manifold $\mathcal{M}$ is an important factor towards characterizing the number m of random measurements needed. At the same time, it is also known that the condition number is closely related to classical notions of curvature in differential geometry via the second fundamental form. In particular, Niyogi *et al.* in [16, Prop. 6.1] show that the condition number $1/\tau$ is an upper bound of the principal curvature (defined below) at any point on the manifold.

Here, we first derive an upper bound of the principal curvature of parametric transformation manifolds $\mathcal{M}$ defined in (9). When there are no (near) self-intersections, this upper bound can be used instead of the condition number for characterizing the manifold. It further indicates that the transformation manifold is well conditioned, which means that the number of measurements required to solve the Problem (1) is clearly bounded. Furthermore, based on the obtained developments, we additionally provide an efficient numerical algorithm for computing in practice the principal curvature at a certain point on $\mathcal{M}$.

For notational convenience, we will denote the transformation parameters as follows: $\eta = (\eta_1, \eta_2, \eta_3, \eta_4) = (b_x, b_y, \alpha, \omega)$. The metric tensor $G \in \mathbb{R}^{4\times 4}$ is then given by

$$[G]_{ij} = \langle t_i, t_j \rangle \tag{16}$$

where t_i, t_j are the ith and jth tangent vectors, defined as

$$t_i = \frac{\partial s(\eta)}{\partial \eta_i}$$

and assumed to be linearly independent. The tangent space $T_\eta \mathcal{M}$ at point $s(\eta) \in \mathcal{M}$ is defined as

$$T_\eta \mathcal{M} = \operatorname{span}\{t_1, t_2, t_3, t_4\}.$$

Note that $d = \dim T_\eta \mathcal{M} = 4$ although the transformation manifold $\mathcal{M}$ is a submanifold of $\mathbb{R}^n$. The codimension of $T_\eta \mathcal{M}$ is therefore given by $n-4$. Consider the direct sum

$$\mathbb{R}^n = T_\eta \mathcal{M} \oplus T_\eta \mathcal{M}^\perp$$

and let $\{N_1, N_2, \ldots, N_{n-4}\}$ be an orthonormal basis of $T_\eta \mathcal{M}^\perp$. Then any (unit) normal vector can be written as

$$N = \sum_{i=1}^{n-4} \zeta^i N_i \tag{17}$$

with coefficients $\zeta^i = \langle N_i, N \rangle$. Fig. 8 provides a graphical illustration of the scenario that we consider.

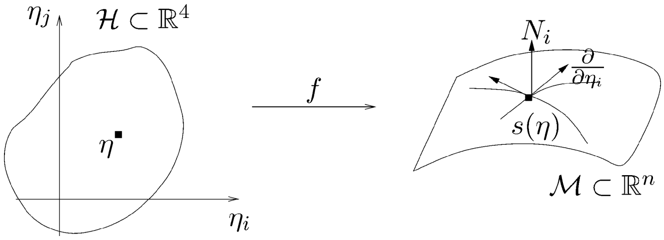


Fig. 8. Parameter space $\mathcal{H}$ provides a parametrization of the transformation manifold $\mathcal{M}$.

In what follows, we show how one can compute the linear operator $L_\zeta : T_\eta \mathcal{M} \to T_\eta \mathcal{M}$ associated with the second fundamental form. According to the standard definition [22, Prop. 2.3]

$$L_\zeta(X) = -(\nabla_X N)^T \tag{18}$$

where $X \in T_\eta \mathcal{M}$, ∇_X denotes the covariant derivative in $\mathbb{R}^n$, and $(\,\cdot\,)^T$ denotes the projection on the tangent space. Take a tangent vector $X = \sum_{j=1}^4 x^j t_j$. Then, it holds that

$$-\nabla_X N = -\nabla_{\sum_{j=1}^4 x^j t_j} N = \sum_{j=1}^4 x^j \left(-\nabla_{t_j} N\right). \tag{19}$$

Hence, it suffices to study the operator $\nabla_{t_j} N$. It holds that

$$\begin{aligned} -\nabla_{t_j} N &= -\nabla_{t_j} \left(\sum_{i=1}^{n-4} \zeta^i N_i \right) \\ &= -\sum_{i=1}^{n-4} \nabla_{t_j} \zeta^i N_i - \sum_{i=1}^{n-4} \zeta^i \nabla_{t_j} N_i. \end{aligned}$$

Now observe that

$$\left(-\nabla_{t_j} N\right)^T = \sum_{i=1}^{n-4} \zeta^i \left(-\nabla_{t_j} N_i\right)^T. \tag{20}$$

The covariant derivative $\nabla_{t_j} N_i \in \mathbb{R}^n$ can be decomposed as

$$-\nabla_{t_j} N_i = -\frac{\partial N_i}{\partial \eta_j} = \sum_{k=1}^4 L_{ij}^k t_k + \sum_{k=1}^{n-4} P_{ij}^k N_k \tag{21}$$

for some coefficients L_{ij}^k with $i = 1, \ldots, n-4$ and $j, k = 1, \ldots, 4$. This directly gives

$$\left(-\nabla_{t_j} N_i\right)^T = \sum_{k=1}^4 L_{ij}^k t_k. \tag{22}$$

In what follows, we show how one can compute the coefficients L_{ij}^k in (21), from which we obtain

$$-\left\langle \frac{\partial N_i}{\partial \eta_j}, t_l \right\rangle = \sum_{k=1}^4 L_{ij}^k g_{kl}.$$

Unfortunately, it is not easy to compute $(\partial N_i)/(\partial \eta_j)$ in practice, as the normal vectors $N_i, i = 1, \ldots, n-4$ are typically obtained by a Gram–Schmidt orthogonalization process. However, it is known [23] that

$$\left\langle \frac{\partial N_i}{\partial \eta_j}, t_l \right\rangle = -\langle N_i, t_{lj} \rangle$$

where t_{lj} is the mixed partial derivative, i.e.,

$$t_{lj} = \frac{\partial^2 s(\eta)}{\partial \eta_l \partial \eta_j}.$$

The proof is provided in the Appendix for the sake of completeness.

This has the advantage that t_{lj} is much easier to compute in practice than $(\partial N_i)/(\partial \eta_j)$. Therefore, for fixed i, j, the coefficients $L_{ij}^k, k = 1, \dots, 4$, can be obtained by solving a 4×4 linear system

$$G \begin{bmatrix} L_{ij}^1 \\ \vdots \\ L_{ij}^4 \end{bmatrix} = \begin{bmatrix} \langle N_i, t_{j1} \rangle \\ \vdots \\ \langle N_i, t_{j4} \rangle \end{bmatrix}$$

where G is the 4×4 metric tensor defined in (16). In more compact form, we have

$$L_{ij}^k = \sum_{l=1}^{4} g^{kl} \langle N_i, t_{jl} \rangle \tag{23}$$

where $g^{kl} = [G^{-1}]_{kl}$ denotes the (k, l) entry of the inverse of the metric tensor.

Combining (20) and (22) yields

$$\begin{aligned} \left(-\nabla_{t_j} N\right)^T &= \sum_{i=1}^{n-4} \zeta^i \left(\sum_{k=1}^{4} L_{ij}^k t_k \right) \\ &= \sum_{k=1}^{4} \underbrace{\left(\sum_{i=1}^{n-4} \zeta^i L_{ij}^k \right)}_{=:\tilde{L}_j^k} t_k = \sum_{k=1}^{4} \tilde{L}_j^k t_k. \end{aligned} \tag{24}$$

Using (24), (19) becomes

$$\begin{aligned} -(\nabla_X N)^T &= \sum_{j=1}^{4} x^j \left(-\nabla_{t_j} N\right)^T = \sum_{j=1}^{4} x^j \sum_{k=1}^{4} \tilde{L}_j^k t_k \\ &= \sum_{k=1}^{4} \left(\sum_{j=1}^{4} \tilde{L}_j^k x^j \right) t_k. \end{aligned} \tag{25}$$

The above equation implies that

$$Y = -(\nabla_X N)^T =: \sum_{k=1}^{4} y^k t_k$$

with components

$$y^k = \sum_{j=1}^{4} \tilde{L}_j^k x^j.$$

Therefore, the linear operator L_ζ has the following matrix representation:

$$L_\zeta = \begin{bmatrix} \tilde{L}_1^1 & \tilde{L}_2^1 & \tilde{L}_3^1 & \tilde{L}_4^1 \\ \tilde{L}_1^2 & \tilde{L}_2^2 & \tilde{L}_3^2 & \tilde{L}_4^2 \\ \tilde{L}_1^3 & \tilde{L}_2^3 & \tilde{L}_3^3 & \tilde{L}_4^3 \\ \tilde{L}_1^4 & \tilde{L}_2^4 & \tilde{L}_3^4 & \tilde{L}_4^4 \end{bmatrix}. \tag{26}$$

It is important to mention at this point that the operator L_ζ is self-adjoint with respect to the induced metric in the tangent space [22] and therefore its eigenvalues are real. The maximum eigenvalue of L_ζ is usually called the *principal curvature*.

Algorithm 1 Numerical estimation of the principal curvature

```
Input: normal direction $\zeta^1, \dots, \zeta^{n-4}$
Output: linear operator $L_\zeta$, estimate $\lambda_\zeta$ of the principal curvature.
Compute the tangent vectors $t_i = (\partial s)/(\partial \eta_i), i = 1, \dots, 4$.
Compute the mixed second order partial derivatives
    $t_{ij} = \frac{\partial^2 s}{\partial \eta_i \partial \eta_j}, i, j = 1, \dots, 4.$
Compute the metric tensor $[G]_{ij} = \langle t_i, t_j \rangle$ as well as its inverse $g^{ij}$.
Build an orthonormal basis $\{N_1, N_2, \dots, N_{n-4}\}$ of $T_\eta \mathcal{M}^\perp$ using Gram-Schmidt orthogonalization
Compute $N = \sum_{i=1}^{n-4} \zeta^i N_i$.
for $j = 1, \dots, 4$ do
  for $k = 1, \dots, 4$ do
    Compute $\tilde{L}_j^k = \sum_{l=1}^{4} g^{kl} \langle N, t_{jl} \rangle$.
  end for
end for
Set $L_\zeta = [\tilde{L}_j^k]_{j,k=1,\dots,4}$.
Compute the maximum eigenvalue $\lambda_\zeta$ of $L_\zeta$.
```

In what follows, we provide an upper bound on the principal curvature.

Proposition 3:

$$\lambda_{\max}(L_\zeta) \leq 4 \frac{\sup_{l,j} \|t_{lj}\|}{\sigma_{\min}(G)}. \tag{27}$$

Proof: It is well known that

$$\lambda_{\max}(L_\zeta) \leq \|L_\zeta\|_2. \tag{28}$$

In light of (23), the entries $\tilde{L}_j^k = \sum_{i=1}^{m} \zeta^i L_{ij}^k$ of L_ζ satisfy

$$\begin{aligned} \tilde{L}_j^k &= \sum_{i=1}^{m} \zeta^i \sum_{l=1}^{4} g^{kl} \langle N_i, t_{jl} \rangle \\ &= \sum_{l=1}^{4} g^{kl} \sum_{i=1}^{m} \zeta^i \langle N_i, t_{jl} \rangle \\ &= \sum_{l=1}^{4} g^{kl} \langle N, t_{jl} \rangle \end{aligned} \tag{29}$$

where we have used (17). Hence, $L_\zeta = G^{-1} \mathcal{N}$ with

$$\mathcal{N} = \begin{bmatrix} \langle N, t_{11} \rangle & \cdots & \langle N, t_{41} \rangle \\ \vdots & & \vdots \\ \langle N, t_{14} \rangle & \cdots & \langle N, t_{44} \rangle \end{bmatrix}.$$

This implies

$$\begin{aligned}\|L_\zeta\|_2 &\le \|G^{-1}\|_2\|\mathcal{N}\|_F \\ &\le \frac{4}{\sigma_{\min}(G)}\sup_{l,j}|\langle N, t_{lj}\rangle| \\ &\le \frac{4}{\sigma_{\min}(G)}\sup_{l,j}\|t_{lj}\|\end{aligned}$$

using the fact that N is a unit vector. ■

Observe that $\|t_{jl}\|$ is finite, since the mother function ϕ is in C^∞. Also, G has full rank, which implies that $\sigma_{\min}(G) > 0$. Therefore, the upper bound (27) is finite. Assuming that there are no (near) self-intersections, this bound on the principal curvature can be used instead of the condition number for analyzing dimensionality reduction of manifolds with random measurements. This implies that the transformation manifolds (9), which we consider in this work, cannot be too much curved and are generally expected to be well behaved. This is also verified in practice as we will show in the experiments section below.

Based on the developments above, Algorithm 1 provides an efficient numerical procedure for computing the principal curvature at a point $s(\eta)$ on the manifold along a certain normal direction ζ. The algorithm makes use of the compact equation (29) for computing the entries of L_ζ. It is important to stress that the tangent vectors t_i as well as the mixed partial derivatives t_{ij} can be computed analytically (i.e., without any approximation) thanks to the closed-form expression of the manifold equation; see (11). Hence, one completely avoids the drawbacks arising from a finite difference approximation, such as noise sensitivity. The details of the computation are given in the Appendix.

VI. EXPERIMENTAL RESULTS

Here, we evaluate the performance of our image alignment algorithm in different problems. We first present illustrative alignment experiments of a facial image, along with the estimated transformations and their corresponding illustrations. Second, we provide alignment experiments that show the alignment performance of the proposed algorithm with respect to the number of measurements. Finally, we fix the number of random measurements and we study the behavior of the proposed approach in the context of facial image alignment towards transformation-invariant face recognition. In all experiments, the entries of the measurement matrices Z follow a standard Gaussian distribution $\mathcal{N}(0,1)$.

A. Illustrative Alignment Experiments

In the first experiments, the pattern s is the facial image shown in Fig. 5. We consider $\eta^* = (b^*, \alpha^*, \omega^*)$ to be a synthesis of translation $b_x^* = 2$ and $b_y^* = 6$, isotropic scaling $\alpha^* = 0.8$, and rotation $\omega^* = 5\pi/4 \approx 3.93$. We build the query image p using (11), by applying the exact transformation η^* to s. Then, we use the cutting plane method to estimate the transformation for increasing number of random measurements m.

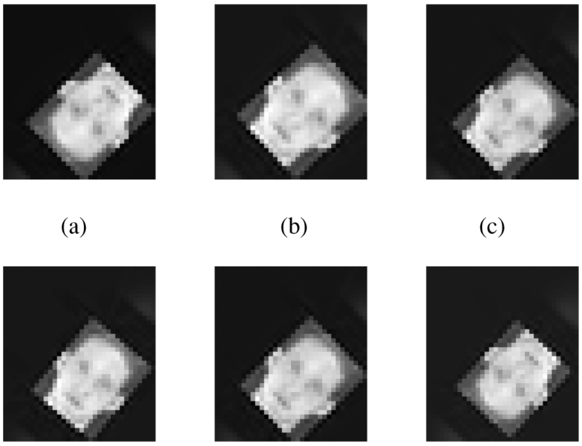

(a) (b) (c)
(d) (e) (f)

Fig. 9. Transformed images corresponding to the Table I. (a) Exact transformed image $s(\eta^*)$. (b)–(f) Estimated transformed images $s(\hat{\eta})$ along with the corresponding number of measurements m. (a) $s(\eta^*)$. (b) $m = 20$. (c) $m = 40$. (d) $m = 60$. (e) $m = 80$. (f) $m = 100$.

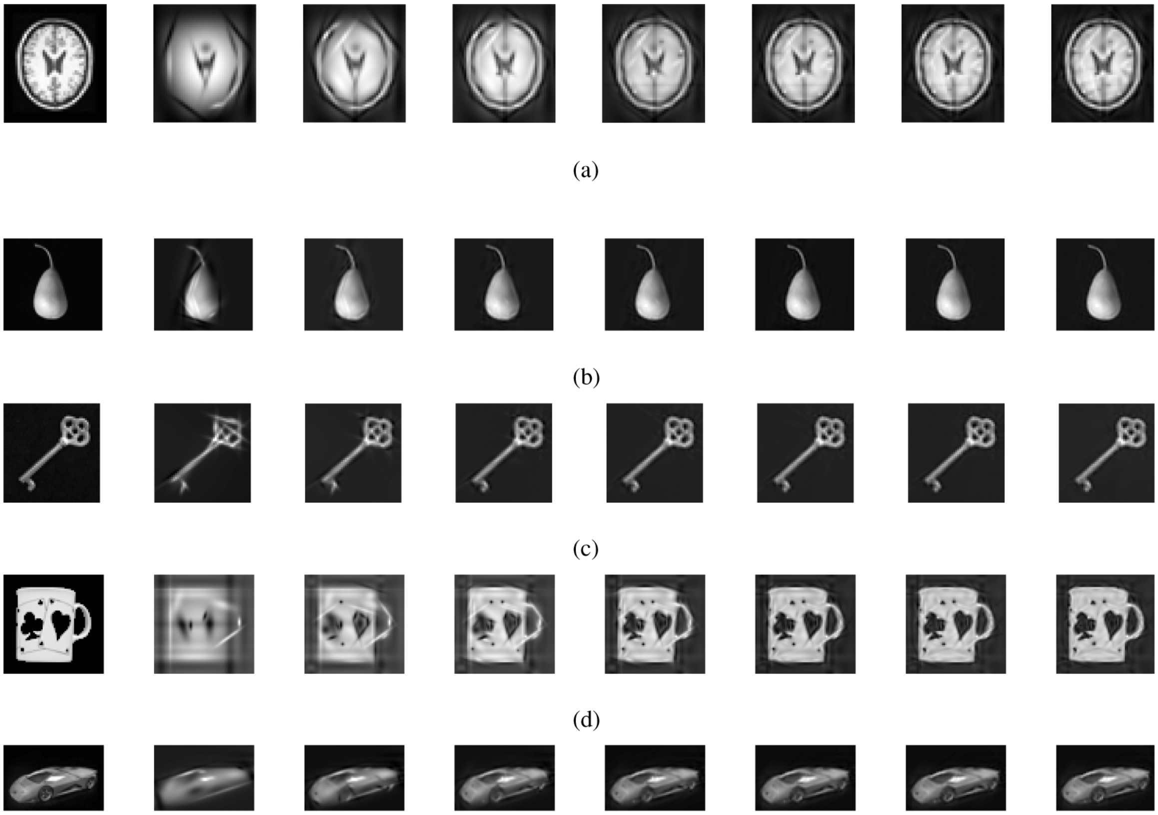
TABLE I
TRANSFORMATION ESTIMATES FOR THE FACE MANIFOLD WITH RESPECT TO THE NUMBER OF MEASUREMENTS m

m	$\eta^* = (\beta_x^*, \beta_y^*, \alpha^*, \omega^*)$	$\hat{\eta} = (\hat{\beta}_x, \hat{\beta}_y, \hat{\alpha}, \hat{\omega})$
20	(2, 6, 0.8, 3.93)	(2.68, 3.62, 0.85, 0.82)
40	(2, 6, 0.8, 3.93)	(2.26, 5.3, 0.79, 0.8)
60	(2, 6, 0.8, 3.93)	(2.28, 5.06, 0.76, 0.7)
80	(2, 6, 0.8, 3.93)	(2.26, 5.3, 0.79, 0.80)
100	(2, 6, 0.8, 3.93)	(1.97, 6.09, 0.73, 3.88)

We report in Table I the estimates $\hat{\eta}$ of the exact transformation η^* with respect to m, obtained after 1000 iterations of the cutting plane method. Fig. 9 shows the corresponding transformed images $s(\hat{\eta})$ (using 200 atoms for better illustration). One can see that the rotation angle ω seems to be the toughest parameter to be estimate accurately, since the algorithm already provides reasonably good estimates for the rest of the transformation parameters as soon as $m \geq 40$ measurements are provided. Observe also that the transformation estimate $\hat{\eta}$ obtained with $m = 100$ measurements is very close to the exact transformation η^*, which verifies in practice the global optimality properties of the proposed algorithm when a sufficient number of measurements is provided.

B. Alignment With Random Measurements

We provide below systematic alignment experiments, where we use the facial image shown in Fig. 5 as well as five additional images collected from the Web (Google image collection) and from the ETH-80 dataset. Fig. 10 shows the images on the leftmost panels and their corresponding progressive approximations obtained using OMP with increasing number of Gaussian atoms. Note that, for the purpose of alignment, one does not need a very accurate sparse representation of the pattern. What is really important is to obtain a representation of the pattern, even a crude one, which captures its salient geometric features.



(a)

(b)

(c)

(d)

(e)

Fig. 10. Progressive approximation of the tested images using Gaussian atoms. (a) MRI brain image (leftmost) approximated progressively when the number of atoms ranges from 20 to 140 with step 20. (b) Pear image (leftmost) approximated progressively when the number of atoms ranges from 20 to 200 with step 30. (c) Key image (leftmost) approximated progressively when the number of atoms ranges from 20 to 200 with step 30. (d) Cup image (leftmost) approximated progressively when the number of atoms ranges from 20 to 200 with step 30. (e) Car image (leftmost) approximated progressively when the number of atoms ranges from 20 to 200 with step 30.

Observe in Fig. 10 that, although the tested images have different characteristics, the obtained sparse representations based on Gaussian atoms are rather successful in capturing the salient geometric features that are crucial for the alignment problem. In all our experiments, the number of Gaussian atoms is set to $K = 40$ (for all images).

1) Numerical Estimation of the Principal Curvature: First, we estimate numerically the principal curvature of the six transformation manifolds, using the numerical algorithm presented in Section V. We uniformly discretize the transformation parameter space $[0, 2\pi) \times [0.5, 1.5]$ using 10 rotation angles and 11 scaling levels, respectively. Then, for each grid point (ω, α), we estimate the principal curve using 40 random realizations of the normal direction ζ. Fig. 11 shows the histogram of the computed curvature values of all image manifolds, as well as the corresponding median value.

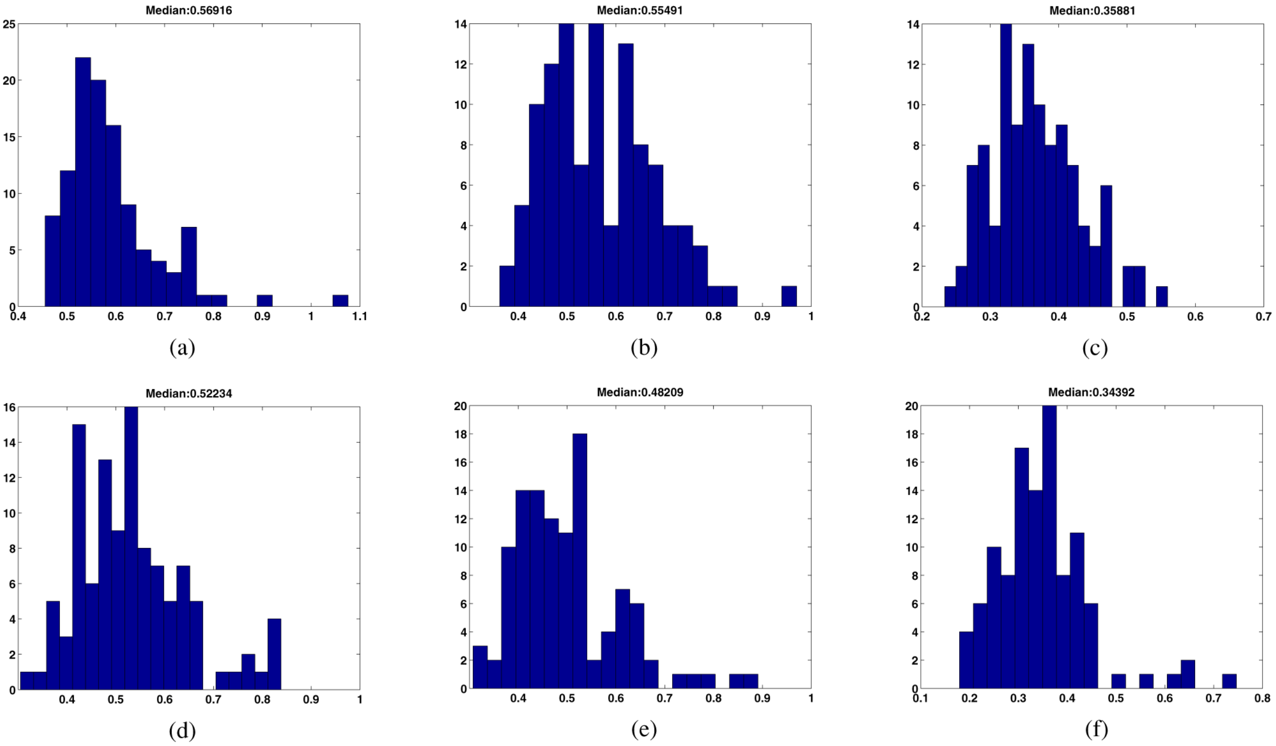
One sees that the curvature values are small, implying that the parametric transformation manifolds studied in this work are not very curved, and therefore are well behaved (see also Proposition 3). Notice that all images have similar curvature values, i.e., of the same scale. One possible explanation for this is the common structure of the image transformation manifolds, i.e., each point on the manifold consists of a linear combination of Gaussians that are very smooth functions.

2) Alignment Experiments: We consider η to be a synthesis of an isotropic scaling $\alpha \in [0.5, 1.5]$ and rotation $\omega \in [0, 2\pi)$. Each query image p is built by image warping on s, using the exact geometric transformation η^*. We run 40 random experiments with random transformations η^* and different random realizations of the measurement matrix Z. Then, for each random experiment, 100 iterations of the cutting plane method are employed to align s with p. We compute the alignment error of the estimated transformation $\hat{\eta}$ as follows:

$$e = \min\{2\pi - |\hat{\omega} - \omega^*|, |\hat{\omega} - \omega^*|\} + |\hat{\alpha} - \alpha^*|. \tag{30}$$

Fig. 12 shows the statistics of the alignment error (30) with respect to the number m of random projections used for image alignment, for all images.

We observe that the alignment error drops quickly when the number of random measurements increases and then it saturates around 0.1. The nonzero alignment error is due to the fact that only 100 iterations of the cutting plane are employed. Furthermore, the experimental results show that for all manifolds a few measurements (e.g., $m \geq 50$) are sufficient to enable the cutting plane method to reach the vicinity of the exact transformation in the vast majority of cases. Given the fact that all images have similar curvature values, it comes to no surprise that the alignment performance of the algorithm is similar in all images. This



(a) (b) (c)

(d) (e) (f)

Fig. 11. Histogram of the principal curvature values of all tested manifolds. (a) Face. (b) Brain. (c) Cup. (d) Car. (e) Pear. (f) Key.

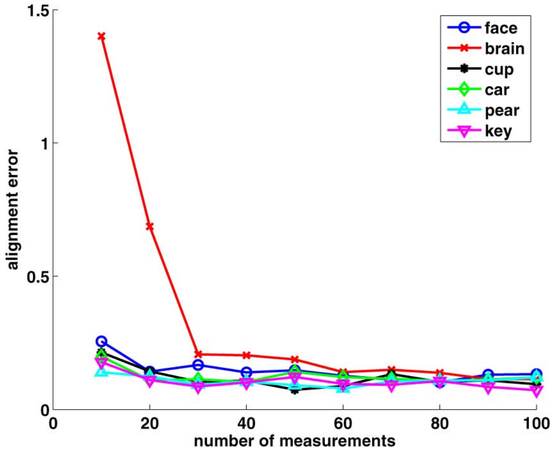


Fig. 12. Median of the alignment error (30) of all tested image manifolds as a function of the number of measurements.

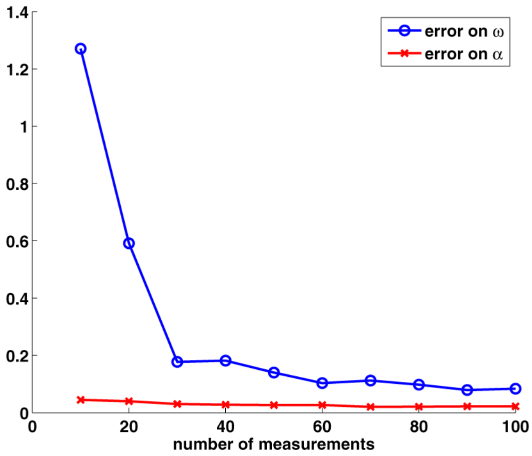


Fig. 13. Median of the two error components: rotation angle ω and scaling α in (30) for the brain image manifold, as a function of the number of measurements.

is consistent with Theorem 1. The only exception is the brain image. In what follows, we discuss further this interesting case.

We see that the pattern geometry of the brain image [see Fig. 10(a)] is almost symmetrical for rotation angles that differ by π. Hence, resolving the correct rotation angle requires more random measurements. To see why this is the case for the brain manifold, we plot the rotation component and the scaling component of the alignment error (30) of the brain image experiment in two different curves in Fig. 13. Notice that the scaling parameter is estimated rather accurately even with very few measurements. However, the rotation angle parameter requires more measurements.

The experiment of the brain manifold implies that, in general, there may exist cases where the manifold nearly intersects itself; see Fig. 14 for a pictorial illustration of such a situation. In such cases, estimating the correct transformation requires more random measurements, as it imposes stricter isometry conditions. Note that the definition of the manifold condition number in [16] does actually take this situation into account. Therefore, this experiment is also consistent with Theorem 1.

To summarize, the bound of Theorem 1 provides qualitative insights into the alignment behaviour, even though it might not admit tight bounds on the number m of required measurements. Note that empirical algorithms are typically used to determine m directly (see, e.g., [8]). Overall, we have seen that for all

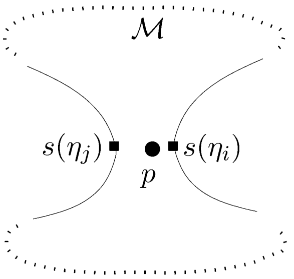


Fig. 14. Case where the manifold $\mathcal{M}$ nearly intersects itself.

image manifolds that have been tested, the curvature is bounded; hence a few random measurements are in general sufficient to reach the vicinity of the global minimizer with the cutting plane method in the image alignment problem.

C. Facial Image Alignment With Random Measurements

In these final experiments, we analyze the performance of the image alignment method in the context of invariant face recognition. Baseline face recognition schemes use nearest neighbor (NN) classifiers in order to predict the correct identity of a novel test facial image, given a set of training facial images associated with the corresponding people identities. Typically, the NN classifier computes the Euclidean distance of the test image from each of the training images and simply predicts the identity of the test image by that of the closest training image. Clearly, successful image alignment prior to distance computation has a major impact on the recognition performance.

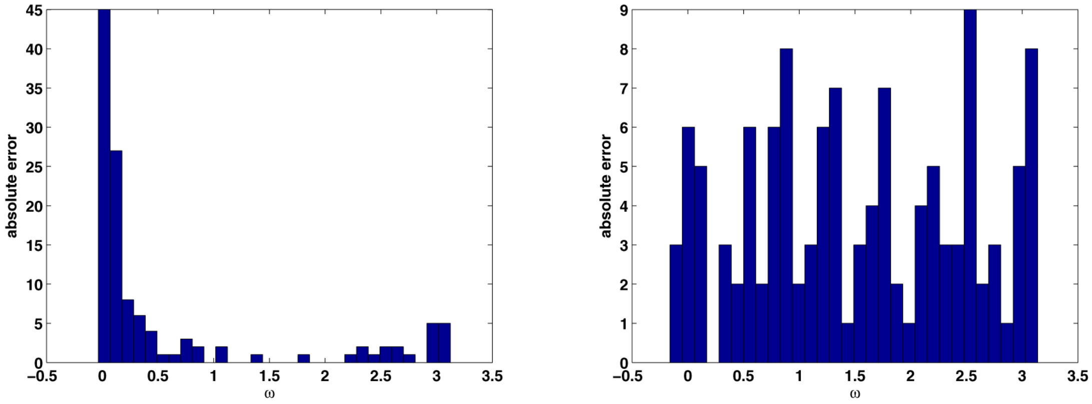
We use the ORL database [24], which is a well-known real-world data set from the face recognition literature. It contains 40 individuals with ten facial images per person. The image set of each individual includes variations in facial expression (smiling/nonsmiling) and head pose (see Fig. 15 for an illustration). We assign the first seven images from each person to the training set and the remaining three to the test set. Hence, the test set has in total $3 \times 40 = 120$ images. Note that, in this setup, the training and test image involved in each alignment experiment, correspond to different facial expressions and poses of the same person or different persons. This makes their alignment even more challenging. We compute first the sparse representation of each training image using $K = 50$ Gaussian atoms. No further processing is done on the training images. Each test image is geometrically transformed by image warping with a random transformation consisting of scale α and rotation ω. Scale and rotation are uniformly distributed in $[0.5, 1.5]$ and $[0, 2\pi)$, respectively.

We study the benefit of image alignment in the face recognition problem by comparing two different methods for the image alignment. We first use a standard Newton algorithm to solve the optimization problem in Section II. Then, we form a hybrid method coined "DC + Newton"that simply consists of the cutting plane method proposed in this paper, followed by Newton's method. This algorithm permits to relax part of the complexity of the cutting plane methods by limiting the number of iterations and to use a fast Newton method for finer alignment. In particular, the hybrid technique first estimates the unknown transformation using the cutting plane method and then provides it as an initial guess to the Newton's method in order to get a more refined estimate. In the hybrid DC + Newton method, we use 100 iterations of the cutting plane before switching to Newton. We use $m = 50$ random measurements for image alignment in both alignment methods. We note finally that for computational convenience, the test image is aligned with only one training image from each person, instead of aligning it with all available training images of the same person. The obtained transformation is then used to align the test image for comparison with the remaining training images of the same person. With this setup, the experiment involves $120 \times 40 = 4800$ image alignment experiments in total.

Fig. 15. Samples from the ORL data set.

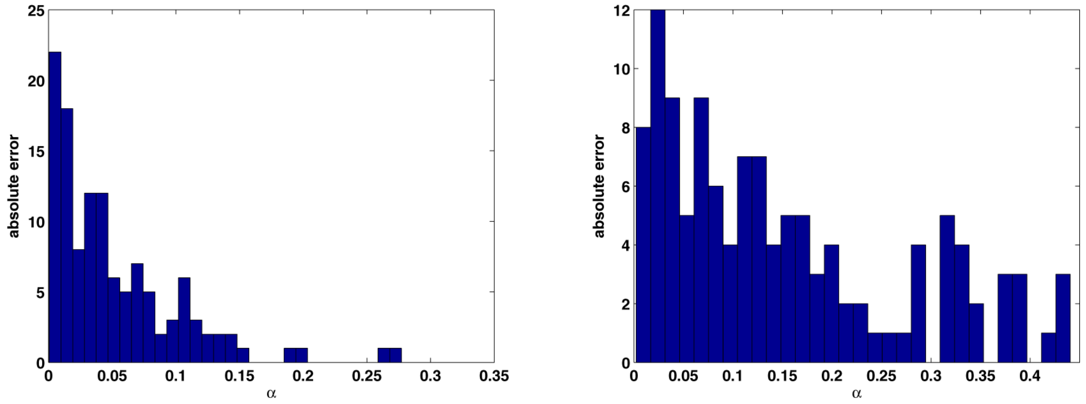
We first compare the alignment performance of the proposed hybrid method with respect to Newton's method. Let $e_r = \min\{2\pi - |\hat{\omega} - \omega^*|, |\hat{\omega} - \omega^*|\}$ and $e_s = |\hat{\alpha} - \alpha^*|$ be the rotation and scale component, respectively, of the registration error (30). Figs. 16 and 17 show the histograms (with 30 bins) of e_r and e_s for each method. We have observed that, when two facial images from *different* persons are aligned, it happens quite often that the global minimizer is far from the exact transformation parameters. For this reason, in the histogram computation we have included only the alignment experiments corresponding to pairs of images of the *same* individual. As expected, one can see from the figures above that Newton's method is stuck in local minima resulting in high errors. On the contrary, the DC + Newton method is successful in finding the global minimizer in the large majority of the cases. The few outlier cases are due to the fact that we use only 100 iterations of the cutting plane method (i.e., before switching to Newton) or due to the low number of random measurements ($m = 50$ in our experiments).

For the sake of completeness, we finally report the face recognition performances, although the main purpose of this experiment is to study the image alignment performance of the methods and not their face recognition one. The NN classifier without any alignments results in 97.5% misclassification rate, which is simply the percentage of test images that have been misrecognized. This is reasonable given the presence of large transformations in the test image set. When standard Newton is employed for alignment, the misclassification rate slightly improves to 87.5%. Finally, when the hybrid DC+Newton method is used for alignment, it dramatically drops to 45.8% since the latter is much more effective in image alignment. This figure should be compared with a similar experiment performed in [11], where the hybrid method achieves an error rate of 26.6% when using full images instead of random measurements. Therefore, one could try to increase the number of random measurements used in the hybrid DC + Newton method, such that its face recognition performance comes closer to the one reported in [11]. However, this is out of the scope of the current



(a) (b)

Fig. 16. Facial image alignment experiment: histogram of the absolute error of the rotation parameter ω. (a) DC + Newton. (b) Newton.



(a) (b)

Fig. 17. Facial image alignment experiment: histogram of the absolute error of the scale parameter α. (a) DC + Newton. (b) Newton.

work whose main objective is to show that image alignment can be performed from random measurements, and the last experiment verifies in practice that successful alignment has a significant impact on the performance of the NN classifier towards invariant face recognition.

VII. RELATED WORK

In what follows, we review the most relevant work from the literature and put it in perspective with the work proposed in this paper.

We first mention the approaches in [25] and [26] for image alignment based on random projections. In particular, the approach in [26] is used in the context of compressive classification, where the images are considered to be geometrically transformed and the problem is to perform invariant classification based on their random projections. Although the two approaches in [25] and [26] may look different at a first glance, they are both based on exhaustive search, by discretizing the transformation parameter space and estimating the transformation by nearest neighbor search among the images after random projection. This not only leads to high memory requirements, but it also provides no theoretical guarantee for the optimality of the attained solution. In contrast, our DC approach enjoys global optimality guarantees.

The authors in [8] propose a linear dimensionality reduction methodology based on random projections. They show that a few random projections are sufficient to estimate the intrinsic dimension of the manifold. They also provide an empirical procedure for estimating the number of necessary random projections in the context of ISOMAP. However, their procedure is particularly designed for the ISOMAP algorithm and does not easily extend to other manifold learning algorithms, or more generally, to other image analysis problems.

An upper bound somewhat similar to (27) has been derived by Jacques and De Vleeschouwer in [27], where it is used towards alleviating the dictionary discretization effects in Matching Pursuit algorithms. Note, however, that the bound in [27] assumes a different definition of curvature and only holds for the case where the manifold represents a parametric dictionary of the form (4).

The authors in [28] study the nondifferentiability of manifolds spanned by natural images. They show that nondifferentiability arises due to the movement of sharp edges (causing global nondifferentiability) and due to the occlusion (causing local nondifferentiability). The manifolds considered in this work are differentiable, thanks to their parametric nature and the smoothness of the Gaussian mother function used in our framework.

Finally, we mention the approach of [29], which has been shown to be globally optimal in determining general nonrigid transformations in the case where one has access to the test images and not only linear measurements. The method employs a set of training images and their (known) corresponding deformation parameters in order to estimate the deformation parameters of a new test image. This is done by iteratively computing the closest training image in the image space and switching between the deformation parameter space and image space in each iteration. The intuition is that each iteration of the algorithm results in a less deformed image and the authors prove that their algorithm is globally optimal, under the assumption that the maximum allowed deformation is bounded. However, in the case of large distortions a very high number of training images is required, as pointed out in [29, Sec. 6]. On the contrary, the computational requirements of our method are independent of the magnitude of the transformation.

VIII. CONCLUSION

We have proposed a globally optimal method for image alignment with random projections, when the geometric transformation consists of a synthesis of shift, rotation and isotropic scaling. We build on previous work and use sparse geometric expansions to represent the transformation manifold, which describes the transformed versions of a pattern. We formulate the image alignment problem with random projections as a DC program, by proving that the objective function is DC when the dictionary is built on Gaussian functions. In addition, we provide theoretical as well as numerical insights into the geometric properties of transformation manifolds, by deriving an upper bound on the principal curvature as well as establishing an efficient numerical algorithm for computing it in practice. We show that the transformation manifolds are well behaved, so that the image alignment problem can be solved with a bounded number of measurements. This is confirmed by experimental results where the proposed method is shown to be successful in finding the global minimizer in practice, even with a small number of random projections. Finally, the proposed method can be combined with local optimization in order to refine the estimated transformation or to cope with more complicated transformation groups such as nonrigid transformations.

APPENDIX

A. *Proof of* $\langle (\partial N_i)/(\partial \eta_j), t_l \rangle = -\langle N_i, t_{lj} \rangle$

Proof: Observe first that $\langle N_i, (\partial s)/(\partial \eta_l) \rangle = 0$, by definition of the normal vector. Therefore

$$\begin{aligned} 0 &= \frac{\partial}{\partial \eta_j} \left\langle N_i, \frac{\partial s}{\partial \eta_l} \right\rangle \\ &= \left\langle \frac{\partial N_i}{\partial \eta_j}, \frac{\partial s}{\partial \eta_l} \right\rangle + \left\langle N_i, \frac{\partial^2 s}{\partial \eta_l \partial \eta_j} \right\rangle \\ &= \left\langle \frac{\partial N_i}{\partial \eta_j}, t_l \right\rangle + \langle N_i, t_{lj} \rangle \end{aligned}$$

which implies the result. ■

B. *Computation of* t_j

By definition $s = \sum_{k=1}^{K} \xi_k \phi_{\gamma_k}$, and therefore

$$\frac{\partial s}{\partial \eta_i} = \sum_{k=1}^{K} \xi_k \frac{\partial \phi_{\eta \circ \gamma_k}}{\partial \eta_i}.$$

Recall that $\phi_{\eta \circ \gamma_k} = \phi(\tilde{x}, \tilde{y})$,, where $\tilde{x}, \tilde{y}$ are the transformed coordinates, and therefore

$$\frac{\partial \phi}{\partial \eta_i} = \frac{\partial \phi}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial \eta_i} + \frac{\partial \phi}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial \eta_i}.$$

In the case that ϕ is Gaussian, then

$$\begin{aligned} \frac{\partial \phi}{\partial \tilde{x}} &= -2\tilde{x} \exp(-(\tilde{x}^2 + \tilde{y}^2)) \\ \frac{\partial \phi}{\partial \tilde{y}} &= -2\tilde{y} \exp(-(\tilde{x}^2 + \tilde{y}^2)). \end{aligned}$$

C. *Computation of* t_{ij}

We have

$$\begin{aligned} \frac{\partial^2 \phi}{\partial \eta_i \partial \eta_j} = &\left[\frac{\partial^2 \phi}{\partial^2 \tilde{x}} \frac{\partial \tilde{x}}{\partial \eta_j} + \frac{\partial^2 \phi}{\partial \tilde{x} \partial \tilde{y}} \frac{\partial \tilde{y}}{\partial \eta_j} \right] \frac{\partial \tilde{x}}{\partial \eta_i} + \frac{\partial \phi}{\partial \tilde{x}} \frac{\partial^2 \tilde{x}}{\partial \eta_i \partial \eta_j} \\ &+ \left[\frac{\partial^2 \phi}{\partial \tilde{y} \partial \tilde{x}} \frac{\partial \tilde{x}}{\partial \eta_j} + \frac{\partial^2 \phi}{\partial \tilde{y}^2} \frac{\partial \tilde{y}}{\partial \eta_j} \right] \frac{\partial \tilde{y}}{\partial \eta_i} + \frac{\partial \phi}{\partial \tilde{y}} \frac{\partial^2 \tilde{y}}{\partial \eta_i \partial \eta_j}. \end{aligned}$$

In the case that ϕ is Gaussian, then

$$\begin{aligned} \frac{\partial^2 \phi}{\partial \tilde{x}^2} &= (4\tilde{x}^2 - 2) \exp(-(\tilde{x}^2 + \tilde{y}^2)) \\ \frac{\partial^2 \phi}{\partial \tilde{y}^2} &= (4\tilde{y}^2 - 2) \exp(-(\tilde{x}^2 + \tilde{y}^2)) \\ \frac{\partial^2 \phi}{\partial \tilde{x} \partial \tilde{y}} = \frac{\partial^2 \phi}{\partial \tilde{y} \partial \tilde{x}} &= (4\tilde{x}\tilde{y}) \exp(-(\tilde{x}^2 + \tilde{y}^2)). \end{aligned}$$

D. *Computation of* $(\partial \tilde{x})/(\partial \eta_i)$ *and* $(\partial y)/(\partial \eta_i)$

Recall that $\eta = (b, \alpha, \omega)$ denotes transformation parameters while $\gamma_i = (b_i, a_i, \omega_i)$ denotes atom parameters. The group law is given by

$$\eta \circ \gamma = (b + \alpha R(-\omega) b_i, \alpha a_i, \omega + \omega_i). \tag{31}$$

Denoting

$$\begin{aligned} A &= \begin{bmatrix} 1/a_{ix} & 0 \\ 0 & 1/a_{iy} \end{bmatrix} \\ R(\omega_i) &= \begin{bmatrix} \cos(\omega_i) & \sin(\omega_i) \\ -\sin(\omega_i) & \cos(\omega_i) \end{bmatrix} \end{aligned}$$

we define the matrix $C := AR(\omega_i)$, which does not depend on the transformation parameters. Then, the transformed coordinates of the ith atom are given by

$$\begin{aligned} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} &= C \frac{1}{\alpha} R(\omega) \left[\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} b_x \\ b_y \end{pmatrix} - \alpha R(-\omega) \begin{pmatrix} b_{ix} \\ b_{iy} \end{pmatrix} \right] \\ &= C \left[\frac{1}{\alpha} R(\omega) \begin{pmatrix} x - b_x \\ y - b_y \end{pmatrix} - \begin{pmatrix} b_{ix} \\ b_{iy} \end{pmatrix} \right] \\ &= C \begin{bmatrix} \frac{1}{a}(\cos(\omega)(x - b_x) + \sin(\omega)(y - b_y)) - b_{ix} \\ \frac{1}{a}(-\sin(\omega)(x - b_x) + \cos(\omega)(y - b_y)) - b_{iy} \end{bmatrix}. \end{aligned}$$

Differentiation gives

$$\begin{bmatrix}\partial\tilde{x}/\partial\omega\\ \partial\tilde{y}/\partial\omega\end{bmatrix} = C\begin{bmatrix}\frac{1}{a}(-\sin(\omega)(x-b_x)+\cos(\omega)(y-b_y))\\ \frac{1}{a}(-\cos(\omega)(x-b_x)-\sin(\omega)(y-b_y))\end{bmatrix}$$

$$\begin{bmatrix}\partial\tilde{x}/\partial b_x\\ \partial\tilde{y}/\partial b_x\end{bmatrix} = C\begin{bmatrix}-\frac{1}{a}\cos(\omega)\\ \frac{1}{a}\sin(\omega)\end{bmatrix}$$

$$\begin{bmatrix}\partial\tilde{x}/\partial b_y\\ \partial\tilde{y}/\partial b_y\end{bmatrix} = C\begin{bmatrix}-\frac{1}{a}\sin(\omega)\\ -\frac{1}{a}\cos(\omega)\end{bmatrix}$$

$$\begin{bmatrix}\partial\tilde{x}/\partial\alpha\\ \partial\tilde{y}/\partial\alpha\end{bmatrix} = C\begin{bmatrix}-\frac{1}{a^2}(-\cos(\omega)(x-b_x)+\sin(\omega)(y-b_y))\\ -\frac{1}{a^2}(-\sin(\omega)(x-b_x)+\cos(\omega)(y-b_y))\end{bmatrix}.$$

E. Computation of $(\partial^2 x)/(\partial\eta_i\partial\eta_j)$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial\omega^2\\ \partial^2\tilde{y}/\partial\omega^2\end{bmatrix} = C\begin{bmatrix}\frac{1}{a}(-\cos(\omega)(x-b_x)-\sin(\omega)(y-b_y))\\ \frac{1}{a}(\sin(\omega)(x-b_x)-\cos(\omega)(y-b_y))\end{bmatrix}$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial\omega\partial b_x\\ \partial^2\tilde{y}/\partial\omega\partial b_x\end{bmatrix} = C\begin{bmatrix}\frac{1}{a}\sin(\omega)\\ \frac{1}{a}\cos(\omega)\end{bmatrix}$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial\omega\partial b_y\\ \partial^2\tilde{y}/\partial\omega\partial b_y\end{bmatrix} = C\begin{bmatrix}-\frac{1}{a}\cos(\omega)\\ \frac{1}{a}\sin(\omega)\end{bmatrix}$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial\omega\partial\alpha\\ \partial^2\tilde{y}/\partial\omega\partial\alpha\end{bmatrix} = C\begin{bmatrix}-\frac{1}{a^2}(-\sin(\omega)(x-b_x)+\cos(\omega)(y-b_y))\\ -\frac{1}{a^2}(-\cos(\omega)(x-b_x)-\sin(\omega)(y-b_y))\end{bmatrix}$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial b_x^2\\ \partial^2\tilde{y}/\partial b_x^2\end{bmatrix} = 0$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial b_x\partial b_y\\ \partial^2\tilde{y}/\partial b_x\partial b_y\end{bmatrix} = 0$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial b_x\partial\alpha\\ \partial^2\tilde{y}/\partial b_x\partial\alpha\end{bmatrix} = C\begin{bmatrix}\frac{1}{a^2}\cos(\omega)\\ -\frac{1}{a}\sin(\omega)\end{bmatrix}$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial b_y^2\\ \partial^2\tilde{y}/\partial b_y^2\end{bmatrix} = 0$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial b_y\partial\alpha\\ \partial^2\tilde{y}/\partial b_y\partial\alpha\end{bmatrix} = C\begin{bmatrix}\frac{1}{a^2}\sin(\omega)\\ \frac{1}{a^2}\cos(\omega)\end{bmatrix}$$

$$\begin{bmatrix}\partial^2\tilde{x}/\partial\alpha^2\\ \partial^2\tilde{y}/\partial\alpha^2\end{bmatrix} = C\begin{bmatrix}\frac{2}{a^3}(\cos(\omega)(x-b_x)+\sin(\omega)(y-b_y))\\ \frac{2}{a^3}(-\sin(\omega)(x-b_x)+\cos(\omega)(y-b_y))\end{bmatrix}.$$

ACKNOWLEDGMENT

The authors would like to thank Dr. X. Claeys for insightful discussions on differential geometry and E. Vural for very fruitful discussions on the alignment problem and helpful comments on the manuscript.

REFERENCES

[1] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[2] R. Muise, "Compressive imaging: An application," *SIAM J. Imaging Sci.*, vol. 2, no. 4, pp. 1255–1276, 2009.

[3] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[4] B. Rinner, T. Winkler, W. Schriebl, M. Quaritsch, and W. Wolf, "The evolution from single to pervasive smart cameras," in *Proc. 2nd ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC)*, 2008, pp. 1–10.

[5] B. Rinner and W. Wolf, "An introduction to distributed smart cameras," *Proc. IEEE*, vol. 96, no. 10, pp. 1565–1575, Oct. 2008.

[6] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "Wireless multimedia sensor networks: Applications and testbeds," *Proc. IEEE*, vol. 96, no. 10, pp. 1588–1605, Oct. 2008.

[7] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found Comput. Math.*, vol. 9, no. 1, pp. 51–77, Feb. 2009.

[8] C. Hegde, M. B. Wakin, and R. G. Baraniuk, "Random projections for manifold learning," in *Proc. Neural Inf. Process. Syst. (NIPS) Conf.*, 2007.

[9] M. B. Wakin and R. G. Baraniuk, "Random projections of signal manifolds," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2006, vol. V, pp. 941–944.

[10] E. Kokiopoulou, D. Kressner, and P. Frossard, "Optimal image alignment with random measurements," in *Proc. 17th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2009.

[11] E. Kokiopoulou and P. Frossard, "Minimum distance between pattern transformation manifolds: Algorithm and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1225–1238, Jul. 2009.

[12] J. Maintz and M. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–16, Mar. 1998.

[13] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.

[14] R. Horst and N. V. Thoai, "DC programming: Overview," *J. Optim. Theory Appl.*, vol. 102, no. 1, pp. 1–43, Oct. 1999.

[15] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of johnson and lindenstrauss," *Random Struct. Alg.*, vol. 22, no. 1, pp. 60–65, Jan. 2003.

[16] P. Niyogi, S. Smale, and S. Weinberger, "Finding the homology of submanifolds with high confidence from random samples," *Discrete Computat. Geometry*, vol. 39, no. 1–3, pp. 419–441, Mar. 2008.

[17] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1998.

[18] R. Figueras i Ventura, P. Vandergheynst, and P. Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 726–739, Mar. 2006.

[19] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, 2nd ed. Norwell, MA: Kluwer, 2000, vol. 48, Nonconvex Optimization and Its Applications.

[20] R. Horst and P. M. Pardalos, *Handbook of Global Optimization*. Norwell, MA: Kluwer, 1995, vol. 2, Nonconvex Optimization and Its Applications.

[21] L. Thi Hoai An and P. Dinh Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Ann. Oper. Res.*, vol. 133, pp. 23–46, 2005.

[22] M. P. Do Carmo, *Riemannian Geometry*. Cambridge, MA: Birkhäuser, 1992.

[23] S. Lang, *Fundamentals of Differential Geometry*. Berlin, Germany: Springer, 1999.

[24] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *2nd IEEE Workshop Applicat. Comput. Vis.*, Sarasota, FL, Dec. 1994, pp. 138–142.

[25] D. M. Healy and G. K. Rohde, "Fast global image registration using random projections," in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, 2007, pp. 476–479.

[26] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. SPIE*, 2007, p. 6498.

[27] L. Jacques and C. De Vleeschouwer, "A geometrical study of matching pursuit parametrization," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2835–2848, Jul. 2008.

[28] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. SPIE Wavelets XI*, Jul. 2005.

[29] Y. Tian and S. G. Narasimhan, "A globally optimal data-driven approach for image distortion estimation," in *IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2010, pp. 1277–1284.

Effrosyni Kokiopoulou (S'05–M'09) received the Diploma in engineering from the University of Patras, Patras, Greece, in 2002, the M.Sc. degree in computer science from the University of Minnesota, Minneapolis, in 2005, and the Ph.D. degree from the Signal Processing Laboratory (LTS4), Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2008.

In September 2005, she joined as a Ph.D. student the Signal Processing Laboratory (LTS4), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. Since 2009, she has been a Postdoctoral Researcher with the Seminar for Applied Mathematics, ETH, Zurich, Switzerland. Her research interests include multimedia data mining, pattern recognition, computer vision, and numerical linear algebra.

Dr. Kokiopoulou was the recipient of the 2010 ACM Special Interest Group on Multimedia (SIGMM) Award for Outstanding Ph.D. Thesis in Multimedia Computing, Communications and Applications and the EPFL Doctorate Award in 2010.

Daniel Kressner received the Diploma in mathematics from the Technical University of Chemnitz, Chemnitz, Germany, in 2001, and the Ph.D. degree from the Technical University of Berlin, Berlin, Germany, in 2004.

He joined the Seminar for Applied Mathematics, ETH Zurich, Zurich, Switzerland, in 2007 as an Assistant Professor of applied mathematics. His research interests include numerical linear algebra and high-performance computing, with applications in control theory and data analysis. He is the author of one book and approximately 30 journal articles.

Pascal Frossard (S'96–M'01–SM'04) received the M.S. and Ph.D. degrees from the Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 1997 and 2000, respectively, both in electrical engineering.

Between 2001 and 2003, he was a Member of the Research Staff with the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he worked on media coding and streaming technologies. Since 2003, he has been a Professor with EPFL, where he heads the Signal Processing Laboratory (LTS4). His research interests include image representation and coding, visual information analysis, distributed image processing and communications, and media streaming systems.

Dr. Frossard was the General Chair of IEEE ICME 2002 and Packet Video 2007. He has been the Technical Program Chair of EUSIPCO 2008, and a member of the organizing or technical program committees of numerous conferences. He has been an associate editor of the IEEE TRANSACTIONS ON MULTIMEDIA (2004–2010), the IEEE TRANSACTIONS ON IMAGE PROCESSING (2010-), and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2006-). He is an elected member of the IEEE Image and Multidimensional Signal Processing Technical Committee (2007-), the IEEE Visual Signal Processing and Communications Technical Committee (2006-), and the IEEE Multimedia Systems and Applications Technical Committee (2005-). He has served as Vice-Chair of the IEEE Multimedia Communications Technical Committee (2004–2006) and as a member of the IEEE Multimedia Signal Processing Technical Committee (2004–2007). He was the recipient of the Swiss NSF Professorship Award in 2003, the IBM Faculty Award in 2005, and the IBM Exploratory Stream Analytics Innovation Award in 2008.