OPTIMAL IMAGE ALIGNMENT WITH RANDOM MEASUREMENTS

Effrosyni Kokiopoulou[1], *Daniel Kressner*[1] *and Pascal Frossard*[2]

[1]Seminar for Applied Mathematics
Department of Mathematics, ETH
Zurich - 8092, Switzerland

[2]Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4)
Lausanne - 1015, Switzerland

ABSTRACT

We consider the problem of image alignment using random measurements. More specifically, this paper is concerned with estimating a transformation that aligns a given reference image with a query image, assuming that not the images themselves but only random measurements are available. According to the theory behind compressed sensing, random projections of signal manifolds nearly preserve pairwise Euclidean distances when the reduced space is sufficiently large. This suggests that image alignment can be performed effectively based on a sufficient number of random measurements. We build on our previous work in order to show that the corresponding objective function can be decomposed as the difference of two convex functions (DC). Thus, the optimization problem becomes equivalent to a DC program that can be solved by an outer-approximation cutting plane method, which always converges to the globally optimal solution.

1. INTRODUCTION

The problem of image alignment is of paramount importance and enjoys numerous applications in various fields including pattern recognition, computer vision and medical image analysis, to name just a few [1]. The comparison of two visual patterns is generally only meaningful if they are aligned first, so that their distance reflects their structural and geometric differences. Image alignment consists in estimating the relative transformation between patterns. The transformed version of a pattern can be described as a point of a (possibly nonlinear) manifold in a high dimensional space, which is usually called the *transformation manifold*. The manifold distance (MD) is the minimum distance between the query image p and the manifold generated by the reference image s, see Figure 1.

At the same time, a theory of sparse signals has recently emerged, often referred to as compressed sensing (CS). According to this theory, a few random projections of a sparse (or nearly sparse) signal are sufficient to preserve its salient information. Moreover, in [2, 3] it is shown that random projections of signal manifolds result into approximately isometric embeddings i.e., pairwise Euclidean distances are nearly preserved in the reduced space. This framework allows to work with random measurements of a signal rather than the full signal itself. In the context of image alignment, we use CS to estimate the geometric transformation of the reference image and, optionally, to reconstruct the query image with the transformation compensated for. An advantage of using random measurements is reduced computational complexity. Moreover, in certain applications we might not have access to the full images (e.g., due to bandwidth, complexity or storage space constraints), but only to such measurements, which moreover provide a non expensive way to get some minimal information about the images.

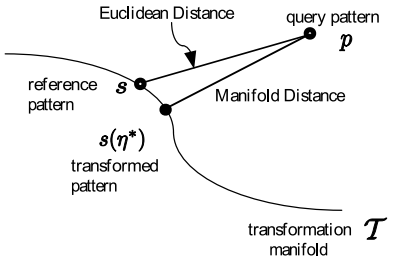


Figure 1: Manifold distance is the minimum distance from a query point p to the transformation manifold $\mathscr{T}$ spanned by the transformed versions of s.

In this paper, we propose a new method for image alignment, which is able to estimate the globally optimal transformation of the reference image s, given a sufficient number of random measurements. We represent s as a sparse linear combination of geometric primitives, called atoms, which are chosen from a parametric, possibly redundant dictionary. The proposed framework also allows for a closed form representation of the transformed image in terms of the transformation parameters. We build on our previous work [4] and formulate the pattern alignment problem with random measurements as a DC program by showing that the objective function is DC, i.e., that it can be written as a *difference of convex functions*. DC programs are non-convex problems that can be solved globally and efficiently by exploiting their special structure. Our approach therefore provides a feasible way to perform image alignment with random measurements. In contrast, existing solutions [5, 6] are based on exhaustive search, which is computationally expensive and provides no theoretical guarantee for the optimality of the attained solution.

The rest of this paper is organized as follows. In Sec. 2 we discuss the representation of transformation manifolds using sparse geometric expansions and in Sec. 3 we formulate the problem of matching random measurements as an optimization problem. We show in Sec. 4 that the objective function is DC. Finally, experimental results are presented in Sec. 5.

2. TRANSFORMATION MANIFOLDS AND RANDOM PROJECTIONS

2.1 Visual pattern representation

In the following, we explain the representation of a pattern as a linear combination of geometric functions (usually

called *atoms*), taken from a structured parametric and possibly redundant dictionary $\mathscr{D} = \{\phi_\gamma,\ \gamma \in \Gamma\}$ spanning the input space. This representation typically allows to capture 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

$$\mathscr{D} = \{\ \phi_\gamma = U(\gamma)\phi,\ \gamma \in \Gamma\}. \tag{1}$$

A transformation γ_i, applied to 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 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 in the two directions i.e., $U(a_i)\phi(x,y) = \phi(\frac{x}{a_{ix}}, \frac{y}{a_{iy}})$.

Composing these transformations yields a transformation $\gamma_i = (b_i, a_i, \omega_i) \in \Gamma$, a synthesis of translations, anisotropic scalings and rotations. It can be observed that applying a transformation on the mother function is equivalent to transforming the coordinate system from $\{x,y\}$ to $\{\tilde{x},\tilde{y}\}$ before applying $\phi(\cdot)$. In particular, when the ith atom in the parametric dictionary (1) is built as $\phi_{\gamma_i} = U(\gamma_i)\phi(x,y)$, where $\gamma_i = (b_i, a_i, \omega_i) \in \Gamma$, it forms the same 2D function as $\phi(\tilde{x},\tilde{y})$, where

$$\begin{aligned}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} &= \underbrace{\begin{bmatrix} \frac{1}{a_{ix}} & 0 \\ 0 & \frac{1}{a_{iy}} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \cos\omega_i & \sin\omega_i \\ -\sin\omega_i & \cos\omega_i \end{bmatrix}}_{R(\omega_i)} \underbrace{\begin{bmatrix} x - b_{ix} \\ y - b_{iy} \end{bmatrix}}_{t} \\ &= AR(\omega_i)t. \end{aligned} \tag{2}$$

The approximation of a pattern s with atoms from the dictionary $\mathscr{D}$ can be obtained in different ways. Even if finding the sparsest approximation of s is generally a hard problem, effective sub-optimal 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) [7, Sec. 9.5.3], which is a simple but 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 jth step the atom ϕ_{γ_j} that best matches the residual r_{j-1} i.e., $\gamma_j = \arg_{\gamma\in\Gamma}\max|\langle r_{j-1}, \phi_\gamma\rangle|$. At each step, it updates the residual by orthogonal projection on the span of the selected atoms (i.e., $r_j = (I - P)r_{j-1}$, where P is the orthogonal projector on the span$\{\phi_{\gamma_1}, \ldots, \phi_{\gamma_j}\}$). After K steps of OMP, the pattern s is approximated by a sparse linear combination of a few atoms i.e.,

$$s \approx \sum_{k=1}^{K} \xi_k \phi_{\gamma_k}. \tag{3}$$

We propose the use of a dictionary of two-dimensional atoms that capture the geometrical information in an image. The generating function ϕ of $\mathscr{D}$ used in this paper is the Gaussian

$$\phi(x,y) = \frac{1}{\rho}\exp(-(x^2 + y^2)). \tag{4}$$

Figure 2 shows the progressive approximation of the digit '5' from a Gaussian dictionary using OMP. Observe that only a few atoms are sufficient to capture the main geometric characteristics of the pattern and the representation (3) does not need to be very accurate before it is useful for alignment purposes.

2.2 Transformation manifolds

In the following, we discuss geometric transformations of the the reference image s. In contrast to the previous section, we restrict scalings to be isotropic, i.e., the geometric transformation η takes the form $\eta = (b, \alpha, \omega)$; that is, it consists of a synthesis of translation $b = [b_x, b_y]$, isotropic scaling α and rotation ω. The manifold $\mathscr{T}$ of all such transformed images can be expressed mathematically as

$$\mathscr{T} = \big\{s(\eta) = U(\eta)s,\ \eta = (b, \alpha, \omega)\big\}. \tag{5}$$

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. Figure 3 shows a few samples from the transformation manifold of the digit "5", when the transformation is a rotation.

All such transformations η form a group, namely the similitude group SIM(2) on the 2D plane. As in (2), we denote

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

as the rotation matrix for the angle ω in the 2D plane. If (b, α, ω) and (b', α', ω') are two elements of the SIM(2) group, then the group law [8] is

$$(b, \alpha, \omega) \circ (b', \alpha', \omega') = (b + \alpha R(-\omega)b', \alpha\alpha', \omega' + \omega). \tag{6}$$

Replacing the reference image s by its approximation (3), applying the transformation η results in

$$s(\eta) = U(\eta)s = \sum_{k=1}^{K} \xi_k U(\eta)\phi_{\gamma_k} = \sum_{k=1}^{K} \xi_k \phi_{\eta\circ\gamma_k}, \tag{7}$$

where $\eta \circ \gamma_k$ is a composition of transformations. In words, the transformation is applied to each constituent atom individually. Furthermore, the group law (6) can be employed to determine the updated parameters of the transformed atoms. Let us emphasize the importance of equation (7): it allows to express the manifold equation (5) in closed form with respect to the transformation parameters η. This is essential for the computation of the manifold distance, and in particular for the applicability of the DC programming methodology that is proposed in the next section.

3. PROBLEM FORMULATION

In this paper, we are interested in estimating the transformation η^* that matches best two visual patterns. As above, let p denote the query pattern and $\mathscr{T}$ the transformation manifold described by equation (5). The main idea is to perform matching between their random projections instead of

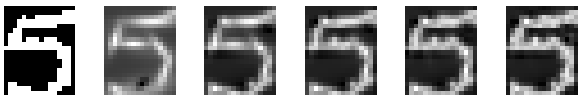


Figure 2: Progressive OMP approximation of the digit "5" (leftmost) with 10, 20, 30, 40 and 50 Gaussian atoms (from left to right).

Figure 3: Samples from the transformation manifold of the digit "5". From left to right, the samples correspond to rotation angles from 0 to 2π with step $\pi/4$.

the visual patterns themselves. Assume that we have M random projections of both the reference pattern s and the query pattern p, obtained by computing inner products with the same M random signals $z_1,\ldots,z_M$. In what follows, we provide first some background material on random projections of manifolds and then we formulate the transformation estimation from random projections as an optimization problem.

3.1 Random projections of manifolds

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 the matching between points in the reduced space is equivalent to matching corresponding points in the initial high-dimensional space, the embedding should be (nearly) isometric i.e., pairwise Euclidean distances should be (nearly) preserved. Only if this is the case then one can reliably perform image alignment in the reduced space and estimate the unknown transformation.

What is a good choice of M? Recall that d is the intrinsic dimension of a manifold (in our case $d=4$, the number of transformation parameters). Recently, Baraniuk and Wakin [2] provide an estimate of M that is linear in d and logarithmic in N, the number of pixels in the image. We revisit the main result from [2].

Theorem 1 *Let $\mathscr{T}$ be a compact d-dimensional manifold in $\mathbb{R}^N$ having volume V and condition number $1/\tau$. Fix $0<\varepsilon<1$ and $0<\rho<1$. Let Z be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ and*

$$M \geq O\Big(\frac{d\log(NV\tau^{-1})\log(\rho^{-1})}{\varepsilon^2}\Big). \tag{8}$$

Suppose $M<N$. Then, with probability exceeding $1-\rho$, the following statement holds: For every pair of points $x,y\in\mathscr{T}$,

$$(1-\varepsilon)\sqrt{\frac{M}{N}} \leq \frac{\|Zx-Zy\|_2}{\|x-y\|_2} \leq (1+\varepsilon)\sqrt{\frac{M}{N}}. \tag{9}$$

Roughly speaking, the theorem is proved by determining a high-resolution sampling on the manifold and then applying the Johnson-Lindenstrauss lemma [9] to the sampled points. The above theorem implies that M depends logarithmically on other properties of the manifold, such as its volume V and condition number $1/\tau$. However, in practice it is hard to know or to estimate these parameters, and therefore empirical algorithms are used for determining M (see e.g., [3]).

3.2 Transformation estimation problem

We formulate the parameter estimation problem as follows[1]

$$\eta^* = \arg\min_{\eta=(b,\alpha,\omega)} f(\eta), \text{ where}$$
$$f(\eta) = \sum_{i=1}^{M} |\langle s(\eta), z_i\rangle - \langle p, z_i\rangle|. \tag{10}$$

Recall that $s(\eta)\in\mathscr{T}$ denotes the transformation of s subject to $\eta=(b,\alpha,\omega)$. We assume that the pattern s has been approximated by a sparse expansion over the dictionary $\mathscr{D}$, see (3). When the atom parameters γ_k in this expansion are fixed, the transformed pattern $s(\eta)$ in (7) provides a parametric pattern model with respect to η. Note that in the optimization problem (10), the query pattern p is *not* expanded but used in its original form.

The optimization problem (10) for determining the best transformation parameters η^* is typically nonlinear and nonconvex [10]. 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. However, it will be shown in Section 4 that the above objective function is a difference of two convex (DC) functions. This allows us to formulate the optimization problem as a DC program and solve it globally by a cutting plane method [11, Thm 5.3].

4. DC DECOMPOSITION

The purpose of this section is to show that the objective function (10) is DC.

4.1 Properties of DC functions

We start with some definitions and basic properties about DC functions [11, 12, 13]. 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

$$f(x) = g(x) - h(x). \tag{11}$$

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 $f(x)=(g(x)+c(x))-$

[1] Note that, in our formulation $f(\eta)$ is based on the 1-norm distance and, strictly speaking, Theorem 1 does not apply. However, by the equivalence of norms in finite-dimensional spaces, a good match in the 1-norm yields a good match in the 2-norm, to which Theorem 1 applies.

$(h(x)+c(x))$ is also DC. We will make use of the following two properties.

Proposition 1 (Properties of DC functions [12, 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) $\sum_{i=1}^{m} \lambda_i f_i = \left[\sum_{\{i:\lambda_i \geq 0\}} \lambda_i g_i - \sum_{\{i:\lambda_i<0\}} \lambda_i h_i\right] - \left[\sum_{\{i:\lambda_i \geq 0\}} \lambda_i h_i - \sum_{\{i:\lambda_i<0\}} \lambda_i g_i\right]$.

(b) $|f| = 2\max\{g,h\} - (g+h)$.

4.2 DC form of the objective function

We now combine Proposition 1 with our previous results [4] to prove the main result of this paper.

Theorem 2 *The objective function f in (10) is DC.*

Proof: Recall that

$$\begin{aligned} f(\eta) &= \sum_{i=1}^{M} |\langle s(\eta), z_i\rangle - \langle p, z_i\rangle| \\ &= \sum_{i=1}^{M} |\sum_{k=1}^{K} \xi_k \langle \phi_{\eta_k}, z_i\rangle - \langle p, z_i\rangle|, \end{aligned} \tag{12}$$

where $\eta_k = \eta \circ \gamma_k$. In [4] 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 1(b), the absolute value of a DC function is DC and hence

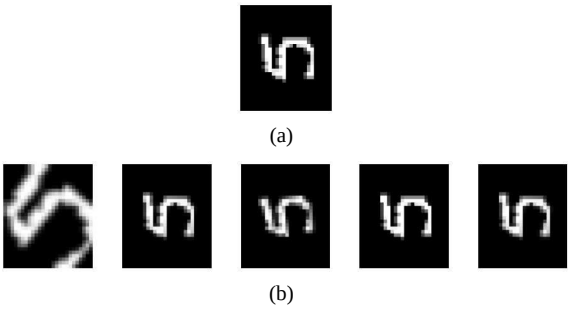
$$|f_i(\eta)| = 2\max\{g_i, h_i\} - (g_i + h_i) = \tilde{g}_i(\eta) - \tilde{h}_i(\eta).$$

is also DC. Finally, the objective function in (12) is DC since it is simply a sum of M DC functions:

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

□

Theorem 2 allows the application of DC programming methods for finding the global minimizer of f. However, a closed form symbolic expression of the DC decomposition of f would be both too inefficient and too complicated to obtain. In contrast, the values of g and h at a specific value of η can be obtained much more easily, by sequentially evaluating the construction in the proof of Theorem 2. In fact, only one of the two functions needs to be evaluated (for example, the evaluation of g yields the corresponding value of h for free and vice versa).



(a)

(b)

Figure 4: (a) Pattern warped with the exact transformation $\omega^* = 3\pi/2$ and $\alpha^* = 0.8$. (b) Estimated transformations $\hat{\eta}$ obtained with M ranging from 10 (leftmost) up to 50 (rightmost) random measurements.

4.3 DC programs

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

$$\begin{aligned} \min_{x} \quad & f(x) = g(x) - h(x), \\ \text{s.t.} \quad & x \in X = \{x \in \mathbb{R}^n : \delta(x) \leq 0\}, \end{aligned} \tag{13}$$

where $g, h : X \to \mathbb{R}$ are convex functions and $\delta : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Assume that (13) is solvable and denote its global minimum by ω^*. The next proposition provides an optimality condition for (13).

Proposition 2 ([11]) *The point $x^* \in X$ is an optimal solution to the DC problem (13) if and only if there exists $t^* \in \mathbb{R}$ such that*

$$0 = \inf\{-h(x) + t :\ x \in X,\ t \in \mathbb{R},\ g(x) - t \leq g(x^*) - t^*\}. \tag{14}$$

In this work, we have chosen to solve the DC Program (13) by the outer approximation cutting plane algorithm proposed in [11, 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 [11, Sec 5.1, Sec 5.2] and DCA [14].

5. EXPERIMENTAL RESULTS

We use as pattern s the handwritten digit image '5' shown in Fig. 2. We build a pattern model of s using OMP with $K = 20$ Gaussian atoms. Observe that already a few atoms are sufficient to capture the main geometric structure of the pattern. In our experiments we consider η to be a synthesis of an isotropic scaling $\alpha \in [0.5, 1.5]$ and rotation $\omega \in [0, 2\pi)$.

We run 40 random experiments with random transformations η and different random realizations of the measurement matrix Z. Each query image q is built with image warping by applying the geometric transformation on s. Then, for each random experiment, 200 iterations of the cutting plane method are employed to align s with q. We compute the relative error of the estimated transformation $\hat{\eta}$ as follows

$$e_r = \frac{|\hat{\omega} - \omega^*|}{\omega^*} + \frac{|\hat{\alpha} - \alpha^*|}{\alpha^*}, \tag{15}$$

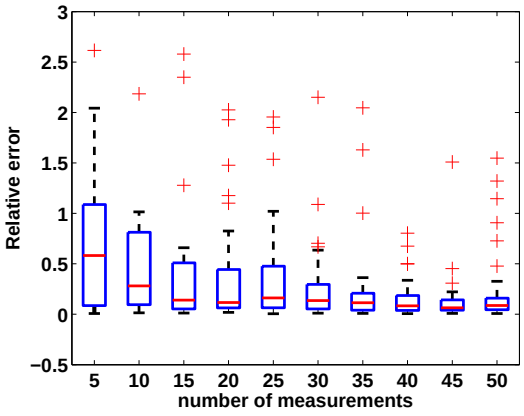


Figure 5: Statistics of the relative error e_r with Rademacher random measurements.

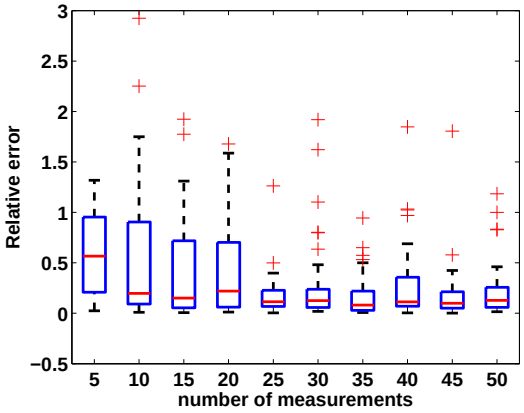


Figure 6: Statistics of the relative error e_r with Gaussian random measurements.

where η^* is the exact transformation.

We consider two different choices of distribution for building the measurement matrices Z; (i) the Rademacher distribution i.e., ± 1 with equal probability and (ii) standard Gaussian distribution $\mathcal{N}(0,1)$. Figures 5 and 6 show the statistics of the relative error e_r (see eq. (15)) in boxplot notation, with Rademacher and Gaussian distributions respectively. For each value of N we run 40 random experiments. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers and represented by crosses. Observe that 5 random measurements are not enough in practice to lead to reliable alignment of this image. Furthermore, the experimental results show that 15 or more measurements are sufficient to enable the cutting plane method to reach the vicinity of the exact transformation in the vast majority of cases.

6. CONCLUSIONS

We have proposed a globally optimal method for image alignment with random measurements. We build on previous work and use sparse geometric expansions to represent the transformation manifold, which is spanned by transformed version of a pattern. A few atoms are sufficient to capture the main geometric structure of the pattern, which is further used for alignment. We formulate the image alignment problem with random measurements as a DC program, by proving that the objective function is DC. The experimental results show that the proposed method is successful in finding the global minimizer in practice, when sufficient number of measurements are provided.

REFERENCES

[1] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

[2] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*. to appear.

[3] C. Hegde, M. B. Wakin, and R. G. Baraniuk. Random projections for manifold learning. *Neural Information Processing Systems (NIPS)*, 2007.

[4] E. Kokiopoulou and P. Frossard. Minimum distance between pattern transformation manifolds: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008. to appear.

[5] 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. *Proceedings of SPIE*, 6498, 2007.

[6] D. M. Healy and G. K. Rohde. Fast global image registration using random projections. *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 476–479, 2007.

[7] S. Mallat. *A Wavelet Tour of Signal Processing, 2nd edn*. Academic Press, 1998.

[8] R. Figueras i Ventura, P. Vandergheynst, and P. Frossard. Low rate and flexible image coding with redundant representations. *IEEE Transactions on Image Processing*, 15(3):726–739, March 2006.

[9] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random structures and algorithms*, 22(1):60–65, 2003.

[10] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[11] R. Horst and N. V. Thoai. DC Programming: Overview. *Journal of Optimization Theory and Applications*, 102(1):1–43, October 1999.

[12] R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization*, volume 48 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 2nd edition, 2000.

[13] R. Horst and P. M. Pardalos. *Handbook of Global Optimization*, volume 2 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 1995.

[14] 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. *Annals of Operations Research*, 133:23–46, 2005.