LEARNING PATTERN TRANSFORMATION MANIFOLDS WITH PARAMETRIC ATOM SELECTION

Elif Vural and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4)
Switzerland-1015 Lausanne

ABSTRACT

We address the problem of building a manifold in order to represent a set of geometrically transformed images by selecting a good common sparse approximation of them with parametric atoms. We propose a greedy method to construct a representative pattern such that the total distance between the transformation manifold of the representative pattern and the input images is minimized. In the progressive construction of the pattern we select atoms from a continuous dictionary by optimizing the atom parameters. Experimental results suggest that the representative pattern built with the proposed method provides an accurate representation of data, where the invariance to geometric transformations is achieved due to the transformation manifold model.

***Keywords*—** Manifold learning, pattern transformation manifolds, dimensionality reduction, sparse signal approximations.

1. INTRODUCTION

The selection of low dimensional structures in high-dimensional data sets is generally referred to as manifold learning. Dimensionality reduction methods such as ISOMAP [1] compute a global parameterization of data based on the preservation of geodesic distances, while the manifold structure in signal sets can also be retrieved via locally linear embeddings as in LLE [2]. In this study, we build on our previous work [3] and examine the problem of constructing a manifold from parametrizable atoms, which efficiently approximates observations of a geometrically transformed signal. In particular, we consider the learning of a pattern transformation manifold, which is a family of image signals that are geometrically transformed versions of a reference pattern. This manifold learning process becomes equivalent to identifying a good reference pattern that is representative of the input images.

We propose a procedure for building the reference pattern progressively by iterative selection of parametric atoms. The atoms are obtained from a mother function via rotation, translation and anisotropic scaling; this corresponds to a manifold-structured dictionary with infinite cardinality and an intrinsic dimension of five. The selection of atoms in the construction of the representative pattern is formulated as an optimization problem in six variables, where five variables stand for atom parameters and the sixth variable is the atom coefficient. In order to minimize the total distance of input images to the computed pattern transformation manifold, we define a cost function in the form of a DC (Difference-of-Convex) function. DC functions have the special property that their global minima can be computed using DC solver algorithms [4]. We perform a preliminary optimization using DC programming and then refine this solution under local linearity assumptions of the transformation manifold. Experimental results show that the data approximation error is minimized gradually with the progressive construction of the representative pattern, and the constructed pattern transformation manifold approximates the input data accurately.

Since we achieve a greedy construction of the representative pattern, our method bears resemblance to sparse approximation algorithms such as Matching Pursuit (MP) [5] or Simultaneous Orthogonal Matching Pursuit (SOMP) [6]. However, the main contributions of this work lie in the following. Firstly, we obtain a transformation-invariant sparse approximation of signals due to the transformation manifold model. Secondly, we perform an optimization of the atom parameters, which corresponds to learning atoms from a dictionary manifold, while MP and SOMP pick atoms in a predefined discrete dictionary. On the other hand, dictionary learning methods such as K-SVD [7] involve alternating applications of a sparse coding stage and a codebook update stage, and the latter corresponds to the optimization of atoms used in signal approximation. The optimization procedure that we apply for selecting an atom in a dictionary manifold resembles this dictionary update stage, which forms a bridge between our manifold building framework and dictionary learning. Finally, we note that our work is also related to transformation-invariant dictionary design, where a sparse representation of signals is sought not only in terms of the original dictionary atoms but also in their geometrically transformed versions. So far transformation-invariant dictionary design has been mostly studied for shift-invariance as in [8] and [9]. We examine the problem of generating atoms that not only contribute to the sparse approximation of a group of signals but also assure invariance to a wider range of geometric pattern transformations. Hence, our work may also provide some insight into transformation-invariance in dictionary learning.

2. PROBLEM FORMULATION

Let $p \in \mathbb{R}^n$ be a visual pattern. We define the transformation manifold of p as

$$\mathcal{M}_p = \{U_\lambda(p), \lambda \in \Lambda\} \subset \mathbb{R}^n, \quad (1)$$

where $\lambda = (\theta, t_x, t_y, s_x, s_y)$ is a transformation parameter vector, Λ is the parameter domain, and $U_\lambda(p)$ is a geometrically transformed version of p specified by λ. Here θ denotes rotation, t_x and t_y represent the translations in x and y directions, and s_x and s_y define an anisotropic scaling in x and y directions. The relation between $U_\lambda(p)$ and p can be expressed as $U_\lambda(p)(x, y) = p(x', y')$, where

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \quad (2)$$

Let $\mathcal{U} = \{u_i\}_{i=1}^N \subset \mathbb{R}^n$ be a set of observations of a geometrically transformed visual signal. We would like to describe the observations $\{u_i\}$ by the transformations of a representative pattern p as $u_i = U_{\lambda_i}(p) + e_i$, where the term e_i denotes the deviation of u_i from the transformation manifold $\mathcal{M}_p$ of p. Here, the representative pattern p is to be learned, as well as the parameter vectors

This work has been partly funded by the Swiss National Science Foundation under grant number 200021_120060.

$\{\lambda_i\}$ corresponding to the observations $\{u_i\}$. We construct p as a linear combination of some atoms selected from a dictionary manifold $\mathcal{D} = \{\phi_\gamma = U_\gamma(\phi), \gamma \in \Gamma\} \subset \mathbb{R}^n$, where γ is an atom parameter vector and Γ is the parameter domain defining $\mathcal{D}$. In our case, we derive atoms ϕ_γ from a Gaussian mother function $\phi(x, y) = \sqrt{2/\pi}\exp(-x^2 - y^2)$ by applying the geometric transformation defined by $\gamma = (\psi, \tau_x, \tau_y, \sigma_x, \sigma_y)$, where ψ is a rotation parameter, τ_x and τ_y denote translations, and σ_x and σ_y represent anisotropic scalings in x and y directions. We construct the representative pattern $p = \sum_{j=1}^{K} c_j \phi_{\gamma_j}$ as a combination of K atoms $\{\phi_{\gamma_j}\}$ with coefficients $\{c_j\}$. We would like to optimize the atom parameters $\{\gamma_j\}$ and the coefficients $\{c_j\}$ such that the total data approximation error

$$E = \sum_{i=1}^{N} \|u_i - U_{\lambda_i}\big(\sum_{j=1}^{K} c_j \phi_{\gamma_j}\big)\|^2 \tag{3}$$

is minimized, where the notation $\|.\|$ stands for the l_2-norm and $U_{\lambda_i}(p)$ corresponds to the transformation of the reference pattern p that has the smallest distance to u_i.

3. MANIFOLD BUILDING ALGORITHM

In order to minimize the approximation error E, we use an approximation of the manifold near each data point. We begin with an arbitrary initialization of the representative pattern as $p = c_0\phi_{\gamma_0}$, where ϕ_{γ_0} can be chosen as an atom that has high correlation with some typical images in the data. Then, we assign a tentative set of parameter vectors $\{\lambda_i\}$ to the points $\{u_i\}$ by projecting them onto $\mathcal{M}_p$ such that $\lambda_i = \arg\min_{\lambda\in\Lambda} \|U_\lambda(p) - u_i\|$. We construct the pattern gradually by adding an atom at each iteration. Let us denote the pattern consisting of $j-1$ atoms by p_{j-1}. In the j^{th} iteration we would like to optimize the parameters of the new atom γ_j and its coefficient c_j such that the new pattern $p_j = p_{j-1} + c_j\,\phi_{\gamma_j}$ minimizes E. Now we focus on the j^{th} iteration and denote $\gamma = \gamma_j$, $c = c_j$ for notational convenience. The error E has a complicated dependence on γ and c, mainly due to the fact that the values of parameter vectors $\{\lambda_i\}$ determined by the previous state of the pattern p_{j-1} are not necessarily the same as their values corresponding to the updated pattern p_j. Hence, we propose the following approach for the minimization of E in two steps. In the first step, we define a coarse approximation $\hat{E}$ of E as

$$\hat{E} = \sum_{i=1}^{N} \|\hat{e}_i\|^2 = \sum_{i=1}^{N} \|u_i - U_{\lambda_i}\big(p_{j-1} + c\,\phi_\gamma\big)\|^2, \tag{4}$$

where the vectors $\{\lambda_i\}$ are determined according to the previous pattern p_{j-1}. Assuming that the parameter vectors $\{\lambda_i\}$ do not change too much upon the addition of a new atom to p, $\hat{E}$ is expected to give an acceptable approximation of E. We have the following proposition, whose proof is given at the end of this section.

Proposition 1. *The error $\hat{E}$ is a DC (Difference-of-Convex) function of the atom parameters γ and the atom coefficient c, i.e., it can be represented in the form $\hat{E}(\gamma, c) = f(\gamma, c) - g(\gamma, c)$, where the functions f and g are convex in γ and c.*

We use the cutting plane algorithm discussed in [4] to approach the vicinity of the globally optimum solution (γ, c) that minimizes $\hat{E}$. Then in the second step of our method, we employ a more refined approximation ϵ of E. We use a locally linear approximation of the manifold $\mathcal{M}_{p_j}$ around each point $U_{\lambda_i}(p_j)$ and derive ϵ as the total distance between data points u_i and their projections onto the first order approximation of $\mathcal{M}_{p_j}$. It can be shown that ϵ is given by [3]

$$\epsilon = \sum_{i=1}^{N} \|\varepsilon_i\|^2 = \sum_{i=1}^{N} \|v_i - T_i(T_i^{\mathrm{T}}\,T_i)^{-1}\,T_i^{\mathrm{T}} v_i\|^2, \tag{5}$$

where $v_i = u_i - U_{\lambda_i}(p_j)$, and T_i is an $n \times 5$ matrix whose columns are the tangent vectors to the manifold $\mathcal{M}_{p_j}$ at the point $U_{\lambda_i}(p_j)$. Starting with the solution (γ, c) that minimizes $\hat{E}$, we finalize the atom selection by minimizing ϵ over (γ, c) using a gradient-descent algorithm.

Once the new atom ϕ_{γ_j} and its coefficient c_j corresponding to the j^{th} iteration are determined by these two optimization steps, the learned pattern is updated as $p_j = p_{j-1} + c_j\,\phi_{\gamma_j}$. Then the parameter vectors $\{\lambda_i\}$ are updated by recomputing the projections of the points $\{u_i\}$ onto the updated pattern transformation manifold $\mathcal{M}_{p_j}$. The projections can be updated by performing a search in a small region around their previous locations. We repeat this procedure to select and add atoms to the pattern p until the data approximation error E converges. We name this method Parameterized Atom Selection (PATS) and summarize it in Algorithm 1.

Algorithm 1 Parameterized Atom Selection (PATS)

1: **Input:**
   $\mathcal{U} = \{u_i\}_{i=1}^{N}$: Set of observations
2: **Initialization:**
3: $j = 0$.
4: Initialize the representative pattern $p_0 = c_0\phi_{\gamma_0}$ with a reference atom $\phi_{\gamma_0} \in \mathcal{D}$.
5: Determine a set of parameter vectors $\{\lambda_i\}$ by projecting $\{u_i\}$ on the transformation manifold $\mathcal{M}_{p_0}$.
6: **repeat**
7: $j = j + 1$.
8: Optimize the atom parameters γ_j and the coefficient c_j with DC programming such that the approximate error $\hat{E}$ in (4) is minimized.
9: Further optimize γ_j and c_j with gradient descent in order to minimize the refined approximate error ϵ in (5).
10: Update $p_j = p_{j-1} + c_j\phi_{\gamma_j}$.
11: Update parameter vectors $\{\lambda_i\}$ by projecting $\{u_i\}$ onto the transformation manifold $\mathcal{M}_{p_j}$.
12: **until** the approximation error E converges
13: **Output**:
   $p = p_j$: A representative pattern whose transformation manifold $\mathcal{M}_p$ fits the data $\mathcal{U}$

Proof of Proposition 1: In order to show the DC property of the objective function $\hat{E}$, we build on the results from [4], where a DC decomposition of the distance between a query pattern and the 4-dimensional transformation manifold of a reference pattern consisting of Gaussian atoms is derived. We make use of the following properties of DC functions, which are explained in more detail in [4] and [10].

Proposition 2. *(a) Let $\{f_i\}_{i=1}^{m}$ be DC functions and $\{\lambda_i\}_{i=1}^{m}$ be real numbers. Then $\sum_{i=1}^{m} \lambda_i f_i$ has a DC decomposition [10], [4].*

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a DC function and $q : \mathbb{R} \to \mathbb{R}$ be a convex function. Then $q(f(x))$ has a DC decomposition [10], [4].

(c) Let f_1 and f_2 be DC functions with nonnegative convex parts. Then the product $f_1\,f_2$ also has a DC decomposition with nonnegative convex parts [10], [4].

(d) Let $\psi \in [0, 2\pi)$ and $\sigma \in \mathbb{R}^+$. Then the following functions have DC decompositions with nonnegative convex parts: $\cos(\psi)$, $\sin(\psi)$, $\frac{\cos(\psi)}{\sigma}$, $\frac{\sin(\psi)}{\sigma}$ [4].

With a simplification of notation, let $\hat{e} = u - U_\lambda\big(p + c\,\phi_\gamma\big)$ denote the difference vector between the data point u and the pattern $p + c\,\phi_\gamma$ transformed by λ. Then one can write $\hat{e}$ in the form $\hat{e} = v - c\,U_\lambda(\phi_\gamma) = v - c\,\phi_{\lambda\circ\gamma}$, where $v = u - U_\lambda(p)$ is a constant

with respect to γ and c, and $\lambda \circ \gamma$ denotes the composition of the transformations λ and γ such that $U_\lambda(\phi_\gamma) = \phi_{\lambda\circ\gamma}$. We express the relation between the transformed Gaussian atom and the mother function with the change of variables

$$\phi_{\lambda\circ\gamma}(x,y) = \phi(\tilde{x},\tilde{y}) = \sqrt{\frac{2}{\pi}}\exp(-\tilde{x}^2-\tilde{y}^2).$$

Then the coordinates $(\tilde{x},\tilde{y})$ of the point in the mother function that corresponds to a point with coordinates (x,y) in the transformed atom can be derived in the form

$$\tilde{x} = \alpha_1\cos(\psi)/\sigma_x + \alpha_2\sin(\psi)/\sigma_x - \cos(\psi)\tau_x/\sigma_x + \sin(\psi)\tau_y/\sigma_x$$
$$\tilde{y} = \beta_1\cos(\psi)/\sigma_y + \beta_2\sin(\psi)/\sigma_y + \cos(\psi)\tau_y/\sigma_y + \sin(\psi)\tau_x/\sigma_y.$$

Here α_1, α_2, β_1, β_2 are functions of the transformation parameters λ and coordinates (x,y) but they stay constant with respect to the atom parameters γ. One can obtain a DC decomposition of the term $\cos(\psi)\tau_x/\sigma_x$ by using the DC decomposition of $\cos(\psi)/\sigma_x$ (Proposition 2.d) and the DC decomposition of τ_x as $\tau_x = 0.5\,(\tau_x+1)^2 - 0.5\,(\tau_x^2+1)$, together with the product property in Proposition 2.c. The decompositions of the terms $\sin(\psi)\tau_y/\sigma_x$, $\cos(\psi)\tau_y/\sigma_y$ and $\sin(\psi)\tau_x/\sigma_y$ are obtained similarly. Then, as $\tilde{x}$ and $\tilde{y}$ are weighted sums of DC-decomposable terms, one can obtain their decompositions using Proposition 2.a. Following this, we define $\tilde{z} = \tilde{x}^2 + \tilde{y}^2$, whose DC decomposition is given by Propositions 2.c and 2.a. Then expressing the mother function $\phi(\tilde{x},\tilde{y}) = \sqrt{2/\pi}\exp(-\tilde{z})$ as a convex function of $\tilde{z}$, Proposition 2.b provides the decomposition of $\phi(\tilde{x},\tilde{y})$.

Now in order to decompose the approximation error $\hat{E}$, we expand $\|\hat{e}\|^2$ in terms of the errors at individual pixels as

$$\begin{aligned}\|\hat{e}\|^2 &= \|v - c\,\phi_{\lambda\circ\gamma}\|^2 \\ &= \sum_{m=1}^{n}\big(v(m)^2 - 2\,v(m)\,c\,\phi_{\lambda\circ\gamma}(m) + c^2\,\phi^2_{\lambda\circ\gamma}(m)\big),\end{aligned} \tag{6}$$

where $v(m)$ and $\phi_{\lambda\circ\gamma}(m)$ are the m^{th} entries of respectively v and $\phi_{\lambda\circ\gamma}$. The first term $v(m)$ is constant with respect to the optimization variables γ and c. Using the decompositions of $\phi_{\lambda\circ\gamma}$ and $c = 0.5\,(c+1)^2 - 0.5\,(c^2+1)$, Propositions 2.c and 2.a provide the DC decomposition of the second term $-2\,v(m)\,c\,\phi_{\lambda\circ\gamma}(m)$. Then, we define $\tilde{z}(m) = \tilde{x}(m)^2 + \tilde{y}(m)^2$, where $\tilde{x}(m)$ and $\tilde{y}(m)$ are the coordinates corresponding to the m^{th} pixel. Observing that $\phi^2_{\lambda\circ\gamma}(m) = 2/\pi\exp(-2\,\tilde{z}(m))$ is directly decomposable by Proposition 2.b, one can obtain the decomposition of the last term $c^2\,\phi^2_{\lambda\circ\gamma}(m)$ by applying Proposition 2.c. Finally, the DC decompositions of $\|\hat{e}\|^2$ and $\hat{E} = \sum_{i=1}^{N}\|\hat{e}_i\|^2$ simply follow from Proposition 2.a. Note that although we have shown the DC property of $\hat{E}$ for the Gaussian mother function here, one can obtain similar DC decompositions for some other types of generating functions as well, such as the anisotropic refinement function (AnR) as shown in [4].

4. EXPERIMENTAL RESULTS

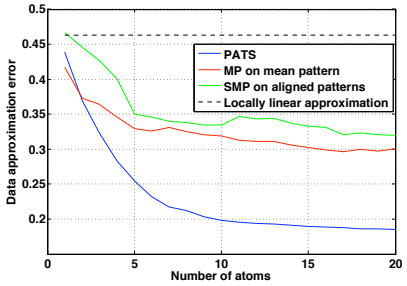
We test the proposed algorithm on several data sets with geometric transformations. In the first experiment we generate a data set by picking 30 random images of the digit '5' from the MNIST handwritten digits database[1] and applying to each image a random geometric transformation according to our transformation model of Eq. (1). Some example data set images are shown in Fig. 1(a). We construct a representative pattern with 20 atoms with the proposed PATS algorithm. The resulting pattern is shown in Fig. 1(b). It is seen that the constitutive characteristics of the input data are successfully captured in the learned pattern. In order to evaluate the performance of the proposed algorithm, we compare it with some baseline approaches. In the first approach (MP on mean pattern), we determine a representative pattern by aligning all data set images and then picking the aligned image that is closest to the centroid of all aligned images. We call this representative pattern as the mean pattern and obtain progressive approximations of the mean pattern with up to 20 atoms via the Matching Pursuit [5] algorithm using a redundant dictionary consisting of samples from our dictionary manifold $\mathcal{D}$. Then in a second approach (SMP on aligned patterns), we again align all data set images but then obtain a progressive simultaneous approximation of the aligned images with up to 20 atoms via the Simultaneous Matching Pursuit [6] algorithm using the same dictionary. We construct a representative pattern gradually by adding the atoms chosen by the SMP algorithm, where each atom is weighted by its average coefficient in all images. In order to compare the PATS method with these approaches we approximate the input images with the transformation manifolds of the representative patterns obtained with all three approaches. In Fig. 1(c) we plot the data approximation error, i.e., the average distance between the data set images and the transformation manifold, with respect to the number of atoms used in the progressive generation of the representative patterns with PATS, MP on the mean pattern and SMP on the aligned patterns. We also compute the locally linear approximation error of the data in order to demonstrate the advantage that our manifold building approach provides over classical manifold learning algorithms. The locally linear approximation error is indicated in Fig. 1(c), which is the average distance between a point and its projection onto the plane passing through its nearest neighbors. Many manifold learning algorithms as [2] and [11] make use of such a linear approximation of the manifold in order to compute a global parameterization of the data. The results in Fig. 1(c) show that the transformation manifold of the representative pattern constructed with the proposed method provides a more accurate data representation compared to the MP and SMP approaches. Moreover, the locally linear approximation error is significantly higher than the approximation errors obtained with transformation manifolds, which is due to the fact that the data is quite sparsely sampled on the manifold in this experiment, i.e., it is not possible to estimate the tangent vectors accurately from the data. Since our method is specially adapted for the pattern transformation model, it is not affected by such a sparse sampling.

Then, we repeat the same experiment on 30 randomly chosen handwritten '2' digits from the same database. Some images used in the experiment are shown in Fig. 2(a), the pattern constructed with 20 atoms using the PATS algorithm is shown in Fig. 2(b), and the data approximation errors are plotted in Fig. 2(c). The results are consistent with the results of the previous experiment and the proposed algorithm yields the most accurate representation of the data.

Finally, we test the algorithm on face images. We use a data set consisting of 35 geometrically transformed face images of a single subject, which also includes some variation in facial expressions [12]. Some sample data set images are shown in Fig. 3(a). Note that as our pattern transformation manifold model does not account for the facial expression changes, the diversity caused by these is regarded as the deviation of the data from the representative manifold. We repeat the same experiment as in the previous setups. The pattern built using the PATS method with 30 atoms is shown in Fig. 3(b) and the data approximation errors yielded by the different manifold construction approaches are plotted in Fig. 3(c). In this experiment, all face images belong to the same subject and the transformation manifold of the mean pattern constitutes a better approximation of all patterns when compared to the handwritten digits experiments. Therefore, while our method still provides the best approximation accuracy, the performances of the MP and SMP methods are closer to that of ours as compared to the previous experiments.

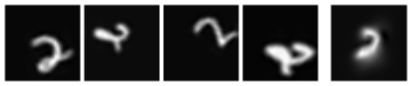
[1] http://yann.lecun.com/exdb/mnist/

(a) Data set images (b) Built pattern

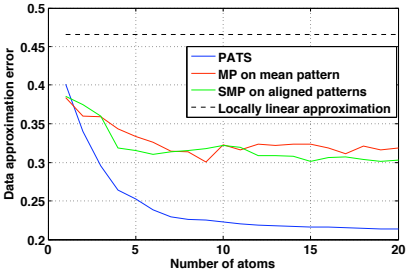


(c) Convergence of error

Fig. 1. Results obtained on handwritten '5' digits

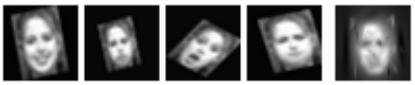


(a) Data set images (b) Built pattern

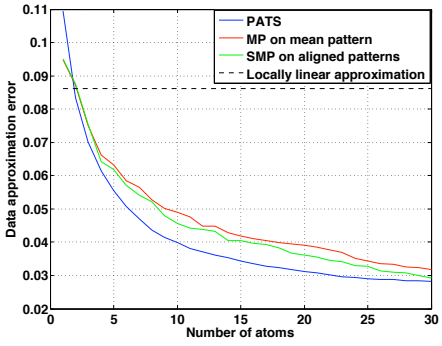


(c) Convergence of error

Fig. 2. Results obtained on handwritten '2' digits



(a) Data set images (b) Built pattern



(c) Convergence of error

Fig. 3. Results obtained on face images

5. CONCLUSIONS

We have presented a method for building a pattern transformation manifold for the approximation of a set of geometrically transformed signals. We construct a representative pattern gradually by selecting atoms from an analytical dictionary manifold. The experimental results suggest that the proposed algorithm can be used effectively for dimensionality reduction, modeling and registration of geometrically transformed data conforming to an implicit manifold structure. Moreover, the optimization of atom parameters on a continuous dictionary manifold yields more accurate results compared to reference pursuit methods that use a fixed discrete dictionary. Unlike traditional manifold learning algorithms, our transformation manifold building approach provides an analytical representation of data, which brings favorable properties such as smoothness, ease of encoding and allowing the generation of new samples.

6. REFERENCES

[1] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction.," *Science*, vol. 290, no. 5500, pp. 2319–2323, December 2000.

[2] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[3] E. Vural and P. Frossard, "Approximation of pattern transformation manifolds with parametric dictionaries," Accepted to *IEEE Conference on Acoustics, Speech and Signal Processing*, 2011.

[4] E. Kokiopoulou and P. Frossard, "Minimum distance between pattern transformation manifolds: Algorithm and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 7, pp. 1225–1238, 2009.

[5] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[6] Joel A. Tropp, Anna C. Gilbert, and Martin J. Strauss, "Algorithms for simultaneous sparse approximation part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[7] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[8] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," in *Proceedings of European Signal Processing Conference*, 2008.

[9] P. Jost, S. Lesage, P. Vandergheynst, and R. Gribonval, "MoTIF: An efficient algorithm for learning translation-invariant dictionaries," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006.

[10] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[11] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 5591–5596, May 2003.

[12] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.