Image alignment with rotation manifolds built on sparse geometric expansions

Effrosyni Kokiopoulou and Pascal Frossard
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute - ITS
CH- 1015 Lausanne, Switzerland
{effrosyni.kokiopoulou,pascal.frossard}@epfl.ch

Abstract— In this paper we discuss the problem of alignment of patterns under arbitrary rotation. When a generic image pattern is geometrically transformed, it typically spans a (possibly nonlinear) manifold in a high dimensional space. When the pattern of interest is given by a sparse approximation over a structured dictionary of geometric atoms, we show that the rotation manifold can be expressed analytically as a function of the transformation parameters. At the same time, its high order derivatives are also given in a closed form when the pattern is represented as a sparse linear combination of a few differentiable basis functions. In this framework, the alignment problem is formulated as the minimization of the distance between the reference pattern and the manifold, which boils down to a nonlinear least squares optimization problem. We propose to solve this problem by a Newton-type method, whose solution is facilitated by the analytical expressions of the manifold derivatives. We further derive a global optimization heuristic algorithm based on Newton, and provide sufficient conditions for computing the global minimizer. Experimental results demonstrate the effectiveness of the proposed methodology for image alignment and rotation invariant pattern recognition.

I. INTRODUCTION AND MOTIVATION

Quite of lot of research is devoted nowadays to bridging the semantic gap that refers to the shortcomings of low level features to address high level mining problems. The problem of image alignment is one of the most fundamental steps in that direction. For applications in multimedia content analysis and computer vision, the comparison of two pattern is generally meaningful if they are aligned first.

In this paper, we focus on the problem of alignment under rotation. When a pattern undergoes rotation it spans a (possibly nonlinear) manifold in a high dimensional space, which is usually called the *rotation manifold*. When the distance between two patterns is truly invariant to rotation, it becomes equivalent to the distance between their corresponding rotation manifolds, which is called the two sided manifold distance or the one sided manifold distance (MD) [1] when one of the pattern is fixed. However, alignment based on the computation of the manifold distance is a hard optimization problem, mostly due to the non convex nature of its objective function. One solution consists in the local linearization of the manifold, in order to compute the distance between tangent spaces instead of the true manifold distance [2]. However, the effectiveness of this approach is highly sensitive to the nonlinearity of the manifold. Another approach is to compute the manifold distance

using Newton's algorithm [3, Sec3.3], where derivatives are generally approximated by finite differences methods. This approach however only provides local rotation invariance due to the local convergence properties of Newton, and therefore leads to sub-optimal solutions when the rotation angle between patterns becomes large.

In this paper, we introduce a method to determine the analytical expression of the rotation manifold, and its derivatives in order to solve the alignment problem. We propose to represent the pattern of interest as a linear combination of geometric primitives, called atoms, which are chosen from a structured and possibly redundant dictionary. A dictionary is called *structured*, when it is built by atoms that are constructed by applying geometric transformations on mother functions [4] (e.g., Gaussian functions). When the pattern is transformed, the transformation is essentially applied on each constituent atom individually, resulting in a synthesis of transformations. Using group theory of transformations we show that the proposed methodology allows for a closed form computation of the manifold equation as well as its derivatives, which is an important advantage compared to the traditional way of approximating the latter with finite differences. We build on this framework to extend the Newton-based algorithm by proposing an optimization heuristic. In particular, we use global information to carefully select the initial guess in Newton, and we provide sufficient conditions which guarantee the computation of the global minimizer. The contribution of the paper is twofold; (i) it proposes the use of sparse expansions over structured dictionaries for image alignment under rotation, which has not been investigated before and (ii) it proposes a practical optimization heuristic which is able to find in many cases the global minimizer. We finally provide experimental results which show the effectiveness of the proposed algorithm in image alignment and pattern recognition applications, where it outperforms existing solutions typically based on manifold, tangent or euclidian distance computation.

II. PATTERN REPRESENTATION

We propose to represent the pattern of interest as a linear combination of geometric functions, taken from a structured and possibly redundant dictionary $\mathcal{D} = \{\phi_\gamma, \gamma \in \Gamma\}$ spanning the input space. The atoms in a *structured* 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 a unitary operator $U(\gamma)$, so that a structured dictionary takes the following form,

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

In the case of images, the simplest transformation γ_i, applied to the i-th atom, may be one of the following three types.

- *Translation* by $\vec{b}_i = [b_{ix}\ b_{iy}]^\top$. $U(\vec{b}_i)$ moves the generating function across the image i.e., $U(\vec{b}_i)\phi(x,y) = \phi(x - b_{ix}, y - b_{iy})$.
- *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 $\vec{a}_i = [a_{ix}\ a_{iy}]^\top$. $U(\vec{a}_i)$ scales the generating function anisotropically in the two directions i.e., $U(\vec{a}_i)\phi(x,y) = \phi(\frac{x}{a_{ix}}, \frac{y}{a_{iy}})$.

Composing all the above transformations yields a transformation with parameters $\gamma_i = \{\vec{b}_i, \vec{a}_i, \omega_i\} \in \Gamma$, which denotes a synthesis of translations, scalings and rotations. This set of transformations forms a group, namely the similitude group SIM(2) on the 2D plane. Let denote by

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

the rotation matrix by angle ω in the 2D plane. If $(\vec{b}, a, \omega)$ and $(\vec{b}', a', \omega')$ are two elements of the group, then the group law is

$$(\vec{b}, a, \omega) \circ (\vec{b}', a', \omega') = (\vec{b} + aR(\omega)\vec{b}', aa', \omega' - \omega). \tag{2}$$

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 $\phi(\cdot)$. When the i-th atom in the structured dictionary (1) is built as $\phi_{\gamma_i} = U(\gamma_i)\phi(x,y)$, it forms the same 2D function as $\phi(\tilde{x}, \tilde{y})$, where

$$
\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{a_{ix}} & 0 \\ 0 & \frac{1}{a_{iy}} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \cos\omega_i & \sin\omega_i \\ -\sin\omega_i & \cos\omega_i \end{bmatrix}}_{R(\omega_i)} \underbrace{\begin{bmatrix} x - b_{ix} \\ y - b_{iy} \end{bmatrix}}_{t}
$$
$$= AR(\omega_i)t. \tag{3}$$

Finally, the approximation of the pattern of interest 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 sub-optimal solutions are usually sufficient to capture the salient semantic and geometric structure of a pattern with only a few atoms. In this paper, we propose to use the Orthogonal Matching Pursuit (OMP) algorithm [5, Sec9.5.3], which is an efficient solution for computing sparse approximations in practice. Initially, OMP sets the residual $r_0 = s$ and then proceeds iteratively by selecting in the j-th 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.,

$$s = \sum_{i=1}^{k} \xi_i \phi_{\gamma_i} + r_k, \tag{4}$$

where r_k is the residual of the approximation.

III. ALIGNMENT UNDER ARBITRARY ROTATION

When a pattern s is rotated by angle θ, equation (4) implies that the rotated pattern has the following form

$$s(\theta) \triangleq U(\theta)s \approx \sum_{i=1}^{k} \xi_i U(\theta \circ \gamma_i)\phi = \sum_{i=1}^{k} \xi_i \phi_{\eta_i}, \tag{5}$$

where $\eta_i \triangleq \theta \circ \gamma_i$. In the above formula we assume that k is sufficiently large such that all the principal geometric components of s have been captured; hence, the residual r_k is small and may be ignored. If $\gamma_i = (\vec{b}_i, \vec{a}_i, \omega_i)$ are the parameters of the atom ϕ_{γ_i}, then using the group law (2), we observe that the parameters of the atom after rotation become

$$\eta_i = \theta \circ \gamma_i = (R(\theta)\vec{b}_i, a_i, \omega_i - \theta). \tag{6}$$

The rotation of the pattern s with different angles θ forms the rotation manifold and Eq. (5) is the manifold equation.

Suppose now that we are given a *reference pattern* p and we want to align with s, or equivalently that we want to find the smallest distance from p to the rotation manifold built on $s(\theta)$. The optimization problem that we need to solve can be formulated as follows,

$$\theta^* = \arg\min_{\theta \in [0, 2\pi)} f(\theta), \quad \text{where} \tag{7}$$

$$f(\theta) = \frac{1}{2}\|s(\theta) - p\|_2^2 = \frac{1}{2}\sum_{m=1}^{M}[s_m(\theta) - p_m]^2, \tag{8}$$

where M is the total number of pixels in the images. The above problem is a nonlinear least squares problem and in general the objective function $f(\theta)$ is non convex. Therefore, computing the global minimizer θ^* is hard due to the presence of an unknown number of local minima.

If we use *Newton*'s method [3, Sec3.3] to attack the above problem, we need to pick an initial guess θ_0 and then use the iteration

$$\theta_{n+1} = \theta_n - \frac{f'(\theta)|_{\theta=\theta_n}}{f''(\theta)|_{\theta=\theta_n}}, \quad n = 0, 1, 2, \ldots \tag{9}$$

The first and second derivatives of f are

$$f'(\theta) = \sum_{m=1}^{M}[s_m(\theta) - p_m]\frac{\partial s_m(\theta)}{\partial \theta} = (s(\theta) - p)^\top \dot{s},$$

$$f''(\theta) = \dot{s}^\top \dot{s} + (s(\theta) - p)^\top \ddot{s},$$

where $\dot{s}$ and $\ddot{s}$ are the first and second order derivatives of s. Observe that $f''(\theta)$ has a special structure, since the first part does not involve second order derivatives. Gauss-Newton relies upon iteration (9) where $f''(\theta)$ is approximated by its first part only.

Equation (5) implies that $\dot{s}(\theta)$ is computed as a linear combination of the individual tangent vectors of the transformed atoms i.e.,

$$\dot{s}(\theta) = \sum_{i=1}^{k} \xi_i \dot{\phi}_{\eta_i}(x,y) = \sum_{i=1}^{k} \xi_i \dot{\phi}(\tilde{x}, \tilde{y}), \qquad (10)$$

where we have used $\phi_{\eta_i}(x,y) = \phi(\tilde{x}, \tilde{y})$, with the new coordinate system for the atom ϕ_{η_i} written by the group law (6), as

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = AR(\omega_i)R(-\theta)\left(\begin{bmatrix} x \\ y \end{bmatrix} - R(\theta) \begin{bmatrix} b_{ix} \\ b_{iy} \end{bmatrix} \right)$$

$$= AR(\omega_i) \begin{bmatrix} \cos\theta x - \sin\theta y - b_{ix} \\ \sin\theta x + \cos\theta y - b_{iy} \end{bmatrix}. \qquad (11)$$

The tangent vector of $\phi(\tilde{x}, \tilde{y})$ with respect to θ is

$$\dot{\phi}(\tilde{x}, \tilde{y}) = \frac{\partial\tilde{x}}{\partial\theta}\frac{\partial\phi}{\partial\tilde{x}} + \frac{\partial\tilde{y}}{\partial\theta}\frac{\partial\phi}{\partial\tilde{y}}, \qquad (12)$$

where $[\frac{\partial\phi}{\partial\tilde{x}} \ \frac{\partial\phi}{\partial\tilde{y}}]^\top$ is the spatial gradient of $\phi(\tilde{x}, \tilde{y})$. By differentiating Eq. (11) with respect to θ, we further have that

$$\begin{bmatrix} \partial\tilde{x}/\partial\theta \\ \partial\tilde{y}/\partial\theta \end{bmatrix} = AR(\omega_i) \begin{bmatrix} -\sin\theta x - \cos\theta y \\ \cos\theta x - \sin\theta y \end{bmatrix}. \qquad (13)$$

The second order derivative can finally be computed by differentiating Eq. (10) as

$$\ddot{s}(\theta) = \sum_{i=1}^{k} \xi_i \ddot{\phi}(\tilde{x}, \tilde{y}). \qquad (14)$$

It is built on the second order derivatives of the atoms $\phi(\tilde{x}, \tilde{y})$, and we have

$$\ddot{\phi}(\tilde{x}, \tilde{y}) = \frac{\partial^2\tilde{x}}{\partial\theta^2}\frac{\partial\phi}{\partial\tilde{x}} + \frac{\partial\tilde{x}}{\partial\theta}\frac{\partial^2\phi}{\partial\tilde{x}\partial\theta} + \frac{\partial^2\tilde{y}}{\partial\theta^2}\frac{\partial\phi}{\partial\tilde{y}} + \frac{\partial\tilde{y}}{\partial\theta}\frac{\partial^2\phi}{\partial\tilde{y}\partial\theta}. \qquad (15)$$

Finally, by differentiation of equation (13) with respect to θ, we have that

$$\begin{bmatrix} \partial^2\tilde{x}/\partial\theta^2 \\ \partial^2\tilde{y}/\partial\theta^2 \end{bmatrix} = AR(\omega_i) \begin{bmatrix} -\cos\theta x + \sin\theta y \\ -\sin\theta x - \cos\theta y \end{bmatrix}.$$

From the equations above, it is clear that one may compute $\dot{s}$ and $\ddot{s}$ in closed form when ϕ is also given in closed form, as it is usually the case in structured dictionaries. It leads to important advantages in the construction of Newton-based algorithms, where common methods are usually forced to rely on approximations based on finite differences.

IV. Optimization heuristic for rotation invariance

As Gauss-Newton methods are only locally convergent, we propose now an optimization heuristic called OHMD, which allows to select carefully the initial estimate used in Gauss-Newton algorithm in order to extend it into a global minimizer. Before, we delve into the details of the algorithm, we review briefly the second order sufficient conditions for optimality.

Theorem 1: (Second-order Sufficient Conditions, [3, sec2.1]) Suppose that $\nabla^2 f$ is continuous in an open neighborhood of θ^* and that $\nabla f(\theta^*) = 0$ and $\nabla^2 f(\theta^*)$ is positive definite. Then θ^* is a strict local minimizer of f.

Now we are ready to describe our OHMD. The main idea is to sample the rotation interval of interest $[\theta_1, \theta_2]$ and use Theorem 1 to detect all sample points being close to local minimizers i.e., having small f' and positive f''. From these points pick the one θ_0 resulting in the smallest value of f, and use Gauss-Newton with θ_0 as initial guess. Assuming that f has a unique global minimizer θ_0^*, we expect θ_0 be close to it.

Algorithm: **OHMD**
Input: Patterns p and $s(\theta)$.
Output: Estimate $\hat{\theta}$ of the correct alignment angle.
1. Set $h = \frac{\theta_2 - \theta_1}{L}$ and define a grid on $[\theta_1, \theta_2]$
   $\mathcal{G}_\mathcal{L} = \{\theta_l, \ \theta_l = \theta_1 + lh, \ l = 0, \ldots, L\}$.
2. Use Theorem 1 to select the grid points being
   close to local minimizers.
4. Evaluate f on the selected grid points.
5. Pick the one, call it θ_0, with the smallest value of f.
6. Call Gauss-Newton with θ_0 as initial guess
   and set $\hat{\theta}$ to its output.

In what follows, we provide a sufficient condition which guarantees convergence to the global minimizer.

Condition 1: Suppose that f is locally strictly convex in a region around θ_0^* and call $\mathcal{C}$ the largest such region. Call also θ_1^* the second best local minimizer. Denote by θ^ℓ and θ^r the left and right points around θ_0^* where the line $y = f(\theta_1^*)$ intersects f. If the following conditions are met

* $h < |\theta^\ell - \theta^r|$ and
* the points θ^ℓ, θ^r fall within $\mathcal{C}$,

then OHMD will find the global minimizer.

Proof: If the first condition is satisfied, there will be at least one grid point lying in the interval (θ^ℓ, θ^r). Call θ_l the grid point closest to θ_0^* and decompose it as $\theta_l = \lambda\theta^\ell + (1-\lambda)\theta^r$, where $\lambda \in (0,1)$. From the second condition, f is strictly convex in the interval (θ^ℓ, θ^r) implying that

$$f(\theta_l) < \lambda f(\theta^\ell) + (1-\lambda)f(\theta^r)$$
$$= \lambda f(\theta_1^*) + (1-\lambda)f(\theta_1^*) = f(\theta_1^*). \qquad (16)$$

The above inequality guarantees that OHMD in step 5 will select θ_l as the initial guess to Gauss Newton, and not another grid point. Finally, Gauss-Newton will converge to θ_0^* due to the strict convexity of f in $\mathcal{C}$. ∎

The computational cost of OHMD is roughly $O(Lk)$, where L is the grid size and k is the number of atoms in (4).

V. Experimental Results

In the experiments we build the structured dictionary $\mathcal{D}$ using the *Gaussian* as a mother function i.e., $\phi(x,y) = \frac{1}{\rho}\exp(-(x^2+y^2))$ where ρ is a normalization factor. Hence, the first order and mixed derivatives used in equations (12) and (15) are

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

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

and similarly for the derivatives with respect to y. For the construction of the dictionary $\mathcal{D}$ we used 10 orientation angles in $[0, \pi)$, 5 scales in $[1, \sqrt{M}/2]$ (M is the image size in pixels) and the pixel locations for translation parameters. We also use two datasets; (i) a *handwritten digit collection*[1], which contains binary images of "0" through "9" and each class contains 39 samples, (ii) the ORL *face database* [6] which contains 40 individuals and 10 different images for each individual. Finally in the OMP approximation we used $k = 30$ atoms for the digits and $k = 50$ atoms for the faces.
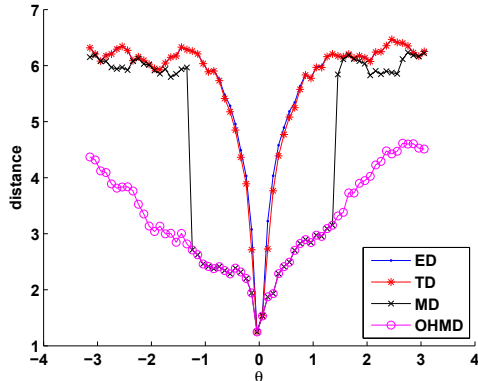

Fig. 1.   Comparison of different methods.

In the first experiment illustrated in Figure 1, we first compute the sparse expansion of a handwritten digit '5' over $\mathcal{D}$ using OMP, to form $s(0)$. Then we rotate the digit in all possible angles θ in $[-\pi, \pi]$ with step 0.1, using image warping to form different reference images p. For each of these images, we report the Euclidian distance (ED) $\frac{1}{2}\|s(0) - p\|_2$ with respect to the approximation $s(0)$. In addition, we also compute the distance between reference images and the rotation manifold at $\theta = 0$, using the common tangent distance method (TD) [2]. The tangent vector is computed with Eq. (10), and the tangent distance is reported in the Figure 1. Finally, for all the reference images, we align the pattern s using manifold distance computation based on both the Newton algorithm (MD), and on the OHMD algorithm, as described previously. For each reference image, we report the value of cost function f after convergence of the alignment algorithms.

We can observe that ED and TD are very sensitive to rotations, and that ED is clearly the worst method. MD allows for a dramatic increase in performance. However, due to its local convergence properties, for large θ it results in large distances since the alignment algorithm does not converge to the right rotation angle in these cases. On the contrary, the proposed method (OHMD) provides a sustained performance. Finally, note that the non-zero distance at $\theta = 0$ is due to the OMP approximation error.

In the second part of the experiments, we evaluate the performance of the nearest neighbor (NN) classifier combined

[1] http://www.cs.toronto.edu/~roweis/data/binaryalphadigs.mat

Class. error rate(%)	ED	TD	MD	OHMD
Digits, $[-\pi/2, \pi/2]$	60.69	57.93	27.59	25.51
Faces, $[-\pi/2, \pi/2]$	83	82.5	34.5	29
Faces, $[-\pi, \pi]$	87	87	58	27.50

TABLE I

ROTATION INVARIANT PATTERN RECOGNITION RESULTS.

with each one of the above methods, on handwritten digit and face recognition. For the digits, we form the training set by the first 10 samples from each class and the last 29 samples are assigned in the test set and for the faces we split it into 5 samples in the training and 5 in the test set. The classification error rate of the NN classifier applied on the non-rotated digits is 25.52% and on the non-rotated faces is 15%. We rotate (independently) each test image by a random angle θ uniformly drawn from the interval of interest, to form the reference image p. We assume that the sparse expansion of each training image has been already computed off-line, to form $s(0)$. In each pairwise comparison of the NN classifier, the training sample is aligned to the test sample in the MD and OHMD algorithms.

Table I summarizes the classification error rates of all methods along with the interval of interest. In all cases, we used $h = 0.1$ in OHMD. Note that the ED classifier completely breaks down while the TD classifier behaves slightly better. On the other hand, MD and OHMD seem to be quite robust to rotations. The former undergoes a drop of its performance when the rotation interval becomes large, but the latter is quite robust in all cases. This suggests that the optimization heuristic seems to work well in practice.

VI. CONCLUSIONS

In this paper we proposed the use of sparse expansions over structured dictionaries as a tool for alignment under arbitrary rotation. We showed that such a framework results in closed form expressions for the high order derivatives as well as the manifold equation, which facilitates the application of Newton-based methods for manifold distance computation. We also proposed an optimization heuristic for global optimization and demonstrated that it works nicely in alignment and rotation invariant pattern recognition problems.

REFERENCES

[1] N. Vasconcelos and A. Lippman. A multiresolution manifold distance for invariant image similarity. *IEEE Transactions on Multimedia*, 7(1):127–142, February 2005.
[2] P. Simard, Y. Le Cun, J. Denker, and B. Victorri. Transformation invariance in pattern recognition - tangent distance and tangent propagation. *Neural networks: tricks of the trade*, 1524:239–274, 1998. LNCS, Springer.
[3] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.
[4] 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.
[5] S. Mallat. *A Wavelet Tour of Signal Processing, 2nd edn*. Academic Press, 1998.
[6] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, December 1994.