

A GEOMETRIC FRAMEWORK FOR REGISTRATION OF SPARSE IMAGES

Alhussein Fawzi, Pascal Frossard

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

ABSTRACT

We examine the problem of image registration when images have a sparse representation in a dictionary of geometric features. We propose a novel algorithm for aligning images by pairing their sparse components. We show numerically that this algorithm works well in practice and analyze key properties on the dictionary that drive the registration performance. We compare these properties to existing characterizations of redundant dictionaries (i.e., coherence, restricted isometry property) and show that the newly introduced properties finely capture the behaviour of our registration algorithm.

Index Terms— Image alignment, sparse approximation, parametric dictionary, dictionary properties.

1. INTRODUCTION

Image *registration* is the process of aligning two or more images, taken from various viewpoints, at different times, or by different sensors. It represents a crucial preprocessing step in many image processing and computer vision applications, such as object detection, localization and classification to name a few.

A popular class of methods are *feature based* approaches [1]. A comprehensive and widely used feature based method is the Scale Invariant Feature Transform (SIFT) [2]. The SIFT features provide a translation, rotation and scale invariant representation of the image. Even though these features have been a great success in many computer vision applications, they are mostly built on empirical results. Furthermore, SIFT keypoints are not well suited for estimating large rotations as shown in [3]. An alternative to building invariant feature-based representations of the images is to use an *invariant distance measure*. In this approach, the transformation invariance is included in the measure of distance. However, computing the transformation invariant distance is in general a difficult problem. The authors in [4] tackle the problem by approximating the transformation invariant distance with the distance between the tangent (linear) spaces of the two manifolds of transformed images. The main disadvantage of this method is that it achieves only *local* invariance. Along the same lines, the work in [5] achieves invariance by approximating the transformation invariant distance; this method is however computationally expensive since it requires the minimization of a difference of two convex functions. For a complete survey on image registration, the reader is referred to [6, 7].

In this paper, we introduce a novel algorithm for aligning input images that have a sparse representation in a well chosen geometric dictionary of features. We examine the performance of the proposed algorithm on illustrative examples and show that the algorithm compares favorably with related methods in the literature. Moreover, we study the performance of this algorithm theoretically. In order to do so, we introduce two novel properties that characterize the registration error of our algorithm: the *robust linear independence* of the dictionary, which relates the distance between sparse signals to the distance between their features, and the *transformation inconsistency*, which measures how different features in the dictionary are

affected by the same transformation. We note that widely used properties of the dictionary (i.e., coherence, restricted isometry property [8]) are not the relevant quantities in our setting, since they are very close to one independently of how good the dictionary is for the purpose of image alignment. Through its detailed analysis, the proposed framework provides useful insight on the connections between image registration problems and sparse signal processing.

2. IMAGE REGISTRATION

2.1. Problem formulation

Let (\mathcal{T}, \circ) be a transformation group and its associated composition rule, and define L^2 to be the Hilbert space of square integrable functions from \mathbb{R}^2 to \mathbb{R} . For any $\eta \in \mathcal{T}$, we denote by $U(\eta)$ the unitary operator that maps an image $f \in L^2$ to its transformed version $U(\eta)f \in L^2$. To avoid heavy notations, we also use f_η to denote $U(\eta)f$.

Let \mathcal{D} be the parametric dictionary of features constructed as follows:

$$\mathcal{D} = \{\phi_\gamma : \gamma \in \mathcal{T}_d\} \subset L^2, \quad (1)$$

where $\phi \in L^2$ is a generating function and $\mathcal{T}_d \subset \mathcal{T}$ is a discretization of \mathcal{T} . We refer to any element ϕ_γ in \mathcal{D} as a *feature* or *atom*. We consider here for simplicity that the generating function ϕ is chosen so that $\gamma \mapsto \phi_\gamma$ defines a one-to-one mapping¹. Besides, the generating function ϕ is chosen to be non negative and has a finite effective support. Consequently, each atom in the dictionary corresponds to a potential part of the image. We assume in this paper that the two patterns p and q we wish to align are exactly K -sparse in the dictionary \mathcal{D} :

$$p = \sum_{i=1}^K c_i \phi_{\gamma_i}, \quad (2)$$

$$q = \sum_{i=1}^K d_i \phi_{\delta_i}. \quad (3)$$

Since the dictionary \mathcal{D} contains features that represent potential parts of the image, we can assume that coefficients c_i and d_i are all non negative as the different features do not cancel each other. Our aim is to find the transformation in \mathcal{T} that minimizes the alignment error between the two patterns p and q . Formally, we consider the problem (P) defined as follows:

$$(P): \text{Find } \eta_0 = \underset{\eta \in \mathcal{T}}{\operatorname{argmin}} \|U(\eta)p - q\|_2.$$

We denote by $d(p, q) = \|U(\eta_0)p - q\|_2$ the *transformation invariant distance* between p and q . It corresponds to the regular Euclidean distance when the images are optimally aligned (in the sense of the L^2 norm).

Computing the transformation η_0 and the transformation invariant distance $d(p, q)$ is a hard problem since the objective function

¹For a complete study of the case where we relax this assumption, we refer the reader to our technical report in [3].

is typically non convex and exhibits many local minimas. Hence, solving this optimization problem using gradient or Newton based algorithms will generally provide only a local minima. Therefore, we propose below a simple algorithm that aligns images based on relative transformations between atoms in the respective sparse representations.

2.2. Registration algorithm

Our approach for aligning p and q is conceptually simple. We exploit the fact that p and q are K -sparse in the geometric dictionary \mathcal{D} , and estimate the transformation based on the relative transformation between the features in both images. More precisely, let $\mathcal{T}_a^{p,q} = \{\delta_i \circ \gamma_j^{-1} : 1 \leq i, j \leq K\}$ define the set of relative transformations between pairs of features taken respectively in p and q . We estimate the transformation by solving the following relaxed problem of (P) :

$$(\hat{P}) : \hat{\eta} = \underset{\eta \in \mathcal{T}_a^{p,q}}{\operatorname{argmin}} \|U(\eta)p - q\|_2.$$

Note that the only difference between (P) and (\hat{P}) is that we replaced the initial search space \mathcal{T} with the more tractable search space $\mathcal{T}_a^{p,q}$. We define $d_a(p, q) = \|U(\hat{\eta})p - q\|_2$ to be the *approximate transformation-invariant distance* between p and q . Since the cardinality of $\mathcal{T}_a^{p,q}$ is at most K^2 , (\hat{P}) can be solved efficiently by a full search over the elements of $\mathcal{T}_a^{p,q}$ when K is relatively small.

2.3. Illustrative experimental results

We compare here our method with several baseline algorithms in the task of computing transformation-invariant distances. We consider \mathcal{T} to be the similarity group², and build the dictionary \mathcal{D} on the model of Eq. (1) with an anisotropic Gaussian generating function defined by:

$$\phi(x, y) = \frac{1}{\xi} \exp\left(-\left(\frac{x}{4}\right)^2 - y^2\right),$$

where ξ is chosen to have $\|\phi\|_2 = 1$ ³. Besides, we have generated \mathcal{T}_d by a discretization of the parameters of \mathcal{T} to guarantee a low sparse approximation error on typical images. Moreover, to reduce the errors due to the discretization of the dictionary, we perform a gradient descent starting from $\hat{\eta}$ in order to refine the estimated transformation in a final step of the algorithm.

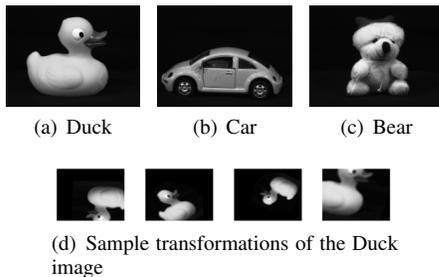


Fig. 1. Test images [9], and transformed versions of the Duck image. Note that the transformations can imply partial occlusions in the image. All images are resized to be of dimension 75×75 pixels.

²The similarity group of the plane $SIM(2)$ contains transformations that are combinations of translation, rotation and isotropic dilation.

³The chosen mother function does not satisfy the one-to-one mapping assumption of $\gamma \mapsto U(\gamma)\phi$. We circumvent this by slightly modifying the definition of $\mathcal{T}_a^{p,q}$. We define the stabilizer of ϕ to be the set that keeps the mother function unchanged: $\mathcal{S}_\phi = \{\gamma : U(\gamma)\phi = \phi\}$. Then, we define $\mathcal{T}_a^{p,q} = \{\delta_i \circ \pi \circ (\gamma_j)^{-1} : 1 \leq i, j \leq K, \pi \in \mathcal{S}_\phi\}$ [3].

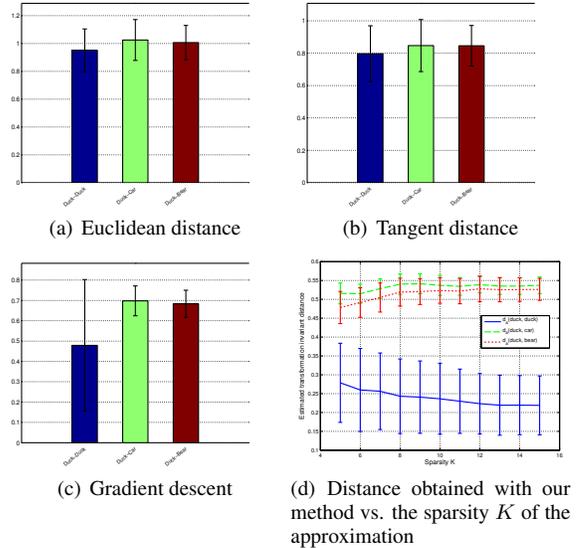


Fig. 2. Average and standard deviation of intra- and inter-class distances for different methods. The blue color denotes the distance between the original Duck image and a transformed version (intra-class), while the green and red colors refer respectively to the distance between the original Duck image and a transformed version of the Car and Bear images (inter-class). Best viewed in color.

We consider the test images collected from the ALOI dataset [9] shown in Fig. 1 (a)-(c). We generate 100 random transformations in $\mathcal{T} = SIM(2)$ and show some of these transformations applied to the Duck image in Fig. 1 (d). We construct the sparse patterns of the test images using a modified version of the Matching Pursuit algorithm [10] where we choose atoms that have highest *positive* correlation with the residual signal.

We compare our method with the regular euclidean distance, the tangent distance [4], and a local gradient descent approach that starts from the identity transformation and estimates, through a gradient descent procedure, the transformation that minimizes $\|U(\eta)I_1 - I_2\|_2$, where I_1 and I_2 are the original images that we wish to align. In Fig. 2, we show the average estimated transformation invariant distances for intra- and inter-class images computed with the different methods. Ideally, the distance between images of the same class (intra-class distance) should be smaller than the distance between images of different classes (inter-class distance). One can see that the euclidean distance cannot discriminate between intra-class images and inter-class images. Similarly, the tangent distance and the gradient descent approaches improve the performance only slightly since these methods provide local invariance. In our method, one can see that the intra-class distance is significantly smaller than the inter-class distance even for small values of the sparsity K . This shows that our registration algorithm works well even in the cases where the original images are not exactly K -sparse in \mathcal{D} .

3. THEORETICAL ANALYSIS

In this section, we examine the penalty of relaxing the problem (P) into (\hat{P}) . In other words, our goal is to find an upper bound on the registration error $E(p, q) = d_a(p, q) - d(p, q)$. In order to do so, we do the following simplifying assumptions:

$$\forall i \in \{1, \dots, K\}, \eta_0 \circ \gamma_i \in \mathcal{T}_d, \quad (4)$$

$$\forall i \in \{1, \dots, K\}, \eta_0^{-1} \circ \delta_i \in \mathcal{T}_d. \quad (5)$$

These hypotheses state that the atoms of $U(\eta_0)p$ and $U(\eta_0^{-1})q$ belong to the dictionary. Recall that $U(\eta_0)p$ is the optimal alignment of p with q and $U(\eta_0^{-1})q$ is the optimal alignment of q with p . We can assume that Eq. (4) and Eq. (5) hold when the parameter space used to design \mathcal{D} is discretized finely.

3.1. Performance with exactly transformed pattern

As a warm up, we consider the special case where $d(p, q) = 0$. In other words, we suppose that there exists a transformation $\eta_0 \in \mathcal{T}$ for which $q = U(\eta_0)p$ (i.e., the images can be aligned exactly). We impose in this case the linear independence of any subset of size $2K$ in \mathcal{D} , which in turn guarantees that any K -sparse signal has a unique decomposition in \mathcal{D} [11]. If this condition is satisfied, the following proposition guarantees a perfect registration:

Proposition 1. *Suppose that any subset of size $2K$ in \mathcal{D} is linearly independent. If $d(p, q) = 0$, then $d_a(p, q) = d(p, q) = 0$.*

The linear independence assumption guarantees that when two K -sparse signals are equal, they have at least one atom in common. If this condition is violated, the patterns $U(\eta_0)p$ and q can have several decompositions in the dictionary with disjoint supports. In this case, all the features of the transformed pattern $U(\eta_0)p$ and q are distinct, which generally leads to $d_a(p, q) \neq d(p, q)$.

3.2. Bound on the registration error in the general case

We study now the general case $d(p, q) < \epsilon\sqrt{\|c\|_2^2 + \|d\|_2^2}$, where ϵ accounts for the innovations between the patterns (other than a transformation in \mathcal{T}). We recall that c and d denote respectively the coefficients of patterns p and q . In order to account for the innovations in the performance analysis of the registration algorithm, we introduce two new key properties of the dictionary, namely *robust linear independence* and *transformation inconsistency*.

3.2.1. Robust linear independence

The linear independence assumption introduced in Section 3.1 is no longer sufficient to bound the registration performance in the case where $d(p, q) \neq 0$ (but infinitesimally close to zero). To see this, we construct a linearly independent dictionary \mathcal{D} and two sparse patterns p and q for which $d(p, q)$ can be made arbitrarily close to zero (i.e., $\epsilon \rightarrow 0$) yet the registration error is large. As illustrated in Fig. 3, we consider a dictionary \mathcal{D} containing 4 square atoms and an additional big square atom parametrized by its position κ with respect to ϕ_{γ_1} . Clearly, when $\kappa \neq 0$, the dictionary is linearly independent. We consider the patterns $p = \frac{1}{2} \sum_{i=1}^4 \phi_{\gamma_i}$ and $q = \phi_{\gamma_5}$ ⁴. All relative transformations between features in p and q are dilations composed with translations which result in a large registration error $d_a(p, q) - d(p, q)$. Since our algorithm estimates the global transformation by computing the best transformation between the features of both patterns, we obtain a large registration error.

This example shows that the linear independence assumption in Section 3.1 is fragile: it does not allow us to bound the registration error even when $d(p, q)$ is infinitesimally small. One needs a stronger condition than mere linear independence to guarantee a small registration error. We thus extend the notion of linear independence as follows:

⁴The optimal transformation η_0 in this case is a translation that aligns exactly p and q (i.e., $U(\eta_0)p = q$). However, η_0 does not satisfy the assumptions in Eq. (4), (5). For small κ , the transformation that best aligns the two patterns and satisfies the assumptions is the identity transformation.

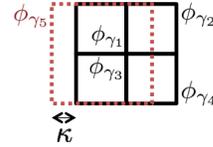


Fig. 3. Example of a linear independent dictionary \mathcal{D} for which our algorithm has a large registration error $d_a(p, q) - d(p, q)$, when $p = 1/2(\phi_{\gamma_1} + \phi_{\gamma_2} + \phi_{\gamma_3} + \phi_{\gamma_4})$ and $q = \phi_{\gamma_5}$. Note that ϵ (i.e., the innovation between p and q) in this particular example is equal to $\sqrt{\kappa}$; by choosing small values of κ , ϵ can thus be made arbitrarily small. Note also that this dictionary is linearly independent, but not robustly linearly independent for $K = 5$ (Definition 1).

Definition 1. *Let $(H, \|\cdot\|)$ be a normed space and $K \geq 1$. A family of vectors $(v_1, \dots, v_K) \in H^K$ is (ϵ, α) -robustly linearly independent (RLI) if any set $a \in \mathbb{R}^K$ satisfies:*

$$\left\| \sum_{i=1}^K a_i v_i \right\| < \epsilon \|a\|_2 \quad (6)$$

$$\implies \exists i, j \text{ with } a_i, a_j \neq 0, \left\| \frac{a_i v_i}{\|a_i v_i\|} + \frac{a_j v_j}{\|a_j v_j\|} \right\| \leq \alpha. \quad (7)$$

In simple words, when ϵ and α are small, any linear combination of vectors that nearly vanishes in a RLI family contains at least two vectors which approximately cancel each other. Armed with this notion, we introduce the following definition:

Definition 2. \mathcal{D} is (K, ϵ, α) -RLI if any subset of size K in \mathcal{D} is (ϵ, α) -RLI.

Since we focus in this analysis on the performance of the registration algorithm for small enough innovations, we examine the behaviour of α when ϵ is chosen to be small. We require in this case α to be close to zero. We say that a dictionary \mathcal{D} is K -RLI if α tends to zero when ϵ approaches zero.

The following dictionary provides a simple example of an infinite-size RLI dictionary.

Example 1 (Translations of a box function). *Let $H = L^2(\mathbb{R})$ and define the box function $v(t) = \begin{cases} 1 & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$. We define the dictionary $\mathcal{D}_{\text{box}} = \{T_\tau v = v_\tau : \tau \in \mathbb{R}\}$, where T_τ is the translation operator by τ .*

Let $K \geq 1$ and $\epsilon \in (0, \sqrt{\frac{3}{4^K - 1}})$. \mathcal{D}_{box} is $(K, \epsilon, \epsilon\sqrt{\frac{2}{3}(4^K - 1)})$ -RLI.

Using a straightforward extension of the definition of the Restricted Isometry Property (RIP) [8] to the case where vectors are in L^2 , it is not hard to see that the RIP constant of \mathcal{D}_{box} is equal to 1. Indeed, two box functions can be made arbitrarily close to each other. Similarly, the coherence [12] of \mathcal{D}_{box} is equal to 1. Nevertheless, the dictionary \mathcal{D}_{box} is K -RLI for any $K \geq 1$, since α goes to zero when ϵ tends to zero. One can understand the fact that \mathcal{D}_{box} is RLI intuitively; if a linear combination of box functions have a small norm, there exist at least two box functions that nearly cancel each other.

Even if the dictionary \mathcal{D}_{box} hardly satisfies the RIP and is highly coherent, it is still an interesting one in our framework. Indeed, it satisfies the key property that two sparse signals that are close in the

L^2 sense have at least two approximately similar features. When applied to our registration problem, this guarantees the existence of two features related approximately by a transformation η_0 when $d(p, q)$ remains small. This property is at the core of our registration algorithm since we infer the global transformation by looking at the local transformations between the features.

3.2.2. Transformation inconsistency

The second dictionary property that is important to study the performance of our algorithm is *the transformation inconsistency*, which measures the difference in the effect of the same transformation on distinct atoms in the dictionary. It is formally defined as follows for parametric dictionaries given by Eq. (1).

Definition 3. *The transformation inconsistency ρ of a parametric dictionary \mathcal{D} is defined by:*

$$\rho = \sup_{\gamma, \gamma' \in \mathcal{T}_d} \sup_{\eta \in \mathcal{T} \setminus \{\mathbb{I}\}} \left\{ \frac{\|U(\eta)\phi_{\gamma'} - \phi_{\gamma'}\|_2}{\|U(\eta)\phi_{\gamma} - \phi_{\gamma}\|_2} \right\},$$

where \mathbb{I} is the identity transformation. It is not hard to see that the transformation inconsistency ρ is always larger than or equal to 1. Furthermore, when \mathcal{T} is commutative group, the transformation inconsistency is equal to 1. On the other hand, a large value of ρ (i.e., $\rho \gg 1$) means that there exist two atoms in the dictionary that behave very differently when they are subject to the same transformation. In order to show the importance of the transformation inconsistency in our registration framework, we give the following illustrative example.

Example 2 (Dictionary built with translation and isotropic dilation, $\mathcal{T} = \mathcal{T}_d = \mathbb{R}^2 \times \mathbb{R}_+^*$). *Let \mathcal{T} to be the group of translations and isotropic dilations. The generating function of the dictionary could have any form, as long as its support is much smaller than the dimension of the image. For example, we can choose a circle-shaped mother function, as depicted in Fig 4 (a). Then, we consider the scenario where two atoms ϕ_{γ} and $\phi_{\gamma'}$ that are separated by z (where z is considered to be very large) as illustrated in Fig.4 (b). A transformation η that consists of a small isotropic dilation has a very different effect on both atoms. Indeed, the transformation η applied to $\phi_{\gamma'}$ results in an atom that has no intersection with $\phi_{\gamma'}$, while the same transformation has almost no effect on ϕ_{γ} , i.e., $U(\eta)\phi_{\gamma} \approx \phi_{\gamma}$. Thus, the transformation inconsistency is very high according to Definition 3. In Fig. 4 (c), we illustrate why this may cause a problem in our registration algorithm: we consider the two sparse patterns p and q composed of two features each, where the coefficients of all the atoms are equal. It is not hard to see in this case that the optimal global transformation between both patterns is the identity. At the same time, our algorithm can only estimate a global transformation that is a dilation (combined possibly with a translation) since all transformations between pairs of atoms in p and q consist in combinations of dilation and translation.*

This example provides insight on the importance of the transformation inconsistency parameter for our algorithm. As the global transformation between two sparse patterns is estimated from one of the relative transformations between features, it is preferable that transformations act in a similar way on all the features of the sparse patterns for more consistent registration. That means that dictionaries with small transformation inconsistency provide better registration performance.

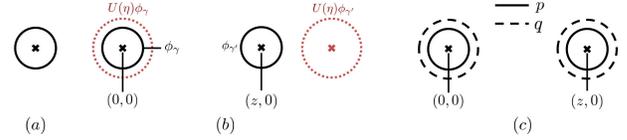


Fig. 4. Example of a dictionary where the transformation inconsistency ρ is large. (a): Mother function of the dictionary (b): Atoms ϕ_{γ} , $\phi_{\gamma'}$, and transformation η that causes ρ to be large. (c): Examples of patterns p (atoms represented with solid line) and q (atoms represented with dashed line) where our algorithm has a large registration error $d_a(p, q) - d(p, q)$.

3.2.3. Bound on the registration error

We now bound the performance of our registration algorithm in the general case.

Theorem 1. *If $d(p, q) < \epsilon \sqrt{\|c\|_2^2 + \|d\|_2^2}$ with $\epsilon > 0$, then:*

$$d_a(p, q) - d(p, q) \leq \alpha \rho \min(\|c\|_1, \|d\|_1),$$

when \mathcal{D} is $(2K, \epsilon, \alpha)$ -RLI for some $\alpha \in [0, \sqrt{2})$, and ρ is the transformation inconsistency of \mathcal{D} .

Theorem 1 shows that robust linear independence with a small α (for a fixed ϵ that depends on $d(p, q)$) and a small transformation inconsistency are key properties of the dictionary in order to guarantee the success of our algorithm. Note that these conditions on the dictionary are essentially tight, since we can construct example images where the algorithm fails, whenever the conditions are not met [3].

This bound guarantees the success of our registration algorithm for *all* sparse signals in the dictionary, when the RLI and transformation inconsistency conditions are satisfied. However, it does not predict the performance of the algorithm on a particular signal or class of signals.

Finally, this bound should be interpreted more in a qualitative way than a quantitative way. It provides two rather intuitive conditions for which our algorithm obtains low registration error. We stress here on the fact that such a bound could not have been established with traditional properties of the dictionary. In order to use this bound quantitatively, one has however to be able to compute explicitly the newly defined properties on generic dictionaries. We leave this interesting question for future work.

4. DISCUSSION

We have proposed in this paper a simple registration algorithm based on the sparse representation of the input images in a well chosen parametric dictionary. By construction, this dictionary is not well-behaved in terms of traditional dictionary properties: the coherence and RIP constants are very close to 1 when the dictionary is densely discretized. We circumvent this issue by introducing two dictionary properties; namely the robust linear independence (RLI) and transformation inconsistency and show that our algorithm has low registration error when these quantities remain small. To the best of our knowledge, this paper constitutes the first theoretically motivated work for image registration through sparse representation in redundant dictionaries. It provides understanding of the dictionary properties that drive the registration performance. A promising future direction is the use of the newly introduced quantities for dictionary learning.

5. REFERENCES

- [1] R. Szeliski, *Computer vision: algorithms and applications*, Springer, 2010.
- [2] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [3] A. Fawzi and P. Frossard, “Image registration with sparse approximations in parametric dictionaries,” *Arxiv preprint arXiv:1301.6646*, 2013.
- [4] P. Simard, Y. Le Cun, J. Denker, and B. Victorri, “Transformation invariance in pattern recognition: Tangent distance and propagation,” *International Journal of Imaging Systems and Technology*, vol. 11, no. 3, pp. 181–197, 2000.
- [5] 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.
- [6] B. Zitova and J. Flusser, “Image registration methods: a survey,” *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [7] R. Szeliski, “Image alignment and stitching: A tutorial,” *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.
- [8] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [9] J.M. Geusebroek, G.J. Burghouts, and A.W.M. Smeulders, “The Amsterdam library of object images,” *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005.
- [10] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
- [11] D.L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [12] J.A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.