View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Infoscience - École polytechnique fédérale de Lausanne

Vol. xx, pp. x x–x

Analysis of Image Registration with Tangent Distance

Elif Vural and Pascal Frossard *

Abstract. The computation of the geometric transformation between a reference and a target image, known as image registration or alignment, corresponds to the projection of the target image onto the transformation manifold of the reference image (the set of images generated by its geometric transformations). It often takes a nontrivial form such that exact computation of projections on the manifold is difficult. The tangent distance method is an effective alignment algorithm that exploits a linear approximation of the transformation manifold of the reference image. As theoretical studies about the tangent distance algorithm have been largely overlooked, we present in this work a detailed performance analysis of this useful algorithm, which can eventually help the selection of algorithm parameters. We consider a popular image registration setting using a multiscale pyramid of lowpass filtered versions of the (possibly noisy) reference and target images, which is particularly useful for recovering large transformations. We first show that the alignment error has a nonmonotonic variation with the filter size, due to the opposing effects of filtering on manifold nonlinearity and image noise. We then study the convergence of the multiscale tangent distance method to the optimal solution. We finally examine the performance of the tangent distance method in image classification applications. Our theoretical findings are confirmed by experiments on image transformation models involving translations, rotations and scalings. Our study is the first detailed study of the tangent distance algorithm that leads to a better understanding of its efficacy and to the proper selection of design parameters.

Key words. Image registration, tangent distance, image analysis, hierarchical registration methods, performance analysis.

1. Introduction. The estimation of the geometric transformation that gives the best match between a target image and a reference image is known as image registration or image alignment. This operation is commonly used in many problems in image processing or computer vision, such as image analysis, biomedical imaging, video coding and stereo vision. The set of images generated by the geometric transformations of a reference pattern is called a transformation manifold. In several image registration problems, it is possible to represent the geometric transformation between the reference and target images by a few parameters, e.g., translation, rotation, and affine transformation parameters. In this case, the image registration problem can be geometrically regarded as the projection of the target image onto the transformation manifold of the reference image. The transformation parameters that best align the image pair are then given by the transformation parameters of the manifold point that has the smallest distance to the target image. By extension, in image analysis problems where different classes are represented by different transformation manifolds, classification can be achieved by measuring the distance of the query image to the transformation manifold of each class.

Even if the image registration problem is generally not easy to solve exactly, its geometric interpretation allows for efficient alignment solutions. A well-known alignment method consists of constructing a first-order approximation of the transformation manifold of the

*E. Vural and P. Frossard are with Ecole Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Laboratory - LTS4, CH-1015 Lausanne, Switzerland (elif.vural@epfl.ch, pascal.frossard@epfl.ch).

reference image by computing the tangent space of the manifold at a reference point. The transformation parameters are then estimated by calculating the orthogonal projection of the target image onto the tangent space of the manifold. This method is known as the tangent distance method. The tangent distance method has been proposed by Simard et al. and its efficiency has been demonstrated in numerous settings, like handwritten digit recognition applications [17], [16] for example. Since then, many variations on the tangent distance method have been presented. The work in [7], for example, proposes the joint manifold distance for transformation-invariance in clustering, which is a similarity measure that is based on the prior distributions of the images and the distance between the linear approximations of their manifolds. The recent work [6] utilizes the tangent distance for motion compensation in video compression. In fact, some early examples of image alignment using manifold linearizations are found in the motion estimation literature, which are called gradient-based optical flow computation methods [18], [3]. Gradient-based methods exploit a linear approximation of the image intensity function in the estimation of the displacement between two image blocks. Applying a first-order approximation of the intensity function of the reference image block and then computing the displacement in a least-squares manner is actually equivalent to projecting the target image block onto the linear approximation of the manifold formed by the translations of the reference image block.

In image alignment with the tangent distance method, the reference manifold point around which the manifold is linearized is required to be sufficiently close to the exact projection of the target image onto the manifold, which corresponds to the optimal transformation parameters. In that case, the linear approximation of the manifold is valid and the optimal transformation parameters can be estimated accurately. When the distance between the reference and optimal transformation parameters is large, an efficient way to get around this limitation is to apply the tangent distance method in a hierarchical manner [16], [19]. In hierarchical alignment, a pyramid of low-pass filtered and downsampled versions of the reference and target images is built, and the alignment is achieved in a coarse-to-fine manner, which is illustrated in Figure 1.1. The transformation parameters are first roughly estimated using the smoothest images in the pyramid, and then progressively refined by passing to the fine scales. The low-pass filtering applied to generate the coarse-scale images helps to reduce the nonlinearity of the manifold, which renders the linear approximation more accurate and allows the recovery of relatively large transformations. Once the transformation parameters are estimated roughly from coarse scale images, the adjustment in the transformation parameters to be computed in fine scales is relatively small and the linear approximation of the manifold is therefore accurate. The study presented in [19] applies the multiresolution tangent distance method in image registration and image classification problems and experimentally shows that the dissimilarity measure obtained with the multiresolution tangent distance outperforms those obtained with the Euclidean distance and the single-scale tangent distance. The hierarchical estimation of transformation parameters using manifold linearizations is also very common in motion estimation [18], [3], and stereo vision [12]. While the efficiency of the hierarchical alignment strategy has been observed in many applications, a true characterization of the performance of this family of algorithms for general geometric transformation models is still missing in the literature.

In this work, we present a theoretical analysis of the properties of the tangent distance

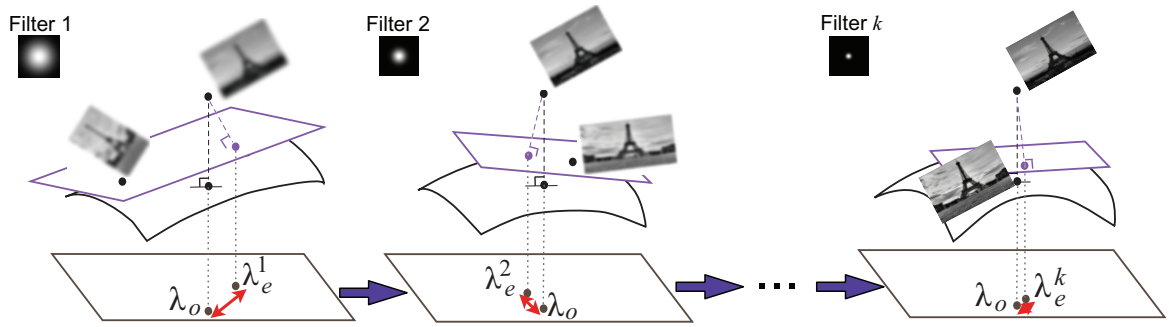


Figure 1.1. *Image alignment with the coarse-to-fine tangent distance method. The target image is a noisy and transformed version of the reference image. The optimal transformation parameters λ_o that best align the images are estimated in a coarse-to-fine manner with a pyramid of low-pass filtered versions of the images. The estimate λ_e^k of each stage is obtained by linearizing the transformation manifold of the reference image around the reference point given by the estimate λ_e^{k-1} of the previous stage. The sizes of the low-pass filters are decreased throughout the alignment algorithm as the estimates $\lambda_e^1, \lambda_e^2, \ldots, \lambda_e^k$ are refined progressively. (Photos in illustration borrowed from [1].)*

method in image alignment and image classification applications. The examination of the effect of filtering in the hierarchical alignment on the accuracy of the solution is especially important, so that the size of the low-pass filter can be properly selected at each stage of the multi-resolution representation. Therefore, an essential step in our study of the tangent distance method is the characterization of the alignment error as a function of the filter size. The second important parameter in our study is the influence of the additive noise that affects images, on the performance of the registration algorithm. We provide a complete analysis of the hierarchical tangent distance algorithm as a function of the manifold properties, the smoothing filter size and the image noise level, and observe the impact of these parameters in both image registration and image classification problems.

In particular, we consider a setting where the reference image is noiseless and the target image is a noisy and transformed version of the reference image. We first analyze the tangent distance method in the original image space (without filtering the images) and derive an upper bound for the alignment error, which is defined as the parameter-domain distance between the optimal transformation parameters that align the image pair perfectly, and their estimate computed with the tangent distance method. The upper bound for the alignment error is obtained in terms of the noise level of the target image, the parameter-domain distance between the reference manifold point (around which the manifold is linearized) and the actual projection onto the manifold, and some geometric parameters of the transformation manifold such as the curvature and the metric tensor. The alignment error bound linearly increases with the manifold curvature and the noise level, and monotonically increases with the parameter-domain distance between the reference and the optimal transformation parameters.

Next, we study the relation between the alignment error and the size of the low-pass filter in a hierarchical registration setting. We consider that both the reference and the target images are smoothed with a low-pass filter before alignment. We introduce an analytic and parametric

representation of the reference pattern and analyze the dependence of the alignment error on the size of the low-pass filter kernel. We show that the alignment error decreases with the filter size ρ for small filter kernels at a rate of $O(1+(1+\rho^2)^{-1/2})$. This is due to the fact that filtering does smooth the manifold and decrease its nonlinearity, which improves the accuracy of the linear approximation of the manifold. However, as one keeps increasing the filter size, the decrease in the alignment error due to the improvement of the manifold nonlinearity converges, and the error starts to increase with filtering at an approximate rate of $O(\rho)$ for relatively large values of the filter size. The increase in the error stems from the adverse effect of filtering, which amplifies the alignment error caused by image noise. Therefore, we show that, in a noisy setting where the target image is not exactly on the transformation manifold of the reference image, there is an optimal size for the filter kernel where the alignment error takes its minimum value.

We then build on our analysis of the alignment error and study the convergence of the hierarchical tangent distance method. We show that the tangent distance is guaranteed to converge to the optimal solution provided that (i) the product of the noise level and the manifold curvature is below a threshold that depends on the manifold dimension, and (ii) the amount of transformation between the reference and the target images is sufficiently small. Furthermore, we determine the optimal value of the filter size that minimizes the alignment error in each iteration of the hierarchical alignment algorithm. Our analysis shows that, the optimal update of the filter size ρ between adjacent iterations $k-1$ and k is approximately given by $\rho_k = \sqrt{\alpha}\,\rho_{k-1}$, where the geometric decay factor $\alpha < 1$ increases linearly with the noise level, the manifold curvature and the initialization error of the hierarchical alignment algorithm (i.e., the amount of transformation at the beginning of the algorithm). This result theoretically justifies the common strategy of reducing the filter size at a geometric rate, which is used very often in coarse-to-fine image registration. Meanwhile, although it is very common to update the filter size as $\rho_k = 1/2\,\rho_{k-1}$ with a constant decay factor of $1/2$ in practice [19], [5], our result suggests that the noise level, the expected amount of transformation, and the frequency characteristics of the images to be aligned must be taken into account in determining the best filter size updates.

Finally, we study the accuracy of image classification based on the manifold distance estimates obtained with the tangent distance method. In an image classification application where a query image is classified with respect to its distance to the transformation manifold of each class, the accuracy of classification largely depends on the accuracy of the estimation of the projection of the query image onto the manifolds. Therefore, one expects the classification performance to vary similarly to the alignment performance. We consider a setting where the query image and the reference images representing different classes are smoothed with low-pass filters. Then, we approximate the projection of the query image onto the transformation manifolds of the reference images with the tangent distance method. We determine the relation between the accuracy of classification and the size of the low-pass filter used for smoothing the images. Our result shows that, assuming bounded and non-intersecting distributions of the images around the transformation manifolds of their classes, the variation of the misclassification probability with the filter size is similar to that of the alignment error. Therefore, the filter size that minimizes the alignment error also minimizes the misclassification probability.

Our theoretical results about the alignment and classification performance of the tangent

distance method are confirmed by experiments conducted on transformation manifolds generated with rotations, translations and scale changes, both with synthetic smooth images and natural images. Our study provides insight into the principles behind the efficacy of the hierarchical alignment strategy in image registration and motion estimation and may be helpful for optimizing the performance of numerous image analysis algorithms that rely on first-order approximations of transformation manifolds.

The rest of the text is organized as follows. In Section 2, we introduce the notation, give an overview of the tangent distance algorithm, and formulate the problem. In Section 3, we present a theoretical analysis of the tangent distance method. We first look at the image registration problem, state an upper bound for the alignment error, examine its variation with the noise level and filtering, and study the convergence of the coarse-to-fine tangent distance method. We then extend these results to analyze the performance of image classification with the tangent distance. In Section 4, we evaluate our theoretical findings with some experiments. In Section 5, we give a discussion of our results in comparison with previous works. Finally, we conclude in Section 6.

2. Image Registration with Tangent Distance.

The computation of the exact projection of a target image onto a reference transformation manifold is a complicated optimization problem, especially when the manifold is high-dimensional and generated by complex geometric transformations. The tangent distance method proposes to solve this problem by using a first-order approximation of the transformation manifold, which is illustrated in Figure 2.1. In the figure, $\mathcal{M}(p)$ is the transformation manifold of the reference pattern p defined over the parameter domain Λ, and q is the target image to be aligned with p. The exact projection of q on $\mathcal{M}(p)$ is the point p_{λ_o}, so that λ_o is the optimal transformation parameter vector that best aligns p with q. In order to estimate λ_o with the tangent distance method, a first order approximation $\mathcal{S}_{\lambda_r}(p)$ of the manifold $\mathcal{M}(p)$ is computed at a reference point p_{λ_r}, which is preferably not too distant from p_{λ_o}. The distance of q to $\mathcal{S}_{\lambda_r}(p)$ can be easily computed with a least squares solution and the point of projection on $\mathcal{S}_{\lambda_r}(p)$ gives the transformation parameter vector λ_e, which is the estimate of λ_o.

Previous works such as [16] and [7] using the tangent distance in image classification and clustering compute the distance in a symmetric fashion; i.e., they linearize the transformation manifolds of both the reference and the target images and compute the subspace-to-subspace distance. In our analysis of the tangent distance method, we consider the point-to-subspace distance obtained by linearizing the transformation manifold of only the reference image. The point-to-subspace distance is more suitable than the subspace-to-subspace distance in image registration applications [19], [12], while it can also be used effectively in image analysis [19].

In the following, we first settle the notations and describe the tangent distance method formally. We then formulate the registration analysis problem studied in this work.

2.1. Notation.

Let $p \in L^2(\mathbb{R}^2)$ be a reference pattern that is C^2-smooth with square-integrable derivatives and $q \in L^2(\mathbb{R}^2)$ be a target pattern. Let $\Lambda \subset \mathbb{R}^d$ denote a compact, d-dimensional transformation parameter domain and $\lambda = [\lambda^1\ \lambda^2\ \cdots\ \lambda^d] \in \Lambda$ be a transformation parameter vector. We denote the pattern obtained by applying p the geometric transformation specified by λ as $A_\lambda(p) \in L^2(\mathbb{R}^2)$. Defining the spatial coordinate variable $X = [x\ y]^T$ in $\mathbb{R}^2$,

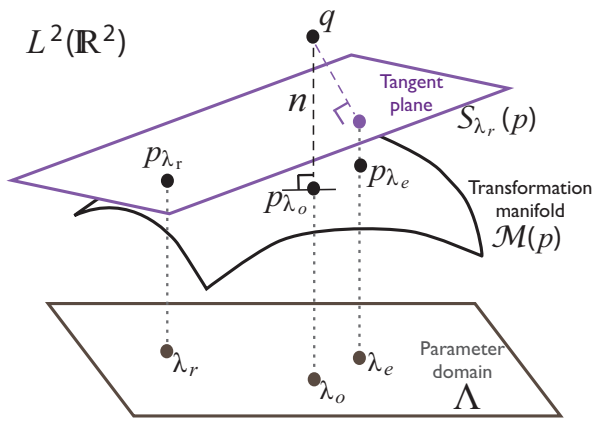


Figure 2.1. *Illustration of image alignment with the tangent distance method. $\mathcal{S}_{\lambda_r}(p)$ is the first-order approximation of the transformation manifold $\mathcal{M}(p)$ around the reference point p_{λ_r}. The estimate λ_e of the optimal transformation parameters λ_o is obtained by computing the orthogonal projection of the target image q onto $\mathcal{S}_{\lambda_r}(p)$.*

we can express the relation between $A_\lambda(p)$ and p as

$$A_\lambda(p)(X) = p(a(\lambda, X)) \tag{2.1}$$

where $a : \Lambda \times \mathbb{R}^2 \to \mathbb{R}^2$ is a C^2-smooth function representing the change of coordinates defined by the geometric transformation λ. We also assume that the coordinate change function $a_\lambda : \mathbb{R}^2 \to \mathbb{R}^2$ such that $a_\lambda(X) := a(\lambda, X)$, is a bijection for a fixed λ.

Let us write $p_\lambda = A_\lambda(p)$ for convenience. Then, the transformation manifold $\mathcal{M}(p)$ of the pattern p is given by

$$\mathcal{M}(p) = \{p_\lambda : \lambda \in \Lambda\} \subset L^2(\mathbb{R}^2)$$

which consists of transformed versions of p over the parameter domain Λ. Since a and p are C^2-smooth, the local embedding of $\mathcal{M}(p)$ in $L^2(\mathbb{R}^2)$ is C^2-smooth. Therefore, the first and second-order derivatives of manifold points with respect to the transformation parameters exist. Let us denote the derivative of the manifold point p_λ with respect to the i-th transformation parameter λ^i as $\partial_i\, p_\lambda$, where

$$\partial_i\, p_\lambda(X) = \frac{\partial\, p_\lambda(X)}{\partial \lambda^i}.$$

The derivatives $\partial_i\, p_\lambda$ are also called tangent vectors. Similarly, we denote the second-order derivatives by

$$\partial_{ij}\, p_\lambda(X) = \frac{\partial^2 p_\lambda(X)}{\partial \lambda^i \partial \lambda^j}.$$

Then, the tangent space $T_\lambda \mathcal{M}(p)$ of the manifold at a point p_λ is the subspace generated by the tangent vectors at p_λ

$$T_\lambda \mathcal{M}(p) = \left\{ \sum_{i=1}^{d} \partial_i\, p_\lambda\, \zeta^i : \zeta \in \mathbb{R}^d \right\} \subset L^2(\mathbb{R}^2) \tag{2.2}$$

where $\{\partial_i p_\lambda\}_{i=1}^d$ are the basis vectors of $T_\lambda \mathcal{M}(p)$, and $\{\zeta^i\}_{i=1}^d$ are the coefficients in the representation of a vector in $T_\lambda \mathcal{M}(p)$ in terms of the basis vectors.

Now, given the reference pattern p and a target pattern q, the image registration problem consists of the computation of an optimal transformation parameter vector λ_o that gives the best approximation of q on $\mathcal{M}(p)$,

$$\lambda_o = \arg\min_{\lambda \in \Lambda} \|q - p_\lambda\|^2 \tag{2.3}$$

where $\|\cdot\|$ denotes the L^2-norm for vectors in the continuous space $L^2(\mathbb{R}^2)$ and the ℓ^2-norm for vectors in the discrete space $\mathbb{R}^n$. Then, the transformed pattern p_{λ_o} is called a projection of q on $\mathcal{M}(p)$. However, the exact calculation of λ_o is difficult in general, since the nonlinear and highly intricate geometric structure of pattern transformation manifolds renders the distance minimization problem quite complicated. The tangent distance method simplifies this problem to a least squares problem, which is described below.

2.2. Tangent distance algorithm. In alignment with the tangent distance, transformation parameters are estimated by using a linear approximation of the manifold $\mathcal{M}(p)$ and then computing λ_o by minimizing the distance of q to the linear approximation of $\mathcal{M}(p)$ [19]. The first-order approximation of $\mathcal{M}(p)$ around a reference manifold point p_{λ_r} is given by

$$\mathcal{S}_{\lambda_r}(p) = \{p_{\lambda_r} + \sum_{i=1}^d \partial_i p_{\lambda_r}(\lambda^i - \lambda_r^i) : \lambda \in \mathbb{R}^d\} \subset L^2(\mathbb{R}^2). \tag{2.4}$$

Then, the estimate λ_e of λ_o with the tangent distance method is given by the solution of the least squares problem

$$\lambda_e = \arg\min_{\lambda \in \mathbb{R}^d} \|q - p_{\lambda_r} - \sum_{i=1}^d \partial_i p_{\lambda_r}(\lambda^i - \lambda_r^i)\|^2. \tag{2.5}$$

The solution of the above problem can be obtained as

$$\lambda_e = \lambda_r + [\mathcal{G}_{ij}(\lambda_r)]^{-1}[\langle q - p_{\lambda_r}, \partial_i p_{\lambda_r}\rangle] \tag{2.6}$$

where $[\mathcal{G}_{ij}(\lambda)] \in \mathbb{R}^{d\times d}$ is the matrix representation of the metric tensor $\mathcal{G}_{ij}(\lambda) = \langle \partial_i p_\lambda, \partial_j p_\lambda\rangle$ induced from the standard inner product on $L^2(\mathbb{R}^2)$. Hence, the (i,j)-th entry of $[\mathcal{G}_{ij}(\lambda)]$ is $\mathcal{G}_{ij}(\lambda)$. Similarly, $[\langle q - p_{\lambda_r}, \partial_i p_{\lambda_r}\rangle]$ represents the $d \times 1$ matrix whose i-th entry is $\langle q - p_{\lambda_r}, \partial_i p_{\lambda_r}\rangle$. The estimate λ_e of the transformation parameters obtained by solving (2.5) is expected to be closer to the optimal solution λ_o than the reference parameters λ_r; therefore, λ_e can be regarded as a refinement of λ_r if the reference parameters λ_r are considered as an initial guess for the optimal ones λ_o. The estimation of transformation parameters with the tangent distance method is illustrated in Figure 2.1.

2.3. Problem formulation. From (2.3), we can decompose the target image q as $q = p_{\lambda_o} + n$ where p_{λ_o} is the projection onto the manifold $\mathcal{M}(p)$ and $n \in L^2(\mathbb{R}^2)$ is the noise representing the deviation of q from $\mathcal{M}(p)$. We define the noise level parameter $\nu = \|n\|$ as the distance of the target pattern to the translation manifold of the reference pattern.

We can now formulate the problems that we study in this chapter. Our first purpose is to examine the deviation between the optimal transformation parameter vector λ_o and its estimate λ_e, which defines the alignment error of the tangent distance method. In particular, we would like to find an upper bound for the alignment error $\|\lambda_e - \lambda_o\|$ in terms of the noise level ν of the target image, the known geometric parameters of the manifold $\mathcal{M}(p)$ that can be computed from p (such as its curvature and metric tensor), and the distance $\|\lambda_o - \lambda_r\|$ between the optimal and the reference transformation parameters. This states a bound on how much the initial guess λ_r for λ_o can be improved, given the proximity of λ_r with respect to λ_o. We thus present an upper bound for the alignment error $\|\lambda_e - \lambda_o\|$ in Section 3.1. Note that it is also possible to formulate the alignment error as the manifold distance estimation error measured in the ambient space $L^2(\mathbb{R}^2)$. However, in this study, we characterize the error in the parameter space Λ instead of the ambient space $L^2(\mathbb{R}^2)$ because of the following reason. The errors in the parameter domain and the ambient space are expected to have similar behaviors. Meanwhile, since we examine the problem in a multiscale setting, it is easier to characterize the error in the parameter domain as the distances in the ambient space are not invariant to smoothing.

Next, our second and main goal is to examine how the alignment error varies when the reference and target patterns are smoothed with a low-pass filter. We formalize this problem as follows. We consider a Gaussian kernel for the low-pass filter, since it is a popular smoothing kernel whose distinctive properties have been well-studied in scale-space theory [11]. Let $\phi(X) = e^{-X^T X} = e^{-(x^2+y^2)}$ denote a Gaussian mother function. Then, the family of functions

$$\frac{1}{\pi\rho^2}\phi_\rho(X) \tag{2.7}$$

define variable-sized, unit L^1-norm Gaussian low-pass filters, where $\phi_\rho(X) = \phi(\Upsilon^{-1}(X))$ is a scaled version of the mother function $\phi(X)$ with

$$\Upsilon = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix}. \tag{2.8}$$

Here, the scale parameter ρ corresponds to the radius of the filter kernel, which controls the filter size. When the tangent distance method is used in a multiscale registration setting, the transformation parameters are estimated using the filtered versions of the reference and target patterns

$$\hat{p}(X) = \frac{1}{\pi\rho^2}(\phi_\rho * p)(X) \qquad \hat{q}(X) = \frac{1}{\pi\rho^2}(\phi_\rho * q)(X)$$

where $*$ denotes a convolution.

We write the parameters that are associated with the filtered versions of the reference and target patterns with the notation $\hat{(\cdot)}$. Now let $\hat{\lambda}_o$ be the transformation parameter vector corresponding to the projection of $\hat{q}$ onto the transformation manifold $\mathcal{M}(\hat{p})$ of the filtered reference pattern $\hat{p}$

$$\hat{\lambda}_o = \arg\min_{\lambda\in\Lambda} \|\hat{p}_\lambda - \hat{q}\|^2. \tag{2.9}$$

Hence, $\hat{\lambda}_o$ is the optimal transformation parameter vector that aligns $\hat{p}$ with $\hat{q}$. Let $\partial_i \hat{p}_\lambda$ and $\hat{\mathcal{G}}_{ij}$ denote respectively the first derivatives and the metric tensor of the manifold $\mathcal{M}(\hat{p})$. From

(2.6), the transformation estimate $\hat{\lambda}_e$ obtained with the filtered versions of the reference and target patterns by linearizing the manifold $\mathcal{M}(\hat{p})$ is given by

$$\hat{\lambda}_e = \lambda_r + [\hat{\mathcal{G}}_{ij}(\lambda_r)]^{-1}[\langle \hat{q} - \hat{p}_{\lambda_r}, \partial_i \hat{p}_{\lambda_r} \rangle]$$

where λ_r is the reference parameter vector. The alignment error obtained with the smoothed patterns is given as $\|\hat{\lambda}_e - \hat{\lambda}_o\|$, which we are interested in in this study. In particular, we would like to characterize the variation of $\|\hat{\lambda}_e - \hat{\lambda}_o\|$ with the size ρ of the low-pass filter used for smoothing the images in multiscale alignment, and the initial noise level ν of the target image before filtering. We thus examine in Section 3.2 the variation of the alignment error with noise and filtering.

3. Analysis of Tangent Distance Methods.

3.1. Upper bound for the alignment error. We now present an upper bound for the error of the alignment computed with the tangent distance method. We can assume that the parameter domain Λ is selected sufficiently large, so that p_{λ_o} is not on the boundary of $\mathcal{M}(p)$. Then, the noise pattern n is orthogonal to the tangent space of $\mathcal{M}(p)$ at p_{λ_o}. In other words, we have

$$\langle n, \partial_i p_{\lambda_o} \rangle = 0, \quad \forall i = 1, \cdots, d. \tag{3.1}$$

The deviation of the target image from the transformation manifold model impairs the estimation of transformation parameters. In our analysis of the alignment error, this deviation is characterized by the distance ν between q and $\mathcal{M}(p)$. Then, there is another source of error that causes the deviation of the estimated parameters λ_e from the optimal ones λ_o. It is related to the nonzero curvature of the manifold, as a result of which $\mathcal{M}(p)$ diverges from its linear approximation $\mathcal{S}_{\lambda_r}(p)$. In the derivation of the component of the alignment error associated with manifold nonlinearity, we make use of a quadratic approximation of the manifold around the reference point p_{λ_r}

$$p_\lambda \approx p_{\lambda_r} + \sum_{i=1}^{d} \partial_i p_{\lambda_r}(\lambda^i - \lambda_r^i) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \partial_{ij} p_{\lambda_r}(\lambda^i - \lambda_r^i)(\lambda^j - \lambda_r^j). \tag{3.2}$$

This approximation is treated as an equality in the derivation of the alignment error bound in Theorem 3.1. Equation (3.2) shows that the nonlinearity of the manifold can be characterized with an upper bound $\mathcal{K}$ on the norm of the second derivatives of the manifold

$$\mathcal{K} := \max_{i,j=1,\cdots,d} \sup_{\lambda \in \Lambda} \|\partial_{ij} p_\lambda\|. \tag{3.3}$$

Since $\mathcal{K}$ is an upper bound for the norms of the derivatives of tangent vectors, it can be regarded as a uniform curvature bound parameter for $\mathcal{M}(p)$.

We can now state our result that defines an upper bound on the alignment error.

Theorem 3.1. *The parameter-domain distance between the optimal transformation λ_o and its estimate λ_e given by the tangent distance method can be upper bounded as*

$$\|\lambda_e - \lambda_o\| \leq \; E := \mathcal{K}\; \lambda_{\min}^{-1}\left([\mathcal{G}_{ij}(\lambda_r)]\right)\left(\frac{1}{2}\sqrt{\mathrm{tr}([\mathcal{G}_{ij}(\lambda_r)])}\; \|\lambda_o - \lambda_r\|_1^2 + \sqrt{d}\;\nu\; \|\lambda_o - \lambda_r\|_1\right) \tag{3.4}$$

where $\lambda_{\min}(\cdot)$ *and* $\mathrm{tr}(.)$ *denote respectively the smallest eigenvalue and the trace of a matrix, and the notation* $\|\cdot\|_1$ *stands for the* ℓ^1*-norm in* $\mathbb{R}^n$*.*

Theorem 3.1 is proved in [21, Appendix A.1], which is a technical report where more details on our study are available. The result is obtained by examining the effects of both the nonlinearity of the manifold and the image noise on the alignment error. The theorem shows that the alignment error augments with the increase in the manifold curvature parameter $\mathcal{K}$ and the noise level ν, as expected. Moreover, another important factor affecting the alignment error is the distance $\|\lambda_o - \lambda_r\|$ between the reference and the optimal transformation parameters. If the reference manifold point p_{λ_r} around which the manifold is linearized is sufficiently close to the true projection of the target image onto the manifold, the tangent distance method is more likely to give a good estimate of the registration parameters.

3.2. Alignment error with low-pass filtering. We now analyze the influence of the low-pass filtering of the reference and target patterns on the accuracy of alignment with the tangent distance method as it is the case in multiscale registration algorithms. We consider a setting where the reference pattern p and the target pattern q are low-pass filtered and the transformation parameters are estimated with the smoothed versions of p and q. The purpose of this section is then to analyze the variation of the alignment error bound given in Theorem 3.1 with respect to the kernel size of the low-pass filter used in smoothing.

We first remark the following. The optimal transformation parameter vector $\hat{\lambda}_o$ corresponding to the smoothed patterns is in general different from the optimal transformation parameter vector λ_o corresponding to the unfiltered patterns p and q. This is due to the fact that both the image noise and the filtering cause a perturbation in the global minimum of the function $f(\lambda) = \|q - p_\lambda\|^2$, which represents the distance between the target pattern q and the transformed versions of the reference pattern p. Note that the overall error in the transformation parameter estimation is $\|\hat{\lambda}_e - \lambda_o\|$ and it can be upper bounded as

$$\|\hat{\lambda}_e - \lambda_o\| \leq \|\hat{\lambda}_e - \hat{\lambda}_o\| + \|\hat{\lambda}_o - \lambda_o\|.$$

Here, the first error term $\|\hat{\lambda}_e - \hat{\lambda}_o\|$ results from the linearization of the manifold, whereas the second error term $\|\hat{\lambda}_o - \lambda_o\|$ is due to the shift in the global minimum of the distance function $f(\lambda)$. The second error term $\|\hat{\lambda}_o - \lambda_o\|$ depends on the geometric transformation model. In our recent work [20], this error is examined for the transformation model of 2-D translations and its dependence on the noise level and low-pass filtering is studied. In this study, we analyze how the linearization of the manifold affects the estimation of the transformation parameters for generic transformation models. Therefore, we focus on the first error term $\|\hat{\lambda}_e - \hat{\lambda}_o\|$ associated particularly with the registration of the images using the tangent distance, and examine its variation with the noise level and the filtering process. The error term $\|\hat{\lambda}_e - \hat{\lambda}_o\|$ caused by the manifold linearization is in general expected to be dominant over the error term $\|\hat{\lambda}_o - \lambda_o\|$ unless the reference parameters λ_r are really close to the optimal parameters λ_o.

The filtered target pattern can be decomposed as

$$\hat{q} = \hat{p}_{\hat{\lambda}_o} + \tilde{n}$$

where the noise pattern $\tilde{n}$ is orthogonal to the tangent space $T_{\hat{\lambda}_o}\mathcal{M}(\hat{p})$ at $\hat{p}_{\hat{\lambda}_o}$. Let $\partial_{ij}\,\hat{p}_\lambda$ and $\hat{\mathcal{K}}$ denote the second order derivatives and the curvature bound parameter of the manifold

$\mathcal{M}(\hat{p})$. Then, from Theorem 3.1, the alignment error obtained with the smoothed patterns can be upper bounded as $\|\hat{\lambda}_e - \hat{\lambda}_o\| \leq \hat{E}$, where

$$\hat{E} = \hat{\mathcal{K}}\ \lambda_{\min}^{-1}\ \left([\hat{\mathcal{G}}_{ij}(\lambda_r)]\right) \left(\frac{1}{2}\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda_r)])}\ \|\hat{\lambda}_o - \lambda_r\|_1^2 + \sqrt{d}\ \|\tilde{n}\|\ \|\hat{\lambda}_o - \lambda_r\|_1\right). \tag{3.5}$$

In order to analyze the variation of $\hat{E}$ with filtering and noise, we examine the dependence of each term in the expression of $\hat{E}$ in (3.5) on the filter size ρ and the initial noise level ν of the unfiltered target image. First, the curvature parameter $\hat{\mathcal{K}}$ of the smoothed manifold is given by

$$\hat{\mathcal{K}} = \max_{i,j=1,\cdots,d}\ \sup_{\lambda\in\Lambda} \|\partial_{ij}\,\hat{p}_\lambda\|.$$

Hence, if a uniform estimate that is valid for all λ and (i,j) can be found for the rate of variation of $\|\partial_{ij}\,\hat{p}_\lambda\|$ with the filter size ρ, the curvature parameter $\hat{\mathcal{K}}$ then also has the same order of variation with ρ.

Next, the metric tensor of the smoothed manifold is given by $\hat{\mathcal{G}}_{ij}(\lambda_r) = \langle \partial_i\,\hat{p}_{\lambda_r}, \partial_j\,\hat{p}_{\lambda_r}\rangle$, and its trace is

$$\mathrm{tr}\big([\hat{\mathcal{G}}_{ij}(\lambda_r)]\big) = \sum_{i=1}^{d} \|\partial_i\,\hat{p}_{\lambda_r}\|^2.$$

Therefore, if the variation of $\|\partial_i\,\hat{p}_{\lambda_r}\|^2$ with the filter size ρ can be characterized uniformly (in a way that is valid for all λ_r and i), the trace $\mathrm{tr}\big([\hat{\mathcal{G}}_{ij}(\lambda_r)]\big)$ of the metric tensor will also have the same order of variation with ρ as $\|\partial_i\,\hat{p}_{\lambda_r}\|^2$.

The smallest eigenvalue $\lambda_{\min}\ \big([\hat{\mathcal{G}}_{ij}(\lambda_r)]\big)$ of the metric tensor is also expected to have the same variation with ρ. This can be observed, for instance, by decomposing the metric tensor into its diagonal and off-diagonal components and regarding the off-diagonal component as a perturbation on the diagonal one. The smallest eigenvalue $\lambda_{\min}\ \big([\hat{\mathcal{G}}_{ij}(\lambda_r)]\big)$ can then be lower bounded as in [15] in terms of the smallest diagonal element $\min_i \|\partial_i\,\hat{p}_{\lambda_r}\|^2$ and the spectral radius of the off-diagonal component of the metric tensor consisting of the terms $\langle \partial_i\,\hat{p}_{\lambda_r}, \partial_j\,\hat{p}_{\lambda_r}\rangle$. As the variation of the off-diagonal elements is upper bounded by the variation of the diagonal elements due to Cauchy-Schwarz inequality, the smallest eigenvalue $\lambda_{\min}\ \big([\hat{\mathcal{G}}_{ij}(\lambda_r)]\big)$ decays with ρ at the same rate as $\|\partial_i\,\hat{p}_{\lambda_r}\|^2$.

Finally, the norm $\|\tilde{n}\|$ of the noise component of $\hat{q}$ depends on both the filter size ρ and the initial noise level ν before filtering.

We study now Equation (3.5) in more details and derive first a relation between the norms $\|\partial_i\,\hat{p}_\lambda\|$, $\|\partial_{ij}\,\hat{p}_\lambda\|$ of the first and second-order manifold derivatives and the norms $\|N_\nabla\hat{p}\|$, $\|N_h\hat{p}\|$ of the gradient and Hessian magnitudes of the filtered reference pattern $\hat{p}$. We state the dependences of $\|N_\nabla\hat{p}\|$ and $\|N_h\hat{p}\|$ on the filter size ρ in Lemma 3.3, which is then used to obtain the variation of the manifold derivatives $\|\partial_i\,\hat{p}_\lambda\|$, $\|\partial_{ij}\,\hat{p}_\lambda\|$ with ρ in Corollary 3.4. Next, we establish the dependence of the norm $\|\tilde{n}\|$ of the noise component on ρ and ν in Lemma 3.5. Finally, all of these results are put together in our main result Theorem 3.6, where we present the rate of variation of the alignment error bound $\hat{E}$ with the filter size ρ and the initial noise level ν of the target image.

3.2.1. Analysis of $\|\partial_i \hat{p}_\lambda\|$ and $\|\partial_{ij} \hat{p}_\lambda\|$. Let us begin with the computation of the terms $\|\partial_i \hat{p}_\lambda\|$ and $\|\partial_{ij} \hat{p}_\lambda\|$. First, from the relation (2.1), we have $p_\lambda(X) = p(X')$ where $X' = a_\lambda(X)$. Let us denote the transformed coordinates as $X' = [x' \; y']^T$ and write the derivatives of the transformed coordinates with respect to the transformation parameters as

$$\partial_i x' = \frac{\partial x'}{\partial \lambda^i}, \qquad \partial_i y' = \frac{\partial y'}{\partial \lambda^i}, \qquad \partial_{ij} x' = \frac{\partial^2 x'}{\partial \lambda^i \, \partial \lambda^j}, \qquad \partial_{ij} y' = \frac{\partial^2 y'}{\partial \lambda^i \, \partial \lambda^j}.$$

Also, let

$$\partial_x p(X') = \left.\frac{\partial p(X)}{\partial x}\right|_{X=X'}, \qquad \partial_y p(X') = \left.\frac{\partial p(X)}{\partial y}\right|_{X=X'}$$
$$\partial_{xx} p(X') = \left.\frac{\partial^2 p(X)}{\partial x^2}\right|_{X=X'}, \qquad \partial_{xy} p(X') = \left.\frac{\partial^2 p(X)}{\partial x \, \partial y}\right|_{X=X'}, \qquad \partial_{yy} p(X') = \left.\frac{\partial^2 p(X)}{\partial y^2}\right|_{X=X'}$$

denote the partial derivatives of the reference pattern p evaluated at the point X'. Then, the derivatives of the manifold $\mathcal{M}(p)$ at p_λ are given by

$$\begin{aligned}
\partial_i p_\lambda(X) &= \partial_x p(X') \partial_i x' + \partial_y p(X') \partial_i y' \\
\partial_{ij} p_\lambda(X) &= \partial_{xx} p(X') \, \partial_i x' \partial_j x' + \partial_{xy} p(X') \, (\partial_i x' \partial_j y' + \partial_j x' \partial_i y') + \partial_{yy} p(X') \, \partial_i y' \partial_j y' \\
&\quad + \partial_x p(X') \, \partial_{ij} x' + \partial_y p(X') \, \partial_{ij} y'.
\end{aligned}$$

One can generalize this to the smoothed versions $\hat{p}$ of the reference pattern as

$$\begin{aligned}
\partial_i \hat{p}_\lambda(X) &= \partial_x \hat{p}(X') \partial_i x' + \partial_y \hat{p}(X') \partial_i y' \\
\partial_{ij} \hat{p}_\lambda(X) &= \partial_{xx} \hat{p}(X') \, \partial_i x' \partial_j x' + \partial_{xy} \hat{p}(X') \, (\partial_i x' \partial_j y' + \partial_j x' \partial_i y') + \partial_{yy} \hat{p}(X') \, \partial_i y' \partial_j y' \\
&\quad + \partial_x \hat{p}(X') \, \partial_{ij} x' + \partial_y \hat{p}(X') \, \partial_{ij} y'.
\end{aligned} \tag{3.6}$$

Notice that, in the above equations, the filtering applied on the reference pattern influences only the spatial derivatives of the reference pattern ($\partial_x \hat{p}$, $\partial_y \hat{p}$, $\partial_{xx} \hat{p}$, $\partial_{xy} \hat{p}$, $\partial_{yy} \hat{p}$), whereas the derivatives of the transformed coordinates ($\partial_i x'$, $\partial_i y'$, $\partial_{ij} x'$, $\partial_{ij} y'$) depend solely on the transformation model λ and are constant with respect to the filter size ρ. Therefore, the variation of $\|\partial_i \hat{p}_\lambda\|$ and $\|\partial_{ij} \hat{p}_\lambda\|$ with ρ is mostly determined by the variation of the spatial derivatives of the pattern with the filter size. We denote the gradient of $\hat{p}$ as $\nabla \hat{p}(X) = [\partial_x \hat{p}(X) \; \partial_y \hat{p}(X)]^T$ and the vectorized Hessian of $\hat{p}$ as

$$(h\hat{p})(X) = [\partial_{xx} \hat{p}(X) \; \partial_{xy} \hat{p}(X) \; \partial_{xy} \hat{p}(X) \; \partial_{yy} \hat{p}(X)]^T. \tag{3.7}$$

We then define the functions $N_\nabla \hat{p}$, $N_h \hat{p} : \mathbb{R}^2 \to \mathbb{R}$ such that $N_\nabla \hat{p}\,(X) = \|\nabla \hat{p}(X)\|$ and $N_h \hat{p}\,(X) = \|(h\hat{p})(X)\|$, which give the ℓ^2-norms of the gradient and the Hessian of $\hat{p}$ at X. Since we assume that the spatial derivatives of the pattern are square-integrable, the functions $N_\nabla \hat{p}$ and $N_h \hat{p}$ are in $L^2(\mathbb{R}^2)$. The equations in (3.6) show that the first derivatives of the manifold are proportional to the first derivatives of the pattern; and the second derivatives of the manifold depend linearly on both the first and the second derivatives of the pattern $p(X)$. One thus expects the L^2-norms of the manifold derivatives to be related to the L^2-norms of $N_\nabla \hat{p}$ and $N_h \hat{p}$ as

$$\begin{aligned}
\|\partial_i \hat{p}_\lambda\| &= O\left(\|N_\nabla \hat{p}\|\right) \\
\|\partial_{ij} \hat{p}_\lambda\| &= O\left(\|N_\nabla \hat{p}\| + \|N_h \hat{p}\|\right)
\end{aligned} \tag{3.8}$$

from the perspective of their dependence on the filter size ρ. These relations indeed hold and they are formally shown in [21, Appendix B.1].

Since we have established the connection between the manifold derivatives and the pattern spatial derivatives, it suffices now to determine how the spatial derivatives $\|N_\nabla \hat{p}\|$ and $\|N_h \hat{p}\|$ depend on the filter size ρ. In order to examine this, we adopt a parametric representation of the reference pattern p in an analytic dictionary. Let

$$\mathcal{D} = \{\phi_\gamma : \gamma = (\psi, \tau_x, \tau_y, \sigma_x, \sigma_y) \in \Gamma\} \subset L^2(\mathbb{R}^2) \tag{3.9}$$

be a parametric dictionary manifold such that each atom ϕ_γ in $\mathcal{D}$ is derived from an analytic mother function ϕ by a geometric transformation specified by the parameter vector γ. Here ψ is a rotation parameter, τ_x and τ_y denote translations in x and y directions, and σ_x and σ_y represent an anisotropic scaling in x and y directions. The dictionary is defined over the continuous parameter domain Γ, and an atom ϕ_γ is given by

$$\phi_\gamma(X) = \phi(\sigma^{-1}\,\Psi^{-1}\,(X - \tau)), \tag{3.10}$$

where

$$\sigma = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}, \quad \Psi = \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix}, \quad \tau = \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix} \tag{3.11}$$

denote respectively the scale change, rotation and translation matrices defining the atom ϕ_γ. We may consider that the parameter domain Γ is defined over the range of parameters $\psi \in [0, 2\pi)$, $\tau_x, \tau_y \in \mathbb{R}$, and $\sigma_x, \sigma_y \in \mathbb{R}^+$. It is shown in [2] (in the proof of Proposition 2.1.2) that the linear span of a dictionary $\mathcal{D}$ generated with respect to the transformation model in (3.9) is dense in $L^2(\mathbb{R}^2)$ if the mother function ϕ has nontrivial support; i.e., unless $\phi(X) = 0$ almost everywhere.

In our analysis, we select the generating mother function as the Gaussian function $\phi(X) = e^{-X^T X}$. The Gaussian function has good time-localization properties, it is easy to treat in derivations due to its well-studied properties, and it ensures that $Span(\mathcal{D})$ is dense in $L^2(\mathbb{R}^2)$. Therefore, any pattern $p \in L^2(\mathbb{R}^2)$ can be represented as the linear combination of a sequence of atoms in $\mathcal{D}$. In the rest of our analysis, we adopt a representation of p in $\mathcal{D}$

$$p(X) = \sum_{k=1}^{\infty} c_k\, \phi_{\gamma_k}(X) \tag{3.12}$$

where γ_k are the atom parameters and c_k are the atom coefficients. Our derivation of the variations of $\|N_\nabla \hat{p}\|$ and $\|N_h \hat{p}\|$ is based on this representation and we use some properties of Gaussian atoms in our analysis. Nevertheless, the conclusions of our analysis are general and valid for all reference patterns in $L^2(\mathbb{R}^2)$ since any square-integrable pattern can be represented in the Gaussian dictionary $\mathcal{D}$.

Now, applying the Gaussian filter in (2.7) on the reference pattern in (3.12), we obtain the filtered pattern as

$$\frac{1}{\pi\rho^2}\,(\phi_\rho * p)(X) = \frac{1}{\pi\rho^2} \sum_{k=1}^{\infty} c_k\,(\phi_\rho * \phi_{\gamma_k})(X)$$

from the linearity of the convolution operator. In order to evaluate the convolution of two Gaussian atoms, we use the following proposition [22].

Proposition 3.2. *Let* $\phi_{\gamma_1}(X) = \phi(\sigma_1^{-1}\,\Psi_1^{-1}\,(X-\tau_1))$ *and* $\phi_{\gamma_2}(X) = \phi(\sigma_2^{-1}\,\Psi_2^{-1}\,(X-\tau_2))$. *Then*

$$(\phi_{\gamma_1} * \phi_{\gamma_2})(X) = \frac{\pi|\sigma_1\sigma_2|}{|\sigma_3|}\phi_{\gamma_3}(X) \tag{3.13}$$

where $\phi_{\gamma_3}(X) = \phi(\sigma_3^{-1}\,\Psi_3^{-1}\,(X-\tau_3))$ *and the parameters of* ϕ_{γ_3} *are given by*

$$\tau_3 = \tau_1 + \tau_2, \qquad \Psi_3\,\sigma_3^2\,\Psi_3^{-1} = \Psi_1\,\sigma_1^2\,\Psi_1^{-1} + \Psi_2\,\sigma_2^2\,\Psi_2^{-1}.$$

Proposition 3.2 implies that, when an atom ϕ_{γ_k} of p is convolved with the Gaussian kernel, it becomes

$$\frac{1}{\pi\rho^2}\,(\phi_\rho * \phi_{\gamma_k})(X) = \frac{|\sigma_k|}{|\hat{\sigma}_k|}\phi_{\hat{\gamma}_k}(X) \tag{3.14}$$

where $\phi_{\hat{\gamma}_k}(X) = \phi(\hat{\sigma}_k^{-1}\,\hat{\Psi}_k^{-1}\,(X-\hat{\tau}_k))$ and

$$\hat{\tau}_k = \tau_k, \quad \hat{\Psi}_k = \Psi_k, \quad \hat{\sigma}_k = \sqrt{\Upsilon^2 + \sigma_k^2}. \tag{3.15}$$

Hence, when p is smoothed with a Gaussian filter, the atom $\phi_{\gamma_k}(X)$ with coefficient c_k is replaced by the smoothed atom $\phi_{\hat{\gamma}_k}(X)$ with coefficient

$$\hat{c}_k = \frac{|\sigma_k|}{|\hat{\sigma}_k|}c_k = \frac{|\sigma_k|}{\sqrt{|\Upsilon^2 + \sigma_k^2|}}c_k = \frac{\sigma_{x,k}\,\sigma_{y,k}}{\sqrt{(\rho^2 + \sigma_{x,k}^2)(\rho^2 + \sigma_{y,k}^2)}}c_k \tag{3.16}$$

where $\sigma_k = \mathrm{diag}(\sigma_{x,k},\, \sigma_{y,k})$. This shows that the change in the pattern parameters due to filtering can be captured by substituting the scale parameters σ_k with $\hat{\sigma}_k$ and replacing the coefficients c_k with $\hat{c}_k$. Then, the smoothed pattern $\hat{p}$ has the following representation in the dictionary $\mathcal{D}$

$$\hat{p}(X) = \sum_{k=1}^{\infty} \hat{c}_k\,\phi_{\hat{\gamma}_k}(X). \tag{3.17}$$

One can observe from (3.16) that the atom coefficients $\hat{c}_k$ of the filtered pattern $\hat{p}$ change with the filter size ρ at a rate

$$\hat{c}_k = O((1+\rho^2)^{-1}). \tag{3.18}$$

Also, from (3.15), the atom scale parameters of $\hat{p}$ are given by

$$\hat{\sigma}_{x,k} = \sqrt{\sigma_{x,k}^2 + \rho^2}, \qquad \hat{\sigma}_{y,k} = \sqrt{\sigma_{y,k}^2 + \rho^2} \tag{3.19}$$

which have the rate of increase

$$\hat{\sigma}_{x,k},\ \hat{\sigma}_{y,k} = O((1+\rho^2)^{1/2}) \tag{3.20}$$

with the filter size ρ.

We are now equipped with the necessary tools for examining the variations of $\|N_\nabla \hat{p}\|$ and $\|N_h \hat{p}\|$ with the filter size ρ. We state these in the following lemma.

Lemma 3.3. *The norms $\|N_\nabla \hat{p}\|$ and $\|N_h \hat{p}\|$ of the first and second-order variations of the pattern decrease with the filter size ρ at the following rates*

$$\|N_\nabla \hat{p}\| = O((1+\rho^2)^{-1}), \qquad \|N_h \hat{p}\| = O((1+\rho^2)^{-3/2}).$$

The proof of Lemma 3.3 is given in [21, Appendix B.2]. The above dependences are shown by deriving approximations of $\|N_\nabla \hat{p}\|$ and $\|N_h \hat{p}\|$ in terms of the atom parameters $\{\gamma_k\}$ and coefficients $\{c_k\}$. Their variations with the filter size ρ are then determined by building on the relations (3.20) and (3.18). The lemma not only confirms the intuition that the norms of the pattern gradient and Hessian should decrease with filtering, but also provides expressions for their rate of decay with the filter size ρ.

An immediate consequence of Lemma 3.3 is the following.

Corollary 3.4. *The norms $\|\partial_i \hat{p}_\lambda\|$, $\|\partial_{ij} \hat{p}_\lambda\|$ of the first and second-order manifold derivatives decrease with the filter size ρ at the following rates*

$$\|\partial_i \hat{p}_\lambda\| = O((1+\rho^2)^{-1}), \qquad \|\partial_{ij} \hat{p}_\lambda\| = O\left((1+\rho^2)^{-3/2} + (1+\rho^2)^{-1}\right).$$

Proof. The corollary follows directly from Lemma 3.3 and the relation between the manifold derivatives and the pattern derivatives given in (3.8). ■

Note that for large values of ρ, the second additive term of $O(1+\rho^2)^{-1}$ in $\|\partial_{ij} \hat{p}_\lambda\|$ dominates the first term of $O(1+\rho^2)^{-3/2}$, therefore $\|\partial_{ij} \hat{p}_\lambda\| = O((1+\rho^2)^{-1})$ for large ρ. However, we keep both additive terms in $\|\partial_{ij} \hat{p}_\lambda\|$ as we will see that the first term is important for characterizing the behavior of the alignment error bound for small values of the filter size. Corollary 3.4 will be helpful for determining the dependences of the curvature bound $\hat{\mathcal{K}}$ and the parameters related to the metric tensor $\hat{\mathcal{G}}_{ij}$ on the filter size. We will use it in our main result of Theorem 3.6.

3.2.2. Analysis of $\|\tilde{n}\|$. In the following lemma, we summarize the dependence of the noise level $\|\tilde{n}\|$ in the filtered target pattern, on the noise level ν in the original target pattern and the size ρ of the smoothing filter.

Lemma 3.5. *The distance $\|\tilde{n}\|$ between the filtered target pattern $\hat{q}$ and the transformation manifold $\mathcal{M}(\hat{p})$ of the filtered reference pattern $\hat{p}$ has a rate of variation of*

$$\|\tilde{n}\| = O\left((\nu+1)(1+\rho^2)^{-1/2}\right)$$

with the filter size ρ and the initial noise level ν for geometric transformation models that allow the change of the scale of the pattern p. The variation of $\|\tilde{n}\|$ is however given by

$$\|\tilde{n}\| = O\left(\nu(1+\rho^2)^{-1/2}\right)$$

if the geometric transformation model does not include a scale change.

The proof of Lemma 3.5 is given in [21, Appendix B.3]. The presented dependences are obtained by deriving a relation between the norm of the noise component $\tilde{n} = \hat{q} - \hat{p}_{\hat{\lambda}_o}$ and the filtered version $\hat{n}$ of the initial noise component $n = q - p_{\lambda_o}$. The lemma states that $\|\tilde{n}\|$ decreases with the filter size ρ at a rate of $O\left((1+\rho^2)^{-1/2}\right)$. Meanwhile, its dependence on the initial noise level ν differs slightly between transformation models that include a scale change or not. The noise term $\|\tilde{n}\|$ increases at a rate of $O(\nu)$ for transformations without a scale change; however, transformations with a scale change introduce an offset to the initial noise level to yield a variation of $O(\nu+1)$. This is due to the following reason. The initial noise level before filtering is given by the norm of $n = q - p_{\lambda_o}$, where $p_{\lambda_o} \in \mathcal{M}(p)$. Meanwhile, when the transformation model λ includes a scale change, the actions of filtering and transforming a pattern do not commute, and the filtered version $\widehat{p_{\lambda_o}}$ of p_{λ_o} does not lie on the transformation manifold $\mathcal{M}(\hat{p})$ of the filtered reference pattern $\hat{p}$ (see [21, Appendix B.3] for more details). The "lifting" of the base point $\widehat{p_{\lambda_o}}$ of $\hat{q}$ (with the decomposition $\hat{q} = \widehat{p_{\lambda_o}} + \hat{n}$) from the manifold $\mathcal{M}(\hat{p})$ further increases the distance between $\hat{q}$ and $\mathcal{M}(\hat{p})$, in addition to the deviation $\hat{n}$. The overall noise level in case of filtering is therefore larger than the norm of the filtered version $\hat{n}$ of n. Note that, for transformations involving a scale change, even if the initial noise level ν is zero, which means that $q \in \mathcal{M}(p)$, we have $\hat{q} \notin \mathcal{M}(\hat{p})$ after filtering. This creates a source of noise when the filtered versions of the image pair are used in the alignment.

3.2.3. Analysis of $\hat{E}$. We are now ready to present our main result, which states the dependence of the alignment error $\hat{E}$ on the initial noise level of the target pattern and the filter size.

Theorem 3.6. *The alignment error bound $\hat{E}$ obtained when the smoothed image pair is aligned with the tangent distance method is given by*

$$\hat{E} = \hat{E}_1 + \hat{E}_2$$

where the error component $\hat{E}_1$ resulting from manifold nonlinearity decreases at rate

$$\hat{E}_1 = O\left(1 + (1+\rho^2)^{-1/2}\right)$$

with the size ρ of the low-pass filter kernel used for smoothing the reference and target images. The second component $\hat{E}_2$ of the alignment error associated with image noise has the variation

$$\hat{E}_2 = O\left((\nu+1)\,(1+\rho^2)^{1/2}\right)$$

with the filter size ρ and the noise level ν if the geometric transformation model includes a scale change. The variation of $\hat{E}_2$ with ρ and ν is

$$\hat{E}_2 = O\left(\nu\,(1+\rho^2)^{1/2}\right)$$

if the geometric transformation model does not change the scale of the pattern.

Proof. Remember from (3.5) that the alignment error bound is given by

$$\hat{E} = \hat{E}_1 + \hat{E}_2$$

where the error terms

$$\begin{aligned}\hat{E}_1 &= \frac{1}{2}\,\hat{\mathcal{K}}\;\lambda_{\min}^{-1}\;\left([\hat{\mathcal{G}}_{ij}(\lambda_r)]\right)\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda_r)])}\;\|\hat{\lambda}_o-\lambda_r\|_1^2\\ \hat{E}_2 &= \sqrt{d}\;\hat{\mathcal{K}}\;\lambda_{\min}^{-1}\;\left([\hat{\mathcal{G}}_{ij}(\lambda_r)]\right)\;\|\tilde{n}\|\;\|\hat{\lambda}_o-\lambda_r\|_1\end{aligned} \tag{3.21}$$

are associated respectively with the nonzero manifold curvature (lifting of the manifold from the tangent space) and the noise on the target image. Also, remember that the variation of $\hat{\mathcal{K}}$ with ρ is the same as that of $\|\partial_{ij}\,\hat{p}_\lambda\|$, and that $\lambda_{\min}\left([\hat{\mathcal{G}}_{ij}(\lambda_r)]\right)$ and $\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda_r)])$ have the same variation with ρ as $\|\partial_i\,\hat{p}_{\lambda_r}\|^2$. Hence, using Corollary 3.4, we obtain

$$\hat{\mathcal{K}}\;\lambda_{\min}^{-1}\;\left([\hat{\mathcal{G}}_{ij}(\lambda_r)]\right) = O\left(1+\;(1+\rho^2)^{-1/2}\right)O(1+\rho^2) \tag{3.22}$$

$$\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda_r)])} = O\left((1+\rho^2)^{-1}\right) \tag{3.23}$$

which gives

$$\hat{E}_1 = O\left(1+\;(1+\rho^2)^{-1/2}\right).$$

Then, from Lemma 3.5 and Equation (3.22), we determine the variation of $\hat{E}_2$ as

$$\hat{E}_2 = O\left((\nu+1)\,(1+\rho^2)^{1/2}\right)O\left(1+(1+\rho^2)^{-1/2}\right) \approx O\left((\nu+1)\,(1+\rho^2)^{1/2}\right)$$

for transformations involving a scale change, and as

$$\hat{E}_2 = O\left(\nu\,(1+\rho^2)^{1/2}\right)O\left(1+(1+\rho^2)^{-1/2}\right) \approx O\left(\nu\,(1+\rho^2)^{1/2}\right)$$

for transformations without a scale change, which finishes the proof of the theorem. ■

Theorem 3.6 can be interpreted as follows. The first error component $\hat{E}_1$ related to manifold nonlinearity is of $O\left(1+\;(1+\rho^2)^{-1/2}\right)$. Since filtering the patterns makes the manifold smoother and decreases the manifold curvature, it improves the accuracy of the first-order approximation of the manifold used in tangent distance. Therefore, the first component of the alignment error decreases with the filter size ρ. Then, we observe that the second error component $\hat{E}_2 = O\left((\nu+1)\,(1+\rho^2)^{1/2}\right)$ resulting from image noise, is proportional to the noise level, as expected, but it also increases with the filter size ρ. The increase of the error with smoothing is due to the fact that filtering has the undesired effect of amplifying the alignment error caused by the noise. This result is in line with the findings of our previous study [20], and previous works such as [14], [23] that examine the Crámer-Rao lower bound in image registration. This is discussed in more detail in Section 5.

The dependence of the overall alignment error on the filter size can be interpreted as follows. For reasonably small values of the image noise level, the overall error $\hat{E}$ first decreases with the filter size ρ at small filter sizes due to the decrease in the first term $\hat{E}_1$, since filtering improves the manifold linearity. As one keeps increasing the filter size, the first error term $\hat{E}_1 = O\left(1+\;(1+\rho^2)^{-1/2}\right)$ gradually decreases and finally converges to a constant value. After that, the second error term $\hat{E}_2$ takes over and the overall alignment error $\hat{E}$ starts to increase with the filter size. The amplification of the registration error resulting from the image noise

then becomes the prominent factor that determines the overall dependence of the error on the filter size. As the alignment error first decreases and then increases with filtering, there exists an optimal value of the filter size ρ for a given noise level ν. In the noiseless case where $\nu = 0$, our result shows that applying a big filter is favorable as it flattens the manifold, provided that the transformation model does not involve a scale change. Meanwhile, for geometric transformations involving a scale change, there exists a nontrivial optimal filter size even in the noiseless case $\nu = 0$, which is due to the secondary source of noise discussed in Lemma 3.5 arising in such transformation models.

The results obtained in this section provide a characterization of the alignment error of the tangent distance method in multiscale image registration. The understanding of the behavior of the error in case of low-pass filtering provides a means for optimizing the performance of the tangent distance algorithm by adapting the filter size to the characteristics of the image data. In the next section, we examine the implications of our findings in hierarchical image registration applications.

3.3. Convergence analysis of tangent distance. We now use the results obtained in Sections 3.1 and 3.2 to analyze the convergence behavior of the tangent distance method in a general setting where the target image is a noisy transformed version of the reference image. We first examine the conditions under which the tangent distance converges to the correct solution at a single scale without filtering. We then generalize this to the convergence of the coarse-to-fine tangent distance method and derive some practical guidelines for optimal filter selection in each scale of the hierarchical alignment process.

3.3.1. Convergence of the single-scale registration algorithm. Consider that the tangent distance method is applied in an iterative manner, starting with the reference parameter vector λ_r and then refining it gradually by taking the estimate from the previous iteration as the reference transformation parameter vector in each iteration. In this way, we obtain a sequence of estimates $\lambda_e^0, \lambda_e^1, \ldots, \lambda_e^k$ where the initial estimate is $\lambda_e^0 = \lambda_r$ and each subsequent estimate λ_e^k is computed by linearizing the manifold around the point given by the previous parameter estimate λ_e^{k-1}.

First, based on the alignment error bound (3.4) in Theorem 3.1, we define the following geometric constants on $\mathcal{M}(p)$:

$$C_1 := \sup_{\lambda \in \Lambda} \sqrt{\operatorname{tr}([\mathcal{G}_{ij}(\lambda)])}\ , \qquad C_2 := \mathcal{K}\ \sup_{\lambda \in \Lambda} \lambda_{\min}^{-1}\ \big([\mathcal{G}_{ij}(\lambda)]\big). \tag{3.24}$$

The parameter C_1 is a constant bounding the magnitude of the tangent vectors since it scales with the supremum of the tangent norms. Similarly, the parameter C_2 is a normalized curvature constant, as the inverse of the metric tensor $[\mathcal{G}_{ij}(\lambda)]$ normalizes the inner products with tangent vectors in the least-squares estimation of transformation parameters in (2.6). The geometric constants C_1 and C_2 thus bound the magnitudes of the first-order and second-order variations of the manifold.

In the next theorem, we focus on a single-scale setting where no filtering is done throughout the iterations. We state conditions guaranteeing that the estimates $\lambda_e^0, \lambda_e^1, \ldots, \lambda_e^k$ converge to the optimal transformation parameters λ_o.

Theorem 3.7. *Let the product of the noise level ν and the curvature constant C_2 be upper bounded as follows*

$$\nu\, C_2 < \frac{1}{d}. \tag{3.25}$$

Furthermore, let

$$\|\lambda_o - \lambda_r\| < \frac{2}{C_1}\left(\frac{1}{dC_2} - \nu\right) \tag{3.26}$$

denote an upper bound on the distance between the reference transformation parameters λ_r and the optimal transformation parameters λ_o. If the initialization of the tangent distance algorithm is sufficiently accurate to satisfy the above upper bound, then the successive estimates given by the iterative application of the tangent distance method at a single scale converge to the optimal solution λ_o

$$\lim_{k\to\infty} \lambda_e^k = \lambda_o.$$

Theorem 3.7 is proved in [21, Appendix C.1] by using the error bound in Theorem 3.1. Theorem 3.7 can be interpreted as follows. First, we observe from the condition in (3.25) that the noise level - curvature product must be below a certain level to recover the correct solution. It has been seen in Theorem 3.1 that the alignment error is affected by both the manifold nonlinearity and the noise level. The condition (3.25) thus excludes the case where both the curvature and the noise level take large values, in order to ensure that the tangent distance method yields an accurate estimation.

Next, the inequality (3.26) implies that the accuracy of the initial solution must satisfy

$$\|\lambda_o - \lambda_r\| \leq O\left(\frac{1}{\mathcal{K}} - \nu\right).$$

This condition requires the initial alignment error to be inversely proportional to the manifold curvature in a noiseless setting. Meanwhile, in a noisy setting, the increase in the noise level also brings a restriction on the accuracy of the initial solution λ_r in order to preserve the convergence guarantee. In particular, the initialization error $\|\lambda_o - \lambda_r\|$ must decrease linearly with the increase in the noise level ν. The overall dependence of the initialization error $\|\lambda_o - \lambda_r\|$ on $\mathcal{K}$ and ν is intuitive in the sense that, as the curvature of the manifold approaches 0, the accuracy of the linear approximation of the manifold increases, and the tangent distance method can recover the correct solution for arbitrarily large values of the initialization error even in the presence of noise.

3.3.2. Convergence of the coarse-to-fine registration algorithm. We now study the convergence of the tangent distance method when it is implemented in a hierarchical coarse-to-fine manner, with image filtering at each successive level. Let the estimation λ_e^k be obtained by linearizing the manifold around the point corresponding to the parameter λ_e^{k-1} as above. Consider however that, in iterations $1, 2, \ldots, k$, the reference and the target images are filtered with low-pass Gaussian filters of size $\rho_1, \rho_2, \ldots, \rho_k$. We would like to investigate under which conditions the hierarchical alignment process converges to the correct solution. Additionally, we would like to derive practical guidelines for selecting the size of the filters in hierarchical alignment algorithms.

First, we observe from (3.21) that the alignment error in iteration k can be upper bounded as follows:

$$\begin{aligned}\|\lambda_e^k - \hat{\lambda}_o\| \leq & \frac{1}{2}\hat{\mathcal{K}}\ \lambda_{\min}^{-1}\ \left([\hat{\mathcal{G}}_{ij}(\lambda_e^{k-1})]\right)\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda_e^{k-1})])}\ \|\hat{\lambda}_o - \lambda_e^{k-1}\|_1^2 \\ & + \sqrt{d}\,\hat{\mathcal{K}}\ \lambda_{\min}^{-1}\ \left([\hat{\mathcal{G}}_{ij}(\lambda_e^{k-1})]\right)\ \|\tilde{n}\|\|\hat{\lambda}_o - \lambda_e^{k-1}\|_1.\end{aligned}$$

Ignoring the small perturbation $\|\hat{\lambda}_o - \lambda_o\|$ due to filtering in the projection of the target pattern onto the manifold, we can approximate $\hat{\lambda}_o \approx \lambda_o$. Also, bounding the ℓ^1-norms in the above expression in terms of ℓ^2-norms, we obtain

$$\begin{aligned}\|\lambda_e^k - \lambda_o\| \leq & \frac{1}{2}d\,\hat{\mathcal{K}}\ \lambda_{\min}^{-1}\ \left([\hat{\mathcal{G}}_{ij}(\lambda_e^{k-1})]\right)\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda_e^{k-1})])}\ \|\lambda_o - \lambda_e^{k-1}\|^2 \\ & + d\,\hat{\mathcal{K}}\ \lambda_{\min}^{-1}\ \left([\hat{\mathcal{G}}_{ij}(\lambda_e^{k-1})]\right)\ \|\tilde{n}\|\|\lambda_o - \lambda_e^{k-1}\|.\end{aligned} \tag{3.27}$$

Remember that, for any fixed $\lambda \in \Lambda$, the terms $\hat{\mathcal{K}}\ \lambda_{\min}^{-1}\ \left([\hat{\mathcal{G}}_{ij}(\lambda)]\right)$ and $\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda)])}$ have a variation with the filter size ρ as given in (3.22) and (3.23). Moreover, at $\rho = 0$, the definitions in (3.24) give the suprema of these terms attained over Λ. From these two relations, we deduce that the following inequalities

$$\begin{aligned}\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}(\lambda)])} &\leq \beta_1 C_1\,(1+\rho^2)^{-1} \\ \hat{\mathcal{K}}\ \lambda_{\min}^{-1}\ \left([\hat{\mathcal{G}}_{ij}(\lambda)]\right) &\leq \beta_2 C_2\left(1 + (1+\rho^2)^{-1/2}\right)(1+\rho^2)\end{aligned} \tag{3.28}$$

hold for some constants β_1 and β_2. The above expressions capture the dependence of these two terms on the filter size ρ as well as on the tangent magnitude and curvature constants C_1 and C_2. In the above inequalities, we omit the constants appearing in the exact variations of these terms with the filter size for the sake of simplicity. From the definitions of C_1 and C_2 in (3.24), we observe that taking $\beta_1 = 1$ and $\beta_2 = 1/2$ results in equalities in (3.28) for the case $\rho = 0$. In the following, we adopt these values for the constants β_1 and β_2. Although this choice does not guarantee the inequalities in (3.28) for all values of ρ, this approximation simplifies our analysis and allows us to obtain an approximate expression for the variation of the alignment error with the filter size ρ that holds up to a multiplication by a constant. Evaluating the expressions in (3.28) at ρ_k and using them in (3.27), we obtain

$$\begin{aligned}\|\lambda_e^k - \lambda_o\| \leq & \frac{1}{4}d\,C_1C_2\left(1 + (1+\rho_k^2)^{-1/2}\right)\ \|\lambda_o - \lambda_e^{k-1}\|^2 \\ & + \frac{1}{2}d\,C_2\left(1 + (1+\rho_k^2)^{-1/2}\right)(1+\rho_k^2)\,\|\tilde{n}\|\,\|\lambda_o - \lambda_e^{k-1}\|.\end{aligned} \tag{3.29}$$

We now define an effective noise level parameter ν_e such that

$$\nu_e = \begin{cases} \nu + \nu_s & \text{if the transformation model includes a scale change} \\ \nu & \text{otherwise}\end{cases}$$

where ν_s is a constant that represents the secondary noise term seen in Lemma 3.5; it results from the non-commutativity of filtering and scaling. Using the result in Lemma 3.5, we can approximate the noise term $\|\tilde{n}\|$ in iteration k as

$$\|\tilde{n}\| \approx \nu_e\,(1+\rho_k^2)^{-1/2}.$$

Finally, using this relation in (3.29) gives the following upper bound E_k for the alignment error in iteration k

$$\|\lambda_e^k - \lambda_o\| \leq E_k$$

where

$$\begin{aligned} E_k := & \frac{1}{4} d\, C_1 C_2 \left(1 + (1+\rho_k^2)^{-1/2}\right) \|\lambda_o - \lambda_e^{k-1}\|^2 \\ & + \frac{1}{2} d\, C_2 \nu_e \left(1 + (1+\rho_k^2)^{-1/2}\right) (1+\rho_k^2)^{1/2} \|\lambda_o - \lambda_e^{k-1}\|. \end{aligned} \tag{3.30}$$

We proceed by determining the optimal value of the filter size ρ_k to be used in iteration k. From (3.30), we obtain the value of ρ that minimizes E_k as follows.

$$\rho_k = \sqrt{\frac{C_1 \|\lambda_o - \lambda_e^{k-1}\|}{2\,\nu_e} - 1} \quad \text{if } \|\lambda_o - \lambda_e^{k-1}\| \geq \frac{2\,\nu_e}{C_1} \tag{3.31}$$

$$\rho_k = 0 \qquad \text{if } \|\lambda_o - \lambda_e^{k-1}\| < \frac{2\,\nu_e}{C_1} \tag{3.32}$$

This indicates that the optimal filter size must be chosen large if the current estimation error $\|\lambda_o - \lambda_e^{k-1}\|$ at the beginning of iteration k is large. The noise level of the target image also influences the optimal filter size. It must be chosen inversely proportional to the square root of the noise level, because of the increase of the alignment error with filtering. The above relations provide a justification of the strategy of reducing the filter size gradually in coarse-to-fine alignment, since the successive estimates $\{\lambda_e^k\}$ approach the optimal solution progressively and the estimation error $\|\lambda_o - \lambda_e^k\|$ decreases throughout the iterations of the hierarchical alignment algorithm. In particular, the result in (3.31)-(3.32) shows that, when the estimation error decreases below a threshold that depends on the noise level, it is better to stop filtering the images and to use their original versions in the alignment process.

Now it is easy to generalize our result in Theorem 3.7 to convergence conditions for the coarse-to-fine tangent distance method.

Corollary 3.8. *Let the product of the effective noise level ν_e and the curvature constant C_2 be upper bounded as follows*

$$\nu_e\, C_2 < \frac{1}{d}. \tag{3.33}$$

Furthermore, let the initialization error of the hierarchical tangent distance algorithm be bounded as

$$\|\lambda_o - \lambda_r\| < \frac{2}{C_1}\left(\frac{1}{dC_2} - \nu_e\right). \tag{3.34}$$

Then, if the filter size ρ_k in each iteration is chosen to be between 0 and the optimal filter size given in (3.31)-(3.32), the successive estimates of the hierarchical tangent distance method converge to the optimal solution

$$\lim_{k\to\infty} \lambda_e^k = \lambda_o.$$

The proof of Corollary 3.8 is given in [21, Appendix C.2]. The corollary builds on the observation that the above selection of the filter size yields an error that is not larger than

the error obtained by applying no filtering and then follows the same lines as in the proof of Theorem 3.7.

In a practical implementation of the tangent distance method, it is not easy to exactly compute the optimal value of the filter size in (3.31)-(3.32) since the alignment error $\|\lambda_o - \lambda_e^k\|$ in an arbitrary iteration is not exactly known. However, using our results, we can deduce a suitable rule for updating the filter sizes ρ_k in practice. As shown in the proof of Corollary 3.8, if the noise level and the distance between the reference and optimal transformation parameters are sufficiently small, the alignment error upper bounds $\{E_k\}$ in the iterative registration process decay at a geometric rate such that

$$E_k \leq \alpha\, E_{k-1}, \tag{3.35}$$

where

$$\alpha = \frac{1}{2} d C_1 C_2 E_0 + d\ \nu_e C_2 < 1 \tag{3.36}$$

and $E_0 = \|\lambda_o - \lambda_r\|$ denotes the initialization error. Now, from (3.35), the alignment error bound E_k in iteration k is bounded as $E_k \leq \alpha^k E_0$, which gives

$$\|\lambda_o - \lambda_e^k\| \leq E_k \leq \alpha^k E_0 = \alpha^k \|\lambda_o - \lambda_e^0\|.$$

Due to the relation $\|\lambda_o - \lambda_e^k\| \leq \alpha^k \|\lambda_o - \lambda_e^0\|$ for all k, one may expect the actual alignment errors $\|\lambda_o - \lambda_e^k\|$ to decay at the same rate α as well. Thus, a reasonable approximation for the relation between the alignment errors in adjacent iterations is given by

$$\|\lambda_o - \lambda_e^k\| \approx \alpha\, \|\lambda_o - \lambda_e^{k-1}\|.$$

Applying this approximation in the expressions of the optimal filter sizes in (3.31)-(3.32), we then get the following for the update of the filter size

$$\rho_k \approx \sqrt{\alpha}\, \rho_{k-1}. \tag{3.37}$$

Notice that, at the early stages of the alignment, the alignment error is large. Then, ignoring the subtractive constant in (3.31) yields the above approximation. Meanwhile, in the late stages of the iterative alignment, the error is small; the geometric decay of the filter sizes in the update rule (3.37) makes ρ_k approach 0, which approximates well the selection $\rho_k = 0$ in (3.32).

The filter size update rule in (3.37) is in agreement with the common practice of reducing the filter size in a geometric manner. While it is typical to reduce the filter size by a factor of $\alpha = 1/2$ in the implementation of hierarchical image registration algorithms [19], [5], we can now reinterpret the selection of the factor α in the light of our results. First, an immediate consequence of the linear proportion between the decay factor α in (3.36) and the curvature parameter C_2 is that α should increase with manifold nonlinearity. This is in agreement with the expectation that applying large filters throughout the iterations improves the accuracy of the linear approximation of the manifold. Similarly, the decay factor α is seen to increase linearly with the initialization error E_0. This shows that adapting α to the accuracy of the initial solution helps to mitigate the influence of the initialization error, which propagates

and affects the estimates of the algorithm throughout the iterations. Finally, regarding the dependence of the filter update strategy on the noise level, we observe the following. From (3.31), we observe that the initial filter size ρ_1 in iteration 1 must be chosen as

$$\rho_1 \approx \sqrt{\frac{C_1 E_0}{2\nu_e}}.$$

Therefore, at small values of the noise level ν_e, one can begin with a relatively large filter size ρ_1 in the first iteration. The decay factor α takes a small value in this case, which is useful for speeding up the convergence of the algorithm. On the other hand, at high noise levels, the above expression for ρ_1 suggests that the initial filter size should be chosen small. The factor α becomes larger in this case; therefore, the decay in the filter size between adjacent iterations needs to be slower.

We have studied in this section the convergence of the multiscale tangent distance method and shown that the convergence of the algorithm is guaranteed if the noise level, the curvature and the initialization error are sufficiently small. Moreover, we have shown that, in the coarse-to-fine tangent distance method, the optimal choice of the filter size depends on the data and transformation model characteristics. Providing an insight into the performance of multiscale image registration, the results of this section can be used in devising effective tools for image registration and analysis.

3.4. Analysis of the error in classification problems. We have so far studied the registration performance of the tangent distance method. Meanwhile, the tangent distance method is also used commonly in image analysis problems for the transformation-invariant estimation of the similarity between a query image and a set of image manifold models representing different classes. A typical similarity measure is the distance between the query image and the class-representative transformation manifolds. Since the distances to the manifolds are computed by estimating the projection of the query image onto the manifolds, the accuracy of the distance estimation is highly influenced by the accuracy of the estimation of the transformation parameters. The classification performance is thus quite related to the registration performance.

In this section, we study the link between the image classification and registration problems and extend our results on the registration analysis to study the performance of the tangent distance method in image classification. Consider a setting with M class-representative patterns $\{p^m\}_{m=1}^M$ whose transformation manifolds

$$\mathcal{M}(p^m) = \{p^m_\lambda : \lambda \in \Lambda\} \subset L^2(\mathbb{R}^2)$$

are used for the classification of query patterns $q \in L^2(\mathbb{R}^2)$ in the image space. We assume that the correct class label $l(q)$ of a query pattern q is given by the class label of the manifold $\mathcal{M}(p^m)$ with smallest distance to it, i.e.,

$$l(q) = \arg \min_{m \in \{1,\ldots,M\}} \|q - p^m_{\lambda^m_o}\| \tag{3.38}$$

where $\lambda^m_o = \arg\min_{\lambda\in\Lambda} \|q - p^m_\lambda\|$ is the optimal transformation parameter vector corresponding to the projection of q on $\mathcal{M}(p^m)$.

Our purpose is then to study in this context the performance penalty when the class label of a query pattern is estimated by employing first-order approximations of the manifolds. Obviously, if the transformation parameters are estimated with an iterative application of the tangent distance method (at a single scale or in a coarse-to-fine manner), the convergence guarantees to the optimal solution established in Theorem 3.7 and Corollary 3.8 ensure that the target pattern be correctly classified. Hence, in this section, we focus on the accuracy of classifying a query image with a one-step application of the tangent distance method, i.e., by estimating the transformation parameters $\{\lambda_o^m\}$ with a single linearization of each manifold, possibly by filtering the target and reference images. We study the performance of classification in this setting and its dependence on the choice of the filter size.

Let λ_e^m denote the estimate of λ_o^m computed with the tangent distance method as in (2.6) by linearizing the manifold $\mathcal{M}(p^m)$ around a reference point with parameter vector λ_r^m. The class label of q is then estimated with the tangent distance method as follows[1]

$$\tilde{l}(q) = \arg \min_{m \in \{1,\dots,M\}} \|q - p^m_{\lambda_e^m}\|. \tag{3.39}$$

Comparing the estimated class label in (3.39) and the true class label in (3.38), it can be observed that the performance of classification depends on the accuracy of the estimation of the transformation parameters. In particular, if the estimate $\|q - p^m_{\lambda_e^m}\|$ of the distance between the query pattern and the manifold is sufficiently close to the true manifold distance $\|q - p^m_{\lambda_o^m}\|$ for each one of the manifolds, the estimated class label $\tilde{l}(q)$ in (3.39) is the same as the true class label $l(q)$. Based on this observation, we study the classification performance of the tangent distance method as follows. First, given a reference pattern p and a target pattern q, we derive a relation between the distance estimation error $\big|\|q - p_{\lambda_o}\| - \|q - p_{\lambda_e}\|\big|$ and the alignment error $\|\lambda_o - \lambda_e\|$ in the parameter domain in the following lemma.

Lemma 3.9. *The distance estimation error of the tangent distance method can be upper bounded in terms of its alignment error as*

$$\big|\|q - p_{\lambda_o}\| - \|q - p_{\lambda_e}\|\big| \leq \mathcal{T}\, \|\lambda_o - \lambda_e\|_1, \tag{3.40}$$

where $\mathcal{T}$ denotes the supremum of the tangent norms on $\mathcal{M}(p)$

$$\mathcal{T} := \max_{i=1,\dots,d} \sup_{\lambda \in \Lambda} \|\partial_i\, p_\lambda\|. \tag{3.41}$$

The proof of Lemma 3.9 is given in [21, Appendix D.1]. Lemma 3.9 provides a link between the accuracy of the alignment measured in the parameter domain Λ, and in the ambient space $L^2(\mathbb{R}^2)$, respectively. It shows that the distance estimation error can be upper bounded with a linear function of the alignment error.

The relation in (3.40) suggests that one may expect the classification performance of the tangent distance method to vary linearly with the accuracy of alignment in the parameter

[1] Note that the class label of a query image can also be estimated by comparing its distance to the first-order approximation $\mathcal{S}_{\lambda_r^m}(p^m)$ of each manifold defined in (2.4). While Simard et al. use this subspace distance for classification [16], the estimate in (3.39) is also commonly used in image analysis problems (e.g., as in [19]). We base our analysis on the definition in (3.39) since it is likely to give more accurate estimates, especially when it is generalized to a multiscale setting as in (3.42).

domain. In order to construct a more precise relation, we now consider a setting where the query images of class m have a distribution that is concentrated around the manifold $\mathcal{M}(p^m)$. We then examine the probability of correctly classifying q based on the distance estimates given by the tangent distance method.

Using the notation of Section 2.3, let $\nu_j = \|q - p^j_{\lambda^j_o}\|$ denote the deviation of a query image q from the manifold $\mathcal{M}(p^j)$ of class j. Furthermore, let q belong to class m. The distance of q to $\mathcal{M}(p^m)$ is the smallest among the distances of q to all manifolds; therefore, $\nu_m < \nu_j$ for all $j \neq m$. Let us assume that the distributions of the images belonging to different classes have bounded and non-intersecting supports around the manifolds, so that the classification rule in (3.38) always gives the true class label. We can then define the following parameters. Let

$$\mathcal{V}_m := \sup_{q:\ l(q)=m} \left\{ \|q - p^m_{\lambda^m_o}\| \right\}$$

denote the maximal distance of query patterns of class m to the manifold $\mathcal{M}(p^m)$ of their own class and

$$\epsilon := \min_{m=1,\dots,M;\ j\neq m} \ \inf_{q:\ l(q)=m} \left\{ \|q - p^j_{\lambda^j_o}\| - \|q - p^m_{\lambda^m_o}\| \right\}$$

define a distance margin that is a measure of the minimum separation between different classes. Finally, let $\mathcal{T}_m$ and $\mathcal{K}_m$ denote the suprema of the tangent norm and the curvature on the manifold $\mathcal{M}(p^m)$, as defined in (3.41) and (3.3) respectively. We then have the following result, which provides an upper bound for the probability of misclassifying a target image of class m.

Theorem 3.10. *Let q be a query pattern of class m. Assume that the optimal transformation parameters λ^m_o aligning q with p^m are within a Δ-neighborhood of the reference transformation parameters λ^m_r around which $\mathcal{M}(p^m)$ is linearized, such that $\|\lambda^m_o - \lambda^m_r\|_1 \leq \Delta$. Then, the probability of misclassifying q with the tangent distance method is upper bounded as*

$$P\left(\tilde{l}(q) \neq l(q)\right) \leq \frac{(M-1)}{\epsilon} \mathcal{T}_m \sqrt{d}\, \mathcal{K}_m\ \lambda^{-1}_{\min}\ \left([\mathcal{G}^m_{ij}(\lambda^m_r)]\right) \left(\frac{1}{2} \sqrt{\mathrm{tr}([\mathcal{G}^m_{ij}(\lambda^m_r)])}\ \Delta^2 + \sqrt{d}\ \mathcal{V}_m\ \Delta \right)$$

where d is the dimension of the manifolds and $[\mathcal{G}^m_{ij}(\lambda^m_r)]$ denotes the metric tensor of manifold $\mathcal{M}(p^m)$ at the point corresponding to λ^m_r.

The proof of Theorem 3.10 is given in [21, Appendix D.2]. The above result is obtained by upper bounding the probability of misclassification in terms of the distance estimation error. The distance estimation error is linked to the alignment error in the parameter domain using Lemma 3.9, which is then upper bounded using Theorem 3.1.

Theorem 3.10 shows how the probability of misclassification when the manifold distances are estimated with the tangent distance method, depends on the geometric properties of the manifolds and on the deviation Δ between the reference transformation parameters λ^m_r used in the linearization of the manifold and the optimal transformation parameters λ^m_o corresponding to the projection of q onto the manifold. In particular, for any non-intersecting and bounded distribution of class samples, the misclassification probability increases at most linearly with the increase in the manifold curvature and the maximal distance of the images to their own representative manifold. The deviation Δ between the parameters used in the linearization and

the parameters corresponding to the exact projection affects the misclassification probability due to its influence on the alignment accuracy. We also observe that better separation of manifolds (i.e., increase in the distance margin ϵ) reduces the probability of misclassification, as expected.

We now discuss the classification of images with the tangent distance method in a multi-scale setting and study the selection of the filter size in order to minimize the misclassification probability. Consider that the transformation parameters are estimated by filtering the query image $\hat{q}$ and the reference images $\hat{p}^m$. From (2.6), the following estimates $\{\hat{\lambda}_e^m\}$ are obtained for the classes $m = 1, \ldots, M$ by registering the query image on each class manifold with the tangent distance method

$$\hat{\lambda}_e^m = \lambda_r^m + [\hat{\mathcal{G}}_{ij}^m(\lambda_r^m)]^{-1}[\langle \hat{q} - \hat{p}_{\lambda_r}^m, \partial_i \hat{p}_{\lambda_r}^m \rangle].$$

Here $\hat{\mathcal{G}}_{ij}^m$ and $\partial_i \hat{p}_{\lambda_r}^m$ are respectively the metric tensor and the tangent vectors on the manifold $\mathcal{M}(\hat{p}^m)$. Once the transformation parameters are estimated, we assume that the unfiltered versions of the reference images and the query image are used in the computation of the actual distances to the manifolds for estimating the class label of the query image. It is preferable to compare the distances in the original image space rather than the space of filtered images, as it yields more accurate estimates. The class label estimate of the query pattern is thus given by

$$\tilde{l}(q) = \arg \min_{m \in \{1,\ldots,M\}} \|q - p_{\hat{\lambda}_e^m}^m\|. \tag{3.42}$$

Repeating the steps in the proof of Theorem 3.10 by replacing the estimates $\{\lambda_e^m\}$ with the ones $\{\hat{\lambda}_e^m\}$ obtained after filtering the reference and target patterns, one can upper bound the misclassification probability as

$$P\left(\tilde{l}(q) \neq l(q)\right) \leq \frac{(M-1)}{\epsilon} \mathcal{T}_m \sqrt{d}\, \hat{\mathcal{K}}_m\, \lambda_{\min}^{-1}\left([\hat{\mathcal{G}}_{ij}^m(\lambda_r^m)]\right) \left(\frac{1}{2}\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}^m(\lambda_r^m)])}\, \Delta^2 + \sqrt{d}\, \hat{\mathcal{V}}_m\, \Delta\right) \tag{3.43}$$

when the filtered images are used for estimating the transformation parameters. We have neglected the perturbation $\|\lambda_o - \hat{\lambda}_o\|$ due to filtering in the projection of patterns onto the manifold. The above expression for the misclassification probability is in the same form as the alignment error bound in (3.5); they only differ by a multiplicative factor (note, however, that the value of this factor depends on the geometric properties of the manifolds through the parameters $\mathcal{T}_m$ and ϵ). Therefore, the misclassification probability bound has the same non-monotonic variation with the filter size as the alignment error. Moreover, the optimal value of the filter size that minimizes the alignment error is a minimizer of the misclassification probability upper bound as well. In an image classification application where a one-step linear approximation of the manifolds is employed, one may thus choose the optimal filter size by minimizing the alignment error. The model parameters should then be selected with respect to the expected characteristics of the data. The maximal distance $\mathcal{V}_m$ is related to the internal variation (noise level) of the data samples within the same class and depends on how well the reference pattern p^m approximates the samples of its own class, whereas the parameter Δ can be set according to the maximum amount of transformation that the data samples are likely to undergo in the application at hand.

4. Experimental Results.

4.1. Image alignment. We now present experimental results that illustrate our alignment error bounds. In all settings, we experiment on three different geometric transformation models, namely a two-dimensional translation manifold

$$\mathcal{M}(p) = \{A_\lambda(p) : \lambda = (t_x, t_y) \in \Lambda\}, \tag{4.1}$$

a three-dimensional manifold given by the translations and rotations of a reference pattern

$$\mathcal{M}(p) = \{A_\lambda(p) : \lambda = (\overline{\theta}, t_x, t_y) \in \Lambda\}, \tag{4.2}$$

and a four-dimensional manifold generated by the translations, rotations and isotropic scalings of a reference pattern

$$\mathcal{M}(p) = \{A_\lambda(p) : \lambda = (\overline{\theta}, t_x, t_y, \overline{s}) \in \Lambda\}. \tag{4.3}$$

In the above models, t_x and t_y represent translations in x and y directions, $\overline{\theta}$ denotes a rotation parameter, and $\overline{s}$ is a scale change parameter. The parameters $\overline{\theta}$ and $\overline{s}$ are normalized versions of the actual rotation angle θ and scale change factor s, so that the magnitudes of the manifold derivatives with respect to t_x, t_y, $\overline{\theta}$, and $\overline{s}$ are proportional.

In all experiments, several target patterns are generated from a reference pattern by applying a random geometric transformation according to the above models. The target patterns are then corrupted with additive noise patterns at different noise levels ν. For each reference and target pattern pair (p, q), a sequence of image pairs $(\hat{p}, \hat{q})$ are obtained by smoothing p and q with low-pass filters with different kernel sizes ρ. Then, the target pattern $\hat{q}$ in each image pair is aligned with the reference pattern $\hat{p}$ using the tangent distance method, where the reference parameter vector λ_r is taken as identity such that $\hat{p}_{\lambda_r} = \hat{p}$. The experimental alignment error is measured as the parameter domain distance $\|\hat{\lambda}_e - \hat{\lambda}_o\|$ between the optimal transformation parameter vector $\hat{\lambda}_o$ and its estimate $\hat{\lambda}_e$. Then, the experimental alignment error is compared to its theoretical upper bound $\hat{E}$ given in Theorem 3.1. The curvature parameter $\mathcal{K}$ is computed numerically in the implementation of Theorem 3.1.

In the first set of experiments, we experiment on 50 different reference patterns that consist of 20 atoms randomly selected from the Gaussian dictionary $\mathcal{D}$. The atom parameters are randomly drawn from the intervals $\psi \in [-\pi, \pi)$; $\tau_x, \tau_y \in [-4, 4]$; $\sigma_x, \sigma_y \in [0.3, 2.3]$; and the atom coefficients are randomly selected within the range $[-1, 1]$. Then, for each one of the models (4.1)-(4.3), 10 target patterns are generated for each reference pattern. The transformation parameters of target patterns are selected randomly within the ranges $\overline{\theta} \in [-0.4, 0.4]$; $t_x, t_y \in [-0.4, 0.4]$; and $\overline{s} \in [0.4, 1.6]$. The above ranges for the normalized rotation and scale parameters $\overline{\theta}$ and $\overline{s}$ correspond to the actual rotation angles $\theta \in [-0.04\pi, 0.04\pi]$ and scale change factors $s \in [0.87, 1.13]$. Each target pattern is corrupted with a different realization of a noise pattern that consists of 100 small-scale Gaussian atoms with random coefficients drawn from a normal distribution, which represents a random noise pattern in the continuous domain. The noise patterns are normalized to match a range of noise levels ν.

The results obtained for the transformation models (4.1), (4.2), and (4.3) are presented respectively in Figures 4.1, 4.2 and 4.3, where the performance is averaged over all reference and target patterns. In all figures, the experimental alignment errors and their theoretical

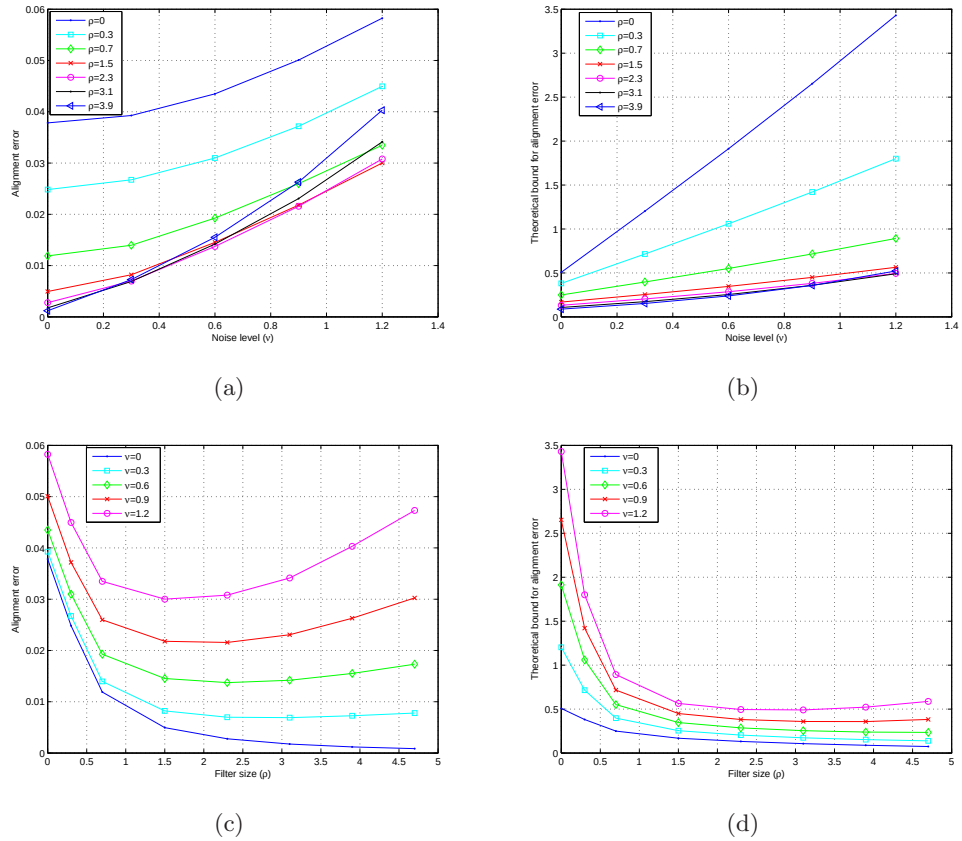


Figure 4.1. *Alignment errors of random patterns for 2-D manifolds generated by translations.*

upper bounds are plotted with respect to the noise level ν in panels (a) and (b), where the noise level ν is normalized with the norm $\|p\|$ of the reference pattern. The same experimental errors and theoretical bounds are plotted as functions of the filter size ρ in panels (c) and (d) of all figures.

The results of this experiment can be interpreted as follows. First, the plots in panels (a) and (b) of Figures 4.1-4.3 show that the variation of the alignment error with the noise level ν approaches an approximately linear rate for large values of ν both in the empirical and the theoretical plots. This confirms the estimations $\hat{E} = O(\nu)$, $\hat{E} = O(\nu + 1)$ of Theorem 3.6. Next, the plots in (c) and (d) of the figures show that the actual alignment error and its theoretical upper bound decrease with filtering at small filter sizes ρ, as smoothing decreases the nonlinearity of the manifold. The error then begins to increase with the filter size ρ at larger values of ρ in the presence of noise. This confirms that the filter size has an optimal value when the target image is noisy, as predicted by Theorem 3.6. The shift in the optimal value of the filter size with the increase in the noise level is observable especially in Figures 4.1 and 4.2, which is in agreement with the approximate relation between the optimal filter size and the noise level given in (3.31). Moreover, in most plots, the optimal value of the filter size that minimizes the theoretical upper bound in (d) is seen to be in the vicinity of

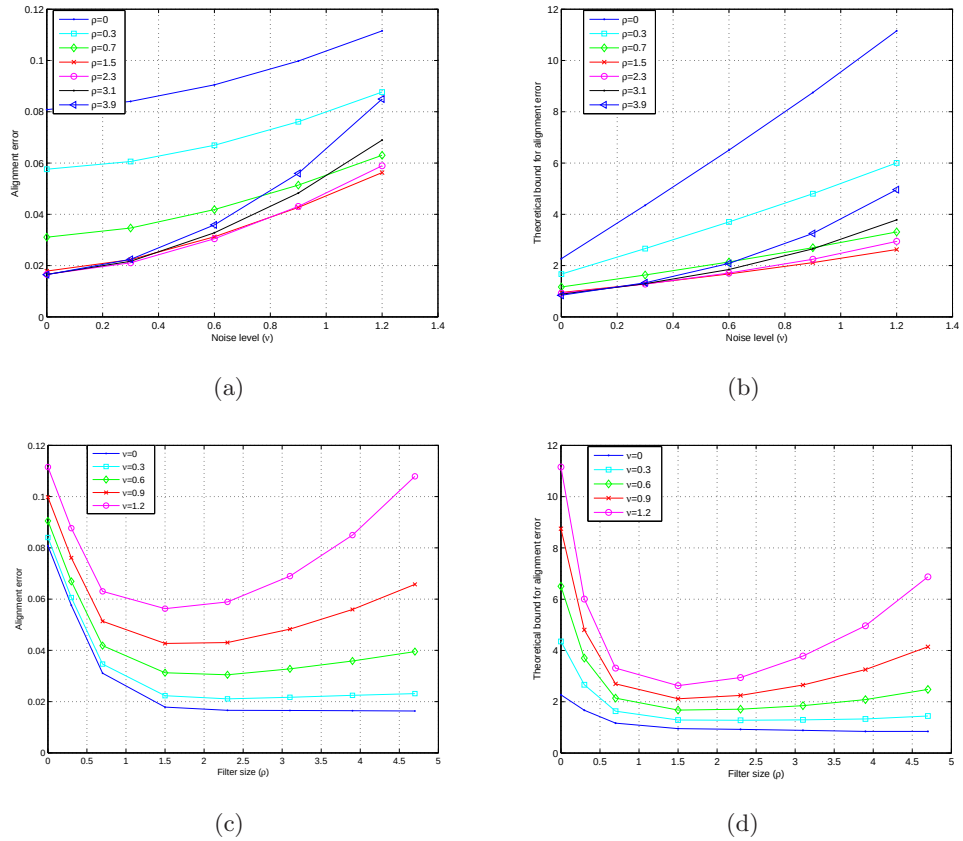


Figure 4.2. *Alignment errors of random patterns for 3-D manifolds generated by translations and rotations.*

the optimal filter size minimizing the actual alignment error in (c), which shows that the theoretical bound provides a good prediction of suitable filter sizes in alignment. The results also show that the variation of the alignment error with the filter size approximately matches the rate $\hat{E} = O\left((1+\rho^2)^{1/2}\right) \approx O(\rho)$ at large filter sizes in most plots.

It is also interesting to compare the behavior of the alignment error between different transformation models. To begin with, one can observe in Figures 4.1(c) and 4.1(d) that, for two-dimensional translation manifolds, the alignment error and its theoretical bound asymptotically approach 0 when the filter size ρ increases in the noiseless setting $\nu = 0$. The monotonic decay of the error with filtering is expected since Theorem 3.6 predicts a variation of $\hat{E} = O\left(1 + (1+\rho^2)^{-1/2}\right)$ for the noiseless case. Meanwhile, the convergence of the error to 0 for the specific transformation model of translations can be explained as follows. In this special case, the variation of the second derivatives of the manifold with the filter size is given by $\|\partial_{ij}\, \hat{p}_\lambda\| = O\left((1+\rho^2)^{-3/2}\right)$, which follows from the fact that the second derivatives of the transformed coordinates in (3.6) vanish, i.e., $\partial_{ij}\, x', \partial_{ij}\, y' = 0$. This gives the rate of decrease of the alignment error with ρ as $\hat{E} = O\left((1+\rho^2)^{-1/2}\right)$ for translation manifolds in the noiseless case. Therefore, the alignment error approaches 0 as ρ increases. Meanwhile, Figures 4.2(c) and 4.2(d) obtained with three-dimensional manifolds generated by translations and rotations

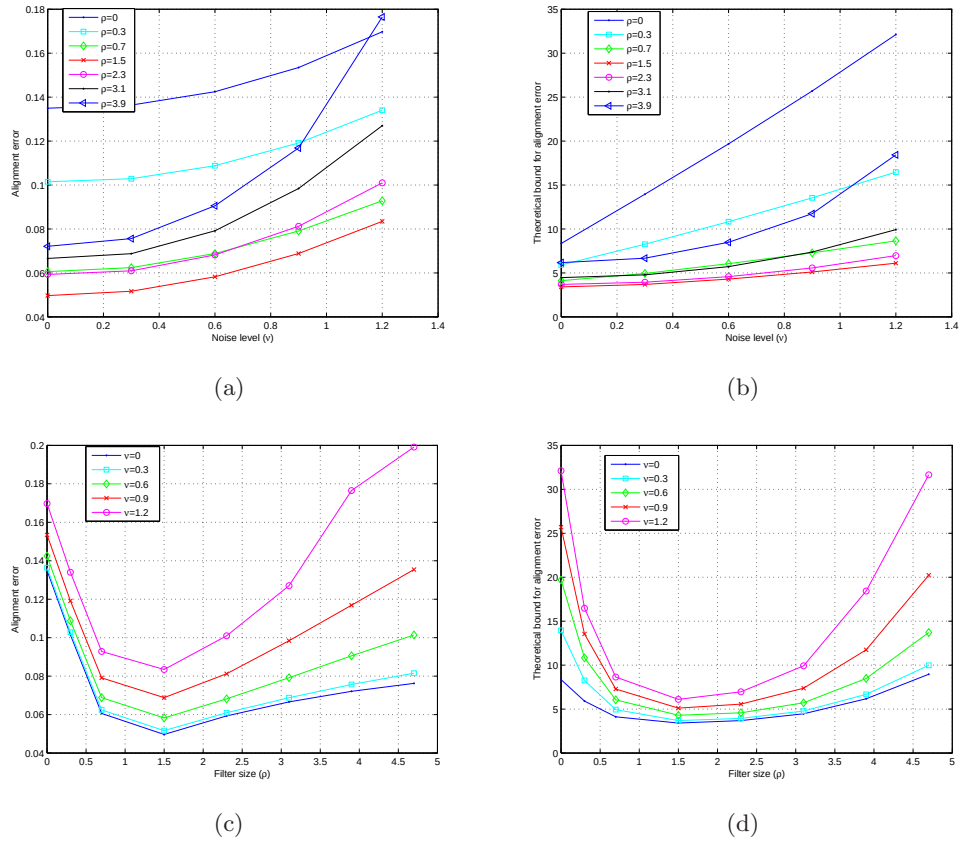


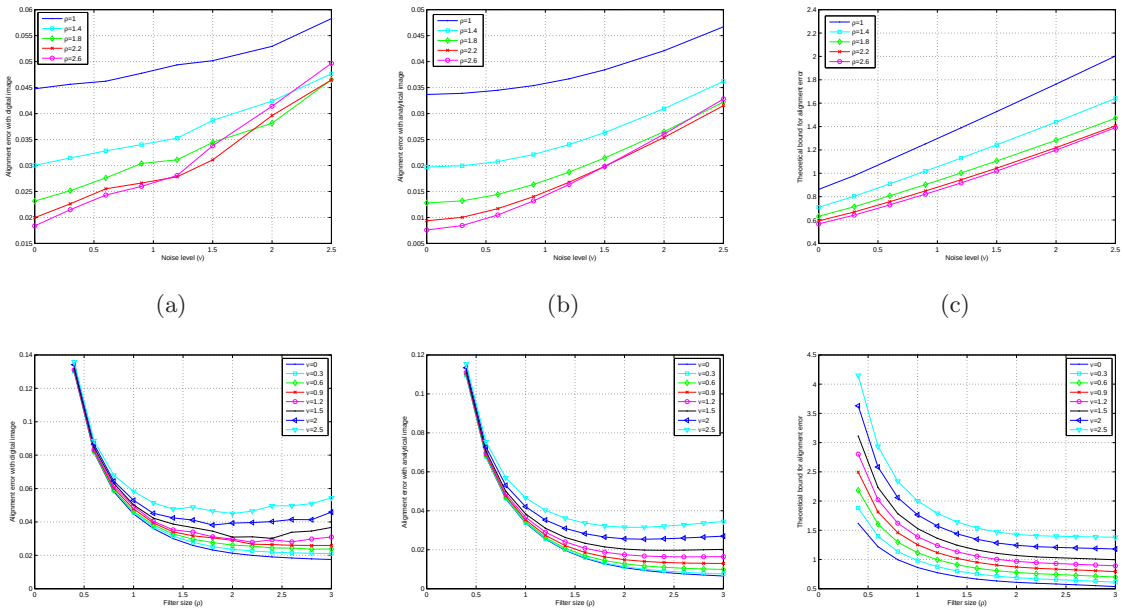
Figure 4.3. *Alignment errors of random patterns for 4-D manifolds generated by translations, rotations, and scale changes.*

show that the experimental and theoretical alignment errors approach a nonzero value in the noiseless case $\nu = 0$ as suggested by the prediction $\hat{E} = O(1 + \left(1 + \rho^2\right)^{-1/2})$.

Lastly, we comment on the plots in Figure 4.3 obtained for four-dimensional transformation manifolds generated by translations, rotations, and isotropic scale changes. One can observe in Figures 4.3(c) and 4.3(d) that both the experimental alignment error and its theoretical upper bound increase significantly with the filter size ρ in the noiseless case $\nu = 0$ when transformations include scale changes. This is due to the secondary source of noise demonstrated in Lemma 3.5. Theorem 3.6 suggests that the error increases with filtering at a rate $\hat{E} = O\left((\nu + 1)(1 + \rho^2)^{1/2}\right)$ at large values of ρ, which corresponds to a variation $\hat{E} = O\left((1 + \rho^2)^{1/2}\right)$ in the noiseless case.

We perform a second set of experiments on five real images, which are shown in Figure 4.4. The images are resized to the resolution of 60×60 pixels, and for each image an analytical approximation in the Gaussian dictionary $\mathcal{D}$ is computed with 100 atoms. The dictionary is defined over the parameter domain $\psi \in [-\pi, \pi)$; $\tau_x, \tau_y \in [-6, 6]$; $\sigma_x, \sigma_y \in [0.05, 3.5]$. Two reference patterns are considered for each image; namely, the digital image itself, and its analytical approximation in $\mathcal{D}$. For each one of the transformation models (4.1)-(4.3), 40 test

Figure 4.4. *Images used in the second set of experiments*

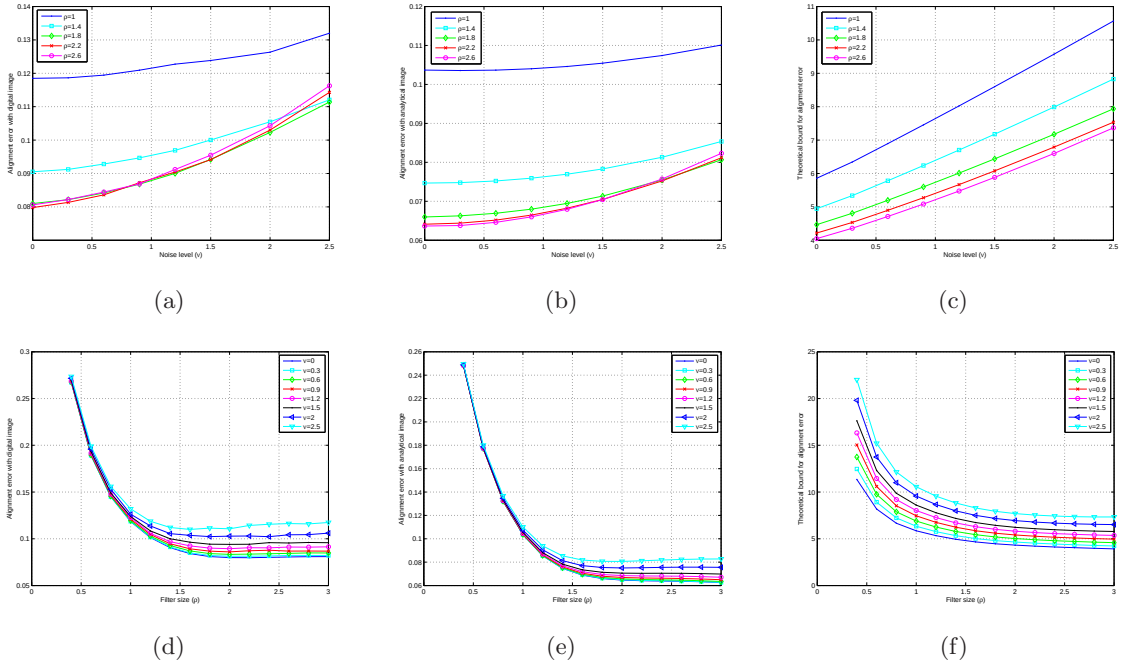


(a) (b) (c)

(d) (e) (f)

Figure 4.5. *Alignment errors of real images for 2-D manifolds generated by translations.*

patterns are generated for each reference pattern by applying a geometric transformation and adding a digital Gaussian noise image that is i.i.d. for each pixel. The geometric transformations are randomly selected from the transformation parameter domain $\overline{\theta} \in [-0.6, 0.6]$; $t_x, t_y \in [-0.6, 0.6]$; $\overline{s} \in [0.1, 2.1]$. The normalized rotation and scale parameters $\overline{\theta}$ and $\overline{s}$ correspond to the actual rotation angle and scale change factors $\theta \in [-0.07\pi, 0.07\pi]$ and $s \in [0.89, 1.13]$. The experimental alignment errors $\|\hat{\lambda}_e - \hat{\lambda}_o\|$ are computed by aligning the target patterns with the reference patterns, for both the original digital images and their approximations in the analytical dictionary $\mathcal{D}$. The theoretical upper bounds $\hat{E}$ are computed based on the analytical representations of the reference patterns. The alignment errors are plotted in Figures 4.5-4.7, which are averaged over all reference and target patterns. Figures 4.5, 4.6, and 4.7 show the errors obtained with the 2-D, 3-D and 4-D manifold models given respectively in (4.1), (4.2), and (4.3). In all figures, the alignment errors of the digital images, the alignment errors of the analytical approximations of images, and the theoretical upper bounds for the alignment error are plotted with respect to the noise level ν in panels (a)-(c), and with respect to the filter size ρ in panels (d)-(f).



(a) (b) (c)

(d) (e) (f)

Figure 4.6. *Alignment errors of real images for 3-D manifolds generated by translations and rotations.*

The results of the experiment show that the behavior of the alignment error for digital image representations is similar to the behavior of the error obtained with the analytical approximations of the images in $\mathcal{D}$. They mostly agree with the theoretical curves as well. The plots confirm that the increase in the alignment error with the noise level approaches an approximately linear rate at large values of the noise level as predicted by the theoretical results. The variation of the error with filtering is also in agreement with Theorem 3.6, and different transformation models lead to different behaviors for the alignment error as in the previous set of experiments. Meanwhile, it is observable that the dependence of the alignment error $\hat{E}$ on the filter size ρ in these experiments is mostly determined by its first component $\hat{E}_1$ related to manifold nonlinearity, even at large filter sizes. This is in contrast to the results obtained in the first setup with synthetically generated random patterns. The difference between the two setups can be explained as follows. Real images generally contain more high-frequency components than synthetical images generated in the smooth dictionary $\mathcal{D}$. These are captured with fine, small-scale atoms in the analytical approximations (the smallest atom scale used in this setup is 0.05, while it is 0.3 in the previous setup). The high-frequency components increase the manifold nonlinearity, which causes the error $\hat{E}_1$ to be the determining factor in the overall error. In return, the positive effect of filtering that reduces the alignment error is more prominent in these experiments, while the non-monotonic variation of the error with the filter size is still observable at large noise levels or for the transformation model (4.3) involving a scale change. The comparison of the two experimental setups shows that the exact variation of the error with filtering is influenced by the frequency characteristics of the reference patterns.

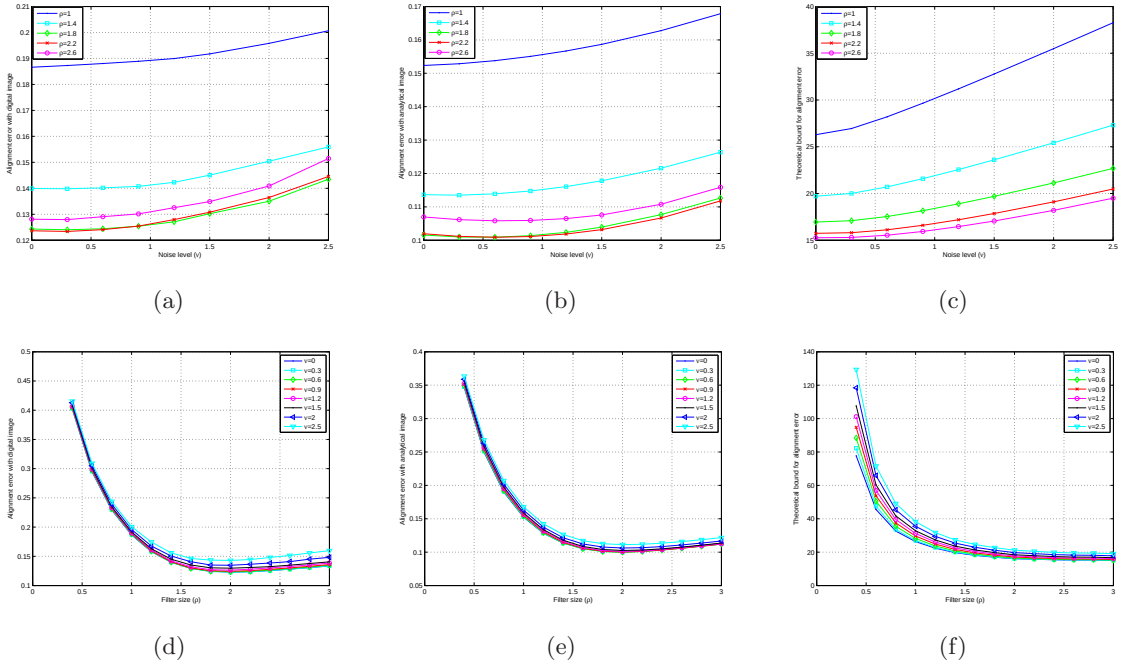


Figure 4.7. *Alignment errors of real images for 4-D manifolds generated by translations, rotations, and scale changes.*

The plots in panels (d)-(f) of the figures also show that, at small filter sizes, experimental errors are relatively high and very similar for different noise levels, while this is not the case in the theoretical plots. This suggests that numerical errors in the estimation of the tangent vectors with finite differences must have some influence on the overall error in practice, which is not taken into account in the theoretical bound. This error is higher for images with stronger high-frequency components and diminishes with smoothing (see the study in [4] for example). Lastly, one can observe that the alignment errors obtained with digital images are slightly larger than the alignment errors given by the analytic approximations of the images. This can be explained by the difference in the numerical computation of the tangent vectors in these two experimental settings. The analytic representation of the images in terms of parametric Gaussian atoms permits a more accurate computation of the tangent vectors, while the numerical interpolations employed in the computation of the tangents in the digital setting create an additional error source.

The overall conclusions of the experiments can be summarized as follows. The theoretical alignment error upper bound given in Theorem 3.1 gives a numerically pessimistic estimate of the alignment error as it is obtained with a worst-case analysis. However, it reflects well the actual dependence of the true alignment error both on the noise level and the filter size, and the results confirm the approximate variation rates given in Theorem 3.6. The theoretical upper bounds can be used in the determination of appropriate filter sizes in hierarchical image registration with tangent distance.

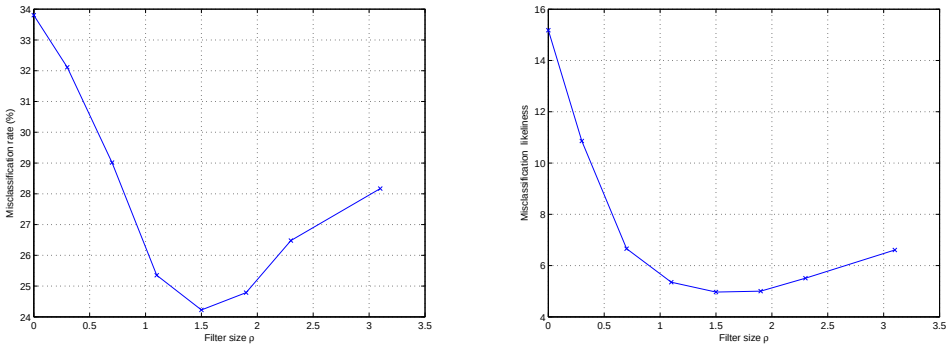
4.2. Image classification. We now study experimentally the image classification performance when manifold distances are computed with registration based on the tangent distance method.

In the first experiment, we classify a data set of synthetic images. We experiment on two classes of images. The reference pattern of each class consists of 20 randomly chosen Gaussian atoms such that 16 of the atoms are common between the two classes and 4 atoms are specific to each class. This configuration has the purpose of simulating a setting where the distinction between different classes stems from class-specific features, meanwhile different classes have some common features as well, which poses a challenge for classification. We then generate a set of test patterns that lie between the transformation manifolds of the two reference patterns. The test patterns are generated such that their true class labels are given by the class label of the closer manifold as in (3.38). We then classify the test patterns with the tangent distance method by estimating the transformation parameters in one step using the low-pass filtered versions of the reference and test patterns. The class labels of the test patterns are then estimated as in (3.42). We conduct the experiment on the transformation models in (4.1)-(4.3) and test the classification accuracy at different filter sizes. In Figures 4.8(a), 4.9(a) and 4.10(a), the percentage of misclassified test patterns is plotted with respect to the filter size, for these three transformation models respectively. Each plot is obtained by averaging the results of 400 repetitions of the experiment with randomly generated reference and test patterns. In order to interpret the variation of the experimental misclassification rate with the filter size in light of the results in Section 3.4, we define a function

$$\mathcal{T}_m \hat{\mathcal{K}}_m \ \lambda_{\min}^{-1} \ \left([\hat{\mathcal{G}}_{ij}^m(\lambda_r^m)]\right) \left(\frac{1}{2}\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}^m(\lambda_r^m)])}\ \|\hat{\lambda}_o - \lambda_r\|_1^2 + \sqrt{d}\ \|\tilde{n}_m\|\ \|\hat{\lambda}_o - \lambda_r\|_1\right) \tag{4.4}$$

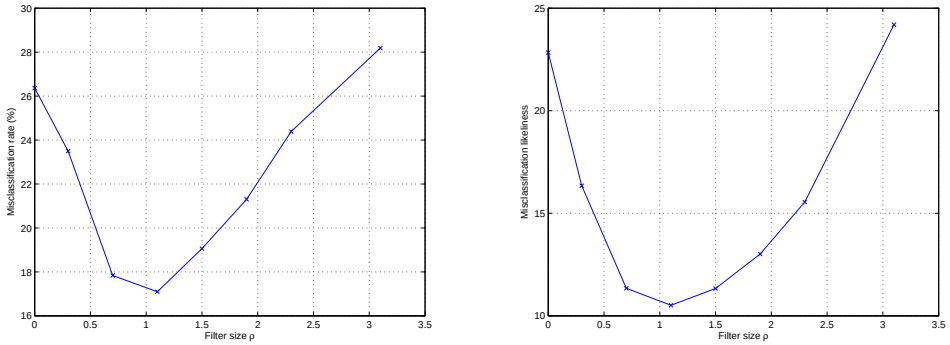
for the test patterns, where $\|\tilde{n}_m\|$ is the distance between the filtered test pattern $\hat{q}$ and the transformation manifold $\mathcal{M}(\hat{p}^m)$ of the filtered reference pattern representing class m. Comparing the function in (4.4) with the misclassification probability bound in (3.43), one can observe that they have the same variation with the filter size ρ, while it is easier to compute (4.4) experimentally. As it provides a measure for the misclassification probability, we call the expression in (4.4) the "misclassification likeliness" function. The average value of the misclassification likeliness (4.4) is plotted in Figures 4.8(b), 4.9(b) and 4.10(b), respectively for the transformation models in (4.1)-(4.3). Comparing panels (a) and (b) of Figures 4.8 - 4.10, we observe that the variation of the experimental misclassification probability with filtering agrees with that of the analytical misclassification likeliness (4.4). This shows that the misclassification probability upper bound in (3.43) captures well the behavior of the actual misclassification probability. Furthermore, as the misclassification likeliness is linearly proportional to the alignment error bound, we observe that the classification performance of the tangent distance method is indeed closely related to its alignment performance. The experimental results confirm that the misclassification probability has a non-monotonic variation with the filter size as predicted by the theoretical results of Section 3.4, and the optimal filter size minimizing the misclassification probability is in the vicinity of the filter size that minimizes the misclassification likeliness.

Next, we study the classification performance of the tangent distance method on a data set of handwritten digit images taken from the MNIST database [9]. We experiment on the



(a) (b)

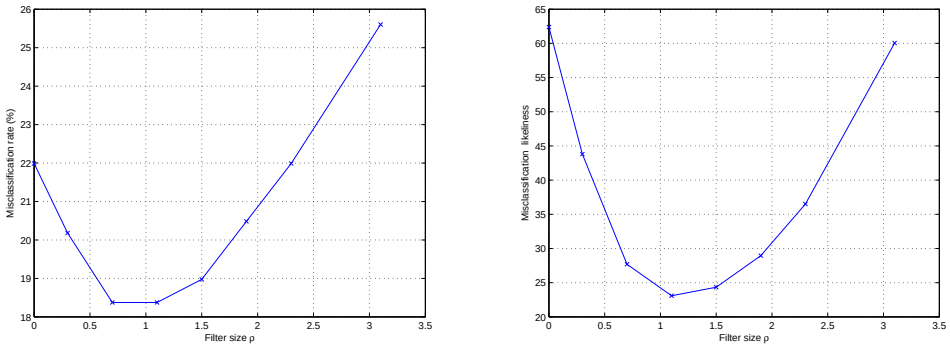
Figure 4.8. *Classification results for random patterns and 2-D manifolds generated by translations.*



(a) (b)

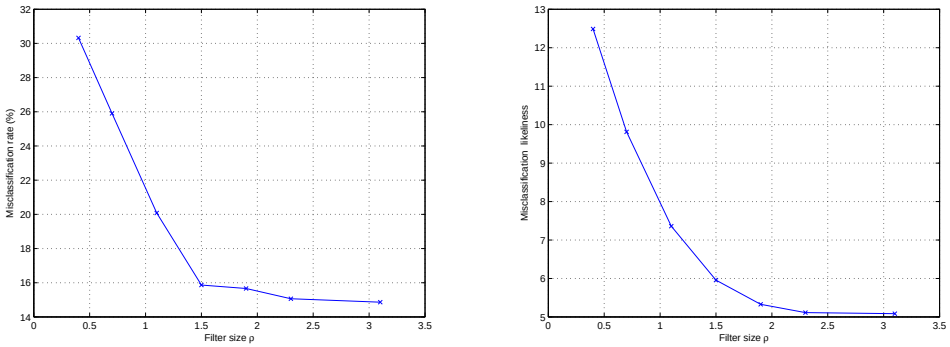
Figure 4.9. *Classification results for random patterns and 3-D manifolds generated by translations and rotations.*

images of the 2, 3, 5, 8, and 9 digits, each of which represents a different class. We randomly choose a reference image among the training samples of each class. The test images are formed by applying a random geometric transformation on randomly selected test samples in the database. We classify the test images by estimating their distance to the transformation manifolds of the reference images with the tangent distance method for different filter sizes as in (3.42). The results obtained for the geometric transformation models in (4.1)-(4.3) are presented respectively in Figures 4.11-4.13. Panels (a) and (b) of the figures show the experimental misclassification probability and the misclassification likeliness function (4.4), which are the average of 1000 repetitions of the experiment with different reference and test images. The behavior of the experimental misclassification probability as a function of the filter size is seen to be similar to that of the misclassification likeliness. Meanwhile, in contrast to the results obtained on synthetic smooth patterns (Figures 4.8-4.10), the best classification performances are obtained at large filter sizes for the digit images. This is in line with the results of the image alignment experiments with real images in Section 4.1, where the error



(a) (b)

Figure 4.10. *Classification results for random patterns and 4-D manifolds generated by translations, rotations, and scale changes.*



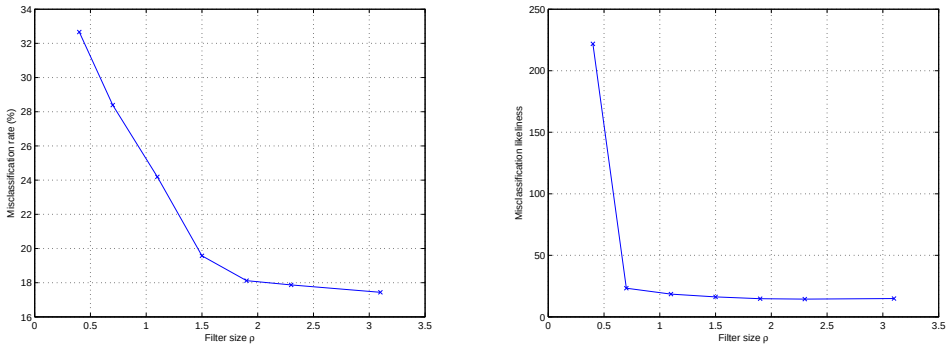
(a) (b)

Figure 4.11. *Classification results for digit images and 2-D manifolds generated by translations.*

resulting from manifold nonlinearity has been seen to be the determining factor in the overall behavior of the alignment error. Indeed, the high-frequency components may be prominent in real images. Since the digit images used in the experiments of Figures 4.11-4.13 also have quite nonlinear manifolds as a result of their frequency characteristics, their misclassification rate, as well as their alignment error, reaches its minimum value at large values of the filter size.

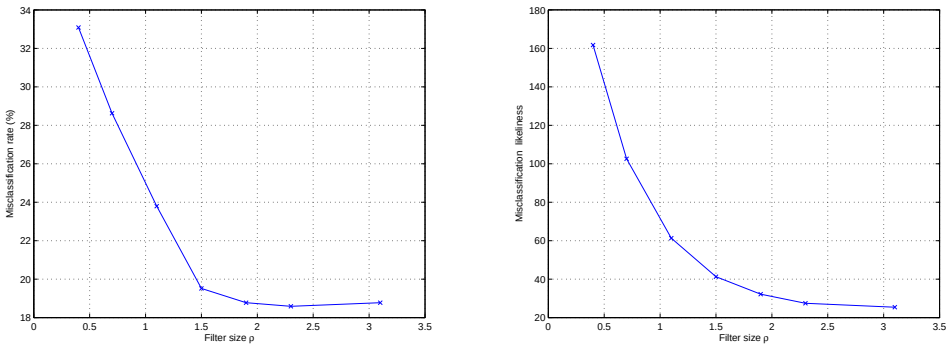
5. Discussion of Related Work. Although the tangent distance method is frequently used in image registration and image analysis applications, to the best of our knowledge, its performance has not been theoretically studied for general transformation models before. A brief overview of the related literature is as follows.

We begin with the works that analyze the dependence of the alignment error on noise. First, the study in [14] derives the Crámer-Rao lower bound (CRLB) for the registration of two images that differ by a 2-D translation. The CRLB gives a general lower bound for the



(a) (b)

Figure 4.12. *Classification results for digit images and 3-D manifolds generated by translations and rotations.*



(a) (b)

Figure 4.13. *Classification results for digit images and 4-D manifolds generated by translations, rotations, and scale changes.*

MSE of any estimator; therefore, the lower bounds derived in [14] are valid for all registration algorithms that aim to recover the translation between two images. A Gaussian noise model is assumed in [14], and the CRLB of a translation estimator is shown to be proportional to the noise variance. One can consider the noise standard deviation in the analysis in [14] to be proportional to our noise level parameter ν, which implies that the alignment error has a lower bound of $O(\nu)$. Then, the study in [23] explores the CRLB of registration for a variety of geometric transformation models and shows that the linear variation of the CRLB with the noise level derived in [14] for translations can be generalized to several other models such as rigid, shear and affine transformations. Being a generic bound valid for any estimator, the Crámer-Rao lower bound is also valid for the tangent distance method. In our main result Theorem 3.6, the second component $\hat{E}_2$ of the alignment error, which is related to image noise, increases at a rate of $O(\nu)$ with the noise level ν for any geometric transformation model. Therefore, the results in [14] and [23] are consistent with ours. Finally, let us remark the following about the variation of $\hat{E}_2$ with the filter size. The studies [14] and [23] show

that the CRLB of transformation estimators increases when the magnitudes of the spatial derivatives of patterns decrease. Since low-pass filtering reduces the magnitudes of spatial derivatives, it increases the MSE of estimators that compute the transformation parameters between an image pair. Similar results can be found in our previous work [20], where we show that the error due to noise in the estimation of 2-D translations with descent-type algorithms is amplified with filtering. Our main result, which indicates that the error component $\hat{E}_2$ associated with image noise increases with filtering, is in line with these previous works.

Next, the scope of the previous studies that examine the effect of manifold linearizations (e.g., [14], [4], [13]) is confined to the context of gradient-based optical flow estimation. Indeed, block-based optical flow estimation methods can be regarded as the restriction of the tangent distance method to estimate 2-D translations between image patches. Our study differs from these analyses in that it considers arbitrary transformation models while characterizing the influence of the image noise on the alignment performance in a multiscale setting (by including the effect of filtering in the analysis). We now briefly discuss some of these results in relation with our work.

The work [14] studies the bias on gradient-based estimators, which employ a first-order approximation of the image intensity function. The bias is the difference between the expectation of the translation parameter estimates and the true translation parameters, and it results from the first-order approximation of the image intensity function. It is therefore associated with the first error term $\hat{E}_1$ in Theorem 3.6 in our analysis. Note that the second error term $\hat{E}_2$ results from image noise and is related to the variance of the estimator when a zero-mean random noise model is assumed. It is shown in [14] that the bias is more severe if the image has larger bandwidth, i.e., if it has stronger high-frequency components. Hence, as smoothing the images with a low-pass filter reduces the image bandwidth, it decreases the bias. The studies in [8] and [4] furthermore report that smoothing diminishes the systematic error in the estimation of the image gradients from finite differences in optical flow computation, as it reduces the second and higher-order derivatives of the image intensity function. The results in [14] are consistent with our analysis, which shows that the component of the alignment error associated with manifold nonlinearity decreases with the filter size ρ. Our result is however valid not only for translations, but for other transformation models as well. Moreover, it provides an exact rate of decrease for the error, which is given by $O\left((1+\rho^2)^{-1/2}\right)$ for translations, and $O\left(1+(1+\rho^2)^{-1/2}\right)$ for other transformation models. The analysis in [14] reports that the bias due to series truncation has a polynomial dependence on the amount of translation. In the bound given in Theorem 3.1, the alignment error term E_1 associated with manifold nonlinearity is seen to be proportional to the square $\|\lambda_o - \lambda_r\|_1^2$ of the distance between the transformation parameters. This quadratic dependence is due to the fact that we have used a second-order approximation of the transformation manifold; a higher-order approximation clearly yields a polynomial dependence of higher-degree as obtained in [14].

Finally, the analysis in [10] studies the convergence of multiscale gradient-based registration methods where the image pair is related with a 2-D translation. It is shown that, for sufficiently small translations, coarse-to fine gradient-based registration algorithms converge to the globally optimal solution if the images are smoothed with ideal low-pass filters such that the filter bandwidth is doubled in each stage of the pyramid. However, this convergence guarantee is limited to an ideal noiseless setting where the target image is exactly a translated

version of the reference image, whereas the convergence guarantee derived in our study is valid for also noisy settings and arbitrary geometric transformation models.

6. Conclusion. We have presented a first complete performance analysis of the tangent distance method, which uses a first-order approximation of the transformation manifold in the estimation of the geometric transformation between a pair of images. We have first derived an upper bound for the alignment error and analyzed its variation with the noise level and the size of the low-pass filter used for smoothing the images in hierarchical registration algorithms. We have shown that the alignment error generally has a non-monotonic variation with the filter size due to the effects of smoothing on the image noise and the transformation manifold curvature. We have then used these results in order to establish some convergence guarantees for the hierarchical tangent distance algorithm. We have also derived some guidelines to choose the filter sizes optimally throughout the algorithm. Our results show that, in order to optimize the performance of the hierarchical alignment method, the initial filter size in the beginning of the algorithm should increase with the amount of transformation and decrease with the noise level. The optimal geometric decay factor of the filter size (which is usually taken as 1/2 in practice) then increases with the manifold curvature, the amount of transformation and the noise level. Finally, we have studied the classification performance of the tangent distance method and shown that the classification accuracy is expected to vary similarly to the alignment error. Our treatment is generic and valid for arbitrary geometric transformation models, and the theoretical results are confirmed by experiments. The presented study provides important insights for the understanding of multiscale registration methods that are based on manifold linearizations, and is helpful for optimizing the performance of such methods in image registration and image analysis applications.

REFERENCES

[1] *Online photo gallery "Copyright-free Paris"*. [Online]. Available: http://www.flickr.com/photos/54156444@N05/galleries/72157624889329591.

[2] J. Antoine, R. Murenzi, P. Vandergheynst, and S. Ali, *Two-Dimensional Wavelets and their Relatives*, Signal Processing, Cambridge University Press, 2004.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, *Performance of optical flow techniques*, International Journal of Computer Vision, 12 (1994), pp. 43–77.

[4] J. W. Brandt, *Analysis of bias in gradient-based optical flow estimation*, in 1994 Conf. Rec. of the 28th Asilomar Conf. on Signals, Systems and Computers, vol. 1, 1994, pp. 721–725.

[5] P. J. Burt and E. H. Adelson, *The laplacian pyramid as a compact image code*, IEEE Transactions on Communications, 31 (1983), pp. 532–540.

[6] J. Fabrizio, S. Dubuisson, and D. Béréziat, *Motion compensation based on tangent distance prediction for video compression*, Sig. Proc.: Image Comm., 27 (2012), pp. 153–171.

[7] A. W. Fitzgibbon and A. Zisserman, *Joint manifold distance: a new approach to appearance based clustering*, IEEE Conference on Computer Vision and Pattern Recognition, 1 (2003), p. 26.

[8] J. K. Kearney, W. B. Thompson, and D. L. Boley, *Optical flow estimation: An error analysis of gradient-based methods with local optimization*, IEEE Trans. Pattern Anal. Machine Intel., (1987).

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, 86 (1998), pp. 2278–2324.

[10] M. Lefébure and L.D. Cohen, *Image registration, optical flow and local rigidity*, Journal of Mathematical Imaging and Vision, 14 (2001), pp. 131–147.

[11] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, 1994.

[12] B. D. Lucas and T. Kanade, *An iterative image registration technique with an application to stereo vision*, in Proc. 7th Intl. Joint Conf. on Artificial Intelligence, 1981, pp. 674–679.

[13] T. Q. Pham, M. Bezuijen, L. J. van Vliet, K. Schutte, and C. L. Luengo, *Performance of optimal registration estimators*, in Proc. SPIE, 2005, pp. 133–144.

[14] D. Robinson and P. Milanfar, *Fundamental performance limits in image registration*, IEEE Trans. Img. Proc., 13 (2004), pp. 1185–1199.

[15] J. Rohn, *Bounds on eigenvalues of interval matrices. Zeitschrift für Angewandte Mathematik und Mechanik, Supplement 3, 78*, http://www.cs.cas.cz/rohn/publist/97.ps, 1998.

[16] P. Y. Simard, Y. LeCun, J. S. Denker, and B. Victorri, *Transformation invariance in pattern recognition – tangent distance and tangent propagation*, International Journal of Imaging Systems and Technology, 11 (2001).

[17] P. Y. Simard, B. Victorri, Y. LeCun, and J. S. Denker, *Tangent prop: a formalism for specifying selected invariances in adaptive networks*, in Advances in Neural Information Processing Systems (NIPS 1991), J. M. Moody, S. J. Hanson, and R. P. Lippman, eds., vol. 4, Denver, CO, 1992, Morgan Kaufman.

[18] G. Tziritas and C. Labit, *Motion Analysis for Image Sequence Coding*, Elsevier Science Inc., New York, NY, USA, 1994.

[19] N. Vasconcelos and A. Lippman, *A multiresolution manifold distance for invariant image similarity*, IEEE Transactions on Multimedia, 7 (2005), pp. 127–142.

[20] E. Vural and P. Frossard, *Analysis of Descent-Based Image Registration*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 2310–2349.

[21] ———, *Analysis of image registration with tangent distance*, EPFL-REPORT-195659, (2014). Available at: http://infoscience.epfl.ch/record/195659.

[22] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall/CRC, 1995.

[23] İ. Ş. Yetik and A. Nehorai, *Performance bounds on image registration*, IEEE Trans. Signal Proc., 54 (2006), pp. 1737 – 1749.