Analysis of Descent-Based Image Registration[*]

Elif Vural and Pascal Frossard [†]

Abstract. We present a performance analysis for image registration with gradient descent. We consider a typical multiscale registration setting where the global 2-D translation between a pair of images is estimated by smoothing the images and minimizing the distance between them with gradient descent. Our study particularly concentrates on the effect of noise and low-pass filtering on the alignment accuracy. We analyze the well-behavedness of the image distance function by estimating the neighborhood of translations for which it is free of undesired local minima. This is the neighborhood of translations that are correctly computable with a simple gradient descent minimization. We show that the area of this neighborhood increases at least quadratically with the smoothing filter size. We then examine the effect of noise on the alignment accuracy and derive an upper bound for the alignment error in terms of the noise properties and filter size. Our main finding is that the error increases at a rate that is at least linear with respect to the filter size. Therefore, smoothing improves the well-behavedness of the distance function; however, this comes at the cost of amplifying the alignment error in noisy settings. Our results provide a mathematical insight about why hierarchical techniques are effective in image registration, suggesting that the multiscale alignment strategy of these techniques is very suitable from the perspective of the trade-off between the well-behavedness of the objective function and the registration accuracy. To the best of our knowledge, this is the first such study for descent-based image registration.

Key words. Image registration, hierarchical registration methods, image smoothing, gradient descent, performance analysis.

1. Introduction. The estimation of the relative motion between two images is one of the important problems of image processing. The necessity for registering images arises in many different applications like image analysis and classification [23], [9], [25]; biomedical imaging [14], stereo vision [13], motion estimation for video coding [24]. The alignment of an image pair typically requires the optimization of a dissimilarity (or similarity) measure, whose common examples are sum-of-squared difference (SSD), approximations of SSD, and cross-correlation [7], [4]. Many registration techniques adopt, or can be coupled with, a multiscale hierarchical search strategy. In hierarchical registration, reference and target images are aligned by applying a coarse-to-fine estimation of the transformation parameters, using a pyramid of low-pass filtered and downsampled versions of the images. Coarse scales of the pyramid are used for a rough estimation of the transformation parameters. These scales have the advantage that the solution is less likely to get trapped into the local minima of the dissimilarity function as the images are smoothed by low-pass filtering. Moreover, the search complexity is lower at coarse scales as the image pair is downsampled accordingly. The alignment is then refined gradually by moving on to the finer scales. Since it offers a good compromise between complexity and accuracy, the coarse-to-fine alignment strategy has been widely used in many registration and motion estimation applications [25], [17], [5], [4], [2], [16].

In this work, we present a theoretical study that analyzes the effect of smoothing on the performance of image registration. One of our main goals is to understand better the mathematical principles behind multiscale registration techniques. Most theoretical results in the image registration literature (e.g., [21], [32], [31]) investigate how image noise affects the registration accuracy. However, the analysis of the effect of smoothing on the registration performance has generally been given little attention in the literature. Although it is widely known as a practical fact that smoothing an image pair is helpful for overcoming the undesired local minima of the dissimilarity function [1], [24], to the best of our knowledge, this has not been extensively studied on a mathematical basis yet. Some of the existing works examine how smoothing influences the bias on the registration with gradient-based methods [21], and the bias and model conditioning in optical flow [10], [6], whose scopes are however limited to methods employing a linear approximation of the image intensity function. Hence, the understanding of the exact relation between smoothing and the well-behavedness of the image dissimilarity function constitutes the first objective of this study. Our second objective is to characterize the effect of noise on the performance of multi-scale image registration, i.e., to derive the noise performance in the registration of noisy images as a function of the smoothing parameter.

We consider a setting where the geometric transformation between the reference and target images is a global 2-D translation. Although the registration problem is formulated for an image pair in this work, one can equivalently assume that the considered reference and target patterns are image patches rather than complete images. For this reason, our study is of interest not only for registration applications where the transformation between the image pair is modeled by a pure translation (e.g., as in satellite images), but also for various motion estimation techniques, such as block-matching algorithms and region-based matching techniques in optical flow that assign constant displacement vectors to image subregions. We adopt an analytic and parametric model for the reference and target patterns and formulate the registration problem in the continuous domain of square-integrable functions $L^2(\mathbb{R}^2)$. We use the squared-distance between the image intensity functions as the dissimilarity measure. This distance function is the continuous domain equivalent of SSD. We study two different aspects of image registration in this work; namely, alignment regularity and alignment accuracy.

We first look at *alignment regularity*; i.e., the well-behavedness of the distance function, and estimate the largest neighborhood of translations such that the distance function has only one local minimum, which is also the global minimum. Then we study the influence of smoothing the reference and target patterns on the neighborhood of translations recoverable with local minimizers such as descent-type algorithms without getting trapped in a local minimum. In more details, we consider the set of patterns that are generated by the translations of a reference pattern, which forms the translation manifold of that pattern. In the examination of the alignment regularity, we assume that the target pattern lies on the translation manifold of the reference pattern. We then consider the distance function $f(U)$ between the reference and target patterns, where U is the translation vector. The global minimum of f is at the origin $U = 0$. Then, in the translation parameter domain, we consider the largest open neighborhood around the origin within which f is an increasing function along any ray starting out from the origin. We call this neighborhood the Single Distance Extremum Neighborhood (SIDEN). The SIDEN of a reference pattern is important in the sense that it defines

the translations that can be correctly recovered by minimizing f with a descent method. We derive an analytic estimation of the SIDEN. Then, in order to study the effect of smoothing on the alignment regularity, we consider the registration of low-pass filtered versions of the reference and target patterns and examine how the SIDEN varies with the filter size. Our main result is that the volume (area) of the SIDEN increases at a rate of at least $O(1 + \rho^2)$ with respect to the size ρ of the low-pass filter kernel, which controls the level of smoothing. This formally shows that, when the patterns are low-pass filtered, a wider range of translation values can be recovered with descent-type methods; hence, smoothing improves the regularity of alignment. Then, we demonstrate the usage of our SIDEN estimate for constructing a regular multiresolution grid in the translation parameter domain with exact alignment guarantees. Based on our estimation of the neighborhood of translations that are recoverable with descent methods, we design an adaptive search grid in the translation parameter domain such that large translations can be recovered by locating the closest solution on the grid and then refining this estimation with a descent method.

Then we look at *alignment accuracy* and study the effect of image noise on the accuracy of image alignment. We also characterize the influence of low-pass filtering on the alignment accuracy in a noisy setting. This is an important question, as the target image is rarely an exactly translated version of the reference image in practical applications. When the target pattern is noisy, it is not exactly on the translation manifold of the reference pattern. The noise on the target pattern causes the global minimum of the distance function to deviate from the solution $U = 0$. We formulate the alignment error as the perturbation in the global minimum of the distance function, which corresponds to the misalignment between the image pair due to noise. We focus on two different noise models. In the first setting, we look at Gaussian noise. In the second setting, we examine arbitrary square-integrable noise patterns, where we consider general noise patterns and noise patterns that have small correlation with the points on the translation manifold of the reference pattern. We derive upper bounds on the alignment error in terms of the noise level and the pattern parameters in both settings. We then consider the smoothing of the reference and target patterns in these settings and look at the variation of the alignment error with the noise level and the filter size. It turns out that the alignment error bound increases at a rate of $O\left(\eta^{1/2} \left(1 - \eta\right)^{-1/2}\right)$ and $O\left(\nu^{1/2} \left(1 - \nu\right)^{-1/2}\right)$ in respectively the first and second settings with respect to the noise level, where η is the standard deviation of the Gaussian noise, and ν is the norm of the noise pattern. Another observation is that the alignment error is small if the noise pattern has small correlation with translated versions of the reference pattern. Moreover, the alignment error bounds increase at the rates $O\left(\rho^{3/2} \left(1 - \rho\right)^{-1/2}\right)$ and $O\left((1 + \rho^2)^{1/2}\right)$ in the first and second settings, with respect to the filter size ρ. Therefore, our main finding is that smoothing the image pair tends to increase the alignment error when the target pattern does not lie on the translation manifold of the reference pattern. The experimental results confirm that the behavior of the theoretical bound as a function of the noise level and filter size reflects well the behavior of the actual error.

The results of our analysis show that smoothing has the desirable effect of improving the well-behavedness of the distance function; however, it also leads to the amplification of the alignment error caused by the image noise. This suggests that, in the development of multiscale image registration methods, one needs to take the noise level into account as a

design parameter.

The rest of the text is organized as follows. In Section 2, we give an overview of related work. In Section 3, we focus on the alignment regularity problem, where we first derive an estimation of the SIDEN and then examine its variation with filtering. Then in Section 4, we look into the alignment accuracy problem and present our results regarding the influence of noise on the alignment accuracy. In Section 5, we present experimental results. In Section 6, we give a discussion of our results and interpret them in comparison with the previous studies in the literature. Finally, we conclude in Section 7.

2. Related Work. The problem of estimating the displacement between two images has been studied extensively in the image registration [33] and motion estimation [24] literatures. Here we limit our discussion mostly to region-based methods. We first give a brief overview of some hierarchical multiscale registration and motion estimation methods and then mention some theoretical results about image alignment.

The coarse-to-fine alignment strategy has been used in various types of image registration applications; e.g., registration for stereo vision [13], alignment with multiresolution tangent-distance for image analysis [25], biomedical image registration [17]. The hierarchical search strategy has proved useful in motion estimation, since it accelerates the algorithm and leads to better solutions with reduced sensitivity to local minima [24], [2], [5]. It is also used very commonly in gradient-based optical flow techniques such as those in [4], [16], which apply a first-order approximation of the variations in the image intensity function. The hierarchical search that filters and downsamples the images permits the design of gradient-based methods that remain in the domain of linearity.

Region-based registration and motion estimation methods use different dissimilarity measures and optimization techniques. Many methods use SSD (sum-of-squared difference) as the dissimilarity measure [24]. SSD corresponds to the squared-norm of what is usually called the displaced frame difference (DFD) in motion estimation. The direct correlation is also widely used as a similarity measure, and it can be shown to be equivalent to SSD [21].

In this work, we consider (the continuous domain equivalent of) SSD as the dissimilarity measure. We will essentially consider gradient descent as the minimization technique in our analysis; however, our main motivation is to understand to what extent local minimizers are efficient in image registration. Hence, the implications of our study concern a wide range of registration and motion estimation techniques that minimize SSD (or its approximations) with local minimizers, e.g., [18], [8], [28], [5], [25], and fast block-matching methods relying on convexity assumptions [24].

We now overview some theoretical studies about the performance of registration algorithms. The work by Robinson et al. [21] studies the estimation of global translation parameters between an image pair corrupted with additive Gaussian noise. The authors first derive the Cramér-Rao lower bound (CRLB) on the translation estimation. Given by the inverse of the Fisher information matrix, the CRLB is a general lower bound for the MSE of an estimator that computes a set of parameters from noisy observations. The authors then examine the bias on multiscale gradient-based methods. A detailed discussion of the results in [21] is given in Section 6 along with a comparison to our results. Another work that examines Cramér-Rao lower bounds in registration is given in [32], where the bounds are derived for several models

with transformations estimated from a set of matched feature points with noisy positions.

The studies in [10], [6] have also examined the bias of gradient-based shift estimators and shown that presmoothing the images reduces the bias on the estimator. However, smoothing also has the undesired effect of impairing the conditioning of the linear system to be solved in gradient-based estimators [10]. Therefore, this tradeoff must be taken into account in the selection of the filter size in coarse-to-fine gradient-based registration. The papers [21], [20] furthermore show that the bias on gradient-based estimators increases as the amount of translation increases. Robinson et al. [21] use this observation to explain the benefits of multiscale gradient-based methods. At large scales, downsampling, which reduces the amount of translation, and smoothing help to decrease the bias on the estimator. Then, as the change in the translation parameters is small at fine scales, the estimation does not suffer from this type of bias anymore. Moreover, at fine scales, the accuracy of the estimation increases as high-frequency components are no more suppressed. This is due to the fact that the CRLB of the estimation is smaller when the bandwidth of the image is larger.

Next, in the article [11] the convergence of gradient-based registration methods is examined. It is shown that, if the images are smoothed with ideal low-pass filters by doubling the bandwidth in each stage, coarse-to fine gradient-based registration algorithms converge to the globally optimal solution provided that the amount of shift is sufficiently small. However, this convergence guarantee is obtained for an ideal noiseless setting.

Lastly, the article [22] is a recent theoretical study on the accuracy of subpixel block-matching in stereo vision, which has relations to our work. The paper first examines the relation between the discrete and continuous block-matching distances, and then presents a continuous-domain analysis of the effect of noise on the accuracy of disparity estimation from a rectified stereo pair corrupted with additive Gaussian noise. An estimation of the disparity that globally minimizes the windowed squared-distance between blocks is derived. A comparison of the results presented in [22] and in our work is given in Section 6.

The previous works that have studied the alignment accuracy of multiscale registration by examining the effect of smoothing are limited to gradient-based methods, i.e., methods that employ a linear approximation of the image intensity. Moreover, none of these studies focus on the alignment regularity aspect of image registration. In this work, we address both of these issues and derive bounds on both alignment regularity and alignment accuracy in multiscale registration.

3. Analysis of Alignment Regularity.

3.1. Notation and Problem Formulation. Let $p \in L^2(\mathbb{R}^2)$ be a visual pattern with a non-trivial support on $\mathbb{R}^2$ (i.e., $p(X)$ is not equal to 0 almost everywhere on $\mathbb{R}^2$). In order to study the image registration problem analytically, we adopt a representation of p in an analytic and parametric dictionary manifold

$$\mathcal{D} = \{\phi_\gamma : \gamma = (\psi, \tau_x, \tau_y, \sigma_x, \sigma_y) \in \Gamma\} \subset L^2(\mathbb{R}^2). \tag{3.1}$$

Here, each atom ϕ_γ of the dictionary $\mathcal{D}$ is derived from an analytic mother function ϕ by a geometric transformation specified by the parameter vector γ, where 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. Γ is the transformation parameter domain over which the

dictionary is defined. By defining the spatial coordinate variable $X = [x\ y]^T \in \mathbb{R}^{2\times1}$, we refer to the mother function as $\phi(X)$. Then an atom ϕ_γ is given by

$$\phi_\gamma(X) = \phi(\sigma^{-1}\,\Psi^{-1}\,(X - \tau)), \tag{3.2}$$

where

$$\sigma = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}, \quad \Psi = \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix}, \quad \tau = \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix}. \tag{3.3}$$

It is shown in [3] (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 (3.1) is dense in $L^2(\mathbb{R}^2)$ if the mother function ϕ has nontrivial support (unless $\phi(X) = 0$ almost everywhere). In our analysis, we choose ϕ to be the Gaussian function $\phi(X) = e^{-X^T X} = e^{-(x^2+y^2)}$ as it has good time-localization and it is easy to treat in derivations due to its well-studied properties. This choice also 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 approximated in $\mathcal{D}$ with arbitrary accuracy. We assume that a sufficiently accurate approximation of p with finitely many atoms in $\mathcal{D}$ is available; i.e.,

$$p(X) \approx \sum_{k=1}^{K} c_k\,\phi_{\gamma_k}(X) \tag{3.4}$$

where K is the number of atoms used in the representation of p, γ_k are the atom parameters and c_k are the atom coefficients.[1]

Throughout the discussion, $T = [T_x\ T_y]^T \in S^1$ denotes a unit-norm vector and S^1 is the unit circle in $\mathbb{R}^2$. We use the notation tT for translation vectors, where $t \geq 0$ denotes the magnitude of the vector (amount of translation) and T defines the direction of translation. Then, the translation manifold $\mathcal{M}(p)$ of p is the set of patterns generated by translating p

$$\mathcal{M}(p) = \{p(X - t\,T) : T \in S^1,\, t \in [0, +\infty)\} \subset L^2(\mathbb{R}^2). \tag{3.5}$$

We consider the squared-distance between the reference pattern $p(X)$ and its translated version $p(X - tT)$. This distance is the continuous domain equivalent of the SSD measure that is widely used in registration methods. The squared-distance in the continuous domain is given by

$$f(tT) = \|p(\cdot) - p(\cdot - tT)\|^2 = \int_{\mathbb{R}^2} (p(X) - p(X - tT))^2 dX \tag{3.6}$$

where the notation[2] $\|.\|$ stands for the L^2-norm for vectors in $L^2(\mathbb{R}^2)$ and the ℓ^2-norm for vectors in $\mathbb{R}^2$.

---

[1]In the practical computation of the representation of a digital image in the Gaussian dictionary, the number of atoms should be chosen such that a substantial part of the energy of the image is captured in the approximation. Atom parameters and coefficients can be computed in various ways, for instance, by using pursuit algorithms such as [15], [19] with a redundant sampling of the Gaussian dictionary, or with DC (difference-of-convex) optimization as in [27].

[2]The notations $\|p\|$ or $\|p(\cdot)\|$ are always used to refer to the L^2-norm of p, considered as an element of the vector space of functions. Since it is then clear whether the $L^2(\mathbb{R}^2)$-norm or the ℓ^2-norm is meant, we denote these in the same way for simplicity of notation.

Note that in image registration, windowed versions of the image pair centered around a point of interest can be used as well as the entire images [33]. If the reference and target images are windowed around a point X_0, the distance function becomes

$$\int_{\mathbb{R}^2} \left(\varphi(X - X_0)p(X) - \varphi(X - X_0)p(X - tT) \right)^2 dX$$

where $\varphi : \mathbb{R}^2 \to [0, 1]$ is a window function. If the window is chosen sufficiently large such that it covers the region where the reference pattern p has significant intensity as well as its translated versions $p(X - tT)$ for realistic values of tT, one can approximate this function with the function $f(tT)$ in (3.6). In this section, we base our analysis on the non-windowed distance function $f(tT)$ in order to keep the derivations simple.

The global minimum of f is at the origin $tT = 0$. Therefore, there exists a region around the origin within which the restriction of f to a ray tT_a starting out from the origin along an arbitrary direction T_a is an increasing function of $t > 0$ for all T_a. This allows us to define the Single Distance Extremum Neighborhood (SIDEN) as follows.

Definition 3.1. *We call the set of translation vectors*

$$\mathcal{S} = \{0\} \cup \{\omega_T T : T \in S^1, \omega_T > 0, \ and \ \frac{df(tT)}{dt} > 0 \ \ for \ all \ 0 < t \leq \omega_T\} \tag{3.7}$$

the Single Distance Extremum Neighborhood (SIDEN) of the pattern p.

Note that the origin $\{0\}$ is included separately in the definition of SIDEN since the gradient of f vanishes at the origin and therefore $df(tT)/dt|_{t=0} = 0$ for all T. The SIDEN $\mathcal{S} \subset \mathbb{R}^2$ is an open neighborhood of the origin such that the only stationary point of f inside $\mathcal{S}$ is the origin. We formulate this in the following proposition.

Proposition 3.2. *Let $tT \in \mathcal{S}$. Then $\nabla f(tT) = 0$ if and only if $tT = 0$.*

Proof. Let $\nabla f(tT) = 0$ for some $tT \in \mathcal{S}$. Then, $\nabla_T f(tT) = 0$, which is the directional derivative of f along the direction T at tT. This gives

$$\nabla_T f(tT) = \frac{d}{du} f(tT + uT)\bigg|_{u=0} = \frac{d}{du} f\left((t + u)T\right)\bigg|_{u=0} = \frac{d}{du} f(uT)\bigg|_{u=t} = \frac{df(tT)}{dt} = 0$$

which implies that $t = 0$, as $tT \in \mathcal{S}$. The second part $\nabla f(0) = 0$ of the statement also holds clearly, since the global minimum of f is at 0. ∎

Proposition 3.2 can be interpreted as follows. The only local minimum of the distance function f is at the origin in $\mathcal{S}$. Therefore, when a translated version $p(X - tT)$ of the reference pattern is aligned with $p(X)$ with a local optimization method like a gradient descent algorithm, the local minimum achieved in $\mathcal{S}$ is necessarily also the global minimum.

The goal of our analysis is now the following. Given a reference pattern p, we would like to find an analytical estimation of $\mathcal{S}$. However, the exact derivation of $\mathcal{S}$ requires the calculation of the exact zero-crossings of $df(tT)/dt$, which is not easy to do analytically. Instead, one can characterize the SIDEN by computing a neighborhood $\mathcal{Q}$ of 0 that lies completely in $\mathcal{S}$; i.e., $\mathcal{Q} \subset \mathcal{S}$. $\mathcal{Q}$ can be derived by using a polynomial approximation of f and calculating, for all unit directions T, a lower bound δ_T for the supremum of ω_T such that $\omega_T T$ is in $\mathcal{S}$. This does not only provide an analytic estimation of the SIDEN, but also defines a set that is known to be completely inside the SIDEN. The regions $\mathcal{S}$ and $\mathcal{Q}$ are illustrated in Figure 3.1.
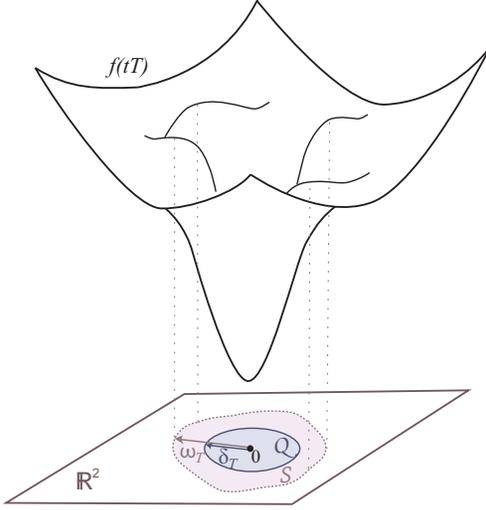
Figure 3.1. *SIDEN $\mathcal{S}$ is the largest open neighborhood around the origin within which the distance f is increasing along all rays starting out from the origin. Along each unit direction T, $\mathcal{S}$ covers points $\omega_T T$ such that $f(tT)$ is increasing between 0 and $\omega_T T$. The estimate $\mathcal{Q}$ of $\mathcal{S}$ is obtained by computing a lower bound δ_T for the first zero-crossing of $df(tT)/dt$.*

In Section 3.2 we derive $\mathcal{Q}$. In particular, $\mathcal{Q}$ is obtained in the form of a compact analytic set and f is a differentiable function. This guarantees that, if the translation that aligns the image pair perfectly is in the set $\mathcal{Q}$, the distance function f can be minimized with gradient descent algorithms; the solution converges to a local minimum of f in $\mathcal{Q}$, which is necessarily the global minimum of f, resulting in a perfect alignment. Moreover, we will see in Section 5 that the knowledge of a set $\mathcal{Q} \subset \mathcal{S}$ permits us to design a registration algorithm that can recover large translations perfectly.

Finally, as $\mathcal{Q}$ is obtained analytically and parametrically, it is simple to examine its variation with the low-pass filtering applied to p. This is helpful for gaining an understanding of the relation between the alignment regularity and smoothing. We study this relation in Section 3.3.

3.2. Estimation of SIDEN. We now derive an estimation $\mathcal{Q}$ for the Single Distance Extremum Neighborhood $\mathcal{S}$. In the following, we consider T to be a fixed unit direction in S^1. We derive $\mathcal{Q} \subset \mathcal{S}$ by computing a δ_T which guarantees that $df(tT)/dt > 0$ for all $0 < t \leq \delta_T$. In the derivation of $\mathcal{Q}$, we need a closed-form expression for $df(tT)/dt$. Since f is the distance between the patterns $p(X)$ and $p(X - tT)$ that are represented in terms of Gaussian atoms (see Eq. 3.4), it involves the integration of the multiplication of pairs of Gaussian atoms. We will use the following proposition about the integration of the product of Gaussian atoms [29].

Proposition 3.3. *Let $\phi_{\gamma_j}(X) = \phi(\sigma_j^{-1} \Psi_j^{-1} (X - \tau_j))$ and $\phi_{\gamma_k}(X) = \phi(\sigma_k^{-1} \Psi_k^{-1} (X - \tau_k))$. Then*

$$\int_{\mathbb{R}^2} \phi_{\gamma_j}(X)\phi_{\gamma_k}(X)dX = \frac{\pi |\sigma_j \sigma_k|}{2\sqrt{|\Sigma_{jk}|}} \exp\left(-\frac{1}{2}(\tau_k - \tau_j)^T \Sigma_{jk}^{-1} (\tau_k - \tau_j)\right)$$

where $\Sigma_{jk} := \frac{1}{2}\left(\Psi_j \sigma_j^2 \Psi_j^{-1} + \Psi_k \sigma_k^2 \Psi_k^{-1}\right)$.

The symbol Σ_{jk} defined in Proposition 3.3 is a function of the parameters of the j-th and k-th atoms. We also denote

$$a_{jk} := \frac{1}{2} T^T \Sigma_{jk}^{-1} T, \qquad b_{jk} := \frac{1}{2} T^T \Sigma_{jk}^{-1} (\tau_k - \tau_j)$$

$$c_{jk} := \frac{1}{2} (\tau_k - \tau_j)^T \Sigma_{jk}^{-1} (\tau_k - \tau_j), \qquad Q_{jk} := \frac{\pi |\sigma_j \sigma_k| e^{-c_{jk}}}{\sqrt{|\Sigma_{jk}|}}. \tag{3.8}$$

Notice that $a_{jk} > 0$ and $c_{jk} \geq 0$ since $\|T\| = 1$ and Σ_{jk}, Σ_{jk}^{-1} are positive definite matrices. By definition, $Q_{jk} > 0$ as well. Note also that a_{jk} and b_{jk} are functions of the unit direction

T; however, for the sake of simplicity we avoid expressing their dependence on T explicitly in our notation.

We can now give our result about the estimation of the SIDEN.

Theorem 3.4. *The region $\mathcal{Q} \subset \mathbb{R}^2$ is a subset of the SIDEN $\mathcal{S}$ of the pattern p if*

$$\mathcal{Q} = \{tT : T \in S^1, \ 0 \le t \le \delta_T\}$$

where δ_T is the only positive root of the polynomial $|\alpha_4|t^3 - \alpha_3 t^2 - \alpha_1$ and

$$\alpha_1 = \sum_{j=1}^{K}\sum_{k=1}^{K} c_j c_k \, Q_{jk} \left(2\, a_{jk} - 4\, b_{jk}^2\right)$$

$$\alpha_3 = \sum_{j=1}^{K}\sum_{k=1}^{K} c_j c_k \, Q_{jk} \left(-\frac{8}{3} b_{jk}^4 + 8\, b_{jk}^2\, a_{jk} - 2\, a_{jk}^2\right)$$

$$\alpha_4 = -1.37 \sum_{j=1}^{K}\sum_{k=1}^{K} |c_j c_k| \, Q_{jk} \, \exp\left(\frac{b_{jk}^2}{a_{jk}}\right) a_{jk}^{5/2}$$

are constants depending on T and on the parameters γ_k of the atoms in p.

The proof of Theorem 3.4 is given in [26, Appendix A.1]. The proof applies a Taylor expansion of $df(tT)/dt$ and derives δ_T such that $df(tT)/dt$ is positive for all $t \le \delta_T$. Therefore, along each direction T, δ_T constitutes a lower bound for the first zero-crossing of $df(tT)/dt$ (see Figure 3.1 for an illustration of δ_T). By varying T over the unit circle, one obtains a closed neighborhood $\mathcal{Q}$ of 0 that is a subset of $\mathcal{S}$. This region can be analytically computed using only the parametric representation of p and provides an estimate for the range of translations tT over which $p(X)$ can be exactly aligned with $p(X - tT)$.

3.3. Variation of SIDEN with Smoothing. We now examine how smoothing the reference pattern p with a low-pass filter influences its SIDEN. We assume a Gaussian kernel for the filter. The Gaussian function is a commonly used kernel for low-pass filtering and its distinctive properties has made it popular in scale-space theory research [12] (see Section 6 for a more detailed discussion). We assume that p is filtered with a Gaussian kernel of the form $\frac{1}{\pi\rho^2}\phi_\rho(X)$ with unit L^1-norm. The function $\phi_\rho(X) = \phi(\Upsilon^{-1}(X))$ is an isotropic Gaussian atom with scale matrix

$$\Upsilon = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix}. \tag{3.9}$$

The scale parameter ρ controls the size of the Gaussian kernel. We denote the smoothed version of the reference pattern $p(X)$ by $\hat{p}(X)$, which is given as

$$\hat{p}(X) = \frac{1}{\pi\rho^2} \left(\phi_\rho * p\right)(X) = \sum_{k=1}^{K} c_k \, \frac{1}{\pi\rho^2} \left(\phi_\rho * \phi_{\gamma_k}\right)(X) \tag{3.10}$$

by linearity of the convolution operator. In order to calculate $\hat{p}$, we use the following proposition, which gives the expression for the Gaussian atom obtained from the convolution of two Gaussian atoms [29].

Proposition 3.5. 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

$$(\phi_{\gamma_1} * \phi_{\gamma_2})(X) = \frac{\pi|\sigma_1\sigma_2|}{|\sigma_3|}\phi_{\gamma_3}(X) \tag{3.11}$$

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

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

From Proposition 3.5, we obtain

$$\frac{1}{\pi\rho^2}\,(\phi_\rho * \phi_{\gamma_k})(X) = \frac{|\sigma_k|}{|\hat{\sigma}_k|}\phi_{\hat{\gamma}_k}(X) \tag{3.12}$$

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

$$\hat{\tau}_k = \tau_k, \quad \hat{\Psi}_k = \Psi_k, \quad \hat{\sigma}_k = \sqrt{\Upsilon^2 + \sigma_k^2}. \tag{3.13}$$

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

$$\hat{c}_k = \frac{|\sigma_k|}{|\hat{\sigma}_k|}c_k = \frac{|\sigma_k|}{\sqrt{|\Upsilon^2 + \sigma_k^2|}}c_k = \frac{\sigma_{x,k}\,\sigma_{y,k}}{\sqrt{(\rho^2 + \sigma_{x,k}^2)(\rho^2 + \sigma_{y,k}^2)}}c_k \tag{3.14}$$

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

$$\hat{p}(X) = \sum_{k=1}^{K}\hat{c}_k\phi_{\hat{\gamma}_k}(X). \tag{3.15}$$

Considering the same setting as in Section 3.1, where the target pattern $p(X - tT)$ is exactly a translated version of the reference pattern $p(X)$, we now assume that both the reference and target patterns are low-pass filtered as it is typically done in hierarchical image registration algorithms. The above equations show that, when a pattern is low-pass filtered, the scale parameters of its atoms increase and the atom coefficients decrease proportionally to the filter kernel size, leading to a spatial diffusion of the image intensity function. The goal of this section is to show that this diffusion increases the volume of the SIDEN. We achieve this by analyzing the variation of the smoothed SIDEN estimate $\hat{\mathcal{Q}}$ corresponding to the smoothed distance

$$\hat{f}(tT) = \int_{\mathbb{R}^2}\left(\hat{p}(X) - \hat{p}(X - tT)\right)^2 dX \tag{3.16}$$

with respect to the filter size ρ. Since the smoothed pattern has the same parametric form (3.15) as the original pattern, the variation of $\hat{\mathcal{Q}}$ with ρ can be analyzed easily by examining

the dependence of the parameters involved in the derivation of $\hat{\mathcal{Q}}$ on ρ. In the following, we express the terms in Section 3.2 that have a dependence on ρ with the notation $\hat{(\cdot)}$, such as $\hat{a}_{jk}$, $\hat{b}_{jk}$, $\hat{c}_k$, $\hat{\sigma}_k$. We write the terms that do not depend on ρ in the same way as before; e.g., t, T, τ_k, Ψ_k.

Now, we can apply the result of Theorem 3.4 for the smoothed pattern $\hat{p}(X)$. For a given kernel size ρ, the smoothed versions $\hat{a}_{jk}$, $\hat{b}_{jk}$, $\hat{c}_{jk}$, $\hat{Q}_{jk}$ of the parameters in (3.8) can be obtained by replacing the scale parameters σ_k with $\hat{\sigma}_k$ defined in (3.13). Then, the smoothed SIDEN corresponding to ρ is given as $\hat{\mathcal{Q}} = \{tT : T \in S^1,\ 0 \le t \le \hat{\delta}_T\}$ where $\hat{\delta}_T$ is the positive root of the polynomial $|\hat{\alpha}_4|t^3 - \hat{\alpha}_3 t^2 - \hat{\alpha}_1$ such that

$$\hat{\alpha}_1 = \sum_{j=1}^{K}\sum_{k=1}^{K} \hat{c}_j \hat{c}_k \, \hat{Q}_{jk} \left(2\,\hat{a}_{jk} - 4\,\hat{b}_{jk}^2\right), \qquad \hat{\alpha}_3 = \sum_{j=1}^{K}\sum_{k=1}^{K} \hat{c}_j \hat{c}_k \, \hat{Q}_{jk} \left(-\frac{8}{3}\hat{b}_{jk}^4 + 8\,\hat{b}_{jk}^2\,\hat{a}_{jk} - 2\,\hat{a}_{jk}^2\right)$$

$$\hat{\alpha}_4 = -1.37 \sum_{j=1}^{K}\sum_{k=1}^{K} |\hat{c}_j \hat{c}_k| \, \hat{Q}_{jk} \, \exp\left(\frac{\hat{b}_{jk}^2}{\hat{a}_{jk}}\right) \hat{a}_{jk}^{5/2}.$$

Similarly to the derivation in Section 3.2, the terms $\hat{a}_{jk}$, $\hat{b}_{jk}$, $\hat{c}_{jk}$, $\hat{Q}_{jk}$ are associated with the integration of the product of smoothed Gaussian atom pairs, and they appear in the closed-form expression of $d\hat{f}(tT)/dt$.

We are now ready to give the following result, which summarizes the dependence of the smoothed SIDEN estimate on the filter size ρ.

Theorem 3.6. *Let $V(\hat{\mathcal{Q}})$ denote the volume (area) of the SIDEN estimate $\hat{\mathcal{Q}}$ for the smoothed pattern $\hat{p}$. Then, the order of dependence of the volume of $\hat{\mathcal{Q}}$ on the filter size ρ is given by $V(\hat{\mathcal{Q}}) = O(1 + \rho^2)$. Moreover, the distance $\hat{\delta}_T$ of the boundary of $\hat{\mathcal{Q}}$ to the origin increases at a rate of $O(\sqrt{1 + \rho^2})$ with ρ along any direction T.*

Theorem 3.6 is proved in [26, Appendix A.2]. The proof is based on the examination of the order of variation of $\hat{a}_{jk}$, $\hat{b}_{jk}$, $\hat{c}_{jk}$, $\hat{Q}_{jk}$ with ρ, which is then used to derive the dependence of $\hat{\delta}_T$ on ρ.

Theorem 3.6 is the main result of this section. It states that the volume of the SIDEN estimate increases with the size of the filter applied on the patterns to be aligned. The theorem shows that the volume of the region of translations for which the reference pattern $\hat{p}(X)$ can be perfectly aligned with $\hat{p}(X - tT)$ using a descent method, expands at the rate $O(1 + \rho^2)$ with respect to the increase in the filter size ρ. Here, the order of variation $O(1 + \rho^2)$ is obtained for the estimate $\hat{\mathcal{Q}}$ of the SIDEN. Hence, one may wonder if the volume $V(\hat{\mathcal{S}})$ of the SIDEN $\hat{\mathcal{S}}$ has the same dependence on ρ. Remembering that $\hat{\mathcal{Q}} \subset \hat{\mathcal{S}}$ for all ρ, one immediate observation is that the rate of expansion of $\hat{\mathcal{S}}$ must be at least $O(1 + \rho^2)$; otherwise, there would exist a sufficiently large value of ρ such that $\hat{\mathcal{Q}}$ is not included in $\hat{\mathcal{S}}$. One can therefore conclude that $V(\hat{\mathcal{S}}) \ge V(\hat{\mathcal{Q}}) = O(1 + \rho^2)$. Theorem 3.6 also states that the SIDEN boundary along any direction T expands at a rate of at least $O(\sqrt{1 + \rho^2})$. However, this only gives a lower bound for the rate of expansion of $\hat{\mathcal{S}}$ and the exact rate of expansion of $\hat{\mathcal{S}}$ may be larger. In the following, we give a few comments about the variation of $\hat{\mathcal{S}}$ with ρ.

Remark. As shown in the proof of Theorem 3.4, the derivative of the distance function $f(tT)$ is of the form

$$\frac{df(tT)}{dt} = \sum_{j=1}^{K}\sum_{k=1}^{K} c_j c_k \ Q_{jk} \ s_{jk}(t) \tag{3.17}$$

where

$$s_{jk}(t) = e^{-\left(a_{jk}\, t^2 + 2\, b_{jk}\, t\right)} \left(a_{jk}\, t \ + \ b_{jk}\right) \ + \ e^{-\left(a_{jk}\, t^2 - 2\, b_{jk}\, t\right)} \left(a_{jk}\, t \ - \ b_{jk}\right). \tag{3.18}$$

In order to derive $\mathcal{S}$, one needs to exactly locate the smallest zero-crossing of $\frac{df(tT)}{dt}$. This is not easy to do analytically due to the complicated form of the functions $s_{jk}(t)$, which we have handled with polynomial approximations in the derivation of $\mathcal{Q}$. However, in order to gain an intuition about how the zero-crossings change with filtering, one can look at the dependence of the extrema of the two additive terms in $s_{jk}(t)$ on ρ. The function $e^{-\left(a_{jk}\, t^2 + 2\, b_{jk}\, t\right)} \left(a_{jk}\, t \ + \ b_{jk}\right)$ has two extrema at

$$\mu_0 = \frac{1}{a_{jk}} \left(-\sqrt{\frac{a_{jk}}{2}} - b_{jk}\right), \qquad \mu_1 = \frac{1}{a_{jk}} \left(\sqrt{\frac{a_{jk}}{2}} - b_{jk}\right) \tag{3.19}$$

and $e^{-\left(a_{jk}\, t^2 - 2\, b_{jk}\, t\right)} \left(a_{jk}\, t \ - \ b_{jk}\right)$ has two extrema at

$$\mu_2 = \frac{1}{a_{jk}} \left(-\sqrt{\frac{a_{jk}}{2}} + b_{jk}\right), \qquad \mu_3 = \frac{1}{a_{jk}} \left(\sqrt{\frac{a_{jk}}{2}} + b_{jk}\right). \tag{3.20}$$

Now replacing the original parameters a_{jk}, b_{jk} with their smoothed versions $\hat{a}_{jk}$, $\hat{b}_{jk}$ and using the result from the proof of Theorem 3.6 that $\hat{a}_{jk}$ and $\hat{b}_{jk}$ decrease at a rate of $O\left((1+\rho^2)^{-1}\right)$, it is easy to show that the locations of the extrema $\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ change with a rate of $O\left((1+\rho^2)^{1/2}\right)$. One may thus conjecture that the zero-crossings of $df(tT)/dt$ along a fixed direction T might also move at the same rate, which gives the volume of $\hat{\mathcal{S}}$ as $V(\hat{\mathcal{S}}) = O(1+\rho^2)$.

On the other hand, $V(\hat{\mathcal{S}})$ may also exhibit a different type of variation with ρ depending on the atom parameters of p. In particular, $V(\hat{\mathcal{S}})$ may expand at a rate greater than $O(1+\rho^2)$ for some patterns. For example, as shown in [26, Proposition 4], there exists a threshold value ρ_0 of the filter size such that for all $\rho > \rho_0$, $\hat{\mathcal{S}} = \mathbb{R}^2$ and thus $V(\hat{\mathcal{S}}) = \infty$ for patterns that consist of atoms with coefficients of the same sign. In addition, patterns whose atoms with positive (or negative) coefficients are dominant over the atoms with the opposite sign are likely to have this property due to their resemblance to patterns consisting of atoms with coefficients of the same sign.

Theorem 3.6 describes the effect of smoothing images before alignment. One may then wonder what the optimal filter size to be applied to the patterns before alignment is, given a reference and a target pattern. Theorem 3.6 suggests that, if the target pattern is on the translation manifold of the reference pattern, applying a large filter is always preferable as it provides a large range of translations recoverable by descent algorithms. The accuracy of alignment does not change with the filter size in this noiseless setting, since a perfect alignment

is always guaranteed with descent methods as long as the amount of translation is inside the SIDEN. However, the assumption that the target pattern is exactly of the form $p(X - tT)$ is not realistic in practice; i.e., in real image processing applications, the target image is likely to deviate from $\mathcal{M}(p)$ due to the noise caused by image capture conditions, imaging model characteristics, etc. Hence, we examine in Section 4 if filtering affects the accuracy of alignment when the target image deviates from $\mathcal{M}(p)$.

4. Analysis of Alignment Accuracy in Noisy Settings. We now analyze the effect of noise and smoothing on the accuracy of the estimation of translation parameters. In general, noise causes a perturbation in the location of the global minimum of the distance function. The perturbed version of the single global minimum of the noiseless distance function f will remain in the form of a single global minimum for the noisy distance function with high probability if the noise level is sufficiently small. The noise similarly introduces a perturbation on the SIDEN as well. The exact derivation of the SIDEN in the noisy setting requires the examination of the first zero-crossings of the derivative of the noisy distance function along arbitrary directions T around its global minimum. At small noise levels, these zero-crossings are expected to be perturbed versions of the first zero-crossings of $df(tT)/dT$ around the origin, which define the boundary of the noiseless SIDEN $\mathcal{S}$. The perturbation on the zero-crossings depends on the noise level. If the noise level is sufficiently small, the perturbation on the zero-crossings will be smaller than the distance between $\mathcal{S}$ and its estimate $\mathcal{Q}$. This is due to the fact that $\mathcal{Q}$ is a worst-case estimate for $\mathcal{S}$ and its boundary is sufficiently distant from the boundary of $\mathcal{S}$ in practice, which is also confirmed by the experiments in Section 5. In this case, the estimate $\mathcal{Q}$ obtained from the noiseless distance function f is also a subset of the noisy SIDEN. Therefore, under the small noise assumption, $\mathcal{Q}$ can be considered as an estimate of the noisy SIDEN as well and it can be used in the alignment of noisy images in practice.[3] Our alignment analysis in this section relies on this assumption. Since we consider that the reference and target patterns are aligned with a descent-type optimization method, the solution will converge to the global minimum of the noisy distance function in the noisy setting. The alignment error is then given by the change in the global minimum of the distance function, which we analyze now.

The selection of the noise model for the representation of the deviation of the target pattern from the translation manifold of the reference pattern depends on the imaging application. It is common practice to represent non-predictable deviations of the image intensity function from the image model with additive Gaussian noise. This noise model fits well the image intensity variations due to imperfections of the image capture system, sensor noise, etc. Meanwhile, in some settings, one may have a prior knowledge of the type of the deviation of the target image from the translation manifold of the reference image. For instance, the deviation from the translation manifold may be due to some geometric image deformations, non-planar scene structures, etc. In such settings, one may be able to bound the magnitude of the deviation of the image intensity function from the translation model. Considering these, we examine two different noise models in our analysis. We first focus on a setting where the target pattern is corrupted with respect to an analytic noise model in the continuous space $L^2(\mathbb{R}^2)$. The analytic noise model is inspired by the i.i.d. Gaussian noise in the discrete space $\mathbb{R}^n$. In

---

[3]The validity of this approximation is confirmed by the numerical simulation results in Section 5.

Section 4.1, we derive a probabilistic upper bound on the alignment error for this setting in terms of the parameters of the reference pattern and the noise model. Then, in Section 4.2, we generalize the results of Section 4.1 to arbitrary noise patterns in $L^2(\mathbb{R}^2)$ and derive an error bound in terms of the norm of the noise pattern. In Section 4.3 we discuss how these results can be adapted to a setting where both the reference image and the target image are corrupted with noise. Lastly, the influence of smoothing the reference and target patterns on the alignment error is studied in Section 4.4.

Throughout Section 4, we use the notations $\overline{(\cdot)}$ and $\underline{(\cdot)}$ to refer respectively to upper and lower bounds on the variable $(\cdot)$. The parameters corresponding to smoothed patterns are written as $\hat{(.)}$ as in Section 3.3. The notations $R_{(.)}$ and $C_{(.)}$ are used to denote important upper bounds appearing in the main results, which are associated with the parameter in the subscript.

4.1. Derivation of an Upper Bound for Alignment Error for Gaussian Noise. We consider the noiseless reference pattern p in (3.4) and a target pattern that is a noisy observation of a translated version of p. We assume an analytical noise model given by

$$w(X) = \sum_{l=1}^{L} \zeta_l \, \phi_{\xi_l}(X), \tag{4.1}$$

where the noise units $\phi_{\xi_l}(X)$ are Gaussian atoms of scale ϵ. The coefficients ζ_l and the noise atom parameters ξ_l are assumed to be independent. The noise atoms are of the form $\phi_{\xi_l}(X) = \phi\big(E^{-1}(X - \delta_l)\big)$ where

$$E = \begin{bmatrix} \epsilon & 0 \\ 0 & \epsilon \end{bmatrix}, \quad \delta_l = \begin{bmatrix} \delta_{x,l} \\ \delta_{y,l} \end{bmatrix}.$$

The vector δ_l is the random translation parameter of the noise atom ϕ_{ξ_l} such that the random variables $\{\delta_{x,l}\}_{l=1}^{L}, \{\delta_{y,l}\}_{l=1}^{L} \sim U[-b, b]$ have an i.i.d. uniform distribution. Here, b is a fixed parameter used to define a region $[-b, b] \times [-b, b] \subset \mathbb{R}^2$ in the image plane, which is considered as a support region capturing a substantial part of the energy of reference and target images. The centers of the noise atoms are assumed to be uniformly distributed in this region. In order to have a realistic noise model, the number of noise units $L \gg K$ is considered to be a very large number and the scale $\epsilon > 0$ of noise atoms is very small. The parameters L and ϵ will be treated as noise model constants throughout the analysis. The coefficients $\zeta_l \sim N(0, \eta^2)$ of the noise atoms are assumed to be i.i.d. with a normal distribution of variance η^2.

The continuous-space noise model $w(X)$ is chosen in analogy with the digital i.i.d. Gaussian noise in the discrete space $\mathbb{R}^n$. The single isotropic scale parameter ϵ of noise units bears resemblance to the 1-pixel support of digital noise units. The uniform distribution of the position δ_l of noise units is similar to the way digital noise is defined on a uniform pixel grid. The noise coefficients ζ_l have an i.i.d. normal distribution as in the digital case. If our noise model $w(X)$ has to approximate the digital Gaussian noise in a continuous setting, the noise atom scale ϵ is chosen comparable to the pixel width and L corresponds to the resolution of the discrete image.

Let now p_n be a noisy observation of p such that $p_n(X) = p(X) + w(X)$, where w and p are independent according to the noise model (4.1). We assume that the target pattern is a

translated version of $p_n(X)$ so that it takes the form $p_n(X - tT)$. Then, the noisy distance function between $p(X)$ and $p_n(X - tT)$ is given by

$$g(tT) = \int_{\mathbb{R}^2} \big(p(X) - p_n(X - tT)\big)^2 dX = \int_{\mathbb{R}^2} \big(p(X) - p(X - tT) - w(X - tT)\big)^2 dX. \quad (4.2)$$

This can be written as $g(tT) = f(tT) + h(tT)$, where

$$h(tT) := -2 \int_{\mathbb{R}^2} \big(p(X) - p(X - tT)\big) w(X - tT) \, dX + \int_{\mathbb{R}^2} w^2(X - tT) \, dX. \quad (4.3)$$

The function h represents the deviation of g from f. We call h the distance deviation function. The expected value of h is independent of the translation tT and given by $\mu_h := E[h(tT)] = \frac{\pi}{2} L \eta^2 \epsilon^2$ where $E[.]$ denotes the expectation [26, Appendix B.1]. Therefore, $E[g(tT)] = f(tT) + \mu_h$ and the global minimum of $E[g(tT)]$ is at $tT = 0$. However, due to the probabilistic perturbation caused by the noise w, the global minimum of g is not at $tT = 0$ in general. We consider g to have a single global minimum and denote its location by $t_0 T_0$. However, the single global minimum assumption is not a strict hypothesis of our analysis technique; i.e., the upper bound that we derive for the distance between $t_0 T_0$ and the origin is still valid if g has more than one global minimum. In this case, the obtained upper bound is valid for all global minima.

We now continue with the derivation of a probabilistic upper bound on the distance t_0 between the location $t_0 T_0$ of the global minimum of g and the location 0 of the global minimum of f. We show in [26, Appendix B.2] that t_0 satisfies the equation

$$\frac{t_0^2}{2} \left(\frac{d^2 f(tT_0)}{dt^2} \bigg|_{t=t_1} + \frac{d^2 f(tT_0)}{dt^2} \bigg|_{t=t_2} + \frac{d^2 h(tT_0)}{dt^2} \bigg|_{t=t_1} \right) = |h(0) - h(t_0 T_0)| \quad (4.4)$$

for some $t_1 \in [0, t_0]$ and $t_2 \in [0, t_0]$. Our derivation of an upper bound for t_0 will be based on (4.4). The above equation shows that t_0 can be upper bounded by finding a lower bound on the term

$$\frac{d^2 f(tT_0)}{dt^2} \bigg|_{t=t_1} + \frac{d^2 f(tT_0)}{dt^2} \bigg|_{t=t_2} + \frac{d^2 h(tT_0)}{dt^2} \bigg|_{t=t_1} \quad (4.5)$$

and an upper bound for the term $|h(0) - h(t_0 T_0)|$. However, h is a probabilistic function; i.e., $h(tT)$ and its derivatives are random variables. Therefore, the upper bound that we will obtain for t_0 is a probabilistic bound given in terms of the variances of $h(0) - h(t_0 T_0)$ and $d^2 h(tT_0)/dt^2$.

In the rest of this section, we proceed as follows. First, in order to be able to bound $|h(0) - h(tT)|$ probabilistically, in Lemma 4.1 we present an upper bound on the variance of $h(0) - h(tT)$. Next, in order to bound the term in (4.5), we state a lower bound for $d^2 f(tT)/dt^2$ in Lemma 4.2 and an upper bound for the variance of $d^2 h(tT)/dt^2$ in Lemma 4.3. These results are finally put together in the main result of this section, namely Theorem 4.4, where an upper bound on t_0 is obtained based on (4.4). Theorem 4.4 applies Chebyshev's inequality to employ the bounds derived in Lemmas 4.1 and 4.3 to define probabilistic upper bounds on the terms $|h(0) - h(t_0 T_0)|$ and $|d^2 h(tT_0)/dt^2|$. Then, this is combined with the

bound on $d^2 f(tT)/dt^2$ in Lemma 4.2 to obtain a probabilistic upper bound on t_0 from the relation (4.4).

In the derivation of this upper bound, the direction T_0 of the global minimum of g is treated as an arbitrary and unknown unit-norm vector. Moreover, the variances of $h(0) - h(tT)$ and $d^2 h(tT)/dt^2$ have a complicated dependence on t, which makes it difficult to use them directly in (4.4) to obtain a bound on t_0. In order to cope with the dependences of these terms on t and T, the upper bounds presented in Lemmas 4.1 and 4.3 are derived as uniform upper bounds over the closed ball of radius $\overline{t}_0 > 0$, $B_{\overline{t}_0}(0) = \{tT : T \in S^1, 0 \leq t \leq \overline{t}_0\}$. The upper bounds are thus independent of tT and valid for all tT vectors in $B_{\overline{t}_0}(0)$. In these lemmas, the parameter $\overline{t}_0$ is considered to be a known threshold for t_0, such that $t_0 \leq \overline{t}_0$. This parameter will be assigned a specific value in Theorem 4.4.

We begin with bounding the variance of term $h(0) - h(tT)$ in order to find an upper bound for the right hand side of (4.4). Let us denote $\Delta h(tT) := h(0) - h(tT)$. From (4.3), $\Delta h(tT) = h(0) - h(tT) = 2 \int_{\mathbb{R}^2} (p(X) - p(X - tT)) w(X - tT) dX$, where we have used the fact that $\int_{\mathbb{R}^2} w^2(X - tT) dX = \int_{\mathbb{R}^2} w^2(X) dX$. Let $\sigma^2_{\Delta h(tT)}$ denote the variance of $\Delta h(tT)$. In the following lemma, we state an upper bound on $\sigma^2_{\Delta h(tT)}$. Let us define beforehand the following constants for the k-th atom of p

$$\Phi_k := \Psi_k(\sigma_k^2 + E^2)^{-1} \Psi_k^{-1}, \qquad \kappa_k := \frac{\pi |\sigma_k| |E|}{\sqrt{|\sigma_k^2 + E^2|}}.$$

Also, let $J^- = \{(j,k) : c_j c_k < 0\}$ and $J^+ = \{(j,k) : c_j c_k > 0\}$ denote the set of (j,k) indices with negative and positive coefficient products.

Lemma 4.1. *Let $\overline{t}_0 > 0$, and let $tT \in B_{\overline{t}_0}(0)$. Then, the variance $\sigma^2_{\Delta h(tT)}$ of $\Delta h(tT)$ can be upper bounded as*

$$\sigma^2_{\Delta h(tT)} < R_{\sigma^2_{\Delta h}} := C_{\sigma^2_{\Delta h}} \eta^2 \tag{4.6}$$

where

$$C_{\sigma^2_{\Delta h}} := 4 L \left(\sum_{(j,k) \in J^+} c_j \kappa_j c_k \kappa_k \overline{\mathbb{c}}_{jk} + \sum_{(j,k) \in J^-} c_j \kappa_j c_k \kappa_k \underline{\mathbb{d}}_{jk} \right).$$

Here the terms $\overline{\mathbb{c}}_{jk}$ and $\underline{\mathbb{d}}_{jk}$ are constants depending on $\overline{t}_0$ and the atom parameters of p. In particular, $\overline{\mathbb{c}}_{jk}$ and $\underline{\mathbb{d}}_{jk}$ are bounded functions of $\overline{t}_0$, given in terms of exponentials of second-degree polynomials of $\overline{t}_0$ with negative leading coefficients.

The proof of Lemma 4.1 is presented in [26, Appendix B.4]. In the proof, a uniform upper bound $\overline{\mathbb{c}}_{jk}$ and a uniform lower bound $\underline{\mathbb{d}}_{jk}$ are derived for the additive terms[4] constituting the variance of $\Delta h(tT)$. The exact expressions of $\overline{\mathbb{c}}_{jk}$ and $\underline{\mathbb{d}}_{jk}$ are given in Appendix A.

We have thus stated a uniform upper bound $R_{\sigma^2_{\Delta h}}$ for the variance of $\Delta h(tT)$ which will be used to derive an upper bound for the right hand side of (4.4) in Theorem 4.4. We now continue with the examination of the left hand side of (4.4). We begin with the term $d^2 f(tT)/dt^2$. The following lemma gives a lower bound on the second derivative of the noiseless distance function $f(tT)$ in terms of the pattern parameters.

---

[4]$\overline{\mathbb{c}}_{jk}$ and $\underline{\mathbb{d}}_{jk}$ are upper and lower bounds for the terms $\mathbb{c}_{jk}$, $\mathbb{d}_{jk}$ used in [26, Lemma 1].

Lemma 4.2. *The second derivative of $f(tT)$ along the direction T can be uniformly lower bounded for all $t \in [0, t_0]$ and for all directions $T \in S^1$ as follows*

$$\frac{d^2 f(tT)}{dt^2} \geq \underline{r}_0 + \underline{r}_2 t_0^2 + \underline{r}_3 t_0^3. \tag{4.7}$$

Here $\underline{r}_0 > 0$, $\underline{r}_2 \leq 0$, and $\underline{r}_3 < 0$ are constants depending on the atom parameters of p. In particular, $\underline{r}_0$, $\underline{r}_2$, $\underline{r}_3$ are obtained from the eigenvalues of some matrices derived from the parameters c_j, τ_j, Q_{jk}, Σ_{jk}.

The proof of Lemma 4.2 is given in [26, Appendix B.5] and the exact expressions of $\underline{r}_0$, $\underline{r}_2$, $\underline{r}_3$ are given in Appendix B. The above lower bound on the second derivative of $f(tT)$ is independent of the direction T and the amount t of translation, provided that t is in the interval $[0, t_0]$. In fact, the statement of Lemma 4.2 is general in the sense that t_0 can be any positive scalar. However, in the proof of Theorem 4.4, we use Lemma 4.2 for the t_0 value that represents the deviation between the global minima of f and g.

The result of Lemma 4.2 will be used in Theorem 4.4 in order to lower bound the second derivative of f in (4.4). We now continue with the term $d^2 h(tT)/dt^2$ in (4.4). Let $h''(tT) := d^2 h(tT)/dt^2$ denote the second derivative of the deviation function h along the direction T. Since $h''(tT)$ can take both positive and negative values, in the calculation of a lower bound for the term (4.5), we need a bound on the magnitude $|h''(tT)|$ of this term. It can be bounded probabilistically in terms of the variance of $h''(tT)$. We thus state the following uniform upper bound on the variance of $h''(tT)$.

Lemma 4.3. *Let $\bar{t}_0 > 0$, and let $tT \in B_{\bar{t}_0}(0)$. Then, the variance $\sigma^2_{h''(tT)}$ of $h''(tT)$ can be upper bounded as*

$$\sigma^2_{h''(tT)} < R_{\sigma^2_{h''}} := C_{\sigma^2_{h''}} \eta^2 \tag{4.8}$$

where

$$C_{\sigma^2_{h''}} := 4L \left(\sum_{(j,k) \in J^+} c_j c_k \kappa_j \kappa_k \, \overline{\mathbb{e}}_{jk} + \sum_{(j,k) \in J^-} c_j c_k \kappa_j \kappa_k \, \underline{\mathbb{f}}_{jk} \right).$$

Here $\overline{\mathbb{e}}_{jk}$ is a constant depending on the atom parameters of p; and the term $\underline{\mathbb{f}}_{jk}$ depends on the atom parameters of p and $\bar{t}_0$. In particular, $\overline{\mathbb{e}}_{jk}$ is given in terms of rational functions of the eigenvalues of Φ_k matrices; and $\underline{\mathbb{f}}_{jk}$ is a bounded function of $\bar{t}_0$ given in terms of exponentials of second-degree polynomials of $\bar{t}_0$ with negative leading coefficients.

The proof of Lemma 4.3 is given in [26, Appendix B.7]. The proof derives uniform upper and lower bounds $\overline{\mathbb{e}}_{jk}$, $\underline{\mathbb{f}}_{jk}$ for the additive terms[5] in the representation of $\sigma^2_{h''(tT)}$. The exact expressions for $\overline{\mathbb{e}}_{jk}$ and $\underline{\mathbb{f}}_{jk}$ are given in Appendix C.

Now we are ready to give our main result about the bound on the alignment error. The following theorem states an upper bound on the distance between the locations of the global minima of f and g in terms of the noise standard deviation η and the atom parameters of p, provided that η is smaller a threshold η_0. The threshold η_0 is obtained from the bounds

---

[5] $\overline{\mathbb{e}}_{jk}$ and $\underline{\mathbb{f}}_{jk}$ are upper and lower bounds for the terms $\mathbb{e}_{jk}$, $\mathbb{f}_{jk}$ used in [26, Lemma 3].

derived in Lemmas 4.1, 4.2 and 4.3 such that the condition $\eta < \eta_0$ guarantees that the assumption $t_0 < \bar{t}_0$ holds. In the theorem, the parameter $\bar{t}_0$, which is treated as a predefined threshold on t_0 in the previous lemmas, is also assigned a specific value in terms the constants $\underline{r}_0$, $\underline{r}_2$, $\underline{r}_3$ of Lemma 4.2.

Theorem 4.4. *Let*

$$\bar{t}_0 := \sqrt{\frac{r_0}{2|\underline{r}_2| + 2^{2/3} \underline{r}_0^{1/3} |\underline{r}_3|^{2/3}}}. \tag{4.9}$$

Let $R_{\sigma_{\Delta h}} := \sqrt{R_{\sigma^2_{\Delta h}}}$ and $R_{\sigma_{h''}} := \sqrt{R_{\sigma^2_{h''}}}$, where $R_{\sigma^2_{\Delta h}}$ and $R_{\sigma^2_{h''}}$ are as defined in (4.6) and (4.8), and evaluated at the value of $\bar{t}_0$ given above. Also, let $C_{\sigma_{\Delta h}} := \sqrt{C_{\sigma^2_{\Delta h}}}$ and $C_{\sigma_{h''}} := \sqrt{C_{\sigma^2_{h''}}}$.

Assume that for some $s > \sqrt{2}$, the noise standard deviation η is smaller than η_0 such that

$$\eta \le \eta_0 := \frac{\bar{t}_0^2 \, r_0}{2 \, s \, C_{\sigma_{\Delta h}} + \bar{t}_0^2 \, s \, C_{\sigma_{h''}}}. \tag{4.10}$$

Then, with probability at least $1 - \frac{2}{s^2}$, the distance t_0 between the global minima of f and g is bounded as

$$t_0 < R_{t_0} := \sqrt{\frac{2s \, R_{\sigma_{\Delta h}}}{r_0 - s \, R_{\sigma_{h''}}}}. \tag{4.11}$$

The proof of Theorem 4.4 is given in [26, Appendix B.8]. In the proof, we make use of the upper bounds $R_{\sigma^2_{\Delta h}}$, $R_{\sigma^2_{h''}}$ on $\sigma^2_{\Delta h(tT)}$, $\sigma^2_{h''(tT)}$, and the lower bound on $d^2 f(tT)/dt^2$ given in (4.7). The upper bound R_{t_0} in (4.11) shows that the alignment error increases with the increase in the noise level, since $R_{\sigma_{\Delta h}}$ and $R_{\sigma_{h''}}$ are linearly proportional to the noise standard deviation η. The increase of the error with the noise is expected. It can also be seen from (4.11) that the increase in the term r_0, which is proportional to the second derivative of the noiseless distance f, reduces the alignment error; whereas an increase in the term $R_{\sigma_{h''}}$, which is related to the second derivative of h, increases the error. This can be explained as follows. If f has a sharp increase around its global minimum at 0, i.e., f has a large second derivative, the location of its minimum is less affected by h. Likewise, if the distance deviation function h has a large second derivative, it introduces a larger alteration around the global minimum of f, which causes a bigger perturbation on the position of the minimum.

Theorem 4.4 states a bound on t_0 under the condition that the noise standard deviation η is smaller than the threshold value η_0, which depends on the pattern parameters (through the terms $\underline{r}_0$, $\underline{r}_2$, $\underline{r}_3$, $\bar{t}_0$) as well as the noise parameters L and ϵ (through the terms $C_{\sigma_{\Delta h}}$ and $C_{\sigma_{h''}}$). The threshold η_0 thus defines an admissible noise level such that the change in the location of the global minimum of f can be properly upper bounded. This admissible noise level is derived from the condition $R_{t_0} \le \bar{t}_0$, which is partially due to our proof technique. However, we remark that the existence of such a threshold is intuitive in the sense that it states a limit on the noise power in comparison with the signal power. Note also that the

denominator $\underline{r}_0 - s\, R_{\sigma_{h''}}$ of R_{t_0} should be positive, which also yields a condition on the noise level

$$\eta < \eta_0' = \frac{r_0}{s\, C_{\sigma_{h''}}}.$$

However, this condition is already satisfied due to the hypothesis $\eta \leq \eta_0$ of the theorem, since $\eta_0 < \eta_0'$ from (4.10). Lastly, the fact that the theoretical upper bound tends to infinity at large values of the noise level should be interpreted in the way that the alignment error bound is not informative at high noise levels. This is mainly because when the noise level gets too high, the noise component becomes the determining factor in the estimation of the translation parameters rather than the noiseless component of the image. As this estimate is unreliable, no theoretical guarantee can be obtained for the performance of registration algorithms.

4.2. Generalization of the Alignment Error Bound to Arbitrary Noise Models. Here, we generalize the results of the previous section in order to derive an alignment error bound for arbitrary noise patterns. In general, the characteristics of the noise pattern vary depending on the imaging application. In particular, while the noise pattern may have high correlation with the reference pattern in some applications (e.g., noise resulting from geometric deformations of the pattern), its correlation with the reference pattern may be small in some other settings where the noise stems from a source that does not depend on the image. We thus focus on two different scenarios. In the first and general setting, we do not make any assumption on the noise characteristics and bound the alignment error in terms of the norm of the noise pattern. Then, in the second setting, we consider that the noise pattern has small correlation with the points on the translation manifold of the reference pattern and show that the alignment error bound can be made sharper in this case.

We assume that the reference pattern $p(X)$ is noiseless and we write the target pattern as $p_g(X - tT)$, where $p_g(X) = p(X) + z(X)$ is a generalized noisy observation of p such that $z \in L^2(\mathbb{R}^2)$ is an arbitrary noise pattern. Then, the generalized noisy distance function is

$$g_g(tT) = \int_{\mathbb{R}^2} \big(p(X) - p_g(X - tT)\big)^2 \, dX$$

and the generalized deviation function is $h_g = g_g(tT) - f(tT)$. Let us call $u_0 U_0$ the point where g_g has its global minimum. Then the distance between the global minima of g_g and f is given by u_0.

We begin with the first setting and state a generic bound for the alignment error u_0 in terms of the norm of the noise $\nu := \|z\|$. In our main result, we denote by $R_p := \|p\|$ the norm of the pattern p, and make use of an upper bound $R_{p''}$ for the norm $\|d^2 p(\cdot + tT)/dt^2\|$ of the second derivative of $p(X + tT)$. It has been derived in terms of the atom parameters of p in [26, Lemma 4]. We state below our generalized alignment error result for arbitrary noise patterns.

Theorem 4.5. *Let $\bar{t}_0$ be defined as in (4.9). Assume that the norm ν of z is smaller than ν_0 such that*

$$\nu \leq \nu_0 := \frac{\bar{t}_0^2 r_0}{8 R_p + 2 R_{p''} \bar{t}_0^2} \tag{4.12}$$

where $\underline{r}_0$ is the constant in Lemma 4.2. Then, the distance u_0 between the global minima of f and g_g is bounded as

$$u_0 \leq R_{u_0} := \sqrt{\frac{8R_p\nu}{\underline{r}_0 - 2R_{p''}\nu}}. \tag{4.13}$$

Theorem 4.5 is proved in [26, Appendix C.2]. The theorem states an upper bound on the alignment error for the general case where the only information used about the noise pattern is its norm. The alignment error bound R_{u_0} is a generalized and deterministic version of the probabilistic bound R_{t_0} derived for the Gaussian noise model. In the proof of the theorem, the change $h_g(0) - h_g(u_0U_0)$ in the deviation function is bounded by $4R_p\nu$. The second derivative of the noiseless distance function f is captured by $\underline{r}_0$ as in the Gaussian noise case. Finally, the term $2R_{p''}\nu$ bounds the second derivative of the deviation h_g. Based on these, the above result is obtained by following similar steps as in Section 4.1.

We now continue with the second setting where the noise pattern z has small correlation with the points on the translation manifold $\mathcal{M}(p)$ of p. We characterize the correlation of two patterns with their inner product. Assume that a uniform correlation upper bound r_{pz} is available such that

$$\left| \int_{\mathbb{R}^2} p(X + tT)z(X)dX \right| \leq r_{pz} \tag{4.14}$$

for all t and T. The following corollary builds on Theorem 4.5 and states that the bound on the alignment error can be made sharper if the correlation bound is sufficiently small.

Corollary 4.6. Let $\bar{t}_0$ be defined as in (4.9) and let a uniform upper bound r_{pz} for the correlation be given such that $r_{pz} < \bar{t}_0^2 \underline{r}_0 / 8$.

Assume that the norm ν of z is smaller than ν_0 such that

$$\nu \leq \nu_0 := \frac{\bar{t}_0^2 \underline{r}_0 - 8r_{pz}}{2R_{p''}\bar{t}_0^2}. \tag{4.15}$$

Then, the distance u_0 between the global minima of f and g_g is bounded as

$$u_0 \leq Q_{u_0} := \sqrt{\frac{8\,r_{pz}}{\underline{r}_0 - 2R_{p''}\nu}}. \tag{4.16}$$

The proof of Corollary 4.6 is given in [26, Appendix C.3]. One can observe that the alignment error bound Q_{u_0} approaches zero as the uniform correlation bound approaches zero. Therefore, if r_{pz} is sufficiently small, Q_{u_0} will be smaller than the general bound R_{u_0}. This shows that, regardless of the noise level, the alignment error is close to zero if the noise pattern z is almost orthogonal to the translation manifold $\mathcal{M}(p)$ of the reference pattern.

Remark. In the alignment error bounds derived in this paper, we assume that entire images are used in the registration rather than windowed regions in the reference and target images. If the reference and target images are windowed in an image registration application, this also causes a perturbation in the distance function, in addition to the perturbation caused by the image noise. In [26, Appendix C.4], we examine how windowing affects the alignment

error. The analysis in [26, Appendix C.4] shows that the increase in the alignment error due to windowing can be mitigated by choosing a window function with a sufficiently slow variation.

4.3. Adaptation of the error bounds for two-sided noise models. Throughout our analysis, we have assumed that the reference image is noiseless and the image noise (w or z) acts only on the target image. However, in practice the reference image can also be contaminated with noise. We now discuss how our results can be adapted to this case.

First, it is easy to show that one gets exactly the same error bounds if the noisy distance function $g(tT)$ in (4.2) is slightly modified as

$$\int_{\mathbb{R}^2} \big(p(X) - p(X - tT) - w(X)\big)^2 dX. \tag{4.17}$$

This corresponds to the situation where the target image $p(X - tT) + w(X)$ is first translated and then corrupted by the additive noise w. Now, if the reference and the target images are corrupted respectively with the noise instances w_1 and w_2, the distance between the reference image $p(X) + w_1(X)$ and the target image $p(X - tT) + w_2(X)$ is given by

$$\int_{\mathbb{R}^2} \big(p(X) - p(X - tT) - (w_2(X) - w_1(X))\big)^2 dX.$$

One can observe that this distance function is the same as the distance function in (4.17) with $w = w_2 - w_1$. Therefore, if w_1 and w_2 have identical distributions conforming to the Gaussian noise model in (4.1), the compound noise $w = w_2 - w_1$ can also be represented with the same model such that the number of atoms L is twice the number of atoms in w_1 and w_2, and the other model parameters are the same. Hence, one can easily adapt the main result in Theorem 4.4 to obtain an alignment error bound for the two-sided noise model. We observe from Lemmas 4.1 and 4.3 that doubling the number of atoms L would increase the terms $C_{\sigma_{\Delta h}}$, $C_{\sigma_{h''}}$, $R_{\sigma_{\Delta h}}$, $R_{\sigma_{h''}}$ by a factor of $\sqrt{2}$. The alignment error bound for the two-sided Gaussian noise model is thus given as

$$t_0 < R_{t_0} = \sqrt{\frac{2\sqrt{2}\, s\, R_{\sigma_{\Delta h}}}{\underline{r}_0 - s\sqrt{2}\, R_{\sigma_{h''}}}}$$

which holds for the admissible noise threshold

$$\eta_0 = \frac{\bar{t}_0^2\, \underline{r}_0}{2\sqrt{2}\, s\, C_{\sigma_{\Delta h}} + \sqrt{2}\, \bar{t}_0^2\, s\, C_{\sigma_{h''}}}$$

with the same probability as in Theorem 4.4.

Next, the analysis in Section 4.2 for arbitrary noise patterns can also be adapted to the two-sided noise assumption in a similar way. The distance function between the noisy reference pattern $p(X) + z_1(X)$ and the noisy target pattern $p(X - tT) + z_2(X)$ is equal to

$$\int_{\mathbb{R}^2} \big(p(X) - p(X - tT) - z(X)\big)^2 dX$$

where $z(X) = z_2(X) - z_1(X)$. If the norms of z_1 and z_2 are bounded by ν, the norm of z is bounded by 2ν by the triangle inequality. Increasing the norm of the noise level ν by a factor of 2 in Theorem 4.5, we get the following alignment error bound for the two-sided noise model

$$u_0 \leq R_{u_0} = \sqrt{\frac{16 R_p \nu}{r_0 - 4 R_{p''} \nu}}$$

which holds if

$$\nu \leq \nu_0 = \frac{\bar{t}_0^2 r_0}{16 R_p + 4 R_{p''} \bar{t}_0^2}.$$

4.4. Influence of Filtering on Alignment Error. In this section, we examine how the alignment error resulting from the image noise is affected when the reference and target patterns are low-pass filtered. We consider the Gaussian kernel $\frac{1}{\pi \rho^2} \phi_\rho(X)$ defined in Section 3.3 and analyze the dependence of the alignment error bounds obtained for the Gaussian noise and generalized noise models in Sections 4.1 and 4.2 on the filter size ρ and the noise level parameters η and ν.

We begin with the Gaussian noise model $w(X)$. The filtered reference pattern $\hat{p}(X)$ and the filtered noisy observation $\hat{p}_n(X)$ of the reference pattern are given by

$$\hat{p}(X) = \sum_{k=1}^{K} \hat{c}_k \phi_{\hat{\gamma}_k}(X), \qquad \hat{p}_n(X) = \hat{p}(X) + \hat{w}(X) = \sum_{k=1}^{K} \hat{c}_k \phi_{\hat{\gamma}_k}(X) + \sum_{l=1}^{L} \hat{\zeta}_l \, \phi_{\hat{\xi}_l}(X).$$

Remember from Section 3.3 that the rotation and translation parameters of the atoms of $\hat{p}(X)$ do not depend on ρ; and the scale matrices vary with ρ such that $\hat{\sigma}_k^2 = \sigma_k^2 + \Upsilon^2$. The parameters of the smoothed noise atoms can be obtained similarly to the parameters of the atoms in $\hat{p}$; i.e., $\phi_{\hat{\xi}_l}(X) = \phi\big(\hat{E}^{-1}(X - \delta_l)\big)$, where $\hat{E}^2 = E^2 + \Upsilon^2$. This gives the scale parameter of smoothed noise atoms, which is written as

$$\hat{\epsilon} = \sqrt{\epsilon^2 + \rho^2}. \tag{4.18}$$

The smoothed noise coefficients are given by $\hat{\zeta}_l = \zeta_l |E|/|\hat{E}| = \zeta_l \epsilon^2/(\epsilon^2 + \rho^2)$. Since all the coefficients ζ_l are multiplied by a factor of $\epsilon^2/(\epsilon^2 + \rho^2)$, the variance of the smoothed noise atom coefficients is

$$\hat{\eta}^2 = \left(\frac{\epsilon^2}{\epsilon^2 + \rho^2}\right)^2 \eta^2. \tag{4.19}$$

As the noise atom units are considered to have very small scale, one can assume first that $\rho \gg \epsilon$ for typical values of the filter size ρ. Then the relations in (4.18) and (4.19) give the joint variations of $\hat{\epsilon}$ and $\hat{\eta}$ with η and ρ as

$$\hat{\epsilon} = O(\rho), \qquad \hat{\eta} = O(\eta \rho^{-2}). \tag{4.20}$$

We now state the dependence of the bound $\hat{R}_{t_0}$ on ρ and η in the following main result.

Theorem 4.7. The joint variation of the alignment error bound $\hat{R}_{t_0}$ for the smoothed image pair with respect to η and ρ is given by

$$\hat{R}_{t_0} = O\left(\sqrt{\frac{\eta\,\rho^{-1}}{(1+\rho^2)^{-2} - \eta\,\rho^{-3}}}\right) = O\left(\sqrt{\frac{\eta\,\rho^3}{1 - \eta\,\rho}}\right).$$

Therefore, for a fixed noise level, $\hat{R}_{t_0}$ increases at a rate of $O\left(\rho^{3/2}(1-\rho)^{-1/2}\right)$ with the increase in the filter size ρ. Similarly, for a fixed filter size, the rate of increase of $\hat{R}_{t_0}$ with the noise standard deviation η is $O\left(\eta^{1/2}(1-\eta)^{-1/2}\right)$.

The proof of Theorem 4.7 is presented in [26, Appendix D.2]. The stated result is obtained by using the relations in (4.20) to determine how the terms $\hat{R}_{\sigma_{\Delta h}}$, $\hat{\underline{r}}_0$, $\hat{R}_{\sigma_{h''}}$ in the expression of $\hat{R}_{t_0}$ vary with ρ and η. We omit the constants in the variation of the error bound for the sake of simplicity. However, due to the relations (4.6) and (4.8), the constants multiplying the noise level parameter η in the numerator and the denominator are respectively related with the parameters $C_{\sigma_{\Delta h}}$ and $C_{\sigma_{h''}}$. Similarly, the positive constant in the denominator is associated with the parameter $\underline{r}_0$.

Theorem 4.7 constitutes the summary of our analysis about the effect of filtering on the alignment accuracy for the Gaussian noise model. While the aggravation of the alignment error with the increase in the noise level is an intuitive result, the theorem states that filtering the patterns under the presence of noise decreases the accuracy of alignment as well. Remember that this is not the case for noiseless patterns. The result of the theorem can be interpreted as follows. Smoothing the reference and target patterns diffuses the perturbation on the distance function, which is likely to cause a bigger shift in the minimum of the distance function and hence reduce the accuracy of alignment. The estimation $\hat{R}_{t_0} = O\left(\rho^{3/2}(1-\rho)^{-1/2}\right)$ of the alignment error suggests that the dependence of the error on ρ is between linear and quadratic for small values of ρ, whereas it starts to increase more dramatically when ρ takes larger values. Similarly, $\hat{R}_{t_0}$ is proportional to the square root of η for small η and it increases at a sharper rate as η grows.

Next, we look at the variation of the bounds $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ for arbitrary noise patterns, which are respectively obtained for the general and small-correlation cases. We present the following theorem, which is the counterpart of Theorem 4.7 for arbitrary noise models.

Theorem 4.8. The alignment error bounds $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ for arbitrary noise patterns have a variation of

$$O\left(\sqrt{\frac{\nu\,(1+\rho^2)}{1-\nu}}\right) \tag{4.21}$$

with the noise level ν and the filter size ρ. Therefore, for a fixed noise level, the errors $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ increase at a rate of $O\left((1+\rho^2)^{1/2}\right)$ with the increase in the filter size ρ. Similarly, for a fixed filter size, $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ increase at a rate of $O\left(\nu^{1/2}(1-\nu)^{-1/2}\right)$ with respect to the noise norm ν.

The proof of Theorem 4.8 is given in [26, Appendix E.1]. The dependence of the generalized bounds $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ on the noise norm ν is the same as the dependence of $\hat{R}_{t_0}$ on η. However,

the variation of $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ with ρ is seen to be slightly different from that of $\hat{R}_{t_0}$. This stems from the difference between the two models. In the generalized noise model z, we have treated the norm ν of z as a known fixed number and we have characterized the alignment error in terms of ν. On the other hand, w is a probabilistic Gaussian noise model; therefore, it is not possible to bound its norm with a fixed parameter. For this reason, the alignment error for w has been derived probabilistically in terms of the standard deviations of the involved parameters. Since the filter size ρ affects the norm of z and the standard deviations of the terms related to w in different ways, it has a different effect on these two type of alignment error bounds. The reason why the two error bounds have the same kind of dependence on the noise level parameters η and ν can be explained similarly. The standard deviations of the terms related to w have a simple linear dependence on η, which is the same as the dependence of the counterparts of these terms in the generalized model on ν.

5. Experiments.
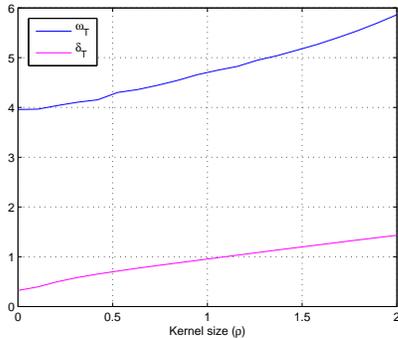
Figure 5.1. *The variations of the true distance $\hat{\omega}_T$ of the boundary of $\hat{S}$ to the origin and its estimation $\hat{\delta}_T$ with respect to the filter size*

5.1. Evaluation of Alignment Regularity Analysis.
We first evaluate our theoretical results about SIDEN estimation with an experiment that compares the estimated SIDEN to the true SIDEN. We generate a reference pattern consisting of 40 randomly selected Gaussian atoms with random coefficients, and choose a random unit direction T for pattern displacement. Then, we determine the distance $\hat{\omega}_T$ of the true SIDEN boundary from the origin along T, and compare it to its estimation $\hat{\delta}_T$ for a range of filter sizes ρ (With an abuse of notation, the parameter denoted as $\hat{\omega}_T$ here corresponds in fact to $\sup \hat{\omega}_T$ in the definition of SIDEN in (3.7)). The distance $\hat{\omega}_T$ is computed by searching the first zero-crossing of $d\hat{f}(tT)/dt$ numerically, while its estimate $\hat{\delta}_T$ is computed according to Theorem 3.4. We repeat the experiment 300 times with different random reference patterns p and directions T and average the results of the cases where $d\hat{f}(tT)/dt$ has zero-crossings for all values of ρ (i.e., 56% of the tested cases). The distance $\hat{\omega}_T$ and its estimation $\hat{\delta}_T$ are plotted in Figure 5.1. The figure shows that $\hat{\delta}_T$ has an approximately linear dependence on ρ. This is an expected behavior, since $\hat{\delta}_T = O\left((1+\rho^2)^{1/2}\right) \approx O(\rho)$ for large ρ. The estimate $\hat{\delta}_T$ is smaller than $\hat{\omega}_T$ since it is a lower bound for $\hat{\omega}_T$. Its variation with ρ is seen to capture well the relative variations of the true SIDEN boundary $\hat{\omega}_T$ with ρ.

5.2. Evaluation of Alignment Accuracy Analysis.
We now present experimental results evaluating the alignment error bounds derived in Section 4. We conduct the experiments on reference and target patterns made up of Gaussian atoms, where the target pattern is generated by corrupting the reference pattern with noise and applying a random translation tT. In all experiments, an estimation $t_e T_e$ of tT is computed by aligning the reference and

(a)
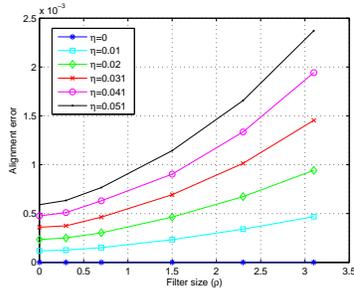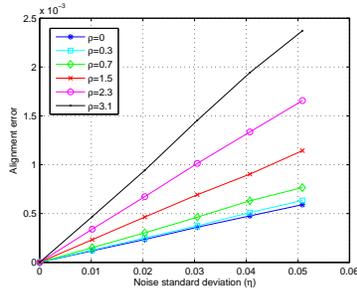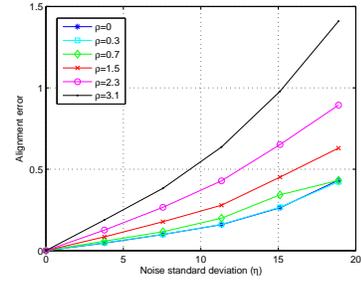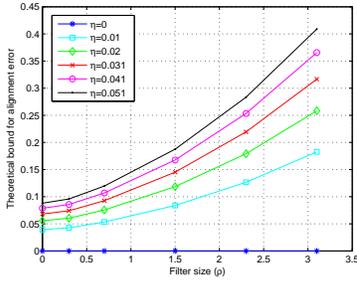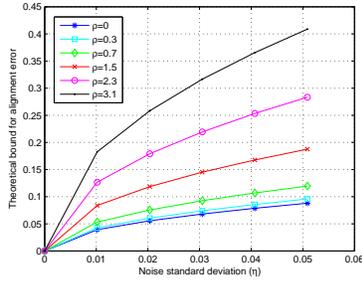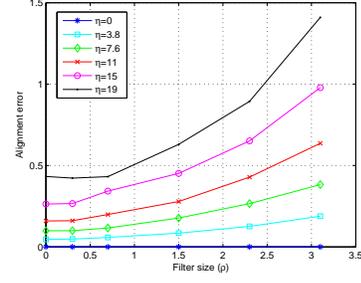
(b)

Figure 5.2. *Alignment error of random patterns as a function of filter size ρ.*

(a)

(b)

Figure 5.3. *Alignment error of random patterns as a function of noise standard deviation η.*

(a)

(b)

Figure 5.4. *Alignment error of random patterns as functions of the noise standard deviation and the filter size, at high noise levels.*

target images with a gradient descent algorithm[6], which gives the experimental alignment error as $\|tT - t_e T_e\|$. The experimental error is then compared to the theoretical bounds derived in Section 4.

5.2.1. Gaussian noise model. In the first set of experiments, we evaluate the results for the Gaussian noise model. We compare the experimental alignment error to the theoretical bound given in Theorem 4.4. [7] In all experiments, the parameter s in Theorem 4.4, which controls the probability, is chosen such that $t_0 < R_{t_0}$ holds with probability greater than 0.5. For each reference pattern, the experiment is repeated for a range of values for noise variances η^2 and filter sizes ρ. The maximum value of the noise standard deviation is taken as the admissible noise level η_0 in Theorem 4.4.

---

[6]In the computation of $t_e T_e$, in order to be able to handle large translations, before the optimization with gradient descent we first do a coarse preregistration of the reference and target images with a search on a coarse grid in the translation parameter domain, whose construction is explained in Section 5.3.

[7]The bound $R_{\sigma^2_{h''}}$ given in Lemma 4.3 is derived from the preliminary bound $R_{\sigma^2_{h''(tT)}}$ in [26, Lemma 3]. In the implementation of Theorem 4.4, in order to obtain a sharper estimate of $R_{\sigma^2_{h''}}$, we compute it by searching the maximum value of $\sigma^2_{h''(tT)}$ over t and T from the expressions for $\mathcal{E}_j$ and $\mathcal{F}_j$ used in the derivation of $R_{\sigma^2_{h''(tT)}}$.

We first experiment on reference patterns built with 20 Gaussian atoms with randomly chosen parameters. The atom coefficients c_k in the reference patterns are drawn from a uniform distribution in $[-1, 1]$; and the position and scale parameters of the atoms are selected such that $\tau_x, \tau_y \in [-4, 4]$ and $\sigma_x, \sigma_y \in [0.3, 2]$. The noise model parameters are set as $L = 750$, $\epsilon = 0.1$. The experiment is repeated on 50 different reference patterns. Then, 50 noisy target patterns are generated for each reference pattern according to the Gaussian noise model w in (4.1) with a random translation tT in the range $tT_x, tT_y \in [-4, 4]$. The results are averaged over all reference and target patterns. In Figure 5.2, the experimental and theoretical values of the alignment error are plotted with respect to the filter size ρ, where different curves correspond to different η values. Figures 5.2(a) and 5.2(b) show respectively the experimental value $\|tT - t_e T_e\|$ and the theoretical upper bound R_{t_0} of the alignment error. Figure 5.3 shows the same results, where the error is given as a function of η. The experimental values and the theoretical bounds are given respectively in Figures 5.3(a) and 5.3(b). Note that, due to the range of atom translation parameters τ_x, τ_y, the energy of the reference pattern is concentrated in the region $[-4, 4] \times [-4, 4]$. Therefore, the maximum filter size $\rho = 3.1$ tested in this experiment is close to the half of the image width. The maximum noise level $\eta = 0.051$ corresponds to an SNR of approximately 26 dB.

The results in Figure 5.2 show that, although the theoretical upper bound is pessimistic (which is due to the fact that the bound is a worst-case analysis), the variation of the experimental value of the alignment error as a function of the filter size is in agreement with that of the theoretical bound. The experimental behavior of the error conforms to the theoretical prediction $\hat{R}_{t_0} \approx O\left(\rho^{3/2} (1-\rho)^{-1/2}\right)$ of Theorem 4.7. Next, the plots of Figure 5.3 suggest that the variation of the theoretical bound R_{t_0} as a function of η is consistent with the result of Theorem 4.7, which can be approximated as $\hat{R}_{t_0} \approx O(\sqrt{\eta})$ for small values of η. On the other hand, the experimental value of the alignment error seems to exhibit a more linear behavior. However, this type of dependence is not completely unexpected. Theorem 4.7 predicts that $\hat{R}_{t_0}$ is of $O(\sqrt{\eta})$ for small η; and $O\left(\eta^{1/2}(1-\eta)^{-1/2}\right)$ for large η, while the experimental value of the error can be rather described as $\|tT - t_e T_e\| = O(\eta)$, which is between these two orders of variation. In order to examine the dependence of the error on η in more detail, we have repeated the same experiments with much higher values of η. The experimental alignment error is given in Figure 5.4, where the error is plotted with respect to the noise standard deviation in Figure 5.4(a) and the filter size in Figure 5.4(b). The maximum noise level $\eta = 19$ in this experiments corresponds to an SNR of approximately -22.5 dB. The results show that, at high noise levels, the variation of the error with η indeed increases above the linear rate $O(\eta)$. The noise levels tested in this high-noise experiment are beyond the admissible noise level derived in Theorem 4.4; therefore, we cannot apply Theorem 4.4 directly in this experiment. However, in view of Theorem 4.7, which states that the error is of $O\left(\eta^{1/2}(1-\eta)^{-1/2}\right)$, these results can be interpreted to provide a numerical justification of our theoretical finding: at relatively high noise levels, the error is expected to increase with η at a sharply increasing rational function rate above the linear rate. The variation of the error with ρ at high noise levels plotted in Figure 5.4(b) is seen to be similar to that of the previous experiments.

We now evaluate our alignment accuracy results under Gaussian noise on face and digit images. First, the reference face pattern is obtained by approximating the face image shown in Figure 5.5(a) with 50 Gaussian atoms. The average atom coefficient magnitude of the

(a)
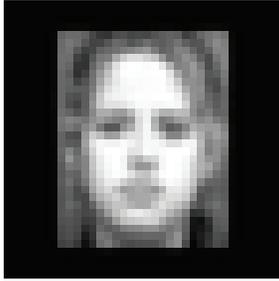
(a) Face image

(a) Digit image

(b)

(b)

(b)

(c)

(c)

(c)

(c)

(d)
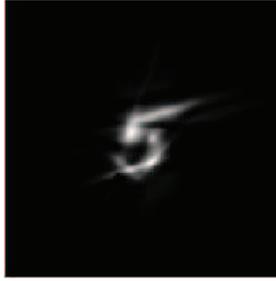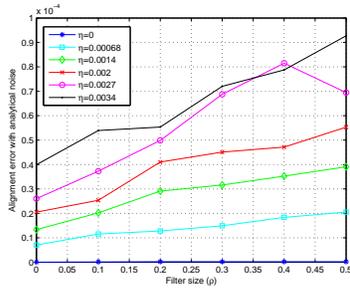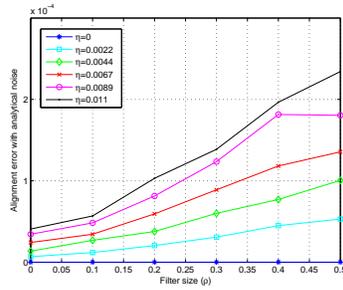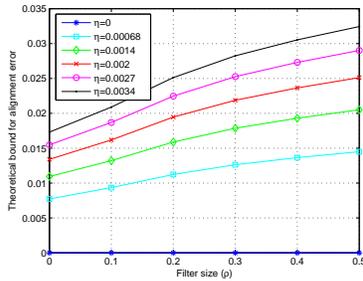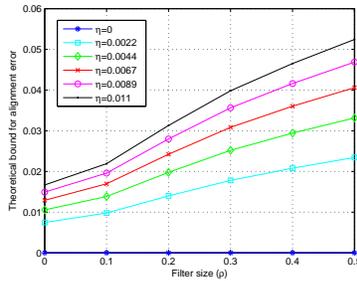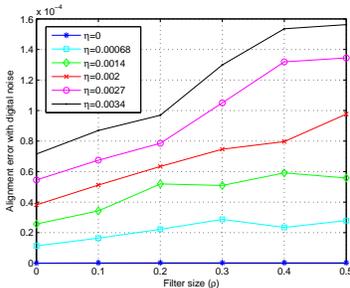
(d)
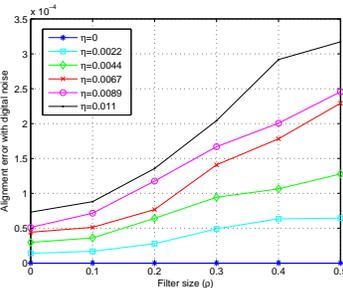
(d)

Figure 5.5. *Face pattern and alignment error as a function of filter size ρ.*

Figure 5.6. *Digit pattern and alignment error as a function of filter size ρ.*
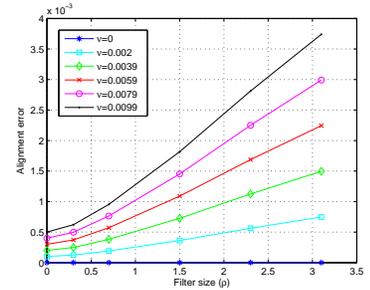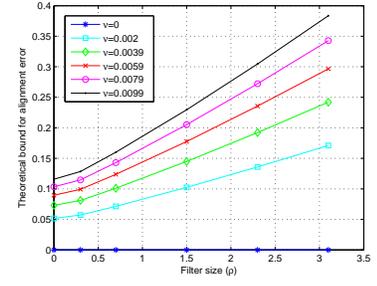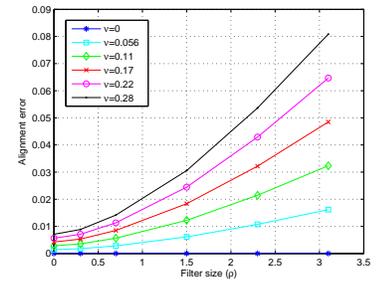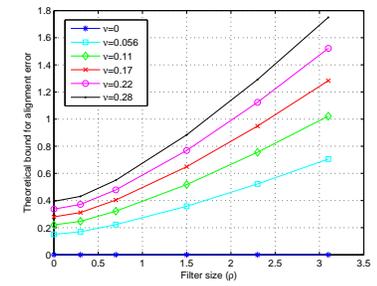
Figure 5.7. *Alignment error for random patterns and generic noise, as a function of filter size ρ. (a) and (c) show the error for noise patterns respectively with high and low correlation with p. Corresponding theoretical bounds R_{u_0} and Q_{u_0} are given respectively in (b) and (d).*

face pattern is $|c| = 0.14$, and the position and scale parameters of the pattern are in the range $[-0.9, 0.9]$ for τ_x, τ_y, and $[0.04, 1.1]$ for σ_x, σ_y. For the digit experiments, the reference pattern shown in Figure 5.6(a) is the approximation of a handwritten "5" digit with 20 Gaussian atoms. The pattern parameters are such that the average atom coefficient magnitude is $|c| = 0.87$, and the position and scale parameters are given by $\tau_x, \tau_y \in [-0.7, 0.7]$, and $\sigma_x, \sigma_y \in [0.05, 1.23]$. The atom parameters for the face and digit images are computed with the matching pursuit algorithm [15]. The range of translation values tT_x and tT_y is selected as $[-1, 1]$ for both settings. In both settings, two different noise models are tested. First, the target patterns are corrupted with respect to the analytical Gaussian noise model w of (4.1), where the noise parameters are set as $L = 750$, $\epsilon = 0.04$. Then, a digital Gaussian noise model is tested, where the pixels in the discrete representation of the images are corrupted with additive i.i.d. Gaussian noise having the same standard deviation η as w. The digital Gaussian noise model is supposed to be well-approximated by the analytical noise model. Again, 50 target patterns are generated with random translations. The alignment errors are plotted with respect to ρ in Figures 5.5 and 5.6 respectively for the face and digit patterns. The experimental error with the analytical noise model, the theoretical upper bound obtained for the analytical noise model, and the experimental error with the digital noise model are plotted respectively in (b), (c) and (d) in both figures. The results are averaged over all target patterns. In the face image experiment, the maximum noise level $\eta = 0.0034$ corresponds to SNRs of 46.8 dB and 34.7 dB for the analytical and digital noise models. In the digit image experiment, the maximum noise level $\eta = 0.011$ yields the respective SNRs of 39.3 dB and 28.1 dB.

The plots in Figures 5.5 and 5.6 show that the experimental and theoretical errors have a similar variation with respect to ρ. The dependence of the error on ρ in these experiments seems to be different from that of the previous experiment of Figure 5.2. Although the theory predicts the variation $\hat{R}_{t_0} = O\left(\rho^{3/2}\left(1 - \rho\right)^{-1/2}\right)$, this result is average and approximate. The exact variation of the error with ρ may change between different individual patterns, as the constants of the variation function are determined by the actual pattern parameters. The similarity between the plots for the analytical and digital noise models suggests that the noise model w used in this study provides a good approximation for the digital Gaussian noise, which is often encountered in digital imaging applications. In [26], the alignment error for face and digit patterns is also plotted with respect to η, and the results are similar to those of the previous experiment with random patterns.

5.2.2. Generic noise model. In the second set of experiments we evaluate the results of Theorem 4.5 and Corollary 4.6 for the generic noise model z. In each experiment, the target patterns are generated by corrupting the reference pattern p with a noise pattern z and by applying random translations in the range $tT_x, tT_y \in [-4, 4]$. In order to study the effect of the correlation between z and the points on $\mathcal{M}(p)$ on the actual alignment error and on its theoretical bound, we consider two different settings. In the first setting, the noise pattern z is chosen as a pattern that has high correlation with p. In particular, z is constructed with a subset of the atoms used in p with the same coefficients. The general bound R_{u_0} is used in this setting. In the second setting, the noise z is constructed with randomly selected Gaussian atoms so that it has small correlation with p. The bound Q_{u_0} for the small correlation

case is used in the second setting, where the correlation parameter r_{pz} in (4.14) is computed numerically for obtaining the theoretical error bound. In both cases, the atom coefficients of z are normalized such that the norm ν of z is below the admissible noise level ν_0. The theoretical bounds[8] are then compared to the experimental errors for different values of the filter size ρ and the noise level ν.

We conduct the experiment on the random patterns used above, in Figures 5.2-5.4. The noise pattern z is constructed with 10 atoms. The average alignment errors are plotted with respect to the filter size ρ in Figure 5.7. The noise levels $\nu = 0.0099$ and $\nu = 0.28$ correspond respectively to the approximate SNRs of 65.7 dB and 36.7 dB. The plots in Figure 5.7 show that the variation of the theoretical upper bounds with ρ fits well the behavior of the actual error in both settings. The results are in accordance with Theorem 4.8, which states that $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ are of $O((1 + \rho^2)^{1/2})$. The results of the experiment show that $\hat{Q}_{u_0}$ is less pessimistic than $\hat{R}_{u_0}$ as an upper bound since it makes use of the information of the maximum correlation between z and the points on $\mathcal{M}(p)$. Moreover, comparing Figures 5.7(b) and 5.7(d) we see that, when a bound r_{pz} on the correlation is known, the admissible noise level increases significantly (from around $\nu_0 = 0.01$ to $\nu_0 = 0.28$). The plots of the errors with respect to ν are also available in [26]; in short, they show that the variation of the alignment error with ν bears resemblance to its variation with η observed in the previous experiment of Figure 5.3. This is an expected result, as it has been seen in Theorem 4.8 that the dependences of $\hat{R}_{u_0}$ and $\hat{Q}_{u_0}$ on ν are the same as the dependence of $\hat{R}_{t_0}$ on η. These plots show also that, at the same noise level, the actual alignment error is slightly smaller when z has small correlation with the points on $\mathcal{M}(p)$. In [26], this experiment is also repeated with the face and digit patterns, with results that are similar to those obtained with random test patterns.

The overall conclusion of the experiments is that increasing the filter kernel size results in a bigger alignment error when the target image deviates from the translation manifold of the reference image due to noise. The results show also that the theoretical bounds for the alignment error capture well the order of dependence of the actual error on the noise level and the filter size, for both the Gaussian noise model and the generalized noise model. Also, the knowledge of the correlation between the noise pattern and the translated versions of the reference pattern is useful for improving the theoretical bound for the alignment error in the general setting. Although the actual values of the theoretical upper bounds are quite pessimistic, they are roughly the same as the actual alignment errors up to a multiplicative factor. In a practical image registration application, the value of this factor can be numerically estimated. The normalization of the theoretical bounds with this factor then yields an accurate model for the multiscale alignment error.

5.3. Application: Design of an optimal registration algorithm. We now demonstrate the usage of our SIDEN estimate in the construction of a grid in the translation parameter domain that is used for image registration. In Section 3.2, we have derived a set $\mathcal{Q}$ of translation vectors that can be correctly computed by minimizing the distance function with descent methods, where $\mathcal{Q}$ is a subset of the SIDEN $\mathcal{S}$ corresponding to the noiseless distance function. As

---

[8]We compute the bound for the second derivative of p numerically by minimizing $\|d^2 p(X + tT)/dt^2\|$ over t and T. While the bound $R_{p''}$ is useful for the theoretical analysis as it has an open-form expression, the numerically computed bound is sharper.

discussed in the beginning of Section 4, in noisy settings, one can assume that $\mathcal{Q}$ is also a subset of the perturbed SIDEN corresponding to the noisy distance function, provided that the noise level is sufficiently small. Therefore the estimate $\mathcal{Q}$ can be used in the registration of both noiseless and noisy images; small translations that are inside $\mathcal{Q}$ can be recovered with gradient descent minimization. However, the perfect alignment guarantee is lost for relatively large translations that are outside $\mathcal{Q}$ and the descent method may terminate in a local minimum other than the global minimum. Hence, in order to overcome this problem, we propose to construct a grid in the translation parameter domain and estimate large translations with the help of the grid. In particular, we describe a grid design procedure such that any translation vector tT lies inside the SIDEN of at least one grid point. Such a grid guarantees the recovery of the translation parameters if the distance function is minimized with a gradient descent method that is initialized with the grid points. In order to have a perfect recovery guarantee, each one of the grid points must be tested. However, as this is computationally costly, we use the following two-stage optimization instead, which offers a good compromise with respect to the accuracy-complexity tradeoff. First, we search for the grid vector that gives the smallest distance between the image pair, which results in a coarse alignment. Then, we refine the alignment with a gradient descent method initialized with this grid vector. In practice, this method is quite likely to give the optimal solution, which has been the case in all of our simulations.

We now explain the construction of the grid. First, notice from (3.8) that $a_{jk}(T) = a_{jk}(-T)$ and $b_{jk}(T) = -b_{jk}(-T)$. Therefore, the function $s_{jk}(t)$ given in (3.18) is the same for T and $-T$ by symmetry. As Q_{jk} does not depend on T, from the form of $df(tT)/dt$ in (3.17) we have $df(tT)/dt = df(-tT)/dt$. Hence, the SIDEN is symmetric with respect to the origin. It is also easy to check that the estimation δ_T of the SIDEN boundary along the direction T satisfies $\delta_T = \delta_{-T}$. One can easily determine a grid unit in the form of a parallelogram that lies completely inside the estimate $\mathcal{Q}$ of the SIDEN and tile the (tT_x, tT_y)-plane with these grid units. This defines a regular grid in the (tT_x, tT_y)-plane such that each point of the plane lies inside the SIDEN of at least one grid point. Note that the complexity of image registration based on a grid search is given by the number of grid points. In our case, the number of grid points is determined by the area of $\mathcal{Q}$; and therefore, the alignment complexity depends on the well-behavedness of the distance function f. In particular, as $V(\mathcal{Q})$ increases with the filter size, the area of the grid units expand at the rate $O(1+\rho^2)$ and the number of grid points decrease at the rate $O\left((1+\rho^2)^{-1}\right)$ with ρ. Therefore, the alignment complexity with the proposed method is of $O\left((1+\rho^2)^{-1}\right)$.

The construction of a regular grid in this manner is demonstrated for the image of the "5" digit used in the experiments of Section 5.2. In Figure 5.8(a), the reference pattern and its translated versions corresponding to the neighboring grid points in the first and second directions of sampling are shown. In Figure 5.8(b), the reference pattern is shown when smoothed with a filter of size $\rho = 0.15$, as well as the neighboring patterns in the smoothed grid. The original grid and the smoothed grid for $\rho = 0.15$ are displayed in Figures 5.9(a) and 5.9(b), where the SIDEN estimates $\mathcal{Q}$, $\hat{\mathcal{Q}}$ and the grid units are also plotted. One can observe that smoothing the pattern is helpful for obtaining a coarser grid that reduces the computational complexity of image registration in hierarchical methods. The corresponding distance functions $f(tT)$ and $\hat{f}(tT)$ are plotted in Figure 5.10, which shows that smoothing

(a) Original pattern and grid images

(b) Smoothed pattern and grid images

Figure 5.8. *Neighboring grid patterns for the original and smoothed images.*
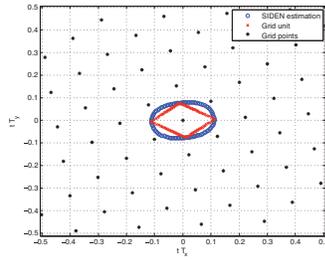
(a) Original grid
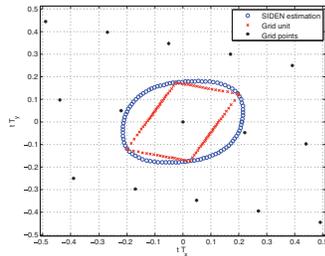
(b) Smoothed grid

Figure 5.9. *Grid construction*
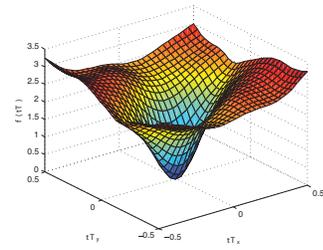
(a) Original distance function $f(tT)$
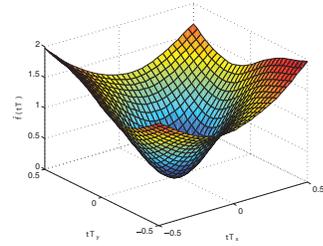
(b) Smoothed distance function $\hat{f}(tT)$

Figure 5.10. *The variation of the distance function with smoothing.*
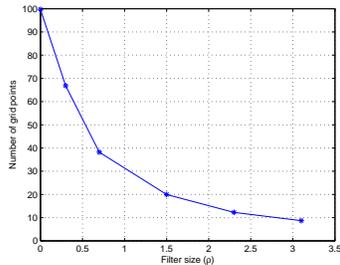
(a) Random patterns
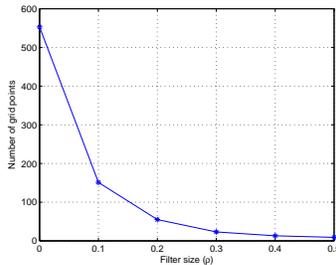
(b) Face pattern
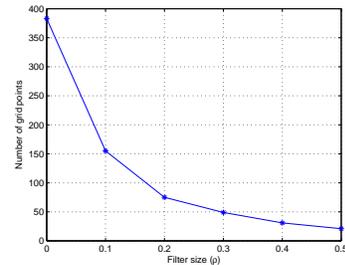
(c) Digit pattern

Figure 5.11. *Number of grid points. The decay rate is of* $O\left((1+\rho^2)^{-1}\right)$.

eliminates undesired local extrema of the distance function and therefore expands the SIDEN.

Then, in order to demonstrate the relation between alignment complexity and filtering, we build multiresolution grids corresponding to different filter sizes and plot the variation of the number of grid points with the filter size. The results obtained with the random patterns and the face and digit patterns used in Section 5.2 are presented in Figure 5.11. The results show that the number of grid points decreases monotonically with the filter size, as predicted by Theorem 3.6, which suggests that the number of grid points must be of $O\left((1+\rho^2)^{-1}\right)$.

The performance guarantee of this two-stage registration approach is confirmed by the experiments of Section 5.2, which use this registration scheme. In the plots of Figures 5.2-5.7, where the coarse alignment in each experiment is done with the help of a grid adapted to the filter size using the grid design procedure explained above, we see that the proposed registration technique results in an alignment error of 0 for the noiseless case ($\eta = 0$ or $\nu = 0$) for all values of the filter size ρ. These alignment error results, together with the grid size plots of Figure 5.11, show that increasing the filter size reduces the alignment complexity while retaining the perfect alignment guarantee in the noiseless case. Figures 5.2-5.7 show also that the proposed grid can be successfully used in noisy settings. The alignment error in these experiments stems solely from the change in the global minimum of the distance function due to noise, and not from the grid; otherwise, we would observe much higher alignment errors that are comparable to the distance between neighboring grid points. This confirms that the estimate $\mathcal{Q}$ remains in the perturbed SIDEN and its usage does not lead to an additional alignment error if the noise level is relatively small.

Finally, we note that, due to the computations needed for obtaining the SIDEN estimates, the construction of an optimal grid that is specifically adapted to a reference image is favorable especially in an application where the reference image is fixed and various target images are to be aligned with it. In an application where target images are aligned with different reference images, a more practical solution may be preferred by constructing a non-adaptive, "average" grid. One may, for instance, compute the SIDEN estimates for some typical patterns that have similar properties (e.g. frequency characteristics, size, support) with the reference images at hand and construct a non-adaptive grid based on these typical SIDEN estimates. The multiscale version of this grid can then be computed by adjusting the grid density to match a rate of $O((1 + \rho^2)^{-1})$ as discussed previously.

6. Discussion of Results. The results of our analysis show that smoothing improves the regularity of alignment by increasing the range of translation values that are computable with descent-type methods. However, in the presence of noise, smoothing has a negative influence on the accuracy of alignment as it amplifies the alignment error caused by the image noise; and this increases with the increase in the filter size. Therefore, considering the computation cost - accuracy tradeoff, the optimal filter size in image alignment with descent methods must be chosen by taking into account the deviation between the target pattern and the translation manifold of the reference pattern; i.e., the expected noise level.

Our study constitutes a theoretical justification of the principle behind hierarchical registration techniques that use local optimizers. Coarse scales are favorable at the beginning of the alignment as they permit the computation of large translation amounts with low complexity using simple local optimizers; however, over-filtering decreases the accuracy of alignment as the target image is in general not exactly a translated version of the reference image. This is compensated for at finer scales where less filtering is applied, thus avoiding the amplification of the alignment error resulting from noise. Since images are already roughly registered at coarse scales, at fine scales the remaining increment to be added to the translation parameters for fine-tuning the alignment is small; it can be achieved at a relatively low complexity in a small search region.

In the following, we first take these observations one step further to sketch some rules for

designing a coarse-to-fine registration algorithm that recovers the correct translation parameters. Next, we discuss the links between our results and previous works.

6.1. Design of hierarchical registration algorithms. Consider a descent-type coarse-to-fine registration method that computes the translation parameters by aligning smoothed versions of the image pair with a series of low-pass filters of size $\rho_1, \rho_2, \dots, \rho_n$ such that the initial solution in stage k is taken as the estimate obtained in the previous stage $k-1$. Let $t_* T_*$ denote the optimal translation vector that best aligns the reference image and the target image; i.e., the target image is a noisy observation of the translated version $p(X - t_* T_*)$ of the reference image $p(X)$. We now state some conditions under which such a registration algorithm converges for the Gaussian noise model and the generic noise model.

Gaussian noise model. Let the noise level η be upper bounded inversely proportionally to the amount of translation t_* between the reference and target images

$$\eta \leq O\left(\frac{1}{t_*}\right).$$

Then, if the size ρ_1 of the filter in the first stage is chosen proportionally to the amount of translation t_* such that

$$\rho_1 = O\left(\sqrt{t_*^2 - 1}\right) \approx O(t_*)$$

and if the filter sizes ρ_{k-1}, ρ_k in adjacent stages are selected such that

$$\rho_k = O\left(\sqrt{\frac{\eta \, \rho_{k-1}^3}{1 - \eta \, \rho_{k-1}} - 1}\right) \tag{6.1}$$

then the hierarchical registration algorithm converges to the correct solution $t_* T_*$.

Generic noise model. Assume that the noise level ν is below some sufficiently small constant threshold. If the filter size ρ_1 in the first stage is chosen proportionally to t_* such that

$$\rho_1 = O\left(\sqrt{t_*^2 - 1}\right) \approx O(t_*)$$

and if the filter sizes in adjacent stages are selected such that

$$\rho_k = O\left(\sqrt{\left(\frac{\nu}{1 - \nu}\right)(1 + \rho_{k-1}^2) - 1}\right) \tag{6.2}$$

then the hierarchical registration algorithm converges to the correct solution $t_* T_*$.

These results are derived in [26, Appendix F]. The proposed strategies to select the filter size ρ_k for both noise models ensure that the SIDEN in stage k is sufficiently large to correct the alignment error of the previous stage $k-1$. The bounds on the noise levels then guarantee that the alignment error decays gradually to zero during the hierarchical alignment.

It is interesting to compare the conditions (6.1) and (6.2) to common filter selection strategies used in practice. The filter size update relations in (6.1) and (6.2) can be approximated[9] in the form $\rho_k \approx C\,\rho_{k-1}$. This is consistent with the practical filter size selection strategy $\rho_k = 1/2\,\rho_{k-1}$, which is used quite commonly in hierarchical registration and optical flow estimation algorithms. Our results imply, however, that the noise level must be taken into account in the selection of the filter size reduction factor C, such that C should be set to a larger value if the noise level is higher.

6.2. Comparison of results with previous studies. We now interpret the findings of our work in comparison with some previous results. We start with the article [21] by Robinson et al., which studies the Cramér-Rao lower bound (CRLB) of the registration. Since the CRLB is related to the inverse of the Fisher information matrix (FIM) J, the authors suggest to use the trace of J^{-1} as a general lower bound for the MSE of the translation estimation. Therefore, the square root of $tr(J^{-1})$ can be considered as a lower bound for the alignment error. It has been shown in [21] that $\sqrt{tr(J^{-1})} = O(\eta)$, where η is the standard deviation of the Gaussian noise. In fact, this result tells that the alignment error with any estimator is lower bounded by $O(\eta)$, i.e., its dependence on the noise level is at least linear. Meanwhile, our study, which focuses on estimators that minimize the SSD with local optimizers, concludes that the alignment error is at most $O(\sqrt{\eta/(1-\eta)})$ for these estimators. Notice that for small η, $O(\sqrt{\eta/(1-\eta)}) \approx O(\sqrt{\eta}) > O(\eta)$; and for large η, we still have $O(\sqrt{\eta/(1-\eta)}) > O(\eta)$ due to the sharply increasing rational function form of the bound. Therefore, the result in [21] and our results are consistent and complementary, pointing together to the fact that the error of an estimator performing a local optimization of the SSD must lie between $O(\eta)$ and $O(\sqrt{\eta/(1-\eta)})$. Note also that, as it has been seen in the experiments of Figure 5.4(a), the error of this type of estimators may indeed increase with η at a rate above $O(\eta)$ in practice, as predicted by our upper bound. Next, as for the effect of filtering on the estimation accuracy, the authors of [21] experimentally observe that $tr(J^{-1})$ decreases as the image bandwidth increases, which suggests that the lower bound on the MSE of a translation estimator is smaller when the image has more high-frequency components. This is stated more formally in [20]. It is shown that the estimation of the x component of the translation has variance larger than $\eta^2/\|(\partial_x p(\cdot))^2\|^2$, and similarly for the y component, where $\partial_x p(X) = \partial p(X)/\partial x$ is the partial derivative of the pattern p with respect to the spatial variable x.[10] Therefore, as smoothing decreases the norm of the partial derivatives of the pattern, it leads to an increase in the variance of the estimation. These observations are also consistent with our theoretical results.

Next, we discuss some results from the recent article [22], which presents a continuous-domain noise analysis of block-matching. The blocks are assumed to be corrupted with additive Gaussian noise and the disparity estimate is given by the global minimum of the noisy

---

[9]The more precise condition on the noise level $\eta \leq O(\alpha^2/(1+\alpha^2)\,t_*^{-1})$ derived in [26, Appendix F] implies that the filter size in (6.1) satisfies $\rho_k \leq O(\sqrt{\alpha^2\rho_{k-1}^2 - 1})$. Therefore, the selection $\rho_k = O(\sqrt{\alpha^2\rho_{k-1}^2 - 1}) \approx C\,\rho_{k-1}$ assures that the SIDEN at stage k is sufficiently large, while it can also be shown to retain the convergence guarantee.

[10]This bound is obtained by assuming Gaussian noise on both reference and target patterns and employing CRLB.

distance function as in our work. Although there are differences in their setting and ours, such as the horizontal and non-constant disparity field assumption in [22], it is interesting to compare their results with ours. We consider the disparity of the block to be constant in [22], such that it fits our global translation assumption. In [22], an analysis of the deviation between the estimated disparity and the true disparity is given, which is similar to the distance t_0 between the global minima of f and g in our work. Sticking to the notation of our text, let us denote this deviation by t_0. In Theorem 3.2 of [22], t_0 is estimated as the sum of three error terms, where the variances of the first and second terms are respectively of $O(\eta^2)$ and $O(\eta^4)$ with respect to the noise standard deviation η. These two terms are stated as dominant noise terms. Then, the third term represents the high-order Taylor terms of some approximations made in the derivations, which however depends on the value of t_0 itself. As the overall estimation of t_0 is formulated using the term t_0 itself, their main result is interesting especially for small values of t_0, since the third term is then negligible. It can be concluded from [22] that $t_0 \approx O(\eta + \eta^2 + H.O.)$, where $H.O.$ represents high-order terms. This result is consistent with the CRLB of t_0 in [21] stating that t_0 is at least of $O(\eta)$, and our upper bound $O(\sqrt{\eta}/\sqrt{1-\eta})$. The analysis in [22] and ours can be compared in the following way. First, since our derivation is rather rigorous and does not neglect high-order terms, these terms manifest themselves in the rational function form of the resulting bound. Meanwhile, they are represented as $H.O.$ and not explicitly examined in the estimation $O(\eta + \eta^2 + H.O.)$ in [22]. For small η, this result can be approximated as $t_0 = O(\eta)$, while our result states that $t_0 \leq O(\sqrt{\eta})$. As $\sqrt{\eta} > \eta$ for small η, this also gives a consistent comparison since our estimation is an upper bound and the one in [22] is not. Indeed, the experimental results in [22] suggest that their derivation gives a slight underestimation of the error. Lastly, our noise analysis treats the image alignment problem in a multiscale setting and analyzes the joint variation of the error with the noise level η and the filter size ρ, whereas the study in [22] only concentrates on the relation between the error and the noise level.

Finally, we mention some facts from scale-space theory [12], which may be useful for the interpretation of our findings regarding the variation of the SIDEN with the filter size ρ. The scale-space representation of a signal is given by convolving it with kernels of variable scale. The most popular convolution kernel is the Gaussian kernel, as it has been shown that under some "well-behavedness" constraints, it is the unique kernel for generating a scale-space. An important result in scale-space theory is [30], which states that the number of local extrema of a 1-D function is a decreasing function of ρ. This provides a mathematical characterization of the well-known smoothing property of the Gaussian kernel. However, it is known that this cannot be generalized to higher-dimensional signals; e.g., there are no nontrivial kernels on $\mathbb{R}^2$ with the property of never introducing new local extrema when the scale increases [12]. One interesting result that can possibly be related to our analysis is about the density of local extrema of a signal as a function of scale. In order to gain an intuition about the behavior of local extrema, the variation of the local extrema is examined in [12] for 1-D continuous white noise and fractal noise processes. It has been shown that the expected density of the local minima of these signals decreases at rate ρ^{-1}. In the estimation of the SIDEN in our work, we have analyzed how the first zero crossing of $d\hat{f}(tT)/dt$ along a direction T around the origin varies with the scale. Therefore, what we have examined is the distance $\hat{f}$ between the scale space $\hat{p}(X)$ of an image and the scale space $\hat{p}(X - tT)$ of its translated version. Since this is

different from the scale-space of the distance function f itself, it is not possible to compare the result in [12] directly to ours. However, we can observe the following. Restricting $\hat{f}$ to a specific direction T_a so that we have a 1-D function $\hat{f}(tT_a)$ of t as in [12], our estimation for the stationary point of $\hat{f}(tT_a)$ closest to the origin expands at a rate of $O((1 + \rho^2)^{1/2})$, which is $O(\rho)$ for large ρ. One can reasonably expect this distance to be roughly inversely proportional to the density of the local extrema of $\hat{f}$. This leads to the conclusion that the density of the distance extrema is expected to be around $O(\rho^{-1})$, which interestingly matches the density obtained in [12].

7. Conclusion. We have presented a theoretical analysis of image alignment with descent-type local minimizers, where we have specifically focused on the effect of low-pass filtering and noise on the regularity and accuracy of alignment. First, we have examined the problem of aligning with gradient descent a reference and a target pattern that differ by a two-dimensional translation. We have derived a lower bound for the range of translations for which the reference pattern can be exactly aligned with its translated versions, and investigated how this region varies with the filter size when the images are smoothed. Our finding is that the volume of this region increases quadratically with the filter size, showing that smoothing the patterns improves the regularity of alignment. Then, we have considered a setting with noisy target images and examined Gaussian noise and arbitrary noise patterns, which may find use in different imaging applications. We have derived a bound for the alignment error and searched the dependence of the error on the noise level and the filter size. Our main results state that the alignment error bound is proportional to the square root of the noise level at small noise, whereas this order of dependence increases at larger noise levels. More interestingly, the alignment error is also significantly affected by the filter size. The probabilistic error bound obtained with the Gaussian noise model has been seen to increase with the filter size at a sharply increasing rational function rate, whereas the deterministic bound obtained for arbitrary noise patterns of deterministic norm increases approximately linearly with the filter size. These theoretical findings are also confirmed by experiments. To the best of our knowledge, none of the previous works about image registration has studied the alignment regularity problem. Meanwhile, our alignment accuracy analysis is consistent with previous results, provides a more rigorous treatment, and studies the problem in a multiscale setting unlike the previous works. The results of our study show that, in multiscale image registration, filtering the images with large filter kernels improves the alignment regularity in early phases, while the use of smaller filters improves the accuracy at later phases. From this aspect, our estimations of the regularity and accuracy of alignment in terms of the noise and filter parameters provide insight for the principles behind hierarchical registration techniques and may find use in the design of efficient, low-complexity registration algorithms.

Appendix A. Exact expressions for $\overline{\mathbb{c}}_{jk}$ and $\underline{\mathbb{d}}_{jk}$.

Here we present the expressions for the terms $\overline{\mathbb{c}}_{jk}$ and $\underline{\mathbb{d}}_{jk}$ used in Lemma 4.1. Let $\alpha_k = \lambda_{\min}(\Phi_k)$ and $\beta_k = \lambda_{\max}(\Phi_k)$ denote respectively the smaller and greater eigenvalues of Φ_k.

Since Φ_k is a positive definite matrix, $\beta_k \geq \alpha_k > 0$. Let $\tau_k = [\tau_{x,k} \ \tau_{y,k}]^T$,

$$\mathfrak{b}_{jk}^x = (\beta_j + \beta_k) \max\left\{ \left(b + \overline{t}_0 - \frac{\beta_j \tau_{x,j} + \beta_k \tau_{x,k}}{\beta_j + \beta_k}\right)^2, \left(-b - \overline{t}_0 - \frac{\beta_j \tau_{x,j} + \beta_k \tau_{x,k}}{\beta_j + \beta_k}\right)^2\right\}$$

$$\mathfrak{c}_{jk}^x = (\beta_j + \beta_k) \max\left\{ \left(b + \overline{t}_0 \frac{\beta_j}{\beta_j + \beta_k} - \frac{\beta_j \tau_{x,j} + \beta_k \tau_{x,k}}{\beta_j + \beta_k}\right)^2, \left(-b - \overline{t}_0 \frac{\beta_j}{\beta_j + \beta_k} - \frac{\beta_j \tau_{x,j} + \beta_k \tau_{x,k}}{\beta_j + \beta_k}\right)^2\right\}$$

$$\mathfrak{d}_{jk}^x = \frac{\beta_j \beta_k}{\beta_j + \beta_k} \max\left\{(-\overline{t}_0 + \tau_{x,k} - \tau_{x,j})^2, (\overline{t}_0 + \tau_{x,k} - \tau_{x,j})^2\right\}$$

and $\mathfrak{b}_{jk}^y$, $\mathfrak{c}_{jk}^y$ and $\mathfrak{d}_{jk}^y$ be defined similarly by replacing x with y in the above expressions. Let

$$\underline{H}_{jk}^x = -\frac{\beta_j \tau_{x,j} + \beta_k \tau_{x,k}}{\beta_j + \beta_k}, \qquad \underline{G}_{jk}^x = \frac{\beta_j \tau_{x,j}^2 + \beta_k \tau_{x,k}^2}{\beta_j + \beta_k}$$

and $\underline{H}_{jk}^y$, $\underline{G}_{jk}^y$ be defined similarly by replacing x with y in the above expressions. Also, let $\overline{H}_{jk}^x$, $\overline{H}_{jk}^y$, $\overline{G}_{jk}^x$, $\overline{G}_{jk}^y$ denote the terms obtained by substituting β_j, β_k with α_j, α_k in the expressions of respectively $\underline{H}_{jk}^x$, $\underline{H}_{jk}^y$, $\underline{G}_{jk}^x$, $\underline{G}_{jk}^y$. [11] Then, let

$$\underline{\mathcal{D}}_{jk}^x = \frac{\sqrt{\pi}}{4b} \frac{1}{\sqrt{\beta_j + \beta_k}} \exp\left(-(\beta_j + \beta_k)\left(\underline{G}_{jk}^x - (\underline{H}_{jk}^x)^2\right)\right)$$
$$\cdot \left[\text{erf}\left(\sqrt{\beta_j + \beta_k}\,(b + \underline{H}_{jk}^x)\right) - \text{erf}\left(\sqrt{\beta_j + \beta_k}\,(-b + \underline{H}_{jk}^x)\right)\right]$$

and $\underline{\mathcal{D}}_{jk}^y$ be defined similarly, and let us denote by $\overline{\mathcal{D}}_{jk}^x$, $\overline{\mathcal{D}}_{jk}^y$ the terms obtained by replacing β_j, β_k with α_j, α_k in the expressions of $\underline{\mathcal{D}}_{jk}^x$, $\underline{\mathcal{D}}_{jk}^y$. Then defining

$$\underline{\mathcal{B}}_{jk} = \exp\left(-\mathfrak{b}_{jk}^x - \mathfrak{b}_{jk}^y - \frac{\beta_j \beta_k \|\tau_k - \tau_j\|^2}{(\beta_j + \beta_k)}\right), \qquad \overline{\overline{\mathcal{B}}}_{jk} = \frac{\pi}{4b^2 (\alpha_j + \alpha_k)} \exp\left(-\frac{\alpha_j \alpha_k \|\tau_k - \tau_j\|^2}{(\alpha_j + \alpha_k)}\right)$$

$$\underline{\mathcal{C}}_{jk} = \exp\left(-\mathfrak{c}_{jk}^x - \mathfrak{c}_{jk}^y - \mathfrak{d}_{jk}^x - \mathfrak{d}_{jk}^y\right), \quad \overline{\overline{\mathcal{C}}}_{jk} = \frac{\pi}{4b^2 (\alpha_j + \alpha_k)}, \quad \underline{\mathcal{D}}_{jk} = \underline{\mathcal{D}}_{jk}^x \underline{\mathcal{D}}_{jk}^y, \quad \overline{\mathcal{D}}_{jk} = \overline{\mathcal{D}}_{jk}^x \overline{\mathcal{D}}_{jk}^y$$

the parameters $\overline{c}_{jk}$ and $\underline{d}_{jk}$ are given by $\overline{c}_{jk} = \overline{\overline{\mathcal{B}}}_{jk} - 2\underline{\mathcal{C}}_{jk} + \overline{\mathcal{D}}_{jk}$, $\underline{d}_{jk} = \underline{\mathcal{B}}_{jk} - 2\overline{\overline{\mathcal{C}}}_{jk} + \underline{\mathcal{D}}_{jk}$.

Appendix B. Exact expressions for $\underline{r}_0$, $\underline{r}_2$ and $\underline{r}_3$.

We now give the expressions for the terms $\underline{r}_0$, $\underline{r}_2$ and $\underline{r}_3$ used in Lemma 4.2. Let $\underline{r}_0$ be the smaller eigenvalue of the following positive definite matrix

$$R_0 = \sum_{j=1}^{K} \sum_{k=1}^{K} c_j c_k\, Q_{jk} \left(\Sigma_{jk}^{-1} - \Sigma_{jk}^{-1}(\tau_k - \tau_j)(\tau_k - \tau_j)^T \Sigma_{jk}^{-1}\right).$$

Next, as shown in [26], the following upper and lower bounds can be obtained

$$a_{jk}^2 \geq \underline{a}_{jk}^2 = \frac{1}{4}\lambda_{\min}^2(\Sigma_{jk}^{-1}), \qquad a_{jk}^2 \leq \overline{a}_{jk}^2 = \frac{1}{4}\lambda_{\max}^2(\Sigma_{jk}^{-1})$$

$$b_{jk}^2 a_{jk} \leq \overline{b}_{jk}^2 \overline{a}_{jk} = \frac{1}{8}\lambda_{\max}(R_2^{jk})\lambda_{\max}(\Sigma_{jk}^{-1}), \qquad b_{jk}^4 \leq \overline{b}_{jk}^4 = \frac{1}{16}\lambda_{\max}^2(R_2^{jk})$$

---

[11] The terms written as $\underline{H}_{jk}^x, \ldots, \overline{G}_{jk}^y$ here correspond to the terms $\underline{H}_{jk}^x(0,0), \ldots, \overline{G}_{jk}^y(0,0)$ in [26].

where $R_2^{jk} = \Sigma_{jk}^{-1}(\tau_k - \tau_j)(\tau_k - \tau_j)^T \Sigma_{jk}^{-1}$. Then, we have

$$\underline{r}_2' = \sum_{(j,k)\in J^+} c_j c_k Q_{jk}(-8\,\overline{b}_{jk}^4 + -6\,\overline{a}_{jk}^2) \quad + \sum_{(j,k)\in J^-} c_j c_k Q_{jk}(24\,\overline{b}_{jk}^2\,\overline{a}_{jk} - 6\,\underline{a}_{jk}^2)$$

and $\underline{r}_2$ is given by $\underline{r}_2 = \min(\underline{r}_2', 0)$. Finally,

$$\underline{r}_3 = -\sum_{j=1}^{K}\sum_{k=1}^{K}\frac{5.46}{2^{5/2}}\,|c_j c_k|\,\frac{\pi|\sigma_j \sigma_k|}{\sqrt{|\Sigma_{jk}|}}\,(\lambda_{\max}(\Sigma_{jk}^{-1}))^{5/2}.$$

Appendix C. Exact expressions for $\overline{\mathbb{e}}_{jk}$ and $\underline{\mathbb{f}}_{jk}$.
Now we present the terms $\overline{\mathbb{e}}_{jk}$ and $\underline{\mathbb{f}}_{jk}$ used in Lemma 4.3. Let

$$\overline{L}_j = \left(\frac{(3^{3/4} + 3^{5/4})e^{-\frac{\sqrt{3}}{2}}}{16} + \frac{3\sqrt{\pi}}{2^{9/2}}\right)\frac{1}{\alpha_j^{5/2}}, \quad \overline{M}_j = \sqrt{\frac{\pi}{2\alpha_j}}, \quad \overline{N}_j := \left(\frac{e^{-\frac{1}{2}}}{4} + \frac{\sqrt{\pi}}{2^{5/2}}\right)\frac{1}{\alpha_j^{3/2}}$$

$$\overline{\overline{\mathcal{E}}}_j = \frac{\beta_j^4}{4b^2}\left(2\overline{L}_j\overline{M}_j + 2\overline{N}_j^2\right), \qquad \overline{\overline{\mathcal{F}}}_j = \frac{1}{4b^2}\overline{M}_j^2.$$

Then, the terms $\overline{\mathbb{e}}_{jk}$ and $\underline{\mathbb{f}}_{jk}$ are given by

$$\overline{\mathbb{e}}_{jk} = 16\sqrt{\overline{\overline{\mathcal{E}}}_j\overline{\overline{\mathcal{E}}}_k} + 4\beta_j\beta_k\,\overline{\overline{\mathcal{B}}}_{jk}, \qquad \underline{\mathbb{f}}_{jk} = -8\,\beta_k\sqrt{\overline{\overline{\mathcal{E}}}_j\overline{\overline{\mathcal{F}}}_k} - 8\,\beta_j\sqrt{\overline{\overline{\mathcal{F}}}_j\overline{\overline{\mathcal{E}}}_k} + 4\alpha_j\alpha_k\,\underline{\underline{\mathcal{B}}}_{jk}.$$

REFERENCES

[1] L. ALVAREZ, J. WEICKERT, AND J. SÁNCHEZ, *A scale-space approach to nonlocal optical flow calculations*, in Proc. 2nd Inter. Conf.on Scale-Space Theories in Computer Vision, SCALE-SPACE '99, London, UK, UK, 1999, Springer-Verlag, pp. 235–246.

[2] P. ANANDAN, *A computational framework and an algorithm for the measurement of visual motion*, International Journal of Computer Vision, 2 (1989), pp. 283–310.

[3] J. ANTOINE, R. MURENZI, P. VANDERGHEYNST, AND S. ALI, *Two-Dimensional Wavelets and their Relatives*, Signal Processing, Cambridge University Press, 2004.

[4] 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.

[5] J. R. BERGEN, P. ANANDAN, K. J. HANNA, AND R. HINGORANI, *Hierarchical model-based motion estimation*, Springer-Verlag, 1992, pp. 237–252.

[6] 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.

[7] L. G. BROWN, *A survey of image registration techniques*, ACM Comput. Surv., 24 (1992), pp. 325–376.

[8] C. CAFFORIO AND F. ROCCA, *The differential method for image motion estimation.*, Springer-Verlag, 1983.

[9] 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.

[10] 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).

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

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

[13] 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.

[14] J. MAINTZ AND M. VIERGEVER, *A survey of medical image registration*, Medical Image Analysis, 2 (1998), pp. 1–36.

[15] S. G. MALLAT AND Z. ZHANG, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Process., 41 (1993), pp. 3397–3415.

[16] E. MÉMIN AND P. PÉREZ, *Hierarchical estimation and segmentation of dense motion fields*, International Journal of Computer Vision, 46 (2002), pp. 129–155.

[17] O. NESTARES AND D. J. HEEGER, *Robust multiresolution alignment of MRI brain volumes*, Magnetic Resonance in Medicine, 43 (2000), pp. 705–715.

[18] A. N. NETRAVALI AND J. D. ROBBINS, *Motion-compensated television coding: Part I*, Bell System Tech. Jour., 58 (1979), pp. 631–670.

[19] Y.C. PATI, R. REZAIIFAR, AND P. S. KRISHNAPRASAD, *Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition*, in 1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993, pp. 40–44.

[20] 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.

[21] D. ROBINSON AND P. MILANFAR, *Fundamental performance limits in image registration*, IEEE Trans. Img. Proc., 13 (2004), pp. 1185–1199.

[22] N. SABATER, J. M. MOREL, AND A. ALMANSA, *How accurate can block matches be in stereo vision?*, SIAM Journal on Imaging Sciences, 4 (2011), pp. 472–500.

[23] P. SIMARD, Y. LECUN, J. S. DENKER, AND B. VICTORRI, *Transformation invariance in pattern recognition-tangent distance and tangent propagation*, in Neural Networks: Tricks of the Trade, New York: Springer-Verlag, 1998.

[24] G. TZIRITAS AND C. LABIT, *Motion Analysis for Image Sequence Coding*, Elsevier Science Inc., New York, NY, USA, 1994.

[25] N. VASCONCELOS AND A. LIPPMAN, *A multiresolution manifold distance for invariant image similarity*, IEEE Transactions on Multimedia, 7 (2005), pp. 127–142.

[26] E. VURAL AND P. FROSSARD, *Analysis of descent-based image registration.* Available at: http://infoscience.epfl.ch/record/183845.

[27] ——, *Learning Smooth Pattern Transformation Manifolds*, IEEE Transactions on Image Processing, 22 (2013), pp. 1311–1325.

[28] D. WALKER AND K. RAO, *Improved pel-recursive motion compensation*, IEEE Trans. Communications, 32 (1984), pp. 1128 – 1134.

[29] M. P. WAND AND M. C. JONES, *Kernel Smoothing*, Chapman and Hall/CRC, 1995.

[30] A. P. WITKIN, *Scale-space filtering.*, in 8th Int. Joint Conf. Artificial Intelligence, vol. 2, Karlsruhe, Aug. 1983, pp. 1019–1022.

[31] M. XU, H. CHEN, AND P. K. VARSHNEY, *Ziv-zakai bounds on image registration*, IEEE Trans. Signal Proc., 57 (2009), pp. 1745–1755.

[32] İ. Ş. YETIK AND A. NEHORAI, *Performance bounds on image registration*, IEEE Trans. Signal Proc., 54 (2006), pp. 1737 – 1749.

[33] B. ZITOVÁ AND J. FLUSSER, *Image registration methods: a survey*, Image and Vision Computing, 21 (2003), pp. 977–1000.