ANALYSIS OF HIERARCHICAL IMAGE ALIGNMENT WITH DESCENT METHODS

Elif Vural and Pascal Frossard

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

ABSTRACT

We present a performance analysis for image registration with gradient descent methods. We consider a 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 their intensity functions with gradient descent. We focus in particular on the effect of low-pass filtering on the alignment performance. We adopt an analytic representation for images and analyze the well-behavedness of the distance function by estimating the neighborhood of translations for which the distance function is free of undesired local minima. This corresponds to the set of translation vectors that are correctly computable with a simple gradient descent minimization. We show that the area of this neighborhood increases at least quadratically with the filter size, which justifies the use of smoothing in image registration with local optimizers. We finally use our results in the design of a regular multiscale grid in the translation parameter domain that has perfect alignment guarantees.

***Keywords*—** Image registration, image smoothing, gradient-descent, performance analysis.

1. INTRODUCTION

The estimation of the transformation that best aligns two images is one of the important problems of image processing. The necessity for registering images arises in many different applications; e.g., image analysis and classification [1], [2], stereo vision [3], motion estimation for video coding [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.

In this work, we analyze the effect of smoothing on the performance of registration. It is commonly admitted that smoothing an image pair is helpful for overcoming the undesired local minima of the distance function between images. In practice, filtering is commonly used in hierarchical registration and motion estimation methods [4]. However, to the best of our knowledge, the analytical relation between filtering and the well-behavedness of the image dissimilarity function has not been extensively studied. Most theoretical results in the image registration literature investigate how image noise affects the registration accuracy, e.g., [5], [6]. However, the analysis of the effect of smoothing on the registration performance has generally been given less attention in the literature. Some of the existing works examine how smoothing influences the bias on the registration with gradient-based methods [5], [7]. Also, there are some results in scale-space theory that examine the variation of the local minima of 1-D and 2-D functions with filtering [8], which however does not exactly have the same setting as in the image registration problem.

In this paper, we consider a setting where the geometric transformation between the reference and target patterns is a global 2-D translation. In particular, we examine the neighborhood of translation vectors in which the only local minimum of the distance function is also the global minimum in the alignment problem. This neighborhood defines the translations between a pair of images, which can be estimated correctly with a descent algorithm. We call this neighborhood the Single Distance Extremum Neighborhood (SIDEN) of the reference pattern. For the ease of derivations, we formulate the registration problem in the continuous domain of square-integrable functions $L^2(\mathbb{R}^2)$ and adopt an analytic and parametric model for the reference and target patterns. We derive an analytic estimation of the SIDEN in terms of the pattern parameters. Then, in order to study the effect of smoothing on the registration performance, we consider the alignment 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 filter size ρ. 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 well-behavedness of the distance function. Finally, we demonstrate the usage of our SIDEN estimate in sampling the translation parameter domain to construct a grid such that any translation between the image pair can be exactly recovered by locating the closest solution on the grid and then locally refining this estimation with a descent method. This can be achieved by adjusting the grid units with respect to the SIDEN of the pattern.

2. IMAGE REGISTRATION ANALYSIS

2.1. Notation and Problem Formulation

Let $p \in L^2(\mathbb{R}^2)$ be a visual pattern. In order to study the image registration problem analytically, we adopt a representation of p in an analytic and parametric dictionary manifold

$$\mathcal{D} = \{\phi_\gamma : \gamma = (\psi, \tau_x, \tau_y, \sigma_x, \sigma_y) \in \Gamma\} \subset L^2(\mathbb{R}^2). \qquad (1)$$

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. Defining the spatial coordinate variable $X = [x\ y]^T \in \mathbb{R}^{2\times 1}$, we will refer to the mother function as $\phi(X)$. Then an atom ϕ_γ is given by $\phi_\gamma(X) = \phi(\sigma^{-1}\,\Psi^{-1}\,(X-\tau))$, where

$$\sigma = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix},\ \Psi = \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix},\ \tau = \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix}. \qquad (2)$$

This work has been partly funded by the Swiss National Science Foundation under Grant 200020_132772.

It is shown in [9] (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 (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

$$p(X) \approx \sum_{k=1}^{K} \lambda_k \, \phi_{\gamma_k}(X) \tag{3}$$

where K is the number of atoms used in the representation of p, γ_k are the atom parameters and λ_k are the atom coefficients.

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. We consider the squared-distance between the reference pattern $p(X)$ and its translated version $p(X - tT)$, which 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

$$f(tT) = \|p(X) - p(X - tT)\|^2 = \int_{\mathbb{R}^2} (p(X) - p(X - tT))^2 dX \tag{4}$$

where the notation[1] $\|.\|$ stands for the L^2-norm for vectors in $L^2(\mathbb{R}^2)$ and the ℓ^2-norm for vectors in $\mathbb{R}^2$.

The global minimum of f is at the origin $tT = 0$. Therefore, there exists an open neighborhood of 0 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 1. *We call the set of translation vectors*

$$\mathcal{S} = \{0\} \cup \{\omega_T T : T \in S^1, \; \omega_T > 0, \; \textit{and} \; \frac{df(tT)}{dt} > 0 \; \textit{for all} \; 0 < t \leq \omega_T\} \tag{5}$$

the Single Distance Extremum Neighborhood (SIDEN) of 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. 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.

Given a reference pattern p, we would like now 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 not only provides 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 1.

[1]Since it is clear from the context which one of these norms is meant, we denote these two norms in the same way for simplicity of notation.

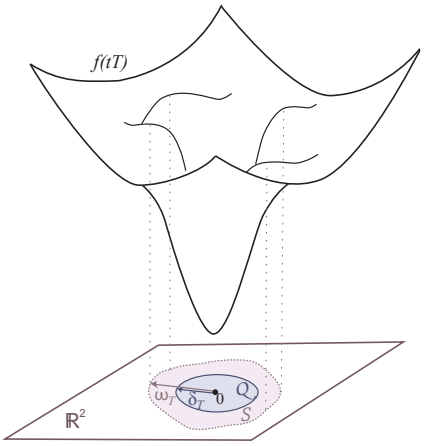


Fig. 1. SIDEN $\mathcal{S}$ is the largest open neighborhood around the origin within which the distance f is increasing along all rays starting out from the origin. Along each unit direction T, $\mathcal{S}$ covers points $\omega_T T$ such that $f(tT)$ is increasing between 0 and $\omega_T T$. The estimate $\mathcal{Q}$ of $\mathcal{S}$ is obtained by computing a lower bound δ_T for the first zero-crossing of $df(tT)/dt$.

2.2. Estimation of SIDEN

We now derive $\mathcal{Q}$ in an analytic and parametric form. 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 two patterns represented in terms of Gaussian atoms, it involves the integration of the product of pairs of Gaussian atoms. These integrations yield the following terms, which are explained in more detail in [10]

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

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

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

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. We are now ready to state our result about the estimation of the SIDEN.

Theorem 1. *The region $\mathcal{Q} \subset \mathbb{R}^2$ is a subset of the SIDEN $\mathcal{S}$ of the pattern p if $\mathcal{Q} = \{tT : T \in S^1, \; 0 \leq t \leq \delta_T\}$, where δ_T is the only positive root of the polynomial $|\alpha_4| t^3 - \alpha_3 t^2 - \alpha_1$ and*

$$\alpha_1 = \sum_{j=1}^{K} \sum_{k=1}^{K} \lambda_j \lambda_k \, Q_{jk} \, (2 \, a_{jk} - 4 \, b_{jk}^2)$$

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

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

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

The proof of Theorem 1 is given in Appendix A.1 of [10], which is an accompanying technical report. The proof applies a Taylor expansion of $df(tT)/dt$, and derives a δ_T such that $df(tT)/dt$ is positive for

all $t \leq \delta_T$. Therefore, along each direction T, δ_T constitutes a lower bound for the first zero-crossing of $df(tT)/dt$. Varying T over the unit circle, one obtains a closed neighborhood $\mathcal{Q}$ of 0 that is a subset of $\mathcal{S}$. This analytic estimate provides a guarantee for the range of translations tT over which $p(X)$ can be exactly aligned with $p(X - tT)$.

2.3. Variation of SIDEN with Smoothing

In this section, we examine how smoothing the reference pattern p with a low-pass filter influences its SIDEN. We assume a Gaussian kernel for the filter. As the reference pattern is sparsely represented in a parametric form in a Gaussian dictionary, its convolution with a Gaussian filtering function is also sparsely representable in the same dictionary. Therefore, the choice of the Gaussian kernel provides an immediate interpretation of our SIDEN estimation results for smoothed versions of the reference pattern. 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 $\phi_\rho(X) = \phi(\Lambda^{-1}(X))$ is an isotropic Gaussian atom with the diagonal scale matrix Λ having ρ on the diagonal entries. The scale parameter ρ controls the size of the Gaussian kernel. The smoothed version of the reference pattern $p(X)$ is given by

$$\hat{p}(X) = \frac{1}{\pi\rho^2}\,\phi_\rho(X) * p(X) = \sum_{k=1}^{K} \lambda_k \frac{1}{\pi\rho^2}\,\phi_\rho(X) * \phi_{\gamma_k}(X) \qquad (6)$$

by linearity of the convolution operator. As shown in [10], the filtered pattern is obtained as $\hat{p}(X) = \sum_{k=1}^{K} \hat{\lambda}_k \phi_{\hat{\gamma}_k}(X)$, where the smoothed atom $\phi_{\hat{\gamma}_k}(X)$ has parameters

$$\hat{\tau}_k = \tau_k, \quad \hat{\Psi}_k = \Psi_k, \quad \hat{\sigma}_k = \sqrt{\Lambda^2 + \sigma_k^2}, \quad \hat{\lambda}_k = \frac{|\sigma_k|}{|\hat{\sigma}_k|}\lambda_k. \qquad (7)$$

Therefore, the change in the pattern parameters due to the filtering can be captured by substituting the scale parameters σ_k of atoms with $\hat{\sigma}_k$ and replacing the coefficients λ_k with $\hat{\lambda}_k$. Now, considering the same setting as in Section 2.1, where the target pattern $p(X - tT)$ is exactly a translated version of the reference pattern $p(X)$, we examine how the volume of the SIDEN changes when the reference and target patterns are low-pass filtered as it is typically done in multiscale image registration algorithms. Hence, we analyze the variation of the smoothed SIDEN estimate $\hat{\mathcal{Q}}$ corresponding to the distance $\hat{f}(tT)$ between $\hat{p}(X)$ and $\hat{p}(X - tT)$ with respect to the filter size ρ. Since the smoothed pattern has the same parametric form as the original pattern, the variation of $\hat{\mathcal{Q}}$ with ρ can be analyzed easily by examining how the parameters involved in the derivation of the SIDEN, e.g., $\hat{a}_{jk}$, $\hat{b}_{jk}$, $\hat{\lambda}_k$, $\hat{\sigma}_k$, depend on ρ. We use the notation $\hat{(.)}$ for referring to the parameters corresponding to the filtered versions of the Gaussian atoms. We now give our main result, which summarizes the dependence of the smoothed SIDEN estimate on the filter size ρ.

Theorem 2. *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 ρ is given by $V(\hat{\mathcal{Q}}) = O(1 + \rho^2)$.*

Theorem 2 is proved in [10, 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. The theorem shows that the neighborhood of translation vectors inside 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. Since $\hat{\mathcal{Q}} \subset \hat{\mathcal{S}}$ for all ρ, one immediate observation is that the rate of expansion of the SIDEN $\hat{\mathcal{S}}$ must be at least of $O(1+\rho^2)$; i.e., $V(\hat{\mathcal{S}}) \geq V(\hat{\mathcal{Q}}) = O(1+\rho^2)$. Note that the dependence of $V(\hat{\mathcal{S}})$ on ρ may get above this rate for particular reference patterns. For instance, for patterns that consist only of atoms with coefficients of the same sign, 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 $V(\hat{\mathcal{S}}) = \infty$ [10, Proposition 4].

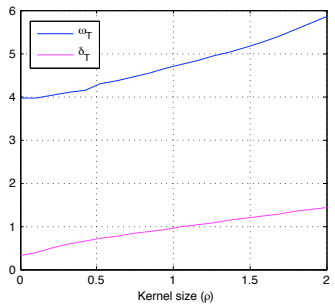


Fig. 2. The variations of the true distance $\hat{\omega}_T$ of the boundary of $\hat{\mathcal{S}}$ to the origin and its estimation $\hat{\delta}_T$ with respect to the filter size

2.4. Evaluation of SIDEN by experiments

We now 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. Then, we determine the distance[2] $\hat{\omega}_T$ 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. 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 1. We repeat the experiment 300 times with different random reference patterns p and directions T; and average the results. In 44% of the trials, $d\hat{f}(tT)/dt$ has been experimentally seen to have no zero-crossings when the pattern is filtered sufficiently. The distance $\hat{\omega}_T$ and its estimation $\hat{\delta}_T$ are plotted in Figure 2 for the remaining 56% of the patterns. The figure shows that both $\hat{\omega}_T$ and $\hat{\delta}_T$ have 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 variation of the true SIDEN boundary $\hat{\omega}_T$.

3. APPLICATION TO PARAMETER DOMAIN SAMPLING

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. We have shown that small translations, i.e., vectors in $\mathcal{Q}$, can be perfectly recovered by minimizing the distance function with descent methods. However, the perfect alignment guarantee is lost for relatively large translations that are outside $\mathcal{Q}$. Hence, we propose to construct a grid in the translation parameter domain and estimate large translation vectors 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

[2]With an abuse of notation, the parameter denoted as $\hat{\omega}_T$ in Section 2.4 corresponds in fact to $\sup \hat{\omega}_T$ in the definition of SIDEN in (5).

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 propose to 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.

We now explain the grid construction. From Theorem 1, one can verify that the estimation δ_T of the SIDEN boundary along the direction T is symmetric and it satisfies $\delta_T = \delta_{-T}$. Therefore, one can easily determine a grid unit in the form of a parallelogram that lies completely inside $\mathcal{Q}$ 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. As the SIDEN 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.

The construction of a regular grid in this manner is demonstrated for a digit pattern. In Figure 3(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 3(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 corresponding grids are displayed in Figures 3(c) and 3(d), where the SIDEN estimates $\mathcal{Q}$, $\hat{\mathcal{Q}}$ and the grid units are also plotted. One can observe that smoothing the pattern results in a coarser grid. In Figure 4, we plot the variation of the number of grid points with the filter size for the random patterns of the previous experiment and the digit pattern. The results confirm that the number of grid points decreases monotonically with the filter size, as stated by Theorem 2, which suggests that the number of grid points must be of $O\left((1+\rho^2)^{-1}\right)$. Finally, the experiments in [10] show that this registration method indeed has an optimal alignment performance.

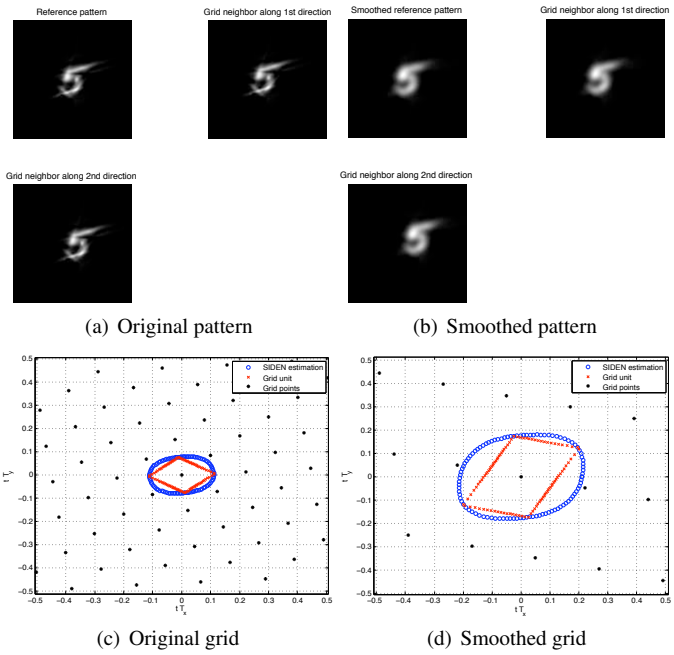
4. CONCLUSION

We have presented an analysis of hierarchical image registration with descent-type local minimizers. We have examined the problem of aligning a reference and a target pattern that differ by a two-dimensional translation. We have derived an estimation of the neighborhood of translations for which the image pair can be exactly aligned with a local optimizer. Then we have investigated how the area of this neighborhood varies with the size of the filter used in the coarse-to-fine registration process. Our finding is that the area of this neighborhood increases quadratically with the filter size, therefore, smoothing the patterns improves the well-behavedness of the distance function. We have used our results in the construction of a multiscale regular grid in the translation parameter domain that guarantees the exact alignment of a reference pattern with its translated versions. The fact that the number of grid points is inversely proportional to the square of the filter size shows that filtering is useful for decreasing the computational complexity of image alignment.

5. REFERENCES

[1] 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*. 1998, New York: Springer-Verlag.

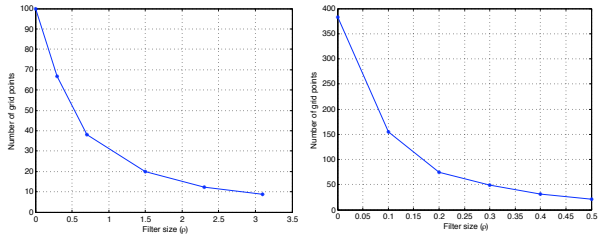
[2] A. W. Fitzgibbon and A. Zisserman, "Joint manifold distance: a new approach to appearance based clustering," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 26, 2003.



(a) Original pattern (b) Smoothed pattern

(c) Original grid (d) Smoothed grid

Fig. 3. Construction of a regular grid in parameter domain with an exact alignment guarantee



(a) Random patterns (b) Digit pattern

Fig. 4. Number of grid points. The decay rate is of $O\left((1+\rho^2)^{-1}\right)$.

[3] 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.

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

[5] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Trans. Img. Proc.*, vol. 13, no. 9, pp. 1185–1199, Sept. 2004.

[6] İ. Ş. Yetik and A. Nehorai, "Performance bounds on image registration," *IEEE Trans. Signal Proc.*, vol. 54, no. 5, pp. 1737 – 1749, May 2006.

[7] 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.*, Mar. 1987.

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

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

[10] E. Vural and P. Frossard, "Analysis of descent-based image registration," Available at: http://infoscience.epfl.ch/record/183845.