A PERFORMANCE STUDY OF THE TANGENT DISTANCE METHOD IN TRANSFORMATION-INVARIANT IMAGE CLASSIFICATION

Elif Vural [1] *and Pascal Frossard* [2]

[1] Centre de Recherche INRIA Rennes Bretagne Atlantique, Rennes, France.
[2] Ecole Polytechnique Fédérale de Lausanne, Signal Processing Laboratory LTS4, Lausanne, Switzerland.

ABSTRACT

A common problem in image analysis is the transformation-invariant estimation of the similarity between a query image and a set of reference images representing different classes. This typically requires the comparison of the distance between the query image and the transformation manifolds of the reference images. The tangent distance algorithm is a popular method that estimates the manifold distance by employing a linear approximation of the transformation manifolds. In this paper, we present a performance analysis of the tangent distance method in image classification applications for general transformation models. In particular, we characterize the misclassification error in terms of the geometric properties of the individual manifolds such as their curvature, as well as their relative properties such as the separation between them. We then extend our results to a multi-scale analysis where the images are smoothed with a low-pass filter and study the effect of smoothing on the misclassification error. Our theoretical results are confirmed by experiments and may find use in the selection of algorithm parameters in multi-scale transformation-invariant image analysis methods.

Index Terms— Tangent distance, image classification, hierarchical image registration, performance analysis.

1. INTRODUCTION

In image analysis problems where different classes are represented by different manifolds generated by the geometric transformations of a reference image, classification can be achieved by measuring the distance of the query image to the transformation manifold of each class. In order to compute the manifold distance, the query image is aligned with the reference class-representative images, which is generally not an easy problem due to the nontrivial form of the transformation manifold. The tangent distance method is a well-known alignment algorithm that constructs a first-order approximation of the transformation manifolds of the reference images by computing the tangent space of the manifold at a reference point. The transformation parameters are then estimated by calculating the orthogonal projection of the target image onto the tangent space of the manifold.

The tangent distance method has been proposed by Simard et al. and its efficiency has been demonstrated in numerous settings, such as handwritten digit recognition applications [1], [2]. Following [1], many variations on the tangent distance method have been presented. The work in [3], for example, introduces the joint manifold distance for transformation-invariance in clustering, which is a similarity measure that is based on the prior distributions of the images and the distance between the linear approximations of their manifolds. It is common to apply the tangent distance method in a hierarchical manner, by using a sequence of smoothed versions of the images with low-pass filters [2], [4]. The multiscale tangent distance method aims to get around the limitation that the reference point around which the manifold is linearized should be sufficiently close to the exact projection of the target image onto the manifold.

The efficiency of the hierarchical tangent distance algorithm has been observed in many applications. However, the previous works performing a theoretical study of this algorithm are confined to optical flow literature (which becomes the equivalent of the tangent distance method for 2-D translation models) [5], [6], [7], and the classification performance of the tangent distance method for general geometric transformation models has not been theoretically studied yet.

In this paper, we build on our previous work [8], where we study the performance of image registration with the tangent distance method. We present a theoretical analysis of the properties of the multiscale tangent distance method in image classification applications. We consider a setting where the query image and the reference images representing different classes are smoothed with low-pass filters, and the projection of the query image onto the transformation manifolds of the reference images is estimated with the tangent distance method. We first derive an upper bound on the misclassification probability of a query image for generic geometric transformation models, where we assume a bounded and non-intersecting distribution of the query images around the transformation manifolds of their classes. We characterize the misclassification probability in terms of the geometric properties of the manifolds, the deviation of query images from the class-representative manifolds, and the separation between the classes. We then examine the variation of the misclassification probability with the size of the low-pass filter in a setting where the smoothed versions of the reference and query images are used in the registration. Our analysis shows that, the misclassification probability has a non-monotonic variation with the filter size. Hence, there is an optimal value of the filter size that minimizes the misclassification error. Our results are confirmed by experiments on transformation models consisting of translations, rotations, and scale changes.

2. OVERVIEW OF THE TANGENT DISTANCE METHOD

2.1. Image Registration with the Tangent Distance Method

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

The work was performed while the first author was at EPFL.

ing the spatial coordinate variable $X = [x\ y]^T$ in $\mathbb{R}^2$, we have

$$p_\lambda(X) = p(a(\lambda, X)) \quad (1)$$

where a is a function representing the change of coordinates defined by the transformation λ. In this work, we treat general transformation models and assume that λ can represent any parametrizable geometric transformation such that $a : \Lambda \times \mathbb{R}^2 \to \mathbb{R}^2$ is a C^2-smooth function and $a_\lambda(X) := a(\lambda, X)$ is a bijection for a fixed λ.

The transformation manifold $\mathcal{M}(p)$ of the pattern p is given by

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

which consists of transformed versions of p over the parameter domain Λ. Since a and p are C^2-smooth, the first and second derivatives of manifold points exist. We denote the derivative of the manifold point p_λ with respect to the i-th transformation parameter λ^i as $\partial_i\, p_\lambda$, where $\partial_i\, p_\lambda(X) = \partial\, p_\lambda(X)/\partial\lambda^i$. The derivatives $\partial_i\, p_\lambda$ correspond to the tangent vectors of $\mathcal{M}(p)$ on p_λ. Similarly, we denote the second-order derivatives by $\partial_{ij}\, p_\lambda(X) = \partial^2 p_\lambda(X)/\partial\lambda^i\partial\lambda^j$. Then, the tangent space $T_\lambda\mathcal{M}(p)$ of the manifold at a point p_λ is the subspace generated by the tangent vectors at p_λ

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

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

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

$$\lambda_o = \arg\min_{\lambda\in\Lambda} \|q - p_\lambda\|^2 \quad (3)$$

where $\|\cdot\|$ denotes the L^2-norm for vectors in the continuous space $L^2(\mathbb{R}^2)$ and the ℓ^2-norm for vectors in the discrete space $\mathbb{R}^n$. Then, the transformed pattern p_{λ_o} is called a projection of q on $\mathcal{M}(p)$.

The exact calculation of λ_o is difficult in general, because of the nonlinear and highly intricate geometric structure of pattern transformation manifolds. The tangent distance method simplifies this problem to a least squares problem, where the transformation parameters are estimated by using a linear approximation of the manifold $\mathcal{M}(p)$ and then computing λ_o by minimizing the distance of q to the linear approximation of $\mathcal{M}(p)$ [4], which is illustrated in Figure 1. The estimate λ_e of λ_o with the tangent distance method is given by the solution of the following least squares problem

$$\lambda_e = \arg\min_{\lambda\in\mathbb{R}^d} \|q - p_{\lambda_r} - \partial_i\, p_{\lambda_r}(\lambda^i - \lambda_r^i)\|^2 \quad (4)$$

whose solution is obtained as

$$\lambda_e^i = \lambda_r^i + \mathcal{G}^{ij}(\lambda_r)\langle q - p_{\lambda_r}, \partial_j\, p_{\lambda_r}\rangle \quad (5)$$

where $\mathcal{G}_{ij}(\lambda) = \langle\partial_i\, p_\lambda, \partial_j\, p_\lambda\rangle$ is the metric tensor induced from the standard inner product on $L^2(\mathbb{R}^2)$, $[\mathcal{G}_{ij}(\lambda)] \in \mathbb{R}^{d\times d}$ is the matrix representation of the metric tensor, and $\mathcal{G}^{ij}$ represents the entries of the inverse $[\mathcal{G}_{ij}(\lambda)]^{-1}$ of the metric. We can then define the alignment error of tangent distance as $\|\lambda_e - \lambda_o\|$, which represents the deviation between the estimated and the optimal transformation parameters. In [8], we present an upper bound on the alignment error $\|\lambda_e - \lambda_o\|$ and study how it varies with the smoothing of the reference and target images with a Gaussian low-pass filter.

[1] In (2) and several other equations, we use the Einstein notation to simplify the writing. In Einstein notation, an index variable appearing twice in a term (once in a superscript and once in a subscript) indicates a summation; i.e., $\sum_{i=1}^d v_i w^i$ is simply written as $v_i w^i$.

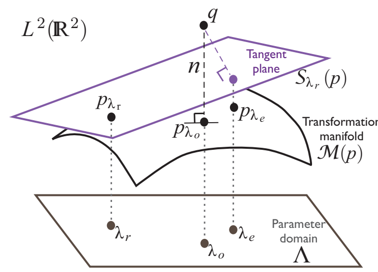


Fig. 1. Image alignment with the tangent distance method. The estimate λ_e of the optimal transformation parameters λ_o is obtained by computing the orthogonal projection of the target image q onto the linear approximation $\mathcal{S}_{\lambda_r}(p)$ of the manifold around p_{λ_r}.

2.2. Image Classification with the Tangent Distance Method

We now consider a setting with M class-representative patterns $\{p^m\}_{m=1}^M$ whose transformation manifolds

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

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

$$l(q) = \arg\min_{m\in\{1,\dots,M\}} \|q - p_{\lambda_o^m}^m\| \quad (6)$$

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

Let λ_e^m denote the estimate of λ_o^m computed with the tangent distance method as in (5) by linearizing the manifold $\mathcal{M}(p^m)$ around a reference point with parameter vector λ_r^m. The class label of q is then estimated with the tangent distance method as follows[2]

$$\tilde{l}(q) = \arg\min_{m\in\{1,\dots,M\}} \|q - p_{\lambda_e^m}^m\|. \quad (7)$$

Our purpose is then to study the performance penalty when the class label of a query pattern is estimated as above by using first-order approximations of the manifolds. We focus on the accuracy of classifying a query image with a one-step application of the tangent distance method, i.e., by estimating the transformation parameters $\{\lambda_o^m\}$ with a single linearization of each manifold, possibly by low-pass filtering the query and reference images (a detailed study of the hierarchical estimation of transformation parameters is given in [8]). We study the performance of classification in this setting and its dependence on the low-pass filter size.

3. ANALYSIS OF THE TANGENT DISTANCE METHOD

3.1. Single-scale analysis

We analyze the classification performance by considering a setting where the query images of class m have a distribution that is concentrated around the manifold $\mathcal{M}(p^m)$. We then examine the probability of correctly classifying q based on the distance estimates given by the tangent distance method.

[2] Note that the class label of a query image can also be estimated by comparing its distance to the first-order approximations of the manifolds. While Simard et al. use this subspace distance for classification [2], the estimate in (7) is also commonly used in image analysis problems (e.g., as in [4]). We base our analysis on the definition in (7) since it is likely to give more accurate estimates, especially with multiscale generalizations as in (11).

Let $\nu_j = \|q - p^j_{\lambda^j_o}\|$ denote the deviation of a query image q from the manifold $\mathcal{M}(p^j)$ of class j. Furthermore, let q belong to class m. The distance of q to $\mathcal{M}(p^m)$ is the smallest among the distances of q to all manifolds; therefore, $\nu_m < \nu_j$ for all $j \neq m$. Let us assume that the distributions of the images belonging to different classes have bounded and non-intersecting supports around the manifolds, so that the classification rule in (6) always gives the true class label. We can then define the following parameters. Let

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

denote the maximal distance of query patterns of class m to the manifold $\mathcal{M}(p^m)$ of their own class, which can be considered as the maximal noise level. Let also

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

define a distance margin that is a measure of the minimum separation between different classes. Finally, let

$$\mathcal{T}_m := \max_{i=1,\dots,d} \; \sup_{\lambda\in\Lambda} \|\partial_i \, p^m_\lambda\| \tag{8}$$

denote the supremum of the tangent norm on $\mathcal{M}(p^m)$ and $\mathcal{K}_m$ be a curvature parameter of $\mathcal{M}(p^m)$ given by

$$\mathcal{K}_m := \max_{i,j=1,\cdots,d} \; \sup_{\lambda\in\Lambda} \|\partial_{ij} \, p^m_\lambda\|. \tag{9}$$

We then have the following result, which provides an upper bound for the probability of misclassifying a target image of class m.

Theorem 1. *Let q be a query pattern of class m. Assume that the optimal transformation parameters λ^m_o aligning q with p^m are within a Δ-neighborhood of the reference transformation parameters λ^m_r around which $\mathcal{M}(p^m)$ is linearized, such that $\|\lambda^m_o - \lambda^m_r\|_1 \leq \Delta$. Then, the probability of misclassifying q with the tangent distance method is upper bounded as*

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

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

The proof of Theorem 1 is given in [9, Appendix E.2]. The above result is obtained by deriving a relation between the the probability of misclassification and the alignment error of tangent distance. The misclassification probability is then bounded using the alignment error upper bounds derived in [8].

Theorem 1 shows how the probability of misclassification when the manifold distances are estimated with the tangent distance method, depends on the geometric properties of the manifolds and on the deviation Δ between the reference and the optimal transformation parameters. In particular, for any non-intersecting and bounded distribution of class samples, the misclassification probability increases at most linearly with the increase in the manifold curvature and the maximal distance of the images to their own representative manifold. The deviation Δ between the reference and optimal transformation parameters affects the misclassification probability since it influences the alignment accuracy. Note that, if a prior estimation of the transformation parameters is available (e.g., as in a hierarchical alignment setting), the reference transformation parameters λ^m_r can be set accordingly. Otherwise λ^m_r is taken as the identity transformation. It is also observed that better separation of manifolds (i.e., increase in the distance margin ϵ) reduces the misclassification probability, as expected. We finally remark that the result in Theorem 1 is based on the tangent and curvature definitions in (8)-(9), which depend on the chosen coordinate system, and the parameter domain distance Δ between the transformation parameters. These coordinate-dependent definitions are particularly interesting in the image registration context as these parameters have immediate physical interpretations. For instance, in an application where rotated images are classified, Δ would quantify the maximum degree of rotation, while $\mathcal{T}_m$ and $\mathcal{K}_m$ would measure the sensitivity of the rotated images to the amount of rotation.

3.2. Multi-scale analysis

We now discuss the classification of images with the tangent distance method in a multiscale setting and study the effect of smoothing. Consider that the transformation parameters are estimated by low-pass filtering the query image and the reference images with a Gaussian filter kernel

$$\frac{1}{\pi\rho^2}\phi_\rho(X) = \frac{1}{\pi\rho^2} e^{-\frac{x^2+y^2}{\rho^2}} \tag{10}$$

of size ρ, which has unit L^1-norm. We thus consider that the tangent distance method uses the filtered versions

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

of the reference images p^m and the query image q for aligning q with p^m. From (5), the estimates of the transformation parameters are obtained as $(\hat{\lambda}^i_e)^m = (\lambda^i_r)^m + (\hat{\mathcal{G}}^{ij})^m(\lambda^m_r)\langle \hat{q} - \hat{p}^m_{\lambda_r}, \partial_j\, \hat{p}^m_{\lambda_r}\rangle$, where the notation $\hat{(\cdot)}$ stands for the counterparts of the parameters in the previous section that correspond to the filtered versions of the patterns. Once the transformation parameters are estimated, we assume that the unfiltered versions of the reference images and the query image are used in the computation of the actual distances to the manifolds for estimating the class label of the query image. It is preferable to compare the distances in the original image space rather than the space of filtered images, as it yields more accurate estimates. The class label of the query pattern is thus estimated as

$$\tilde{l}(q) = \arg\min_{m\in\{1,\dots,M\}} \|q - p^m_{\hat{\lambda}^m_e}\|. \tag{11}$$

In image classification with multiscale tangent distance, the filter size should be selected to make the above estimate $\tilde{l}(q)$ as accurate as possible. Hence, it is important to characterize the variation of the misclassification probability with the size ρ of the low-pass filter. In the following theorem, we provide an upper bound on the rate of variation of the classification error with ρ.

Theorem 2. *The probability of misclassifying q with the multiscale tangent distance method is upper bounded as*

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

The misclassification probability bound E_ρ varies with the filter size ρ as $E_\rho = O\left(1 + (1+\rho^2)^{-1/2} + (\mathcal{V}_m + 1)\,(1+\rho^2)^{1/2}\right)$ if the transformation model includes a scale change of the pattern, and $E_\rho = O\left(1 + (1+\rho^2)^{-1/2} + \mathcal{V}_m\,(1+\rho^2)^{1/2}\right)$ if the transformation model does not include a scale change of the pattern.

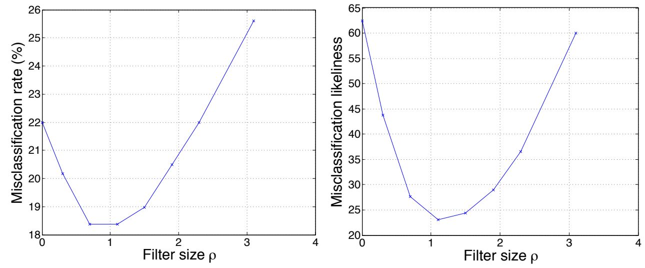
Proof. The upper bound E_ρ on the misclassification probability is a direct implication of Theorem 1. Comparing the expression of E_ρ and the alignment error upper bound in [8] (presented in [9, Theorem 1]), one can observe that they only differ by a multiplicative factor (note, however, that the value of this factor depends on the geometric properties of the manifolds through the parameters $\mathcal{T}_m$ and ϵ). Therefore, the misclassification probability upper bound has the same variation with the filter size as the alignment error upper bound in [9, Theorem 1]. Combining this observation and the variation of the alignment error with the noise level and the filter size presented in [8] (available in [9, Theorem 2]), we obtain the stated result. □

Theorem 2 shows that the misclassification probability has a nonmonotonic variation with the filter size. The first component of E_ρ related to the manifold curvature decreases at a rate of $O\left(1 + (1+\rho^2)^{-1/2}\right)$ with the filter size ρ. Filtering makes the manifold smoother and decreases the manifold curvature, which improves the accuracy of the first-order approximation of the manifold. However, the second component of the misclassification probability associated with noise increases with the filter size and the noise level at a rate of $O\left((\mathcal{V}_m + 1)\,(1+\rho^2)^{1/2}\right)$. This is due to the fact that filtering has the undesired effect of amplifying the alignment error caused by noise. This result is in line with the findings of our previous study [10], and previous works such as [5], [11] that examine the Crámer-Rao lower bound in image registration.

For sufficiently small values of the image noise level, the misclassification probability E_ρ first decreases with the filter size ρ due to the first term, and then starts to increase with the filter size due to the second term. Therefore, there exists an optimal value of the filter size ρ that minimizes the misclassification probability. The optimal value of ρ depends on the parameters of the classification problem. The maximal distance $\mathcal{V}_m$ is related to the internal variation (noise level) of the data samples within the same class and depends on how well the reference pattern p^m approximates the samples of its own class, whereas the parameter Δ can be set according to the maximum amount of transformation that the data samples are likely to undergo in the application at hand.

4. EXPERIMENTAL RESULTS

We now experimentally study the image classification performance when manifold distances are computed with the tangent distance method. We experiment on two classes of synthetic images. The reference pattern of each class consists of 20 randomly chosen Gaussian atoms (generated by applying random geometric transformations to a Gaussian mother function) such that 16 of the atoms are common between the two classes and 4 atoms are specific to each class. This configuration simulates a setting where different classes contain class-specific features as well as common features. We then generate a set of test patterns that lie between the transformation manifolds of the two reference patterns. The test patterns are generated such that their true class labels are given by the class label of the closer manifold as in (6). We classify the test patterns with the tangent distance method as in (11), where transformation parameters are estimated from the low-pass filtered versions of the images. We experiment on a transformation model consisting of a 2-D translation, rotation and a scale change; and test the classification accuracy at different filter sizes. In Figure 2(a), the percentage of misclassified test patterns is plotted with respect to the filter size. In order to interpret the variation of the experimental misclassification rate with the filter size in light of our theoretical results, we define a function $\mathcal{T}_m\,\hat{\mathcal{K}}_m\,\eta_{\min}^{-1}\,\left([\hat{\mathcal{G}}_{ij}^m(\lambda_r^m)]\right)\left(\frac{1}{2}\sqrt{\mathrm{tr}([\hat{\mathcal{G}}_{ij}^m(\lambda_r^m)])}\,\|\hat{\lambda}_o - \lambda_r\|_1^2 + \sqrt{d}\,\|\tilde{n}_m\|\,\|\hat{\lambda}_o - \lambda_r\|_1\right)$ for the test patterns, where $\|\tilde{n}_m\|$ is the distance between the filtered test pattern $\hat{q}$ and the transformation manifold $\mathcal{M}(\hat{p}^m)$ of the filtered reference pattern representing class m. This function has the same variation with the filter size ρ as the misclassification probability bound E_ρ, while it is easier to compute experimentally. As it provides a measure for the misclassification probability, we call this function the "misclassification likeliness". The average value of the misclassification likeliness is plotted in Figure 2(b). Comparing panels (a) and (b) of Figure 2, we observe that the variation of the experimental misclassification probability with filtering agrees with that of the analytical misclassification likeliness. This shows that the misclassification probability upper bound in Theorem 2 captures well the behavior of the actual misclassification probability. The experimental results confirm that the misclassification probability has a non-monotonic variation with the filter size as predicted by Theorem 2, and the optimal filter size minimizing the misclassification probability is in the vicinity of the filter size that minimizes the misclassification likeliness. In [9], we present further experiments on the classification of digit images, which we skip here due to lack of space. These experiments also confirm the findings of our theoretical analysis and suggest that the best classification performance is obtained at large filter sizes for real images, where the main source of misclassification is the manifold nonlinearity due to the high-frequency components prominent in real images.



(a) (b)

Fig. 2. Classification results for synthetic patterns.

5. CONCLUSIONS

We have presented an analysis of image classification with the tangent distance method, which aligns the query image and the class-representative reference images by linearizing the transformation manifolds of the reference images. We have presented an upper bound on the probability of misclassification for generic transformation models. The misclassification probability depends on the individual geometric properties of the manifolds such as their metric and curvature, as well as the separation between the manifolds and the deviation of query images from the manifold of their own class, i.e., noise level. We have then studied the variation of the misclassification probability when the tangent distance method is applied in a multiscale setting, and shown that there exists an optimal value of the low-pass filter size that minimizes the misclassification probability. Our study provides useful insight for optimizing the performance of multiscale methods relying on first-order manifold approximations in the analysis and classification of images.

6. REFERENCES

[1] P. Y. Simard, B. Victorri, Y. LeCun, and J. S. Denker, "Tangent prop: a formalism for specifying selected invariances in adaptive networks," in *Advances in Neural Information Processing Systems (NIPS 1991)*, Denver, CO, 1992, vol. 4.

[2] P. Y. Simard, Y. LeCun, J. S. Denker, and B. Victorri, "Transformation invariance in pattern recognition – tangent distance and tangent propagation," *International Journal of Imaging Systems and Technology*, vol. 11, no. 3, 2001.

[3] 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.

[4] N. Vasconcelos and A. Lippman, "A multiresolution manifold distance for invariant image similarity," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 127–142, 2005.

[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] 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*, 1994, vol. 1, pp. 721–725.

[7] 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.

[8] E. Vural and P. Frossard, "Analysis of image registration with tangent distance," Submitted to SIAM Journal on Imaging Sciences.

[9] E. Vural and P. Frossard, "A study of image analysis with tangent distance," 2014, Available at: http://arxiv.org/abs/1401.2529.

[10] E. Vural and P. Frossard, "Analysis of Descent-Based Image Registration," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2310–2349, 2013.

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