Introducing error estimation in the shape learning framework : Spline based registration and non-parametric density estimator in the space of higher order polynomials.

Maxime TARON
Nikos PARAGIOS
Marie-Pierre JOLLY

Research Report 05-08
March 2005

Ecole Nationale Des Ponts et Chaussées
Centre d'Enseignement et de Recherche en Technologies de l'Information et Systèmes

CERTIS, ENPC,
77455 Marne la Vallee, France,

Introducing error estimation in the shape learning framework : Spline based registration and non-parametric density estimator in the space of higher order polynomials.[1]

Evaluation des incertitudes liées au recalage de formes : Transformations polynomiale par morceaux et apprentissage en vue de la création d'un modèle statistique non paramétrique adapté. [1]

Maxime TARON[2]
Nikos PARAGIOS[2]
Marie-Pierre JOLLY[3]

[1]Research work supported by the Imaging & Visualization Department of Siemens Corporate Research http://www.scr.siemens.com

[2]CERTIS, ENPC, 77455 Marne la Vallee, France, http://www.enpc.fr/certis/
{taron,paragios}@certis.enpc.fr

[3]Imaging & Visualization Dept. , Siemens Corporate Research. , Princeton USA
marie-pierre.jolly@siemens.com

Abstract

In this report, we introduce a new technique to shape modelling in the space of implicit polynomials. Registration consists of recovering an optimal one-to-one transformation of a higher order polynomial along with uncertainties measures that are determined according to the covariance matrix of the correspondences at the zero isosurface. Such measures are used to weight the importance of the training samples in the modelling phase according to a variable bandwidth non-parametric density estimation process. The selection of the most appropriate kernels to represent the training set is done through the maximum likelihood criterion. Exceptional results for patterns of digits, related with the registration and the modelling aspects of our approach demonstrate the potentials of our method.

* The author gratefully acknowledge Dr. Jean-Yves Audibert for fruitful discussion regarding the probabilistic models.

Résumé

Ce rapport introduit une nouvelle technique de modélisation des déformations d'une forme dans l'espace des transformations polynomiales.
Le recalage d'une forme, représentée dans l'espace des fonctions implicites, a pour but de déterminer la transformation polynomiale par morceaux inversible et optimale selon des critères de similarité et de régularité. Les mesures d'incertitudes quantifiant la qualité de la transformation obtenue sont estimées sur la courbe de niveau zéro. Ces mesures calculées sur un vaste ensemble d'apprentissage permettent de déterminer l'importance relative de chaque élément lors de la phase de modélisation. Un modèle non paramétrique basé sur des estimateurs a noyaux à covariance variable sera introduit. Les spécimens de recalage avec leur mesure d'incertitude les plus représentatifs de l'ensemble d'apprentissage sont sélectionnés selon un critère de maximisation de la vraisemblance de l'échantillon.
Des résultats exceptionnels sur des caractères numériques démontrent tout le potentiel de notre méthode, tant en termes de recalage de forme que de modélisation statistique.

Contents

1 Introduction

Domain knowledge is often available in computational vision and therefore efficient techniques are to be developed to account for it. To this end, once registration of data (shapes, appearance, motion, etc.) to a common pose is completed, its statistical characterisation according to a compact model is to be recovered that is used to impose constraints when solving the inference problem.

One can define the registration problem as follows: recover a transformation between a source and a target shape that results in meaningful correspondences between their basic elements. To this end, one (i) should select an appropriate representation for the structures of interest, (ii) define the set and the nature of plausible transformations, and (iii) determine an appropriate mathematical framework to recover the optimal registration parameters.

Point-based global and local registration [20] through low cost optimisation techniques like the ICP [4] algorithm is the most primitive approach to shape registration. One can also refer to more advanced methods like diffeomorphic matching [5]. More advanced representations of shapes refer to B-splines as well as other form of continuous interpolation functions [19], shocks, skeletons [10] and distance transforms [1].

Registration can be either global or local. Global parametric transformations are within a restricted group, like rigid, similarity, affine, etc. The term local registration is often used in a narrower sense and refer to a transformation with infinite degrees of freedom. Such a deformation can potentially map any finite number of points to the same number of points. However, non-rigid registration is often an under constrained problem. Therefore in order to find a unique non-rigid transformation, we need further constraints to be introduced through a regularization of the registration field.

A different approach consists of addressing registration as a statistical estimation problem [9] through successive steps. Within each step the uncertainty in the estimates is being computed [16] and is used to guide further steps of the overall algorithm [12]. In [15] the covariance matrix is used within an ICP algorithm to sample the correspondences so that registration is well-constrained in all directions in parameter space. Last, but not least in in [14] local deformation and uncertainties are simultaneously recovered for the optical flow estimation problem through a Gaussian noise assumption on the observation.

Similar to the registration problem, the modelling aspect consists of (i) selecting the nature of the density function, and (ii) recovering the parameters of such a function so it approximates the registered data. Parametric linear models like Gaussian densities are often employed through either an EM algorithm or a singular value decomposition. One can claim that such models refer to an efficient compact approximation when the selected model fits to the data. Non-parametric approaches of fixed bandwidth kernels like Parzen windows [8] are a more efficient technique to approximate data that do not obey a particular rule. Their tradeoff of is being a computationally expensive approach while important attention is to be paid on the selection of their bandwidth.

Modelling the geometric form of objects is a challenging task of computational vision. Such a task consists of two steps, (i) registration, and (ii) statistical modelling. Prior work consists of addressing registration and modelling in a sequential fashion. Within such an approach registration errors are not accounted for and often lead to incorrect and erroneous models.

In this paper we propose a novel technique to shape modelling that exploits registration uncertainties. To this end shapes are represented in an implicit fashion and are registered using a thin-plate spline deformation model according to a topology-preservation algorithm. This approach can also provide uncertainty measures according to the covariance estimation matrix at the zero iso-surface. Upon dimensionality reduction, through a maximum likerlyhood criterion that dictates the most representative kernel set, these measures are used within a variable bandwidth kernel-based density function. Given a new example once registration and uncertainties estimation have been completed, appropriate metrics are designed that do explicitly encode the estimates and their uncertainties to evaluate the probability of the subject under consideration being part of the family of the model.

The reminder of the paper is organized in the following fashion. In section 2 we briefly present shape registration in the space of implicit polynomials while the estimation of uncertainties is part of section 3. The objective of building compact non-parametric densities to describe shapes is addressed in section 4. Results and discussion appear in section 5.

2 Registration through Implicit Polynomials

Smoothness and in particular topology preservation are desirable properties in registration. A transformation is said to be smooth if all partial derivatives, up to certain orders, exist and are continuous. At the same time it is said to preserve the topology if the source and the transformed source have the same topology.

In the present framework, a shape $\mathcal{S}$ is represented in an implicit fashion using the Euclidean distance transform $\mathcal{D}$ [1]. In the 2D case, we consider the function defined on the image domain Ω and $\mathcal{R}_\mathcal{S}$ is the region enclosed by $\mathcal{S}$:

$$\phi_\mathcal{S}(x,y) = \begin{cases} 0, & (x,y) \in \partial\mathcal{S} \\ +\mathcal{D}((x,y),\mathcal{S}), & (x,y) \in \mathcal{R}_\mathcal{S} \\ -\mathcal{D}((x,y),\mathcal{S}), & (x,y) \in \bar{\mathcal{R}_\mathcal{S}} \end{cases}$$

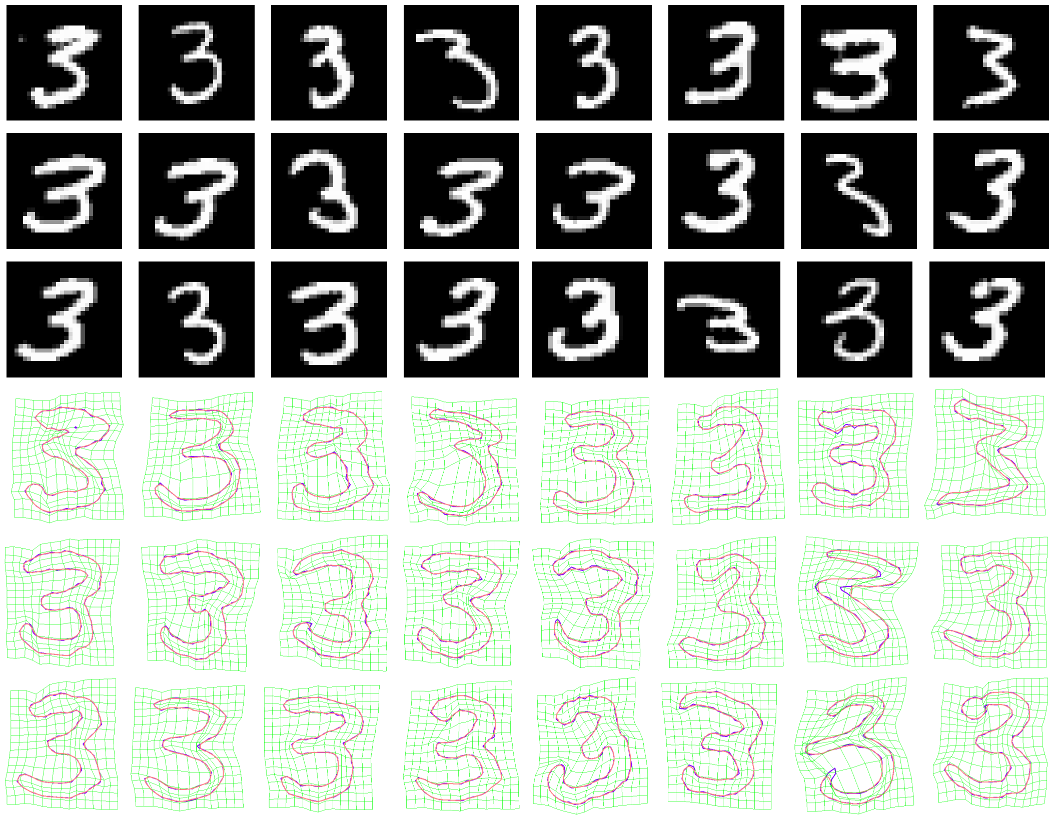

Figure 1: *Examples of registration to a common pose of various samples of '3' using an* 16×14 *ffd grid.*

Such a space is invariant to translation and rotation and can also be modified to account for scale variations. In the most general case an apparent relation between the distance function of the source and the target is not present.

Now consider a smooth 2D diffeomorphism depending upon a vector of parameters $\Theta \in \mathbb{R}^n$ and defines an image transformation on Ω :

$$\mathcal{L}(\Theta, .) : \Omega \rightarrow \Omega$$

Standard point-based curve registration consists of applying $\mathcal{L}$ to the source shape $\mathcal{S}$ and minimize the curve integral along $\partial\mathcal{S}$:

$$E_0(\mathcal{L}(\Theta)) = \oint_{\partial\mathcal{S}} \rho(\phi_{\mathcal{T}}(\mathcal{L}(\Theta, \mathbf{x}))d\mathbf{s}$$

Such that some metric error between the transformed source and the target is minimal. One can extend registration within a band of information along numerous image isophotes :

$$E_\alpha(\mathcal{L}(\Theta)) = \iint_\Omega \chi_\alpha(\phi_\mathcal{S}(\mathbf{x}))\rho\left(\phi_\mathcal{S}(\mathbf{x}) - \phi_\mathcal{T}(\mathcal{L}(\Theta, \mathbf{x}))\right) d\mathbf{x}$$

and introduce the indicator function :

$$\chi_\alpha(x) = \begin{cases} 1/(2\alpha) & \text{if} x \in [-\alpha, \alpha] \\ 0 & \text{else} \end{cases}$$

Within such a process the selection of the α is crucial since to some extent it refers to the scale of the shapes to be registered and eliminates the risk of convergence to local minima.

On the other hand, it is natural when converging to the optimal solution that $\alpha \approx 0$. To this end, we assume a finite number of decreasing set of radius $\{\alpha_0 > \alpha_1 > ... > \alpha_n \approx 0\}$ that is equivalent to a scale-space decomposition of the process. However, it shall also be noticed that the complexity of the transformation $\mathcal{L}$ and therefore the size of Θ has to be increased progressively as α_k decreases in order to prevent the convergence to local minima. At the scale $t-1$, minimum will be obtained for the parameter Θ_{t-1} defining the transformation $\mathcal{L}_{t-1} = \mathcal{L}(\Theta_{t-1}, .)$. Also let $\mathcal{S}^{t-1} = \mathcal{L}_{t-1} \circ \mathcal{S}$, the registration between shapes is then equivalent with iteratively minimizing :

$$\begin{aligned} & E_{\alpha_t}(\mathcal{L}(\Theta)) \\ = \iint_\Omega \chi_{\alpha_t}(\phi_\mathcal{S}(\mathbf{x}))\rho(\phi_{\mathcal{S}^{t-1}}(\mathcal{L}_{\sqcup-\infty}(\mathbf{x})) - \phi_\mathcal{T}(\mathcal{L}(\Theta, \mathbf{x}))) d\mathbf{x} \end{aligned}$$

where a correction process is applied when refining scales through the modification of the distance transform that describes the source shape $\phi_{\mathcal{S}^{t-1}}()$. Within such a formulation the integration domain is always related to the initial source shape and does not depend on the number of iteration or the parameter α_t. Moreover when using Euclidean distance and $\alpha_t \to 0$, $E_{\alpha_t}(\mathcal{L}(\Theta))$ is equivalent to the point based registration ($E_{\alpha_\infty}(\mathcal{L}(\Theta)) = E_0(\mathcal{L}(\Theta))$).

Such an objective function can be used to address global registration as well as local deformations. Affine models are used to globally align shapes with six degrees of freedom, while as proposed in [7] we refine the transformation using free form deformations to address local registration.

Cubic B-spline based free form deformations are an efficient way to model locally smooth transformations on images [13]. Deformations of shapes (and their implicit representation $\phi_\mathcal{S}$) are recovered by evolving a square control lattice $\mathbf{P}$ that is overlaid on the initial distance transform structure. Let us consider the control lattice points $\{\mathbf{P}_{m,n}\}$ defining the initial regular grid. The displacement of any of control points will induce a local and $\mathcal{C}^2$ field of deformation:

$$\mathcal{L}(\Theta, \mathbf{x}) = \sum_{k=-1}^{2}\sum_{l=-1}^{2} B_k(u)B_l(v)(\mathbf{P}_{i+k,j+l} + \delta\mathbf{P}_{i+k,j+l})$$

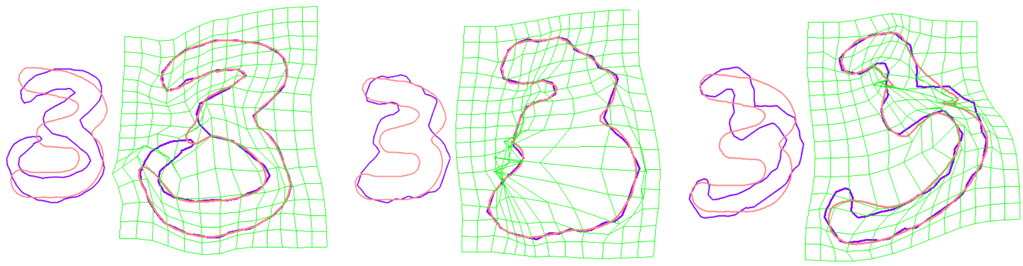

Figure 2: *Examples of problematic registration where the recovered transformation present certain irregularities. Such cases represent only 1.8 % of the MNIST OCR database.*

Where B_k is the k^th basis function of the Cubic B-spline. This local transformation is a compromise between global and local registration and its parameters consist of the displacement of control points ($\Theta = \{\delta \mathbf{P}_{m,n}\}$). In [7], such a framework is introduced using implicit functions defined on the complete domain Ω.

Such a transformation accounts for smoothness to a certain degree in an implicit fashion. In order to avoid folding and recover a wiggle free transformation one can consider the use of additional smoothness terms aim to constraint the spatial variation of the displacement. Opposite to [7], we adopt a regularization term motivated by the thin plate energy functional [18] :

$$E_{smooth}(\mathcal{L}(\Theta)) = \iint_{\Omega} \left(|\mathcal{L}_{xx}|^2 + 2\,|\mathcal{L}_{xy}|^2 + |\mathcal{L}_{yy}|^2 \right) d\Omega$$

that can be further simplified in the case of the cubic B-spline and reduced to the quadratic form $[E_{smooth}(\mathcal{L}(\Theta)) = \Theta^T.\mathrm{C}.\Theta]$ with C a symmetric matrix.

The objective function $[E_{\alpha_\infty}(\mathcal{L}(\Theta)) + wE_{smooth}(\mathcal{L}(\Theta))]$ is optimized using a standard gradient descent method leading to exceptional results as shown in [FIG. (1)]. The method was tested for approx 2000 digits of the number '3' from the MNIST database and the registration ratio was 98.2%. Some examples of cases where the method has failed are shown in [FIG. (1)] for demonstration purposes.

However, one can claim that the local deformation field is not sufficient to characterize the registration between two shapes. Often data is corrupted by noise while at the same time outliers exist in the training set. Therefore recovering measurements of the quality of the registration is an eminent condition for accurate shape modelling.

3 Estimation of Registration Uncertainties

Several attempts to build statistical models on noisy set of data in order to infer the properties of a certain model were proposed in the former literature. In [9], various techniques to extract feature points in images along with uncertainties due to the inherent noise were reported while in [12] an iterative estimation method was proposed to handle uncertainties estimates of rigid motion on sets of matched points. Last, but not least in [15] an iterative technique to determine uncertainties within the Iterated Closest Point [4] registration algorithm was proposed. In a quite different context, [14] introduced uncertainties within the estimation of dense optical flow, that can be seen as a form of registration between images.

In the present case curves are considered using implicit representation, therefore uncertainty does not lie in the relative position of points but of an isosurface and therefore one can seek for equivalences with "aperture problem" in optical flow estimation. Inspired by the work in [2, 15] we aim to recover uncertainties on the vector Θ while being able to use only the zero iso-surface, defining the shape itself. To this end, we use a discrete formulation of the Energy $E_0 = E_{\alpha_\infty}$:

$$E_0(\Theta) = \sum_{i=1}^{K} \rho(\phi_{\mathcal{T}}(\mathcal{L}(\Theta, \mathbf{x}_i)) = \sum_{i=1}^{K} \rho(\phi_{\mathcal{T}}(\mathbf{x}'_i))$$

Let us consider $\mathbf{q}_i$ to be the closest point on the target contour from $\mathbf{x}'_i$. Since $\phi_{\mathcal{T}}$ is assumed to be an Euclidean distance transform, it satisfies the condition $\|\nabla\phi_{\mathcal{T}}(\mathbf{x}'_i)\| = 1$. Therefore one can express the values of $\phi_{\mathcal{T}}(\mathbf{x}'_i)$:

$$\phi_{\mathcal{T}}(\mathbf{x}'_i) = (\mathbf{x}'_i - \mathbf{q}_i) \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}'_i)$$

Then, one has a first order approximation of $\phi_{\mathcal{T}}(\mathbf{x})$ in the neighborhood of $\mathbf{x}'_i$, in the form :

$$\begin{aligned} \phi_{\mathcal{T}}(\mathbf{x}'_i + \delta\mathbf{x}'_i) &= \phi_{\mathcal{T}}(\mathbf{x}'_i) + \delta\mathbf{x}'_i \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}'_i) \\ &= (\mathbf{x}'_i + \delta\mathbf{x}'_i - \mathbf{q}_i) \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}'_i) \end{aligned}$$

that reflects the condition that a point to curve distance is adopted rather than a point to point. Under the assumption that $E_0(\mathcal{L}(\Theta)) = \circ(1)$ we can neglect the second order term in the development of $\phi_{\mathcal{T}}$ and therefore write the following second order approximation of E_0 in quadratic form :

$$E(\mathcal{L}(\Theta)) = \sum \left[(\mathcal{L}(\Theta, \mathbf{x}_i) - \mathbf{q}_i) \cdot \nabla\phi_{\mathcal{T}}(\mathbf{x}'_i) \right]^2$$

Free form deformations is a linear transformation with respect to the parameters $\Theta = \delta\mathbf{P}_{i,j}$. Therefore one can rewrite this transformation over the image domain in a rather compact form:

$$\begin{aligned}\mathcal{L}(\Theta, \mathbf{x}) &= \mathbf{x} + \sum_{k=-1}^{2} \sum_{l=-1}^{2} B_k(u) B_l(v) \delta \mathbf{P}_{i+k,j+l} \\ &= \mathbf{x} + \mathcal{X}(\mathbf{x}).\end{aligned}$$

where $\mathcal{X}(\mathbf{x})$ is a matrix of dimensionality $2 \times N$ with N being the size of Θ. One now can substitute this term in the objective function towards :

$$E(\Theta) = (\mathcal{X} \cdot \Theta - \mathbf{y})^T (\mathcal{X} \cdot \Theta - \mathbf{y})$$

with

$$\mathcal{X} = \begin{pmatrix} \eta_1^T \mathcal{X}(\mathbf{x}_1) \\ \vdots \\ \eta_K^T \mathcal{X}(\mathbf{x}_K) \end{pmatrix} \text{ and } \mathbf{y} = \begin{pmatrix} \eta_1^T (\mathbf{q}_1 - \mathbf{x}_1) \\ \vdots \\ \eta_K^T (\mathbf{q}_K - \mathbf{x}_K) \end{pmatrix}$$

and $[\eta_i = \nabla \phi_{\mathcal{T}}(\mathbf{x}'_i)]$ due to the distance transform nature of the implicit function. We assume that $\mathbf{y}$ is the only random variable. Such assumption is equivalent with saying that errors in the point positions are only quantified along the normal direction. This accounts for the fact that the point set is treated as samples extracted from a continuous manifold. One can take the derivative of the objective function in order to recover a linear relation between Θ and $\mathbf{y}$:

$$\mathcal{X}^T \mathcal{X} \Theta = \mathcal{X}^T \mathbf{y}$$

Last, assume that the components of $\mathbf{y}$ are independent and identically distributed. In that case, the covariance matrix of $\mathbf{y}$ has the form $\sigma^2 \mathbf{I}$ of magnitude σ^2 with $\mathbf{I}$ being the identity. In the most general case one can claim that the matrix $\mathcal{X}^T \mathcal{X}$ is not invertible because due to the fact that the registration problem is underconstrained. Additional constraints are to be introduced towards the estimation of the covariance matrix of Θ through the use of an arbitrarily small positive parameter γ :

$$E(\Theta) = (\mathcal{X}\Theta - \mathbf{y})^T (\mathcal{X}\Theta - \mathbf{y}) + \gamma \, \Theta^T \Theta$$

Then the covariance matrix of the parameter estimate is :

$$\Sigma_\Theta = \sigma^2 (\mathcal{X}^T \mathcal{X} + \alpha \mathbf{I})^{-1}$$

Some example of such estimates are shown in [FIG 3]

Modelling the registered examples according to some density function is a step further to registration. To this end, two critical issues are to be addressed : the form of the *pdf* as well as the procedure to determine the corresponding parameters. In the most general case deformations of shapes that refer to objects of particular interest cannot be modeled with simple parametric models likes Gaussians. Therefore within our approach we propose a non-parametric form of the *pdf*.

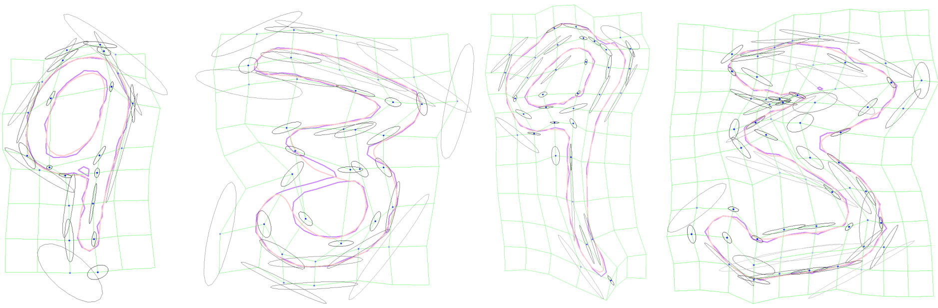

Figure 3: *Projection of the covariance matrix* Σ_Θ *on the grid points. The projections are* 2×2 *matrices and are represented as ellipses. Displaying the importance of the covariance coefficients between* 2 *different points is not possible.*

4 Variable Bandwidth Density Estimation

Let $t\{\mathbf{x}_i\}_{i=1}^M$ denote a random sample with common density function f. The fixed bandwidth kernel density estimator consists of:

$$\hat{f}(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M} \mathrm{K}_{\mathbf{H}}\left(\mathbf{x}-\mathbf{x}_i\right)$$
$$= \frac{1}{M}\sum_{i=1}^{M} \frac{1}{\|\mathbf{H}\|^{1/2}} \mathrm{K}\left(\mathbf{H}^{-1/2}(\mathbf{x}-\mathbf{x}_i)\right)$$

where $\mathbf{H}$ is a symmetric definite positive - often called a bandwidth matrix - that controls the width of the kernel around each sample point $\mathbf{x}_i$. The fixed bandwidth approach often produces an undersmoothing in areas with sparse observations and oversmoothing in the opposite case. Usefulness of varying bandwidths is widely acknowledged to estimate long-tailed or multi-modal density functions with kernel methods.

In the literature, Kernel density estimation methods that do rely on such varying bandwidths are generally referred to as " adaptive kernel " density estimation methods [17]. An adaptive kernel approach adapts to the sparseness of the data by using a broader kernel over observations located in regions of low density. Two useful state of-the-art variable bandwidth kernels consists of the *sample point estimator* and the *balloon estimator*.

The first one refers to a covariance matrix depending on the repartition of the points constituting the sample :

$$\hat{f}_S(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{\|\mathbf{H}(\mathbf{x}_i)\|^{1/2}} \mathrm{K}\left(\mathbf{H}(\mathbf{x}_i)^{-1/2}(\mathbf{x}-\mathbf{x}_i)\right)$$

where a common selection of $\mathbf{H}$ refers to

$$\mathbf{H}(\mathbf{x}_i) = h(\mathbf{x}_i) \cdot \mathbf{I}$$

with $h(\mathbf{x}_i)$ being the distance of point $\mathbf{x}_i$ from the k^{th} nearest point. One can consider various alternatives to determine the bandwidth function. Such estimator may be directly used with the uncertainties calculated in section 3 and $\mathbf{H}(\mathbf{x}_i) = \mu\Sigma_{\Theta_i}$ as proposed in [3, 6].

In the present paper we may have an estimation of the uncertainty on the point to be evaluated. In order to make use of this information we first introduce another standard variable bandwidth kernel method knows as *balloon estimator*. It adapts the measures to the point of estimation depending on the shape of the sampled data according to:

$$\hat{f}_B(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{\|\mathbf{H}(\mathbf{x})\|^{1/2}} \mathrm{K}\left(\mathbf{H}(\mathbf{x})^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right)$$

with $\mathbf{H}(\mathbf{x})$ may be chosen with the same model as *sample point estimator*. Such function may be seen as the average of a density associated to the estimation point $\mathbf{x}$ on all the sample points $\mathbf{x}_i$. One should point out that such a process could lead to estimates on $\hat{f}(\mathbf{x})$ that do not refer to density function in terms of discontinuity, integration to infinity, etc.

Let us consider $\{\mathbf{x}_i\}_{i=1}^{M}$ a multi-variate set of measurements where each sample $\mathbf{x}_i$ exhibits uncertainties in the form of a covariance matrix Σ_i. Our objective can be stated as follows: estimate the probability of a new measurement $\mathbf{x}$ that is associated with covariance matrix Σ.

Let $\mathbf{X}$ be the random variable associated to the training set and assume a density function f. f may be estimated with $\hat{f}$ in a similar fashion to *sample point* estimator. Therefore f may be expressed in the form $f = \sum f_i$ where f_i are densities associated to a single kernel $\mathbf{x}_i$. Let $\mathbf{Y}$ be the be a random variable for the new sample with density g.

Then one can claim that in order to estimate the probability of the new sample, one should first determine for all possible $\mathbf{u} \in \mathbb{R}^N$ their *distance* from the existing kernels of the trainning set $\mathbf{X}$, $f(\mathbf{u})$ in a similar fashion as *sample point estimator* and weight them according to their fit with the density function of $\mathbf{Y}$:

$$\begin{aligned} f(\mathbf{x}) &= \int f(\mathbf{u})g(\mathbf{u})d\mathbf{u} \\ &= \int \left[\sum_{i=1}^{M} f_i(t)\right] g(t)dt = \sum_{i=1}^{M}\left[\int f_i(t)g(t)dt\right] \end{aligned}$$

In that case of gaussian kernels for g and the f_i the following expression is recovered:

$$\hat{f}_G(\mathbf{x}) = \frac{1}{M(2\pi)^{d/2}} \sum_{i=1}^{M} \frac{1}{\|\Sigma + \Sigma_i\|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_i))^\top (\Sigma + \Sigma_i)^{-1}(\mathbf{x} - \mathbf{x}_i)\right)$$

Such an expression has a simple mathematical interpretation: Consider two points $\{\mathbf{x}_1, \mathbf{x}_2\}$ with associated uncertainty $\{\Sigma_1, \Sigma_2\}$. Assuming that these are the parameters (mean and variance) of two independent random variables with normal distribution

$$\{\mathbf{X}_1 \sim N(\mathbf{x}_1, \Sigma_1), \mathbf{X}_2 \sim N(\mathbf{x}_2, \Sigma_2)\}$$

Then the random variable $\mathbf{Z} = \mathbf{X}_1 - \mathbf{X}_2$ follows a distribution $N(\mathbf{x}_1 - \mathbf{x}_2, \Sigma_1 + \Sigma_2)$ and the density at $\mathbf{Z} = 0$ is given by

$$p(\mathbf{X}_1 = \mathbf{X}_2) = \frac{1}{(2\pi)^{d/2}\|\Sigma_1 + \Sigma_2\|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2))^T (\Sigma_1 + \Sigma_2)^{-1}(\mathbf{x}_1 - \mathbf{x}_2)\right)$$

The present concept could be relaxed to address the case of non-gaussians kernels according to a *hybrid* estimator that is considered in the present paper :

$$\hat{f}_H(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{\|\mathbf{H}(\Sigma, \Sigma_i)\|^{1/2}} \mathrm{K}(\mathbf{H}(\Sigma, \Sigma_i)^{1/2}(\mathbf{x} - \mathbf{x}_i))$$

Such a density estimator takes into account the uncertainty estimates both on the sample points themselves as well as on the estimation of point $\mathbf{x}$ as introduced in [11]. The outcome of this estimator may be seen as the average of the probabilities that estimation measurement is equal to the sample measurement, calculated over all sample measurements. Consequently, in directions of important uncertainties the density estimation decreases more slowly when compared to the other directions.

This metric can now be used to assess for a new sample the probability of being part of the training set through a process that evaluates the probability for each of the examples in the training set. The resulting approach can account for the non-parametric form of the observed density while the limitation of being time consuming since the cost is linear to the number of samples in the training set. Therefore, there is an eminent need on decreasing the cardinality of the set of retained kernels.

4.1 Kernels Selection

The maximum likelihood criterion expresses the quality of approximation from the model to the data.

Consider a set $\mathcal{Z}_K = \{X_1, X_2, \ldots, X_K\}$ of kernels extracted from the training set. These have associated mean and uncertainties $\{\mathbf{x}_i \Sigma_i\}_{i=1}^{|\mathcal{Z}_K|}$. Then the probability of any registered shape with associated kernel Y has the form :

$$P_{\mathcal{Z}_K}()(Y) = \frac{1}{|Z_K|} \sum_{X \in Z_K} K(X, Y)$$

and K(X,Y) correspond to the calculation of the hybrid kernel estimator. For such a selection of kernels, one can evaluate the log-likelihood for the entire training set with the associated kernels $\{Y_i\}_{i=1}^{N}$:

$$C_K = \sum_{i=1}^{N} \log(P_{\mathcal{Z}_K}(Y_i))$$

We use an efficient sub-optimal iterative algorithm to update the set $\mathcal{Z}_K$. A new kernel Y is extracted from the training set as the one maximizing the quantity C_{K+1} associated to $\mathcal{Z}_{K+1}$ with : $\mathcal{Z}_{K+1} = \mathcal{Z}_K \bigcup Y$. One kernel may be chosen several times in order to preserve a decreasing order of C_K when adding new kernels. Consequently the selected kernels Y_i used to evaluate the global density probability have prior weight.

4.2 Validation

The proposed method is indented to provide efficient models for family of shapes with important variation. Digits, is an example where the shape of the characters varies along individuals and therefore one can claim important variability on the training set. Based on this observation and using an important training set from the database, we have considered two digits (random variables of 2000 samples each) that have a quite similar structure, the 3 and 9.

Upon intra-class registration two models have been built of 100 kernels each according to the maximum likelihood principle. Then, a cross validation task was performed where for all samples of the database (3 & 9) the probability of being part of the classes 3 & 9 was estimated according to the presented variable bandwidth density function. In [FIG. (4)] one can see in a logarithmic scale the performance of the method using the model built for 3 and applied also to the samples of 9 while the opposite case is presented in [FIG. (4)].

In both cases one can see a clear separation of the two classes and a substantial difference in terms of probabilities between the true and the non-true case. It is

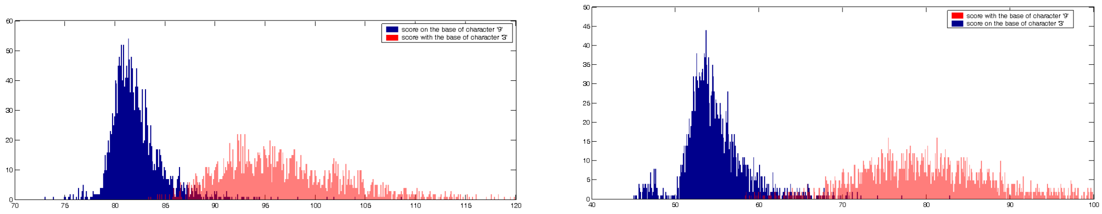

Figure 4: *(left) Distribution of the distance of the training set from the kernel based model build for '3' in logarithmic scale. (right)Distribution of the distance of the training set from the kernel based model build for '9' in logarithmic scale.*

important to note that the presented method is not indented for such an application. However, given this validation we can claim that such a model can capture samples of increasing complexity and the use of deformations along with uncertainties provide efficient density estimators.

5 Discussion

We have introduced an original framework to estimate uncertainty in the process of registration on shapes. We take advantage of this additional knowledge to build an efficient probabilistic descriptor of a certain class of shapes that can be registered to a common pose.

Future directions exists in the registration as well as the modeling aspect of our approach.

First, in the registration process, uncertainties could be propagated through scale when updating the transformation. We shall also notice that the uncertainties calculated on a certain ffd-grid could be extended to any finer grid and therefore qualify the density probability of any image transformation without the limitation of the choice of parameters.

Another path will be the exploration of the kernel used to make a Parzen-Window like density estimation into more advanced kernel-based learning methods such as kernel-PCA. The issue of defining the right Mercer kernel has however to be addressed.

Last but not least, this evaluation of densities using uncertainty has to be exported to the more general problem of image registration with prior knowledge. Consider an original image used as a model with the region of interest manually delineated. Then, registration can be performed with a shape term that directly handle the parameters of the transformation $\mathcal{L}$. Eventually, a calculation of uncertainties qualifying the present image registration may enhance the confidence for this term when using the *hybrid* estimator.

References

[1] G. Borgefors. Distance Transformations in Digital Images. *Computer Vision, Graphics, and Image Processing*, 34:344–371, 1986.

[2] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum. A Feature-Based, Robust, Hierarchical Algorithm for Registering Pairs of Images of the Curved Human Retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:347–364, 2002.

[3] H. Chen and P. Meer. Robust computer vision through kernel density estimation. In *ECCV (1)*, pages 236–250, 2002.

[4] Y. Chen and G. Medioni. Object Modeling by Registration of Multiple Range Images. *Image and Vision Computing*, 10:145–155, 1992.

[5] H. Chui and A. Rangarajan. A New Algorithm for Non-Rigid Point Matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages II: 44–51, 2000.

[6] D. Comaniciu. Nonparametric information fusion for motion estimation. In *CVPR*, volume 1, pages 59–66, 2003.

[7] X. Huang, N. Paragios, and D. Metaxas. Establishing Local Correspondences towards Compact Representations of Anatomical Structures. In *Medical Imaging Copmuting and Computer-Assisted Intervention*, 2003.

[8] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[9] K. Kanatani. Uncertainty modeling and model selection for geometric inference. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(10):1307–1319, 2004.

[10] B. Kimia, A. Tannenbaum, and S. Zucker. Shocks and Deformations i: The Components of two-dimensional Shape and the reaction-diffusion space. *International Journal of Computer Vision*, 15:189–224, 1995.

[11] A. Mittal and N. Paragios. Motion-based background substraction using adaptive kernel density estimation. In *Computer Vision and Pattern Recognition*, volume 2, pages 302–309, 2004.

[12] X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3-d registration methods based on points and frames. *Int. J. Comput. Vision*, 25(3):203–229, 1997.

[13] T. Sederberg and S. Parry. Free-form deformation of solid geometric models. *Proceedings SIGGRAPH '86*, 20:151–160, 1986.

[14] E. Simoncelli. Bayesian multi-scale differential optical flow. In B Jähne, H Haussecker, and P Geissler, editors, *Handbook of Computer Vision and Applications*, volume 2, chapter 14, pages 397–422. Academic Press, San Diego, April 1999.

[15] C. Stewart, C.-L. Tsai, and B. Roysam. The dual bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans. Med. Img.*, 22:1379–1394, 2003.

[16] A. Stoddart, S. Lemke, A. Hilton, and T. Renn. Estimating Pose Uncertainty for Surface Registration. *Image and Vision Computing*, 16:111–120, 1998.

[17] M. Wand and M. Jones. *Kernel Smoothing*. Chapman & Hall, 1995.

[18] Z. Xie and G. Farin. Deformation with hierarchical b-splines. *Mathematical Methods for Curves and Surfaces: Oslo 2000*, pages 545–554, 2001.

[19] L. Younes. Computable elastic distances between shapes. *SIAM Journal of Applied Mathematics*, 58:565–586, 1998.

[20] Z. Zhang. Iterative Point Matching for Registration of Free-Form Curves and Surfaces. *International Journal of Computer Vision*, 13:119–152, 1994.