Chapter 11
Applying Random Forests to the Problem of Dense Non-rigid Shape Correspondence

Matthias Vestner, Emanuele Rodolà, Thomas Windheuser, Samuel Rota Bulò, and Daniel Cremers

Abstract We introduce a novel dense shape matching method for deformable, three-dimensional shapes. Differently from most existing techniques, our approach is general in that it allows the shapes to undergo deformations that are far from being isometric. We do this in a supervised learning framework which makes use of training data as represented by a small set of example shapes. From this set, we learn an implicit representation of a shape descriptor capturing the variability of the deformations in the given class. The learning paradigm we choose for this task is a random forest classifier. With the additional help of a spatial regularizer, the proposed method achieves significant improvements over the baseline approach and obtains state-of-the-art results while keeping a low computational cost.

11.1 Introduction

Matching three-dimensional shapes is a pervasive problem in computer vision, computer graphics and several other fields. Nevertheless, while the advances made by works such as [2, 4, 10, 14, 23, 29] have been dramatic, the problem is far from being solved.

Many methods in shape matching use a notion of similarity that is defined on a very general set of possible shapes. Due to the highly ill-posed nature of the shape matching problem, it is very unlikely that a general method will reliably find good matchings between arbitrary shapes. In fact, while many matching methods (such as methods based on metric distortion [4, 20, 22] and eigen-decomposition of the Laplacian [2, 23, 29]) mostly capture near-isometric deformations, others might consider too general deformations which are not consistent with the human

M. Vestner (✉) • E. Rodolà • T. Windheuser • D. Cremers
Technische Universität München, Munich, Germany
e-mail: matthias.vestner@in.tum.de

S.R. Bulò
Fondazione Bruno Kessler, Trento, Italy

© Springer International Publishing Switzerland 2016

M. Breuß et al. (eds.), *Perspectives in Shape Analysis*, Mathematics and Visualization, DOI 10.1007/978-3-319-24726-7_11

intuition of correspondence. In applications where the class of encountered shapes is in-between, adapting the matching methods at hand is often very tedious.

In this paper we try to bridge the gap between general shape matching methods and application-specific algorithms by taking a learning-by-examples approach.

In our scenario, we assume to have a set of training shapes which are equivalent up to some class of non-isometric deformations. Our goal is to learn from these examples how to match two shapes falling in the equivalence class represented by the training set. To this end, we treat the shape matching problem as a classification problem, where input samples are points on the shape manifold and the output class is an element of a canonical label set, which might e.g. coincide with the manifold of one of the shapes in the training set. A first contribution of this paper consists in a new random forest classifier, which can tackle this unconventional classification problem in an efficient and effective way, starting from a general parametrizable shape descriptor. Our classifier is designed in a way to randomly explore the descriptor's parametrization space and find the most discriminative features that properly recover the transformation map characterizing the shape category at hand. In this work, we consider the *wave kernel signature* (WKS) [2] as the shape descriptor. This descriptor is known to be invariant to isometric transformations, but the forest can effectively exploit it to match shapes that undergo non-rigid and non-isometric deformations.

In some sense, the output of the random forest can be seen as a new descriptor by itself that is tuned to the shapes and deformations appearing in the training set. In this respect, the proposed method is complementary to existing shape descriptors as it can improve the performance of a given descriptor [11, 12, 32]. Early attempts to apply machine learning techniques to the problem of non-rigid correspondence [25, 28] consider shapes represented by signed distance functions. We follow the intrinsic view point, considering shapes given by their boundary surface, seen as a Riemannian manifold.

One of the main benefits of our approach is the fact that the random forest classifier gives for each point on the shape an ordered set of matching candidates, hence delivering a dense point-to-point matching. Since such a descriptor does not include any spatial regularity, we propose to use a regularization technique along the lines of the *functional maps framework* [16]. We experimentally validate that the proposed learning approach improves the underlying general descriptor significantly, and it performs better than other state-of-the-art matching algorithms on equivalent benchmarks.

An earlier version of this work was published in [21].

11.1.1 Intrinsic Point Descriptors

We consider 3D shapes that are represented by their boundary surface, a two-dimensional Riemannian manifold (M, g) without boundary. A point descriptor is a function ϕ that assigns to each point on the surface an element of a metric space

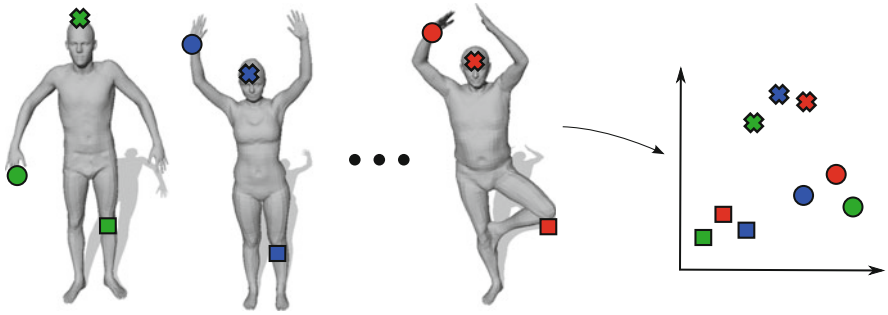

Fig. 11.1 A good point descriptor should at the same time assign similar values to corresponding points on deformed shapes and dissimilar values to non-corresponding points

D, the descriptor space. A good point descriptor should satisfy two competing properties (Fig. 11.1):

- **deformation-invariance:** it should assign similar values to corresponding points on deformed shapes
- **discriminativity:** it should well distinguish non-corresponding points

While it is in principle possible to construct a descriptor that is invariant under an arbitrary large class of deformations (e.g. the constant function), it is evident that there will always be a tradeoff between deformation-invariance and discriminativity.

The descriptors we consider are based on the spectrum of the *Laplace-Beltrami operator* $\Delta_M = -\mathrm{div}_M(\nabla_M)$. Being a symmetric operator the spectrum of Δ_M consists of real eigenvalues $\lambda_1, \lambda_2, \ldots$ and the corresponding eigenfunctions $\gamma_1, \gamma_2, \ldots$ can be chosen to be real valued and orthonormal. Moreover, Δ_M is a non-negative operator with a one-dimensional kernel and a compact pseudo-inverse, so we can order the eigenvalues $0 = \lambda_1 < \lambda_2 \leq \ldots$ and assign to each point $x \in M$ a vector $p \in \mathbb{R}^{2K}$, $p = (\lambda_1, \ldots, \lambda_K, \gamma_1(x), \ldots, \gamma_K(x))$. The Laplace Beltrami Operator is purely intrinsic as it is uniquely determined by the metric tensor $g = (g_{ij})_{i,j=1}^2$ (respectively its inverse $(g^{ij})_{i,j=1}^2$):

$$\Delta_M = \frac{1}{\sqrt{\det g}} \sum_{i,j=1}^{2} \frac{\partial}{\partial x_i} \left(g^{ij} \sqrt{\det g} \frac{\partial}{\partial x_j} \right). \tag{11.1}$$

As a consequence the eigenvalues λ_k as well as the corresponding eigenspaces do not change whenever a shape undergoes an isometric deformation. The eigenbases however are not uniquely determined, even in the case of one dimensional eigenspaces the normalized eigenvectors are only unique up to sign. Nevertheless from the representation p it is possible to construct descriptors that are invariant under isometric deformations. Given a collection $(t_i)_{i=1}^n$ of positive numbers, the

Heat Kernel Signature (HKS)

$$HKS(p) = \left(\sum_k \exp(-\lambda_k t_i) \gamma_k(x)^2 \right)_{i=1}^{n} \in \mathbb{R}^n \tag{11.2}$$

is a n-dimensional intrinsic point-descriptor [29]. From a physical point of view each component tells us how much heat $u(x, t_i)$ remains at point x after time t_i when the initial distribution of heat is a unit heat source at the very same point:

$$\Delta u = u_t \tag{11.3}$$

$$u(0, \cdot) = \delta_x \tag{11.4}$$

Since the class of isometric deformations includes reflections, any intrinsic descriptor will assign identical values to a point and its symmetric counterpart, whenever shapes exhibit bilateral intrinsic symmetries. Using information about the symmetry [18] or making use of extrinsic information as in [27] would overcome this problem.

From a signal processing viewpoint HKS can be seen as a collection of low-pass filters and thus it is not appropriate to localize features, see Fig. 11.2. Motivated by this observation Aubry et al. [2] introduced the Wave Kernel Signature (WKS), a descriptor where the low-pass filters are replaced by band pass filters:

$$WKS(p) = \left(\sum_k f_{(e_i, \sigma_i^2)}(\lambda_k)^2 \gamma_k(x)^2 \right)_{i=1}^{n} \in \mathbb{R}^n \tag{11.5}$$

Here the parameters (e_i, σ_i^2) correspond to mean and variance of the log-normal energy distributions

$$f_{(e,\sigma^2)}(\lambda) \propto \exp(-\frac{(\log e - \log \lambda)^2}{2\sigma^2}) \tag{11.6}$$

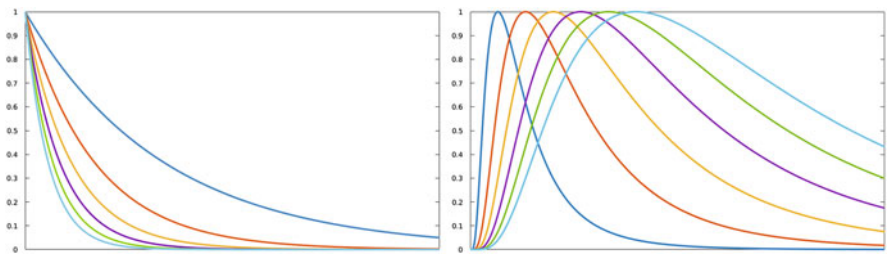

Fig. 11.2 The weighting functions of the heat kernel signature (*left*) can be seen as low-pass filters, the ones of the wave kernel signature (*right*) in contrary behave like band-pass filters

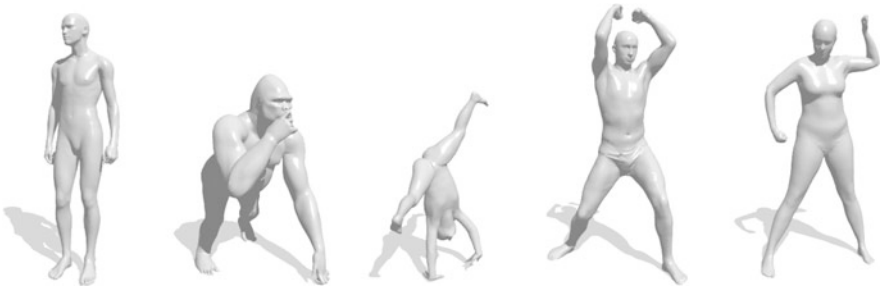

Fig. 11.3 Finding a correspondence between shapes should be feasible even if they are far from being isometric

The authors propose fixed values for the parameters (e_i, σ_i) depending on the truncated spectrum of the Laplace-Beltrami-operator. Moreover they equip the descripor with a metric related to the L^1-distance.

In this work the parameters will be learned from training data, a distance function between vector valued descriptors is unneeded since descriptors are compared component wise in a hierarchical manner (Sects. 11.2.1.1 and 11.2.1.3).

Both, HKS and WKS, are invariant under isometric deformations. However the human notion of similarity by far exceeds the class of isometries. Asking for a correspondence between an adult and a child or even an animal like a gorilla is a feasible task for us. Figure 11.3 shows examples of shapes taken from different datasets [1, 5, 19, 21] that could in principle be put into correspondence. By choosing application dependent parameters one can achieve descriptors that are less sensitive to the type of deformation one is interested in. In this work we implicitly determine optimal parameters when the deformation class is represented by a set of training shapes with known ground truth correspondence.

11.1.2 Discretized Surfaces and Operators

In practice the shapes are given as triangular meshes $M = (V_M, F_M)$. We will henceforth identify a shape M by the set of it vertices V_M. A one-to-one correspondence between two shapes can then be represented by a permutation matrix, a fuzzy correspondence, i.e. a function that assigns to each point a probability distribution over the other shape, respectively as a left-stochastic matrix. Functions defined on a shape become vectors and linear operators acting on them, e.g. the Laplace-Beltrami operator can be written as matrices. Inner products between functions are calculated via an area-weighted inner product between the vectors representing them. We chose the popular cotangent scheme [15] as the discretization of the Laplacian.

11.2 Dense Correspondence Using Random-Forests

In this work we treat the shape matching problem as a classification problem, where input samples are points on the shape and the output class is an element of a canonical label set, which might e.g. coincide with one of the shapes in the training set (the *reference shapes*). The classifier we choose is a Random forest, designed in a way to randomly explore the descriptor's parametrization space and find the most discriminative features that properly recover the transformation map characterizing the shape category at hand. In this work, we consider the *wave kernel signature* (WKS) as the parametrizable point descriptor (weak classifier). In general other choices of parametrizable descriptors, e.g. HKS, are possible. As mentioned in Sect. 11.1.1 any classifier based on isometry-invariant point descriptors can not distinguish a point from its symmetric counterpart. Thus the fuzzy outcome of the Random forest classifier has to be regularized in order to get a consistent correspondence.

11.2.1 Learning and Inference Using Random Forests

Random forests [3] are ensembles of decision trees that have become very popular in the computer vision community to solve both classification and regression problems with applications ranging from object detection, tracking and action recognition [9] to semantic image segmentation and categorization [26], and 3D pose estimation [30], to name just a few. The forest classifier is particularly appealing because its trees can be trained efficiently and techniques like bagging and randomized feature selection allow to limit the correlation among trees and thus ensure good generalization. We refer to [7] for a detailed review.

11.2.1.1 Inference

In the context of shape matching, a decision tree comprised by the forest routes a point m of a test shape M from the root of the tree to a leaf node, where a probability distribution defined on a discrete label set L is assigned to the point. The path from the root to a leaf node is determined by means of binary decision functions called *split functions* located at the internal nodes, which given a shape point return L or R depending on whether the point should be forwarded to the left or to the right with respect to the current node. According to this inference procedure, each tree $t \in \mathscr{F}$ of a forest $\mathscr{F}$ provides a posterior probability $\mathrm{P}(\ell|m, t)$ of label $\ell \in L$, given a point $m \in M$ in a test shape M (Fig. 11.4).

This probability measure is the one associated with the leaf of tree $t \in \mathscr{F}$ that the shape point would reach. The prediction of the whole forest $\mathscr{F}$ is finally obtained

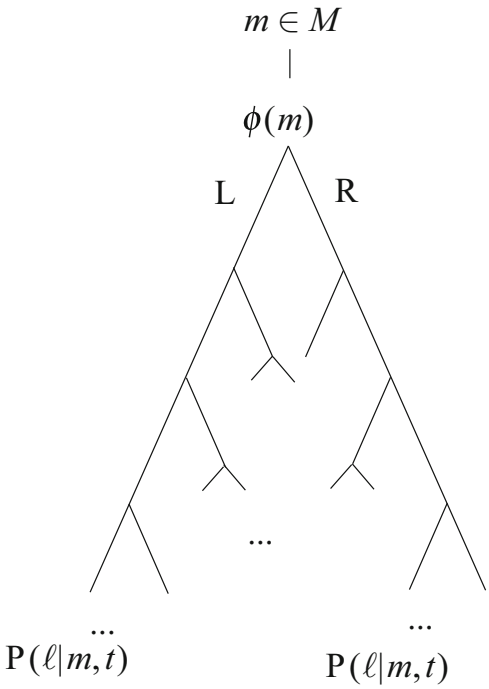

Fig. 11.4 At each inner node of a decision tree a binary split function is evaluated. Depending on the result the point m is either routed to the left or to the right. Leafs of the tree correspond to probability distributions in the label space. A random forest is a collection of mulitple decision trees

by averaging the predictions of the single trees:

$$\mathrm{P}(\ell|m,\mathscr{F}) = \frac{1}{|\mathscr{F}|}\sum_{t\in\mathscr{F}} \mathrm{P}(\ell|m,t) \ . \tag{11.7}$$

The outcome of the prediction over an entire shape M can be represented as a left-stochastic matrix X_M encoding the probabilistic canonical transformation, where

$$(\mathrm{X}_M)_{ij} = \mathrm{P}\left(\ell_i|m_j,\mathscr{F}\right) \tag{11.8}$$

for each $\ell_i \in L$ and $m_j \in M$. Using Bayes' theorem we can further construct a fuzzy correspondence between two previously unseen shapes (i.e. no members of the training set).

11.2.1.2 Learning

During the learning phase, the structure of the trees, the split functions and the leaf posteriors are determined from a training set. Let $\{(R_i, T_i)\}_{i=1}^{\mathsf{m}}$ be a set of m reference shapes R_i each equipped with a canonical transformation, i.e. a bijection $T_i : R_i \to L$ between the vertex set of the reference shape and the label set L. A training set $\mathbb{T}$

for the random forest is given by the union of the graphs of the mappings T_i, i.e.

$$\mathbb{T} = \{(\boldsymbol{r}, T_i(\boldsymbol{r})) \,:\, \boldsymbol{r} \in R_i,\ 1 \leq i \leq \mathsf{m}\} \ . \tag{11.9}$$

The learning phase that creates each tree forming the forest consists in a recursive procedure that starting from the root iteratively splits the actual terminal nodes. During this process each shape point of the training set is routed through the tree in a way to partition the whole training set across the terminal nodes. The decision whether a terminal node has to be further split and how the splitting will take place is purely local as it involves exclusively the shape points that have reached that node. A terminal node typically becomes a leaf of the tree if the depth of the node exceeds a given limit, if the size of the subset of training samples reaching the node is small enough, or if the entropy of the sample's label distribution is low enough. If this is the case, then the leaf node is assigned the label distribution of subset $\mathbb{S}$ of training samples that have reached the leaf, i.e.

$$\mathrm{P}\left(\ell|\mathbb{S}\right) = \frac{|\{(\boldsymbol{r}, \ell) \in \mathbb{S}\}|}{|\mathbb{S}|} \ . \tag{11.10}$$

The probability distribution $\mathrm{P}\left(\cdot|\mathbb{S}\right)$ will become the posterior probability during inference for every shape point reaching the leaf. Consider now the case where the terminal node is split. In this case, we have to select a proper split function $\psi(r) \in \{\mathsf{L}, \mathsf{R}\}$ that will route a point r reaching the node to the left or right branch. An easy and effective strategy for guiding this selection consists in generating a finite pool Ψ of random split functions and retaining the one maximizing the information gain with respect to the label space L. The information gain $\mathrm{IG}\left(\psi\right)$ due to split function $\psi \in \Psi$ is given by the difference between the entropy of the node posterior probability defined as in (11.10) before and after having performed the split. In detail, if $\mathbb{S} \subseteq \mathbb{T}$ is the subset of the training set that has reached the node to be split and $\mathbb{S}^{\mathsf{L}}, \mathbb{S}^{\mathsf{R}}$ is the partition of $\mathbb{S}$ induced by the split function ψ then $\mathrm{IG}\left(\psi\right)$ is given by

$$\mathrm{IG}\left(\psi\right) = \mathrm{H}\left(\mathrm{P}\left(\cdot|\mathbb{S}\right)\right) - \mathrm{H}\left(\mathrm{P}\left(\cdot|\mathbb{S}\right)|\psi\right) \ , \tag{11.11}$$

where $\mathrm{H}\left(\cdot\right)$ denotes the entropy and

$$\mathrm{H}\left(\mathrm{P}\left(\cdot|\mathbb{S}\right)|\psi\right) = \frac{|\mathbb{S}^{\mathsf{L}}|}{|\mathbb{S}|}\mathrm{H}\left(\mathrm{P}\left(\cdot|\mathbb{S}^{\mathsf{L}}\right)\right) + \frac{|\mathbb{S}^{\mathsf{R}}|}{|\mathbb{S}|}\mathrm{H}\left(\mathrm{P}\left(\cdot|\mathbb{S}^{\mathsf{R}}\right)\right) \ . \tag{11.12}$$

Intuitively the information gain of a split function is higher, the better it seperates members belonging to different classes (see Fig. 11.5).

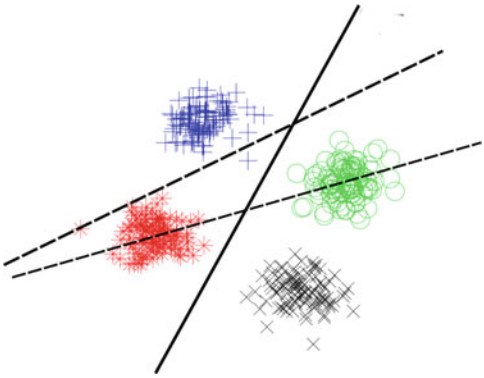

Fig. 11.5 The split function visualized as a *solid line* has the highest information gain (IG) among the three candidates

11.2.1.3 Choice of Decision Functions

During the build up of the forest the randomized training approach allows us to vary the parametrization of the shape descriptor for each point of the shape. In fact, we can in principle let the forest automatically determine the optimal discriminative features of the chosen descriptor for the matching problem at hand. In this work we have chosen the Wave Kernel Signature (WKS) but as mentioned above, in principle any parametrizable feature descriptor (e.g. HKS) can be considered. From a practical perspective, it can be shown [2] that the sum in (11.5) can be restricted to the first $\overline{k} < \infty$ components. We make explicit in (11.5) the dependency on $\overline{k}$ by writing:

$$p(m; e, \overline{k}) = \sum_{k=1}^{\overline{k}} f_e^2(\nu_k)\phi_k^2(m) \,. \tag{11.13}$$

We are now in the position of generating at each node of a tree during the training phase a pool of randomized split functions by sampling an energy level e_i, a number of eigenpairs $\overline{k}_i$ and a threshold τ_i. Accordingly, the split functions will take the form:

$$\psi_i(m) = \begin{cases} \mathsf{L} & \text{if } p(m; e_i, \overline{k}_i) > \tau_i \\ \mathsf{R} & \text{otherwise} \,. \end{cases} \tag{11.14}$$

11.2.2 Interpretation and Regularization of the Forests Prediction

The simplest way to infer a correspondence from a forest prediction consists in assigning each point $m \in M$ to the most likely label according to its final distribu-

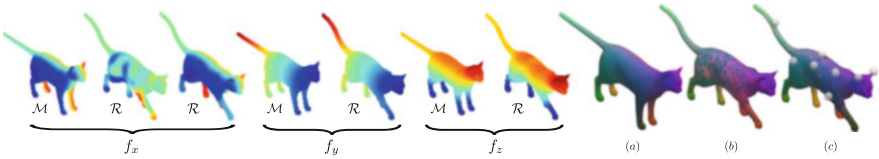

Fig. 11.6 The coordinate functions from a test shape M (*standing cat*) are transferred to a reference shape R (*walking cat*) via the functional map $T_{X_{M,R}}$ induced by the forest prediction. Most of the ambiguities arise in f_x, and are due to the global intrinsic symmetry of the cat. The *first column* shows the map f_x on the test cat, while the *second* and *third columns* are obtained by mapping f_x without and with regularization respectively. The remaining four columns show the mappings of f_y and f_z *without* regularization. The symmetric ambiguities disappear as a result of the regularization process (columns (*a*)–(*c*), matches encoded by color)

tion, i.e., the label maximizing $\mathrm{P}(\ell|\boldsymbol{m}, \mathscr{F})$. If we are also given a reference shape R from the training set, the maximum a posteriori estimate of ℓ can be transformed into a point-to-point correspondence from M to R via the known bijection $T : R \to L$. Figure 11.6a, b show an example of this approach. The resulting correspondence is exact for about 50 % of the points, whereas it induces a large metric distortion on the rest of the shape. However, this is not a consequence of the particular criterion we adopted when applying the prediction. Indeed, the training process can not distinguish symmetric points and is oblivious to the underlying manifolds as it is only based on pointwise information: the correspondence estimates are taken *independently* for each point and thus the metric structure of the test shape is not taken into account during the regression. Nevertheless, as we shall see, the predicted distributions carry enough information that can be exploited to obtain a consistent matching.

11.2.2.1 Functional Maps

Multiplying X_M (as defined in (11.8)) from the left with the permutation matrix associated to the known bijection $T : L \to R$ between the label space L and a reference shape R gives raise to another left-stochastic matrix $X_{M,R}$. As pointed out in [16] this (fuzzy) correspondence $X_{M,R}$ can be interpreted as a linear map $T_{X_{M,R}} : L^2(M) \to L^2(R)$. In Fig. 11.6 (first 7 columns) we use such a construction to map the coordinate functions $f_i : M \to \mathbb{R}$ (where $i \in \{x, y, z\}$) to scalar functions on R. Specifically, we plot $\boldsymbol{f}_i$ and their reconstructions $\boldsymbol{g}_i = T_{X_{M,R}}\boldsymbol{f}_i$. Note that the reference shape is axis-aligned, so that the x coordinates of its points grow from the right side (blue) to the left side of the model (red).

As in [16] from now on we consider $T_{X_{M,R}}$ in the truncated harmonic bases on the respective shapes and by that dramatically reduce the size of the problem. Since the LB-eigenfunctions are chosen to form orthonormal bases, the norms considered in the following section are invariant under this basis-transform. For simplicity we will still denote the associated matrix by $X_{M,R}$.

11.2.2.2 Metric Distortion Using Functional Maps

The plots we show in Fig. 11.6 tell us that most of the error in the correspondence arises from the (global) intrinsic symmetries of the shape. As mentioned previously, this is to be expected since the training process does not exploit any kind of structural information about the manifolds.

This suggests the possibility to regularize the prediction by introducing metric constraints on the correspondence. Specifically, we consider an objective of the form

$$E(\mathtt{X}) = c(\mathtt{X}_{M,R}, \mathtt{X}) + \rho(\mathtt{X}) \,, \tag{11.15}$$

where $\mathtt{X}$ is a correspondence between shapes M and R. The first term (or *cost*) ensures closeness to the prediction given by the forest, while the second term is a regularizer giving preference to geometrically consistent solutions.

A functional map is assumed to be geometrically consistent if it approximately preserves distance maps. Suppose for the moment we are given a sparse collection of matches $O \subset M \times R$. Then for each $(\boldsymbol{p}, \boldsymbol{q}) \in O$ we can define the two distance maps $d_{\boldsymbol{p}} : M \to \mathbb{R}$ and $d_{\boldsymbol{q}} : R \to \mathbb{R}$ as

$$d_{\boldsymbol{p}}(\boldsymbol{x}) = d_M(\boldsymbol{p}, \boldsymbol{x}) \,, \qquad d_{\boldsymbol{q}}(\boldsymbol{y}) = d_R(\boldsymbol{q}, \boldsymbol{y}) \,. \tag{11.16}$$

With these definitions, we can express the regularity term $\rho(\mathtt{C})$

$$\rho(\mathtt{C}) = \sum_{(\boldsymbol{p},\boldsymbol{q}) \in O} \omega_{\boldsymbol{pq}} \| X_{M,R} \boldsymbol{d}_{\boldsymbol{p}} - \boldsymbol{d}_{\boldsymbol{q}} \|_2^2 \,, \tag{11.17}$$

with weights $\omega_{\boldsymbol{pq}} \in [0, 1]$ (Fig. 11.7).

In order for the regularization to work as expected, the provided collection of matches should constrain well the solution, in the sense that it should help to

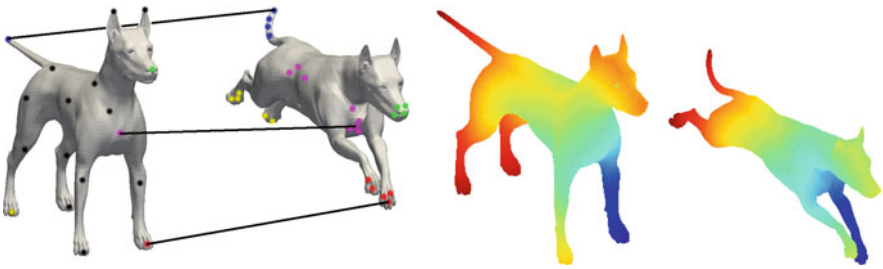

Fig. 11.7 In the regularization step first a coarse subsampling of the shape is constructed via Euclidean farthest point sampling (dots on the left shape). In the small set of predicted matches O (cross product of dots on the two shapes) a sparse correspondence is obtained using an l^1 constrained optimazation technique. We expect a consistent correspondence to approximately preserve the distance maps d_p

disambiguate the intrinsic symmetries of the shape. For example, matches along the tail of the cat would bring little to no information on what solution to prefer. In practice, we can seek for a few matches that cover the whole shape and be as accurate as possible. To this end, we generate evenly distributed samples $V_{\text{fps}} \subset M$ on the test shape via farthest point sampling [13] by using the extrinsic Euclidean metric. Then, we construct a matching problem restricted to the set of *predicted* matches

$$O = \{(\boldsymbol{m}, \boldsymbol{r}) \in V_{\text{fps}} \times R \mid (\mathrm{X}_{M,R})_{\boldsymbol{rm}} > 0\}\,. \tag{11.18}$$

In practice this set is expected to be small, since the prediction given by the forest is very sparse and we select around 50 farthest samples per test shape (≈0.2 % of the total number of points on the adopted datasets). This results in a small matching problem that we solve via game-theoretic matching [20], a ℓ_1-regularized technique that allows to obtain sparse, yet very accurate solutions in an efficient manner. Once a sparse set of matches is obtained, we solve (11.15) as the weighted least-squares problem

$$\min_{\mathrm{X}} \|\mathrm{X}_{M,R} - \mathrm{X}\|_F^2 + \sum_{(\boldsymbol{p},\boldsymbol{q}) \in O} \omega_{\boldsymbol{pq}} \|\mathrm{X}\boldsymbol{d_p} - \boldsymbol{d_q}\|_2^2\,, \tag{11.19}$$

where $\omega_{\boldsymbol{pq}} \in [0, 1]$ are weights (provided by the game-theoretic matcher) giving a measure of confidence for each match $(\boldsymbol{p}, \boldsymbol{q}) \in O$. Figure 11.6c shows the result of the regularization performed using 25 sparse matches (indicated by small spheres).

Notice that the distance between functional maps is yet not well understood. The authors of [6] suggest to replace the Frobenius norm in (11.19) with a regularized l^0 norm of the vector of singular values:

$$\|A\|_\varepsilon = \sum_i \frac{\sigma(A)_i^2}{\sigma(A)_i^2 + \varepsilon} \tag{11.20}$$

Assuming the shapes to be (nearly) isometric one can expect the Laplace Beltrami operators on the shapes to commute with the functional map, i.e. (in the harmonic bases):

$$X\Lambda_M = \Lambda_R X \tag{11.21}$$

where Λ_M and Λ_R are the diagonal matrices of the singular values. A measure of deviation from (11.21) can be used as an alternative regularity cost.

11.2.3 Experimental Results

In all our experiments we used the WKS as pointwise descriptor for the training process. As in [16], we limited the size of the bases on the shapes to the first 100 eigenfunctions of the Laplace-Beltrami operator, computed using the cotangent scheme [15].

11.2.3.1 Comparison with Dense Methods

In this set of experiments we compare with the state of the art techniques in (dense) non-rigid shape matching, namely the functional maps pipeline [16], blended intrinsic maps (BIM) [10], and the coarse-to-fine combinatorial approach of [24]. We perform these comparisons on the TOSCA high-resolution dataset [5]. The dataset consists of 80 shapes belonging to different classes, with resolutions ranging in 4–52K points. Shapes within the same class have the same connectivity and undergo nearly-isometric deformations. Ground-truth point mapping among shapes from the same class is available. In particular, given a predicted map $f : M \to N$ and the corresponding ground-truth $g : M \to N$, we define the *error* of f as

$$\varepsilon(f, g) = \sum_{\boldsymbol{m} \in M} d_N(f(\boldsymbol{m}), g(\boldsymbol{m})) \,, \tag{11.22}$$

where d_N is the geodesic metric on N, normalized by $\sqrt{Area(N)}$ to allow inter-class comparisons. Similarly, we define the average (pointwise) geodesic error as $\frac{\varepsilon(f, g)}{|M|}$.

Although the methods considered in these experiments do not rely on any prior learning, the comparison is still meaningful as it gives an indication of the level of accuracy that our approach can attain in this class of problems. The experiments were designed on the same benchmark and following a procedure similar to the one reported in [10, 16]. Specifically, for each model M of a class (e.g., the class of dogs), we randomly picked other 6 models from the same class (not including M), and trained a random forest with them (thus, we only considered classes with at least 6 shapes). Then we predicted a dense correspondence for M according to the technique described in Sect. 11.2.2.

We show the results of this experiment in Fig. 11.8 (right). Each curve depicts the percentage of matches that attain an error below the threshold given on the x-axis. Our method (red line) detects 90 % correct correspondences within a geodesic error of 0.05. Almost all correct matches are detected within an error of 0.1. This is compatible with and even improves the results given by the other methods on the same data. Note that our training process only makes use of pointwise information (namely, the WKS); in contrast, the functional maps pipeline (blue line) adopts several heuristics (WKS preservation constraints in addition to orthogonality of $\mathtt{C}$, region-wise features, etc.) in order to constrain the solution optimally [16]. Upon

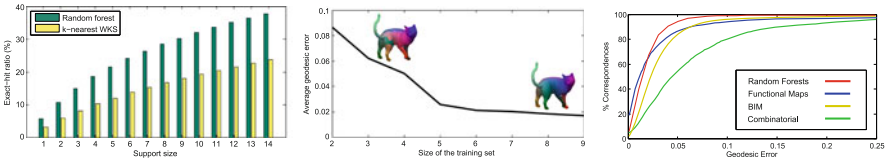

Fig. 11.8 *Left*: Fraction of exact matches predicted by a random forest vs. maximum support size of the probability distributions on a test shape. The forest was trained with 9 shapes. *Middle*: Sensitivity to number of shapes used in the training set. Note how the correspondence predicted using little training data (*top-left* model) is only partially regularized. *Right*: Comparison with the state-of-the-art methods on nearly-isometric shapes (TOSCA). Symmetric correspondences are considered correct solutions for all methods

visual inspection, we observed that most of the errors in our method were due to the poor choice of points made in the regularization step. This is analogous to what is reported for the BIM method [10]. Typically, we observed that around 20 well-distributed points are sufficient to obtain accurate results.

11.2.4 Sensitivity to Training Parameters

We performed a sensitivity analysis of our method with respect to the parameters used in the training process, namely the size of the training set and the number of trees in the forest. In these experiments we employed the cat models from the TOSCA dataset (28K vertices) with the corresponding ground-truth.

In Fig. 11.8 (middle) we plot the average geodesic error obtained by a test shape (depicted along the curve) as we varied the number of shapes in the training set. The geodesic error of the correspondence stabilizes when at least 6 shapes are used for training. This means that only 6 samples per label are sufficient in order to determine an optimal parametrization of the nearly-isometric deformations occurring on the shape. This result contrasts the common setting in which random forests are trained with copious amounts of data [8, 30], making the approach rather practical when only limited training data is available.

Figure 11.8 (left) shows the change in accuracy as we increase the number of trees in the forest. Note that increasing the number of trees directly induces a larger support of the probability distributions over L. In other words, each point of the test shape receives more candidate matches if the forest is trained with more trees (see Eq. (11.7)). The hit ratio in the bar plot is defined as the fraction of *exact* predictions given by the forest over the entire test shape. We compare the results with the hit ratio obtained by looking for k-nearest neighbors in WKS descriptor space, with k equal to the maximum support size employed by the forest at each level. From this plot we see that the forest predictions are twice as accurate as WKS predictions for equal support sizes. In particular, random forest predicts the *exact* match for almost half (around 14K points) of the shape when trained with 15 trees.

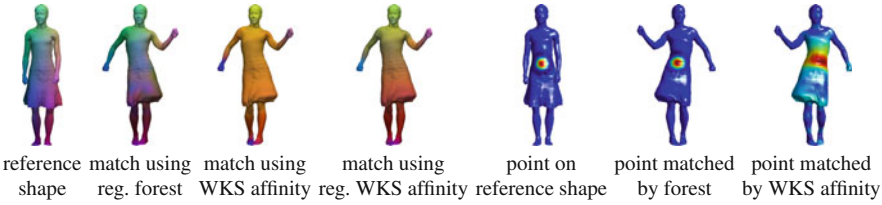

Fig. 11.9 Comparison between our method and an approach based on WKS affinity using shapes from the dataset of Vlasic et al. Columns one to four show the predicted and regularized solutions for both approaches. The last three columns show how the indicator function at one point gets functionally mapped to a second shape, by using the (non-regularized) X obtained from the forest, and by X_{WKS}

Finally, in Fig. 11.9 we show a qualitative comparison between our method and an approach based on WKS. The rationale of this experiment is to show that the prediction given by the forest gives better results than what can be obtained without prior learning within the same pipeline (i.e., prediction followed by regularization). Specifically, for each point in one shape we construct a probability distribution on the other shape based on a measure of descriptor affinity in WKS space. We then estimated a functional map $\mathrm{C}_{\mathrm{WKS}}$ from the resulting set of constraints, and plotted a final correspondence before and after regularization.

11.2.5 Learning Non-isometric Deformations

In this section we consider a scenario in which the shapes to be matched may undergo more general (i.e., far from isometric) deformations. Examples of such deformations include local and global changes in scale, topological changes, resampling, partiality, and so forth. Until now, few methods have attempted to tackle this class of problems. Most dense approaches [10, 16, 17, 24] are well-defined in the quasi-isometric and conformal cases only; instances of inter-class matching were considered in [10], but the success of the method depends on the specific choice of (usually hand-picked) feature points used in the subsequent optimization. Sparse methods considering the general setting from a metric perspective [4, 20, 22] attempt to formalize the problem by using the language of quadratic optimization, leading to difficult and highly non-convex formulations. An exception to the general trend was given in [31], where the matching is formulated as a linear program in the product space of manifolds. The method allows to obtain dense correspondences for more general deformations, but it assumes consistent topologies and is computationally expensive ($\sim$2 h to match around 10K vertices). Another recent approach [11] attempts to model deviation from isometry in the framework of functional maps, by seeking compatible harmonic bases among two shapes. However, it relies on a (sparse) set of matches being given as input and it shares with [31] the high computational cost.

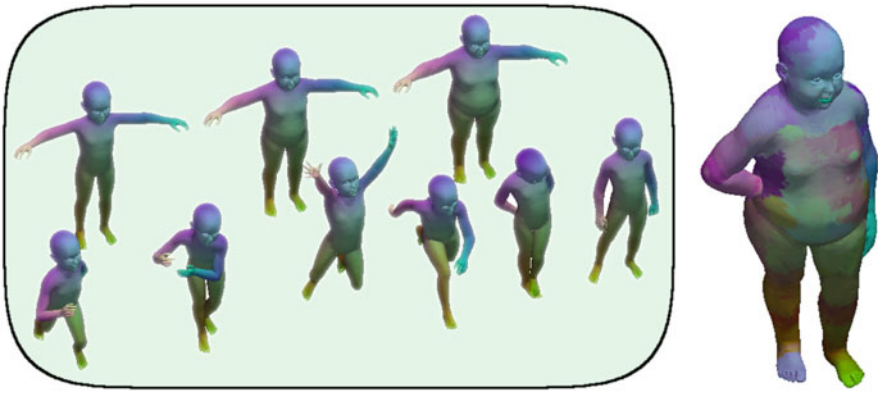

Fig. 11.10 Example of dense shape matching using random forests under non-isometric deformations. Shapes in the shaded area are a subset of the training set. The forest is trained with wave kernel descriptors and consists of 80K training classes with 19 samples per class. Matches are encoded by color

As described in Sect. 11.2, the forest does not contain any explicit knowledge of the type of deformations it is asked to parametrize. This means that, in principle, one could feed the learning process with training data coming from any collection of shapes, with virtually no restrictions on the transformations that the shapes are allowed to undergo. Clearly, an appropriate choice of the pointwise descriptor should be made in order for the forest to provide a concise and discriminative model. To test this scenario, we constructed a synthetic dataset consisting of 8 high-resolution (80K vertices) models of a kid under different poses (quasi-isometries), and 11 additional models of increasingly corpulent variants of the same kid (local scale deformations) with a fixed pose (see Fig. 11.10). The shapes have equal number of points and point-to-point ground-truth is available. We test the trained random forest with a plump kid having a previously unseen pose.

Note that the result is reasonably accurate if we keep in mind the noisy setting: the forest was trained with WKS descriptors, which are originally designed for quasi-isometric deformations, and thus not expected to work well in the more general setting [12]. Despite being just a qualitative evaluation, this experiment demonstrates the generality of our approach. The matching process we described can still be employed in general non-rigid scenarios if provided with limited, yet sufficiently discriminative training data.

11.3 Conclusions

In this article, we showed how the random forest learning paradigm can be employed for problems of dense correspondence among deformable 3D shapes. To our knowledge, this is among the first attempts at introducing a statistical learning view

on this family of problems. The effectiveness of our approach is demonstrated on a standard benchmark, where we obtain comparable results with respect to the state of the art, and very low prediction times for shapes with tens of thousands of vertices. The approach is flexible in that it provides a means to model deformations which are far from isometric, and it consistently obtains high predictive performance on all tested scenarios.

References

1. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. In: ACM Transactions on Graphics (TOG), vol. 24, pp. 408–416. ACM, New York (2005)
2. Aubry, M., Schlickewei, U., Cremers, D.: The wave kernel signature: a quantum mechanical approach to shape analysis. In: ICCV Workshops, Barcelona (2011)
3. Breiman, L.: Random forests. In: Machine Learning, vol. 45. Springer, Berlin/Heidelberg (2001)
4. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. PNAS **103**(5), 1168–1172 (2006)
5. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Numerical Geometry of Non-Rigid Shapes. Springer, New York (2008). Incorporated, 1 edition
6. Corman, É., Ovsjanikov, M., Chambolle, A.: Supervised descriptor learning for non-rigid shape matching. In: European Conference on Computer Vision, pp 283–298 (2014)
7. Criminisi, A., Shotton, J., Konukoglu, E.: Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-supervised Learning. Foundations and Trends in Computer Graphics and Vision. Now, Boston (2012)
8. Fanelli, G., Gall, J., Van Gool, L.: Real time head pose estimation with random regression forests. In: CVPR, Colorado Springs (2011)
9. Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V.: Hough forests for object detection, tracking, and action recognition. PAMI **33**(11), 2188–2202 (2011)
10. Kim, V.G., Lipman, Y., Funkhouser, T.: Blended intrinsic maps. In: SIGGRAPH 2011, Vancouver (2011)
11. Kovnatsky, A., Bronstein, M.M., Bronstein, A.M., Glashoff, K., Kimmel, R.: Coupled quasi-harmonic bases. Comput. Graph. Forum **32**(2pt4), 439–448 (2013)
12. Litman, R., Bronstein, A.M.: Learning spectral descriptors for deformable shape correspondence. TPAMI **36**(1), 170–180 (2013)
13. Mémoli, F.: Gromov-Wasserstein distances and the metric approach to object matching. Found. Comput. Math. **11**, 417–487 (2011)
14. Mémoli, F., Sapiro, G.: A theoretical and computational framework for isometry invariant recognition of point cloud data. Found. Comput. Math. **5**(3), 313–347 (2005)
15. Meyer, M., Desbrun, M., Schröder, P., Barr, A.H.: Discrete differential-geometry operators for triangulated 2-manifolds. In: Proceedings of VisMath, Berlin (2002)
16. Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L.: Functional maps: a flexible representation of maps between shapes. ACM Trans. Graph. **31**(4), 30:1–30:11 (2012)
17. Ovsjanikov, M., Mérigot, Q., Mémoli, F., Guibas, L.: One point isometric matching with the heat kernel. Comput. Graph. Forum **29**(5), 1555–1564 (2010)
18. Ovsjanikov, M., Sun, J., Guibas, L.: Global intrinsic symmetries of shapes. In: Computer Graphics Forum, vol. 27, pp. 1341–1348. Wiley Online Library (2008)
19. Pickup, D., Sun, X., Rosin, P., Martin, R., Cheng, Z., Lian, Z., Aono, M., Ben Hamza, A., Bronstein, A., Bronstein, M., et al.: Shrec14 track: shape retrieval of non-rigid 3d human models. 3DOR **4**(7), 8 (2014)

20. Rodolà, E., Bronstein, A.M., Albarelli, A., Bergamasco, F., Torsello, A.: A game-theoretic approach to deformable shape matching. In: CVPR, Providence (2012)
21. Rodolá, E., Rota Bulo, S., Windheuser, T., Vestner, M., Cremers, D.: Dense non-rigid shape correspondence using random forests. In: CVPR, Columbus, pp. 4177–4184 (2014)
22. Rodolà, E., Torsello, A., Harada, T., Kuniyoshi, T., Cremers, D.: Elastic net constraints for shape matching. In: ICCV, Sydney (2013)
23. Rustamov, R.M.: Laplace-beltrami eigenfunctions for deformation invariant shape representation. In: SGP, Barcelona. Eurographics Association (2007)
24. Sahillioğlu, Y., Yemez, Y.: Coarse-to-fine combinatorial matching for dense isometric shape correspondence. Comput. Graph. Forum **30**(5), 1461–1470 (2011)
25. Schölkopf, B., Steinke, F., Blanz, V.: Object correspondence as a machine learning problem. In: ICML, Bonn, pp. 776–783. ACM (2005)
26. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: CVPR, Anchorage (2008)
27. Shtern, A., Kimmel, R.: Matching lbo eigenspace of non-rigid shapes via high order statistics. arXiv preprint arXiv:1310.4459 (2013)
28. Steinke, F., Blanz, V., Schölkopf, B.: Learning dense 3d correspondence. In: Advances in Neural Information Processing Systems, pp. 1313–1320. MIT Press, Cambridge/London (2006)
29. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: SGP, Berlin. Eurographics Association (2009)
30. Taylor, J., Shotton, J., Sharp, T., Fitzgibbon, A.: The vitruvian manifold: inferring dense correspondences for one-shot human pose estimation. In: CVPR, Pittsburgh (2012)
31. Windheuser, T., Schlickewei, U., Schmidt, F.R., Cremers, D.: Geometrically consistent elastic matching of 3d shapes: a linear programming solution. In: ICCV, Barcelona (2011)
32. Windheuser, T., Vestner, M., Rodola, E., Triebel, R., Cremers, D.: Optimal intrinsic descriptors for non-rigid shape analysis. In: BMVC, Nottingham. BMVA Press (2014)