Learning Shape Metrics based on Deformations and Transport

Guillaume Charpiat
Pulsar Project
INRIA Sophia-Antipolis - France
Guillaume.Charpiat@sophia.inria.fr

Abstract

Shape evolutions, as well as shape matchings or image segmentation with shape prior, involve the preliminary choice of a suitable metric in the space of shapes. Instead of choosing a particular one, we propose a framework to learn shape metrics from a set of examples of shapes, designed to be able to handle sparse sets of highly varying shapes, since typical shape datasets, like human silhouettes, are intrinsically high-dimensional and non-dense. We formulate the task of finding the optimal metrics on an empirical manifold of shapes as a classical minimization problem ensuring smoothness, and compute its global optimum fast.

First, we design a criterion to compute point-to-point matching between shapes which deals with topological changes. Then, given a training set of shapes, we use these matchings to transport deformations observed on any shape to any other one. Finally, we estimate the metric in the tangent space of any shape, based on transported deformations, weighted by their reliability. Experiments on difficult sets are shown, and applications are proposed.

Introduction

The notion of shape is important in many fields of computer vision, from tracking to scene understanding. As for usual object features, it can be used as a prior, as in image segmentation, or as a source of information, as in gesture classification. When image classification or segmentation tasks require high discriminative power or precision, the shape of objects naturally appear relevant to our human minds. However, shape is a complex notion which cannot be dealt with directly like a simple parameter in $\mathbb{R}^n$. Modeling shape manually is tedious, and one arising question is the one of learning shapes automatically.

Various statistical models of shapes exist in the literature. Most of them consist in estimating a mean pattern and characteristic deformations of a given set of shapes, under the assumption that the shape variability in the training set is not too high, in order to be able to consider sensible deformations from one shape to another one, and to compute sensible statistics on these deformations (usually with principal component analysis). Deformations can be given by the user, fully or with a few landmarks [17], or can be computed via a search for best diffeomorphisms [5, 19] or matchings [8, 7] for hand-designed metrics or criteria. They can also be computed by algorithms suited to particular shape representations [11, 4] or just be distance gradients [3]. In all cases, the statistics computed on deformations can be turned into a new metric in the tangent space of the mean pattern, acting as a deformation prior on one particular shape.

When the shape variability is too high, this paradigm fails because the automatic computation of deformations between very different shapes is not reliable, and because linear approximations of the space of shapes are not meaningful anymore. Distance-based algorithms, such as kernel methods, were proposed [12, 6] to handle high variability, but at the price of considering only distances between shapes, instead of deformations, thus losing the crucial information they carry. These methods consider training sets as graphs, whose nodes are shapes and whose edges are distances (for a particular metric chosen). They assume the neighborhood of any shape to be representative of the intrinsic dimensionality of the space of shapes, and require consequently relatively high sampling densities, which are not affordable in the case of datasets with high intrinsic dimension, like human silhouettes with at least 30 degrees of freedom. Some other interesting methods are based on shape patches or parts [10, 2], but they sacrifice the notion of continuous, global shape. Another approach consists in searching the training set for the closest shapes to the one of interest, based on shape distances [13] or, for videos, on time order [16], and then in applying classical approaches to this neighborhood only. Again, the neighborhood representativeness issue arises. Moreover, the local metrics thus computed are not guaranteed to be globally coherent as a function of the shape of interest.

This paper aims at extending the approaches based on continuous deformations, to the case of high shape variability, where the notion of mean shape is not relevant anymore, or where deformations cannot be estimated throughout the whole training set even if they still can be computed between close enough samples. In our approach, we compute point-to-point matchings between close shapes, and use them as a way to propagate information within the training set. Thus, we can transport a deformation of a shape

to any other shape, with an associated reliability weight. Metrics, *i.e.* inner products on deformation spaces, are then estimated on the tangent space of any shape, while taking into account deformations observed at other locations, thus decreasing dramatically the sample density required and ensuring the global coherence of the manifold.

The paper is organized as follows: First, we propose a shape matching algorithm which handles topological changes. Second, we use pairwise matchings to define transport and to build a manifold-like structure. Third, we transport deformations and propose a method to estimate metrics. We show results on video datasets and, finally, in a theoretical section, we study different tracks for metric estimation, and prove that the method presented in the previous part computes the optimal metrics for a natural criterion.

1. Shape matching

In order to compare quantities defined on different shapes, like deformations, we need a way to transport them from shape to shape, and to do this we need point-to-point correspondences between close shapes. Since in typical shape datasets, like walking human silhouettes, topological changes are very frequent (fig.1, left), we need a matching algorithm able to consider pairs of shapes with different topologies. The only one we found in the literature relies on successive bipartite graph matchings and spline estimations [1]. It imposes similar sampling rates on both shapes, which limits the precision and the regularity of the matching (fig.1, right). Better regularity and precision (for a given time cost) will be ensured here by oversampling the target.

In the case of contours in images, a shape is a union of 1D curves. If shapes have only one connected component, *i.e.* if they are topologically equivalent to a circle or a segment, then dynamic time warping helps find quickly the global optima of simple matching energies [9, 18, 14]. In recent works, graph-cuts or minimum cycle search in graphs have also been used [15]. We will adapt here dynamic time warping to minimize an energy which favors smooth matchings and deals with several connected components, topological changes, and optionally with vanishing parts. Such an algorithm is needed in the sequel, but our work is not specific to the particular matching algorithm presented here.

1.1. Matching criterion

Let us start with the case of simple, closed curves A and B, seen as functions from the circle $\mathbb{S}_1$ to $\mathbb{R}^2$, parameterized by their arc length s. We search for the best matching from A to B, *i.e.* for the best function m from $\mathbb{S}_1$ to $\mathbb{S}_1$ so that $A \simeq B \circ m$. Let us note $\mathbf{f} = B \circ m - A$ so that $\mathbf{f}(s)$ stands for the vector $\overrightarrow{A_{(s)}B_{(m(s))}}$ linking a point of A to its correspondent on B. The deformation $\mathbf{f}$ should be as small and as smooth as possible (see figure 1 for explanation),

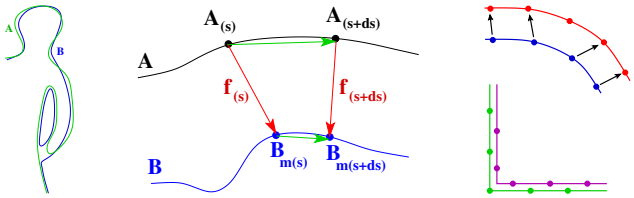

Figure 1. **(Left)** Example of topological change: hand in the pocket. **(Middle)** Matching energy explanation. We search for small linking vectors $\mathbf{f}(s)$, and for small difference between the two green vectors, or between the red ones, to ensure spatial coherence. These two differences have same norm: $\|\partial \mathbf{f}/\partial s\|$. **(Right)** Imposing similar sampling rates limits precision and regularity.

so we minimize the following criterion over deformations $\mathbf{f}$ that can be written as $\mathbf{f} = B \circ m - A$ for some m:

$$\|\mathbf{f}\|^2_{H^1_\alpha} := \int_{\mathbb{S}_1} \|\mathbf{f}(s)\|^2 + \alpha \left\| \frac{\partial \mathbf{f}}{\partial s} \right\|^2 ds.$$

When shapes have thin parts, more information is required in order to distinguish nearby parallel sides. We include in the criterion the relative angle between outgoing normals at corresponding points $\angle(\mathbf{n}_{A(s)}, \mathbf{n}_{B\circ m(s)})$:

$$E_{match}(m) = \|B \circ m - A\|^2_{H^1_\alpha} + \gamma \|\angle(\mathbf{n}_A, \mathbf{n}_{B\circ m})\|^2_{L^2}$$

Note that we require the deformation $\mathbf{f}$ to be smooth, and not the parameterization correspondence m itself. Now if B as several connected components, say $B = \cup_i B^i$, then we can still search for an optimal matching m between A and B, with $m : \mathbb{S}_1 \to \coprod_i \mathbb{S}_1$ possibly pointing to any of the parameterization supports of connected components B^i. The matching cannot be guaranteed to be one-to-one anymore, since some parts of B may have no antecedent through m. In order to reciprocally account for points on A that cannot be matched to B (disappearing parts), we may allow a match to nothing, i.e. $m(s)$ may have the value $\varnothing$.

1.2. Optimization with dynamic time warping

In practice, shapes are unions of polygons, and the energy can be discretized accordingly. In the case where A has only one connected component, E_{match} can be minimized efficiently. For each vertex $A(s)$ of A, we define the set $N(s) = \{\varnothing\} \cup \{s' \ s.t. \ \|B(s') - A(s)\| \leqslant d_{max}\}$ of its possible matches, *i.e.* the set of possible values of $m(s)$, as points of B relatively close and $\varnothing$. The consideration of a maximum distance d_{max} speeds up the process significantly. Choosing any initial point $A(s_0 = 1)$, and considering the set of ordered vertices along A, the problem reduces to finding an optimal function from $\{1, \dots, \#A\}$ to $N(1) \times N(2) \cdots \times N(\#A)$, which can be seen as a search for an optimal path in a graph, with costs on graph edges $\big((s, m(s)), \ (s+1, m(s+1))\big)$ derived from the energy E_{match}. A constant high cost is assigned to edges involving two $\varnothing$, and an even higher one if involving only one $\varnothing$. This problem is solved by dynamic time warping.

If A has several components, each of them is treated independently. This process is not symmetric in the sense that the matching obtained from A to B can differ from the one from B to A. The quality and the precision of the results increase when the discretization of the target B is finer than the template A, so we oversample targets. This is corroborated by a convergence study when discretizations get finer, included in the supplementary materials.

The computational cost is low, only a fraction of a second to match shapes with hundreds of points on a standard PC. The value of $E_{match}(m)$ reflects the quality of the matching computed: the lower the energy is, the more similar the two shapes are, and the more reliable the matching found is. We noticed that allowing matchings to $\varnothing$ gives more accurate correspondence fields, but unluckily also less significant values $E_{match}(m)$ (because $\varnothing$ induces a saturation cost). Because of the importance of these reliability values in the sequel, we remove the possibility of matching to $\varnothing$.

2. Transport and information propagation

We now use the matching tool to define transport in training sets of shapes, and to propagate information with reliability weights. Points on shapes will now be confused with their parameterizations, so that functions can be defined on shapes rather than on the parameterizations thereof.

2.1. Local transport

Let A and B be two shapes, and $m_{A\to B}$ the matching from A to B (so that $A \simeq B \circ m_{A\to B}$). Any function h defined along B, with values in any space $\mathcal{X}$, can be transported to A, or more exactly to the points of A linked with points of B, since $m_{A\to B}$ may be not one-to-one. More generally, all quantities in the sequel will be computed over matching domains. The local transport $T^L_{B\to A}$ is defined by:

$$\forall\, h : B \to \mathcal{X}, \quad \begin{array}{rcl} T^L_{B\to A}(h) : \ A & \to & \mathcal{X} \\ \left(T^L_{B\to A}(h)\right)(s) & = & h\left(m_{A\to B}(s)\right) \end{array}$$

We can compare any two functions h_A, h_B defined on different shapes, by $h_A - T^L_{B\to A}(h_B)$ or $T^L_{A\to B}(h_A) - h_B$.

When the functions to be transported are deformations, other transports may be defined. For example, in the case of rotations, one may prefer the angle between the normal to the shape $\mathbf{n}_{B(s)}$ and the vector $h(s)$ to be kept constant during transport. However it is not obvious whether such a transport would be sensible for all deformations h and all pairs (A, B). In the sequel we keep the former transport.

2.2. Global transport

Given a training set of shapes $\mathcal{S} = (S_i)$, we compute all possible pairwise matchings $m_{i\to j}$ and associate to each of them the matching cost $C^m_{ij} = E_{match}(m_{i\to j})$. A low cost C^m_{ij} means that the shapes S_i and S_j were close and that the matching $m_{i\to j}$ is reliable, whereas a high cost reveals an unsatisfying matching. For any pair (i_0, j_0) we search for the best path from S_{i_0} to S_{j_0} in the graph whose nodes are shapes and whose edges are matching costs C^m_{ij}. We then denote by $C^G_{i_0 j_0}$ the cost of this path $(i_0, i_1, \ldots, i_k = j_0)$ and by $T^G_{i_0\to j_0}$ the composition of local transports along it:

$$T^G_{i_0\to j_0} = T^L_{i_{k-1}\to j_0} \circ \cdots \circ T^L_{i_1\to i_2} \circ T^L_{i_0\to i_1}.$$

This gives the optimal transport from S_{i_0} to S_{j_0}. Since E_{match} is a quadratic energy, the optimal path will prefer series of small, reliable steps to big, uncertain jumps.

The computation of all best paths is affordable with standard shortest path algorithms. In the sequel, $T_{i\to j}$ will stand for the global transport $T^G_{i\to j}$, and $w^G_{ij} = e^{-\alpha_T C^G_{ij}}$ for the associated transport reliability, for a fixed positive α_T. We will also denote by $w^L_{ij} = e^{-\alpha_T C^m_{ij}}$ the confidence in the direct matching between shapes S_i and S_j .

2.3. Individualized transport

One could also consider individualized transports, in the sense that the best path may depend on the transported quantity. Indeed the cost of a local transport $T^L_{i\to j}$ is a sum over vertices of S_i : $E_{match}(m_{i\to j}) = \int_{S_i} E^v(m_{i\to j})(s)ds$ so it would make sense to consider the following individual cost, for any deformation h to be transported:

$$C^{m,ind.}_{ij}(h) = \int_{S_i} E^v(m_{i\to j})(s)\ \|h(s)\|\, ds$$

so that bad matchings along a shape are not significant if they occur where there is no information to transmit.

2.4. Propagating information

This structure $(T_{i\to j}, w^G_{ij})$ is useful to propagate information along the training set. A local matching $m_{i\to j}$ between two close shapes can be seen as a deformation from S_i to S_j. Because training sets are relatively small and reliable deformations are scarce, it makes sense to complete the set of observed deformations on one shape S_k by deformations observed at other locations S_i for which the transport $T_{i\to k}$ is reliable. In the ideal case of a rigid object with d articulations and perfect local matchings, only one observation of each articulation moving, at any position, is sufficient to realize the full complexity of all possible movements at all positions, by transports and linear combinations.

Usual articulated human models have about $d = 30$ degrees of freedom, and consequently the intrinsic dimension of typical shape datasets is potentially high. This implies that there is no hope in obtaining a dense training set, even with a loose grid (say N bins for each degree of freedom, which makes N^d bins), even with billions of examples. Consequently, methods involving only distances [12] or nearest neighbors [6, 13, 16] are not likely to be successful for high d, whereas our approach based on transport of

Figure 2. Colors on one shape, propagated via transports to other shapes from a same video sequence (see text for details).

deformations reduces the number of required samples from N^d to Nd in the ideal case.

2.5. Suitability for video analysis

In video sequences, each frame is relatively similar to the next one, so that any temporal succession of observed shapes is a natural good path to transport through. In particular it is not possible to find a shape without close neighbors, and consequently information can be shared.

We present an example for a sequence of walking human silhouettes, from the ViHASi dataset[1]. We compute automatically pairwise matchings, and from them the best transport paths (without temporal order information). The paths obtained are strongly correlated to temporal ordering (forwards or backwards). We pick one shape, add colors randomly along it, and propagate color (as a function with values (r, g, b) in $\mathbb{R}^3$) through transports to all training shapes, in order to visualize correspondences. The result, in figure 2 (see the supplementary materials for the whole video), is correct, validating the method. Of course a few matching mistakes are sometimes observed when auto-occultation happens, but these errors are few, and they are not propagated because these mismatchings cost more. Considering individualized transports (part 2.3) would allow the transmission of information not related to the precise location where difficulties occur. For instance a hand gesture could be transmitted to another silhouette with similar arm positions even if the legs have been crossed in the meanwhile.

2.6. Learning with shapes and transport

With such a structure, one can learn functions from a space of shapes $\mathcal{S}$ to any vector space $\mathcal{X}$, in particular to spaces of functions defined on shapes. For example, let us consider a training set of shapes with appearance $(S_i, A_i) \in \mathcal{S} \times \mathcal{F}(S_i \to \mathbb{R}^n)$. The appearance $A_i(s)$ could be, for any point s of any shape S_i, an image patch centered on s taken from the image from which S_i was segmented. The transport structure transmits examples of patches $A_j(m_{i\to j}(s))$ observed at similar locations on other shapes, and thus local statistics on patches can be performed, leading to a prior of the image given the shape. Thus, in a Bayesian framework, this structure allows the learning of segmentation/detection criteria with shape and appearance priors.

For scene understanding purposes, given videos where a few objects interact, one could also learn which parts of objects interact (by propagating contact locations through the set of shapes), and how (by propagating gestures). The perspectives opened by this kind of framework are wide.

[1] http://dipersec.king.ac.uk/VIHASI/

3. Learning metrics

Using previous sections, we now learn metrics, show how to use them as priors, and how to learn the whole structure.

3.1. Metric estimation with weighted H^1_α PCA

Given a training set of shapes $\mathcal{S} = (S_i)$, we are now able to compute matchings $m_{i\to j}$ between close shapes, which can be seen as deformations

$$\mathbf{f}_{i\to j} \;=\; S_j \circ m_{i\to j} - S_i.$$

We are also able to transport these deformations to any other shape S_k :

$$\mathbf{f}^k_{i\to j} \;=\; T_{i\to k}(\mathbf{f}_{i\to j})$$

with a weight combining reliability about deformation and transport computation (section 2.2):

$$w^k_{i\to j} \;=\; w^L_{ij}\; w^G_{ik}.$$

We now estimate the metric in the tangent space of the shape S_k, *i.e.* we search for a relevant inner product in the space of deformations that can be applied to S_k, based on the set of weighted deformations $(\mathbf{f}^k_{i\to j}, w^k_{i\to j})$. Principal Component Analysis (PCA) seems a reasonable way to compute statistics but we need to adapt it to probability weights and to the H^1_α product, which favors smoothness and is more coherent with the matching energy E_{match} than the standard L^2 product. PCA is derived from an energy minimization problem: the search for the best orthonormal axes $\mathbf{e}_n$ to project data. Here, the projection error to be minimized is:

$$\inf_{\langle \mathbf{e}_n | \mathbf{e}_{n'} \rangle_{H^1_\alpha} = \delta_{n=n'}} \sum_{i,j} w^k_{i\to j} \left\| \mathbf{f}^k_{i\to j} - \sum_n \left\langle \mathbf{f}^k_{i\to j} \,|\mathbf{e}_n \right\rangle_{H^1_\alpha} \mathbf{e}_n \right\|^2_{H^1_\alpha}$$

This is equivalent to the maximization problem:

$$\sup_{\langle \mathbf{e}_n | \mathbf{e}_{n'} \rangle_{H^1_\alpha} = \delta_{n=n'}} \sum_n \sum_{i,j} w^k_{i\to j} \left\langle \mathbf{f}^k_{i\to j} \,|\mathbf{e}_n \right\rangle^2_{H^1_\alpha}$$

and to:

$$\sup_{\langle \mathbf{e}_n | \mathbf{e}_{n'} \rangle_{H^1_\alpha} = \delta_{n=n'}} \sum_n \mathbf{e}_n \, HFH \, \mathbf{e}_n$$

where $F = \sum_{i,j} w^k_{i\to j}\, \mathbf{f}^k_{i\to j} \otimes \mathbf{f}^k_{i\to j}$ is the weighted covariance matrix, and where $H = Id - \alpha\Delta$ is the symmetric definite operator such that $\langle a \,|b\rangle_{H^1_\alpha} = \langle H\, a \,|b\rangle_{L^2}$. If we note $\mathbf{d}_n = H^{1/2}\mathbf{e}_n$ then the problem becomes:

$$\sup_{\langle \mathbf{d}_n | \mathbf{d}_{n'} \rangle_{L^2} = \delta_{n=n'}} \sum_n \mathbf{d}_n \, H^{1/2} F H^{1/2} \, \mathbf{d}_n$$

so that the optimal $\mathbf{d}_n$ are the eigenvectors of $H^{1/2}FH^{1/2}$ with highest eigenvalues. As with usual PCA, we diagonalize the weighted correlation matrix M instead, given by:

$$M_{(i,j),(i',j')} = \left\langle \sqrt{w^k_{i\to j}}\,\mathbf{f}^k_{i\to j} \,\middle|\, \sqrt{w^k_{i'\to j'}}\,\mathbf{f}^k_{i'\to j'} \right\rangle_{H^1_\alpha}.$$

Let γ_n be the eigenvectors of M, and λ_n the eigenvalues. One can prove that $\mathbf{d}_n = \sum_{ij} \gamma^{(i,j)}_n H^{1/2} \sqrt{w^k_{i\to j}}\mathbf{f}^k_{i\to j}$ so that

$$\mathbf{e}_n = \sum_{ij} \gamma^{(i,j)}_n \sqrt{w^k_{i\to j}}\;\mathbf{f}^k_{i\to j}$$

Thus, PCA of the set of weighted deformations $(\mathbf{f}^k_{i\to j}, w^k_{i\to j})$ in the tangent space of a shape S_k leads to modes of deformation $\mathbf{e}_n$, with eigenvalues λ_n, computed easily from the correlation matrix M. Based on the associated Mahalanobis distance, we set the natural inner product P between any two deformations $\mathbf{f}_1$ and $\mathbf{f}_2$ of S_k:

$$\langle \mathbf{f}_1 \,|\, \mathbf{f}_2 \rangle_P = \sum_n \frac{1}{\lambda_n^2} \langle \mathbf{f}_1 \,|\mathbf{e}_n\rangle_{H^1_\alpha} \langle \mathbf{e}_n \,|\, \mathbf{f}_2 \rangle_{H^1_\alpha}.$$

In practice we replace λ_n by $\max(\lambda_n, \lambda_{noise})$ for a chosen level of noise. In case the distributions obtained along significant eigenmodes would not be Gaussian, we build histograms along eigenmodes, *i.e.* of $\left\langle \mathbf{f}^k_{i\to j} \,|\, \mathbf{e}_n \right\rangle_{H^1_\alpha}$. The probability distribution rebuilt from histograms, as if eigenmodes were independent, was found to be relatively close to the real one in many cases, but biases between eigenmodes may also be observed, especially for small training sets.

In figure 3 we consider a complex dancing silhouette sequence[2] with high variability and fast moves, and show first eigenmodes and histograms computed for a few shapes. Since linear combinations of first eigenmodes are most probable deformations, they can be seen as deformation priors. The priors obtained here are sensible, intuitively related to articulations and cloth moves, while the usual mean-and-modes model performs poorly and kernel methods do not lead to explicit deformation priors. The advantages over neighborhood-based methods are explained in figure 4.

3.2. An example of how to use the learned metric

The metric learned can be used as a prior on shape matching. Let A be a shape from the training set, to be matched, and B the new target. Any possible matching $\mathbf{f} = B\circ m - A$ is the sum of its projection $p(\mathbf{f}) = \sum_n a_n \mathbf{e}_n$ on modes estimated in the tangent space of shape A, and of the remaining part $\mathrm{noise}(\mathbf{f})$. From the metric P estimated before, we can derive a prior in H^1_α, which associates to $\mathbf{f}$ its cost $\|p(\mathbf{f})\|^2_P + \frac{1}{\lambda^2_{\mathrm{noise}}}\|\mathrm{noise}(\mathbf{f})\|^2_{H^1_\alpha}$.

Histograms along first eigenmodes can also be used, their negative log-likelihood be turned into a cost, and a similar noise term be added. In both cases we want to minimize an energy of the form $C((a_n)) + \frac{1}{\lambda^2_{\mathrm{noise}}}\|\mathrm{noise}(\mathbf{f})\|^2_{H^1_\alpha}$, which can be expressed as:

$$\inf_{m,\,\overrightarrow{a}} C(\overrightarrow{a}) + \frac{1}{\lambda^2_{\mathrm{noise}}} \left\| B\circ m - A - \sum_n a_n \mathbf{e}_n \right\|^2_{H^1_\alpha}$$

Given any $\overrightarrow{a}$, the optimal m can be found by the method described in section 1. A classical gradient descent can then be performed on $\overrightarrow{a}$. A few steps only are needed, the process consists mainly in cutting the part of $B\circ m - A$ which belongs to the span of the eigenmodes to add it to $p(\mathbf{f})$.

3.3. Learning the whole structure: second pass

The process to estimate metrics from a training set of shapes consists in three steps: computation of matchings between close shapes (section 1), transport of deformations (section 2), and turning statistics on deformations with reliability weights into metrics (section 3). If we re-run the whole process a second time, we can replace the shape matching algorithm of section 1 by the matching prior deriving from the learned metric (section 3.2). In the same spirit, in section 2, we could choose the geodesic transport associated to the learned metric. These geodesics would be easy to obtain, with a path-straightening method, since we already know point-to-point matchings as well as the metric. Thus, the choices made in these two sections can be seen as reasonable initializations aimed to be replaced with learned quantities in a second pass. Concerning the third section, one could wonder whether there could be other ways to estimate metrics, given matchings and transports. This is the subject of next part, where we will show that the method we presented already computes the optimal metrics.

4. Criteria on shape metrics and optima

Given a training set of shapes $\mathcal{S} = (S_i)$, and, for any shape S_k, an empirical distribution $\mathcal{D}_{emp}$ of transported deformations weighted by their reliability, one can wonder what the possible ways to estimate metrics are, whether there would be an objective criterion to assess how much a metric is suited to the set of shapes, and whether it is possible to find the optimal metrics.

4.1. Criterion in one tangent space

Let us consider first the case of the tangent space to one shape only. We are given a set of deformations $\mathbf{f}_j$ in this tangent space T, with probability weights w_j whose total sum is 1, *i.e.* we are given the empirical distribution:

$$\mathcal{D}_{emp} \;=\; \sum_j \; w_j \; \delta_{\mathbf{f}_j}$$

where $\delta_{\cdot}$ are Dirac peaks, and we would like to find a suitable inner product P for T. To any inner product P can be associated a probability distribution over deformations $\mathbf{f}$:

$$\mathcal{D}_P(\mathbf{f}) \;\propto\; e^{-\|\mathbf{f}\|^2_P}$$

[2] From Grimage platform, `https://charibdis.inrialpes.fr`

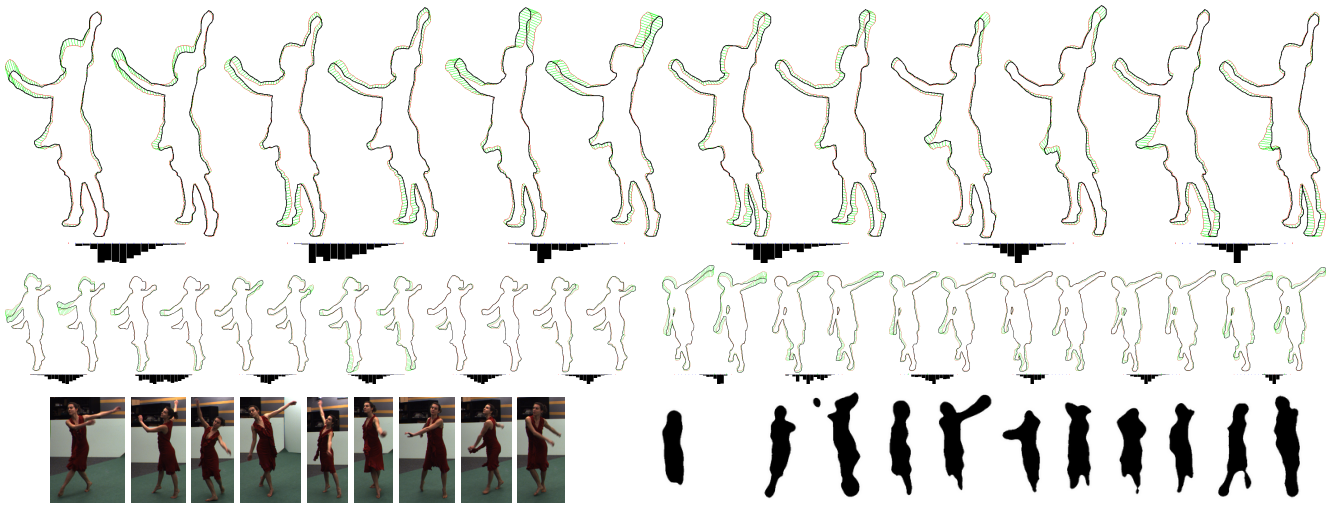

Figure 3. Dancing sequence (9s, 24Hz). **Top & middle**: first modes of deformation for various postures of the dancing sequence. Each mode is drawn twice, with amplitude $\pm\lambda_n$, and associated histogram is shown. Note how the modes are sensible, related to articulations (arms, legs, elbows, ...) or dress moves. Full resolution images can be found in the supplementary materials. **Bottom left**: some frames of the video sequence; **right**: mean and first modes with the classical PCA approach on level-sets[11, 4] : limbs are not correctly treated.

up to a normalizing constant. Let us restrict P to be continuous with respect to an inner product P_0 proposed by default, for instance H^1. The class of all such P is still huge, and, thanks to Riesz representation theorem, for any such P there exists a linear symmetric continuous operator A s.t.:

$$\forall\, \mathbf{f}_1, \mathbf{f}_2 \in T, \qquad \langle\, \mathbf{f}_1 \,|\, \mathbf{f}_2 \,\rangle_P = \langle A\, \mathbf{f}_1 \,|\, \mathbf{f}_2 \,\rangle_{P_0}\,.$$

Since such an operator can be diagonalized, there exists an orthonormal basis $(\mathbf{e}_n)$ for P_0, and real, positive coefficients (α_n) such that $A = \sum_n \alpha_n\, \mathbf{e}_n \otimes \mathbf{e}_n$, and consequently:

$$\forall\, \mathbf{f}_1, \mathbf{f}_2 \in T, \quad \langle\, \mathbf{f}_1 \,|\, \mathbf{f}_2 \,\rangle_P = \sum_n \alpha_n \,\langle\, \mathbf{f}_1 \,|\mathbf{e}_n\,\rangle_{P_0} \,\langle \mathbf{e}_n \,|\, \mathbf{f}_2 \,\rangle_{P_0}$$

$$\text{which implies} \quad \forall\, \mathbf{f} \in T, \quad \|\mathbf{f}\|_P^2 = \sum_n \alpha_n \,\langle\, \mathbf{f} \,|\, \mathbf{e}_n \,\rangle_{P_0}^2$$

so that the associated distribution

$$\mathcal{D}_P(\mathbf{f}) := \prod_n \left(\frac{\alpha_n}{\pi}\right)^{\frac{1}{2}} e^{-\alpha_n \langle \mathbf{f} | \mathbf{e}_n \rangle_{P_0}^2}$$

is Gaussian. Reciprocally, any Gaussian distribution relates to a definite positive quadratic form, *i.e.* an inner product on T. Thus, a search over probability distributions derived from inner products is a search over Gaussian distributions.

We would like the inner product P to be relevant to the set of deformations $\mathbf{f}_j$. One possible way is to pick the one whose associated distribution $\mathcal{D}_P$ is the closest to $\mathcal{D}_{emp}$.

Proposition 1. *The inner product P which leads to the probability distribution $\mathcal{D}_P$ the closest to the empirical distribution $\mathcal{D}_{emp} = \sum_j w_j\, \delta_{\mathbf{f}_j}$ for the Kullback-Leibler divergence, is the one obtained by weighted PCA on $(\mathbf{f}_j, w_j)$.*

Proof. The Kullback-Leibler divergence between any two probability distributions p_1 and p_2 is defined by:

$$KL(p_2|p_1) = \int p_1 \ln \frac{p_1}{p_2}.$$

Minimizing the Kullback-Leibler divergence between p_1 and p_2 with respect to p_2 consequently leads to the minimization of $E(p_2|p_1) = -\int p_1 \ln p_2$. In our case this gives:

$$\begin{aligned} E(\mathcal{D}_P|\mathcal{D}_{emp}) \;&=\; -\sum_j w_j \,\ln \mathcal{D}_P(\mathbf{f}_j) \qquad (1)\\ &=\; \sum_j \sum_n w_j \left(\alpha_n \,\langle\, \mathbf{f}_j \,|\, \mathbf{e}_n \,\rangle_{P_0}^2 - \frac{1}{2}\ln \alpha_n + \frac{1}{2}\ln \pi\right). \end{aligned}$$

If we denote by F the covariance matrix $\sum_j w_j\, \mathbf{f}_j \otimes \mathbf{f}_j$, the energy becomes, up to a constant:

$$\sum_n \left(\alpha_n \,\langle\, \mathbf{e}_n \,|F\,|\, \mathbf{e}_n \,\rangle_{P_0} - \frac{1}{2}\ln \alpha_n\right) \qquad (2)$$

The minimization with respect to α_n gives:

$$\partial_{\alpha_n} E \;=\; \langle\, \mathbf{e}_n \,|F\,|\, \mathbf{e}_n \,\rangle_{P_0} - \frac{1}{2\alpha_n} \;=\; 0. \qquad (3)$$

At the minimum, the derivative with respect to the unit-normed deformation $\mathbf{e}_n$ is 0 in all directions except for pure norm variation:

$$\partial_{\mathbf{e}_n} E \;=\; 2\,\alpha_n\, F\, \mathbf{e}_n \;\propto\; \mathbf{e}_n \qquad (4)$$

which implies that $\mathbf{e}_n$ is an eigenvector of F, say with eigenvalue λ_n. Together with (3) it gives: $\alpha_n = \frac{1}{2\lambda_n^2}$ and consequently the optimal inner product that induces the closest distribution to $\mathcal{D}_{emp}$ is the one related to the norm:

$$\|\mathbf{f}\|_P^2 = \frac{1}{2}\sum_n \frac{\langle\, \mathbf{f} \,|\, \mathbf{e}_n \,\rangle_{P_0}^2}{\lambda_n^2}.$$

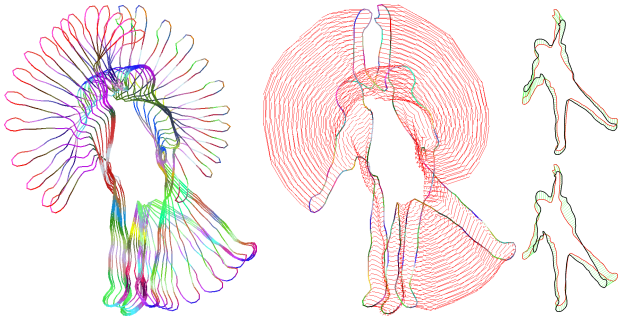

Figure 4. Waving and changing posture: set of 30 hand-segmented shapes from a video. **(Left)** Transport path between two shapes, **(middle)** correspondence flow to another shape. A waving sign can be transported to a different posture. **(Right)** Mean (in black) of the direct matchings from a shape (in red) towards its 5 and 10 nearest neighbors, respectively. Despite small neighborhood size, the mean is irrelevant. Neighborhoods made of reliable matchings are small and cannot include gestures observed at other postures.

This is, up to a constant factor, the Mahalanobis distance associated to weighted PCA on $(\mathbf{f}_j, w_j)$, which is precisely the algorithm developed in section 3.1 with $P_0 = H^1_\alpha$. □

One might however wonder whether minimizing the Kullback-Leibler distance to a sum of Dirac peaks makes sense. Luckily, the previous proposition can be extended to the case of symmetric translation-invariant unit-mass kernels $\mathcal{K}(\cdot - \cdot)$ defined on the space T of deformations. We replace $\mathcal{D}_{emp}$ by the kernel-smoothed empirical distribution

$$\mathcal{D}^{\mathcal{K}}_{emp}(\mathbf{f}) \;=\; \sum_j w_j \, \mathcal{K}(\mathbf{f}_j - \mathbf{f}).$$

Note : The family $(\mathbf{f}_j)$ is finite, so we work in a finite-dimensioned subspace of the tangent space T, and $\mathcal{K}$ can be understood, in the simple case, as a real function multiplied by the usual Lebesgue measure $d\mathbf{f}$. In the infinite dimension case, $\mathcal{K}$ cannot be isotropic (because it has finite mass).

Proposition 2. *The inner product P which leads to the probability distribution $\mathcal{D}_P$ the closest to the empirical distribution $\mathcal{D}^{\mathcal{K}}_{emp} = \sum_j w_j \mathcal{K}(\mathbf{f}_j - \cdot)$ for the Kullback-Leibler divergence, is obtained by diagonalization of the sum of the correlation matrix F and the second moment of $\mathcal{K}$.*

Proof. Full details in the supplementary materials. After computations, we find a similar expression to (2) except that F is replaced by $F + M_{\mathcal{K}}$ where $M_{\mathcal{K}} = \int_T \mathbf{f} \otimes \mathbf{f} \quad \mathcal{K}(\mathbf{f})$. Note that when the kernel $\mathcal{K}$ gets closer to a Dirac peak, $M_{\mathcal{K}}$ gets closer to 0, and we obtain proposition 1 again. □

We have defined a criterion, based on the Kullback-Leibler divergence, to quantify how suitable an inner product is for a tangent space given with an empirical distribution of deformations, and we have shown how to compute the optimal one. Next sections face coherency issues.

4.2. No best smooth direction field

We would like to compute a suitable inner product P_i for each tangent space T_i as previously, but in a coherent way: we would like P_i to be close to P_k if shapes S_i and S_k are close. One approach would be to compute a joint PCA in the tangent spaces T_i of all shapes S_i simultaneously, with a regularity criterion imposing that, after transport, the eigenmodes $\mathbf{e}^i_n$ in T_i should not differ too much from the ones $\mathbf{e}^k_n$ in T_k (with a weight w^G_{ik}). In the continuous case, this problem can be stated as searching for smooth vector fields $\mathbf{e}_n$ over a manifold $\mathcal{S}$. But the hairy ball theorem tells us that even in the simple case where the manifold is a sphere, there exists no non-vanishing continuous tangent vector field on the sphere. Which means that there are manifolds for which we cannot find smooth fields $\mathbf{e}_n$ whose norm is never 0, so that global modes of deformation do not always exist.

4.3. Criterion for a smooth metric

In fact we do not need the eigenmodes $\mathbf{e}^k_n$ to be smooth with respect to the shape S_k, we only need the probability distributions $\mathcal{D}_{P_k}$ related to them to be smooth. One way to ensure this consists in requiring the distribution $\mathcal{D}_{P_k}$ to be close not only to the empirical distribution $\mathcal{D}_{emp_k}$ in the tangent space of S_k, but also to the transported empirical distributions $T_{i\to k}(\mathcal{D}_{emp_i})$ from neighboring shapes S_i, with a weight w^G_{ik} depending on transport reliability.

Proposition 3. *The metric computed in part 3.1 is the optimal metric deriving from an inner product, for the criterion:*

$$\sum_{i,k} \; w^G_{ik} \;\; KL\left(\mathcal{D}_{P_k} \;\middle|\; T_{i\to k}(\mathcal{D}_{emp_i})\right)$$

where $\mathcal{D}_{emp_i} = \sum_j w^L_{ij} \, \delta_{\mathbf{f}_{i\to j}}$ *and* $T_{i\to k}(\delta_{\mathbf{f}}) = \delta_{T_{i\to k}(\mathbf{f})}$.

Proof. This criterion rewrites $\sum_{ijk} w^L_{ij} \, w^G_{ik} \;\ln \mathcal{D}_{P_k}(\mathbf{f}^k_{i\to j})$ which is $\sum_k KL(\mathcal{D}_{P_k} | \mathcal{D}^T_{emp_k})$, a sum of independent terms, where $\mathcal{D}^T_{emp_k} = \sum_{i,j} w^k_{i\to j} \; \delta_{\mathbf{f}^k_{i\to j}}$ is the empirical distribution of transported deformations considered in section 3.1. The optimal P_k are given by proposition 1 applied independently to each tangent space T_k with the distribution $\mathcal{D}^T_{emp_k}$. Which is precisely the content of section 3.1. □

4.4. Criterion for smooth probability distributions

We could also ask for smooth distributions with an explicit regularizer term. For the sake of simplicity, let us assume that the tangent spaces T_i are finite-dimensioned, so that probability distributions $\mathcal{D}_{P_i}$ are just functions g_i defined over T_i (times the Lebesgue measure). Similarly we denote by g^0_i the density functions related to the empirical distributions $\mathcal{D}^{\mathcal{K}}_{emp_i}$ (smoothed by a kernel in order to avoid Dirac peaks). Let us consider the usual L^2 norm between these density functions over T_i, and denote by $T_{i\to j}$

any choice of transport of functions defined over T_i, to T_j. A natural criterion to minimize would be:

$$E'(g) = \sum_i \|g_i - g_i^0\|^2_{L^2(T_i)} + \sum_{ij} w_{ij} \|T_{i\to j}(g_i) - g_j\|^2_{L^2(T_j)}$$

so that the desired distributions g_i are close to the empirical ones g_i^0 observed in the same tangent space T_i, but also so that they do not vary much when transported to close shapes S_j. At the minimum of E', we have: $\forall i,\ \partial_{g_i} E' = 0 =$

$$g_i - g_i^0 + \sum_j w_{ij}(T_{i\to j}(g_i) - g_j)T^*_{i\to j} + w_{ji}(g_i - T_{j\to i}(g_j))$$

where $T^*_{i\to j}$ is the adjoint of the linear application $T_{i\to j}$. This linear system in g can be rewritten as $Ag = g^0$, where A is a matrix of linear operators:

$$\begin{cases} A_{ii} = 1 + \sum_j w_{ij}\, T^*_{i\to j}\, T_{i\to j} + w_{ji} \\ A_{ij} = -w_{ij}\, T^*_{i\to j} - w_{ji}\, T_{j\to i} \qquad \text{for } i \neq j \end{cases}$$

In fact $A = Id + \varepsilon\Delta$ where Δ is the usual graph Laplacian, but with transports since one cannot compare directly quantities defined on different tangent spaces, and ε is related to the norm of w. Thus, A is symmetric positive definite and

$$g = A^{-1}g^0 \ = \ (Id + \varepsilon\Delta)^{-1} g^0 \ \simeq (Id - \varepsilon\Delta) g^0 \ \simeq \ \mathcal{N}_\varepsilon * g^0.$$

Thus, the optimal distribution is, in first order approximation, the empirical one smoothed over the set of shapes with a Gaussian kernel $\mathcal{N}_\varepsilon$. Moreover, up to renormalization, $g = (Id - \varepsilon\Delta)\, g^0$ coincides with the $\mathcal{D}^T_{emp}$ of the previous paragraph. The inner products (P_i) which suit $g = (g_i)$ the best in the sense of proposition 1 are precisely the ones obtained in proposition 3 and thus the ones in section 3.1. This is consequently another validation of our approach.

5. Conclusion

We proposed an approach to learn shape metrics from small training sets of highly-varying shapes, particularly suited to video analysis. The structure we build on sets of shapes relies on deformations and transport, on the contrary to distance-based methods, and allows the consideration of non-dense sample sets. We compute pairwise matchings between close shapes with possibly different topologies, transport deformations with reliability weights, and estimate smooth shape metrics in the whole training set. Thus, we generalize statistical approaches based on deformations, to the case of shape datasets with high variability, where the notion of mean pattern is not relevant anymore.

We studied several ways to estimate metrics, to propose criteria on metrics. We showed that the metric computed in our approach is the optimal one for these criteria, because of a link between Kullback-Leibler divergence and PCA.

We emphasized the new perspectives in segmentation or learning based on shapes, offered by such a transport-based structure. We showed how the metric learned can be turned into a shape matching prior. We also pointed out how to learn all notions (matching, transport) with a second pass, whose completed implementation remains future work.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *TPAMI*, Apr. 2002

[2] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *Proc. of ECCV'02*, pages 639–641, 2002.

[3] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *F.of.Comp.Math.*, 5(1), Feb. 2005.

[4] D. Cremers and M. Rousson. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In J. S. Suri and A. Farag, editors, *Parametric and Geometric Deformable Models: An application in Biomaterials and Medical Imagery*. Springer, May 2007.

[5] I. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley & Son, 1998.

[6] P. Etyngier, F. Ségonne, and R. Keriven. Shape priors using manifold learning techniques. In *ICCV*, Brazil, 2007.

[7] V. Ferrari, F. Jurie, and C. Schmid. Accurate object detection with deformable shape models learnt from images. In *Proc. of CVPR'07, Minneapolis, Minnesota*, June 2007.

[8] Y. Gdalyahu and D. Weinshall. Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes.*TPAMI*, 21(12):1312–1328, 1999.

[9] D. Geiger, A. Gupta, L. A. Costa, and J. Vlontzos. Dynamic programming for detecting, tracking, and matching deformable contours. *TPAMI*, 17(3):294–302, 1995.

[10] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Obj cut. In *Proc. of CVPR'05*, volume 1, pages 18–25 vol. 1, 2005.

[11] M. Leventon, E. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. In *Proc. of CVPR'00*, pages 316–323, South Carolina, June 2000.

[12] Y. Rathi, S. Dambreville, and A. Tannenbaum. Comparative analysis of kernel methods for statistical shape learning. In *Proc. of CVAMIA'06, Graz*, pages 96–107, 2006.

[13] Y. Rathi, N. Vaswani, and A. Tannenbaum. A generic framework for tracking using particle filter with dynamic shape prior. *Transactions on Image Processing*, 16(5), 2007.

[14] F. R. Schmidt, D. Farin, and D. Cremers. Fast matching of planar shapes in sub-cubic runtime. In *Pr. of ICCV*, 2007.

[15] T. Schoenemann and D. Cremers. Matching non-rigidly deformable shapes across images: A globally optimal solution. In *Proc. of CVPR'08*, Anchorage, Alaska, June 2008.

[16] S. Soatto and A. J. Yezzi. Deformotion - deforming motion, shape average and the joint registration and segmentation of images. *IJCV*, 53:153–167, 2002.

[17] C. J. Taylor. Active shape models - 'smart snakes. In *Proc. of BMVC'92*, pages 266–275. Springer-Verlag, 1992.

[18] A. Trouvé and L. Younes. Diffeomorphic matching problems in one dimension: Designing and minimizing matching functionals. In *Proc. of ECCV '00*, pages 573–587, 2000.

[19] M. Vaillant, M. Miller, A. Trouvé, and L. Younes. Statistics on diffeomorphisms via tangent space representation. *Neuroimage*, 2005.