Gauge invariant framework for trajectories analysis

Hassen Drira[a], Alice Barbara Tumpach[b], Mohamed Daoudi[a]

[a] *LIFL (UMR CNRS 8022), Institut Mines Télécom*
[b] *Laboratoire Painlevé (UMR 8524), université Lille1*

Abstract

We investigate the analysis of trajectories that are observed under arbitrary time evolutions. We propose a new Riemannian metric on the shape space of trajectories. The resulting metric is invariant with respect to time-warpings (or temporal re-parameterizations) of trajectories. We apply this framework to human actions recognition and to vehicle trajectory analysis.

1. Introduction

In this paper we focus on problems that deal with comparison of shapes of trajectories. One motivation comes from action recognition where features extracted from video (or RGB Depth video) frames are naturally represented by elements of nonlinear manifolds, and where temporal evolutions of an action can be modeled by trajectories on those manifolds. However, as mentionned by [1], [2] and [3], the execution rate (velocity) of activities may often vary. It follows that, without the execution invariance, two identical actions can be viewed as very different trajectories. Typical approaches for accounting for variations in execution rate are either directly based on the dynamic time warping (DTW) algorithm or some variation of this algorithm.

One promising idea is to formulate the features motion as trajectories. Matikainen et al. [4] present a method for using the trajectories of tracked feature points in a bag of words paradigm for video action recognition. Despite of the promising results obtained, the authors do not take into account the geometric information of the trajectories.

More recently, in the case of human skeleton in RGB-Depth images, Devianne et al. [5] propose to formulate the actions recognition as the problem of computing a distance between trajectories generated by the joints moving during the action. An action is a parameterized path on the shape space of the human skeleton. Similar to the ideas of Devianne et al., Su et al. [6] propose a metric which takes into account time-warping on a Riemannian manifold. They propose a metric, which allows the regitration of trajectories and compute statistics of the trajectories. Su et al. [7] apply this framework to the problem of visual speech recognition. All these approachs require a registration of trajectories. In the present paper, we propose a new theoretical framework which uses the shape information of trajectories. The main contributions of this paper are:

- The proposed framework is independent of time-re-parameterization of trajectories in $\mathbb{R}^3$.
- A new rate-invariant metric on the shape space of trajectories is proposed. No trajectories registration is required.
- We demonstrate the use of this framework theory in two computer vision applications.

The rest of this paper is organized as follows. Section 2 presents the gauge invariant framework for comparing shapes. Section 3 presents applications of the proposed approach to action recognition and to vehicle trajectories recognition.

2. Mathematical framework

The trajectory of a point in $\mathbb{R}^n$ can describe the evolution of a hurricane on the Earth, of a car driver in a city, or of the joints of a tennis player. There are two principal characteristics of the movement: the velocity of movement and the route. In this paper, we are only interested in the route used by the point, i.e. we want to be able to compare the routes of two different points irrespective of the velocity of the movements. The corresponding mathematical objects are the following : a point-trajectory will be synonymous with a parameterized curve $f : [0,1] \to \mathbb{R}^n$ and the route used by the

Preprint submitted to Workshop on DIFFerential Geometry in Computer Vision for Analysis of Shapes, Images and Trajectories

point $f(t)$ will be synonymous with the shape $[f]$ of the curve f, which is the equivalence class of f modulo the action of the re-parameterization group $\mathrm{Diff}^+([0,1])$. The shape $[f]$ of the parameterized curve f will be also called the un-parameterized curve corresponding to f. We recall this quotient construction in the next section.

2.1. Space of Trajectory Shapes

In this section the space of interest is the space of un-parameterized smooth curves in $\mathbb{R}^n$. A curve in $\mathbb{R}^n$ can be represented by a smooth function $f : [0, T_E] \to \mathbb{R}^n$ in the following way : given f, the corresponding curve is the trajectory of the point $f(t)$ when t ranges in $[0, T_E]$. The maximal value T_E of time is the duration of execution of the movement and will be set equal to $T_E = 1$. However, we will keep writing T_E instead of 1 in the body of the paper in order to avoid confusion with another duration T_D which will be the duration of deformation of a movement into another and will appear below. Two functions f_1 and f_2 represent the same shape or route if their images are the same (see Figure 1), and this happens if and only if $f_2 = f_1 \circ \gamma$, where γ is a re-parameterization of the interval $[0, T_E]$. To be fully precise the space $\mathcal{F}$ of functions considered in this paper is

$$\mathcal{F} = \{f \in C^\infty([0, T_E], \mathbb{R}^n) \,|\, f'(t) \wedge f''(t) \neq 0,\ \forall t \in [0, T_E]\},$$

and the space of shapes is denoted by $\mathcal{S}$, and is the quotient space of the space $\mathcal{F}$ by the group of time re-parameterizations $\Gamma = \mathrm{Diff}^+([0, T_E])$:

$$\text{Shape space } \mathcal{S} = \mathcal{F}/\Gamma \quad \begin{array}{l}\text{space of functions}\\ \text{modulo re-parameterization.}\end{array}$$

Both $\mathcal{F}$ and $\mathcal{S}$ are infinite-dimensional smooth Fréchet manifolds.

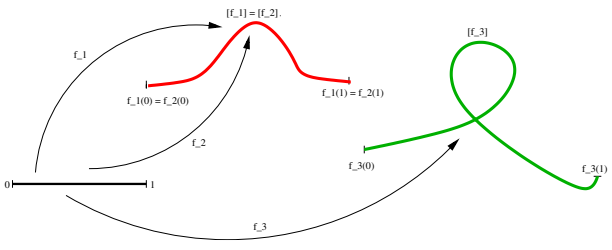

Figure 1: Two parameterized curves f_1 and f_2 corresponding to the same shape $[f_1] = [f_2]$, and a parameterized curve f_3 corresponding to a different shape $[f_3]$.

2.2. Comparison of shapes

In order to compare two shapes, i.e. two un-parameterized curves S_1 and S_2 in $\mathcal{S}$, we will quantify the minimal energy needed to deform one shape into the other. To define an appropriate energy function, we will endow $\mathcal{S}$ with a Riemannian metric $(\!(\cdot,\cdot)\!)$. A Riemannian metric allows to compute the norm of a tangent vector to $\mathcal{S}$. Note that a tangent vector to the space $\mathcal{S}$ at some curve S_1 is an infinitesimal deformation of S_1 which is given by a vector field along S_1. A deformation $s \mapsto \Psi(s)$ of one un-parameterized curve S_1 into another one S_2 is a metamorphosis of un-parameterized curves such that at deformation-time 0, the shape is S_1, i.e. $\Psi(0) = S_1$, and at deformation-time T_D the shape is S_2, i.e. $\Psi(T_D) = S_2$. Given a deformation Ψ relating two shapes S_1 and S_2, one can compute the energy of deformation $E(\Psi)$ using the Riemannian metric by integrating along the deformation the squared-norm of the velocity vector $\Psi_s = \frac{d\Psi}{ds}(s)$ of the deformation :

$$E(\Psi) = \int_0^{T_D} (\!(\Psi_s(s), \Psi_s(s))\!)_{\Psi(s)}\, ds. \qquad (1)$$

Analogously, one can compute the length $L(\Psi)$ of the deformation Ψ by integrating the norm of the velocity vector :

$$L(\Psi) = \int_0^{T_D} \sqrt{(\!(\Psi_s(s), \Psi_s(s))\!)_{\Psi(s)}}\, ds. \qquad (2)$$

Of course, there are many possible deformations of S_1 into S_2, but the energy being positive (since it is the integral of a positive function), it has a minimum. A deformation having the minimal value of energy is remarkable and is a geodesic. The minimal value of the length is called the geodesic distance between S_1 and S_2. The geodesic distance between two curves S_1 and S_2 is therefore given by :

$$d_{\mathcal{S}}(S_1, S_2) = \inf_{\Psi:[0,T_D]\to\mathcal{S},\, \Psi(0)=S_1, \Psi(1)=S_2} L(\Psi),$$

where the infimum is taken over all deformations from S_1 to S_2.

2.3. From shapes to functions

Remark that to define the distance between two un-parameterized curves using a Riemannian framework, we needed to speak about deformations of un-parameterized curves or shapes. In practice, instead of handling shapes directly, it is more convenient to handle functions f representing shapes. Indeed the curves we consider are given by the positions of sensors with respect to time, i.e. by functions f on a time-interval with

values in Euclidean space. Therefore, instead of working on the space $\mathcal{S}$ directly, we will work on the space of functions $\mathcal{F}$, and the deformation space which will play a predominant role will be the space of deformations of functions. Let us therefore introduce the space of deformations :

$$\mathcal{D} := \mathcal{C}^\infty([0, T_D], \mathcal{F}),$$

which is a smooth Fréchet manifold.

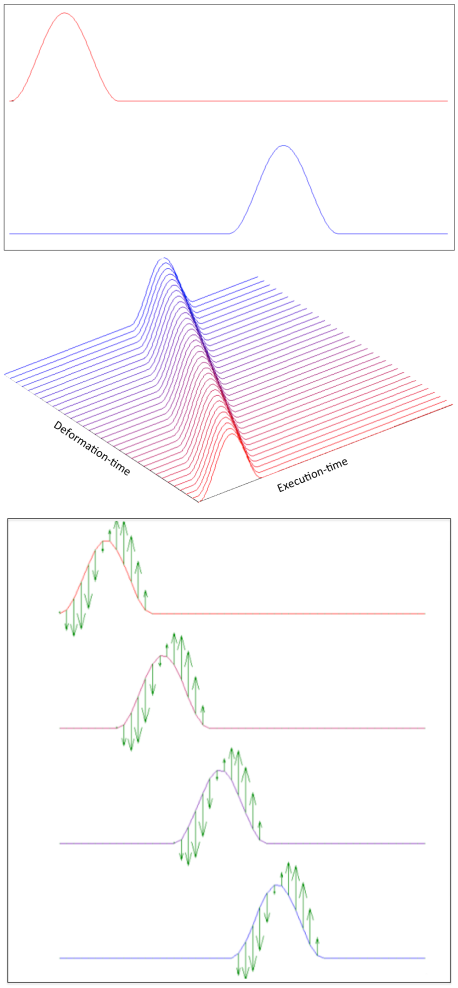

Figure 2: Upper frame : initial curve in red, final curve in blue; middle frame : a deformation of the red curve into the blue one; lower frame : velocity vector field of the deformation at 4 deformation-times (initial time, 2 intermediate times, and final time).

Note that, at this stage, there are two parameters representing an evolution : the parameter t corresponding to the velocity of execution of a movement represented by a function $f(t)$, which will be called the execution-time and ranges from 0 to T_E, and the parameter s corresponding to the deformation Ψ of a movement (more precisely of a function representing a movement) into another, which will be called the deformation-time and ranges from 0 to T_D. To have a picture in mind consider the two curves depicted in the upper frame of Figure 2. A metamorphosis from the red curve into the blue one is depicted in the middle frame of Figure 2. It is a continuous deformation of the first curve into the other. The execution-time axis is the same as the one used in the upper frame of Figure 2. The deformation-time axis is the one used to draw intermediate curves, interpolating between the red and blue curves. In the last frame of Figure 2, the starting red curve, the ending blue curve, and two intermediate curves are depicted, as well as the velocity vector field of the deformation (in green) at these four deformation-times.

2.4. *Drawback of using functions*

The drawback of using functions to encode the variation of the shape of curves is that it introduces variability in the way curves are parameterized. Recall that two parameterized curves f_1 and f_2 correspond to the same shape if and only if $f_1(t) = f_2(\gamma(t))$ for any $t \in [0, T_E]$, where γ belongs to the reparameterization group $\Gamma := \mathrm{Diff}^+([0, T_E])$. Analogously, two deformations Ψ_1 and Ψ_2 correspond to the same parameterized metamorphosis of un-parameterized curves if

$$\Psi_1(s, t) = \Psi_2(s, \gamma(s, t)), \tag{3}$$

for any $t \in [0, T_E]$ and $s \in [0, T_D]$, where this time γ belongs belongs to the group $\mathcal{G} := \mathcal{C}^\infty([0, T_D], \Gamma)$ of time-dependent reparameterizations. The relation (3) will be written

$$\Psi_1 = \gamma^{-1} \cdot \Psi_2 \tag{4}$$

for short. Note that at each deformation-time $s \in [0, T_D]$, the function $t \mapsto \gamma(s, t)$ belongs to Γ and is a reparameterization of the curve $t \mapsto \Psi_2(s, t)$. The group $\mathcal{G}$ is called the *gauge group*, and one says that $\mathcal{G}$ acts by *gauge transformations* on the space of deformations $\mathcal{D}$.

2.5. *Choice of a Riemannian metric on Shape space*

Recall that the unit tangent vector field to the parameterized curve f is defined as $\vec{v} = \frac{f'}{\|f'\|}$, the unit binormal is $\vec{b} = \frac{f' \wedge f''}{\|f' \wedge f''\|}$ and the unit normal is $\vec{n} = \vec{b} \wedge \vec{v}$. Any parameterized curve f can be re-parameterized according to the arc-length $l = \int_0^{T_E} \|f'(t)\| dt$ into a curve with constant speed. If δf is a function taking any $t \in [0, T_E]$ to a vector of $\mathbb{R}^n$ based at $f(t)$ (i.e. $\delta f \in T_f \mathcal{F}$), then its derivative with respect to arc-length is defined as

$$D_l(\delta f)(t) := \frac{(\delta f)'(t)}{\|f'(t)\|}.$$

Let us introduce the following Riemannian metric on the space of parameterized curves $\mathcal{F}$:

$$G^{a,b,c}(\delta f, \delta f) = \int_0^{T_E} \left(a\langle D_l \delta f, \vec{v}\rangle^2 + b\langle D_l \delta f, \vec{n}\rangle^2 + c\langle D_l \delta f, \vec{b}\rangle^2\right) \|f'\| dt \quad (5)$$

where $\delta f \in T_f\mathcal{F}$ and a, b, c are positive constants. For $c = 0$, this metric was introduced in [8] and called an elastic metric. It was shown in [9] that it is, in the case of plane curves, a flat Riemannian metric.

The important property of this metric on $\mathcal{F}$ is that it is Γ-invariant, that is $G^{a,b,c}(\delta f, \delta f) = G^{a,b,c}(\delta f \circ \gamma, \delta f \circ \gamma)$, for any re-parameterization $\gamma \in \Gamma = \text{Diff}^+([0, T_E])$. One consequence of this property is that it induces a Riemannian metric on the quotient space $\mathcal{S}$ such that the quotient map is a Riemannian submersion. However to compute the quotient Riemannian metric on $\mathcal{S}$, an optimization over the infinite-dimensional group of re-parameterizations is needed, leading to extra computational costs. The comparison of shapes using the quotient elastic metric via an optimization over the re-parameterization group has been implemented in [8].

In the present paper, we want to avoid the optimization step in order to reduce computational cost. For this purpose, we will use another consequence of the Γ-invariance of the metric $G^{a,b,c}$. Namely that its restriction to the normal vector fields defines a Riemannian metric on the quotient space, which is different from the quotient metric but as good as the quotient metric for comparing shapes, and which has the advantage of giving the same 'distance' $d(f_1, f_2) = d(f_1', f_2')$ for any parameterized curves $f_1' \in [f_1]$ and $f_2' \in [f_2]$. Let us first explain what we mean by normal vector fields.

Definition 1. A vector field V along a parameterized curve f is said to be a *normal vector field* if $V(t)$ is orthogonal to the unit tangent vector field $\vec{v}(t) = \frac{f'(t)}{\|f'(t)\|}$, for any $t \in [0, T_E]$:

$$\langle V(t), \vec{v}(t)\rangle = 0, \quad \forall t \in [0, T_E],$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product of $\mathbb{R}^n$.

In a complementary manner, we define the space of tangent vector fields as follows.

Definition 2. A vector field V along a parameterized curve f is said to be a *tangent vector field* if

$$V(t) = u(t)\, \vec{v}(t), \quad \forall t \in [0, T_E]$$

for a real-valued function $u : [0, T_E] \to \mathbb{R}$.

Note that the tangent vector fields are precisely the vector fields generated by the infinitesimal action of the re-parameterization group Γ. Indeed, an infinitesimal re-parameterization of a parameterized curve f does not change the shape of f, hence can only result in a rearrangement of the points along the curve f. Note also that, given a vector field V along f, we will denote by V^T the component of V tangent to f, and by $V^\perp$ the component of V orthogonal to f. One has :

$$V^T = \langle V, \vec{v}\rangle \vec{v}$$

and

$$V^\perp = V - \langle V, \vec{v}\rangle \vec{v}.$$

Hence $V = V^\perp + V^T$. Then we have the following :

Proposition 1. *The non-negative semi-definite inner product on $\mathcal{F}$ defined by*

$$((\delta f, \delta f))_f := G^{a,b,c}(\delta f^\perp, \delta f^\perp) \quad (6)$$

induces a Riemannian metric on the quotient space $\mathcal{S}$ and satisfies the gauge invariance condition

$$L(\Psi) = L(\gamma \cdot \Psi), \quad (7)$$

for any time-dependent re-parameterization $\gamma \in \mathcal{G} = C^\infty([0, T_D], \Gamma)$, where

$$L(\Psi) = \int_0^{T_D} \sqrt{((\Psi_s(s), \Psi_s(s)))_{\Psi(s)}}\, ds \quad (8)$$

is the length of the deformation Ψ.

Proof. The idea of the proof is that the product $((\delta f, \delta f))_f$ is zero precisely when $\delta f^\perp = 0$, or equivalently when δf is tangent to the parameterized curve f. This happens if and only if δf is generated by an infinitesimal re-parameterization of the parameterized curve f. Taking the quotient by the re-parameterization group Γ amounts therefore precisely to cancelling out the tangent vector fields. Hence the resulting inner product on shape space is positive-definite and defines a Riemannian metric on $\mathcal{S}$. The gauge-invariance of the length is a direct consequence of the fact that the gauge group generates vector fields at which the inner product (6) vanishes.

Corollary 1. *The Riemannian distance $d_{\mathcal{S}}$ on the Shape space $\mathcal{S}$ for the Riemannian metric induced by* (6) *satisfies :*

$$d_{\mathcal{S}}([f_1], [f_2]) = d(f_1', f_2'),$$

for any parameterized curves $f_1' \in [f_1]$ and $f_2' \in [f_2]$, where

$$d(f_1', f_2') = inf_{\Psi\, :[0,T_D]\to\mathcal{F},\, \Psi(0)=f_1', \Psi(1)=f_2'}\, L(\Psi). \quad (9)$$

The practical consequence of the previous corollary is that, contrary to the Riemannian framework on the space of functions $\mathcal{F}$, *any* pair of functions f_1 and f_2 representing two given shapes S_1 and S_2 respectively is good enough to compute the geodesic distance in Shape space $\mathcal{S}$. Recall that when $\mathcal{F}$ is endowed with a Riemannian metric, the geodesic distance in shape space between S_1 and S_2 is obtained by minimizing over the reparameterization group the geodesic distance between f_1 and $f_2 \circ \gamma$ where f_1 and f_2 are such that $[f_1] = S_1$ $[f_2] = S_2$, and where $\gamma \in \Gamma$. This minimization is necessary in the Riemannian framework on $\mathcal{F}$, since the geodesic distance varies when moving in the fiber of the quotient map $\mathcal{F} \to \mathcal{S}$. In the present framework, the function $d(\cdot,\cdot)$ defined by (9) is not properly speaking a distance function on $\mathcal{F}$ since $d(f_1, f_1 \circ \gamma) = 0$ for any $\gamma \in \Gamma$, so the inner product (6) on $\mathcal{F}$ does not define a Riemannian metric on $\mathcal{F}$ since it has a kernel. But the kernel has been chosen to ensure that the function $d(f_1', f_2')$ does not vary when f_1' and f_2' are moving independently in $[f_1'] = S_1$ and $[f_2'] = S_2$ (contrary to the geodesic distance in a Riemannian framework). This property is a consequence of the vanishing of the inner product on tangent vector fields together with the Γ-invariance inherited from the Γ-invariance of the elastic metric $G^{a,b,c}$ and of the space of normal vector fields.

2.6. Implementation of the Rate-invariant comparison of shapes

As mentioned before, the advantage of this gauge-invariant construction is that there is no need to optimize over the re-parameterization group. In practice, the deformation cost function used to compare two shapes $S_1 = [f_1]$ and $S_2 = [f_2]$ is :

$$E = \inf_{\Psi:[0,T_D]\to\mathcal{F}|,\, \Psi(0)=f_1, \Psi(T_D)=f_2} E(\Psi),$$

where

$$E(\Psi) = \int_0^{T_D} G^{a,b,c}\left(\frac{d\Psi}{ds}^{\perp}, \frac{d\Psi}{ds}^{\perp}\right) ds,$$

and

$$\frac{d\Psi}{ds}^{\perp} = \frac{d\Psi}{ds} - \left\langle \frac{d\Psi}{ds}, \vec{v} \right\rangle \vec{v}.$$

The infimum in the definition of the cost function E will be approximated using a path-straightening method explained in detail in Algorithm 1. The algorithm has as input two parameterized curves f_1 and f_2 representing two shapes $[f_1]$ and $[f_2]$, and a basis of perturbation B. An element of B is a vector field on the path connecting f_1 to f_2.

Input: Two trajectories f_1 and f_2, a basis of perturbation B.
Output: The minimal energy needed to deform f_1 into f_2 given by the value of the cost function E.

Set $\triangledown E = 1$.
1- Pre-processing step : $i = 1, 2$ decompose f_i into Fourier series $f_i(t) = a_0 + \sum_m a_m \cos\left(\frac{2\pi m}{T_E} t\right) + b_m \sin\left(\frac{2\pi m}{T_E} t\right)$ where a_0, a_m and b_m are constant vectors in $\mathbb{R}^n$ and cut the high-frequency components, i.e. replace f_i by $\tilde{f_i} = a_0 + \sum_m^M a_m \cos\left(\frac{2\pi m}{T_E} t\right) + b_m \sin\left(\frac{2\pi m}{T_E} t\right)$ for some chosen M.
2- for $i = 1, 2$ parameterize $\tilde{f_i}$ by arc-length and resample it uniformly using a spline interpolation.
3- Initialize the path Ψ between $\tilde{f_1}$ and $\tilde{f_2}$ by the linear interpolation between $\tilde{f_1}$ and $\tilde{f_2}$.
while $\triangledown E > 10^{-3}$ **do**
  4- Compute the energy E of the path Ψ.
  5- Set $\Psi_{upd} = 0$.
  for *ind* $\leftarrow 1$ **to** *size*(B) **do**
    6- Add a perturbation to the current path Ψ: $\Psi_{eps} = \Psi + \epsilon_1 B(ind)$, where $B(ind)$ is the element of the pertubation basis B of index *ind*.
    7- Compute the energy E_{eps} of the perturbed path Ψ_{eps}.
    8- Compute the gradient of energy $\triangledown E_{eps}$ in the direction $B(ind)$ using the approximation $\triangledown E_{eps} \sim \frac{E_{eps} - E}{\epsilon_1}$.
    9- Compute the updating path: $\Psi_{upd} = \Psi_{upd} + \triangledown E \cdot G(ind)$
  end
  10- Update the path: $\Psi = \Psi - \epsilon_2 \Psi_{upd}$
end

Algorithm 1: Computation of the cost function E.

3. Applications

3.1. Human actions recognition

An action is a sequence of frames forming a movie. Typical actions are walking, running, rotating. The speed of execution of the action induces variability that one would like to remove. Indeed, the walk can be slow or fast, but the movements accomplished in a fast walk are closer to the ones accomplished in a slow walk then in a run. In the context of action recognition, one major challenge is to be able to distinguish actions like a fast walk and a slow run.

The distribution of activity-specific temporal warpings, represents the space of all permissible time-

warping transformations for each activity [1]. During actions, trajectories of skeleton's joints are built and can be analyzed for classifing the actions. An example of a skeleton extracted from the Kinect stream is reported in Figure 3.1 (at the right). The depth and the color stream are also reported in this Figure.

Several authors [10, 11, 12] use the skeleton extracted using the Kinect for action recognition. In Figure 4, we report the skeleton's tracking during human actions. The first two rows report the same action with different rates. It is clear that the trajectories generated by joints present time-warping transformations and this is a challenging task in action recognition.

Using the Kinect, the human skeleton can be extracted from depth images in real-time thanks to the work proposed by Shotton et al.[13] where a real-time method is defined to accurately predict 3D positions of body joints (20 joints) in individual depth maps without using any temporal information.

For given human action sequences, we propose to compare the pairwise trajectories of corresponding joints using the gauge invariant framework described in the previous section. The sum over all joints of the resulting distances represents the dissimilarity score. This score is used for human action recognition. We notice that we detect the specific case when the two corresponding joints do not move considerably. In this case, the distance is forced to zero and the calculation is avoided. This idea is illustrated in Fig. 4. The trajectories illustrated in the upper two rows correspond to the same action executed with different rates. However, the action corresponding to the trajectories in the two lower row is different. The trajectories generated by the same action present similar shapes with different time execution. In this case, using the gauge invariant framework presented previously, we argue that the distance between two trajectories corresponding to the same action will be small without any need for time alignment.

We propose to use data from a public dataset : MSR Action 3D dataset [10] on which many methods have been evaluated. This dataset includes 20 actions performed by 10 persons facing the camera. Each action is performed 2 or 3 times. In total, 567 sequences are available. The different actions are high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw X, draw tick, draw circle, hand clap, two hand wave, side-boxing, bend,..
We perform our test on a subset of 30 human actions from MSR dataset with different rates. Examples of the actions are reported in Figure 4; the two upper rows represent the same action with different rates. The down two rows reports another action with different rates. We success to correctly classify 29 actions over the 30 used in this small experiment.

For further understanding of the limitation of our method, the action that was not recognized by our approach is reported in the first row of Figure 5. The closest action is reported in the lower row. The two actions are quite similar.

3.2. Vehicle Trajectories recognition

Here we study the problem of classifying vehicle trajectories into broad motion patterns using data obtained from traffic videos. While the general motion of a vehicle at a traffic intersection is predicable: left turn, right turn, U turn, or straight line, the travel speeds of vehicles may be different in different instances due to traffic variations. Since we are interested in tracking position and orientation of a vehicle, we will consider individual tracks as parameterized trajectories. The data for this experiment comes from traffic videos available at the Image Sequence Server website[1]. Examples of this dataset are illustrated in Fig. 6. From left to right, are illustrated a left turn, a right turn and a straight line.

Experiments show that the distances corresponding to similar trajectories are much smaller in terms of the geodesic distances, when compared to the distances between different trajectories. The 14×14 matrix shows pairwise distances between 14 trajectories, associated with three classes (left turn, right turn, U turn, or straight line) corresponding to right turn indexed from 1 to 5, 5 trajectories of straight line indexed from 6 to 10, and 4 trajectories of left turn indexed from 11 to 14. The resulting distances are shown as a matrix in Figure 7. In this visualization of the matrix, the lightness of each element (i;j) is proportional to the magnitude of the distances between actions i and j. That is, each row and column represent the distances for a trajectorie when compared to all 14 trajectories. Darker elements represent better matches, while lighter elements indicate worse matches. Shown in the first row of Figure 7 are the resulting pairwise distance matrices computed by our approach (c) and from [6] with and without temporal alignment; respectively (a) and (b).

We have performed a dendrogram clustering vehicule trajectories using pairwise distances. The dendrogram can be analyzed by slicing them with an horizontal line and see the resulting clustering. The classification is raffined as well as the horizontal line moves down. To analyze the results, we consider the case of the classification into 3 classes. As these pictures illustrate, the

[1] http://i21www.ira.uka.de/image sequences/

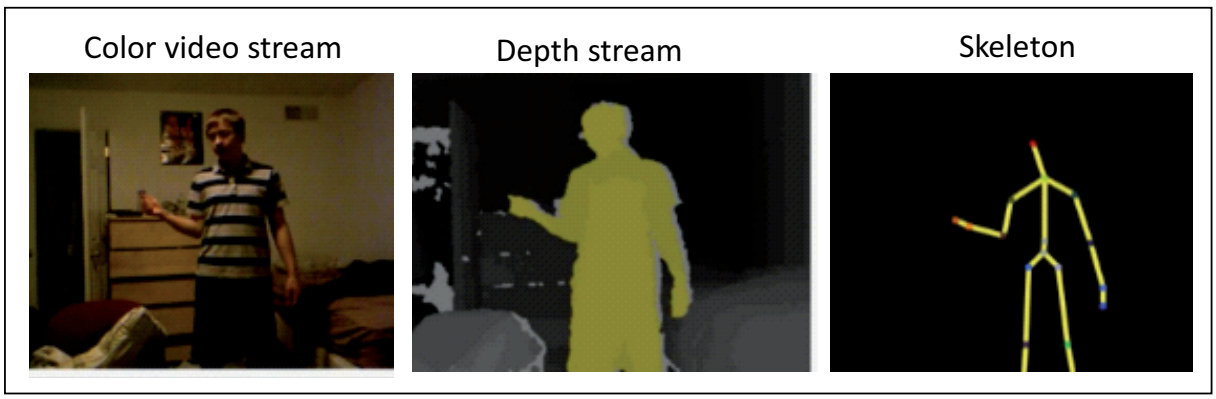

Figure 3: An example of human captured by the Kinect. From left to right are reported the color stream, the depth stream and the skeleton.

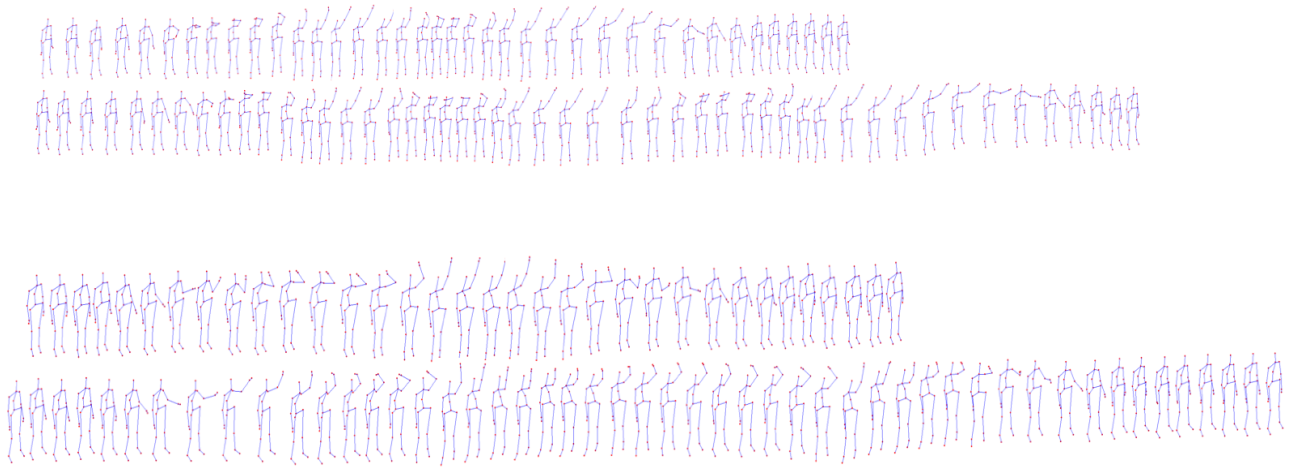

Figure 4: Illustration of skeleton's tracking during human actions. The upper two rows report the same action conveyed in different rates. The remaining two rows illustrate another action with different rates.

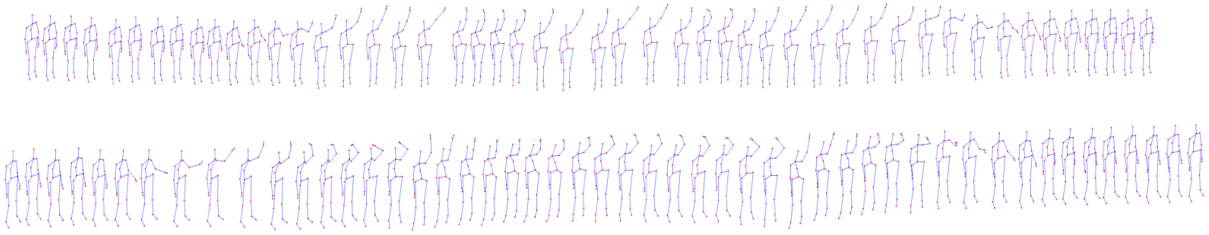

Figure 5: Illustration of non recognized action (upper row) and the closest one (lower row).

clustering is quite successful by our approach although no time registration is performed. The same trajectory are mostly clustered together, away from the clusters for trajectories. The classification results obtained by our approach are comparable to the results obtained by [6] which required trajectories registration.

It can be clearly seen that the temporal alignment proposed by [6] helps in revealing the underlying patterns of the data and improves the clustering performance. Actually, with temporal alignment, the classes were correctly distinguished. The proposed approach performed similar accuracy without any temporal alignment. We demonstrate here that by defining the metric directly on the shape space, we are able to perform as good as with

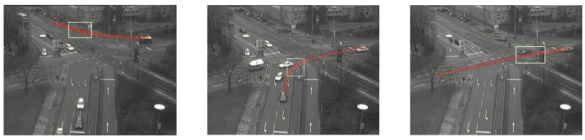

Figure 6: An example vehicle trajectories : right turn (left pannel), left turn (middle pannel) and straignt line (right pannel).

temporal alignment.

4. Conclusion

Analysis of trajectories are very important in many areas, e.g. medical imaging, computer vision etc. In this paper, we have provided a gauge invariant framework for comparing trajectories while being invariant to time-warping.

Specifically, we have defined a proper Riemannian metric directly on the quotient (shape) space, rather than inheriting it from pre-shape space. We have used it to formulate a path energy that measures only the normal components of velocities along the path. The geodesic computation is based on a path-straightening technique that iteratively corrects paths between curves until geodesics are achieved.

Both theoretical proofs and experimental results on trajectories from human action and vehicles trajectories are provided to validate this framework. For future work, we would like to extend it to other applications with different underlying manifolds.

Appendix .1. Expression of the Riemannian metric on Shape space

To be fully complete, let us give the expression of the Riemannian metric obtained on Shape space by the Gauge invariant construction. Let $\delta f \in T_f\mathcal{F}$ be given by $\delta f = h_v\vec{v} + h_n\vec{n} + h_b\vec{b}$, where h_v, h_n and h_b are real functions. Then $(\delta f)^\perp = h_n\vec{n} + h_b\vec{b}$ and

$$D_l(\delta f)^\perp = -h_n\kappa\vec{v} + (D_l h_n - h_b\tau)\vec{n} + (D_l h_b + h_n\tau)\vec{b},$$

where κ denotes the curvature, and τ the torsion of the parameterized curve f. Therefore $\langle D_l\delta f, \vec{v}\rangle^2 = h_n^2\kappa^2$, $\langle D_l\delta f, \vec{n}\rangle^2 = (D_l h_n - h_b\tau)^2$ and $\langle D_l\delta f, \vec{b}\rangle^2 = (D_l h_b + h_n\tau)^2$. It follows that

$$G^{a,b,c}(\delta f^\perp, \delta f^\perp) = \int_0^{T_E} \left(ah_n^2\kappa^2 + b(D_l h_n - h_b\tau)^2 + c(D_l h_b + h_n\tau)^2\right)\|f'\|dt$$

In other word, the Riemannian metric on shape space we used in this paper is the following : given a shape $[f]$, and real functions h_n and h_b of the arc-length parameter l of $[f]$, the vector field $h = h_n v + h_b b$ defines uniquely a infinitesimal deformation of the shape $[f]$ whose norm is given by

$$((h,h))_{[f]} = \int_0^{T_E} \left(ah_n^2\kappa^2 + b(D_l h_n - h_b\tau)^2 + c(D_l h_b + h_n\tau)^2\right)\|f'\|dt. \quad (.1)$$

The Riemannian metric on $\mathcal{S}$ defined by this formula is a H^1-type metric, depending smoothly on the curvature and the torsion.

References

[1] A. Veeraraghavan, A. Srivastava, A. K. Roy-Chowdhury, R. Chellappa, Rate-invariant recognition of humans and their activities, Image Processing, IEEE Transactions on 18 (6) (2009) 1326–1339. doi:10.1109/TIP.2009.2017143.

[2] F. Zhou, F. De la Torre Frade, Generalized time warping for multi-modal alignment of human motion, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[3] R. Slama, H. Wannous, M. Daoudi, Grassmannian representation of motion depth for 3d human gesture and action recognition, in: International Conference On Pattern Recognition ICPR, 2014.

[4] P. Matikainen, M. Hebert, R. Sukthankar, Trajectons: Action recognition through the motion analysis of tracked features, in: Workshop on Video-Oriented Object and Event Classification, ICCV 2009, 2009.

[5] M. Devanne, H. Wannous, S. Berretti, P. P. Pala, M. Daoudi, A. D. Bimbo, Space-time pose representation for 3d human action recognition (2013) 1.

[6] J. Su, S. Kurtek, E. Klassen, A. Srivastava, Statistical analysis of trajectories on riemannian manifolds: Bird migration, hurricane tracking and video surveillance, Annals of Applied Statistics 8 (1).

[7] J. Su, A. Srivastava, F. D. M. de Souza, S. Sarkar, Rate-invariant analysis of trajectories on riemannian manifolds with application in visual speech recognition, 2014.

[8] W. Mio, A. Srivastava, S. H. Joshi, On shape of plane elastic curves, International Journal of Computer Vision 73 (3) (2007) 307–324.

[9] M. Bauer, M. Bruveris, S. Marsland, P. W. Michor, Constructing reparametrization invariant metrics on spaces of plane curves, Differential Geometry and its Applications 34 (2014) 139–165.

[10] W. Li, Z. Zhang, Z. Liu, Action recognition based on a bag of 3d points, in: Workshop on Human Communicative Behavior Analysis, 2010, p. 914.

[11] L. Xia, C. Chen, J. K. Aggarwal, View invariant human action recognition using histograms of 3d joints, in: Workshop on Human Activity Understanding from 3D Data, Providence, Rhode Island, USA, 2012, p. 2027.

[12] X. Yang, Y. Tian, Eigenjoints-based action recognition using naive-bayes-nearest-neighbor, in: Workshop on Human Activity Understanding from 3D Data, Providence, Rhode Island, USA, 2012, p. 1419.

[13] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: Workshop of IEEE Conf. on Computer Vision and Pattern Recognition, Colorado Springs, Colorado, USA, 2011, pp. 1–8.

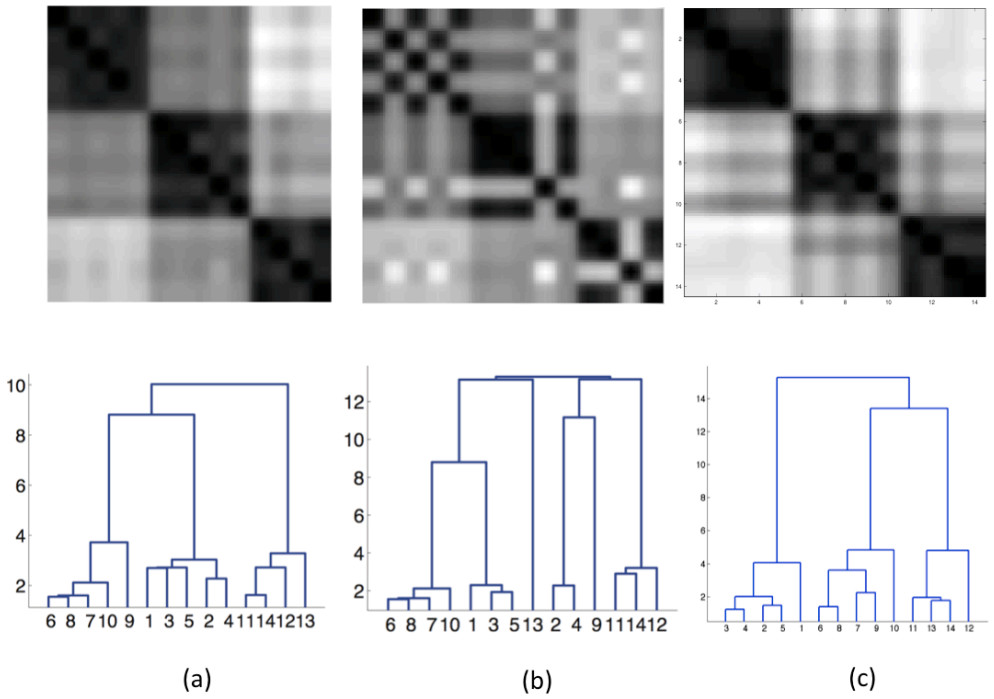

Figure 7: Comparison of vehicle trajectories distance matrices and dendrograms resulting by [6] with temporal alignment (a), [6] without temporal lignment (b) and our approach (c).