On Signature Invariants for Effective Motion Trajectory Recognition

Shandong Wu and Y.F. Li

The International Journal of Robotics Research 2008; 27; 895
DOI: 10.1177/0278364908091678

Shandong Wu
Y. F. Li
Department of Manufacturing Engineering and
Engineering Management,
City University of Hong Kong,
83 Tat Chee Avenue, Kowloon, Hong Kong
s.d.wu@student.cityu.edu.hk
meyfli@cityu.edu.hk

On Signature Invariants for Effective Motion Trajectory Recognition

Abstract

Motion trajectory can be an informative and descriptive clue that is suitable for the characterization of motion. Studying motion trajectory for effective motion description and recognition is important in many applications. For instance, motion trajectory can play an important role in the representation, recognition and learning of most long-term human or robot actions, behaviors and activities. However, effective trajectory descriptors are lacking and most reported work just uses motion trajectory in its raw data form. In this paper, we propose a novel motion trajectory signature descriptor and study its rich descriptive invariants which benefit effective motion trajectory recognition. These invariants are key measures of the flexibility and effectiveness of a descriptor. Substantial descriptive invariants can be deduced from the proposed trajectory signature, which is attributed to the computational locality of the signature components. We first present the signature definition and its robust implementation. Then the signature's invariants are elaborated. A non-linear inter-signature matching algorithm is developed to measure the signature's similarity for trajectory recognition. Experiments are conducted to recognize human sign language, in which both synthetic and real data are used to verify the signature's invariants, and to illustrate the effectiveness in the signature recognition.

KEY WORDS—invariants, signature, trajectory descriptor, motion trajectory recognition, robot vision.

1. Introduction

A motion trajectory here consists of a set of position vectors of sampled points of interest in spatiotemporal motion. It is an informative and meaningful description due for the characterization of motion. Motion trajectory based motion description is the foundation for many applications such as robot motion analysis, motion modeling for robot learning (Bennewitz et al. 2005) and human behavior recognition (Psarrou et al. 2002). A typical situation is that most long-term human or robot actions, behaviors and activities can be well described by the underlying motion trajectories. Therefore, studying motion trajectory for effective motion description and recognition is of practical importance.

Motion trajectories have been extensively explored in recent research. Yang and Ahuja (1999) investigated hand gesture recognition using extracted gestures motion trajectory. Meyer et al. (1998) performed gait classification using a hidden Markov model with multiple motion trajectories extracted from several human body parts. As the motion of human body is usually small in relation to a full human body, Min and Kasturi (2004) proposed a significant motion points (SMPs) method to track stable trajectories from human hands and feet for motion recognition. In the work by Black and Jepson (1998), human motion was modeled as motion trajectory and the condensation algorithm was applied for trajectory matching in gesture and expression recognition. Aleotti and Caselli (2005) studied robot task representation via modeling motion trajectory using NURBS (Non-Uniform Rational B-spline). Ude et al. (2000) investigated the use of a B-spline wavelet to describe humanoid joint motion trajectory for robot learning.

However, in most existing work, motion trajectory was either used directly in its raw data form or represented by simple descriptors without considering their effectiveness. Raw trajectory data are quite inflexible to use because they rely much on absolute positions. The effectiveness of a trajectory descriptor can be measured in terms of three criteria. First, a descriptor is supposed to be able to capture complete shape and detailed motion features for motion description and (high-level) analysis. Second, high accuracy and efficiency in computation and classification are desired. Third, the invariants (Forsyth et al. 1991) are important measures for a flexible and adaptable in

The International Journal of Robotics Research
Vol. 27, No. 8, August 2008, pp. 895–917
DOI: 10.1177/0278364908091678

motion description. Improvements on any of the abovementioned aspects will enhance the performance of a descriptor.

In this paper, we propose a novel and generic signature mechanism to serve as an effective descriptor for free form motion trajectory modeling. As a computationally local descriptor, the signature can fully describe a trajectory's shape and capture the detailed motion features. More importantly, the signature has its advantage in descriptive invariants. In addition to the rigid invariant and metric invariant, the signature is also invariant with respect to visual projection and viewpoint changes. In addition, the speed invariant and occlusion invariant can also be deduced from the signature. Once attributed to the six invariants, motion trajectory can be described and recognized more effectively using the signature descriptor.

In the existing work, some shape descriptors have been built for both motion trajectories and still curves. Simple contour functions such as chain code, centroid–contour distance (CCD) and R–S curve just admit limited invariants (Kindratenko 2003). Geometric moments based descriptors are widely adopted for region oriented shape representation (Hu 1962), which can be regulated in translation, rotation, metric and skew invariant forms. However, moment functions suffer from occlusion since they use global features.

Mathematical descriptors based on algebraic curve, spline (Chand and Doty 1985), B-spline (Cohen et al. 1995; Charlebois et al. 1999) and NURBS (Aleotti and Caselli 2005) can be used to model the motion trajectory, object contour or surface. These kinds of descriptors need data fitting to obtain descriptive parameters for shape representation, inevitably causing inaccuracy in the fitting. The B-spline based description may even result in recognition ambiguity as it is hard to compare B-spline parameters directly for recognition because a piece of trajectory is not uniquely described by a single set of control points (Cohen et al. 1995). An algebraic curve is not invariant to rigid transformation unless it is centering the polynomial at the origin of the coordinate system (Kindratenko 2003). Both B-spline and NURBS are affine invariant. NURBS is even invariant under perspective transformation as studied by Alferez and Wang (1999). However, this kind of invariant is just a structure invariant, in which the fitting and operation of the control points and weights are complicated.

The descriptors based on principal components analysis (PCA; Jiang and Motai (2005)), Fourier descriptor (FD; Harding and Ellis (2004)), wavelet transform (Chuang and Kuo 1996) and curvature scale space (CSS; Mokhtarian et al. 1996) can represent shape in a coarse-to-fine or multi-scale manner, in which only partially salient features are of concern in the shape description. This explains why they are incapable of fully capturing motion properties because much of the less-important information is ignored. In particular, in Fourier transform, it is difficult to perform local motion analysis in the frequency domain because time information is lost. In terms of invariants, while linear PCA is rotation, translation and metric invariant, the invariants of the non-linear PCA rely on the property of the kernel used. FD is translation, rotation and metric invariant subject to simple normalization, provided that the zero-frequency component (shape center) is ignored. Wavelet transform can remain rigid invariant, but it still depends on an extra normalization process. Curvature in Euclidean space is rigid invariant, but the CSS representation is inexact as the trajectory's length shrinks with trajectory evolution.

It is feasible to combine more than one of the abovementioned descriptors to design multi-level descriptors, such as FD plus wavelet (Kunttu et al. 2003), B-spline plus moments (Huang and Cohen 1996) as well as moments plus wavelet and FD (Bui and Chen 1997). Also, FD, wavelet, CSS and moments can be further regulated by developing affine invariants (Khalil and Bayoumi 2002; Chaker et al. 2003). As for the visual projective (imaging) invariant, several geometric invariants (Weiss 1993) were investigated such as the cross ratio, four coplanar lines (meeting at one point) and some mathematical elements such as conic sections. However, they only work for simple shapes and it is hard to see benefit for the description for complex shapes.

For a comparison of more descriptors, readers are referred to the surveys by Forsyth et al. (1991) and Kindratenko (2003). In essence, a descriptor's performance, especially the invariant, depends much on its structure and the kind of the feature we are interested in. Instead of a global description, the intrinsic features such as differential invariants (Kehtarnavaz and Figueiredo 1988) and integral invariants (Manay et al. 2006) admit richer invariants. The signature descriptor proposed here is based on local differential invariants, which is the basis for deducing the abovementioned six descriptive signature invariants.

The paper is organized as follows. In Section 2, the signature descriptor is formulated and implemented. Section 3 is dedicated to the elaboration of the signature invariants. After that, a metric for signature recognition is developed in Section 4. Section 5 proceeds with experiments, followed by a general discussion and conclusion in Section 6.

2. Motion Trajectory Signature Descriptor

2.1. Signature Principle

Motivated by the two-dimensional still curve representation in differential geometry (Calabi et al. 1998), we propose the following three-dimensional trajectory signature definition using differential invariants: for a free form space motion trajectory $\Gamma(t)$ parameterized by

$$\Gamma(t) = \{X(t), Y(t), Z(t) \mid t \in [1, N]\}, \tag{1}$$

where N is the trajectory length (number of the sampled points), its three-dimensional Euclidean signature S is defined

in terms of four differential invariants, say, curvature (κ), torsion (τ) and their first-order derivatives with respect to Euclidean arc-length parameter s (κ_s and τ_s), in the following form:

$$S = \{[\kappa(t), \kappa_s(t), \tau(t), \tau_s(t)] \mid t \in [1, N]\}. \quad (2)$$

For a point sampled at temporal index t, its curvature $\kappa(t)$ and torsion $\tau(t)$ are calculated by

$$\kappa(t) = \left\|\dot{\Gamma}(t) \times \ddot{\Gamma}(t)\right\| / \left\|\dot{\Gamma}(t)\right\|^3, \quad (3)$$

$$\tau(t) = (\dot{\Gamma}(t) \times \ddot{\Gamma}(t)) \cdot \dddot{\Gamma}(t) / \left\|\dot{\Gamma}(t) \times \ddot{\Gamma}(t)\right\|^2. \quad (4)$$

As the Euclidean arc-length is calculated by $s(t) = \int_0^t \left\|\dot{\Gamma}(t)\right\| dt$, we have $ds(t)/dt = \left\|\dot{\Gamma}(t)\right\|$. Hence, $\kappa_s(t)$ and $\tau_s(t)$ can be derived as follows:

$$\kappa_s(t) = \frac{d\kappa(t)}{ds} = \frac{d\kappa(t)}{dt} \cdot \frac{dt}{ds} = \frac{d\kappa(t)}{dt} \cdot \frac{1}{\left\|\dot{\Gamma}(t)\right\|}, \quad (5)$$

$$\tau_s(t) = \frac{d\tau(t)}{ds} = \frac{d\tau(t)}{dt} \cdot \frac{dt}{ds} = \frac{d\tau(t)}{dt} \cdot \frac{1}{\left\|\dot{\Gamma}(t)\right\|}. \quad (6)$$

The signature descriptor models a discretely sampled motion trajectory in terms of points and arc-lengths. First, the Euclidean curvature and torsion are two intrinsic curve features that are sufficient to describe a three-dimensional point in Euclidean space. Intuitively, the curvature measures how far a point is from being on a straight line and torsion measures how far it is from being in a plane. Second, the modeling of a trajectory needs to take into account the description of the entire arc-length or, in other words, the arc-lengths between consecutive points. To this end, the local quantities $[\kappa_s(t), \tau_s(t)]$ are introduced as the other two signature components that can measure trajectory arc-length. As can be seen from equations (3)–(6), all four of the signature components are computed in a local sense without involving global constraints. This is the computational locality of the signature and it is the foundation of the six signature invariants.

Within the signature based description, the representation of motion trajectory is switched to the signature space. To ensure the mathematical existence of the signature, equation (1) is assumed to be regular which means that there is no stationary point along a trajectory, that is, for all t, $\dot{\Gamma}(t) \neq \vec{0}$. For an irregular trajectory, a prior segmentation can be carried out at the stationary points to obtain multiple regular segments.

The above signature principle is actually applicable to both three- and two-dimensional motion trajectories. As a special case, we give the two-dimensional signature by simplifying equation (2) for two-dimensional trajectory description. For a planar trajectory parameterized by $\Gamma_2(t) = \{X(t), Y(t) \mid t \in [1, N]\}$, its signature S_2 is defined by $S_2 = \{[\kappa(t), \kappa_s(t)] \mid t \in [1, N]\}$. It is observed that S_2 just has two components and the torsion pair $[\tau(t), \tau_s(t)]$ is no longer there. This is because, in essence, the curvature feature is sufficient to fully describe a two-dimensional point.

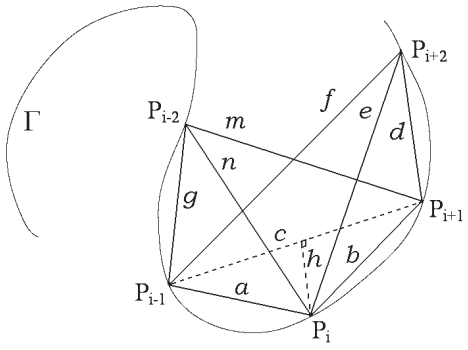

Fig. 1. Signature approximation with multiple points.

2.2. Numerical Signature Approximation

The above-formulated accurate signature depends on high-order derivatives that are usually sensitive to noise and round-off errors. To reduce this effect and make the signature robust, a so-called approximate signature is numerically implemented for practical use. To avoid high-order derivatives, the differential invariants in the accurate signature are replaced by the joint differential invariants which are computed based on multiple consecutive points rather than only a single point for the differential invariants. Note that the joint invariants are not sensitive to noise because they involve only the lowest-order derivatives (Calabi et al. 1998). Following this, the joint Euclidean invariants (i.e. the inter-point Euclidean distances) are employed to approximate the four signature components. As joint Euclidean invariants are local features, the approximate signature admits the computational locality too.

As illustrated in Figure 1, let P_{i-2}, P_{i-1}, P_i, P_{i+1} and P_{i+2} be five consecutively sampled points along a three-dimensional trajectory Γ, in which the inter-point Euclidean distances are denoted by from a to n. Denote by $H^+(H^-)$ the height of the tetrahedron with sides a, b, c, d, e, f (a, b, c, g, n, m) with respect to the point $P_{i+2}(P_{i-2})$. Then at P_i the signature quaternion of the approximate signature $S^* = \{\kappa^*, \kappa_s^*, \tau^*, \tau_s^*\}$ can be obtained by extending the derivations in Boutin (2000).

First, the curvature κ^* at point P_i can be approximated by the curvature of the circle passing through points P_{i-1}, P_i and P_{i+1}. It has been proved that the curvature of a circle is equal to the reciprocal of its radius. Denote by Δ_{abc} the area of the triangle with sides a, b and c, and define $\bar{s} = (a + b + c)/2$; the curvature κ^* is approximated by

$$\kappa^*(P_i) = 4\frac{\Delta_{abc}}{abc} = 4\frac{\sqrt{\bar{s}(\bar{s}-a)(\bar{s}-b)(\bar{s}-c)}}{abc}. \quad (7)$$

Based on Taylor series manipulations, the torsion τ^* is calculated by

$$\tau^*(P_i) = \frac{1}{2}\left(6\frac{H^+}{def \cdot \kappa^*(P_i)} + 6\frac{H^-}{gnm \cdot \kappa^*(P_i)}\right). \quad (8)$$

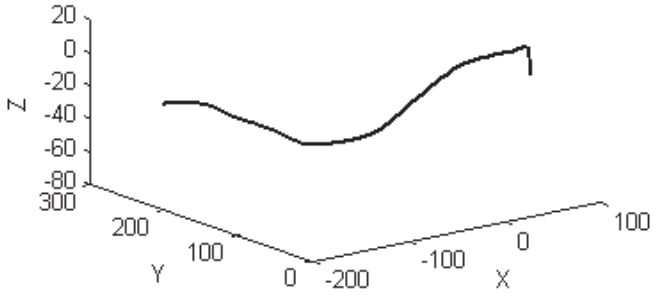

Fig. 2. A three-dimensional motion trajectory Π (length: 183 frames).

Here $H^+(H^-)$ can be derived via the calculation of the tetrahedron's (signed) volume, as formulated in the following equation:

$$\frac{1}{3!}\cdot\begin{vmatrix} x_i & y_i & z_i & 1 \\ x_{i-1} & y_{i-1} & z_{i-1} & 1 \\ x_{i+1} & y_{i+1} & z_{i+1} & 1 \\ x_{i+2} & y_{i+2} & z_{i+2} & 1 \end{vmatrix} = V_{abcdef} = \frac{\Delta_{abc}\cdot H^+}{3}, \quad (9)$$

where $\{x_j, y_j, z_j\}_{j=i-1}^{i+2}$ denote the coordinates of vertex P_j.

Furthermore, κ_s^* and τ_s^* are approximated as follows:

$$\kappa_s^*(P_i) = 3\frac{\kappa^*(P_{i+1})-\kappa^*(P_{i-1})}{2a+2b+d+g}, \quad (10)$$

$$\tau_s^*(P_i) = 4\frac{\tau^*(P_{i+1})-\tau^*(P_{i-1})+r\cdot(\tau^*(P_i)\kappa_s^*(P_i)/6\kappa^*(P_i))}{2a+2b+2d+h+g}, \quad (11)$$

where $r = 2a + 2b - 2d - 3h + g$.

Thus, combining equations (7), (8), (10) and (11), we obtain an approximate version of the trajectory signature,

$$S^* = \{[\kappa^*(P_i), \kappa_s^*(P_i), \tau^*(P_i), \tau_s^*(P_i)] \mid P_i \in \Gamma\}. \quad (12)$$

Visually, the signature curve is shown by two sub-signatures. One is the curvature sub-signature coordinated by curvature in the horizontal axis versus its derivative in vertical axis, and the other is the torsion sub-signature coordinated by torsion in the horizontal axis versus its derivative in vertical axis. As an example, for the trajectory Π shown in Figure 2, the basic forms of its accurate and approximate signatures are illustrated in Figures 3 and 4, respectively. We have two observations by comparing their signature curves. First, the high similarity in the signature shapes verifies that the approximate algorithms are able to preserve the signature features. Second, the approximate signature looks smoother and more stable than the accurate signature, as the sawteeth effects and disturbing saltation in Figure 3, which were caused by calculating the high-order derivatives, are reduced. Hence, the approximate signature is more robust.

2.3. Trajectory Smoothing and Noise Reduction

In addition to the approximate signature, trajectory smoothing is also helpful for stable signature calculation. Noise and unsteadiness usually exist in the acquired trajectories. Therefore, using raw trajectory data directly would not be advisable for signature calculation. Rather, a smoothing preprocessing should be employed to improve the signature's computational stability. The problem is that the trajectory shape may also be affected by the smoothing. Therefore, a compromise needs to be made between shape preservation and trajectory smoothing. For example, the anisotropic Perona–Malik diffusion smoothing algorithm (Perona and Malik 1990) is widely used to reduce noise to smooth the image while preserving the key edge features.

Here, two trajectory smoothers are designed to reduce trajectory noise. The first is an adaptive moving average filter, which smoothes trajectories by using the weighted average of the neighbor points. Smoothing and shape preservation can be balanced by interactively setting a tunable span parameter. The other trajectory smoother is a wavelet smoother. Although wavelet descriptor would become inaccurate when much detailed information is lost, wavelet analysis is still effective for noise reduction. To try to preserve the underlying trajectory shape, the wavelet decomposition level for approximate coefficients exaction can be tuned according to the noise strength. In normal situations when we do not know the noise strength in advance, in principle the decomposition level should be restricted to a relatively low range (e.g. 1–3) to avoid over-influencing the trajectory shape. The two filters can be used individually or combined in practice (see Section 5.3 for the application and comparison of the two smoothers).

3. Signature Invariants

In this section, six signature invariants are investigated. The proof is based on the accurate signature, but is verified using the approximate signature. With these invariants, trajectories can be treated independently of the condition changes in the space position, viewing distance (the distance between human motion and camera), viewpoint, speed and occlusion.

3.1. Euclidean Group Invariant

Trucco and Verri (1998) defined invariant as the property of geometric configurations which do not change under a certain class of transformations. For the Lie group based transformation, Cartan's theory (Olver 2001) states that the solution to the equivalence and symmetry problems of a submanifold is based on the functional interrelationships among the fundamental differential invariants. The equivalence means that the unique representation of a submanifold and the symmetry

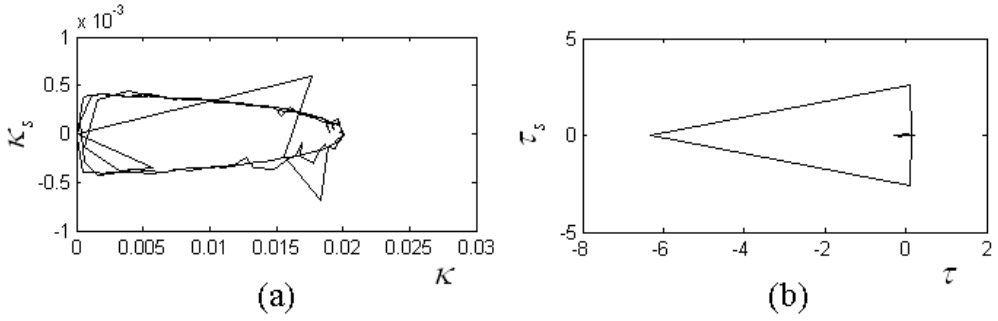

Fig. 3. The accurate signature of trajectory Π: (a) curvature sub-signature (κ versus κ_s); (b) torsion sub-signature (τ versus τ_s).

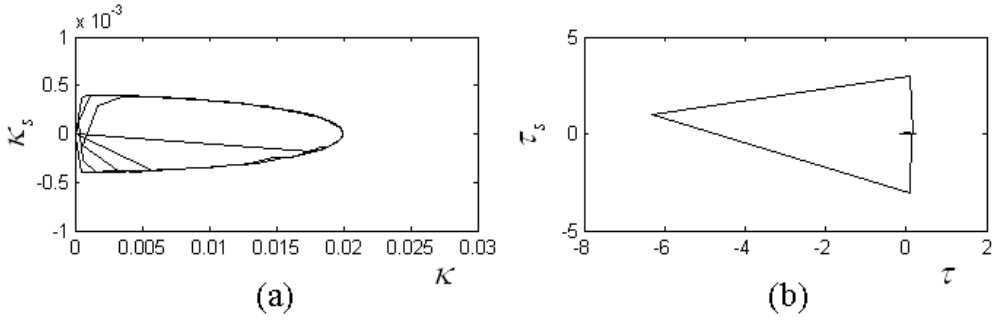

Fig. 4. The approximate version of the accurate signature shown in Figure 3.

corresponds to invariant representation with respect to group transformations. Since the signature components are based on four typical Euclidean differential invariants, the trajectory signature can remain invariant with respect o Euclidean (rigid) group transformations (translation, rotation or both).

As can be seen from equations (3)–(6), all of the signature components are based on dot and cross products of derivatives, which are purely local quantities of the trajectory features. Hence, we can also conclude that the signature is Euclidean invariant.

3.2. Metric Group Invariant

Following the Euclidean group invariant, the metric group invariant can be easily achieved by normalization. Denote the metric transformations by

$$X^{\text{Metric}} = \delta \cdot X + T_{\text{Euclidean}} \cdot X \tag{13}$$

where δ is a scaling factor. In view of the Euclidean invariant, only the scaling factor has an influence on the signature's variability. Applying a scaling of δ to a trajectory $\gamma(t)$, we obtain a new trajectory $\hat{\gamma}(t) = \delta \cdot \gamma(t)$. Then we have $\dot{\hat{\gamma}}(t) = \delta \cdot \dot{\gamma}(t)$, $\ddot{\hat{\gamma}}(t) = \delta \cdot \ddot{\gamma}(t)$, $\dddot{\hat{\gamma}}(t) = \delta \cdot \dddot{\gamma}(t)$, $\left\|\dot{\hat{\gamma}}(t)\right\| = \delta \cdot \|\dot{\gamma}(t)\|$ and $\hat{s}(t) = \delta \cdot s(t)$. Based on these quantities and referring to equations (3)-(6), the signature of $\hat{\gamma}(t)$ are derived as follows:

$$\hat{\kappa}(t) = \frac{\|(\delta \cdot \dot{\gamma}(t)) \times (\delta \cdot \ddot{\gamma}(t))\|}{\|\delta \cdot \dot{\gamma}(t)\|^3} = \frac{1}{\delta} \cdot \kappa(t), \tag{14}$$

$$\begin{aligned} \hat{\tau}(t) &= \frac{(\delta \cdot \dot{\gamma}(t)) \times (\delta \cdot \ddot{\gamma}(t)) \cdot (\delta \cdot \dddot{\gamma}(t))}{\|(\delta \cdot \dot{\gamma}(t)) \times (\delta \cdot \ddot{\gamma}(t))\|^2} \\ &= \frac{1}{\delta} \cdot \tau(t), \end{aligned} \tag{15}$$

$$\hat{\kappa}_{\hat{s}}(t) = \frac{d(1/\delta \cdot \kappa(t))}{d(\delta \cdot s)} = \frac{1}{\delta^2} \cdot \kappa_s(t), \tag{16}$$

$$\hat{\tau}_{\hat{s}}(t) = \frac{d(1/\delta \cdot \tau(t))}{d(\delta \cdot s)} = \frac{1}{\delta^2} \cdot \tau_s(t). \tag{17}$$

The above equations reveal that the signature is relatively invariant with respect to metric actions. In trajectory recognition, the effect of scaling can be removed by normalizing the four items $\hat{\kappa}(t)$, $\hat{\tau}(t)$, $\sqrt{\hat{\kappa}_{\hat{s}}(t)}$ and $\sqrt{\hat{\tau}_{\hat{s}}(t)}$ (see Section 4.2 for details). The metric relative invariant is useful in making the recognition independent of the viewing distance of the tracked motion trajectories.

The metric invariant is illustrated in Figures 5–7 using a scaling trajectory Π. Compared with the original signature in

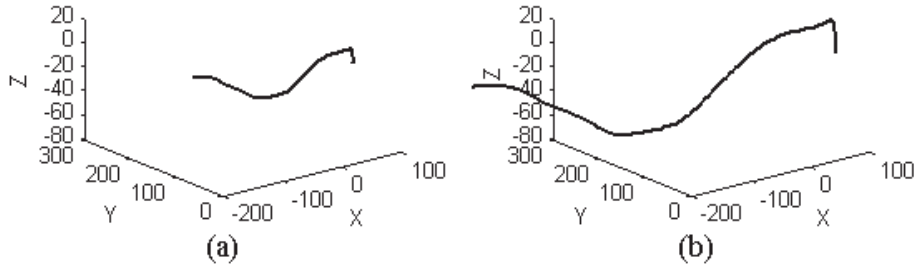

Fig. 5. Metric transformations of trajectory Π by scaling factors of (a) 0.7 and (b) 1.6.

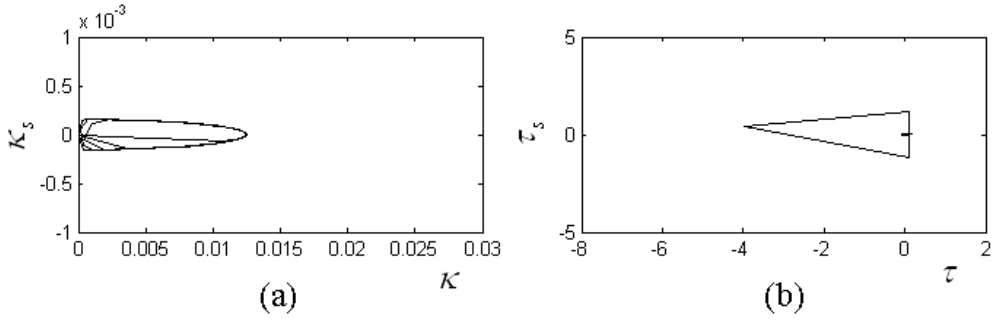

Fig. 6. The signature of the trajectory shown in Figure 5(a).

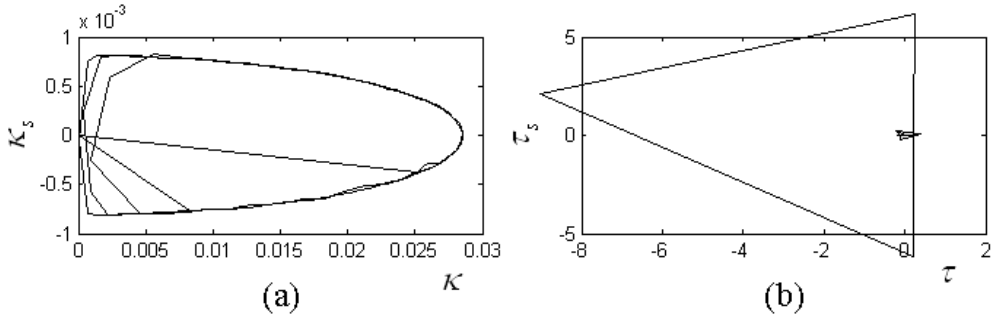

Fig. 7. The signature of the trajectory shown in Figure 5(b).

Figure 4, it is observed that the signatures of the scaled trajectories are either zoomed in or zoomed out. Actually, applying the formulae in (14)–(17) backwards, the scaled signatures in Figures 6 and 7 can be restored to be identical to the signature in Figure 4.

3.3. Visual Projective Invariant

For the visual imaging process, we need camera calibration for space data reconstruction from projected images (views). If the description is projective invariant, calibration and reconstruction can be skipped because projective invariance ensures that the projected image can be used directly. Geometric invariants have been widely explored to achieve projective invariants. Classical geometric invariants are based on relation-constrained geometric elements such as a number of interrelated points, lines and/or planes. Unfortunately, the over-constrained conditions imposed on the invariant limit its applications. They may be applied to simple and regular man-made objects, but it is hard to use them to represent complex subjects. On the contrary, the curve-oriented descriptor with pro-

jective invariant will be more generic because curve is capable of representing general shapes.

Since it has been proved that there is no real three-dimensional projective invariant from a single view (Burns et al. 1992), we explore the signature's projective invariant in two-dimensional space, which is useful for pure two-dimensional motion, when the relative depth of a motion is much smaller than the viewing distance or when the distance between the camera projection centers is negligible compared with the viewing distance (in the case of camera motion). In these cases, motion projection can be described by the weak-perspective camera model and, consequently, it is possible to deduce the signature's two-dimensional projective invariant.

Assume that the projected image of space two-dimensional motion trajectory $\Gamma_2(t)$ is represented by $\gamma_2(t) = \{x(t), y(t)\}$, the weak-perspective model gives the following relations:

$$x(t) = (l \cdot f)\frac{X(t)}{\bar{Z}} + x_0, \quad y(t) = (l \cdot f)\frac{Y(t)}{\bar{Z}} + y_0, \quad (18)$$

where $\bar{Z}$ is the average depth of the trajectory, $f \cdot l$ is the focal parameter and $[x_0, y_0]$ is the image center. It is observed that the projection from $\Gamma_2(t)$ to $\gamma_2(t)$ is just a metric action ($(l \cdot f)/\bar{Z}$ is a constant). Therefore, according to the above-deduced metric invariant, we can infer that the relations between the signature $\hat{S}_2$ of $\gamma_2(t)$ and the signature S_2 of $\Gamma_2(t)$ satisfy equations (14)–(17). Hence, projective invariance can be claimed for the two-dimensional trajectory signature, which enables the space trajectory $\Gamma_2(t)$ to be recognized invariantly by using the image signature $\hat{S}_2$.

3.4. Viewpoint Invariant

For the visual projection, the same trajectory view may result from different space motions, and the same motion may result in different projected trajectory views, both of which depend on the viewpoint (camera pose) through which the motion is seen. An effective trajectory descriptor should be able to build a consistent description for all of the different views of the same motion. Since it is not efficient to register all of the possible views, the viewpoint invariant is important.

3.4.1. Discussion

Viewpoint invariant was studied mostly in two ways. The first is to derive the invariant based on projective invariant, and the second is a so-called canonical view method (Rao and Shah 2001). While the first relies on the projective invariant, the second merely targets the projected two-dimensional views.

For the first method, the recognition is about the space data of a subject (motion or object), and the description of the subject is supposed to remain invariant in the case of viewpoint (camera pose) changes. Hence, making use of the projective invariant features, we can prove that the descriptions based on the projected views are invariant with respect to viewpoint changes. Relying on projective invariance, this method is a little restrictive for two reasons: (1) since there is no general three-dimensional to two-dimensional projective invariant, we have to use a weak-perspective model or focus on two-dimensional space subjects (Parameswaran and Chellappa 2006); (2) it is not always easy to find an effective projective invariant for complex subjects.

For the second method, however, the operational entity for labeling and recognition is the projected two-dimensional views rather than the (reconstructed) space data. This occurs often in database-oriented image storage and retrieval. Focusing on the two-dimensional views, the transformation relations among different views (such as rough affine transformation (Rao and Shah 2001), homography (Malis et al. 2003) and fundamental matrix (Caspi et al. 2006)) are studied for inter-view matching. This can give rise to a viewpoint (or view) invariant by measuring the matching error of the transformation matrix between a canonical view and a transformed view. This method is able to take into account the complicated projective models such as the perspective (in essence, it does not care about the type of projective model). However, as no projective invariant is required to ensure unique correspondence between a projected view and a space subject, the recognition may not be the same as the case for recognizing space subjects.

In the following, we study the viewpoint invariant for both two- and three-dimensional signatures. In particular, the three-dimensional viewpoint invariant is developed based on stereo views.

3.4.2. Two-dimensional Viewpoint Invariant

Extending the two-dimensional projective invariant in Section 3.3, we can prove the signature's two-dimensional viewpoint invariant. The viewpoint change results from a rigid action applied to the space trajectory or the observing cameras. Assume that at viewpoint ϕ, the space and image signatures of trajectory $\Gamma_2^{\phi}(t)$ are denoted by S_2^{ϕ} and $\hat{S}_2^{\phi}$, respectively. By applying a rigid action M_{ϕ}^{φ} to the trajectory, we obtain a new trajectory variation $\Gamma_2^{\varphi}(t) = M_{\phi}^{\varphi} \cdot \Gamma_2^{\phi}(t)$, and the viewpoint is now changed to φ, where we denote the space and image signatures by S_2^{φ} and $\hat{S}_2^{\varphi}$, respectively. From the rigid invariant and considering that M_{ϕ}^{φ} is a rigid action, we have

$$S_2^{\phi} = S_2^{\varphi}. \quad (19)$$

Meanwhile, according to the projective invariant, we have the following relations ($\Delta = (l \cdot f)/\bar{Z}$):

$$\hat{S}_2^{\phi} = \frac{1}{\Delta} \cdot S_2^{\phi}, \quad \hat{S}_2^{\varphi} = \frac{1}{\Delta} \cdot S_2^{\varphi}. \quad (20)$$

Combining equations (19) and (20), we have

$$\hat{S}_2^{\phi} = \hat{S}_2^{\varphi}. \quad (21)$$

This relation shows that the 2-D signature remains invariant when viewpoint (camera pose) changes. Although it is based on the weak-perspective model, the viewpoint invariant is still advantageous because the signature is capable of describing complex and irregular trajectories.

3.4.3. Three-dimensional Viewpoint Invariant

Here, a stereo vision based viewpoint invariant is developed for the three-dimensional signature. Since it is impractical to calculate signature directly from the stereo image pairs, we have to reconstruct the space trajectory from two-dimensional stereo images to calculate three-dimensional signature. For a motion Ω, assuming $\Gamma_\phi(t)$ is its space trajectory reconstructed from viewpoint ϕ, and $\Gamma_\varphi(t)$ another trajectory counterpart reconstructed from a different viewpoint φ, we have the following relations:

$$\Gamma_\phi(t) = P_\phi^{-1}(\Omega), \quad \Gamma_\varphi(t) = P_\varphi^{-1}(R(\Omega) + T(\Omega)), \tag{22}$$

where P denotes the projective process, R a rotation action and T a translation action. This equation indicates that the viewpoint change is equivalent to a corresponding action $(R(\Omega) + T(\Omega))$ applied to the space trajectory. As $(R(\Omega) + T(\Omega))$ is of the Euclidean group, we can claim that the signatures of $\Gamma_\phi(t)$ and $\Gamma_\varphi(t)$ are identical according to the Euclidean invariant. Note that this three-dimensional viewpoint invariant is insensitive to the projective transformation P. Therefore, it is generic for projective models.

3.5. Speed Invariant

It is observed that velocity, speed and acceleration are constitutive factors of the signature components as formulated in Section 2.1. The factors $\dot{\Gamma}(t)$ and $\ddot{\Gamma}(t)$ are actually the velocity and acceleration features. This implies that these factors may influence the signature in certain ways. Hence, the signature's variability is analyzed in the following with respect to the changes in speed and velocity, respectively.

3.5.1. Speed Invariant

First, assume that the motion is not in a straight line and the whole trajectory shape is unchanged. In other words, no matter how the speed profile changes, the motion directions at all trajectory points are fixed.

For a piece of motion trajectory $\gamma(t)$ with fixed arc-length, at a point P_i, assume that its original speed is $\dot{\gamma}(t)$, if its new speed differs from the original by a scaling factor δ, then the new speed at P_i will be in the form of $\dot{\hat{\gamma}}(\hat{t}) = \delta \cdot \dot{\gamma}(t)$. As speed change will lead to corresponding change in motion time, say, $t = \delta \cdot \hat{t}$, we have $\dot{\hat{\gamma}}(\hat{t}) = \delta \cdot \dot{\gamma}(\delta \cdot \hat{t})$. Thus, we can derive the following equations:

$$\ddot{\hat{\gamma}} \cdot (\hat{t}) = \delta^2 \cdot \ddot{\gamma}(\delta \cdot \hat{t}), \quad \dddot{\hat{\gamma}} \cdot (\hat{t}) = \delta^3 \cdot \dddot{\gamma}(\delta \cdot \hat{t}). \tag{23}$$

Meanwhile, according to the arc-length formula $s(t) = \int_0^t \left\|\dot{\Gamma}(t)\right\| dt$, the passing arc-length up to point P_i is invariant. Referring to equations (3)–(6), we obtain the following relations regarding the signature quaternion at point P_i:

$$\begin{aligned}
\hat{\kappa}(\hat{t}) &= \frac{\left\|(\delta \cdot \dot{\gamma}(\delta \cdot \hat{t})) \times (\delta^2 \cdot \ddot{\gamma}(\delta \cdot \hat{t}))\right\|}{\left\|\delta \cdot \dot{\gamma}(\delta \cdot \hat{t})\right\|^3} \\
&= \kappa(\delta \cdot \hat{t}) = \kappa(t), && (24)\\
\hat{\tau}(\hat{t}) &= \tau(\delta \cdot \hat{t}) = \tau(t), && (25)\\
\hat{\kappa}_{\hat{s}}(\hat{t}) &= \kappa_s(\delta \cdot \hat{t}) = \kappa_s(t), && (26)\\
\hat{\tau}_{\hat{s}}(\hat{t}) &= \tau_s(\delta \cdot \hat{t}) = \tau_s(t). && (27)
\end{aligned}$$

The above equations show that speed change from $\dot{\gamma}(t)$ to $\dot{\hat{\gamma}}(\hat{t}) = \delta \cdot \dot{\gamma}(\delta \cdot \hat{t})$ only results in the temporal traversing moment of point P_i changing from t to t/δ. The signature quaternion remains unchanged at point P_i. This gives rise to the signature's speed invariant: the signature quaternion at a trajectory point has nothing to do with the traversing speed at that point.

However, at the same time we should notice that speed change causes simultaneous changes in trajectory sampling. When the sampling rate of a tracking system is fixed, the number of sampled points depends on the motion speed. A higher speed leads to sparser data sampled and a lower speed results in denser data. Figure 8 illustrates the trajectory sampling for different speeds with a fixed sampling rate. As a result, the signatures of two trajectories with the same shape but different speed profiles are not completely invariant (identical). They are actually partially identical but differ in signature length (see Section 4.2 for how they are matched for recognition). Therefore, the speed invariant is not an absolute invariant for the entire signature, but it is the basis to enable the same shape of trajectories with different speed profiles to be recognized as the same. This can deal with a typical case when the same motion is demonstrated in multiple demonstrations with various speed profiles.

3.5.2. Signature Variability with Respect to Velocity Change

Following the speed invariant, in a more general sense, here we explore the effect to the signature components caused by arbitrary changes in motion velocity or acceleration in direction and magnitude. This exploration can help to clarify the role of velocity or acceleration played in the signature. Similar problems have been studied in Rao et al. (2002) for a specific

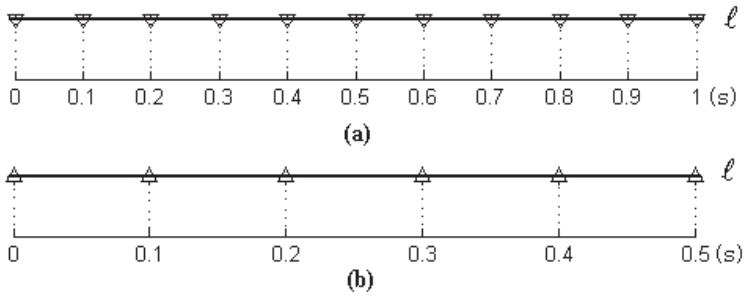

Fig. 8. The effect of trajectory sampling caused by speed change: (a) speed v; (b) speed $2 \cdot v$.

one-dimensional trajectory, but we will show a more general result.

Analyze this problem by evaluating the changing trend of the signature quaternion between two consecutive points P_1 and P_2. Assume that their velocity vectors are v_1 and v_2, respectively, and $v_1 \neq \vec{0}$, then the acceleration at P_2 is $a_2 = v_2 - v_1$. According to equations (3) and (4), the curvature $\kappa(P_2)$ and torsion $\tau(P_2)$ at P_2 can be formulated in terms of v_1 and v_2 as follows:

$$\kappa(P_2) = \frac{\|v_2 \times a_2\|}{\|v_2\|^3} = \frac{\|v_2 \times (v_2 - v_1)\|}{\|v_2\|^3}, \quad (28)$$

$$\tau(P_2) = \frac{(v_2 \times a_2) \cdot a_2'}{\|v_2 \times a_2\|^2} = \frac{(v_2 \times (v_2 - v_1)) \cdot (v_2 - v_1)'}{\|v_2 \times (v_2 - v_1)\|^2}. \quad (29)$$

Let us see several special cases first. If $\|v_2\| = 0$, then $\kappa(P_2)$ and $\tau(P_2)$ have no meaning because the denominators of equations (28) and (29) are zero. As mentioned in Section 2.1, zero speed violates the regularity constraint of the trajectory that is parameterized. In the case $a_2 = \vec{0}$, say, $v_2 = v_1$, $\kappa(P_2) = 0$ but $\tau(P_2)$ has no meaning. This case can be further generalized. Assume that $v_2 \neq v_1$ and v_2 can be represented by a linear transformation of v_1, that is, $v_2 = \delta \cdot v_1$, where δ ($\delta \neq 1$) is a real number. Then $a_2 = v_2 - v_1 = (\delta - 1) \cdot v_1$, which leads to $v_2 \times a_2 = \delta \cdot v_1 \times (\delta - 1) \cdot v_1 = \vec{0}$. So $\kappa(P_2) = 0$ too and $\tau(P_2)$ has no meaning either. This reveals that as long as the velocity's direction remains unchanged from P_1 to P_2, no matter how speed changes, we always have $\kappa(P_2) = 0$. This actually is the special case that the trajectory between P_1 and P_2 is a straight line.

More generally, in the case when v_2 differs from v_1 in both motion direction and magnitude, equations (28) and (29) imply that the variability of curvature or torsion is not monotonic with respect to the velocity. Similar conclusions can be drawn for acceleration. This non-monotonic claim can be verified by some exemplary data. Assume that a three-dimensional motion trajectory γ that consists of nine points ($i \in [1, 9]$), as shown in Figure 9(a). Applying equations $v_i = \gamma_i - \gamma_{i-1}$, $a_i = v_i - v_{i-1}$ and $a_i' = a_i - a_{i-1}$, we can obtain the velocity v, acceleration a and the derivative of acceleration a' as $\{[v_i, a_i, a_i']|i \in [3, 8]\}$. Based on these vectors, the corresponding curvatures and torsions are calculated according to equations (3) and (4). Figure 9(b) shows the finally obtained four profiles of curvature, torsion, magnitudes of velocity and acceleration. It is seen that no monotonic relation can be traced.

Usually, the salient features such as speed or the curvature profile are used for motion segmentation and analysis. The above non-monotonic conclusion indicates that between the curvature (torsion) and velocity (acceleration) profiles, no one is replaceable and no one is able to replace another to cover the relevant three-dimensional motion features.

3.6. Occlusion Invariant

Occlusion is a common phenomenon in motion trajectory representation. Occlusion cannot be dealt with by global descriptors such as moments. In contrast, the proposed signature descriptor is capable of handling occlusion naturally (Bruckstein et al. 1993). Attributed to the computational locality, the partial signature of the non-occluded portion of the full trajectory can still be generated invariantly. This is defined as the occlusion invariant and it enables a trajectory to be recognized under partial occlusion. Note that this invariant imposes an extent constraint on occlusion. Heavy occlusion will result in recognition ambiguity for similar trajectories.

The occlusion invariant is demonstrated in Figures 10–12 using trajectory Π. Referring to the full signature of Π in Figure 4, we find that the signatures (Figures 11 and 12) of the occluded trajectories can be completely localized in shape within the full signature. In fact, occlusion will only make the signature shorter than the original in length. Hence, the occluded trajectory can be recognized from the partially invariant matching of the trajectory signatures.

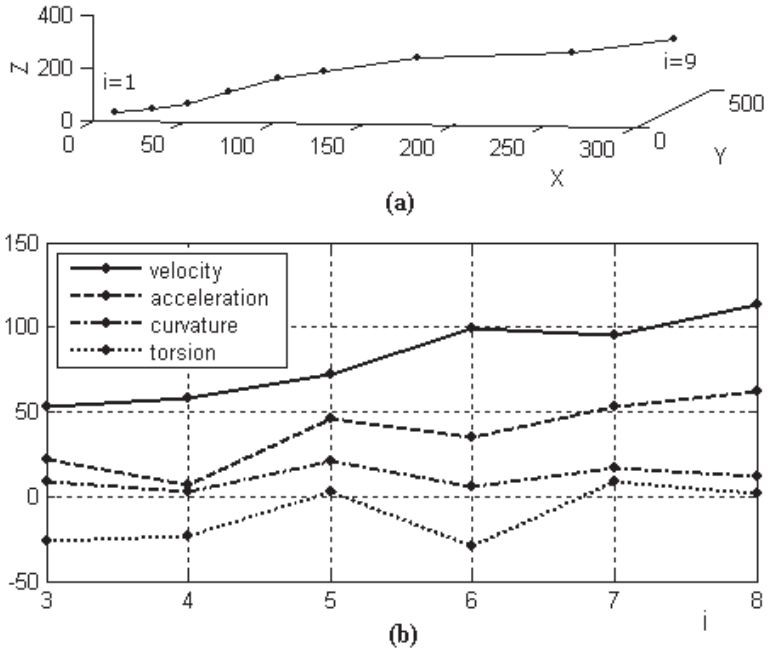

Fig. 9. Signature variability with respect to velocity and acceleration changes. (a) Trajectory γ consisting of nine points. (b) The changing trend of curvature (torsion) with respect to velocity (acceleration) changes (note that the curvatures and torsions are scaled by 3000 for a more explicit display).

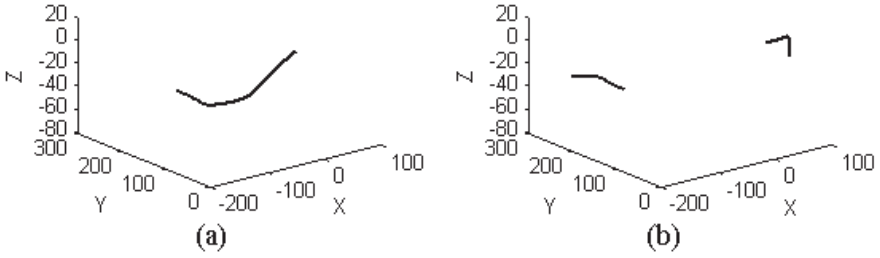

Fig. 10. Two occluded cases of trajectory Π: (a) frames 33–120 visible; (b) frames 1–26 plus frames 124–183 visible.

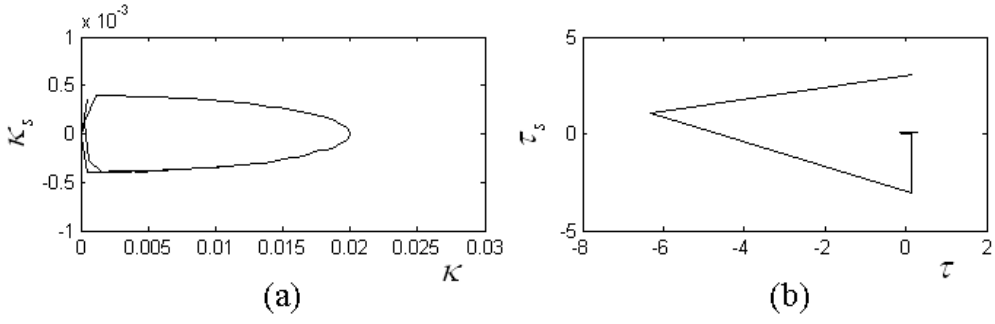

Fig. 11. The signature of the trajectory shown in Figure 10(a).

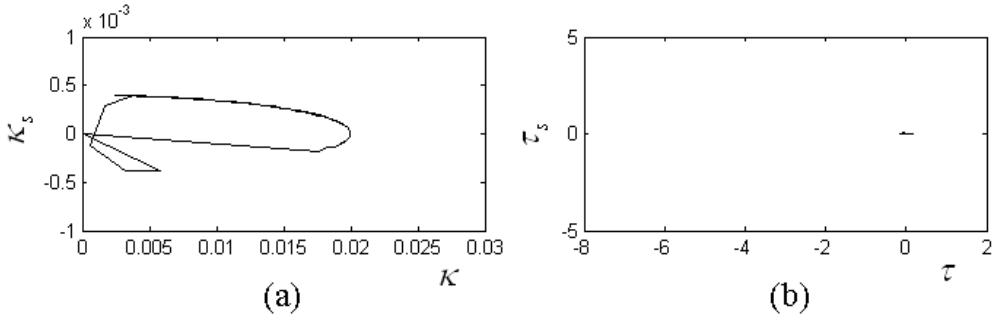

Fig. 12. The signature of the trajectory shown in Figure 10(b).

3.7. Invariants from the Approximate Signature

In addition to the accurate signature, the above invariants can also be verified from the approximate signature. This consistence is important because it is the approximate signature that is used in practice.

As can be seen from equations (7), (8), (10) and (11), only the quantities a, b, c, d, e, f, g, m, n, h, $\bar{s}$, H^+ and H^- are the possible factors influencing the invariants. In fact, a, b, c, d, e, f, g, m and n are Euclidean joint invariants, which are essentially invariant with respect to Euclidean actions. As a result, h, $\bar{s}$, H^+ and H^- are also invariant since they are completely determined by a, b, c, d, e, f, g, m and n. Therefore, it can be concluded that the approximate signature is Euclidean invariant.

The scaling relations defined by equations (14)–(17) also hold for the approximate signature. Assume that a trajectory is scaled by factor δ, which means that the Euclidean distances a, b, c, d, e, f, g, m and n are also scaled by δ. Hence, it can be further derived that h and $\bar{s}$ are scaled by δ, and H^+ and H^- are scaled by δ^2. Substituting the relevant scaled quantities into equations (7), (8), (10) and (11), the scaling relations are obtained in the same manner as those defined in equations (14)–(17). Hence, the approximate signature is proved to be relatively invariant with respect to metric transformations.

Following the above claims of rigid and metric invariants and referring to Sections 3.3 and 3.4, we can obtain the two-dimensional projective invariant and viewpoint invariant for the approximate signature.

For the speed invariant, the same argument also holds for the approximate signature because no matter what the motion speed is, the final lengths of a, b, c, d, e, f, g, m, n, h, $\bar{s}$, H^+ and H^- are unchanged/fixed.

As mentioned before, as the approximate signature is also based on local features, it also admits occlusion invariance.

4. Signature Recognition Metric

With the signature based descriptions, the trajectory's recognition is switched to the similarity measurement of the corresponding signatures. To do this, a suitable recognition metric is required. As the signature is based on the features extracted from all of the trajectory points, we have to consider several factors related to the signature matching.

4.1. Factors Influencing Signature Matching

Signature matching has to take two important factors into account, that is, trajectory length and point distribution along the trajectory. Motion speed, sampling rate, occlusion and the user are possible factors that could cause the trajectories to have diverse lengths and arbitrary distributions of the trajectory points. As mentioned in Section 3.5.1, with a fixed sampling rate, a lower speed results in denser points and vice versa. The sampling rate may change proportionally or randomly in trajectory tracking. Occlusions make a trajectory shorter in length. The user is also a source of inconsistencies in the trajectory length and point distribution among multiple instances as a user cannot perform a motion in exactly the same way each time.

Six trajectory instances of a motion class are shown in Figure 13. It is observed that straightforward pointwise signature matching is infeasible. Hence, we need to design a suitable metric to measure the similarity among signatures.

4.2. Dynamic Time Warping Based Signature Recognition

To take the diversity in trajectory length and point distribution into account, we develop a non-linear inter-signature matching method aiming at achieving the most reasonable alignment of two signatures to obtain a similarity measurement.

The signature matching problem is actually analogous to elastic time series/sequences comparison (Vlachos et al. 2005). In our view, the precondition of appropriate similarity measurement lies in finding the best match of the element pairs along two signatures. Following this principle, the dynamic time warping (DTW) method is employed to calculate the inter-signature distance.

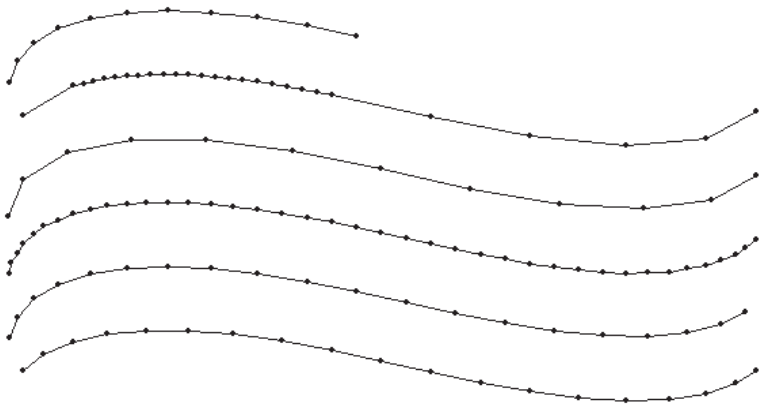

Fig. 13. Illustration of the trajectory sampling under various conditions.

Dynamic programming is used in DTW algorithm to calculate the DTW distance, i.e. the minimum cost of the best alignment between two signatures. For two signatures S^{*i} and S^{*j} with respective length P and Q, let

$$S^{*i} = \{[\kappa^{*i}, \kappa_s^{*i}, \tau^{*i}, \tau_s^{*i}]_p \mid p \in [1, P]\}$$

and

$$S^{*j} = \{[\kappa^{*j}, \kappa_s^{*j}, \tau^{*j}, \tau_s^{*j}]_q \mid q \in [1, Q]\}$$

represent the two sequences of signature quaternions. The cost function $d(p, q)$ reflecting the similarity between S^{*ip} and S^{*jq} is defined based on the Euclidean distance, and is formulated as follows:

$$d(p,q) = \Delta S^{*ip,jq} = \frac{\Delta\kappa^{*ip,jq} \cdot \Delta\tau^{*ip,jq}}{\sqrt{(S^{*ip})^2} \cdot \sqrt{(S^{*jq})^2}}, \tag{30}$$

where (defining the operator $\sqrt[+]{x} = |x| / x \cdot \sqrt{|x|}$)

$$\Delta\kappa^{*ip,jq} = \left\| \left(\kappa^{*ip}, \sqrt[+]{\kappa_s^{*ip}}\right) - \left(\kappa^{*jq}, \sqrt[+]{\kappa_s^{*jq}}\right) \right\|, \tag{31}$$

$$\Delta\tau^{*ip,jq} = \left\| \left(\tau^{*ip}, \sqrt[+]{\tau_s^{*ip}}\right) - \left(\tau^{*jq}, \sqrt[+]{\tau_s^{*jq}}\right) \right\|, \tag{32}$$

$$(S^{*ip})^2 = (\kappa^{*ip})^2 + \left(\sqrt[+]{\kappa_s^{*ip}}\right)^2 + (\tau^{*ip})^2 + \left(\sqrt[+]{\tau_s^{*ip}}\right)^2, \tag{33}$$

$$(S^{*jq})^2 = (\kappa^{*jq})^2 + \left(\sqrt[+]{\kappa_s^{*jq}}\right)^2 + (\tau^{*jq})^2 + \left(\sqrt[+]{\tau_s^{*jq}}\right)^2. \tag{34}$$

The accumulative minimum cost of aligning up to S^{*ip} and S^{*jq} is represented by $u(p, q)$, which is determined by the minimum cost among its three neighbors' plus the cost of itself, as expressed by

$$u(p,q) = \min(u(p-1,q-1), u(p,q-1), u(p-1,q)) + d(p,q). \tag{35}$$

Following the above and working from $u(1, 1)$ to $u(P, Q)$, the best alignment of the two signatures is found giving rise to the DTW distance at $u(P, Q)$, which can serve as the similarity measurement between S^{*i} and S^{*j}. It should be noted that equation (30) is a normalized measurement transparent to scaling actions (see Section 3.2). Therefore, the DTW matching algorithm can accommodate metric transformation in trajectory recognition.

Applying the above-formulated DTW algorithm, we obtain the non-linear paths warping among the six trajectory instances given in Figure 13, with the result shown in Figure 14. It is observed that the trajectory alignments are quite reasonable. In addition, we can find that the warping paths offer an intuitive interface to perceive the consistency and difference between two trajectories. For example, the one-to-many point correspondence reflects the difference in speed profiles between two trajectories.

In general, while the DTW distance can give a quantitative measurement for the overall similarity of two signatures, motion features can be perceived qualitatively by visualizing the inter-trajectory warping paths. In this sense, the DTW based inter-signature matching method generally behaves better than the matching method proposed by Gueziec and Ayache (1994) which is applicable to still curve matching. At the same time, note that as DTW always matches all of the points along trajectories, the DTW based recognition may become ambiguous when occlusion is severe. A matching method called Minimal Variance Matching (MVM) proposed in Latecki et al. (2005)

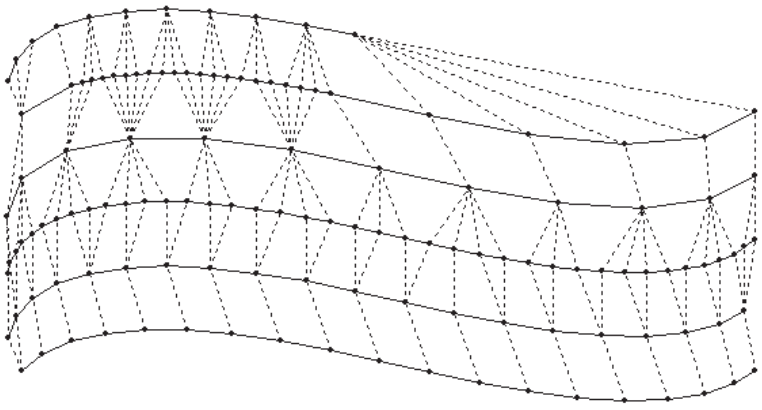

Fig. 14. DTW based non-linear paths warping (in dotted lines) among the six trajectory instances shown in Figure 13.

can offer better recognition specifically for occluded trajectories. However, the MVM method imposes a length constraint to the query and the reference. Hence, its application scope is narrower than the DTW method.

5. Experiments

Experiments are conducted to implement the trajectory signature, confirm the signature invariants and show the signature's recognition performance. Here we use sign language, which is characterized by the extracted motion trajectories. Sign language is a special human communication modality, and it can be introduced as an effective method for human–robot interaction.

5.1. Sign Motion Trajectory Acquisition

In our experiments, when an operator performs a sign word by hand, the hand movement is tracked by visual sensors to acquire the spatiotemporal motion trajectory to represent the sign word. In the experiments, only the sign language performed by a single human hand is used and each sign admits a regular trajectory (for all t, $\dot{\Gamma}(t) \neq \vec{0}$). A stereo vision system with two TM-765 cameras is employed for the three-dimensional trajectory tracking. It should be pointed out that to simplify the trajectory tracking, we do not track the hand directly but track a distinctive rigid object held by the hand. This makes trajectory tracking easier and more stable. The issue of pursuing robust human hand tracking is out of the scope of this paper. The object of interest is tracked using the CAMShift (Continuously Adaptable Mean-Shift) algorithm (Bradski 1998) under common scene background. The sampling rate is 15 fps with a resolution of 480×480. Figure 15 shows the system setup and several stereo snapshots in the trajectory tracking.

The two-dimensional trajectory is the concatenation of the position vectors of the tracked object centroid, and the three-dimensional trajectory is reconstructed using the stereo triangular algorithm. Based on the two-dimensional object centroids (x_1, y_1) and (x_2, y_2) in the stereo views, the space coordinates (X, Y, Z) are calculated by

$$\begin{cases} X = T_b \cdot (x_1 - x_0)/(x_1 - x_2), \\ Y = T_b \cdot (y_1 - y_0)/(x_1 - x_2), \\ Z = T_b \cdot (l \cdot f)/(x_1 - x_2). \end{cases} \tag{36}$$

In our experiments the baseline length T_b is fixed. Hence, following the signature's metric invariant, T_b can be eliminated from equation (36) without influencing the trajectory recognition. The image center (x_0, y_0) and focal parameter $l \cdot f$ are calibrated using the method of Zhang (2000).

5.2. Invariant Signature Recognition

The first experiment aims at verifying the invariants in the sign signature description and recognition. We define 10 different sign classes characterized by their motion trajectories whose shapes are close to the characters from "0" to "9". These classes are labeled from (00) to (09) sequentially. Note that we are not performing character recognition but illustrating the free forms of sign trajectory. Firstly, five series of samples of the 10 classes are demonstrated by user 1 with normal viewpoints, viewing distances and motion speed profiles. Among the five samples of each class, we manually pick a representative and stable one as the canonical sample for this class. Figure 16 shows the selected canonical samples for the 10 classes.

Next, four series of test samples were acquired by repeating the 10 sign classes by varying the conditions in viewpoints (Figure 17), viewing distances (Figure 18), speed profiles (Figure 19) and occlusions (Figure 20). The viewpoint is changed

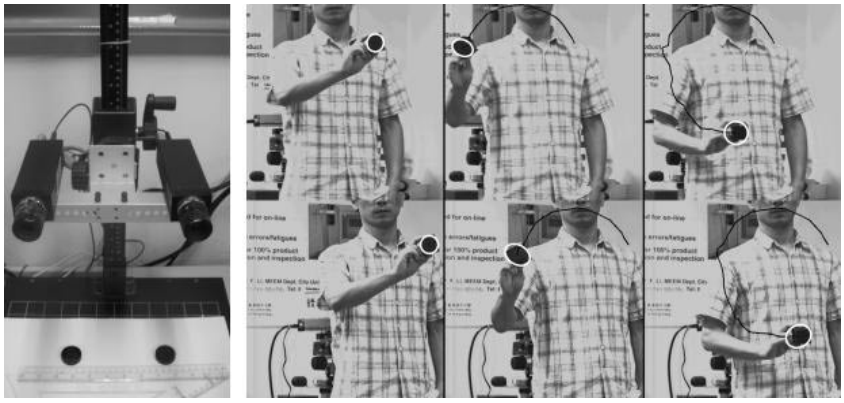

Fig. 15. Vision system setup and snapshots of the stereo trajectory tracking.

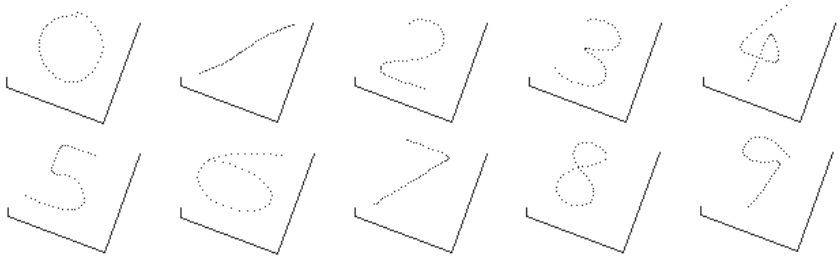

Fig. 16. The canonical samples of the 10 sign classes.

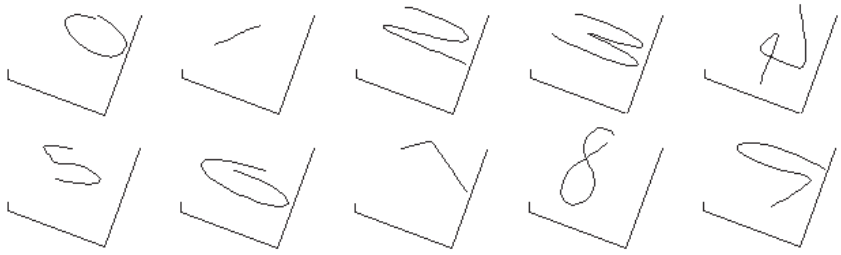

Fig. 17. The sign samples demonstrated with diverse viewpoints (indexed by series A).

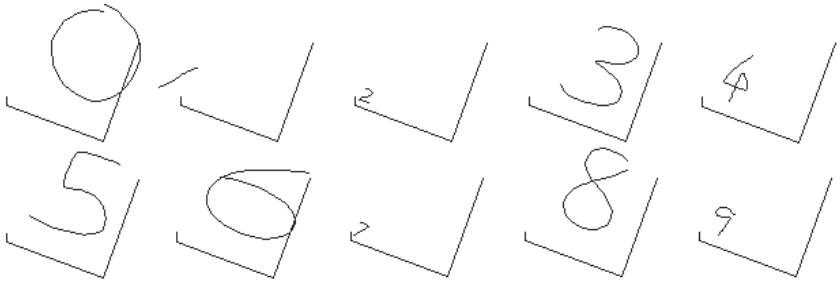

Fig. 18. The sign samples demonstrated with diverse viewing distances (indexed by series B).

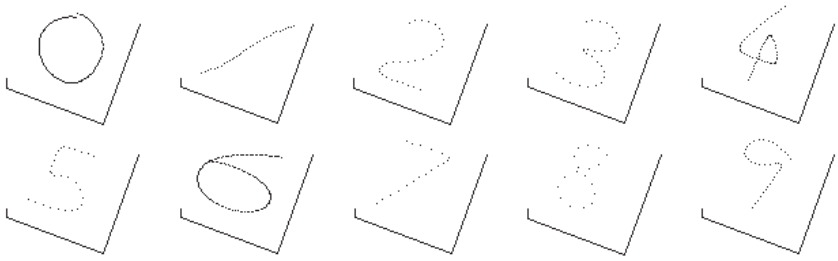

Fig. 19. The sign samples demonstrated with diverse motion speed profiles (indexed by series C).

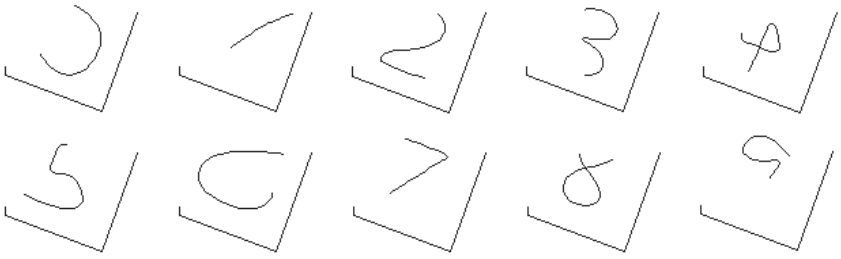

Fig. 20. The sign samples with diverse extents of occlusions (indexed by series D).

by changing the camera pose relative to the sign demonstration. Note that the viewpoint invariant can confirm the rigid invariant at the same time. The viewing distance is changed by varying the distance between the camera and the demonstrator. The motion speed changes result from faster or slower movements of the user's hand in the course of sign demonstration. Because humans cannot perform exactly the same motion each time, the repeated signs are not exactly the same shape as the canonical samples. However, user 1 was asked to try his best to repeat the signs exactly. In particular, the samples in Figures 16 and 19 are displayed as dotted lines to enable the difference in point distribution caused by speed changes to be observed.

Now look at the signature's variability with respect to the four series of test samples. According to the signature's invariants proved in Section 3, the signatures of the test samples should be the same as those of the canonical samples. The DTW based inter-signature matching is applied to measure the similarity between the test samples and the canonical samples.

Figure 21 shows the comparison results of series A and B in two tables. For each test sample (row), its similarity to every canonical sample (column) is normalized to a relative distance range of [0, 1]. The inter-signature distances of the series samples are then scaled with bounds [0, 0.2] to a grayscale image to form a similarity table. Here, the darker the cell is, the smaller the distance (denoted by α) between the corresponding test and canonical sample (black $\alpha = 0$; white $\alpha \geq 0.2$; gray $\alpha \in (0, 0.2)$). The two tables reveal that the test samples' signatures are very close to those of the canonical samples.

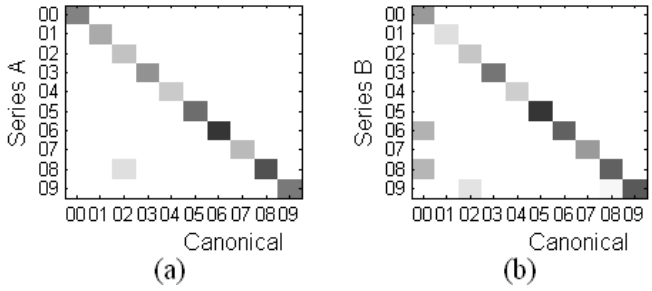

Fig. 21. The signatures' similarity of (a) series A (different viewpoints) and (b) series B (different viewing distances) compared with the canonical samples.

To demonstrate the viewpoint and metric invariant more clearly, we manually synthesize a series of test samples (indexed by A*) as shown in Figure 22, which are the counterparts of the canonical samples after applying arbitrary rigid transformations. This kind of data synthesis can simulate simultaneous tracking of the same sign trajectory by two or more sets of visual sensors at different viewpoints. In addition, some synthetic databases also contain many such rigid transformed trajectory variations. Hence, the data synthesis of series A* is meaningful for practical applications and the experiment result will be representative. The signatures of series A* are compared in Figure 23 with scaling bounds [0, 0.0001]. We can observe that the cells at the table diagonal are almost black,

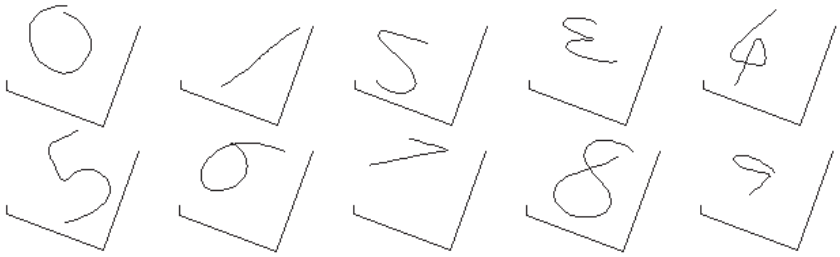

Fig. 22. The sign samples synthesized manually by applying arbitrary rigid actions (indexed by series A*).

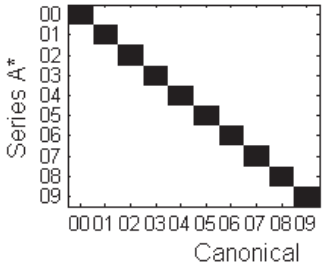

Fig. 23. The signatures' similarity of series A* (arbitrary rigid actions applied) compared with the canonical samples.

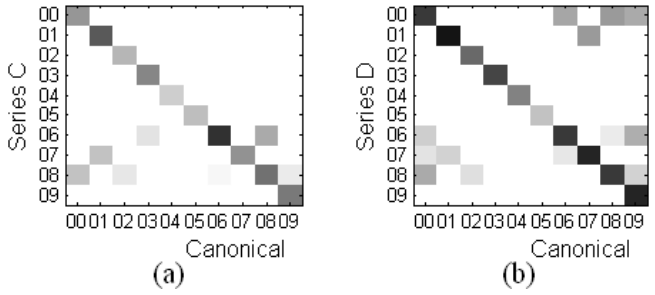

Fig. 24. The signatures' similarity of (a) series C (different speed profiles) and (b) series D (diverse occlusions) compared with the canonical samples.

which indicates extremely small values because the upper scaling bound is $0.0001 << 1$. We believe that the signature deviations just come from digital computing errors. That is, if the same sign word is tracked with different viewpoints (camera poses), the signature will be identical (ignoring the digital computing errors). Hence, the viewpoint (rigid) invariant is verified using both real and synthetic data. Likewise, the metric invariant can also be explicitly testified by synthetic data.

Figure 24 shows the similarity of series C and D with the canonical samples (scaling bounds [0, 0.2]). We observe that the results in Figure 24 are a little different to those of Figure 21. Since the speed invariant and occlusion invariant are just the invariants of partial signature data rather than the entire signature, the values of the diagonal cells are a little uneven. However, the inter-signature distances are distinctive enough to make the test samples recognizable. The invariants with respect to diverse motion speeds and acceptable occlusions can also be seen.

Note that heavy occlusions may ruin the occlusion invariant as can be seen from Table 1, which shows more about the recognition with respect to various extents of occlusions. The inputs here are the occluded variations of the canonical samples by making adjustable degree of occlusion in sign's moving direction. The results reveal that as the occlusion increases, the recognition becomes ambiguous. Therefore invariant recognition necessitates some limit on the occlusion degree.

From the above experimental results, the signature's invariants can be verified. Taking advantage of the invariants, the sign demonstrator can perform the signs freely without being limited by the conditions on the viewpoint (camera pose), viewing distance, motion speed or occlusion within a reasonable degree.

5.3. Small-scale Trajectory Recognition Test

Two kinds of query samples are customized to evaluate the recognition with respect to diverse trajectory accuracy. The first kind of query sample is the real signs demonstrated by different users. The second is synthetic data simulating the trajectories with increasing shape variance.

First, some real sign trajectory data are acquired. Three naive users (user 2, 3 and 4) were asked to perform the 10 sign classes individually in front of the camera. No training was conducted for these three users and they just demonstrate the signs in their own way. The users are allowed to select various viewpoints, viewing distances and motion speeds. Figure 25 shows the trajectories of a set of sign samples produced by the three users.

The second kind of query sample is synthesized based on the canonical samples by attaching white noise with increasing strengths. As emphasized in Section 2.3, noisy data will be

Table 1. Recognition results with respect to various extents of occlusions.

	Occlusion extent in normal (+) or reverse (–) direction							
Samples	+10%	+30%	+50%	+80%	–10%	–30%	–50%	–80%
"0"	"0"	"0"	"6"	"2"	"0"	"0"	"9"	"3"
"1"	"1"	"7"	"4"	"7"	"1"	"4"	"7"	"7"
"2"	"2"	"2"	"4"	"9"	"2"	"2"	"3"	"5"
"3"	"3"	"3"	"3"	"8"	"3"	"3"	"8"	"8"
"4"	"4"	"4"	"4"	"5"	"4"	"4"	"7"	"7"
"5"	"5"	"5"	"3"	"6"	"5"	"5"	"2"	"4"
"6"	"6"	"0"	"9"	"3"	"6"	"0"	"9"	"0"
"7"	"7"	"7"	"1"	"9"	"7"	"7"	"2"	"1"
"8"	"8"	"6"	"9"	"1"	"8"	"9"	"6"	"2"
"9"	"9"	"7"	"7"	"1"	"9"	"9"	"3"	"0"

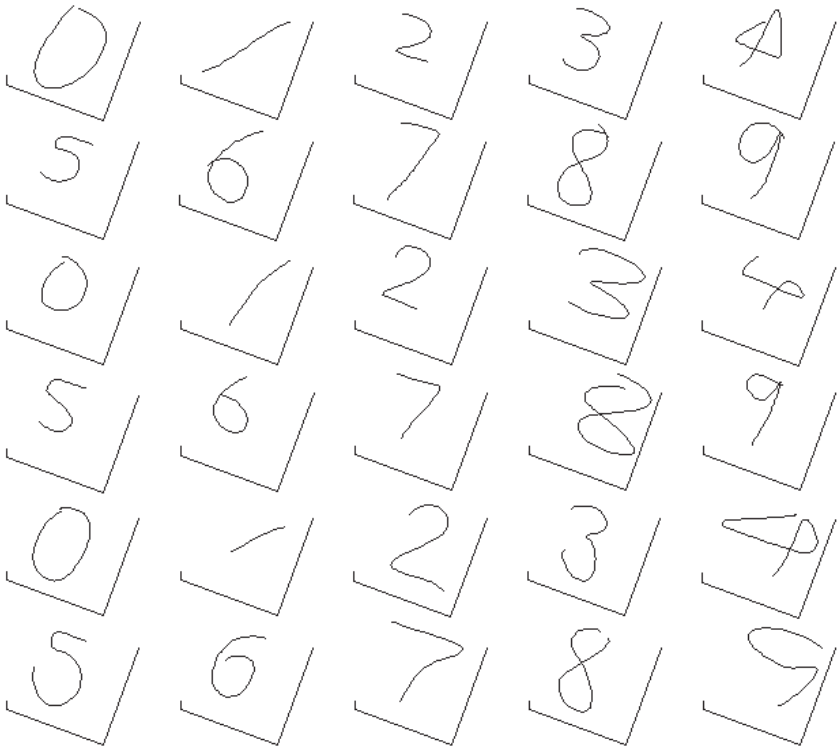

Fig. 25. Three series of query samples demonstrated by user 2 (rows 1 and 2), 3 (rows 3 and 4) and 4 (rows 5 and 6).

smoothed first. Hence, the two trajectory smoothers are used. For the three series of noisy sign trajectories shown in Figures 26, in Figures 27 and 28 we show the smoothed samples using the moving average filter (span parameter 9, 17 and 31) and wavelet smoother (wavelet DB4 and the coefficients at level 4–6), respectively. It is observed that the smoothing effects are good and the shape deviations are acceptable with the smoothing. This shows that the noisy trajectory can be handled nicely by the smoothers to improve the signature's robustness.

The second kind of samples can simulate a kind of real situation where the sign demonstration has irregular and un-

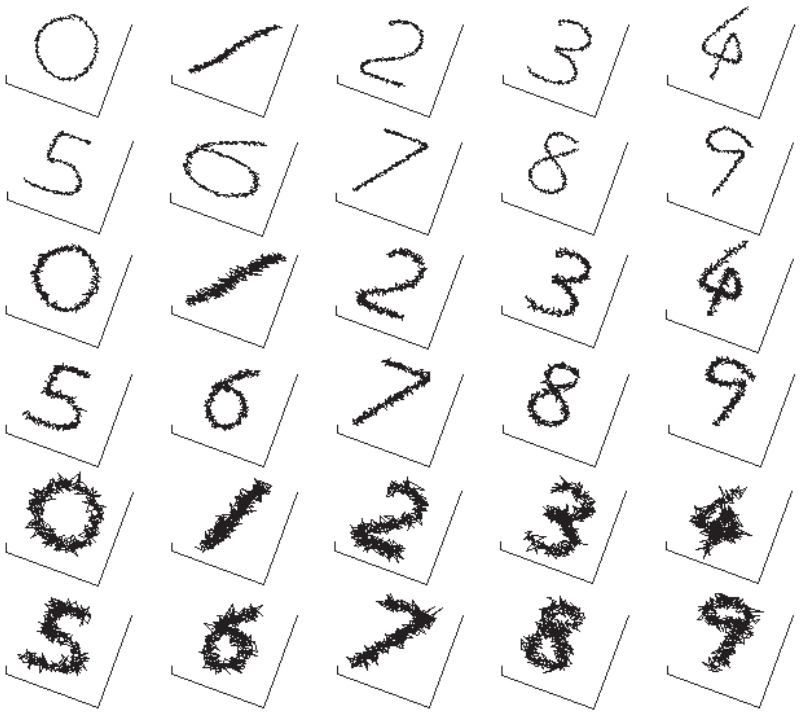

Fig. 26. The query series E (rows 1 and 2), F (rows 3 and 4) and G (rows 5 and 6) with increasing strength of white noise.

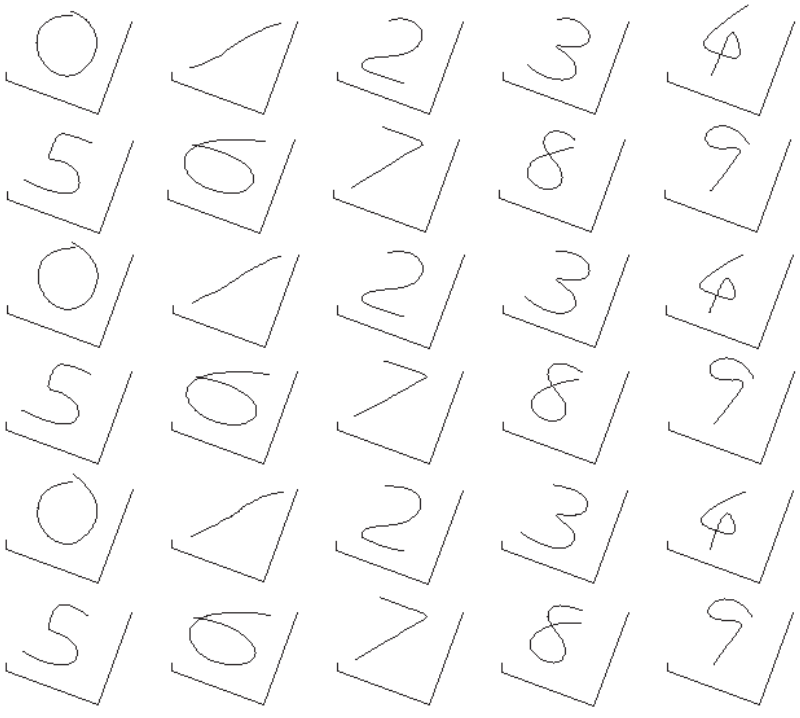

Fig. 27. The smoothed samples of the query series E, F and G by using the moving average filter.

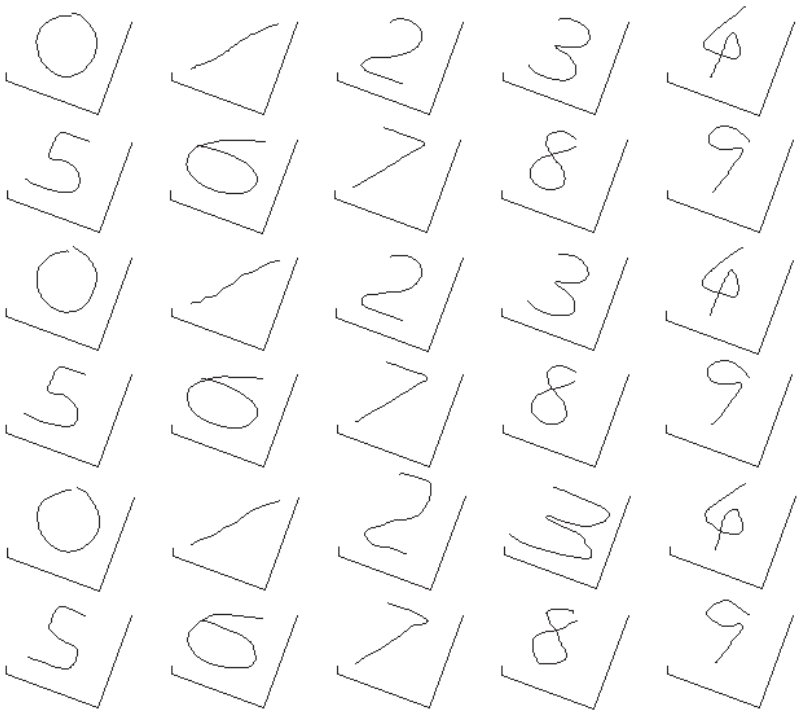

Fig. 28. The smoothed samples of the query series E, F and G by using the wavelet smoother.

predictable shape variance. For example, the sign samples in Figure 26 could be some unsteady demonstrations with much vibration by certain users. The results in Figures 27 and 28 show that the underlying trajectory shapes can be outlined reasonably from the noisy samples, which will lead to better performance than raw data in trajectory recognition. However, in some extreme cases, the smoothing may result in serious shape deviations from the original trajectory. This is understandable because it is essentially hard to perceive the original shape exactly from some very noisy data set. In summary, when the trajectories are not well acquired (with noise or disturbing vibrations), they would not be used in raw data form, but the smoothing preprocess proves to be an effective way to improve the quality of the trajectory.

A 1-NN classifier is designed to recognize the two kinds of query trajectories based on the DTW inter-signature distances. A query is recognized as the labeled canonical sample (Figure 16) with minimal signature distance. The trajectories in Figures 25 and 28 are recognized from the scaled distance tables in Figures 29 and 30, respectively (scaling bounds [0, 0.5]). According to the distinctive results in the table diagonals, we can obtain an intuitive confirmation that all of the query samples are recognized correctly.

Although the scale of the dataset is small in this experiment, the results reveal that the signature is effective in dealing with most sign trajectories acquired in real circumstances.

5.4. Large-scale Trajectory Recognition Test

In the third experiment, we test the signature's recognition using the large-scale UCI KDD ASL[1] dataset, and compare the performance with FD. Referring to the review in Section 1, the comparison result to FD is typical for that kind of descriptors using partial features.

ASL trajectory archive consists of 6650 samples of 95 sign classes, and 70 samples are collected for each class from five signers. The sign's length ranges from 40 to 200 points. Figure 31 shows three instances of the sign word "all" and "exit". To reduce noise and vibration, the wavelet smoother is applied to the sign trajectories using wavelet DB5 and extracting the third level coefficients.

A one-dimensional trajectory vector $x(t)$ with length M can be described by the Fourier coefficients based on the following Fourier transform (Harding and Ellis 2004),

$$X_f = \sum_{t=1}^{M} x(t) e^{-i2\pi(t-1)(f-1)/M}, \quad 1 \le f \le M. \tag{37}$$

As usual, setting $\hat{f} = [0, 1, 2, 3, 4]$, we can obtain a Fourier feature vector $F_l = \{X_{\hat{f}}, Y_{\hat{f}}, Z_{\hat{f}}\}$ to represent a three-dimensional trajectory defined by equation (1). Because F_l has

1. Available at: http://kdd.ics.uci.edu/databases/auslan/auslan.html.

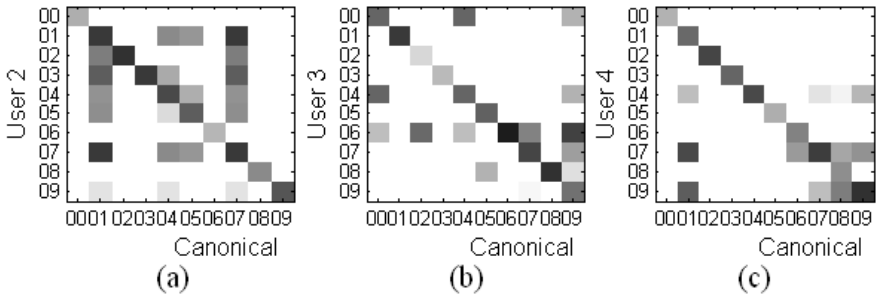

Fig. 29. The inter-signature distances comparison of the query series demonstrated by (a) user 2, (b) user 3 and (c) user 4.

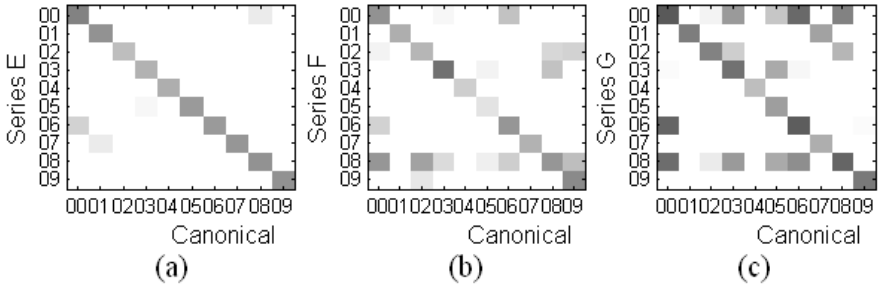

Fig. 30. The inter-signature distances comparison of the query (a) series E, (b) series F and (c) series G.

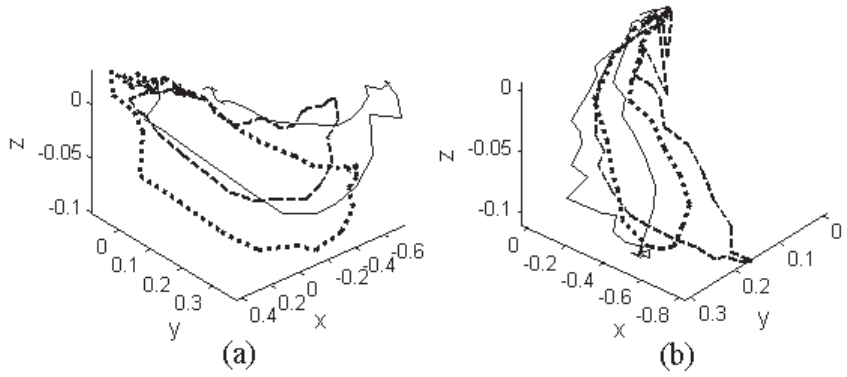

Fig. 31. Sign instances of words (a) “all” and (b) “exit”.

fixed length, two Fourier representations F_l^i and F_l^j are compared directly using the mean Euclidean distance

$$D_F = \frac{1}{l}\sum_{n=1}^{l}\left\|F_l^i - F_l^j\right\|.$$

The recognition experiment was carried out in two steps. First, half of the sign samples of a class are selected and resampled to the same length (100 frames) to average out a reference representation. Then, the other half of the samples are input to perform recognition using 1-NN classifier. Different classes and samples are picked randomly from the dataset to repeat the two-step trajectory recognition more than 50 times on a PC (Pentium 4 CPU 3.00 GHz, 512 MB RAM). An average result of the recognition performance in terms of accuracy and efficiency is compared in Tables 2 and 3, respectively. Obviously, the signature descriptor outperforms FD in recognition accuracy. However, this is achieved by sacrificing efficiency to some extent as can be seen from the query time in Table 3. As the signature is a full trajectory description and the

Table 2. Recognition accuracy comparison.

	Number of classes		
	2	4	8
Approximate signature	89.83%	83.79%	75.92%
Fourier descriptor	82.08%	63.70%	52.31%

Table 3. Recognition efficiency comparison (units of milliseconds per query).

	Number of classes				
	2	4	8	16	32
Approximate signature	996	1618	2981	5423	8724
Fourier descriptor	306	333	358	421	497

DTW algorithm is time-consuming, the recognition efficiency remains to be improved via signature optimization and a faster recognition engine in the future.

6. Discussion and Conclusion

Invariance is an important quality in measuring the effectiveness of a shape descriptor. In this paper, we have proposed a trajectory signature descriptor and focused on the investigation of its invariants to show how the substantial signature invariants can benefit effective trajectory recognition.

Compared with the existing descriptors, our signature has richer invariants. The signature's computational locality is the basis for deducing the six descriptive invariants. The rigid invariant and viewpoint invariant can make the trajectory description independent of observing viewpoints (camera poses), and the metric invariant enables the signature to be insensitive to the viewing distance. In particular, camera calibration and reconstruction can be avoided by use of the projective invariant. Although the speed invariant is not an absolute invariant, the developed non-linear inter-signature matching can recognize the same trajectories with diverse speed profiles. The occlusion invariant is also a relative invariant subject to the degree of occlusion.

It should be pointed out that the performance of a trajectory recognition system relies on two key factors: the trajectory description and the recognition engine. As verified in Section 5.2, the signature's invariants enable a trajectory to be represented and recognized invariantly. This can enhance the effectiveness in trajectory recognition. Meanwhile, noticing the experiment results in Section 5.4, there is still room for improvement in the recognition efficiency for large-scale trajectories.

The proposed trajectory signature is a generic descriptor for free form motion trajectory. Hence, it serves as a foundation for wide applications where motion trajectory is involved, for instance, trajectory based robot motion recognition and interaction, object trajectory based motion retrieval from a video database and the learning and recognition of trajectory based human motions (actions, behaviors and activities). For complex motions, multiple spatially parallel and temporally continuous trajectories can be extracted for joint description and recognition.

The signature used in experiments in this paper was for sign language recognition which could be used to enhance human–robot interaction. It can be extended to robot learning, particularly for robot programming by demonstration (PbD; Friedrich et al. (1996)). Unlike short-term robot tasks such as grasping (Bernadin et al. 2005), motion trajectory is an effective element for describing many kinds of PbD human demonstrations (Sato et al. 2002) especially for the long-term motions. Further, PbD demands systematic representation to demonstrations (Acosta Calderon and Hu 2005). The invariant signature description actually admits generalized representation to the task to be learned. Hence, the signature's substantial invariants can offer much flexibility and adaptability to the reproduction of the learned task by the robot. For example, the Euclidean invariant enables the signature to be independent of absolute coordinates. Thus, given a learned signature, we can instantiate arbitrary trajectory samples at any relative position. Similarly, the metric invariant makes it possible to produce diverse sizes of trajectories from an individual signature. In addition, the developed DTW matching algorithm provides an appropriate learning metric to measure the quality of a reproduced trajectory by robot by comparison with a reference signature. In general, the signature's invariants can not only make recognition more effective, but can also boost robot learning.

The focus of this paper has been on building the signature descriptor and highlighting its advantage in descriptive invariants. From the experiment results and analysis of both synthetic and real motion trajectories, it can be concluded that the novel signature is rich in invariants and it can enhance the effectiveness in trajectory recognition.

Acknowledgments

This work was fully supported by the Research Grants Council of Hong Kong under Project CityU117605. The authors would like to thank Professor Jianwei Zhang of University of Hamburg, Germany for the valuable discussion on this research topic. Thanks are also due to Songtao Guo, He Huang and Yangyang Zhao for their assistance in the experiments.

References

Acosta Calderon, C. A., and Hu, H. (2005). Robot imitation from human body movements. *Proceedings of the 3rd International Symposium on Imitation in Animals and Artifacts, AISB Convention*, Hatfield, UK, pp. 1–9.

Aleotti, J. and Caselli, S. (2005). Trajectory clustering and stochastic approximation for robot programming by demonstration. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Alberta, Canada.

Alferez, R. and Wang, Y. F. (1999). Geometric and illumination invariants for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(6): 505–535.

Bennewitz, M., Burgard, W., Cielniak, G. and Thrun, S. (2005). Learning motion patterns of people for compliant robot motion. *The International Journal of Robotics Research*, **24**(1): 31–48.

Bernadin, K., Ogawara, K., Ikeuchi, K. and Dillmann, R. (2005). A sensor fusion approach for recognizing continuous human grasping sequences using Hidden Markov models. *IEEE Transactions on Robotics*, **21**(1): 47–57.

Black, M. J. and Jepson, A. D. (1998). A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. *European Conference on Computer Vision (ECCV)*, Freiburg, Germany, vol. 1, pp. 909–924.

Boutin, M. (2000). Numerically invariant signature curves. *International Journal of Computer Vision*, **40**(3): 235–248.

Bradski, G. R. (1998). Computer vision face tracking as a component of a perceptual user interface. *Workshop on Applications of Computer Vision*, Princeton, NJ, pp. 214–219.

Bruckstein, A. M., Holt, R., Netravali, A. and Richardson, T. (1993). Invariant signatures for planar shape recognition under partial occlusion. *Computer Vision, Graphics, and Image Processing: Image Understanding*, **58**(1): 49–65.

Bui, T. D. and Chen, G. Y. (1997). Multiresolution moment-Fourier-wavelet descriptor for 2D pattern recognition. *Proceedings of SPIE*, **3078**: 552–557.

Burns, J. B., Weiss, R. S. and Riseman, E. M. (1992). The non-existence of general-case view-invariants. *Geometric Invariance in Computer Vision*, Mundy, J. L. and Zisserman, A. (eds). Cambridge, MA, MIT Press.

Calabi, E., Olver, P. J., Shakiban, C., Tannenbaum, A. and Haker, S. (1998). Differential and numerically invariant signature curves applied to object recognition. *International Journal of Computer Vision*, **26**(2): 107–135.

Caspi, Y., Simakov, D. and Irani, M. (2006). Feature-based sequence-to-sequence matching. *International Journal of Computer Vision*, **68**(1): 53–64.

Chaker, F., Bannour, T. and Ghorbel, F. (2003). A complete and stable set of affine-invariant Fourier descriptors. *International Conference in Image Analysis and Processing*, Mantova, Italy, pp. 578–581.

Chand, S. and Doty, K. L. (1985). On-line polynomial trajectories for robot manipulators. *The International Journal of Robotics Research*, **4**(2): 38–48.

Charlebois, M., Gupta, K. and Payandeh, S. (1999). Shape description of curved surfaces from contact sensing using surface normals. *The International Journal of Robotics Research*, **18**(8): 779–787.

Chuang, G. and Kuo, C. C. (1996). Wavelet descriptor of planar curves: theory and applications. *IEEE Transaction on Image Processing*, **5**(1): 56–70.

Cohen, F. S., Huang, Z. and Yang, Z. (1995). Invariant matching and identification of curves using B-splines curve representation. *IEEE Transactions on Image Processing*, **4**(1): 1–10.

Forsyth, D., Mundy, J. L., Zisserman, A., Coelho, C., Heller, A. and Rothwell, C. (1991). Invariant descriptors for 3D object recognition and pose. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **13**(10): 971–991.

Friedrich, H., Munch, S., Dillmann, R., Bocionek, S. and Sassin, M. (1996). Robot programming by demonstration (RPD): supporting the induction by human interaction. *Machine Learning*, **23**(2–3): 163–189.

Gueziec, A. and Ayache, N. (1994). Smoothing and matching of 3-D space curves. *International Journal of Computer Vision*, **12**(1): 79–104.

Harding, P. R. G. and Ellis, T. J. (2004). Recognizing hand gesture using Fourier descriptors. *International Conference on Pattern Recognition (ICPR)*, Cambridge, UK, vol. 3, pp. 286–289.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, **8**(2): 179–187.

Huang, Z. and Cohen, F. S. (1996). Affine-invariant B-spline moments for curve matching. *IEEE Transactions on Image Processing*, **5**(10): 1473–1480.

Jiang, X. and Motai, Y. (2005). Learning by observation of robotic tasks using on-line PCA-based Eigen behavior. *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Espoo, Finland, pp. 391–396.

Kehtarnavaz, N. and Figueiredo, R. J. P. (1988). A 3-D contour segmentation scheme based on curvature and torsion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**(5): 707–713.

Kindratenko, V. V. (2003). On using functions to describe the shape. *Journal of Mathematical Imaging and Vision*, **18**(3):225–245.

Khalil, M. I. and Bayoumi, M. M. (2002). Affine invariants for object recognition using the wavelet transform. *Pattern Recognition Letters*, **23**(1–3): 57–72.

Kunttu, I., Lepisto, L., Rauhamaa, J. and Visa, A. (2003). Multiscale Fourier descriptor for shape classification. *International Conference on Image Analysis and Processing*, Mantora, Italy, pp. 536–541.

Latecki, L. J., Megalooikonomou, V., Wang, Q., Lakaemper, R., Ratanamahatana, C. A. and Keogh, E. (2005). Elastic partial matching of time series. *European Conference on Principles and Practice in Knowledge Discovery in Databases (PKDD)*, Porto, Portugal, pp. 577–584.

Malis, E., Chesi, G. and Cipolla, R. (2003). 2-1/2D visual servoing with respect to planar contours having complex and

unknown shapes. *The International Journal of Robotics Research*, **22**(10–11): 841–854.

Manay, S., Cremers, D., Hong, B. W., Yezzi, A. J. and Soatto, S. (2006). Integral invariants for shape matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**(10): 1602–1618.

Meyer, D., Psl, J. and Niemann, H. (1998). Gait classification with HMMs for trajectories of body parts extracted by Mixture densities. *British Machine Vision Conference (BMVC)*, Southampton, England, pp. 459–468.

Min, J. and Kasturi, R. (2004). Extraction and temporal segmentation of multiple motion trajectories in human motion. *IEEE CVPR Workshop on Detection and Recognition of Events in Video*, Washington, DC.

Mokhtarian, F., Abbasi, S. and Kittler, J. (1996). Robust and efficient shape indexing through curvature scale space. *British Machine Vision Conference (BMVC)*, Edinburgh, UK, vol. 1, pp. 53–62.

Olver, P. J. (2001). Moving frames—in geometry, algebra, computer vision, and numerical analysis. *Foundations of Computational Mathematics*, DeVore, R., Iserles, A. and Suli, E. (eds) (Mathematical Society Lecture Note Series, vol. 284). Cambridge, Cambridge University Press, pp. 267–297.

Parameswaran, V. and Chellappa, R. (2006). View invariance for human action recognition. *International Journal of Computer Vision*, **66**(1): 83–101.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(7): 629-639.

Psarrou, A., Gong, S. and Walter, M. (2002). Recognition of human gestures and behaviour based on motion trajectories. *Image and Vision Computing*, **20**(5–6): 349–358.

Rao, C. and Shah, M. (2001). View-invariant representation and learning of human action. *IEEE Workshop on Detection and Recognition of Events in Video*, Vancouver, BC, Canada, pp. 55–63.

Rao, C., Yilmaz, A. and Shah, M. (2002). View-invariant representation and recognition of actions. *International Journal of Computer Vision*, **50**(2): 203–226.

Sato, Y., Bernardin, K., Kimura, H. and Ikeuchi, K. (2002). Task analysis based on observing hands and objects by vision. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Lausanne, Switzerland, pp. 1208–1213.

Trucco, E. and Verri, A. 1998. *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ, Prentice Hall.

Ude, A., Atkeson, C. G. and Riley, M. (2000). Planning of joint trajectories for humanoid robots using B-spline wavelets. *IEEE International Conference on Robot and Automation (ICRA)*, San Francisco, CA, vol. 3, pp. 2223–2228.

Vlachos, M., Kollios, G. and Gunopulos, D. (2005). Elastic translation invariant matching of trajectories. *Machine Learning*, **58**(2–3): 301–334.

Weiss, I. (1993). Geometric invariants and object recognition. *International Journal of Computer Vision*, **10**(3): 207–231.

Yang, M. H. and Ahuja, N. (1999). Recognizing hand gestures using motion trajectories. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Fort Collins, CO, pp. 468–472.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11): 1330–1334.