To Appear in *Pattern Recognition* 1996

Invariant Representation, Matching and Pose Estimation of 3D Space Curves under Similarity Transformations

Stan Z. Li [1]

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798

This paper presents a system for matching and pose estimation of 3D space curves under the similarity transformation composed of rotation, translation and uniform scaling. The system makes use of constraints not only on the feature points but also on the curve segment. A representation called the similarity-invariant coordinate system (SICS) is presented for deriving semi-local invariants of 3D curves. The SICS also enables an efficient exploration of constraints on the shape of the entire curve. Model-based curve matching is performed in the principle of maximum *a posteriori* (MAP) probability in which Markov random fields (MRFs) are used to model the prior probability. Results are given to demonstrate the matching with missing and extra feature points. The extensive exploration of the curve shape constraints in both the matching and the pose estimation stages are shown to significantly improve the quality of the recognition and pose estimation.

Key words: Invariants, interpolation, partial matching, pose estimation, similarity transformations, 3D curves.

[1] Email address: szli@szli.eee.ntu.ac.sg

1 Introduction

Recognition and pose determination of objects in 3D space is important for industry automation, robotics, navigation, and medical applications. The task is to identify the categories of objects in the scene and to determine their locations relative to the modeled objects. 3D objects can be represented by space curves, such as those formed by boundaries of regions, ridge lines, lines of curvature and parabolic lines of 3D surfaces [1,2] which are usually extracted from CAD/CAM object models, range images and MR medical images [3–10]. Such a representation is abstract and yet highly descriptive. The matching and pose of 3D objects is then performed based on the space curves [3–6,9,10].

Curve representation and matching has attracted many researchers in the past. Planar curve matching has been most widely investigated [11–17]. Works are also done for 3D space curves with known scales [3–5,18,9]. The following situations increase difficulties in solving the problem: (i) when the scale of curves is unknown. (ii) when only partial curves (with respect to the modeled curves) are present, (iii) when the data contains several partial curves, each of which is due to one modeled curve, and these partial curves are un-separated *a priori*, and (iv) when the data is subject to noise. The problem considered in this paper is representation, matching and pose estimation of 3D space curves under the class of similarity transformations (denoted by $\mathcal{T}_s$ hereafter) composed of rotations, translations and unknown scales, with all the above difficult features. The problem in such situations has not been well researched.

Representation is the basis on which the matching and pose estimation is performed. The use of invariant representations makes the direct comparison of shapes possible so that object matching can be performed independently of the object pose. An invariant representation for curves may be classified as global, local or "semi-local". In a local representation, a curve is decomposed into segments at local feature points such as curvature extrema, and then represented by the points and their (desirably invariant) attributes. Based on this representation, the matching of curves is reduced to the matching of attributed point patterns. Such a representation is suitable for handling partial and unsegmented (*i.e.* connected) curves. A drawback is that it treats a curve as a pattern of the feature points and thus makes no use of the information about curve segments between the feature points. This is wasteful because such information provides substantial constraint on curve matching. Another drawback is that the computation of local invariants usually requires high order derivatives.

A global representation can be non-feature based, for example using polynomials [5,18]. The matching thereby is usually done by using a least squares fitting method. The advantage of such scheme is that it needs neither the extraction of local features nor the estimation of derivatives, thus avoiding difficulties therein. A disadvantage is that it is less suitable to handle the partial curves and scale changes. Moreover, it requires that the curve data should contain only one modeled curve before fitted to the model. When the data contains several modeled curves, different curves have to be separated *a priori*.

A semi-local representation offers a compromise. It requires lower order derivatives than purely local representations in the computation of invariants. Also, it makes fuller use of curve information than a local representation and hence provides more constraints on the matching. An example is the affine invariant coordinates for flat objects [12,16,19]. There, a planar curve segment is mapped into an invariant coordinate frame based on which the invariants can be computed. A disadvantage of such a representation is that it is less local than a truly local representation.

In this paper, we present a system for partial matching and pose estimation of 3D space curves under $\mathcal{T}_s$. A two level invariant representation is used in which the lower level representation deals with curve segments and the higher level representation does the whole curve. A curve is decomposed into curve segments at feature points such as curvature maxima and tangent discontinuities which are invariant to viewpoints. We use a semi-local representation called the similarity-invariant coordinate system (SICS) as the lower level representation. The SICS requires only up to second order derivatives to achieve the similarity invariance. An SICS is constructed for each curve segment bounded by a pair of feature points. In the SICS, the 3D geometry of the segment, *i.e.* the set of the coordinates of the segment, is invariant under $\mathcal{T}_s$. This enables an extensive exploration of information about the curve segment. A vector of invariant quantities can be derived as the descriptor of the segment. They are used in matching as relational constraints for the two feature points bounding the segment. At the higher level, the whole curve is represented by an attributed relational graph (ARG) in which the nodes correspond to the feature points and the links do to the vectors of invariant relations. Curve matching is then reduced to ARG matching.

An optimal matching solution is sought in terms of an optimality criterion. We formulate the matching problem as finding the maximum *a posteriori* (MAP) probability labeling of the feature points on the observed curve in terms of the feature points on a modeled curve. Markov random fields (MRFs) are used to model the prior contextual constraints and define the prior distribution of the labeling [20]. The contextual constraints due to the relational quantities derived from the SICS's are encoded into the distribution of the observation conditioned on the labeling. By these, the MAP labeling is defined. It is found by using an optimization algorithm.

In pose estimation, curve information is again explored not only about the corresponding feature points but also about the curve segments bounded by these points. The set of matched point correspondences is augmented by interpolating the corresponding curve segments. The augmentation increases the accuracy of pose estimation and can help detecting accidental false matches, which is demonstrated by our results.

The rest of the paper is organized as follows. The next section presents the SICS representation. Section 3 explores curve constraints based on SICS, derives similarity invariants, and builds the higher level ARG description. Section 4 addresses model-based curve matching based on the invariants. Section 5 deals with pose estimation of curves. Results are presented in Section 6. This is followed by conclusion in Section 7.

2 Similarity Invariant Coordinator System

A curve is broken at invariant feature points and segmented into curve segments. An SICS is established for each of the curve segments. This provides a similarity-invariant curve representation at the low level.

2.1 Invariant Feature Points

Initially, a curve is in the form of an un-directed chain of (x, y, z) coordinates of 3D points. Curvature maxima and tangent discontinuities are detected from the curve after a smoothing pre-processing and used as similarity invariant feature points[2]. However, the set of useful true feature points will be empty if all the feature points happen to be occluded. It is assumed in this paper that this does not happen and a sufficient number of feature points are present. Some regular curves such as a planar circle or a space helix are not considered here because they cannot be characterized by the above features. Missing and false detection of some points is tolerated by our matching algorithm. In the following, we illustrate how to detect the feature points.

Figure 1 illustrates a curve before and after smoothing and shows the detected feature points. The curvature estimation undergoes three steps: curve re-parameterization, derivative estimation and curvature computation. Firstly, the input curve is re-parameterized by arc-length using linear interpolation, which makes curve points evenly spaced in arc-length. This enables us to choose a smoothing kernel size relative to the arc-length and to control the

[2] Torsion extrema and curvature discontinuities could also have been used, but not because they are less reliable.

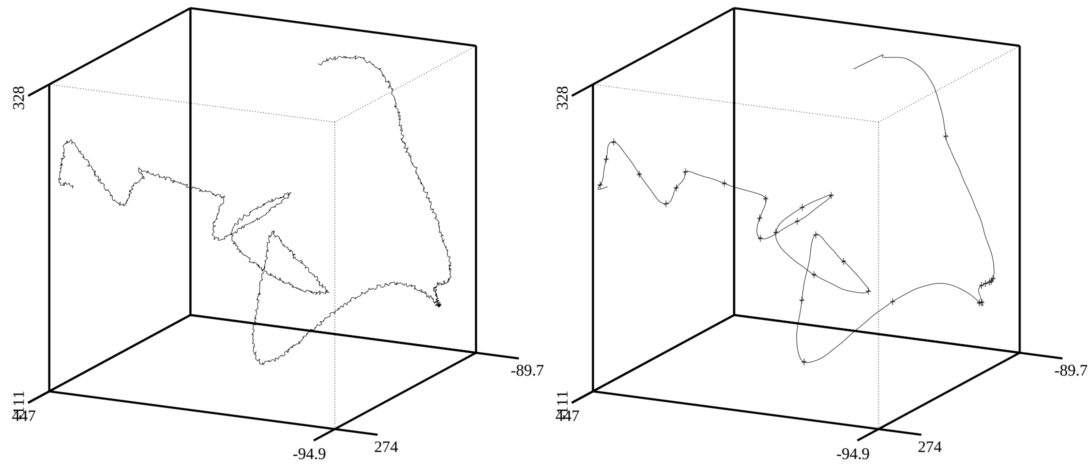

Fig. 1. A noisy curve before (left) and after (right) smoothing. Curvature-maxima (in crosses), *i.e.* corner points, are detected after smoothing. Mid-points are inserted (also in crosses).

degree of smoothing. Repeated re-parameterization is found to have an effect of removing noise. In the second step, the first and second derivatives in three directions are estimated by convolution using derivative operators [21]. Gaussian kernels are used for the smoothing where the kernel size is also chosen relative to the arc-length used for the re-parameterization. Currently, the size is chosen on an *ad hoc* basis[3]. Compensation is performed on the smoothed curve coordinates to reduce the effect of curve shrinking due to smoothing [22]. The curvature is computed using the estimated derivatives.

Corner feature points are detected after the curvature estimation. A method proposed in [23] is used to detect reliable curvature maxima from candidate maxima: First, the "region of support" [23] is determined for each candidate maximum point. Then the perpendicular distance from the point to the line passing through the two end points of the support is measured. A candidate is considered to be a true maximum if the measured distance is larger than a threshold, say 0.1 in our implementation.

If there are too few feature points detected, say only two, the mid-point is inserted between each pair of consecutive corner points to augment the set of feature points (see Figure 1). The mid-point seems to be a natural choice because when the shape of the segment is unknown, dividing the segment into two equal length pieces is better than two uneven ones. The mid-points are considered to be of different type from the corner points in the matching stage. The use of the mid-points in addition to the corner points improves the reliability of matching[4]. The improvement is significant when there are too

[3] Choosing the right size for a smoothing kernel is an unsolved problem. A fixed size may not be good for all scales. Multi-scale based methods may provide better solutions.

[4] We may have chosen to insert more than one point, but we found the increased

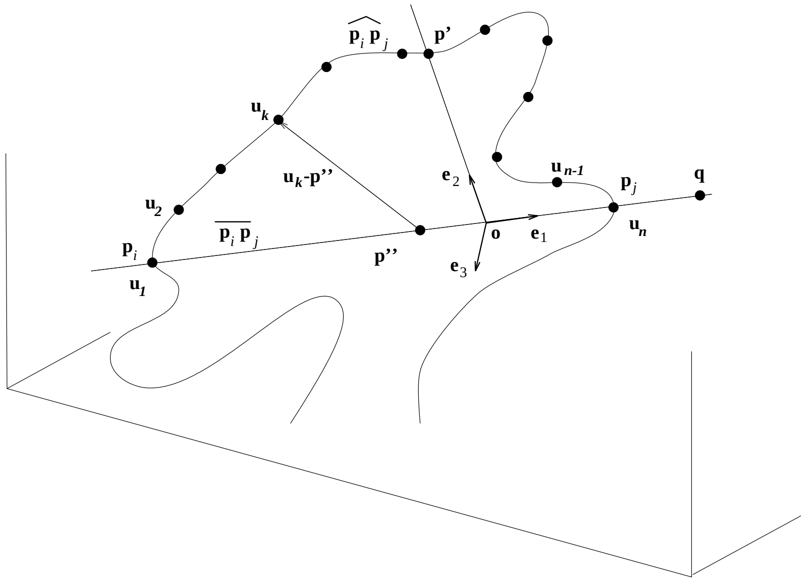

Fig. 2. An SICS is constructed for each curve segment bounded by two feature points p_i and p_j (which can be curvature extrema or of other types).

few feature points to impose sufficient constraints.

2.2 Similarity-Invariant Coordinate System (SICS)

An SICS in 3D is then established for each curve segment bounded by a pair of feature points (the detected points and the inserted points if any). Refer to the curve in Fig.2. Assume there are a set of m feature points $\{p_i = (x_i, y_i, z_i)\}$. Consider the curve segment $\widehat{p_ip_j}$ between a pair of points p_i and p_j and the straight line $\overline{p_ip_j}$ passing through them [5]. Find the mid-point [6], denoted by p', on $\widehat{p_ip_j}$ such that curve segments $\widehat{p_ip'}$ and $\widehat{p'p_j}$ have the equal arc-length. Obviously, p' is unique for $\widehat{p_ip_j}$. Next, find the point, denoted by $\mathbf{o}$, on $\overline{p_ip_j}$ such that line $\overline{\mathbf{o}p'}$ is perpendicular to $\overline{\mathbf{o}p_j}$. There are two degenerated cases where the SICS cannot be constructed: when $\mathbf{o} = p'$ and when $\mathbf{o} = p_j$. Some protocol can be established to deal with such cases. But in this paper we consider only non-degenerated situations.

Now, $\overline{\mathbf{o}p'}$, $\overline{\mathbf{o}p_j}$ and $\overline{\mathbf{o}p'} \times \overline{\mathbf{o}p_j}$, where "$\times$" denotes the cross product, are three orthogonal vectors which are rotation- and translation-invariant *with respect*

complexity un-necessary.

[5] In the case of *closed* curves, there are two possible ways of defining $\widehat{p_ip_j}$ bounded by p_i and p_j. It is assumed in this paper that $\widehat{p_ip_j}$ is defined by the shorter one. This is tricky.

[6] Unlike those the mid-points introduced previously as additional feature points, a mid-points here is not as a feature point but for constructing the SICS.

to the curve segment $\widehat{p_i p_j}$ in the original 3D space. They can be normalized into unit length

$$\mathbf{e}'_1 = \frac{p_j - \mathbf{o}}{||p_j - \mathbf{o}||} \tag{1}$$

$$\mathbf{e}'_2 = \frac{p' - \mathbf{o}}{||p' - \mathbf{o}||} \tag{2}$$

and

$$\mathbf{e}'_3 = \mathbf{e}'_1 \times \mathbf{e}'_2 \tag{3}$$

where $||\cdot||$ denotes the Euclidean norm. They are orthonormal vectors in that space.

To achieve the invariance to scaling, we re-scale the vectors by a factor which is determined by the curve segment itself. The factor can be the length between any two distinct known points on the curve segment, which tells us the scaling factor of the curve. We use $||p' - \mathbf{o}||$ as the factor and obtain the following three rotation- , translation-, and scale-invariant orthogonal vectors

$$\mathbf{e}_1 = (||p' - \mathbf{o}||)\mathbf{e}'_1 = \frac{||p' - \mathbf{o}||}{||p_j - \mathbf{o}||}(p_j - \mathbf{o}) \tag{4}$$

$$\mathbf{e}_2 = (||p' - \mathbf{o}||)\mathbf{e}'_2 = (p' - \mathbf{o}) \tag{5}$$

and

$$\mathbf{e}_3 = (||p' - \mathbf{o}||)\mathbf{e}'_3 = \frac{1}{||p_j - \mathbf{o}||}(p_j - \mathbf{o}) \times (p' - \mathbf{o}) \tag{6}$$

They are uniquely defined for the curve in non-degenerated cases.

The SICS is then denoted by quadruple

$$\{\mathbf{o}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} \tag{7}$$

where $\mathbf{o}$ is the origin. In this system, a point p is expressed as

$$p = \mathbf{o} + u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3 \tag{8}$$

where $(u_1, u_2, u_3) \triangleq u(p)$ are the coordinates of p in the SICS. The coordinates of p', p_j and p_i are $(0, 1, 0)$, $(\frac{||p_j - \mathbf{o}||}{||p' - \mathbf{o}||}, 0, 0)$, and $(-\frac{||p_i - \mathbf{o}||}{||p' - \mathbf{o}||}, 0, 0)$, respectively.

In general, the SICS coordinates of any point p on the curve segment $\widehat{p_i p_j}$ are invariant under $\mathcal{T}_s$. This is illustrates as follows. Assume that $\widehat{p_i p'}$ is translated, rotated and uniformly scaled into a new segment $\mathcal{T}_s(\widehat{p_i p'})$ and thus a point p on it is transformed into $\mathcal{T}_s(p)$. Establish two SICS's, denoted SICS_1 and SICS_2, one for $\widehat{p_i p'}$ and the other for $\mathcal{T}_s(\widehat{p_i p'})$. Because each SICS is established relative to the respective segment, the coordinates $u(p)$ of p in SICS_1 and the coordinates $u(\mathcal{T}_s(p))$ of $\mathcal{T}_s(p)$ in SICS_2 should be exactly the same. For this reason, we may say that the SICS makes the 3D geometry of the curve segment unique.

Both the SICS proposed here and the affine coordinate systems in [12,19] are aimed to establish some invariant reference frame. The differences between the latter and the former are the following: Objects of interest in those works are flat (2D) whereas true 3D space curves are considered here. Those work explore affine invariance whereas this work similarity invariance, resulting in different types of basis.

Both types of coordinate systems are semi-local. They require that (i) the curve under consideration should not break into too small segments; (ii) there should be at least one continuous segment in the scene; and (iii) the segment should contain at least a minimum number of consecutive feature points, the minimum number being 3 in [12], 4 in [19] and 2 in this work. However, under $\mathcal{T}_s$, the orthogonal basis of the SICS may be better than general affine basis as the feature quantities in orthogonal basis are least correlated.

3 Invariant Relational Constraints and ARG

Similarity invariant quantitative relations are derived between each pair of feature points. They will be used as a similarity invariant descriptor of the curve segment bounded by the two points. An ARG is then constructed to encapsulate the feature points and the invariant relations. This provides an invariant representation for the whole curve at the higher level.

3.1 Exploring Curve Constraints

The SICS enables invariant representation of the feature points in the form of invariant coordinates. It allows a full exploration of the information about each curve segment. This is illustrated as follows.

Firstly, sample a curve segment $\widehat{p_i p_j}$ into n equally spaced (in arc-length) points. This is equivalent to inserting $n-2$ points between p_i and p_j. The n

points in the SICS, denoted by $u_k = (u_{k,1}, u_{k,2}, u_{k,3})$ $(k = 1, ..., n)$ (see Fig.2), provide $3n$ similarity invariant quantities; of those u^k $(k = 2, ..., n-1)$ are useful for matching (u^1 and u^n corresponding to p_i and p_j are the same for all the segments). These coordinates constitute a complete and independent set of invariants for the n points which are used as the vector of invariant relations between p_i and p_j. It is a rich descriptor of the curve segment and provides much stronger constraints for matching than merely the locations of the two end points.

One can derive the relational vector for $\widehat{p_j p_i}$ from the vector for $\widehat{p_i p_j}$. But the order of the components in these two vectors is different when the segment is symmetric about the mid-point p'. Therefore, the vector for $\widehat{p_j p_i}$ is different from the vector for $\widehat{p_i p_j}$ as long as the segment is non-symmetric.

An important special case is when the curve segment $\widehat{p_i p_j}$ is a straight line and hence the orthogonal system can not be established. In this case, the relations can be defined as follows: only the coordinates in the direction of $\mathbf{e}_1$ are effective and those in the other two directions are always zero.

Invariant angles can also be derived as follows: Consider the mid-point of the straight line segment $\overline{p_i p_j}$, denoted by p'', and obtain n vectors $u_k - p''$. The angle between each of these vectors and each of the three orthogonal axes $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ is similarity invariant[7]. There are $3n$ such angles, denoted by angle vectors θ_k $(1 \leq k \leq n)$, of which $3(n-2)$ are useful for the matching for the same reason as mention earlier for the u_k coordinates. The angles are dependent on the coordinates through some nonlinear relationship. They are also useful for matching.

Other quantities may also be explored as similarity invariant relations between two concerned feature points in addition to the invariant coordinates u_k and angles θ_k. Consider the tangents to the curve at points u_k. The angle between each tangent and each axis is also an invariant. The disadvantage of using such angles is that they require the computation of one order higher derivatives than the coordinates and thus more susceptible to noise. Additional useful quantitative relations include the length of $\widehat{p_i p_j}$ over the length l^0 of $\overline{p_i p_j}$ and the ratio of curvature magnitude at p_i and p_j.

[7] The choice of the midpoint p'' instead of $\mathbf{o}$ as the reference point for the angles is for stability reason: The angles can change drastically as point $\mathbf{o}$ shifts from between p_i and p_j to the outside of the segment, say between p_j and q in Figure 2, due to noise; however this instability can be eliminated by using the midpoint p''.

A curve as an object in the scene or in the model-base can be described by an attributed relational graph (ARG) [24–29]. The scene ARG is denoted by a triplet $\mathcal{G} = (\mathcal{S}, d_1, d_2)$. In this notation, $\mathcal{S} = \{1, \ldots, m\}$ indexes the set of m feature points in the scene curve; $d_1 = \{d_1(i) \mid \forall i \in \mathcal{S}\}$ is a set of unary properties; each $d_1(i) = [d_{1,1}(i), \ldots, d_{1,K_1}(i)]$ is a vector consisting of K_1 different types of unary properties of node i; $d_2 = \{d_2(i,j) \mid \forall i, j \in \mathcal{S}, i \neq j\}$ is a set of binary relations; each $d_2(i,j) = [d_{2,1}(i,j), \ldots, d_{2,K_2}(i,j)]$ is a vector consisting of K_2 different types of binary relations made between nodes i and j. In 3D curve matching, the unary property here is the type of the feature point in the set {corner-point, mid-point} and is useful only when the mid-points are used as additional feature points. Two nodes can be matched to each other only when they are of the same type. There are a total number of $6(n-2)$ relational quantitaties for $\widehat{p_i p_j}$, consisting of the coordinate vectors $[u_2, \ldots, u_{n-1}]$ and the angle vectors $[\theta_2, \ldots, \theta_{n-1}]$.

A model object is similarly represented by $\mathcal{G}' = (\mathcal{L}, D_1, D_2)$ in which

$$\mathcal{L} = \{1, \ldots, M\} \tag{9}$$

is a set of M model nodes each of which indexes a feature of the model object, D_1 is the set of unary properties and D_2 is a set of binary relations. Now introduce a virtual feature, called the NULL and indexed by 0. The NULL node represents everything not modeled by $\mathcal{G}'$, such as features due to all un-modeled objects and noise. Let

$$\mathcal{L}^+ = \{0\} \cup \mathcal{L} = \{0, 1, \ldots, M\} \tag{10}$$

denote the augmented set of object nodes. It is the set of labels to be assigned to the scene features in $\mathcal{S}$.

4 MAP-MRF Curve Matching

The matching establishes correspondences between a subset of nodes, *i.e.* feature points, in the scene ARG and those in the model ARG. In this paper, this is posed as a labeling problem which involves assigning a label from $\mathcal{L}^+$ to each of the sites in $\mathcal{S}$. The maximum *a posteriori* (MAP) principle is used to evaluate the assignment of labels. Markov random fields (MRFs) are used to derive the posterior distribution of a labeling [20].

A natural representation in the study of MRFs is labeling [30]. Object matching, like many other image analysis problems, can be posed as a labeling problem. A labeling problem is specified in terms of a set of sites and a set of labels. Let

$$\mathcal{S} = \{1, \ldots, m\} \tag{11}$$

be a set of m discrete sites representing the features in the scene, and

$$\mathcal{L} = \{1, \ldots, M\} \tag{12}$$

be a set of M discrete labels representing the features in a model. Object matching is to assign a label from $\mathcal{L}$ to each site in $\mathcal{S}$.

Let $F = \{F_1, \ldots, F_m\}$ be a family of random variables, or a random field, defined on $\mathcal{S}$, in which each random variable F_i assumes a value in $\mathcal{L}$. A realization of F is called a configuration or a labeling, and is denoted by $f = \{f_1, \ldots, f_m\}$. In connection to the image matching problem, $f_i \in \mathcal{L}$ is the model feature matched to the image feature i. The admissible configuration space for f is denoted by $\mathbb{F}$. An MRF is always defined with respect to a neighborhood system. A neighborhood system for $\mathcal{S}$ is given by

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{S}\} \tag{13}$$

where $\mathcal{N}_i$ is the collection of sites neighboring i. Denote $f_{\mathcal{S}-\{i\}} = \{f_{i'} \mid i' \in \mathcal{S} - \{i\}\}$ and $f_{\mathcal{N}_i} = \{f_{i'} \mid i' \in \mathcal{N}_i\}$. A random field F is said to be an MRF on $\mathcal{S}$ with respect to $\mathcal{N}$ if the Markovianity condition

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = P(f_i \mid f_{\mathcal{N}_i}) \tag{14}$$

and the positivity condition

$$P(f) > 0 \qquad \forall f \in \mathbb{F} \tag{15}$$

are satisfied.

In the above, an MRF is defined according to the local conditional distributions. However, the global joint distribution $P(f)$ is often required for applications such as the MAP-MRF matching in this paper. Yet, defining the joint distribution from the conditional distributions is a difficult problem [30]. Fortunately, there is a MRF-GD (Gibbs distribution) theorem [31,30,32] which can be used for conveniently obtaining the joint distribution. The theorem states that F is an MRF on $\mathcal{S}$ with respect to $\mathcal{N}$ if and only if its joint distribution $P(F = f)$ is a Gibbs distribution with respect to $\mathcal{N}$. In other words, if F is locally an MRF, then globally its joint distribution is of Gibbs; and vice versa.

A GD of F with respect to $\mathcal{N}$ is of the following form

$$P(f) = Z^{-1} \times e^{-U(f)/T} \tag{16}$$

In the above, Z is a normalizing constant, T is a global control parameter called the temperature and $U(f)$ is the energy of the form

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) \tag{17}$$

where $\mathcal{C}$ is the set of "cliques" and $V_c(f)$ are the clique potential functions. In essence, a GD is featured by two things: it belongs to the exponential family; the energy is defined on cliques. A clique c for $(\mathcal{S}, \mathcal{N})$ is a subset of $\mathcal{S}$ consisting of either a single site or several neighboring sites. When cliques containing up to two sites are considered, the energy has the following form

$$U(f) = \sum_{\{i\} \in \mathcal{C}_1} V_1(f_i) + \sum_{\{i,i'\} \in \mathcal{C}_2} V_2(f_i, f_{i'}) \tag{18}$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are the set of single-site and the set of pair-site cliques, respectively, and V_1 and V_2 are single- and pair-site potential functions.

In the MAP-MRF framework, the optimal solution is the one which maximizes the posterior probability. The posterior probability is usually also a GD

$$P(F = f \mid d) = Z_E^{-1} \times e^{-E(f)} \tag{19}$$

where d is the observation and, $E(f) = \frac{1}{T} U(f) + U(d \mid f)$, called the posterior energy, is a weighted sum of the prior energy $U(f)$ (the energy in the prior probability) and the likelihood energy $U(d \mid f)$ (the energy in the conditional probability). The MAP solution is equivalently found by minimizing the posterior energy

$$f^* = \arg \min_{f \in \mathbb{F}} E(f) \tag{20}$$

When $\mathcal{L}$ consists of (un-ordered) discrete labels as in our case (as opposed *e.g.* to continuous labels in a real interval), the minimization of the posterior energy is a combinatorial optimization problem. The interested reader is referred to [20] for a detailed exposure to labeling problems in image analysis, the MAP-MRF framework and minimization algorithms for MRFs.

4.2 The Posterior Distribution

The prior distribution of f is modeled by MRFs. The neighborhood $\mathcal{N}_i$ of i comprises of other feature points which are within the arclength of r from i

$$\mathcal{N}_i = \{i' \in \mathcal{S} \mid \text{arclength}(\widehat{p_i p_{i'}}) \leq r, i' \neq i\} \tag{21}$$

The set of first order cliques is

$$\mathcal{C}_1 = \{\{i\} \mid i \in \mathcal{S}\} \tag{22}$$

The set of second order cliques is

$$\mathcal{C}_2 = \{\{i, i'\} \mid i' \in \mathcal{N}_i, i \in \mathcal{S}\} \tag{23}$$

The single-site potential is defined as

$$V_1(f_i) = \begin{cases} v_{10} \text{ if } f_i = 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{24}$$

where v_{10} is a constant. If f_i is the NULL label, it incurs a penalty of v_{10}. The pair-site potential is defined as

$$V_2(f_i, f_{i'}) = \begin{cases} v_{20} \text{ if } f_i = 0 \text{ or } \ f_{i'} = 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{25}$$

where v_{20} is a constant. If either f_i or $f_{i'}$ is the NULL , it incurs a penalty of v_{20}. The definitions of the above prior potentials are a generalization of that penalizing line process variables [32,33]. The energy in the prior Gibbs distribution is then

$$U(f) = \sum_{i \in \mathcal{S}} V_1(f_i) + \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'}) \tag{26}$$

The conditional p.d.f., $p(d \mid f)$, of the observed data d, also called the *likelihood function* when viewed as a function of f given d fixed, has the following characteristics:

- It is conditioned on pure non-NULL matches $f_i \neq 0$,
- It is independent of the neighborhood system $\mathcal{N}$, and
- It depends on how the model object is observed in the scene which in turn depends on the underlying transformations and noise.

Assume (i) that D and d are composed of types of features which are invariant under the considered class of transformations [8]; (ii) that they are related via the observation model

$$d_1(i) = D_1(f_i) + e_1(i), \qquad d_1(i,i') = D_2(f_i, f_{i'}) + e_2(i,i') \tag{27}$$

where e_1 and e_2 are additive independent zero mean Gaussian noise vectors. The assumptions of the independent and Gaussian noise may not be accurate but offers an approximation when an accurate observation model is not available.

Then the likelihood function is a Gibbs distribution with the energy

$$U(d \mid f) = \sum_{i \in \mathcal{S}, f_i \neq 0} V_1(d_1(i) \mid f_i) + \sum_{i \in \mathcal{S}, f_i \neq 0} \ \sum_{i' \in \mathcal{S} - \{i\}, f_{i'} \neq 0} V_2(d_2(i,i') \mid f_i, f_{i'}) \tag{28}$$

where the constraints, $f_i \neq 0$ and $f_{i'} \neq 0$, restrict the summations to take over the non-NULL matches. The likelihood potentials are

$$V_1(d_1(i) \mid f_i) = \begin{cases} \sum_{k=1}^{K_1} [d_1^{(k)}(i) - D_1^{(k)}(f_i)]^2 / \{2[\sigma_1^{(k)}]^2\} & \text{if } f_i \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

and

$$V_2(d_2(i,i') \mid f_i, f_{i'}) = \begin{cases} \sum_{k=1}^{K_2} [d_2^{(k)}(i,i') - D_2^{(k)}(f_i, f_{i'})]^2 / \{2[\sigma_2^{(k)}]^2\} \\ \qquad \qquad \text{if } i' \neq i \& f_i \neq 0 \ \& \ f_{i'} \neq 0 \\ 0 \qquad \qquad \text{otherwise} \end{cases} \tag{30}$$

[8] The discovery and computation of visual invariants is an active area of research. See [34].

where $[\sigma_n^{(k)}]^2$ ($k = 1, \ldots, K_n$ and $n = 1, 2$) are the variances of the corresponding noise components. The vectors $D_1(f_i)$ and $D_2(f_i, f_{i'})$ are the "conditional mean vectors" of the random vectors $d_1(i)$ and $d_2(i, i')$, respectively.

Given d, $P(f \mid d)$ is a Gibbs distribution with the energy function

$$
\begin{aligned}
U(f \mid d) = & \textstyle\sum_{i \in \mathcal{S}} V_1(f_i) + \\
& \textstyle\sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'}) + \\
& \textstyle\sum_{i \in \mathcal{S}: f_i \neq 0} V_1(d_1(i) \mid f_i) + \\
& \textstyle\sum_{i \in \mathcal{S}: f_i \neq 0} \; \sum_{i' \in \mathcal{S} - \{i\}: f_{i'} \neq 0} V_2(d_2(i, i') \mid f_i, f_{i'})
\end{aligned}
\tag{31}
$$

The MAP labeling is equivalently found by minimizing $U(f \mid d)$.

The MAP configuration f^* derived in the above is the optimal mapping from the scene to the model object under consideration. In other words, it is the optimal labeling of the scene in terms of the model object. When there are L potential model objects, L MAP solutions, $f^{(1)}, \ldots, f^{(L)}$, can be obtained after matching the scene to each of the models in turn. However, any feature in the scene can have only one corresponding model feature. The match with the minimum local cost is chosen as the final one [20].

5 Verification and Pose Estimation

Suppose in the result of matching, there are a set of N matches between the scene and a certain model

$$
S = \{(p_k, P_k) \mid k = 1, ..., N\} \tag{32}
$$

where p_k and P_k denote 3D points in the scene and in the model, respectively. These correspondences are verified and clustered before used for pose estimation. The verification and clustering are performed for two reasons: First, there may be false matches in S and these should be identified for correct estimation of poses. Second, there may be more than one instance of a same model in the scene, each having a set of different transformation parameters, and they should be separated. The process of pose computation consists of three steps: (i) pose clustering to verify the matches and to find the maximal consistent subset(s) out from S, (ii) augmenting the consistent subset by adding more correspondences through interpolation and (3) pose estimation.

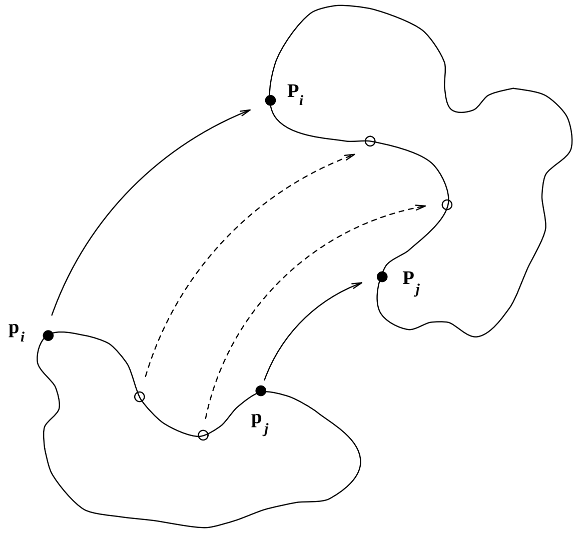

Fig. 3. The interpolation of matched correspondences. The filled dots are existing feature points and the hollow dots are interpolated ones. The solid arrows show two existing point correspondences and the dashed arrows are the interpolated ones.

5.1 *Pose Clustering*

The constraint for the pose clustering is the rigidity under $\mathcal{T}_s$. To check this, at least three pairs of corresponding points must be considered. In the exact sense, a triplet of point correspondences $((p_i, P_i), (p_j, P_j), (p_k, P_k))$ is consistent if

$$\frac{||p_i - p_j||}{||P_i - P_j||} = \frac{||p_j - p_k||}{||P_j - P_k||} = \frac{||p_k - p_i||}{||P_k - P_i||} \tag{33}$$

The above equations test the conditions for the two triangles (p_i, p_j, p_k) and (P_i, P_j, P_k) to be similar [35]. To deal with inaccuracy and noise, lower and upper bounds (0.8 and 1/0.8 in our implementation) are used in testing the above conditions. The test is done for all $1 \leq i \leq N-2$, $i+1 \leq j \leq N-1$ and $j+1 \leq k \leq N$, yielding a set of pose candidates. Best pose candidates are selected using a clustering algorithm. This finds the maximally consistent subset(s). The computation in this step is exhaustive search over all combinations of n correspondences. Because N is usually not big, the exhaustive search is acceptable.

5.2 Interpolating Correspondences

The second step, interpolating correspondences augments the set of consistent correspondences. It in effect increases reliability and accuracy of the pose estimation. Refer to Fig.3. Let (p_i, P_i) and (p_j, P_j) be two consistent correspondences. These define two consistent curve segments $\widehat{p_i p_j}$ and $\widehat{P_i P_j}$. From the two corresponding curve segments, many corresponding point pairs can be derived by sampling the segments in equal arc-length. The n points u_k used in the precedent derivation of the invariant coordinates can be re-used for this purpose.

The corresponding points on the two segments form the interpolated correspondences. The interpolation is useful for increasing reliability and accuracy only when the two segments are not straight lines; otherwise, the interpolated pairs are linearly dependent. In this way, two corresponding curve segments suffice to determine the pose uniquely because any number of point pairs can be obtained when necessary by using the interpolation.

The interpolation of the corresponding feature points can also help verify the correctness of curve matching. If two correspondences (p_i, P_i) and (p_j, P_j) happen to be good in terms of the feature points but not of the curve segments, then they may be accidental false matches. Such matches cannot be detected by using the previous pose clustering method. However, the residual error of pose computed by using post-interpolation correspondences will be significantly larger than that by using pre-interpolation correspondences. By checking this, these false matches can be detected.

5.3 Pose Estimation

In the third step, the pose is estimated by using the least squares (LS) method performed on the augmented set of the corresponding points after the interpolation. The LS solution can be obtained in close-form using the singular value decomposition method [36–38].

Under $\mathcal{T}_s$, a point p_k in a model is transformed into the corresponding point P_k in the scene according to the following

$$p_k = sRP_k + T \tag{34}$$

where s is a scale factor, R is a 3×3 orthogonal rotation matrix and T is a 3×1 translation vector. Assume there are a number of n correspondences after the interpolation. The pose estimate (s, R, T) minimizes the following weighted squared error

$$e^2(R,T,s) = \sum_{k=1}^{n} w_k \|p_k - sRP_k - T\|^2 \tag{35}$$

where $w_k \geq 0$ are normalized weights, $\sum_{k=1}^{n} w_k = 1$, which may be computed based on some matching confidence measure or simply set to $1/n$. The minimal solution can be obtained in close-form using the singular value decomposition method [36–38]. Let

$$\bar{p} = \sum_{k=1}^{n} w_k p_k \ , \qquad \bar{P} = \sum_{k=1}^{n} w_k P_k \tag{36}$$

be 3×1 weighted mean vectors,

$$\sigma_p^2 = \sum_{k=1}^{n} w_k \|p_k - \bar{p}\|^2 \ , \qquad \sigma_P^2 = \sum_{k=1}^{n} w_k \|P_k - \bar{P}\|^2 \tag{37}$$

be scalar variances, and

$$H = \sum_{k=1}^{n} w_k (p_k - \bar{p})(P_k - \bar{P})^T \tag{38}$$

be a 3×3 covariance matrix. Let the singular value decomposition of H be

$$H = UDV^T \tag{39}$$

where U and V are orthogonal and D is diagonal. When $rank(H) \geq 3$, the minimum squared error estimate of pose is uniquely determined as follows:

$$R = UV^T \ , \qquad s = tr(D)/\sigma_P^2 \ , \qquad T = \bar{p} - sR\bar{P} \tag{40}$$

where $tr(D)$ is the trace of the matrix D.

6 Results

The following simulations demonstrate the ability of the present approach to perform 3D curve match under the conditions of noise, missing and extra feature points, and unsegmented curves. The model base consists of sixteen simulated model curves (Figure 4).

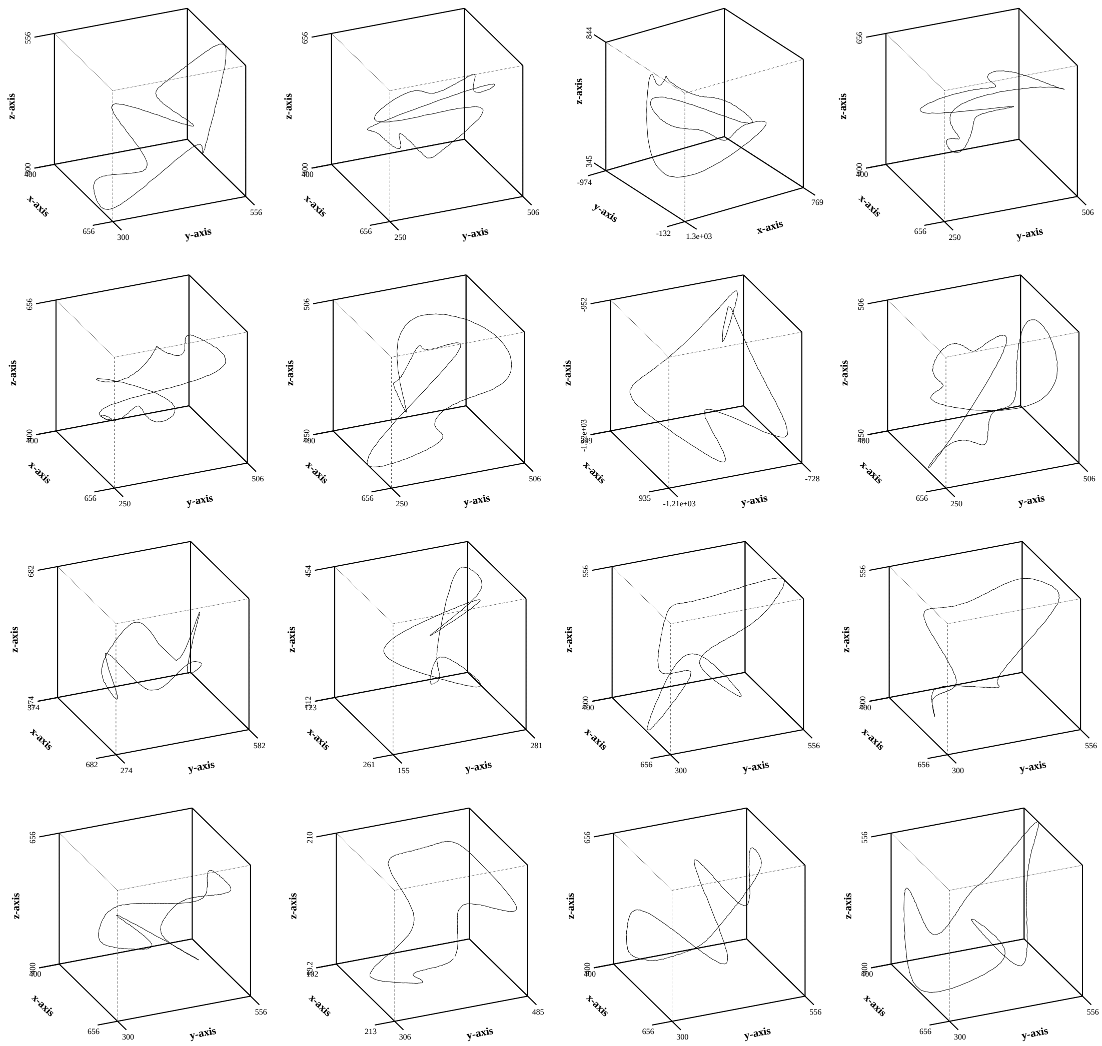

Fig. 4. Model curves 1 through 16 (from left to right, top to bottom).

The first simulation is illustrated in Fig.5. The model curve with its feature points is shown in (a). The observed scene curve (b) is a partial and noisy version of (a). Feature points are detected after the scene curve is smoothed. Note that due to the partialness and noise, there are missing and extra feature points in (c) as compared to the model curve in (a). Despite this, the two curves are still successfully matched and the pose found as in (d). Both the estimated pose and the true pose are shown in Table 1 for comparison. In this simulation, the true orthogonal rotation matrix is purposely set to the identity matrix so that the missing and extra feature points can be observed clearly in the plots (a) and (c) of Fig.5. However, the information about the scaling, rotation and translation is not known to the system *a priori*.

The second simulation is illustrated in Fig.6. The observed curve consists of two connected partial curves. In (a) is a simulated noisy scene curve. It was generated by (i) rotating, translating and scaling sub-parts of some of the

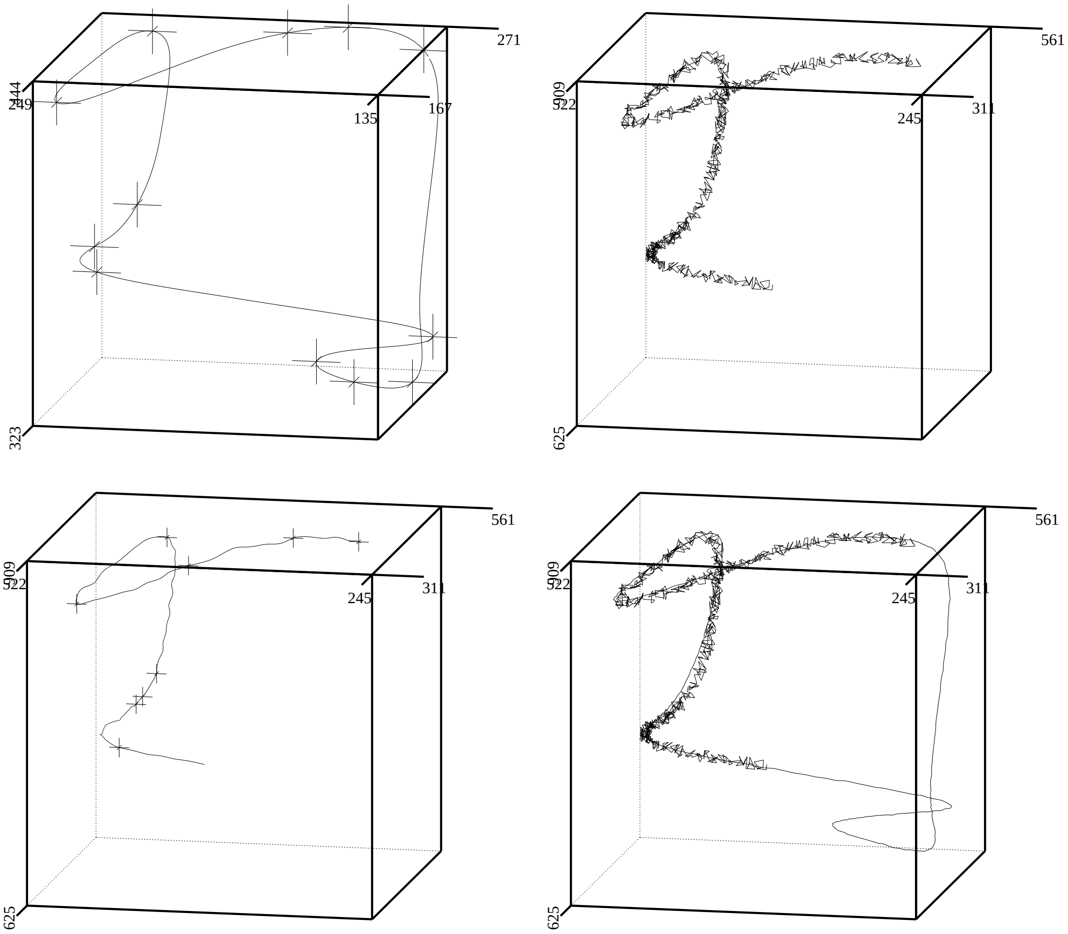

Fig. 5. Matching with missing and extra feature points. From left to right, top to bottom are: (a) the model curve and the feature points; (b) a partial and very noisy version of the model curve in the scene; (c) the feature points detected from the smoothed scene curve (note that feature points cannot be detected from the unobserved part and extra feature points emerge due to noise); (d) the observed curve is matched and align with the model curve.

model curves, (ii) concatenating the transformed sub-curves and (iii) adding uniformly distributed noise. The detected corner points and the inserted mid-points for this curve were shown in Fig.1. In (b)-(d), the matched model curves (the smooth ones) aligned to the (noisy) scene curve are shown. The numbers of matched correspondences given by the matching stage were 5, 7 and 10 for the three curves, respectively. Table 2 gives the estimated poses as compared to the real poses. The errors are smaller than those in the first simulation because the noise level in the observed curve is lower in this simulation.

The interpolated information is very useful for the matching and pose estimation. The matching would have failed if only information about the feature points were used ; the accuracy of pose estimation is significantly improved with the interpolation information. The residual errors of estimation is reduced as the number of interpolation matches is increased. Table 3 numerically illustrates the errors with 0 (no interpolation), 3 and 7 interpolated matches

	True Pose			Estimated Pose		
s	2.0			2.041		
sR	2	0	0	2.036	0.139	-0.003
	0	2	0	-0.139	2.036	-0.011
	0	0	2	0.003	0.011	2.041
T	0	0	0	21.208	-40.189	-13.511

Table 1
The true and estimated poses for the matching with missing and extra feature points.

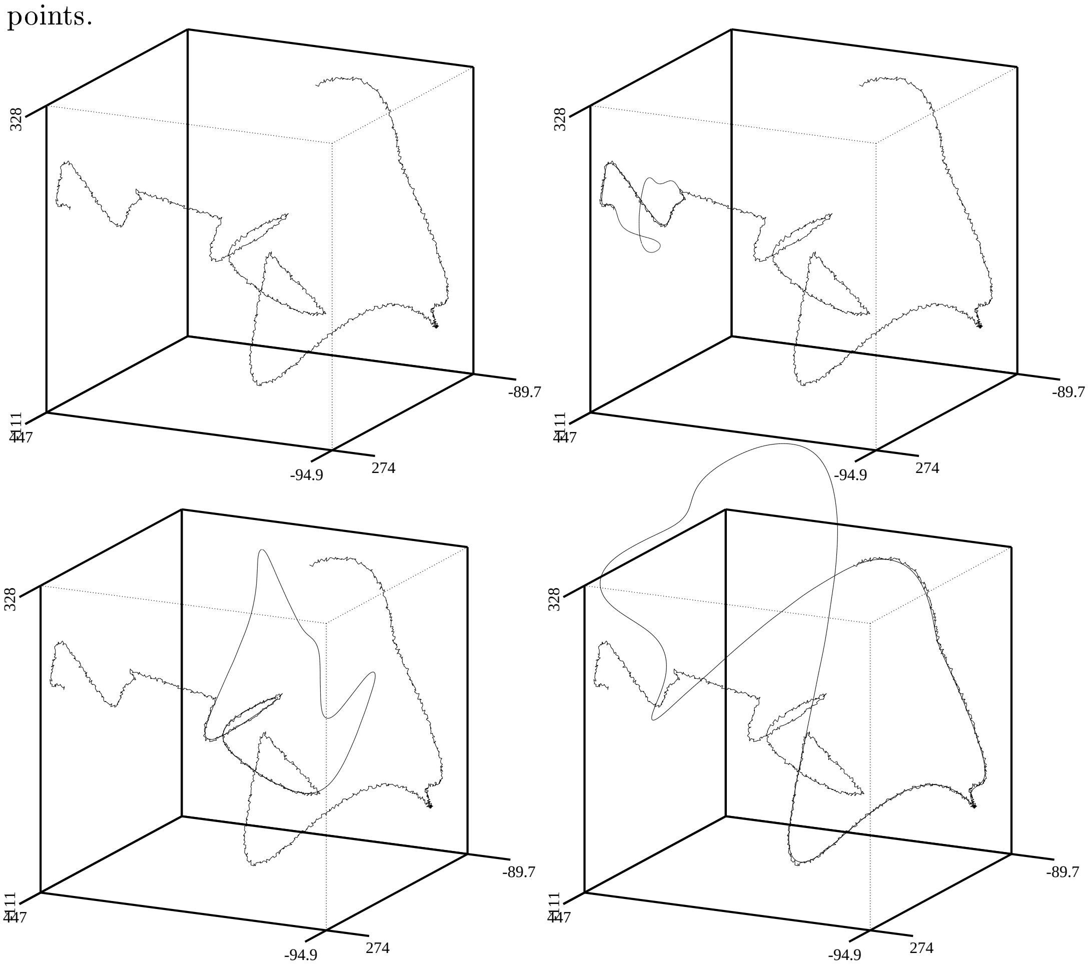

Fig. 6. Matching with unsegmented curve data. From left to right, top to bottom are: (a) the scene curve, (b) matched curve 1 (which is model curve 6) aligned with the scene curve, (c) matched curve 2 (model curve 16) and (d) matched curve 3 (model curve 2).

(between each pair of consecutive matches) for the above three matched curves. The improvement of the pose accuracy is visually demonstrated in Fig.7. There, the input to the pose estimation procedure is the set of correspondences for Fig.6(d) but with the location of some of the points deliberately distorted to deteriorate the quality of the set. The result on the left is obtained

	True Pose			Estimated Pose		
s	0.5			0.496556		
sR	0.463	0.082	-0.171	0.459	0.070	-0.177
	0.010	-0.461	-0.193	-0.001	-0.461	-0.185
	-0.189	0.175	-0.429	-0.190	0.171	-0.426
T	200.00	300.00	500.00	205.61	307.01	499.86
s	1.0			0.978958		
sR	0.866	0.000	-0.500	0.860	-0.003	-0.468
	0.374	0.664	0.647	0.350	0.656	0.637
	0.332	-0.748	0.575	0.312	-0.727	0.577
T	-600.00	100.00	-100.00	-577.64	97.04	-115.76
s	2.0			2.001439		
sR	1.628	-0.940	0.684	1.607	-0.950	0.721
	0.923	0.330	-1.743	0.945	0.278	-1.742
	0.706	1.734	0.702	0.727	1.739	0.672
T	-1400.00	-500.00	100.00	-1410.67	-474.06	96.23

Table 2
The estimate poses as compared to the real poses. A pose is represented in a scale s, a scaled rotation matrix sR (3×3) and a translation vector T $(3 \times 1$ transposed). From the top to the bottom of the table are the poses for the three matching curves. The true pose and the estimated one are shown for each case. (Top) for matched curve 1, (middle) for matched curve 2, and (bottom) for matched curve 3.

	Curve 1	Curve 2	Curve 3
0	1.800387	1.516043	3.320574
3	1.780630	1.072547	2.603601
7	0.990901	1.002981	2.290454

Table 3
Residual error of estimation vs. number of interpolated points.

using the matche points only. The pose on the right is estimated by including 7 interpolated matches between each consecutive pair of distorted matches . The improvement in pose accuracy is significant.

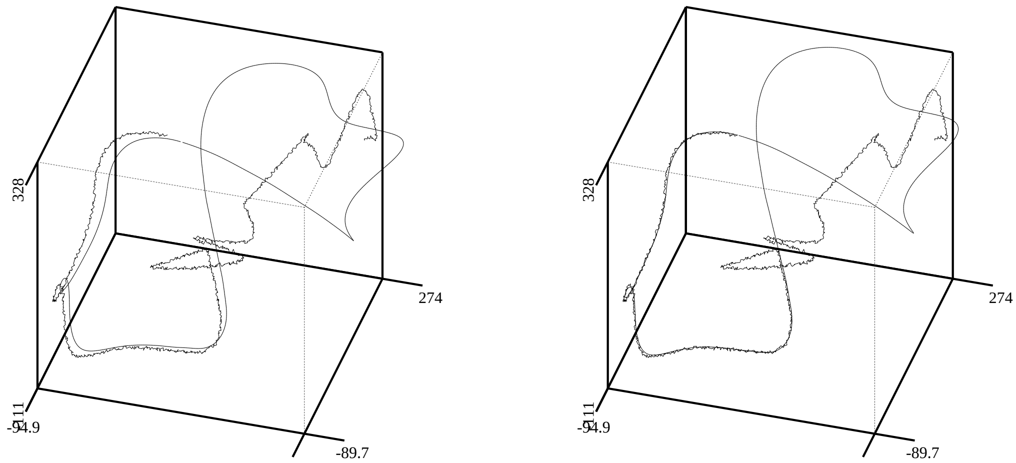

Fig. 7. The effect of using interpolated information on pose estimation. Left: pose estimated without using the interpolation. Right: the pose improved by using the interpolation.

7 Conclusion

A system has been presented for the representation, matching and pose estimation of 3D space curves under the similarity transformation with missing and extra curve features. The main contributions are (i) the two level similarity invariant curve representation based on the similarity invariant coordinate system (SICS), and (ii) the idea of the extensive exploration of curve information for the matching and pose. The results show that the system is able to perform matching and pose in the presence of noise, with missing and spurious feature points, and with unsegmented curves. A limitation of the system is that in the scene, there should be at least one continuous curve segment containing at least two feature points. Feature-less curves cannot be handled.

References

[1] P. J. Besl and R. C. Jain. "Three-Dimensional object recognition". *Computing Surveys*, 17(1):75–145, March 1985.

[2] M. Brady, J. Ponce, A. Yuille, and H. Asada. "Describing surfaces". *Computer Vision, Graphics and Image Processing*, 32:1–28, 1985.

[3] R. C. Bolles and P. Horaud. "3DPO: A three-dimensional part orientation system". *International Journal of Robotic Research*, 5(3):3–26, Fall 1986.

[4] O. D. Faugeras and M. Hebert. "The representation, recognition and locating of 3D objects". *International Journal of Robotic Research*, 5(3):27–52, Fall 1986.

[5] J. T. Schwartz and M. Sharir. "Identification of partially obscured objects in two and three dimensions by matching noisy characteristic curves". *International Journal of Robotic Research*, 6(2):29–44, 1987.

[6] R. E. Sampson. "3D range sensor-phase detection". *Computer*, 20:23–24, 1987.

[7] O. Monga, R. Deriche, and J.M. Rocchisani. "3D edge detection using recursive filtering: application to scanner images". *CVGIP: Image Understanding*, 53(1):76–87, 1991.

[8] L. Robert and O.D. Faugeras. "Curve-based stereo: Figural continuity and curvature". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 57–62, Maui, HI, 1991.

[9] A. Gueziec and N. Ayache. "Smoothing and matching of 3-d curves. *International Journal of Computer Vision*, 12(1):79–104, 1994.

[10] T. Joshi, J. Ponce, B. Vijaykumar, and D.J. Kriegman. "HOT curves for modelling and recognition of smooth curved 3d objects". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 876–880, Los Alamitos, CA, 1994.

[11] N. Ayache and O. D. Faugeras. "HYPER: A new approach for the representation and positioning of two-dimensional objects". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):44–54, January 1986.

[12] Y. Lamdan, J.T. Schwartz, and H.J. Wolfson. "Object recognition by affine invariant matching". In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 335–344, 1988.

[13] W.E.L. Grimson. On the recognition of curved objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:632–643, 1989.

[14] N. Ansari and E.J. Delp. "Partial shape recognition: A landmark-based approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):470–483, May 1990.

[15] H.-C. Liu and M. D. Srinath. "Partial shape classification using contour matching in distance transformation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1072–1079, November 1990.

[16] H. J. Wolfson. "On curve matching". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):483–489, May 1990.

[17] A. M. Bruckstein, N. Katzir, M. Lindenbaum, and M. Porat. "Similarity-invariant signatures for partially occluded planar shapes. *International Journal of Computer Vision*, 7:271–285, 1992.

[18] P. J. Besl and N. D. McKay. "A method for registration of 3-D shapes". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.

[19] C.A. Rothwell, A. Zisserman, A. Forsyth, and J.L. Mundy. "Canonical frames for planar object recognition". In *Proceedings of the European Conference on Computer Vision*, pages 757–772, 1992.

[20] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.

[21] F. Mokhtarian and A. Mackworth. "Scale-based description and recognition of planar curves and two-dimensional shapes". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:34–44, 1986.

[22] D.G. Lowe. "Organization of smooth image curves at multiple scales". In *Proceedings of Second International Conference on Computer Vision*, pages 558–567, 1988.

[23] A. Rattarangsi and R. T. Chin. "Scale-based detection of corners of planar curves". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(4):430–449, April 1992.

[24] M. Fischler and R. Elschlager. "The representation and matching of pictorial structures". *IEEE Transactions on Computers*, C-22:67–92, 1973.

[25] A. P. Ambler, H. G. Barrow, C. M. Brown, R. M. Burstall, and R. J. Popplestone. "A versatile computer-controlled assembly system". In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 298–307, 1973.

[26] J. K. Cheng and T. S. Huang. "Image registration by matching relational structures". *Pattern Recognition*, 17(1):149–159, 1984.

[27] B. Radig. "Image sequence analysis using relational structures". *Pattern Recognition*, 17(1):161–167, 1984.

[28] S. Z. Li. " Towards 3D vision from range images: An optimization framework and parallel networks". *CVGIP: Image Understanding*, 55(3):231–260, May 1992.

[29] S. Z. Li. "Matching: invariant to translations, rotations and scale changes". *Pattern Recognition*, 25(6):583–594, June 1992.

[30] J. Besag. "Spatial interaction and the statistical analysis of lattice systems" (with discussions). *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.

[31] J. M. Hammersley and P. Clifford. "Markov field on finite graphs and lattices". unpublished, 1971.

[32] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[33] J. L. Marroquin. *Probabilistic Solution of Inverse Problems.* PhD thesis, MIT AI Lab, 1985.

[34] J. L. Mundy and A. Zisserman, editors. *Geometric Invariants in Computer Vision.* MIT Press, Cambridge, MA, 1992.

[35] G. Stockman. "Object recognition and localization via pose clustering". *Computer Vision, Graphics and Image Processing*, 40:361–387, 1987.

[36] K.S. Arun, T.S. Huang, and S.D. Blostein. "Least squares fitting of two 3-D point sets". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:698–700, 1987.

[37] R. M. Haralick, H. Joo, C.N. Lee, X. Zhuang, V.G. Vaidya, and M.B. Kim. " Pose estimation from corresponding point data". *IEEE Transactions on Systems, Man and Cybernetics*, 19:1426–1446, 1989.

[38] S. Umeyama. "Least squares estimation of transformation parameters between two point patterns". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:376–380, 1991.

Stan Z. Li received the B.Sc degree from Hunan University, China, in 1982, M.Sc degree from the National University of Defense Technology, China, in 1985 and Ph.D degree from the University of Surrey, UK, in 1991. All degrees are in EEE. He is currently a lecturer at Nanyang Technological University. His research interests include computer vision, image processing, pattern recognition and optimization methods.

Mrs. Blaire V. Mossman
Managing Editor
Pattern Recognition
Scottsdale
AZ 85267-3177, USA