Chapter 1

PROJECTIVE JOINT INVARIANTS FOR MATCHING CURVES IN CAMERA NETWORKS

Raman Arora
Department of Electrical Engineering
University of Washington
Seattle, WA 98105
rmnarora@u.washington.edu

Charles R. Dyer
Department of Computer Sciences
University of Wisconsin-Madison
Madison, WI 53706
dyer@cs.wisc.edu

Abstract A novel method is presented for distributed matching of curves across widely varying viewpoints. The fundamental projective joint-invariants for curves in the real projective space are the volume cross-ratios. A curve in m-dimensional projective space is represented by a signature manifold comprising n-point projective joint invariants, where n is at least $m + 2$. The signature manifold can be used to establish equivalence of two curves in projective space. However, without correspondence information, matching signature manifolds is a computational challenge. Our approach in this chapter is to first establish best possible correspondence between two curves using sections of the invariant signature manifold and then perform a simple test for equivalence. This allows fast computation and matching while keeping the descriptors compact. The correspondence and equivalence of curves is established independently at each camera node. Experimental results with simulated as well as real data are provided.

Keywords: Camera networks, curve matching, projective invariants, joint invariant signatures, cross-ratios

1. Introduction

Object recognition in automated visual surveillance systems must be capable of matching features which represent distinctive parts of objects (such

as people or vehicles) in complex environments in an online fashion across multiple view-points. Commercial, law enforcement, and military applications abound, including detection of loiterers, monitoring vehicles on highways, patrolling borders, measuring traffic flow, counting endangered species, and activity monitoring in airports. As costs for cameras and computers continue to drop while the desire for security and other applications increases, research in this area has been developing rapidly over the last decade [3; 5; 9; 11; 20]. Matching curves across widely varying viewpoint requires local image features that are invariant to changes in pose, occlusion, illumination, scale, and intrinsic differences between cameras.

This chapter describes a method that uses projective joint invariants to match curves across multiple views. Given a pair of images taken from unknown viewpoints, a set of curves is extracted from each image that are projections of unknown 3D curves in the scene. The objective is to determine if any two given curves, one from each image, match, i.e. if they correspond to the same curve in the scene.

An invariant is defined to be a function on the set of points (or a subset) of an image, of a planar or 3D object, that remains constant under a collection of transformations of the object. Two images (or sub-images) with the same values of the invariant are identified as images of the same object under a transformation, thereby making the problem of multiple hypothesis detection direct. Due to the utility of transformation-invariant features in their ability to reduce the set of possible matches and speed up the search for similar classes or objects, invariant-based approaches to problems in computer vision have been well studied [16; 17].

Invariant-based methods may be classified as global or local: global invariants utilize the entire image to compute feature values whereas local invariants are computed from much smaller subsets. Local invariants are more desirable due to their robustness to occlusion and noise. However, one of the fundamental problems with the use of local invariants is that they must be computed on corresponding subsets of points in each view.

Related Work

Projective invariants have been applied to various computer vision tasks such as localization [13; 21], autonomous navigation [26], 3D reconstruction [24], and surveillance [27]. A few researchers have focused on the probabilistic analysis of projective invariants. In [2], a probability distribution was derived for the four-point cross-ratio, a classical planar projective invariant, under various assumptions on the distribution of the four points. The distribution of cross ratios was further examined in [10] with constraints on relative distances of the four points. The performance of cross ratios was described quantitatively in terms of probability of rejection and false alarm in [15]. Unfortunately, in all these works, the correspondence of points between images was given a priori

or external makers were used to assist with the correspondence. Without correspondence information, the classification methodology breaks down since the cross ratios are not unique.

In other related work, Scale-invariant feature transform (SIFT) [12] was used to compute a large set of local feature vectors from each image and the correspondence between two images was established using RANSAC [6]. This approach has been used for 3D model reconstruction in the Photo Tourism system [25]. However, the computational complexity of SIFT and RANSAC makes it difficult to use for real-time video surveillance applications with non-overlapping field-of-view camera networks.

Rothwell et. al. [22; 23] presented an approach for planar object recognition by constructing a canonical frame for determining projectively invariant indexing functions for planar curves. The idea is to identify four distinguished points on the curve and then compute the projective transformation that maps these points to the four corners of a square. The distinguished points are chosen using tangency conditions that are preserved under projective transformations. Similar ideas have been put forth by Hann and Hickman [7; 8] and by Orrite and Herrero [19]. These methods also utilize bi-tangency points on curves to learn a projective map. The key difference in the more recent work [7; 8; 19] from the algorithms presented in mid-nineties [22; 23] is that they learn the best projective transformation between two given planar curves in an iterative fashion whereas the earlier work focussed on solving for a projective transformation that best maps the bi-tangency points to four corners of the unit square.

There are several shortcomings of existing methods for matching curves across multiple viewpoints that preclude their deployment to applications like video surveillance in camera networks. The methods based on learning projective transformations [7; 8; 19; 22; 23] between given curves are inherently centralized and computationally expensive. In order to deal with differences in sampling of images (resulting from different grids in imaging devices), existing methods employ an iterative scheme where the learnt projective transformation is corrected-for based on the resulting mismatch in the image domain. For specialization to matching curves in video streams, this will require complete exchange of images at each iteration and repeated estimation of projective transformations. Furthermore, the methods depend on the ability to consistently identify bi-tangents. But, due to possible occlusions, the curves extracted from the images may not admit any bi-tangents. Finally, the methods based on detecting interest points and representing images in a visual dictionary obtained by clustering SIFT descriptors [4] are inherently offline. In applications such as video surveillance (where the object to be matched may be moving), these methods require frame synchronization across video feeds from different cameras as well as dictionary computation and exchange every few frames.

Our Approach

We present a method for matching curves in different views without assuming any knowledge of the relative positions and orientations of the viewpoints. Our approach is based on the computation and comparison of projective invariants that are expressed as volume cross ratios of curves extracted from images of an arbitrary 3D scene. Signatures based on these cross ratios are computed from each image and a clustering-based method is presented for distributed matching. This work was inspired by recent advances in joint invariants [18] and probabilistic analysis of random cross ratios.

Joint invariant signatures were recently studied by Olver [18] for various transformation groups. However, due to the sheer size and global nature of the signatures, they cannot be directly employed for curve-matching. The novel ideas in this chapter include generating compact local signatures independently from each image and clustering-based matching. We systematically reduce the size and computational complexity of the matching by reformulating the problem and offering a tradeoff between the size of feature space and size of the search space for registration parameters.

Our method alleviates the aforementioned shortcomings of existing methods. Unlike existing methods that match curves in the image domain, the proposed method matches curves in an invariant domain. The classification rule is based on comparing the projective invariants of a given pair of curves. The invariants are known to be complete and therefore uniquely represent the corresponding curve. The joint-invariants are also robust to noise as a small perturbation of points on the curve results in a small relative error in the invariant domain [1]. The matching of two curves can be performed efficiently in the presence of bi-tangents on the given curves. It should be remarked that unlike previous work [7; 8; 19; 22; 23], the proposed method does not critically depend on the existence of bi-tangents. However, whenever bi-tangents are present, our approach utilizes them to speed up the search for registration parameters.

It is important to note that a description of a curve using projective joint invariants remains invariant to Euclidean transformations as well. Therefore, in video surveillance applications, the representation is redundant across frames of the video-feed. when the object undergoes rigid-body motion. This saves network and computational resources and allows for robust matching of curves between two cameras without frame synchronization.

The chapter is organized as follows. Section 2 describes the curve matching problem in a multiview setting and briefly presents mathematical preliminaries in projective joint invariants. Section 3 describes the joint invariant signature and challenges associated with distributed matching of signature manifolds. In Section 4, we discuss the use of sub-manifolds and local signatures to establish correspondence and equivalence of curves across view-points. Section 5 presents experimental results with simulated as well as real datasets.

Table 1.1: Notation Table

Symbols	Description
C_i	i^{th} camera or i^{th} viewpoint
S_{ij}	j^{th} continuous curve at viewpoint C_i
I	an interval of $\mathbb{R}$
D_{ij}	j^{th} discrete curve j in image plane of camera C_i
G	transformation group
M	manifold
$\mathbf{F}^{(n)}$	$n-$point joint-invariant vector-function $[\mathbf{F}_1^{(n)}, \ldots, \mathbf{F}_l^{(n)}]$
$V(z_i, z_j, z_k)$	area of the triangle with vertices z_i, z_j, z_k
$CR(z_1;\ z_2, z_3, z_4, z_5)$	volume cross-ratio given in eqn. (5)
S_{ij}^n	$n-$times Cartesian product of the curve S_{ij}
J_{ij}	invariant signature manifold associated with curve S_{ij}
d_{kl}^{ij}	distance function used for matching; see eqn. (10)
t^*	pivot points $(t_1^*, t_2^*, \ldots, t_{n-p}^*)$
π	a permutation of n-points
$\mathcal{A}_{t^*,\pi}$	slice of the signature manifold determined by t^*, π
$\mathcal{U}_{t^*,\pi}$	a section of the slice $\mathcal{A}_{t^*,\pi}$

2. Problem Formulation and Preliminaries

This section introduces the notation and presents problem formulation for pairwise curve matching across different viewpoints. Let C_i represent an unknown view-point or camera location. Let S_{ij} denote the j^{th} continuous planar-curve observed at viewpoint C_i. The curve S_{ij} is obtained from a space curve S under an unknown projective transformation. Formally, S_{ij} is defined to be a parametric curve

$$\begin{aligned} S_{ij} : I_{ij} &\rightarrow \mathbb{R}^m \\ t &\mapsto S_{ij}(t), \end{aligned} \tag{1}$$

where $I_{ij} \subset \mathbb{R}$ is a real interval and m is the dimensionality of the ambient Euclidean space for the curve. For $t \in T$, $S_{ij}(t) = [x_1(t) \ \cdots \ x_m(t)]^T$ gives coordinates in $\mathbb{R}^m$ of the corresponding point on the curve. For planar curves, $m = 2$ and for space curves $m = 3$. The n-dimensional Cartesian product of the curve S_{ij} is written as,

$$\begin{aligned} S_{ij}^n : I_{ij}^n &\rightarrow \mathbb{R}^{m \times n} \\ t^n &\mapsto S_{ij}^n(t^n), \end{aligned} \tag{2}$$

where $I_{ij}^n \subset \mathbb{R}^n$ is an n-dimensional interval, $t^n = (t_1, \ldots, t_n)$ and $S_{ij}^n(t^n) = (S_{ij}(t_1), \ldots, S_{ij}(t_n))$. For a pair of curves S_{ij}, S_{kl} observed at viewpoints

C_i, C_k respectively, the objective is to determine if the two curves in the image space at the two unknown viewpoints represent the same space curve in the observed scene. The sampled, discretized version of S_{ij} is denoted as D_{ij}.

Lets denote the manifold associated with the given curve as M. A given curve may undergo various transformations like rotation, translation, scaling and projection. These transformation can be described as a Lie group, denoted G, acting on the manifold M. The *joint action of a group on a manifold* describes how the group transforms any given n-tuple on the manifold. Formally, the joint action of the group G on the Cartesian product M^n is a map $(G \times M^n) \rightarrow M^n$ given as:

$$g \cdot (z_1, \ldots, z_n) = (g \cdot z_1, \ldots, g \cdot z_n), \tag{3}$$

for $g \in G$ and $\mathbf{z} = (z_1, \ldots, z_n) \in M^n$. With a slight abuse of notation we use z_i to represent a point on the manifold $M \subseteq \mathbb{R}^m$ as well as its Euclidean coordinates in the ambient space $\mathbb{R}^m$. An $n-$*point joint invariant* of the transformation group G on M^n is defined to be a function

$$\begin{aligned} \mathbf{F}^{(n)} : \mathbb{R}^{m \times n} &\rightarrow \mathbb{R}^l \\ \mathbf{z} &\mapsto [\mathbf{F}_1^{(n)}(\mathbf{z}), \ldots, \mathbf{F}_l^{(n)}(\mathbf{z})], \end{aligned} \tag{4}$$

that is invariant to the joint action of the group on the manifold: $\mathbf{F}^{(n)}(g \cdot \mathbf{z}) = \mathbf{F}^{(n)}(\mathbf{z})$.

The projective transformation group, $G = \text{PSL}(m+1, \mathbb{R})$, which is the subject of study in this chapter, acts on the projective space $\mathbb{RP}^m$ as $w = g \cdot z = \frac{Az+b}{c \cdot z+d}$, where A is an $m \times m$ matrix, b, c are $m \times 1$ vectors, and d is a scalar. The transformation g maps the point $z \in \mathbb{R}^m \subset \mathbb{RP}^m$ to $w \in \mathbb{R}^m$.

For planar curves ($m = 2$) consider $n = 5$. The fundamental invariants for 5-point joint action of the projective transformation group $\text{PSL}(3, \mathbb{R})$,are given by the volume cross ratios [18]:

$$CR(z_1;\ z_2, z_3, z_4, z_5) = \frac{V(z_1, z_2, z_3)V(z_1, z_4, z_5)}{V(z_1, z_2, z_5)V(z_1, z_3, z_4)}, \tag{5}$$

and

$$CR(z_2;\ z_1, z_3, z_4, z_5) = \frac{V(z_1, z_2, z_3)V(z_2, z_4, z_5)}{V(z_1, z_2, z_5)V(z_2, z_3, z_4)}, \tag{6}$$

where $V(z_i, z_j, z_k)$ is the area of the triangle defined by z_i, z_j and z_k. The cross ratio defined in (5) is described as the ratio of the product of the areas of the non-shaded triangles in Figure 1.1 and the product of areas of the shaded triangles. Therefore, for $\text{PSL}(3, \mathbb{R})$, $\mathbf{F}^{(n)}$ is given as

$$\mathbf{F}^{(n)}(\mathbf{z}) = [\ CR(z_1;\ z_2, z_3, z_4, z_5),\ CR(z_2;\ z_1, z_3, z_4, z_5)\ ]. \tag{7}$$

For $m = 3$, we consider 6-point joint invariants. There are three fundamental volume cross-ratios: $CR(z_1, z_6;\ z_2, z_3, z_4, z_5)$, $CR(z_1, z_3;\ z_2, z_4, z_5, z_6)$,

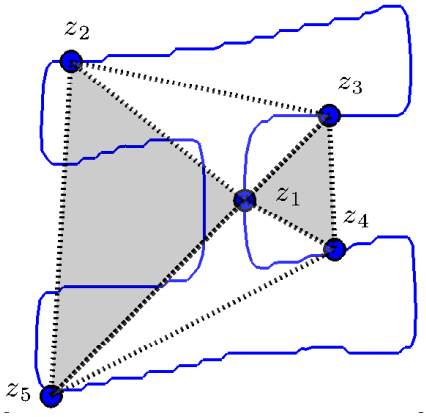


Figure 1.1: The five-point projective joint-invariant is the ratio of the product of areas of the non-shaded triangles and the product of areas of the shaded triangles.

and $CR(z_3, z_6;\ z_1, z_2, z_4, z_5)$. Geometrically, $CR(z_1, z_6;\ z_2, z_3, z_4, z_5)$ is the ratio of the volumes of four tetrahedrons:

$$CR(z_1, z_6;\ z_2, z_3, z_4, z_5) = \frac{V(z_1, z_2, z_3, z_4)V(z_2, z_3, z_5, z_6)}{V(z_1, z_2, z_4, z_5)V(z_3, z_4, z_5, z_6)}. \qquad (8)$$

Fig. 1.1 shows a bird's eye view of the double pyramid with common base resulting from the union of the tetrahedrons [18].

The probabilistic analysis of random five point cross-ratios reveals that no single cross-ratio is unique on smooth manifolds [1]. Consequently, for planar curves, the matching schemes based on comparing single cross-ratios are not discriminative. The six-point joint-invariants for 3D space curves lend themselves to the same analysis as the empirical distributions are found to exhibit characteristics similar to the planar case. Furthermore, it is argued using jitter-analysis that the cross-ratios are robust to noise. For more details the reader is referred to [1].

3. Joint Invariant Signatures

The non-uniqueness of any single cross ratio value implies that no single cross ratio value can be used for matching without establishing correspondence of points [1]. However, the joint invariant signature defined to be the manifold comprising cross-ratio values generated by *all possible* n-point sets on a curve represents the curve uniquely up to a projective transformation [18].

Let $\mathbf{F}^{(n)}$ be the n-point joint invariant map for the projective transformation group PSL$(3,\mathbb{R})$ given in eqn. (7). Consider the composition maps $J_{ij} = \mathbf{F}^{(n)} \circ S^n_{ij}$,

$$\begin{aligned} J_{ij} : I^n &\rightarrow \mathbb{R}^l \\ t^n &\rightarrow \mathbf{F}^{(n)}(S^n_{ij}(t^n)). \end{aligned} \tag{9}$$

The *invariant signature manifold* at viewpoint C_i for j^{th} curve is defined to be $J_{ij}(I^n_{ij})$. We now focus on restriction of the curves to a common domain. Consider $I \subseteq I_{ij} \cap I_{kl}$. Note that the intervals I_{ij}, I_{kl} can be translated, flipped and scaled appropriately so that $I \neq \emptyset$. We denote the restriction of the curves to the common interval I by $\tilde{S}_{ij}$ and $\tilde{S}_{kl}$.

Now, if the curves $\tilde{S}_{ij}, \tilde{S}_{kl}$ are related by a projective transformation, i.e., $\tilde{S}_{ij} = g \cdot \tilde{S}_{kl}$ for some $g \in G =$ PSL$(3,\mathbb{R})$ on interval $I \subset \mathbb{R}$, then from the definition of joint action, $\tilde{S}^n_{ij} = g \cdot \tilde{S}^n_{kl}$. This implies that the invariant signature manifold for the two curves coincide: For all $t^n \in I^n$,

$$J_{ij}(t^n) = (\mathbf{F}^{(n)} \circ \tilde{S}^n_{ij})(t^n) = (\mathbf{F}^{(n)} \circ (g \cdot \tilde{S}^n_{kl}))(t^n) = (\mathbf{F}^{(n)} \circ \tilde{S}^n_{kl})(t^n) = J_{kl}(t^n).$$

More importantly, $\tilde{S}_{ij} = g \cdot \tilde{S}_{kl}$ for some $g \in G$ if $J_{ij}(I^n) = J_{kl}(I^n)$[18]. Therefore, two curves are equivalent up to a projective transformation *if and only if* their signature manifolds coincide. Consider the following distance function that measures the degree of mismatch between two curves in terms of the mismatch in the associated invariant signature manifolds,

$$d^{ij}_{kl}(U) \equiv \int_U ||J_{ij}(t^n) - J_{kl}(t^n)||^2_2 \, dt^n, \;\; U \subseteq I^n, \tag{10}$$

where $||\cdot||_2$ represents the standard Euclidean norm. For robustness to noise in observation, we adopt the following test for matching two curves,

$$d^{ij}_{kl}(I^n) < \epsilon, \tag{11}$$

where threshold $\epsilon > 0$ depends on the amount of noise resulting from differences in quantization and the sampling grids of two cameras. The matching criterion in (11) is straightforward if the two curves are defined on the same domain and the true correspondence between the curves is known. However, without the correspondence information between the two curves there is no direct method to compute the distance function in (10). This problem is further compounded by various practical aspects of matching in a camera network. First and foremost, the curves observed at each camera are discrete curves (D_{ij} obtained from sampling S_{ij}). We assume that the curves are sampled dense enough to admit a good approximation in an interpolating basis. Secondly, due to differences in the sampling grids for different cameras, the discretization of the interval I differs for each camera. Finally, the size of signature manifold

associated with each curve grows exponentially with the number of samples on the curve. Therefore, estimating correspondence between the two curves using entire signature manifolds is not a feasible solution due to computational constraints. In this chapter we first establish correspondence efficiently by restricting ourselves to certain sub-manifolds of the signature manifold and then test for equivalence by estimating d_{kl}^{ij} from given observations (D_{ij}, D_{kl}).

4. Toward Local Signatures

The joint invariant signature manifold is a global descriptor. Owing to the lack of robustness of global signatures to occlusion, we restrict our attention to sections of invariant signature sub-manifolds. We first define a slice of the signature manifold and then discuss local signatures computed on sections of slices.

Slices and Sections of Signature Manifold

A *slice* of the signature manifold may be generated by freezing one or more coordinate directions of I^n (at pivot points denoted as t^*) while at least one of the coordinate direction spans the entire interval. Matching slices of the signature manifold across varying viewpoints provides an efficient method for establishing correspondence of curves. Such slices (or sub-manifolds) may not be generated arbitrarily: If the pivot points in $\mathbb{RP}^2$ ($\mathbb{RP}^3$) result in collinear (coplanar) points on the curve, then the resulting slice may comprise all zero cross-ratio values. A p-dimensional sub-manifold of the n-point joint invariant signature manifold is obtained by pivoting $n-p$ of the n points. Consider a canonical p-dimensional slice of the interval $I^n \subset \mathbb{R}^n$,

$$\mathcal{A}_{t^*} = \{t_1^*\} \times \{t_2^*\} \times \cdots \times \{t_{n-p}^*\} \times I^p, \tag{12}$$

where first $(n-p)$ coordinates for $t^n \in \mathcal{A}_{t^*}$ are fixed: $t^n = (t^*, t^p)$. In order to generate all possible slices, we need to introduce the notion of permutation of coordinates. Let π denote a permutation of integers $\{1, \ldots, n\}$,

$$\begin{array}{rcl} \pi : \{1, \ldots, n\} & \rightarrow & \{1, \ldots, n\} \\ i & \mapsto & \pi(i). \end{array} \tag{13}$$

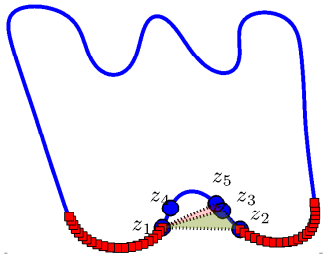
The permutation π acts on $\mathcal{A}_{t^*}$ to give another slice,

$$\mathcal{A}_{t^*,\pi} = \left\{ \tilde{t} \in I^n : \tilde{t}_{\pi(i)} = t_i, \text{ for } i = 1, \ldots, n, \text{ where } (t_1, \ldots, t_n) \in \mathcal{A}_{t^*} \right\}. \tag{14}$$

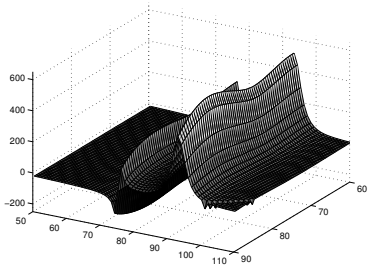
Given a slice $\mathcal{A}_{t^*}$, consider a local section $\mathcal{U}_{t^*} \subseteq \mathcal{A}_{t^*}$ defined as

$$\mathcal{U}_{t^*} = \{t_1^*\} \times \{t_2^*\} \times \cdots \times \{t_{n-p}^*\} \times U^p, \tag{15}$$

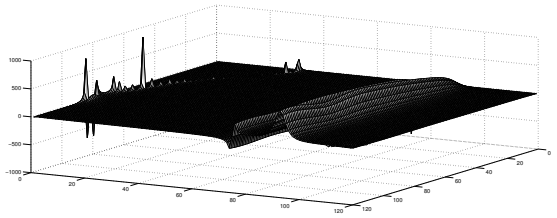
where $U^p = U_{n-p+1} \times U_{n-p+2} \times \cdots \times U_n \subseteq I^p$. A section of $\mathcal{A}_{t^*,\pi}$ is denoted as $\mathcal{U}_{t^*,\pi}$.



(a) Smoothed contour of letter "W".



(b) Surface plot of a section of 2D slice.



(c) Surface plot of a 2D slice under $\mathbf{F}_1^{(n)}$.

Figure 1.2: The points z_3, z_4 and z_5 are the pivot points and points z_1, z_2 span the 2D slice of the signature sub-manifold associated under $\mathbf{F}_1^{(n)}$.

Using sections of signature manifolds offers a tradeoff between the size of the invariant descriptors versus the computational load associated with searching pivot points. Consider the special case of projective transformations of planar curves ($m = 2, n = 5$). A 2D slice of signature manifold for the letter "W" from the license plate dataset is shown in Fig. 1.2(c) with permutation $\begin{pmatrix} 1\,2\,3\,4\,5 \\ 3\,4\,5\,1\,2 \end{pmatrix}$. The pivot points are $z_{\pi(i)} = S_{ij}(t_i^*)$ for $i = 1, 2, 3$ and $z_{\pi(4)}, z_{\pi(5)}$ span the entire interval I. A section of the slice is shown in Fig. 1.2(b) with free points restricted to the region marked by squares in Fig. 1.2(a).

Correspondence and Equivalence from Matching Sections

Correspondence between curves can be established efficiently using the sections of the signature manifold. Lets denote the section of the signature manifold J_{ij} collected on $\mathcal{U}_{t^*,\pi}$ as $J_{ij}(\mathcal{U}_{t^*,\pi})$. Camera C_k reconstructs the curve S_{ij} from samples of the curve received from camera C_i and the curve S_{kl} from the samples captured locally. Then pivot points $\hat{t}$ that generate the section of J_{kl} which is the best possible match to section $J_{ij}(\mathcal{U}_{t^*,\pi})$ can be found in an iterative manner as outlined below.

For $t^n \in I^n$, define

$$\Delta(t^n) = \inf_{J \in J_{ij}(\mathcal{U}_{t^*,\pi})} ||J_{kl}(t^n) - J||_2. \tag{16}$$

Consider the set V comprising all n-point sets in the domain of J_{kl} that result in cross-ratios within an ϵ-neighbourhood of the given section $J_{ij}(\mathcal{U}_{t^*,\pi})$:

$$V = \{t^n \in I^n : \Delta(t^n) < \epsilon^*\}. \tag{17}$$

The set V is learnt as follows. Starting with a uniform distribution on I^n, sample-points t^n are generated and tested for inclusion: $t^n \in V$ as given by (17). Alternatively, picking pivot points as points of inflections on the curve allows initializing with a prior distribution on I^n that is a Gaussian mixture model with components centered at inflection points. This yields tremendous speed-up in matching. See next section for more details on picking pivot points. As the set V grows, the clusters in the set V are learnt using Density-Based Scan Algorithm with Noise (DBSCAN). For each cluster, Hausdorff distance is computed between the given section $J_{ij}(\mathcal{U}_{t^*,\pi})$ and the signature manifold J_{kl} evaluated on points comprising the cluster. The distribution on I^n is updated in accordance with the score given by the Hausdorff distance and the sampling process is repeated. A concentrated cluster with small Hausdorff distance indicates a matching section between the two signature manifolds. Multiple matching sections provide significant statistical evidence for the equivalence of curves at two viewpoints.

Picking Pivot Points

A robust method for choosing pivot points and sections consistently across varying view-points is based on the identification of inflection points of curves. Inflection points are defined to be the points on the curve at which the curvature changes sign. Consider the motion of the tangent to a given planar curve at a point as the point moves along the curve. The tangent either rotates clockwise or anti-clockwise in the plane. The rate of rotation of the tangent is given by the curvature of the curve. The points at which the rotation changes direction (clockwise to anti-clockwise or vice versa) are the inflection points of the curve. It is well known that inflection points are invariant to projective transformations. Thus they can be found consistently across different perspectives and result in the same segmentation of the curve.

However, inflection points are very sensitive to noise. Figure 1.3 shows inflection points for various contour images extracted from the license plate dataset. Due to the quantized nature of the contours and associated noise or discontinuities, a simple test for inflection points results in a host of possible candidates as seen in Fig. 1.3(a). Smoothing the curve using a simple low-pass filter eliminates most of the noisy candidates (Fig. 1.3(b)). Further elimination based on the area under the curvature plot about each candidate

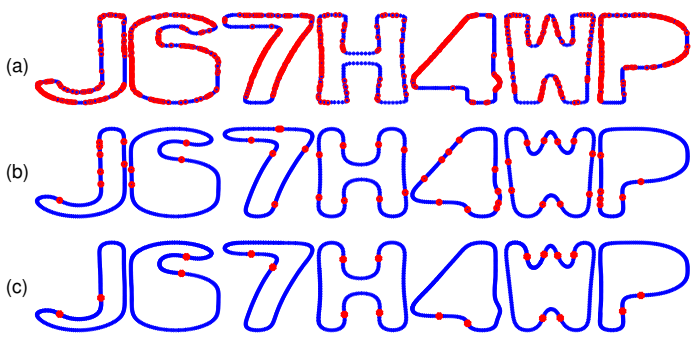


Figure 1.3: Inflection points marked with dots on (a) original contours, (b) smoothed contours and (c) post-elimination based on the amount of rotation of tangent about inflection points.

point reveals the significant inflection points as seen in Fig. 1.3(c). This preprocessing method is robust to severely noisy conditions as well as widely varying perspectives. It allows for robust segmentation of curves.

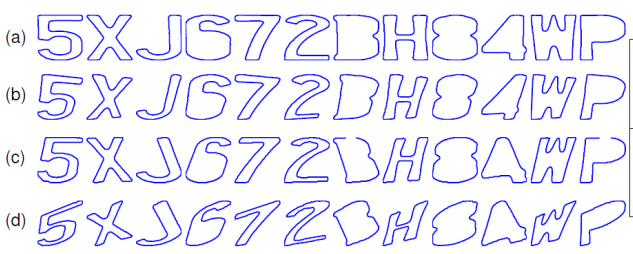
It should be remarked that most interesting shapes admit inflection points in the resulting contours. However, in the case where no inflection points are observed, the pivot points can be picked arbitrarily without affecting the matching accuracy of the iterative clustering method described in the previous section. Thus, the matching methodology does not depend critically on the presence of inflection points.

5. Matching Performance

This section discusses performance of the clustering-based algorithm on simulated data using the Epipolar Geometry Toolbox [14] as well as on a license plate image database [28].

Case: $m = 2, n = 5$

Figure 1.2 shows contour plots from license plate dataset along with invariant signatures. Figure 1.2(a) shows the contour of the letter "W" (extracted from images of the license plate WMP619). The set of five points on contours that generated the invariant signature (in Fig. 1.2(c)), are highlighted with dots. The points z_3, z_4 and z_5 are the pivot points and points z_1, z_2 span a 2D slice of the signature sub-manifold associated with $\mathbf{F}_1^{(n)}$ in eqn. (7). The surface plot of the 2D slice is shown in Fig. 1.2(c) and surface plot of a section is shown in Fig. 1.2(b).



(e)

True Positives 1106	False Positives 3
False Negatives 94	True Negatives 6597

Figure 1.4: Contour images from the license plate database [28]. The digits extracted from license plates as seen at (a) Camera 1, (b) at a randomly generated viewpoint of the curves observed at Camera 1, (c) at Camera 2, and (d) at a randomly generated viewpoint of the curves observed at Camera 2. (e) Confusion matrix for 100 random projective transformations of curves in the license plate database.

The images from the license plate dataset captured at two different viewpoints are shown in Figures 1.4(a,c). The test dataset comprising the 12 contour images was enlarged by generating random projective transformations of the given contours. The confusion matrix for this experiment involving 100 random projective transformations is shown in Fig. 1.4(e). It is evident from the experimental results that the method enjoys a good specificity as well as sensitivity. The number of detected inflection points for the test images ranged from 0 (for the digit "8" after smoothing) to 8 (for the letter "X").

Case: $m = 3, n = 6$

The Epipolar Geometry Toolbox [14] was used to simulate a 3D scene with moving space curves being tracked by two pinhole cameras. Since the signature sub-manifolds are invariant to Euclidean as well as perspective transformations, they uniquely describe these space-time curves. Therefore, the signature sub-manifold need not be recomputed at every frame of the trajectory of the curve. This allows for robust matching of curves in the image planes of the two cameras at every instant, without frame synchronization. For more details and results on matching space curves, refer to [28].

6. Discussion

This chapter presented an efficient algorithm for matching curves across widely varying viewpoints using joint invariants. The equivalence of curves under projective transformation is established by matching sections of invari-

ant signature manifold which are local, compact descriptors of curves. Matching results are provided using both simulated data and a license plate dataset.

References

[1] R. Arora, Y. H. Hu and C. R. Dyer. Estimating correspondence between multiple cameras using joint invariants. In *Proc. Int. Conf. Acoust., Speech and Signal Process.*, 2009.

[2] K. Astrom and L. Morin. Random cross ratios. In *Proc. 9th Scand. Conf. Image Anal.*, 1995, pp. 1053–1061.

[3] P. Clarot, E.B. Ermis, P.M. Jodoin and V. Saligrama. Unsupervised camera network structure estimation based on activity. In *Proc. 3rd ACM/IEEE Int. Conf. Dist. Smart Cameras*, 2009.

[4] O. Chum, J. Philbin and A. Zisserman. Near Duplicate Image Detection: min-Hash and tf-idf Weighting. In *Proc. British Mach. Vision Conf.*, 2008.

[5] R. Collins, A. Lipton, H. Fujiyoshi and T. Kanade. Algorithms for cooperative multisensor surveillance. In *Proc. IEEE*, 89(10):1456–1477, 2001.

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24:381–395, 1981.

[7] C. E. Hann. *Recognizing two planar objects under a projective transformation*. PhD dissertation, Univ. of Canterbury, 2001.

[8] C. E. Hann and M. S. Hickman. Recognising two planar objects under a projective transformation. *Kluwer Academic Publishers*, 2004.

[9] W. Hu, T. Tan, L. Wang and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. Systems, Man and Cybernetics, Part C: Applications and Reviews*, 34(3):334–352, 2004.

[10] D. Q. Huynh. The cross ratio: A revisit to its probability density function. In *Proc. 11th British Machine Vision Conf.*, 2000.

[11] L. Lee, R. Romano and G. Stein. Monitoring activities from multiple video streams: establishing a common coordinate frame. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):758–767, 2000.

[12] D. Lowe. Object recognition from local scale-invariant features. In *Proc. 7th Int. Conf. Computer Vision*, 1999, pp. 1150–1157.

[13] B. M. Marhic, E. M. Mouaddib and C. Pegard. A localisation method with an omnidirectional vision sensor using projective invariant. In *Proc. Int. Conf. Intelligent Robots Systems*, 1998, pp. 1078–1083.

[14] G. Mariottini and D. Prattichizzo. EGT for multiple view geometry and visual serving. *IEEE Robotics and Automation Mag.*, 12(4):26–39, 2005.

[15] S. J. Maybank. Probabilistic analysis of the application of the cross ratio to model based vision. *Int. J. Computer Vision*, 14:199–210, 1995.

[16] J. L. Mundy and A. Zisserman, editors. *Geometric Invariance to Computer Vision*. MIT Press, 1992.

[17] J. L. Mundy, A. Zisserman and D. Forsyth, editors. *Applications of Invariance in Computer Vision*. Springer-Verlag, 1993.

[18] P. J. Olver. Joint invariant signatures. *Foundations of Computational Mathematics*, 1:3–67, 2001.

[19] C. Orrite, S. Blecua and J. E. Herrero. Shape matching of partially occluded curves invariant under projective transformation. *Comp. Vis. Image Understanding*, 93:34–64, 2004.

[20] P. Remagnino, S. Velastin, G. Foresti and M. Trivedi. Novel concepts and challenges for the next generation of video surveillance systems. *Machine Vision and Applications*, 18(3):135–137, 2007.

[21] K. S. Roh, W. H. Lee and I. S. Kweon. Obstacle detection and self-localization without camera calibration using projective invariants. In *Proc. Int. Conf. Intelligent Robots and Systems*, 1997, pp. 1030–1035.

[22] C. A. Rothwell, A. Zisserman, D. A. Forsyth and J. L. Mundy. Canonical frames for planar object recognition. In *Proc. 2nd European Conf. Computer Vision*, 1992.

[23] C. A. Rothwell, A. Zisserman, D. A. Forsyth and J. L. Mundy. Planar object recognition using projective shape representation. *Int. J. Computer Vision*, 16:57–99, 1995.

[24] A. Shashua. A geometric invariant for visual recognition and 3D reconstruction from two perspective/orthographic views. In *Proc. IEEE Workshop on Qualitative Vision*, 1993, pp. 107–117.

[25] N. Snavely, S. Seitz and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 25(3):835–846, 2006.

[26] V. S. Tsonis, K. V. Chandrinos and P. E. Trahanias. Landmark-based navigation using projective invariants. In *Proc. Int. Conf. Intell. Robots Systems*, 1998, pp. 342–347.

[27] S. Velipasalar and W. Wolf. Frame-level temporal calibration of video sequences from unsynchronized cameras by using projective invariants. In *Proc. Advanced Video Signal-based Surveillance*, 2005, pp. 462–467.

[28] Wisconsin Computer Vision Group. Repository for joint-invariant matching. http://www.cae.wisc.edu/∼sethares/links/raman/JICRvid/space.html.