Distributed Curve Matching in Camera Networks using Projective Joint Invariant Signatures

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

Yu Hen Hu
Department of Electrical & Computer Eng.
University of Wisconsin-Madison
Madison, WI 53706
hu@engr.wisc.edu

Nigel Boston
Department of Mathematics
University of Wisconsin-Madison
Madison, WI 53706
boston@math.wisc.edu

ABSTRACT

An efficient method based on projective joint invariant signatures is presented for distributed matching of curves in a camera network. 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 between the two curves, matching signature manifolds is a computational challenge. In this paper we overcome this challenge by finding discriminative sections of signature manifolds consistently across varying viewpoints and scoring the similarity between these sections. This motivates a simple yet powerful method for distributed curve matching in a camera network. Experimental results with real data demonstrate the classification performance of the proposed algorithm with respect to the size of the sections of the invariant signature in various noisy conditions.

Keywords

Camera networks, distributed 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 viewpoints. 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 [5, 6, 10, 12, 21]. Matching curves across widely varying viewpoints requires local image features that are invariant to changes in pose, occlusion, illumination, scale, and intrinsic differences between cameras.

This paper 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.

A function defined on an image, of a planar or 3D object, is said to be *invariant* to a collection of transformations if its value remains unchanged despite the transformation 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.

1.1 Related Work

Projective invariants have been applied to various computer vision tasks such as localization [14, 22], autonomous navigation [27], 3D reconstruction [25], and surveillance [28]. The statistical distribution of random four-point cross-ratios has been studied numerically [11] as well as analytically [4] and performance of a classifier based on cross-ratios was

ICDSC 2010 August 31 – September 4, 2010, Atlanta, GA, USA

characterized quantitatively in terms of probability of rejection and false alarm [15]. However, in all of the works mentioned above, the correspondence of points between images was either given a priori or was obtained using external markers. Without correspondence information, the classification methodology breaks down since the cross ratios are not unique.

In other related work, Scale-invariant feature transform (SIFT) [13] was used to compute a large set of local feature vectors from each image and the correspondence between two images was established using RANSAC [7]. This approach has been used for 3D model reconstruction in the Photo Tourism system [26]. However, the computational complexity of SIFT and RANSAC makes it difficult to use for real-time video surveillance applications in camera networks with non overlapping field-of-view.

Rothwell et al. [24, 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 [8, 9] and by Orrite and Herrero [20]. These methods also utilize bi-tangency points on curves to learn a projective map. The key difference in the more recent work [8, 9, 20] from the algorithms presented in mid-nineties [24, 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 [8, 9, 20, 24, 23] between given curves are inherently centralized and computationally expensive. In order to deal with differences in sampling of images (resulting from different 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. Specializing these methods to matching curves in video streams will require transmitting full frames 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 [18] 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.

1.2 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 computationally efficient algorithm is presented for distributed matching. This work was inspired by recent advances in joint invariants [19] and probabilistic analysis of random cross ratios. We build on our previous work [2] which focusses on first establishing correspondence of points on the given curves and then comparing the associated projective joint invariant signatures. This requires exchange of images of the curves between cameras and solving an optimization problem to search for matching pivot points for each pair of curves [2]; therefore, the algorithm does not lend itself to distributed matching.

Joint invariant signatures were studied by Olver [19] 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 paper include generating compact local signatures independently from each image and fast distributed matching suited for camera networks. We systematically reduce the size and computational complexity of the matching by reformulating the problem and offering a tradeoff between the size of the feature space and the 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. Therefore, it is only required to exchange these invariant descriptors of the curve rather than the entire curve itself.

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 [3]. The matching of two curves can be performed efficiently in the presence of bi-tangents on the given curves. Unlike previous work [8, 9, 20, 24, 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 bandwidth and computational resources and allows for robust matching of curves between two cameras without frame synchronization.

The paper 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 submanifolds and local signatures to establish correspondence and equivalence of curves across viewpoints. Experimental results are presented in Section 5.

2. PROBLEM FORMULATION AND PRELIMINARIES

This section introduces notation and presents the problem formulation for pairwise curve matching across different viewpoints. Let C_i represent an unknown viewpoint or camera location. Let S_{ij} denote the j^{th} continuous planar-curve captured at viewpoint C_i. The curve S_{ij} is obtained from a 3D 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,

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

where $I^n_{ij} \subset \mathbb{R}^n$ is an n-dimensional interval, $t^n = (t_1, \ldots, t_n)$ and $S^n_{ij}(t^n) = (S_{ij}(t_1), \ldots, S_{ij}(t_n))$. 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 scene. The sampled, discretized version of S_{ij} is denoted D_{ij}.

We denote the manifold associated with the given curve as M. A given curve may undergo various transformations like rotation, translation, scaling and projection. These transformations can be described by 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

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

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

The projective transformation group, $G = \mathrm{PSL}(m+1, \mathbb{R})$, which is the subject of study in this paper, 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 ($m = 2$), the fundamental 5-point joint invariants for the projective transformation group $\mathrm{PSL}(3, \mathbb{R})$, are given by the volume cross ratios [19]:

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

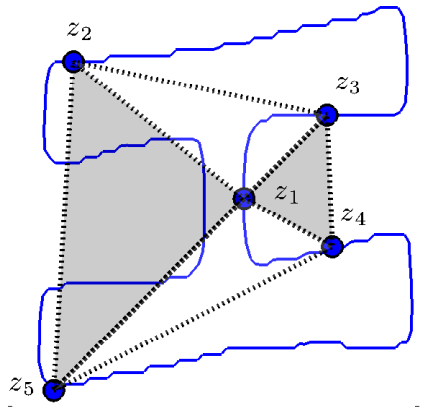


Figure 1: The five-point projective joint-invariant is the ratio of the product of areas of the non-shaded triangles and the product of areas of the shaded triangles.

and

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

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 Fig. 1 and the product of areas of the shaded triangles. Therefore, for $\mathrm{PSL}(3, \mathbb{R})$, $\mathbf{F}^{(n)}$ is given by

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

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

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

Fig. 1 shows a bird's eye view of the double pyramid with common base resulting from the union of the tetrahedrons [19].

The probabilistic analysis of random five point cross-ratios reveals that no single cross-ratio is unique on smooth manifolds [3]. Consequently, for planar curves, the matching schemes based on comparing single cross-ratios are not discriminative. The six-point joint-invariants for 3D 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 [3].

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 [3]. However, the joint invariant signature defined to be the manifold comprising cross-ratio values generated by *all possible* n-point

Table 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 in image plane of camera C_i
G	transformation group
M	manifold
$\mathbf{F}^{(n)}$	$n-$point joint-invariant 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 for 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 signature manifold determined by t^*, π
$\mathcal{U}_{t^*,\pi}$	a section of the slice $\mathcal{A}_{t^*,\pi}$

sets on a curve represents the curve uniquely up to a projective transformation [19].

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

$$\begin{aligned} J_{ij} : I^n &\rightarrow \mathbb{R}^l \\ t^n &\rightarrow \mathbf{F}^{(n)}(S_{ij}^n(t^n)). \end{aligned} \tag{8}$$

The *invariant signature manifold* at viewpoint C_i for the j^{th} curve is defined to be $J_{ij}(I_{ij}^n)$. 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 $g \in G = \mathrm{PSL}(3,\mathbb{R})$ on interval $I \subset \mathbb{R}$, then from the definition of joint action, $\tilde{S}_{ij}^n = g \cdot \tilde{S}_{kl}^n$. This implies that the invariant signature manifold for the two curves coincide: For all $t^n \in I^n$,

$$\begin{aligned} J_{ij}(t^n) &= (\mathbf{F}^{(n)} \circ \tilde{S}_{ij}^n)(t^n) \\ &= (\mathbf{F}^{(n)} \circ (g \cdot \tilde{S}_{kl}^n))(t^n) \\ &= (\mathbf{F}^{(n)} \circ \tilde{S}_{kl}^n)(t^n) \\ &= J_{kl}(t^n). \end{aligned} \tag{9}$$

More importantly, $\tilde{S}_{ij} = g \cdot \tilde{S}_{kl}$ for some $g \in G$ if $J_{ij}(I^n) = J_{kl}(I^n)$[19]. 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,

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

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

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

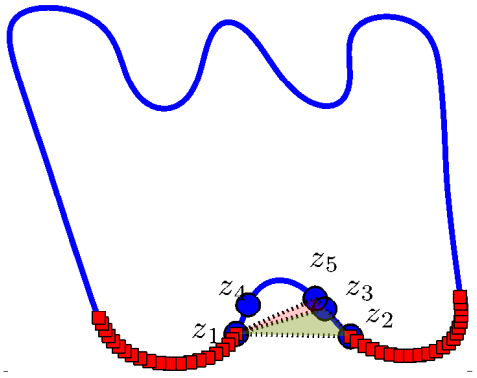
where threshold $\epsilon > 0$ depends on the amount of noise resulting from differences in quantization and the sampling grids of the two cameras. The matching criterion in (11) is straightforward if the two curves are defined on the same domain and the true correspondence between the curves is known. 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}). 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. Instead we restrict invariant descriptors of the curves to be certain sections of the signature manifold. Then the matching proceeds with estimating d_{kl}^{ij} between given sections J_{ij} and J_{kl}.

4. TOWARD LOCAL SIGNATURES

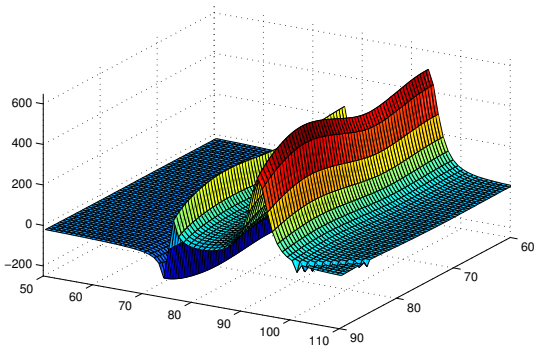
The joint invariant signature manifold is a global descriptor. Owing to the lack of robustness of global signatures to occlusion, we restrict our attention to sections of invariant signature sub-manifolds. We first define a slice of the signature manifold and then discuss local signatures computed on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

4.1 Slices and Sections of Signature Manifold

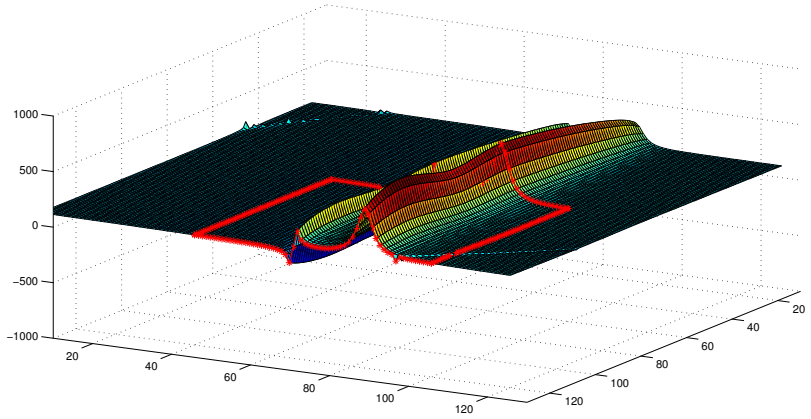
A *slice* of the signature manifold may be generated by freezing one or more coordinate direction of I^n (at pivot points denoted as t^*) while at least one coordinate direction spans the entire interval. Matching slices of the signature manifold across varying viewpoints provides an efficient method 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



(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 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 $\mathbf{F}_1^{(n)}$.

(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,

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

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 $\mathcal{U}_{t^*,\pi}$.

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 the signature manifold for the letter "W" from a license plate dataset [1] is shown in Fig. 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. 2(b) with free points restricted to the region marked by squares in Fig. 2(a).

4.2 Distributed Matching

An important issue in an implementation of a local signa-

ture based matching algorithm is that of non-uniform sampling of the curves in the two images. That is, some parts of the curve may be more densely sampled in C_1 than in C_2. One solution is to re-sample each extracted contour from an image (as a preprocessing step) on to a uniform grid. This realizes a uniform resolution over the sampled curve. Another possible solution is to exchange images of the curves between cameras and solve an optimization problem to search for matching pivot points for each pair of curves [2]. Then, for each possible match of pivot points the joint invariant signature is computed locally and compared against a pre-determined threshold [2]. Although effective, this algorithm does not lend itself to distributed matching.

In this section we propose a distributed matching method that avoids the overhead of reconstructing or resampling the images on a uniform grid. The key idea behind the matching algorithm is that local sections on the curve correspond to tight clusters in the domain of the signature manifold. This motivates the following algorithm for generating sections of the signature manifold that are discriminative as well as compact descriptors of the given curve.

---

def $Jsig_1 = \text{genSIG}(S_1, N, lo, hi)$
Input: S_1 - local planar curve, N - size of the section of invariant signature, (lo, hi) - terminal indices for a segment of the curve
Output: An $N \times 2$ matrix representing a section of the invariant signature
L = number of samples on S between lo and hi;
Pivot point $t^* = (lo + \lfloor \frac{L}{4} \rfloor, lo + \lfloor \frac{2L}{4} \rfloor, lo + \lfloor \frac{3L}{4} \rfloor, hi, lo)$;
$CR_1 = \mathbf{F}_1^{(5)}(S_1^5(t^*))$;
$CR_2 = \mathbf{F}_2^{(5)}(S_1^5(t^*))$ (see Eqns. (4), (5), (6));
$Jsig_1(1,1) = CR_1$, $Jsig_1(1,2) = CR_2$;
Define a neighbourhood $U \subset I^5$ around t^*;
for $i = 2, 3, \ldots, N$ **do**
  Generate uniformly random sample t^5 from U;
  $(CR_1, CR_2) = \mathbf{F}^{(5)}(S_1^5(t^5))$;
  $Jsig_1(i,1) = CR_1$, $Jsig_1(i,2) = CR_2$;
end

Algorithm 1: Generating sectional signature

It should be remarked that at each iteration of the inner loop in Algorithm 1, a new point in U is sampled (without replacement). Also, the drawn sample is discarded if the corresponding cross-ratios are not well defined or lie outside a desired interval. Recall that the probabilistic analysis of random cross ratios [3, 4, 11, 15] suggests that smaller cross ratios are not discriminative and very large cross ratios are sensitive to noise.

Given sections of two invariant descriptors, one computed on a curve captured locally and another received from a different node in the camera network, the goal is to score the similarity of the invariant descriptors. A simple scoring scheme follows from identification of the best possible match between the subsets (of sufficient size) of the sections of the two descriptors. The scoring scheme is described in Algorithm 2.

---

def $Score = \text{matchSIG}(Jsig_1, Jsig_2, N, \alpha)$
Input: $Jsig_1$ - Invariant descriptor received, $Jsig_2$ - Invariant descriptor for the curve captured locally, N - size of the invariant signature, overlap parameter $0 < \alpha \leq 1$
Output: Similarity score between $Jsig_1$ and $Jsig_2$
$Score = 0$;
$J = \{1, 2, \ldots, N\}, K = \{1, 2, \ldots, N\}$;
for $i = 1, 2, \ldots, \lfloor \alpha N \rfloor$ **do**
  $(\hat{j}, \hat{k}) = \arg\min_{j \in J, k \in K} ||Jsig_1(j,:) - Jsig_2(k,:)||_2$;
  $Score = Score + ||Jsig_1(\hat{j},:) - Jsig_2(\hat{k},:)||_2$;
  $J = J \backslash \hat{j}$, $K = K \backslash \hat{k}$
end

Algorithm 2: Matching sectional signature

In order to facilitate fast matching of sections of invariant descriptors, one should be able to generate the same sets of pivot points consistently across varying viewpoints, even in noisy conditions. This is accomplished by identification of inflection points on the curve which are left invariant by projective transformations. The segment of the curve delimited by a pair of inflection points is used to generate a section of the invariant signature as described in Algorithm 1. For more details on finding inflection points, see Sec. 4.3.

4.3 Picking Pivot Points

A robust method for choosing pivot points and sections consistently across varying viewpoints 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 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 viewpoints and result in the same segmentation of the curve.

However, inflection points are sensitive to noise. Fig. 3 shows inflection points for various contour images extracted

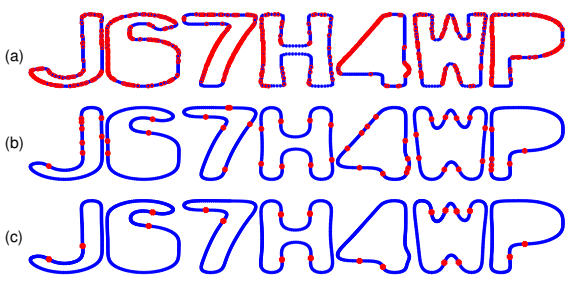
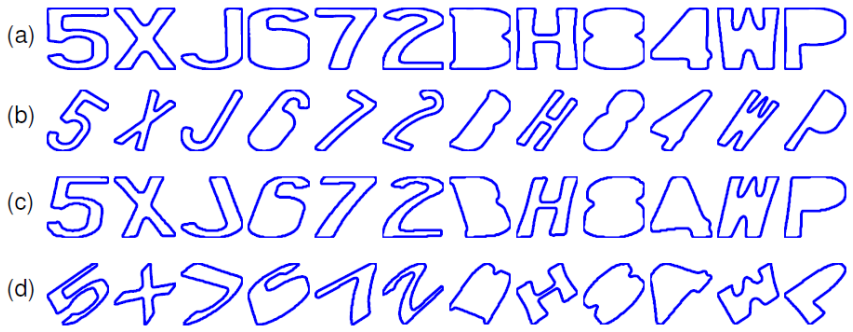


Figure 3: Inflection points marked with dots on (a) original contours, (b) smoothed contours and (c) post-elimination based on the amount of rotation of tangent about inflection points.



(e)

True Positives 1179	False Positives 21
False Negatives 21	True Negatives 13179

Figure 4: Contour images from the license plate database [1]. 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.

from the license plate dataset [1]. 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. 3(a). Smoothing the curve using a simple low-pass filter eliminates most of the noisy candidates (Fig. 3(b)). Further elimination based on the area under the curvature plot about each candidate point reveals the significant inflection points as seen in Fig. 3(c). This pre-processing method is robust to severely noisy conditions as well as widely varying viewpoints. Thus it allows for robust segmentation of curves.

Note 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 to ensure that the associated cross-ratio values satisfy certain design constraints. As long as the pivot points are identified consistently, it does not affect the accuracy of the distributed matching scheme described in the previous section. This is also evident from experimental results in severe noisy conditions.

5. EXPERIMENTAL RESULTS

This section discusses performance of the distributed curve-matching algorithm proposed in this paper on a license plate image database [1].

Fig. 2 shows one of the contour plots from the license plate dataset along with its invariant signatures. Fig. 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. 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 submanifold associated with $\mathbf{F}_1^{(n)}$ in Eq. (7). The surface plot of the 2D slice is shown in Fig. 2(c) and surface plot of a section is shown in Fig. 2(b).

The images from the license plate dataset captured at two different viewpoints are shown in Figs. 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 for 100 random projective transformations is shown in Fig. 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").

Finally, the performance of matching was studied with respect to the size of the exchanged sections of the invariant signature in various noisy conditions. A family of curves representing the probability of correct classification was generated for various SNR levels as shown in Fig. 5. Results are reported for the 12 contours of digits and letters (shown in Fig. 4) extracted from the multi-view license-plate dataset [1]. Our method enjoys better accuracy than the clustering-based method presented in [2] on the same dataset. At the same time the algorithm is considerably faster with each comparison (including the generation of the invariant signature) taking roughly 0.5s on a 2.4GHz machine. The source code along with a detailed report on experimental settings and extended results is available online [1].

6. DISCUSSION

This paper 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 the invariant signature manifold which are local, compact descriptors of curves. Experimental results with a license plate dataset show superior accuracy and speed over existing methods.

7. REFERENCES

[1] R. Arora. Repository for curve-matching using joint-invariant signatures. `http://www.cae.wisc.edu/~sethares/links/raman/JICRvid/space.html`, 2009.

[2] R. Arora and C. R. Dyer. Projective joint invariants for matching curves in camera networks. In B. Bhanu,

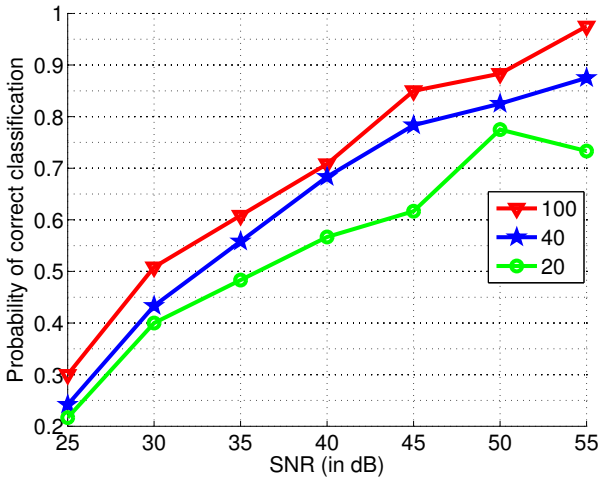


Figure 5: Probability of correct classification versus the additive noise levels ($25dB - 55dB$) in the contour images for various lengths of the invariant descriptor $(20, 40, 100)$.

C. V. Ravishankar, A. K. Roy-Chowdhury, D. Terzopoulos, and H. Aghajan, editors, *Distributed Video Sensor Networks*. Springer-Verlag, 2010.

[3] R. Arora, Y. H. Hu, and C. Dyer. Estimating correspondence between multiple cameras using joint invariants. In *Proc. Int. Conf. Acous., Speech and Sig. Proc.*, 2009.

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

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

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

[7] 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.

[8] C. E. Hann. *Recognizing two planar objects under a projective transformation*. PhD Thesis, University of Canterbury, 2001.

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

[10] 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.

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

[12] 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.

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

[14] 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 and Systems*, pages 1078–1083, 1998.

[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] J. P. O. Chum and A.Zisserman. Near duplicate image detection: min-hash and tf-idf weighting. In *Proc. British Mach. Vision Conf.*, 2008.

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

[20] 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.

[21] 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(3-4):135–137, 2007.

[22] 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*, pages 1030–1035, 1997.

[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] C. A. Rothwell, A. Zisserman, D. A. Forsyth, J. L. Mundy, and J. L. Canonical frames for planar object recognition. In *Proc. 2nd European Conf. on Comp. Vision*, pages 757–772, 1992.

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

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

[27] V. S. Tsonis, K. V. Chandrinos, and P. E. Trahanias. Landmark-based navigation using projective invariants. In *Proc. Int. Conf. Intelligent Robots Sys.*, pages 342–347, 1998.

[28] 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*, pages 462–467, 2005.