Chapter 1

JOINT PROJECTIVE INVARIANTS FOR DISTRIBUTED 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 across different viewpoints. The fundamental perspective invariants for curves in the real projective space are the volume cross-ratios. Probabilistic analysis of projective invariants shows that they are not unique and therefore not discriminative. However, a curve in m-dimensional Euclidean space is completely prescribed by the signature manifold of joint invariants generated by taking all possible combinations of n points on the projective curve where n is at least $m + 2$. Furthermore, sub-manifolds given by the projection of the signature manifold also represent the curve uniquely. Sections of the sub-manifolds that admit large enough variation of cross ratios are found to be sufficient, statistically, for matching of curves. Such sectional signatures allow fast computation and matching of features while keeping the descriptors compact. These features are computed independently at cameras with different viewpoints and shared, thereby achieving the matching of objects in the image. Experimental results with simulated as well as real data are provided.

Keywords: Camera networks, correspondence-less matching, joint projective invariant signatures

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 [4; 8; 18]. 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 joint projective 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 the projections of unknown 3D curves in the scene. The problem is to determine if two curves in two images 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 [14; 15].

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

The object correspondence problem arises in many different contexts. Projective invariants have been applied to various computer vision tasks such as localization [11; 19], autonomous navigation [24], 3D reconstruction [22], and surveillance [25]. 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 different assumptions on the distribution of the four points. The distribution of cross ratios was further examined in [9] as more constraints on relative distances of

the four points are imposed. The performance of cross ratios was described quantitatively in terms of probability of rejection and false alarm in [13]. Unfortunately, as noted by [1], 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) [10] features have been used to compute a large set of local feature vectors from each image, and the correspondence between two images established using RANSAC [5]. This approach has been used for 3D model reconstruction in the Photo Tourism system [23]. 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. [21; 20] proposed 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. The representation of the curve in the canonical frame is claimed to be semi-local. The method extends to model-based matching of shapes but is not suitable for a camera network setting because different distinguished points may be selected under different occlusion conditions in the two cameras. To tackle this issue, the authors introduced the notion of redundancy in their later work [20]. The idea is to have multiple canonical frames from various segments of the curve. However, the bi-tangency condition required for distinguished points may not hold in such scenarios.

Similar ideas have been put forth by Hann and Hickman [7; 6] and by Orrite and Herrero [17]. These methods also utilize bi-tangency points on curves to learn a projective map. The key difference in the more recent work [7; 6; 17] from the algorithms proposed in mid-nineties [21; 20] 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 curve matching across multiple viewpoints that preclude their employment to applications like video surveillance in camera networks. Firstly, the schemes proposed are either inherently centralized or are too expensive in terms of network resources. The classification rule based on these methods relies on the Hausdorff distance between the curve from one point to the image of the curve from the other viewpoint under the best estimate of the projective transformation given by identification of bi-tangency points. In order to compute this Hausdorff distance, each sensor node in the network will require images from other nodes. Secondly, 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. This requires complete exchange of images at each iteration. Thirdly, 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 [3] 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 and 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 space curves extracted from images of an arbitrary 3D scene. Signatures based on these cross ratios are computed independently from each image. The invariant signatures are semi-local and compact. A clustering-based method is presented for distributed matching of signatures between viewpoints that simultaneously solves the correspondence and matching problems. This work was inspired by recent advances in joint invariants [16] and analysis of uniqueness of joint projective invariant signatures [1].

Joint invariant signatures were recently studied by Olver [16] 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 here include generating compact semi-local signatures independently from each image and clustering-based unsupervised matching. We systematically reduce the size and computational complexity of the matching by reformulating the problem and offering a tradeoff between size of feature space and complexity of optimization.

Our method alleviates the aforementioned shortcomings of existing methods. Unlike existing methods that approach curve matching 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. Our scheme deals with differences in image sampling by regularization. Furthermore, we utilize a recent result from [1] whereby it was established that a small perturbation of points on the curve results in a small relative error in the invariant

domain. The proposed method does not critically depend on the existence of bi-tangents. However, whenever bi-tangents are present, our approach utilizes them to reduce the size of the invariant descriptors that need to be exchanged between camera nodes. To the best of our knowledge, this is the first time projective invariant curve matching has been employed on a real dataset with extensive performance characterization. Most of the existing work has been limited to experimentation with toy examples [1; 6; 7; 21; 20]. Finally, the description of a curve using joint projective invariants is invariant to Euclidean transformations too. Therefore, in video surveillance applications, the representation is redundant across frames of the video-feed. 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 joint projective invariants. Section 3 describes the joint invariant signature and the computational complexity associated with using the entire signature manifold. Although correspondence-less matching is possible with the invariant signature manifold, its use is not desirable due to the global nature of the descriptor and the size of the computational problem. In Section 4, the size of the problem is reduced by using the fact that a slice (or sub-manifold) of the signature manifold, if chosen properly, prescribes the curve completely. Furthermore, local signatures can be extracted from this sub-manifold as shown in Section 5. These local signatures consist of regions of the prescribing slice with high enough variation of cross ratios to encode important information content and represent the curve uniquely with high probability. Finally, Section 6 presents a distributed, clustering-based method for simultaneous matching and correspondence of image curves, and Section 7 presents experimental results with simulated as well as real datasets.

2. Problem Formulation and Preliminaries

This section introduces the notation (largely adopted from [16; 1]) and presents the problem formulation for pairwise curve matching across different viewpoints. Let S_1, S_2 be two images taken by uncalibrated cameras from unknown viewpoints of an arbitrary 3D scene. Let $\{C_{1,i}\}_{i=1}^{M}, \{C_{2,j}\}_{j=1}^{N}$ denote the collection of planar curves extracted from images S_1 and S_2, respectively. For each pair of curves $(C_{1,i}, C_{2,j})$, the objective is to either establish or reject the possibility that the two curves in the image space at different viewpoints represent the same space curve in the observed scene. The sampled, discretized version of a curve C_i is denoted as $C_i^{(d)}$ with the total number of samples denoted $|C_i^{(d)}|$. The n-dimensional Cartesian product of the curve C_i is denoted as C_i^n.

A smooth planar curve in Euclidean space is a one-dimensional manifold and a space curve is a two-dimensional manifold. We denote the finite-dimensional

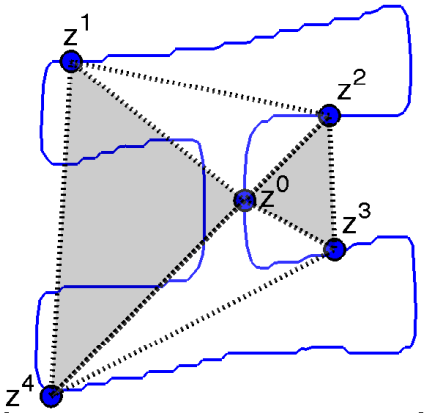

Figure 1.1. The five-point projective joint-invariant is the ratio of the product of areas of the shaded triangles and the product of areas of the non-shaded triangles.

smooth manifold associated with a 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 the group on the manifold describes how the group transforms any given n-tuplet on the manifold. Formally, the joint action of the group G on the Cartesian product M^n is given as:

$$g \cdot (z_1, \ldots, z_n) = (g \cdot z_1, \ldots, g \cdot z_n)$$

for $g \in G$ and $(z_1, \ldots, z_n) \in M^n$.

An $n-$point joint invariant of the transformation group G on M is defined to be a function $I(z_1, \ldots, z_n)$ that is invariant to the joint action of the group on the manifold. 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 $z \in \mathbb{R}^m$ to $w \in \mathbb{R}^m$.

For an n-point joint action of the projective transformation group on the projective space, the fundamental m-dimensional projective joint invariants are given by the volume cross ratios [16]:

$$CR(i_0, \ldots, i_{m-2}; j, k, l, n) = \frac{V(i_0, \ldots i_{m-2}, j, k)V(i_0, \ldots i_{m-2}, l, n)}{V(i_0, \ldots i_{m-2}, j, l)V(i_0, \ldots i_{m-2}, k, n)} \quad (1)$$

where $V(i_0, \ldots, i_{m-2}, j, k)$ is the volume of the simplex with vertices given by $z_0, \ldots, z_{m-2}, z_j, z_k$ in $\mathbb{R}^m$:

$$V(i_0, \ldots, i_{m-2}, j, k) = \det \begin{bmatrix} z_0 & z_1 & \cdots & z_{m-2} & z_j & z_k \\ 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix}.$$

This implies that any invariant of n points on a curve in $\mathbb{RP}^m$, such that $n \geq m+2$, can be expressed as volume cross ratios defined in Eq. (1). For the case $m = 2$, $n = 5$, the expression above reduces to the ratios

$$CR(0; 1, 2, 3, 4) = \frac{V(0, 1, 2)V(0, 3, 4)}{V(0, 1, 4)V(0, 2, 3)} \quad (2)$$

and

$$CR(1;0,2,3,4) = \frac{V(0,1,2)V(1,3,4)}{V(0,1,4)V(1,2,3)} \tag{3}$$

where $V(i,j,k)$ is the area of the triangle defined by z_i, z_j and z_k. The cross ratio defined in (2) is described as the ratio of the product of the areas of the shaded triangles in Figure 1.1 and the product of areas of the non-shaded triangles. For $m = 3, n = 6$, there are three fundamental volume cross-ratios: $CR(0,1;2,3,4,5)$, $CR(0,2;1,3,4,5)$, and $CR(1,2;0,3,4,5)$. Geometrically, $CR(0,1;2,3,4,5)$ is the ratio of the volumes of four tetrahedrons: $V(0,1,2,3)V(0,1,4,5)/V(0,1,2,4)V(0,1,3,5)$. Fig. 1.1 shows a bird's eye view of the double pyramid with common base resulting from the union of the tetrahedrons [16].

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 presented in [1]. 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 *does* represent the curve uniquely up to a projective transformation [16].

Let I be an n-point joint invariant map (see Section 2) for the projective transformation group, $I : C_i^n \to \mathbb{R}$. Denote the n-dimensional[1] invariant signature manifold given by the images of curves C_1 and C_2 under the map I by $J_{C_1^n}$ and $J_{C_2^n}$ respectively:

$$J_{C_1^n} = I(C_1^n), \quad J_{C_2^n} = I(C_2^n)$$

where $C_i^n = C_i \times \cdots \times C_i$. If the given curves are related by a projective transformation, i.e., $C_1 = g \circ C_2$ for some $g \in G = \text{PSL}(m+1, \mathbb{R})$, then from the definition of joint action, $C_1^n = g \circ C_2^n$. This implies that the set of all cross ratio values observed on the two curves are the same:

$$J_{C_1^n} = I(C_1^n) = I(g \circ C_2^n) = I(C_2^n) = J_{C_2^n}.$$

More importantly, $C_1 = g \circ C_2$ for some $g \in G$ if and only if $J_{C_1^n} = J_{C_2^n}$ [16]. Therefore, two curves are equivalent up to a projective transformation if and only if the Hausdorff distance between the sets of cross-ratios is 0, i.e.

$$d(J_{C_1^n}, J_{C_2^n}) = 0.$$

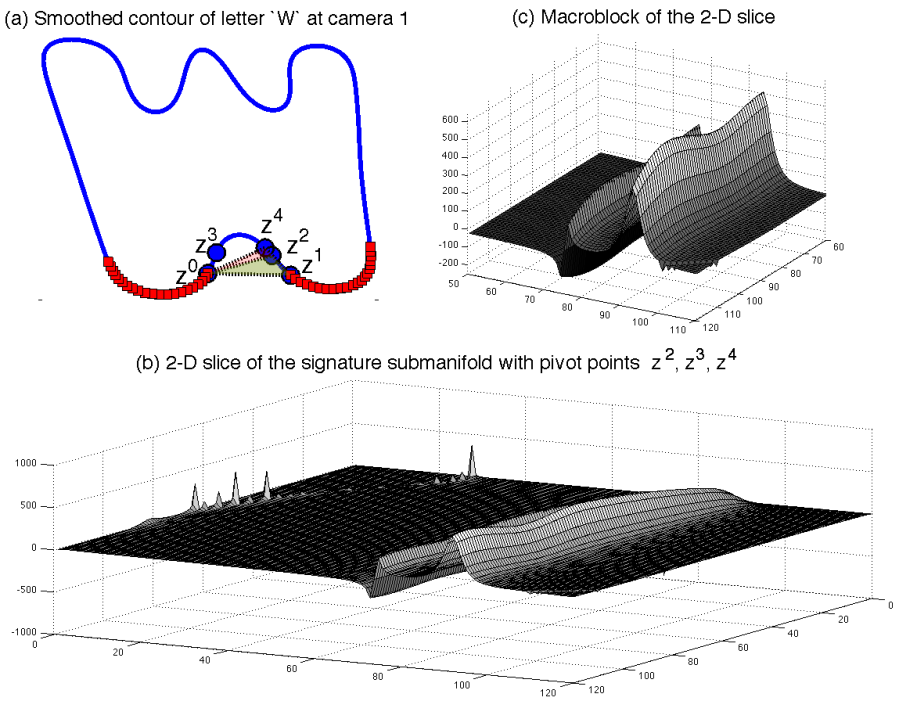

Figure 1.2. (a) Smoothed contour of letter "W" from the license plate dataset [26]. (b) Surface plot of a 2D slice. The points z_2, z_3 and z_4 are the pivot points and points z_0, z_1 span the 2D slice of the 5D signature sub-manifold as they traverse the entire curve. (c) Surface plot and contour image of a informative local section of the 2D slice.

In general the sets $J_{C_1^n}$ and $J_{C_2^n}$ are not bounded. Thus the Hausdorff criterion needs to be restated as:

$$d(U \cap J_{C_1^n}, U \cap J_{C_2^n}) = 0, \tag{4}$$

where $U \subset \mathbb{R}$ is any closed and bounded subset of real numbers. Also, in practice, C_1, C_2 are sampled discrete curves and the sampling grids for the two images may not correspond, resulting in non-zero Hausdorff distance. But, given the jitter analysis above, it is guaranteed that

$$d(U \cap J_{C_1^n}, U \cap J_{C_2^n}) < \epsilon \tag{5}$$

where $\epsilon > 0$ depends on the amount of jitter resulting from noise, quantization differences, and mismatches between the two images' pixel arrays.

Establishing the relation in (5) for sufficiently small ϵ yields correspondence-less matching. However, solving for the Hausdorff distance is computationally intensive and the rest of this chapter deals with systematically reducing the size of the problem, both in terms of the algorithmic complexity of computing the joint invariants as well as solving the correspondence problem. Various approaches are compared based on the size of the problem. From the perspective of space requirements, another aspect of the problem size is the size of the invariant descriptor. For correspondence-less matching, using the complete joint invariant signature manifold, $|C_1^{(d)}|^n + |C_2^{(d)}|^n$ cross ratio values must be computed and compared using the Hausdorff distance and solving Eq. (5), has a complexity of $\mathcal{O}(|C_1^{(d)}|^n \cdot |C_2^{(d)}|^n)$. This approach is clearly not practical.

4. Slicing through the Signature Manifold

A sub-manifold of the signature manifold may be generated by taking a slice of the manifold by pivoting one or more points while at least one of the points takes values over the entire curve. An important result in the theory of joint invariant signatures for projective transformations is that certain sub-manifolds also describe the curve uniquely. Such slices (or sub-manifolds) may not be generated arbitrarily: If the pivot (fixed) points in $\mathbb{RP}^2$ ($\mathbb{RP}^3$) are collinear (coplanar), then the resulting slice comprises all zero cross-ratio values. A p-dimensional sub-manifold of the signature manifold is obtained by pivoting $n-p$ of the n points. Let $P \subset \{1, 2, \ldots, n\}$ denote the set of indices of pivot points, with $|P| = p$. Define

$$A_1 = \{z \in C_1^n | z_i = x_i \text{ for } i \in P, z_i \in C_1 \text{ for } i \notin P\}.$$

A_1 is a p-dimensional slice of the Cartesian product C_1^n. Let J_A denote the p-dimensional signature sub-manifold obtained by restricting the signature manifold J to the slice A. For instance, let $P = \{2, 3, \ldots, n\}$. A_1 is 1D with z_1 taking all values on the curve C_1 and the corresponding 1D signature sub-manifold J_{A_1} defined to be the signature collected over this 1D slice of C_1^n:

$$J_{A_1} \equiv I(\bullet, x_2, \ldots, x_n) = I(C_1^n \cap A_1)$$

From [16] we know that any of the slices of the signature sub-manifold also prescribes the entire curve. Thus

$$C_1 = g \circ C_2 \text{ for some } g \in G \iff d(J_{A_1}, J_{A_2}) < \epsilon$$

for $\epsilon > 0$. This implies that if $d(J_{A_1}, J_{A_2}) < \epsilon$ for sufficiently small ϵ, then curves C_1 and C_2 match. To use this test for matching, we need to find the pivot points $\tilde{x}_i \in C_2$, for $i \in P$, that determine the slice A_2 and minimize the Hausdorff distance between J_{A_1} and J_{A_2},

$$\Delta = \min_{\tilde{x}_i \in C_2,\ i \in P} d(J_{A_1}, J_{A_2}).$$

and test for $\Delta < \epsilon$. In contrast to the formulation in the previous section that allows for correspondence-less matching, testing for $\Delta < \epsilon$ simultaneously achieves correspondence and matching. Also note the tradeoff between the size of the invariant descriptors versus the computational load.

Consider the special case of projective transformations of planar curves ($m = 2, n = 5$). A 1D slice allows for much smaller numbers of invariants ($\mathcal{O}(|C_i|)$) to be compared but at the same time requires solving a 4D optimization problem ($\mathcal{O}(|C_i|^4 \times |C_i|)$):

$$\Delta = \min_{\tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5 \in C_2} d(I(A), I(\bullet, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5))$$

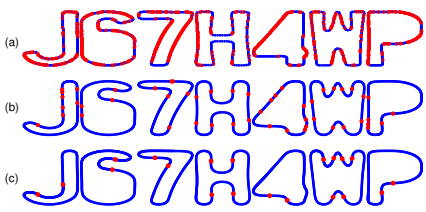

Figure 1.3. Inflection points marked with dots on (a) original contours, (b) smoothed contours and (c) post-elimination based on the amount of rotation of tangent about inflection points.

A more balanced trade-off between time and space may result from using 2D slices. A 2D sub-manifold of the signature manifold $J_{C_1^5}$ is defined to be the collection of invariants collected over the 2D slice of the set C_1^n. This signature requires $\mathcal{O}(|C_i|^2)$ cross ratio values to be computed and compared. For $P = \{2, 4, 5\}$, the corresponding discrete optimization problem is

$$\Delta = \min_{\tilde{x}_2, \tilde{x}_4, \tilde{x}_5 \in C_2} d(I(A_1), I(\bullet, \tilde{x}_2, \bullet, \tilde{x}_4, \tilde{x}_5))$$

which requires a search in $\mathcal{O}(|C_i|^3)$ dimensional space.

5. Toward Local Signatures

All the signatures discussed so far are global in nature. Owing to the lack of robustness of global signatures to occlusion, we now restrict our attention to sections of invariant signature sub-manifolds. This requires that the n points in the domain of the sub-manifold are selected over a short segment of the given curve. For curves in $\mathbb{RP}^2$, a section of a 2D slice is shown in Figure 1.2(e).

A method to choose segments of a curve consistently across varying viewpoints is based on the identification of inflection points. 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

point reveals the significant inflection points as seen in Fig. 1.3(c). This pre-processing method is robust to significant amounts of noise and 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, we have only one segment - the whole curve itself - and the distributed clustering method described in the next section still produces a good match. Thus, our matching methodology does not depend on the presence of inflection points.

Given segments of a curve, the next step is to identify good pivot points in the segment and sections of the curve that yield informative feature values. As discussed in [1], high cross ratio values on a curve are rare and therefore a collection of cross ratio values with large enough mean and variance is unique with high probability. Thus the segments of the curve are scanned via random sampling for pivot points that generate such discriminative features. The cross-ratios observed in a tight neighborhood around pivot points constitute the feature vectors.

6. Distributed, Clustering-based Matching

An important issue in an implementation of the local signature-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. Alternatively, this pre-processing step can be avoided by using a distributed, clustering-based method. Given a section of the signature sub-manifold from image S_1, the n-point sets in image S_2 with volume cross-ratios lying in the neighborhood of the given section are clustered iteratively to learn the matches between the two images. 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. The five-point sets corresponding to the received feature vector at camera 2 from camera 1 is centered around a mean $z = (z_1, \ldots, z_5)$ and enclosed in an epsilon ball $(z - \epsilon, z + \epsilon) \subset C_1^5$. This motivates the method described in Algorithm 1.

Another important aspect of the algorithm is to deal with the sensitivity of high cross-ratio values in noisy settings. High cross ratio values close to singularity points are saturated to a predetermined threshold. Consistent with the ideas in Section 3 (also see remarks leading up to Eq. (4)), the cross ratio map is restricted to a closed and bounded set of real numbers, $U = [-l, r]$, clipping the cross ratios that lie outside this interval.

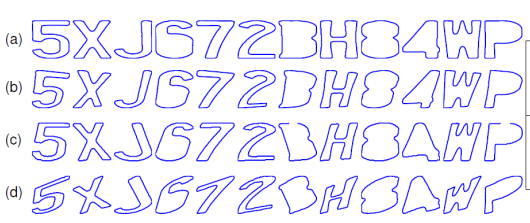

Figure 1.4. Contour images from the license plate database [26]. The digits extracted from license plates as seen at (a) Camera 1, (b) at a randomly generated viewpoint of the curves observed at camera 1, (c) at Camera 3, and (d) at a randomly generated viewpoint of the curves observed at camera 2. (e) Confusion matrix for 100 random projective transformations of curves in the license plate database.

---

def $ClusterCenter$ = matchSIG($Jsig$, C)
Input: Section of Joint invariant signature $Jsig$, Local planar curve C
Output: Center of the best cluster $ClusterCenter$
repeat
  Generate M random 5-point sets on C;
  for $i = 1, 2, \ldots M$ **do**
    Compute CR = cross ratio of i^{th} set;
    if $\min\left(|CR - Jsig|^2\right) < \delta$ **then**
      update the clusters $Clust = update(Clust, i^{th}$ set$)$
    end
  end
  foreach *cluster* $iClust$ *in* $Clust$ **do**
    Compute the distance $d(iClust, Jsig)$;
  end
  Find best cluster $bestClust = iClust$ with minimum $d()$;
  Obtain $Dmin = d(bestClust)$
until $Dmin < \epsilon$ *or max iterations* ;
return center of $bestClust$.;

Algorithm 1: Algorithm for distributed curve matching.

---

7. Matching Performance

This section discusses performance of the clustering-based algorithm on simulated data using the Epipolar Geometry Toolbox [12] as well as on a license plate image database [26].

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 digit "6" (extracted from images of the license plate 67724QB). The set of five points on contours that generated the invariant signatures (in Fig. 1.2(c)), are highlighted with symbols. The points z_2, z_3 and z_4 are the pivot points and points z_0, z_1 span a 2D slice of the 5D signature manifold. The surface plot of the 2D slice is shown in Fig. 1.2(b) and the grayscale contour image is shown in Fig. 1.2(c). The local section of the 2D slice consists of a 30×30 macroblock with surface plot shown in Fig. 1.2(d). Numerous such signatures were generated for contour (a) and the matching signatures in contour (b) were unique with high probability.

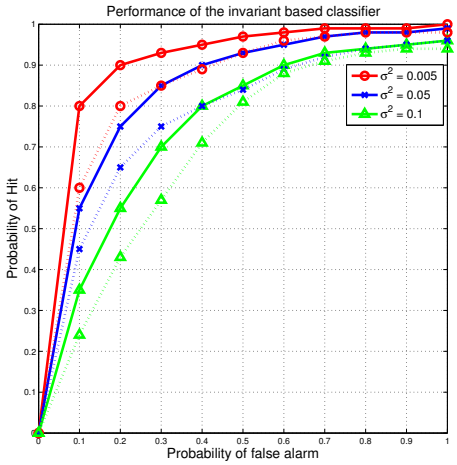

Figure 1.5. Receiver operating characteristic curves at various noise levels in the contour images based on joint projective invariant signatures (solid lines). Compared with Hann and Hickman's method [6; 7] (dashed lines)

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(f). 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, to study the performance of matching in noisy conditions, the receiver operating characteristic curve was generated for various SNR levels as shown in Fig. 1.5. The results are reported for the entire multi-view license-plate dataset available online at [26]. The performance is compared with the method proposed by Hann and Hickman [6; 7] which was found to perform better than Rothwell's canonical frames method [21]. Clearly, our method outperforms the canonical-frame method at all SNR levels. The computing time for the matching algorithm was considerably less than that proposed in [6; 7]. Each comparison with the proposed method took $\sim$4-8s on a 2.4GHz machine.

Case: $m = 3, n = 6$

The Epipolar Geometry Toolbox [12] 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. In other words, 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 [26].

8. Discussion

This chapter presented a new algorithm for curve matching using slices of the joint invariant signature. While invariant-based methods traditionally suffer

from the lack of uniqueness of any single cross ratio value, with high probability image curves can be uniquely described by multiple cross ratio values, chosen judiciously. The novel contributions in this chapter are in computing semi-local, compact descriptors of curves from each image and developing an efficient, distributed algorithm for simultaneous correspondence and matching. Matching results are provided using both simulated data and a license plate dataset.

Notes

1. Note that the manifold lies in $m \times n$ dimensional Euclidean space but since it is parameterized by the curve, the effective dimension is n

References

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

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

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

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

[5] 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.

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

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

[8] 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.

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

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

[11] 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.

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

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

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

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

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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*, 1992.

[22] 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.

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

[24] 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.

[25] 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.

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