Fourier Domain Representation of Planar Curves for Recognition in Multiple Views

Sujit Kuthirummal, C. V. Jawahar and P. J. Narayanan
Centre for Visual Information Technology
International Institute of Information Technology
Gachibowli, Hyderabad, India. 500 019
{sujit@gdit.,jawahar@,pjn@}iiit.net

Abstract

Recognition of a planar shape in different views from the boundary contour is an important problem. Traditional shape recognition deals with views that differ only by simple rotations, translations, and scaling. Shapes suffer more serious deformation between two general views. For instance, two views of a plane are related by a general projective homography. Many such relations between matching primitives in multiple views have been identified recently. In this paper, we explore how shape properties and multiview relations can be combined to recognize planar shapes across multiple views. We propose novel recognition constraints that a planar shape boundary must satisfy in multiple views. The constraints are on the rank of a Fourier-domain measurement matrix computed from the points on the shape boundary. Our method can additionally compute the correspondence between the curve points after a match is established. We demonstrate the applications of these constraints experimentally on a number of synthetic and real images.

Keywords

Computer Vision, Stereo, Projective Geometry, Fourier Transform

1 Introduction

Planar shape recognition has immense applications in surveillance and robotic vision. Two dimensional objects can be recognised based on their boundary information. In many situations, three dimensional objects can also be recognised similarly, by assuming that the images are the orthographic projections of the objects. An example is the recognition of objects from satellite images. The cameras on board the satellites are far away from the objects and therefore any distances within the objects become negligible. The objects, in this case, can be considered to be planar.

Planar shape recognition has received widespread attention in literature for many years. In this class of recognition problems, boundaries of objects are extracted using appropriate and a set of appropriate features are extracted in a suitable domain such that recognition can be carried out successfully. Traditional shape recognition work has, however, concentrated on the situation where the image-to-image transformation between views is limited to translation, rotation, and scaling [1, 2, 3, 4].

A planar object is recognised by comparing it with the set of *a priori* known shapes. Recognition by alignment was attempted by Huttenlocher and Ullman [5]. They computed a match by determining the existence of a transformation that when applied to a model would result in the given view. Comparison can also be carried out by generating a geometric model of the boundary, as is done in algorithms based on polygonal approximation [4]. Linear or other parametric approximations of the boundary need not be appropriate for many situations. This resulted in the development of algorithms based on geometrically invariant features computed out of the discrete set of boundary points [6]. These features can be curvatures, compactness, moments, etc. A third class of algorithms integrate the advantages of both by modelling the boundary in a transform domain like the Fourier one as was done by Zahn and Roskies [1]. In these algorithms, the reference and test images are related by similarity transformations, involving in-plane rotations, translations and scaling. However, the transformation between the reference and test images is more complex in typical problems of interest. When a planar object is imaged from multiple viewpoints, the image-to-image transformation is a general projective homography [7]. The conventional approaches based on Euclidean and similarity frameworks are insufficient in this situation.

There exists a notably different approach for recognition across multiple views. These class of algorithms consider recognition as establishing one-to-one relationships between shapes, in the presence of unknown image-to-image transformations. The first algorithm in this category is due to Ullman and Basri [8] who formulated mechanisms for recognition of objects using linear combination of models for orthographic views. Shashua [9] extended this recognition strategy for perspective cameras. These results hint that the various views of an object lie in a lower dimensional linear subspace. The performance of these algorithms depend on the accuracy of the feature-to-feature correspondences. In this paper, we propose a novel recognition strategy by formulating the problem in the Fourier Domain. Arbter *et al.* [**?**] formulated techniques for affine invariant recognition in the Fourier Domain. Their emphasis was on choosing a suitable set of affine invariant features and performing matching using those features.

In this paper, we present a novel method for recognizing planar shape boundaries in multiple views. The recognition is performed in a Fourier domain representation of the boundary points. We derive recognition constraints satisfied by matching contours using a complex vector representation in the Fourier domain. These are in the form of rank constraints on a measurement matrix computed in the Fourier domain. It is not necessary to know the correspondence between the shapes ahead of time. Our method can, instead, compute the pixel-to-pixel correspondences once the match is established using the rank constraints. The correspondence translates to a shift in the point sequences and can be recovered from the peak of the inverse Fourier transform of an appropriate measure computed in the Fourier domain. Some of this work appeared in an earlier paper [11]. Here we present a comprehensive analysis of the multiview recognition problem under different classes of image-to-image transformations. We present mathematical proof of recognizability constraints for the affine case and present experimental evidence for the projective case.

In Section 2 we present the problem formulation, necessary background, and notation. The complex vector representation of the shape boundary we use and its Fourier domain representation are presented in this section. In Section 3, we present algebraic constraints for recognising shapes in views related by similarity transforms. The theory for an algebraic affine invariant recognition scheme that does not need explicit pixel to pixel correspondence is described in Section 4. Section 5 presents the results of several

experiments on both synthetic and real images. We also show how the recognition constraints for affine homographies hold good for the general situation in practice using a few examples. Section 6 presents a few concluding remarks.

2 Problem Formulation

2.1 Recognition vs Recognisability Constraints

Pattern Recognition is concerned with the grouping of similar feature vectors and assigning an appropriate label to the test sample. The basic assumption has been the existence of a physical process which provides these measurements with a particular probability distribution. Classical pattern recognition algorithms emphasize minimising the misclassification by an appropriate selection of features and classifiers. The emphasis of these approaches has been mainly on recognising similar objects.

The problem of recognising the same object in multiple views is conceptually different from the conventional recognition problem, due to the additional geometric transformation that exists from one image to the other. Moreover, the statistical model of pattern distortion may not be appropriate to characterise the deviation of planar shapes from one image to another. The geometric transformations that exist among multiple views is known precisely in terms of algebraic relations between matching scene primitives – points, lines, etc. [7, 12, 13]. The variability in feature measurements can be restricted, by using available geometric information, providing tighter constraints for recognition. For example, given the image location of a particular world point in two views, the locus of its corresponding point in any view can be expressed in the form of a trilinear relationship [9]. We can come up with recognizability constraints based on algebraic relationships between measurements in multiple views, exploiting the known algebraic multiview relations.

The problem of recognising an object in multiple views may be formally stated as follows : Given a set of L views of an object, identify a view independent function $\mathbf{f}(\cdot)$ such that $\mathbf{f}(\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^{L-1}) = 0$, $\mathbf{x}^l$ being the image measurements made in view l, This recognition constraint can be linear or nonlinear in image coordinates. The algebraic relation given by $\mathbf{f}(\cdot)$ can be used to answer the question whether the L observed views were of the same object. Arriving at such constraints is the focus of this paper.

2.2 Classes of Image-to-Image Homographies

When a planar scene is imaged from multiple view points or when a scene is imaged by cameras having the same optical centre, the images are related by homographies. A homography or a collineation is a mapping from one plane to another such that the collinearity of any set of points is preserved [7]. In other words, a homography is an invertible mapping h from $\mathbb{P}^2$ to itself such that three points x_1, x_2 and x_3 lie on the same line if and only if $h(x_1)$, $h(x_2)$ and $h(x_3)$ do.

Plane-to-plane homographies can be categorised into isometry, similarity, affine and projective [7]. The later classes subsume the earlier ones, i.e., isometry $\subset$ similarity $\subset$ affine $\subset$ projective.

Isometry: An Isometry is a transformation of the plane $\mathbb{R}^2$ that preserves Euclidean distance. Such a transformation is represented as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \epsilon\cos\theta & -\sin\theta & t_x \\ \epsilon\sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where $\epsilon = \pm 1$. If $\epsilon = 1$ then the isometry is orientation preserving and is a Euclidean Transformation. If $\epsilon = -1$ then the isometry reverses orientation and involves a reflection. The above can be expressed more compactly as $\mathbf{x}' = \mathbf{H_E x}$ where $\mathbf{H_E} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0^T & 1 \end{bmatrix}$ where $\mathbf{R}$ is a 2×2 orthonormal rotation matrix and $\mathbf{t}$ is a translational 2-vector.

Similarity: A similarity transformation is an isometry with isotropic scaling. Such a transformation can be written as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

or more compactly $\mathbf{x}' = \mathbf{H_S x}$ where $\mathbf{H_S} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ 0^T & 1 \end{bmatrix}$ and s is the isotropic scaling factor. A similarity transformation is also known as an equi-form transformation as its preserves the shape form.

Affine: An affine transformation is a non-singular linear transformation followed by a linear translation. In the form of a matrix it can be represented as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

or more compactly $\mathbf{x}' = \mathbf{H_A x}$ where $\mathbf{H_A} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ 0^T & 1 \end{bmatrix}$ and $\mathbf{A}$ is a non-singular 2×2 matrix.

Projective: A projective transformation is a general non-singular linear transformation of homogeneous coordinates. This generalizes an affine transformation, which is the composition of a general non-singular linear transformation of inhomogeneous coordinates and a translation.

A projective transformation can be expressed as

$$\begin{aligned} \mathbf{x}' &= \mathbf{H}_P \mathbf{x} \\ \mathbf{H}_P &= \begin{bmatrix} \mathbf{A} & t \\ \mathbf{V}^T & v \end{bmatrix} \text{ and } \mathbf{V} \text{ is a vector } [v_1, v_2]^T \end{aligned} \tag{1}$$

The image-to-image homography is projective (a) when the object being imaged is planar and (b) when the scene is imaged with cameras having the same optical centre.

Figure 1 shows various views of a hexagon under different image-to-image homographies. View (a) is the reference view from which other views were generated using appropriate homographies. Views (a) and

(b) are related by isometric homographies, (c) and (d) by similarity transformations, (e) and (f) by affine homographies, while general projective homographies relate views (g) and (h). It can be seen that all lengths and angles are preserved in the views related by isometries. The hexagons in the views related by similarity transforms look similar (hence the name similarity) with all angles preserved; lengths however are not preserved. In the views related by affine homographies, neither lengths nor angles are preserved, but parallelism is maintained. While, in the views related by projective transformations none of lengths, angles and parallelism are maintained.

2.3 Complex Vector Representation of a Boundary

The notation that we use for the rest of this discussion is given below. Let $\mathbf{O}$ be a set of N points on the boundary of a planar object and let $\mathbf{P}_l$ be its images in views $\mathbf{V}_l$ where l is the view index. Let $(u^l[i], v^l[i], w^l[i])$ be the homogeneous coordinates of points on the closed boundary in view $\mathbf{V}_l$. We represent this shape using a sequence $\mathbf{x}^l[i]$ of complex vectors as given below.

$$\mathbf{x}^l[i] = \begin{bmatrix} u^l[i] + j0 \\ v^l[i] + j0 \\ w^l[i] + j0 \end{bmatrix}$$

We define the Fourier domain representation of the complex vectors as another complex vector $\mathbf{X}^l[k]$ given by

$$\mathbf{X}^l[k] = \begin{bmatrix} U^l[k] \\ V^l[k] \\ W^l[i] \end{bmatrix} \tag{2}$$

where $U^l[i], V^l[i], W^l[i]$ are respectively the Fourier transforms of the sequences $u^l[i], v^l[i], w^l[i]$. The sequence $\mathbf{X}^l[i]$ is periodic and conjugate symmetric, as $\mathbf{x}^l[i]$ is real.

Let the image-to-image transformation of these points from view 0 to view l be given by a 3×3 matrix $\mathbf{M}_l$.

$$\mathbf{x}^l[i] = \mathbf{M}_l \mathbf{x}^0[i] \tag{3}$$

Taking the Fourier transform on both sides we get,

$$\mathbf{X}^l[k] = \mathbf{M}_l \mathbf{X}^0[k] \tag{4}$$

where $\mathbf{X}^0$ and $\mathbf{X}^l$ are the Fourier transform sequences of $\mathbf{x}^0$ and $\mathbf{x}^l$, respectively.

The matrix $\mathbf{M}_l$ in Equation 3 is a homography relating the image planes in views l and 0. The homography $\mathbf{M}_l$ has at most eight degrees of freedom as overall scale is unimportant.

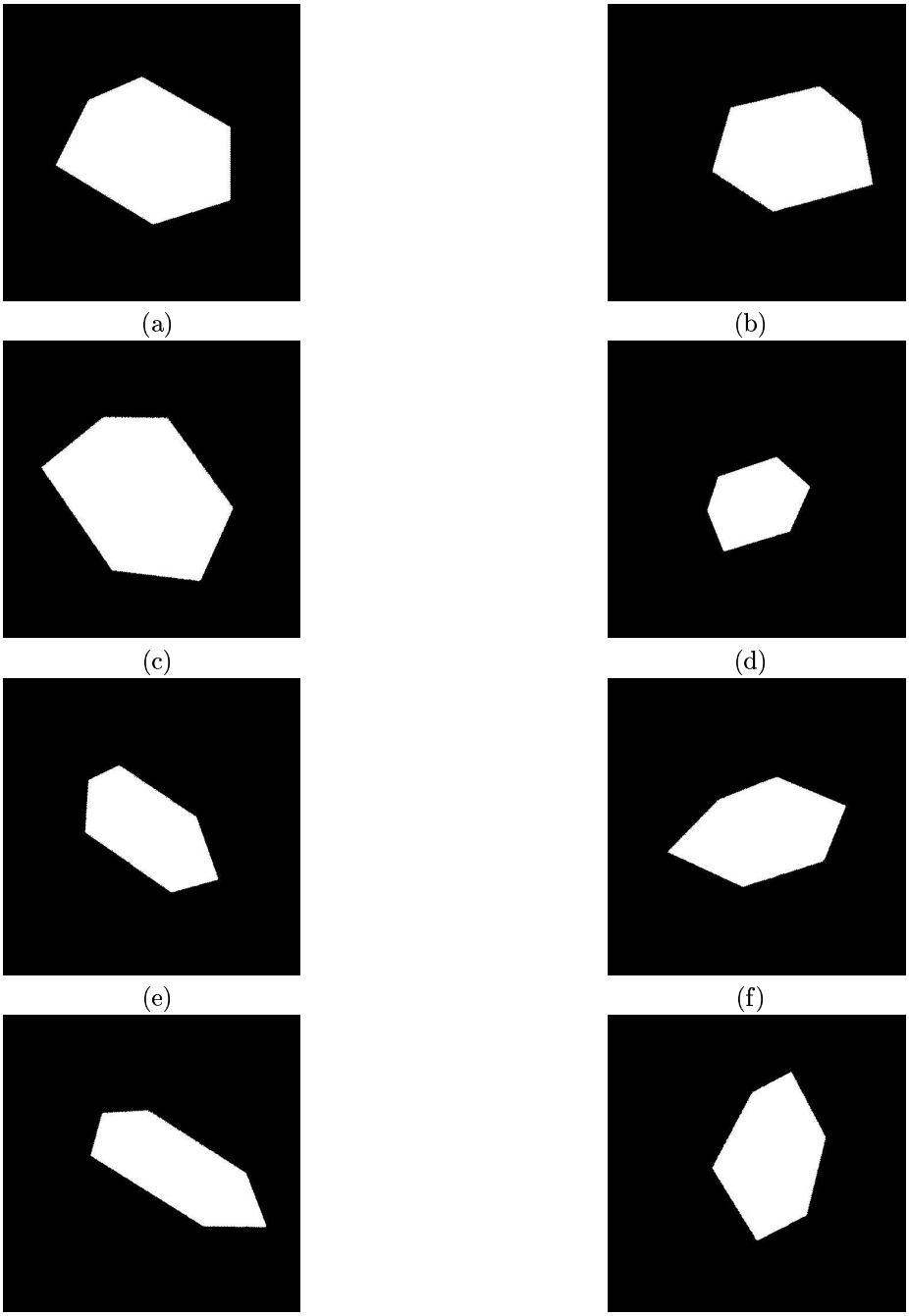

Figure 1: Several views of a hexagon for different image-to-image homographies

2.4 Planar Shape Recognition Problem

We can now formulate the problem formally. The problem of planar shape recognition in two views can be formulated as the identification of the *existence of* an appropriate image-to-image homography between them. Two cases arise based on the information available about the scene or the transformation

1. If the homography is known, recognition involves projection of the reference view into the questioned view and matching or correlating the shapes using an appropriate measure.
2. If point-to-point correspondences are known, the homography can be computed using a suitable number of corresponding points. Matching or recognition can then follow as in the previous case.

The interesting case, however, is when neither the homography nor the correspondence is known. Can we match a planar shape in multiple views if neither the transformation nor the pixel-to-pixel correspondence is known? We offer a few solutions to this question in this paper. In fact, our solution can provide the point to point correspondence and hence the explicit homography matrix as a side products if the boundary curves match. Our solutions are in the form of constraints satisfied by matching boundaries in multiple views. We discuss the simpler case of similarity homography first followed by the more general case of affine homography. We later show that the affine measures work well in practice for general projective homography also.

3 Algebraic Recognisability Constraints for Similarity Homographies

3.1 Without Pixel to Pixel Correspondence

The algebraic relationships between scene primitives – points, lines, conics etc. – in different views that have come to light in recent years are for corresponding features. However, identification of corresponding features in practice is not trivial. In this subsection, we show how the Fourier domain representation is able to achieve recognition based on an algebraic constraint, when correspondence information is not available.

Let us start with the simplest case where we have one view of a shape but its boundary representation starts from a different point each time. M sequences of boundary points are generated by starting the representation in sequence l from a boundary position λ_l away from the starting point in sequence 0. ($\lambda_0 = 0$). Therefore, we have

$$\mathbf{x}^l[i] = \mathbf{x}^0[i + \lambda_l]$$

where λ_l is the unknown shift. A shift in the spatial domain translates into a rotation in the Fourier domain. Taking the Fourier transform of the above expression gives

$$\bar{\mathbf{X}}^l[k] = \bar{\mathbf{X}}^0[k]\, e^{j2\pi\lambda_l k/N}\ ,\ 0 \leq k < N \tag{5}$$

The M boundary representations of the same scene, result in the Fourier vector sequences

$$\bar{\mathbf{X}}^0[k],\ \bar{\mathbf{X}}^1[k],\ \bar{\mathbf{X}}^2[k],\ \ldots,\bar{\mathbf{X}}^{M-1}[k],$$

From Equation 5, it can be seen that $\bar{\mathbf{X}}^l[k]$ has a phase difference of $\frac{2\pi\lambda_l k}{N}$ from $\bar{\mathbf{X}}^0[k]$. Let θ_0, θ_1, ..., θ_{N-1} be the phases of the Fourier coefficients $\bar{\mathbf{U}}^0[k]$ ($\mathbf{U}$ is a component of $\bar{\mathbf{X}}$ from Equation 2). We can form a $M \times N$ measurement matrix Θ with row l consisting of the phase angles of the Fourier coefficients $\bar{\mathbf{U}}^l$ of view l.

$$\Theta = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \theta_3 & \ldots & \theta_{N-1} \\ \theta_0 & \theta_1+\phi_1 & \theta_2+2\phi_1 & \theta_3+3\phi_1 & \ldots & \theta_{N-1}+(N-1)\phi_1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \theta_0 & \theta_1+\phi_{M-1} & \theta_2+2\phi_{M-1} & \theta_3+3\phi_{M-1} & \ldots & \theta_{N-1}+(N-1)\phi_{M-1} \end{bmatrix} \tag{6}$$

where ϕ_l is $\frac{2\pi\lambda_l}{N}$. It can be observed that any row of the above matrix can be expressed as a linear combination of two other rows. For instance, if R_i is the i^{th} row,

$$R_3 = R_1 + (R_2 - R_1)\,\phi_2/\,\phi_1$$

Therefore, Θ is a rank deficient matrix with a fixed rank of 2, irrespective of the number of views M. Therefore the condition for recognition of a shape in such a case is

$$rank(\Theta) = 2 \tag{7}$$

The shift values λ_l can be recovered using the Cross Power Spectrum, which is defined as

$$\frac{\bar{\mathbf{U}}^l[k])^{*\mathrm{T}}\bar{\mathbf{U}}^0[k]}{|(\bar{\mathbf{U}}^l[k])^{\mathrm{T}}\bar{\mathbf{U}}^0[k]|} = e^{\frac{-2\pi j\lambda_l k}{N}}$$

The cross power spectrum is a complex sinusoid. If we take the Inverse Fourier Transform of this sinusoid, it will exhibit a peak at λ_l. (Note: The same result can be achieved using $\bar{\mathbf{V}}$ instead of $\bar{\mathbf{U}}$.)

Recognition of planar shapes under similarity transformation is popular in literature. In the rest of this section we demonstrate recognisability constraints for similarity transformations based on the rank constraint.

3.2 Similarity Transformations

We now show how a Fourier Domain representation is capable of handling the image to image homographies induced by translation, scaling, and rotation.

3.2.1 Translation

The translation transformation would look like

$$\mathbf{x}^l[i] = \mathbf{x}^0[i] + \mathbf{T}_l$$

where $\mathbf{T}_l$ is the translation vector. In the absence of knowledge of correspondence across views this would become

$$\mathbf{x}^l[i] = \mathbf{x}^0[i + \lambda_l] + \mathbf{T}_l$$

where cyclic shifting the order of points in view l by λ_l would align it with the ordering of points in view 0. The Fourier domain form of the above expression is

$$\bar{\mathbf{X}}^l[k] = \bar{\mathbf{X}}^0[k]\, e^{j2\pi\lambda_l k/N} + \delta(0)\mathbf{T}_l \ , \ 0 \leq k < N$$

Ignoring the DC component (spatial frequency of zero) would give

$$\bar{\mathbf{X}}^l[k] = \bar{\mathbf{X}}^0[k]\, e^{j2\pi\lambda_l k/N} \ , \ 0 < k < N \tag{8}$$

The form in Equation 8 is similar to Equation 5 and the same recognition mechanism would be valid. Translating the shape, such that the origin is the centroid of the shape would also provide invariance to translation.

3.2.2 Scaling

The scaling transformation would look like

$$\mathbf{x}^l[i] = \mathbf{M}_l \mathbf{x}^0[i]$$

where

$$\mathbf{M}_l = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

s is the isotropic scaling factor. In the absence of correspondence, scaling becomes

$$\mathbf{x}^l[i] = \mathbf{M}_l \mathbf{x}^0[i + \lambda_l]$$

where cyclic shifting the order of points in view l by λ_l would align it with the ordering of points in view 0. The Fourier domain representation of this is

$$\bar{\mathbf{X}}^l[k] = \mathbf{M}_l \bar{\mathbf{X}}^0[k]\, e^{j2\pi\lambda_l k/N} \tag{9}$$

which in terms of U and V is

$$\bar{\mathbf{U}}^l[k] = s\bar{\mathbf{U}}^0[k]\, e^{j2\pi\lambda_l k/N}$$
$$\bar{\mathbf{V}}^l[k] = s\bar{\mathbf{V}}^0[k]\, e^{j2\pi\lambda_l k/N}$$

The technique described in the previous subsection depends on the phases of $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ and from the above, its evident that the phases are unaffected by the scaling. Hence, we conclude that scaling can be accounted for in this framework.

3.2.3 Rotation

Rotation is yet another important similarity transformation. Rotations can be handled by conventional Fourier Descriptors [4] by representing the boundary points in polar coordinates, as opposed to cartesian coordinates. A point (x, y) in cartesian coordinates can be represented by (r, θ) in the polar coordinates, where $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Rotation of a point assumes the form of a translation in polar coordinates. $(r, \theta) \leftrightarrow (r, \theta + \phi)$, where ϕ represents the angle of rotation. In 3.2.1 we have described how translations can be handled making use of this Fourier domain representation of planar curves. The geometric "distortion" brought about by rotation can be handled using the same technique. The major contribution of this paper is the derivation of an algebraic recognition constraint for affine homographies which we describe in the next section.

4 Algebraic Recognisability Constraints for Affine Homographies

Let us now look at the case when the homography between two views is affine. In this case, the image to image mapping is given by

$$\mathbf{x}^l[i] = \mathbf{M}_l \mathbf{x}^0[i]$$

where

$$\mathbf{M}_l = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ 0 & 0 & 1 \end{bmatrix}$$

The above equation can be rewritten in inhomogeneous coordinates as

$$\begin{aligned} \mathbf{x}^l[i] &= \mathbf{A}_l \mathbf{x}^0[i] + \mathbf{b}_l \\ \bar{\mathbf{X}}^l[k] &= \mathbf{A}_l \bar{\mathbf{X}}^0[k] + \mathbf{b}_l \delta(0) \end{aligned}$$

in the spatial domain and Fourier domain respectively, where $\mathbf{A}_l$ is a 2×2 matrix ($\begin{bmatrix} m_1 & m_2 \\ m_4 & m_5 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$) and $\mathbf{b}_l$ is a translation vector ($\begin{bmatrix} m_3 \\ m_6 \end{bmatrix}$). We can discard the effect of vector $\mathbf{b}_l$ by computing the image coordinates with respect to the centroid of the shape. Discarding the Fourier DC coefficient corresponding to $k = 0$ also has the same effect. In the rest of this discussion, we write this transformation as $\mathbf{x}^l[i] = \mathbf{A}_l \mathbf{x}^0[i]$ in the time domain and $\bar{\mathbf{X}}^l[k] = \mathbf{A}_l \bar{\mathbf{X}}^0[k]$ in the Fourier domain without any loss in generality. The scale factors of the homogeneous representation of the points are assumed to be unity and ignored in the representation of $\mathbf{x}^l$. From here on both $\mathbf{x}^l$ and $\mathbf{X}^l$ are 2×1 matrices of complex numbers.

If pixel-to-pixel correspondences are not known

$$\bar{\mathbf{X}}^l[k] = \mathbf{A}_l \bar{\mathbf{X}}^0[k]\, e^{j2\pi\lambda_l k/N} \tag{10}$$

where λ_l is the unknown shift in view l.

A measure similar to the cross power spectrum can be defined to compute the shift values λ_l in this case also. The cross-conjugate product (CCP) of the Fourier representations of two views is defined as

$$\psi(0,l) = (\bar{\mathbf{X}}^0[k])^{*\mathrm{T}} \bar{\mathbf{X}}^l[k] = (\bar{\mathbf{X}}^0[k])^{*\mathrm{T}} \mathbf{A}_l \bar{\mathbf{X}}^0[k]\, e^{j2\pi\lambda_l k/N}. \tag{11}$$

The measure $\psi(\cdot)$ provides a mechanism for estimation of correspondence and thereby possible recognition. In the next two subsections we study the characteristics of $\psi(\cdot)$ under affine transformations.

4.1 Affine and Symmetric

The measure $\psi(\cdot)$ is quadratic in Fourier coefficients $\bar{\mathbf{X}}^0$. For a set of real vectors, a quadratic form $X^{\mathrm{T}}AX$ is equivalent to $X^{\mathrm{T}}BX$, where B is a symmetric matrix, without any loss in generality. This result is, however, not true for complex vectors $\bar{\mathbf{X}}$.

For a complex vector $\bar{\mathbf{X}} = [(p_1 + jq_1) \quad (p_2 + jq_2)]^{\mathrm{T}}$,

$$\bar{\mathbf{X}}^{*\mathrm{T}} A \bar{\mathbf{X}} = \begin{bmatrix} p_1 & p_2 \end{bmatrix} A \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} q_1 & q_2 \end{bmatrix} A \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} + j\,(a_{21} - a_{12})(p_1 q_2 - p_2 q_1)$$

If A is symmetric, i.e., when $a_{12} = a_{21}$, the imaginary component in the previous expression vanishes and the expression becomes real $= c$.

If correspondence information is not available, $\psi(\cdot)$ becomes a complex sinusoid. $\psi(\cdot) = ce^{\frac{j2\pi\lambda_l}{N}}$. The frequency of this sinusoid is directly related to the shift λ_l in the sequence, which can be determined by looking for a peak in the Inverse Fourier Transform of $\psi(\cdot)$

If we have multiple views, Equation 11 states that the phase of $\psi(0,l)$ differs from the phase of the auto-correlation $\psi(0,0)$ by $\frac{2\pi\lambda_l k}{N}$. The phases of the auto-correlation terms is zero. Hence the phases of the terms in $\psi(0,l)$ is $\frac{2\pi\lambda_l k}{N}$.

If we have M views, then we can form a $M \times (N-1)$ matrix Θ with row l, corresponding to view l consisting of the phase angles of $\psi(0,l)$.

$$\Theta = \begin{bmatrix} \phi_1 & 2\phi_1 & 3\phi_1 & \dots & (N-1)\phi_1 \\ \phi_2 & 2\phi_2 & 3\phi_2 & \dots & (N-1)\phi_2 \\ \phi_3 & 2\phi_3 & 3\phi_3 & \dots & (N-1)\phi_3 \\ \phi_M & 2\phi_M & 3\phi_M & \dots & (N-1)\phi_M \end{bmatrix} \tag{12}$$

where ϕ_l is $\frac{2\pi\lambda_l}{N}$. It is evident that the rows of the matrix differ only by a scale factor. Therefore, Θ is a rank deficient matrix with a fixed rank of 1, irrespective of the number of views. Therefore a necessary condition for recognition in multiple views related by symmetric affine homographies is

$$rank(\Theta) = 1 \tag{13}$$

Experimental Results: To numerically validate the above results, two views of a planar object (an aircraft) were generated with a random symmetric affine image-to-image homography. The IDFT of ψ was computed. This is depicted in Figure 2. The graph shows a distinct and unique peak at the optimal

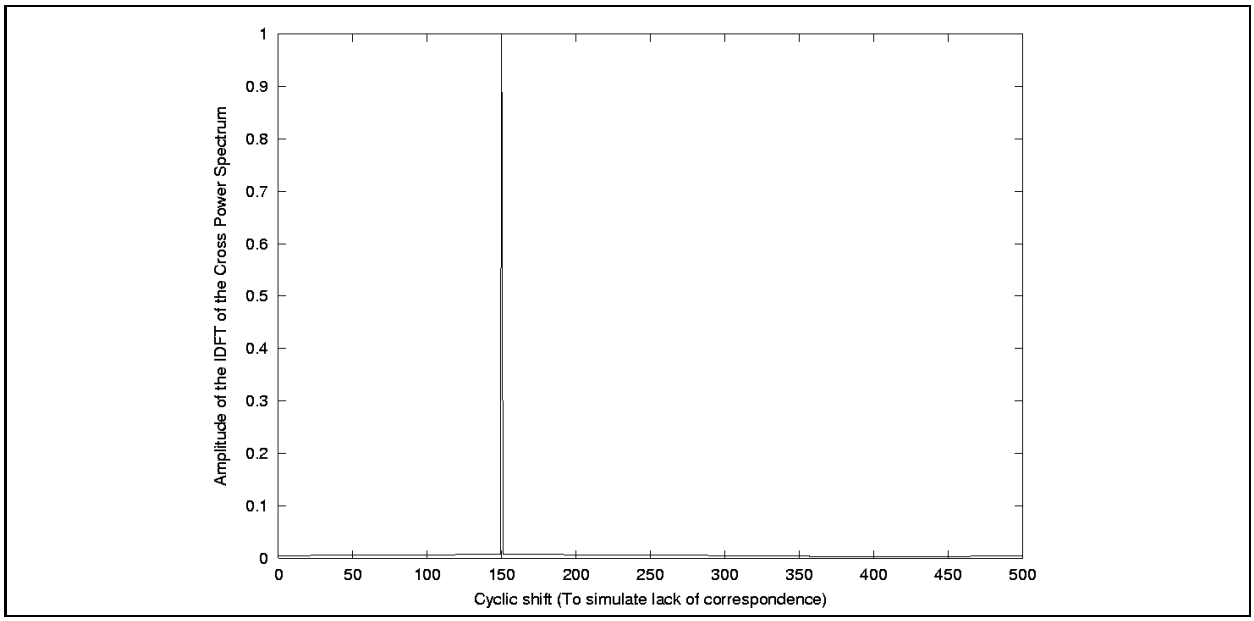

Figure 2: The IDFT of $\bar{\mathbf{X}}^0[k]^{*T}\bar{\mathbf{X}}^l[k]$ when A is symmetric and the shift is 150

λ (150 in this case) to align and recognize the sequences. The rank of matrices was determined using Singular Value Decomposition (SVD). The number of non-zero singular values of a matrix gives the rank of the matrix. In a four-view situation with random symmetric affine transformations and random cyclic shifts in the order of points (to simulate lack of correspondence) the two largest singular values of Θ were 320749 and 0.0575142. The rank of the matrix was practically 1 for all such experiments. This experiment was repeated for various planar shapes with the same result.

If A is not symmetric, $\bar{\mathbf{X}}^{*\mathrm{T}}A\bar{\mathbf{X}}$ will not be real and ψ will no longer be a pure complex sinusoid. For a random non-symmetric affine homography, the above experiment was repeated. The magnitude spectrum of the IDFT of $\psi(0,l)$ is shown in Figure 3. Interestingly, the magnitude spectrum has a peak at the right shift value, though the graph is noisy.

A series of experiments were carried out to study the performance of this technique when the affine homography is not symmetric. A plot of the ratio of the second highest singular value of Θ to the highest singular value against the ratio of the off diagonal elements is shown in Figure 4.

4.2 General Affine

It is well known that any square matrix can be expressed as a sum of a symmetric and a skew symmetric matrix. We can decompose the matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{A}_s + \mathbf{A}_{sk}$$

where

$$\mathbf{A}_s \quad = \quad \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}}), \text{ is symmetric and}$$

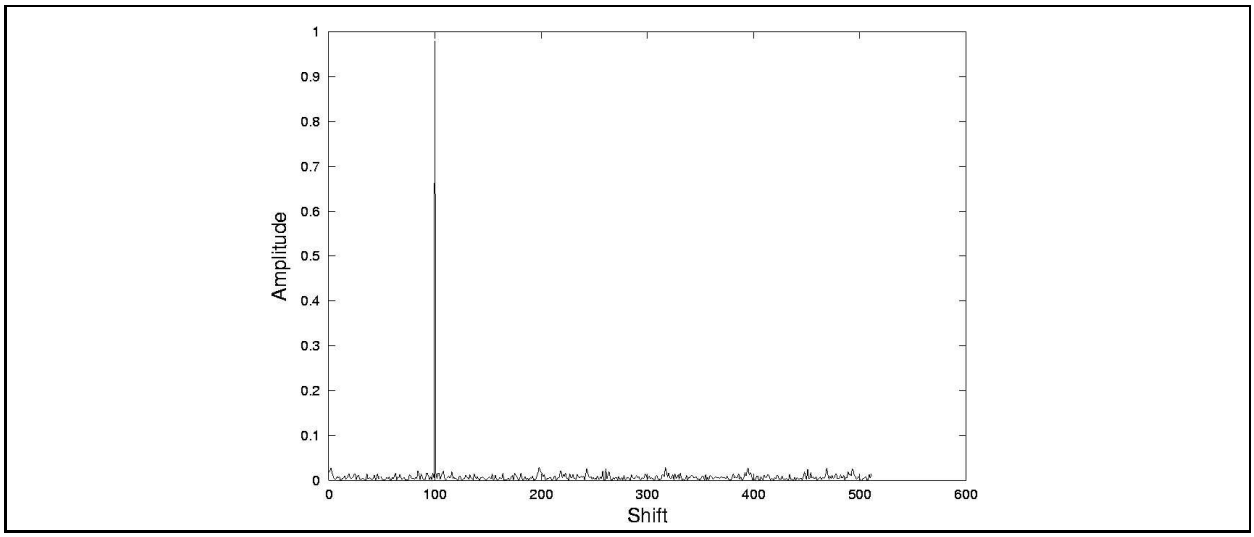

Figure 3: The IDFT of $\bar{\mathbf{X}}^0[k]\bar{\mathbf{X}}^l[k]$ when A is asymmetric and the shift is 100

$$\mathbf{A}_{sk} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}}) \text{ is skew symmetric.}$$

In the case of an affine homography relating images of a planar shape, the skew symmetric matrix will reduce to

$$\begin{aligned} &= \begin{bmatrix} 0 & a_{12} - a_{21} \\ a_{21} - a_{12} & 0 \end{bmatrix} \\ &= c \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \end{aligned}$$

where $c = a_{12} - a_{21}$. We can now write Equation 11 as

$$\psi(0,l) = \bar{\mathbf{X}}^{*\mathrm{T}}\mathbf{A}\bar{\mathbf{X}}\, e^{-j2\pi\lambda_l k/N} = \bar{\mathbf{X}}^{*\mathrm{T}}\left(\mathbf{A}_s + c\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right)\bar{\mathbf{X}}\, e^{-j2\pi\lambda_l k/N} = \psi_1 + \psi_2. \tag{14}$$

The term $\bar{\mathbf{X}}^{*\mathrm{T}}\mathbf{A}_s\bar{\mathbf{X}}$ of the above equation is purely real and the term $\bar{\mathbf{X}}^{*\mathrm{T}}c\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\bar{\mathbf{X}}$ – which corresponds to a rotation by 90° followed by scaling by c – is purely imaginary. The phases of ψ_1 and ψ_2 depend only on the shift λ_l. Thus, λ_l can be recovered from the inverse Fourier transform of ψ_1 or ψ_2, if known. However, we can only compute $\psi(0,l)$, a combination of ψ_1 and ψ_2, which is not useful to recover the shift.

We observe that the effect of the transformation matrix $\mathbf{A}$ in ψ_2 is restricted to a scaling factor c. We ignore c, and define a new measure κ for the sequence $\bar{\mathbf{X}}^l$ as

$$\kappa(l) = \bar{\mathbf{X}}^l[k]^{*\mathrm{T}}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\bar{\mathbf{X}}^l[k]. \tag{15}$$

It can be shown that

$$\kappa(l) = (\bar{\mathbf{X}}^l[k])^{*\mathrm{T}}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\bar{\mathbf{X}}^l[k]$$

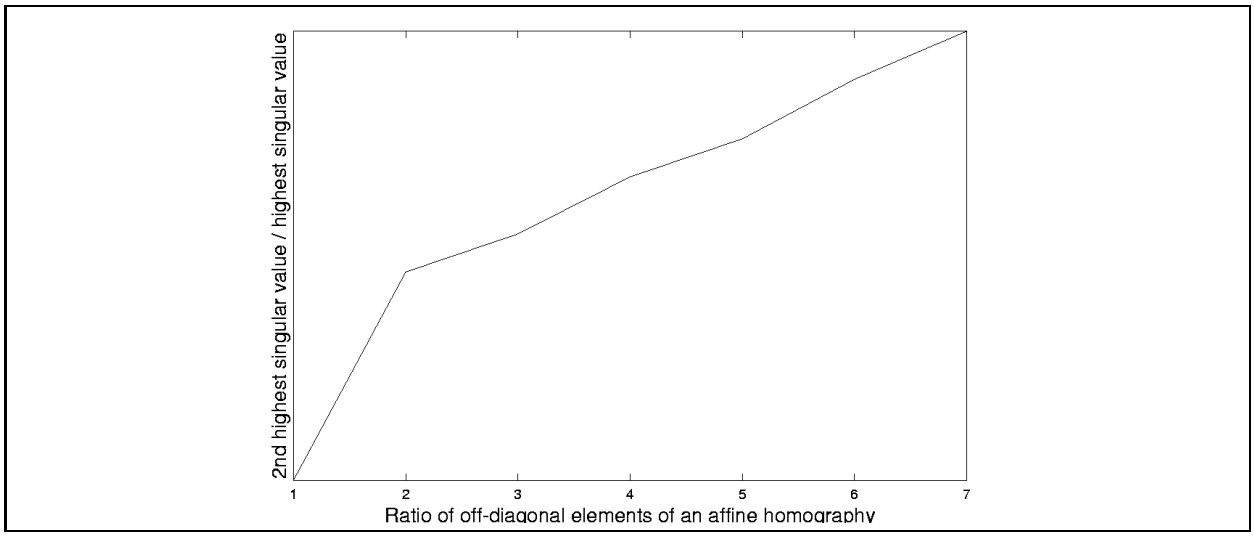

Figure 4: Plot of the ratio of the the second highest singular value of Θ to the highest singular value vs the ratio of the off-diagonal elements of an affine homography (Θ is computed using the technique for symmetric affine homographies)

$$\begin{aligned} &= (A\bar{\mathbf{X}}^0[k]\, e^{j2\pi\lambda_l k/N})^{*\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} A\bar{\mathbf{X}}^0[k] e^{j2\pi\lambda_l k/N} \\ &= (\bar{\mathbf{X}}^0[k])^{*\mathrm{T}}\, A^{\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} A\, \bar{\mathbf{X}}^0[k]\, e^{-j2\pi\lambda_l k/N} e^{j2\pi\lambda_l k/N} \\ &= |A|\, \kappa(0) \end{aligned} \tag{16}$$

Equation 16 gives a necessary condition for the sequences $\bar{\mathbf{X}}^l$ and $\bar{\mathbf{X}}^0$ to be two affine-transformed views of the same planar shape, namely, that the coefficients of the measure κ should be scaled versions of each other. This extends to multiple views also. Consider the $M \times (N-1)$ matrix formed by the coefficients of the κ measures for M different views.

$$\Theta = \begin{bmatrix} \kappa(0)[1] & \cdots & \kappa(0)[N-1] \\ \kappa(1)[1] & \cdots & \kappa(1)[N-1] \\ \cdots & \cdots & \cdots \\ \kappa(M-1)[1] & \cdots & \kappa(M-1)[N-1] \end{bmatrix} \tag{17}$$

The necessary condition for matching of the planar shape in M views then reduces to

$$\mathrm{rank}(\Theta) = 1. \tag{18}$$

It should be noted that the recognition condition does not require correspondence between views and is valid for any number of views.

Equation 16 eliminates the shift λ_l from the recognition condition. How can we also recover the shift corresponding to each view if the boundaries match? We can modify the definition of κ as below.

$$\kappa^l(l,p) = (X^l[k])^{*\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} X^l[p] \tag{19}$$

The measure $\kappa'(\cdot)$ correlates each Vector Fourier Coefficient with a fixed one within each view (p). Following reasoning similar to Equation 16, we can show that

$$\begin{aligned}\kappa'(l,p) &= (X^l[k])^{*\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} X^l[p] \\ &= |A|\ \kappa'(0,p)\ e^{-j2\pi\lambda_l(k-p)/N}. \end{aligned} \tag{20}$$

Equation 20 states that the phases of $\kappa'(l,p)$ and $\kappa'(0,p)$ differ by an amount proportional to the shift λ_l and the differential frequency $k-p$. Therefore, the ratio $\frac{\kappa'(l,p)}{\kappa'(0,p)}$ will be a complex sinusoid $ce^{-j2\pi\lambda_l(k-p)/N}$. The value of λ_l can be computed from the inverse Fourier transform of the quotient series.

We can also form a $M \times (N-1)$ matrix Θ', similar to the one above, that stacks the phases of $\kappa'(l,1)$ (taking $p=1$). It will have the form

$$\Theta' = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \dots & \theta_{N-1} \\ \theta_1 & \theta_2+\phi_1 & \theta_3+2\phi_1 & \dots & \theta_{N-1}+(N-2)\phi_1 \\ \dots & \dots & \dots & \dots & \dots \\ \theta_1 & \theta_2+\phi_{M-1} & \theta_3+2\phi_{M-1} & \dots & \theta_{N-1}+(N-2)\phi_{M-1} \end{bmatrix} \tag{21}$$

where θ_i are the phases of $\kappa'(0,1)$ and $\phi_l = -2\pi\lambda_l/N$. This matrix will have a rank of 2 irrespective of M. The rank constraint on the above matrix, which is a necessary condition for recognition of shapes in views related by affine image-to-image homographies, is

$$\mathrm{rank}(\Theta') = 2. \tag{22}$$

We have defined two necessary conditions κ and κ' for affine invariant recognition. Recognition using κ involves finding the rank of a matrix, while κ' can also be used to compute point-to-point correspondence between corresponding shapes.

In the next section we present the results of a number of experiments that we had conducted to verify our claims.

5 Experimental Results and Discussion

Experiments were first conducted on synthetic views and then on real images. Figure 1 shows a synthetic hexagon under various image-to-image homographies. For experiments on synthetic images, two kinds of boundary representations were considered – when the points on the boundary are described using real coordinates (floating point numbers) and when the locations of the boundary points are in terms of integer coordinates. Boundary representation using real coordinates preserves the mathematical basis of the formulations discussed above and the rank constraint can be strictly enforced. When the boundary representation is in the form of integer coordinates, discretization noise introduces 'errors' that make the rank constraint an approximation, but nonetheless enforceable. The real images were taken using a Sony digital camera and had dimensions of 1024×768. From these images, the objects of interest were segmented out and their boundaries sampled to have 1024 boundary points.

5.1 Isometry and Similarity Homographies

The views (a) and (b) of Figure 1 are related by isometries, while (c) and (d) are related by similarity homographies. The performance of the κ measure is analysed for these. The ratio of the highest singular value to the next highest singular value of the Θ matrix of κ values (Equation 17) of these two views was found to be very high, and hence the rank can be considered to be 1. Both cases were considered – when the points on the boundary are real values and when the points are discretized. Table 1 shows the performance for both cases.

	Isometry		Similarity	
Boundary Points	Highest	Next	Highest	Next
Real	256420	0.00386031	256310	0.00486031
Discrete	256423	1.96133	256398	4.10475

Table 1: Singular values of the Θ matrix of κ measures of views (of the hexagon) related by isometry and similarity homographies

5.2 Affine Homographies

In this subsection we demonstrate the application of the κ measure on views related by affine image-to-image homographies. Views (e) and (f) of Figure 1 are related by affine homographies. The two greatest singular values of the Θ matrix of the κ measures for these two views for both real and discrete representations of the boundary are given in Table 2.

	Singular Values	
Boundary Points	Highest	Next
Real	231124	0.00817599
Discrete	231123	2.91271

Table 2: Singular values of the Θ matrix of κ measures of views (e) and (f) of Figure 1

The next set of experiments were performed on the boundary of an aircraft. Four views of an aircraft related by affine homographies are shown in Figure 5. The shape boundaries in the various views were sampled so that each shape was represented by 1024 points.

The Θ matrix for all the four views was formed using the κ measures for each view as described earlier. The rank of this matrix Θ was found to be 1 using SVD, as the largest two singular values were 247476 and 0.00186574.

Experiments were then conducted on the boundary of the logo of the International Institute of Information Technology. Four views were generated using random affine image-to-image homographies. Tables 3 and 4 present the ratio of the highest singular value to the second highest singular value for various combinations of views shown in Figure 6 for real and discretized boundary descriptions respectively.

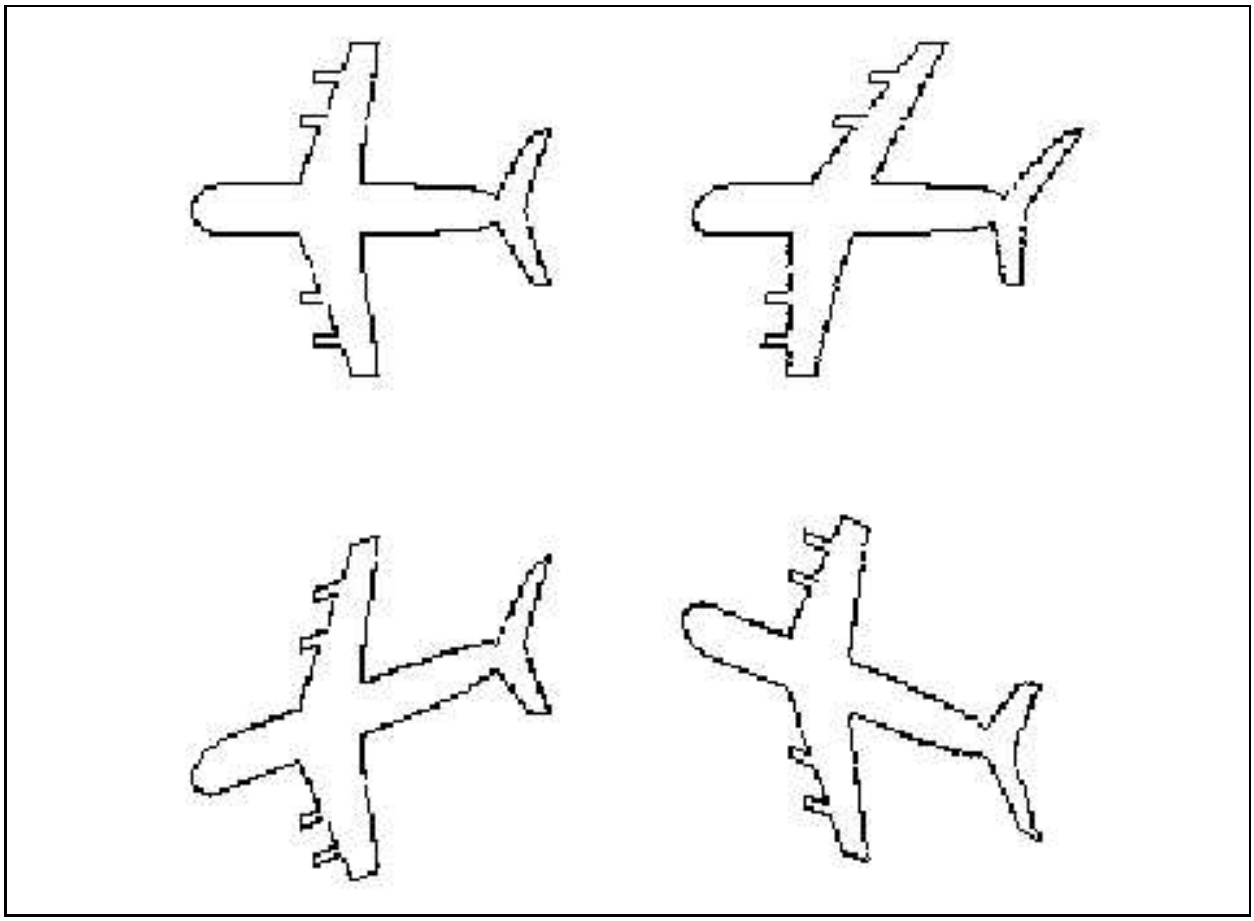

Figure 5: Four affine-transformed views of an aircraft

When all the four views were stacked to form the Θ matrix the ratio of the highest singular value to the second highest singular value was 5.53749e+06 and 738.366 for the real and discrete boundary representations respectively.

Determining Point Correspondence: We tested the effectiveness of our technique for estimating correspondences through the shift λ_l. Figures 7 and 8 show the inverse Fourier spectrum of the ratio $\kappa'(l,1)/\kappa'(0,1)$, when the shifts aligning corresponding points in the two affine views are 150 and 300 respectively.

We have achieved recognition between two planar shapes under the assumption that the homography between them has a specific form, without knowing the correspondence between points. We were also able to estimate the correspondence.

Views	a	b	c	d
a	-	2.94429e+07	1.91431e+07	5.09852e+06
b	2.94429e+07	-	9.19418e+07	4.63504e+0
c	1.91431e+07	9.19418e+07	-	4.14435e+06
d	5.09852e+06	4.63504e+06	4.14435e+06	-

Table 3: Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of views shown in Figure 6 for real point boundary descriptions

Views	a	b	c	d
a	-	6913.86	880.174	1615.47
b	6913.86	-	1698.57	1424.83
c	880.174	1698.57	-	598.581
d	1615.47	1424.83	598.581	-

Table 4: Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of views shown in Figure 6 for discretized boundary descriptions

5.3 Projective Homographies

In Section 4.1 we had seen how the theory developed for a symmetric affine image-to-image homography worked in practice for an asymmetric affine image-to-image homography. In this subsection we show that the above theory developed for general affine image-to-image homographies described in Section 4.2 works well for reasonable projective homographies.

Though the homography relating two views of a plane is projective in general, an affine approximation seems to be sufficient for most practical cases as a number of real life configurations of imaging a scene from multiple view points, possess structure that are very similar to that of affine homographies. In fact, a random projective homography would cause such a distortion in the shape that it would be near impossible for the human eye to recognise the two shapes to be the same! This approximation of a perspective camera by a weak perspective camera - an affine camera is quite popular [14]. We conducted experiments to examine the validity of this approximation, the results of which are presented next. We first present the results of experiments on synthetic data.

Experiments on Synthetic Images The views (g) and (h) of Figure 1 are related by projective homographies. The two highest singular values of the Θ matrix of κ measures for these two views are given in Table 5 for both cases when the boundary description is in terms of real points and when the boundary representations are discretized.

	Singular Values	
Boundary Points	Highest	Next
Real	216817	110.982
Discrete	216803	113.986

Table 5: Singular values of the Θ matrix of κ measures of views (g) and (h) of Figure 1

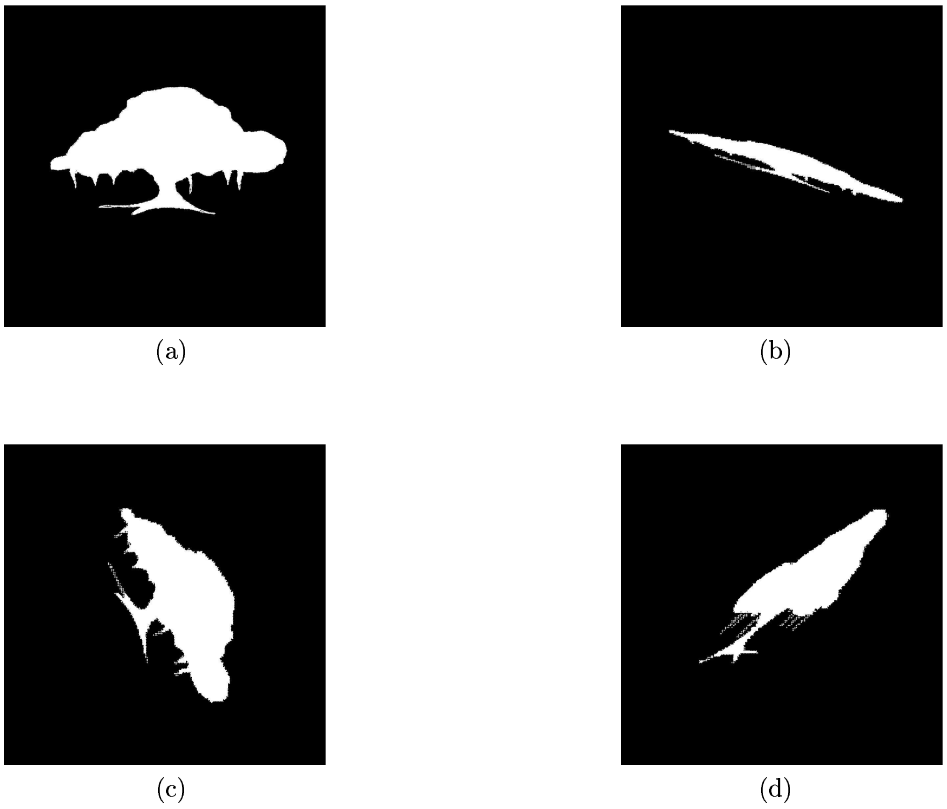

Figure 6: Four views of IIIT's Logo related by random affine image-to-image homographies

Next we present the results of experiments on real images.

Experiments on Real Images Figure 9 shows three views of the logo of the International Institute of Information Technology (IIIT). The boundaries of the logo in the various views were extracted and sampled to have 1024 points. Tests were then carried out to determine the efficacy of the κ measure in determining whether all the views were of the same object. Table 6 shows the ratio of the highest singular value to the second highest singular value of the Θ matrix of κ measures for various combinations of views. As can be seen from the table. the highest singular value is greater than the second highest by more than an order of 2 and so the rank of the Θ matrix is essentially 1.

Views	a	b	c
a	-	431.048	505.847
b	431.048	-	292.71
c	505.847	292.71	-

Table 6: Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of views shown in Figure 9.

Similar experiments were carried out on a number of shapes. Here we show results of experiments

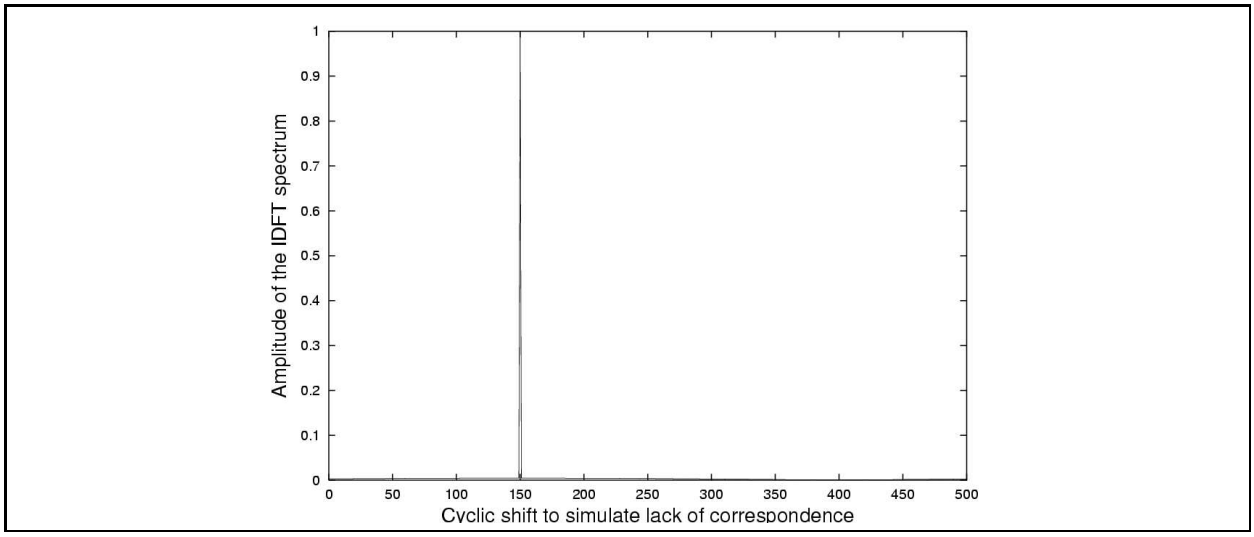

Figure 7: Graph showing the amplitude of the IDFT of $\frac{\kappa'(l,1)}{\kappa'(0,1)}$ against the shift for an affine homography when the synthetic shift is 150.

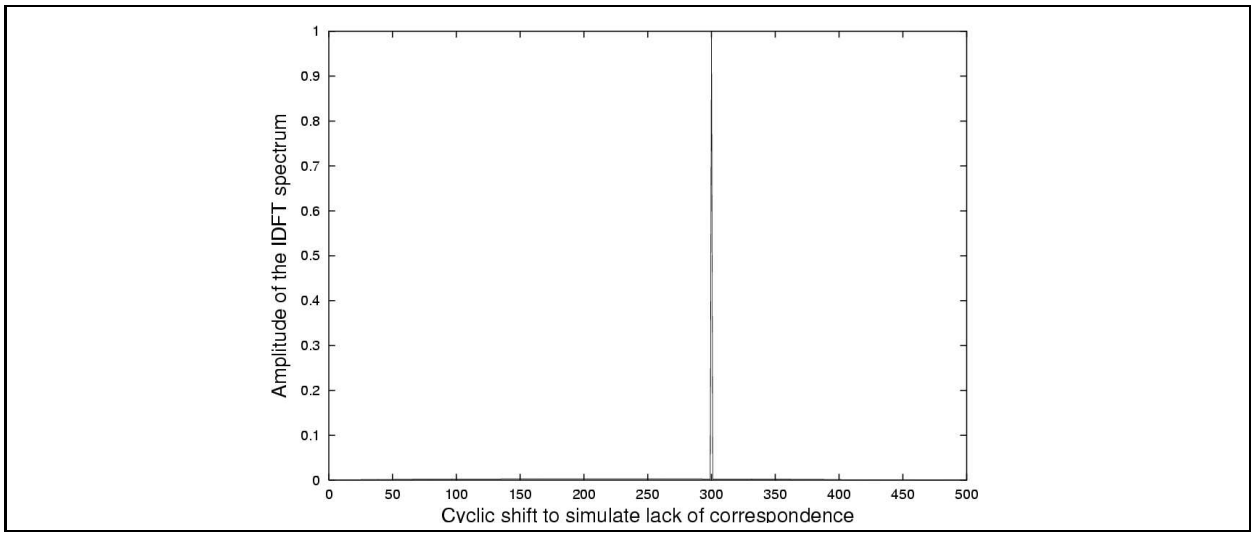

Figure 8: Graph showing the amplitude of the IDFT of $\frac{\kappa'(l,1)}{\kappa'(0,1)}$ against the shift for an affine homography when the synthetic shift is 300.

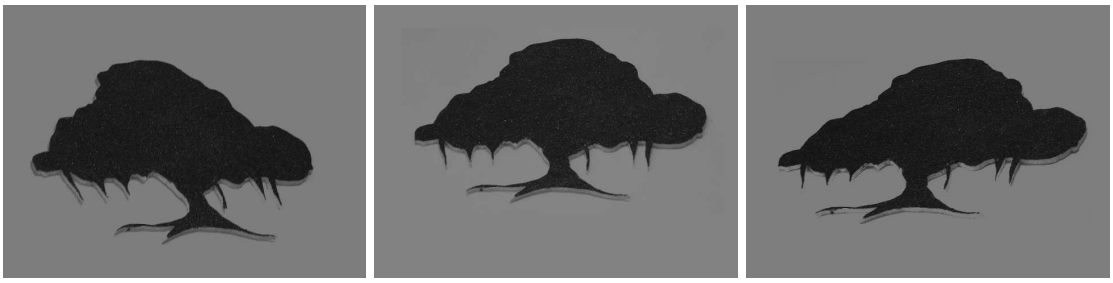

Figure 9: Three views of the logo of IIIT

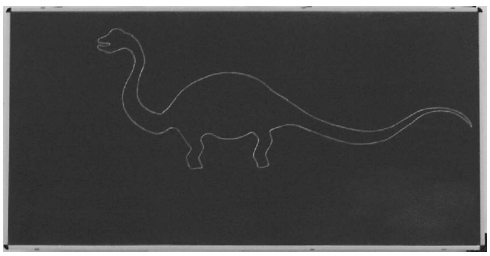

Figure 10: A view of a dinosaur

conducted on four views each of four planar shapes of – a dinosaur, a lizard, a hammer and a floppy (one view of each are shown in Figures 10, 11, 12, and 13). The performance of the κ measure – the ratio of the highest singular values for different combinations of views for each shape are shown in Tables 7, 8, 9, and 10 respectively. These results indicate that the rank of the Θ matrix is essentially 1.

The rank of the Θ matrix obtained on stacking the κ measures for all the views of the same shape was also found to be 1 as can be observed from the results shown in Table 11, which gives the two highest singular values and the ratio of the highest singular value to the second highest singular value of the Θ matrix for all views of the same shape, for the shapes shown in Figures 9, 10, 11, 12, and 13. The highest singular value is greater than the second highest singular value by more than order of 2.

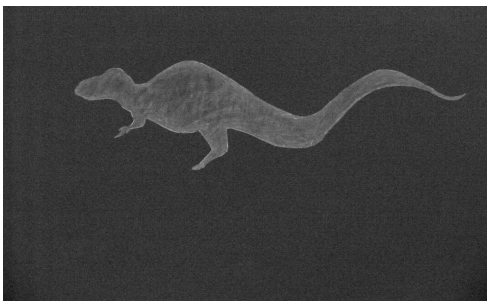

Figure 11: A view of a lizard

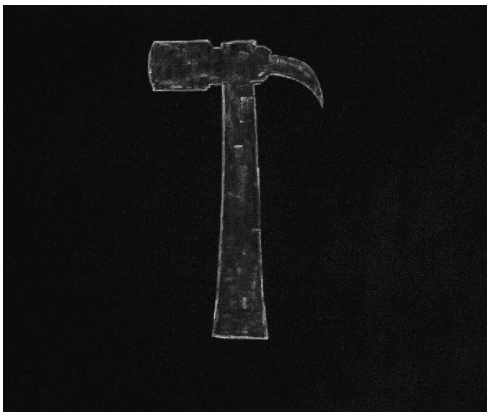

Figure 12: A view of a hammer

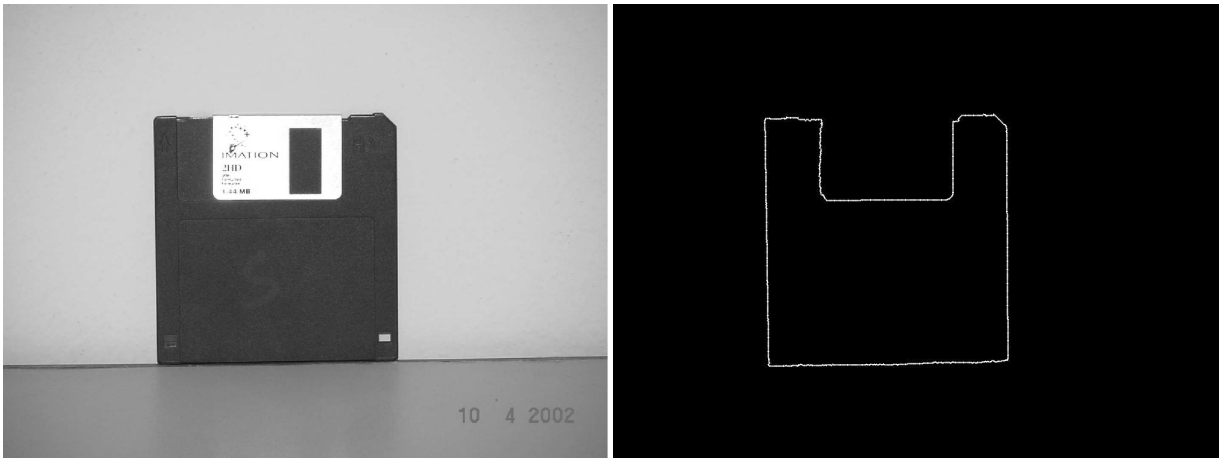

Figure 13: A view of a floppy disk and the extracted boundary

Views	a	b	c	d
a	-	503.341	149.383	256.191
b	503.341	-	123.839	239.291
c	149.383	123.839	-	157.725
d	256.191	239.291	157.725	-

Table 7: Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of 4 views (a-d) of the shape shown in Figure 10

.

Views	a	b	c	d
a	-	265.959	208.302	100.214
b	265.959	-	379.928	107.613
c	208.302	379.928	-	136.592
d	100.214	107.613	136.592	-

Table 8: Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of 4 views (a-d) of the shape shown in Figure 11

5.4 Discriminatory Power

To examine the capability of this technique to distinguish between shapes, tests were carried out to evaluate the κ measure for views of different shapes. A view each of four objects were chosen - IIIT's logo(9), a dinosaur(10), a lizard(11), and a floppy(13). The ratio of the highest singular value to the next highest singular value of the Θ matrix for various combinations of views (shapes) is shown in Table 12.

It is interesting to observe that the dinosaur and lizard shapes exhibit greater similarity to each other than other shapes.

5.5 Robustness of Recognition

Affine Image-to-Image Homographies: We now study the recognition accuracy when a zero mean random noise is added to the position of the synthetically transformed shape for an affine homography. The highest two singular values for different maximum noise levels are shown in Table 13. The ratio of the highest to the next highest singular values does suffer, but there was still more than an order of magnitude separation between the top two even with a noise of 20% in the positions of the boundary

Views	a	b	c	d
a	-	127.263	224.673	578.354
b	127.263	-	245.475	128.397
c	224.673	245.475	-	257.312
d	578.354	128.397	257.312	-

Table 9: Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of 4 views (a-d) of the shape shown in Figure 12

Views	a	b	c	d
a	-	829.15	1103.81	878.434
b	829.15	-	1399.44	564.725
c	1103.81	1399.44	-	847.364
d	878.434	564.725	847.364	-

Table 10: Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of 4 views (a-d) shown in Figure 13

	Singular Values		
Shape	Highest	Next Highest	Ratio of the two highest
IIIT Logo	1.026e+06	2878.12	356.482
Dinosaur	641940	3907.43	164.287
Lizard	786130	7065.8	111.258
Hammer	1.029e+06	8375.58	122.857
Floppy	1.203e+06	1196.99	1005.021

Table 11: The two highest singular values and their ratio of the Θ matrix obtained by stacking the values of the κ measure of all views of the same shape (shown in 9, 10, 11, 12 and 13)

points.

Projective Image-to-Image Homographies: When a zero mean random noise is added to the positions of the points on the boundaries of the shape in the various views, the performance of the κ measure deteriorates, but even with a noise of 20% in the positions of boundary points, there is more than an order of separation between the two highest singular values of the matrix of κ measures, as is demonstrated by Table 14

Clearly, the recognition is excellent in all cases with the degradation in performance along expected lines.

Experiments were also conducted to obtain an idea of how the performance of the κ measure falls with an increase in the projective component. A plot showing how the performance of the κ measure falls with an increase in the projective component v_2 (Equation 1) keeping v_1 fixed, is shown in Figure 14.

Views	IIIT Logo	Dinosaur	Lizard	Floppy
IIIT Logo	-	9.95088	12.1399	18.1149
Dinosaur	9.95088	-	35.7616	14.9745
Lizard	12.1399	35.7616	-	20.1782
Floppy	18.1149	14.9745	20.1782	-

Table 12: Discriminatory Power : Ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of shapes.

	Real		Discrete	
Noise	Singular Values		Singular Values	
Level	Highest	Next	Highest	Next
0	247476	0.00186574	213036	73.0211
0.5%	232918	63.6448	229286	124.335
3%	211296	356.347	228500	483.168
5%	208896	839.34	209417	1233.88
10%	193925	1424.26	197214	2069.28
15%	190745	2324.85	176999	3251.64
20%	180199	3887.51	166523	4931.72

Table 13: Impact of noise on singular values of the matrix of κ measures for floating point (real) and integer (discrete) representations of boundary points in views related by affine image-to-image homographies

Noise	Singular Values		
Level	Highest	Next	Ratio of the highest
0	1.02679e+06	2878.12	356.757
0.5%	1.0268e+06	2996.6	342.655
3%	1.02323e+06	3601.61	284.103
5%	1.01308e+06	3689.42	274.590
10%	982443	3920.95	250.562
15%	923793	6235.78	148.143
20%	854356	14580.3	58.597

Table 14: Impact of noise on singular values of the matrix of κ measures in real images of the same planar object imaged from multiple view points

6 Conclusions

We formulated Fourier domain constraints combining shape properties and multiview relations for planar shape recognition in this paper. These constraints serve as the necessary recognition conditions for a planar shape in multiple views. Our method does not need to know any corresponding points on the shape boundary. The Fourier domain measurement matrices used for recognition are based on simple measures that can be computed easily. The recognition constraints are rank constraints on these measurement matrices We derive the conditions for the cases of the image-to-image transformations between the multiple views being similarity or affine. The method works well experimentally when the image-to-image transformation is projective.

It is now possible to answer if the shape observed in multiple cameras correspond to the same world object, without requiring correspondence between points on the shape. The correspondence information can be calculated if the shapes are seen to be matching. The experiments demonstrate our algorithm works well for most common types of multiview situations.

We propose to work on extending the philosophy to general projective image-to-image transformations. Also of interest is the multiview properties of other collections of primitives, such as textures and 3D shape . These collections have properties that hold in each view individually. Multiview relations lay

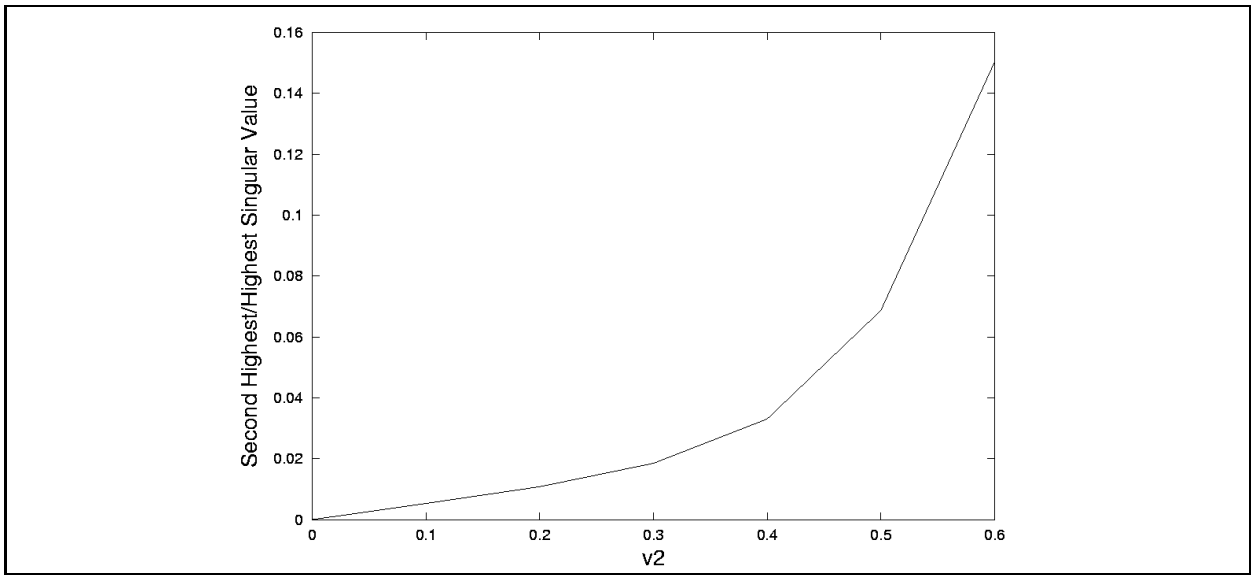

Figure 14: Graph showing the variation in the performance of the κ measure as the projective component v_2 (Equation 1)is increased

down conditions on multiple views of the same primitives. Thus, their combination has great potential in deriving rich constraints and invariants that can aid in matching, recognition, `etc.`

References

[1] C. Zahn and R. Roskies, "Fourier Descriptors for Planar curves," *IEEE Trans. Comput.*, vol. C-21, 1972.

[2] T. Crimmins, "A complete set of Fourier descriptors for two dimensional shapes," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-12, pp. 195–201, 1982.

[3] G. Granlund, "Fourier preprocessing for hand printed character recognition," *IEEE Trans. Comput.*, vol. C-21, 1972.

[4] T. Pavlidis, *Structural Pattern Recognition.* Springer-verlag, 1977.

[5] D. P. Huttenlocher and S. Ullman, "Object Recognition using Alignment," *Proc. International Conference on Computer Vision*, pp. 102–111, 1987.

[6] A. K. Jain, *Fundamentals of Digital Image Processing.* Prentice-Hall, 1989.

[7] R. Hartley and A. Zisserman, *Multiple View Gemoetry in Computer Vision.* Cambridge University Press, 2000.

[8] S. Ullman and R. Basri, "Recognition by Linear Combination of Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, pp. 992–1006, 1991.

[9] A. Shashua, "Trilinear tensor: The Fundamental Construct of Multiple-view Geometry and its applications," *Int. Worskshop on AFPAC*, 1997.

[10] K. Arbter, "Affine-invariant Fourier descriptors, in From Pixels to Features," *Elsevier Science*, 1989.

[11] Sujit Kuthirummal, C. V. Jawahar, and P. J. Narayanan, "Planar Shape Recognition across Multiple Views," *International Conference on Pattern Recognition*, 2002.

[12] O. Faugeras and Q. Luong, *The Geometry of Multiple Images*. MIT Press, 2001.

[13] J. Mundy and A. Zisserman, *Geometric Invariances in Computer Vision*. MIT Press, 1992.

[14] L. Shapiro, A. Zisserman, and M. Brady, "3D Motion Recovery via Affine Epipolar Geometry," *IJCV*, vol. 16(2), pp. 147–182, 1995.