PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Pattern Recognition 37 (2004) 739–754

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Fourier domain representation of planar curves for recognition in multiple views☆

Sujit Kuthirummal, C.V. Jawahar*, P.J. Narayanan

Centre for Visual Information Technology, International Institute of Information Technology, Gachibowli, Hyderabad 500 019, India

Received 18 September 2002; accepted 4 June 2003

Abstract

Recognition of planar shapes is an important problem in computer vision and pattern recognition. The same planar object contour imaged from different cameras or from different viewpoints looks different and their recognition is non-trivial. Traditional shape recognition deals with views of the shapes that differ only by simple rotations, translations, and scaling. However, shapes suffer more serious deformation between two general views and hence recognition approaches designed to handle translations, rotations, and/or scaling would prove to be insufficient. Many algebraic 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: Planar shape recognition; Fourier transform; Computer vision; Affine homography; Projective geometry

1. Introduction

Planar shape recognition has immense applications in surveillance, robotic vision, etc. Two-dimensional objects can be recognised based on their boundary information. In some 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. The objects can then be considered to be planar as any distances within the objects become negligible.

Planar shape recognition has received widespread attention in literature for many years. In this class of recognition problems, boundaries of objects are extracted using relevant techniques and a set of appropriate features are extracted in a suitable domain such that recognition can be carried out successfully. Many planar object recognition efforts have been reported earlier for the simple case of similarity transformation between views [1–4]. A planar object can be 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 can also be used. Algorithms based on computing

☆ This work was partially supported by the Advanced Data Processing Research Institute, Department of Space, Government of India.

* Corresponding author. Fax: +91-33-23001413.

E-mail addresses: sujit@gdit.iiit.net (S. Kuthirummal), jawahar@iiit.net (C.V. Jawahar), pjn@iiit.net (P.J. Narayanan).

0031-3203/$30.00 © 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2003.06.002

geometrically invariant features from the discrete set of boundary points have also been developed [6]. These features can be curvatures, compactness, moments, etc. Another 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. 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. One of the algorithms 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. This result hints 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. Arbter et al. [9] 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 recognising planar shape boundaries in multiple views. We derive recognition constraints satisfied by matching contours using a complex vector representation of the boundary points 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. Preliminary results were shown in an earlier paper [10]. 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. 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 focus on 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. 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. [7,11,12]. 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 one view, the locus of its corresponding point in another view can be expressed in the form of a bilinear relationship [13,14]. Similar relations hold for three and more views of points and lines [15–19]. We can come up with recognizability constraints based on algebraic relationships between measurements in multiple views, exploiting the known algebraic multiview relations.

A multiview recognizability constraint can be defined as follows. Given a set of M views of an object, identify a view independent function $\mathbf{f}(\cdot)$ such that $\mathbf{f}(\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^{M-1}) = 0$, $\mathbf{x}^l$ 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 M 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 object is imaged from multiple view points or when it is imaged by cameras having the same optical

centre, the images are related by homographies. A homography or a collineation is a linear 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

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

where $\varepsilon = \pm 1$. If $\varepsilon = 1$ then the isometry is orientation preserving and is a Euclidean transformation. If $\varepsilon = -1$ then the isometry reverses orientation and involves a reflection. The above can be expressed more compactly as $\mathbf{x}' = \mathbf{H}_{\mathrm{E}}\mathbf{x}$ where

$$\mathbf{H}_{\mathrm{E}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}$$

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 as $\mathbf{x}' = \mathbf{H}_{\mathrm{S}}\mathbf{x}$ where

$$\mathbf{H}_{\mathrm{S}} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}$$

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 as $\mathbf{x}' = \mathbf{H}_{\mathrm{A}}\mathbf{x}$ where

$$\mathbf{H}_{\mathrm{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}$$

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

$$\mathbf{x}' = \mathbf{H}_{\mathrm{P}}\mathbf{x},$$

$$\mathbf{H}_{\mathrm{P}} = \begin{bmatrix} \mathbf{A} & t \\ \mathbf{V}^{\mathrm{T}} & v \end{bmatrix} \quad \text{and } \mathbf{V} \text{ is a vector } [v_1, v_2]^{\mathrm{T}}. \tag{1}$$

The vector $\mathbf{V}$ is the projective component of the homography that makes the transformation nonlinear in inhomogeneous coordinates [7]. The image-to-image homography is projective (a) when the object being imaged is planar, or (b) when the scene is imaged with cameras having the same optical centre.

Fig. 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; however, lengths 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 representation 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. We assume that the object is imaged in its entirety in all views, i.e. all of the object is visible in all views; no part of the object is occluded. 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.

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

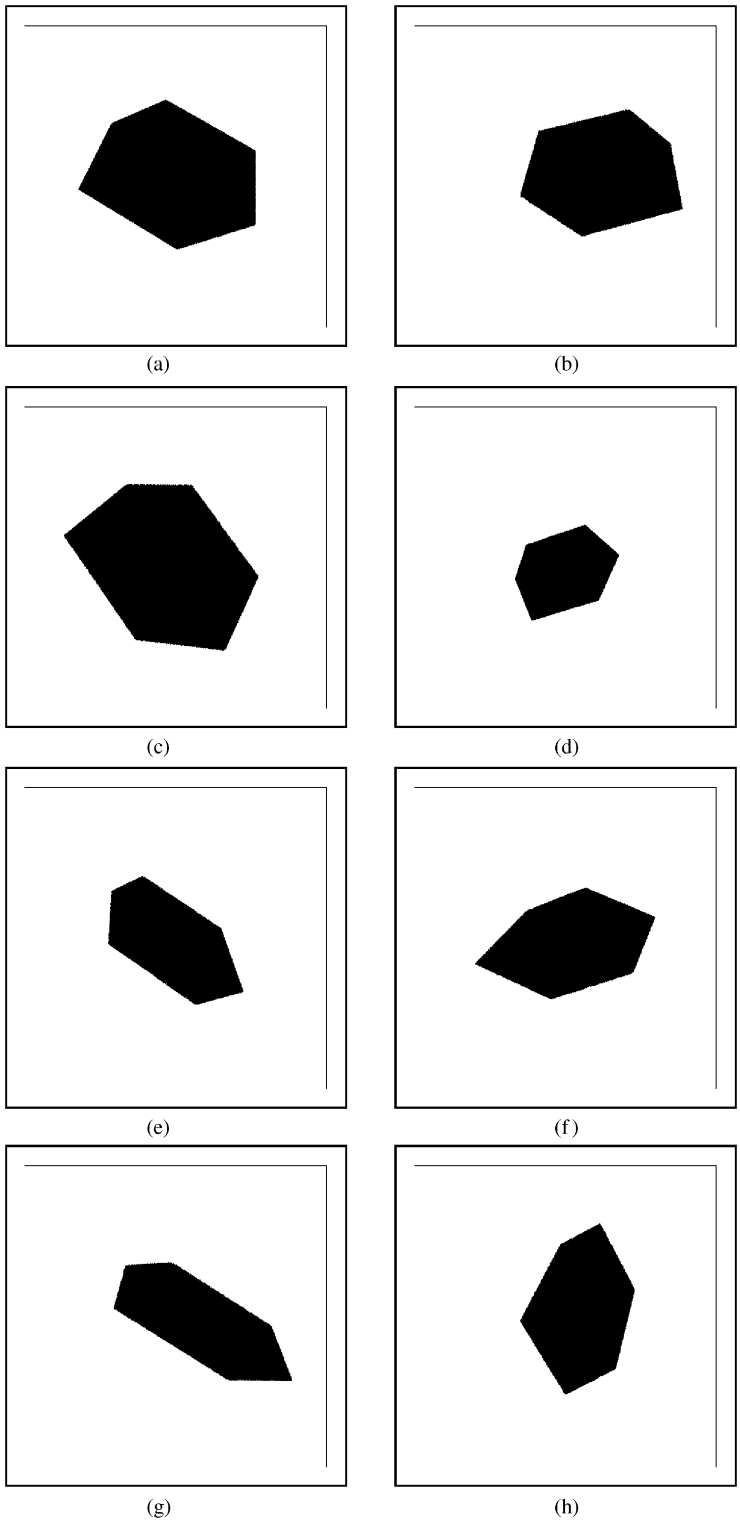

Fig. 1. Several views of a hexagon under different image-to-image homographies.

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

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

where $U^l[k], V^l[k], W^l[k]$ are, respectively, the Fourier transforms of the sequences $u^l[i], v^l[i], w^l[i]$. The sequence $\bar{\mathbf{X}}^l[k]$ 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 obtain

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

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

The matrix $\mathbf{M}_l$ in Eq. (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.

2.4. Planar shape recognition problem

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 recognize 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 cases.

3. Algebraic recognisability constraints under similarity

3.1. Unknown starting point

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 the views of a shape are identical, but its boundary representation starts from a different point in each view. M sequences of boundary points can result 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, for view l,

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

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

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

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 Eq. (5), it can be seen that $\bar{\mathbf{X}}^l[k]$ has a phase difference of $2\pi\lambda_l k/N$ from $\bar{\mathbf{X}}^0[k]$. Let $\theta_0, \theta_1, \ldots, \theta_{N-1}$ be the phases of the Fourier coefficients $U^0[k]$ (U is a component of $\bar{\mathbf{X}}$ from Eq. (2)). We can form an $M \times N$ measurement matrix Θ with row l consisting of the phase angles of the Fourier coefficients U^l of view l.

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

where ϕ_l is $2\pi\lambda_l/N$. 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 ith 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 algebraic 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

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

The crosspower 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 V instead of U.)

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

$$\bar{\mathbf{X}}^l[k] = \bar{\mathbf{X}}^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N} + \delta(0)\mathbf{T}_l, \quad 0 \leqslant k < N.$$

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

$$\bar{\mathbf{X}}^l[k] = \bar{\mathbf{X}}^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N}, \quad 0 < k < N. \tag{8}$$

The form in Eq. (8) is similar to Eq. (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

$$\bar{\mathbf{X}}^l[k] = \mathbf{M}_l \bar{\mathbf{X}}^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N} \tag{9}$$

which in terms of U and V is

$$U^l[k] = sU^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N},$$

$$V^l[k] = sV^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N}.$$

The technique described in the previous subsection depends on the phases of U and V and from the above, it is 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. We can use the complex vector notation of

$$\mathbf{x}^l[i] = \begin{bmatrix} r^l[i] + j0 \\ \theta^l[i] + j0 \end{bmatrix}, \quad 0 \leqslant k < N.$$

A rotation transformation, in the absence of knowledge of correspondence, assumes the form

$$\mathbf{x}^l[i] = \mathbf{x}^0[i + \lambda_l] + [\,0 \quad \phi\,]^{\mathrm{T}}, \quad 0 \leqslant k < N,$$

which in the Fourier domain becomes

$$\bar{\mathbf{X}}^l[k] = \bar{\mathbf{X}}^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N} + [\,0 \quad \phi\,]^{\mathrm{T}}\delta(0), \quad 0 \leqslant k < N.$$

On dropping the spatial frequency corresponding to $k = 0$, we get

$$\bar{\mathbf{X}}^l[k] = \bar{\mathbf{X}}^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N}, \quad 0 < k < N.$$

This transformation form is the same as the ones in Eqs. (5) and (8), techniques for handling which we have discussed above.

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. 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

$$\mathbf{x}^l[i] = \mathbf{A}_l \mathbf{x}^0[i] + \mathbf{b}_l,$$

$$\bar{\mathbf{X}}^l[k] = \mathbf{A}_l \bar{\mathbf{X}}^0[k] + \mathbf{b}_l \delta(0),$$

in the spatial domain and Fourier domain, respectively, where $\mathbf{A}_l$ is a 2×2 matrix

$$\left(\begin{bmatrix} m_1 & m_2 \\ m_4 & m_5 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right)$$

and $\mathbf{b}_l$ is a translation vector

$$\left(\begin{bmatrix} m_3 \\ m_6 \end{bmatrix} \right).$$

We can eliminate 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 spatial 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 $\bar{\mathbf{X}}^l$ are 2×1 matrices of complex numbers.

If pixel-to-pixel correspondences are not known

$$\bar{\mathbf{X}}^l[k] = \mathbf{A}_l \bar{\mathbf{X}}^0[k] \mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N}, \tag{10}$$

where λ_l is the unknown shift in view l.

A measure similar to the crosspower 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

$$\psi(0,l) = (\bar{\mathbf{X}}^0[k])^{*\mathrm{T}} \bar{\mathbf{X}}^l[k] = (\bar{\mathbf{X}}^0[k])^{*\mathrm{T}} \mathbf{A}_l \bar{\mathbf{X}}^0[k] \mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N}. \tag{11}$$

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 $\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}$ is equivalent to $\mathbf{X}^{\mathrm{T}}\mathbf{B}\mathbf{X}$, where $\mathbf{B}$ 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 $\bar{\mathbf{X}} = [(p_1 + \mathrm{j}q_1)\ (p_2 + \mathrm{j}q_2)]^{\mathrm{T}}$,

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

If $\mathbf{A}$ 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. $\psi(\cdot) = c\mathrm{e}^{(\mathrm{j}2\pi\lambda_l/N)}$. 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, Eq. (11) states that the phase of $\psi(0,l)$ differs from the phase of the auto-correlation $\psi(0,0)$ by $2\pi\lambda_l k/N$. The phases of the auto-correlation terms is zero. Hence the phases of the terms in $\psi(0,l)$ is $2\pi\lambda_l k/N$.

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 $\psi(0,l)$.

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

where ϕ_l is $2\pi\lambda_l/N$. 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

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

Experimental results: To numerically validate the above results, two views of a planar object (an image of an aircraft) were generated with a random symmetric affine image-to-image homography. The IDFT of ψ was computed. This is depicted in Fig. 2. The graph shows a distinct and unique peak at the optimal λ (150 in this case) to align and recognise 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 essentially 1. This experiment was repeated for various planar shapes with the same result.

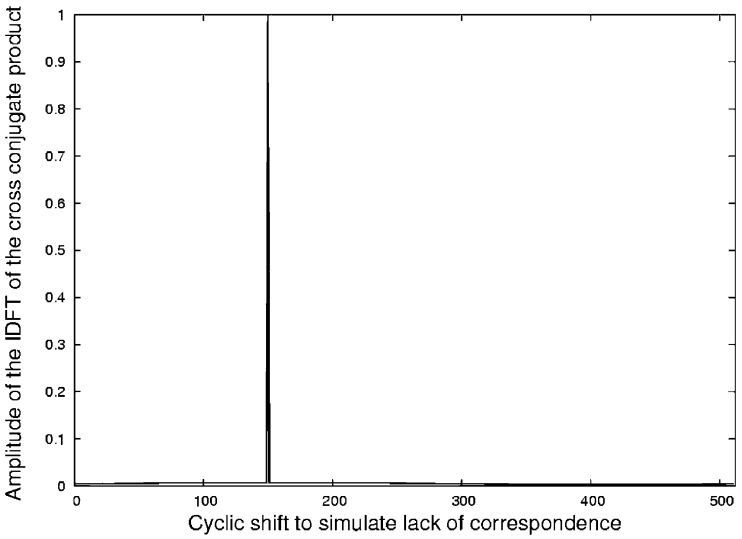

Fig. 2. The IDFT of $\bar{\mathbf{X}}^0[k]^{*\mathrm{T}}\bar{\mathbf{X}}^l[k]$ when A is symmetric and the shift is 150.

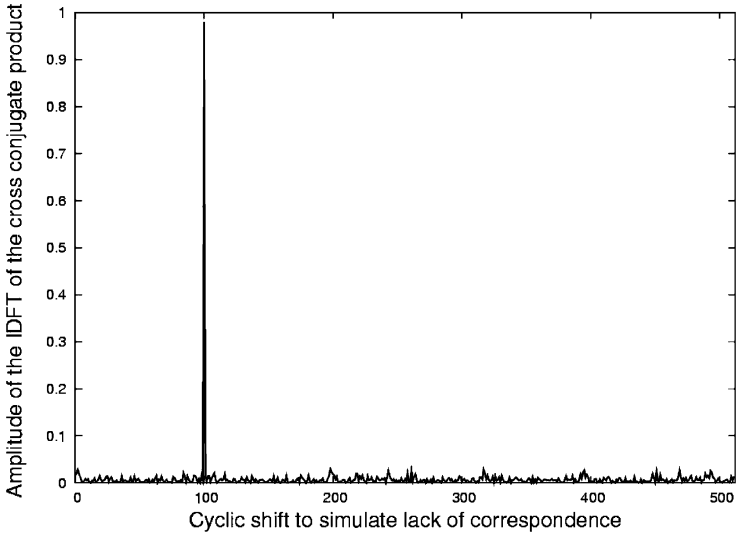

Fig. 3. The IDFT of $\bar{\mathbf{X}}^0[k]^{*\mathrm{T}}\bar{\mathbf{X}}^l[k]$ when A is asymmetric and the shift is 100.

If $\mathbf{A}$ is not symmetric, $\bar{\mathbf{X}}^{*\mathrm{T}}\mathbf{A}\bar{\mathbf{X}}$ 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 Fig. 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 Fig. 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

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

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

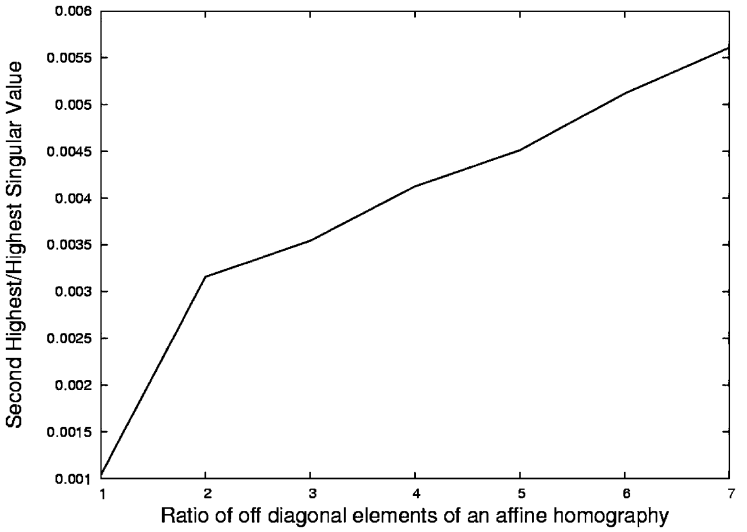

Fig. 4. Plot of the ratio of the second highest singular value of Θ to the highest singular value vs the ratio of the off-diagonal elements of an affine homography (Θ is computed using the technique for symmetric affine homographies).

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

$$= \frac{1}{2}\begin{bmatrix} 0 & a_{12}-a_{21} \\ a_{21}-a_{12} & 0 \end{bmatrix},$$

$$= c\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

where $c = \frac{1}{2}a_{12} - a_{21}$. We can now write Eq. (11) as

$$\begin{aligned}\psi(0,l) &= \bar{\mathbf{X}}^{*\mathrm{T}}\mathbf{A}\bar{\mathbf{X}}\mathrm{e}^{-\mathrm{j}2\pi\lambda_l k/N} \\ &= \bar{\mathbf{X}}^{*\mathrm{T}}\left(\mathbf{A}_s + c\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right)\bar{\mathbf{X}}\mathrm{e}^{-\mathrm{j}2\pi\lambda_l k/N} \\ &= \psi_1 + \psi_2. \end{aligned} \tag{14}$$

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

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

We can see that

$$\begin{aligned}\kappa(l)[k] &= (\bar{\mathbf{X}}^l[k])^{*\mathrm{T}}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\bar{\mathbf{X}}^l[k] \\ &= \left(\mathbf{A}_l\bar{\mathbf{X}}^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N}\right)^{*\mathrm{T}}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\mathbf{A}_l\bar{\mathbf{X}}^0[k]\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N} \\ &= (\bar{\mathbf{X}}^0[k])^{*\mathrm{T}}\mathbf{A}_l^{\mathrm{T}}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\mathbf{A}_l\bar{\mathbf{X}}^0[k]\mathrm{e}^{-\mathrm{j}2\pi\lambda_l k/N}\mathrm{e}^{\mathrm{j}2\pi\lambda_l k/N} \\ &= |\mathbf{A}_l|\kappa(0). \end{aligned} \tag{16}$$

Eq. (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.

Eq. (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.

$$\kappa'(l,p)[k] = (\bar{\mathbf{X}}^l[k])^{*\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{X}}^l[p]. \tag{19}$$

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

$$\begin{aligned}\kappa'(l,p)[k] &= (\bar{\mathbf{X}}^l[k])^{*\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{X}}^l[p] \\ &= |\mathbf{A}_l|\kappa'(0,p)[k]\mathrm{e}^{-\mathrm{j}2\pi\lambda_l(k-p)/N}. \end{aligned} \tag{20}$$

Eq. (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 $\kappa'(l,p)/\kappa'(0,p)$ will be a complex sinusoid $c\mathrm{e}^{-\mathrm{j}2\pi\lambda_l(k-p)/N}$. The value of λ_l can be computed from the inverse Fourier transform of the quotient series.

We can also form an $M \times (N-1)$ 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. Fig. 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 constraints are 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. Ease in using Fourier transform routines was the motivation for sampling the boundaries to have 1024 boundary points.

5.1. Isometry and similarity homographies

The views (a) and (b) of Fig. 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 (Eq. (17)) of 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 discretised. Table 1 shows the performance for both cases.

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 Fig. 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.

Table 1
Singular values of the Θ matrix computed from the κ measures of views of the hexagon in Fig. 1a–d which are related by isometry and similarity homographies

	Isometry		Similarity	
Boundary points	Highest	Next	Highest	Next
Real	256 420	0.00386031	256 310	0.00486031
Discrete	256 423	1.96133	256 398	4.10475

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

	Singular values	
Boundary points	Highest	Next
Real	231 124	0.00817599
Discrete	231 123	2.91271

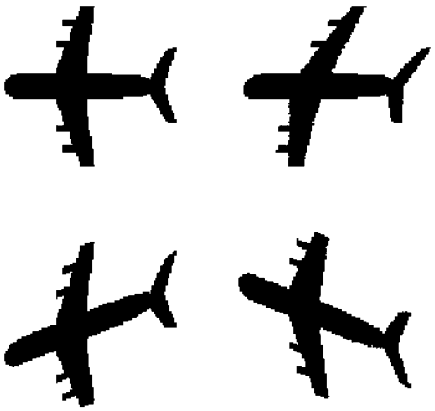

Fig. 5. Four affine-transformed views of an aircraft.

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 Fig. 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 Fig 6 for real and discretised boundary descriptions, respectively.

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. Figs. 7(a) and (b) 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 100 and 300, respectively.

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

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 4
Ratio of the highest singular value to the second highest singular value of the matrix of κ measures for different combinations of views shown in Fig. 6 for discretised 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	—

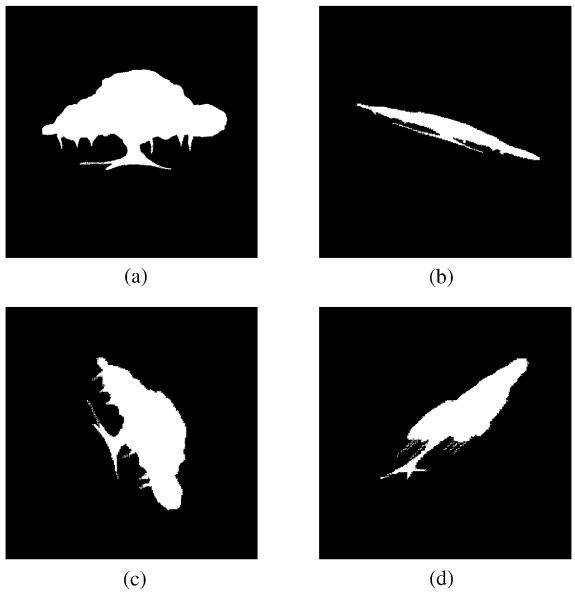

Fig. 6. Four views of IIIT's logo related by random affine image-to-image homographies.

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.

5.3. Projective homographies

We tested the κ measure (Eq. (15)) on views related by general plane-to-plane homographies on both synthetic and

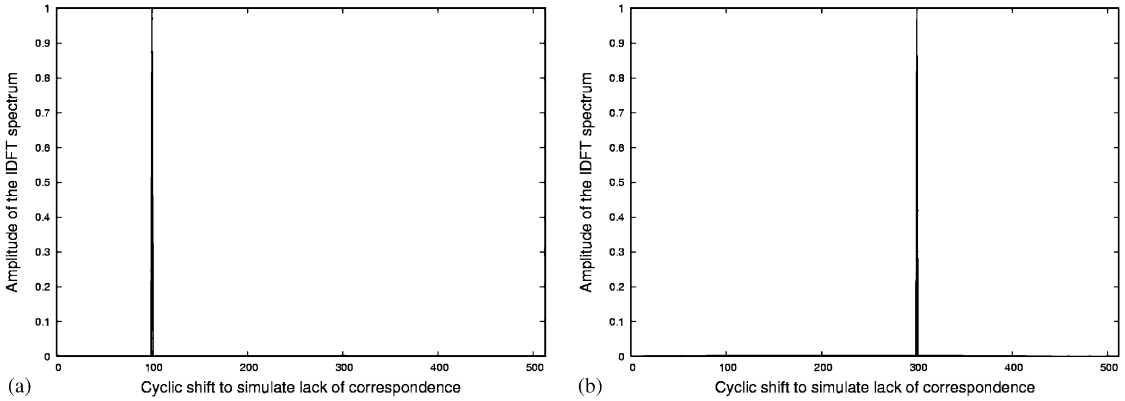

Fig. 7. Graph showing the amplitude of the IDFT of $\kappa'(l,1)/\kappa'(0,1)$ against the shift for an affine homography when the synthetic shift is (a) 100 and (b) 300.

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

	Singular values	
Boundary points	Highest	Next
Real	216 817	110.982
Discrete	216 803	113.986

Table 6
Ratio of the highest singular value to the second highest singular value of the matrix of κ measures for different combinations of views shown in Fig. 8

Views	a	b	c
a	—	431.048	505.847
b	431.048	—	292.71
c	505.847	292.71	—

real images. The experimental results presented below, show that we are able to recognise shapes even under such situations. The image-to-image homography is projective in general, but in practice 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! Approximation of a perspective camera by a weak perspective camera (which is affine) is popular [20]. 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 Fig. 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 discretised.

Next we present the results of experiments on real images.

Experiments on real images: Fig. 8 shows three views of our Institute's logo. 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.

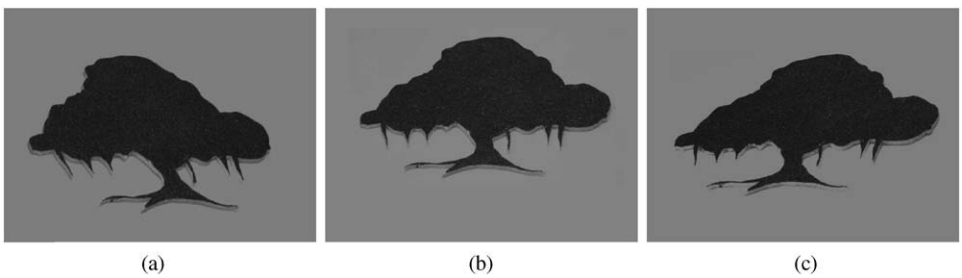

Fig. 8. Three views of the logo of IIIT.

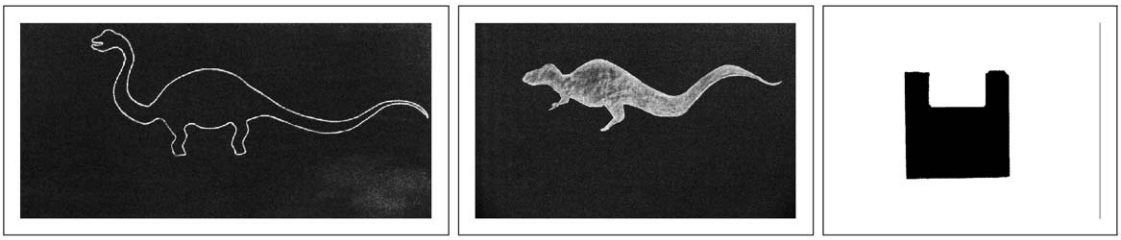

Fig. 9. A view each of drawings of a dinosaur and a lizard and a floppy diskette that has been segmented out.

Table 7
Ratio of the highest singular value to the second highest singular value of the matrix of κ measures for different combinations of four views (a–d) of the shape shown in Fig. 9a

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 8
Ratio of the highest singular value to the second highest singular value of the matrix of κ measures for different combinations of four views (a–d) of the shape shown in Fig. 9b

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 9
Ratio of the highest singular value to the second highest singular value of the matrix of κ measures for different combinations of four views (a–d) shown in Fig. 9c

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
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 Figs. 8, 9a–c)

	Singular values		
Shape	Highest	Next highest	Ratio
IIIT Logo	1.026e+06	2878.12	356.482
Dinosaur	641940	3907.43	164.287
Lizard	786130	7065.8	111.258
Floppy	1.203e+06	1196.99	1005.021

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.

Similar experiments were carried out on a number of shapes. Here we show results of experiments conducted on four views each of three planar shapes—a dinosaur, a lizard, and a floppy (one view of each is shown in Fig. 9a–c). 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 7–9, 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 10, 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 Fig. 9a–c. The highest singular value is greater than the second highest singular value by more than order of 2 in all cases.

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—IIITs logo (Fig. 8b), a dinosaur (Fig. 9a), a lizard(Fig. 9b), and a floppy (Fig. 9c). 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 11.

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

Table 11
Discriminatory power: ratio of highest singular value to the second highest singular value of the matrix of κ measures for different combinations of shapes

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
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 level	Real		Discrete	
	Singular values		Singular values	
	Highest	Next	Highest	Next
0	247 476	0.00186574	213 036	73.0211
0.5%	232 918	63.6448	229 286	124.335
3%	211 296	356.347	228 500	483.168
5%	208 896	839.34	209 417	1233.88
10%	193 925	1424.26	197 214	2069.28
15%	190 745	2324.85	176 999	3251.64
20%	180 199	3887.51	166 523	4931.72

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 boundary to be recognised for an affine homography. The highest two singular values for different maximum noise levels are shown in Table 12. 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 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 13.

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

Table 13
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

Noise level	Singular values		
	Highest	Next	Ratio
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

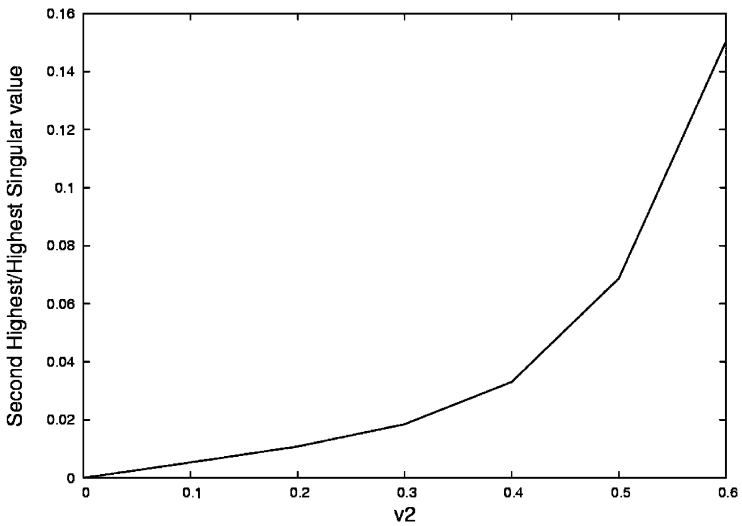

Fig. 10. Graph showing the variation in the performance of the κ measure as the projective component v_2 (Eq. (1)) is increased.

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 (Eq. (1)), keeping v_1 fixed, is shown in Fig. 10.

6. Conclusions

We formulated Fourier domain constraints combining shape properties and multiview relations for planar shape recognition in this paper. These serve as algebraic constraints for recognition of a planar shape in multiple views. Our method does not need correspondence of 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 recognition constraints 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. We are currently extending this philosophy to general projective image-to-image transformations.

7. Summary

Recognition of planar shapes is an important problem in computer vision and pattern recognition and has received widespread attention in literature for many years. The same planar object when imaged from multiple view points and/or under different camera parameters looks different and its recognition is non-trivial. Many planar object recognition efforts have been reported for the simple case of similarity transformations—translation, rotation, and scaling between views. However, the transformation between views of a planar object are more complex in typical problems of interest and hence recognition approaches designed to handle translations, rotations, and/or scaling would prove to be insufficient.

Many algebraic 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 recognise 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 Fourier-domain measurement matrices computed from the points on the shape boundary. It is significant that our techniques do not need point-to-point correspondence which is needed by many recognition algorithms and rarely available in real-life situations. 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.

References

[1] C. Zahn, R. Roskies, Fourier descriptors for planar curves, IEEE Trans. Comput. C-21 (1972) 269–281.
[2] T. Crimmins, A complete set of Fourier descriptors for two dimensional shapes, IEEE Trans. Syst. Man. Cybern. SMC-12 (1982) 848–855.
[3] G. Granlund, Fourier preprocessing for hand printed character recognition, IEEE Trans. Comput. C-21 (1972) 195–201.
[4] T. Pavlidis, Structural Pattern Recognition, Springer, Berlin, 1977.
[5] D.P. Huttenlocher, S. Ullman, Object recognition using alignment, Proceedings of International Conference on Computer Vision, London, UK, 1987, pp. 102–111.
[6] A.K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ, 1989.
[7] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2000.
[8] S. Ullman, R. Basri, Recognition by linear combination of models, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 992–1006.
[9] K. Arbter, W. Snyder, H. Burkhardt, G. Hirzinger, Application of affine-invariant Fourier descriptors to recognition of 3d objects, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1990) 640–647.
[10] S. Kuthirummal, C. V. Jawahar, P. J. Narayanan, Planar shape recognition across multiple views, Proceedings of International Conference on Pattern Recognition 1, Quebec City, Canada, 2002, pp. 456–459.
[11] O. Faugeras, Q. Luong, The Geometry of Multiple Images, MIT Press, Cambridge, MA, 2001.
[12] J. Mundy, A. Zisserman, Geometric Invariances in Computer Vision, MIT Press, Cambridge, MA, 1992.
[13] H.C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, Nature 293 (1981) 133–135.
[14] O. Faugeras, Three Dimensional Computer Vision, MIT Press, Cambridge, MA, 1992.
[15] A. Shashua, Trilinear tensor: the fundamental construct of multiple-view geometry and its applications, International Worskshop on AFPAC, Kiel, Germany, 1997, pp. 190–206.
[16] J. Weng, T.S. Huang, N. Ahuja, Motion and structure from line correspondences: closed-form solution, uniqueness and optimization, IEEE Trans. Pattern Anal. Mach. Intell. 14 (1992) 318–336.
[17] M. Spetsakis, J. Aloimonos, Structure from motion using line correspondences, Int. J. Comput. Vision 4 (3) (1990) 171–183.
[18] R. Hartley, Lines and points in three views: a unified approach, Proceedings of ARPA Image Understanding Workshop, Monterey, Canada, Vol. II, 1994, pp. 1009–1016.
[19] B. Triggs, Matching constraints and the joint image International Conference on Computer Vision, Boston, USA, 1995, pp. 338–343.
[20] L. Shapiro, A. Zisserman, M. Brady, 3D Motion Recovery via affine epipolar geometry, Int. J. Comput. Vision 16 (2) (1995) 147–182.

About the Author—SUJIT KUTHIRUMMAL is currently an MS student at the International Institute of Information Technology, Hyderabad, India. His areas of interest include computer vision, image and video processing, and computer graphics.

About the Author—C.V. JAWAHAR received his Ph.D. from the Indian Institute of Technology, Kharagpur in 1997. He was with the Centre for Artificial Intelligence and Robotics, Bangalore, India, as a scientist till December 2000. At present, he is an Assistant Professor at the International Institute of Information Technology, Hyderabad, India. His areas of interest include computer vision, soft computing, image processing, multimedia systems, and pattern recognition.

About the Author—P.J. NARAYANAN got his Ph.D. in Computer Science from the University of Maryland, College Park in 1992. From 1992 to 1996, he was a research faculty member at the Carnegie Mellon University and worked on the Virtualized Reality project. He was with the Centre for Artificial Intelligence and Robotics (CAIR) from 1996 to 2000. He is currently an Associate Professor at the International Institute of Information Technology. His research interests include computer vision, computer graphics, and virtual reality.