Multiview Constraints for Recognition of Planar Curves in Fourier Domain

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

Abstract

Multiview relations such as the Fundamental matrix and the Trilinear tensor provide scene-independent characterization of a combination of views in the form of algebraic constraints. In this paper, we present a number of multiview constraints for collections of primitives, such as a planar shape boundary. The rich domain of Fourier transforms helps us to combine the properties of the collection with the multiview situation. We derive a number of view-independent algebraic constraints under the assumption of affine image-to-image homography. These constraints provide useful tools to match and recognize planar boundaries across multiple views without the knowledge of the camera parameters or pixel-to-pixel correspondence. We present the results of shape matching in a number of synthetic and real situations in this paper.

1. Introduction

Analysis of multiple views of the same scene is an area of active research in computer vision. The study of the structure of points and lines in two views received much attention in the eighties and early nineties [2, 4, 8]. Studies on the constraints existent in three and more views have followed since then [3, 10, 11, 12]. These multiview studies have concentrated on how geometric primitives like points, lines and planes are related across views. Specifically, the algebraic constraints satisfied by the projections of such primitives in different views have been the focus of intense studies.

An important issue in multi-view analysis is recognition. The problem of recognition in the context of multiple views is as follows: Given the image of an object in one or more views, can we recognise the object in novel views, specifically, when the viewing parameters of the camera are unknown.

An object can be recognised either based on the object-boundary or the textural and structural content inside the boundary. In this paper, we limit the recognition problem to that of recognising planar objects from their boundaries. A notable object recognition approach is due to Ullman and Basri [13] who formulated the recognition problem using linear combination of models, for orthographic views. This recognition differs from the conventional shape recognition approaches. This provides algebraic constraints between different views of the same object in contrast to a set of features invariant to similarity transformations. This algorithm was later generalised by Shashua [11] for perspective views. These results demonstrate that the various views of an object lie in a lower dimensional linear subspace and there can exist some algebraic constraints for recognition of objects in multiple views.

A number of approaches have been proposed for planar shape recognition. Algorithms for planar shape recognition include recognition by alignment [5], polygonal approximation [9], based on geometrically invariant features [6],etc. . Boundaries are also recognised by modelling the boundary in a transform domain like the Fourier one [14]. All these algorithms limited their attention to similarity transformation between views. In most practical situations, the image to image homography is more general than the simple similarity transformation. When a planar object is imaged from multiple viewpoints, the image to image transformation is general projective and the conventional algorithms based on Euclidean and similarity frameworks will not work for them. Klaus Arbter *et al.* [1] formulated techniques for affine invariant recognition. Their emphasis was on choosing a suitable set of affine invariant features and then perform matching in an affine invariant space.

In this paper, we try to analyse the properties of a collection of points, such as a planar object's contour, in multiple views instead of analysing them as independent points. Collections of points such as a boundary have more information than isolated points. The sequencing inherent in such a collection makes a transform domain approach, such as the Fourier one, a good tool to study their properties. The linear image-to-image relationships combined with the properties of the contour in the Fourier domain enable rich constraints that essentially characterize the contour independent of the viewpoint. We come up with a number of view-independent characterizations of the planar shape boundary using measures computed in the Fourier domain. Recognition of a shape in different views is a natural consequence if the description is invariant to the types of view transformations. We present the problem of recognition in multiple views in the form of a rank constraint on a matrix computed from the contours. Some preliminary results were presented in [7].

We formulate the basic problem of view-independent characterization and give our notation in Section 2. Section 3 presents the main results in terms of a number of rank constraints on the measurement matrices computed from the Fourier domain representation of contours. Results of experiments on synthetic and real data are given in Section 4. Some concluding remarks are in Section 5.

2. Problem Formulation

When a planar object is imaged from multiple view points or when a scene is imaged by cameras having the same optical centre, the images are related by a unique homography [4]. A homography or a collineation is a mapping from one plane to another such that the collinearity of a set of points is preserved. In other words, a homography, more precisely a projective 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 $h(x_3)$ are also collinear.

Plane-to-plane homographies can be categorised into isometry, similarity, affine and projective [4]. The later classes subsume the earlier ones, i.e., `isometry` $\subset$ `similarity` $\subset$ `affine` $\subset$ `projective`. Projective homography is mathematically most general. In this paper, we derive the rank constraints for affine homographies, and later show that the constraints are valid in most practical situations of imaging a scene from multiple points, when the image to image homography is projective. Let the image-to-image transformation of points from view 0 to view l be given by a 3×3 matrix $\mathbf{M}_l$.

$$\mathbf{x}^l[i] = \mathbf{M}_l \mathbf{x}^0[i] \tag{1}$$

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 $\mathcal{V}_l$ 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 $\mathcal{V}_l$. This shape is represented by a sequence of vectors of complex numbers as shown below.

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

($w^l[i]$ need not be 1). Let the Fourier domain representation of the sequence $\mathbf{x}^l[i], 0 \le i < N$ be $\mathbf{X}^l[k], 0 \le k < N$ such that

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

where $U^l[k] = U^l_R[k] + jU^l_I[k]$, $V^l[k] = V^l_R[k] + jV^l_I[k]$, $W^l[k] = W^l_R[k] + jW^l_I[k]$ are respectively the Fourier transforms of the individual sequences $(u^l[i] + j0)$, $(v^l[i] + j0)$, $(w^l[i] + j0)$. The subscripts R and I denote the real and imaginary components of the corresponding complex number. Note that the sequences $\mathbf{X}^l[k]$ are periodic and conjugate symmetric.

Theorem 1: The Fourier transform and the collineation commute with the above representation. That is, if points are transformed between views 0 and l using Equation 1, the same homography will transform corresponding frequency terms in the Fourier domain also. In other words,

$$\mathbf{X}^l[k] = \mathbf{M}_l \mathbf{X}^0[k]\ , 0 \le k < N. \tag{2}$$

Proof: Let $\mathbf{M}_l = m_{lij}, 1 \le i, j \le 3$. Expanding Equation 1 for the u term,

$$u^l[i] = m_{l11} u^0[i] + m_{l12} v^0[i] + m_{l13} w^0[i]$$

Taking the Fourier transform of the above equation and using the linearity property of Fourier transforms, we get

$$U^l[k] = m_{l11} U^0[k] + m_{l12} V^0[k] + m_{l13} W^0[k]$$

Similarly for $V^l[k]$ and $W^l[k]$. It is now easy to see that

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

giving us the desired result. □

Given a set of L views, the recognition problem can be formulated as the identification of a view-independent function $f(\cdot)$ such that $f(\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^L) = 0$. This recognition constraint can be linear or nonlinear in image coordinates. The algebraic relation given by $f(\cdot)$ can then be used to settle the question whether the L observed views were of the same object.

3. Rank Constraints for Recognition

If the image to image homography is affine, the transformation matrix has $m_{l31} = m_{l32} = 0$ and $m_{l33} = 1$. The transformation can be expressed in terms of inhomogeneous coordinates as

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

where $\mathbf{x}^l[i]$ is the inhomogeneous representation of the ith point on the contour in view l, $\mathbf{A}_l$ is the upper 2×2 minor of $\mathbf{M}_l$ and $\mathbf{b}_l$ is the upper two elements of the last column of $\mathbf{M}_l$.

The above expression is valid for the scenarios when correspondence between points across views is known. However in practice, correspondence is rarely available. In case correspondence information is not available, Equation 3 assumes the form

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

where shifting $\mathbf{x}^0$ by λ_l would align the corresponding points of $\mathbf{x}^0$ and $\mathbf{x}^l$. The frequency domain representation can be given by

$$\mathbf{X}^l[k] = \mathbf{A}_l \mathbf{X}^0[k] \exp(\frac{j2\pi\lambda_l k}{N}), \quad 0 < k < N \qquad (4)$$

if the $\mathbf{b}$ term is eliminated by omitting the $k = 0$ term in the Fourier domain.

3.1. Affine Invariant

The study of invariants has been pursued actively for many years. Invariants provide us with the ability to come up with representations of the features in a scene that do not depend on the view, and can prove to be extremely handy for purposes of recognising objects from multiple views. In this Subsection we explore the possibility of deriving an affine invariant for a contour.

Let us define a measure called the *cross-conjugate product (CCP)* on the Fourier representations of two views as

$$\begin{aligned}\psi(0,l)[k] &= (\bar{\mathbf{X}}^0[k])^{*T} \bar{\mathbf{X}}^l[k], \quad 0 < k < N \\ &= (\bar{\mathbf{X}}^0[k])^{*T} \mathbf{A}_l \bar{\mathbf{X}}^0[k] \exp(\frac{j2\pi\lambda_l k}{N}). \quad (5)\end{aligned}$$

The matrix $\mathbf{A}_l$ can be expressed as a sum of a symmetric matrix and a skew symmetric matrix as $\mathbf{A}_l = \mathbf{A}_l^s + \mathbf{A}_l^{sk}$ where $\mathbf{A}_l^s = \frac{1}{2}(\mathbf{A}_l + \mathbf{A}_l^T)$ and $\mathbf{A}_l^{sk} = \frac{1}{2}(\mathbf{A}_l - \mathbf{A}_l^T)$. The skew symmetric matrix reduces to

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

where $c = m_{l12} - m_{l21}$ is the difference of the off-diagonal elements of $\mathbf{A}_l$. We now have

$$\psi(0,l)[k] = \bar{\mathbf{X}}^0[k]^{*T} \left(\mathbf{A}_l^s + \mathbf{A}_l^{sk}\right) \bar{\mathbf{X}}^0[k] \exp(\frac{j2\pi\lambda_l k}{N})$$

The first term of the above equation is purely real and the second term is purely imaginary. We observe that the effect of the transformation matrix A_l on the second term is restricted to a scaling by a factor c. We can define a new measure κ, ignoring scale, for the sequence $\bar{\mathbf{X}}^l$ in view l as

$$\kappa(l)[k] = \bar{\mathbf{X}}^l[k]^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{X}}^l[k]. \qquad (6)$$

It can be shown [7] that

$$\kappa(l)[k] = |\mathbf{A}_l| \, \kappa(0)[k], \quad 0 < k < N \qquad (7)$$

Equation 7 gives a necessary condition for the sequences $\bar{\mathbf{X}}^l$ and $\bar{\mathbf{X}}^0$ to be two different views of the same planar shape, or in other words, the values of the measure $\kappa(\cdot)$ in the two views 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 $\kappa(\cdot)$ measures for M different views.

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

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

$$\text{rank}(\Theta) = 1. \qquad (8)$$

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

Since, the κ measures in the various views are only scaled versions of each other, if we normalize the κ measure terms in each view with respect to a fixed one then

$$\begin{aligned}\gamma(l)[k] &= \kappa(l)[k] \ / \ \kappa(l)[p]\,, \quad \text{p is fixed} \\ &= (|\mathbf{A}_l|\kappa(0)[k]) \ / \ (|\mathbf{A}_l|\kappa(0)[p]) \\ &= \kappa(0)[k] \ / \ \kappa(0)[p] \\ \gamma(l)[k] &= \gamma(0)[k] \qquad (9)\end{aligned}$$

These terms of the normalized κ measure – the γ measure are independent of the view. Hence, γ is an affine view invariant of a contour, whose computation does not need correspondence information across views.

3.2. Constraints based on Phases

If $\mathbf{A}_l$ is a symmetric matrix (in Equation 5), it can be shown that the auto-correlation $\psi(0,0)$ is real. This implies that that the phase of $\psi(0,l)$ would be $\frac{2\pi\lambda_l k}{N}$.

If we have M views, then we can form a $M \times (N-1)$ matrix Θ' with the phase angles of $\psi(0,l)[k], 0 < k < N$ forming the row l. It is clear 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

$$rank(\Theta') = 1 \qquad (10)$$

This rank-one constraint implies that the phases of CCP in different views are linearly related. Thus, the phases form a signature of the shape that is invariant to affine transformations. Unfortunately this relationship is valid only if the affine transformation ($\mathbf{A}_l$) is symmetric. Also, the CCP is computed between each view and a fixed reference view; it thus depends on two views.

For affine transformations when $\mathbf{A}_l$ can be arbitrary, we derive a rank-two constraint as follows. $\mathbf{A}_l$ now contains a skew-symmetric component in addition to a symmetric one. We define a new measure $\kappa'(\cdot)$, which correlates each Vector Fourier coefficient with a fixed one within each view.

$$\begin{aligned}
\kappa'_p(l)[k] &= (\mathbf{X}^l[k])^{*\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{X}^l[p], \quad 0 < k < N \\
&= (A\mathbf{X}^0[k]\, e^{j\frac{2\pi\lambda_l k}{N}})^{*\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} A\mathbf{X}^0[k] e^{j\frac{2\pi\lambda_l p}{N}} \\
&= (\mathbf{X}^0[k])^{*\mathrm{T}} A^{\mathrm{T}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} A\mathbf{X}^0[k] e^{-j\frac{2\pi\lambda_l (k-p)}{N}} \\
&= |A|\, \kappa'_p(0)[k]\, e^{-j2\pi\lambda_l(k-p)/N} \qquad (11)
\end{aligned}$$

for any fixed $p \neq 0$. Equation 11 states that the phases of $\kappa'_p(l)$ and $\kappa'_p(0)$ differ by an amount proportional to the shift λ_l and the differential frequency $k - p$. Therefore, the ratio $\frac{\kappa'_p(l)}{\kappa'_p(0)}$ 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. Thus, the phases of $\kappa'_p(l)$ can be used as a signature for the contour.

We can also form a $M \times (N-1)$ matrix Θ'', similar to the one above, that stacks the phases of $\kappa'_1(l)$ (taking $p = 1$). It will have the form $\Theta'' =$

$$\begin{bmatrix}
\theta_1 & \theta_2 & \theta_3 & \dots & \theta_{N-1} \\
\theta_1 & \theta_2 + \phi_1 & \theta_3 + 2\phi_1 & \dots & \theta_{N-1} + (N-2)\phi_1 \\
\dots & \dots & \dots & \dots & \dots \\
\theta_1 & \theta_2 + \phi_{M-1} & \theta_3 + 2\phi_{M-1} & \dots & \theta_{N-1} + (N-2)\phi_{M-1}
\end{bmatrix} \qquad (12)$$

where θ_i are the phases of $\kappa'_1(0)$ 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

$$\mathrm{rank}(\Theta'') = 2. \qquad (13)$$

We see that κ' can be computed from a single view. Thus, the phases of κ' values provide a truly view-independent description of the boundary.

Experiments were conducted to affirm the validity of the above constraint. Figure 1 shows four affine transform related views of a dinosaur. When the Θ'' was constructed from the κ' measures of the four views (a), (b), (c) and (d), with random shifts applied to the boundary representations in each view, the rank of Θ'' was found to be essentially 2, the three greatest singular values being 33952.7, 58.8366, and 0.00242446.

3.3. Constraints based on Magnitudes

Unless properly taken care, the phase based algebraic constraints can have problems with the phase wrap around. We now present a rank-three constraint based on magnitudes of the vector Fourier coefficients. We start with Equation 4. This equation can be rewritten as

$$U^l[k] = (m_{l11}U^0[k] + m_{l12}V^0[k]) \exp(\frac{j2\pi\lambda_l k}{N})$$
$$V^l[k] = (m_{l21}U^0[k] + m_{l22}V^0[k]) \exp(\frac{j2\pi\lambda_l k}{N})$$

Writing in terms of the real and imaginary components of the complex numbers

$$\begin{aligned}
U^l_R[k] + jU^l_I[k] &= ((m_{l11}U^0_R[k] + m_{l12}V^0_R[k]) + \\
&\quad j(m_{l11}U^0_I[k] + m_{l12}V^0_I[k]))e^{\frac{j2\pi\lambda_l k}{N}}
\end{aligned}$$

Taking the square of the magnitudes of both sides, we get

$$\begin{aligned}
|U^l[k]|^2 &= (U^l_R[k])^2 + (U^l_I[k])^2 \\
&= (m_{l11}U^0_R[k] + m_{l12}V^0_R[k])^2 + \\
&\quad (m_{l11}U^0_I[k] + m_{l12}V^0_I[k]))^2 \\
&= m^2_{l11}(U^0_R[k])^2 + m^2_{l12}(V^0_R[k])^2 + \\
&\quad 2(m_{l11}U^0_R[k])(m_{l12}V^0_R[k]) + \\
&\quad m^2_{l11}(U^0_I[k])^2 + m^2_{l12}(V^0_I[k])^2 + \\
&\quad 2(m_{l11}U^0_I[k])(m_{l12}V^0_I[k]) \\
&= m^2_{l11}[(U^0_R[k])^2 + (U^0_I[k])^2] + \\
&\quad m^2_{l12}[(V^0_R[k])^2 + (V^0_I[k])^2] + \\
&\quad 2m_{l11}m_{l12}[U^0_R[k]V^0_R[k] + U^0_I[k]V^0_I[k]] \quad (14)
\end{aligned}$$

Similarly,

$$\begin{aligned}
|V^l[k]|^2 &= (V^l_R[k])^2 + (V^l_I[k])^2 \\
&= m^2_{l21}[(U^0_R[k])^2 + (U^0_I[k])^2] + \\
&\quad m^2_{l22}[(V^0_R[k])^2 + (V^0_I[k])^2] + \\
&\quad 2m_{l21}m_{l22}[U^0_R[k]V^0_R[k] + U^0_I[k]V^0_I[k]] \quad (15)
\end{aligned}$$

Its evident from equations 14 and 15 that the magnitude of the components of the Fourier domain representation in any view can be expressed in terms of the components in a reference view.

This result can also be expressed in the following manner. Given M views, we can construct a $(2M+1) \times (N-1)$ matrix as follows. The first row consists of the sum of products $(U^0_R[k]V^0_R[k] + U^0_I[k]V^0_I[k])$, 0 being the reference view. Every view contributes two rows to this matrix (except the reference view, which contributes 3 rows) the magnitudes of U in one row and the magnitudes of V in the

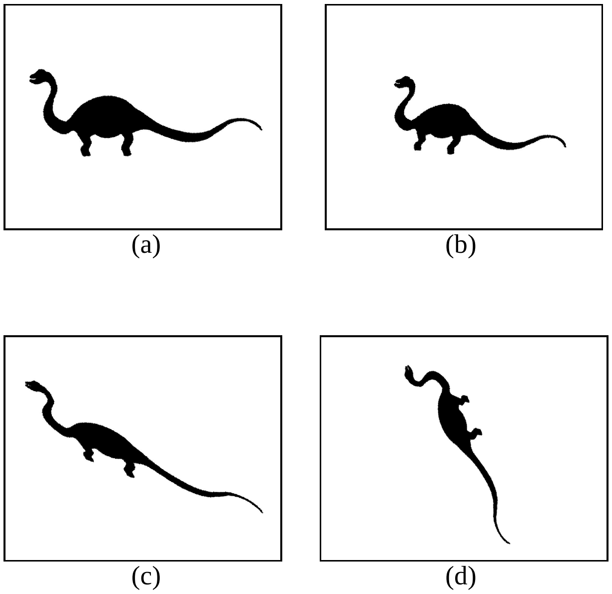

Figure 1: Four affine transformed views of a dinosaur

other. Let $\Theta''' =$

$$\begin{bmatrix} (U_R^0[1]V_R^0[1] + U_I^0[1]V_I^0[1])) & \cdots & (U_R^0[G]V_R^0[G] + U_I^0[G]V_I^0[G]) \\ ((U_R^0[1])^2 + (U_I^0[1])^2) & \cdots & ((U_R^0[G])^2 + (U_I^0[G])^2) \\ ((V_R^0[1])^2 + (V_I^0[1])^2) & \cdots & ((V_R^0[G])^2 + (V_I^0[G])^2) \\ ((U_R^1[1])^2 + (U_I^1[1])^2) & \cdots & ((U_R^1[G])^2 + (U_I^1[G])^2) \\ ((V_R^1[1])^2 + (V_I^1[1])^2) & \cdots & ((V_R^1[G])^2 + (V_I^1[G])^2) \\ ((U_R^2[1])^2 + (U_I^2[1])^2) & \cdots & ((U_R^2[G])^2 + (U_I^2[G])^2) \\ ((V_R^2[1])^2 + (V_I^2[1])^2) & \cdots & ((V_R^2[G])^2 + (V_I^2[G])^2) \\ \cdots & \cdots & \cdots \\ ((U_R^M[1])^2 + (U_I^M[1])^2) & \cdots & ((U_R^M[G])^2 + (U_I^M[G])^2) \\ ((V_R^M[1])^2 + (V_I^M[1])^2) & \cdots & ((V_R^M[G])^2 + (V_I^M[G])^2) \end{bmatrix} \tag{16}$$

(using G for $(N-1)$)

From equations 14 and 15, one can conclude that the rank of Θ'' is 3, irrespective of the number of views. Therefore, the constraint,

$$rank(\Theta''') = 3 \tag{17}$$

is a necessary condition for recognition in multiple views related by affine image-to-image homographies. This observation is consistent with the notion that the various views of a shape lie in a lower dimensional linear subspace. We can also say that the squares of the magnitudes of the Fourier Domain representation of a contour can be used as a signature of the boundary. These are, naturally, view independent as they can be computed from a single view.

4. Results and Discussions

We conducted a number of experiments to affirm the validity of the formulations in the previous section. Extensive experimentations were carried out on synthetic images, natural images with simulated transformations and real natural images. In the rest of this section, we demonstrate the performance of the proposed schemes with quantitative results. For simulation of views, transformations were applied on a reference view and then the boundary representations were shifted by random amounts in each view to simulate lack of correspondence. In experiments on real images, objects of interest were segmented out and their boundaries were sampled to 1024 boundary points. The ranks of the matrices Θ, Θ'' and Θ''' were determined using the Singular Value Decomposition algorithm, wherein the number of non-zero singular values gives the rank of the matrix. When the boundary representation is in the form of integer coordinates, discretization introduces quantization noise that make the rank constraint an approximation of the true one derived, but nonetheless enforceable. To verify whether a matrix has an approximate rank r, we give the ratio of rth to $(r+1)$th singular values (arranged in descending order). This ratio is high if the matrix has an approximate rank of r. Also all the following singular values are very small in magnitude.

In the first example, we considered four views of a dinosaur as in Figure 1(a), (b), (c) and (d). These views are related by affine transformations. One may observe that the Euclidean measures are no longer preserved under these transformation. However, the rank constraints allow us to recognise a dinosaur image given another view. Ranks of Θ and Θ''' were computed. Ratio of the rth and $(r+1)$th singular values are used to verify the ranks. All constraints provided the ratio to be much more than 100 in all cases. The ratios of singular values for each pairs of images for the invariant and magnitude constraints are arranged in Table 1. In all cases, the rth singular value (r = 1 for Θ and r = 3 for Θ''') was found to be greater than the $(r+1)$th one by a factor of 100 or 1000.

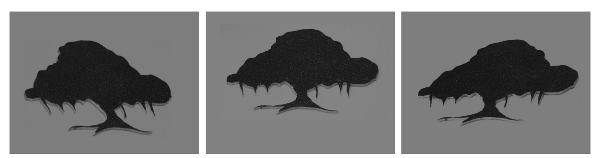

Figure 2: Three views of the logo of our institute

Now, we demonstrate the performance when a zero mean random noise is added to the position of the synthetically transformed shape for an affine homography. The two singular values of interest of matrix Θ for different noise levels for real(without quantization) and discrete (integers) boundary representations are shown in Table 2. This ratio does deteriorate with noise, however, there was still more than an order of magnitude separation between them even with a noise of 20% in the positions of the boundary points.

The recognition is clearly very good in all cases with the degradation in performance along expected lines. We have achieved recognition between two planar shapes under the

	Dinosaur 1	Dinosaur 2	Dinosaur 3	Dinosaur 4
Dinosaur 1	—-	43176.5, 504.423	23988.5, 322.283	35453.9, 439.72
Dinosaur 2	43176.5, 504.423	—-	25733.7, 312.512	35352.6, 322.338
Dinosaur 3	23988.5, 322.283	25733.7, 312.512	—-	17548, 137.258
Dinosaur 4	35453.9, 439.72	35352.6, 322.338	17548, 137.258	—-

Table 1: Results on Dinosaur Images. Ratios of the relevant consecutive singular values for Θ and for Θ''' are shown

	Real		Discrete	
Noise	Singular Values		Singular Values	
Level	Highest	Next	Highest	Next
0	247476	0.0019	213036	73.02
0.5%	232918	63.65	229286	124.34
3%	211296	356.35	228500	483.17
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 2: Impact of noise on singular values

Views	a	b	c
a	-	431.0	505.8
b	431.0	-	292.7
c	505.8	292.7	-

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 2.

assumption that the homography between them has a specific form, without knowing the correspondence between points.

Though the theory was primarily developed for affine homographies, the rank constraints are practically valid for images under projective transformation. The logo of the International Institute of Information Technology was imaged from various viewing positions. These images are known to be related by projective homographies. Three views are shown in Figure 2. Ratios of the two highest singular values of the Θ matrix for various combinations of those views are given in Table 3. All pairs are clearly recognisable and the ratios are more than 250 in all cases.

5. Conclusions and Future Work

In this paper, we presented a number of view-independent constraints for matching and recognition of planar boundaries under affine homographies. The constraints were presented as the rank constraints on a matrix that can be computed from the images alone. We presented results to demonstrate the effectiveness of our scheme to a number of planar shape recognition problems. Extending such constraints to the general projective image-to-image homography is the next step in this direction.

References

[1] K. Arbter, W. Snyder, H. Burkhardt, and G. Hirzinger. Application of Affine-Invariant Fourier Descriptors to Recognition of 3d Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12, 1990.

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

[3] R. Hartley. Lines and points in three views: An integrated approach. *Proc. ARPA Image Understanding Workshop*, 1994.

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

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

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

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

[8] H. C. Longuet-Higgins. A Computer Algorithm for Reconstructing a Scene from two Projections. *Nature*, 293:133–135, 1981.

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

[10] A. Shashua. Algebraic Functions for Recognition. *IEEE Tran. Pattern Anal. Machine Intelligence*, 16:778–790, 1995.

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

[12] B. Triggs. Matching Constraints and the Joint Image. *International Conference on Computer Vision*, pages 338–343, 1995.

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

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