Planar Shape Recognition across Multiple Views

Sujit Kuthirummal, C. V. Jawahar, 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

Multiview studies in Computer Vision have concentrated on the constraints satisfied by individual primitives such as points and lines. Not much attention has been paid to the properties of a collection of primitives in multiple views, which could be studied in the spatial domain or in an appropriate transform domain. We derive an algebraic constraint for planar shape recognition across multiple views based on the rank of a matrix of Fourier domain descriptor coefficients of the shape in different views. We also show how correspondence between points on the boundary can be computed for matching shapes using the phase of a measure for recognition.

1 Introduction

Multiview analysis of scenes is an active area in Computer Vision today. The structure of points and lines as seen in two views attracted the attention of computer vision researchers in the eighties and early nineties [5, 1, 3]. Similar studies on the underlying constraints in three views followed [6, 2]. The structure of greater than four views has also been studied [7, 8]. Two excellent textbooks have recently appeared focusing on multiview geometry for Computer Vision[1, 3]. The mathematical structure underlying multiple views has been studied with respect to projective, affine, and Euclidean frameworks of the world with amazing results.

Multiview studies have focussed on how geometric primitives such as points, lines and planes are related across views. Specifically, the algebraic constraints satisfied by the projection of such primitives in different views have been the focus of intense studies. The multilinear relationships that were discovered have been found to be useful for a number of tasks, such as view transfer, geometric reconstruction and self calibration. The richness of the information present among the geometric primitives in a collection of them has not attracted a lot of attention. Such properties are difficult to capture in the spatial domain but can be extracted with relative ease in a transform domain. The analysis of boundary shapes in multiple views using Fourier domain descriptors can provide structure not explicit in the geometric space and provide interesting handles for solving problems like object recognition and view transfer.

The properties of collections of primitives in multiple views are studied in this paper. Specifically, we look at the situation of viewing a planar shape from different viewpoints. Recognizing objects from diverse viewpoints is essential to interpreting the structure and meaning of a scene. We use a Fourier domain representation for the boundary of the object and derive recognition constraints the projections of the object must satisfy in multiple views. These constraints are in the form of the rank of the matrix of the descriptor coefficient values.

We present the basic problem formulation in the next section. Numerical results to validate the theoretical claims are presented in Section 3, along with some discussions on the underlying issues. Section 4 presents a few concluding remarks.

2 Problem Formulation

We are interested in exploiting the relationships between points on the shape boundary in the domain of a Fourier descriptor. Affine homographies have been studied in the Fourier domain [4], in which the boundary points were represented as complex numbers. We need a richer representation to linearize the affine homography relation and so use a vector of complex numbers as our descriptor for points on the boundary of a shape.

Let $P[i] = (u[i], v[i], w[i])$ be the homogeneous coordinates of points on the closed boundary of a planar shape. The shape is represented by a sequence of vectors of com-

plex numbers as shown below.

$$\mathbf{x}[i] = \begin{bmatrix} u[i] + j\,0 \\ v[i] + j\,0 \\ w[i] + j\,0 \end{bmatrix}$$

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

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

Taking the Fourier transform on both sides we get,

$$\bar{\mathbf{X}}^l[k] = \mathbf{M}\bar{\mathbf{X}}^0[k] \tag{2}$$

where $\bar{\mathbf{X}}^0$ and $\bar{\mathbf{X}}^l$ are the Fourier transforms of $\mathbf{x}^0$ and $\mathbf{x}^l$, respectively. The sequences $\bar{\mathbf{X}}^l[k]$ are periodic and conjugate symmetric.

2.1 Affine Homography

We first look at the case when the transformation between two views is affine. The third row of the matrix M in Equation 1, has the special form $[0\ 0\ 1]$ for affine transformations. We can write Equation 1 in this case as

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

where $\mathbf{A}$ is a 2×2 matrix, $\mathbf{b}$ is a translation vector and l is the view index, with $l = 0$ being considered to be the reference view. We can discard the effect of vector $\mathbf{b}$ by discarding the DC component - the Fourier coefficient corresponding to $k = 0$ (or by shifting the origin to the centroid of the shape). The affine transformation can now be written as

$$\begin{aligned} \mathbf{x}^l[i] &= \mathbf{A}\mathbf{x}^0[i] \text{ in the spatial domain, and} \\ \bar{\mathbf{X}}^l[k] &= \mathbf{A}\bar{\mathbf{X}}^0[k] \text{ in the Fourier domain} \end{aligned} \tag{3}$$

without any loss in generality. In general, correspondence between images, i.e., information as to which image points in different views are projections of the same 3D point, is not available. This implies that when the boundary is seen in two views, the description may not start from the same point. In other words, there is an unknown shift between the sequences of boundary points in different views. This shift in the spatial domain translates into a rotation in the Fourier domain. Equation 3 can now be written as

$$\bar{\mathbf{X}}^l[k] = \mathbf{A}\bar{\mathbf{X}}^0[k]\, e^{\omega_k \lambda_l} \tag{4}$$

where, λ_l is the unknown shift in view l, N is the number of points on the boundary of the shape, and $\omega_k = -j2\pi k/N$.

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

$$\begin{aligned} \psi(0,l) &= (\bar{\mathbf{X}}^0[k])^{*T}\bar{\mathbf{X}}^l[k] \\ &= (\bar{\mathbf{X}}^0[k])^{*T}\mathbf{A}\bar{\mathbf{X}}^0[k]\, e^{\omega_k \lambda_l}. \end{aligned} \tag{5}$$

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

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

where $c = a_{12} - a_{21}$ is the difference of the off-diagonal elements of $\mathbf{A}$. We now have

$$\begin{aligned} \psi(0,l) &= \gamma_1 + \gamma_2 \\ &= \bar{\mathbf{X}}^0[k]^{*T}\,(\mathbf{A}_s + \mathbf{A}_{sk})\,\bar{\mathbf{X}}^0[k]\, e^{\omega_k \lambda_l}. \end{aligned} \tag{6}$$

The term $\bar{\mathbf{X}}^0[k]^{*T}\mathbf{A}_s\bar{\mathbf{X}}^0[k]$ of the above equation is purely real and the term $\bar{\mathbf{X}}^0[k]^{*T} c \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{X}}^0[k]$ is purely imaginary. The phases of γ_1 and γ_2 depend only on the shift λ_l. Thus, λ_l can be recovered from the inverse Fourier transform of γ_1 or γ_2, if known. However, we can only compute $\psi(0,l)$, a combination of γ_1 and γ_2, which is not directly useful to recover the shift.

We observe that the effect of the transformation matrix A on γ_2 is restricted to a scaling by a factor c. We can define a new measure κ, ignoring scale, for the sequence $\bar{\mathbf{X}}^l$ as

$$\kappa(l) = \bar{\mathbf{X}}^l[k]^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{X}}^l[k]. \tag{7}$$

It can be shown that

$$\begin{aligned} \kappa(l) &= (\bar{\mathbf{X}}^l[k])^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{X}}^l[k] \\ &= (\mathbf{A}\bar{\mathbf{X}}^0[k]\, e^{\omega_k \lambda_l})^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{A}\bar{\mathbf{X}}^0[k] e^{\omega_k \lambda_l} \\ &= (\bar{\mathbf{X}}^0[k])^{*T}\, \mathbf{A}^T \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{A}\, \bar{\mathbf{X}}^0[k] \\ &= |\mathbf{A}|\, \kappa(0) \end{aligned} \tag{8}$$

Equation 8 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. We can express it differently in multiple views. 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

$$\text{rank}(\Theta) = 1. \tag{9}$$

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

Can we also estimate the shift λ_l that would align corresponding points in two views? The answer is yes, using a measure κ_2 for a fixed p given by

$$\kappa_2(l,p) = (X^l[k])^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} X^l[p] \qquad (10)$$

κ_2 correlates each Fourier descriptor coefficient with a fixed one within each view. Following a chain of reasoning similar to the one above, we can show that

$$\kappa_2(l,p) \quad = \quad |A|\ \kappa_2(0,p)\ e^{j2\pi\lambda_l(k-p)/N}. \qquad (11)$$

Equation 11 states that the phases of $\kappa_2(l,p)$ and $\kappa_2(0,p)$ differ by a value proportional to the shift λ_l and the differential frequency $k-p$. Therefore, the ratio $\frac{\kappa_2(l,p)}{\kappa_2(0,p)}$ will be a complex sinusoid. The value of λ_l can be computed from the inverse Fourier transform of the quotient series. Thus, we can compute the correspondence between points on the shape boundary across views by determining λ_l as above, starting with no prior knowledge.

The measure κ_2 can also be adopted for the purpose of recognition. However, this approach would not be discussed in this paper for want of space.

The general projective homography relating two different views of the same planar shape can be reasonably approximated by an affine homography. This approximation seems to be a practical one as most real life configurations of imaging a scene from multiple view points, possess structure that are very close that of affine homographies. This assumption is also validated by the results for general homographies, which are presented in the next section.

3 Results and Discussions

We present the results from a number of experiments conducted to affirm the validity of the formulations in the previous section. For the first experiment, we use the planar boundary of an aircraft in a reference view for the study. Other views were generated using affine homographies to map points in the reference view into the new views. The shape boundaries in the views were sampled so that each shape was represented by 1024 points.

Measure for Recognition(κ): Four affine homography related views of an aircraft are shown in figure 1. The Θ matrix for these four views was formed using the measures of κ for each view as described earlier. The rank of this matrix Θ was found to be 1 using SVD (the number of non-zero singular values gives the rank of the matrix) , as the largest four singular values were 51138.4, 0.0056, 0.0028, and 0.0026.

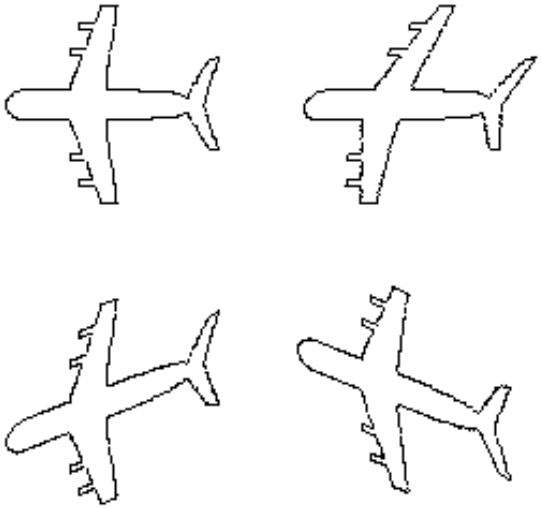

Figure 1. Four affine-transformed views of an aircraft.

Measure for Determining Point Correspondence(κ_{mod}): We test the effectiveness of our technique for estimating correspondences through the shift λ_l. Figures 2 and 3 show the inverse Fourier spectrum of the ratio $\kappa_{mod}(l,1)/\kappa_{mod}(0,1)$, when the shifts aligning corresponding points in the two affine views are 150 and 300 respectively.

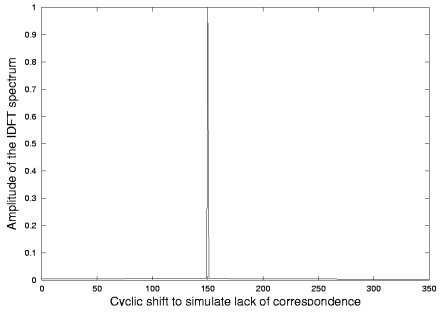

Figure 2. Graph showing the amplitude of the IDFT of $\frac{\kappa_{mod}(l,1)}{\kappa_{mod}(0,1)}$ against the shift for an affine homography when the synthetic shift is 150.

Robustness of Recognition: In the next experiment, we study the recognition accuracy when a zero mean random noise is added to the position of the synthetically transformed shapes related by affine homographies. The highest two singular values for different maximum noise levels

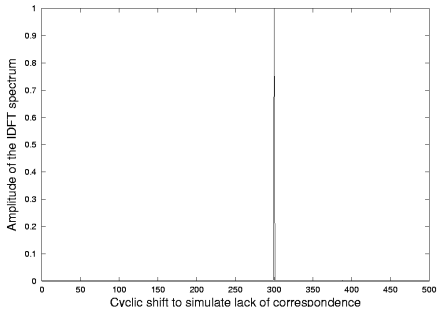

Figure 3. Graph showing the amplitude of the IDFT of $\frac{\kappa_{mod}(l,1)}{\kappa_{mod}(0,1)}$ against the shift for an affine homography when the synthetic shift is 300.

are shown in Table 1, for both cases when the image positions are real numbers and when the image positions are discretized. The ratio of the highest to the next 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. Clearly, the

	Real		Discrete	
Noise	Singular Values		Singular Values	
Level	Highest	Next	Highest	Next
0	247476	0.00187	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 1. Impact of noise on singular values.

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

In the next experiment we demonstrate the performance on a real projective homography. Figure 4 shows three different views of the logo of the International Institute of Information Technology. The ratio of the highest singular value to the next highest singular value of the Θ matrix for various combinations of views is presented in Table 2. When the Θ matrix was constructed for all the three views the two highest singular values were 1.02679e+06 and 2878, i.e. the rank of the matrix can be considered to be 1.

(a) (b) (c)

Figure 4. Three different views of IIIT's logo.

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

Table 2. 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 4.

4 Conclusions

We derived multiview relations for a collection of points using Fourier domain descriptors in this paper and demonstrated a new multiview recognition strategy for planar shapes without explicit correspondences. This scheme can also compute the correspondence between matching shapes. This philosophy can be extended to other collections of points such as textures in multiple views. We are also working on extending the recognition to non-planar objects.

References

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

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

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

[4] H. B. Klaus Arbter, Wesley Snyder and G. Hirzinger. Application of Affine-Invariant Fourier Descriptors to Recognition of 3D Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12, 1990.

[5] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[6] A. Shashua. Algebraic functions for recognition. *IEEE Tran. Pattern Anal. Machine Intelligence*, 16:778–790, 1995.

[7] A. Shashua. Trilinear tensor:The fundamental construct of multipleview geometry and its applications. *Int. Worskshop on AFPAC*, 1997.

[8] B. Triggs. Matching constraints and the joint image. *International Conference on Computer Vision*, 1995.