

PLANAR HOMOGRAPHY FROM FOURIER DOMAIN REPRESENTATION

M. Pawan Kumar, Sujit Kuthirummal, C. V. Jawahar and P. J. Narayanan

Centre for Visual Information Technology
International Institute of Information Technology
Gachibowli, Hyderabad 500019. INDIA.
jawahar@iiit.net

ABSTRACT

Computing the transformation between two views of a planar scene is an important step in many computer vision applications. Spatial approaches to solve this problem need corresponding sets of primitives – points, lines, conics, etc. Identification of corresponding primitives in two images is non-trivial, limiting the applicability of such approaches. In this paper, we present a novel Fourier domain based approach that makes use of image intensities for computing the image-to-image transformation. Our approach transforms the images to the Fourier domain and then represents them in a coordinate system in which the affine transformation is reduced to an anisotropic scaling. The anisotropic scale factors can be computed using cross correlation methods, and working backwards from this, we compute the entire transformation. It does not require any correspondences thereby making it practically very useful. Applications to registration and recognition are discussed.

1. INTRODUCTION

A number of spatial and frequency domain approaches have been proposed to estimate the image-to-image transformation between two views of a planar scene. Most of them are limited to similarity transformations. Spatial domain methods need corresponding points, lines, conics, etc [3, 5, 6], whose identification in many practical situations is non-trivial, thereby limiting their applicability. Many approaches to this problem are based on the tonal information present in the scene [2, 13]. They compute features using the intensity values of the image. Scale, rotation, and translation invariant features have been popular, facilitating recognition under these transformations.

Geometry of multiple views of the same scene has been a subject of extensive research over the past decade. Important results relating corresponding entities such as points and lines can be found in [3, 5]. Recent work has also focused on more complex entities such as conics and higher order algebraic curves [6]. Invariants have been formulated

for contours in multiple views [10]. However, these approaches depend upon extracting corresponding entities such as points, lines or contours and do not utilize the abundant information present in the form of the tonal values in the multiple views of the scene.

In this paper, we present an approach to recognize planar scenes undergoing affine transformations between multiple views. Our technique does not need any correspondence information across views. Kruger and Calway [7] described a similar multiresolution approach to register images under affine transformations. Lucchese [8, 9], presented a Fourier domain based technique to compute the affine transformation between a pair of images of a planar scene. They used radial projections of the energy and a series of non-linear optimization procedures to estimate the transformation. The performance of the algorithm, thus, depends on the choice of good initial estimates. In comparison, the method presented here is linear and hence more robust. Frequency domain methods are in general superior to methods based on spatial features [5] as the whole of the image information is used for matching. They also avoid the crucial issues regarding the selection of the best features.

Section 2 describes some notations used in the paper. The technique for recovering the planar homography relating two views of the same planar scene is discussed in Section 3, while computation of the complete projective transformation is described in Section 4. In Section 5, we discuss the applications of this approach to the problems of rectification, registration and recognition. A few concluding remarks are presented in Section 6.

2. PRELIMINARIES

When a planar object is imaged from multiple views, the transformation which relates these views is known as a projective transformation or homography. It can be represented by a 3×3 non-singular matrix, \mathbf{H} , such that $\mathbf{x}' = \mathbf{H}\mathbf{x}$, where \mathbf{x} and \mathbf{x}' are the homogeneous coordinates of the projections in two views of the same world point [5].

The subgroup of affine transformations is an important

specialization of the projective transformation group as general projective homographies can be approximated in many practical cases by affine homographies (eg. when the scene object is far, the imaging system has high focal length [12], etc.). An affine homography is a linear transformation in inhomogeneous coordinates followed by a translation. It has 6 degrees of freedom accommodating for in-plane rotation, scaling, translation, and shearing. Similarity transformations form a subgroup of affine transformations encapsulating in-plane rotation, scaling and translation. Many problems related to tonal images such as image registration have been solved using Fourier Transforms for the case of similarity transformations [11]. In most applications of practical interest the transformation between views is more complex, rendering these algorithms ineffective.

The Fourier Transform is used for a number of applications in image processing [4]. Given a $N \times N$ image $g(x, y)$, its frequency domain representation is obtained by taking its 2D Discrete Fourier Transform which is defined as

$$G(u, v) = \sum_{x,y=0}^{N-1} g(x, y)W^{(ux+vy)},$$

where $W = \exp(\frac{-j2\pi}{N})$.

3. PLANAR HOMOGRAPHY COMPUTATION

Let a homography \mathbf{H} transforms the point $[x \ y \ 1]^T$ to $[x' \ y' \ 1]^T$. We first describe the relationship between the transformation that relates the images and the transformation that relates their Fourier Magnitude Spectra. This relation is same as the result reported in [1, 7] in a non-homogeneous coordinate system.

Claim 1: If \mathbf{H} is the affine transformation between the spatial representations of two images, then the upper leftmost 2×2 minor of \mathbf{H}^{-T} relates their 2D Fourier Magnitude Spectra.

Proof: Let the affine homography between the two views

$$\text{be } \mathbf{H} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{H}^{-1} = \begin{bmatrix} a' & b' & c' \\ d' & e' & f' \\ 0 & 0 & 1 \end{bmatrix} \text{ be its}$$

inverse. Then, $x = a'x' + b'y' + c'$ and $x' = ax + by + c$. Similarly for y and y' . We can now relate two intensity distributions as $g_2(x, y) = g_1(ax + by + c, dx + ey + f)$ and its transform $G_2(u, v)$ as

$$\begin{aligned} &= \sum_{x',y'=0}^{N-1} g_1(x', y')W^{(u(a'x'+b'y'+c')+v(d'x'+e'y'+f'))} \\ &= G_1(ua' + vd', ub' + ve')W^{uc'+vf'}. \end{aligned}$$

Comparing the magnitudes, we have

$$|G_2(u, v)| = |G_1(ua' + vd', ub' + ve')|. \quad (1)$$

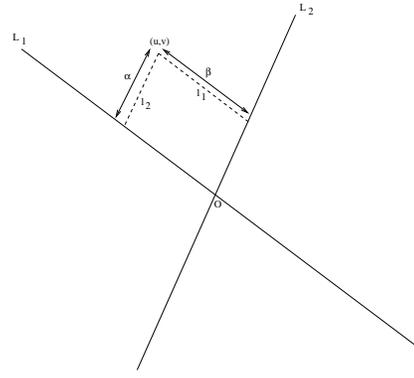


Fig. 1. Transforming from uv space to $\alpha\beta$ space

Thus, the transformation relating the Fourier magnitude spectra is the upper left 2×2 minor of \mathbf{H}^{-T} .

Transformation to $\alpha\beta$ space We have shown above that the Fourier Magnitude Spectra are related by a 2×2 transformation in in-homogeneous coordinates. Therefore, for every line in one spectrum, there is a corresponding line in the other spectrum with the same intensity values. We identify two corresponding lines passing through the origin in the Fourier Magnitude Spectra of the two views and use them to define a new coordinate space – the $\alpha\beta$ space – in which the chosen lines form the axes of the coordinate system. As we will see shortly, this helps us in reducing the problem of finding the affine transformation to a simpler problem of computing the anisotropic scale factors relating the two transformed images. The coordinates of a point in the new system are defined as follows. Let L_i^j be the i th line in view j (Figure 1). To determine the location of a point in the Fourier Magnitude Spectra space (u^j, v^j) in view j , in the $\alpha\beta$ space, we draw lines l_1^j and l_2^j parallel to L_1^j and L_2^j respectively passing through (u^j, v^j) . Let the distance of (u^j, v^j) from L_1^j along l_2^j be α^j and the distance from L_2^j along l_1^j be β^j . We represent the point (u^j, v^j) in the Fourier Magnitude Spectra by (α^j, β^j) in $\alpha\beta$ space.

Claim 2: In the $\alpha\beta$ space, the views differ only by anisotropic scale factors.

Proof Say we have (α^1, β^1) , $(0, r^1)$ and $(s^1, 0)$ in the first view and their corresponding points (α^2, β^2) , $(0, r^2)$ and $(s^2, 0)$ in the second view. (The actual values of r^1 , r^2 , s^1 , and s^2 are irrelevant) Ratios of lengths on parallel lines are preserved in affine transformed views [5]. Therefore,

$$\begin{aligned} \frac{\alpha^1}{s^1} &= \frac{\alpha^2}{s^2} \text{ and } \frac{\beta^1}{r^1} = \frac{\beta^2}{r^2} \\ \alpha^1 &= \psi_1 \alpha^2 \text{ and } \beta^1 = \psi_2 \beta^2 \\ \begin{bmatrix} \alpha^1 \\ \beta^1 \end{bmatrix} &= \begin{bmatrix} \psi_1 \alpha^2 \\ \psi_2 \beta^2 \end{bmatrix} \end{aligned} \quad (2)$$

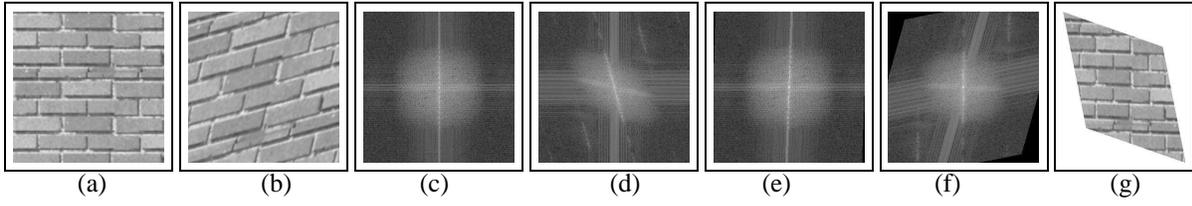


Fig. 2. (a)(b):Images of two views of a brick texture related by an affine homography. (c)(d):Fourier Magnitude Spectra of (a) and (b). (e)(f):Magnitude Spectra of (a) and (b) in $\alpha\beta$ space (g): Image obtained on applying the computed homography to (b)

Here, ψ_1 and ψ_2 are s^1/s^2 and r^1/r^2 respectively.

Therefore, by converting the coordinate system to the $\alpha\beta$ space, we obtain two spectra whose intensity values are related by the unknown anisotropic scale factors ψ_1 and ψ_2 . If we can determine these scale factors, working backwards we can determine the transformation which relates the Fourier Magnitude Spectra in the (u, v) space and hence the transformation between the input images.

Determination of Scale Factors We convert anisotropic scaling into translation by taking the logarithm on both sides of Equation 2 to get

$$\begin{bmatrix} \log \alpha^1 \\ \log \beta^1 \end{bmatrix} = \begin{bmatrix} \log \alpha^2 \\ \log \beta^2 \end{bmatrix} + \begin{bmatrix} \log \psi_1 \\ \log \psi_2 \end{bmatrix} \quad (3)$$

The translation component is $[\log \psi_1 \quad \log \psi_2]$. It can be recovered by looking for an impulse in the inverse Fourier transform of the cross power spectrum of the Log $\alpha\beta$ space representations of the two images. To see this, consider two images f_1 and f_2 related by a translation (x_1, y_1) . Therefore, $f_2(x, y) = f_1(x - x_1, y - y_1)$ Their corresponding Fourier Transforms are related as

$$F_2(u, v) = e^{-j2\pi(ux_1+vy_1)} F_1(u, v)$$

Let $F_2^*(\cdot)$ be the conjugate of $F_2(\cdot)$. Thus, their cross power spectrum would be

$$\frac{F_1(u, v)F_2^*(u, v)}{|F_1(u, v)F_2(u, v)|} = e^{j2\pi(ux_1+vy_1)}$$

By taking the inverse Fourier Transform of the cross power spectrum, we get an impulse at (x_1, y_1) , i.e. we have a function which is approximately zero everywhere except at (x_1, y_1) . The antilog of the coordinates of the location of the impulse gives the non-linear scale factors ψ_1 and ψ_2 . This method of determining ψ_1 and ψ_2 is accurate and tolerant to high levels of noise.

Computation of the Transformation Once the non-linear scale factors have been identified, we get corresponding points in $\alpha\beta$ space. Converting them to uv space results in dense correspondence in the Fourier Magnitude Spectra. From

them we can obtain the upper leftmost 2×2 minor of \mathbf{H}^{-1} in a linear least squares manner using the DLT algorithm [5]. This transformation is then applied to the second image to get the transformed image $g_t(x, y)$. Now all that remains unknown is the translation component. As described earlier, the translation relating two images can be obtained from the location of the impulse in the inverse Fourier Transform of the Cross Power Spectrum. Using $g_1(x, y)$ and $g_t(x, y)$, we obtain the translation component as well. The affine homography \mathbf{H} relating the original images $g_1(x, y)$ and $g_2(x, y)$ is thus obtained.

Implementation Details From Equation 1, we can see that every line in the Fourier Magnitude Spectrum gets transformed to another line in the other view. The intensity value of the corresponding points on the two lines remains the same. However, in practice, due to noise, this might not be the case. Hence, identification of corresponding lines in two Fourier Magnitude Spectra is not trivial when the spectra have many dominant lines. We overcome this problem by taking the two most dominant lines in the first view and the five most dominant lines in the second. Even with high levels of noise, two of these five lines correspond to the two lines chosen in the first view. Therefore, we try all permutations of the five lines in the second view taken two at a time as possible line correspondences, thus resulting in a total of 20 permutations. These 20 ‘correspondences’ would generate 20 homographies, which are used to map the second view into 20 new views. We compute the cross correlation of each of these new views with the first view and declare that transformation as the result for which corresponding cross correlation value is highest.

Computation of Projective Transformation In [9], the computed affine approximation of the transformation [8] is used as an initial estimate in a non-linear optimization procedure to compute the projective components of the transformation. Our technique produces a far more robust estimate of the affine transformation (See comparison of recognition accuracies on a Brodatz texture dataset in the next Section) and hence provides as good if not better initial estimates for an identical non-linear optimization procedure to

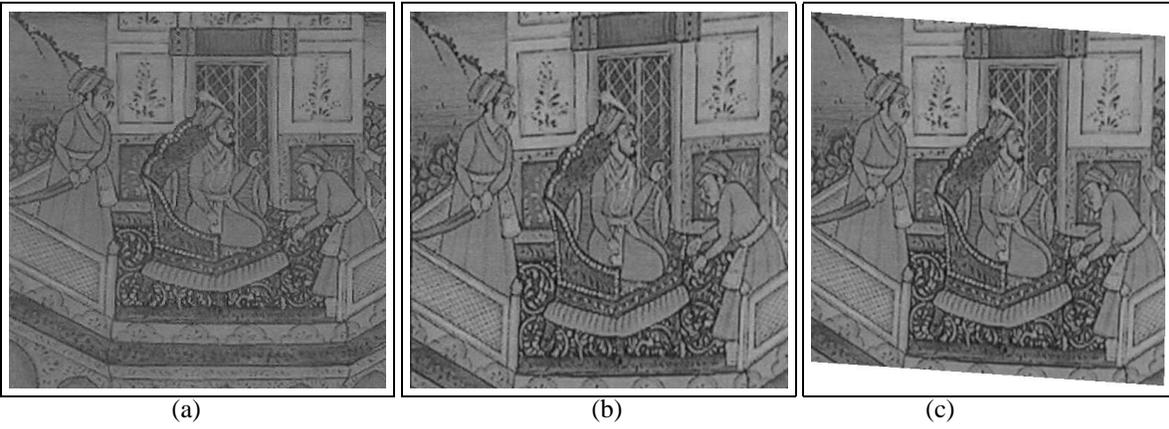


Fig. 3. (a),(b):Images of a painting taken from two different view points (c) Image obtained on applying the computed affine transformation to (b). Notice that the rectangular door and wall tiles appear slanted in (b), but not so in (c).

compute the complete projective transformation. We refer the reader to [9] for details of this step. Our technique can thus provide a sounder basis for recovery of the complete transformation than previous methods.

4. APPLICATIONS

4.1. Homography Calculation

We now present the results of two of the many experiments we conducted to evaluate the efficacy of our approach for homography computation described in 3. Figures 2(a) and (b) show two views of a brick wall texture related by an affine transform. Their Fourier Magnitude Spectra are shown in Figures 2(c) and (d). Using two corresponding dominant lines of these spectra the coordinate axes are changed to result in Figures 2(e) and (f). The anisotropic scale factors are then determined from the inverse Fourier Transform of the cross power spectrum of the log images constructed from these images. Point correspondences were then obtained using these scale factors and the affine transformation relating the two views was determined as described in Section 3. Applying the transformation to the second view results in the view shown in Figure 2(g).

Figures 3(a) and (b) show two views of a painting taken from two different viewpoints and hence related by a projective transformation. We calculate the affine approximation of this projective transformation and apply it to the second view to obtain Figure 3(c). Notice that the rectangular door and wall tiles appear slanted in (b). This ‘distortion’ does not appear in (a) and (c). An affine approximation of the projective homography seems sufficient in this case.

4.2. Registration and Mosaicing

Image registration involves identifying the relative orientations of two or more overlapping images of the same scene.

Most of the direct image registration [11] algorithms are restricted to handling similarity transformations which though useful, prove to be insufficient in many real life situations wherein the image-to-image transformations are more complex. The technique that we have designed can handle a more general class of transformations between views making it more applicable to solving the problem of registration.

Once registration has been achieved, we can perform automatic image mosaicing. The homography between any two consecutive images of the given image sequence is computed and then used to align and stitch the images together to obtain a large field of view image. Mosaicing deals with obtaining an image with a large field of view by aligning and stitching a sequence of images with a small field of view and considerable overlap. The images in Figure 4(a) and (b) were mosaiced to get the large field of view image shown in Figure 4(c). Note that the two input images are related by a general homography.

4.3. Planar Scene Recognition

The problem of planar scene recognition can be stated as follows : Given the reference view of various planar scenes and a test (distorted) view of one of them, identify the correct reference scene. To compare the efficacy of our technique of computing the transformation and hence perform recognition, with the previous work that is closest to ours [8, 9], we tested on a dataset of 200 views of 5 Brodatz textures. The views are generated synthetically by applying general projective transformations to the Brodatz textures. Figure 5 shows some of the images used as input in our recognition experiment. For each texture class there is one representative sample. Given a texture view to be classified, we compute the transformations between the test view and each of the reference views using both methods. For each computed transformation, we calculate a confidence value

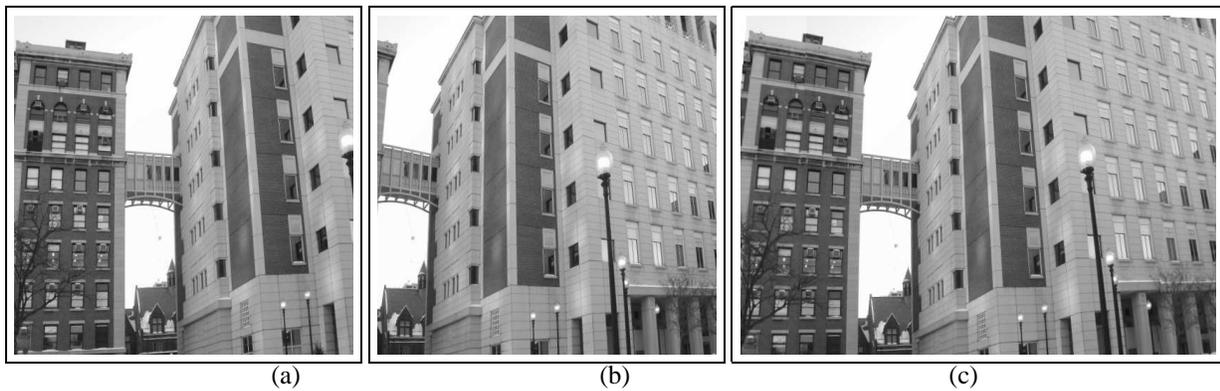


Fig. 4. (a): Input image -1, (b):Input image -2, (c): Mosaiced image

– the cross correlation of the corresponding reference view and the test view transformed using the computed transformation. The test view is labeled as belonging to the class of the reference view for which the confidence value is highest. Recognition accuracies using the two techniques is given in Table 1. Clearly our technique provides better recognition accuracies.

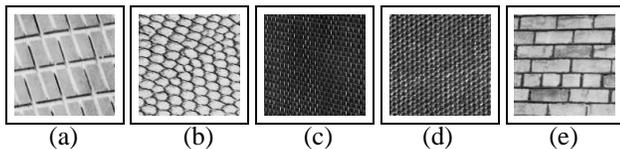


Fig. 5. Examples of Brodatz textures used as input for the recognition experiment

Texture	1	2	3	4	5
Method [8]	75	70	72.5	75	65
This	100	100	97.5	100	97.5

Table 1. Recognition accuracies (%) obtained for various Brodatz textures using the technique in [8] and this paper

5. CONCLUSIONS

This paper presents a novel technique for determining the affine transform relating two views of a planar scene using the geometric information encoded in the multiple views. Unlike most of the previously reported geometric approaches that usually use entities such as corresponding points, lines and other contours, our method uses the tonal information present in them. It does not need correspondence across views. Furthermore, our technique is linear and hence more robust than earlier tonal based methods which are non-linear. The computed affine transformation can then be used as an

initial estimate in a non-linear optimization procedure to recover the complete projective transformation. Our proposed scheme is found to be effective for rectification, registration and recognition.

6. REFERENCES

- [1] R. N. Bracewell, K. Y. Chang, A. K. Jha, and Y. H. Wang. Affine theorem for two-dimensional fourier transform. *Electronic Letters*, 29:304, 1993.
- [2] M. Chantler, G. McGuinnigle, and J. Wu. Surface rotation invariant texture classification using photometric stereo and surface magnitude spectra. In *BMVC*, pages 486–495, 2000.
- [3] O. Faugeras and Q. Luong. *The Geometry of Multiple Images*. MIT Press, 2001.
- [4] R. Gonzalez and R. Woods. *Digital Image Processing*. Addison-Wesley, 1992.
- [5] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2000.
- [6] J. Kaminski and A. Shashua. Multiple view geometry of algebraic curves. *IJCV*, 2003.
- [7] S. Kruger and A. Calway. Image registration using multiresolution frequency domain correlation. In *Proc. BMVC*, pages 316–325, 1998.
- [8] L. Lucchese. A frequency domain technique based on energy radial projections for robust estimation of global 2d affine transformations. *Computer Vision and Image Understanding*, 81:72–116, 2001.
- [9] L. Lucchese. A hybrid frequency-space domain algorithm for estimating projective transformations of color images. *ICIP*, 2:913–916, 2001.
- [10] J. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992.
- [11] B. S. Reddy and B. N. Chatterji. An FFT-based technique for translation, rotation and scale-invariant image registration. *IEEE Trans. Image Processing*, 5:1266–1271, Aug. 1996.
- [12] L. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *IJCV*, 16:147–182, 1995.
- [13] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *ECCV*, volume 3, pages 255–271. Springer-Verlag, 2002.