Estimation of the Fundamental Matrix Based on Complex Wavelets

Tao Hong, Nick Kingsbury
Signal Processing Lab, Cambridge University,
Cambridge, Cambridgeshire, UK,
CB2 1PZ
th315@cam.ac.uk, ngk@eng.cam.ac.uk

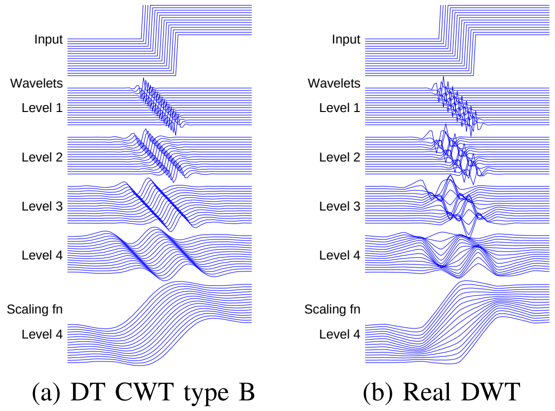
Abstract— **In this paper, an automatic fundamental matrix estimation method based on complex wavelets is presented. The fundamental matrix is considered important because it reflects the intrinsic projective geometry of the scene. It is widely used in computer vision areas, such as camera calibration, object reconstruction, visual navigation, stereo vision etc. In comparison with the Discrete Wavelet Transform (DWT), the dual-tree complex wavelet transform (DT CWT) possesses two key properties for computer vision: shift invariance, which makes it possible to extract stable local features in an image; and good directional selectivity, making it possible to measure image energy accurately in multiple directions. First, a feature detector based on complex wavelets is used to find the points of interest, and then complex-wavelet-based polar matching is used to find putative correspondences. Compared with the classic 'Harris corner' interest point detector, the interest point detector based on DT CWT is a multiscale interest point detector, able to detect different kinds of features, including corners, edges, blobs etc. and the number of interest points can be made scale-dependent. Polar matching is a rotation invariant descriptor derived from the DT CWT coefficients; and scale invariance is induced by adjusting the wavelet levels and sampling radius according to the scale estimated by the detector. A minimum of only 7 correspondence points are needed to compute the fundamental matrix. Preliminary tests on some classic building scene images show that the method works well.**

Index Terms— **Complex Wavelets, Fundamental Matrix Estimation, Feature Extraction, Polar Matching**

I. Introduction

In order to obtain 3D information from multiple views of a scene, there are two main methods. In 1986, Tsai[1], and Faugeras and Toscani [2] built a model with the 3D pixel coordinates in their camera calibration problem. However, there are 11 parameters in this projection matrix. So in 1992, Mundy and Zisserman [3] proposed to directly use projection information without computing specific camera parameters. Compared with the first method, it has the advantage that 7 parameters are sufficient for the projection, whose information is entirely encapsulated in the fundamental matrix. The fundamental matrix reflects the intrinsic projective geometry between two views of the same scene and it is not related to the structure of scene. Incorrect correspondence points between the projection images of a scene can be searched for and deleted because the correct correspondences are constrained to lie on an epipolar lines computed from the fundamental matrix. Hence the computation of the fundamental matrix is a necessary step for many tasks in computer vision, such as camera calibration, object reconstruction, visual navigation, stereo vision etc.

Since Mallat first demonstrated wavelets as the foundation of multi-resolution theory for signal processing and analysis in 1987[4], [5], the Discrete Wavelet Transform (DWT) has been widely and successfully used in many areas of image processing including denoising and compression (JPEG 2000) etc. However, for object recognition, there is an important shortcoming in the DWT: the lack of shift invariance. This means that the distribution of energy between coefficients at different scales may vary sharply as the input signal shifts. The dual-tree complex wavelet transform (DT CWT), Kingsbury[6], overcomes this disadvantage by introducing limited redundancy into the transform, which makes it possible to extract stable local features. Fig.1 shows an example of shift invariance. The input signal is a unit step and it is shifted to 16 adjacent sampling instances in turn. Output signals are reconstructed from the wavelet coefficients, one level at a time. From the figure, we can see the approximate shift invariance in 1-D of the subband transfer functions of the DT CWT compared with those of the DWT.



(a) DT CWT type B (b) Real DWT

Fig. 1. Wavelet and scaling function components, at levels 1 to 4, of 16 shifted step responses of the DT CWT (a) and the DWT (b).[6]

In [7], experiments have shown that aliasing effects due to decimation within the transform are small enough to be

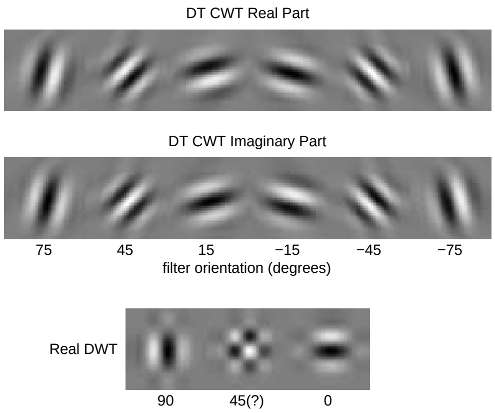


Fig. 2. Basis functions of the 2-D DT CWT at level 4, showing real and imaginary parts for 6 directional subbands, contrasted with basis functions for the 3 bands of the DWT.

neglected for most image processing purposes. When we use a transform method to capture object features, the method should be shift invariant. Since a feature may present itself in an image under varying shifts, it should be processed in a shift-invariant way, or otherwise it will become more difficult to recognize the feature. The practical advantages of shift invariant transforms can be seen more clearly when we reconstruct a signal from only a limited number of transform levels.

The DT CWT also provides true directional selectivity. For example, a 2-D DT CWT can provide six sub-bands of complex coefficients, each of which is oriented at a different angle. It is illustrated by the level 4 impulse responses in Fig.2. The ability to discriminate an object's directional energy distribution is very powerful. Studies have shown that six subbands are a sufficient number for most texture recognition and modelling tasks[8].

Interest point detection is a necessary initial step for automatic fundamental matrix estimation. A typical interest point detector is the Harris corner detector[9], which detects corners by calculating the differential of the corner score with respect to direction. In [10], Fauqueur and Kingsbury proposed a multi-scale interest point detector by using an accumulated energy map, which is built by accumulating the interest point energies achieved from the geometric means of the 6 oriented DT CWT coefficients at each scale. Compared with the Harris corner detector, this has the important advantage of being a multi-scale detector, and so it is not necessary to define the scale in advance.

In [11], Kingsbury proposed a rotation-invariant local feature descriptor, known as polar matching, in which the coefficients of DT CWT are formed into a rotationally symmetric polar matching matrix whose elements are the interpolated coefficients of the 12 sampled points around a ring, plus the centre point. Polar matching is used to find the putative correspondences to compute the initial fundamental matrix and is unaffected by rotations (usually perspective-induced) between the two images.

In this paper, an automatic fundamental matrix estimation method based on complex wavelets is presented. Preliminary tests on the algorithm using real images show that it works well. The following is the structure of the paper: section 2 introduces the details of the algorithm; section 3 shows the experiment results; and section 4 discusses the test results and gives the conclusions.

II. Algorithm Description

The fundamental matrix is important because it depends only on the pose and internal parameters of the camera. Let x and x' be a pair of of corresponding points in homogeneous image coordinates and let the fundamental matrix F be a 3×3 matrix. Then they should satisfy the following equation:

$$x'^T F x = 0 \tag{1}$$

Let I' define an epipolar line satisfying $x'^T I' = 0$, such that Fx defines the epipolar line of point x by $I' = Fx$. The epipolar line is the image in one camera of the ray from the other camera's optical center to the 3D world coordinate P. The corresponding point x' on the other image should lie on the epipolar line. Similarly, the epipolar line I corresponding to x' can be described by $I = F^T x'$.

The fundamental matrix has the advantage that it does not require the cameras to be calibrated, so it is not necessary to know the essential matrix E. Let u, v be the image coordinates of point x, and u', v' be the image coordinates of the corresponding point x'. We then get the following constraint on the 3×3 fundamental matrix F:

$$\begin{pmatrix} u' & v' & 1 \end{pmatrix} \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0 \tag{2}$$

Each pair of correspondence points can provide one such constraint. If there are n pairs of correspondence points, then we get

$$\begin{pmatrix} u'_1 u_1 & u'_1 v_1 & u'_1 & u_1 v'_1 & v_1 v'_1 & v'_1 & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u'_n u_n & u'_n v_n & u'_n & u_n v'_n & v_n v'_n & v'_n & u_n & v_n & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = 0 \tag{3}$$

If there are 8 or more perfect correspondence points, then the fundamental matrix F can be determined uniquely up to scale. However, the image measurements will usually be noisy, so least squares can be used to solve the overdetermined system of equations if $n > 8$. The linear criterion for the estimate of the fundamental matrix can then be defined as the solution of

$$\min_F \sum_{i=1}^{n} (x_i'^T F x_i)^2 \tag{4}$$

According to equation 1, the value of determinant F is required to be zero, that is $det(F) = 0$. If the 8 correspondence points are noisy, then the estimated value of F will not have zero determinant and the epipolar lines will not meet at a point. In this condition, we can solve for the fundamental matrix by making use of the singularity constraint, which means that the fundamental matrix is singular, in fact of rank 2. In this case, the minimum n to compute the fundamental matrix only requires 7 correspondence points, and nonlinear criteria can be used to estimate the fundamental matrix F. There are two measurement methods for the nonlinear criteria: the first is the distance to epipolar lines and the the second is the gradient criterion. In the first method, the sum of squares of distances of a point to the corresponding epipolar line is used as the cost function. The second measurement method uses a surface fitting between the data and the surface defined by equation 1. The nonlinear criteria can be expressed as follows:

$$\min_{det(F)=0} \sum_{i=1}^{n} w(F, x_i', x_i)({x_i'}^T F x_i)^2 \tag{5}$$

where $w(F, x_i', x_i)$ is a weighting function. If the projection is pure planar motion, then it becomes an affine transform. The affine transform has 6 degrees of freedom. Generally speaking, there are 3 kinds of methods to compute the fundamental matrix, the normalized 8-point algorithm, the 7-point algorithm and the affine transform algorithm [12]. Here, we propose a fundamental matrix estimation method based on complex wavelets which can be used for any of the 3 methods of fundamental matrix estimation. For 3D scenes, the 7-point method is very popular because it requires fewer correspondence points and it is suitable for all conditions including non-affine movement. Hence we use the 7-point algorithm as an example. The details of the algorithm can be described as follows:

1) Interest points. Interest points are extracted with the DT CWT feature detector.
2) Putative correspondences. Putative correspondences are detected by polar matching.
3) RANSAC robust estimation. F is estimated from 7 random corespondences and then re-estimated from all corresponding points that are classified as inliers. This is repeated K times to find the best solution for F.
4) Guided matching: With the estimated fundamental matrix F, the correspondence points can be determined within a search strip about the epipolar line.

In the automatic computation of the fundamental matrix F, the first two steps are to detect the interest points and find the putative correspondences. In the the classic method[12], [13], typically the Harris corner detector is used to detect the interest points and the putative correspondences are chosen based on similarity of their intensity neighbourhoods. The Harris corner detector[9] is based on a Gaussian kernel filter, so it is necessary to decide the scale of the Gaussian kernel at the start. Meanwhile, it is hard to choose the number of the interest points. The Harris corner detector also can only detect corners, not other types of features (blobs, edges etc.). Compared with the Harris corner detector, our interest point detector based on DT CWT[10] is a multiscale interest point detector because it is based on the accumulated energy map, which is achieved by accumulating energy from different levels of the DT CWT. The number of the interest points can then be chosen by sorting the energy of the candidate interest points and picking the largest N of these. Moreover, the DT CWT detector can detect several different types of features (corner, edge, blob etc.) which have edge energy in multiple directions at the same spatial location.

Classic methods for finding putative correspondences are based on proximity and similarity of their intensity neighbourhoods. The polar matching method has the additional advantage of being a rotation invariant interest point descriptor. We improved the scale invariance of polar matching by adjusting the sampling radius and wavelet level selection of polar matching according to the estimated scale from the interest point detector. In step 4, we tried two kinds of methods, least squares and weighted least squares to re-estimate the fundamental matrix F. From our experiments, we found that the least squares method is simpler and more stable, so it is adopted in our algorithm.

III. Experimental Results

The algorithm has been tested on some classic building scene images, as Fig 3 where the left image is thought as the first image and the right image is the second image. Our aim is to estimate the fundamental matrix between them based on complex wavelets. The first step is to extract the interest points from the left image and right image. The size of the image is 1024×768. DT CWT levels 2, 3 and 4 are used to form the accumulated map. The second step is to find the putative correspondences of the features between the template and the test image. In each image, we detect the 400 strongest interest points with the DT CWT detector because this is found to be sufficient to represent the image features. Here, we use the feature of the DT CWT detector which allows us to choose the number of interest points.

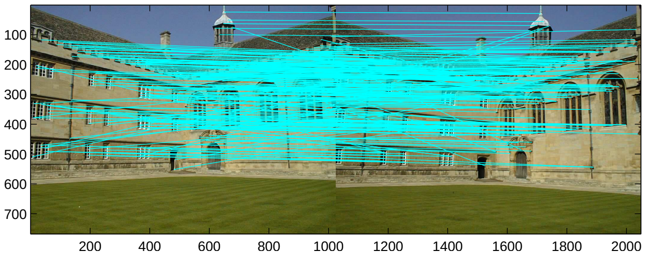


Fig. 3. Putative matches from polar matching. Initially, 400 interest points are detected in each image with the DT CWT interest point detector and then 297 matched corresponding points are found with polar matching.

If the score from a polar matching correlation is larger than a threshold, then it is thought to be a correct match. Here, we set the threshold at 0.7, where the polar matching vectors are normalized so the score must lie between 1 and -1. The threshold is selected from experiments where it is found that 0.7 can provide enough correspondences for the object recognition and localisation, while keeping the proportion of correct matches in the correspondences reasonably high. The result of polar matching can be shown in Fig 3, which shows the interest points and the putative matches with DT CWT. Here, we found 297 matched correspondence points, which is plenty to estimate the fundamental matrix. The blue lines show the putative matches, and the ends of the blue lines stand for the locations of the interest points.

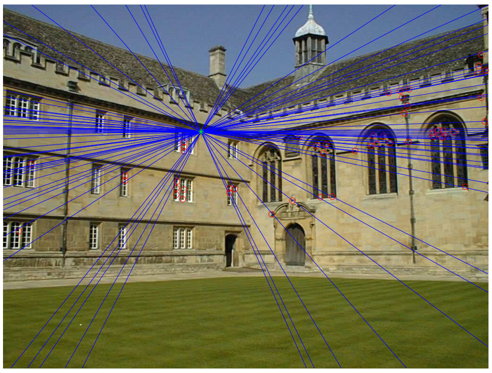


Fig. 4. 50 Epipolar lines on the first image. The red circles denote the interest points and blue lines are the epipolar lines. The green $*$ is the epipole.

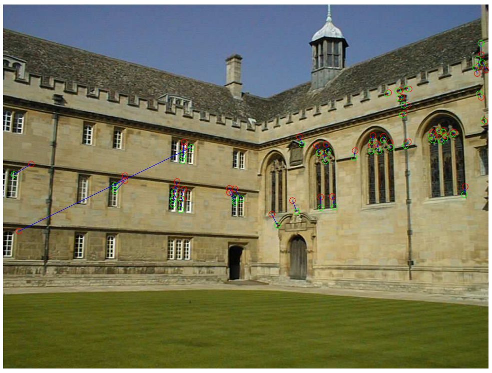


Fig. 5. 50 inlying matched interest points in the second image. The red circle denotes the 50 interest points of Fig 4. The green circle denotes the corresponding 50 interest points and their positions in the second image.

After this, the RANSAC algorithm is used to find 7 correspondence points to compute the fundamental matrix F and least squares is used to the compute the error, where the nonlinear criteria, the distance to epipolar lines is used as the measurement of the error. It is repeated $K = 20$ times to estimate the best fundamental matrix F. A distance threshold between each data point and the corresponding epipolar line is used to decide whether a point is an inlier or not. The point coordinates are normalized to that their mean distance from the origin is $\sqrt{2}$. If the Sampson distance of a data point to its corresponding line is less than 0.002, then it is regarded as inlier point, otherwise it is regarded as outlier. There are 145 inliers in the example. In order to show it more clearly, 50 inlier points and their corresponding epipolar lines are chosen. Fig.4 displays the corresponding epipolar lines on the first image. The red circles denote the interest points from the first image, which are detected by the DT CWT interest point detector. Blue lines show the epipolar lines, which are computed by multiplying the F matrix with the corresponding interest points from second image. If a red circle is near to the corresponding epipolar line, it means the F matrix fits well. Ideally, the blue line should cross the corresponding red circle. The green $*$ is the epipole.

Fig.5 displays the first image with 50 inlying matched feature points. The red circle denotes the 50 interest points of Fig.4. The green circle denotes the corresponding 50 interest points and their positions in the second image. According to the Fig.4 and Fig.5, we can see that most of the interest points matched correctly.

Fig.6 shows the distance from the interest points to epipolar lines. The width of the search strip is 8 pixels and 116 correspondences are found. The size of the image is 1024 × 768. The distance from each interest point to its corresponding epipolar line is regarded as the error. Figure 6 is the histogram of the error, from which we can see that most of the error is less than ±4 pixels. If the error is near to 0, then it means that the F matrix fits well. The mean error is 1.92 pixel, which we judge to be an encouraging result.

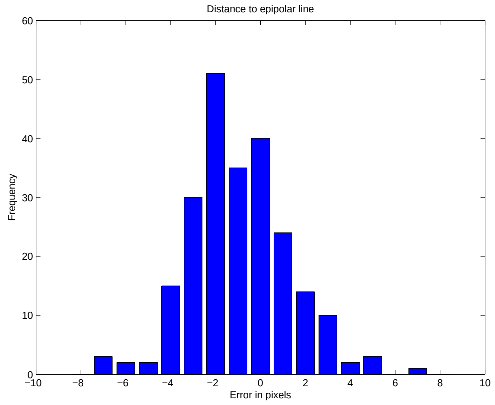


Fig. 6. The histogram of the error, which is the distance from each interest point to its corresponding epipolar line.

IV. Conclusion

In this paper, we have presented an automatic fundamental matrix estimation method based on dual-tree complex wavelets. We use a multiscale interest point detector based on complex wavelets, which has the advantage that it is able to detect different kinds of features, including corners, edges and blobs and it allows easy choice of the number of interest points. Polar matching is a rotation-invariant descriptor, and scale invariance is achieved by adjusting the sampling radius of polar matching according to the estimated scale. A RANSAC algorithm with least squares minimisation is used to compute the fundamental matrix. Some classic building scene images are adopted to test the algorithm and preliminary results show that it works well. The fundamental matrix can be applied in many areas i.e. camera calibration, object reconstruction, visual navigation, stereo vision etc. Our future work is aimed at localising objects in the image more accurately based on the estimated fundamental matrix.

References

[1] R.Y. Tsai et al., "An efficient and accurate camera calibration technique for 3D machine vision," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Miami Beach: FL, 1986, vol. 374.

[2] O. Faugeras and G. Toscani, "The calibration problem for stereo," in *Proceedings*. IEEE Computer Society Press, 1986, p. 15.

[3] J.L. Mundy and A. Zisserman, "Geometric invariance in computer vision," *MIT Press Cambridge, MA, USA*, p. 540, 1992.

[4] S. Mallat, "A compact multiresolution Representation: The Wavelet Model," *Proc.IEEE Computer Society Workshop on Computer Vision*, pp. 2–7, 1987.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[6] N. Kingsbury, "Shift invariant properties of the dual-tree complex wavelettransform," *Acoustics, Speech, and Signal Processing, 1999. ICASSP'99. Proceedings., 1999 IEEE International Conference on*, vol. 3, 1999.

[7] N. Kingsbury, "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, 2001.

[8] BS Manjunath and WY Ma, "Texture features for browsing and retrieval of image data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 8, pp. 837–842, 1996.

[9] C. Harris and M. Stephens, "A combined corner and edge detector," *Alvey Vision Conference*, vol. 15, pp. 50, 1988.

[10] J. Fauqueur, N. Kingsbury, and R. Anderson, "Multiscale keypoint detection using the dual-tree complex wavelet transform," *Proceedings of the IEEE International Conference on Image Processing (ICIP06)*, 2006.

[11] N. Kingsbury, "Rotation invariant local feature matching with complex wavelets," *Proceedings of 14th European Signal Processing Conference (EUSIPCO06)*, 2006.

[12] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, vol. PART 2: Two-View Geometry, Chapter 11, Computation of the Fundamental Matrix F, Cambridge Univ Press, 2003.

[13] PHS Torr and A. Zisserman, "Feature Based Methods for Structure and Motion Estimation," in *Vision algorithms: theory and practice: International Workshop on Vision Algorithms, Corfu, Greece, September 21-22, 1999: proceedings*. Springer Verlag, 2000, p. 278.