CS6790: Project report

Robust trifocal tensor estimation using six points correspondences

Santhosh Kumar R, *EE12B101,* and Vignesh Krishnakumar, *CS10B054,*

Abstract

A robust algorithm for computing trifocal tensor has been implemented. The trifocal tensor is computed from six point correspondences, and a parameterization which enforces the constraints between tensor elements has been used. The input to the algorithm is three images of the same scene, and the output is the estimated tensor. The algorithm further computes a Maximum Likelihood Estimate of the trifocal tensor, assuming a Gaussian noise model in the estimation of point matches.

I. Introduction

The trifocal tensor is a measure of geometric consistency for three views, an analogue of the fundamental matrix for two view geometry. It captures all the projective constraints between three views. The tensor depends on the relative motion between the three views and internal parameters of the camera, but it can be computed from image correspondences without using the knowledge of the motion or calibration.

The trifocal tensor can be used for point and line transfer. Given a pair of corresponding points in two views, we can estimate the corresponding point in the third view. This is called point transfer. We can similarly transfer lines to the third view. The advantage with trifocal tensor based point transfer is that it does not fail for points on the trifocal plane. Epipolar transfer fails for points on or close to the trifocal plane.

The algorithm consists of three stages.

- Corners are detected in each image and matching triplets are obtained
- Using RANSAC and six point correspondence, we obtain an estimate of the Trifocal tensor
- The output of RANSAC is further refined to obtain a Maximum Likelihood Estimate of the tensor.

II. ALGORITHMIC DESCRIPTION

Our implementation consists of three stages as mentioned in the introduction. We explain each of the stages below.

A. Corner points detection

Corner points are obtained in each image using Harris corner detector. Corners are detected to sub-pixel accuracy using this detector. Putative correspondences are obtained over the three images using similarity of features.

B. Trifocal tensor estimation using RANSAC

A random sample of six corner correspondences to estimate the Trifocal tensor. The inlier rato is computed for the estimate, and correspondence which gives sufficient inlier ratio is selected.

1) Trifocal tensor computation for six point correspondences: The six 3D points are assigned canonical projective coordinates as follows: $(1,0,0,0)^T$, $(0,1,0,0)^T$, $(0,0,1,0)^T$, $(1,1,1,1)^T$ and $(X,Y,Z,W)^T$ where X,Y,Z,W are unknown. The six image points can be assigned the homogeneous coordinates $(1,0,0)^T$, $(0,1,0)^T$, $(0,0,1)^T$, $(1,1,1)^T$, $(x_5^{(i)}, y_5^{(i)}, w_5^{(i)})^T$, $(x_6^{(i)}, y_6^{(i)}, w_6^{(i)})^T$. It can be shown the any given set of coordinates can be converted to the canonical frame of reference, using the transform $\mathbf{B}^{(i)}$. This can be computed efficiently as follows

$$\mathbf{B}^{(i)} = \begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \lambda_3 x_3 \end{bmatrix}^{-1}$$

where $(\lambda_1\ \lambda_2\ \lambda_3)^T = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^{-1} x_4$

Given points in the canonical frame of reference, we obtain the following equation:

$$\begin{bmatrix} 1 & 0 & 0 & 1 & x_5^{(i)} & x_6^{(i)} \\ 0 & 1 & 0 & 1 & y_5^{(i)} & y_5^{(i)} \\ 0 & 0 & 1 & 1 & w_5^{(i)} & w_6^{(i)} \end{bmatrix} = \begin{bmatrix} \alpha^{(i)} & 0 & 0 & \delta^{(i)} \\ 0 & \beta^{(i)} & 0 & \delta^{(i)} \\ 0 & 0 & \gamma^{(i)} & \delta^{(i)} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & X \\ 0 & 1 & 0 & 0 & 1 & Y \\ 0 & 0 & 1 & 0 & 1 & Z \\ 0 & 0 & 0 & 1 & 1 & W \end{bmatrix}$$

for each image $i = 1,2,3$. Thus the trifocal tensor estimation is equivalent to estimating the coordinates of the sixth point $(X,Y,Z,W)^T$ and $(\alpha^{(i)}, \beta^{(i)}, \gamma^{(i)}, \delta^{(i)})$ for each camera.

Using the above equation, we can obtain the following equation:

$$\begin{bmatrix} w_5^{(i)} & 0 & -x_5^i & w_5^i - x_5^i \\ 0 & w_5^{(i)} & -y_5^{(i)} & w_5^i - y_5^{(i)} \\ w_6^{(i)} X & 0 & -x_6^{(i)} Z & w_6^{(i)} W - x_6^{(i)} W \\ 0 & w_6^{(i)} Y & -y_6^{(i)} Z & w_6^{(i)} W - y_6^{(i)} W \end{bmatrix} \begin{pmatrix} \alpha^{(i)} \\ \beta^{(i)} \\ \gamma^{(i)} \\ \delta^{(i)} \end{pmatrix} = 0$$

The 4×4 matrix on the left has rank 3. By equating the determinant to 0, we obtain

$$(-x_5^{(i)}y_6^{(i)}+x_5^{(i)}w_6^{(i)})(WX-YZ)+(x_6^{(i)}y_5^{(i)}-y_5^{(i)}w_6^{(i)})(WY-YZ)+(-x_6^{(i)}w_5^{(i)}+y_6^{(i)}w_5^{(i)})(WZ-YZ)+(-x_5^{(i)}w_6^{(i)}+y_5^{(i)}w_6^{(i)})(XY-YZ)+$$

Using this equation in all the 3 images, we obtain

$$A\mathbf{t} = 0$$

where A is the matrix of coefficients and $\mathbf{t} = (WX - YZ, WY - YZ, WZ - YZ, XY - YZ, XZ - YZ)$. The vector $\mathbf{t}$ thus lies in the null space of A which has rank 3 in general. By using SVD, we can obtain the $\mathbf{t}$ vector. After obtaining a solution for $\mathbf{t}$, we can find $(X, Y, Z, W)^T$ as follows:

$$\frac{X}{W} = \frac{t_4 - t_5}{t_2 - t_3}$$

$$\frac{Y}{W} = \frac{t_4}{t_1 - t_3}$$

$$\frac{Z}{W} = \frac{t_5}{t_1 - t_2}$$

(where $t = [t_1, t_2, t_3, t_4, t_5]^T$) assuming W is not 0. In the case where $W = 0$, we can trivially solve for X, Y, Z by setting $W = 0$ in the beginning.

Given $(X, Y, Z, W)^T$, we can solve for the camera parameters using the relations described above by applying SVD and selecting the eigenvector corresponding to the 0 eigenvalue. We can convert this back to the original coordinate system and obtain the camera matrix.

$$\mathbf{P}^{(i)} = \mathbf{B}^{-1} \begin{bmatrix} \alpha^{(i)} & 0 & 0 & \delta^{(i)} \\ 0 & \beta^{(i)} & 0 & \delta^{(i)} \\ 0 & 0 & \gamma^{(i)} & \delta^{(i)} \end{bmatrix}$$

To recover the trifocal tensor, the first camera is set to $[\mathbf{I}|0]$.This is done by taking the homography H,

$$H = \begin{bmatrix} P_1^{(1)} \\ P_2^{(1)} \\ P_3^{(1)} \\ P_1^{(1)} \times P_2^{(1)} \times P_3^{(1)} \end{bmatrix}^{-1}$$

Then the elements of the trifocal tensor are given by:

$$T_{ijk} = P_{ji}^2 P_{k4}^3 - O_{j4}^2 P_{ki}^3$$

2) Point transfer: Point transfer can be accomplished using the 9 expressions given by triliniearities, four of which are linearly independent.

$$x_l^3 = x_i^2 \sum_{k=1}^{k=3} x_k T_{kjl} - x_j^2 \sum_{k=1}^{k=3} x_k T_{kil}$$

for $i, j = 1, 2, 3$.

C. Maximum Likelihood Estimate of the trifocal tensor

It is assumed the each coordinate of each point has a Normal noise associated with it. Therefore, given a true correspondence $x_i - x_i^2 - x_i^3$, the probability density function of the data is

$$Pr(\hat{x_i}^{1,2,3}|\mathbf{T}) = \prod_j \frac{1}{\sqrt{2\pi}\sigma} \exp(-(x_i^j - \hat{x_i}^j)^2/(2\sigma^2))$$

There are 9 trilinearities given, out of which 4 are independent. They are

$$x_i T_{i11} - x_i x_1^3 T_{i13} - x_i x_1^2 T_{i31} + x_i x_1^2 x_1^3 T_{i33} = 0$$

$$x_i T_{i21} - x_i x_1^3 T_{i23} - x_i x_2^2 T_{i31} + x_i x_2^2 x_1^3 T_{i33} = 0$$

$$x_i T_{i12} - x_i x_2^3 T_{i13} - x_i x_1^2 T_{i32} + x_i x_1^2 x_2^3 T_{i33} = 0$$

$$x_i T_{i22} - x_i x_2^3 T_{i23} - x_i x_2^2 T_{i32} + x_i x_2^2 x_2^3 T_{i33} = 0$$

Based on Taylor series expansion of the trilinearities about the true correspondence, we get

$$\mathbf{t}(\hat{x}^{1,2,3}) = \mathbf{t}(x^{1,2,3}) + \mathbf{J}(\hat{x}^{1,2,3} - x^{1,2,3}) = r + \mathbf{J}(\hat{x}^{1,2,3} - x^{1,2,3})$$

(where $\mathbf{t}$ refers to the set of trilinearities) which is an approximation to the first order. $\mathbf{J}$ is the Jacobian of the set of trilinearities. Since $\mathbf{t}(\hat{x}^{1,2,3} = 0$, it folllows that

$$r = -\mathbf{J}(\hat{x}^{1,2,3} - x^{1,2,3})$$

Taking the pseudo-inverse of $\mathbf{J}$, we obtain a least square estimate

$$\hat{x}^{1,2,3} = x^{1,2,3} - \mathbf{J}^{+} r$$

This is the estimated set of matches for this correspondence. We can similarly estimate the matches for 6 such correspondences. The Trifocal tensor is re-estimated using this set of correspondences. This is the Maximum Likelihood Estimate of the Trifocal tesnor.

III. RESULTS

A robust estimate of the trifocal tensor has been obtained. Using this estimate, the RANSAC inlier percentage crosses 85% in 200 iterations on an average. The images below represent the 3 views used to compute the trifocal tensor. The red points are the best set of matches. These are used to compute the trifocal tensor. Another set of 6 matches were randomly selected in 2 images and the corresponding points in the 3rd image were computed using the estimate. These points are represented in blue.

REFERENCES

[1] P H S Torr, A Zisserman *Robust Parameterization and Computation of the Trifocal Tensor*.

[2] Jia Li *The Trifocal Tensor and its applications in augmented reality*.

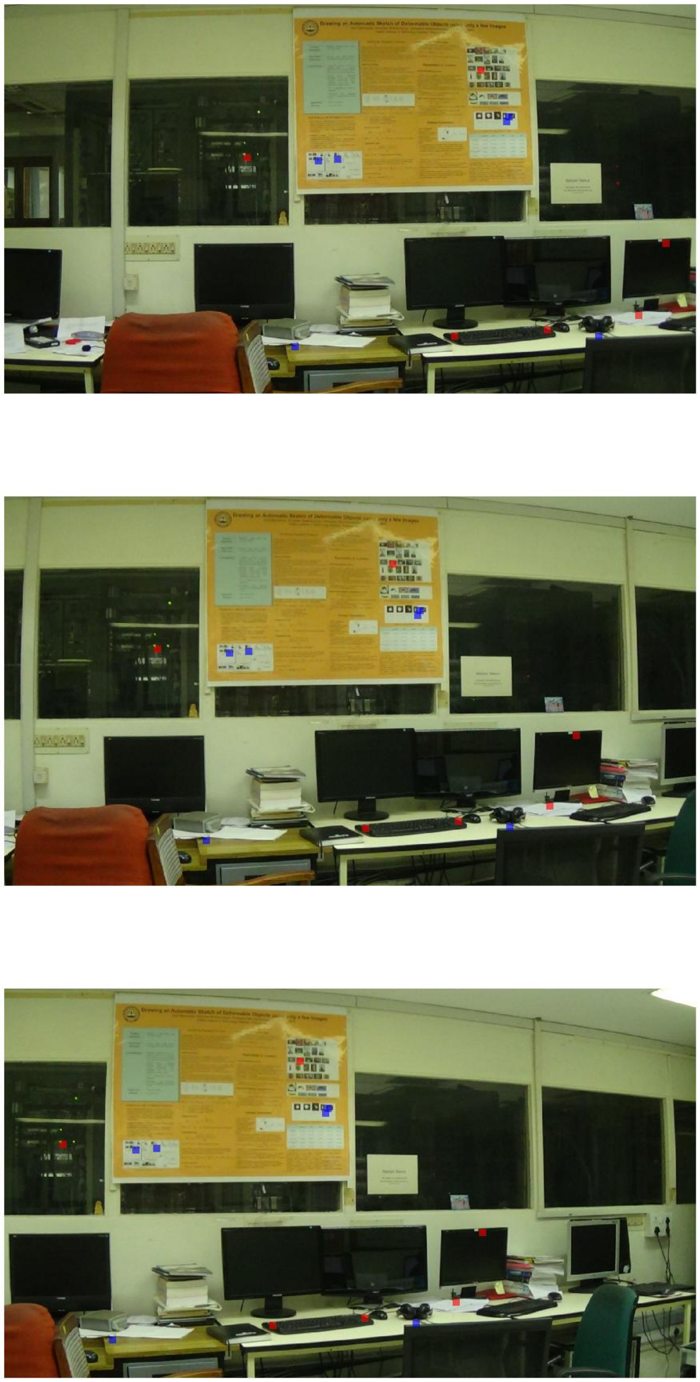

Fig. 1: The three view are shown in the figures. The red points are the best set of matches used to compute the trifocal tensor. The blue points are the points selected in the first 2 views and computed in the third view using the estimated trifocal tensor.