

# CONSISTENT AVERAGING OF MULTI-CAMERA EPIPOLAR GEOMETRIES

A PROJECT REPORT  
SUBMITTED IN PARTIAL FULFILMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
**Master of Engineering**  
IN  
SYSTEM SCIENCE AND AUTOMATION

by

**Jaishanker.K.Pillai**



Electrical Engineering  
Indian Institute of Science  
BANGALORE – 560 012

JUNE 2008

©Jaishanker.K.Pillai

JUNE 2008

All rights reserved

# Acknowledgements

It is my pleasure and privilege to express my deep gratitude to my guide Dr Venu Madhav Govindu for his thorough guidance and insightful suggestions, during the course of the project. I am also thankful to the faculties of the Electrical Department especially Prof K.R.Ramakrishnan and Prof P.S.Sastry for their excellent teaching and help. Last but not least, I would like to thank my 'System Science and Automation' batchmates and the other friends, I made in the past two years, who made my tenure in the campus a truly remarkable one.

# Publications based on this Thesis

1. Venu Madhav Govindu, Jaishanker.K.Pillai, Consistent Averaging of Multi-Camera Epipolar Geometries, submitted to the European Conference Of Computer Vision 2008.

# Abstract

*The geometric relationships, existing between point correspondences across multiple images can be utilized to estimate the three dimensional scene structure and camera motion, and also to transfer points from one view to the others. But due to the noise in the image point locations, these geometric relationships will not be satisfied exactly. The information redundancies existing in image sequences can be utilized to obtain a consistent set of points from the observed noisy point locations. We have developed an efficient iterative scheme which computes a set of image points, satisfying the geometric relationships exactly, without taking recourse to a full three dimensional reconstruction. The algorithm can be performed even when the cameras are not calibrated. Experimental results on simulated and real image datasets illustrate the feasibility of deriving such geometrically consistent relationships.*

# Contents

<b>Acknowledgements</b>	<b>2</b>
<b>Publications based on this Thesis</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>1 Introduction</b>	<b>8</b>
1.1 The Proposed Method . . . . .	8
1.2 Thesis Road Map . . . . .	9
<b>2 Multiview Geometry</b>	<b>10</b>
2.1 Camera Geometry . . . . .	10
2.2 Epipolar Geometry . . . . .	13
2.2.1 Fundamental Matrix . . . . .	13
2.2.2 Essential Matrix . . . . .	15
2.2.3 Computing Relative Motions from Essential Matrix . . . . .	16
<b>3 Averaging Techniques For Global Motion Estimation</b>	<b>17</b>
3.1 Advantages of motion averaging algorithms . . . . .	17
3.2 Motion Consistency . . . . .	18
3.3 Linear Fitting for motion estimation [2] . . . . .	19
3.3.1 Global Rotation Estimation . . . . .	19
3.3.2 Global Translation Estimation . . . . .	20
3.4 Lie-Algebraic Averaging for globally consistent motion estimation [3] . .	21

---

3.4.1	Lie Groups . . . . .	21
3.4.2	Averaging on the Lie group . . . . .	22
3.4.3	Lie Algebraic Motion Averaging . . . . .	22
<b>4</b>	<b>Epipolar Averaging</b>	<b>25</b>
4.1	Consistency of epipolar geometries . . . . .	25
4.2	Averaging to improve Epipolar geometry consistency . . . . .	26
4.2.1	Notations used . . . . .	27
4.2.2	Algorithm in detail . . . . .	27
4.2.3	Advantages of our averaging scheme . . . . .	29
<b>5</b>	<b>Results</b>	<b>31</b>
5.1	Description of the Image Sequences Used . . . . .	31
5.1.1	Dinosaur Sequence . . . . .	31
5.1.2	Valbonne Church Sequence . . . . .	31
5.1.3	IISc Main Building . . . . .	32
5.2	Epipolar Lines Before and After Averaging For Different Image Sequences	32
5.2.1	Valbonne Church Sequence . . . . .	32
5.2.2	IISc Main Building Sequence . . . . .	32
5.3	Distribution Of Geometric Errors Before And After Averaging . . . . .	34
5.4	Point Transfer . . . . .	34
5.5	Effect of averaging on estimated rotations . . . . .	34
5.6	Results On Simulated Data . . . . .	37
5.6.1	Effect On Geometric Error . . . . .	37
5.6.2	Effect on Epipolar Lines . . . . .	37
<b>6</b>	<b>Conclusions And Future Work</b>	<b>39</b>
	<b>References</b>	<b>41</b>

# List of Figures

2.1	Pinhole camera geometry . . . . .	11
2.2	Epipolar geometry . . . . .	14
4.1	Epipolar Lines in third view corresponding to the other two views . . . . .	26
5.1	Results for the Valbonne Church Sequence . . . . .	33
5.2	Results for the IISc Main Building Sequence . . . . .	35
5.3	Distribution Of Geometric Error for the IISc Main Building Sequence . . . . .	35
5.4	Point Transfer Results for the Dinosaur Sequence . . . . .	36
5.5	Estimated rotations for Dinosaur Sequence before and after averaging . . . . .	36
5.6	Variation of Geometric Error with iterations for simulated data . . . . .	38
5.7	Variation of sum of third singular values of the matrices of epipolar lines . . . . .	38

# Chapter 1

## Introduction

With the developments in multi-view geometry [1], the underlying geometric relationships between two and three cameras are well understood. There exists efficient algorithms for their estimation. But these relationships cannot be extended to longer sequences. So for such sequences, techniques like Bundle Adjustment [9] are used, which try to explain the observed data by estimating both the geometry and the structure. Such methods are computationally expensive as the extra variables of structure too have to be estimated in a non-linear optimization routine.

### 1.1 The Proposed Method

Because of the noise in the point locations, the estimated parameters will not satisfy the geometric constraints exactly. In this report, we look at a method to utilize the information redundancy existing in image sequences, to solve for a global geometric representation that reconciles the individual errors in multi-view relationships. Since the proposed method does not use three dimensional structure information, it is computationally efficient. The image points are perturbed iteratively until they satisfy the geometric relationships exactly. The resultant epipolar geometries are consistent allowing us to carry out many geometric operations that are also consistent with the three dimensional geometry. For example, the consistent epipolar geometries allow us to carry

---

out point transfer across a long sequence of images. Also the camera motion and the three dimensional structure can be solved in a straightforward manner, with out going for computationally expensive methods like the Bundle Adjustment [9].

## 1.2 Thesis Road Map

The report is organised as follow. In Chapter (2), we look at the basics of the pin hole camera model and the epipolar geometry, which are essential for the proper understanding of our algorithm. We look at existing averaging techniques for global motion estimation in Chapter (3), where consistent global motion is estimated from noisy relative motions. Our proposed averaging scheme for obtaining consistent multi-camera epipolar geometries is discussed in Chapter (4). In Chapter (5), we evaluate the performance of our algorithm on simulated and real data. Finally Chapter (6) concludes the report and looks at the possible work, that can be carried out in the future.

# Chapter 2

## Multiview Geometry

In this chapter, we will be discussing only the concepts in Multi-view geometry, which are essential for understanding our algorithm clearly. For a detailed description of the same, refer [1].

### 2.1 Camera Geometry

A camera is a mapping between the 3D world and a 2D image. In this section, we will be looking at the pin hole camera model alone. It models the projection of points in space to the image plane. Let the centre of projection be the origin of a Euclidean coordinate system and let the focal length of the camera be  $f$ . The plane  $z = f$  is called the focal plane or the image plane. Then under the pin hole camera model, a point in space with coordinates  $\mathbf{X} = (X, Y, Z)^\top$  gets mapped to the point on the image plane, where the ray joining the centre of projection to the point  $\mathbf{X}$  meets the image plane.

By the principle of similar triangles, we can compute that the point  $(X, Y, Z)$  gets mapped to  $(fX/Z, fY/Z, f)$  on the image plane. This is shown in Fig. 2.1. Since all such points on the image plane have the same  $Z$  coordinate (which only indicates that the point is on the focal plane), it can be ignored. Hence the point on the image plane is  $(fX/Z, fY/Z)$

The centre of projection is called the camera centre or the optic centre. The line

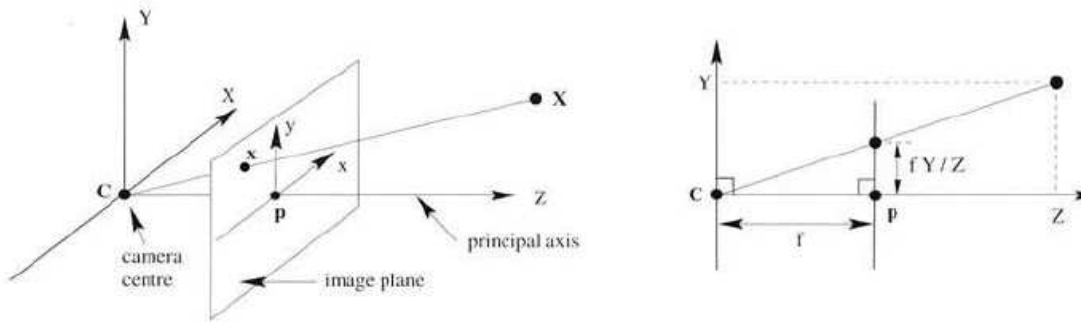


Figure 2.1: Pinhole camera geometry

from the camera centre perpendicular to the focal plane is called the principal ray or the principal axis. The point where the principal ray meets the focal plane is called the principal point. The plane through the camera centre parallel to the focal plane is called the principal plane.

If the world and image coordinates are represented in homogeneous coordinates, the projection of the three dimensional point to the image plane can be represented by a linear mapping.

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX/Z \\ fY/Z \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Let  $\mathbf{X}$  denote the world point represented by the homogeneous 4-vector  $(X, Y, Z, 1)^\top$ . Let  $\mathbf{x}$  be the corresponding image point represented in homogeneous coordinates. Then the above equation can be written as

$$\mathbf{x} = P\mathbf{X}$$

where  $P$  denotes the homogeneous  $3 \times 4$  camera projection matrix.

$$P = \text{diag}(f, f, 1)[I \mid 0]$$

In the above expression, we assumed that the principal point is the origin in the image coordinates. But it need not be the case in general. Let the principal point be at  $p_x$  and  $p_y$  respectively in the image coordinates. Then the above equation becomes

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX + p_x Z \\ fY + p_y Z \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Let

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

Now we can write the above equation in a concise form as

$$x = K[I \mid 0]X$$

The matrix  $K$  is called the camera calibration matrix.

In general, the points in space will be expressed in a different Euclidean coordinate frame called the world coordinate frame. The camera coordinate frame and the world coordinate frame are related by a rotation and a translation. Let  $\tilde{\mathbf{X}}$  be the inhomogeneous 3 vector representing the coordinates of a point in the world coordinate system and  $\tilde{\mathbf{X}}_{cam}$  be the corresponding point in the camera coordinate system. So  $\tilde{\mathbf{X}}_{cam} = R(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})$ , where  $\tilde{\mathbf{C}}$  represents the coordinates of the camera centre in the world coordinate frame and  $R$  is a  $3 \times 3$  matrix representing the orientation of the camera coordinate frame. This can

be written in homogeneous coordinates as

$$\mathbf{X}_{cam} = \begin{bmatrix} R & -R\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \mathbf{X}$$

Combining the two results above, we get

$$\mathbf{x} = KR[I \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

where  $\mathbf{X}$  is the world coordinates. In general the pin hole camera projection matrix  $P$  has 9 degrees of freedom, 3 for the entries in  $K$ , 3 for  $R$  and 3 for  $\tilde{\mathbf{C}}$ . The parameters contained in the camera matrix  $K$  are called the internal parameters and the parameters  $R$  and  $\tilde{\mathbf{C}}$  which relate the camera orientation and position to a world coordinate system are called the extrinsic parameters.

## 2.2 Epipolar Geometry

The epipolar geometry describes the intrinsic projective geometry between two views. It is independent of the scene structure and depends only on the camera's relative pose and internal parameters. This intrinsic geometry is captured by a  $3 \times 3$  matrix called the fundamental matrix.

### 2.2.1 Fundamental Matrix

If a point  $\mathbf{X}$  in the 3D world gets mapped to  $x$  in the first view and  $x'$  in the second, then the points satisfy the relationship

$$x'^{\top} F x = 0$$

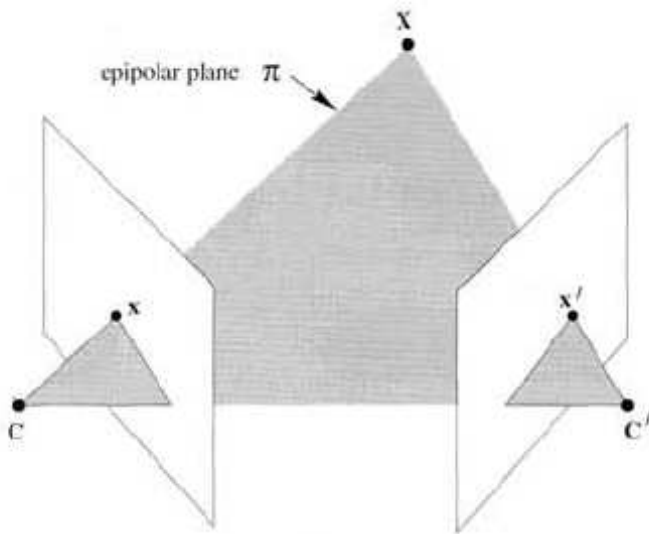


Figure 2.2: Epipolar geometry

where  $F$  is the fundamental matrix. The fundamental matrix can be estimated from the image correspondences alone, the simplest method of estimation being the eight point algorithm [15]. Given the fundamental matrix, the two camera matrices can be estimated upto a projective transformation. Also if the cameras are calibrated, their euclidean motion can be estimated upto a fixed number of ambiguities.

Let a 3D point  $\mathbf{X}$  in the world space gets mapped to a point  $x$  in the first view and a point  $x'$  in the second view. Then as shown in Fig 2.2 below, the points  $X, x, x'$  and the camera centres are coplanar. Let this plane be denoted by  $\pi$ , called the epipolar plane. Clearly the rays back projected from  $x$  and  $x'$  intersect at  $X$  and are coplanar, lying in  $\pi$ .

Suppose that we know a point  $x$  in the first image and the two camera centres. Then the plane  $\pi$  is fully determined as we know three points lying on it, namely the camera centres and the point  $x$ . The ray from the second camera centre, passing through the point  $x'$  lies in the plane  $\pi$ . So the point  $x'$  has to lie on the line of intersection of the image plane and the plane  $\pi$ . This line is the epipolar line corresponding to the point  $x$ . It is the image in the second view of the ray back projected from the point  $x$ . The point of intersection of the line joining the two camera centres with the image plane is called the epipole. Hence the epipole in one view is the image of the camera centre in

the other view.

The Fundamental matrix has rank 2. If the Fundamental matrix from first view to the second is denoted by  $F$ , the one from the second to the first view will be given by  $F^\top$ . For any point  $x$  in the first image, the corresponding epipolar line is  $l' = Fx$ . Similarly,  $l = F^\top x'$  represents the epipolar line corresponding to the point  $x'$  in the second image. The epipoles in the two views can be found out as the null vectors of  $F$  and  $F^\top$  respectively. Let  $e$  and  $e'$  be the epipoles in the first and second image respectively. Then

$$Fe = 0$$

$$F^\top e' = 0$$

### 2.2.2 Essential Matrix

The essential matrix is a specialization of the fundamental matrix for normalized image coordinates (when the camera calibration is known). The essential matrix  $E$  can be expressed in the form

$$E = [t]_\times R$$

where  $R$  and  $t$  are the relative rotation and translation of the camera between the two views. The fundamental matrix and the essential matrix are related by the formula

$$E = K'^\top FK$$

The essential matrix has only five degrees of freedom: both rotation  $R$  and translation  $t$  have three degrees of freedom, but there is an overall scale ambiguity. A  $3 \times 3$  matrix becomes an essential matrix only if its first two singular values are the same and the third one is zero. This information is useful in decomposing it into rotation and translation, explained below.

### 2.2.3 Computing Relative Motions from Essential Matrix

As mentioned in the previous sub section, the essential matrix can be expressed in the form

$$E = [t]_{\times} R$$

Since the first two singular values of the essential matrix have to be same and the third one should be zero, the singular value decomposition of the essential matrix will of the form

$$E = U \text{diag}(1, 1, 0) V$$

where  $U$  and  $V$  are orthogonal matrices. There are two possible factorizations for  $E$  into the product of a skew symmetric matrix and an orthogonal matrix. They are

$$E = SR$$

with

$$S = UZU^{\top}, R = UWV^{\top}$$

or

$$S = UZU^{\top}, R = UW^{\top}V^{\top}$$

$$\text{where } Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The relative translation can be obtained upto scale from the skew symmetric matrix  $S$ . The relative rotation will be one of the two  $R$  matrices mentioned above. The ambiguity in relative rotation can be resolved by computing the depth of any image point and choosing that  $R$  which gives positive depth.

# Chapter 3

## Averaging Techniques For Global Motion Estimation

In this chapter, we will look at existing algorithms which utilize averaging techniques to compute consistent global motion from noisy relative motion. In a sequence of  $N$  images, the global motion of the entire sequence can be parameterized using  $N - 1$  independent motions. However as many as  $\frac{N(N-1)}{2}$  pairwise relative motions can be estimated. Each of these estimated relative motions provide a constraint on the global motions. Thus we have an over determined system of equations which can be averaged efficiently to obtain consistent global motions.

### 3.1 Advantages of motion averaging algorithms

The advantages of using averaging techniques for global motion estimation as opposed to other existing techniques are the following.

1. Other algorithms based on the algebraic elimination of structure variables are applicable only upto four views. So they have limited accuracy. But motion averaging can be applied over all the existing views.

2. The “optimal” method based on Bundle Adjustment minimizes the distance between the observed feature points and their estimated reprojections. [9]. But it is computationally expensive as it involves a non linear optimization over both motion and structure variables. It also requires a good initialization. Since motion averaging does not use structure information, it is significantly faster.
3. Other methods for multi frame motion estimation are [11] and [12], where structure and motion are solved simultaneously using rank based factorization techniques. But they require the points to be tracked across all the images. Motion averaging uses all the available information and does not suffer from this drawback.

## 3.2 Motion Consistency

Let the first image be chosen as the reference. So the motions of all the other images with respect to this reference is to be estimated. Let the motion between the frame  $i$  and the reference be denoted by  $M_i$  and the relative motion between the frames  $i$  and  $j$  be denoted by  $M_{ij}$ . Therefore

$$M_{ij} = M_j M_i^{-1}$$

This relationship captures the notion of consistency, i.e. the composition of any sequence of relative motions from  $i$  to  $j$  should be identical to  $M_{ij}$ . But due to the presence of noise in the observed relative motions  $\hat{M}_{ij}$ , the consistency constraint will not be satisfied exactly. Thus for observed noisy relative motions

$$\hat{M}_{ij} \neq M_j M_i^{-1}$$

However each of these relationships can be written as a constraint on the global motion to be estimated. Thus we have an over determined system of equations

$$\hat{M}_{ij} M_i - M_j = 0, \forall i \neq j$$

where the variables  $\{M_i\}$  are the unknowns to be estimated. Intuitively, we want to estimate  $\{M_i\}$  that are most consistent with the measurements  $\{\hat{M}_{ij}\}$  in a least-square sense, so that the errors in individual estimated of  $\hat{M}_{ij}$  will get averaged out. The over determined system mentioned above should not be directly solved as the  $M_i$  thus obtained will be general matrices in  $R^9$  and may not represent valid motion matrices. Methods for obtaining consistent global motions are described in the sections below.

### 3.3 Linear Fitting for motion estimation [2]

This methods first averages the relative rotations after representing it in the quaternion form. The global translations are then estimated from the relative rotations and the relative translations.

#### 3.3.1 Global Rotation Estimation

The consistency relationship for rotation is given by

$$R_{ij} = R_j R_i^{-1}$$

But because of the noise in the relative rotations, these relationships will not be satisfied exactly. Each of these relationships form a constraint on the global rotation, which is to be estimated. Let  $\hat{R}_{ij}$  denote the noisy relative rotations. So the linear system of equations can be written as

$$\hat{R}_{ij} R_i - R_j = 0$$

Since the rotation matrices are constrained to  $SO(3)$  where as any linear method will result in a solution in  $R^9$ , we rewrite the linear solutions using the quaternion representation of rotation ( $\mathbf{q} = \{q_0, q_1, q_2, q_3\}$ ). The linear relationship can be written as

$$Q_{ij} q^i = q^j$$

where

$$Q = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix}$$

In the above system of equations,  $Q_{ij}$  terms are known where as  $q^i$  are the unknowns to be estimated. This system of equations can be solved linearly to estimate the global rotation matrices.

### 3.3.2 Global Translation Estimation

In the case of translation estimation, the consistency equations are of the form

$$T_{ij} = T_j - R_{ij}T_i$$

However, from the fundamental matrices, the relative translations can be computed only upto a scale factor. So we have equations of the form

$$t_{ij} = \lambda_{ij}(T_j - R_{ij}T_i)$$

where  $\lambda_{ij}$ 's are the unknown scale factors. We can utilize the crossproduct constraints, described by

$$t_{ij} \times (T_j - R_{ij}T_i) = 0$$

The algorithm to estimate global translation is given by

1. Initialise scalar weights  $\lambda^0 = 1$
2. At step n, solve

$$\left[ t_{ij} \lambda_{ij}^{n-1} \right]_{\times} (T_j - R_{ij}T_i) = 0$$

3. Update  $\lambda^n = \frac{1}{\|T_j - R_{ij}T_i\|}$

4. Repeat till convergence

### 3.4 Lie-Algebraic Averaging for globally consistent motion estimation [3]

In this section, we look at estimating global motion by utilizing the underlying Lie group structure of the motion representation.

#### 3.4.1 Lie Groups

As a prerequisite background for this section, basics of Lie groups are discussed below. For a detailed description on the use of Group Theory in Computer Vision, see [16]. A group  $G$  is a set whose elements satisfy the relationships

$$X.(Y.Z) = (X.Y).Z(\textit{associativity})$$

$$\exists E \ni X.E = E.X = X(\textit{identity})$$

$$\exists X^{-1} \ni X.X^{-1} = X^{-1}.X = E(\textit{inverse})$$

A Lie group is a group for which the operations  $X \times Y \mapsto XY$  and  $X \mapsto X^{-1}$  are differential mappings. Intuitively, Lie groups can be locally viewed as topologically equivalent to the vector space  $R^n$ . The local neighborhood of any group element can be adequately described by its tangent space. The elements of this vector space form a Lie algebra.

All finite dimensional Lie groups have matrix representations. The Lie algebra and the associated Lie group are related by the exponential mapping. Three dimensional rotation  $R$  belongs to the Special Orthogonal group  $SO(3)$  and the three dimensional Euclidean motion consisting of rotations followed by translation belongs to the Special Euclidean group  $SE(3)$ .

### 3.4.2 Averaging on the Lie group

Let  $\{X_1, X_2, \dots, X_N\}$  be the elements of a Lie Group  $G$ . Let  $d(.,.)$  be the intrinsic distance (i.e. the Riemannian distance) between the points on the manifold. Then the intrinsic average can be defined as

$$\mu = \arg \min_{X \in G} \sum_{k=1}^N d^2(X_k, X)$$

In case of Lie groups, the intrinsic average can be computed efficiently. For matrix groups, the Riemannian distance is defined by the matrix logarithm operation, i.e. for matrix group elements  $X$  and  $Y$ , we have

$$d(X, Y) = \|\log(YX^{-1})\|$$

This distance can be approximated as

$$d(X, Y) \approx \|\log(Y) - \log(X)\| = \|y - x\|$$

where  $x$  and  $y$  are the logarithms of matrices  $X$  and  $Y$  respectively. Thus the Riemannian distance between elements of a Lie group can be approximated by the Euclidean distance in its Lie algebra. For a set of group elements  $\{X_1, X_2, \dots, X_N\}$ , the minimizer of  $\sum_{k=1}^N d^2(X_k, X)$  can be estimated from the sample average of the Lie algebra  $\{x_1, x_2, \dots, x_N\}$ . Given the estimate of the average  $\mu = \exp(\frac{1}{N} \sum_{i=1}^N x_i)$ , we can remap the samples by left multiplying by the inverse of  $\mu$ . This operation can be repeated till the estimate converges to a local minima.

### 3.4.3 Lie Algebraic Motion Averaging

The estimation of global motion by Lie Algebraic averaging is similar in spirit to the averaging algorithm mentioned in the previous section. Let the motion between the frame  $i$  and the reference be denoted by  $M_i$  and the relative motion between the frames

$i$  and  $j$  be denoted by  $M_{ij}$ . Therefore

$$M_{ij} = M_j M_i^{-1}$$

Let  $m = \log(M)$ . So by applying the first order approximation to the Riemannian distance, the constraint  $M_{ij} = M_j M_i^{-1}$  becomes

$$m_{ij} = m_j - m_i$$

The matrix  $m$  can be described using three parameters for  $m \in so(3)$  and six parameters for  $m \in se(3)$ . If the parameters are arranged in the form of a column vector  $v$ , the same relationship holds, i.e.  $v_{ij} = v_j - v_i$ . If we stack all the column vectors  $v_i$  to form one big vector  $V$

$$V = \begin{bmatrix} v_1, v_2, \dots, v_N \end{bmatrix}$$

Then

$$v_{ij} = \underbrace{\begin{bmatrix} \dots, -I, \dots, -I, \dots \end{bmatrix}}_{=D_{ij}} V$$

where  $I$  denotes an identity matrix of dimensions  $N_{dim} \times N_{dim}$  where  $N_{dim}$  can be 3 or 6.  $D_{ij}$  is a matrix of size  $N_{dim} \times (N_{dim} \times N - 1)$ . Let us stack all the relative motion vectors  $v_{ij}$  into one big vector  $V_{ij}$ . Similarly, let all the matrices  $D_{ij}$  be stacked into one big matrix  $D$ . So we have

$$M_j M_i^{-1} = M_{ij}$$

$$\Rightarrow DV = V_{ij}$$

$$\Rightarrow V = D^\dagger V_{ij}$$

where  $D^\dagger$  represents the pseudo-inverse operation,  $D^\dagger = (D^\top D)^{-1} D^\top$ .

An iterative scheme can be developed where we update the current estimate from the observed motion values to improve the estimate, as indicated below.

$$\text{Input : } \left\{ M_{ij1}, M_{ij2}, \dots, M_{ijn} \right\} \text{ (n relative motions)}$$

Output :  $\{ M_2, M_3, \dots, M_N \}$  ( $N$  global motions)

Set global motions to an initial guess.

Do

1.  $\Delta M_{ij} = M_j^{-1} M_{ij} M_i$
2.  $\Delta m_{ij} = \log(\Delta M_{ij})$
3.  $\Delta v_{ij} = \text{vec}(m_{ij})$
4.  $\Delta V = D^\dagger \Delta V_{ij}$
5.  $M_k = M_k \exp(\Delta m_k), \forall k \in [2, N]$
6. Repeat till  $\|\Delta V\| < \epsilon$

Emperically the method was found to converge in 5-8 iterations.

# Chapter 4

## Epipolar Averaging

### 4.1 Consistency of epipolar geometries

Due to the noise in the point correspondences, the epipolar geometries estimated in a sequence will not all be consistent with each other. For the epipolar geometries in a sequence to be consistent, two criteria have to be satisfied as explained below.

Consider the triplet of images shown in Fig 4.1. Let the triplet of correspondences across the images be denoted by  $(p_1, p_2, p_3)$ . Let the epipolar geometries between the images be denoted by  $F_{12}$ ,  $F_{23}$  and  $F_{13}$ . If the points are consistent, the algebraic distance of the points from the corresponding epipolar lines will be zero. In other words, the algebraic error  $p_j^\top F_{ij} p_i$  will be zero for all pairs  $(i, j)$  of images. Also the point transferred from first and second image should coincide with the corresponding point in the third image. The epipolar line in the third image due to a point  $p_1$  in the first image is given by  $l_{13} = F_{13} p_1$ . Similarly the epipolar line in the third image due to the corresponding point  $p_2$  in the second image is given by  $l_{23} = F_{23} p_2$ . Since the point  $p_3$  has to lie on both these epipolar lines, in the ideal scenario  $p_3$  should be equal to  $l_{13} \times l_{23}$ . If the points are consistent, this point transfer from the first and second images to the third image will be exact.

But in actual case, due to the noise in the point locations, the individual epipolar geometries will not be consistent with each other.

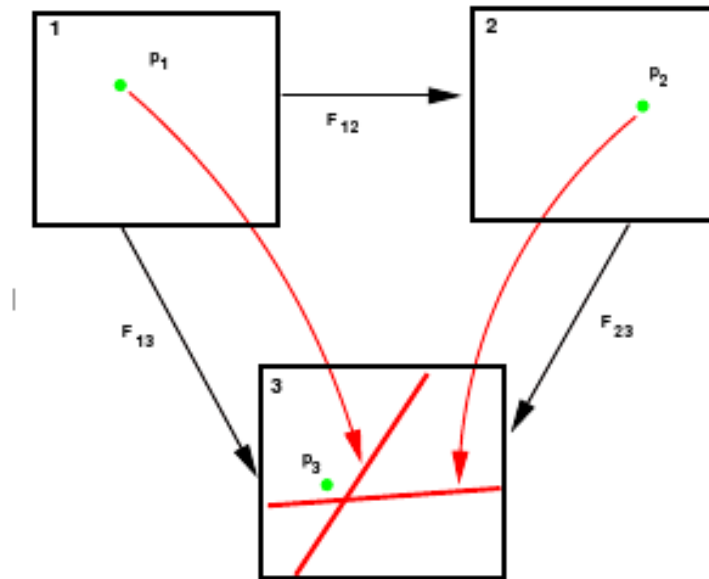


Figure 4.1: Epipolar Lines in third view corresponding to the other two views

## 4.2 Averaging to improve Epipolar geometry consistency

In case of the motion averaging discussed in Chapter 3, the underlying geometric structure of the matrix group can be used to average them and derive consistent global motion estimates. But such an averaging scheme is not available for fundamental matrices. The matrix representation of the relative motion can be written in terms of the global motion as  $M_{ij} = M_{kj}M_{ki}^{-1}$ . No such relationship can be established between the equivalent epipolar geometries, represented by the fundamental matrices. This problem can be solved by utilising the epipolar constraints that algebraically relate the point correspondences to the fundamental matrices. Since all the epipolar geometries have to consistent with each other, the possible positions for the point correspondences are constrained. We have developed an iterative algorithm which pertubes the correspondence points location such that it satisfies the epipolar geometry exactly.

### 4.2.1 Notations used

Let there be  $N$  images and a total of  $T$  matched points. The  $k^{th}$  point in the  $i^{th}$  image is denoted as  $p_i^k$ , where  $i \in \{1, \dots, N\}$  denote the image index and  $k \in \{1, \dots, T\}$  denote the point index. The index  $k$  identifies the image projections corresponding to the same 3D point. The set of images where the point  $k$  is visible is denoted by  $I^k$ . For the algorithm to be applied to a point, it should be visible in atleast two images. So  $|I^k| \geq 2$ . The set of points visible in the image  $i$  is denoted by  $P_i = \{p_i^k | k \in S_i\}$ , where  $S_i$  denotes the set of indices of points present in the  $i^{th}$  image. The set of point matches between the  $i^{th}$  and  $j^{th}$  image is denoted by  $S_{ij} = S_i \cap S_j$ . Since fundamental matrix computation requires atleast seven points,  $S_{ij} \geq 7$  for an edge to exist in the view graph from  $i$  to  $j$ . The edges of the view graph are given by  $E = \{(i, j) | |S_{ij}| \geq 7\}$ . The set of all epipolar geometries are denoted by  $F = \{F_{ij} | (i, j) \in E\}$  and the full set of all points in all images are represented as  $P = \{p_i^k | \forall i = 1, \dots, N; k \in S_i\}$

### 4.2.2 Algorithm in detail

The point coordinates are assumed to be corrupted by additive iid white Gaussian noise. Let  $p_i^{0k}$  denote the noiseless actual point and  $p_i^k$  denote the observed noisy points. Let  $n_i^k$  be the additive white gaussian noise affecting the  $k^{th}$  point in the  $i^{th}$  image.

$$p_i^k = p_i^{0k} + n_i^k$$

$$n_i^k \sim N(0, \sigma^2)$$

Let the vectors  $P, P^0$  and  $N$  be formed by stacking together the noisy image points, the true image points and the noise values respectively. So

$$P = P^0 + N$$

The maximum likelihood estimate of the true points can be obtained as

$$\arg \max_{P^0} Prob(P|P^0)$$

subject to

$$\|p_j^{0k\top} F_{ij} p_i^{0k}\| = 0, \forall (i, j) \in E, \forall k \in S_{ij}$$

Since the noise is assumed to be gaussian, this can be written as

$$\arg \min_{P^0} \|P - P^0\|^2$$

subject to

$$\|p_j^{0k\top} F_{ij} p_i^{0k}\| = 0, \forall (i, j) \in E, \forall k \in S_{ij}$$

In other words, given the noisy point correspondences  $P$ , we have to find a set of true points which are consistent with the epipolar geometries and also are closest to the noisy points in a least square sense. Finding a globally optimal solution for the above problem is hard. So we derived a greedy iterative approach.

In each iterations, our algorithm has two main steps. They are

### Updating the Epipolar Geometries

In this step, we compute the fundamental matrices between all possible pairs of images, using the image points available in the current iteration. We have used the MAPSAC based estimation method of Dr Phill Torr [7] for computing the fundamental matrices.

### Updating the Image Points

In this step, we compute the new image point locations from the current image points and the updated fundamental matrices. For the  $k^{th}$  point in the image pair  $(i, j)$ , the epipolar constraint requires the point  $p_j^k$  to lie on the line  $F_{ij} p_i^k$ . This relationship has to be satisfied by all points across all image pairs.

There are two steps in updating a point  $p_j^k$ . At first, we estimate all possible epipolar

lines generated by the matches for the  $k^{\text{th}}$  image point in the other images, i.e.  $F_{ij}p_i^k$  where  $i \in \{I^k - j\}$ . In case of a single epipolar line,  $p_j^k$  can be made consistent by projecting it onto the epipolar line. But in case of multiple images, we can compute a point projection for each available epipolar line. Due to noise in the image points, these projections will not be identical. So we average these point projections to get the new estimate of the point  $p_j^k$ . This procedure is carried out for all the image points to get the updated point locations.

We continue updating the fundamental matrices and the image points until convergence. At convergence, the epipolar geometries will be consistent with one another, i.e. the points will lie on the corresponding epipolar lines and further updation will not change the image points. The residual algebraic error will be zero and all the epipolar lines will intersect at a single point corresponding to the respective image point.

### 4.2.3 Advantages of our averaging scheme

1. The motion averaging methods mentioned in Chapter 3 need relative motion as input, which require the camera to be accurately calibrated. But since we are averaging directly the epipolar geometries, we donot require the intrinsic parameters of the cameras.
2. After the points become consistent, all the epipolar lines in an image corresponding to a particular  $3D$  point will intersect at a single point. So even if the actual image point location is not known due to occlusion, we can compute the same as the point of intersection of the corresponding epipolar lines. So if a point is visible in two views, we can always transfer it to a third view, given that the epipolar geometries between the images are known. So by repeated application of this method, we can transfer a point to any other view where it is originally not present.
3. Since the estimated points are now consistent, all geometric relationships between three dimensional points, camera matrices and image points will be exact. So the computation of camera geometries and three dimensional points are straight

forward and does not involve an estimation procedure. So we can solve the camera geometry from a minimal set of relative motions by cascading the relative motions along a spanning tree in the view graph. Also the three dimensional structure can be obtained by triangulation. Thus the camera geometry and structure can be computed significantly faster after performing the averaging algorithm than using computationally expensive methods like the Bundle Adjustment [9].

# Chapter 5

## Results

The proposed averaging algorithm was tested on three different data sets, in addition to simulated data. They are the Dinosaur sequence, Valbonne Church sequence and the IISc Main Building sequence. The various results obtained are illustrated in this chapter.

### 5.1 Description of the Image Sequences Used

#### 5.1.1 Dinosaur Sequence

This data set is available online at

<http://www.robots.ox.ac.uk/vgg/data/data-mview.html>

The data set consists of 36 images of a toy dinosaur, rotated on a turn table. The angle of rotation is roughly 10 degrees.

#### 5.1.2 Valbonne Church Sequence

This data set can be accessed online through the link

<http://www.robots.ox.ac.uk/vgg/data/data-mview.html>

The data set consists of 15 images of the Valbonne Church taken from different viewpoints. The focal length of the camera is not changed while taking the images.

### 5.1.3 IISc Main Building

This data set consists of 9 images of the Indian Institute Of Science (IISc) Main Building. The sequence is captured using three different focal lengths.

## 5.2 Epipolar Lines Before and After Averaging For Different Image Sequences

### 5.2.1 Valbonne Church Sequence

Fig 5.1 shows the epipolar lines before and after averaging for key points in the Valbonne Church sequence. The first column shows the original image, with the red rectangle indicating the area which is zoomed in, in the second and third column. Each of these rectangles indicate the region around a keypoint. In the second and third column, the key point is indicated as a yellow dot and the epipolar lines due to other images are shown as red lines. As can be clearly observed from the examples, before averaging is performed, the epipolar lines donot intersect at a point. But after the epipolar averaging algorithm is performed, all the new epipolar lines intersect at a single point, which corresponds to the new image point. Hence the image sequence satisfies the epipolar geometry exactly.

### 5.2.2 IISc Main Building Sequence

Fig 5.2 shows the epipolar lines before and after averaing for the IISc Main Building sequence. The interpretation of the figures is same as that for the Valbonne Church Sequence results shown in Fig 5.1. Observe that our algorithm works well and gives consistent epipolar geometries even when the focal lengths of the cameras are different.

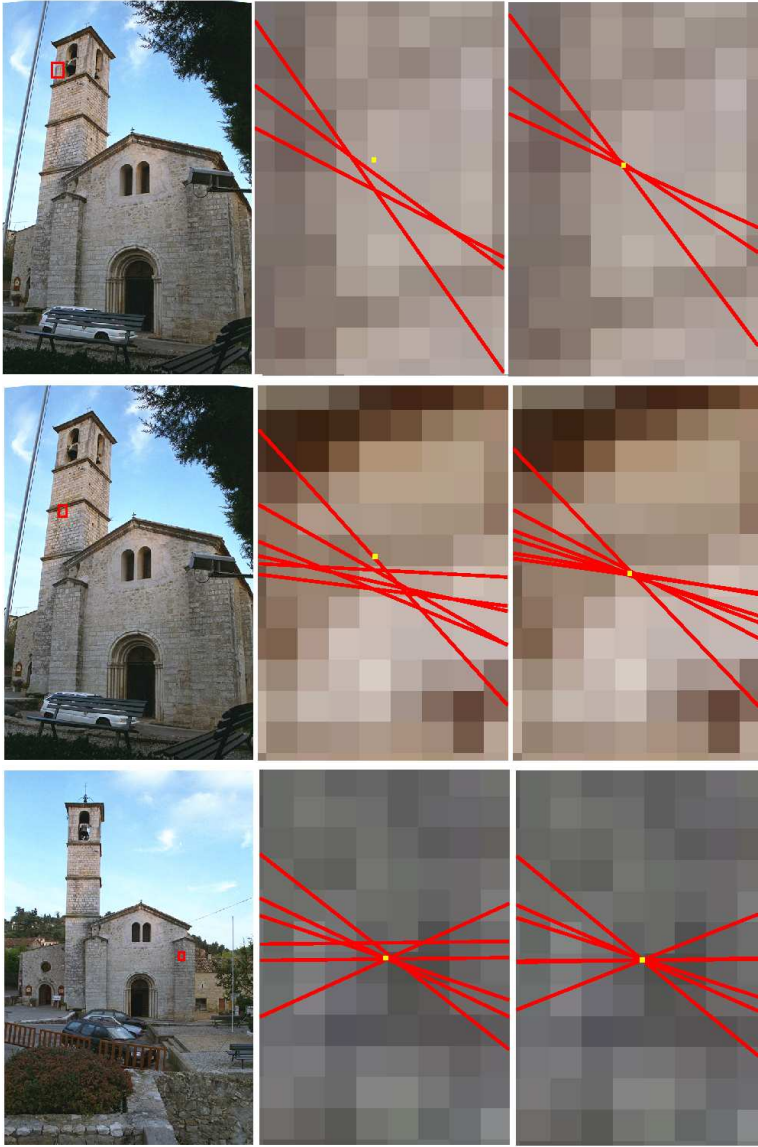


Figure 5.1: Results for the Valbonne Church Sequence

### 5.3 Distribution Of Geometric Errors Before And After Averaging

Fig 5.3 illustrates the reduction of geometric error due to the epipolar averaging. The plot shows the distribution of the geometric error (the perpendicular distance of the image points from the corresponding epipolar lines) for the IISc Main Building before averaging in red and after averaging in blue. The horizontal axis denotes the logarithm of the error values. Note the clear leftward shift on the logarithmic scale which indicates a dramatic reduction in geometric errors.

### 5.4 Point Transfer

In this section, we show the effect of our algorithm on transfer of points from one view to others. In Fig 5.4, image points from the first nine views of the Dinosaur sequence are transferred to the first view. The red dots show the points transferred to the first view. The blue curve shows the dinosaur silhouette. The corresponding image is shown in the left corner. The green dots show the points originally visible in the first view. If the point transfer is performed, without performing our algorithm, the transferred points can be far away from the dinosaur silhouette due to the noise in the individual point locations. This can be clearly observed in the left image below. But once the algorithm is performed, the epipolar geometries become consistent and the transferred points come close to the dinosaur silhouette as required.

### 5.5 Effect of averaging on estimated rotations

In this section, we look at the effect of our averaging scheme on rotation estimation. Fig 5.5 shows the estimated global camera rotation for the dinosaur sequence. The actual relative rotation between adjacent images for the sequence is 10 degrees and hence the actual global camera rotation angles are linear, indicated by the blue line.



Figure 5.2: Results for the IISc Main Building Sequence

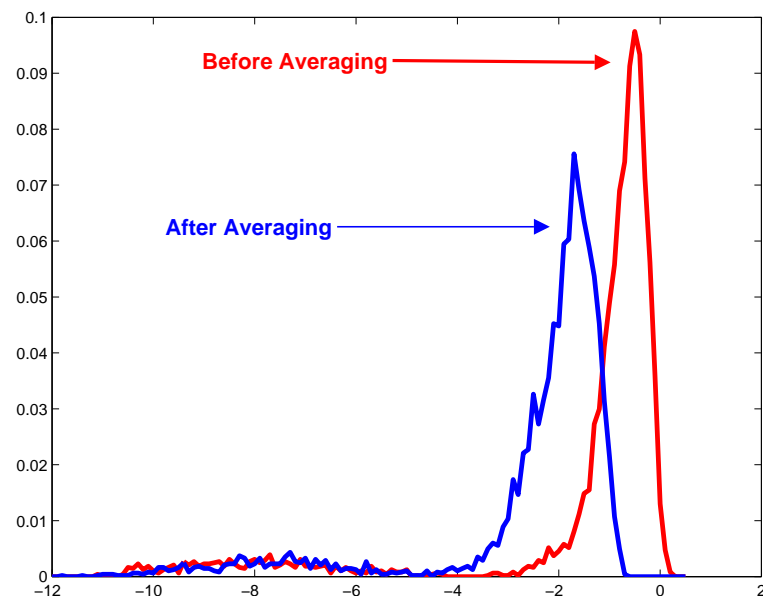


Figure 5.3: Distribution Of Geometric Error for the IISc Main Building Sequence

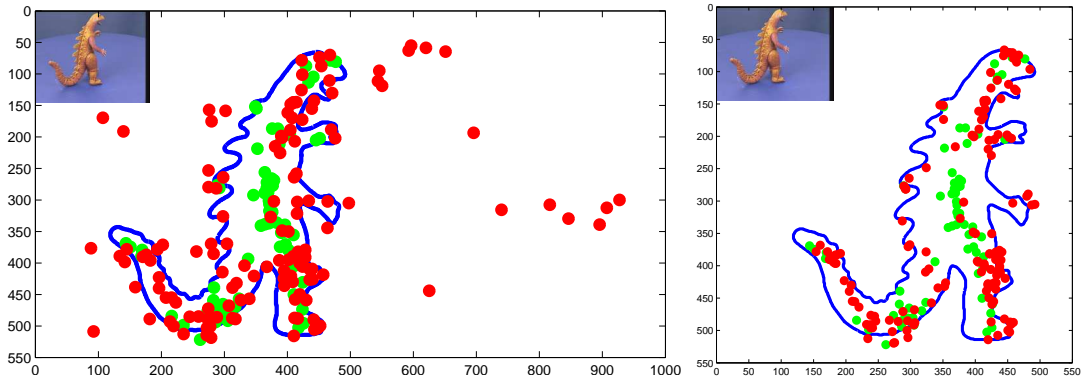


Figure 5.4: Point Transfer Results for the Dinosaur Sequence

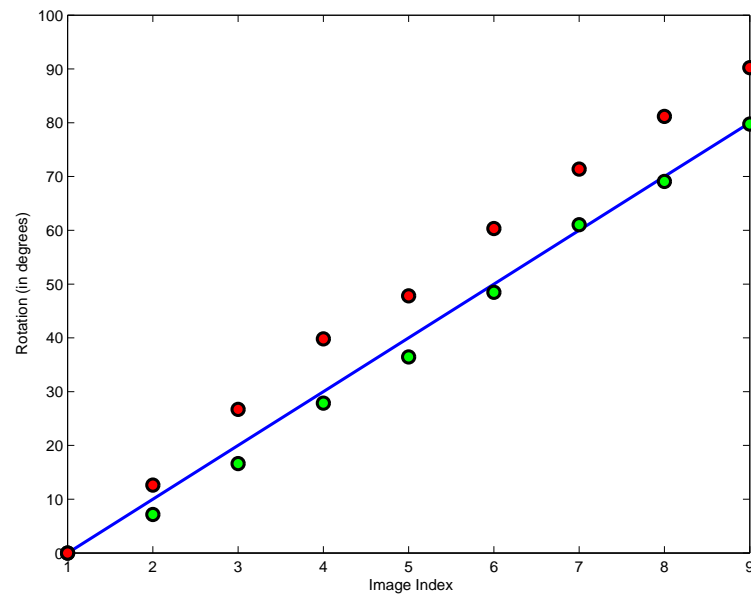


Figure 5.5: Estimated rotations for Dinosaur Sequence before and after averaging

The red circles indicate the estimated global motions without the epipolar averaging and the green circles represent the corresponding estimated global rotations after epipolar averaging. One can easily notice that the green dots are closer to the blue line than the red dots. Hence our averaging improves the estimated global rotations significantly.

## 5.6 Results On Simulated Data

In this section, we illustrate some results for the simulated data cases. Four images were simulated, each containing 40 points, corresponding to 40 three dimensional points and random motion of the camera.

### 5.6.1 Effect On Geometric Error

Fig 5.6 shows the L2 norm of the geometric error for the points as a function of the iteration number. Clearly, successive iterations of the algorithm reduces the geometric error and hence improves the consistency of the points.

### 5.6.2 Effect on Epipolar Lines

For each image point, we obtained a matrix by stacking all the corresponding epipolar lines due to the other images. Then the third singular value of the matrix was computed. Ideally, all the lines should intersect at a single point. So the matrix should have a non zero null vector and hence the third singular value should be zero. But due to the noise in the image point locations, the third singular value of the matrix described above will not be zero. On performing our averaging algorithm, the third singular value is found to decrease steadily with iterations as illustrated in Fig 5.7. On convergence of our algorithm, the third singular value is found to be close to zero which means that all the epipolar lines corresponding to an image point intersects at a single point.

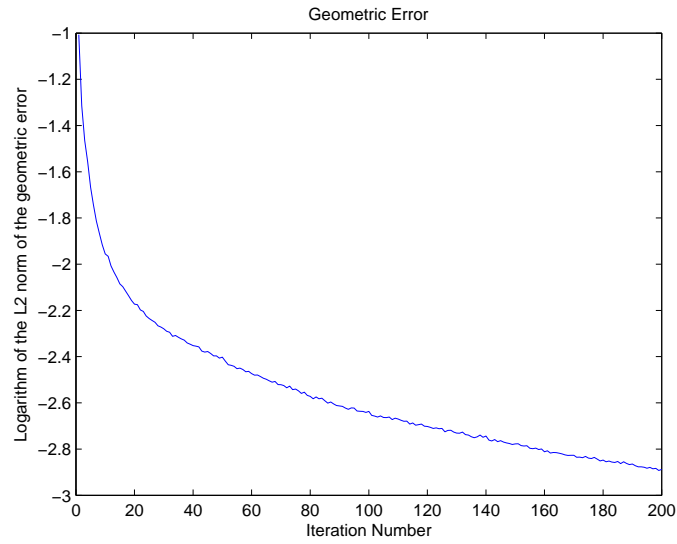


Figure 5.6: Variation of Geometric Error with iterations for simulated data

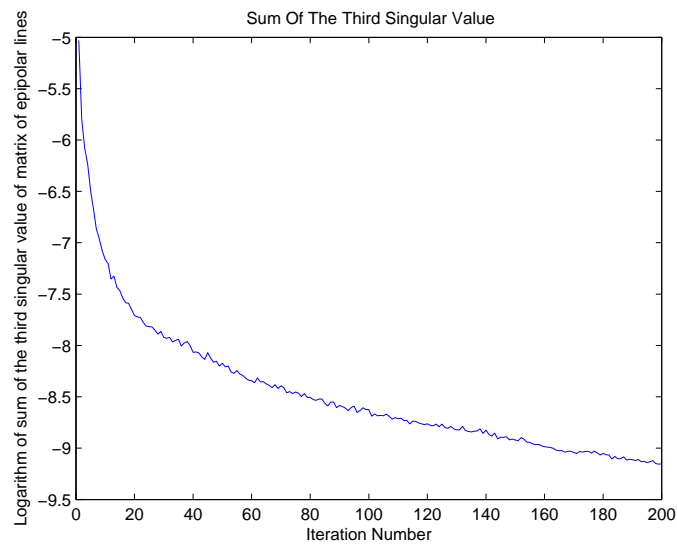


Figure 5.7: Variation of sum of third singular values of the matrices of epipolar lines

# Chapter 6

## Conclusions And Future Work

Summarizing our work, we have developed an averaging scheme for making the epipolar geometries consistent. Our method is extremely efficient as it does not require any three structure computation. Unlike the existing motion averaging schemes which require the camera calibration to be known exactly, our method does not require information about the internal parameters of the camera. Once consistent geometries are obtained by running our algorithm, camera motion and structure can be computed directly without using computationally expensive methods like the Bundle Adjustment. Also points from an image can be transferred to any other view, which can be used to identify object parts occluded in a particular view.

But certain issues remain unexplored due to time restrictions.

1. In our method, we are using Fundamental matrix for averaging. In a calibrated case, one should be able to do a similar algorithm using Essential matrices. This should improve the rotation and translations, estimated from Essential matrices significantly. But in order to avoid initial drift in the points, we require to explore techniques for estimating essential matrices accurately from noisy point locations. We have not looked at this aspect in this project.
2. Once our averaging algorithm converges and image points consistent with the

epipolar geometries are obtained, there exists a set of cameras and three dimensional points for which reprojection error is zero. So we should be able to solve a set of camera matrices from a spanning tree of fundamental matrices. Given a fundamental matrix, we can always obtain a pair of camera projection matrix which satisfies it [1]. But for multiple fundamental matrices, obtaining consistent camera projection matrices involve solving for a homography matrix, which is not addressed in our work.

3. It is possible to estimate calibration parameters from the fundamental matrix [10]. Since our algorithm changes the Fundamental matrix, the estimated focal length will be altered by our algorithm. One can study the effects of our algorithm on the estimated focal length and the changes it makes to the energy landscape of the cost function defined in [10].

# References

- [1] Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)
- [2] Govindu, V.M.: Combining two-view constraints for motion estimation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2001) 218–225
- [3] Govindu, V.M.: Lie-algebraic averaging for globally consistent motion estimation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Volume 1. (2004) 684–691
- [4] Govindu, V.M.: Robustness in Motion Averaging. In: ACCV 2006, Lecture Notes In Computer Science, pp 457-466, Springer-Verag 2006.
- [5] Anandan, P., Avidan, S.: Integrating Local Affine into Global Projective Images in the Joint Image Space. In: Proceedings of European Conference on Computer Vision. Volume 1. (2000) 907–921.
- [6] McLachlan, G., Krishnan, T.: The EM Algorithm and Extensions. John Wiley & Sons (1997).
- [7] Torr, P., Bayesian model estimation and selection for epipolar geometry and generic manifold fitting International Journal of Computer Vision **50:1** (2002) 35–61.
- [8] Avidan, S., Shashua, A., Novel View Synthesis by Cascading Trilinear Tensors IEEE Transactions on Visualization and Computer Graphics **4:4**, (1998) 293–306.

- 
- [9] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, Bundle Adjustment-A Modern Synthesis Lecture Notes in Computer Science, vol. 1883, pp. 298-375, 2000.
- [10] P.R.D.S. Mendonca and R. Cipolla. A simple technique for self-calibration. In CVPR99, pp. 500 - 505, 1999.
- [11] C. Tomasi and T. Kanade, Shape and Motion From Image Streams Under Orthography: A Factorization Method, International Journal Of Computer Vision, vol. 9, no. 2, pp. 137-154, 1992.
- [12] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In European Conference Computer Vision, pages 709-20, Cambridge, U.K., 1996. Springer-Verlag.
- [13] Lowe, D.G., Distinctive image features from scale-invariant keypoints International Journal of Computer Vision **60:2** (2004), 91–110.
- [14] Levi, N., Werman, M.: The viewing graph. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Volume 2. (2003) 599–606.
- [15] Hartley, R.: In defence of the 8-point algorithm. In: Proceedings of the 5th International Conference on Computer Vision. (1995) 1064–1070.
- [16] Kanatani, K., Group-Theoretical Methods in Image Understanding, Springer-Verag, 1990.