RELATIVE CONSISTENCY OF PROJECTIVE RECONSTRUCTIONS OBTAINED FROM AN IMAGE PAIR

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURÇAK OTLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF COMPUTER ENGINEERING

June 2011

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Tayfur Öztürk
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Ayşe Kiper
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Volkan Atalay
Supervisor

Examining Committee Members

Prof. Dr. Kemal Leblebicioğlu

Assoc. Prof. Dr. Volkan Atalay

Assoc. Prof. Dr. İsmail Hakkı Toroslu

Assoc. Prof. Dr. Göktürk Üçoluk

Y. Müh. Adem Yaşar Mülayim

ABSTRACT

RELATIVE CONSISTENCY OF PROJECTIVE RECONSTRUCTIONS OBTAINED FROM AN IMAGE PAIR

M.Sc., Department of Computer Engineering

Supervisor: Assoc. Prof. Dr. Volkan Atalay

June 2011, 49 pages

This thesis works on the projective reconstruction of an object or a scene from an image pair and their relative consistency. In this study 3D points are estimated from image pairs using projective geometry and epipolar geometry. Two measures are presented for verification of projective reconstructions with each other. These measures are based on the equality of ratio between the x, y and z coordinates of two 3D reconstructed points obtained from the same corresponding points. This information is used for measuring the relative consistency of projective reconstructions obtained from the same image pair.

Keywords: Projective reconstruction, epipolar geometry, relative consistency

ÖZ

GÖRÜNTÜ ÇİFTİNDEN ELDE EDİLEN PROJEKTİF GERİÇATIMLARIN BAĞIL DOĞRULUĞU

Otlu, Burçak

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi: Assoc. Prof. Dr. Volkan Atalay

Haziran 2011, 49 sayfa

Bu tez bir nesne veya manzara görüntü çiftinin projektif geriçatımı ve bunların bağıl doğruluğu üzerinedir. Bu çalışmada projektif geometri ve epipolar geometri kullanılarak bir görüntü çiftinden üç boyutlu noktalar bulunmuştur. Projektif geriçatımları birbiriyle doğrulamak için iki ölçüm yöntemi sunulmuştur. Bu ölçümler aynı karşılık gelen noktalardan elde edilen iki tane üç boyutlu geriçatılmış noktanın x, y ve z koordinatlarının oranının eşitliğine dayanmaktadır. Bu bilgi aynı görüntü çiftinden elde edilen projektif geriçatımların bağıl doğruluklarını ölçmek için kullanılmıştır.

Anahtar Kelimeler: Projektif geriçatımı, epipolar geometri, bağıl doğruluk

To my family

ACKNOWLEDGMENTS

I would like to thank to my supervisor Assoc. Prof. Dr. Volkan Atalay and to my friends for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

There has been a considerable progress on 3D reconstructions of an object or a scene from its image sequence. It is a well-known fact that when the intrinsic parameters, extrinsic parameters or metric information about the object is unknown, only projective reconstruction is possible from an image pair [4]. Therefore, in general to obtain affine, metric or euclidean reconstruction additional information about the intrinsic parameters, extrinsic parameters or object is needed. Constraints on the parameters such as constant or known parameters and constriants on the number of images are used to upgrade the reconstructions [18, 19, 22, 23]. In this thesis, since camera calibration, which is often erroneous, is not known, thus projective reconstruction from an image pair is achieved. Camera is undergoing an unknown motion around an object or a scene. Given an image pair, interest points are computed by Canny edge detector [9] and then these points are matched by the constraint on their normalized-cross-correlation scores and constraint of disparity [10]. More information about matching can be found in [20, 21]. From the uncalibrated image pair, equivalent epipolar geometry is obtained, which is represented by the fundamental matrix, F. There are various techniques for estimating the fundamental matrix, F. Commonly used robust methods are M-Estimators, LMedS (Least Median of Squares) and RANSAC (RANdom SAmple Consensus) [17]. These are detailed in [16, 10, 11].

However, in this study, because of its easy implementation, Normalized 8-point algorithm is used [14], which is linear and SVD (Singular Value Decomposition) is used as a main tool. Camera matrices are computed by weak calibration and using these matrices, projective reconstructions are computed [11, 12, 13]. Therefore two or more 3D data sets are obtained from the same image pair. It has been observed that 3D reconstructed points from the same corresponding points, the ratio between the x, y and z coordinates of the 3D points are the same. This fact is used for measuring the relative consistency of two projective reconstructions. Moreover, in computer vision literature, it seems that no such work describing these measures are presented. In chapter 2 projective geometry is introduced. Camera models and epipolar geometry are reviewed in Chapter 3. Steps of the projective reconstruction, implementation details and theory behind new measures with the experimental results are explained in Chapter 4. Finally in Chapter 5 conclusion and significant issues arising from this thesis are mentioned.

CHAPTER 2

Projective Geometry

Perspective projection is the idealized mathematical model of real cameras. Therefore necessary terminology and notation are described in this chapter.

2.1 The Projective Geometry of 2-space

The projective plane is the projective space P^2 which may be also considered as vector space. There is no difference between the representations of points and lines in projective plane.

2.1.1 Points and Lines

A *point* x=$[x, y]$ in R^2 is represented as a 3-vector (3 element vector) x=$[x, y, 1]$ by adding 1 as the last element. Therefore homogeneous representation of any point x=$(x_1, x_2, x_3)^\top$ represents the point $(x_1/x_3, x_2/x_3)^\top$ in R^2. A *line* in the plane with the equation ax+by+c=0 is represented by a 3-vector $[a, b, c]^\top$. A point x = $[x, y]^\top$ lies on the line l = $[a, b, c]^\top$ if and only if ax+by+c=0. This may be written as the inner product of vectors $[x\, y\, 1][a\, b\, c]^\top$; as a result, the point x lies on the line l if and only if $x^\top l = l^\top x = x \cdot l = 0$. *Intersection of two lines* l and l' is the point x=l$\times l'$, where $\times$ represents vector or cross product. *Line through two points* x and x' is l=x$\times x'$.

2.1.2 Ideal points and the line at infiniy, l_∞

Consider two parallel lines ax+by+c=0 and ax+by+$c^{'}$=0, which are represented by vectors $l = [a, b, c]^\top$ and $l^{'} = [a, b, c^{'}]^\top$. *The intersection of parallel lines* is $l \times l^{'} = [b, -a, 0]^\top$. Inhomogeneous representation of this point, $[b/0, -a/0]^\top$, which doesn't make sense. In general, points with homogeneous coordinates $[x, y, 0]^\top$ do not correspond to any finite point in R^2. Therefore, we can say that parallel lines meet at infinity [2].

Homogeneous vectors $x = [x_1, x_2, x_3]^\top$ such that $x_3 \neq 0$ correspond to finite points in R^2. If we add points with $x_3 = 0$ to R^2, the resulting space is the set of all homogeneous 3-vectors, namely the projective space P^2. The points with the last coordinate $x_3 = 0$ are known as *ideal points* , or points at infinity. The set of all ideal points lies on a single line, *the line at infinity*, $l_\infty = [0, 0, 1]^\top$. This can be verified by $[0, 0, 1][x_1, x_2, 0]^\top = 0$.

A model of the projective plane P^2 may be thought of a set of rays in R^3 through the origin. Points and lines may be obtained by intersecting this set of rays by the plane $x_3 = 1$ (see Figure 2.1). As shown in the Figure 2.1, point **x** is

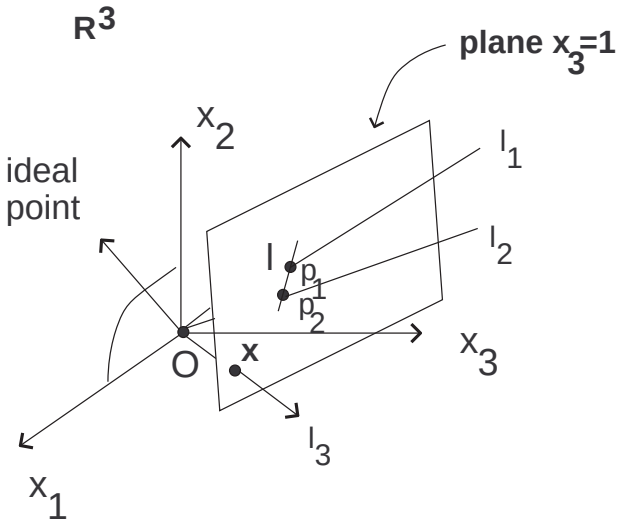

Figure 2.1: A model of projective plane.

obtained by intersection of the line l_3 with the plane $x_3 = 1$ and line **l** passes through points p_1 and p_2 which are obtained by intersection of lines l_1 and l_2 with the plane $x_3 = 1$. The rays representing ideal points and the plane repre-

senting l_∞ are parallel to the plane $x_3 = 1$.

2.1.3 Duality Principle

Formally, there is no difference between points and lines in P^2, therefore any theorem or statement can be rewritten by interchanging the roles of points and lines, which is called duality principle [3].

2.1.4 Conics

A conic is a curve described by a second-degree equation in the plane. In Euclidean geometry conics are of three main types: hyperbola, ellipse and parabola. The equation of a conic in inhomogeneous coordinates is

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

In homogeneous coordinates by replacing $x \longmapsto x_1/x_3$, $y \longmapsto x_2/x_3$ gives

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0$$

or in the matrix form

$$x^\top Cx = 0$$

where the conic coefficient matrix C is given by

$$C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix} \tag{2.1}$$

2.1.5 Projectivity

A projectivity such as the one shown in Figure 2.2 is an invertible mapping H from P^2 to itself such that three points x_1, x_2 and x_3 lie on the same line if and only if after the mapping H, $x_1^{'} = H(x_1)$, $x_2^{'} = H(x_2)$ and $x_3^{'} = H(x_3)$ also lie on

the same line.

A projectivity is also called a *collineation*, a *projective transformation* or a *homography*: these are all synonymous terms.

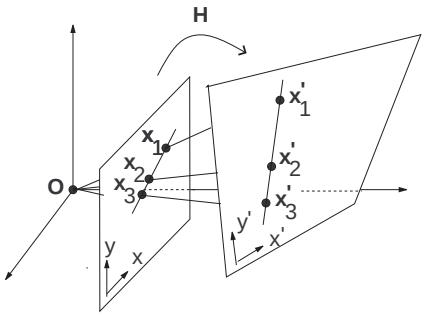

Figure 2.2: Projectivity H maps the points x_1, x_2 and x_3 to the points $x_1^{'}$, $x_2^{'}$ and $x_3^{'}$.

2.2 Projective Geometry of 3-space

In the former section, projective plane, P^2 and its basic concepts are described. These concepts have analogies in projective space, P^3. For example, the duality between points and lines in P^2 is between points and planes in P^3.

In 3-space, homogeneous representation of point X=$[x_1,x_2,x_3,x_4]^\top$ with $x_4 \neq 0$ is

$$[x_1/x_4, x_2/x_4, x_3/x_4, 1]^\top$$

which is $[x_1/x_4, x_2/x_4, x_3/x_4]^\top$ in R^3. Ideal points are points at infinity with $x_4 = 0$. Ideal points lie on a plane at infinity, π_∞ which is the analogy of line at infinity, l_∞ in P^2. [2].

2.2.1 Projectivity (Projective Transformation)

The group of invertible n × n matrices with real elements is the general linear group of n dimensions is *GL(n)*. To obtain the projective linear group, the matrices related by a scalar multiplier are identified, giving *PL(n)* (this is a quotient group of *GL(n)*)[2]. PL(3), Projective transformations of the plane is considered.

Projectivity of 3-space which is an invertible 4x4 matrix that maps points in P^3 to points in P^3 forms projective linear transformations namely *projective*, *affine*, *similarity* and *euclidean*. Every latter is the subgroup of the former. Therefore projective transformation is the most general one whereas euclidean transformation is the most specialized one (see Figure 2.3). Note that transformation and group will be used as synonyms.

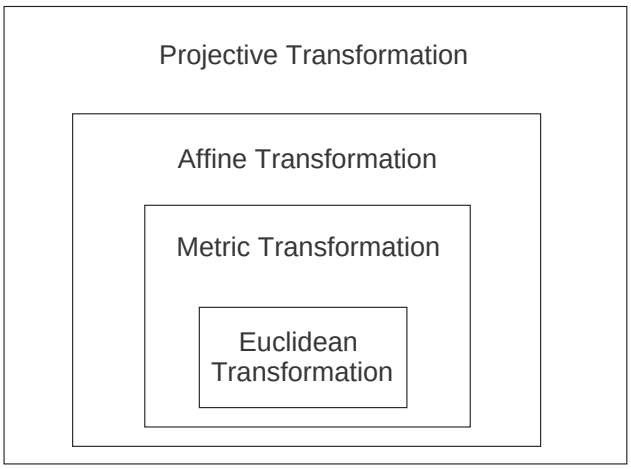

Figure 2.3: Transformations and subgroups.

2.2.2 Euclidean Transformation

Euclidean transformation, which is known as displacement is composed of a translation and rotation. It can be written as

$$x^{'} = H_E x = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} x$$

where R is 3x3 rotation matrix, t is a translation vector, $0 = [0, 0, 0]^T$ is a null 3-vector. Euclidean transformation has 6 dof (degrees of freedom)[1], three for rotation and three for translation. Length (the distance between two points), area, volume and angle (the angle between two lines) are invariants[2] for Euclidean transformation [2].

2.2.3 Metric (Similarity) Transformation

Metric transformation (simply similarity) is an euclidean transformation composed with an uniform scaling. It can be written as

$$x^{'} = H_S x = \begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix} x$$

where the scalar s represents the uniform scaling. It preserves the "shape". Similarity transformation has 7 dof, one more dof than that of Euclidean transformation. Since angle between two lines doesn't change by rotation, translation and uniform scaling, in metric transformation, angle is invariant. In addition, ratio of lengths, ratio of areas and ratio of volumes are invariants for similarity transformation.

2.2.4 Affine Transformation

Affine transformation has the following block form

$$x^{'} = H_A x = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix} x$$

where A is the composition of rotations and non-uniform scalings. Affine transformation has 12 dof, nine for matrix A and three for translation t. Since affine transformation includes non-uniform scalings, ratio of lengths and angle between lines can not be preserved by affine transformation.

[1] Among the n elements of any matrix, number of independent ratios gives the degrees of freedom (dof).

[2] Properties that do not change when the specified transformation is applied.

Table 2.1: Transformations of groups.

	euclidean	similarity	affine	projective
rotation	√	√	√	√
translation	√	√	√	√
uniform scaling		√	√	√
nonuniform scaling			√	√
shear			√	√
perspective projections				√
composition of projections				√

Parallel lines are the lines which intersect at a point at infinity. After affine transformation, parallel lines intersect at another point at infinity so they are parallel after affinity. Parallel lines, length ratios of parallel lines and ratio of areas are invariants for affinity.

2.2.5 Projective Transformation

Projective transformation is the most generalized one and has the following block form

$$x^{'} = H_P x = \begin{bmatrix} A & t \\ v^T & v \end{bmatrix} x$$

where v is a general 3-vector. The 15 dof of a projective transformation are accounted for as seven for a similarity (three for rotation, three for translation, one for scaling), five for affine scalings, and three for the projective part of the transformation. Incidence, ratio of ratios (cross ratio) of lengths on a line are the fundamental invariants of projective transformation.

Allowed transformations in each group are shown in Table 2.1. As it can be seen from Table 2.2, after each transformation some knowledge of the original scene is lost. In other words, each transformation causes its own distortion. This is shown in Figure 2.4. Transformations are applied to a cube and after each transformation the cube is shown. Projective transformation loses the most of knowledge related with the original scene, whereas euclidean transformation

Table 2.2: Invariants of groups.

	euclidean	similarity	affine	projective
length	√			
angle	√	√		
ratio of lengths	√	√		
parallelism	√	√	√	
incidence	√	√	√	√
collinearity	√	√	√	√
tangency	√	√	√	√
cross ratio	√	√	√	√

keeps the most of it. It has been the aim of the researchers to get as much as knowledge as they can.

2.3 The Projective Geometry of n-space

This section is the generalization of the projective geometry and it is only a small introduction. A point in projective n-space, P^n, is given by a vector of (n+1) coordinates x=$[x_1 \dots x_{n+1}]^\top$. All of these coordinates shouldn't be zero at the same time. If this is the case, they are called homogeneous coordinates. A line is also represented by a vector.

Two points represented by (n+1)-vectors x and y are equal if and only if there exists a nonzero scalar λ such that $x_i = \lambda y_i$, for every i $(1 \leq i \leq n+1)$. This will be indicated by x $\sim$ y [1].

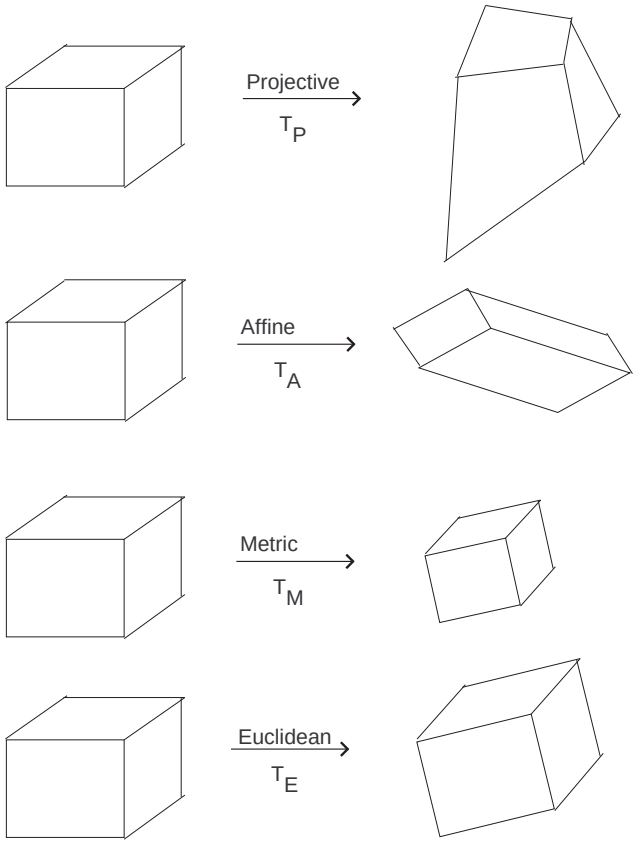

Figure 2.4: Transformations of geometries.

CHAPTER 3

Camera Models and Epipolar Geometry

In this chapter, first the process of image formation is described. After the geometry of image formation, the camera models and their parameters are given.

3.1 Image Formation

Light rays reflected from the object point enters the camera through the camera's aperture and hits the image plane. In this way, object point converges to an image point. If the camera's aperture is not small enough as in Figure 3.1, all

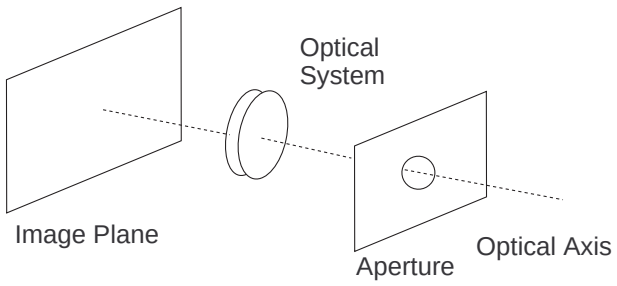

Figure 3.1: A model of pinhole.

light rays coming from the same object point may not hit the image plane at the same point. And this will cause distorted images. To prevent this, camera's aperture is reduced to a point, which is called a pinhole [5]. Furthermore it is assumed that a thin lens in the optical system will focus all the rays from the

object point to the same image point. This is the basic of perspective or pinhole camera.

3.2 Geometric Image Formation

Basic properties of thin lenses help us to understand the geometric aspects of image formation. Basic properties of a thin lens are as follows:

- Any ray entering the lens parallel to the axis on one side goes to through the focus on the other side.
- Any ray entering the lens from the focus on one side emerges parallel to the axis on the other side.

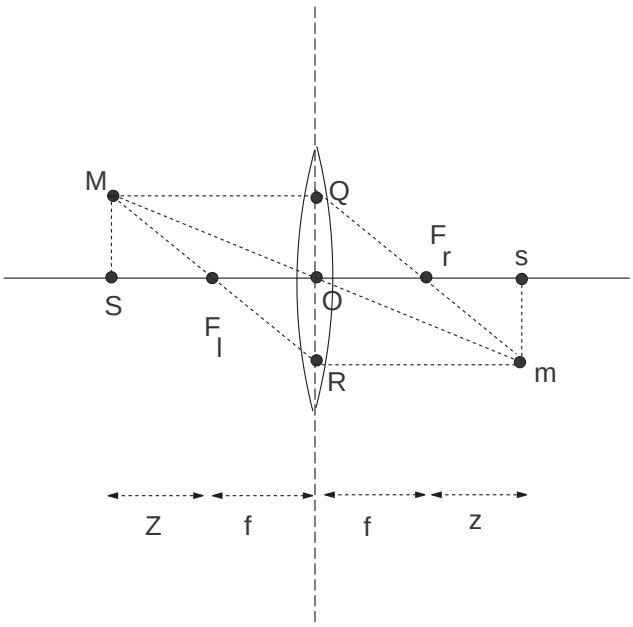

Figure 3.2: Image formation of a thin lens. F_l and F_r are the focuses of the lens, whereas f is the focal length.

As it is shown in Figure 3.2, the point $M = [X, Y, Z]^\top$ in world frame is mapped to the point $m = [x, y]^\top$ in camera frame. From the similar triangles $< MSF_l >$ and $< ROF_l >$ we obtain the equation below:

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \tag{3.1}$$

3.2.1 Pinhole Camera Model

A basic pinhole camera is shown in Figure 3.3. Line from the camera center, O, perpendicular to the image plane is called the principal axis, and the point where the principal axis meets the image plane is called the principal point. Image plane is at $Z = f$. Therefore, the third component of image point m is always equal to the focal length. Hence $m = [x, y, f]^\top$ can be used instead of $m = [x, y]^\top$.

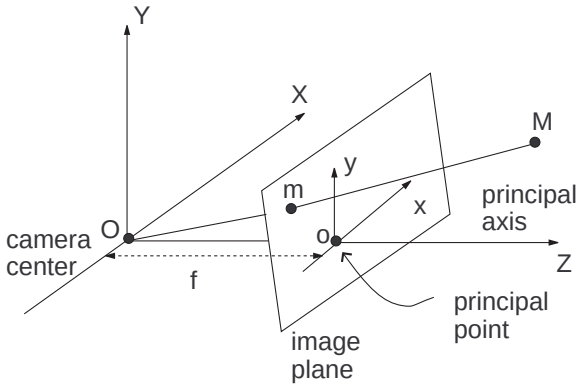

Figure 3.3: Pinhole camera model.

3.3 Camera Parameters

Using a camera, a world point M is mapped to the image pixel m. During this mapping process, two transformations are performed between coordinate systems. The first transformation is from world coordinate system to camera coordinate system. Camera coordinate system can be located with respect to a known world coordinate system. Then the transformation from world frame ($\vec{X_w}$,$\vec{Y_w}$,$\vec{Z_w}$) to camera frame ($\vec{X_c}$,$\vec{Y_c}$,$\vec{Z_c}$) can be specified by the translation vector T and the 3×3 rotation matrix R as shown in Figure 3.4. T and R forms the extrinsic parameters of the camera. The second transformation is from the camera coordinate system to the image coordinate system. Equation 3.1 that is obtained from triangle similarity can be written in linear projection equation using the homogeneous representation of points.

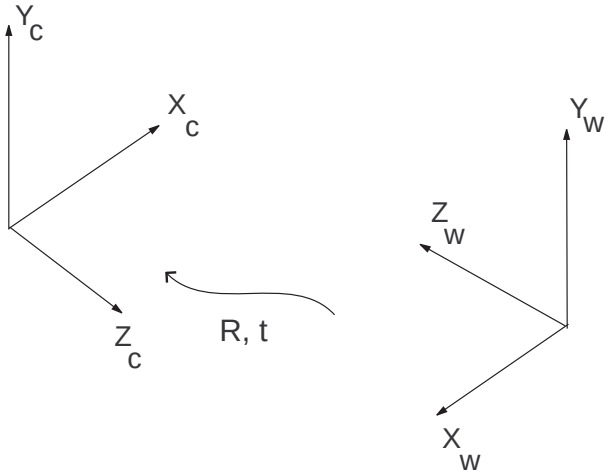

Figure 3.4: Transformation from world frame to camera frame.

$$\begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3.2}$$

This is the simplest case. As it is shown in Figure 3.5, the origin of the image plane ($\vec{u_o}$,$\vec{v_o}$) may not be at the principal point ($\vec{x_c}$,$\vec{y_c}$) in practice. Therefore,

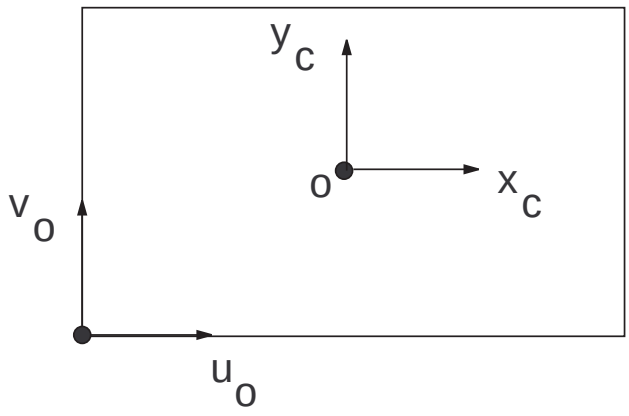

Figure 3.5: Transformation from camera frame to image frame.

in general the mapping is

$$[\mathrm{X},\ \mathrm{Y},\ \mathrm{Z}]^\top \mapsto [f\mathrm{X}/\mathrm{Z}{+}u_o,\ f\mathrm{X}/\mathrm{Z}{+}v_o]^\top$$

and can be written in matrix multiplication form by

$$\begin{bmatrix} fX + Zu_o \\ fY + Zv_o \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & u_o & 0 \\ 0 & f & v_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3.3}$$

Assume that matrix A transforms the camera frame to image frame, which is given as follows

$$A = \begin{bmatrix} f & 0 & u_o \\ 0 & f & v_o \\ 0 & 0 & 1 \end{bmatrix} \tag{3.4}$$

In the pinhole camera model, we assume that image plane is composed of square pixels. However in CCD (Charge Coupled Device) camera model there is the possibility of having non-square pixels. If the number of pixels per unit distance in image coordinates is m_x and m_y in the x and y directions then $\alpha_x = fm_x$ and $\alpha_y = fm_y$ where α_x and α_y are scale factors, that represent the focal length in pixel dimensions in x and y direction respectively. Similarly, principal point in pixel dimensions is $u_o = m_x p_x$ and $v_o = m_y p_y$. Therefore, the general form of calibration matrix of a CCD camera is

$$A = \begin{bmatrix} \alpha_x & 0 & u_o \\ 0 & \alpha_y & v_o \\ 0 & 0 & 1 \end{bmatrix} \tag{3.5}$$

In general, we can consider one more parameter s. It is the skew parameter which is zero for most of the cases. If $s \neq 0$, then it means that x and y axes are not perpendicular and it may arise when the image is being imaged again. In this case,

$$A = \begin{bmatrix} \alpha_x & s & u_o \\ 0 & \alpha_y & v_o \\ 0 & 0 & 1 \end{bmatrix} \tag{3.6}$$

All of these, α_x, α_y, s, u_o, v_o, are the intrinsic parameters of the camera. The problem of estimating intrinsic and extrinsic parameters is called camera calibration.

Finally perspective projection matrix, P can be introduced, P maps the world point M to image pixel m.

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3.7}$$

where s is a scale factor and P is the 3×4 perspective projective matrix, which can be decomposed as

$$P = A \begin{bmatrix} R & T \end{bmatrix} \tag{3.8}$$

where A is the 3×3 matrix that stands for intrinsic camera parameters and transforms the camera frame to image frame. $\begin{bmatrix} R & T \end{bmatrix}$ stands for external camera parameters and transforms the world frame to camera frame.

3.4 Epipolar Geometry

Several relationships exist between two, three or more images that are acquired simultaneously or sequentially of a scene. Relationship of two images is defined by epipolar geometry, which is very important for reconstruction from images.

3.4.1 Epipolar Geometry

Two images of a scene are related by the epipolar geometry. As shown in Figure 3.6, consider two cameras with the camera centers c_1 and c_2, respectively. **M** is the scene point in world coordinate system and its projections on the image planes are m and $m^{'}$, respectively. Hence, m and $m^{'}$ are the corresponding points. Given the point m in the first image, its corresponding point $m^{'}$ in the second image is restricted to line on the line, l_m, which is called the epipolar line

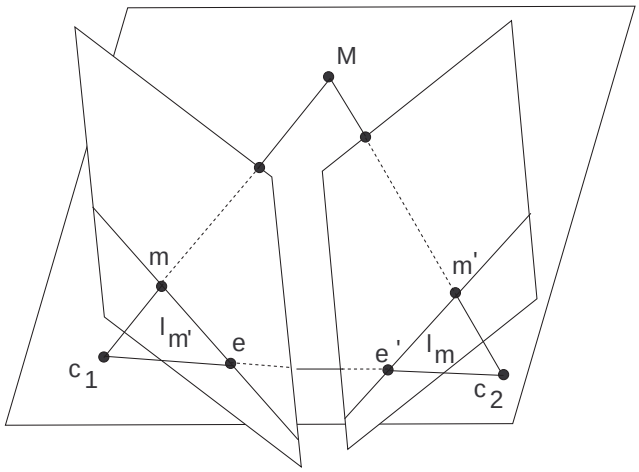

Figure 3.6: Epipolar Geometry.

of m and vice versa.

$$\begin{aligned} & m^{'T} l_m = 0 \\ \text{since} \quad & m^{'T} F m = 0 \\ \text{therefore} \quad & l_m = F m \end{aligned} \tag{3.9}$$

It can be seen that l_m is obtained from the intersection of the second image plane with the plane that passes through M (scene point), c_1 (first camera center) and c_2 (second camera center). Plane that is defined from M, c_1 and c_2 is called epipolar plane. In the same way, epipolar line of $m^{'}$, $l_{m'}$ is obtained from the intersection of the first image plane with the epipolar plane. All of the epipolar lines in each image, intersect at a point, which is called epipole (see Figure 3.7) and each epipole is the projection of camera center in the opposite image.

3.4.2 Fundamental Matrix

Epipolar geometry is represented by a 3×3 matrix called fundamental matrix F. Let's assume that the displacement from the first camera to the second camera is (R, t), where R is a 3×3 rotation matrix and t is a 3×1 translation matrix. Suppose that the world coordinate system is set to the first camera coordinate system. Hence, the scene point M can be expressed in the first camera frame. Then the projection equations of the scene point M to the corresponding points

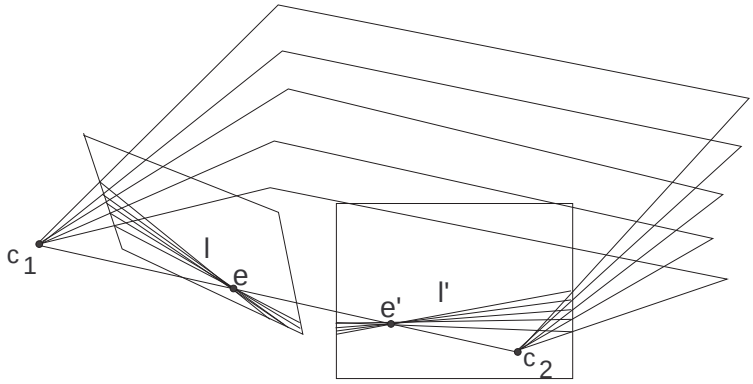

Figure 3.7: Epipolar lines intersect at epipoles e and $e^{'}$.

m and $m^{'}$ can be written as

$$s_1 m = A_1 \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} M \\ 1 \end{bmatrix}$$
$$s_2 m^{'} = A_2 \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} M \\ 1 \end{bmatrix} \tag{3.10}$$

where A_1 and A_2 are the intrinsic parameters of the first and second camera, respectively and s_1 and s_2 are the nonzero scalars. After eliminating M, s_1 and s_2, Equation 3.10 can be written as

$$m^{'T} A_2^{-T} t \times R A_1^{-1} m = 0 \tag{3.11}$$

where $\times$ denotes the cross product of two 3-vectors. Equation 3.11 gives very important constraints between the two images of the same scene. If we denote $F = A_2^{-T} t \times R A_1^{-1}$, then F is called the fundamental matrix. Then, given a set of correspondences $m \leftrightarrow m^{'}$ in two images, the fundamental matrix F satisfies

$$m^{'\top} F m = 0 \tag{3.12}$$

for all $m \leftrightarrow m^{'}$. It gives the all relation between two uncalibrated images and this is the only information that can be obtained.

3.4.3 Computing the Fundamental Matrix

Each point correspondence generates one linear equation in the entries of F. Specifically, for a pair of matching points m=$[x, y, 1]^T$ and $m^{'}$=$[x^{'}, y^{'}, 1]^T$, the equation 3.12 becomes,

$$x^{'}xf_{11} + x^{'}yf_{12} + x^{'}f_{13} + y^{'}xf_{21} + y^{'}yf_{22} + y^{'}f_{23} + xf_{31} + yf_{32} + f_{33} = 0 \quad (3.13)$$

which is inner product of

$$(x^{'}x, x^{'}y, x^{'}, y^{'}x, y^{'}y, y^{'}, x, y, 1)F = 0$$

for n pair matches $m \leftrightarrow m^{'}$

$$\begin{bmatrix} x_1^{'}x_1 & x_1^{'}y_1 & x_1^{'} & y_1^{'}x_1 & y_1^{'}y_1 & y_1^{'} & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^{'}x_n & x_n^{'}y_n & x_n^{'} & y_n^{'}x_n & y_n^{'}y_n & y_n^{'} & x_n & y_n & 1 \end{bmatrix} F = 0 \quad (3.14)$$

which can be written as

$$AF = 0 \quad (3.15)$$

F can be solved up to scale by solving these linear set of equations.

Smallest singular value of A gives F, which is the last column of V, where $SVD(A) = UDV^T$. The matrix F found may not be singular, may have rank 3, then F is replaced with $F^{'}$, the closest singular matrix to F with rank 2. Fundamental matrix is a singular matrix with rank 2 (see Appendix C), this is called singular constraint.

There are several methods for estimating fundamental matrix F [11, 14, 10, 16]. 8-point algorithm given in Algorithm 1, is the simplest method for computing the fundamental matrix.

Algorithm 1 8-point algorithm.

Given at least 8 corresponding points $m_i \leftrightarrow m_i^{'}$

Form the matrix A as defined in Equation 3.14

Determine the F from the singular vector corresponding to the smallest singular value of A.

Replace the F with $F^{'}$ such that $\det(F^{'}) = 0$

If the input data is normalized, this algorithm performs better, which is called normalized 8-point algorithm that is given in algorithm 2.

Algorithm 2 Normalized 8-point algorithm.

Given at least 8 corresponding points $m_i \leftrightarrow m_i^{'}$

Normalize the corresponding points $\hat{m}_i = Tm_i$ and $\hat{m}_i^{'} = T^{'}m_i^{'}$ such that their centroid is at the origin and their RMS (root mean square) is $\sqrt{2}$

Form the matrix A as defined in Equation 3.14

Determine the F from the singular vector corresponding to the smallest singular value of A.

Replace the F with $F^{'}$ such that $\det(F^{'}) = 0$

Denormalize the $F^{'}$ by $\hat{F}^{'} = T^{'T}F^{'}T$, where $\hat{F}^{'}$ is the fundamental matrix of the original data $m_i \leftrightarrow m_i^{'}$.

There are other linear and nonlinear methods for computing F. However they are not mentioned in this thesis.

3.4.4 Computation of the camera matrices

Camera geometry is obtained from fundamental matrix F, which is very important for reconstruction. Given the corresponding points $m_i \leftrightarrow m_i^{'}$, camera matrices P and $P^{'}$ corresponding to a fundamental matrix F may be chosen as

$$\begin{aligned} P_1 &= \left[\begin{array}{c|c} I & 0 \end{array}\right] \\ P_2 &= \left[\begin{array}{c|c} [e^{'}]_x F & e^{'} \end{array}\right] \end{aligned} \tag{3.16}$$

where left and right null spaces of F gives the epipoles in each image. Note that $[e^{'}]_x$ is the skew-symmetric matrix (see Appendix D).

3.5 Scene Geometry

Given the image sequences of a scene, obtaining the 3D information from them is called structure from motion problem. Given the corresponding points $m \leftrightarrow m^{'}$ and camera matrices P and $P^{'}$, computing the 3D position of the scene point

from its projections on images is simply achieved by triangulation. In this way the scene is reconstructed, since these computations are carried with a projective ambiguity, this is called projective reconstruction.

3.5.1 Projective Ambiguity

If the cameras are uncalibrated, namely if the intrinsic and extrinsic parameters of each camera are unknown, nor their location with respect to each other, then reconstruction is possible up to a projective reconstruction.

3.5.2 Triangulation

Triangulation is simply the process of intersecting the backprojected lines from m and $m^{'}$. Given two camera matrices P and $P^{'}$, let m and $m^{'}$ be two points in the two images that satisfy the epipolar constraint, $m^{'T}Fm = 0$.
In particular it means that $m^{'}$ lies on the epipolar line Fm. This means that two rays back-projected from image points m and $m^{'}$ lie in a common epipolar plane which is the plane passing through the two camera centers. Since two rays lie in a plane, they will intersect at some point.
Numerically stable method of actually determining the point **M** at the intersection of the two rays back-projected from **m** and $\mathbf{m}^{'}$ is described in the next section.
The only points in 3-space that cannot be determined from their images are points on the baseline between two images. Baseline is the line connecting the two camera centers. In this case, the back-projected rays are collinear (both being equal to the baseline) and intersect along their whole length. Thus, the point **M** cannot be uniquely determined.

3.5.3 Linear Triangulation Methods

In each image, following equations are valid, $m = PM$ and $m^{'} = P^{'}M$, where $m \leftrightarrow m^{'}$ are corresponding points, P and $P^{'}$ are the projective matrices and M is the scene point in 3D.

These equations are transformed into a form $AM = 0$, which is an linear

equation in M. For the first image $x \times (PM) = 0$, which can be written as

$$\begin{aligned} x(p^{3T}M) - (p^{1T}M) &= 0 \\ y(p^{3T}M) - (p^{2T}M) &= 0 \\ x(p^{2T}M) - y(p^{1T}M) &= 0 \end{aligned} \tag{3.17}$$

where p^{iT} is the i^{th} row of the projection matrix, P. An equation of the form $AM = 0$ can be composed, with

$$A = \begin{bmatrix} xp^{3T} - p^{1T} \\ yp^{3T} - p^{2T} \\ x'p'^{3T} - p'^{1T} \\ y'p'^{3T} - p'^{2T} \end{bmatrix} \tag{3.18}$$

where two equations are included from each image, giving a total of four equations in four homogeneous unknowns. This is a redundant set of equations, since the solution is determined only up to a scale.

Solution is found by homogeneous method. M is assigned to the smallest singular value of A. For more information [2] can be referred.

CHAPTER 4

Relative Consistency of Projective Reconstructions

This chapter presents two measures for relative verification of projective reconstructions obtained from the same image pair. These methods are based on the ratio between the x, y and z coordinates of two 3D projective reconstructions of the same point correspondences. It is known that this ratio is similar for the same point correspondences. This information is used to measure the consistency of projective reconstructions obtained from the same image pair.

4.1 Image acquisition environment

In an ordinary room, two images of an object or a scene are captured with a 2/3" Color Progressive scan CCD camera at a resolution of 1300x1030. These two images are sequential images with a slight difference between them (see Figure 4.1 and Figure 4.2 courtesy of INRIA (France)).

4.2 Projective Reconstruction

When an image is taken by a camera, scene points are projectively transformed onto the image plane. It is assumed that camera calibration is unknown; camera is undergoing an unknown motion around the object or the scene. The steps in the projective reconstruction is shown in Figure 4.3. Given an image pair,

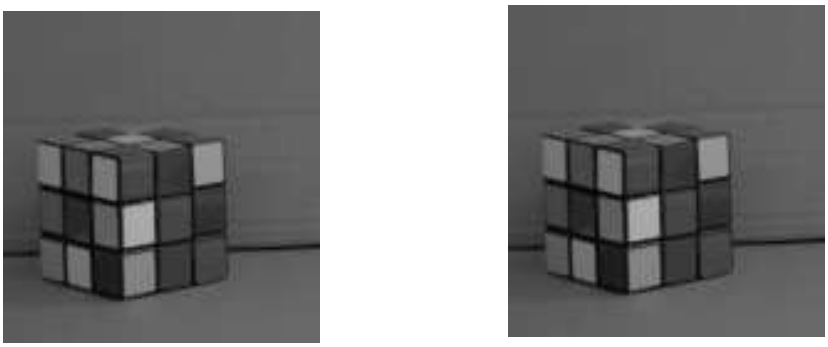

Figure 4.1: An image pair of a rubic cube.

Figure 4.2: An image pair of a scene.

interest points are computed by Canny edge detector [9] and then these interest points are matched. Given the corresponding points, fundamental matrix F is estimated, which is called weak calibration. From F, camera matrices P_1 and P_2 are computed, therefore camera geometry is obtained. Using camera geometry, projective reconstruction is achieved [11, 12, 13]. In other words, scene geometry is projectively defined. In the following sections, these steps are explained in detail.

4.2.1 Matching

Given an image pair, the aim is to match the corresponding points. First, points of interest in each image are extracted by Canny edge detector. Edge detector is preferred rather than corner detector because of the reasons mentioned in [12]. Taking one of the image as base image, corresponding point of each interest point in the base image is searched in the other image. For every interest point in the base image, a search window is defined in the other image centered at the pixel values of this interest point, as illustrated in Figure 4.4. A correlation window

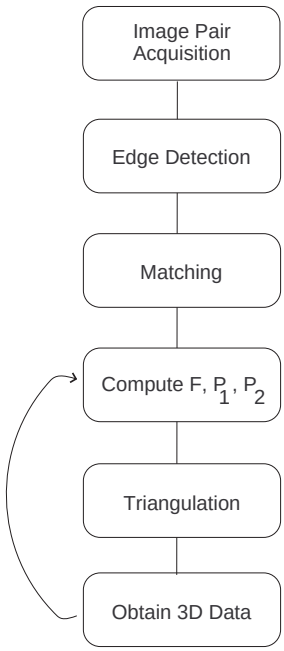

Figure 4.3: Steps in projective reconstruction. Repetition of steps after matching gives projectively equivalent reconstructions.

of size $(2n+1) \times (2m+1)$ is defined over the pixels in this search window. For every correlation window, normalized cross correlation of the intensity values of the local neighborhood is evaluated. Correlation score of points m_1 and m_2 is defined as:

$$C(m_1, m_2) = \frac{\sum_{i=-n}^{n} \sum_{j=-m}^{m} \left[I_1(u_1+i, v_1+j) - \overline{I_1(u_1, v_1)}\right] \times \left[I_2(u_2+i, v_2+j) - \overline{I_2(u_2, v_2)}\right]}{(2n+1)(2m+1)\sqrt{\sigma^2(I_1) \times \sigma^2(I_2)}} \tag{4.1}$$

where $\overline{I_k(u,v)} = \sum_{i=-n}^{n} \sum_{j=-m}^{m} I_k(u+i, v+j)/[(2n+1)(2m+1)]$ is the average intensity value of the correlation window centered at (u, v) and $\sigma(I_k)$ is the standard deviation of the image I_k in the neighborhood $(2n+1) \times (2m+1)$ of (u, v), which is given by:

$$\sigma(I_k) = \sqrt{\frac{\sum_{i=-n}^{n} \sum_{j=-m}^{m} (I_k(u,v) - \overline{I_k(u,v)})^2}{(2n+1)(2m+1)}} \tag{4.2}$$

Correlation window with the highest correlation score is taken as candidate. If this score is greater than a threshold value, these pixels are accepted as corresponding points. To achieve one-to-one matching, the pixels matched previously are not considered in the subsequent matching process.

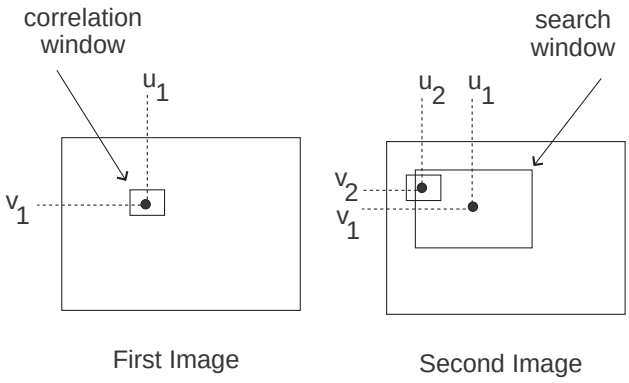

Figure 4.4: Correlation window centered at (u_1, v_1) in the first image and search window centered at (u_1, v_1) in the second image.

After the corresponding points are obtained, for each corresponding points pair, difference between their x coordinate, $disparity_x$ and difference between their y coordinate, $disparity_y$ are computed. Among these values, frequencies of each $disparity_x$ and $disparity_y$ are evaluated. The $disparity_x$ value with the highest frequency is selected as the most voted $disparity_x$, $tmvdisparity_x$ and in the same manner $tmvdisparity_y$ is selected. The corresponding points pairs that have $disparity_x$ and $disparity_y$ in a defined closure of these $tmvdisparity_x$ and $tmvdisparity_y$ are kept, the rest are eliminated.

Furthermore, mutually matching is tested by changing the roles of the images. For example, as shown in Figure 4.5 (a) (u_i, v_i) in the first image is matched to a point (u_j, v_j) in the second image, when the second image is the base image, (u_j, v_j) in the second image should also match to the point (u_i, v_i) in the first image. Therefore $(u_i, v_i) \leftrightarrow (u_j, v_j)$ can be accepted as corresponding points, since they are mutually matching each other. In Figure 4.5(b), (u_i, v_i) is matched to (u_j, v_j), however (u_j, v_j) is matched to (u_k, v_k). Therefore $(u_i, v_i) \leftrightarrow (u_j, v_j)$ is discarded, not accepted as corresponding points.

4.2.2 Computing the epipolar geometry

Epipolar geometry is estimated by the Algorithm 3 given below. Given the corresponding points $m \leftrightarrow m'$, fundamental matrix F is computed by using the

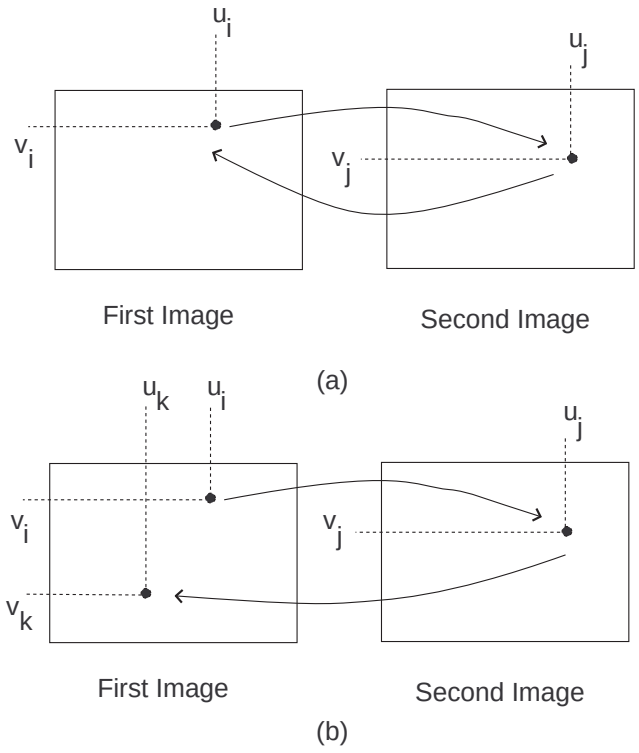

Figure 4.5: Finding the corresponding points. (a), $(u_i, v_i) \leftrightarrow (u_j, v_j)$ are accepted as corresponding points. (b) $(u_i, v_i) \leftrightarrow (u_j, v_j)$ are not accepted as corresponding points.

Algorithm 2. As mentioned before, each corresponding point must satisfy the epipolar constraint:

$$m^{'T} F m = 0 \tag{4.3}$$

Corresponding points satisfying this constraint is called an inlier, otherwise it is called an outlier. Number of corresponding points satisfying epipolar constraint is evaluated. Fundamental matrix, F that gives the smallest number of outliers is chosen as the fundamental matrix of the given image pair. Outliers are discarded and only inliers are carried to the next step. This is continued till the number of outliers is less than a threshold.

Algorithm 3 Estimation of epipolar geometry.

do

for n times do

compute F

evaluate the number of outliers

select the F with the smallest number of outliers

discard the outliers

while (number of outliers $< N$)

4.2.3 Camera geometry and triangulation

After the epipolar geometry is obtained, camera matrices are computed by using the method given in Section 3.4.4. Scene geometry is estimated by the linear triangulation method described in Section 3.5.3. Therefore 3D data set is obtained and projective reconstruction is achieved. Figure 4.6 gives the projective reconstruction obtained from the image pair of the rubic cube in Figure 4.1.

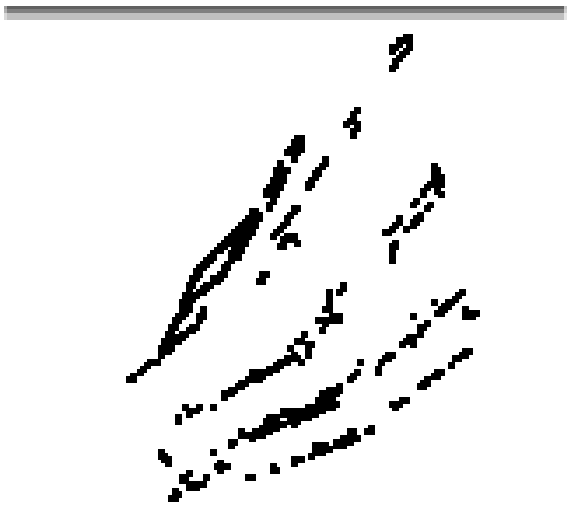

Figure 4.6: Projective reconstruction of the rubic cube.

As illustrated in Figure 4.3, if we repeat the steps of projective reconstruction

after matching, a new 3D data set is obtained from the same image pair.

4.3 Verification Measures

Two verification measures are introduced. But before that the main idea of these measures are given.

4.3.1 Theory behind the measures

For a reconstruction of a scene, when only a set of corresponding points are available, the obtained 3D points and camera matrices are true only up to a projective transform. This means; a group of projective reconstructions may be obtained from an image pair using corresponding points, which are related to each other by 4×4 transform matrices H. In other words any two such reconstructions are projectively equivalent.

Assuming that P_1 and P_2 are the camera matrices corresponding to the first reconstruction and $P_1^{'}$ and $P_2^{'}$ are the camera matrices of the second reconstruction, $P_1 = P_1^{'}H^{-1}$ and $P_2 = P_2^{'}H^{-1}$. Also suppose that M_i and $M_i^{'}$ are the 3D reconstructed points from the same corresponding points using the first camera matrices and second camera matrices, respectively, where $M_i = HM_i^{'}$. From the above relations we have:

$$x_i = P_1 M_i = P_1^{'} H^{-1} H M_i^{'} = P_1^{'} M_i^{'} \tag{4.4}$$

After the matching step is over, if the following steps are repeated (see Figure 4.3)

- compute F, P_1 and P_2.
- by triangulation obtain 3D data set.

different scene geometry are obtained, which are projectively equivalent.

Let's assume that $M = [x_1, x_2, x_3, x_4]^T$ and $M^{'} = [x_1^{'}, x_2^{'}, x_3^{'}, x_4^{'}]^T$ are the two 3D reconstructed points from the same corresponding points, $m \leftrightarrow m^{'}$ as shown in

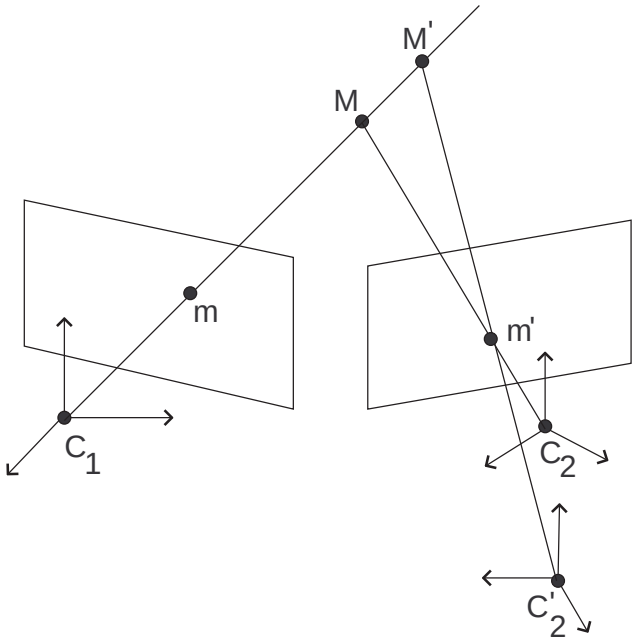

Figure 4.7: M and $M^{'}$ are the 3D reconstructed point of the same corresponding points, $m \leftrightarrow m^{'}$.

Figure 4.7. C_2 and $C_2^{'}$ are two different camera coordinate system of the second image for two different projective reconstructions obtained from the same corresponding points. Since world coordinate system is set to the first camera coordinate system, C_1 and $C_1^{'}$ are same. P_1, P_2 are the camera matrices of the first projective reconstruction and $P_1^{'}$, $P_2^{'}$ are the camera matrices of the second projective reconstruction. Therefore, the following equations are valid:

$$\begin{aligned} s_1 m &= P_1 M \\ s_2 m^{'} &= P_2 M \\ s_1^{'} m &= P_1^{'} M^{'} \\ s_2^{'} m^{'} &= P_2^{'} M^{'} \end{aligned} \tag{4.5}$$

where s_1, s_2, $s_1^{'}$, $s_2^{'}$ are nonzero scale factors. From Equation 4.5,

$$P_1 M = \frac{s_1}{s_1^{'}} P_1^{'} M^{'} \tag{4.6}$$

Since $P_1 = P_1^{'} = [I|0]$, this implies that

$$[x_1\ x_2\ x_3]^T = k[x_1^{'}\ x_2^{'}\ x_3^{'}]^T \tag{4.7}$$

where $M = [x_1\ x_2\ x_3\ 1]^T$, $M^{'} = [x_1^{'}\ x_2^{'}\ x_3^{'}\ 1]^T$ and $k = \frac{s_1}{s_1}$ is a nonzero scalar. This means that both M and $M^{'}$ lie on the same ray from the first camera center. Based on this fact, the ratio between the x, y and z coordinates of M and $M^{'}$ which are 3D reconstructed points of the same corresponding points should be equal. This is the basic idea behind the measures for evaluating the relative consistency of projective reconstructions.

4.3.2 Measure 1: Mean and Standard Deviation

It has been found out that for a given corresponding points, there is a relation between its two 3D reconstructions. If $M_1 = [x_1, y_1, z_1, 1]^T$ and $M_1^{'} =$

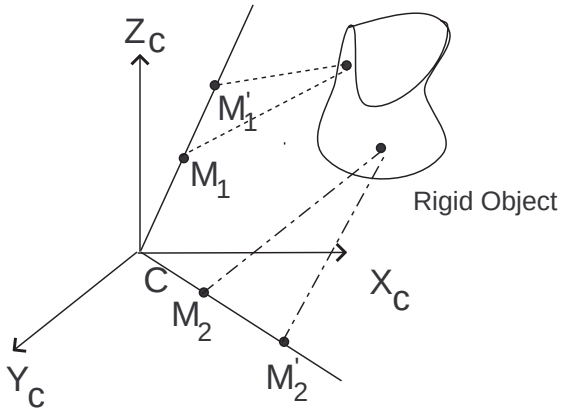

Figure 4.8: M_1 and $M_1^{'}$ are the two 3D reconstructed points of the same object point triangulated from the $m_1 \leftrightarrow m_1^{'}$. In the same manner, M_2 and $M_2^{'}$ that are triangulated from $m_2 \leftrightarrow m_2^{'}$.

$[x_1^{'}, y_1^{'}, z_1^{'}, 1]^T$ are the two 3D reconstructions of the same object point, and in the same manner, if $M_2 = [x_2, y_2, z_2, 1]^T$ and $M_2^{'} = [x_2^{'}, y_2^{'}, z_2^{'}, 1]^T$ are the two 3D reconstructions of another object point (see Figure 4.8), then the ratios between the x,y and z coordinates of the 3D reconstructions of the same object point are

equal.

$$\begin{aligned} \frac{x_1}{x_1'} = \frac{y_1}{y_1'} = \frac{z_1}{z_1'} = k_1 \\ \frac{x_2}{x_2'} = \frac{y_2}{y_2'} = \frac{z_2}{z_2'} = k_2 \\ \vdots \\ \frac{x_n}{x_n'} = \frac{y_n}{y_n'} = \frac{z_n}{z_n'} = k_n \end{aligned} \tag{4.8}$$

which can be expressed as

$$\begin{aligned} r_{x1} = r_{y1} = r_{z1} = k_1 \\ r_{x2} = r_{y2} = r_{z2} = k_2 \\ \vdots \\ r_{xn} = r_{yn} = r_{zn} = k_n \end{aligned} \tag{4.9}$$

These ratios, r_{x1}, r_{y1},..., r_{z1} in theory should be equal but in practice they are very close to each other. Therefore the absolute values of their differences are expected to be zero. For every common object point, in two projective reconstructions, average ratio of these three ratios is evaluated. Let's call it avgr.

$$\begin{aligned} avgr_1 = (r_{x1} + r_{y1} + r_{z1})/3 \\ avgr_2 = (r_{x2} + r_{y2} + r_{z2})/3 \\ \vdots \\ avgr_n = (r_{xn} + r_{yn} + r_{zn})/3 \end{aligned} \tag{4.10}$$

The difference between $avgr$ and any of these three ratios is expected to be zero. However due to error in measurements, it might not be the case always. $avgr$ is subtracted from each ratio. Absolute value of these three subtractions are added up and its average is computed. Let's call this value $avgdiff$.

$$\begin{aligned} avgdiff_1 = (|r_{x1} - avgr_1| + |r_{y1} - avgr_1| + |r_{z1} - avgr_1|)/3 \\ avgdiff_2 = (|r_{x2} - avgr_2| + |r_{y2} - avgr_2| + |r_{z2} - avgr_2|)/3 \\ \vdots \\ avgdiff_n = (|r_{xn} - avgr_n| + |r_{yn} - avgr_n| + |r_{zn} - avgr_n|)/3 \end{aligned} \tag{4.11}$$

Then, mean and standard deviation of avgdiff is computed and used in verification of projective reconstructions. For relatively consistent projective reconstructions they are expected to be zero.

4.3.3 Measure 2: Differences of Ratios

In the second measure, the equality of ratios is used. The difference between the ratios must be zero. In this way, the differences between these ratios are looked at, the ones having the difference less than a threshold are named as goodpoints.

$$\begin{aligned} |r_{x1} - r_{y1}| &< Threshold \\ |r_{x1} - r_{z1}| &< Threshold \\ |r_{y1} - r_{z1}| &< Threshold \end{aligned} \tag{4.12}$$

The ratio of the goodpoints to the common points has been used as a measure to verify the projective reconstructions relatively. Common points are the scene points that are present in both of the projective reconstructions.

$$ratio = \frac{goodpoints}{commonpoints} \tag{4.13}$$

4.4 Experimental Results

Experimental results of the image pair shown in Figure 4.1 and Figure 4.2 are considered separately. The first measure shows the consistency of two projective reconstructions by analyzing the mean and the standard deviation of $avgdiff$. If they are consistent, then the mean and the standard deviation of $avgdiff$ is expected to be close to zero. The second measure shows the consistency of two projective reconstructions by the use of ratio, which is the number of good points to the number of common points in both reconstructions. For consistent projective reconstructions, this value is expected to be 1 or close to 1.

4.4.1 Experimental Results for The Rubic Cube Image Pair

512×512 images are used. In the first image of the rubic cube, 3606 interest points and in the second image of the sequence, 3657 interest points are detected by Canny edge detector. During the matching step, by using normalized-cross-correlation score the number of corresponding points is 3594. This number is reduced to 2470 by disparity check. After the mutually constraint, 2024 corresponding points are remained valid. Epipolar geometry is estimated by randomly chosen corresponding points among these 2024 corresponding points

as described in Algorithm 3. Some of the corresponding points, which are not satisfying the the epipolar constraint

$$m^{'T} F m = 0 \tag{4.14}$$

are discarded. In fact, because of the errors in measurement and computations till this step, epipolar constraint is not expected to be zero as in Equation 4.14. In other words, given the corresponding points $m \leftrightarrow m^{'}$, point $m^{'}$ may not be on the epipolar line l_m, where $l_m = Fm$. It can be allowed number of pixels far away from the epipolar line. That's why epipolar constraint is replaced as follows:

$$distance(m^{'T}, Fm) < \tau \tag{4.15}$$

where the corresponding points $m \leftrightarrow m^{'}$ are accepted as inliers, if the distance between point $m^{'}$ and the epipolar line $l_m = Fm$ is less than τ pixels. Therefore the number of corresponding points is reduced differently each time this step is repeated. Using linear triangulation method in Section 3.5.3, projective reconstruction is accomplished, 3D data set is obtained. Repetitions of steps after matching gives different 3D data sets that are projectively equivalent. Each experiment analyzes the relative consistency two projective reconstructions. Measure 1 and Measure 2 are carried on the 3D points of each projective reconstructions obtained from the common corresponding points.

As shown in Table 4.1, the first projective reconstruction is obtained with the stated τ_1 value and the second projective reconstruction is obtained with the stated τ_2 value. For Measure 1, *mean* and *standard deviation* of $avgdiff$ and for Measure 2, *ratio* are shown. Depending on these experiments, projective reconstructions in the 5^{th} experiment with the $\tau_1 = 0.9$ and $\tau_2 = 0.9$ values, is found to have the lowest *mean*, a very low *standard deviation* and the $ratio = 1$. Therefore among all, these two projective reconstructions are the most relatively consisted ones. Projective reconstructions in the 6^{th} experiment with the $\tau_1 = 1.3$ and $\tau_2 = 1.3$ values, is found to have the highest *mean* with a high *standard deviation* as compared to others and the $ratio = 0.34841$. Therefore among these experiments, these two projective reconstructions are the least relatively consisted ones. However, projective reconstructions in the

Table 4.1: Relative consistency of projective reconstructions obtained from an image pair of a rubic cube in Figure 4.1 with respect to various τ_1 and τ_2 values.

Exp.Num.	τ_1	τ_2	*Mean* (10^{-7})	*Standard Deviation* (10^{-7})	*Ratio*
1	0.5	0.5	6.6	8	1
2	0.5	0.5	258.7	414.5	0.92648
3	0.7	0.7	6.3	5.4	1
4	0.7	0.7	9.5	10	1
5	0.9	0.9	5.9	15.7	1
6	1.3	1.3	54650	302990	0.34841
7	1.3	1.3	9.5	30.7	1
8	1.5	1.5	6.5	7.7	1
9	1.5	1.5	13.7	119.9	0.99505
10	1.7	1.7	6.7	16.2	1
11	1.7	1.7	11.4	56.7	0.99801
12	2.0	2.0	26.1	51.3	1
13	2.0	2.0	11.7	29.1	1
14	2.5	2.5	757.2	12456	0.79088
15	2.5	2.5	8.3	30.4	1
16	3.0	3.0	199.6	258.2	0.98613
17	3.0	3.0	61.2	87.2	1
18	0.5	3.0	78	80.9	1
19	0.7	2.5	8.7	21	1
20	0.9	2.0	7.4	22.7	1
21	1.3	1.7	61.8	96.4	1

7^{th} experiment with the same $\tau_1 = 1.3$ and $\tau_2 = 1.3$ values as in the 6^{th} experiment is found to have a very low *mean* and *standard deviation* with the *ratio* = 1. Therefore, τ_1 and τ_2 values that are used in generation of 3D data set indirectly, do not help us to estimate the relative consistency of two projective reconstructions. Measure 1 and Measure 2 only measure the relative consistency of two given projective reconstructions. If the experiments are repeated with the same τ_1 and τ_2 values, then different results are obtained. This is because of the nature of projective matrices and fundamental matrix estimations. Using different τ values in each projective reconstruction does not change the *mean*, *standard deviation* and *ratio* values as expected. Since these τ_1 and τ_2 values do not change the number of corresponding points for the image pair of rubic cube. Measure 1 and Measure 2 are also verifies each other. When *mean* and *standard deviation* is very low, *ratio* is equal to 1 and when *mean* and *standard deviation* is not very low, *ratio* is less than 1. It can be seen from the experiments that Measure 2 is less sensitive than Measure 1. *Mean* and *standard deviation* varies frequently, whereas *ratio* is equal to 1 most of the time.

4.4.2 Experimental Results for The Scene Image Pair

512×256 images are used. In the first image of the scene, 4622 interest points and in the second image of the sequence, 4570 interest points are detected by Canny edge detector. Using normalized-cross-correlation score 4533 corresponding points are found. This number is reduced to 2576 by disparity check. After the mutually constraint, 2028 corresponding points are remained valid. Depending on these experiments illustrated in Table 4.2, projective reconstructions in the 2^{nd} experiment with the $\tau_1 = 0.5$ and $\tau_2 = 0.5$ values, is found to have the lowest *mean* and *standard deviation* with the *ratio* = 1. Therefore among all, these two projective reconstructions are the most relatively consisted ones. Projective reconstructions in the 5^{th} experiment with the $\tau_1 = 0.9$ and $\tau_2 = 0.9$ values, is found to have the highest *mean* and *standard deviation* with the *ratio* = 0.02854. Therefore among these experiments, these two projective re-

Table 4.2: Relative consistency of projective reconstructions obtained from an image pair of a scene in Figure 4.2 with respect to various τ_1 and τ_2 values.

$Exp.Num.$	τ_1	τ_2	$Mean\ (10^{-7})$	$StandardDeviation\ (10^{-6})$	$Ratio$
1	0.5	0.5	2836.9	1248.7	0.64190
2	0.5	0.5	70.3	4.39	1
3	0.7	0.7	998.9	67.36	0.60183
4	0.7	0.7	420.8	42.78	0.91531
5	0.9	0.9	102080	16344	0.02854
6	1.3	1.3	768	72.55	0.90375
7	1.3	1.3	976	111	0.68525
8	1.5	1.5	140.7	26.36	0.97287
9	1.5	1.5	528.7	76.97	0.89936
10	1.7	1.7	496.3	62.42	0.90774
11	1.7	1.7	71.4	5.4	1
12	2.0	2.0	2448	331.8	0.45064
13	2.0	2.0	298.1	45.29	0.91708
14	2.5	2.5	573	67.49	0.83909
15	2.5	2.5	375.6	65.07	0.89639
16	3.0	3.0	1658.8	218.99	0.57129
17	3.0	3.0	251.1	25.09	0.96201
18	0.5	3.0	1665.3	190.16	0.50822
19	0.7	2.5	188.1	21.56	0.99948
20	0.9	2.0	172.3	20.24	0.98421
21	1.3	1.7	4205.1	672.16	0.26492
22	0.9	0.9	87.5	12.03	1

constructions are the least relatively consisted ones. However, projective reconstructions in the 22^{nd} experiment with the same $\tau_1 = 0.9$ and $\tau_2 = 0.9$ values as in the 5^{th} experiment is found to have a very low *mean* and *standard deviation* with the *ratio* = 1. As stated in 4.4.1, if the experiments are repeated with the same τ_1 and τ_2 values, then different results are obtained. This is because of the nature of projective matrices and fundamental matrix estimations. Again, using different τ values in each projective reconstruction does not change the *mean*, *standard deviation* and *ratio* values as expected. Since these τ_1 and τ_2 values do not change the number of corresponding points for the image pair of scene. Measure 1 and Measure 2 are also consisted with each other. When *mean* and *standard deviation* is very low, *ratio* is equal to 1 and when *mean* and *standard deviation* is not very low, *ratio* is less than 1. Different than image pair of rubic cube, image pair of scene gives higher *mean* and *standard deviation* and lower *ratio* values. Therefore, projective reconstructions obtained from image pair of scene are less relatively consisted than the projective reconstructions obtained from image pair of rubic cube. This shows the difference between the used image pairs. Rubic cube was easier for macthing as compared to scene and this is reflected in results.

CHAPTER 5

Conclusion and Future Work

In this study, projective reconstruction from an image pair is obtained. For the same image pair, projective reconstruction is repeated and two projective reconstructions obtained from the same image pair are checked whether they are verifying each other. To accomplish this, two measures are presented for measuring the relative For consisted projective reconstructions *mean* and *standard deviation* are expected to be very close to zero and *ratio* is expected to be 1. It can be said that these measures confirm each other when *mean* and *standard deviation* is very low, *ratio* gets its maximum value, 1 and when *mean* and *standard deviation* is not very low as compared to others, *ratio* is less than its maximum value, 1. But as stated earlier, measure1 gives more reliable analysis than measure2 since measure1 is more sensitive which involves standard deviation and mean calculations. However, these measures don't measure the precision of projective reconstructions individually. They only measure the consistency of two projective reconstructions with each other. If there is a projective reconstruction, which is known to be a precise one, it can be used for testing the candidate projective reconstructions and choose the most consistent one, which can be assumed to be also precise. The ones having the highest consistency can be chosen for further reconstructions. As seen in Table 4.1 and Table 4.2, in general the *mean* and *standard deviation* of the experiments performed on the rubic cube image pair are less than the *mean* and *standard deviation* of

the experiments performed on the scene image pair. In other words, projective reconstructions obtained from the rubic cube image pair seem more consisted with each other than the projective reconstructions obtained from the scene image pair. This may give us an idea of the projective reconstructions of different image pairs. Since matching is a common source of errors, maybe this result shows that there are less errors in matching of rubic cube image pair. In addition, since collinearity is preserved in projective reconstruction, given a group of three or more points on the same line (for example points belonging to the edge of an object with geometric shape like a cube) we can find the reconstructed points using two different F matrices. These points should be:

- on the same ray from the camera center in both reconstructions,
- on the same line in each reconstruction.

Other projective transformation invariants such as incidence and cross ratio can be also used. These are ideas for future work.

REFERENCES

[1] M. Pollefeys, "Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences", PhD thesis, ESAT-PSI, K.U.Leuven, 1999.

[2] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", *Cambridge University Press*, 2000.

[3] O. Faugeras, "Three-Dimensional Computer Vision, A Geometric Viewpoint", *MIT Press*, 1993.

[4] O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig", *Proc. ECCV*, pp.563–578, 1992.

[5] E. Trucco, A. Verri, "Introductory Techniques for 3-D Computer Vision" *Prentice-Hall*, 1998.

[6] Q. T. Luong, O. D. Faugeras, "Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices", *Int. Journal of Computer Vision*, 22(3), pp.261-289, 1997

[7] P. Beardsley, A. Zisserman and D. Murray, "Sequential Updating of Projective and Affine Structure from Motion", *International Journal of Computer Vision*, Vol.23 No.3, pp.235-259, 1997.

[8] R. Hartley, "Triangulation", *Computer Vision and Image Understanding*, Vol.68, No.2, pp. 146-157, 1997.

[9] J. F. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 679-698, 1986.

[10] R. Deriche, Z. Zhang, Q.-T. Luong and O. Faugeras., "Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig", *Proceedings of the Internat. Conference, European conference on Computer Vision*, Stockholm, Sweden, pp. 567-576, 1994.

[11] Q.T. Luong, O.D. Faugeras, "The fundamental Matrix: Theory, Algorithms and stability Analysis", *IJCV*, Vol.17, pp. 43-75, 1996.

[12] C. Rothwell, O. Faugeras, and G. Csurka, "A comparison of projective reconstruction methods for pairs of views", *Computer Vision and Image Understanding*, Vol.68, No.1, pp. 37-58, 1997.

[13] P. Beardsley, P. Torr, and A. Zisserman, "3D model acquisition from extended image sequences", *In Proc. European Conf. on Computer Vision*, Vol. 2, pp. 683-695, 1996.

[14] R. I. Hartley, "In Defense of the Eight-Point Algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, No.6, pp. 580–593, June 1997.

[15] Stan Birchfield, "Tutorial: An Introduction to Projective Geometry (for computer vision)", 1998.

[16] Z. Zhang, "Determining the Epipolar Geometry and its Uncertainty: A Review", *IJCV The International Journal of Computer Vision*, Vol.27, No.2, pp. 161–195, 1998.

[17] M. A. Fischler, R.C. Bolles, "Random Sample Consensus: A paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of the ACM*, Vol.24, No.6, pp. 381–395, June 1981.

[18] M. Pollefeys, R. Koch, L. V. Gool, "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters", *Technical Report*, Nr. KUL/ESAT/MI2/9707, August 1997.

[19] B. Triggs "Autocalibration and the Absolute Conic", *Conference on Computer Vision and Pattern Recognition*, June 1997.

[20] Z. Zhang, R. Deriche, O. Faugeras, Q. T. Luong, "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry", *Research Report*, No:2273, May 1994.

[21] I. J. Cox, S. L. Hingorani, S. B. Rao, "A Maximum Likelihood Stereo Algorithm", *Research Report*, No:2273, May 1994.

[22] R. I. Hartley, "Self-Calibration of Stationary Cameras", *IJCV International Journal of Computer Vision*, Vol.22, No.1, pp. 5–23, 1997.

[23] Q.T. Luong, O.D. Faugeras, "Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices", *IJCV International Journal of Computer Vision*, Vol.22, No.3 pp. 261-289, 1997.

APPENDIX A

Invariants of Projective Transformation

Projective geometry preserves some of the properties of the scene. These are as follows:

Incidence: If a point p is lying on a line l, after projective transformation point p lies on the line l. *Collinearity*: If there are n points that lie on a line, after projective transformation those points lie on the same line. *Tangency*: If a line l is tangent to a curve at a point p, after projective transformation, line l is tangent to the curve at the point p. *Cross Ratio*: Let p_1, p_2, p_3 and p_4 be collinear (on the same line) points as illustrated in Figure A.1. Crossratio is

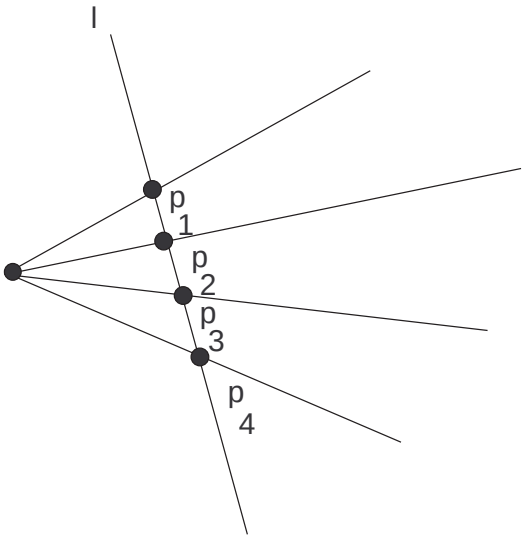

Figure A.1: Cross ratio of points p_1, p_2, p_3, p_4 on the same line l.

ratio of ratios.

$$CrossRatio(p_1, p_2; p_3, p_4) = \frac{\frac{\Delta_{13}}{\Delta_{14}}}{\frac{\Delta_{23}}{\Delta_{24}}} \tag{A.1}$$

which is

$$CrossRatio(p_1, p_2; p_3, p_4) = \frac{\Delta_{13}\Delta_{24}}{\Delta_{14}\Delta_{23}} \tag{A.2}$$

where Δ_{ij} defines the distance between point i and point j.

APPENDIX B

How the geometric structure can be upgraded?

When image sequence of a scene are taken, scene is projected to the images. What we have at hand is the projective structure of the scene. However, using enough number of images, geometric structure of the scene can be upgraded from projective to metric, since metric structure is the highest structure that can be retrieved from images [6].

For better understanding of these concepts ideal points, line at infinity, l_∞, plane at infinity, π_∞ and absolute conic, Ω_∞ must be described in detail.

Under a projective transformation ideal points may be mapped to finite points and consequently l_∞ is mapped to a finite line. However, if the transformation is an affinity, then l_∞ is not mapped to a finite line, but remains at infinity [2].

The line at infinity, l_∞, is a fixed line under the affine transformation. Identification of l_∞ allows the recovery of affine properties (parallelism, ratio of areas) [2].

The plane at infinity π_∞ is fixed under affine transformation. The plane π_∞ is a geometric representation of the 3 degrees of freedom required to specify affine properties in a projective coordinate frame [2]. It has the canonical position

$$\pi_\infty = [0, 0, 0, 1]^T$$

in affine 3-space and enables the identification of affine properties such as parallelism [2]. Two planes are parallel if and only if their line of intersection is on π_∞. A line is parallel to another line, or to a plane, if the point of intersection is on π_∞.

Once π_∞ is identified in projective 3-space, it is then possible to determine the affine properties such as whether geometric entities are parallel if they intersect on π_∞.

The conic at infinity Ω_∞ a is fixed conic under any similarity transformation [2]. The conic Ω_∞ is a geometric representation of the 5 additional degrees of freedom that are required to specify metric properties in an affine coordinate frame. The absolute conic, Ω_∞, is a (point) conic on π_∞. In a metric frame $\pi_\infty = [0, 0, 0, 1]^T$, and points on Ω_∞ satisfy

$$X_1^2 + X_2^2 + X_3^2 = 0$$
$$X_4 = 0$$

This conic is composed of imaginary points on π_∞.

Once Ω_∞ (and its support plane π_∞) have been identified in projective 3-space then metric properties, such as angles and relative lengths, can be measured [2].

APPENDIX C

Singularity Constraint

The $n \times n$ matrix A is non-singular (and thus has an inverse) if and only if the homogeneous linear system $AX = 0$ has only the trivial solution $X = 0$.
Since 3×3 fundamental matrix F is expected to be a non-trivial solution therefore it must be a singular matrix with rank 2.

APPENDIX D

Skew-symmetric Matrix and Cross Product

If $a = [a_1, a_2, a_3]$, then corresponding skew-symmetric matrix is defined as follows:

$$[a]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \tag{D.1}$$

Cross product of two 3-vectors $a \times b$ is defined as:

$$a \times b = [a]_x \, b \tag{D.2}$$