NEW METHODS FOR WIDE-BASELINE $\label{eq:mage_energy} \mbox{IMAGE INTERPOLATION}$

by

JEFF WOOD

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2016

Copyright © by Jeff Wood 2016 All Rights Reserved To all those who were told they couldn't \dots

... but persevered, and did anyways.

TABLE OF CONTENTS

LI	ST O	F ILLU	JSTRATIONS	V
LI	ST O	F TAB	BLES	vi
CO	OMM	ONLY	USED SYMBOLS AND NOTATION	vii
Ch	aptei	ſ	F	Page
1.	Back	kground	d	1
	1.1	Chang	ge of Reference	1
	1.2	Points	s and Lines in the Image Plane	2
	1.3	Epipo	lar Geometry	3
	1.4	Funda	amental Matrix	4
		1.4.1	Theoretical Calculation	5
		1.4.2	Calculation from Data	5
	1.5	Came	ra Calibration Matrix	6
		1.5.1	Pinhole Camera Model	6
		1.5.2	Intrinsic Calibration Matrix	8
		1.5.3	Extrinsic Calibration Matrix	9
	1.6	Essent	tial Matrix	10
2.	Rect	ificatio	on	11
RF	EFER	ENCE	S	12
RΙ	OCR	Δ PHIC	TAI STATEMENT	13

LIST OF ILLUSTRATIONS

Figure		Pa	ge
1.1	Pinhole CameraModel		7

LIST OF TABLES

Table Page

COMMONLY USED SYMBOLS AND NOTATION

Symbol	Description
v	Vectors in lowercase bold
v_a	a -component of vector \mathbf{v}
$v_{a/b}$	The ratio of vector v 's a -component to it's b -component
${f M}$	Matrices in uppercase bold
$ ilde{\mathbf{M}}$	Matrix ${\bf M}$ expressed homogeneously such that ${\it right\ lower-}$
	most entry equals 1
$M_{r,c}$	Entry in row r and column c of matrix \mathbf{M}
\mathbf{m}_c	$Vector$ occurring in column c of matrix \mathbf{M}
x	Generic 3-dimensional spatial coordinate
$ ilde{\mathbf{x}}$	Generic 3-dimensional spatial coordinate (expressed $homo-$
	geneously)
У	Generic 2-dimensionals image coordinate
$ ilde{ ilde{\mathbf{y}}}$	Generic 2-dimensional image coordinate (expressed $homo-$
	geneously)
\mathbf{u}	Pixelized 2-dimensional image coordinate
$ ilde{ extbf{u}}$	Pixelized 2-dimensional image coordinate (expressed ho -
	mogeneously)
$^{A}\mathbf{x}$	Generic 3-dimensional spatial coordinate in reference frame
	A
$^{A}\mathbf{ ilde{x}}$	Generic 3-dimensional spatial coordinate (expressed $homo-$
	geneously) in reference frame A

- ${}^{C}_{B}\tilde{\mathbf{M}}$ Change from of reference frame B to reference frame C
- \cong Equal to to a scale factor. Used in $\mathbf{v} \cong \tilde{\mathbf{v}} \iff \mathbf{v} = s \cdot \tilde{\mathbf{v}}$ or $\mathbf{M} \cong \tilde{\mathbf{M}} \iff \mathbf{M} = s \cdot \tilde{\mathbf{M}}$
- f focal-length
- s Scalar applied to homogeneous vector $\tilde{\mathbf{v}}$ or homogeneous matrix $\tilde{\mathbf{M}}$ such that original $\mathbf{v} = s \cdot \tilde{\mathbf{v}}$ or $\mathbf{M} = s \cdot \tilde{\mathbf{M}}$ is recovered
- $^{D}\mathbb{S}$ Spatial reference frame D
- $[\mathbf{x}]_{\times}$ Skew-symmetric matrix version of vector \mathbf{x} used as *left*-operand in the *cross*-product such that $[\mathbf{x}]_{\times} \cdot \mathbf{y} = \mathbf{x} \times \mathbf{y}$
 - l Epipolar line (expressed as *vector*
 - $\mathbb{P} \qquad \text{Ray (or } pencil) \text{ of all possible vectors } \mathbf{x} \text{ where } \mathbf{x} = s \cdot \tilde{\mathbf{x}} \text{ for}$ some value of s

CHAPTER 1

Background

Oridinarily, real-world data contains 3-dimensions. Because standard images only include 2-dimensional data, information regarding depth is lost (i.e. it is often difficult to judge distance from a single image without visual cues). Stereovision attempts to resolved this by finding the same point in both stereoscopic images (known as a corresponding point), and recovering the depth information. An elementry example of this occurs in stereoscopic images with relatively low distance between cameras (i.e. they are right next to each other). Objects that are farther away from the observer occur closer together in the stereo images, whereas objects closer to the camera appear appear farther appart in the stereo-images.

1.1 Change of Reference

Each view from a pair of stereo-images encompasses its own frame of reference (i.e. the directions of forward or backward are unique to image and may differe considerably depending on camera displacement). As such it is necessary to be able to express on coordinates ${}^{A}\mathbf{x}$ a given reference frame as coordinates ${}^{B}\mathbf{x}$ in another reference frame.

Coordinates given in ${}^A\mathbf{x}$ can be expressed in ${}^B\mathbf{x}$ by the geometric transformation:

$${}^{B}\mathbf{x} = {}^{B}_{A}\mathbf{R} \cdot {}^{A}\mathbf{x} + {}^{B}_{A}\mathbf{t}$$

or

$${}^{B}\tilde{\mathbf{x}} = \begin{bmatrix} {}^{B}\mathbf{R} & {}^{B}\mathbf{t} \\ \hline 0 & 1 \end{bmatrix} \cdot {}^{A}\tilde{\mathbf{x}}$$

$$= {}^{B}_{\Lambda}\mathbf{M} \cdot {}^{A}\tilde{\mathbf{x}}$$

where ${}^B_A\mathbf{M}$ is also the geometric transformation necessary to transform ${}^B\mathbb{S}$ into ${}^A\mathbb{S}$. Without calculating any new quantities, rearranging allows us to express coordinates in ${}^B\mathbf{x}$ in the ${}^A\mathbf{x}$ reference frame as:

$${}_{A}^{B}\mathbf{R}^{\intercal}\cdot({}^{B}\mathbf{x}-{}_{A}^{B}\mathbf{t})={}^{A}\mathbf{x}$$

and similarly transforms ${}^{A}\mathbb{S}$ into ${}^{B}\mathbb{S}$.

1.2 Points and Lines in the Image Plane

Points in world-space of \mathbb{R}^3 are converted to points in the image-plane of \mathbb{R}^2 by homogenization. This occurs when a world-coordinate of $\mathbf{x} = [x_1, x_2, x_3]^{\mathsf{T}}$ is mapped to a homogeneous image coordinate of $\tilde{\mathbf{y}} = [y_1, y_2, 1]^{\mathsf{T}} = [x_1/x_3, x_2/x_3, x_3/x_3]^{\mathsf{T}}$ or a non-homogeneous image coordinate of $\mathbf{y} = [y_1, y_2]^{\mathsf{T}} = [x_1/x_3, x_2/x_3]^{\mathsf{T}}$. Points of the form $\tilde{\mathbf{y}} = [y_1, y_2, 0]^{\mathsf{T}}$ are special case of homogeneous point referred to as a point at infinity.

Lines in \mathbb{R}^2 can be represented in different contexts. The *vector offset* method calculates a line $\mathbf{s}(t)$ between points $\mathbf{y_1}$ and $\mathbf{y_2}$ as

$$\mathbf{s}(t) = (1 - t) \cdot \mathbf{y_1} + t \cdot \mathbf{y_2}$$
$$= \mathbf{y_1} + t \cdot (\mathbf{y_2} - \mathbf{y_1})$$

in which the line is parallel to the vector $\mathbf{y_2} - \mathbf{y_1}$ and offset from the origin by the vector $\mathbf{y_1}$. Lines are also represented by their coefficients as $\mathbf{l} = [a, b, c]^{\mathsf{T}}$ where

$$\mathbf{l}^{\mathsf{T}} \cdot \tilde{\mathbf{y}} = \begin{bmatrix} a & b & c \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ 1 \end{bmatrix}$$
$$= a \cdot y_1 + b \cdot y_2 + c \cdot 1$$
$$= 0$$

This definition lets us say $\tilde{\mathbf{y}}$ is located on line \mathbf{l} if and only if $\mathbf{l}^{\intercal} \cdot \tilde{\mathbf{y}} = 0$. The line \mathbf{l} joining two homogeneous image coordinates $\tilde{\mathbf{y}}_1$ and $\tilde{\mathbf{y}}_2$ is then calculated as the cross product of $\mathbf{l} = \tilde{\mathbf{y}}_1 \times \tilde{\mathbf{y}}_2$.

1.3 Epipolar Geometry

Each point of of interest (also referred to as a *feature*) in a single image occurs in a 2-dimensional space at location $\tilde{\mathbf{y}} = [x, y, 1]^{\mathsf{T}}$. The same point in space when viewed from an image at a similar (though different) angle is referred to as a *corresponding* point with location of $\tilde{\mathbf{y}}' = [x', y', 1]^{\mathsf{T}}$. This set of infinitley many points form a 1-dimensional subspace (also known as a *pencil*) of the 3-dimensional world space.

 $\tilde{\mathbf{y}}' = [x', y', 1]^{\mathsf{T}}$ produces an epipolar line in the *original image* of $\mathbf{l} = [A, B, C]$. The original point of $\tilde{\mathbf{y}} = [x, y, 1]^{\mathsf{T}}$ must lie located on this epipolar line as required by the epipolar constraint, resulting in the *line-point equality* of $\mathbf{l}^{\mathsf{T}} \cdot \tilde{\mathbf{y}} = 0$ for the *original image*.

When viewed in ther respective images, each point ($\tilde{\mathbf{y}}$ and $\tilde{\mathbf{y}}'$) has a pencil that coincides with that point. Since the pencils act as directional-vectors in 3-dimensional space, there is a unique 2-dimensional plane which contain both of these vectors, known as the epipolar plane. It is the intersection of the epipolar plane with the original image-plane and the angled image-plane that results in the epipolar lines of \mathbf{l} and \mathbf{l}' , respectively. In fact, the epipolar plane (in each image's coordinate systems)² has the same vector form as its epipolar line. Specifically, in the original image reference frame $\mathbf{l} = \mathbf{P} = [A, B, C]^{\mathsf{T}}$, and in the angled image reference frame $\mathbf{l}' = \mathbf{P}' = [A', B', C']^{\mathsf{T}}$. This results from the fact that any world-point \mathbf{x} lying on the epipolar plane \mathbf{P} will result in a homogeneous image-point $\tilde{\mathbf{y}}$ that also lies on the plane \mathbf{P} . Specifically, when $\mathbf{x} = s \cdot \tilde{\mathbf{y}}$ for some non-zero value of s, then $\mathbf{P}^{\mathsf{T}} \cdot \mathbf{x} = 0$ implies $\mathbf{P}^{\mathsf{T}} \cdot \mathbf{x} = \mathbf{P}^{\mathsf{T}} \cdot (s \cdot \tilde{\mathbf{y}}) = 0$. Since $s \neq 0$, its true that $\mathbf{P}^{\mathsf{T}} \cdot \tilde{\mathbf{y}} = 0$.

In the majority of images, the sets of epipolar lines will converge at a point known as an *epipole*, denoted as **e** in the *original image* and **e**' in the *angled image*.

1.4 Fundamental Matrix

In stereo vision, points $(\tilde{\mathbf{x}})$ in one image I are related to the epipolar line (l') that contain the corresponding point $(\tilde{\mathbf{x}}')$ by the Fundamental Matrix (\mathbf{F}) .

$$l' = \mathbf{F} \cdot \tilde{\mathbf{x}}$$

²There is a single *epipolar plane* for each pair of corresponding points $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{y}}'$. However, the single plane can be parameterized infinitley many ways, depending on the *frame of reference*

1.4.1 Theoretical Calculation

1.4.2 Calculation from Data

Calculation of the Fundamental Matrix through theoretical metods may be difficult due to the abscense of parameters or through errors introduced in the use of pixelized points common in digital cameras. As previously mentioned, the fundamental matrix \mathbf{F} for points $\tilde{\mathbf{y}}_i = [y_{i1}, y_{i2}, 1]^{\intercal}$ and $\tilde{\mathbf{y}}_i' = [y_{i1}', y_{i2}', 1]^{\intercal}$ (for all i) is represented as the formula $0 = \tilde{\mathbf{y}}_i'^{\intercal} \cdot \mathbf{F} \cdot \tilde{\mathbf{y}}_i$. Rearranging this gives:

$$0 = \tilde{\mathbf{y}}_{i}^{\prime \mathsf{T}} \cdot \mathbf{F} \cdot \tilde{\mathbf{y}}_{i}$$

$$= \begin{bmatrix} y_{i1} & y_{i2} & 1 \end{bmatrix} \begin{bmatrix} f_{1} & f_{2} & f_{3} \\ f_{4} & f_{5} & f_{6} \\ f_{7} & f_{8} & f_{9} \end{bmatrix} \begin{bmatrix} y_{i1} \\ y_{i2} \\ 1 \end{bmatrix}$$

which when the matrices are multiplied, gives

$$0 = \begin{bmatrix} y'_{i1} & y'_{i2} & 1 \end{bmatrix} \begin{bmatrix} f_1 \cdot y_{i1} + f_2 \cdot y_{i2} + f_3 \\ f_4 \cdot y_{i1} + f_5 \cdot y_{i2} + f_6 \\ f_7 \cdot y_{i1} + f_8 \cdot y_{i2} + f_9 \end{bmatrix}$$

$$= \begin{bmatrix} f_1 \cdot y_{i1} \cdot y'_{i1} & + f_2 \cdot y_{i2} \cdot y'_{i1} & + f_3 \cdot y'_{i1} & + \\ f_4 \cdot y_{i1} \cdot y'_{i2} & + f_5 \cdot y_{i2} \cdot y'_{i2} & + f_6 \cdot y'_{i2} & + \\ f_7 \cdot y_{i1} & + f_8 \cdot y_{i2} & + f_9 \cdot 1 \end{bmatrix}$$

and a factored form of

$$0 = \begin{bmatrix} y_{i1} \cdot y'_{i1} & y_{i2} \cdot y'_{i1} & y'_{i1} & y_{i1} \cdot y'_{i2} & y_{i2} \cdot y'_{i2} & y_{i2} & y_{i1} & y_{i2} & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{bmatrix} = \mathbf{m}_i^\intercal \cdot \mathbf{f}$$
where f is then calculated as the f is the f is the calculated as the f is the f i

The fundamental matrix (in vector form) **f** is then calculated as the null space (or closest possible equivalent) of the collection \mathbf{M}_n of all n pairs of points $\mathbf{m}_i^{\mathsf{T}}$ since

$$\mathbf{0} = egin{bmatrix} \mathbf{m}_1^\intercal \ \mathbf{m}_2^\intercal \ dots \ \mathbf{m}_n^\intercal \ \end{pmatrix} \cdot \mathbf{f} = \mathbf{M}_n \cdot \mathbf{f}$$

The vector \mathbf{f} that results in product $\mathbf{M}_i \cdot \mathbf{f}$ closest to the zero vector $\mathbf{0}$ can then be calculated through linear least squares through the singular value decomposition of \mathbf{M}_{i} . The vector \mathbf{f} is calculated as the singular vector of \mathbf{M} with the smallest singular value [1].

1.5 Camera Calibration Matrix

1.5.1 Pinhole Camera Model

A point \mathbf{x} in the camera-coordinate system of \mathbb{R}^3 is projected to the point $\tilde{\mathbf{y}}$ in \mathbb{R}^2 by means of the *pinhole camera model*. The set of all $\tilde{\mathbf{y}}$ are the result of rays passing through the *image plane* located at z = f, and converging at the *optical* center as shown in the figure below:

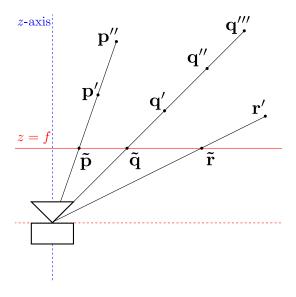


Figure 1.1. Pinhole CameraModel.

The exact location that $\tilde{\mathbf{y}}$ appears on the image plane is determined by utilizing the *similarity of triangles* between \mathbf{x} and $\tilde{\mathbf{y}}$. Specifically, we see that $y_1/f = x_1/x_3$ and $y_2/f = x_2/x_3$ rearranged gives $x_3 \cdot y_1 = f \cdot x_1$ and $x_3 \cdot y_2 = f \cdot x_2$. This lets us relate $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ by the *projection matrix* \mathbf{P} as

$$x_3 \cdot \tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{p} & 0 & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix}$$

Since, for a given $\tilde{\mathbf{x}}$ in camera space, the quantities $x_{1/3} = x_1/x_3$ and $x_{2/3} = x_2/x_3$ are invariant under the scale of $\tilde{\mathbf{x}}$, the location of $\tilde{\mathbf{y}}$ in the image plane depends only

on the ratios $x_{1/3}$ and $x_{2/3}$ and the quantity f. This yields a similar form, obtained from dividing by x_3 , of

$$\tilde{\mathbf{y}} = \frac{l}{x_3} \begin{bmatrix} \mathbf{P} & 0 \\ 0 \end{bmatrix} \cdot \tilde{\mathbf{x}} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1/x_3 \\ x_2/x_3 \\ x_3/x_3 \\ 1 \end{bmatrix}$$

This results in camera space points $\tilde{\mathbf{x}}$ with similar values of x_1 and x_2 , but containing infinitley large values of x_3 being mapped to the same point $\tilde{\mathbf{y}}$ in the image plane. This point $\mathbf{y} = 0$ is referred to as the principal point (or center of projection) in the image plane, and sometimes appears as a vanishing point for fixed values of x_1 and x_2 , but infinitley increasing values of x_3 .

1.5.2 Intrinsic Calibration Matrix

Points $\tilde{\mathbf{y}}$ given in the *image plane* have the same *units of measure* (or *scale*) as the points $\tilde{\mathbf{x}}$ in *camera space*. When dealing with digital images it's often more convenient to express *image coordinates* in terms of units such as *pixels* rather than real world units such as *inches*, *feet*, or *meters*. The matrix \mathbf{K} , where

$$\mathbf{K} = \begin{bmatrix} k_u & 0 & p_u \\ 0 & k_v & p_v \\ 0 & 0 & 1 \end{bmatrix}$$

is used to parameterize an image point $\tilde{\mathbf{u}}$ (in *pixels*), as a function of the coordinates \mathbf{x} in camera space and the camera specific parameters of horizontal pixel resolution k_u , vertical pixel resolution k_v , and principal point $\mathbf{p} = [p_x, p_y]^{\mathsf{T}}$. When combined with

the additional camera specific parameter of focal length f in the projection matrix \mathbf{P} , the result is

$$\mathbf{Q} = \mathbf{K} \cdot \mathbf{P}$$

$$= \begin{bmatrix} k_u & 0 & p_u \\ 0 & k_v & p_v \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} f \cdot k_u & 0 & p_u \\ 0 & f \cdot k_v & p_v \\ 0 & 0 & 1 \end{bmatrix}$$

which lets us relate the pixel point $\tilde{\mathbf{u}}$ to a point $\tilde{\mathbf{x}}$ in camera space as

$$\tilde{\mathbf{u}} = \mathbf{K} \cdot \tilde{\mathbf{y}} \cong \mathbf{K} \cdot \left[\begin{array}{c|c} \mathbf{P} & 0 \end{array} \right] \cdot \tilde{\mathbf{x}} = \left[\begin{array}{c|c} \mathbf{Q} & 0 \end{array} \right] \cdot \tilde{\mathbf{x}} = \mathbf{Q} \cdot \left[\begin{array}{c|c} \mathbf{I} & 0 \end{array} \right] \cdot \tilde{\mathbf{x}}$$

where \mathbf{Q} is referred to as the *camera calibration matrix*. Since \mathbf{Q} is dependant only on parameters *internal to the camera*, its also referred to as the *intrinsic calibration matrix*.

1.5.3 Extrinsic Calibration Matrix

Use of pinhole camera model by itself requires several assumptions being made, namely that the optical center \mathbf{C} occurs at the origin, and that the image plane is placed at z=f (is parallel to xy-plane). This implies the camera space is coincident with world space, or that the camera-coordinate and world-coordinate systems are one and the same. In simple scenes, this may not present a problem. In more complex scenes, including those with multiple cameras, this requires using the pinhole camera model in the context of an arbitary world space. This can be accomplished through the previously discussed change of reference.

As previously discussed, the change of reference from a world coordinate ${}^W\mathbf{x}$ to a camera coordinate ${}^C\mathbf{x}$ is calculated by the formula

$$^{C}\mathbf{x} = {}^{C}_{W}\mathbf{R} \cdot {}^{W}\mathbf{x} + {}^{C}_{W}\mathbf{t}$$

or homogeneously as

$$C\tilde{\mathbf{x}} = \begin{bmatrix} \frac{C}{W}\mathbf{R} & \frac{C}{W}\mathbf{t} \\ \hline 0 & 1 \end{bmatrix} \cdot {}^{W}\tilde{\mathbf{x}}$$
$$= {}^{C}_{W}\tilde{\mathbf{M}} \cdot {}^{W}\tilde{\mathbf{x}}$$

which allows us to project world coordinates ${}^W\tilde{\mathbf{x}}$ to the pixel coordinates $\tilde{\mathbf{u}}$ in the image plane as

$$\tilde{\mathbf{u}} \cong \mathbf{Q} \cdot \left[\mathbf{I} \mid 0 \right] \cdot {}^{C}\tilde{\mathbf{x}} = \mathbf{Q} \cdot \left[\mathbf{I} \mid 0 \right] \cdot {}^{C}_{W}\tilde{\mathbf{M}} \cdot {}^{W}\tilde{\mathbf{x}}$$

Since the matrix ${}^{C}_{W}\tilde{\mathbf{M}}$ is dependent only on the relative position and orientation of the camera (rather than the camera itself) it is commonly referred to as the extrinsic calibration matrix.

1.6 Essential Matrix

When coordinates from a reference frame are expressed as *normalized image* coordinates the range of possible NIC values in the corresponding image are given by the

CHAPTER 2

Rectification

REFERENCES

R. I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision,
 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

BIOGRAPHICAL STATEMENT

Jeff G. Wood was born in Evanston, Illinois, in 1981. He received his B.A. degree in Mathematics from Clarke College (now Clarke University) in Dubuque, Iowa, in 2003. Since that time, has worked as an actuary pricing Universal Life and Longterm Care insurance. He is a member of the Tau Beta Pi and Upsilon Pi Epsilon honor societies as well as the Association of Computing Machinary society.