2003

The mathematics of object recognition in machine and human vision

Sunyoung Kim

THE MATHEMATICS OF OBJECT RECOGNITION IN MACHINE AND HUMAN VISION

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

Sunyoung Kim

December 2003

THE MATHEMATICS OF OBJECT RECOGNITION IN MACHINE

AND HUMAN VISION

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Sunyoung Kim

December 2003

Approved by:

[redacted] 12/4/03

Chetan Prakash, Committee Chair — Date

[redacted]

John Sarli, Committee Member

[redacted]

Belisario Ventura, Committee Member

[redacted] J. T Hallett

Peter William, Chair
Department of Mathematics

Terry Hallett,
Graduate Coordinator
Department of
Mathematics

ABSTRACT

We present the framework of projective geometry. This framework allows us to study the reconstruction of three-dimensional structure and motion from sequences of two-dimensional images of the available features of an object. This theory is derived in the context of the affine camera, which preserves parallelism and generalizes the orthographic, scaled orthographic and para-perspective camera models.

We derive explicit recognition polynomials for the detection of rigid three-dimensional motion from two weak-perspective views by using Kontsevich's equation. In addition to detection of rigid motion, these polynomials can be used to recognize a given three-dimensional object from two-dimensional views, and in fact to reconstruct its depth coordinates.

We also provide some interesting theorems in linear algebra which arise as generalizations of theorems used in developing the recognition polynomials.

ACKNOWLEDGMENTS

I thank God for the intellect to explore the wonderful world of Mathematics. I thank my parents for give me a support and opportunity to study in United State of America. I thank Paul Amaya, who is assistant director of International Student Services, for helped me and understood me about my situation and legal status. Most of all, I thank my advisor Dr. Chetan Prakash for his patient and his interested way of approached Mathematics.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE

PROJECTIVE GEOMETRY

The study of projective geometry is related to the sort of sensors that machines and humans use for vision. It is known from geometric optics that any system of lenses can be approximated by a system that realizes a perspective projection of the world onto a plane. The simplest way to look at such a system is to look at it projectively. Here is the general definition of projective space of any dimension.

Definition: a point of an n dimensional real <u>projective space</u>, P^n is the set $R^n/\sim$, where $[x_1, x_2, \ldots, x_{n+1}] \sim [y_1, y_2, \ldots, y_{n+1}]$, is represented by an $(n+1)$ vector of real coordinates $x=[x_1, x_2, \cdots, x_{n+1}]$, where at least one of the x_i is nonzero. The numbers are called the <u>homogeneous</u> or <u>projective coordinates</u> of the point, and the vector x is called a coordinate vector. Note that the correspondence between points and coordinate vectors is not one to one.

Note that an $(n+1) \times (n+1)$ matrix A such that $\det(A) \neq 0$ defines a linear transformation on R^{n+1}, which may be interpreted as a projective isomorphism of P^n to itself.

Definition: i) $(n+1)\times(n+1)$ matrix A with $\det A \neq 0$ is called a collineation. ii) the set of collineations is a group and this group is also known as the projective group. iii) a projective basis is a set of $(n+2)$ points of P^n such that no $(n+1)$ of them are linearly dependent. Any point x of P^n can be described as a linear combination of the $(n+1)$ points of the standard basis:

$$\mathrm{x} = \sum_{i=1}^{n+1} x_i e_i \,,$$

where x_i are the projective coordinates in this basis.

For $n=1$, projective space P^1 is called the projective line; P^2 is called the projective plane; P^3 is called simply projective space. The space P^1 is the simplest of all projective spaces and many structures embedded in higher-dimensional projective spaces have the same structure as P^1. In P^1, a point on the line can be written as x = $x_1 e_1 + x_2 e_2$ with x_1 and x_2 not both equal to zero.

The space P^2 is used to model the image plane as a projective plane. A point in P^2 is defined by three numbers (x_1, x_2, x_3), not all zero. There are objects other by a triplet of numbers (u_1, u_2, u_3), not all zero. The points and the lines form coordinate vectors x and u defined up to a scale factor. The equation of the line is then $\sum_{i=1}^{3} u_i x_i = 0$ in the standard projective basis of P^2. Formally, there is no difference between points and lines in P^2. This is known as the principle of duality. Among all possible lines, the one whose equation is $x_3 = 0$ is called the line at infinity of P^2, denoted by l_∞. Each line L=(u_1, u_2, u_3) in the projective plane of the form of $\sum_{i=1}^{3} u_i x_i = 0$ intersects l_∞ at the point $(-u_2, u_1, 0)$, which is the point at infinity of the line L.

There is a structure of the projective plane that has numerous applications, especially in stereo and motion: Definition: A pencil of lines is the set of lines in P^2 passing through a fixed point. Any pencil of lines in P^2 is projectively isomorphic to the one-dimensional projective space P^1.

A point x in P^3, known as the projective space, is defined by four numbers (x_1, x_2, x_3, x_4), not all zero. There are objects other than just points and lines in P^2, such as planes. A plane is also defined as a four numbers (u_1, u_2, u_3, u_4), not all zero. The points and the planes form a coordinate vector x and u defined up to a scale factor. The equation of this plane is then $\sum_{i=1}^{4} u_i x_i = 0$ in the standard projective basis $(e_1, e_2, e_3, e_4, e_5)$ of P^3. A line is defined as the set of points that are linearly dependent on two points P_1 and P_2. Among all possible planes, the one whose equation is $x_4 = 0$ is called the plane at infinity or π_∞ of P^3. As in the case of the projective plane, it is often useful to think of the points in the plane at infinity as the set of directions of the underlying affine space. For example, the point of projective coordinates $[x_1, x_2, x_3, 0]$ represents the direction parallel to the vector $[x_1, x_2, x_3]$ and indeed it does not matter whether x_1, x_2 and x_3 are defined up to a scale factor, since the direction does not change.

CHAPTER TWO

CAMERA MODELS

There is a deep relationship between camera models and projective geometry. So I would like to look at camera models. A simple camera model can be considered from two standpoints. One is a geometric model and the other is physical model. In this paper, we are only interest in a geometric model.

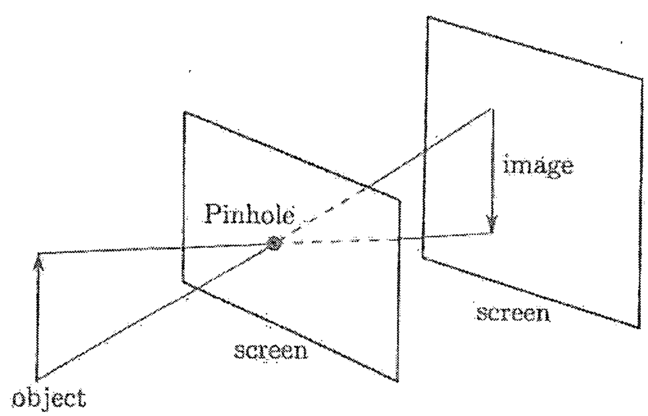

Figure 1. Image Formation in a Pinhole Camera

Let us consider the system consists of two screens. In the first screen, a small hole has been punched and through this hole some rays of light reach on the second screen. We can directly build a geometric model of the pinhole camera that consists of a plane R, called the retinal plane in which the image is formed through an operation called a perspective projection: The distance f from the optical center C to the retinal plane R is called the focal length of the camera.

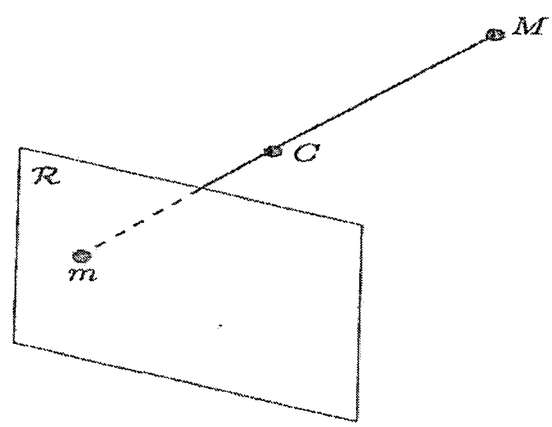

Figure 2. The Pinhole Camera Model

This is used to form the image m in the retinal plane of the three-dimensional point M as the intersection of the line $\overline{CM}$ with the plane R.

Let us take a look at camera model in further detail. We can choose the coordinate system X = (X_1,X_2,X_3) for the 3D space and x = (x_1,x_2) for 2D space, for example we can think of 2D space as the retinal plane. The coordinate system X = (X_1, X_2, X_3) is called the standard coordinate system of the camera. The relationship between image coordinates x and 3D space coordinates can be written in terms of a projection matrix P = $[P_{ij}]$,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \qquad (1)$$

where (x_1, x_2, x_3) and (X_1, X_2, X_3, X_4) are homogeneous coordinates related to x and X by x $= (\frac{x_1}{x_3}, \frac{x_2}{x_3})$ and

X $= (\frac{X_1}{X_4}, \frac{X_2}{X_4}, \frac{X_3}{X_4})$.

Thus a camera can be considered as a system that performs a linear projective transformation from the projective space P^3 into the projective plane P^2. We sometimes refer to the three-dimensional coordinate system X as the <u>world coordinate system</u>. Also, the camera can be considered as a system that depends upon both intrinsic and extrinsic parameters. Intrinsic parameters are those that do not depend on the position and orientation of the camera in space. There are four intrinsic parameters such as the scale factors α_u and α_v and the coordinates u_0 and v_0 of the intersection of the optical axis with the image plane. There are six extrinsic parameters, three for the rotation and three for the translation of the camera, which define the transformation from the world coordinate system to the coordinate system of the camera.

There is a special case of the projective camera called the <u>affine camera</u>. This affine camera can be written using equation (1) with $P_{31} = P_{32} = P_{33} = 0$:

$$\mathrm{P}_{aff} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & P_{34} \end{bmatrix} \tag{2}$$

It corresponds to a projective camera with its optical center at the plane at infinity; consequently, all projection rays are parallel. We can decompose P_{aff}.

$$\mathrm{P}_{aff} = \mathrm{C}\mathrm{P}_{//}\mathrm{G} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ G_{41} & G_{42} & G_{43} & G_{44} \end{bmatrix}$$

The 3×3 matrix C accounts for intrinsic camera parameters and represents a 2D affine transformation (hence $C_{31} = C_{32} = 0$). We assume there is no shear in the camera axes and use four parameters,

$$C=\begin{bmatrix} f\xi & 0 & o_x \\ 0 & f & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

where ξ is the camera aspect ratio, f the focal length and (o_x, o_y) the principal point (where the optic axis intersects the image plane). The 3×4 matrix $P_{//}$ performs the parallel projection operation, and the 4×4 matrix G accounts for extrinsic camera parameters, encoding the relative position and orientation between camera and the standard coordinate system. Therefore the affine camera covers the composed effects of i) a 3D affine transformation between world and camera coordinate system; ii) parallel projection onto the image plane; and iii) a 2D affine transformation of the image.

In terms of inhomogeneous image world coordinates, the affine camera is written

$$x = MX + t$$

where M= $\lfloor M_{ij} \rfloor$ is a 2×3 matrix with elements $M_{ij} = \frac{P_{ij}}{P_{34}}$ and

t = $\left(\frac{P_{14}}{P_{34}}, \frac{P_{24}}{P_{34}}\right)$. is a 2 vector giving the projection of the

origin of the world coordinate frame which is X = 0. A major property of the affine camera is that it preserves parallelism: lines that are parallel in the world remain parallel in the image.

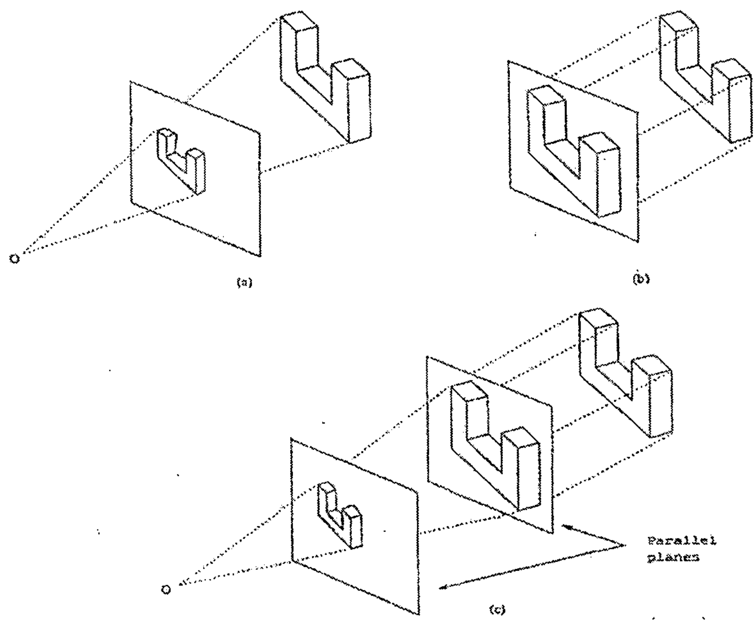

Figure 3. Camera Models: (a) perspective (all rays pass through a single projection point O, and the intersection of the ray star with the image plane generates the image); (b) orthographic(all rays are parallel, with the optical center O at infinity); (c) weak perspective(combined orthographic and perspective projection). For (b) and (c), parallel lines in the scene remain parallel in the image; this isn't truefor (a).

I would like to introduce some special cases of the affine camera, such as orthographic projection, weak perspective projection, and para-perspective projection cameras.

The orthographic projection camera is modeled by rays parallel to the optical axis projected orthographically onto the average depth plane $Z^c = Z^c_{ave}$.

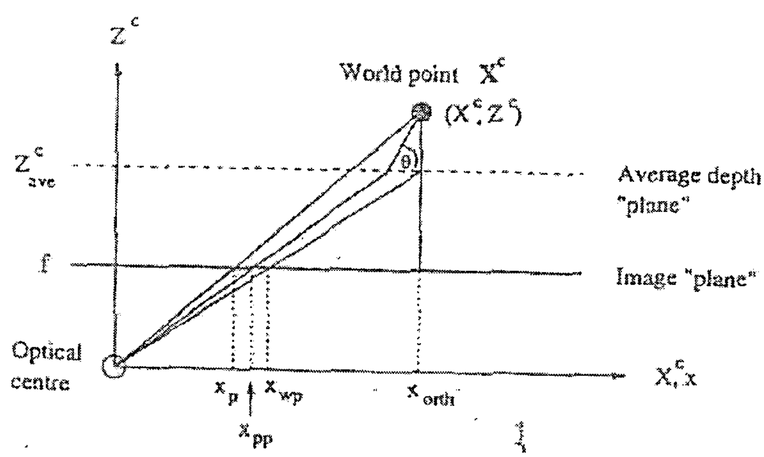

Figure 4. One-dimensional Image Formation
The image is the line $Z^c = f$.
For x_p (perspective) projection is along the ray connecting the world point X^c to the optical center. For x_{orth} (orthographic), projection is perpendicular to the image. For x_{pp} (para-perspective), X^c is first projected onto the average depth plane at angle θ, and then projected perspectively onto the image plane; x_{wp} (weak perspective) is a special case of x_{pp} with $\theta = 90^\circ$ (i.e.,orthographic projection onto the average depth plane).

Next, we look at the weak perspective projection camera. Consider the familiar camera centered perspective equations, where each point is scaled by its individual depth Z_i^c and all projection rays converge to the optical center:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{f}{Z_i^c} \begin{bmatrix} X_i^c \\ Y_i^c \end{bmatrix}$$

In above equation, $X^c = (X^c, Y^c, Z^c)$ denotes coordinates in the camera frame. When the camera field of view is small and the depth variation of the object $\Delta Z_i^c = Z_i^c - Z_{ave}^c$ is small compared to the average distance of the object from the camera Z_{ave}^c, the individual depths Z_i^c maybe approximated by Z_{ave}^c, giving a weak perspective or scaled orthographic camera:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{f}{Z_{ave}^c} \begin{bmatrix} X_i^c \\ Y_i^c \end{bmatrix}$$

We can say the weak perspective camera is a combination of the orthographic and perspective projection. Coordinates measured in a world coordinate system (X) are related to X^c by a rigid transformation $X^c = RX + T$, where R is a 3×3 rotation matrix with rows (R_1, R_2, R_3), and $T = (T_x, T_y, T_z)$ is a translation vector representation the origin of the world frame. The depth of a point X_i measured along the line of sight in the camera frame is then $Z_i^c = R_3^T X_i + T_z$. The center of the point set is denoted X_{ave} and the depth variation of the object is given by $\Delta Z_i^c = R_3^T (X_i - X_{ave})$. The weak perspective projection equations are then

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = C \begin{bmatrix} R_1 X + T_x \\ R_2 x + T_y \\ Z_{ave}^c \end{bmatrix} = C \begin{bmatrix} R_1^T & T_x \\ R_2^T & T_y \\ 0^T & Z_{ave}^c \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Next, we look at <u>para-perspective camera</u>. The para-perspective camera generalizes weak perspective case, such that projection of the scene point onto the average depth plane occurs parallel to the optic axis by projecting direction. Since the average depth plane remains parallel to the image plane, the perspective projection stage simply introduces a scale factor. The 1D case takes the form

$$x_{pp} = \frac{f}{Z_{ave}^c}(X^c - \Delta Z^c \cot\theta),$$

where θ denotes the angle between the projection direction and the positive X-axis. In the 2D case, the projection direction is described by two angles (θ_x, θ_y), where θ_x lies in the $X-Z$ plane and θ_y is the equivalent angle in the $Y-Z$ plane. Factoring in camera calibration parameters and the rigid transformation between the camera and world coordinate frames gives

$$\mathrm{P}_{pp} = \mathrm{C} \begin{bmatrix} 1 & 0 & -\cot\theta_x & \cot\theta_x \\ 0 & 1 & -\cot\theta_y & \cot\theta_y \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & Z^c_{ave} \end{bmatrix}.$$

CHAPTER THREE

EPIPOLAR LINES AND PLANES

The concept of an epipolar lines and plane is familiar in stereo and motion. The epipolar constraint relates a point in one image to a line in the other image depending on the intrinsic and extrinsic camera parameters.
First, we can take look at a very powerful constraint that arises from the geometry of stereo vision.

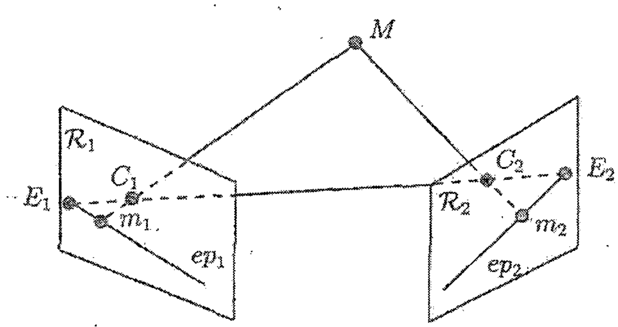

Figure 5. The Epipolar Geometry

Given m_1 in the retina plane R_1, all possible physical points M that may have produced m_1 are on the infinite half-line $\langle m_1, C_1 \rangle$, where C_i is optical center. All possible matches m_2 of m_1 in the plane R_2 are located on the image, through the second imaging system, of this infinite half-line. This image is an infinite half-line ep_2 going through the point E_2, which is the intersection of the line $\langle C_1, C_2 \rangle$ with the plane R_2.

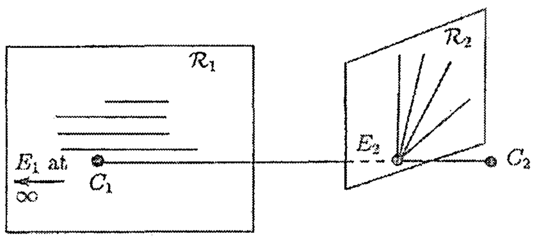

Figure 6. $\langle C_1, C_2 \rangle$ is Parallel to the Plane R_1: E_1 is ∞; the epipolar lines are parallel in the plane R_1 and at intersect at E_2 in the plane R_2.

E_2 is called the epipole of the second camera with respect to the first, and the line ep_2 is called the epipolar line of point m_1 in the retinal plane R_2 of the second camera. The corresponding constraint is that, for a given point m_1 in the plane R_1, its possible matches in the plane R_2 all lies on a line. The epipolar constraint is of course symmetric and for a point m_2 in the plane R_2, its possible matches in the retinal plane R_1 all lie on a line ep_1 through the epipole E_1, which is the intersection on the line $\langle C_1, C_2 \rangle$ with the plane R_1. The lines ep_1 and ep_2 are the intersections of the plane C_1MC_2, called the epipolar plane defined by M, with the planes R_1 and R_2, respectively.

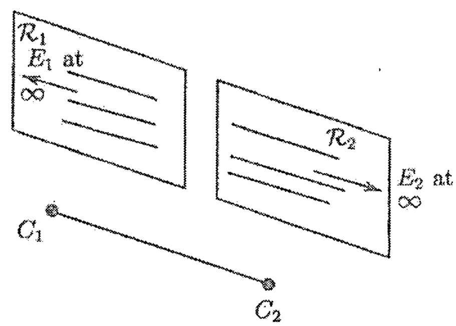

Figure 7. $\langle C_1, C_2 \rangle$ is Parallel to the Planes R_1 and R_2: E_1 and E_2 are at ∞; the epipolar lines are parallel in both planes R_1 and R_2.

When the plane R_1 or the plane R_2, or both, are parallel to the line $\langle C_1, C_2 \rangle$, one or both epipoles go to infinity and the epipolar lines in one plane or both become parallel. The situation where both planes are parallel to the line $\langle C_1, C_2 \rangle$ is often assumed because of its simplicity. Let's compute the epipolar geometry. In this section, the notation ~ is to indicate projective quantities. For example, $\tilde{x}$ denotes a projective coordinate vector, which is defined up to a multiplicative nonzero scalar, and x denotes a vector of R^n. Let $\tilde{m}_1 = \tilde{P}_1 \tilde{M}$ and $\tilde{m}_2 = \tilde{P}_2 \tilde{M}$ be two cameras. The coordinates of the two optical centers, C_i (i= 1,2), in the world reference frame, are obtained by solving the following two systems of linear equations: $\tilde{P}_i \tilde{M} = 0$, where i = 1,2.

Since each epipole E_i is the image by the ith camera of the other camera's optical center C_j ($j \neq i$), the image coordinates of the epipoles E_i are obtained by applying matrices $\tilde{P}_i$ to the vectors $\tilde{C}_j$ (i,j= 1,2, $i \neq j$).

Now, I like to show how, for a given point m_1 in the plane R_1, the corresponding epipolar line ep_1 can be computed. We need to points to determine a line. One of them is the epipole E_2, which is given by $\tilde{e}_2 = \tilde{P}_2 \begin{bmatrix} -P_1^{-1}\tilde{p}_1 \\ 1 \end{bmatrix}$, and another point is the point at infinity of the optical ray $\langle C_1, m_1 \rangle$. The image m_2 of this point in the second retinal plane is given by $\tilde{m}_2 = P_2 P_1^{-1} \tilde{m}_1$. A projective representation of ep_2 is the cross product $\tilde{e}_2 \wedge \tilde{m}_2$. The cross product $\tilde{e}_2 \wedge \tilde{m}_2$ can be written a $F\tilde{m}_1$ where F is a 3×3 matrix. If we let E_2 be the 3×3 antisymmetric matrix representing the cross product with $\tilde{e}_2$. We have $F = \tilde{E}_2 P_2 P_1^{-1}$. Any pixel m_2 on the epipolar line ep_2 of m_1 satisfies the equation $\tilde{m}_2^T F \tilde{m}_1 = 0$. It shows in particular that the roles of m_1 and m_2 are symmetric and that the epipolar line of a pixel m_2 in the first retinal plane is represented by the vector $F^T \tilde{m}_2$.

CHAPTER FOUR

RECOGNITION OF RIGID THREE DIMENSIONAL MOTION FROM SEQUENECS OF TWO DIMENSIONAL IMAGES

A <u>recognition polynomial</u> is a polynomial in the image data (i.e., the coordinates of the points in the given view) that evaluates to zero when the image data are consistent with those from rigid motions of a given 3-D object. Using a Kontsevich's approach, we can derive explicit recognition polynomials for the detection of rigid 3-D motion from two weak-perspective views. I would like to introduce Kontsevich's derivation of the two-view rigidity constraint in weak perspective projection. Kontsevich used some of the same geometric ideas as in Koenderink and van Doorn, notably the decomposition of rigid rotation into a rotation about the viewing direction and a rotation about an axis in the image plane. The motions we consider are rigid rotations in 3-D space or translations along the viewing direction that followed by uniform scaling so we can ignore translations parallel to and reflections through the viewing plane.

Suppose we have an object that has five distinguished points, labelled "0, 1, 2, 3, and 4." Let r_j be the displacement in R^3 between point 0 and point j, in the first view (with j=1, 2, 3, 4). Then r'_j is the corresponding displacement, in R^3, in the second view. And let π be the orthographic projection to the image plane, then the projected displacement can be written

$$p_j = \pi(r_j), \quad p'_j = \pi(r'_j).$$

The fact is given an image plane, any rigid rotation in R^3 can be thought of as a composition of two rotations. The first is a rotation about a unit axis vector v parallel to the image plane and second is a rotation about an axis vector perpendicular to the image plane. The second rotation takes the first axis vector v to a new unit vector v^1 in the image plane. If the uniform scaling factor is s, then we can let $v' = \frac{1}{s}v^1$.

Consider the first of the decomposed rotations, around the axis v. Then the respective projections of r_j and p_j onto v are equal and the second rotation takes v to v^1. If we let r_j^1 be the edge between 0 and j after that rotation, and denote $\pi(r_j^1) = p_j^1$, then the respective projections of r_j^1 and p_j^1 onto v^1 are equal, and that these projections are the same as those rigid of r_j and p_j onto v. Thus $p_j \cdot v = p_j^1 \cdot v^1$. Finally, consider the scaling. Since π is orthographic, the scaling factor of s results in $p_j' = sp_j^1$ and, in equation $v' = \frac{1}{s}v^1$, we can arrive at the linear Kontsevich equations:

$$p_j \cdot v - p_j' \cdot v' = 0,$$

with $\|v\|=1$, $\|v'\| = \frac{\|v\|}{s}$, $\|p'\| = s\|p\|$.

The equation $p_j \cdot v - p_j' \cdot v' = 0$ is a homogeneous system of four linear equations in the four unknown coordinates c_1, c_2, c_1', c_2', where

$$v = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad v' = \begin{bmatrix} c_1' \\ c_2' \end{bmatrix}.$$

If we let $p_j = \begin{bmatrix} x_j \\ y_j \end{bmatrix}$, $p_j' = \begin{bmatrix} x_j' \\ y_j' \end{bmatrix}$, the condition for there to be a nontrivial solution to equation $p_j \cdot v - p_j' \cdot v' = 0$ is that the coefficient matrix have rank less than 4. Thus v and v' can be known only up to an overall scale factor; choosing a solution for which $\|v\|=1$ allows us to compute the scale factor s from the equation $\|v'\| = \frac{\|v\|}{s}$. This completes the exposition of Kontsevitch's two-view derivation.

Now, take look at how a recognition polynomial arises for the two-view weak perspective recognition of rigid motion. This is a polynomial in the 16 data values $\{x_j, y_j, x'_j, y'_j\}_{j=1,\ldots,4}$ which must evaluate to zero for there to be a rigid interpretation. The condition that equation $p_j \cdot v - p'_j \cdot v' = 0$ has a nontrivial solution is then that the determinate of the coefficient matrix vanish. Ignoring the negative signs in the last two columns of this matrix, we get:

$$\det\begin{Bmatrix} x_1 & y_1 & x'_1 & y'_1 \\ x_2 & y_2 & x'_2 & y'_2 \\ x_3 & y_3 & x'_3 & y'_3 \\ x_4 & y_4 & x'_4 & y'_4 \end{Bmatrix} = 0$$

Therefore the two-view recognition polynomial is the determinant in above.

The recognition polynomial is a polynomial in 16 variables, corresponding to the image plane coordinates of the weak perspective projections of four feature points in each of the two views. To use the polynomial for the detection of rigid motion, given two weak perspective views of an actual object with, say, five feature points, an observer can choose one of the points. Then, if the image coordinates of the four remaining feature points in the two views satisfy the recognition polynomial, we may infer that the 3-D motion is rigid with probability one. To use the polynomial for the recognition of a given object, suppose we know the image plane coordinates of five feature points on the object. We can use these coordinates to assign numerical values to those variables in the polynomial which correspond to the first view. This gives a reduced polynomial in eight variables. Then, given a view of a novel object with four feature points, we can plug the image plane coordinates of these points into the reduced polynomial, and if the result is 0, we may infer that the novel object coincides with the memory object with probability one.

Theorem: Let A be a 4×4 matrix with rank 3. Let $\overline{A}$ be the 3×4 matrix obtained from A by deleting the last row. Let $\overline{A}^i$ be the ith column of $\overline{A}$. Then the vector $C=(c_1,c_2,c_1',c_2')^T$ defined by

$$c_1 = -\det(\overline{A}^2, \overline{A}^3, \overline{A}^4)$$

$$c_2 = \det(\overline{A}^1, \overline{A}^3, \overline{A}^4)$$

$$c_1' = -\det(\overline{A}^1, \overline{A}^2, \overline{A}^4)$$

$$c_2' = \det(\overline{A}^1, \overline{A}^2, \overline{A}^3)$$

is a nontrivial solution to $Ax = 0$. Moreover, if both of the 2×2 diagonal minors of A are nonsingular, then in this solution both vectors $(c_1, c_2)^T$ and $(c_1', c_2')^T$ are nonzero.

Lemma: Let $A(k)$ be the 3×4 matrix obtained from A by deleting the kth row. Let $C(k)$ be the vector $(\tilde{c}_1, \tilde{c}_2, \tilde{c}_1', \tilde{c}_2')^T$ obtained from $A(k)$ as $C = (c_1, c_2, c_1', c_2')^T$ is from $\overline{A}$. Then $\widetilde{C}$ is a nonzero scalar multiple of C.

CHAPTER FIVE

RECONSTRUCTION OF THREE DIMENSIONAL MOTION FROM SEQUENCES OF TWO DIMENSIONAL IMAGES

Kontsevich performed a change of coordinates so as to simplify the depth reconstruction. To this end, define

$$\hat{e}_v = (c_1^2 + c_2^2)^{-\frac{1}{2}} (c_1 \quad c_2)^T$$

$$\hat{e}_u = (c_1^2 + c_2^2)^{-\frac{1}{2}} (-c_2 \quad c_1)^T$$

$$\hat{e}_{v'} = (c_1^2 + c_2^2)^{-\frac{1}{2}} (c_1' \quad c_2')^T$$

$$\hat{e}_{u'} = (c_1^2 + c_2^2)^{-\frac{1}{2}} (-c_2' \quad c_1')^T$$

Thus $\hat{e}_u, \hat{e}_v, \hat{e}_z$ (where $\hat{e}_z$ is the unit vector perpendicular to the image plane) forms an orthogonal coordinate system, as does $\hat{e}_{u'}, \hat{e}_{v'}, \hat{e}_z$.

$$p_j \cdot \hat{e}_v = p_j' \cdot \hat{e}_{v'}$$

Now define the u coordinates in the new systems:

$$p_{j,u} = \overset{u}{e}_u \cdot p_j = \frac{-c_2 x_j + c_1 y_j}{(c_1^2 + c_2^2)^{\frac{1}{2}}}$$

$$p'_{j,u'} = \overset{u}{e}_{u'} \cdot p'_j = \frac{-c_2 x'_j + c_1 y'_j}{(c_1^2 + c_2^2)^{\frac{1}{2}}}$$

Note that the v,v' coordinates, defined similarly, are equal for corresponding points. Now let the z-coordinate of the edge r_j be $r_{j,z}$. For each angle α, we have

$$p'_{j,u'} = \begin{pmatrix} \cos\alpha & -\sin\alpha \end{pmatrix} \begin{pmatrix} p_{j,u} \\ r_{j,z} \end{pmatrix}$$

$$= p_{j,u}\cos\alpha - r_{j,z}\sin\alpha$$

Hence

$$r_{j,z} = p_{j,u}\cot\alpha - p'_{j,u'}\csc\alpha \,, \quad \forall\alpha \,.$$

Letting $\lambda = \cot\alpha, \mu = \csc\alpha$, we have:

$$r_{j,z} = \lambda p_{j,u} - \mu p'_{j,u'}$$

$$\mu^2 - \lambda^2 = 1 \,.$$

CHAPTER SIX

SOME THEOREMS IN LINEAR ALGEBRA

Euler's Theorem: Any $R \in SO(3)$ can be written $R = R_\phi^z R_\theta^y R_\psi^z$, where (ϕ, θ, ψ) are Euler angles of R and R_ϕ^Z is a rotation through angle ϕ around the positive Z-axis.

Proof: Let n be the north pole of S^2. If $B \in SO(3)$ with $Bn = Rn$, set $C = RB^{-1}$. Then $C \in SO(3)$ and $R = CB$; $CBn = Bn$. i.e., C is a rotation about Bn (and Rn).

Now $\exists \theta$ such that $R_\theta^y n$ has the same latitude as Rn and $\exists \phi$ such that $R_\phi^z (R_\theta^y n) = Rn$. ($\therefore$ we can set $R_\phi^z R_\theta^y = B$)

i.e., $R = CR_\phi^z R_\theta^y$ where C is a rotation about $Rn = Bn$.

Lemma: If A is a rotation about $\hat{r}$ and $B\hat{r} = \hat{u}$, for some rotation B then BAB^{-1} is a rotation about $\hat{u}$.

Proof:
$$
\begin{aligned}
(BAB^{-1})u &= BAB^{-1}B\hat{r} \\
&= BAI\hat{r} \\
&= BA\hat{r} \\
&= B\hat{r} \\
&= \hat{u}
\end{aligned}
$$

As a consequence, we can say that any rotation C about $u = Br$ is of the form BAB^{-1}, where A is a rotation about r, since by the lemma, $A = B^{-1}CB$ is a rotation about $r = B^{-1}u$. Then $C = BAB^{-1}$.

Now with $R = CB$, $(B = R_\phi^z R_\theta^y)$ and C is a rotation about $Bn(= Rn)$, we have that $C = BDB^{-1}$ with $D = R_\psi^z$.

Finally $R = BDB^{-1}B$

$$= BD$$

$$= R_\phi^z R_\theta^y R_\psi^z \quad \blacksquare$$

Theorem: Let A be a 3×3 matrix with rank 2. Let $\overline{A}$ be the 3×3 matrix obtained from A by deleting the last row. Let $\overline{A}^i$ be the ith column of $\overline{A}$.

Then the vector $C = (c_1, c_2, c_3)^T$ defined by

$$c_1 = -\det(\overline{A}^2, \overline{A}^3)$$

$$c_2 = \det(\overline{A}^1, \overline{A}^3)$$

$$c_3 = -\det(\overline{A}^1, \overline{A}^2)$$

is a nontrivial solution to $Ax = 0$ and $c_i \neq 0$.

Proof: Expanding the determinant of A off the nth row, we get

$-a_{31}\det(\overline{A}^2, \overline{A}^3) + a_{32}\det(\overline{A}^1, \overline{A}^3) - a_{33}\det(\overline{A}^1, \overline{A}^2) = 0$, since A has rank less than n. That is, $a_{31}\, c_1 + a_{32}\, c_2 + a_{33}\, c_3 = 0$.

Thus the vector $C = (c_1, c_2, c_3)^T$ is orthogonal to the row A_3.

By the theory of determinants C is orthogonal to all the rows of A, i.e., is a solution to $Ax = 0$.

Moreover, since A has rank exactly 2, C can not be the zero vector. ∎

Let's check if $AC = 0$.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \stackrel{?}{=} 0$$

$$\begin{bmatrix} a_{11}c_1 + a_{12}c_2 + a_{13}c_3 \\ a_{21}c_1 + a_{22}c_2 + a_{23}c_3 \\ a_{31}c_1 + a_{32}c_2 + a_{33}c_3 \end{bmatrix} = \begin{bmatrix} -a_{11}\det(\overline{A}^2, \overline{A}^3) + a_{12}\det(\overline{A}^1, \overline{A}^3) - a_{31}\det(\overline{A}^1, \overline{A}^2) \\ -a_{21}\det(\overline{A}^2, \overline{A}^3) + a_{22}\det(\overline{A}^1, \overline{A}^3) - a_{32}\det(\overline{A}^1, \overline{A}^2) \\ -a_{31}\det(\overline{A}^2, \overline{A}^3) + a_{32}\det(\overline{A}^1, \overline{A}^3) - a_{33}\det(\overline{A}^1, \overline{A}^2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

If 2×2 minors of A are nonsingular than in this solution vectors c_i's are nonzero.

Let $c_1 = 0$, then from $AC = 0$ we infer

$$\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Since $C \neq 0$ and we are assuming $c_1 = 0$, $(c_2, c_3)^T$ is a nontrivial solution to 3×3 system. This contradicts 2×2 minors of A are nonsingular. Therefore $c_1 \neq 0$. A similar argument holds for $c_2 = 0$ and $c_3 = 0$. Thus c_i's are nonzero

Theorem: Let A be a n×n matrix with rank n-1. Let $\overline{A}$ be the (n-1)×n matrix obtained from A by deleting the last row. Let $\overline{A}^i$ be the ith column of $\overline{A}$. Then the vector $C =$ $(c_1, c_2, c_3, \ldots, c_n)^T$ defined by

$$c_1 = -\det(\overline{A}^2, \overline{A}^3, \overline{A}^4, \ldots, \overline{A}^n)$$

$$c_2 = \det(\overline{A}^1, \overline{A}^3, \overline{A}^4, \ldots, \overline{A}^n)$$

$$c_3 = -\det(\overline{A}^1, \overline{A}^2, \overline{A}^4, \ldots, \overline{A}^n)$$

$$\vdots$$

$$c_i = (-1)^i \det(\overline{A}^1, \overline{A}^2, \ldots, \overline{A}^{i-1}, \overline{A}^{i+1}, \ldots, \overline{A})$$

$$\vdots$$

$$c_n = (-1)^n \det(\overline{A}^1, \overline{A}^2 \overline{A}^3, \ , \ldots \ , \overline{A}^{n-1})$$

is a nontrivial solution to $Ax = 0$.

Proof: Expanding the determinant of A off the nth row, we get

$$-a_{n1}\det(\overline{A}^2, \overline{A}^3, \ldots \ , \overline{A}^n) + a_{n2}\det(\overline{A}^1, \overline{A}^3, \ , \ \ldots, \overline{A}^n) + \ldots +$$

$(-1)^n a_{nn} \det(\overline{A}^1, \overline{A}^2, \ldots, \overline{A}^{n-1}) = 0$, since A has rank less than n.

That is, $a_{n1} c_1 + a_{n2} c_2 + \ldots + a_{nn} c_n = 0$.

Thus the vector $C = (c_1, c_2, c_3, \ldots, c_n)^T$ is orthogonal to the row A_n. By the theory of determinant C is orthogonal to all the rows of A, i.e., is a solution to $Ax = 0$.■

Let's check if $AC = 0$.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \ldots a_{1n} \\ a_{21} & a_{22} & a_{23} \ldots a_{2n} \\ \vdots & & \\ a_{n1} & a_{n2} & a_{n3} \ldots a_{nn} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \overset{?}{=} 0$$

$$\begin{bmatrix} a_{11}c_1 + a_{12}c_2 + a_{13}c_3 \ldots + a_{1n}c_n \\ a_{21}c_1 + a_{22}c_2 + a_{23}c_3 \ldots + a_{2n}c_n \\ \vdots \\ a_{n1}c_1 + a_{n2}c_2 + a_{n3}c_3 \ldots + a_{nn}c_n \end{bmatrix} =$$

$$\begin{bmatrix} -a_{11} \det(\overline{A}^2, \overline{A}^3, \ldots, \overline{A}^n) + a_{12} \det(\overline{A}^1, \overline{A}^3, \ldots, \overline{A}^n) + \ldots + (-1)^n a_{1n} \det(\overline{A}^1, \overline{A}^2, \ldots, \overline{A}^{n-1}) \\ -a_{21} \det(\overline{A}^2, \overline{A}^3, \ldots, \overline{A}^n) + a_{22} \det(\overline{A}^1, \overline{A}^3, \ldots, \overline{A}^n) + \ldots + (-1)^n a_{2n} \det(\overline{A}^1, \overline{A}^2, \ldots, \overline{A}^{n-1}) \\ \vdots \\ -a_{n1} \det(\overline{A}^2, \overline{A}^3, \ldots, \overline{A}^n) + a_{n2} \det(\overline{A}^1, \overline{A}^3, \ldots, \overline{A}^n) + \ldots + (-1)^n a_{nn} \det(\overline{A}^1, \overline{A}^2, \ldots, \overline{A}^{n-1}) \end{bmatrix}$$

$= (0, 0, \ldots, 0)^T$.

Theorem: Let A be a 4×4 matrix with rank 2. Then we provide a general algorithm for finding a basis of the null space of A (i.e, for solutions to $Ax=0$); Pick any pair of non-trivial solution vectors given below.

Proof: Let B be the row-reduced matrix obtained from A. Rearrange if necessary so that the last two rows of B are zero.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Let $\overline{B}$ be the 3×4 matrix obtained from B by deleting its last row. Let $\overline{B}^i$ be the ith column of $\overline{B}$. The null set of A and B are identical (subject to the rearranging described above), so it suffices to find solutions to $Bx = 0$. Solutions can be found as follows.

If we let $x_4 = 0$, then the solution will be

$$\begin{aligned} x_1 &= \det(\overline{B}^2, \overline{B}^3) = a_{12}a_{23} - a_{13}a_{22} \\ x_2 &= -\det(\overline{B}^1, \overline{B}^3) = -(a_{11}a_{23} - a_{13}a_{21}) \\ x_3 &= \det(\overline{B}^1, \overline{B}^2) = a_{11}a_{22} - a_{12}a_{21}) \end{aligned}$$

If we let x_3=0, then the solution will be

$$\begin{aligned} x_1 &= -\det(\overline{B}^2, \overline{B}^4) = a_{12}a_{24} - a_{14}a_{22} \\ x_2 &= \det(\overline{B}^1, \overline{B}^4) = -(a_{11}a_{24} - a_{14}a_{21}) \\ x_4 &= -\det(\overline{B}^1, \overline{B}^2) = a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

If we let $x_2 = 0$, then the solution will be

$$x_1 = -\det(\overline{B}^3, \overline{B}^4) = a_{13}a_{24} - a_{14}a_{23}$$
$$x_3 = \det(\overline{B}^1, \overline{B}^4) = -(a_{11}a_{24} - a_{14}a_{21})$$
$$x_4 = -\det(\overline{B}^1, \overline{B}^3) = a_{11}a_{23} - a_{13}a_{21}$$

If we let $x_1 = 0$, then the solution will be

$$x_2 = -\det(\overline{B}^3, \overline{B}^4) = a_{13}a_{24} - a_{14}a_{23}$$
$$x_3 = \det(\overline{B}^2, \overline{B}^4) = -(a_{12}a_{24} - a_{14}a_{22})$$
$$x_4 = -\det(\overline{B}^2, \overline{B}^3) = a_{12}a_{23} - a_{13}a_{22}$$

In each case it is verified by direct computation that the given vector is a solution.

The question is: are any of these vectors non-trivial?

Since B has rank 2, at least one 2×2 sub-matrix of the matrix consisting of the first two rows has to be nonzero (since the matrix has rank 2 and any other 2×2 sub-matrix has automatically zero determinants).

By inspection of the solutions given above, we see that this particular 2×2 determinant, whatever it is, appears in two distinct vectors, both of which are therefore non-trivial.

Furthermore, these two vectors must be linearly independent, since the zero entries are different from each other. Thus they form a basis of the two-dimensional solution space to $Bx = 0$ and therefore also to $Ax = 0$. ■

REFERENCES

1. Oliver Faugeras, Three-dimensional computer vision: a geometric viewpoint. Massachusetts Institute of Technology press, 1999

2. M. Brady, L. s. Shapiro, and A. Zisserman, "Three-dimensional motion recovery via affine epiploar geometry." Internation Jounal of Computer Vision, 16, 147-182 (1995)

3. B. M. Bennett, D. D. Hoffman, and C. Prakash, "Recognition of objects and detection of rigid motion using weak-perspective polynomials." 1999