Notes on Multi-view Geometry in Computer Vision

Yufei Ye

April 5, 2023

intro

uncalibrate problems hasve achieved great progress in recent decade:

- Given 2 iamges, compute matches between the images, and the 3D position of the points that generate these matches and the cameras that generate the images.
- Given three images, and no other information, similarly compute the matches between images of points and lines, and the position in 3D of these points and lines and the cameras.
- Compute the epipolar geometry of a stereo rig, and trifocal geometry of a trinocularrig, without requiring a calibration object.
- Compute the internal calibration of a camera from a sequence of images of natural scenes (i.e. calibration "on the fly".

Why these achievement?

- the error that hsould be minimized in over-determined system
- robust estimation

Solved problem:

- Estimation of the multifocal tensors from image point correspondences, particularly the fundamental matrix and trifocal tensors (the quadrifocal tensor having not received so much attention).
- Extraction of the camera matrices from these tensors, and subsequent projective reconstruction from two, three and four views.

More to learn:

- bundle adjustment to solve more **general** reconstruction problems.
- Metric (Euclidean) reconstruction given minimal assumptions on the camera matrices.
- Automatic detection of correspondences in image sequences, and elimination of outliers and false matches using the multifocal tensor relationships.

Chapter 0

Intro - a Tour of Multiple View Geometry.

0.1 Projective geo

Projectvie Space: just expansion of Euclidean space. Euclidean space + points at infinity −> projective space. Euclidean space is troublesome in one major respect: keeps making exception, e.g. parrellel lines.

Coordinates: (homogeneous vectors) represnte points by *equivalence class* of coordinate triples. $(kx, ky, k), k \neq 0$!!!. Points at Infinity is repreensted as $(x, y, 0)$.

In Euclidean / projective geometry, all points are the same. It is just accident that particular point are selected to be original / points at infinity have final coordinate 0.

Transformation of Euclidean / projective space is represented by matrix multiplication. However, in projective space, points at infinity are not preserved, it could land on anywhere.

For practical reason, we sometimes treat line at infinity special / equal...

0.1.1 from Projective geo to euclidean space

Affine Geometry : map "line at infinity" to "line at infinity". The geometry of the projective plane and a distinguished line is known as affine geometry and any projective transformation that maps the distinguished line in one space to the distinguished line of the other space is known as affine tranformation.

Euclidean Geometry: Specify line at infinity and two circular points. length ratio / angle can be defined in temrs of ciruclar points

Absolute Conic: related to camera calibration.

in 2D, circular points: $(1, \pm i, 0)$ lie on *every* circle (in regular Euclidean coordinate). circle with homogeneous coordinate (x, y, w) is

$$(x - aw)^2 + (y - bw)^2 = r^2w^2$$

. 5 points define an ecllipes, 3 points + 2 circular points define a circle.

in 3D, all sphere intersect at a second-degree curve(conic) on the plane at infinity, It is called absolute conic. Angle can defined in terms of the absolute conic in ANY ARBITRARY coordinate system.

0.2 Camera Projections

. Mapping from P^3 to P^2. Generally,

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = P_{3\times 4} \begin{bmatrix} X \\ Y \\ Z \\ T \end{bmatrix}$$

If all points lie on a plane, say choose this plane as $z = 0$, then the linear mappign *reduces* to

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = H_{3\times 3} \begin{bmatrix} X \\ Y \\ T \end{bmatrix}$$

Cameras as points All points in a ray passing through the center of projection projects to the same point in an image. Thus, allpoints along such a ray is equal. Thus, the set of all image points = the set of rays through the camera center. Thus, the rays themselves are represented by homogeneous coordinates, specifically, 2-d space of rays in $\mathbb{P}^2$.

camera center as essence. All that is important is the camera cnter, for it alone deternmines the set of rays forming the image. Any two images taken from the same point in space are **projectively equivalent**. (can be tranforemd only by projective transformation without any infromation about 3D points.)

Calibrated cameras. For a camera not located on the plane at infinity, the plane at infinity in the world maps one-to-one onto the iamge plane. it is because any point in the image defines a ray in space that meets the plane at infinity in a single point. Thus**emmmmmm????**, the plane at infinity in the world does not tell us anything new about hte image.

However, the absolute conic does. It projects to **IAC** (Image of the Absolute Conic). location of IAC is known $\equiv$ calibrated camera.

There is dual: The angle between two rays / lines in the world, is determined by where they meet the plane at inf, relative to the absolute conic. The projective relationship between the two image points and ω is exactly equal to the relationship between the intersections of the back-projected rays with the plane at infinity, and ω_∞.

comment: *Analogy in DDG, It's like you define a metric g <,>, in the real space (3D world) but you need to work on another manifold (image). can we further do better?*

If we wanna reconstruct the scene from single view, we normally determin the line at infinity for some observed planes in the image. And upgrade the affine reconstruction to Euclidean by assuming angles observed in the scene (most particularly orthogonal lines / planes.)

0.3 Reconstruction from more than one view

comment: *normally the usual input is a set of point correspondences!! but how do we get the correspondence? how robust are they? how accurate / noisy should the correspondence? in what scale??*

Ambiguity. the ambiguity in the reconstruciotn is expressed by **projective transformations emmmmm????**Because:

$$P_j X_i = (P_j H^{-1})(H X_i)$$

. This projective ambiguity is *unavoidable* for 2-view camera. Up to that, the scene can be reconstructed for ≤ 7 points which do not lie in *critical configurations.*

correspondence: $x_i <-> x_i'$

fundamental matrix & fundamental-matrix method. basic tool for recon of points sets from 2-views!!!

$$x_i'^T F x_i = 0, rank(F) = 2, F \in 3 \times 3$$

.

comment: *almost forget!!! Fundamental matrix is the basic algebraic entity!!*

0.4 Three-view Geometry

Trifocal tensor. $3\times3\times3$ tensor, which relate the coordiates of corresponding in 3 views. It is determined by 3 camera matrices, and determines them, up to projective transformations.

$$\sum_{ijk} x^i l_j^{'} l_k^{''} T_i^{jk} = 0\textbf{emmmmm}????$$

The constraints: interanl constraints.

correspondence:

$$x \leftrightarrow l' \leftrightarrow l^{''}$$

pros of 3-vew over 2 view 1) allow mixture of line and point correspondences, instead of just point-point correspondence. 2) stability.

comment: *if we have more views, say view from category, can we further relax teh correspondence requirement??? how messy can it be?*

0.5 Four view geometry and n-view reconstruction.

In general, quadrifocal is the most. the tensor method does not extend to more than four views.

relax: affine camera. + a set of points are visiable for n views, then, the factorization algorithm can be used.

relax: projective camera. but also requires all points to be visible in all images.

relax: various assumptions.

dominant general method: bundle adjustment. relation to maximum likelihood solution!!

0.6 Euclidean reconstruction

Till now, the camera are all un-calibrated. I fwe have complete calibration of each of the camera, then some ambiguity can be removed.

the distortion / equality in projective space will arise with same probablity. To human, it is not correct in Euclidean sense. Knowledge of the camera calibration is equivalent to being able to determine the Euclidean structure of the scene.

Chapter 1

Projective Geometry and Transformation of 3D

Gemetric distortion arises when a plane is imaged by a *perspective camera*. The imaging process can be modeled by projective geometry.

comment: *1. how to rectify planes? how to remove perspecitve distortion from an image? what is perspecitve camera?*

1.1 Planar Geometry

geometry	algebra
point	vector
line	vector
conic section / conic	symmetric matrix **emmmmm????**

1.2 2D projective plane

homogenieous representation of lines. $ax + by + c = 0$ represents a line, $(a, b, c)^T \equiv k(a, b, c)^T, k \neq 0$. The set of equivalence classes of vectors in $\mathbb{R}^3 + (0, 0, 0)^T$ gives us projective space $\mathbb{P}^2$.

homogenieous representation of points. a point $x = (x, y)^T$ lines on the line (a, b, c) is the same as $(x, y, 1) \cdot (a, b, c)^T = 0$. Homogeneous vector representative of a point

Both line and point in 2-space has 2-dof. **Result** *point x lines on the line l iff $x^T l = 0$*

3-space	P^2
rays	point
plane	line
two non-identcal rays lie on exactly one plane	two points define a line
two non-identical planes intersect in one ray	two lines interesect at a point
image plane at $x_3 = 1$	homogeneous coordinate to represent point

Result *intersection of two lines $l, l^{'}$ is the point $x = l \times l^{'}$.* They are so convinient thanks to homogeneous expression. proof:

$$l.(l \times l^{'}) = l^{'}.(l \times l^{'})$$

. Let $x = l \times l^{'}$, then $l.x = l^{'}.x = 0$, x is at two lines.

Result *a line joining two points x, $x^{'}$ is $l = x \times x^{'}$*

ideal points and the line at infinity. idea points / points at infinity: $x_3 = 0$. The whole set iles on a single line, which is $l_\infty = (0, 0, 1)$. For any line $l = (a, b, c)^T$, its ideal point meets l_∞ at $(b, -a, 0)^T$, which is the direction of lines. Thus, the line at infinity can be thought of as *the set of directions* of lines in the plane.

relation to projective plane / imaging. **Result Duality Principle:** *To any theorem of 2-D projective geometry, there corresponds a dual therem, which may be derived by interchanging the roles of points and lines.*

Conics and dual conics. Conic: 2-degree equation curve. and can be expressed homogeneously as symmetric matrix. *Proof:*
a conic in inhomogeneous coordinate is

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

. Now we homogenize it by $x \rightarrow x_1/x_3, y \rightarrow x_2/x_3$, rewrite it:

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

$$x^T Cx = 0, C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}$$

It has 5-dof. and full rank if it is not degenerated case.

comment: *The method of fitting a geometric entity (or relation) by determining a null space will be used frequently in the computation chapters throughout this book*

Result *l tangent to C at point x is given by $l = Cx$ Proof:*

Dual conics / conic envelopes. C^* The ls that is the dual of xs forms a conic $C^* = C^{-1}$(if full rank) in dual space. C^* is the adjoint matrix of C.

Degenerate conics. C is not full rank. $rank = 2$ two lines, $rank = 1$ repeated line.

comment: *from mathematic view: adjoint matrix of C and dual of C???*

1.3 Projective transformations

Geometry is the study of properties invariant under groups of transformations – Felix Klein.

(Geometry def): Projectivity / collineation / projective transformation / homography. a group of transformations. an invertible mapping $h : \mathbb{P}^2 \to \mathbb{P}^2$ such that three points x_1, x_2, x_3 lie on the same line iff $h(x_i), i = 1, 2, 3$ do. It preserves line!

(Algebraic def) Projectivity. h is Projectivity iff there exists a non-singular 3×3 matrix H, for any point in $\mathbb{P}^2$ represented by a vector x it is true that $h(x) = Hx$. h can be represented by H.

There are two equivalent ways to define, 1) Any invertible linear transformations of homogeneous coordinates is projectivity. 2) Any projectivity arises as such a linear transformations.

Projective transformation. planar projective transformation is a linear transformation on homogeneous 3-vector(coordinates). It can be represented by a non-sigular 3×3 matrix. $x^{'} = Hx$. Scale of H does not matter, so H is homogeneous matrix, with 8-dof.

After projective transformation, the projective properties remain invariance.

Perspectivity. If two coordinate system defined in two planes are both Euclidean coordinate, then the mapping is called perspectivity, with $6 - dof$.

comment: *Figure 2.5 in page 36. If the world lie in a planar, even locally, or camera rotating, or shadow, there is perspective images, where lots of good properties arise.*

Fundamentally different ways to transform line / points / conics. if $x' = Hx$, $l' = H^{-T}l$. $C' = H^{-T}CH^{-1}$, $C^{*'} = HC^*H^T$. **emmmmm????**think more!!

comment: *Conic and dual conic undergos two different transformation!*

class	expression	invariants	dof	#point
Isometries	$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$	length, angle, area	3	2
Similarity	$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$	ratio of length, angle, parallel	4	2
Affinity	$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$	ratio of length of $\|$, $\|$, area ratio	6	3
Projective	$\begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$	cross ratio of 4 collinear points	8	4

1.4 A hierarchy of transformations / projective linear group

Isometries. preserves Euclidean distance (iso=same, metric). Orientation-preserving if upper left $det(R) = 1$. It form a group.

Metric Structure. structure is defined up to a similarity.

Essence of Affinity. 1) scaling in orthogonal direction. $A = R(\theta)R(-\phi)DR(\phi)$ always decomposable! 2) affinity is the most general linear transformation that fix l_∞

Affinity is the middle ground for similarity and projective transformation. in affinity, $det(A)$ fully define scaling anywehre on the plane, and orientation. Also, ideal point remains ideal.

Decomposition of projective transformation. transformation higher in the hierarchy than the previous one. $H_P(2dof)$ moves the line at infinity, $H_A(2dof)$ affects the affine properties. $H_S(4dof)$ is similarity transformation.

$$H = H_{sim}H_AH_P = \begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ v^T & v \end{bmatrix} = \begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$$

emmmmm????think! Where $A = sRK + tv^T$. K is a normalized upper-tri matrix $detK = 1$

comment: *Well structured composition is easier to learn / compute??*

1.5 The projective geometry of 1D

Cross Ratio.

$$cross(x_1, x_2, x_3, x_4) = \frac{|x_1x_2||x_3x_4|}{|x_1x_3||x_2x_4|}, |x_ix_j| = det \begin{bmatrix} x_{i1} & x_{j1} \\ x_{i2} & x_{j2} \end{bmatrix}$$

The definition of $|x_i x_j|$ is the signed distance from i to j, if $x_{last} = 1$. cross ratio is valid for infinity points, invariant with choice of coordinate.

comment: *Once you fix the camera position / points in the world lie in one plane, the cross ratio of those points projected on the plane, is the same!!!, irrelavant of where my imaging plane puts.*

1.6 Topology of hte projective plane

$\mathbb{P}^2$ is equivalent of the set of all homogeneous 3-vector!! $x = (x_1, x_2, x_3)$ can be normalized as $x_1^2 + x_2^2 + x_3^2 = 1$. It lies in a sphere S^2 in $\mathbb{R}^3$. But $x, -x$ is the same point. Thus, **2-1 correspondence from points in $\mathbb{P}^2$, S^2.** line corresponds to great circle

emmmmm????In the language of topology, the sphere S2 is a 2-sheeted covering space of $\mathbb{P}^2$. This implies that $\mathbb{P}^2$ is not simply-connected, which means that there are loops in $\mathbb{P}^2$ which cannot be contracted to a point inside $\mathbb{P}^2$. To be technical, the fundamental group of IP2 is the cyclic group of order 2. **emmmmm????**

$P^2 \equiv$ a disk *with opposite points* on its boundary identified. $\equiv$ glued together.

P^2 is not orientatble.

comment: *will topology constraints helps learn a canonical uv map??* **comment:** *the projection is camera dependent? can we map images from a categories to the same mapping???* **comment:** *what does images from a category really capture???*

1.7 Recovery of affine and metric properties from images

DoF We actually only want to restore to a similarity transformation. so, only 4 dof is needed. line at infifity (2dof) + 2 circular points (2dof). $\equiv$ restore in the decmoposition chain .

vanishing line. affine properties can be recovered from identifing imaged vanishing line.

Length ratio. affine properties can also be recovered by length / cross ratio.

To reconstruct similarity / metric.

Result *circular points I, J are fixed points under H iff H is a similarity.*

Proof:
easy to calculate.

I, J determines the orthogonal direction of Euclidean geometry. Some intuition: $(0,1,0)^T, (1,0,0)^T$ are packed into $I = (1,0,0) + i(0,1,0)$.

The conic dual to the circular points.

$$C^*_\infty = IJ^T + JI^T \leftrightarrow I^T C^*_\infty I = 0, J^T C^*_\infty J = 0$$

. It is a degenerate rank2 conic.

Degenerate rank2 conic. passing two lines l, m, $C = lm^T + ml^T$. **Result** *dual conic C^*_∞ of circular points I, J are fixed points under H iff H is a similarity.* **Result** *junder euclidean coordinate,* $C^*_\infty = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

1. C^*_∞ is generally a symmetric 3×3 matrix, with 5-dof. But $det(C^*_\infty) = 0$, thus, it has 4-dof.
2. l_∞ is in the null space of C^*_∞, $C^*_\infty l_\infty = 0$

Angles on the projective plane

$$\cos\theta = \frac{l_1 m_1 + l_2 m_2}{\sqrt{(l_1^2 + l_2^2)(m_1^2 + m_2^2)}} = \frac{l^T C^*_\infty m}{\sqrt{(l^T C^*_\infty l)(m^T C^*_\infty m)}}$$

The later one to express angle is invariant of projective transformation. bc,

$$l^T C^*_\infty m \to l^T H^{-1} H C^*_\infty H^T H^{-T} m = l^T C^*_\infty m$$

Result *Once the conic C^*_∞ is defined, 1) the Euclidean angles can be measured. 2) the ratio of length can also be measured. (by* $\sin A, \sin B$*).*

Recovery of metric from image

$$C^{*'}_\infty = H C^*_\infty H^T = (H_P H_A H_s) C^*_\infty H^T = \begin{bmatrix} KK^T & KK^T v \\ v^T KK^T & v^T KK^T v \end{bmatrix}$$

So, if we can obeserve $C^{*'}_\infty$, the $H = U$ can be obtained by SVD.

$$C^{*'}_\infty = U \Sigma U^T, \Sigma = \begin{bmatrix} I & 0 \\ 0^T & 0 \end{bmatrix}$$

How to observe $C^{*'}_\infty$? It can either from two orthogonal lines / known angle, or from length ratio.

Othogonal lines. two lines l, m, $C^{*'}_\infty$ can be expressed as $\begin{bmatrix} KK^T & 0 \\ 0^T & 0 \end{bmatrix}$ (because we have affinity, set $v = 0$ in the previous result. $S = KK^T$ to guarentee symmetric.) Therefore, one pair of orthogonal lines forms one constraints:

$$(l_1' m_1', l_1' m_2' + l_2' m_1', l_2' m_2')(s_{11}, s_{12}, s_{22})^T = 0$$

emmmmm????this always has $\sigma = 1, 1, 0$??

length ratio. The image conic (an ellipse) intersect with l_∞ at circular points.

Orthogonal lines. (one step) Suppose we start from perspective image (instead of affine transformation as above.) one pair of othogonal lines provides one constraints on C^*_∞

$$(l_1m_1, (l_1m_2 + l_2m_1)/2, l_2m_2, (l_1m_3 + l_3m_1)/2, (l_2m_3 + l_3m_2)/2, l_3m_3)c = 0$$

$c = (a, b, c, d, e, f)^T$

comment: *but the two-step approach (termed stratified / stratification) also apply in 3D???*

1.8 More properties of conics

Polarity. for any x, C defines a line $l = Cx$. This line has a property: for the two points x_1, x_2it intersects with the conic, their tangent will meet back at x. x is the pole of l w.r.t C, l is the polar of x w.r.t. C.

A map: Correlation between pts and lines. Hence! the conic C induces a map between points and lines of $\mathbb{P}^2$. And interestintly, the mapping is invariant under projective transformation, since it is all based on incidence.

Conjugate points. if point y is on the line of $l = Cx$, then $y^TCx = 0$. y, x are conjugate w.r.t C.

Classification of conics under perspective. C is symmetric thus always has real eigenvalue. under projective transformation, $C' = U^{-T}CU^{-1} = D = diag(\sigma_1 d_1, \sigma_2 d_2, \sigma_3 d_3)$, $\sigma_i = \pm 1, 0, d_i > 0$, $D = diag(s_i)^T diag(\sigma_i) diag(s_i)$. Thus, the type of conics can be enumerated by enumerating σ_i. See table 2.2 in page 60.

Classification of conics under affinity. l_∞ is preserved, thus, has 3 classes, by intersetcting line and conic: 1) ellipse (does not intersect) 2) hyperbola(tangent) 3) parabola(2 points)

1.9 Fixed points and lines

fixed popints corresponds to eigen vector. **comment:** *does not need $\lambda = 1$ because $e, \lambda e$ is the same point. Fixed line are usually not mapped pointwise.*

Euclidean matrix. two circular points I, J, which corresponds to eigenvalues $\{e_{i\theta}, e^{-i\theta}\}$. Third eigenvector is called pole, $\lambda_3 = 1$. The Euclidean transformation is equivalent of rotate θ around pole, without translation. **comment:** *KDC!*

Similarity. ideal fixed points: circular points. eigenvalue $\{1, se^{i\theta}, se^{-i\theta}\}$

Affine. two eigen vector: ideal points $x_3 = 0$ (not circular points?). tird eigen vector is finite in general.

Chapter 2

Projective Geometry and Transformation of 3D

similar properties: l_∞ in P^2, π_∞ in P^3.

different properties: lines always intersects.

2.1 Points and projective transformation

dual , points - planes, lines- lines.

2.1.1 Planes

Incidence and relations. 1) points on the plane: $\pi^T X = 0$, $\pi = (n, d)$. 2) note that two points defines a line is not trivial as in $\mathbb{P}^2, l = x \times y$.

Three points define a plane.

$$\begin{bmatrix} X_1^T \\ X_2^T \\ X_3^T \end{bmatrix}_{3\times 4} \pi = 0$$

rank(3), then 1-dim null-space. To have similar expression as in $\mathbb{P}^2$, $l = x \times y \equiv [x, y]^T \cdot l = 0$. we have $det[X, X_1, X_2, X_3] = 0$, for any points on π.

$$detM = X_1 D_{234} - X_2 D_{134} + X_3 D_{124} - X_4 D_{123} = 0, so\pi = (D_{234}, -D_{134}, D_{124}, -D_{123})^T$$

Three points defines a point. Direct apply dual.

Projective transformation. Under point transformation $X' = HX$, $\pi' = H^{-T}\pi$

Parametrized points on a plane. Points X on the plane π can be written as $X = M_{4\times 3}x$, x in any point in $\mathbb{P}^2$, M generate null-space of π^T ($\pi^T M = 0$.

2.1.2 Lines

DoF and awkward problem. lines have 4-dof. the natural coordinate are 5. homogeneous 5 vetor cannot easily be used in math expressions togetehr with 4-vectors representing points and planes.

emmmmm????comment: *it is counter-intuitive. coz first defines a point, and defines a direction. it's 6-dof. what is misisng here???*

Null-space and span representation. Two points A, B. the line can be represnted as $W = \begin{bmatrix} A^T \\ B^T \end{bmatrix}$. *with the properties:* (1) span of W^T is the pencil of points $\lambda A + \mu B$ on the line (2) *span* of 2d right null-space of W is the pencile of planes with the line as axis. **comment:** *use span to cancel the choice of points! this is not in dual space.* **emmmmm????comment:** *are these represent the same line???*

Two plane P, Q. $W^* = \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$ (1) span of W^{*T} is the pencil of planes with the line as axis. (2) span of 2d right null-space of W^* is the pencil of points. **comment:** *use span to cancel the choice of points! this is not in dual space.*

Result $W^* W^T = 0_{2\times 2}$.

Join and incidence. the join of a point X and a line$W$ is

$$null(M) = null \begin{bmatrix} W \\ X^T \end{bmatrix}$$

the point defined by intersection of line W and plane π is

$$null(M) = \begin{bmatrix} W^* \\ \pi^T \end{bmatrix}$$

Plucker matrices. where line is represented by a 4×4 skew-symmetric homogeneous matrix. line joining 2 points A, B is represented by L

$$L = AB^T - BA^T$$

comment: *formally similar to cross product?*

It has properties: 1) L in rank 2. 2-dimensional null-space is spanned by the pencil of planes with the line as axis. $LW^{*T} = 0$ 2) L has 4dof. because a) skew-symmetric matrix has 6dof. 1 in scale, 1 is constrained by $det(L) = 0$ 3) is gerazalized of $l = x \times y$ 4) L is independent of choice of A, B. because if $C = A + \mu B$, we can derive the same L 5) point transformation, matrix is transformed as $L' = HLH^T$

$$L^* = PQ^T - QP^T, L^{*'} = H^{-T} L H^{-1}$$

Join and incidence.

$$\pi = L^{*}X, X = L\pi$$

Plucker line coordinates : six non-zero elements of $L_{4\times 4}$.

$$\mathcal{L} = \{l_{12}, l_{13}, l_{14}, l_{23}, l_{42}, l_{34}\}, detL = l_{12}l_{34} + l_{13}l_{42} + l_{14}l_{23} = 0$$

every line in P^3 defiens a surface in P^5 – Klein quadric.

comment: *It must satisfy the $detL = 0$ constraints. It is not homogeneous, but a subspace*

Result *Two lines $L\hat{L}$ are coplannar (and thus intersect) iff* $(L|\hat{L}) = 0$ $(L|\hat{L})$ is bilinear product!

some properties: 1) $det[A, B, \hat{A}, \hat{B}] = (L|\hat{L})$. it does not depends on the choice of $A, B, \hat{A}, \hat{B}$ 2) it can also defiend by dual plane $det[P, Q, \hat{P}, \hat{Q}] = (L|\hat{L})$ 3) L only represent a line in P^3 if $(L|L) = 0$ 4) if L is defiend by P, Q, $\hat{L}$ defind by A, B. $(L|\hat{L}) = (P^T A)(Q^T B) - (Q^T A)(P^T B)$ **emmmmm????**

2.1.3 Quadircs and dual quadrics. (conic in P^3)

$$X^T Q X = 0$$

1. 9-dof
2. 9 points defines a quadratic. **comment:** *why???*
3. a quadirc defines as polarity between a point and a plane $\pi = QX$
4. quadric is transformed by $Q' = H^{-T} Q H^{-1}$
5. dual quadrics are equations on planes $\pi^T Q^* \pi = 0, Q^* = adjointQorQ^{-1}$ **comment:** *what is adjoint mean??*
6. Dual quadric is transformed by $Q^{*'} = HQ^*H^T$

2.1.4 classification of quadrics

signature quadric is classified by diagnose matrix D where the element are only $1, 0, -1$. Number of $+1-$ number of -1.

classification. the projective type of a queadric is uniquely determined by rank and signature. Table 3.1 p74.

2.2 Twisted Cubics c

it is basically a parameterized curve!

2-d projective plane as parameterized curve.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = A \begin{bmatrix} 1 \\ \theta \\ \theta^2 \end{bmatrix}$$

a twisted cubic is defined to be a curve in P^3. Similarily we have A in 4×4 for P^3

The standard form / canonical form is $c(\theta) = (1, \theta, \theta^2, \theta^3)^T$

Application. 1) horopter for 2-view geometry. 2) defining the degenerate set for camera resectioning.

2.3 Hierarchy of Tranformation

Hierachy Table 3.2. 7-dof for similarity – (3 rotation, 3 translation, 1 isotropic scaling. ; 5 for affinity ($3 \times 3 - 3 - 1 = 5$ **emmmmm????**other intuition? s); – 3 projective part.

Screw deomposition a decomposition of a euclidean transformation.

comment: *will be useful to discuss special motions.*

Result *Any particular translation and rotation is equivalent to a rotation about a screw axis together with a translation along the screw axis. The screw axis is parallel to the rotation axis.* It can be determined from the fixed points of the 4×4 matrix representing the Euclidean transformation.

2.4 The plane at infinity

π_∞ / l_∞ allow affine properties to be measured. Ω_∞ / circular points allows metric properties to be mesured.

By defining infinity plane, we have $\mathbb{P}^3$ that any two pairs of planes intersect in a line.

comment: *π_∞ has 3 dof. why??? scale???*

Result *the plane at infinity is a fxied plane under H iff H is an affinity.* The plane is fixed as a set, not pointwise.

2.5 The absolute conic

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

Firstly, it means those points lie on the plane at infinity, $X_4 = 0$; Secondly, it means those points are in a conic $\sum_i^3 X_i^2 = 0$ Ω_∞ corresponds to a conic C with $C = I_{3\times3}$. 5dof. The points are all imaginary points / no real points.

Result *Absolute conic, Ω_∞ is a fixed conic under the projective transformation H iff H is a similarity transformation.*

Orthogonality and polarity. d_1, d_2 are orthogonal if $d_1^T \Omega_\infty d_2 = 0$. It is the definition of conjugacy w.r.t Ω_∞.

If image points are conjugate w.r.t IAC, then corresponding rays are orthogonal.

comment: *The 3×3 matrix indictes that the conic lies in the plane at infinity.*

2.6 The absolute dual quadric

comment: *it is actully more important than primal absolute quadric. Because it can be expressed in one algebric form.*

Euclidean / canonical coordinate. The envolope of all plane tangent to Ω_∞:

$$Q_\infty^* = \begin{bmatrix} I_{3\times 3} & 0 \\ 0 & 0 \end{bmatrix}$$

emmmmm????comment: *why Ω_∞ can not be expressed as the same form???*

Proof:
$\pi = (n, d)$. $\pi^T Q_\infty^* \pi = 0$, thus, $n^T n = 0$. On the other hand, n represents the lines that π meets with π_∞. from polarity w.r.t Ω_∞, $n^T I n = 0$ iff it is tangent to the conic.

Another *Proof:*
Consider all conic in form of $Q = diag(1, 1, 1, k)$. when $k \to \infty$, it only contains points $\sum_i^3 X_i^2 = 0, X_4 = 0$ (expand to infinity). The dual of Q is $Q^* = diag(1, 1, 1, k^{-1}) \to diag(1, 1, 1, 0)$.

Result *Q_∞^* is fixed iff H is similarity.*

Result *π_∞ is the null-vector of Q_∞^*.*

Result *angle is given by $\cos\theta = \frac{\pi_1^T Q_\infty^* \pi_2}{\sqrt{()()}}$, which is invariant to H.* **comment:** *this is a metric invariant to $f(X)$*

emmmmm????comment: *p84: should be π_∞, Q_∞^*?*

Chapter 3

Estimation – 2D Projective Transformation

3.1 DLT

Setting. 2D 2D point correspondence.

Key equation.

$$x'_i = Hx_i, or, x'_i \times (Hx_i)$$

comment: *This is a common trick to convert inhomogeneous equation to homogeneous.*

Solution. Inhomogeneous solution (not recommended), vs homogeneous solution

Solution from lines and other entities. $l' = H^{-T}l$. Count dof. Care be taken for mixed type.

3.2 Different costt functions

Algebraic distance. [Bookstein-79] pros: linear, cheap to compute. Starting point for non-linear min of a geometric / statistical cost function. Cons: no geometrically / statistically meaningful. or not expected intuitively. – can be solved by choice of normalization. **comment:** *You might want to consider the following when designing the loss function / supervision!!*

Geometric distance. x: measured imaged coordinates, $\hat{x}$: eistimated value of points, $\bar{x}$: true values of the points.

Error in one image. assume in the first image, $\bar{x} = x$. True in calibration pattern where points are measured to a very high accuracy. **Transfer error**:

$$\sum_i d(x'_i, H\bar{x}_i)^2$$

Symmetric transfer error. errors occur in both images. Forward, backward transormation.

$$\sum_i d(x'_i, Hx_i)^2 + \sum_i d(H^{-1}x'_i, x_i)^2$$

Estimated homography is the one for the above is minimized.

Reprojection error. Estimating a "correction" for each correspondence. Estimate both $\hat{H}$ *and* perfectly matched correspondence $\hat{x}_i, \hat{x}'_i$

$$\sum_i d(x_i, \hat{x}_i)^2 + \sum_i d(x'_i, \hat{x}'_i).s.t. \quad \hat{x}' = \hat{H}\hat{x}$$

MLE of homography and correspondence.

emmmmm????is there sth deep??

Comparison of geometric and algebraic. Fitting V_H on point $X = (x, y, x', y')$.

Conic analogue. fitting conic to 2D points. **emmmmm????**I'm lost...

Sampson error. A middle ground in between algebraic and geometric cost function in terms of complexity. close approximation to geometric error. 1st order method! The key is to consider constraints $Ah = 0$ as a cost depends on $X, C_H(X) + C_H(\hat{X}) = 0$. we wanna solve for $\|\delta_X\|^2$ subject to Taylor Expansion on $X = \delta X + \hat{X}$.

Result *Sampson Error is:*

$$\|\delta_X\|^2 = \epsilon^T (JJ^T)^{-1}\epsilon$$

, where J is ∇C_X

emmmmm????need to run through it yourself.

To find H for all points,

$$D = \sum_i \epsilon_i^T (J_i J_i^T)^{-1}\epsilon_i$$

Both ϵ, J depends on H.

Linear cost function. $C_H(X) = A(X)h$ is *linear* w.r.t X(??) is important. 1) the Taylor expansion is exact – Sampson error is geometric error. 2) Finding H becomes a hyperplane fitting problem.

Aonotehr geometric interpretation. all measurements is represented by a single point in a measurement space $\mathbb{R}^N$.

1. measurement space $\mathbb{R}^N$
2. model: a subset S of points in $\mathbb{R}^N$. if X in the subset, it satisfiy the modle.

Given the X, wanna find vector $\hat{X}$ closes to X that satisfy the model

Error in both image. $N = 4n$, if x and H are selected, the model defines x'. thus, the feasible subset S has $2n + 8$ dof. The geometric distance becomes: given a set of measured point pairs, which corresponds to a point X in $\mathbb{R}^N$, and an esitmated points $\hat{X} \in \mathbb{R}^N$ lying on S,

$$min\|X - \hat{X}\|^2 = \sum_i d(x_i, \hat{x}_i)^2 + d(x'_i, \hat{x}'_i)^2$$

Error in one image. model becomes $x'_i = H\bar{x}_i$

3.3 Statistical cost functions and Maximum Likelihood estimation

Assume noise is Gaussian on each image coord without bias, with uniform standard deviation σ. **comment:** *this might not hold true for imaging reason.*

Error in one image. MLE of the homography $\hat{H}$ maximizes the log-likelihood, which is equivalent as $\sum_i d(x'_i, H\bar{x}_i)^2$. In short, MLE is equivalent to minimizing the geometric error function.

$$logP(\{x'_i\}|H) = -\frac{1}{2\sigma_2}\sum_i d(x'_i, H\bar{x}_i)^2 + constant.$$

Error in both images. MLE is identiacal with minimizing *reprojection* error funciton. $\sum_i d(x_i, \hat{x}_i)^2 + d(x'_i, \hat{x}'_i)^2$

Mahalanobis distance. now we assume covariance matrix Σ. Then max log-likelihood is equivalent to minimizing the Mahalanobis distance

$$\|X - \bar{X}\|^2_\Sigma = (X - \bar{X})^T\Sigma^{-1}(X - \bar{X})$$

3.4 Transformation invariance and normalization

Question: Invariance of the algorithm to different choices of coordinates.

3.4.1 Invariance to image coordinate transormations

To what extent is the result of an algorithm that minimizes a cost funciton to estimate a homography dependent on the choice of coordinate in the image?

coordinates x in one image are replaced by $\tilde{x} = Tx$, and the in the other image $\tilde{x}' = T'x'$. Then, H are transormed correspondingly: $\tilde{H} = T'HT^{-1}$

The next question is, whetehr the outcome of the algorithm is independent of the T, T'.

3.4.2 Non-invariance of the DLT algorithm

Setup: a set of correspondence $x_i \leftrightarrow x'_i$ and H that is the reulst of the DLT. Consider further a related set $Tx_i \leftrightarrow T'x'_i$, let $\tilde{H} := T'HT^{-1}$. Question: does DLT apllied to the $\tilde{x}_i, \tilde{x}'_i$ yield $\tilde{H}$?

Result *T' be a similarity transformation with scale factor s, T is any arbetrary projective transformation. H is any homography and $\tilde{H} := T'HT^{-1}$. Then, $\|\tilde{A}\tilde{h}\| = s\|Ah\|$. In other word,*

$$d_{algebraic}(\tilde{x}'_i, \tilde{H}\tilde{x}_i) = sd_{alg}(x'_i, Hx_i)$$

Remark: **emmmmm????**(miss the argument p106).

1. No one-to-one correpondence between H and $\tilde{H}$.
2. $\|H\| = 1$ is not equivalent to $\|\tilde{H}\| = 1$

comment: *only dependent on T'.* **emmmmm????comment:** *go through the proof yourself. hint: start from $\epsilon = x' \times Hx$*

3.4.3 Invariance of geometric error

Minimizing gemetric error is invariant to similarity transformation.

$$d(\tilde{x}', \tilde{H}\tilde{x}) = d(T'x', T'HT^{-1}Tx) = d(T'x', T'Hx) = d(x', Hx)$$

, hwere Euclidean distance is unchagned under Euclidean transformation T'.

3.4.4 Nomralizing transformation

We have seen in the previous seciton, that there are some corrdinate systems better than others fro computing a 2D homography. Some normalization hsould be carried out before applying the DLT algorithm. Two pros: 1) result is more accurate 2) it undo the arbitrary choice of scale and origin. The algebraic min is carried out in a *canonical frame*.

comment: *here the term canonical comes with reason for computational accuracy.*

Isotropic scaling.

1. translate points to the origin.
2. scale points so that distance from the origin is equal to $\sqrt{2}$
3. apply those transformation to both image independently.

Why is normalization essential? (pre-conditioning) (must not consider as optional) Think in detail of DLT. we essencially solve a SVD for $A = U\Sigma V^T$, to solve h, $Ah = 0$. A is $2n \times 9$ but should have rank 8. However, it is impossible due to noisy data. Thus we want to find h to minimize $\|Ah\|$. It is equivalent to find a rank 8 matrix $\hat{A}$ that satisfy exactly $\|\hat{A}h\| = 0$ and closes to A in Frobenious form. $\hat{A} = U\hat{\Sigma}V^T$. $min_{\hat{A}}\|A - \hat{A}\|_F = \|UDV^T - U\hat{D}V^T\|_F = \|D - \hat{D}\|_F, s.t.rank(\hat{A}) = 8$.

The element of A is just $xx', xy', ww'...$. The xx' will be in the order of 10^4. ww' will be one. The min above is just increase / detrese the value such that the sum of the change is minimum to reach rank 8 matrix. But changing small amount of w (100) will have a huge effect in H but xx' will not.

comment: *It is insightful to switch between linear equation, SVD, null space, rank, optimize matrix, find vector.*

From the condition number aspect, the condition number of DLT is d_1/d_{n-1}. The exatct arithmetic results is independent of normalization, but it will diverge from the correct result in the presence of noise. large condintion number will amplify the effect.

Non-isotropic scaling and variants. experiment suggests it does not lead to significantly better results. Another variant is based on the observation that some column of A are not affected by noise (w, w', thus those column should not be varied to find $\hat{A}$).

Scaling with points near infinity. It makes no sense to normalize the coordinates of points in the infinite plane by setting the centroid at the origin, since centroid may have very large coordinates. ?????

3.5 Iterative minimization methods

methods for minimize the various geometric cost functions. Cons:

1. slower
2. need initial estimation
3. might not converge, or to local minimum
4. stopping criterion.

Overparameterization. Not neccessary: optimizer will 'notice' that its is not neccessary to move to redundant direciton. Not advisable: cost funciton surface is more complicated when minimal parameteraization is used.

Restricted to particular class.

$$H = I + (\mu - 1)\frac{va^T}{v^T a}$$

emmmmm????homography has 5 dof??

Function specification. The minimization is considered as fitting a surface S specified by the model to the measurement space $X \in \mathbb{R}^N$. S is loccaly parameterized. Each iteration is considered as varying the parameters to min the distance.

goal:

$$\|X - f(P)\|_\Sigma^2$$

, where P is parameter in $\mathbb{R}^M$, $f : \mathbb{R}^M \to \mathbb{R}^N$.

Error in one image (9):

$$\sum d(x_i', H\bar{x}_i)^2$$

$$f : h \to (Hx_1, Hx_2, \ldots, Hx_n)$$

Symmetric transfer error(9):

$$\sum d(x_i, H^{-1}x_i')^2 + d(x_i', Hx_i)^2$$

$$f : h \to (H^{-1}x_1', \ldots, H^{-1}x_n', Hx_1, \ldots Hx_n)$$

Reprojection error(2n + 9):

$$f : (h, \hat{x}_1, \ldots \hat{x}_n) \to (\hat{x}_1, \hat{x}_1', \ldots \hat{x}_n, \hat{x}_n'$$

Sampson error(9)

Gold Standard error **emmmmm????**same as geometric error?

comment: *new things: sampson error, golde standard error*

3.6 Experimental comparison of the algorithms

DLT is less robust to noise than Gold Standard algorithm. Always use normlized DLT.

3.7 Robust estimation

Till now only consider the error of measurement but ignore the error of mismatch.

emmmmm????homography dof? & projective dof?

RANSAC: a model is instantiated from a minimal set and is *scored* by the number of data points within a threshold distance. Alternative: median distance.

3.8 Automatic computation of a homography

input: just two images without correspondence. output: homography.

chicken and egg problem: correspondence & interest of point. Idea: first obatin a set of putative correspondences. **comment:** *now deep learning will initialize with almost identity? and optimize, thanks to the large field of view.*

3.9 Closure

ideas will reoccur throughout the rest of the book. 1) minimal number of correspondences; 2) degenerate case that should be avoided. **3)** algebraic and geometric error with more than minimal number of correspondences. **4)** parameterization that enforce internal constraints!

Chapter 4

Algorithm Evaluation and Error Analysis

How to test the performance of certain algorithm? ofetn not sufficient to only estimate variable or transformation, but measure of **confidence or uncertainty** is also required.

There are two methods for computing this uncertainty(covariance). 1) based on linear approximation and involves concatenating various Jacobian. 2) Monte Carlo method

comment: *why uncertainty is related to covariance?*

4.1 Bounds on performance

This section test on synthetic data, and also sketch the methodology.

4.1.1 Error in one image

Residual error.

$$\epsilon_{res} = (\frac{1}{2n}\sum_{i=1}^{n} d(x_i', \hat{x}_i')^2)^{1/2}$$

Note, it is not the true data $\bar{x}$. The value of the residual error is not in itself and absolute mearsure of the quality of the solution obtained. it depends on the number of correspondences. e.g. if given 4 correspondences to compute H, $\hat{H}$ of course will matches the observation exactly. BUT, it has variance σ^2 to the noise-free data.

4.1.2 Error in both images

$$\epsilon_{res} = (\frac{1}{2n}\sum_{i=1}^{n} d(x_i', \hat{x}_i')^2 + d(x_i, \hat{x}_i)^2)^{1/2}$$

4.1.3 Optimal estimators(MLE)

Minimization of geometric error is equivalent to MLE.

comment: *TO be continued... lost context...*

Chapter 5

Part 1: Camera Geometry and Single View Geometry

5.1 Outline

Chapter 6 is about projection of 3D scene onto 2D plane, that is encoded by a matrix. P is a 3×4 matrix to map from homogeneous coordinates of a world point in 3D o homogeneous coordinates of imaged point. The properties of the camera (eg focal length, center) can be extracted from it. Two classes of camera matrix: finite and afffine camera (parallel projection)

Chapter 7 is about

1. estimation of P.
2. Constraints can be incorporated into the estimation.
3. corection for radial lens distortion.

Chapter 8 covers

1. The action of camera on geometric objects other than finite point including lines, conics, quadrics, points at infinity.
2. calibration: K is computed without computing P by imaged absolute conic, or vanishing points and vanishing lines.
3. calibrating conic: simple geometric device for visualizing calibration.

Chapter 6

Camera Models

def: camera is a mapping between 3D world and a 2D image. We focus on central projection. The camera models are examined using the tools of projective geometry. The geometric entities of the camera can be simply computed from matrix representation.

Tow classes of camera: finite camera and affine camera (camera center at infinity). This chapter is principally concerned with the projection of points. Lines and other geometric entities are deferrred till chapter 8

6.1 Finite cameras

The basic pinhole model (with inhomogeneous coordinates). is illustrated in the figure. Consider the plane $Z = f$ – image plane / focal plane. Under such a model, the point in space (X, Y, Z) is mapped to $(fX/Z, fY/Z, f)$. We get the mapping when ignoring the last coordinate in $\mathbb{R}$ (inhomogeneous):

$$(X, Y, Z) :\to (fX/Z, fY/Z)$$

Central projection using homogeneous coordinates.

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

$$x = PX, P = diag(f, f, 1)[I|0]$$

comment: *Now Make the projeciton model (matrix) more general:*

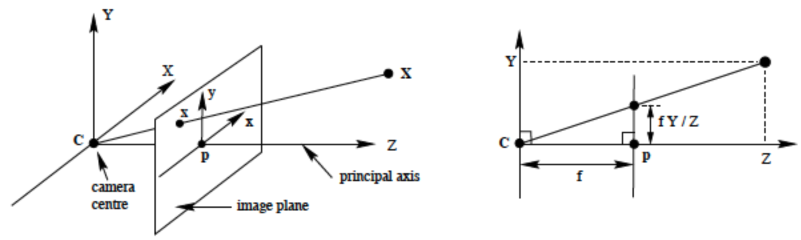

Fig. 6.1. Pinhole camera geometry. **C** *is the camera centre and* **p** *the principal point. The camera centre is here placed at the coordinate origin. Note the image plane is placed in front of the camera centre.*

Figure 6.1: note the def of camera center / optical cneter, principal axis ray, principal point

Principal point offset.

$$(X, Y, Z) :\to (fX/Z + p_x, fY/Z + p_y)$$

$$(X, Y, Z, 1) \to (fX + Zp_x, fY + Zp_y, Z) = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$x = K[I|0]x_{cam}, K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}$$

where X_{cam} is called camera coordinate frame in $\mathbb{P}^3$

rotation and translation. Now, the points in space is expressed in terms of different Euclidean coordinate frame. X_{cam} and X represents the same point in different frames.

$$X_{cam} = \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$x = KR[I| - C]X = K[R|t]X, P = K[R|t], t = -RC$$

dof: 1(f) +2 (p) + 3(R) + 3 (t) = 9

CCD cameras: non-square pixels.

$$K = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}$$

dof: 9 + 1 = 10

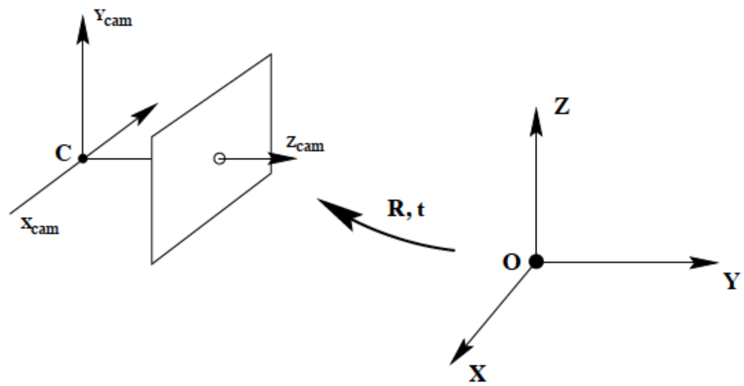

Fig. 6.3. *The Euclidean transformation between the world and camera coordinate frames.*

Figure 6.2: note the meaning of R, t

Finite projective camera. skew param, unusual cases.

$$K = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}$$

dof: $10 + 1 = 11$

Equivalence of finite projective camera and nonsigular projective matrix.

$$P = M[I|M^{-1}p_4] = KR[I| - C]$$

where p_4 is the last column of P. one decomposes M as product of $M = KR$, where K is upper triangular , R is the rotation. The decomposition is essentially RQ decomposition.

the set of camera matrices of finite projective cameras is identical with the set of homogeneous 3×4 matrix for which the left hand 3×3 submatrix is non-singular.

Most general projective cameras. if M can be singular, the mapping would be a line or point, instead of a 2D image. **comment:** *or affine camera??*

6.2 The proctive camera

6.2.1 Camera anatomy

Notation: $P = [M|p_4]$. M is non-singular if it is a finite camera. We'll see how we read the properties of cameras from the matrix P, M.

Camera center. it is the null space. $PC = 0$. *Proof:*
consider the line containing C and any other point A, points on this line will be

$$X(\lambda) = \lambda C + (1-\lambda)A$$

$$x = PX = \lambda PC + (1-\lambda)PA = (1-\lambda)PA$$

Since they are up to scale, so all x are the same point on image plane, which means , the line AC is a ray through camera center.

For finite and affine camera, we can also right the null space as :

$$C = \begin{pmatrix} -M^{-1}p_4 \\ 1 \end{pmatrix}, C = \begin{pmatrix} d \\ 0 \end{pmatrix}, Md = 0$$

Column vectors. columns of vector $p_1, \dots p_3$ are vanishign points of the world coordinates X, Y, Z. p_4 is the imaged world origin. *Proof:*
$(1,0,0,0)$ are mapped to p_1.

Row vectors. Some particualr planes.

The principal plane. The principal plane is *defined* as hte plane through the camera center parallel to the image plane. or defined as the set of points X which are imaged on the line at infinity of the image. $P^{3T}X = 0$. *Proof:*
It consists of the set of points X which are imaged at $PX = (x, y, 0)$. P^3 is the vector representing the principal plane of the camera.

Axis planes. Consider the properties of the set of points X on the plane P^1, $P^{1T}X = 0$. then the plane will be imaged to $PX = (0, y, w)$, which is the image y-axis. Note that 1) the axis plane P^1, P^2 are dependent on the choice of image coordinate system. P^3 is more tightly coupled to the natural camera geometry. 2) the join of axis planes (camera center) algebraically coincide with $PC = 0$.

The principal point and principal axis. The principal axis is the line passing through the camera center (where Z_{cam} points). The axis meets the iamge plane at principal point. *Proof:*
1) What is the direction perpendicualr to the principal plane? $d = (p_{31}, p_{32}, p_{33}, 0)$, which is $\hat{P}^3$ or m^3. 2) what is the imaged direction? Pd. That is the principal point.

$$x_0 = Mm^3$$

6.2.2 Action of a projective camera on points

Forward projection. The points of infinity ($D = (d^T, 0)^T$) only depend on the left 3×3 (M). **comment:** *use it in calibration.*

$$x = PD = [M|p_4]D = Md$$

Back-projection of points to rays. Points P^+x lies on the ray because it projects to x. Thus, the ray is the line

$$X(\lambda) = P^+x + \lambda C$$

If the camera is finite, then M is invertible. $\tilde{C} = -M^{-1}p_4$, $D = (M^{-1}x, 0)$.

$$X(\mu) = \mu \begin{pmatrix} M^{-1}x \\ 0 \end{pmatrix} + \begin{pmatrix} -M^{-1}p_4 \\ 1 \end{pmatrix} = \begin{pmatrix} M^{-1}(\mu x - p_4) \\ 1 \end{pmatrix}$$

6.2.3 Depth of points

First consider the 'normalized' camera, i.e. $|m^3| = 1$, det(M) > 0, plus, $X = (X, Y, Z, 1)$, then ,w is the depth to the camera center. ($x = w(x, y, 1) = PX$)

Proof:
$x = w(x, y, 1) = PX$, consider the 3rd dim.

$$w = P^3X = P^3(X - C) = m^3(\tilde{X} - \tilde{C})$$

Note $PC = 0$. The right side can be interpreted as 'dot product of the ray from the camera center to the point X', with the principal ray direction. if $\|m^3\|$ is unit vector, then the dot product is the depth.

In general sense, here is the result: **Result** *If* $X = (X, Y, Z, T)$*,* $P = [M|p_4]$*,* $PX = w(x, y, 1)$

$$depth(X; P) = \frac{sign(detM)w}{T\|m^3\|}$$

6.2.4 Decompostiion of the camera matrix

The goal is to find camera center C, orientation (R) and internal parameters(K).

Center: Find null space by SVD.

R, K**:** QR decomposition.

6.2.5 Euclidean vs projective spaces

Note that the world and image are essentially Euclidean. *We just borrow ideas from projective geometry such that central projection can be expressed linearly.* Cameras are euclidean devices and simply because we have a projective model of camera it does not mean we should eschew notion of Euclidean goemetry.

6.3 Camera at Infinity

When M is singular, it is not an finite camera. the camera center lies on the plane at infinity. There are two classes: affine and non-affine camera

Definition *Affine camera is one that has camera matrix* P *in which the last row* P^{3T} *is of the form* $0, 0, 0, 1$*.* **comment:** *it is not defined as the camera center is at infinity. it is more specific than that since affine camera also requries principal plane is at plane at infinity.*

6.3.1 affine cameras

Consider the cinematographic tracking back while zooming in techniques. (Vertigo effect / dolly zoom)

Moving back effect: consider camera moving away in the direction of principle axis m^3/r^3.

$$P_t = K \begin{bmatrix} r^1 & -r^1(C - tr^3) \\ r^2 & -r^2(C - tr^3) \\ r^3 & -r^3(C - tr^3) \end{bmatrix} = K \begin{bmatrix} r^1 & -r^1C \\ r^2 & -r^2C \\ r^3 & d^t \end{bmatrix}$$

Zoom in effect: by fatctor $k = d_t/d_0$

$$K \begin{bmatrix} d_t/d_0 & & \\ & d_t/d_0 & \\ & & 1 \end{bmatrix} \begin{bmatrix} r^1 & -r^1C \\ r^2 & -r^2C \\ r^3 & d^t \end{bmatrix} = \frac{d_t}{d_0} K \begin{bmatrix} r^1 & -r^1C \\ r^2 & -r^2C \\ r^3 d_0/d_t & d_0 \end{bmatrix}$$

$$P_\infty = K \begin{bmatrix} r^1 & -r^1C \\ r^2 & -r^2C \\ 0 & d_0 \end{bmatrix}$$

Thus, P_∞ becomes an instance of an affine camera by definition.
comment: *Need derivation myself.*

Chapter 7

Computation of the Camera Matrix P

Given (X_i, x_i) (or line correspondence) correspondences, we want to recover P from by numerical methods. More specifically, if some constraints apply to the matrix P, we will show how to use those constraints.

A important assumption is that central projection is *linear* , i.e. no lens distortion. If there is one, we introduce the correction method at the end of the chapter.

There are in general two ways to compute K. 1. decomposition from P by QR. 2) computed directly without necessitatign estimating P. The second method will be discussed in chapter 8.

7.1 Basic equations

DLT: It derives from $x_i = PX_i$ up to scale, i.e. $x_i^T PX_i = 0$. If we write them to linear equation w.r.t. P, we got

$$\begin{bmatrix} 0 & -w_i X_i^T & y_i X_i^T \\ w_i X_i^T & 0 & -x_i X_i^T \\ -y_i X_i^T & x_i X_i^T & 0 \end{bmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0$$

P^{iT} is the i-th row of P, so it has 12 unknowns, every pair provides 2 constraints (the 3 equations are linear dependent.)

Let A denote the $2n \times 12$ matrix from correspondences:

$$A = \begin{bmatrix} 0 & -w_i X_i^T & y_i X_i^T \\ w_i X_i^T & 0 & -x_i X_i^T \end{bmatrix}_{i=1\dots n} \in \mathbb{R}^{2n\times 12}$$

Minimal solution: $11/\ 2 = 5.5$

Overconstrained solution: when n > 5.5, we can solve a minimization problem rather than an exact solution. For *algebraic* error, we wanna minimize

$$\min \|Ap\|, \quad \text{s.t. 1)} \quad \|p\| = 1 \quad \text{or 2)} \quad p_{31}^2 + p_{32}^2 + p_{33}^2 = 1$$

Degenerate config. Degenerate cases are more involved than in 2D homography. Among those, there are 2 critical configs:

1. the camera and points lie on a twisted cubic.
2. The points all lie on the union of a plane and a single straight line containing the camera centers. **comment:** *Intuitively, you cannot decide the projection behavior which is orthogonal to the plane. :p*

Normalization. When the variance of depth is relatively slight, the centroid of the pointed should be translated to the origin. And the coordinates are scaled so that the RMS distance from the origin in $\sqrt{3}$ (**comment:** *isometric or not are discuessed before.*). However, if the variance of depth is significant, e.g. one near camera, and one at plane at infinity, other methods should be applied.

If we normalized the x by Tx, and X by UX, with brief proof, we got $P = T^{-1}\tilde{P}U$.

Line correspondence. if we know that L is mapped to l. then we also can build 2 constraints. Let us sample two points X_0, X_1 from L. we know that the projected X should also be on the line l. In the formal words,

$$l^T P X_j = 0, j = 0, 1$$

7.2 Geometric error

Same argument as in homography section, we don't actually wanna min the algebraic error $\|Ap\| = [w_i d(x_i, Px_i)]^2$, but $d(x_i, \hat{x}_i)^2 = d(x_i, PX_i)^2$. This is when we assume accurate measurement in the world space X_i. The situation might arise from an accurately machined calibration object.

Note that, we can only solve the min by iterative techniques, where a para of P is required. **comment:** *such that we can walk on the surface of parameter space by gradient descent.* The initialization could come from DLT solution / minimal solution.

comment: *From example 7.1, I'm surprised that how good linear solution can be. And it is much faster than iterative method.*

Errors in the world points. If errors happen in both world, we are optimizing

$$\sum_{i=1}^{n} d_{Mah}(x_i, P\hat{X}_i)^2 + d_{Mah}(X_i, \hat{X}_i)^2$$

d_{Mah} is the Mahalanobis distance w.r.t. known error covariance matrices.

7.2.1 Geometric interpretation of algebraic error

What does DLT is actually optimizing, geometrically? Answer: **Result** *it's optimizing*

$$f \sum_{d} (X_i, X_i')$$

, where X_i' is mapped exactly to x_i with the same depth of X_i. In contrast, $\hat{X}_i$ in the golder rules is the closest point that is mappeed to x_i.

Note that for points X_i not far away from the principal ray, the $\hat{X}$ is well approximated by X'. We also notcie two things to interpretate the optimization goal:

1. $d(X, X')$ is slightly larger than $d(X, \hat{X})$. DLT slightly weights the points farther away from the principal ray??? **emmmmm????**
2. DLT algorithm will be biased towards minimizing focal length at a cost of slight increase in 3D geometric error, since there is an f in the goal.

A loose *Proof:*
, (in terms of we assume a buch of normlized point) DLT is minimizing

$$\sum_i [\hat{w}_i d(x_i, \hat{x}_i)]^2, \hat{w}_i = \pm \|\hat{p}^3\| depth(X; P)$$

If P is normalized such that $\|\hat{p}^3\| = 1$, from simple geometry in Fig 7.2, we know that $\hat{w} = C_1 depth(X; P) = C_2 f d(X', X)^2$

7.2.2 Estimation of an affine camera

for affine camera, algebraic error and geometric error are equal. **emmmmm????**(–proof?). So, geometric error can be minimized linearly.

The basic function becomes:

$$\begin{bmatrix} 0 & -X_i^T \\ X_i^T & 0 \end{bmatrix} \begin{pmatrix} P^1 \\ P^2 \end{pmatrix} + \begin{pmatrix} y_i \\ -x_i \end{pmatrix} = 0$$

7.3 Restricted camera estimation

The constraints could be

1. $s = 0$
2. $\alpha_x = \alpha_y$
3. x_0, y_0 is known.
4. K is known.

We will look into several problems with the example of $s = 0, \alpha_x = \alpha_y$. So, dof = 11 - 1 - 1 = 9.

Minimizing geometric error. Let the set of parameters to estimate is q, then, we work on finding $f : \mathbb{R}^9 \to \mathbb{R}^{2n}$ that minimize the geometric errors, for errors on one side. If errors on both sides, we find $f : \mathbb{R}^{9+3n} \to \mathbb{R}^{2n+3n}$ which is pretty large..

Minimizing algebraic error Solving this is much smaller by *reduced measurement matrix*. If we define $g(q) = p$ that construct P by q, minimizing algebraic error is equivalent to min $\|Ag(q)\}|$ essentially, (**comment:** *which might not be linear, requiring iterative method. So we don't want the dimension too high*

Reduced measurement matrix. If we look at A, it is a $2n \times 12$ matrix, with very large number of rows. To 'reduce' the computation, we could replace A with $\hat{A} \in \mathbb{R}^{12\times 12}$ such that $\|Ap\| = \|\hat{A}p\|$. e.g. $\hat{A} = DV^T$. So $q \to \hat{A}g(q)$ is just a mapping from $\mathbb{R}^9 \to \mathbb{R}^{12}$

Result *Given a set of n correspondence X_i, x_i, the problem of finding a constrained camera matrix P that min algebraic reduces to minimization of function $\mathbb{R}^9 \to \mathbb{R}^{12}$, independent of n.*

Initialization The naive way is to run DLT to get P first, and then clamp the values to desired values which will initialize the variable in the SGD. However, in practice this does not work so well. People instaed will add **soft constraints** to the SGD, instead of clamping them during initialization.

$$\sum d(x_i, PX_i)^2 + ws^2, +w(\alpha_x - \alpha_y)^2.$$

And clamps the value at the end of the optimization.

Covariance estimation We can also estimate propagation of errors into an image, by calculating the covariance matrix.

$$\epsilon_{res} = \sigma(1 - d/2n)^{1/2}$$

d is the number of camera parameters.

comment: *Similar to homography, I'll put them brief for now...*

7.4 Radial distortion

Now, in reality, central projection could be non-linear. It happens if the ray is far aways from the principal axis, distorted by lens. We'll model the nonlinear effect by L, such that

$$(x_d, y_d) = L(\tilde{r})(\tilde{x}, \tilde{y})$$

where $\tilde{x}, \tilde{y}$ are ideal pinhole cameras. $\tilde{r} = \sqrt{\tilde{x}^2 + \tilde{y}^2}$.

choice of distrotion function L and center. We could approximate it by Taylor expansion. let

$$L = 1 + \kappa_1 r + \kappa_2 r^2 + \dots$$

. Then $\kappa_1, \kappa_2, \dots, x_c, y_c$ are considered as part of hte interior calibration of the camera.

Note, the feature extraction should still be in the original image, not corrected image, since artifact / aliasing will be introduced.

Chapter 8

More single View Geometry

This chapter described imaged line / plane / conic / quadric, by developing their forward and backward properties.

The camera matrix P is dissected further, with focus on 1) camera center and 2) image plane (K). There are two key properties about them. 1) the images with the same camera center are *projectively equivalent*. 2) images of entities on the plane at infinity only depends on KR, not on C.

Then, we study the images of entities on π_∞ because they are particularly important **comment:** *in terms of calibration?*. The image of a $X \in \pi_\infty$ is a vanising point, image of $l \subset \pi_\infty$ is vanising line. Their images both depends on K, R. However, ω only depends on $K, \omega = (KK^T)^{-1}$, thus is intimately connected with camera calibration K. ω defines the angle between rays back projected from image points.

8.1 Action of a projective camera on planes, lines, and conics

8.1.1 On planes

Result *Points on plane $x = (X, Y, 1) \in \pi$ and their image x is a planar homography. i.e. $x = Hx_\pi$ Proof:*
Consider we choose the XY plane as π in the world space. The points on the plane $X \in \pi$ is projeted:

$$x = PX = (p_1, p_2, p_3, p_4)\begin{pmatrix} X \\ Y \\ Z = 0 \\ 1 \end{pmatrix} = (p_1, p_2, p_4)\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = Hx_\pi$$

8.1.2 On lines

Forward projection. The line is parameterized by mu and two points A, B

$$x(\mu) = PX(\mu) = P(A + \mu B) = a + \mu b$$

Back projection to a plane. Algebraically, it back projects to plane $P^T l$. (simple proof.) Geometrically, all planes passing through camera center forms a star of plane (2-parameter family). The three rows of P, P^{iT} are a basis for the star. The plane $P^T l$ is just a linear combination of this basis.

8.1.3 On conics

A conic C back project to a cone Q_{co}, from simple calculation,

$$Q_{co} = P^T C P$$

Note that Q_{co} is a degenerate quadric, thus a cone, with null vector C – vertex of the cone.

8.2 Images of smooth surfaces

Definition Γ *is the contour generator in 3-space, apparent countor* γ *is the image of* Γ Note that 1) for finite camera, Γ only depends on camera center, not on image plane (or K). But γ depends on image plane. 2) for affine camera with parallel projection, both Γ and γ only depend on the projection direction k.

Next we describe the projection properties of quadrics, which is a general case for contour generator and apparent contour.

8.3 Action of a projective camera on quadrics

Forward projection. Since it's tangent, it relates to the dual quadric Q^*.
Result *Under camera* P*, the outline of quadric* Q *is the conic* C *defined by* C^**,*

$$C^* = PQ^*P^T$$

Proof:
1) dual conic C^* is illustrated by lines $l, l^T C^* l = 0$. 2) l are back projected to planes by $\pi = P^T l$.

Result *The plane of* Γ *for a quadric* Q *and camera with center* C *is given by* $\pi_\Gamma = QC$. **comment:** *Note that* Γ *of a quadric is a plane, but not in general.*
Proof:
C is the polar, π is the pole.

8.4 The importance of camera center

Images taken by two cameras with the same center are related by a planar projective transformation. 1) They are projectively equivalent 2) cameras can be considered as devices measuring projecitve properteis of the *cone of rays*.

Consider two camera with 1st as reference $P = KR[I|-C], P' = K'R'[I|-C]$. $x' = P'X = (K'R')(KR)^-1(KR)[I|-C]X = (K'R')(KR)^{-1}x$

For the next part, we assume world coordinate is the same as camera coordinate, $P = K[I|0]$

8.4.1 Moving the image plane

Result *the effect of zooming by a factor k is right multiplying K by $diag(k, k, 1)$*
Proof:
from the previous part we have. $x' = K'K^{-1}x, H = K'K^{-1}$ Suppose f is increased to kf while remaining principal point the same as $\tilde{x}_0$

$$K'K^{-1} = \begin{bmatrix} kI & (1-k)\tilde{x}_0 \\ 0 & 1 \end{bmatrix}$$

$$K' = \begin{bmatrix} kI & (1-k)\tilde{x}_0 \\ 0 & 1 \end{bmatrix} K = K \begin{bmatrix} kI & \\ & 1 \end{bmatrix}.$$

We leverage the fact that K, K' are upper triangle matrix.

8.4.2 Camera rotation

easy to show that **Result** $H = KRK^{-1}$. Even further, we can direclty read R / θ from estimating homography H. The conjugate rotation H has several good properties:

1. H has the same eigenvalues with R up to scale $\mu, \mu e^{i\theta}, \mu e^{-i\theta}$.
2. The angle can be comptued from the phase of complex eigenvalues.
3. The direction of rotation axis / vanishign point is μ

It is one of *infinite homography mapping.* It plays key role for auto-calibration.

8.4.3 Applications and examples

Synthetic views and panoramic images.

comment: *Note for synthetic views: by applying H, it is exactly the camera sitting at C with a diff rotation matrix would have seen, regardless if X is on the plane or not.*

8.4.4 Moving the camera center

Note that, if the camera center does not move, sit cannot tell us the 3-space structure!! To measure the 3-space, we actually need the camera center to move around.

8.5 Camera calibration and the image of the absolute conic

Now we study what we'll gain if K is known. In short, the projective device becomes a metric measure device (Euclidean properties).

More conretely, we can measure the *angle between two rays.* (later we'll show this expands to more nice properties).

Result *the camera calibration K is the transformation between x and the ray's diretion*

$$d = K^{-1}x$$

in the camera Euclidean coordinate frame.

Result

8.6 Vanishing points and vanishing lines

8.7 Affine 3D measurements and reconstruction

8.8 Determining camera calibration K from a single view

8.9 Single view reconstrutcion

8.10 The calibrating conic

Outline

Ther are three questions that will be addressed:

1. **Correspondence geometry**. Given an image x in the first view, how does this constrain the position of x' in the second view?
2. **Camera geometry(motion)**. Given a set of corresponding points x_i, x'_i, what are the P P'?
3. **Scene geometry(struction)**. Given x_i, x'_i P, P', what is X?

Chapter 9 describes the epipolar geometry and answer the first question. Note that this epipolar geometry *only* depends on cameras, *not* depend on the scene structure.

Chapter 10 describes one of the most important results in *uncalibrated* MV geometry: a reconstruction of **both** cameras and scene structure can be computed from iamge point correspondence. **comment:** *Here we use reconstruction as term for both camera and scene structure. not only refer to* X.

Chapter 9

Part 2: Epipolar Geometry and the Fundamental Matrix

9.1 Epipolar geometry.

Epipole, epipolar line, baseline, etc.

9.2 F

The fundamantial matrix is the algebraic representation of epipolar geometry. F defines a correlation / map from x to l'. This mapping is represented as F.

$$l' = Fx$$

comment: *$x'^T F x = 0$ is just a famous property, which is commonly used for forming constraints. It is not a definition / geometric meaning.*

9.2.1 Geometric derivation.

Result

$$F = [e']_\times H_\pi$$

. *H_π is the transfer mapping from one iamge to another via any plane π*

An arbitrary π defines a homography mapping each x to x'. where x' is not corresponding point, but a *potential* corresponding point. The epipolar geometry defines that, l' is the join of x' and e'. Thus, $l' = e' \times x' = [e']_\times H_\pi$.

comment: *π is arbitrary, thus H_π is not arbitrary projective transformation. yet, it induce the same F...*

9.2.2 Algebraic derivation.

1. you shoot a ray from C through x. The ray is

$$X(\lambda) = P^{+}x + \lambda C$$

. 2. Select 2 special points on this ray: $\lambda = 0, \lambda = \infty$. Project them by P'. $x'_1 = P'P^{+}x$, $x'_2 = P'C = e'$. 3. The join of x'_1, x'_2 is $l' = x'_1 \times x'_2 = [e']_{\times} P'P^{+}x$.

$$F = [e']_{\times} P'P^{+}$$

Specifically, if we assume world space sits at C, P, P' having special form such that: $P = K[I, 0], P' = K'[R, t]$, then

$$F = [e']_{\times} K'RK^{-1} = K'^{-T}RK^{T}[e]_{\times}$$

Here, the term $K'RK^{-1}$ occur again.

9.2.3 conrespondence condition

. $x^{'T}Fx = 0$

properties of the fundamental matrix

$$e'F = 0, Fe = 0.$$

Proof: e' is always on l', for all x.

F is a orrelation. x defines a line l'. any point on l will map to the same l'.

9.2.4 The epipolar line homography

Epipolar lines forms a pencil of lines. it is a 1-d projective space. So a homography can be derived for two pencil of lines. it has 3 dof.

thus the $F, 7dof$ can be considered as : 2 (e) + 2 (e) + 3 (pencil of lines)

9.3 Fundamental matrices arising from special motions

Consider pure tranlation and planar motion.

9.3.1 Pure translation

$F = [e']_{\times} K'RK^{-1}$ reduce to $[e']_{\times}$, only 2 dof.

$$x' = x + Kt/Z$$

if $x = x, y, 1$

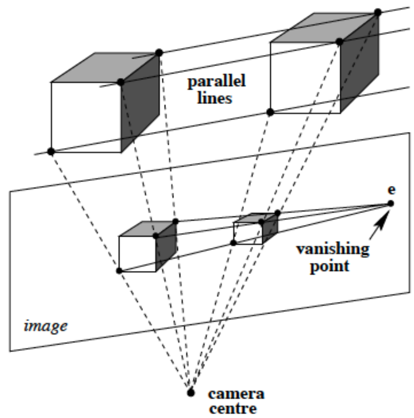

Fig. 9.7. *Under a pure translational camera motion, 3D points appear to slide along parallel rails. The images of these parallel lines intersect in a vanishing point corresponding to the translation direction. The epipole* e *is the vanishing point.*

Figure 9.1: vanishing point is e because all correspondent points move in parallel and intersect at v, which **emmmmm????**..

general motion. first rotate by R (represented by homography H), then apply the pure tranlation. $F = [e']_\times H, H = K'RK^{-1}$.

$$x' = K'RK^{-1}x + Kt/Z$$

First term only depends on x, iamge position alone, not on depth. This part accounts for H / R, K part. the second term depends on depth, but not on image point, accounts for tranlation.

Chapter 10

10

Chapter 11

10