CS4277 3D Computer Vision

Owen Leong
owenleong@u.nus.edu
November 29, 2024

1 2D Projective Geometry

1.1 Points and Lines

Point-line incidence relation $x^T l = 0$
Intersection of two lines is $x = l \times l'$
Line through two points is $l = x \times x'$
Ideal points (points at ∞) have $x_3 = 0$
Points and lines exhibit duality in $\mathbb{P}^2$

1.2 Conics

Conic $C = C^T \in \mathbb{R}^{3\times 3}$
Point incidence relation is $x^T C x = 0$
Tangent line at point x is $l = Cx$
A line is tangent to a conic if $l^T C^* l = 0$
Dual conic is given by $C^* = C^{-1}$
Degenerate conics include:

- Two intersecting lines for rank 2
- Single repeating line for rank 1
- Single repeating point for rank 0

1.3 Projective Transformations

$h(x) = H_{3\times 3} x, \det(H) \neq 0$
Collinearity is preserved in a projectivity
If $x' = Hx$, then $l' = H^{-T} l$ and $C' = H^{-T} C H^{-1}$ and $C^{*\prime} = HC^* H^T$

1.3.1 Isometric Transformation

$$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}, \begin{bmatrix} \epsilon\cos\theta & -\sin\theta & t_x \\ \epsilon\sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Preserves euclidean distance, $\epsilon = 1$ is orientation preserving, and $\epsilon = -1$ reverses orientation

1.3.2 Similarity Transformation

Include an isotropic scaling $\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$

1.3.3 Affine Transformation

$$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}, \det(A) \neq 0$$

Preserves parallel lines, ratio of lengths of parallel segments and ratios of areas
Projective transformations preserves order of contact, tangency and cross ratio

1.3.4 Decomposition of proj. tf

$$H = H_S H_A H_P = \begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ \mathbf{v}^T & v \end{bmatrix}$$

2 1D Projective Geometry

$$Cross(x_1, x_2, x_3, x_4) = \frac{|x_1 x_2||x_3 x_4|}{|x_1 x_3||x_2 x_4|}$$

$$|x_i x_j| = \det \begin{bmatrix} x_{i1} & x_{j1} \\ x_{i2} & x_{j2} \end{bmatrix}$$

Invariant under any proj. tf in $\mathbb{P}^1$

3 3D Projective Geometry

Proj. tf is $X' = H_{4\times 4} X, \det H \neq 0$
Plane is $\pi^T X = 0$
With $X' = HX$, $\pi' = H^{-T}\pi$
Plane can be defined from 3 points using null-space

3.1 Lines in $\mathbb{P}^3$

Given points A and B, a line is

$$\lambda A + \mu B \text{ or } \mathrm{span}(W^T), W = \begin{bmatrix} A^T \\ B^T \end{bmatrix}$$

Right null-space of W defines a family of planes passing through the line
Given planes P and Q, pencil of planes around the line intersected by P and Q can be defined as $\mathrm{span}(W^{*T}), W^* = \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$ and the line itself is the null-space
Plane defined by point X and line W is null-space of $M = \begin{bmatrix} W \\ X^T \end{bmatrix}$
Point defined by intersection of line W with plane π is null-space of $M = \begin{bmatrix} W^* \\ \pi^T \end{bmatrix}$

3.2 Quadrics

$$X^T Q X = 0, Q = Q^T \in \mathbb{R}^{4\times 4}$$

Intersection of plane and quadric is a conic
Dual quadric $\pi^T Q^* \pi = 0$, $Q^* = Q^{-1}$ or $\mathrm{adj}(Q)$ if $\det Q = 0$
For $X' = HX$, $Q' = H^{-T} Q H^{-1}$ and $Q^{*\prime} = HQ^* H^T$

3.3 Line at ∞ and Circ Points ($\mathbb{P}^2$)

In $\mathbb{P}^2$ space, identifying l_∞ recovers affine properties, and identifying circular points recovers metric properties.
Line at infinity is fixed under affine transformations, canonically $[0, 0, 1]^T$
Circular points lie on l_∞ at $I = [1, i, 0]^T$ and $J = [1, -i, 0]^T$ are fixed under similarity transformations
The dual to the circular points is the conic $C^*_\infty = IJ^T + JI^T$ which is $diag(1, 1, 0)$ in euclidean coordinates
C^*_∞ is fixed under similarity transforms
Under projective transformation, euclidean angles can be measured by

$$\cos\theta = \frac{l^T C^*_\infty m}{\sqrt{(l^T C^*_\infty l)(m^T C^*_\infty m)}}$$

C^*_∞ can be identified using constraints of orthogonal lines
Under proj tf $C^{*\prime}_\infty = \begin{bmatrix} KK^T & KK^T\mathbf{v} \\ \mathbf{v}^T KK^T & \mathbf{v}^T KK^T\mathbf{v} \end{bmatrix}$

3.4 Rectification

Remove projective distortion by computing vanishing line (l_∞) of an imaged plane, and compute a homography that maps the vanishing line to $[0, 0, 1]^T$
Remove affine distortion by identifying pairs of orthogonal lines to compute C^*_∞ and map it to $diag(1, 1, 0)$

3.5 Plane at ∞, Abs Conic Ω_∞ ($\mathbb{P}^3$)

Identifying plane at infinity recovers affine properties and identifying the absolute conic recovers metric properties.
π_∞ is fixed under affine transformation, canonically $[0, 0, 0, 1]^T$
The absolute conic lies on π_∞ and has points satisfying $X_1^2 + X_2^2 + X_3^2 = 0$ and $X_4 = 0$, or the conic $\Omega_\infty = I_{3\times 3}$.
The absolute conic is fixed under a similarity transformation.
For directions d_1, d_2,

$$\cos\theta = \frac{d_1^T \Omega_\infty d_2}{\sqrt{(d_1^T \Omega_\infty d_1)(d_2^T \Omega_\infty d_2)}}$$

Absolute dual quadric is $Q^*_\infty = diag(1, 1, 1, 0)$ in a metric frame, is fixed under similarity transforms.
Angle between two planes is

$$\cos\theta = \frac{\pi_1^T Q^*_\infty \pi_2}{\sqrt{(\pi_1^T Q^*_\infty \pi_1)(\pi_2^T Q^*_\infty \pi_2)}}$$

4 Rigid Body Motion

Represented by $\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \in SE(3)$

5 Homographies

If X_1 in C_1 is $X_2 = RX_1 + t$ in C_2 and if N is the unit normal vector representing plane π and d is perpendicular distance from π to C_1 then $x_2 \sim (R - \frac{tN^T}{d})x_1$
For far away scenes, $d \to \infty$ so $H_\infty = R$
H has 8 degrees of freedom, each point correspondence gives 2 constraints, so 4 points needed.
Use Direct Linear Transform algorithm with normalization step.
Geometric distance $\sum d(x_i', Hx_i)^2$
Sym transfer err $\sum d(x_i, H^{-1}x_i')^2 + d(x_i', Hx_i)^2$
Sampson err

$$||\delta_x||^2 = \epsilon^T (JJ^T)^{-1}\epsilon, J\delta_x = -\epsilon$$

Mahalanobis distance
$|X - f(P)|^2_\Sigma = (X_f(P))^T \Sigma^{-1}(X - f(P))$

5.1 RANSAC

Randomly select sample, estimate model from sample, count number of points within threshold from model, and repeat to find the largest consensus set. Num of samples required $N \approx \frac{\log(1-p)}{\log(1-w^s)}$

6 Camera Models

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

C is the coord of the cam centre in world frame, $M = KR$

$$\text{Intrinsics} \quad K = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

s skew parameter, assumed to be 0
assume that pixels are square $f_x = f_y$
$PC = 0$ is undefined point, null-vector
P^{3T}, the third row of P is the principal plane (plane through camera centre $\parallel$ to image plane)
P^{1T} is the axis plane defined by the camera centre and the $x = 0$ line in the image
P^{2T} is the plane defined by camera centre and the $y = 0$ line in the image
Backproj of an image point is the ray

$$X(\lambda) = P^+ x + \lambda C, P^+ = P^T (PP^T)^{-1}$$

6.1 Decomposition of Camera Matrix

$$P = KR[I| - C] = [M| - MC] = K[R|t]$$

Solve for K, R by QR decomp of M

6.2 Calibration of Projective Camera

For a given checkerboard pattern, setting the surface to the $z = 0$ plane, there is a homography between image points and world coordinates
Generate orthonormal constraints

$$h_1^T K^{-T} K^{-1} h_2 = 0$$
$$h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2$$

Solve for $B = K^{-T}K^{-1}$, a symmetric matrix with 6-DOF
Each view generates 2 constraints, hence 3 views needed to solve 6 unknowns in B
K recovered from B by Cholesky decomp

6.3 Radial Distortion

$$x_d = (1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_5 r^6)\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2\kappa_3 xy + \kappa_4(r^2 + 2x^2) \\ \kappa_3(r^2 + 2y^2) + 2\kappa^4 xy \end{bmatrix}$$

6.4 Affine camera

Affine camera is a camera at infinity, has last row $[0, 0, 0, 1]$, M is rank 2.

7 Effects of Projective Cam

Map between points on a plane and their image is a planar homography
A line projects to a line in the image, where plane formed by C and the line intersects with the image plane
A line backproj to plane $\pi = P^T l$
A conic backproj to cone $Q_{co} = P^T C P$
For an image of a smooth surface, the contour generator Γ is the set of points X at which rays are tangent to the surface, and the image apparent contour γ is the image of Γ
A quadric Q projects to the conic C given by $C^* = PQ^*P^T$

$$depth(X; P) = \frac{\mathrm{sign}(\det M)w}{|m^3|}$$

$x = w(x, y, 1)^T = PX$
m^3 is third row of M

7.1 Same Camera Centre

Images with same camera centre are related by a homography

$$x' = (K'R')(KR)^{-1}x$$

7.2 Known Calibration

An image line l defines a plane through camera centre with normal $n = K^T l$ in camera's euclidean coordinate frame

7.3 Image of Absolute Conic

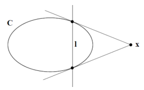
Points on π_∞, $X_\infty = (d^T, 0)^T$ project to $PX_\infty = KRd$, thus the mapping is $H = KR$, thus Ω_∞ is mapped to the image of the absolute conic (IAC) $\omega = (KK^T)^{-1}$
Angle btw rays is

$$\cos\theta = \frac{x_1^T \omega x_2}{\sqrt{(x_1^T \omega x_1)(x_2^T \omega x_2)}}$$

ω can be identified using constraints on vanishing points
Dual image of absolute conic is $\omega^* = \omega^{-1} = KK^T = PQ^*_\infty P^T$
A plane π intersects π_∞ in a line, the line intersects Ω_∞ at the circular points.

7.4 Pole-Polar Relationship

Polar line $l = Cx$



If y is on $l_x = Cx$, then $y^T Cx = 0$ are conjugate points.

7.5 Vanishing Points

$v = \lim_{\lambda\to\infty} x(\lambda) = \lim_{\lambda\to\infty} P(A+\lambda D) = KRd$

$$\cos\theta = \frac{v_1^T \omega v_2}{\sqrt{v_1^T \omega v_1}\sqrt{v_2^T \omega v_2}} \text{ (vanishing point)}$$

$$\cos\theta = \frac{l_1^T \omega^* l_2}{\sqrt{l_1^T \omega^* l_1}\sqrt{l_2^T \omega^* l_2}} \text{ (vanishing line)}$$

8 Epipolar Geometry

Image point x backprojects into a ray that projects onto I' as as epipolar line l'
Epipolar plane π is determined by the baseline and ray
Epipoles are the intersections with the baseline and the image planes, $e' = P'C, e = PC'$

8.1 Fundamental Matrix

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

Epipolar lines $l' = Fx$ and $l = F^T x'$

Correspondence condition $x'^T Fx = 0$

Epipoles are null-vecs $Fe = 0, F^T e' = 0$

For pure translation, $F = [e']_\times$
For F to be the fundamental matrix for P and P', $P'^T FP$ must be skew sym
F is fundamental matrix for (P, P'), F^T is fundamental matrix for (P', P)
Suppose l and l' are corresponding epipolar lines, k is any line not passing through epipole e, then $l' = F[k]_\times l$
Use 8 point correspondences to solve for F based on correspondence condition.
Normalize to $\hat{x}_i = Tx_i$, solve for F based on $\hat{x}_i \leftrightarrow \hat{x}'_i$, enforce singularity constraint by setting last singular value to 0, set $F = T'^T \hat{F}' T$

8.1.1 Solve for P given F

General formula for P, P' given F is $P = [I|0]$ and $P' = [[e']_\times F + e'v^T|\lambda e']$

8.2 Essential Matrix

$$F = K'^{-T} E K^{-1}, E = [t]_\times R$$

Recovery of R, t from E uses SVD, has 4 cases, find which has 3D points appearing in front of both cameras

8.3 Linear Triangulation

Solve for X based on constraints $x \times (PX) = 0$ and $x' \times (P'X) = 0$, each contributes 2 equations from the cross product, 4 equations for 4 homogenous unknowns.

8.4 Projective Ambiguity

$$PX = (PH)(H^{-1}X) \text{ given } x \leftrightarrow x'$$

9 Generalized Camera

Plucker vector is (direction q, moment $q' = P \times q$) for any point P on the line. $q^T q' = 0$
Points on the line are $(q \times q') + \alpha q, \alpha \in \mathbb{R}$
$q = R_{Ci} K_{Ci}^{-1} [x, y, 1]^T$
$q' = t_{Ci} \times q$, R_{Ci}, t_{Ci} is camera rotation and translation, K_{Ci} is internal calibration matrix
Rigid transformation R and t applied to plucker vectors:

$$\begin{bmatrix} R & 0 \\ [t]_\times R & R \end{bmatrix} \begin{bmatrix} q_1 \\ q'_1 \end{bmatrix} = \begin{bmatrix} Rq_1 \\ [t]_\times Rq_1 + Rq'_1 \end{bmatrix}$$

Intersection happens if

$$\begin{bmatrix} q_b \\ q'_b \end{bmatrix}^T \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} q_a \\ q'_a \end{bmatrix} = 0$$

Generalized epipolar constraint

$$\begin{bmatrix} q_2 \\ q'_2 \end{bmatrix}^T \begin{bmatrix} E & R \\ R & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q'_1 \end{bmatrix} = 0, E = [t]_\times R$$

Given R and t and a pair of correspondences, solve for αs in $R(q_1 \times q'_1) + \alpha_1 Rq_1 + t = (q_2 \times q'_2) + \alpha_2 q_2$

10 Pose Estimation (PnP)

Given 2D-3D correspondences, find camera pose R, t.
For unknown K, setup constraints of $x_i \times PX_i = 0$
For known K,

10.1 Gunert 3-point algorithm

First solve for distances from camera to 3D points using angles between rays and distances between 3D position. Solve 4th degree polynomial. Distances give 3D points in camera frame. Remaining is R,t. Done by moving both centroids to origin, then compute $M = \sum r'_i r_i^T$, $R = MQ^{-1/2}, Q = M^T M, t = \bar{p}' - R\bar{p}$, $Q^{-1/2} = V\text{diag}(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \frac{1}{\sqrt{\lambda_3}})V^T, Q = V\Lambda V^T$ is the eigenvector decomposition

10.2 Quan 4-point algorithm

10.3 EPnP n-point algorithm

Express each point as a linear combination of 4-control points. Solve the 4 point PnP.

11 Trifocal Tensor 3-view geom

$$l^T = l'^T [T_1, T_2, T_3] l'', (l'^T [T_i] l'')[l]_\times = 0$$

$$P = [I|0], P' = [A|a_4], P'' = [B|b_4]$$

$$T_i = a_i b_4^T - a_4 b_i^T$$

PLP: $x'' = H_{13}(l')x = [T_i^T] l' x$
PPL: $x' = H_{12}(l'')x = [T_i] l'' x$
PLL: $l'^T (\sum_i x^i T_i) l'' = 0$
PPP: $[x']_\times (\sum_i x^i T_i)[x'']_\times = 0_{3x3}$

$$e'^T [u_1, u_2, u_3] = 0, e''^T [v_1, v_2, v_3] = 0$$

where $u_i^T T_i = 0$ and $T_i v_i = 0$ are left/right null-vectors

$$F_{21} = [e']_\times [T_i] e'', F_{31} = [e'']_\times [T_i^T] e'$$

$$P' = [[T_i]e''|e'], P'' = [(e''e''^T - I)[T_i^T]e'|e'']$$

To compute $[T_i]$, first use algebraic error and constraints, then compute epipoles, then repeat but substituting $e' = a_4, e'' = b_4$ into $T_i = a_i b_4^T - a_4 b_i^T$

12 SfM

First, associate images with overlapping regions. Next, find relative poses. Then, use plane sweeping algorithm to get dense 3D.
Extract keypoints using SIFT/ORB. Train "visual words" from training data via hierarchical clustering. Express each image as frequency of visual words (bag of words).
Find 2D-3D correspondences, then solve PnP problem. Apply bundle adjustment to refine camera poses. $\arg\min_{P,X} \sum\sum ||x_{ij} - \pi(P_i, X_j)||^2$, $\pi(X_j, P_i) = \frac{\hat{x}}{\hat{x}_3}, \hat{x} = P_i X_j$

13 Bundle Adjustment

For $f(x) : \mathbb{R}^n \to \mathbb{R}^m$,

$$J \in \mathbb{R}^{m\times n}, J_{ij} = \frac{\partial f_i}{\partial x_j}$$

$$H \in \mathbb{R}^{n\times n}, H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$P_{i+1} = P_i + \Delta_i$$

Gradient descent $\Delta = -\lambda J^T \epsilon$
Newton's method uses quadratic approximation $J^T J\Delta = -J^T \epsilon$
Levenberg-Marquardt uses the augmented normal equations

$$(J^T J + \lambda I)\Delta = -J^T \epsilon$$

If iteration has reduction in error, λ is divided by factor. If error increases, λ is multiplied by factor. Small λ is similar to Gauss-Newton iteration, large λ is similar to gradient descent.
Sparse LM can be used by partitioning image points and camera parameters.
Fill-in is avoided by reordering A using minimum degree, column approximate minimum degree permutation, reverse cuthill-mckee, nested dissection, etc.

14 Two view stereo

14.1 Stereo rectification

Assuming O, O' are camera centers and $P = K[I|0], P' = K[R|t]$

Compute $e = P\begin{bmatrix} O' \\ 0 \end{bmatrix} = KO'$ and e',

then compute $\hat{e} = K^{-1}e = O'$ and $\hat{e}' = t$

Map $\hat{e}$ to $[1, 0, 0]^T$ using $H = \begin{bmatrix} R_1^T \\ R_2^T \\ R_3^T \end{bmatrix}$

$$R_1 = \frac{O'}{|O'|} \; R_2 = \frac{[-o'_y, o'_x, 0]^T}{\sqrt{o'^2_x + o'^2_y}} \; R_3 = R_1 \times R_2$$

Map $\hat{e}'$ to $[1, 0, 0]^T$ using $H' = HR^T$

14.2 Correspondence Search

- Normalized cross corr $\frac{(f-\bar{f})(g-\bar{g})}{\sigma_f \sigma_g}$
- Sum-of-squared diffs $||f - g||^2$
- Sum-of-absolute diffs $||f - g||_1$
- Mutual info $\sum P(x, y) \log \frac{P(x,y)}{P(x)P(y)}$

14.3 Depth from Disparity

B is baseline (dist between O and O')
disparity $= x - x' = \frac{B\cdot f}{z}$
depth $= z = \frac{B\cdot f}{x-x'}$

14.4 Methods

Scanline optimization stereo minimises $E(d) = \sum D(p, d_p) + \sum_{q\in N(p)} R(d_p, d_q)$
Semi-global matching performs line optimization along multiple directions
Single-view stereo deep learning optimizes appearance matching loss, disparity smoothness loss and left-right disparity consistency loss
Multiview stereo using plane sweeping algorithm considers a discrete set of depth values, for each image warps the image based on the depth and pose to match reference view and computes a consistency between reference view and warped image. Takes best consistency across all depth levels

15 Others

$(Ma) \times (Mb) = (\det M)(M^{-1})^T (a \times b)$

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

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_2 b_3 - b_2 a_3 \\ -(a_1 b_3 - b_1 a_3) \\ a_1 b_2 - b_1 a_2 \end{bmatrix}$$

Data normalization:

$$T = \begin{bmatrix} s & 0 & -sc_x \\ 0 & s & -sc_y \\ 0 & 0 & 1 \end{bmatrix}, s = \frac{\sqrt{2}}{d}$$

where c is centroid, d is mean distance of all points from centroid
For 3D, same formula as above except $\sqrt{3}$ instead of $\sqrt{2}$

$$\begin{bmatrix} c\gamma & -s\gamma & 0 \\ s\gamma & c\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\beta & 0 & s\beta \\ 0 & 1 & 0 \\ -s\beta & 0 & c\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\alpha & -s\alpha \\ 0 & s\alpha & c\alpha \end{bmatrix}$$

Adjoint of matrix is transpose of matrix of cofactors - remove row and col, apply determinant and add checkerboard shape of plus and minus
For 3×4 matrix, $P^+ = P^T (PP^T)^{-1}$ and $PP^+ = 0$