Class Scribe：Structure From Motion (2005/04/20)

網媒所 R93944019 杜明鴻

一、Singular value decomposition (SVD)

1. 每個 matrix 都代表一個 transformation
2.

Theorem 3.2.1 *If A is a real $m \times n$ matrix then there exist orthogonal matrices*

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \in \mathcal{R}^{m\times m}$$
$$V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \in \mathcal{R}^{n\times n}$$

such that

$$U^T AV = \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_p) \in \mathcal{R}^{m\times n}$$

where $p = \min(m, n)$ *and* $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$. *Equivalently,*

$$A = U\Sigma V^T \; .$$

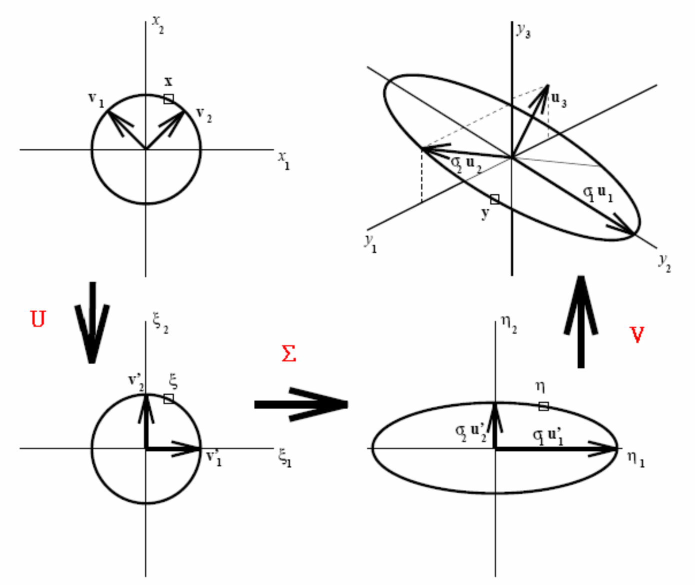
The SVD reveals a great deal about the structure of a matrix. If we define r by

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = 0 \; ,$$

that is, if σ_r is the smallest nonzero singular value of A, then

$$\operatorname{rank}(A) = r$$

3. SVD 三個 componen 的幾何意義：



U：對向量做 rotation 和 shift
Σ ：對向量做 scaling
V：對向量做 rotation 和 shift

4. 用途：

(1) 一個 rank 原本為 r 的 matrix 可能因為 noise 的緣故，造成 rank 大於 r。這時我們可以利用 SVD，找到一個 matrix A’，滿足

$\|A'-A\|$ 為最小，且 rank(A’) = r < rank(A)

=>

$$A' = U\Sigma' V^T \text{ ，where } \Sigma' = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$$

(2) 解 over-determined linear system 時，我們希望能 minimize $\|Ax-b\|$

=>

Theorem. The minimum-norm least squares solution to a linear system Ax = b, that is, the shortest vector x that achieves the

$$\min_x \|Ax-b\|$$

is unique, and is given by

$$\hat{x} = V\Sigma^+ U^T b$$

where

$$\Sigma^\dagger = \begin{bmatrix} 1/\sigma_1 & & & & & 0 & \cdots & 0 \\ & \ddots & & & & & & \\ & & 1/\sigma_r & & & \vdots & & \vdots \\ & & & 0 & & & & \\ & & & & \ddots & & & \\ & & & & 0 & 0 & \cdots & 0 \end{bmatrix}$$

is an n × m diagonal matrix.

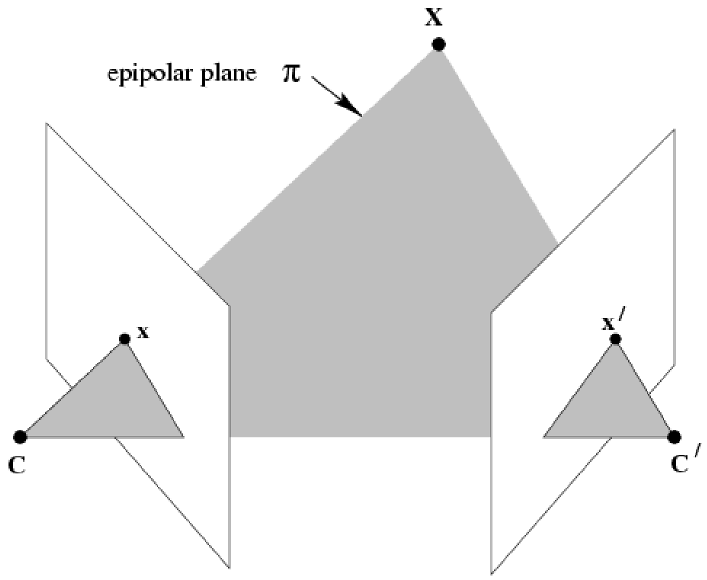
The matrix

$$A^+ = V\Sigma^+ U^T$$

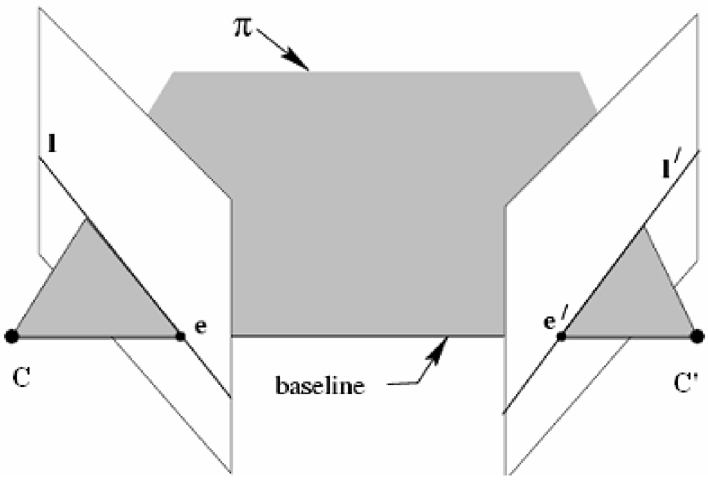
is called the pseudoinverse of A.

二、Epipolar geometry & fundamental matrix

1. The epipolar geometry



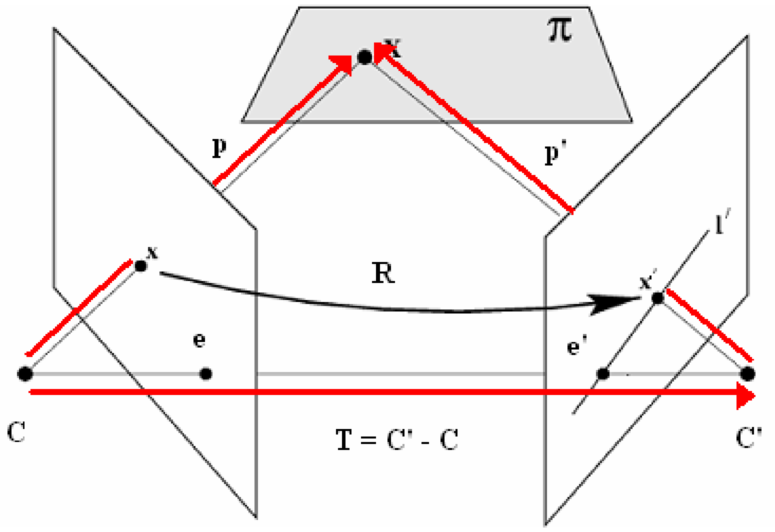
上圖中的 C, C’分別是兩張 image 的 projection center，X 是實際空間上的一點，x, x’是 X 在兩張 image 上的投影。C, x, X 三點共線，C’, x’, X 三點共線，C, C’, x, x’, X 五點共平面。



epipolar pole (e, e’)：projection center 在另一張 image 上的投影。
baseline：epipolar pole 的連線，通過 projection center。
epipolar plane (π **)**：兩張 image 的 projection center 和實際空間上的任一點 X 形成一個 epipolar plane。不同的 X 形成不同的 epipolar plane，但 baseline 都是相同的，換句話說，包含 baseline 的平面都是 epipolar plane。
epipolar line (*l, l’*)：image 與 epipolar plane 相交的直線。

所有 epipolar plane 投影在 epipolar line *l* 上的點, 只會投影在對應的 epipolar line *l’*上，因此做 block matching 只需在 *l’*上做 linear search。

2. The fundamental matrix F



向量 p, p'之間的關係可透過下面方程式來表示

p' = R (p - T) ---(1)

(p – T)是 epipolar plane 上的向量，(T × p)是垂直 epipolar plane 的向量，兩者垂直，內積爲 0，因此通過 X 點的 epipolar plane 則可透過下面方程式表示

$$(p-T)^T(T\times p)=0 \quad \text{---(2)}$$

將(1)式代入(2)式後可得

$$(R^T p')^T(T\times p)=0$$

最後再將(T × p)的外積以矩陣表示，T × p = S p

$$\mathbf{S}=\begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

代入後化簡

$$\begin{aligned} &(R^T p')^T(Sp)=0 \\ \Rightarrow\ &(p'^T R)(Sp)=0 \\ \Rightarrow\ &p'^T (RS)p=0 \\ \Rightarrow\ &p'^T Ep=0 \end{aligned}$$

得到 $p'^T Ep=0$，其中 E = RS 稱爲 **essential matrix**，R 的幾何意義是 rotation，S 的幾何意義是 translation。

令 M, M' 爲內部參數，滿足

$$p=M^{-1}x,\quad p'=M'^{-1}x\ 。$$

將上面兩式代入 $p'^T Ep=0$ 化簡

$$(M'^{-1}x')^T E(M^{-1}x) = 0$$
$$\Rightarrow x'^T [(M'^{-1})^T E M^{-1}] x = 0$$
$$\Rightarrow x'^T F x = 0$$

$F = (M'^{-1})^T E M^{-1}$ 稱為 fundamental matrix，是 epipolar geometry 的代數表示法。

兩張 image 上的任一組對應點(x, x')都會滿足 $x'^T Fx = 0$ 。

fundamental matrix 具有下面特性：

(1) **Transpose**：假設 F 是(P, P')的 fundamental matrix，則 F^T 會是(P', P)的 fundamental matrix。

(2) **Epipolar lines**：*l'* = Fx & *l* = F^T x'。

(3) Epipoles：因為 epipolar 落在所有的 epipolar line 上，所以 $e'^T Fx = 0 \ \forall x$

$\Rightarrow e'^T F = 0$ 。同理， $Fe = 0$ 。

(4) F 有七個自由度。(F 是 homogeneous，且 rank=2)。

(5)F is a correlation, projective mapping from a point x to a line l'=Fx (not a proper correlation, i.e. not invertible)。

fundamental matrix 的用途：

(1) **簡化** matching：因為 $x'^T Fx = 0$ ，當我們知道 F 和 x 時，只需做 linear search。

(2) Detect wrong matches：將 match 的點代入 $x'^T Fx = 0$ ，由等式的成立與否即可判斷是不是正確的 match。

3. Estimation of F - 8-point Algorithm：

令 x= $(u, v, 1)^T$, x' $= (u', v', 1)^T$, **F** $= \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$

每組對應點可以得到一條如下的線性方程式

$$uu' f_{11} + vu' f_{12} + u' f_{13} + uv' f_{21} + vv' f_{22} + v' f_{23} + uf_{31} + vf_{32} + f_{33} = 0$$

雖然 F 的自由度是 7，但是使用 8 點來求 F 比較好解。8 組對應點代入方程式後可以得到下面的 linear system

$$\begin{bmatrix} u_1u_1' & v_1u_1' & u_1' & u_1v_1' & v_1v_1' & v_1' & u_1 & v_1 & 1 \\ u_2u_2' & v_2u_2' & u_2' & u_2v_2' & v_2v_2' & v_2' & u_2 & v_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu_n' & v_nu_n' & u_n' & u_nv_n' & v_nv_n' & v_n' & u_n & v_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

通常我們不去直接解 Af = 0，而是去找一個使得$\|Af\|$最小的 f，這個 f 會是 $A^T A$ 的 least eigenvector。因爲一些誤差，可能造成求出的 F 的 rank 大於 2，我們找一個 rank 爲 2 且使得$\|F - F'\|$最小的 F’代替。首先對 F 做 SVD，令 F=UΣV^T，

其中 $\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$, 令 $\Sigma' = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

則 F’= UΣ’V^T

優缺點：

8-point algorithm 的優點是它是線性的，容易完成且快速。缺點是容易受 noise 的影響，原因是 A 各行的數值大小差異過大，是 ill conditioned，即使只有很小的 noise 也會造成很大的誤差。

解決方法：normalized 8-point algorithm

4. Normalized 8-point algorithm

(1) 透過 transformation T, T’，將 input 的對應點 x, x’ normalize 到[-1, 1]之間。
$\hat{x} = Tx$, $\hat{x}' = T'x'$ 。

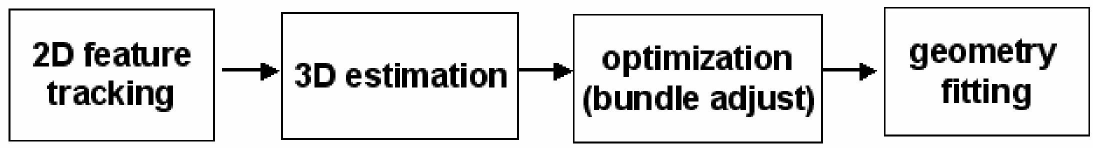
(2) 以 $\hat{x}$, $\hat{x}'$ 當作 input，呼叫 8-point algorithm，得到 $\hat{F}$

(3) $F = T'^T \hat{F} T$

二、Structure from motion

自動從兩張或多張 image 重建出 camera motion 和 scene structure，是一種自我較正的技術，稱爲 automatic camera tracking 或 matchmoving。應用於 computer vision, multiple-view shape reconstruction, novel view synthesis and autonomous vehicle navigation。另外也用於影片製作，將 CGI 插入真實場景中。

SFM pipeline



2D feature tracking 做完後，可以直接做 bundle adjustment，但 bundle adjustment 需要一個好的 initial guess，3D estimation 就用來提供 bundle adjustment 的 initial guess。

Step1：track features

尋找 frame 之間對應的 feature，有一些常用的方法可以使用：

(1)Lucas & Kanade-style motion estimation

window-based correlationSIFT matching 其中 KLT tracking 的效果較差，SIFT 的效果較好，但是 SIFT 沒有用到時間的概念，使用時必須修正原本的演算法。

Step2：Estimate Motion and Structure

(1) Simplified projection model, e.g., [Tomasi 92]

(2) 2 or 3 views at a time [Hartley 00]

Step3：Refine estimates

(1) "Bundle adjustment" in photogrammetry

(2) Other iterative methods

Step4：Recover surfaces

image-based triangulation, silhouettes, stereo…。

Issues in SFM

(1) Track lifetime

通常一段影片中的場景都是一直在變換的，track 的 feature 可能經過幾十個 frame 後就跑出畫面外，應該替每個 feature 加上一個 weighting function，時間過得越久 weight 就越小，如此可避免 track 錯誤的 feature 一直延續下去。

(2) Nonlinear lens distortion

(3) Nonlinear lens distortion

(4) Degeneracy and critical surfaces

(5) Prior knowledge and scene constrains

(6) Multiple motions

三、Factorization methods

利用 image sequence 的資訊重建出 scence geometry 和 camera motion。

1. Notations

n 3D points are seen in m views

q=(u,v,1)：2D image point

p=(x,y,z,1)：3D scene point

Π ：projection matrix

π ：projection function

$$\pi(x, y, z) = (x/z, y/z)$$

q_{ij} ：the projection of the i-th point on image j

$$q_{ij} = \pi(\Pi_j p_i)$$

λ_{ij} ：projective depth of q_{ij}

$$\lambda_{ij} = \mathrm{z}$$

2. 目標

Estimate M_j and p_i to minimize

$$\varepsilon(\mathbf{\Pi}_1, \cdots, \mathbf{\Pi}_m, \mathbf{p}_1, \cdots, \mathbf{p}_n) = \sum_{j=1}^{m}\sum_{i=1}^{n} w_{ij} \log P(\pi(\mathbf{\Pi}_j \mathbf{p}_i); \mathbf{q}_{ij})$$

$$w_{ij} = \begin{cases} 1 & \text{if } p_i \text{ is visible in view j} \\ 0 & \text{otherwise} \end{cases}$$

Assume isotropic Gaussian noise, it is reduced to

$$\varepsilon(\mathbf{\Pi}_1, \cdots, \mathbf{\Pi}_m, \mathbf{p}_1, \cdots, \mathbf{p}_n) = \sum_{j=1}^{m}\sum_{i=1}^{n} w_{ij} \left\| \pi(\mathbf{\Pi}_j \mathbf{p}_i) - \mathbf{q}_{ij} \right\|^2$$

3. SFM under orthographic projection

q = Π p + t

q：2D image point (2×1)

Π ：orthographic projection matrix (2×3)

p：3D scene point (3×1)

t：image offset (2×1)

藉由將 3D scene origin 移到 3D point 的 centroid，2D image origin 移到 2D point 的 centroid，公式中的 t 可以消去，變成 q = Π p。上面公式是將一個 feature 投影到一張 image 上，而將 n 個 feature 投影到一張 image 上的公式可以寫成

$$\begin{bmatrix}\mathbf{q_1} & \mathbf{q_2} & \cdots & \mathbf{q_n}\end{bmatrix} = \prod \begin{bmatrix}\mathbf{p_1} & \mathbf{p_2} & \cdots & \mathbf{p_n}\end{bmatrix}$$
$$2\times n \qquad 2\times 3 \qquad 3\times n$$

將 n 個 feature 投影到 m 張 image 的公式可以寫成

$$\underset{2m\times n}{\begin{bmatrix}\mathbf{q}_{11} & \mathbf{q}_{12} & \cdots & \mathbf{q}_{1n} \\ \mathbf{q}_{21} & \mathbf{q}_{22} & \cdots & \mathbf{q}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_{m1} & \mathbf{q}_{m2} & \cdots & \mathbf{q}_{mn}\end{bmatrix}} = \underset{2m\times 3}{\begin{bmatrix}\mathbf{\Pi}_1 \\ \mathbf{\Pi}_2 \\ \vdots \\ \mathbf{\Pi}_m\end{bmatrix}} \underset{3\times n}{\begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n\end{bmatrix}}$$

W measurement　　M motion　S shape

其中 rank(W) <= 3 (因爲 rank(M)<=3)

4. Factorization

W = MS 公式中，W 是已知，M 和 S 是我們想要求的。

(1) 對 W 做 SVD 得到 $W = U\Sigma V^T$，其中 U 的維度是 2m×2m，Σ的維度是 2m×n，V 的維度是 n×n。

(2) 強制將Σ轉成 3x3 的矩陣得到Σ’，使 rank(Σ’)≦3，並將 U 多餘的行去掉得到 U’(2m×3)，將 V^T 多餘的列去掉得到 V'^T (3×n)

(3) $W' = U'\Sigma' V'^T = (U'\sqrt{\Sigma'})(\sqrt{\Sigma'}V'^T) = M'S'$

(4) S’和 S 的用一個 transformation 表示，$W' = M'S' = (MA^{-1})(AS)$，對 M 加上些 metric constrains，即可解出 A。

(5) 因爲相機的投影是 orthographic，projection matrix Π 的列向量是 orthonormal，因此 M 的列向量也必須是 orthonormal。有了這個條件我們可以解出 A，進一步求出 M(=M’A)。

(6) 通常不使用(5)的方法 (原本 Tomasi/Kanade paper 上的方法)，而是假設 Π=Π’A，則

$$\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix} = \Pi\,\Pi^T = (\Pi' A)(\Pi' A)^T = \Pi'(AA^T)\Pi = \Pi' G\Pi, \quad where\ G = AA^T,$$

列出 M 中所有 Π_i 的 equation，解出 G。之後對 G 做 SVD (U=V)，將 G 拆解成 A A^T，即可得到 A。

5. Factorization with noisy data

$$\underset{2m\times n}{\mathbf{W}} = \underset{2m\times 3}{\mathbf{M}}\ \underset{3\times n}{\mathbf{S}} + \underset{2m\times n}{\mathbf{E}}$$

從 image sequence 中得到的 W 可能帶有一些誤差，利用 SVD 可以得到 optimal rank 3 approximation W’ of W，將誤差分離出來 $\underset{2m\times n}{\mathbf{W}} = \underset{2m\times n}{\mathbf{W}'} + \underset{2m\times n}{\mathbf{E}}$。接下來用 W’去做前面的 factorization，得到的結果會使 image feature 的位置與我們建出的

projection 的 SSD 最小。

6. Extensions to factorization methods
   (1) Projective projection
   (2) With missing data
   (3) Projective projection with missing data