Primer on Image Formation

Javier R. Movellan

July 23, 2006

1 Weak Perspective Projection

Let $v(i)$ be the projection of $q(i)$ on the image plane. Under the weak perspective model we have that

$$v_i(i) = \frac{f}{z} q_1(i) \tag{1}$$

$$v_2(i) = \frac{f}{z} q_2(i) \tag{2}$$

$$v_3(i) = f \tag{3}$$

where $f \in \Re$ is the camera's focal length ad $z \in \Re$ is the *scale* or *principle depth*. Typically we define $s = z/f$ and treat it as an unknown parameter to be estimated from the data. Weak prespective is a reasonable approximation when the depths of the q points are equal and the object is close to the camera's optical axis.

If we know p and v recovering r, t and s reduces to an absolute orientation problem **?**. Defining an error function ρ of the following form

$$\rho(r, t, s) = \frac{1}{s} \sum_{i=1}^{n} \|rp(i) + t - sv(i)\|^2 \tag{4}$$

This is the same objective function as for the absolute orientation problem. **?** presents a classic solution to this problem by modifying the objective function as follows:

$$\rho'(r, t, s) = \frac{1}{s} \rho(r, t, s) \tag{5}$$

This objective function has a unique minimum at

$$\hat{s} = \sqrt{\frac{\sum_{i=1}^{n} \|p(i) - \bar{p}\|^2}{\sum_{i=1}^{n} \|q(i) - \bar{q}\|^2}} \tag{6}$$

$$\hat{r} = VU^T \tag{7}$$

$$\hat{t} = s\bar{v} - \hat{r}\bar{p} \tag{8}$$

where

$$\bar{p} = \frac{1}{n} \sum_{i=1}^{n} p(i) \tag{9}$$

$$\bar{v} = \frac{1}{n} \sum_{i=1}^{n} v(i) \tag{10}$$

and U, V are defined via the svd decomposition of the following matrix M

$$M = \sum_{i=1}^{n} (q(i) - \bar{q})(p(i) - \bar{p})^T = USV^T \tag{11}$$

is the svd decomposition of the cov Let $r \in \Re^3 \otimes \Re^3$ be a rotation matrix and $t \in \Re^3$ a translation vector. Let $y'(i)$ be the 3D coordinates of the projection of point i on the image plane. where

$$\Psi = \begin{pmatrix} \mathrm{f/z} & 0 & 0 \\ 0 & \mathrm{f/z} & 0 \\ 0 & 0 & \mathrm{f} \end{pmatrix} \tag{12}$$

is a projection matrix, with f representing the focal length and z an arbitrary constant representing the depth of an orthographic projection plane. In practice Ψ makes $y(i)$ insensitive to the third row of r' and the last element of t'. Thus it is convenient to drop the last coordinates on y', and t' and the last row of r'. Thus (**??**) simplifies as follows

$$y(i) = ro(i) + t, \;\; i = 1, \cdots, n \tag{13}$$

where

$$y(i) = \begin{pmatrix} y'_1(i) \\ y'_2(i) \end{pmatrix} \tag{14}$$

$$t = \begin{pmatrix} \mathrm{f/z} & 0 & 0 \\ 0 & \mathrm{f/z} & 0 \end{pmatrix} \Psi t' \tag{15}$$

$$r = \begin{pmatrix} \mathrm{f/z} & 0 & 0 \\ 0 & \mathrm{f/z} & 0 \end{pmatrix} \Psi r' \tag{16}$$

2 ParaPerspective Projection

Note in weak projection when object moves parallel to image plane its projection on the image plane shifts but does not change in appearance. The paraperspective transformation is an affine approxiamtion to perspective projection that produces changes in appearance as we move parallel to the image plane. In weak projection we first do an orthographic projection of the object drawing rays parallel to the optic axis (perpendicular to image plane and through the focal point) onto a plane parallel to the image plane and passing through the center of mass of the centroid. Then we do perspective projection of the orthographically projected points.

In paraperspective projection we first project the object onto a plane passigng through the center of mass of the object and parallel to the image plane. In weak projection the projecting rays are parallel to the optical axis. In paraperspective projection the rays are parallel to the line that connects the optic axis and the center of mass of the object. The projected points are then reprojected onto the image plane using perspective projection. Since all the points on the plain are at equal depth, then the scaling factor is the same for all the points in the object.

Here is some useful math for paraperspective projection.

Let p a point in the object, c the center of mass of the object. We let the coordinate origin be at the focal point and have $z = (0, 0, 1)^T$ a unit vector perpendicular to the image plane. We want the intersection p' between the line that passes through p, and is parallel to the vector c and the plane that is is perpendicular to z and contains the point c. Thus since $p' - p$ is parallel to the vector c it follows that

$$p' - p = \alpha c \tag{17}$$

where $\alpha \in \Re$. Note the distance from the origin to the plane we will project onto is c_3, the third coordinate of the center of mass. Thus $p' - c_3 z$ is on the plane, and must be perpencidular to z.

$$(p + \alpha c - c_3 z) \cdot z = p_3 + \alpha c_3 - c_3 = 0 \tag{18}$$

and

$$\alpha = \frac{c_3 - p_3}{c_3} \tag{19}$$

$$p' = p - \frac{p_3 - c_3}{c_3} c \tag{20}$$

The perspective projection $\hat{p}$ of p' is the para-perspective projection of p.

$$\hat{p} = \frac{f}{p'_3} p' = \frac{f}{c_3} p' \tag{21}$$

3 Rigid Transformations

Let $p = \{p(i) \ : i = 1, \cdots, n\}$ represent 3D coordinates of n reference points exrpressed in an object-centered reference frame. Let $q = \{q(i) \ : i = 1, \cdots, n\}$ be the corresponding coordinates in a camera-centered reference frame, i.e,

$$q(i) = rp(i) + t \tag{22}$$

where

$$r = \begin{pmatrix} r_1^T \\ r_2^T \\ r_3^T \end{pmatrix} \tag{23}$$

is a rotation matrix. The row vectors r_1^T, r_2^T, r_3^T are the coordinates of the camera unit axis vectors with respect to the object frame of reference. By convention the camera reference frame is such that the center of projection of the camera is at the origin and the third axis is in the positirection of the camera's optical axis.

4 The Absolute Orientation Problem

The absolute orientation problem consists of recovering r and t from p and q. Horn **?** presents a classic solution to this problem by minimizing the following objective function

$$\rho(r,t,s) = \frac{1}{s}\sum_{i=1}^{n} \|rp(i) + t - sq(i)\|^2 \tag{24}$$

where $s \in \Re$ is an scale parameter introduced to allow for changes in scale, not just rigid transformations.

This function is uniquely minimized (see Apendix A) at

$$\hat{s} = \sqrt{\frac{\sum_{i=1}^{n} \|p(i) - \bar{p}\|^2}{\sum_{i=1}^{n} \|q(i) - \bar{q}\|^2}} \tag{25}$$

$$\hat{r} = UV^T \tag{26}$$

$$\hat{t} = s\bar{q} - \hat{r}\bar{p} \tag{27}$$

where

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p(i) \tag{28}$$

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v(i) \tag{29}$$

and U, V are defined via the svd decomposition of the following matrix M

$$M = \sum_{i=1}^{n} (q(i) - \bar{q})(p(i) - \bar{p})^T = UWV^T \tag{30}$$

5 Appendix A: The Absolute Orientation Problem

We want to minimize the function

$$\rho(r,t,s) = \frac{1}{s}\sum_{i=1}^{n} \|rp(i) + t - sq(i)\|^2 \tag{31}$$

First we take the gradient with respect to t and set it to zero

$$\nabla_t\, \rho = \frac{2}{s}\sum_{i=1}^{n} (rp(i) + t - sq(i)) = 0 \tag{32}$$

thus

$$\hat{t} = s\bar{q} - r\bar{p} \tag{33}$$

and

$$\rho(r,\hat{t},s) = \frac{1}{s}\sum_{i=1}^{n} \|r\tilde{p}(i) - s\tilde{q}(i)\|^2 = \frac{1}{s}\sum_{i=1}^{n} \|\tilde{p}(i)\|^2 - 2\sum_{i=1}^{n} r\tilde{p}(i)\cdot\tilde{q}_i + s\sum_{i=1}^{n} \|q(i)\|^2 \tag{34}$$

where $\tilde{p}(i) = p(i) - \bar{p}$ and $\tilde{q}(i) = q(i) - \bar{q}$. Taking the gradient with respect to t we get

$$\nabla_t \rho = -\frac{1}{s^2} \sum_{i=1}^{n} \|\tilde{p}(i)\|^2 + \sum_{i=1}^{n} \|\tilde{q}(i)\|^2 \tag{35}$$

Thus

$$\hat{t} = \sqrt{\frac{\sum_{i=1}^{n} \|\tilde{p}(i)\|^2}{\sum_{i=1}^{n} \|\tilde{q}(i)\|^2}} \tag{36}$$

All that is left is findign an orthonormal matrix $\hat{r}$ that maximizes

$$\sum_{i=1}^{n} \tilde{q}(i)^T r \tilde{p}(i) \tag{37}$$

Note

$$\sum_{i=1}^{n} \tilde{q}(i)^T r \tilde{p}(i) = \text{trace}(r^T m) = \tag{38}$$

where

$$m = \sum_{i=1}^{n} \tilde{q}(i)^T \tilde{p}(i) \tag{39}$$

Taking the eigen decomposition of $m^T m$, a symmetric psd matrix

$$m^T m = p \lambda p^T \tag{40}$$

where the columns of p are eigenvectors of $m^T m$ and $\lambda_1, \lambda_2, \lambda_3$ are the diagonal elements of the eigenvalue matrix λ. Let $u = m s^{-1}$ where $s = p\sqrt{\lambda} p^T$. Note u is orthonormal, s is psd and $m = us$. We want to maximize

$$\text{trace}(r^T m) = \text{trace}(r^T us) = \text{trace}(r^T up\sqrt{\lambda} p^T) \tag{41}$$

$$= \text{trace}(r^T u \sum_{i=1}^{3} \sqrt{\lambda_i} p_i p_i^T = \sum_{i=1}^{3} \sqrt{\lambda_i} \text{trace}(r^T u p_i p_i^T) \tag{42}$$

$$\leq \sum_{i=1}^{3} \sqrt{\lambda_i} \tag{43}$$

where the last inequality occurs because $p_i^T r^T$ is the transpose of a unit vector and up_i is a unit vector, since both r and u are orthonormal. Now let

$$m = uwv^T \tag{44}$$

be the svd decomposition of m, i.e., u and v are orthonormal and w is diagonal. Note $m^T m = v^T w^2 u$ and thus $w_i = \sqrt{\lambda_i}$ for $i = 1, 2, 3$. Moreover if $r = uv^T$ then

$$r^T m = vu^T uwv^T = vWvT \tag{45}$$

and

$$\text{trace} r^T m = \text{trace}(vwv^T) = \sum w_i \tag{46}$$

showing that r achieves the desired maximum[1].

[1]It can be shown this maximum is unique

5.1 A more general error function

It is useful to generalize Horn's approach to the case in which the objective function is of the form

$$\rho(r,t,s) = \frac{1}{s}\sum_{i=1}^{n}(rp(i)+t-sq(i))^T\sigma^2(i)(rp(i)+t-sq(i)) \tag{47}$$

where $\sigma^2(i) \in \Re^3 \otimes \Re^3$ encodes the relative precission of the coordiantes of the i^{th} point. Let $\sigma(i)$ be a square root matrix of $\sigma^2(i)$, i.e. $\sigma^2(i) = \sigma^T(i)\sigma(i)$. Thus

$$\rho(r,t,s) = \frac{1}{s}\sum_{i=1}^{n}\|\sigma(i)(rp(i)+t-sq(i))\|^2 \tag{48}$$

Taking the gradient of ρ with respect to t and setting it equal to zero

$$\frac{2}{s}\sum_{i=1}^{n}\sigma(i)(rp(i)+t-sq(i)) = 0 \tag{49}$$

Thus

$$\hat{t} = s\bar{q} - r\bar{p} \tag{50}$$

where

$$\bar{p} = \frac{\sum_{i=1}^{n}\sigma(i)p(i)}{\sum_{i=1}^{n}\sigma(i)} \tag{51}$$

$$\bar{q} = \frac{\sum_{i=1}^{n}\sigma(i)q(i)}{\sum_{i=1}^{n}\sigma(i)} \tag{52}$$

Thus,

$$\rho(r,\hat{t},s) = \frac{1}{s}\sum_{i=1}^{n}\|\sigma(i)(r\tilde{p}(i)-s\tilde{q}(i))\|^2 \tag{53}$$

$$= \frac{1}{s}\sum_{i=1}^{n}\|\sigma(i)r\tilde{p}(i)\|^2 - 2\sum_{i=1}^{n}(\sigma(i)r\tilde{p}(i))^T\sigma(i)q(i) \tag{54}$$

$$+ s\sum_{i=1}^{n}\|\sigma(i)\tilde{q}(i)\|^2 \tag{55}$$

where

$$\tilde{p}(i) = p(i) - \bar{p} \tag{56}$$

$$\tilde{q}(i) = q(i) - \bar{q} \tag{57}$$

6 Representing Rotations

6.1 Axis Angle Represntation

The rotation matrix tr hat corresponds to a rotation of θ degrees (with sign determined by the right hand rule) about the unit vector n is as follows.

$$r = Icos\theta + (1-\cos\theta)\begin{pmatrix} n_1^2 & n_1n_2 & n_1n_3 \\ n_2n_1 & n_2^2 & n_2n_3 \\ n_3n_1 & n_3n_2 & n_3^2 \end{pmatrix} + \sin\theta\begin{pmatrix} 0 & -n3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix} \tag{58}$$

For example rotation of θ degrees about the X, Y, and Z axis are as follows

$$r_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \tag{59}$$

$$r_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \tag{60}$$

$$r_z(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{61}$$

6.2 Elevation Azymuth Tilt (roll) Representation

We will use the convention for X representing the horizontal axis, Y the vertical axis and Z representing depth.

The elevation, azymuth, tilt represenation is same as the axis-angle representation but with the unit vector n specified by the pan angle α and the tilt vector β. Thus

$$n_x = a\sin\alpha \tag{62}$$

$$n_y = a\sin\beta \tag{63}$$

$$n_z = \sin\gamma \tag{64}$$

where a is the magnitude of the projection of the unit vector n onto the horizontal plane.

6.3 Euler Angle Representation

The Euler (pronounced "Oiler") angles representation of a rotation matrix is its representation in terms of sequential rotations aboutt the X, Y, Z angles

$$r_{Euler}(\alpha,\beta,\gamma) = r_x(\alpha)r_y(\beta)r_z(\gamma) \tag{65}$$

6.4 Rotations using the Matrix Exponential

Let $\omega = [\omega_1 \;\; \omega_2 \;\; \omega_3]^T$ be a unit vector, and let θ be a scalar. Consider the following real-valued skew-symmetric matrix:

$$u = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}. \tag{66}$$

The matrix exponential $e^{u\theta}$ is defined by its Taylor explansion (**??**):

$$e^{u\theta} = I + u\theta + \frac{(u\theta)^2}{2!} + \frac{(u\theta)^3}{3!} + \cdots \tag{67}$$

It can be readily shown that $u^3 = -u$. Substituting $-u$ for u^3 in the preceding equation gives

$$e^{u\theta} = I + u(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots) + u^2(\frac{\theta^2}{2!} - \frac{\theta^4}{4!} + \frac{\theta^6}{6!} + \cdots) \tag{68}$$

Substituting $\sin\theta$ and $\cos\theta$ for their Taylor series, we get Rodrigues' formula:

$$e^{u\theta} = I + u\sin\theta + u^2(1 - \cos\theta). \tag{69}$$

Now consider the rotation of a point p by an angle θ, about an axis through the origin that is represented by the unit vector $\omega = [\omega_1 \;\; \omega_2 \;\; \omega_3]^T$, as shown in Figure 1. The component of p parallel to ω, $p_\omega^{\|}$, will be unchanged by the rotation. The component of p perpendicular to ω, $p_\omega^{\perp}$, will be rotated by the angle θ about the axis ω. Let p' be the rotated point location.

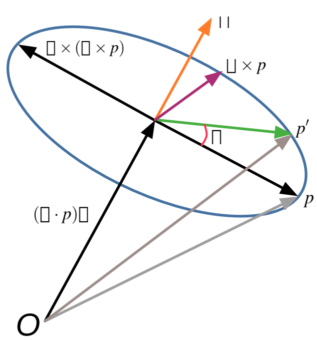

Figure 1: Geometry of rotation

$$\begin{aligned} p' &= p_\omega^{\|} + cos\theta(p_\omega^{\perp}) + sin\theta(\omega \times p) & (70)\\ &= (\omega \cdot p)\omega - cos\theta(\omega \times (\omega \times p)) + sin\theta(\omega \times p) & (71)\\ &= p + sin\theta(\omega \times p) + (1 - cos\theta)\omega \times (\omega \times p) & (72) \end{aligned}$$

Let u be the matrix defined in equation (66). It is easy to verify that u performs the cross product of ω with any 3D vector z, i.e.,

$$uz = \omega \times z \tag{73}$$

Substituting into (72) yields the rotation matrix, r, as a function of u and θ:

$$p' = rp \tag{74}$$

where

$$r = I + u \sin\theta + u^2(1 - \cos\theta) \tag{75}$$

or, by Rodrigues' formula (69),

$$r = e^{u\theta}. \tag{76}$$

Thus every 3D rotation matrix, r, can be expressed as e^m, where m is a 3×3 skew-symmetric matrix. Conversely, if m is any 3×3 skew-symmetric matrix, then e^m is a rotation matrix, since m can be written in the form $u\theta$, where θ is a scalar and u is a matrix that is related to a unit vector ω by (66).