The Motion Field and its Affine Approximation

Konstantinos G. Derpanis
kosta@cs.yorku.ca

December 12, 2004

In this report a summary of the *motion field* and its affine approximation is presented. The motion field is a purely geometric concept that represents the projection of the world velocities of scene points onto the image. The *optical flow field* (often erroneously confused with the motion field) is an estimate of the motion field derived from the apparent motion of the brightness patterns in the image. In the ideal case the optical flow field would correspond to the motion field. For a discussion of the correspondence of the two flow fields see [1].

1 World velocity description

The mathematical description of a 3D point (world point in camera coordinates) undergoing a rigid transformation about the camera axes follows.

Let ω_x, ω_y and ω_z represent the angle of rotation about the X, Y and Z axes respectively (see Fig. 1). An arbitrary rotation $\mathbf{R}$ is represented as:

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\omega_x) & -sin(\omega_x) \\ 0 & sin(\omega_x) & cos(\omega_x) \end{pmatrix} \begin{pmatrix} cos(\omega_y) & 0 & sin(\omega_y) \\ 0 & 1 & 0 \\ -sin(\omega_y) & 0 & cos(\omega_y) \end{pmatrix} \begin{pmatrix} cos(\omega_z) & -sin(\omega_z) & 0 \\ sin(\omega_z) & cos(\omega_z) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

Assuming infinitesimal rotations, the zeroth order terms of the Taylor series expansion of the trigonometric functions *sin* and *cos* provide the following approximations,

$$cos(\theta) \quad \approx \quad 1, \qquad sin(\theta) \quad \approx \quad \theta \tag{2}$$

Using the approximations in (2), $\mathbf{R}$ can be approximated as follows in terms of angular velocity,

$$\mathbf{R} \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\omega_x \\ 0 & \omega_x & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \omega_y \\ 0 & 1 & 0 \\ -\omega_y & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\omega_z & 0 \\ \omega_z & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \approx \begin{pmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{pmatrix} \tag{3}$$

Let the vector $\vec{T} = (t_x, t_y, t_z)^\top$ represent the translational velocity, where the elemental components t_x, t_y and t_z represent the translational velocities in the X, Y and Z directions respectively. The velocity $\vec{V} = (\dot{X}, \dot{Y}, \dot{Z})^\top$ of a point in the world $\vec{P} = (X, Y, Z)^\top$ with respect to camera coordinates undergoing a

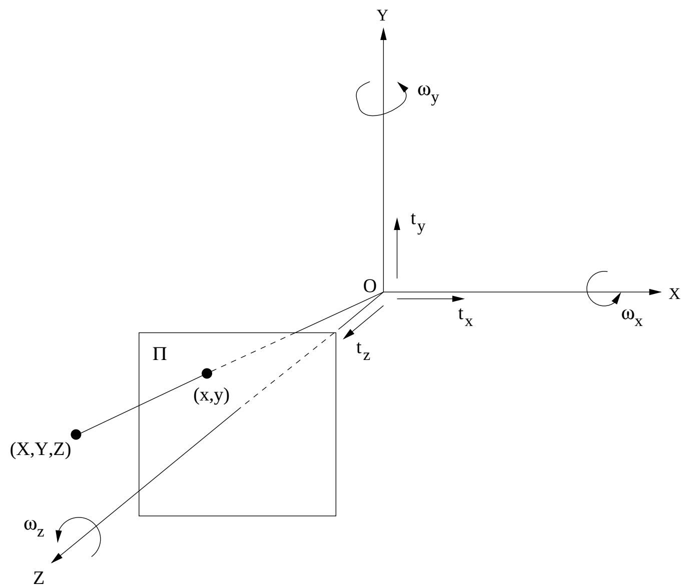

Figure 1: Camera coordinate system. Depicted is the camera coordinate system, with an image plane Π located at $Z = 1$. Perspective projection maps a point (X, Y, Z) to (x, y). The parameters t_x, t_y and t_z represent the translational velocities in the X, Y and Z directions respectively, ω_x, ω_y and ω_z represent the infinitesimal angle of rotation about X, Y and Z conducted about the point $\vec{Q} = (0, 0, 0)^\top$ (i.e. camera origin).

rigid transformation is represented as,

$$\vec{V} = (\mathbf{R} - \mathbf{I})\vec{P} + \vec{T} = \vec{T} + \Omega \times \vec{P} = \begin{pmatrix} \omega_y Z - \omega_z Y + t_x \\ \omega_z X - \omega_x Z + t_y \\ \omega_x Y - \omega_y X + t_z \end{pmatrix} \tag{4}$$

where $\mathbf{I}$ represents a 3 by 3 identity matrix.

In the case where the rotation of the object is about an arbitrary point in space $\vec{Q} = (q_x, q_y, q_z)^\top$ (assuming the coordinate systems are aligned) the transformation is represented by the following,

$$\vec{V} = \mathbf{R}(\vec{P} - \vec{Q}) + \vec{Q} + \vec{T} - \vec{P} = \begin{pmatrix} \omega_y(Z - q_z) - \omega_z(Y - q_y) + t_x \\ \omega_z(X - q_x) - \omega_x(Z - q_z) + t_y \\ \omega_x(Y - q_y) - \omega_y(X - q_x) + t_z \end{pmatrix} \tag{5}$$

2 Motion field description

The mathematical description of the motion field (i.e., image velocity) follows for a rigid transformation about the camera coordinate system as defined in Section 1. Assuming a perspective projection onto a plane parallel the X, Y axes and located at $Z = 1$ (without loss of generality the focal length $f = 1$), the relationship between an image point (x, y) and a scene point (X, Y, Z) is

$$x = \frac{X}{Z}, \qquad y = \frac{Y}{Z} \tag{6}$$

Differentiating (6) with respect to time yields,

$$\begin{aligned} \dot{x} &= u &= \frac{\dot{X}}{Z} - \frac{X\dot{Z}}{Z^2} \\ \dot{y} &= v &= \frac{\dot{Y}}{Z} - \frac{Y\dot{Z}}{Z^2} \end{aligned} \tag{7}$$

Substituting (4) and (6) into (7) results in the following,

$$\begin{aligned} u &= -\omega_x xy + \omega_y(x^2+1) - \omega_z y + \frac{t_x - t_z x}{Z} \\ v &= \omega_y xy - \omega_x(y^2+1) + \omega_z x + \frac{t_y - t_z y}{Z} \end{aligned} \tag{8}$$

Assuming that the imaged surface is a plane with surface normal $\vec{n} = (n_x, n_y, n_z)^\top$ and containing the point (X_0, Y_0, Z_0), provides the following constraint,

$$\alpha X + \beta Y + \gamma Z = 1 \tag{9}$$

or equivalently as

$$\alpha x + \beta y + \gamma = \frac{1}{Z} \tag{10}$$

where

$$\begin{aligned} \alpha &= \frac{n_x}{d} \\ \beta &= \frac{n_y}{d} \\ \gamma &= \frac{n_z}{d} \\ d &= n_x X_0 + n_y Y_0 + n_z Z_0 \end{aligned} \tag{11}$$

Substituting the planar constraint (10) into (8) leads to following formulation for the instantaneous velocities,

$$\begin{aligned} u &= a_0 + a_1 x + a_2 y + a_7 xy + a_6 x^2 \\ v &= a_3 + a_4 x + a_5 y + a_6 xy + a_7 y^2 \end{aligned} \tag{12}$$

where

$$\begin{aligned} a_0 &= t_x\gamma + \omega_y \\ a_1 &= t_x\alpha - t_z\gamma \\ a_2 &= t_x\beta - \omega_z \\ a_3 &= t_y\gamma - \omega_x \\ a_4 &= t_y\alpha + \omega_z \\ a_5 &= t_y\beta - t_z\gamma \\ a_6 &= \omega_y - t_z\alpha \\ a_7 &= -t_z\beta - \omega_x \end{aligned} \tag{13}$$

Through first order in image coordinates, we have the following affine model for the instantaneous velocity,

$$\begin{aligned} u &= a_0 + a_1 x + a_2 y \\ v &= a_3 + a_4 x + a_5 y \end{aligned} \tag{14}$$

References

[1] B.K.P. Horn. *Robot Vision.* MIT Press, Cambridge, MA, 1986.