Table of Contents

Image Motion

Image motion is a strong visual cue to 3D structure and allows to compute useful properties of a 3D scene without a-priori knowledge of the scene. The goal of motion analysis is to estimate 2D motion from sequences of images and then proceed to estimate the 3D translation and rotation of the observer (or sensor). Additionally, the 3D structure of the visible environment may also be extracted.

2D Motion Analysis

The problem is often formulated as an estimation of the apparent motion of local image brightness patterns. One may find correspondences of image features from frame to frame or solve differential equations locally to estimate optical flow.

2D motion analysis may be approached with differential methods (leading to dense motion analysis) or rely on matching methods (leading to sparse motion analysis).

However, for computations to work right, a few assumptions are needed. For instance, it is assumed that there is only one rigid motion between the camera and the scene, subtending the image area under analysis. Also, Lambertian shading is assumed.

The motion field is defined as a 2D vector field of velocities of the image points induced by relative motion.

Fundamentals

Suppose a 3D point $\vec{P}=(X,Y,Z)^T$ expressed in the reference system of the sensor. The perspective projection of this point onto the imaging plane is given by:

$$\vec{p} = f\frac{\vec{P}}{Z} = (x, y, f)^T$$

where f is the focal length of the sensor. Additionally, we can relate the motion between the sensor and point $\vec{P}$ as:

$$\vec{V} = -\vec{T} - \vec{\omega} \times \vec{P}$$

where $\vec{T} = (T_x, T_y, T_z)^T$ is the instantaneous rate of translation and $\vec{\omega} = (\omega_x, \omega_y, \omega_z)^T$ is the instantaneous rate of rotation. This equation can be expressed in vector components as:

$$\begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix} = \begin{pmatrix} -T_x - \omega_y Z + \omega_z Y \\ -T_y - \omega_z X + \omega_x Z \\ -T_z - \omega_x Y + \omega_y X \end{pmatrix}$$

From the perspective projection of the 3D point and the above 3D velocity equation we derive the motion field equations. First, derive the perspective equation with respect to time:

$$f\frac{\vec{P}}{Z}\frac{d}{dt} = f\left(\frac{Z\vec{V} - V_z\vec{P}}{Z^2}\right) = \vec{v}$$

since $\vec{P}\frac{d}{dt} = \vec{V}$ and $\vec{p}\frac{d}{dt} = \vec{v}$. Here $\vec{v}$ is known as optical (or image) flow. The time derivative of the 3D point projection is written in vector components as:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = f\begin{pmatrix} \frac{T_z x - T_x}{Z} - \omega_y + \omega_z y + \omega_x xy - \omega_y x^2 \\ \frac{T_z y - T_y}{Z} + \omega_x - \omega_z x - \omega_y xy + \omega_x y^2 \end{pmatrix}$$

As it may be observed, terms that depend on Z are decoupled from the terms that depend on $\vec{\omega}$. Hence, the rotational part of the motion field does not carry any information on depth.

Special Case: Pure Translation

In the case of pure translation, the rotational components of the motion field equations become null and the motion field reduces to:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} \frac{T_z x - T_x}{Z} \\ \frac{T_z y - T_y}{Z} \end{pmatrix}$$

Assume that the translation in the line of sight is non-null ($T_z \neq 0$), and consider:

$$\vec{p}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} \frac{T_x}{T_z} \\ \frac{T_y}{T_z} \end{pmatrix}$$

As a consequence, we have $T_x = x_0 T_z$ and $T_y = y_0 T_z$ and hence:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} \frac{(x - x_0) T_z}{Z} \\ \frac{(y - y_0) T_z}{Z} \end{pmatrix}$$

which means that the motion field of pure translation is radial around $\vec{p}_0 = (x_{0,} y_0)^T$, which we term the vanishing point.

- If $T_z < 0$: vectors point away from the vanishing point (focus of expansion)
- if $T_z > 0$: vectors point toward the vanishing point (focus of contraction)
- if $T_z = 0$: then we have:

  $$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} \frac{-T_x}{Z} \\ \frac{-T_y}{Z} \end{pmatrix}$$

  and the motion field vectors are parallel, with their magnitude proportional to the depth of the 3D points.

Special Case: A Moving Plane

Let's assume that the sensor is observing a planar surface described by a plane equation such as $\vec{n}^T\vec{P}=d$, where $\vec{n}^T=(n_x, n_y, n_z)$ is the normal vector to the plane, d is the distance between the plane and the origin of the reference system. The plane translates and rotates as described by $\vec{T}$ and $\vec{\omega}$. Hence, $\vec{n}$ and d are functions of time.

We develop the motion field equations for the plane by first noting that

$$\frac{Z\vec{p}}{f}=\vec{P}$$

We then multiply both sides with $\vec{n}^T$ to obtain the following equation:

$$\frac{Z\vec{n}^T\vec{p}}{f} = \vec{n}^T\vec{P} = d$$

We may now solve for Z and put the result back into the motion field equations, from which we obtain:

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{fd}\begin{pmatrix} a_1x^2+a_2xy+a_3fx+a_4fy+a_5f^2 \\ a_1xy+a_2y^2a_6fy+a_7fx+a_8f^2 \end{pmatrix}$$

where

$$\begin{array}{lllllll} a_1 & = & -d\,\omega_y+T_zn_x & & a_2 & = & d\,\omega_x+T_zn_y \\ a_3 & = & T_zn_z-T_xn_x & & a_4 & = & d\,\omega_z-T_xn_y \\ a_5 & = & -d\,\omega_y-T_xn_z & & a_6 & = & T_zn_z-T_yn_y \\ a_7 & = & -d\,\omega_z-T_yn_x & & a_8 & = & d\,\omega_x-T_yn_z \end{array}$$

Hence, the motion field equations are quadratic in the image coordinates (x, y, f) . In addition, this set of equations has a dual solution. In other words, two planar surfaces moving differently may generate an identical motion field. This duality may be expressed by formulating the second solution in terms of the first:

$$d' = d \quad \vec{n}' = \frac{\vec{T}}{\|\vec{T}\|} \quad \vec{T}' = \|\vec{T}\|\vec{n} \quad \vec{\omega}' = \vec{\omega}+\frac{\vec{n}\times\vec{T}}{d}$$

Optical Flow

The motion field equations describe image motion onto the imaging plane as a purely geometric quantity. Optical flow is defined as an approximation to the motion field, as it is measured from images in which photometric effects are the norm.

Under most circumstances, the apparent brightness of moving surfaces remains constant within small spatio-temporal image neighborhoods. A hypothesis such as this mathematically translates into:

$$I \frac{d}{dt} = 0$$

for an image I, over a relatively small subtending spatio-temporal region. We differentiate the above equation with the chain rule and obtain:

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$

Written in concise form, this partial differential equation becomes:

$$\nabla I^T \vec{v} + I_t \; = \; (I_x, I_y)\begin{pmatrix} v_x \\ v_y \end{pmatrix} + I_t \; = \; 0$$

where

$$I_x \; = \; \frac{\partial I}{\partial x} \quad I_y \; = \; \frac{\partial I}{\partial y} \quad I_t \; = \; \frac{\partial I}{\partial t} \quad v_x \; = \; \frac{dx}{dt} \quad v_y \; = \; \frac{dy}{dt}$$

This is generally known as the motion constraint equation and widely used as a base hypothesis to compute optical flow, a well known approximation to the motion field equation (also often called image motion).

The motion constraint equation represents a line in motion space (u, v). There are thus two unknowns and only one equation. This problem is said to be under-constrained. In the computer vision field, this situation is referred to as the aperture problem, where only normal motion v_n (that is, the perpendicular vector to the motion constraint equation line) is computable:

$$\frac{-I_t}{\|\nabla I\|} = \frac{\nabla I^T \vec{v}}{\|\nabla I\|} = v_n$$

As a consequence of this fact, full optical flow cannot be obtained locally in general. More constraints need to be added to the problem in order to compute full optical flow. There are many and one of them is to require that neighboring points on a surface have similar velocities (optical flow vectors). One early method to use this global (as opposed to local) constraint employed a minimization of the square of the optical flow variation over the entire image:

$$\left(\frac{\partial v_x}{\partial x}\right)^2+\left(\frac{\partial v_x}{\partial y}\right)^2+\left(\frac{\partial v_y}{\partial x}\right)^2+\left(\frac{\partial v_y}{\partial y}\right)^2 \to 0$$

Horn & Schunck's Algorithm

Horn and Schunck were the first ones to use a variational calculus approach to the problem of computing optical flow, the approximation to image motion. They devised an equation that included both the motion constraint equation and the smoothness term seen above:

$$\epsilon^2 = \iint \left(\alpha^2 \epsilon_c^2 + \epsilon_b^2\right) dx\, dy$$

where:

$$\epsilon_b = \nabla I^T \vec{v} + I_t$$

and

$$\epsilon_c = \left(\frac{\partial v_x}{\partial x}\right)^2+\left(\frac{\partial v_x}{\partial y}\right)^2+\left(\frac{\partial v_y}{\partial x}\right)^2+\left(\frac{\partial v_y}{\partial y}\right)^2$$

The algorithm is based on finding (u, v) that minimizes the integral.

Optical flow is best computed when the imagery respects a number of constraints, namely:

- Lambertian surfaces (absence of specular reflections)
- Point-wise light source at infinity
- Absence of photometric distortions

Computing Optical Flow with Least-Squares

For each point $\vec{p}_i$ within a small $N \times N$ neighborhood we term region Q , we compute spatio-temporal derivatives at points $\vec{p}_{1,} \vec{p}_{2,} \ldots, \vec{p}_{N^2}$. The task now is to find $\bar{v}$ which minimizes the functional:

$$\psi[\vec{v}] = \sum_{\vec{p}_i \in Q} \left[\nabla I^T \vec{v} + I_t \right]^2$$

This is a least-squares problem, and we can solve it as $A^T A \vec{v} = A^T \vec{b}$ where:

$$A = \begin{bmatrix} \nabla I|_{\vec{p}_1} \\ \nabla I|_{\vec{p}_2} \\ \vdots \\ \nabla I|_{\vec{p}_{N^2}} \end{bmatrix}$$

and

$$\vec{b} = -\left(I_t|_{\vec{p}_1}, I_t|_{\vec{p}_2}, \ldots, I_t|_{\vec{p}_{N^2}} \right)$$

The least-squares solution is obtained as $\bar{v} = (A^T A)^{-1} A^T \vec{b}$ where $\bar{v}$ is an estimate of the optical flow vector at the center of region Q .

Particularities

The 2×2 matrix

$$A^T A = \begin{bmatrix} \sum_Q I_x^2 & \sum_Q I_x I_y \\ \sum_Q I_x I_y & \sum_Q I_y^2 \end{bmatrix}$$

is singular of and only if the spatial gradients in neighborhood Q are null or parallel. This is the aperture problem and, in cases such as this, full optical flow cannot be estimated. The solution with minimum norm is that of the normal flow.

This is not the only problem plaguing differential optical flow techniques. For instance, the presence of multiple motions caused by occlusions or translucent surfaces will create errors. Robust methods are needed to cope with these phenomena. Some of them use variants of the EM algorithm, robust estimators, RANSAC, etc. Additionally, fast motions will require adaptive windowing or image pyramids to effectively deal with the problem. Last but not least, optical flow estimation accuracy has a profound impact on the quality of subsequent 3D structure and motion computations, generally rendering the reconstruction task nearly impossible in the general case.

Solving the Motion Field Equations

Once image motion has been estimated, the next task is to infer 3D motion and structure. Consider a planar patch expressed in 3D coordinates as:

$$Z = Z_0 + p X + q Y$$

and in image coordinates as:

$$Z^{-1} = Z_0^{-1}(1 - p x - q y)$$

Also consider a second-order Taylor series expansion of optical flow in the image coordinates:

$$\vec{v}(x, y) = \sum_{i=0}^{2} \sum_{j=0}^{2} \vec{v}^{(i,j)} \frac{x^i y^i}{i!\, j!}$$

under the condition that $i + j \leq 2$, where

$$\vec{v}^{(i,j)} = \left. \frac{\partial^{i+j} \vec{v}}{\partial x^i \partial y^j} \right|_{(0,0)}$$

represents the optical flow in terms of 12 quantities. Waxman and Ullman related the first 8 of these to the 3D motion and structure parameters for a planar surface, in the following way:

$$\begin{array}{rclrcl}
v_x^{(0,0)} & = & -(T_x + \omega_y) & v_y^{(0,0)} & = & -(T_y - \omega_x) \\
v_x^{(1,0)} & = & T_z + p T_x & v_y^{(0,1)} & = & T_z + q T_y \\
\frac{1}{2}[v_y^{(1,0)} + v_x^{(0,1)}] & = & \frac{1}{2}(p T_y + q T_x) & \frac{1}{2}[v_y^{(1,0)} - v_x^{(0,1)}] & = & -\omega_z + \frac{1}{2}(p T_y - q T_x) \\
v_x^{(2,0)} & = & -2(\omega_y + p T_z) & v_x^{(1,1)} & = & \omega_x - q T_z
\end{array}$$

There is a closed-form solution for $\vec{T}$ and $\vec{\omega}$. By elimination, one obtains a cubic equation for which the middle root yields T_z . The remaining variables follow by back substitution. The cubic equation containing T_z is:

$$T_z^3 + C_1 T_z^2 + C_2 T_z + C_3 = 0$$

where:

$$\begin{aligned} C_1 &= -(D_3+D_4) \\ C_2 &= -\frac{1}{4}\left[\left(D_1-\frac{D_7}{2}\right)^2+(D_2-D_8)^2\right]-D_5^2+D_3 D_4 \\ C_3 &= \frac{1}{4}D_4\left(D_1-\frac{D_7}{2}\right)^2+\frac{1}{4}D_3(D_2-D_8)^2-\frac{1}{2}D_5\left(D_1-\frac{D_7}{2}\right)(D_2-D_8) \end{aligned}$$

and where:

$$\begin{aligned} D_1 &= -(T_x+\omega_y) & D_2 &= -(T_y-\omega_x) \\ D_3 &= T_z+pT_x & D_4 &= T_z+qT_y \\ D_5 &= \frac{1}{2}(pT_y+qT_x) & D_6 &= -\omega_z+\frac{1}{2}(pT_y-qT_x) \\ D_7 &= -2(\omega_y+pT_z) & D_8 &= \omega_x-qT_z \end{aligned}$$

It has been shown that the all the roots of this polynomial are always real, and that the middle root is invariably the solution to T_z .

Finding the planar patch orientation is also a dual solution and it is given by the following equations:

$$(p_1, p_2) = \frac{\left(D_1-\frac{D_7}{2}\right)\pm\sqrt{\left(D_1-\frac{D_7}{2}\right)-4T_z(T_z-D_4)}}{2T_z}$$

$$(q_1, q_2) = \frac{(D_2-D_8)\pm\sqrt{(D_2-D_8)-4T_z(T_z-D_4)}}{2T_z}$$

What remains to be determined is what combinations constitute parts of the dual solution. Consider:

$$S=4D_5T_z+\left(D_1-\frac{D_7}{2}\right)(D_2-D_8)$$

Then, the dual solutions for the planar surface orientations are given by:

- if $S\geq 0$: the set $\{(p_1,q_1),(p_2,q_2)\}$ is the dual solution

- otherwise, the set $\{(p_{1,} q_2), (p_{2,} q_1)\}$ is the dual solution

For each acceptable planar surface solution (p, q) the other 3D motion parameters are obtained as:

$$
\begin{aligned}
T_x &= p\,T_z - \left(D_1 - \frac{D_7}{2}\right)\\
T_y &= q\,T_z - (D_2 - D_8)\\
\omega_x &= D_2 + T_y\\
\omega_y &= -D_1 - T_x\\
\omega_z &= -D_6 + \frac{1}{2}(p\,T_y - q\,T_x)
\end{aligned}
$$

As before, the two solutions $(\vec{T}, \vec{\omega}, p, q)$ and $(\vec{T}', \vec{\omega}', p', q')$ can be obtained from each other:

$$
\begin{aligned}
p' &= -\frac{T_x}{T_z}\\
q' &= -\frac{T_y}{T_z}\\
T'_x &= p\,T_z\\
T'_y &= -q\,T_z\\
T'_z &= T_z\\
\omega'_x &= \omega_x + (T_y + q\,T_z)\\
\omega'_y &= \omega_y + (T_x + p\,T_z)\\
\omega'_z &= \omega_z + (p\,T_y - q\,T_x)
\end{aligned}
$$