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:

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

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:

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

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:

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

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

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

As a consequence, we have $T_x = \frac{x_0 T_z}{f}$ and $T_y = \frac{y_0 T_z}{f}$ and hence:

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

which means that the motion field of pure translation is radial around $\vec{p}_0$, 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:

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

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:

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

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 .

Spectral Analysis of Motion

Spectral analysis of images is an important aspect of image understanding. The spectral characteristics of imagery help in determining filtering and noise-removal modalities. Further, this type of analysis is demonstrably useful in the analysis of image motion.

Special Functions

The Dirac-delta function plays an important role in spectral analysis. One operational way of defining the Dirac-delta function is to consider a Gaussian function in the limit of its variance going toward zero:

$$\delta(x-x_0)=\lim_{b\to 0}\frac{1}{|b|}G\left(\frac{x-x_0}{b}\right)$$

where

$$G\left(\frac{x-x_0}{b}\right)=\exp\left\{-\pi\left(\frac{x-x_0}{b}\right)^2\right\}$$

Properties of Dirac-delta Functions

- $\delta(x-x_0)=0 \qquad x\neq x_0$
- $\int_{x_1}^{x_2} f(\alpha)\delta(\alpha-x_0)d\alpha=f(x_0)$
- $\delta\left(\frac{x-x_0}{b}\right)=|b|\delta(x-x_0)$
- $\delta(ax-x_o)=\frac{1}{|a|}\delta\left(x-\frac{x_0}{a}\right)$
- $\delta(-x+x_0)=\delta(x-x_0)$
- $\delta(-x)=\delta(x)$
- $f(x)\delta(x-x_0)=f(x_0)\delta(x-x_0)$
- $x\delta(x-x_0)=x_0\delta(x-x_0)$
- $\delta(x)\delta(x-x_0)=0 \qquad x_0\neq 0$
- $\delta(x-x_0)\delta(x-x_0)$ is undefined
- $\int \delta(\alpha-x_0)d\alpha=1$
- $\int A\delta(\alpha-x_0)d\alpha=A$
- $\int \delta(\alpha-x_0)\delta(x-\alpha)d\alpha=\delta(x-x_0)$

Derivatives of the Dirac-delta Function

The derivative of the Dirac-delta function written as:

$$\delta^{(k)}=\frac{d^k\delta(x)}{dx^k}$$

If we choose a function $f(x)$ whose k^{th} derivative is known to be:

$$f^{(k)}(x)=\frac{d^k f(x)}{dx^k}$$

then we can define the derivatives of the Dirac-delta function as:

$$\delta^{(k)}(x-x_0)=0 \qquad x\neq x_0$$

and

$$\int f(\alpha)\delta^{(k)}(\alpha-x_0)\,d\alpha=(-1)^k f^{(k)}(x_0)$$

which sifts out the derivative of $f(x)$ at x_0 . Also note that $\int \delta^{(k)}(\alpha)\,d\alpha=0$.

General Dirac-delta Functions

In general, we refer to the one-dimensional Dirac-delta function as a point-mass. In 2D, a point mass Dirac-delta is written as: $\delta(x-x_{0,}\,y-y_0)$ while a line-mass Dirac-delta function is defined as: $\delta(a_1 x+b_1 y+c_1)$ and is non-zero along $a_1 x+b_1 y+c_1=0$. Alternatively, a general point-mass Dirac-delta function can be defined as: $\delta(a_1 x+b_1 y+c_{1,}\,a_2+b_2 y+c_2)$ and is non-zero at the intersection of $a_1 x+b_1 y+c_1=0$ and $a_2 x+b_2 y+c_2=0$, provided that it exists.

Harmonic Analysis

A function $f(x)$ can be expressed as a Fourier series if it satisfies Dirichlet conditions, which are:

- $f(x)$ is single-valued
- it has a finite number of maxima and minima
- it has a finite number of finite discontinuities
- it is absolutely integrable: $\int |f(x)|\,dx<\infty$

Then we can write $f(x)$ as a Fourier series expansion

$$f(x)=\sum_{n=-\infty}^{\infty} C_n e^{ik_0 x n}$$

where $k_0=2\pi f_o$. f_o is the fundamental frequency of $f(x)$. This Fourier series expansion may also be written as:

$$f(x)=\sum_{n=-\infty}^{\infty} C_n(\cos(k_0 x n)+i\sin(k_0 x n))$$

as per Euler's formula: $e^{i\theta}=\cos\theta+i\sin\theta$.

A Simple Fourier Integral

Let's compute the Fourier transform of a function that has the simplest frequency characteristics, such as $f(x)=A\cos(k_0 x)$:

$$\hat{f}(k)=A\int\cos(k_0 x)e^{-ikx}\,dx$$

From Euler's formula, we have:

$$\cos\theta=\frac{e^{i\theta}+e^{-i\theta}}{2}$$

and hence we can write:

$$\begin{aligned}\hat{f}(k) &= \frac{A}{2}\int\left[e^{ik_0x}+e^{-ik_0x}\right]e^{-ikx}\,dx\\ &= \frac{A}{2}\int e^{-ix(k-k_0)}\,dx+\frac{A}{2}\int e^{-ix(k+k_0)}\,dx\\ &= \frac{A}{2}\left[\delta(k-k_0)+\delta(k+k_0)\right]\end{aligned}$$

Consider a more complex function $f(x)$ for which our only requirement is that it satisfies Dirichlet conditions. Then, we can use a Fourier series expansion to express it as follows:

$$f(x)=\sum_{n=-\infty}^{\infty} C_n\left[\cos(k_0 x n)+i\sin(k_0 x n)\right]$$

Or, using Euler's formula:

$$f(x)=\sum_{n=-\infty}^{\infty} C_n e^{ik_0xn}$$

Then, the Fourier transform of this Fourier series expansion is given by:

$$\begin{aligned}\hat{f}(k) &= \int\left[\sum_{n=-\infty}^{\infty} C_n e^{i k_0 x n}\right] e^{-ikx}\,dx \\ &= \sum_{n=-\infty}^{\infty}\int C_n e^{-ix(k-k_0 n)}\,dx \\ &= \sum_{n=-\infty}^{\infty} C_n \delta(k-k_0 n)\end{aligned}$$

Hence, the Fourier transform of a function satisfying Dirichlet conditions is a train of Dirac-delta functions controlled by amplitude coefficients C_n .

Spectral Image Motion

Suppose we have a 1D signal translating with velocity v :

$$f(x)=I_0(x-vt)$$

where $f(x)$ satisfies Dirichlet conditions. Since it is translating in time, we can describe this signal as a 2D signal:

$$I(x,t)=I_0(x-vt)$$

Taking the Fourier transform of this signal yields:

$$\begin{aligned}\hat{I}(k,\omega) &= \iint I_0(x-vt)e^{-ikx}e^{-i\omega t}\,dx\,dt \\ &= \int\left[\int I_0(x-vt)e^{-ikx}\,dx\right]e^{-i\omega t}\,dt \\ &= \int\left[\hat{I}_0(k)e^{-ikvt}\right]e^{-i\omega t}\,dt \\ &= \hat{I}_0(k)\int e^{-it(kv+\omega)}\,dt \\ &= \hat{I}_0(k)\delta(kv+\omega)\end{aligned}$$

where $\hat{I}_0(k)$ is the Fourier transform of the signal (when not translating) and $\delta(kv+\omega)$ is a line-mass Dirac-delta function. As can be easily observed, the slope of $kv+\omega=0$ is the velocity of the translating signal.

We repeat the same exercise this time with a translating 2D image signal $I(\vec{x})$:

$$\begin{aligned}I(\vec{x},t) &= I_0(\vec{x}-\vec{v}t) \\ \hat{I}(\vec{k},\omega) &= \iint I(\vec{x},t)e^{\vec{x}^T\vec{k}+\omega t}\,d\vec{x}\,dt \\ &= \iint I_0(\vec{x}-\vec{v}t)e^{-i(\vec{x}^T\vec{k}+\omega t)}\,d\vec{x}\,dt \\ &= \int\left[\int I_0(\vec{x}-\vec{v}t)e^{-i\vec{x}^T\vec{k}}\,d\vec{x}\right]e^{-i\omega t}\,dt \\ &= \hat{I}_0(\vec{k})\int e^{-it(\vec{k}^T\vec{v}+\omega)}\,dt \\ &= \hat{I}_0(\vec{k})\delta(\vec{k}^T\vec{v}+\omega)\end{aligned}$$

where $\hat{I}_0(\vec{k})$ is the Fourier transform of the 2D signal $I_0(\vec{x})$. This result shows that for a translating signal, all the non-zero frequencies lie on a plane in the frequency domain described by $\vec{k}^T\vec{v}+\omega=0$. The orientation of this plane yields the 2D velocity of the translating signal. In addition, the Fourier transform of the optical flow equation is consonant with this result, as it is obtained in the following way:

$$F[\nabla I\vec{v}+I_t]=i\hat{I}(\vec{k},\omega)\delta(\vec{k}^T\vec{v}+\omega)$$

Gabor Filtering

It would be convenient to be able to sample the frequency spectrum of a sequence of images locally in order to estimate the amount of translation (image motion). The idea behind Gabor filters is to sample the spectrum within a spatio-temporal neighborhood without taking a Fast Fourier Transform of the entire image sequence.

A Gabor filter responds to a pre-determined frequency if it is present under the area of convolution. Gabor filters are Gaussian-windowed complex exponential functions of the form:

$$B(\vec{x},t,\vec{k}_0,\omega_0)=e^{i(\vec{x}^T\vec{k}_0+\omega_0 t)}G(\vec{x},t;\sigma)$$

In the frequency spectrum, such a Gabor filter is an un-normalized Gaussian centered at $(\vec{k}_{0,}\omega_0)$ with standard deviation $\hat{\sigma}=\frac{1}{\sigma}$. Hence, the convolution of a Gabor filter with a spatio-temporal region of an image sequence is the equivalent of multiplying the local frequency spectrum with a Gaussian centered at $(\vec{k}_{0,}\omega_0)$. It is instructive to take the Fourier transform of a Gabor filter to understand its spectral characteristics:

$$F[B(\vec{x},t,\vec{k}_{0,}\omega_0)]=(2\pi)^{\frac{1}{3}}\sigma^3[\int[\int[\int e^{\frac{-x^2}{2\sigma^2}-ix(k_x-k_{x0})}dx]e^{\frac{-y^2}{2\sigma^2}-iy(k_y-k_{y0})}dy]e^{\frac{-t^2}{2\sigma^2}-it(\omega-\omega_0)}dt]$$

We observe that the integral in x solves as:

$$\int\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-x^2}{2\sigma^2}-ix(k_x-k_{x0})}dx = e^{\frac{-(k_x-k_{x0})^2}{2\sigma^2}}$$

(and similarly for y and t). Hence, the Fourier transform of a Gabor filter is a

Gaussian centered at $(\vec{k}_{0,}\omega_0)$ in the frequency spectrum.