CS 280 Computer Vision
Spring 2018 Efros, Malik, Yu Lecture 5

1 Lecture

Optical Flow

In dynamic perspective, we assume it's not the world that moves, but the camera. Also, there's a fixed relationship between the world coordinate system at the optical center and the coordinate system on the image plane. Note that as the camera moves, points in the image will seem to move too.

We call (x, y) a parameterization of the image plane, i.e. an index into the camera sensor array. Meanwhile, (X, Y, Z) represents a scene point seen in the moving world coordinate system. If camera motion is a rigid motion involving a translation t and a rotation ω (each a 3×1 vector with (X, Y, Z) components), then we have the equation

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = -t - \omega \wedge \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

describing the relative motion of a scene point in the 3D world. It is $-t$ because (e.g.) when the camera moves to the left, the scene point appears to move to the right. We also have the equation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} \frac{f}{Z}$$

to describe the projection. f is the focal length of the camera; Z is the Z-coordinate of p (the point in the moving coordinate system). Z is actually a function of X and Y (because the coordinates always move together).

We want to use these equations to derive optical flow (where we see points on the image plane moving in a certain way). If we have the projection of a point onto the 2D film, it's how that point moves (the velocity):

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \Delta x / \Delta t \\ \Delta y / \Delta t \end{bmatrix}$$

From before, we can see that

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \frac{d}{dt} \begin{bmatrix} -fX/Z \\ -fY/Z \end{bmatrix}$$

If we carry out the decomposition, we will be able to write optical flow in terms of the camera's translation and rotation vectors, specifically in terms of the first translation component and the second rotation component.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} + \begin{bmatrix} xy & -(1+x^2) & y \\ 1+y^2 & -xy & -x \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

Optical flow varies with (x, y); different points have different flow vectors.

An optical flow field is just a bunch of velocity vectors, so it won't give us actual distances. On the other hand, it *will* give us quantities like "time to impact." The flow fields are scale-agnostic: as long as the ratios remain unchanged, the actual scale of the scene can be anything and the optical flow will feel the same!

If an optical flow field is off-center (vectors don't radiate outward from the center), we know there's a translation component along X and/or Y. If the vectors always point outward (or inward) from the focus of expansion (FOE), we know it's only pure translation.

If there's only rotation, we can determine ω from the optical flow field. However, this isn't really very interesting to us (we ourselves control the camera, after all), and it tells us nothing about the content of the scene.

Radiometry of Image Formation

A typical image consists of three arrays of brightness values (RGB) – the intensity of light hitting our sensors. The challenge of vision is to see these pixels as groups that are meaningful to us. However, first we should understand how these intensity values correspond to physical quantities in the world.

An image $I(x, y)$ measures how much light is captured at pixel (x, y). The entire imaging process would like to know (a) where a point (X, Y, Z) in the world gets imaged (this is the geometry problem; we've learned the relevant projection equations), and (b) the brightness at the resulting point (x, y). We need to count the photons hitting that point (x, y) in a time interval; $I(x, y)$ will be proportional to this quantity.

We use the scientific term *irradiance* for this concept: the radiant power per unit area, measured in units W/m^2 (watts per meters squared) and denoted by E.