6.801/6.866: Machine Vision, Lecture 2

Professor Berthold Horn, Ryan Sander, Tadayuki Yoshitake
MIT Department of Electrical Engineering and Computer Science
Fall 2020

These lecture summaries are designed to be a review of the lecture. Though I do my best to include all main topics from the lecture, the lectures will have more elaborated explanations than these notes. Therefore, if you're looking for the most rigorous review and treatment of these topics, we encourage you to rewatch the lecture videos. With that said, we hope these summaries are beneficial for your learning. If you have any feedback for these lecture summaries, please submit it **here**.

1 Lecture 2: Image Formation, Perspective Projection, Time Derivative, Motion Field

Definition of perspective projection:

$$\frac{x}{f} = \frac{X}{Z}, \frac{y}{f} = \frac{Y}{Z} \textbf{ (component form)} \tag{1}$$

$$\frac{1}{f}\mathbf{r} = \frac{1}{\mathbf{R}\cdot\hat{\mathbf{z}}}\mathbf{R} \textbf{ (vector form)} \tag{2}$$

If we differentiate these perspective projection equations:

$$\frac{1}{f}\frac{dx}{dt} = \frac{1}{Z}\frac{dX}{dt} - \frac{X}{Z^2}\frac{dZ}{dt}$$

What are these derivatives? They correspond to **velocities**. Let's define some of these velocities:

- $u \triangleq \frac{dx}{dt}$
- $v \triangleq \frac{dy}{dt}$
- $U \triangleq \frac{dX}{dt}$
- $V \triangleq \frac{dY}{dt}$
- $W \triangleq \frac{dZ}{dt}$

Now, rewriting the differentiated perspective projection equations with these velocity terms, we first write the equation for the x component:

$$\frac{1}{f}u = \frac{1}{Z}U - \frac{X}{Z^2}W \tag{3}$$

Similarly, for y:

$$\frac{1}{f}v = \frac{1}{Z}V - \frac{Y}{Z^2}W \tag{4}$$

Why are these equations relevant? They allow us to find parts of the image that don't exhibit any motion - i.e. stationary points. Let's find where $U = V = 0$. Let the point (x_0, y_0) correspond to this point. Then:

$$\frac{x_0}{f} = \frac{U}{W}, \frac{y_0}{f} = \frac{V}{W} \tag{5}$$

Focus of Expansion (FOE): Point in image space given by (x_0, y_0). This point is where the 3D motion vector intersects with the line given by $z = f$.

Why is FOE useful? If you know FOE, you can derive the **direction of motion** by drawing a vector from the origin to FOE.

Additionally, we can rewrite the differentiated perspective projection equations with FOE:

$$\frac{1}{f}u = \frac{x_0 - x}{f}\frac{W}{Z} \textbf{ (x comp.)}, \quad \frac{1}{f}v = \frac{y_0 - y}{f}\frac{W}{Z} \textbf{ (y comp.)} \tag{6}$$

Cancelling out the focal length (f) terms:

$$u = (x_0 - x)\frac{W}{Z} \textbf{ (x comp.)}, \quad v = (y_0 - y)\frac{W}{Z} \textbf{ (y comp.)} \tag{7}$$

A few points here:

- You can draw the vector diagram of the motion field in the image plane.
- All vectors in the motion field expand outward from **FOE**.
- Recall that perspective projection cannot give us absolute distances.

For building intuition, let's additionally consider what each of these quantities mean. The inverse term $\frac{Z}{W} = \frac{Z}{\frac{dZ}{dt}}$ has units of $\frac{\textbf{meters}}{\frac{\textbf{meters}}{\textbf{second}}} = \textbf{seconds}$ - i.e. **Time of Impact**.

Let's now revisit these equations in vector form, rather than in the component form derived above:

$$\frac{1}{f}\frac{d\mathbf{r}}{dt} = \frac{1}{\mathbf{R}\cdot\hat{\mathbf{z}}} - \frac{R}{(\mathbf{R}\cdot\hat{\mathbf{r}})^2}\frac{d}{dt}(\mathbf{R}\cdot\hat{\mathbf{r}}) \tag{8}$$

Let's rewrite this with dots for derivatives. Fun fact: The above notation is Leibniz notation, and the following is Newtonian notation:

$$\frac{1}{f}\dot{\mathbf{r}} = \frac{1}{\mathbf{R}\cdot\hat{\mathbf{z}}}\dot{\mathbf{R}} - \frac{R}{(\mathbf{R}\cdot\hat{\mathbf{z}})^2}(\dot{\mathbf{R}}\cdot\hat{\mathbf{z}}) \tag{9}$$

$$\frac{1}{f}\dot{\mathbf{r}} = \frac{1}{Z}(\dot{\mathbf{R}} - W\frac{1}{f}\mathbf{r}) \tag{10}$$

One way for reasoning about these equations is that **motion is magnified by the ratio of the distance terms**.

Next, we'll reintroduce the idea of **Focus of Expansion**, but this time, for the vector form. FOE in the vector form is given at the point where $\dot{\mathbf{r}} = 0$:

$$\frac{1}{f}\dot{\mathbf{r}} = \frac{1}{W}\dot{\mathbf{R}} \tag{11}$$

We can use a dot product/cross product identity to rewrite the above expression in terms of cross products. The identity is as follows for any $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{c}\cdot\mathbf{a})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c} \tag{12}$$

Using this identity, we rewrite the expression above to solve for FOE:

$$\frac{1}{f}\dot{\mathbf{r}} = \frac{1}{(\mathbf{R}\cdot\hat{\mathbf{z}})^2}(\hat{\mathbf{z}}\times(\dot{\mathbf{R}}\times\mathbf{R})) \tag{13}$$

What is this expression? This is **image motion** expressed in terms of **world motion**. Note the following identities/properties of this motion, which are helpful for building intuition:

- $\dot{\mathbf{r}}\cdot\hat{\mathbf{z}} = 0 \implies$ Image motion is perpendicular to the z-axis. This makes since intuitively because otherwise the image would be coming out of/going into the image plane.
- $\dot{\mathbf{r}} \perp \hat{\mathbf{z}}$
- $\dot{\mathbf{R}} \parallel \mathbf{R} \implies \dot{\mathbf{r}} = 0$ (this condition results in there being no image motion).

1.1 Brightness and Motion

Let's now consider how brightness and motion are intertwined. Note that for this section, we will frequently be switching between continuous and discrete. The following substitutions/conversions are made:

- **Representations of brightness functions**: $E(x, y) \leftrightarrow E[x, y]$
- **Integrals and Sums**: $\int_x \int_y \leftrightarrow \sum_x \sum_y$
- **Brightness Gradients and Finite Differences**: $(\frac{\partial E}{\partial x}, \frac{\partial E}{\partial y}) \leftrightarrow (\frac{1}{\delta x}(E[k, e+1] - E[k, e])$

1.1.1 1D Case

$$\frac{dx}{dt} = U \implies \delta x = U\delta \tag{14}$$

By taking a linear approximation of the local brightness:

$$\delta E = E_x \delta x = u E_x \delta t \quad (\text{note here that } E_x = \frac{\partial E}{\partial x}) \tag{15}$$

Dividing each side by δt, we have:

$$uE_x + E_t = 0 \implies U = -\frac{E_x}{E_t} = -\frac{\frac{\partial E}{\partial t}}{\frac{\partial E}{\partial x}} \tag{16}$$

A couple of points about this:

- This 1D result allows us to recover motion from brightness.
- We can infer motion from a single point. However, this is only true in the 1D case.
- We can estimate from 1 pixel, but frequently, we have much more than 1 pixel, so why use just 1? We can reduce noise by estimating motion from many pixels through regression techniques such as Ordinary Least Squares (OLS).
- From statistics, the standard deviation of the motion estimates will be reduced by a factor of $\frac{1}{\sqrt{N}}$, where N is the number of pixels sampled for estimating motion.

Finite Difference approximation for E is given by:

$$E \approx \frac{1}{\delta x}(E(x + \delta x, t) - E(x, t)) \tag{17}$$

Motion estimation can be done through unweighted averaging:

$$\bar{u}_{\text{unweighted}} = \frac{1}{N} \sum_{i=1}^{N} \frac{-E_{t_i}}{E_{x_i}} \tag{18}$$

As well as weighted averaging:

$$\bar{u}_{\text{weighted}} = \frac{\sum_{i=1}^{N} w_i \frac{-E_{t_i}}{E_{x_i}}}{\sum_{i=1}^{N} w_i} \tag{19}$$

A quick check here: take $w_i = 1 \; \forall \; i \in \{1, ..., N\}$. Then we have that $\bar{u}_{\text{weighted}} = \frac{1}{N} \sum_{i=1}^{N} \frac{-E_t}{E_x} = \bar{u}_{\text{unweighted}}$.

Note that in the continuous domain, the sums in the weighted and unweighted average values are simply replaced with integrals.

1.1.2 2D Case

While these results are great, we must remember that images are in 2D, and not 1D. Let's look at the 2D case. First and foremost, let's look at the brightness function, since it now depends on x, y, and t: $E(x, y, t)$. The relevant partial derivatives here are thus:

- $\frac{\partial E}{\partial x}$ - i.e. how the brightness changes in the x direction.
- $\frac{\partial E}{\partial y}$ - i.e. how the brightness changes in the y direction.
- $\frac{\partial E}{\partial t}$ - i.e. how the brightness changes w.r.t. time.

As in the previous 1D case, we can approximate these derivatives with finite forward first differences:

- $\frac{\partial E}{\partial x} = E_x \approx \frac{1}{\delta x}(E(x + \delta x, y, t) - E(x, y, t))$
- $\frac{\partial E}{\partial y} = E_y \approx \frac{1}{\delta y}(E(x, y + \delta y, t) - E(x, y, t))$
- $\frac{\partial E}{\partial t} = E_t \approx \frac{1}{\delta t}(E(x, y, t + \delta t) - E(x, y, t))$

Furthermore, let's suppose that x and y are parameterized by time, i.e. $x = x(t), y = y(t)$. Then we can compute the First-Order Condition (FOC) given by:

$$\frac{dE(x, y, t)}{dt} = 0 \tag{20}$$

Here, we can invoke the chain rule, and we obtain the result given by:

$$\frac{dE(x, y, t)}{dt} = \frac{dx}{dt}\frac{\partial E}{\partial x} + \frac{dy}{dt}\frac{\partial E}{\partial y} + \frac{\partial E}{\partial t} = 0 \tag{21}$$

Rewriting this in terms of u, v, w from above:

$$uE_x + vE_y + E_t = 0 \tag{22}$$

Objective here: We have a time-varying sequence of images, and our goal is to find and recover motion.

To build intuition, it is also common to plot in velocity space given by (u, v). For instance, a linear equation in the 2D world corresponds to a line in velocity space. Rewriting the equation above as a dot product:

$$uE_x + vE_y + E_t = 0 \leftrightarrow (u, v) \cdot (E_x, E_y) = -E_t \tag{23}$$

Normalizing the equation on the right by the magnitude of the brightness derivative vectors, we obtain the **brightness gradient**:

$$(u, v) \cdot \left(\frac{E_x}{\sqrt{E_x^2 + E_y^2}}, \frac{E_y}{\sqrt{E_x^2 + E_y^2}} \right) = -\frac{E_t}{\sqrt{E_x^2 + E_y^2}} \tag{24}$$

What is the **brightness gradient**?

- A unit vector given by: $\left(\frac{E_x}{\sqrt{E_x^2+E_y^2}}, \frac{E_y}{\sqrt{E_x^2+E_y^2}} \right) \in \mathbb{R}^2$.
- Measures spatial changes in brightness in the image in the image plane x and y directions.

Isophotes: A curve on an illuminated surface that connects points of equal brightness (source: Wikipedia).

As we saw in the previous case with 1D, we don't want to just estimate with just one pixel. For multiple pixels, we will solve a system of N equations and two unknowns:

$$uE_{x_1} + vE_{y_1} + E_{t_1} = 0 \tag{25}$$
$$uE_{x_2} + vE_{y_2} + E_{t_2} = 0 \tag{26}$$

Rewriting this in matrix form:

$$\begin{bmatrix} E_{x_1} & E_{y_1} \\ E_{x_2} & E_{y_2} \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} -E_{t_1} \\ -E_{t_2} \end{bmatrix} \tag{27}$$

Solving this as a standard $Ax = b$ problem, we have:

$$\begin{bmatrix} U \\ V \end{bmatrix} = \frac{1}{(E_{x_1}E_{y_2} - E_{y_1}E_{x_2})} \begin{bmatrix} E_{y_2} & -E_{y_1} \\ -E_{x_2} & E_{x_1} \end{bmatrix} \begin{bmatrix} -E_{t_1} \\ -E_{t_2} \end{bmatrix} \tag{28}$$

Note that the expression given by $\frac{1}{(E_{x_1}E_{y_2} - E_{y_1}E_{x_2})}$ is the determinant of the partial derivatives matrix, since we are taking its inverse (in this case, simply a 2x2 matrix).

When can/does this fail? It's important to be cognizant of edge cases in which this motion estimation procedure/algorithm fails. Some cases to consider:

- When brightness partial derivatives / brightness gradients are parallel to one another $\leftrightarrow$ The determinant goes to zero $\leftrightarrow$ This corresponds to linear dependence in the observations. This occurs when $E_{x_1}E_{y_2} = E_{y_1}E_{x_2} \implies \frac{E_{y_1}}{E_{x_1}} = \frac{E_{y_2}}{E_{x_2}}$.

This issue can be mitigated by weighting the pixels as we saw in the 1D case above. However, a more robust solution is to search for a minima of motion, rather than the point where it has zero motion. The intuition here is that even if we aren't able to find a point of zero motion, we can still get as close to zero as possible. Mathematically, let us define the following objective:

$$J(u, v) \triangleq \int_{x \in \mathbb{X}} \int_{y \in \mathbb{Y}} (uE_x + vE_y + E_t)^2 dxdy \tag{29}$$

Then we now seek to solve the problem of:

$$u^*, v^* = \arg\min_{u,v} J(u, v) = \arg\min_{u,v} \int_{x \in \mathbb{X}} \int_{y \in \mathbb{Y}} (uE_x + vE_y + E_t)^2 dxdy \tag{30}$$

Since this is an unconstrained optimization problem, we can solve by finding the minima of the two variables using two First-Order Conditions (FOCs):

- $\frac{\partial J(u,v)}{\partial u} = 0$
- $\frac{\partial J(u,v)}{\partial v} = 0$

Here, we have two equations and two unknowns. When can this fail?

- When we have linear independence. This occurs when:
  - E = 0 everywhere
  - E = constant
  - $E_x = 0$
  - $E_y = 0$
  - $E_x = E_y$
  - $E_x = kE_y$
- When $E = 0$ everywhere (professor's intuition: "You're in a mine.")
- When $E_x, E_y = 0$ (constant brightness).
- Mathematically, this fails when: $\int_x \int_x E_x^2 \int_y \int_y E_y^2 - (\int_x \int_y E_x E_y)^2 = 0$

When is this approach possible? **Only when isophotes are not parallel straight lines - i.e. want isophote curvature/rapid turning of brightness gradients.**

Noise Gain: Intuition - if I change a value by this much in the image, how much does this change in the result?