2.7 Solving Regularization Problems with the Calculus of Variations

2.7.1 The Optical Flow Constraint Equation

Introduction

In this section we will encounter a different way to solve regularization problems. More specifically, we will see how Horn and Schunk apply a differential framework to the motion problem and how they solve it using the Calculus of Variations. The basic idea is to understand motion by assuming that images are "snapshots" of a continuous intensity function. That is, instead of a series of isolated images $I_1(x,y)$, $I_2(x,y)$, and so on, we assume that there exists a continuous intensity function $I(x,y,t)$ and each image is a sample of that function at discrete times $t_i = t_0 + i * \Delta t$. We will begin by defining the Motion Problem. Next we will discuss the optical flow constraint equation (OFCE) that will lead us to a regularization function that is expressed in terms of differentials. Horn uses the Euler criterion from the Calculus of Variations to minimize this regularization problem and to this end, we will talk a bit about the Calculus of Variations in general. Finally, we come back to images by translating this minimization method from the continuous domain into an algorithm that works on images that are sampled discretely in both the spatial and temporal domain.

The Motion Problem

In the stereo problem, we were given two images and our task was to figure out "what went where", or more technically, which points in the left image corresponded to which points in the right image. This is also called the *Visual Correspondence Problem*.

The Motion Problem is another instance of the Visual Correspondence Problem, i.e. the task is again that of finding corresponding pixels. The main difference is that for the Motion problem, the two images are taken one after each other with the same camera (usually $\frac{1}{30}$ of a second apart). In the Motion Problem, we are interested in reconstructing the motion of objects in the scene, or the motion of the camera with respect to the scene, or both. Another important difference to the Stereo Problem is that the disparity of points is now a two-dimensional vector field.

	Stereo	Motion
images taken by ...	2 cameras at same time t	1 camera at times t and t+Δt
cause of disparity	depth of objects	motion of objects or camera
output	matrix of 1D disparity values	matrix of 2D disparity vectors

Table 1: The major differences between Motion and Stereo.

The Optical Flow Constraint Equation

As mentioned in the introduction, we will now consider a continuous intensity function $I(x,y,t)$. We can define the instantaneous "brightness flow" caused by object or camera motion in terms of two functions $u(x,y)$ and $v(x,y)$ where $u(x,y) = \frac{dx}{dt}$ and $v = \frac{dx}{dt}$ are the

x- and y-components of the velocity, respectively. If a given point has intensity $I(x, y)$ at time t, then at time $t + \delta t$ this point will have moved to $I(x + \delta x, y + \delta y)$ where the displacements δx and δy are given by $\delta x = u\,\delta t$ and $\delta y = v\,\delta t$. Now we make an important assumption, namely that the objects in the image do not change their appearance but only move around. In other words, we assume constant brightness, which is expressed mathematically by the condition $\frac{dI}{dt} = 0$. This leads to the basic equation

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$$

If we perform a Taylor expansion and throw out nonlinear terms we get

$$I(x, y, t) = I(x, y, t) + \delta x \tfrac{\partial I}{\partial x} + \delta y \tfrac{\partial I}{\partial y} + \delta t \tfrac{\partial I}{\partial t}$$

We can rewrite this equation by writing the partial derivatives of I as $I_x = \frac{\partial I}{\partial x}$ etc.

$$I(x, y, t) = I(x, y, t) + \delta x\, I_x + \delta y\, I_y + \delta t\, I_t$$

Cancelling I(x,y,t) on both sides, dividing by δt and taking the limit as $\delta t \to 0$, we obtain

$$I_x \frac{dx}{dt} + I_y \frac{dy}{dt} + I_t = 0$$

which we can write even more concisely as

$$I_x u + I_y v + I_t = 0$$

This key equation is called the *Optical Flow Constraint Equation* (OFCE). It relates the measurable quantities I_x, I_y, and I_t to the image flow (u, v). This is good, but we are still left with a significant problem: at each point, we have one constraint equation, but two unknowns. Therefore, solutions to the OFCE are non-unique, and the problem ill-posed. Another problem is that we are assuming that δt is very small. Since most video cameras only generate images every thirtieth of a second or so, this limit is usually not approached in practice. This means that the dismissal of the higher order terms of the Taylor expansion is only a reasonable assumption for scenes with slow motion[1]. The sampling error in the spatial domain also leads to errors in the computation of the observables I_x and I_y.

Meaning of the OFCE

Consider (u, v) space. Values of (u, v) satisfying the OFCE lie along a line in velocity space (see Fig. 1 below):

$$(I_x, I_y) \cdot (u, v) = -I_t$$

Note that $(I_x, I_y) = \nabla I$ and that the gradient always points in the direction of maximum increase of the intensity. The vector (u, v) lies somewhere along a line perpendicular to ∇I, the so-called *constraint equation.*

The OFCE does not tell us exactly where along the constraint line the actual motion (u,v) lies (since the problem is underconstrained). How far along ∇I the constraint line lies

[1] Experience has shown that object movements of up to 2 pixels every one thirtiest of a second give good results under this assumtion.

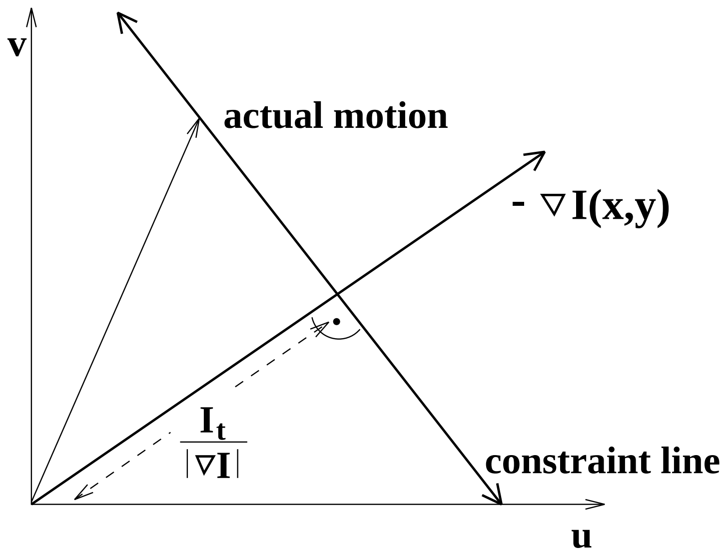

Figure 1: The geometric interpretation of the OFCE.

depends on I_t.

Understanding the details: Why the minus sign?

To test our understanding, we might ask ourselves why there is a minus sign in the equation

$$(I_x, I_y) \cdot (u, v) = -I_t.$$

The answer to this lies in the metaphor of a California Earthquake. We can picture the image plane as a contoured landscape, with large intensities corresponding to high elevations, and small intensities corresponding to low elevations. Image flow is analogous to an earthquake that shifts the landscape around in one of the compass directions. Now suppose that we are sitting in a helicopter that remains at fixed absolute height and position and watch the earthquake move the territory below us. We assume that the shape of the territory remains the same during the earthquake and is only shifted horizontally. Now suppose we observe after a while that we are closer to the ground. (This corresponds to an increase in intensity at a given point in our image.) How much did the landscape move horizontally under us? The answer depends on the local topography. If the area around us is relatively flat, then the horizontal motion must have been large. If the surrounding area is steep, then a small horizontal motion suffices. (See Fig. 2 below.)

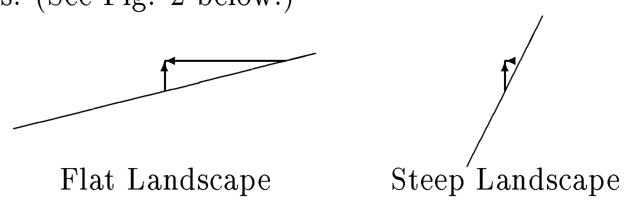

Figure 2. California Earthquakes.

We can now justify the presence of the minus sign in the equation above. $\nabla I = (I_x, I_y)$ points in the direction of maximum increase of the brightness function - "uphill" in our analogy. If the local height (intensity) increases, then the the actual motion of the terrain is opposite to the gradient, or in other words, it is moving in the "downhill" direction. (See Figure 3.)

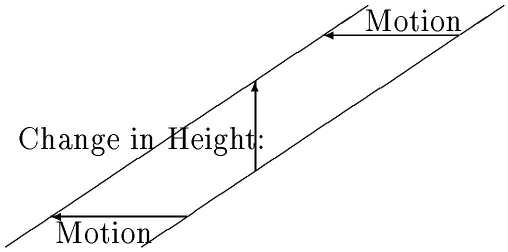

Figure 3. Effect of Motion Opposite the Gradient.

Understanding the details (2): Geometric interpretation of I_t

We have said already that the distance of the constraint line from the origin (as measured along the gradient) depends on I_t. This statement can easily be confirmed by recalling from Linear Algebra that the magnitude of the projection of a vector **a** onto a vector **b** is given by $(a \cdot \frac{b}{|b|})$. Notice that the OFCE tells us

$$(u, v) \cdot \nabla I = -I_t$$

By normalizing ∇I and exploiting the linearity of the inner product, we obtain

$$(u, v) \cdot -\frac{\nabla I}{|\nabla I|} = \frac{I_t}{|\nabla I|}$$

That is, the magnitude of the projection of the velocity vector (u,v) onto the negative gradient is $\frac{I_t}{|\nabla I|}$.

The Aperture Problem

The *aperture problem* focuses on the fact that given the optical flow constraint equation, it is only possible to constrain motion to lie along a single line. An effect of this is that motion that is perpendicular to the gradient of the energy function cannot be perceived. To see this, consider the simple example of looking at a piece of paper which varies evenly from being dark at the left edge of the paper to being light at the right edge of the paper. The paper is symmetric along the vertical axis and only varies horizontally. Now consider observing this piece of paper through a porthole. If we observe the image getting darker, we know that the piece of paper is moving to the right. However, we do not know whether it is also moving with some vertical component to the motion. All motions with equivalent horizontal components will appear identical to us.

Optical Flow

We have been referring to our solution (u, v) to the optical flow constraint equation as the motion occuring between the two images. This solution is also sometimes called the *normal* *flow component* of the optical flow. It should be noted that *optical flow* is an ambiguous term which is used to refer to several different phenomena. The term optical flow can be used to describe the solution (u, v) to the optical flow constraint equation, to describe any two-dimensional map indicating the motion of points in an image, or to describe changes in brightness in an image. These multiple meanings make it important to consider carefully in what sense the term "optical flow" is being used in a given context.

Regularization function for Motion

We introduced the OFCE because it would lead us to a regularization function for motion. Recall that the function to be minimized in a regularization problem is of the general form

$$E = \lambda e_{data} + e_{smooth}$$

where the data term e_{data} measures closeness of the solution to the given data and the smoothness term e_{smooth} measures how smooth the solution is. In the above, λ is a factor of the data term instead of the smoothness term as before. That does not make a big difference since we can still use it to adjust the relative influence of these two terms.

In Horn's formulation of the motion problem, the data term is given by the left hand side of the OFCE

$$e_{data}(u, v) = I_x u + I_y v + I_t$$

Recall that we derived the OFCE starting from the assumption

$$I(x, y, t) - I(x + \delta x, y + \delta y, t + \delta t) = 0$$

which is actually very similar to the data term

$$L(x, y) - R(x - D(x, y), y)$$

that we used for the Stereo problem . The major difference between these two data terms is that we consider two-dimensional displacements and phrase everything in terms of differentials, because we operate now in the continuous domain.

As before, the data term is not sufficient to define a unique solution and we add a smoothness term as additional constraint to make the problem well-posed. Horn is using

$$\begin{aligned} e_{smooth}(u, v) &= (u_x^2 + u_y^2) + (v_x^2 + v_y^2) \\ &= |\nabla u|^2 + |\nabla v|^2 \end{aligned}$$

Finally, we obtain the energy function for the whole image by integrating the data term and the smoothness term over x and y

$$E(u, v) = \iint \lambda e_{data} + e_{smooth}\, dx dy$$

There is also a nice geometrical interpretation of the data term. Remember that the optical flow constraint equation says that the e_{data} term should be equal to 0. In our energy function, it is being used to measure how far away a point is from the line that the motion is constrained to lie along, or how far our latest approximation is from the actual motion.

Horn solves this regularization problem with the help of the Calculus of Variations.

2.7.2 Calculus of Variations

The Calculus of Variations is not all that different from regular Calculus, which is mainly concerned with finding extrema of a function $f(x)$ that takes a value x and returns a value $f(x)$. In the Calculus of Variations however, we want to find extrema of functionals $I(f)$ that take a function f as input and return a value $I(f)$. In this introduction, we will consider functionals $I(f)$ of the form

$$I(f) = \int_{x_1}^{x_2} F(x, f, f')dx.$$

From Calculus we know that a necessary condition of an extremum x is that the function does not change very much locally at this point: $f'(x) = 0$ or $(f(x) - f(x + \epsilon)) \cong 0$ for 'very small' ϵ.

In the Calculus of Variations, there is a similar necessary condition for extrema: if I attains a minimum or maximum for input f, then a small change to f does not change $I(f)$ very much; this means that I is locally flat at 'point' f. A small change to f means that we add a small amount of another function g. Mathematically, we say that

$$|f(x) - (f(x) + \epsilon g(x))| \propto \epsilon^2$$

for an arbitrary function g(x).[2]

Given this, it is possible to derive the Euler equation by formulating the Taylor expansion of the above condition and performing some algebraic transformations.

Euler equations (Euler-Lagrange equations) *Let $I(f)$ be given as before by*

$$I(f) = \int_{x_1}^{x_2} F(x, f, f')dx.$$

Let $F_f = \frac{\partial}{\partial f}F$ and $F_{f'} = \frac{\partial}{\partial f'}F$. Then

$$F_f - \frac{d}{dx}F_{f'} = 0$$

is a necessary condition for extrema.

However, we are really interested in minimizing a more complicated function then $I(f)$, namely a function $I(f)$ of the form $\iint F(u, v, u_x, u_y, v_x, v_y)$. The two-dimensional generalization of the Euler equations for functionals of this form is

$$F_u - \frac{\delta}{\delta x}F_{u_x} - \frac{\delta}{\delta y}F_{u_y} = 0$$
$$F_v - \frac{\delta}{\delta x}F_{v_x} - \frac{\delta}{\delta y}F_{v_y} = 0$$

[2] Actually, there are some constraints on the nature of the function g, but they are not important for our purposes.

In our case,

$$F = \lambda(I_x u + I_y v + I_t)^2 + (u_x^2 + u_y^2) + (v_x^2 + v_y^2)$$

so the Euler equations yield

$$\nabla^2 u = \lambda(I_x u + I_y v + I_t)\, I_x$$
$$\nabla^2 v = \lambda(I_x u + I_y v + I_t)\, I_y$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

Equations like this usually do not have closed-form solutions but have to be solved iteratively.

2.7.3 Using the Calculus of Variations to solve the OFCE

In the preceding section, we have developed a necessary condition for the minimum (u^*, v^*) in the continuous domain. However, this is only of indirect interest since we need an algorithm for the discrete domain.

The Discrete Case

In the discrete case, we can calculate the smoothness of the motion field $(u,\ v)$ at the pixel $(i,\ j)$ as follows:

$$\begin{aligned} S_{ij} &= \frac{1}{4}((u_{i+1,\,j} - u_{i,\,j})^2 + (u_{i,\,j+1} - u_{i,\,j})^2 + && (1)\\ &\quad (v_{i+1,\,j} - v_{i,\,j})^2 + (v_{i,\,j+1} - v_{i,\,j})^2) && (2)\end{aligned}$$

The discrete form of the data term is given by

$$C_{ij} = (I_x\ u_{i,\,j} + I_y\ v_{i,\,j} + I_t)^2$$

where $I_x = I_{i,\,j+1} - I_{i,\,j}$ and $I_y = I_{i+1,\,j} - I_{i,\,j}$. What we would like is for

$$\sum\sum \underbrace{S_{i,\,j}}_{smoothness} + \underbrace{\lambda\, C_{i,\,j}}_{data}$$

to vanish, indicating the best smooth fit to the data. It turns out that the partial derivatives with respect to $u_{i,\,j}$ and $v_{i,\,j}$ vanish when the following conditions hold:

$$\begin{aligned} (1 + \lambda\, {I_x}^2)\, u_{i,\,j} \ + \ \lambda\, I_x\, I_y\, v_{i,\,j} &= \overline{u_{i,\,j}} \ - \ \lambda\, I_x\, I_t \\ (1 + \lambda\, {I_y}^2)\, v_{i,\,j} \ + \ \lambda\, I_x\, I_y\, u_{i,\,j} &= \overline{v_{i,\,j}} \ - \ \lambda\, I_y\, I_t \end{aligned}$$

where $\overline{u_{i,\,j}}$ and $\overline{u_{i,\,j}}$ are *local average velocity values*, or "typical values in the neighborhood" of pixel $(i,\ j)$.

Horn's method

This leads us (finally!) to Horn's iterative method for solving the Motion Problem:

$$u_{i,\,j}{}^{n+1} \;=\; \overline{u_{i,\,j}}{}^{\,n} \;-\; \alpha\, I_x \tag{3}$$

$$v_{i,\,j}{}^{n+1} \;=\; \overline{v_{i,\,j}}{}^{\,n} \;-\; \alpha\, I_y \tag{4}$$

where

$$\alpha \;=\; \frac{I_x\, \overline{u}^{\,n} \;+\; I_y\, \overline{v}^{\,n} \;+\; I_t}{1 \;+\; \lambda\, (I_x{}^2 \;+\; I_y{}^2)} \tag{5}$$

Note that n is not an exponent, but denotes the level of iteration (i.e., $u_{i,\,j}{}^n$ is the value of $u_{i,\,j}$ at iteration n).

One thing that is interesting about this method is that $u_{i,\,j}{}^{n+1}$ does not depend on $u_{i,\,j}{}^n$ but on $\overline{u_{i,\,j}}{}^{\,n}$, the local average around pixel $(i,\, j)$ at iteration n.

Another observation is that the numerator of α is the error term of our data. In other words, it tells us how close the current approximation is to a solution of the OFCE. Note that when $\alpha = 0$, the OFCE is solved.

Geometric Interpretation

There is an interesting interpretation of this method in the by now familiar image of the Optical Flow Constraint (OFC) line in $(u,\, v)$ space: We start with some guess $(\overline{u},\, \overline{v})$, a local average. With each update, we move towards the OFC line a distance that is:

- proportional to the distance from the OFC line
- approximately inversely proportional to $|\,\nabla I\,|$
- approximately inversely proportional to λ

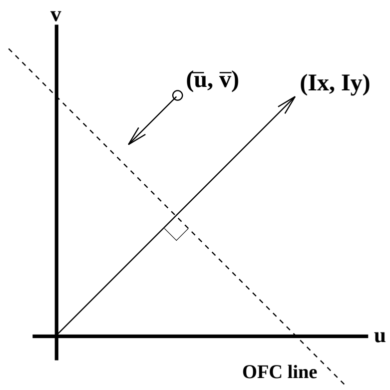

Figure 4: The geometric interpretation of the algorithm