Bayesian Approach to Motion Estimation

Konstantinos G. Derpanis
kosta@cs.yorku.ca

Version 1.2

January 19, 2005

In this note the Bayesian approach to motion estimation is summarized. For a more detailed treatment of the subject see [4, 6].

The central idea behind the Bayesian approach as one might guess is Bayes' theorem, formally stated,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

where, $P(A|B)$ denotes the posterior probability distribution function (PDF), $P(B|A)$ the likelihood PDF, $P(A)$ the prior, and $P(B)$ a normalization factor (hereafter omitted). Bayes' theorem can be thought of as a means of reversing the likelihood statement. The goal of Bayesian approaches for the problem at hand is to calculate the posterior probability of velocity (v_x, v_y) given image data I [6],

$$P(v_x, v_y|I) \propto P(I|v_x, v_y)P(v_x, v_y). \tag{2}$$

A velocity estimate is often attributed to the maximum of the posterior (commonly referred to as the the maximum a posteriori or MAP for short). An advantage of casting the motion estimation problem in a Bayesian framework is that it not only yields a velocity estimate but also allows for the quantization of the uncertainty of the best estimate, where uncertainty is related to the inevitable occurrence of the aperture problem and noise in the observation data.

The starting point of the Bayesian approach to motion estimation is the specification of the likelihood PDF. For the case of translational constant velocity motion, the image data in the form of the gradient of the image sequence $\nabla I(x, y, t)$ is related to image velocity by assuming brightness constancy of image points over time and local linearity of the image sequence in all dimensions [1]. This relationship, commonly termed the brightness constancy constraint, is given formally as [1],

$$\nabla I(x, y, t) \cdot (v_x(x, y, t), v_y(x, y, t), 1)^\top = 0 \tag{3}$$

where $\nabla I(x, y, t) = (I_x(x, y, t), I_y(x, y, t), I_t(x, y, t))^\top$ denotes the gradient that encapsulates the partial derivatives of the image with respect to the spatial x, y and temporal t parameters and $\mathbf{v} = (v_x(x, y, t), v_y(x, y, t))$ represents the image velocity; for simplicity of notation the spatial/temporal parameters of the gradient and velocity are hereafter assumed. This constraint defines a plane in the gradient space (I_x, I_y and I_t). Given velocity $\mathbf{v}$ and assuming perfect precision in the extraction of the gradient, the simplest likelihood function can be defined as an equiprobable plane in the gradient space where all other probabilities are zero. Inevitably in the real-world, noise enters our estimates of the data. A more realistic model (though still an approximation) is that our data, specifically the temporal derivative I_t, is contaminated by additive zero-mean Gaussian noise $N(0, \sigma_n^2)$ while the spatial derivatives are assumed to be exact,

$$\nabla I \cdot (v_x, v_y, 1)^\top = N(0, \sigma_n^2) \tag{4}$$

Correspondingly, the likelihood function of observing data ∇I given velocity $\mathbf{v}$ is written as follows,

$$\begin{aligned} P(\nabla I | v_x, v_y) &\propto e^{\frac{-(\nabla I \cdot (v_x, v_y, 1)^\top)^2}{2\sigma_n^2}} && (5) \\ &\propto e^{\frac{-(I_x v_x + I_y v_y + I_t)^2}{2\sigma_n^2}} && (6) \end{aligned}$$

This can be thought of as Gaussian deviations from the ideal plane (or mean plane) in the gradient space isolated to the I_t dimension.

To combine multiple measurements taken over a small spatial image region it is assumed that each measurement, ∇I^i where $i = 1, \ldots, N$, is independent and the velocity is constant[1] [4], formally,

$$\begin{aligned} P(\nabla I^1, \ldots, \nabla I^N | v_x, v_y) &\propto \prod_{i=1}^{N} e^{\frac{-(I_x^i v_x + I_y^i v_y + I_t^i)^2}{2\sigma_n^2}} && (7) \\ &\propto e^{\sum_{i=1}^{N} \frac{-(I_x^i v_x + I_y^i v_y + I_t^i)^2}{2\sigma_n^2}} && (8) \end{aligned}$$

To complete the definition of the posterior, a prior must be selected. When the prior is assumed uniform (i.e., all velocities are equally likely) the maximum of the posterior corresponds to the least-squares solution [2]. This is what is commonly referred to in the literature as the the maximum likelihood estimate. In [4] the authors propose a zero-mean Gaussian prior that favours slower velocities over larger ones,

$$P(v_x, v_y) \propto e^{\frac{-(v_x^2 + v_y^2)}{2\sigma_v^2}} \tag{9}$$

Combining the likelihood (8) and prior (9) yields the following posterior,

$$\begin{aligned} P(v_x, v_y | \nabla I^1, \ldots, \nabla I^N) &\propto e^{\sum_{i=1}^{N} \frac{-(I_x^i v_x + I_y^i v_y + I_t^i)^2}{2\sigma_n^2}} e^{\frac{-(v_x^2 + u_y^2)}{2\sigma_v^2}} && (10) \\ &\propto e^{\sum_{i=1}^{N} \frac{-(I_x^i v_x + I_y^i v_y + I_t^i)^2}{2\sigma_n^2} - \frac{(v_x^2 + v_y^2)}{2\sigma_v^2}} && (11) \end{aligned}$$

Since the posterior (11) is a Gaussian, the MAP solution in this case is equivalent to the mean of the posterior.

In practice, the assumption of attributing additive noise solely to the temporal derivative may be unrealistic [3]. In [3] the authors counter with the assumption that additive zero-mean Gaussian noise pervades all three derivative measurements. This assumption yields the following likelihood function [3],

$$P(\nabla I^1, \ldots, \nabla I^N | v_x, v_y) \propto e^{\sum_{i=1}^{N} \frac{-(I_x^i v_x + I_y^i v_y + I_t^i)^2}{2\sigma_n^2 (1 + v_x^2 + v_y^2)}} \tag{12}$$

[1] This begs the question, how valid is the local independence assumption?

The maximum likelihood estimate (i.e., prior set to uniform) corresponds to the total-least squares velocity estimate [5].

References

[1] B.K.P. Horn. *Robot Vision.* MIT Press, Cambridge, MA, 1986.

[2] B.D. Lucas and T. Kanade. An interative registration techniqu with an application to stereo vision. In *International Joint Conference on Artificial Intelligences*, pages 674–679, 1981.

[3] O. Nestares, D.J. Fleet, and D.J. Heeger. Likelihood functions and confidence bounds for total-least-squares problems. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages I: 523–530, 2000.

[4] E.P. Simoncelli. Local analysis of visual motion. In L.M. Chalupa and J.S. Werner, editors, *The Visual Neurosciences*, chapter 109. MIT Press, 2003.

[5] J. Weber and J. Malik. Robust computation of optical-flow in a multiscale differential framework. *International Journal of Computer Vision*, 14(1):67–81, January 1995.

[6] Weiss Y. and Fleet D.J. Velocity likelihoods in biological and machine vision. In R.P.N. Rao and M. S. Olshausen, B. A.and Lewicki, editors, *Probabilistic Models of the Brain: Perception and Neural Function*, pages 77–96. MIT Press, 2002.