The Computation of Optical and Affine Flow

Konstantinos G. Derpanis
kosta@cs.yorku.ca
Version 1.1

January 31, 2005

In this note we review the computation of optical and affine flow. We will assume that the velocity within the considered neighbourhood is constant [7, 8]. An alternative formulation is to consider the motion within a neighbourhood (or entire image) as slowly varying [6].

The starting point is the assumption of brightness constancy [5], which assumes that the brightness structures of local-time varying image regions are unchanging under motion for a short period of time. Formally, this is defined as,

$$I(x, y, t) = I(x - u, y - v, t - 1) \tag{1}$$

where $(x, y)^\top$ represents image position in pixel coordinates, t represents the temporal coordinate, $(u, v)^\top$ represents the motion at image position $(x, y)^\top$ over the time $t + 1$ and $I(x, y, t)$ represents the image brightness function.

Using the least-squares criteria, we seek the motion that minimizes the error ϵ over a region R of the image, formally,

$$\epsilon = \sum_{x,t,\in R} [I(x, y, t) - I(x - u, y - v, t - 1)]^2 \tag{2}$$

Assuming that $I(x, y, t)$ is approximately locally linear (2) is simplified by taking a Taylor series expansion and omitting terms higher than first order [5],

$$\begin{aligned} \epsilon &\approx \sum_{x,y,\in R} [I(x, y, t) - (I(x, y, t) - I_x(x, y, t)u - I_y(x, y, t)v - I_t(x, y, t))]^2 && (3) \\ &= \sum_{x,y,\in R} [I_x(x, y, t)u + I_y(x, y, t)v + I_t(x, y, t)]^2 && (4) \end{aligned}$$

where I_x, I_y and I_t represent the partial derivatives of the intensity function with respect to the spatial parameters x, y and temporal parameter t, respectively.

To obtain the estimate of velocity in the pure translation case, the partial derivatives of the error in (4) are taken with respect to the translational velocity $(u, v)^\top$ and set to zero. This yields the following set of linear equations,

$$\begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = - \begin{pmatrix} \sum I_x I_t \\ \sum I_y I_t \end{pmatrix} \tag{5}$$

A minimum of two points are required for the full solution. The underconstrained problem given a single point is commonly referred to in the literature as the aperture problem [5].

In the case where the motion is modeled locally by an affine transformation, corresponding to a composition of rotation, dilation, shear and translation, $(u, v)^\top$ are defined as follows,

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_4 & a_5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_0 \\ a_3 \end{pmatrix} \tag{6}$$

As in the case of translational motion, to obtain the estimates of the affine parameters (4) is differentiated with respect to the six unknown affine parameters (a_i where $i = 0, \dots, 5$) and set to zero, yielding the following set of linear equations [2],

$$\begin{pmatrix} \sum I_x^2 & \sum xI_x^2 & \sum yI_x^2 & \sum I_xI_y & \sum xI_xI_y & \sum yI_xI_y \\ \sum xI_x^2 & \sum x^2I_x^2 & \sum xyI_x^2 & \sum xI_xI_y & \sum x^2I_xI_y & \sum xyI_xI_y \\ \sum yI_x^2 & \sum xyI_x^2 & \sum y^2I_x^2 & \sum yI_xI_y & \sum xyI_xI_y & \sum y^2I_xI_y \\ \sum I_xI_y & \sum xI_xI_y & \sum yI_xI_y & \sum I_y^2 & \sum xI_y^2 & \sum yI_y^2 \\ \sum xI_xI_y & \sum x^2I_xI_y & \sum xyI_xI_y & \sum xI_y^2 & \sum x^2I_y^2 & \sum xyI_y^2 \\ \sum yI_xI_y & \sum xyI_xI_y & \sum y^2I_xI_y & \sum yI_y^2 & \sum xyI_y^2 & \sum y^2I_y^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = - \begin{pmatrix} \sum I_xI_t \\ \sum xI_xI_t \\ \sum yI_xI_t \\ \sum I_yI_t \\ \sum xI_yI_t \\ \sum yI_yI_t \end{pmatrix} \tag{7}$$

In this case, a minimum of six points are required for a unique solution.

For both the translational and affine motion cases the Taylor series approximation is a source of imprecision in the final estimates. The precision can be improved by an iterative alignment procedure proposed in [8]. This procedure can be likened to Newton-Raphson's root finding method. Given the initial velocity estimate the first image is shifted towards the second image. The motion estimation procedure is repeated between the newly shifted image and the second image. This procedure is iterated until convergence. For further details see [8]. Finally, to further extend the range of the estimator for large displacements, one can implement the estimation within a multiresolution image pyramid structure (e.g., Gaussian pyramid [3], Laplacian pyramid [4]); for details see [1].

References

[1] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *European Conference on Computer Vision*, pages 237–252, 1992.

[2] J.R. Bergen, P.J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):886–896, September 1992.

[3] P.J. Burt. Fast filter transforms for image processing. *Computer Graphics Image Processing*, 16(1):20–51, May 1981.

[4] P.J. Burt and E.H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, April 1983.

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

[6] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, August 1981.

[7] J.O. Limb and J.A. Murphy. Estimating velocity of moving images in television signals. *Computer Graphics Image Processing*, 4(3):311–327, December 1975.

[8] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.