Presented at: IEEE Signal Processing Society, Eighth Workshop on Image and Multidimensional Signal Processing, Cannes France, September 1993.

Course-to-fine Estimation of Visual Motion

Eero P. Simoncelli

Computer and Information Science Department
University of Pennsylvania

The recovery of motion information from visual input is an important task for both natural and artificial vision systems. The standard approach to representing motion information is via the image *velocity field*: that is, the projection of the motion of points in the three-dimensional world onto the image plane. As an approximation to this, computer vision techniques typically compute an estimate of the motion field from the spatial and temporal variations of image brightness. This field of approximate velocities is known as the “optical flow”.

One common source of difficulty in optical flow estimation systems is temporal aliasing. Motion picture and video imagery is typically sampled below the Nyquist rate in time. Filter-based (including differential) algorithms will typically give erroneous results in these situations. For optical flow algorithms based on matching, the errors appear as “false matches”.

A number of authors have developed “coarse-to-fine” processing strategies for handling the temporal aliasing problem in the context of motion estimation [7, 9, 5, 1]. These algorithms are typically efficiently implemented using “image pyramids” [3]. The motivating concept in these approaches is that the aliasing only affects the high spatial frequencies of the input sequence. Thus, one can accurately estimate image velocities of a spatially lowpass-filtered copy of the input imagery. These estimates may then be used as initial guesses for estimating the motion at finer scales. Alternatively, one can “undo” the motion in finer scale images by warping these images according to the velocity estimated from a coarser scale. The velocity field of the warped images may then be computed and added as a correction term to the current estimate. This process may be repeated recursively until one reaches the finest scale of the input imagery. Multi-scale approaches also provide an efficient means of imposing regularization or smoothness constraints in the optical flow problem [4, 8].

One shortcoming of the coarse-to-fine approaches mentioned above is that if the coarse scale estimate is incorrect by a substantial amount, the finer scales will be unable to recover from these errors. To correct this, we must have knowledge of the uncertainty in the coarse-scale estimates. We have developed a filter-based coarse-to-fine algorithm, which includes an explicit model for the errors in the estimator. The details of the algorithm are described in [11]. We give the basic solution here.

We define a *state evolution* equation for the image velocity field, $\vec{v}$, with respect to scale:

$$\vec{v}(l+1) = E(l)\vec{v}(l) + \vec{n}_0(l), \qquad \vec{n}_0(l) \sim N(0, \Lambda_0)$$

where l is an index for scale (larger values of l correspond to finer scales). Note that indices for spatial and temporal position are implicit. $E(l)$ is a linear interpolation operator used to extend a coarse scale flow field to finer resolution. [1] $\vec{n_0}$ is a random variable that represents the uncertainty of the prediction of the fine-scale motion from the coarse-scale motion. We assume that the $\vec{n_0}(l)$ are pointwise independent, zero-mean, and normally distributed.

We also define a *measurement* equation, based on the standard brightness gradient constraint [6]:

$$-f_t(l) = \vec{f_s}(l) \cdot \vec{v}(l) + (n_2 + \vec{f_s}(l) \cdot \vec{n}_1),$$

where $\vec{f_s}(l)$ is the spatial gradient of the image and $f_t(l)$ is its temporal partial derivative. This constraint holds at each point in space and time (again, spatial and temporal indices are implicit). The uncertainty model, described in [10], incorporates both an additive and a multiplicative noise term. We again assume that the random variables $\vec{n}_1$ and n_2 are zero-mean, independent and normally distributed.

Given these two equations, we may derive the optimal estimator for $\vec{v}(l+1)$, the velocity at the fine scale, given an estimate for the velocity at the previous coarse scale, $\mu_{\vec{v}}(l)$, and a set of fine scale derivative measurements. The solution is in the form of a standard Kalman filter, but with the time variable replaced by the *scale*, l:

$$\begin{aligned} \mu_{\vec{v}}(l+1) &= E(l)\mu_{\vec{v}}(l) + K(l+1)\nu(l+1) \\ \Lambda(l+1) &= \Lambda'(l+1) - K(l+1)\vec{f_s}^T(l+1)\Lambda'(l+1) \end{aligned}$$

$$
\begin{aligned}
K(l+1) &= \Lambda'(l+1)\vec{f}_s(l+1) \cdot \left[\vec{f}_s^T(l+1)\left(\Lambda'(l+1)+\Lambda_1\right)\vec{f}_s(l+1)+\Lambda_2\right]^{-1} \\
\nu(l+1) &= -f_t(l+1) - \vec{f}_s^T(l+1)E(l)\mu_{\vec{v}}(l) \\
\Lambda'(l+1) &= E(l)\Lambda(l)E(l)^T + \Lambda_0,
\end{aligned}
$$

where $\nu(l)$ corresponds to an *innovations* process.

One important modification must be made to the standard form. We cannot compute the derivative measurements at scale l without making use of the velocity estimate at scale $l-1$, due to the temporal aliasing problem. Therefore, we must write $\nu(l)$ in terms of derivatives of the warped sequence. We can view the definition of the innovations process ν as a temporal partial derivative of the warped sequence:

$$
\begin{aligned}
\nu(l+1) &= -f_t(l+1) - \vec{f}_s^T(l+1)E(l)\mu_{\vec{v}}(l) \\
&\approx -\frac{\partial}{\partial t} f\left(\vec{x} + tE(l)\mu_{\vec{v}}(l), t\right) \\
&= -\frac{\partial}{\partial t}\mathcal{W}\{f(l+1), E(l)\mu_{\vec{v}}(l)\},
\end{aligned}
$$

where the operator $\mathcal{W}\{\}$ "warps" the first argument (an image) according to the second argument (a vector field). Thus, the innovations process is computed as the temporal derivative of the image at scale $l+1$, after it has been warped with the interpolated flow field estimate from scale l.

We have implemented this algorithm using a "Gaussian image pyramid" multi-scale data structure [3]. In order to make the solution computationally feasible, we ignore the off-diagonal elements in $\Lambda'(l+1)$ (i.e., the correlations between adjacent interpolated flow vectors). We have applied the algorithm to a number of synthetic image sequences, including the realistic "Yosemite" sequence which was generated by texture-mapping an aerial photograph onto a range map. The results are excellent, and compare quite favorably to those reported in a recent experimental comparison of optical flow techniques [2].

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Intl. J. Comp. Vis.*, 2:283–310, 1989.

[2] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. TR #RPL-TR-9107, Queen's University, Kingston, Ontario, Robotics and Perception Lab, July 1992.

[3] P. J. Burt. Fast filter transforms for image processing. *Comp. Graphics and Im. Proc.*, 16:20–51, 1981.

[4] K. C. Chou, A. S. Willsky, A. Benveniste, and M. Basseville. Recursive and iterative estimation algorithms for multi-resolution stochastic processes. TR # LIDS-P-1857, MIT Laboratory for Information and Decision Sciences, March 1989.

[5] W. Enkelmann and H. Nagel. Investigation of multigrid algorithms for estimation of optical flow fields in image sequences. *Comp. Vis. Graphics Image Proc.*, 43:150–177, 1988.

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

[7] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. 7th Int. Joint Conf. Artificial Intelligence*, pages 674–679, Vancouver, 1981.

[8] M. Luettgen, W. Karl, and A. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. TR # LIDS-P-2115, MIT Laboratory for Information and Decision Systems, 1992. Submitted to IEEE Trans. Image Proc.

[9] L. Quam. Hierarchical warp stereo. In *Proceedings of the DARPA Image Understanding Workshop*, September 1984.

[10] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition*, Mauii, Hawaii, June 1991.

[11] E. P. Simoncelli. Distributed Analysis and Representation of Visual Motion. PhD Thesis, MIT Department of Electrical Engineering and Computer Science. January 1993.