Recursive Non-Linear Estimation of Discontinuous Flow Fields

Michael J. Black*

Xerox Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304, USA

Abstract. This paper defines a *temporal continuity* constraint that expresses assumptions about the evolution of 2D image velocity, or optical flow, over a sequence of images. Temporal continuity is exploited to develop an *incremental minimization framework* that extends the minimization of a non-convex objective function over time. Within this framework this paper describes an incremental continuation method for recursive non-linear estimation that robustly and adaptively recovers optical flow with motion discontinuities over an image sequence.

1 Introduction

Many approaches for estimating optical flow have focused on the analysis of motion between two frames in an image sequence while others have attempted to deal with spatiotemporal information by processing long sequences in batch mode. More recently, there has been an interest in *incremental* approaches which are more suited to the dynamic nature of motion estimation [4, 9, 11]. This paper addresses the problem of incrementally estimating optical flow when the formulation of the problem accounts for motion discontinuities. In this situation we minimize a non-convex objective function that is changing over time. To do so, we propose a general incremental minimization framework which is illustrated by extending a deterministic continuation method over time.

Our goal is to incrementally integrate motion information from new images with previous optical flow estimates to obtain more accurate information about the motion in the scene over time. There are some general properties that an incremental algorithm should have: *(i) Anytime Access:* motion estimates are always available; *(ii) Temporal Refinement*[1]: flow estimates are refined over time as more data is acquired; *(iii) Computation Reduction* [9]: by exploiting the information available over time, the amount of computation between any pair of frames is reduced; *(iv) Adaptation:* as the motion of the observer and scene changes over time, the algorithm must adapt to changes in the motion and the changing image.

* Portions of this work were performed at the NASA Ames Research Center, Yale University, and the University of Toronto with support from NASA (NGT–50749), ONR(N00014–91–J–1577), and NSERC.

[1] This idea has also been referred to as "quality improvement" [9].

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

In the following section we explore the idea of temporal continuity. We then show how a temporal continuity constraint is added to a robust formulation of the optical flow problem. Section 4 describes the incremental minimization framework and an incremental continuation method, called *IGNC*.[2] The algorithm recovers accurate optical flow estimates, preserves motion discontinuities, requires only a fixed amount of computation between frames, and adapts to scene changes. Experimental results are presented for natural and synthetic image sequences.

2 Temporal Continuity

The predictable motion of surfaces in the world gives rise to a predictable change in image velocity over time which we call *temporal continuity.* This property is exploited by spatiotemporal-filtering approaches [1] and epipolar-plane image analysis [7]. In contrast to these locally batch approaches we are interested in incrementally processing a sequence of images.

Murray and Buxton [10] extend the standard spatial neighborhood system of Markov random field approaches to include neighbors in both space and time and they define a crude temporal continuity constraint, E_T, that assumes that the flow at an image location remains constant over time. We take a different approach in which we treat temporal continuity as a constraint on image velocity, formulate it to account for violations, and incorporate it into the estimation problem. For example, consider the simple assumption that the acceleration of a surface patch is constant over time. Let $\mathbf{u}(x,y,t) = (u(x,y,t), v(x,y,t))$ be the optical flow at a point (x,y) at a particular instant in time t. We can predict what the flow will be at the next instant, $t+\delta t$, as follows:

$$\mathbf{u}^-(x,y,t) = \mathbf{u}(x-u\delta t, y-v\delta t, t-\delta t) + \frac{\partial}{\partial t}\mathbf{u}(x-u\delta t, y-v\delta t, t-\delta t)\delta t, \quad (1)$$

where the acceleration is approximated by

$$\frac{\partial}{\partial t}\mathbf{u}(x,y,t) \approx (\mathbf{u}(x,y,t) - \mathbf{u}^-(x,y,t)), \quad (2)$$

and where $\mathbf{u}^-$ is the "predicted" flow field. This equation corresponds to *warping* the flow field *by our current estimate of the flow.*

3 Estimating Piecewise-Smooth Flow

We formulate the problem of recovering the optical flow, $\mathbf{u}_s = (u_s, v_s)$, at every pixel, s, in the image, as the minimization of an objective function, E, composed of a *data conservation* constraint, E_D, a *spatial coherence* constraint, E_S, and a *temporal continuity* constraint, E_T:

$$E(\mathbf{u}_s, \mathbf{u}_s^-) = \lambda_D E_D(\mathbf{u}_s) + \lambda_S E_S(\mathbf{u}_s) + \lambda_T E_T(\mathbf{u}_s, \mathbf{u}_s^-), \quad (3)$$

[2] IGNC stands for Incremental Graduated Non-Convexity.

where the λ_i control the relative importance of the terms.

To illustrate we adopt a robust gradient-based formulation of the optical flow problem [5] where the data conservation, spatial coherence, and temporal continuity constraints are defined as

$$E_D(\mathbf{u}) = \rho_D(I_x u + I_y v + I_t, \sigma_D), \tag{4}$$

$$E_S(\mathbf{u}_s) = \sum_{n \in \mathcal{G}_s} \rho_S(u_s - u_n, \sigma_S) + \sum_{n \in \mathcal{G}_s} \rho_S(v_s - v_n, \sigma_S), \tag{5}$$

$$E_T(\mathbf{u}, \mathbf{u}^-) = \rho_T(u - u^-, \sigma_T) + \rho_T(v - v^-, \sigma_T). \tag{6}$$

where the ρ_* are robust estimators, the σ_* are continuation parameters described below, and $\mathcal{G}_s$ is the set of nearest neighbors of s. The data term is the standard optical flow constraint equation where I_x, I_y, and I_t are the partial derivatives of the image sequence with respect to both spatial dimensions and time, and the spatial term, E_S, implies a first-order smoothness assumption. The temporal term, E_T, insures that the estimate, $\mathbf{u}$, is close to the prediction, $\mathbf{u}^-$.

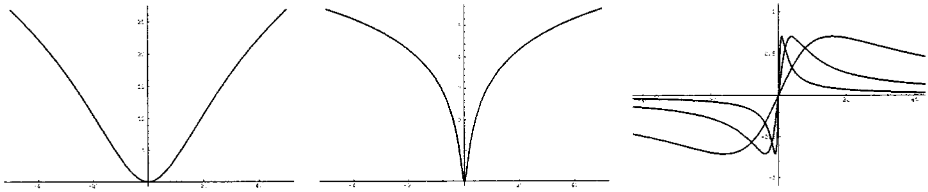
Each of these constraints embodies a set of assumptions about the scene, the motion, and the imaging process. These assumptions are often violated in real scenes and the measurements made by the constraints can be viewed in a statistical context as *outliers.* To reduce the effect of these outlying measurements we adopt the *robust estimation framework* of [3, 5] in which the standard constraints are formulated in terms of robust estimation [8]. We choose the ρ_* to be robust estimators; in this case the Lorentzian estimator:

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \qquad \psi(x, \sigma) = \frac{2x}{2\sigma^2 + x^2}. \tag{7}$$

The ψ–function is the derivative of the estimator and can be used to characterize the "influence" of outliers. In the case of the Lorentzian, the influence of outliers tends to zero. This robust estimation formulation results in a computationally expensive non-convex minimization problem.

3.1 Global Optimization

Local minimization of E is performed using Simultaneous Over-Relaxation (SOR) (see [3] for details of the approach). We focus here on the problem of finding a globally optimal solution when the function is non-convex. The general idea is to take the non-convex objective function and construct a convex approximation. In the case of the Lorentzian estimator, this can be achieved by making the continuation parameters (σ_D, σ_S, σ_T) sufficiently large (see [3] for details). This approximation is then readily minimized using a local technique like SOR. Successively better approximations of the true objective function are constructed by gradually lowering the values of the σ_*. Each successive approximation is minimized starting from the solution of the previous approximation. Figure 1 shows the Lorentzian estimator (Figure 1*a, b*) and its ψ-function (Figure 1*c*) for various values of σ.



a *b* *c*

Fig. 1. Graduated Non-Convexity. Figures *a*, *b* show $\rho(x, \sigma)$ for various values of σ. Figure *c* shows the ψ–functions for three values of σ.

```
Incremental Minimization:
    u, u⁻ ← initially 0 everywhere
    T ← initial value at every site
    n ← fixed, small number of iterations
    for each image
        ;; refinement
        for n iterations
            u ←  minimize (E, u, u⁻, T)     ; peform n iterations beginning at u⁻
            T(x, y) ← f(T(x, y))            ; update the control parameter
        end
        ;; prediction
        u ← u + (u − u⁻)                    ; constant acceleration assumption
        u⁻(x, y) ← u(x − u, y − v)          ; warp flow by current flow
        T(x, y) ← T(x − u, y − v)           ; warp control parameter
        ;; adaptation
        if location (x, y) is occluded or disoccluded then
            T(x, y) ← initial value
            u, u⁻ ← [0, 0]
        end if
    end.
```

Fig. 2. Incremental Minimization algorithm.

4 Recursive Non-Linear Estimation

The traditional recursive estimation techniques for incremental estimation (eg. [11]) are ill-suited to the robust estimation task. Here the problem is to minimize a non-convex objective function that is changing over time and to do so, we develop an new *incremental minimization framework* that performs recursive non-linear estimation. The basic algorithm is summarized in Figure 2.

At any instant in time, the algorithm has a current estimate of the flow field $\mathbf{u}$ and a control parameter $\mathbf{T}$ at each pixel. When a new image is acquired, the constraints are applied to yield a new objective function E and the estimate is refined, beginning with the prediction $\mathbf{u}^-$ as an initial estimate, by performing a fixed number of iterations (usually between 1 and 10) of some continuation method, where an iteration here corresponds to updating all flow vectors in the image.

The assumption of *temporal continuity* is exploited to predict what the flow field and the control parameter will be at the next instant in time. In areas of the image that are undergoing significant change, the values of $\mathbf{T}(x,y)$ must be reset. This can be done by detecting occlusion and disocclusion boundaries in the flow and reinitializing $\mathbf{T}$ in these locations [4]. In our current implementation we reset $\mathbf{T}$ when we detect a violation of any of the three constraints (ie. whenever a measurement is treated as an outlier). Thus, unlike standard continuation methods, for incremental estimation we allow the continuation parameter to vary spatially; this will permit the algorithm to adapt to scene changes. After prediction, a new image is acquired and the process is repeated.

A number of algorithms can be implemented using this general framework. In previous work we have described an *Incremental Stochastic Minimization (ISM)* algorithm [4] in which the minimization is achieved through simulated annealing. Unlike stochastic minimization techniques, continuation methods, such as *Graduated Non-Convexity (GNC)* [6], provide a deterministic minimization strategy for non-convex optimization problems. One benefit of these deterministic approaches is that the coarse approximations provide useful descriptions of the flow field.

5 Experimental Results

SRI Tree Sequence: The first experiment illustrates the dynamic nature of the algorithm by showing the evolution of the horizontal component of the optical flow over time. The SRI tree sequence[3] contains 63 images in which the camera is translating in a direction parallel to the image plane. The maximum displacement between frames is approximately 2 pixels, thus a two-level image pyramid was used. The images were Laplacian filtered and the weights used for this experiment were: ($\lambda_D = 10.0$, $\lambda_S = 1.0$, $\lambda_T = 0.1$). The continuation parameters had the following ranges: $\sigma_D \in [5.0, 0.5]$, $\sigma_S \in [0.5, 0.01]$, $\sigma_T \in [2.5, 0.15]$. These continuation parameters started at the highest value and were reduced by a factor of 0.8 per frame down to the minimum value with only 5 iterations of the method per frame. The results at every tenth frame (starting at frame 32) are shown in Figure 3. At Frame 34 the spatial discontinuities are not yet enforced and the flow is smoothed across the branches of the tree. By Frame 44 the flow becomes more piecewise smooth and this character is maintained throughout the rest of the sequence.

Yosemite Fly-Through: The Yosemite fly-through image sequence[4] consists of 15 synthetic images for which the largest displacement is approximately 4 pixels. For this sequence a three-level pyramid was used and the images were Laplacian filtered. We took the weights $\lambda_D = \lambda_T = 1.0$ and $\lambda_S = 4.0$ to give a

[3] Provided by Bob Bolles and Harlyn Baker.
[4] This sequence was generated by Lynn Quam.

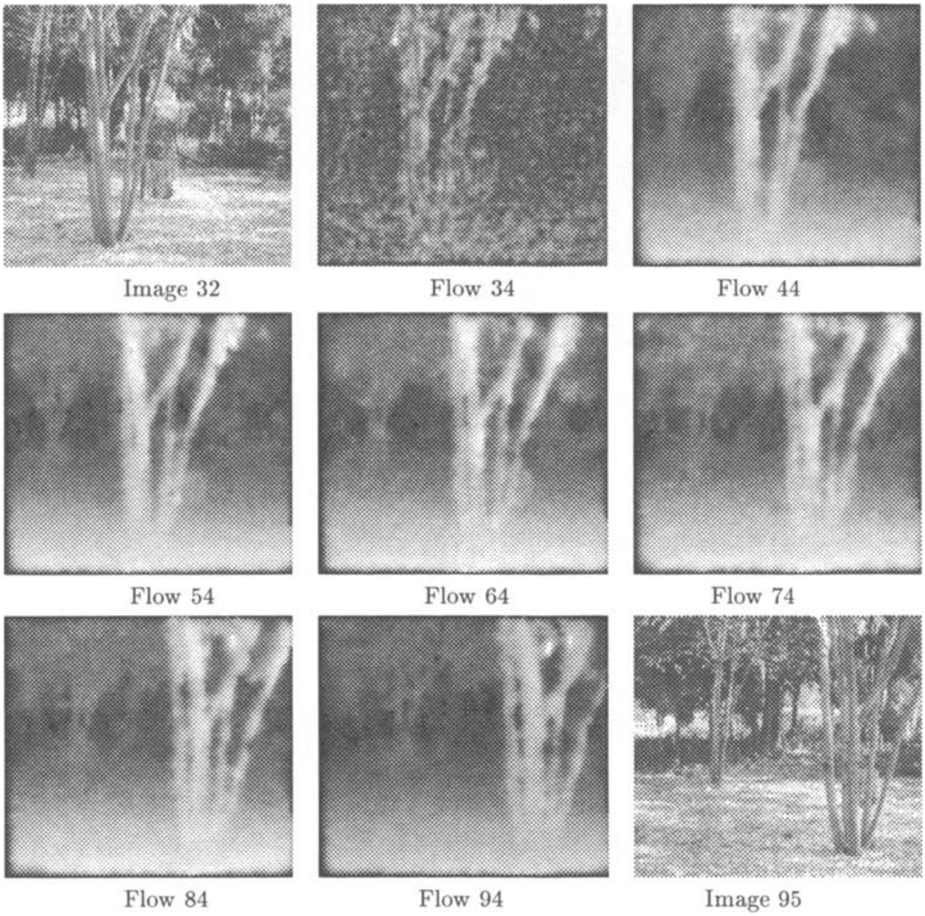


Fig. 3. The SRI Tree Sequence. The horizontal component of the flow at every tenth image is shown. Bright areas are moving faster to the right than dark areas. Discontinuities in flow are gradually introduced over time.

higher weight to the spatial smoothness constraint. The values for the continuation parameters σ_D, σ_S, and σ_T were all taken to be the same with an initial value of 4.0 and a minimum value of 1.0. These parameters were lowered by a factor of 0.8 per frame. Ten iterations (at each level of the pyramid) were used per frame. Figure 4 shows the flow field computed at the end of the sequence.

Since the sequence is synthetic, we can quantify the accuracy of the results using the angular error measure of Barron *et al.* [2]. Table 1 lists the results of a number of algorithms applied to this sequence. The first set of algorithms in the table produce dense flow fields and generally have large average angular errors. The second set of algorithms produce lower average errors but do not provide flow estimates everywhere. The robust formulation results are for a two frame robust

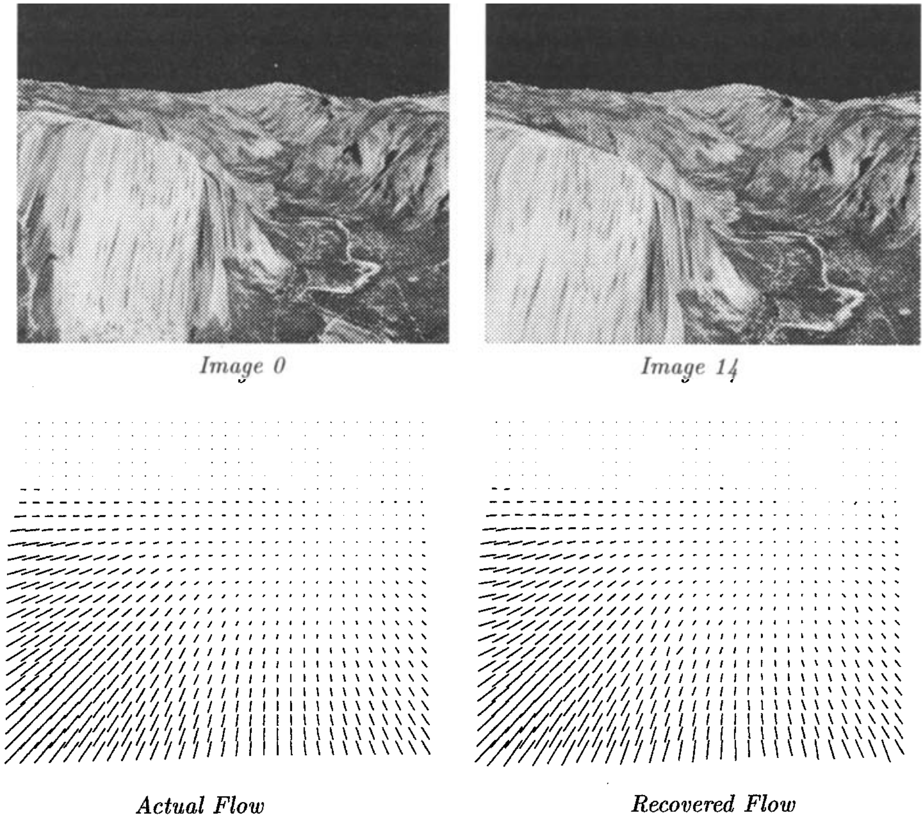


Fig. 4. Yosemite Sequence. The first and last images in the sequence are shown above. The final flow field recovered at the end of the sequence is shown beside the actual flow field.

estimation problem [5] which is identical to the formulation presented here but without temporal continuity.[5] The incremental version (IGNC) achieves better results than the two-frame algorithm and produces errors in the range of the most accurate approaches, but still gives dense estimates. The table on the right shows that the majority of flow vectors have angular errors less than three degrees.

6 Conclusions

We have addressed the problem of incrementally estimating optical flow over a sequence of images in the case where the robust formulation of the optical flow problem results in a computationally expensive non-convex minimization problem. We have developed a framework for solving these problems over time and

[5] Flow errors were not computed in the sky area, because, unlike the Barron *et al.* images which contained clouds, our images were cloudless.

Technique	Average Error	Standard Deviation	Density
Horn and Schunck	32.43°	30.28°	100%
Anandan	15.84°	13.46°	100%
Singh	13.16°	12.07°	100%
Fleet and Jepson	4.17°	11.28°	34.1%
Weber and Malik [12]	3.42°	5.35°	45.2%
Robust Formulation [5]	4.47°	3.90°	100%
IGNC	3.52°	3.25°	100%

% flow vectors with error:	
$< 1°$	13.3%
$< 2°$	38.3%
$< 3°$	56.5%
$< 5°$	79.5%
$< 10°$	96.5%

Table 1. Comparison of various optical flow algorithms (adapted from [2]).

have shown how a deterministic continuation method can be made incremental within this framework. The result is an algorithm which uses a fixed amount of computation per frame, incrementally improves the motion estimates over time, and adapts to scene changes.

References

1. E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2):284–299, February 1985.
2. J.L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. Tech. Rep. No. 299, Univ. of Western Ontario, July 1992.
3. M. J. Black. *Robust Incremental Optical Flow.* PhD thesis, Yale Univ., New Haven, CT, 1992. Research Rep. YALEU/DCS/RR-923.
4. M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *CVPR-91*, pages 296–302, Maui, Hawaii, June 1991.
5. M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *ICCV-93*, pages 231–236, Berlin, Germany, May 1993.
6. A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press, Cambridge, Mass., 1987.
7. R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *IJCV*, 1(1):7–57, 1987.
8. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, NY, 1986.
9. J. Heel. Temporal surface reconstruction. In *CVPR-91*, pages 607–612, Maui, Hawaii, June 1991.
10. D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE PAMI*, 9(2):220–228, March 1987.
11. A. Singh. Incremental estimation of image-flow using a Kalman filter. In *IEEE Workshop on Visual Motion*, pages 36–43, Princeton, NJ, Oct. 1991.
12. J. Weber and J. Malik. Robust computation of optical flow in a mulit-scale differential framework. In *ICCV-93*, pages 12–20, Berlin, Germany, May 1993.