

TECHNICAL RESEARCH REPORT

A Convex Model for the Robust Estimation of Optical Flow
for, Motion-Based Image Segmentation

by R. Haridasan

CSHCN T.R. 97-6
(ISR T.R. 97-17)



The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site <http://www.isr.umd.edu/CSHCN/>

A Convex Model for the Robust Estimation of Optical Flow for Motion-Based Image Segmentation

Radhakrishnan Haridasan
University of Maryland, College Park, MD.

This paper has been submitted to the
IEEE International Conference on Image Processing 97
October 26-29, 1997
Santa Barbara, California, USA

February 18, 1997

1 Introduction

The fundamental problem that needs to be solved in object-based coding is that of decomposing a scene into its constituent objects. The most important cue to segmenting a scene comes from visual motion. Image motion can be used to identify areas corresponding to differently moving objects in space and to distinguish these from the background [1]. The goal of motion-based segmentation is to partition the image into regions that have different characteristics or properties. The problem is compounded owing to the fact that the observable data captured by a camera is image intensity and not image motion. The perceived temporal displacement of intensity patterns in an image sequence owing to relative motion of the objects and the viewer is called *optical flow*. Optical flow provides information pertaining to the spatial arrangement and structure of moving objects [2]. Discontinuities in optical flow arise due to different objects moving at different velocities. Consequently, in the context of segmenting moving images, estimation of optical flow is a fundamental problem. Such segmentation methods are key to content based access of video, which is a requirement of the emerging MPEG-4 video coding standard[3]. Thus, fast, robust and stable estimation of optical flow is the first step towards scene segmentation which is an enabling technology for the next generation of object-based video coding algorithms [4].

2 Robust Estimation of Optical Flow Using a Convex Model

The standard setup for recovering optical flow is to combine the *optical flow constraint*, which follows from the *intensity constancy* assumption with the additional assumption of *spatial smoothness*. The smoothness constraint can be introduced by means of regularization in an analytic framework, or by specifying prior probabilities which favor smoothness in a probabilistic framework. These two criteria taken together provide an *objective function* which measures a goodness or cost of a particular solution. The objective function is in the form of an energy functional which embeds “the correct solution” as its global minimum.

The recovered optical flow is erroneous at pixels where the constraints are used inspite of

the assumptions being violated. The intensity constancy assumption is violated whenever there are occlusions, motion discontinuities or intensity discontinuities (sharp edges or highly textured regions). The spatial smoothness assumption is violated at motion boundaries. These violations manifest as *outliers* in the respective constraints. Thus, we have *data outliers* and *spatial outliers*. Robust estimation methods allow for the rejection of outliers. This makes the recovered flow field less sensitive to violations of assumptions. To do this, given a set of N data samples $d = \{d_i | 1 \leq i \leq N\}$ where $d_i = f + \eta_i$, and the noise process η is assumed to be symmetric and i.i.d., a residual error $\eta_i = d_i - f$, and an error norm $\rho(\eta_i)$ are defined $\forall i$ [5]. The M-estimate f^* is then defined as the minimum of a global error function, i.e.,

$$f^* = \operatorname{argmin}_f E(f) \quad \text{where} \quad E(f) = \sum_{i=1}^N \rho(d_i - f) \quad (1)$$

As opposed to ordinary least squares (LS) regression (equivalent to setting $\rho(\eta) = \eta^2$) which implicitly assumes that the noise process is Gaussian i.i.d. and weights all data points equally, robust M-estimators reduce the weightage given to outliers by reducing the influence from d_i when $|\eta_i| = |d_i - f|$ increases beyond a threshold. In the context of regularization, the same idea was introduced by means of *controlled-continuity stabilizers*. One such stabilizer which became very popular was the truncated quadratic [6]. The truncated quadratic is given by $\rho_\gamma(\eta) = \min\{\eta^2, \sigma\}$. It was shown that the location of discontinuities is determined by means of the line process which could be recovered by means of a simple thresholding operation. The choice of a truncated quadratic for the penalty function makes the energy functional to be minimized non-convex and one has to use deterministic graduated non-convexity (GNC) methods or probabilistic annealing type methods to obtain solutions. These methods are computationally intensive. Also, the stability of such approaches is poor [7]. Small noise in the data can dramatically change the result. In that sense, non-convex robust penalty functions are not “truly robust”. Since estimation of optical flow is the front end of a motion-based image coding system, this is not a desirable feature.

Discontinuity preserving regularization techniques for inverse visual problems have been reviewed by Stevenson et. al. in [7]. The connection between discontinuity adaptivity models and

robust estimation models has been pointed out by S.Z. Li [8]. In the context of regularization, Stevenson et. al. and Li [9] have pointed out the advantages of convex regularization models, namely that of stability and computational efficiency. Convex MRF models are guaranteed to be stable with respect to the input in any situation [10]. The solutions are less sensitive to changes in parameters and data. Convexity helps in the optimization process in terms of ease of implementation and complexity. For a review see Li [9]. Among convex models Huber’s ρ -function is the best in terms of having the least computational complexity. It has been shown by Huber [11] that if we assume the noise process η to be a mixture of Gaussian and unknown p.d.f, the best ρ -function, i.e., the one that minimizes the residual error E for the worst possible unknown p.d.f is given by

$$\rho_\sigma(\eta) = \begin{cases} \frac{\eta^2}{\sigma^2} & |\eta| \leq \sigma \\ \frac{|\eta|}{\sigma} - 1 & |\eta| > \sigma \end{cases} \quad (2)$$

We use the above function to perform robust estimation of optical flow. Since it is strictly convex in $[-\sigma, \sigma]$ and is bounded below, the objective function $E(\mathbf{f})$ is convex and has a unique minimum.

The optical flow equation is not satisfied at pixels where the intensity changes abruptly (occlusions, intensity edge, high texture) and at motion discontinuities. The use of robust estimators provides a mechanism to accommodate for these violations or *outliers*. Thus the data constraint assumes the form of an energy or cost functional,

$$E_D(\mathbf{f})(s) = \rho_{\sigma_D}(ud_x + vd_y + d_t)(s), \quad s \in \mathcal{S} \quad (3)$$

where the data d is specified on a collection of N sites denoted by \mathcal{S} , for time t and $t + 1$. Each member of the site, s , is indexed by (i, j) where $1 \leq i \leq X$, $1 \leq j \leq Y$ and $N = X \times Y$.

The *aperture problem* demands use of an additional constraint for recovering optical flow. For this the important contextual constraint of smoothness is embedded in a robust ρ -function to give

$$E_S(\mathbf{f})(s) = \sum_{n \in \mathcal{N}_s} g_{\sigma_S}(u(s) - u(n)) + g_{\sigma_S}(v(s) - v(n)) \quad (4)$$

where \mathcal{N}_s represents the neighboring sites to site s .

Thus, we obtain the objective function to be minimized as, $E(\mathbf{f}) = E_D(\mathbf{f}) + \lambda E_S(\mathbf{f})$ where λ is a weighting parameter. The M-estimate of the optical flow, f^* at time instant t is given by $\mathbf{f}^* = \operatorname{argmin}_{\mathbf{f}} E(\mathbf{f})$ where $\mathbf{f}^* = [u^* \ v^*]^T$ for particular values of σ_D and σ_S . The convex formulation allows us to compute the M-estimate by using gradient descent methods such as Steepest Descent(SD) and Jacobi Relaxation. More generally a Successive Over-Relaxation (SOR) approach yields highly parallelizable iterative computations of the form

$$u_{i,j}^{(n+1)} = u_{i,j}^{(n)} - \omega \frac{1}{T(u_{i,j})} \frac{\partial E}{\partial u_{i,j}} \quad (5)$$

$$v_{i,j}^{(n+1)} = v_{i,j}^{(n)} - \omega \frac{1}{T(v_{i,j})} \frac{\partial E}{\partial v_{i,j}} \quad (6)$$

where $0 < \omega < 2$ is the over-relaxation parameter and $T(\cdot)$ is an upper bound on the partial derivatives of E.

3 Results

The SOR technique yielded good results on natural test image sequences with just 30 iterations (Fig. 1 & 2). The update equation is local and can be used in conjunction with the “chequerboard” technique to accelerate convergence. This augurs well for the use of such algorithms in motion-based image sequence coding. To estimate large motions multi-grid coarse to fine strategies may be employed.

4 Conclusion

The paper establishes feasibility of using computer vision algorithms for real-time segmentation and compression of motion video sequences. We apply Huber’s regularizer to the problem of optical flow segmentation from a pair of successive frames. The convex formulation of the problem in a robust estimation framework has significant advantages over previous approaches. In particular, no annealing or graduation process is required. Gradient descent procedures or Jacobi relaxation techniques can be used and a high degree of parallelism can be obtained. Unlike previous techniques

our approach guarantees stable, repeatable (or reproducible) segmentations which make real time applications in segmenting video possible.

References

- [1] A. Mitiche and P. Bouthemy, "Computation and Analysis of Image Motion: A Synopsis of Current Problems and Methods", *International Journal of Computer Vision*, Vol. 19, No. 1, pp. 29-55, Jan. 1996.
- [2] S. Ullman, "The interpretation of visual motion", MIT Press, Cambridge, MA, 1979.
- [3] *MPEG-4 Proposal Package Description (PPD) - Revision 2 (Lausanne Revision)*, ISO/IEC JTC1/SC29/WG11, MPEG 95/March 1995.
- [4] K. Zhang, M. Bober and J. Kittler, "Motion based Image Segmentation for Video Coding", IEEE Conference on Image Processing, Washington D.C., pp. 476-479, December 1995.
- [5] S. Z. Li, "Markov Random Field Modeling in Computer Vision", Springer, 1995.
- [6] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.
- [7] R. L. Stevenson, B. E. Schmitz and E. J. Delp, "Discontinuity Preserving Regularization of Inverse Visual Problems", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol24, No.3, March 1994, pp. 455-469.
- [8] S. Z. Li, "Discontinuous MRF Prior and Robust Statistics: A Comparative Study", *Image and Vision Computing*, Vol. 14, No. 4, April 95, pp. 227-233.
- [9] S. Z. Li, "Convex MRF Potential Functions", *International Conference on Image Processing*, ICIP-95, pp. 296-299, December 95, Washington D.C.
- [10] C. Bouman and K. Sauer, "A Generalized Gaussian Image Model for Edge Preserving MAP Estimation", *IEEE Transactions on Image Processing*, Vol. 2, No. 3, July 1993.
- [11] P. Huber, *Robust Statistics*, Wiley 1981.

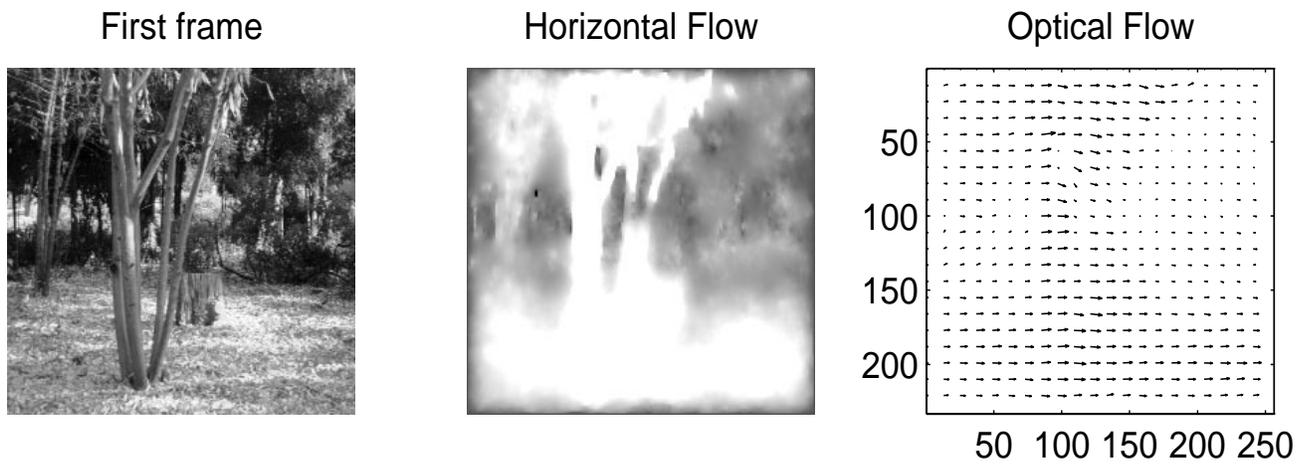


Figure 1: SRI Trees sequence

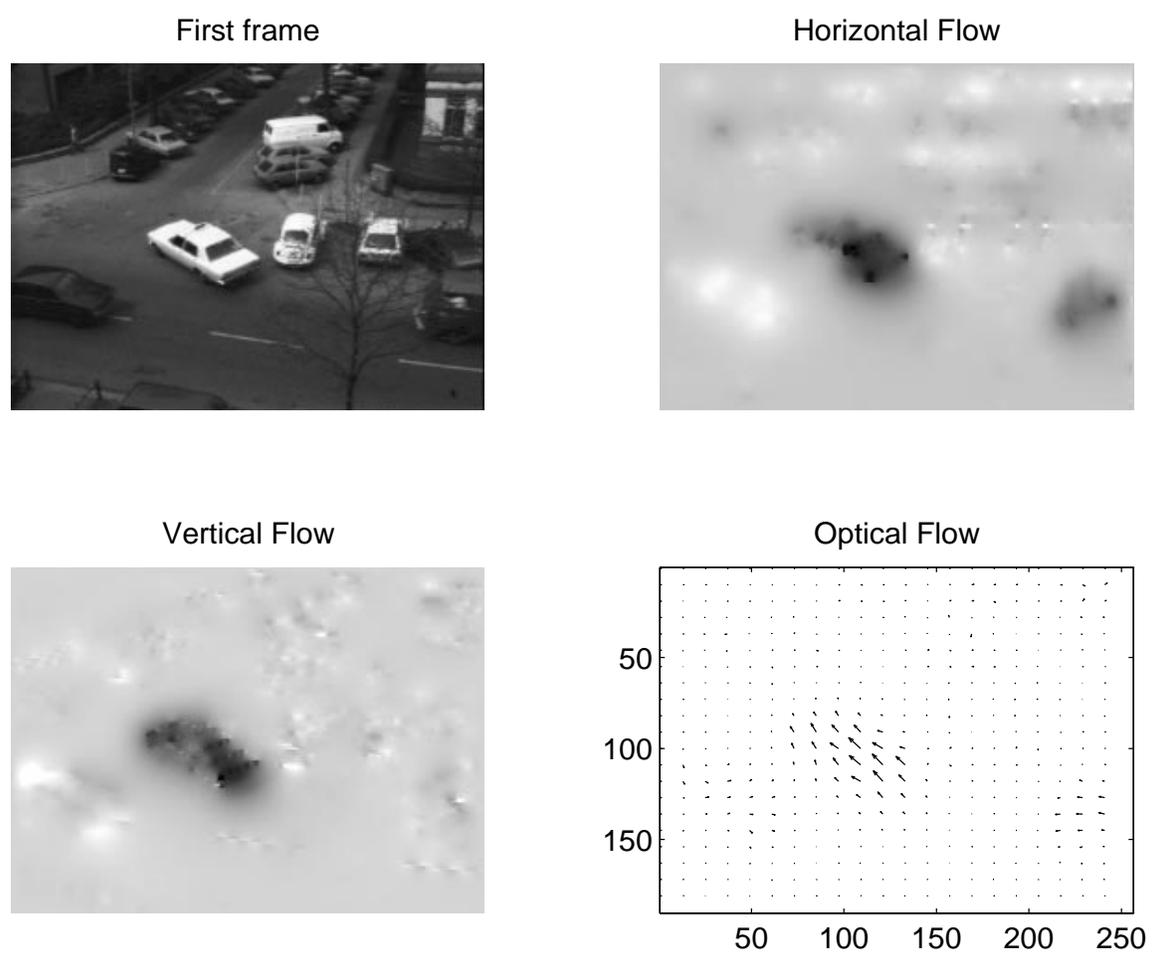


Figure 2: Hamburg Taxi Sequence