Hybrid Recursive Energy-based Method for Robust Optical Flow on Large Motion Fields

Jangheon Kim and Thomas Sikora
Department of Communication Systems
Technical University of Berlin
Einsteinufer 17, Berlin, Germany
{j.kim, sikora}@nue.tu-berlin.de

***Abstract*—We present a new reliable hybrid recursive method for optical flow estimation. The method efficiently combines the advantage of discrete motion estimation and optical flow estimation in a recursive block-to-pixel estimation scheme. Integrated local and global approaches using the robust statistic of anisotropic diffusion remove outliers from the estimated motion field. We separately describe the process with two frameworks i.e. an incremental updating framework and a robust energy minimization framework. With robust error norms of Perona and Marik anisotropic diffusion, the formulation usually leads to non-convex optimization problems. Thus, the solution has many local minima, and convergence to the global minima is not guaranteed. Our hybrid recursive energy-based method employs a hierarchical block-to-pixel estimation concept to prevent this problem. The experimental results prove the excellent performance on several large motion fields.**

Keywords- hybrid recursive method; dense optical flow; discrete motion estimation; block-to-pixel estimation; robust statistic; anisotropic diffusion; very large motion

I. Introduction

To estimate the motion from a sequence of dynamic images is one of the most important tasks in computer vision. Two different approaches, i.e. discrete motion estimation and optical flow estimation, have been separately developed.

The discrete motion estimation [1] establishes the correspondences by measuring similarity using blocks or masks. The advantages are simplicity and reliability for discrete large motion. However, detailed motion of a deformable-body cannot be recovered because block and mask are inherently rigid with translational motion. Blocking artifacts and poor motion prediction along the moving boundaries are serious drawbacks. Hybrid recursive method was first proposed in [1] to solve the problems. The method is recursively refining block motion to arrive at a dense motion field using a gradient technique. It can be hierarchically applied using multi-resolution representation for the accuracy.

On the other hand, optical flow estimation [2-7] aims to obtain a delicate velocity field using the computation of spatial and temporal image derivatives. Most methods define the displacement of pixels based on brightness-conserving assumptions and gradient constancy assumptions or both. One brightness-conserving assumption is that the brightness of objects in subsequent frames does not change by small movements. A gradient consistency assumption determines the displacement with a criterion that is invariant against the change of brightness. Using the partial derivatives, the optical flow efficiently handles the piecewise and detailed variation of displacement. However, the discontinuity from a large motion causes inappropriate matching and ill-posed local minima. In order to overcome these defects, hierarchical algorithms have been proposed based on a multi-resolution representation [2, 3]. A coarse but robust estimation of the motion field is obtained at a low level. Then, it is iteratively refined at higher levels. A motion trajectory can be also considered instead of using low levels [10]. From a motion trajectory, a dense flow field is estimated as a process of interpolation.

Some researches [5-7] have exploited robust parameter estimation to remove outliers. The robust estimation is less sensitive to the outliers using the localization of the motion field. The localization methods are classified into two approaches. The first approach concentrates on removing outliers of a local energy model to determine the best flow within a region. For example, the robust quadratic estimation which uses least squares is solved by a regression [6, 7]. The second approach corresponds to the design of a global energy model with a regularization function which preserves the discontinuity [2-5]. The outliers are removed by smoothing each homogeneous motion area. For example, one model considering a robust M-estimator instead of the quadratic estimator is proposed in [2]. The most recent approach is the combination of the local energy model and global energy model [5].

In this paper, we combine the advantage of the discrete motion estimation and the optical flow estimation in an efficient common scheme, leading to a hybrid recursive energy-based method. The two approaches for the robust estimation are integrativly applied for our method:

- Incremental updating framework : It unites the discrete motion estimation and the optical flow estimation in the concept of displaced frame difference (DFD) [1]. When the discrete motion estimation recursively calculates the displaced block difference (DBD) from a new frame, the optical flow iteratively refines using the DBDs as the initial value. Incremental detailed motion values from the new frame with the previous optical flow are recursively updated using the displaced pixel difference(DPD).
- Robust energy minimization framework : The optical flow should be robustly estimated inside the homogeneous

region of a moving object. It should be recovered without smoothing across the motion discontinuity. Robust error norm of anisotropic diffusion and adaptive spatial-temporal anisotropic regularization are integratively applied in this framework to preserve the motion boundary. The local energy model and the global energy model are used to remove outliers.

II. Hybrid Recursive Energy-based Method For Robust Dense Optical Flow Estimation

A. Incremental updating framework

Incrementally integrated motion has a lot of advantages, i.e. the motion can be accessed at any time. Thus, motion is temporally refined. Only pairs of frames need to be analyzed, therefore the amount of computation is reduced. Moreover, it is adaptive to the changes of motion and luminance over long time [2].

If we assume large motion (dx, dy), integration of incremental motion is defined as translational displacement mapping.

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{1}$$

The best correspondence of intensity $I(x, y)$ is found by matching over the horizontal, vertical and temporal increments (dx, dy, dt). The matching algorithm usually search a candidate set of motion vector (dx, dy) which minimizes some function of displaced frame difference (DFD) [1] as

$$DFD(w, d) = I_t(w) - I_{t+1}(w + d) \tag{2}$$

where d is the total displacement on the image domain $w = (x, y)$. By taking a Taylor series, we linearize (2) as

$$I_t(w) = I_{t+1}(w) - d^T \nabla I_{t+1}(w) - e_{t+1}(w) \tag{3}$$

∇ is the multidimensional gradient operator and $e_{t+1}(w)$ represents the higher order terms of the expansion to be set up at each pixel in a block. Using (2) and (3), the expression can be rearranged involving the DFD with zero displacement as

$$DFD(w, 0) = I_t(w) - I_{t+1}(w) = d^T \nabla I_{t+1}(w) + e_{t+1}(w) \tag{4}$$

In this paper, d is estimated by calculating the sum of displaced block difference (DBD) and displaced pixel difference (DPD) for a pair of frames.

$$d = DBD\left(U_B, V_B\right) + DPD\left(u_P, v_P\right) \tag{5}$$

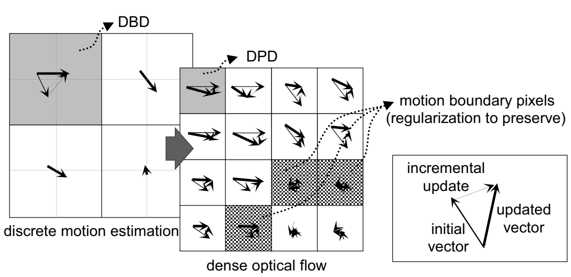


Fig. 1. Incremental updating framework.

DBD and DPD are defined by a large increment (U_B, V_B) and a small increment (u_P, v_P) between any pair of frames. First, the images are pre-filtered, then down-sampled to remove noise and to reduce the system cost. The final DBD is hierarchically refined from the minimization of the correspondence energy E_B with the multidimensional blocks sizes M and N.

$$E_B(U, V) = \sum_{x=0}^{M} \sum_{y=0}^{N} \left| I_t(x, y) - I_{t+1}(x + U_B, y + V_B) \right|^2 \tag{6}$$

When $DBDs$ are calculated in block recursive stage, $DPDs$ are estimated by iteratively refining the $DBDs$ as the initial value of the pixel recursive stage in the gradient solution in (3). The final output vector flow is obtained by recursively updating the increments between the $DBDs$ and $DPDs$. The method achieves highly reliable and per-pixel total displacements to prevent local minima for large motion. Fig. 1 illustrates the method.

B. Robust energy minimization framework

Robust statistics have been applied to remove outliers in many computer vision problems. More recently, anisotropic diffusion was proposed by Perona and Malik [8] and applied for image enhancement. It yields robust results for smoothing homogeneous areas and preserving the discontinuity boundaries. A diffusion function $G(s) = \left[1 + (s/\varepsilon)^2\right]^{-1}$ which is called "edge-stopping function" suppresses diffusion in areas of high gradients. A constant ε controls the level of contrast of edges to affect the smoothing process. The discrete Perona and Malik diffusion is obtained by integrating $G(s) \cdot s$ into a non-convex potential energy as

$$\rho_G = \int [G(s) \cdot s] ds = \int [\psi(s)] ds = \sigma^2 \log\left[1 + 1/2\left(s^2/\sigma^2\right)\right] = \sigma^2 \rho_L \tag{7}$$

where the influence function $\psi(s)$ encloses $G(s) \cdot s$. The derivative $div[G(s) \cdot s]$ modifies the diffusion coefficient at the boundary. Equation (7) shows that ρ_G for $\varepsilon^2 = 2\sigma^2$ can be equivalently treated with Lorentzian error norm ρ_L. Fig. 2 illustrates the relationship of $G(s)$, $\psi(s)$ and ρ_G. In this paper, we integrate the robust error norm into the energy minimization to estimate the optical flow. The edge-preserving robust statistic considers the motion boundaries as those points to be "outliers" between piecewise smooth flow regions.

Considering the brightness-conserving assumption and the gradient constancy assumption for (1), we derive a robust energy function that penalizes deviations in (8)

$$E_D(d) = \int_\Omega \rho_G \left(\left| I(w) - I\left(w + d\right) \right|^2 + \kappa \left| \nabla I(w) - \nabla I\left(w + d\right) \right|^2 \right) dw \tag{8}$$

κ is a proportional weight coefficient between the two terms. Now, the goal of robust estimation of ρ_G is to find the threshold σ to remove outliers. Black [9] achieved the appropriate σ considering Tukey's biweight norm as

$$\rho_G(s, \sigma) = \begin{cases} 1/2\left[1 - (s/\sigma)^2\right]^2 & \text{if } |s| \le \sigma \\ 0 \text{ (i.e. outlier)} & \text{otherwise} \end{cases} \tag{9}$$

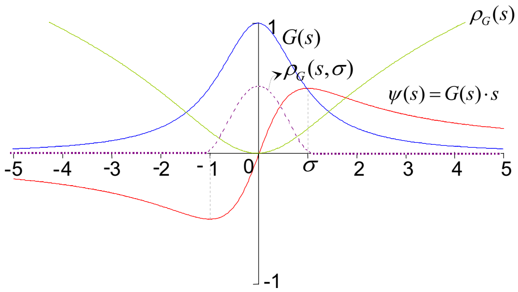


Fig. 2. Robust anisotropic diffusive function.

The robust local flow can be achieved by removing the outliers. Now, we consider a global approach using a regularization function preserving the motion discontinuity. An anisotropic diffusive regularization in (10) is applied to a global energy.

$$E_R(d) = \int_\Omega \psi\left(\nabla I(w), d(w)\right) dw \tag{10}$$

In (10), $\psi\left(\nabla I(w), \nabla d(w)\right) = G\left(\|\nabla_3 d\|\right)\nabla d$ is a modified version of discrete Perona and Marik equation $G(s)\cdot s$ and $\nabla_3 = (\partial_x, \partial_y, \partial_t)$ denotes the spatial-temporal gradient. We sensitivity-adaptively control the diffusion direction by using different step-sizes for the spatial gradient $\nabla I(w)$ and the temporal gradient $\nabla d(w)$ in discrete sense. Finally, the total energy term which combines $E_D(d)$ of (8) with $E_R(d)$ in (10) using the Lagrange multiplier λ is obtained in (11).

$$E(d) = E_D(d) + \lambda E_R(d) \tag{11}$$

III. Numerical Solution

We iteratively solve the minimization problem of (11) by the associated Euler-Lagrange equations as

$$\begin{aligned}&\rho_G\left(I_t^2 + \kappa(I_{xt}^2 + I_{yt}^2)\cdot(I_x I_t + \kappa(I_{xx} I_{xt} + I_{xy} I_{yt})\right)\\ &-\lambda div\left(G\left(\|\nabla_3 d\|\right)\cdot\nabla u\right) = 0\end{aligned} \tag{12}$$

$$\begin{aligned}&\rho_G\left[I_t^2 + \kappa(I_{xt}^2 + I_{yt}^2)\cdot(I_y I_t + \kappa(I_{yy} I_{yt} + I_{xy} I_{xt})\right]\\ &-\lambda div\left(G\left(\|\nabla_3 d\|\right)\cdot\nabla v\right) = 0\end{aligned} \tag{13}$$

where we define the abbreviations of spatial-temporal derivations as

$$\begin{aligned}&I_x = \partial_x I(w+d), I_y = \partial_y I(w+d), I_t = I(w) - I(w+d),\\ &I_{xx} = \partial_{xx} I(w+d), I_{xy} = \partial_{xy} I(w+d), I_{yy} = \partial_{yy} I(w+d),\\ &I_{xt} = \partial_x I(w+d) - \partial_x I(w), I_{yt} = \partial_y I(w+d) - \partial_y I(w)\end{aligned}$$

The asymptotic analysis of a parabolic PDE with natural boundary conditions approximates w^{k+1} using an iteration variable w^k in (14) and (15). High confidence values are calculated much slower than low confidence values.

$$\begin{aligned}&\rho_G\left((I_t^{k+1})^2 + \kappa\left((I_{xt}^{k+1})^2 + (I_{yt}^{k+1})^2\right)\cdot(I_x^k I_t^{k+1} + \kappa(I_{xx}^k I_{xt}^{k+1} + I_{xy}^k I_{yt}^{k+1})\right)\\ &-\lambda div\left(G\left(\|\nabla_3 d^{k+1}\|\right)\cdot\nabla u^{k+1}\right) = 0\end{aligned} \tag{14}$$

$$\begin{aligned}&\rho_G\left((I_t^{k+1})^2 + \kappa\left((I_{xt}^{k+1})^2 + (I_{yt}^{k+1})^2\right)\cdot(I_y^k I_t^{k+1} + \kappa(I_{yy}^k I_{yt}^{k+1} + I_{xy}^k I_{xt}^{k+1})\right)\\ &-\lambda div\left(G\left(\|\nabla_3 d^{k+1}\|\right)\cdot\nabla v^{k+1}\right) = 0\end{aligned} \tag{15}$$

The computational complexity of the nonlinear system is removed retaining the first-order terms of the Taylor approximation in implicit discretization,

$$\begin{aligned}&I_t^{k+1} \approx I_t^k + (I_x^k du^k + I_y^k dv^k),\\ &I_{xt}^{k+1} \approx I_{xt}^k + (I_{xx}^k du^k + I_{xy}^k dv^k),\\ &I_{yt}^{k+1} \approx I_{yt}^k + (I_{xy}^k du^k + I_{yy}^k dv^k)\end{aligned} \tag{16}$$

where $u^{k+1} = u^k + du^k$ and $v^{k+1} = v^k + dv^k$. u^{k+1} and v^{k+1} are easily calculated using the previous iteration steps u^k, v^k and the incremental updates du^k, dv^k. Equation (14) can be rewritten into (17). The case of (15) can be solved in a similar way.

$$\begin{aligned}&\rho_G\left[\left(I_t^k + (I_x^k du^k + I_y^k dv^k)\right)^2 + \kappa\left\{\left(I_{xt}^k + (I_{xx}^k du^k + I_{xy}^k dv^k)\right)^2 +\right.\right.\\ &\left.\left.\left(I_{yt}^k + (I_{xy}^k du^k + I_{yy}^k dv^k)\right)^2\right\}\right]^\tau \cdot I_x^k\left(I_t^k + (I_x^k du^{k,\tau+1} + I_y^k dv^{k,\tau+1})\right)\\ &+\kappa\rho_G\left[\left(I_t^k + (I_x^k du^k + I_y^k dv^k)\right)^2 + \kappa\left\{\left(I_{xt}^k + (I_{xx}^k du^k + I_{xy}^k dv^k)\right)^2 +\right.\right.\\ &\left.\left.\left(I_{yt}^k + (I_{xy}^k du^k + I_{yy}^k dv^k)\right)^2\right\}\right]^\tau \cdot\left[I_{xx}^k\left(I_{xt}^k + (I_{xx}^k du^{k,\tau+1} + I_{xy}^k dv^{k,\tau+1})\right) +\right.\\ &\left. I_{xy}^k\left(I_{yt}^k + (I_{xy}^k du^{k,\tau+1} + I_{yy}^k dv^{k,\tau+1})\right)\right] - \lambda div\left(G\left(\|\nabla_3(u^k + du^k)\|\right)^\tau \cdot\nabla(u^k + du^{k,\tau+1})\right) = 0\end{aligned} \tag{17}$$

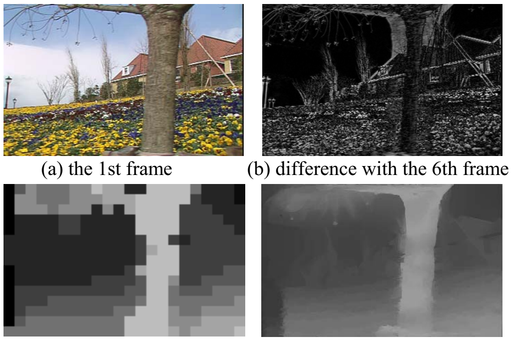
The non-convex robust estimation problem in section II-B converges into a local minimum. A solution is to utilize the previous step as a good initial estimate of the flow and to obtain the incremental value [2, 11]. As stated in section II-A, the motion u^k and v^k are recursively estimated as $(U_B + u_P)^k$ and $(V_B + v_P)^k$. Whenever the updated values U_B and V_B from the block recursion are initialized, the new increments du^k and dv^k are recursively updated in the same way. The recursive block-to-pixel estimation supports the good localization performance. Consequently, an additional iterative minimization strategy with a few steps τ converges quickly into a global minimum.

IV. Experimental Results

First, the performance is evaluated for very large motion of dynamic scene. The *Flower Garden* sequence of size 352×240 pels and 150 frames is shown in Fig. 3. It consists of a tree in the foreground, with a background increasingly vanishing away by the panning of the camera from left to right. The sequence has very large motion in excess of 6 pixels/frame and also has an abundance of fine moving textures. The results in Fig. 3 show the dense motion field between the 1st frame and 6th frame using the magnitude. It proves the robustness and efficiency of our method for very large motion. Comparative results per frame for the whole sequence are shown with some traditional methods in Table 1 using the result in [10]. Our method by far outperforms the above methods.

TABLE I.
Performance Comparison of Flower Garden Sequence

Algorithms	Mean error	Standard Deviation	Density (%)
Horn & Schunck (modified)	16.09	13.64	100
Lucas & Kanade	15.75	13.30	100
Nagel	17.11	14.55	100
Anandan	12.25	10.03	100
Odobez and Bouthemy	13.21	11.12	97.77
Hybrid recursive method	**4.41**	**3.13**	**100**

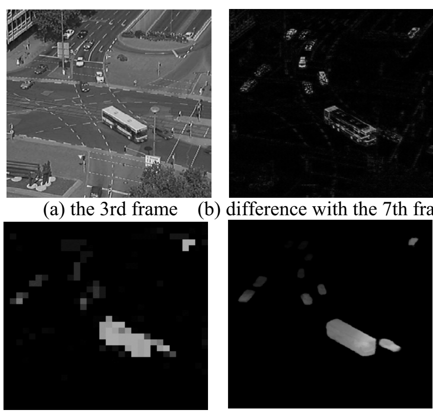


(a) the 1st frame (b) difference with the 6th frame

(c) block initial vector (16×16) (d) final dense motion field

Fig. 3. Result on very large motion of the *Flower Garden* sequence.

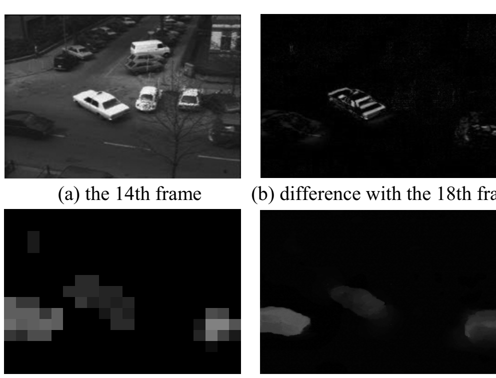
The performance of the proposed method is next evaluated with two natural image sequences from static cameras i.e. *Ettlinger Tor* traffice sequence and *Hamburg taxi* sequence [12]. Fig. 4 shows the 3rd frame and the difference with the 7th frame of *Ettlinger Tor* traffic sequence which has the size of 512×512 pels. Both block initial vector and final dense motion field are adjusted to be more visible with the same scale. The result represents the robustness of the estimated vectors and their good localization into the discontinuity. Fig. 5 represents *Hamburg taxi* sequence which has the size of 256×190 pels. The result also represents the excellent performance of our method. The sharp motion boundaries are well preserved.



(a) the 3rd frame (b) difference with the 7th frame

(c) block initial vector (8×8) (d) final motion field

Fig. 4. Result of *Ettlinger Tor traffice* sequence.



(a) the 14th frame (b) difference with the 18th frame

(c) block initial vector (8×8) (d) final dense motion field

Fig. 5. Result of *Hamburg taxi* sequence.

V. Conclusion

In this paper, we introduced a reliable hybrid recursive energy-based method for optical flow estimation. The method efficiently computes dense optical flow fields using a combination of discrete motion estimation. The method yields the highly accurate and good localized dense motion vectors using the robust statistic of anisotropic diffusion in the energy terms. Outliers of the estimated motion field are well removed. Recursive block-to-pixel estimation scheme efficiently prevents from converging into a local minimum.

Acknowledgment

This work was developed within *3DTV*, *a European Network of Excellence* (*http://www.3dtv-research.net*) funded under the European Commission IST FP6 program.

References

[1] N. Atzpadin, P. Kauff and O. Schreer, "Stereo Analysis by Hybrid Recursive Matching for Real-Time Immersive Video Conferencing," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 3, pp. 321-334, 2004.

[2] M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise smooth flow fields," Computer Vision and Image Understanding, Vol. 63, No. 1 pp. 75-104, 1996.

[3] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," International Journal of Computer Vision, Vol. 2, No. 3, pp. 283–310, 1989.

[4] L. Alvarez, R. Deriche, T. Papadopoulo and J. Sánchez, "Symmetrical Dense Optical Flow Estimation with Oclussions Detection", In Proceeding of European Conference on Computer Vision, pp. 721-736, 2002.

[5] A. Bruhn, J. Weickert and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods," International Journal of Computer Vision, Vol. 61, No. 3, pp. 211-231, 2005.

[6] E. Ong, and M. Spann, "Robust optical flow computation based on least-median-of squares regression," International Journal of Computer Vision, Vol. 31, No. 1, pp. 51-82, 1999.

[7] Charles V. Stewart, "Robust parameter estimation in computer vision," SIAM Reviews, 41(3), pp. 513-537, 1999.

[8] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," IEEE Transactions on Pattern Analysis and Machine Intelligence archive, Vol. 12, pp. 629-639, 1990.

[9] M. J. Black, G. Sapiro, D. Marimont and D. Heeger, "Robust anisotropic diffusion," IEEE Transactions on Image Processing, pp. 421-432, 1998.

[10] D. Gibson and M. Spann, "Robust optical flow estimation based on a sparse motion trajectory set," IEEE Transactions on Image Processing, Vol. 12, pp. 431-445, 2003.

[11] T. Brox, A. Bruhn, N. Papenberg and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," In Proceeding of European Conference on Computer Vision, pp. 25-36, 2004.

[12] Image sequence server of Universität Karlsruhe (http://i21www.ira.uka.de/image_sequences/).