Video motion estimation with temporal coherence

Liang Li, Tianjin University
Wei Feng*, Tianjin University
Jiawan Zhang, Tianjin University
Jianmin Jiang, Tianjin University

Abstract

In this paper, we propose a new method to estimate the optical flow fields of a sequence of images. With the assumption that the velocity is constant from frame to frame along with the flow vectors, which is very common in real scene, we develop a new constraints on the flow field. We formalize the estimation of flow fields as an energy minimization problem and adapt the iterative reweighted least square (IRLS) to solve it. To make the penalties robust to outliers, we mainly consider the Charbonnier penalty. We follow the modern optimization procedure including coarse-to-fine schema, graduated non-convex (GNC) schema to get the final result. The effectiveness of our method is borne out by experiments on both synthetic and real scene.

CR Categories: I.4.8 [Image Processing and Computer vision]: Scene Analysis—Motion;

Keywords: optical flow, temporal coherence

1 Introduction

A key problem in the processing of video is estimating the motion between video frames. Once the optical flow is estimated, a wide variety of tasks such as video compression, super-resolution [Zhao and Sawhney 2002], motion-based segmentation [Kühne et al. 2001], and image registration can be achieved. Optical flow estimation based on two frames has been extensively researched with several classes of methods developed, such as variational framework [Jia and Matsushita 2010], Bayesian framework [Barbu and Yuille 2004] and other methods [Baker et al. 2007]. Many applications estimate the optical flow of a video using every two frames [Chuang et al. 2002] [Zhang et al. 2011].

The basis of differential optical flow is the *motion constraint equation*. It assumes the brightness of the corresponding pixels in two frames remains the same. Modern algorithms for estimating the optical flow often adapt the variational framework [Horn and Schunck 1981] and the coarse-to-fine refinement [Anandan 1989].

In the last two decades, the quality of optical flow estimation has increased drastically. Starting from the original work of Horn and Schunck [Horn and Schunck 1981] as well as Lucas and Kanade [Lucas and Kanade 1981] on optical flow estimation, there are many new concepts for dealing with the shortcomings of previous methods. Non-convex functions has replaced the quadratic function which is used in [Horn and Schunck 1981] to handle the discontinuities in the flow field [Black and Anandan 1996]. Graduated non-convex (GNC) schema is very effective for optimizing the non-convex penalty. It combines a quadratic objective with a robust objective in varying proportions, from fully quadratic to fully robust [Sun et al. 2010]. Coarse-to-fine strategies [Anandan 1989] as well as non-linearize models have been used to tackle with the large displacement.

Comparing to the two-frame optical flow estimation, there is relatively small amount of work on multi-frame optical flow estimation. Most of the applications which involve with estimating optical flow for a video just calculate the flow field between every two frames [Chuang et al. 2002]. Since there are less constraints in two frames comparing to the whole frames, the flow fields is much less accurate than that of using whole frames.

To estimate the optical flows for a video, it is natural to use multi-frame to improve the result. In [Brox et al. 2004] [Bruhn et al. 2006], a 3D gradient ∇_3 is introduced to impose the temporal constraints. In [Liu 2009], C. Liu introduces a more sophistical temporal constraint to improve the estimation of video optical flow and uses iterative reweighted least square (IRLS) to solve the problem. All these methods improves the results.

In this paper, we propose a new temporal constraint on image sequences, which can get more precise result than [Liu 2009]. We formalize the optical flow estimation in energy minimization framework and solve the energy function using IRLS. As suggested in [Sun et al. 2010], we implement our algorithm in coarse-to-fine framework. Besides, we choose Charbonnier $\rho(x) = \sqrt{x^2 + \epsilon^2}$ as the penalty function, which achieves much more results than the quadratic HS penalty function $\rho(x) = x^2$.

The remaining of this paper is organized as follows. In Sect. 2, we talk about the current available algorithms for both two-frame and multi-frame flow estimation. Sect. 3 describes our assumption and the energy functional that used to estimate the video optical flow. In Sect. 4, the optimization details is explained. The results and comparisons with other method are shown in Sect. 5. Finally, we discuss our future work and conclude this paper in Sect. 6.

2 Previous Work

The variational method has proven to be a very effective approach in solving the optical flow estimation problem. Since the starting work of Horn and Schunck [Horn and Schunck 1981], there are many modifications on it. The quadratic penalty is replaced by a seriary of robust penalty, such as L_1 norm [Bruhn et al. 2005] and Lorentzian penalty [Black and Anandan 1996]. The efforts have also been put into improving the optical flow constraints.

3 Optical flow model

Assume $I(x, y, t)$ denotes a rectangular image sequence at time t where $I : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$. Let the image lattice at time t be $\mathbf{p} = (x, y, t)$ and the underlying flow field be $\mathbf{w} = (u(\mathbf{p}), v(\mathbf{p}), 1)$, where $u(\mathbf{p})$ and $v(\mathbf{p})$ is the horizontal and vertical components of the flow field respectively. In the following subsections, we will build a framework for flow estimation.

*Corresponding author, e-mail:wfeng@tju.edu.cn

VRCAI 2011, Hong Kong, China, December 11 – 12, 2011.
© 2011 ACM 978-1-4503-1060-4/11/0012 $10.00

3.1 Robust data function

The basis of the optical flow estimation is by assuming the grey value of a pixel is not changed by the displacement, i.e.,

$$I(\mathbf{p}) = I(\mathbf{p}+\mathbf{w}). \tag{1}$$

If we perform 1^{st} order Taylor series expansion on $I(\mathbf{p}+\mathbf{w})$ at $I(\mathbf{p})$ and omit the higher order terms, we can obtain the famous *motion constraint equation* [Horn and Schunck 1981],

$$I_x u + I_y v + I_t = 0, \tag{2}$$

where I_x and I_y is the horizontal and vertical derivatives respectively, I_t is the temporal derivative. Since Eqn.(2) is only valid under the assumption that the image changes linearly along the displacement, which is in general not the case, especially for large u and v, our model will use the original grey value constancy assumption Eqn. (1).

The grey value constancy assumption is not valid when the brightness of the video is slightly changed, which is very common in the natural scenes. Hence the gradient constancy assumption is proposed, [Brox et al. 2004] [Bruhn and Weickert 2005]

$$\nabla I(\mathbf{p}) = \nabla I(\mathbf{p}+\mathbf{w}), \tag{3}$$

where $\nabla = (\partial_x, \partial_y)$ denotes the spatial gradient.

Combining the grey value constancy and the gradient constancy assumptions on optical flow $\mathbf{w}$, it is straightforward to derive the so called data term of the energy function,

$$E_D(\mathbf{w}) = \int_\Omega \rho_D(|I(\mathbf{p}+\mathbf{w})-I(\mathbf{p})|^2+\lambda|\nabla I(\mathbf{p}+\mathbf{w})-\nabla I(\mathbf{p})|^2). \tag{4}$$

Here λ is a weight between the two assumptions and $\rho_D(\cdot)$ can be any robust function.

3.2 Edge-preserving smoothness function

As suggested by [Sun et al. 2008], [Wedel et al. 2009a] and [Sun et al. 2010], we define our smoothness term as

$$E_S(\mathbf{w}) = \int_\Omega \omega(\mathbf{p})\rho_S(||\nabla \mathbf{w}||_2^2)d\mathbf{p}, \tag{5}$$

where $||\cdot||_2$ is the ℓ_2 norms and ρ_S is any robust function. $\omega(\mathbf{p})$ is the simple structure adaptive map that maintains motion discontinuities: [Werlberger et al. 2009]

$$\omega(\mathbf{p}) = \exp(-||\nabla I(\mathbf{p})||^\kappa). \tag{6}$$

3.3 Temporal constraints

We assume the velocity is constant from frame to frame along with the flow vectors. Formally, we assume,

$$\mathbf{w}(\mathbf{p}) = \mathbf{w}(\mathbf{p}+\mathbf{w}). \tag{7}$$

Here, the $\mathbf{w}(\mathbf{p}+\mathbf{w})$ is the warped flow field at time $t+1$ to time t according to the flow $\mathbf{w}$. For each two successive frames, we prefer the following function to be small,

$$E_T(\mathbf{w}) = \int_\Omega \rho_T(||\mathbf{w}(\mathbf{p})-\mathbf{w}(\mathbf{p}+\mathbf{w})||^2)d\mathbf{p}, \tag{8}$$

where ρ_T is a robust function.

3.4 Energy functional

With the description above, it is straightforward to derive an energy functional that penalizes deviations from these model assumption,

$$E(\mathbf{w}) = \int_t E_D(\mathbf{w}) + \alpha E_S(\mathbf{w}) + \beta E_T(\mathbf{w}) dt, \tag{9}$$

where α and β is the weight. We sum the whole errors over temporal dimension to get a global optima flow field for the video. Note that although ρ_D, ρ_S and ρ_T can be any robust function, here we restrict it to Charbonnier penalty $\rho(x) = \sqrt{x^2+\epsilon^2}$ and set $\epsilon = 0.001$ in our experiments.

4 Minimization

It is difficulty to estimate the flow fields simultaneously using Eqn.(9). We adapt the incremental flow framework [Brox et al. 2004] to estimate the flow field. Assume $\mathbf{w}$ is known, we just need to estimate the incremental $d\mathbf{w} = (du, dv)$. So the objective function in Eqn.(9) can be rewritten as,

$$\begin{aligned} &E(d\mathbf{w}) = \\ &\int_t\int_\Omega \rho_D(|I(\mathbf{p}+\mathbf{w}+d\mathbf{w})-I(\mathbf{p})|^2+\lambda|\nabla I(\mathbf{p}+\mathbf{w}+d\mathbf{w})-\nabla I(\mathbf{p})|^2) \\ &\quad + \omega(\mathbf{p})\rho_S(\nabla(\mathbf{w}+d\mathbf{w})) \\ &\quad + \rho_T(\mathbf{w}(\mathbf{p})+d\mathbf{w}(\mathbf{p})-\mathbf{w}(\mathbf{p}+\mathbf{w}+d\mathbf{w})). \end{aligned} \tag{10}$$

Brox et. al. [Brox et al. 2004] solves the energy function Eqn.(10) using *Euler-Lagrange* variational approach. The mathematical derivation is rather complicated since a function in the continuous spatial domain needs to be optimized. In [Liu 2009], Liu proposed a discrete version to solve the optimization problem, Iterative Reweighted Least Square (IRLS). Furthermore, he has proved the equivalence between IRLS and the variational optimization. Here, we follow the IRLS schema to minimize Eqn.(10).

Let

$$\begin{aligned} I_z(\mathbf{p}) &= I(\mathbf{p}+\mathbf{w}) - I(\mathbf{p}) \\ I_x(\mathbf{p}) &= \partial_x I(\mathbf{p}+\mathbf{w}) \\ I_y(\mathbf{p}) &= \partial_y I(\mathbf{p}+\mathbf{w}) \\ I_{xx}(\mathbf{p}) &= \partial_{xx} I(\mathbf{p}+\mathbf{w}) \\ I_{yy}(\mathbf{p}) &= \partial_{yy} I(\mathbf{p}+\mathbf{w}) \\ I_{xy}(\mathbf{p}) &= \partial_{xy} I(\mathbf{p}+\mathbf{w}) \\ I_{zx}(\mathbf{p}) &= \partial_x I(\mathbf{p}+\mathbf{w}) - \partial_x I(\mathbf{p}) \end{aligned}, \tag{11}$$

then the $I(\mathbf{p}+\mathbf{w})-I(\mathbf{p})$ can be linearized by 1^{st} Taylor expansion,

$$I(\mathbf{p}+\mathbf{w})-I(\mathbf{p}) \approx I_z(\mathbf{p}) + I_x(\mathbf{p})du(\mathbf{p}) + I_y(\mathbf{p})dv(\mathbf{p}). \tag{12}$$

Since

$$\begin{aligned} ||\nabla I(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - \nabla(\mathbf{p})||^2 = &(\partial_x I(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - \partial_x I(\mathbf{p}))^2 \\ &+ (\partial_y I(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - \partial_y I(\mathbf{p}))^2 \end{aligned} \tag{13}$$

and

$$\begin{aligned} \partial_x I(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - \partial_x I(\mathbf{p}) &= I_{zx}(\mathbf{p}) + duI_{xx}(\mathbf{p}) + dvI_{xy}(\mathbf{p}) \\ \partial_y I(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - \partial_y I(\mathbf{p}) &= I_{zy}(\mathbf{p}) + duI_{xy}(\mathbf{p}) + dvI_{yy}(\mathbf{p}) \end{aligned}, \tag{14}$$

we can now rewrite Eqn.(13) as

$$\begin{aligned} ||\nabla I(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - \nabla(\mathbf{p})||^2 = &(I_{zx}(\mathbf{p}) + duI_{xx}(\mathbf{p}) + dvI_{xy}(\mathbf{p}))^2 \\ &+ (I_{zy}(\mathbf{p}) + duI_{xy}(\mathbf{p}) + dvI_{yy}(\mathbf{p}))^2 \end{aligned} \tag{15}$$

We denote the flow field at time t by $\mathbf{w}^t = (u^t, v^t, 1)$, the incremental of flow field by $d\mathbf{w}^t = (du^t, dv^t)$. We now vectorize u^t, v^t, du^t, dv^t into U^t, V^t, dU^t, dV^t respectively. Let I_*^t be the corresponding derivatives in Eqn.(11), where $*$ can be any subscribes in Eqn.(11). Let $\mathbf{I}_x^t = diag(I_x^t)$, $\mathbf{I}_y^t = diag(I_y^t)$, $\mathbf{I}_{xx}^t = diag(I_{xx}^t)$, $\mathbf{I}_{xy}^t = diag(I_{xy}^t)$, $\mathbf{I}_{yy}^t = diag(I_{yy}^t)$, $\Psi = diag(\omega)$ be the diagonal matrices where the diagonals are the image derivatives or the weight. We use $\mathbf{D}_x$ and $\mathbf{D}_y$ to denote the matrix corresponding x- and y-derivative filters, i.e. $\mathbf{D}_x U = u * [0, -1, 1]$, where the $*$ is the convolution operator. We use the $\delta_\mathbf{p}$ to denote a vector that has only one nonzero value (one) at position $\mathbf{p}$ so that $\delta_\mathbf{p} I = I(\mathbf{p})$. Then the continuous function in Eqn.(10) can now be discretized as,

$$\begin{aligned}
&E(dU, dV) = \\
&\sum_t \sum_\mathbf{p} \rho_D\Big(\Big(\delta_p^T(I_z + \mathbf{I}_x dU^t + \mathbf{I}_y dV)\Big)^2 + \\
&\lambda\left((\delta_p^T(I_{zx} + \mathbf{I}_{xx} dU^t + \mathbf{I}_{xy} dV^t))^2 + (\delta_p^T(I_{zy} + \mathbf{I}_{xy} dU^t + \mathbf{I}_{yy} dV^t))^2\right)\Big) \\
&+ \alpha\rho_S\Big(\Big(\delta_p^T \Psi \mathbf{D}_x(U^t + dU^t)\Big)^2 + \Big(\delta_p^T \Psi \mathbf{D}_y(U^t + dU^t)\Big)^2 + \\
&\qquad \Big(\delta_p^T \Psi \mathbf{D}_x(V^t + dV^t)\Big)^2 + \Big(\delta_p^T \Psi \mathbf{D}_y(V^t + dV^t)\Big)^2\Big) \\
&+ \beta\rho_T\left(\delta_p^T(U^t + dU^t - H^t U^{t+1})\right),
\end{aligned} \tag{16}$$

where H^t is the bilinear warping matrix corresponding to the flow field $(U^t + dU^t, V^t + dV^t)$. The main idea of iterative reweighted least square (IRLS) [Meer et al. 2004] is to find the dU, dV so that $(\frac{\partial E}{\partial dU^t}, \frac{\partial E}{\partial dV^t}) = 0$.

Let

$$\begin{aligned}
f_\mathbf{p}^t &= \Big(\delta_p^T(I_z + \mathbf{I}_x dU^t + \mathbf{I}_y dV)\Big)^2 + \\
&\qquad \lambda\Big(\Big(\delta_p^T(I_{zx} + \mathbf{I}_{xx} dU^t + \mathbf{I}_{xy} dV^t)\Big)^2 \\
&\qquad + \Big(\delta_p^T(I_{zy} + \mathbf{I}_{xy} dU^t + \mathbf{I}_{yy} dV^t)\Big)^2\Big) \\
g_\mathbf{p}^t &= \Big(\delta_p^T \Psi \mathbf{D}_x(U^t + dU^t)\Big)^2 + \Big(\delta_p^T \Psi \mathbf{D}_y(U^t + dU^t)\Big)^2 + \\
&\qquad \Big(\delta_p^T \Psi \mathbf{D}_x(V^t + dV^t)\Big)^2 + \Big(\delta_p^T \Psi \mathbf{D}_y(V^t + dV^t)\Big)^2 \\
h_\mathbf{p}^t &= \Big(\delta_p^T(U^t + dU^t - H^t U^{t+1})\Big)^2 \\
l_\mathbf{p}^t &= \Big(\delta_p^T(U^{t-1} - H^{t-1}U^t - H^{t-1}dU^t)\Big)^2,
\end{aligned} \tag{17}$$

we can derive for

$$\begin{aligned}
&\frac{\partial E}{\partial dU^t} = \\
&\sum_\mathbf{p} \rho_D'(f_\mathbf{p}^t)\frac{\partial f_\mathbf{p}^t}{\partial dU^t} + \alpha\rho_S'(g_\mathbf{p}^t)\frac{\partial g_\mathbf{p}^t}{\partial dU^t} + \beta\left(\rho_T'(h_\mathbf{p}^t)\frac{\partial h_\mathbf{p}^t}{\partial dU^t} + \rho_T'(l_\mathbf{p}^t)\frac{\partial l_\mathbf{p}^t}{\partial dU^t}\right) \\
&= 2\sum_\mathbf{p} \rho_D'(f_\mathbf{p})\Big(\mathbf{I}_x^t \delta_\mathbf{p}\delta_\mathbf{p}^T \mathbf{I}_x^t dU^t + \mathbf{I}_x^t \delta_\mathbf{p}\delta_\mathbf{p}^T(I_z^t + \mathbf{I}_y^t dV^t) \\
&\qquad + \mathbf{I}_{xx}^t \delta_\mathbf{p}\delta_\mathbf{p}^T \mathbf{I}_{xx}^t dU^t + \mathbf{I}_{xx}^t \delta_\mathbf{p}\delta_\mathbf{p}^T(I_{zx} + \mathbf{I}_{xy} dV^t) \\
&\qquad + \mathbf{I}_{xy}^t \delta_\mathbf{p}\delta_\mathbf{p}^T \mathbf{I}_{xy}^t dU^t + \mathbf{I}_{xy}^t \delta_\mathbf{p}\delta_\mathbf{p}^T(I_{zy} + \mathbf{I}_{yy} dV^t)\Big) \\
&\quad + \alpha\rho_S'(g_\mathbf{p})(\mathbf{D}_x^T(\Psi^t)^T \delta_\mathbf{p}\delta_\mathbf{p}^T \Psi^t \mathbf{D}_x^T + \mathbf{D}_y^T(\Psi^t)^T \delta_\mathbf{p}\delta_\mathbf{p}^T \Psi^t \mathbf{D}_y^T)(U^t + dU^t) \\
&\quad + \beta(\rho_T'(h_\mathbf{p}^t)\delta_\mathbf{p}\delta_\mathbf{p}^T dU^t + \rho_T'(l_\mathbf{p}^t)\frac{\partial l_\mathbf{p}^t}{\partial dU^t}).
\end{aligned} \tag{18}$$

Note that $\sum_\mathbf{p} \delta_\mathbf{p}\delta_\mathbf{p}^T$ is the identity matrix and $\mathbf{I}_x, \mathbf{I}_y$ are the diagonal matrix. We also define the vector $\mathbf{P}_D^t = diag(\rho_D'(f_\mathbf{p}))$, $\mathbf{P}_S^t = diag(\rho_S'(g_\mathbf{p}))$ $\mathbf{P}_h^t = diag(\rho_T'(h_\mathbf{p}^t))$, $\mathbf{P}_l^t = diag(\rho_T'(l_\mathbf{p}^t))$ and the generalized Laplacian filter $\mathbf{L}^t$,

$$\mathbf{L}^t = \mathbf{D}_x^T(\Psi^t)^T \mathbf{P}_S^t(\Psi^t)\mathbf{D}_x + \mathbf{D}_y^T(\Psi^t)^T \mathbf{P}_S^t(\Psi^t)\mathbf{D}_y, \tag{19}$$

and $\mathbf{M}^t$

$$\mathbf{M}^t = \mathbf{P}_h^t + (H^t)^T \mathbf{P}_l^t H^t. \tag{20}$$

Then the Eqn.(18) can be rewritten as:

$$\begin{aligned}
\frac{\partial E}{\partial dU^t} &= 2\Big(\mathbf{P}_D^t\big((\mathbf{I}_x^t)^2 + \lambda((\mathbf{I}_{xx}^t)^2 + (\mathbf{I}_{xy}^t)^2) + \alpha\mathbf{L} + \beta\mathbf{M}^t\big)dU^t \\
&\quad + \mathbf{P}_D^t\big(\mathbf{I}_x^t\mathbf{I}_y^t + \lambda(\mathbf{I}_{xx}^t\mathbf{I}_{xy}^t + \mathbf{I}_{xy}^t\mathbf{I}_{yy}^t)\big)dV^t \\
&\quad + \mathbf{P}_D^t\big(\mathbf{I}_x^t I_z^t + \lambda(\mathbf{I}_{xx}^t I_{zx}^t + \mathbf{I}_{xy}^t I_{zy}^t)\big) + \alpha\mathbf{L}^t U^t \\
&\quad + \beta\big(\mathbf{M}^t U^t - \mathbf{P}_h^t H^t U^{t+1} - (H^{t-1})^T \mathbf{P}_l^t U^{t-1}\big)\Big).
\end{aligned} \tag{21}$$

Using the same method, one can show that,

$$\begin{aligned}
\frac{\partial E}{\partial dV^t} &= 2\Big(\mathbf{P}_D^t\big(\mathbf{I}_x^t\mathbf{I}_y^t + \lambda(\mathbf{I}_{xx}^t\mathbf{I}_{xy}^t + \mathbf{I}_{xy}^t\mathbf{I}_{yy}^t)\big)dU^t \\
&\quad + \mathbf{P}_D^t\big((\mathbf{I}_y^t)^2 + \lambda((\mathbf{I}_{xy}^t)^2 + (\mathbf{I}_{yy}^t)^2) + \alpha\mathbf{L}^t + \beta\mathbf{M}^t\big)dV^t \\
&\quad + \mathbf{P}_D^t\big(\mathbf{I}_y^t I_z^t + \lambda(\mathbf{I}_{xy}^t I_{zx}^t + \mathbf{I}_{yy}^t I_{zy}^t)\big) + \alpha\mathbf{L}^t V^t \\
&\quad + \beta\big(\mathbf{M}^t V^t - \mathbf{P}_h^t H^t V^{t+1} - (H^{t-1})^T \mathbf{P}_l^t V^{t-1}\big)\Big).
\end{aligned} \tag{22}$$

We consider the non-linear term $\mathbf{P}_D^t, \mathbf{P}_S^t, \mathbf{P}_h^t$ and $\mathbf{P}_h^t$ as a weight, and using the following fixed-point algorithm to solve $(\dfrac{\partial E}{\partial dU^t}, \dfrac{\partial E}{\partial dV^t}) = 0$ for dU^t and dV^t:

(a) Initialize $dU^t = 0, dV^t = 0$ for every t.

(b) Calculate the weight matrix $\mathbf{P}_D^t, \mathbf{P}_S^t, \mathbf{P}_h^t$ and $\mathbf{P}_h^t$.

(c) Solve the following linear equation system

$$\begin{bmatrix} \dfrac{\partial E}{\partial dU^t} \\ \dfrac{\partial E}{\partial dV^t} \end{bmatrix} = 0; \tag{23}$$

(d) If dU and dV converge, stop; otherwise, goto (b).

In [Liu 2009] Liu explained why this algorithm is called iterative reweighted least square. As mentioned in Sect.(1), we here incorporate the graduated non-convex (GNC) schema for the robust data term and smooth term, which linearly combines a quadratic objective with a robust objective in varying proportions. To tackle with large displacement in optical flow, the energy minimization procedure is embedded into a coarse-to-fine approach to avoid convergence to unfavorable local minima. We employ image pyramids with a down-sampling factor of 2 for this purpose. Actually, as mentioned in [Sun et al. 2010]the down-sampling factor dost little matter to the final result. The resulting numerical schema is summarized in Algorithm 1.

5 Evaluation and experimental results

In this section, we present our results and comparisons with other algorithms. Although there are lots of groundtruth data that available in [Baker et al. 2007], there are not any for video optical flow. All the groundtruth data in [Baker et al. 2007] is for two frames.

Algorithm 1 Numerical schema for our algorithm.

Input: A sequence of image $I^1, I^2, \cdots, I^n$,
Output: The corresponding optical flow, $uv_1, uv_2, \cdots, uv_{n-1}$
for $L = 0$ to max_level **do**
  Calculate restricted pyramid images ${}^{L}I^1, {}^{L}I^2, \cdots, {}^{L}I^n$
end for
Perform ROF Decomposition described in [Wedel et al. 2009b]
Initialize ${}^{L}uv^1 = 0, {}^{L}uv^2 = 0, \cdots, {}^{L}uv^{n-1} = 0$ and $\alpha = 1$
while $\alpha \geq 0$ **do**
  for $L = 0$ to max_levels **do**
    for $W = 0$ to max_warps **do**
      Calculate the weight matrix $\mathbf{P}_D^t, \mathbf{P}_S^t, \mathbf{P}_h^t$ and $\mathbf{P}_h^t$;
      Solve the linear system Eqn.(23) using current weight;
    end for
    Upsample ${}^{L}uv$ to next pyramid level;
  end for
  Update α;
end while

	With TC Avg. EPE	Without TC Avg. EPE
Rubble	0.380	0.411
Car	0.315	0.381
Toy	0.334	0.403
Table	0.297	0.335

Table 1: *Average end-point error (EPE) on the annotated motion, which we regards as groundtruth.*

In order to get the groundtruth flow field of a video, we use the method proposed by Liu to generate groundtruth data. Fig. 1 we get by annotating each layers of the video. The first column is the layer labeling and the second column shows the annotated motion, which we regard as the groundtruth data.

The third column in Fig. 1 shows the flow fields that estimated without the temporal constraints, i.e., the flow fields estimated using only two adjacent frames using the method proposed in [Black and Anandan 1996]. The last column in Fig. 1 is the optical flow estimate using our method.

We have measured the temporal constraints quantitatively. Table 1 shows the average end point error (EPE) between the groundtruth and the corresponding optical flow.

Our constraints is quite useful for crowd segmentation. In [Ali and Shah 2007] Saad first utilize crowd flow segmentation to analyze video. It first estimate the flow fields for each two frame, then segment the flow field according to the similarity between each other. Since the flow field of a crowd is thinner than the structure, it is difficult to observe it without temporal constraints. The example is shown in Fig. 2. Fig. 2(a) and Fig. 2(b) are two frames in a crowd video, Fig. 2(c) is the flow field that estimated using temporal constraints we propose and Fig. 2(d) is the one without using temporal constraints. Note the red-rectangle region. There is a small flow in the original image sequence in that area. The algorithm proposed by this paper can easily detect the thin flow field.

6 Conclusion and future work

In this paper, we assume the velocity is constant from frame to frame along with the flow vectors, which is very common in general scene. Based on this assumption, we propose a temporal constraint on flow fields of videos. We formalize this constraint into energy minimization framework and minimize the objective function using IRLS schema. The experiments shows that the results have been improved a lot.

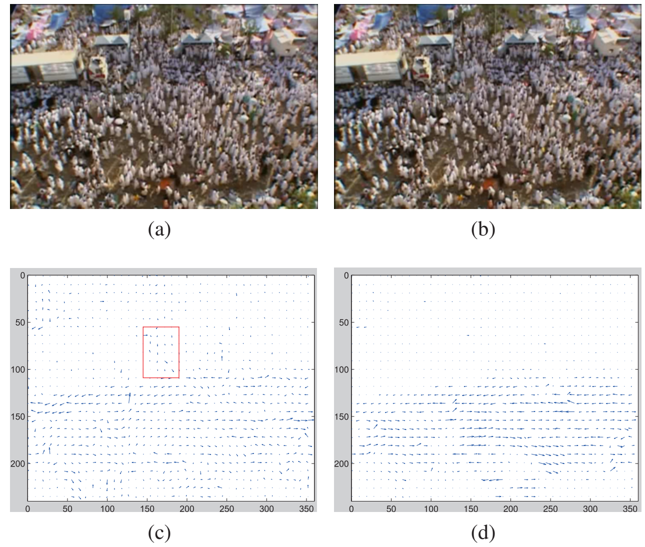

Figure 2: *The application of our constraints on the crowd segmentation.*

However, as mentioned in Sect. 5, there are lots of parameters that need the user to determine. Unlike [Sun et al. 2008], in which a statistical model between grey value constancy and gradient constancy is learned from many, our method is a heuristic formulation.

For future work, we intend to concentrate on the current constraints of our method. Although our algorithm can deal with most scenes, there are a lot of parameters that makes it difficult to get a good result.

References

ALI, S., AND SHAH, M. 2007. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proc. of CVPR*.

ANANDAN, P. 1989. A computational framework and an algorithm for the measurement of visual. *Proc. of IJCV*.

BAKER, S., SCHARSTEIN, D., LEWIS, J., ROTH, S., BLACK, M. J., AND SZELISKI, R., 2007. Opticalflow. http://vision.middlebury.edu/flow/.

BARBU, J., AND YUILLE, A. 2004. Motion estimation by swendsen-wang cuts. In *Proc. of CVPR*.

BLACK, M. J., AND ANANDAN, P. 1996. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding 63* (1), 75–106.

BROX, T., BRUHN, A., PAPENBERG, N., AND J.WEICKERT. 2004. High accuracy optical flow estimation based on a theory for warping. In *Proc. of ECCV*.

BRUHN, A., AND WEICKERT, J. 2005. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *Proc. of ICCV*.

BRUHN, A., WEICKERT, J., KOHLBERGER, T., AND SCHNÖRR, C. 2005. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision 61*, 211–231.

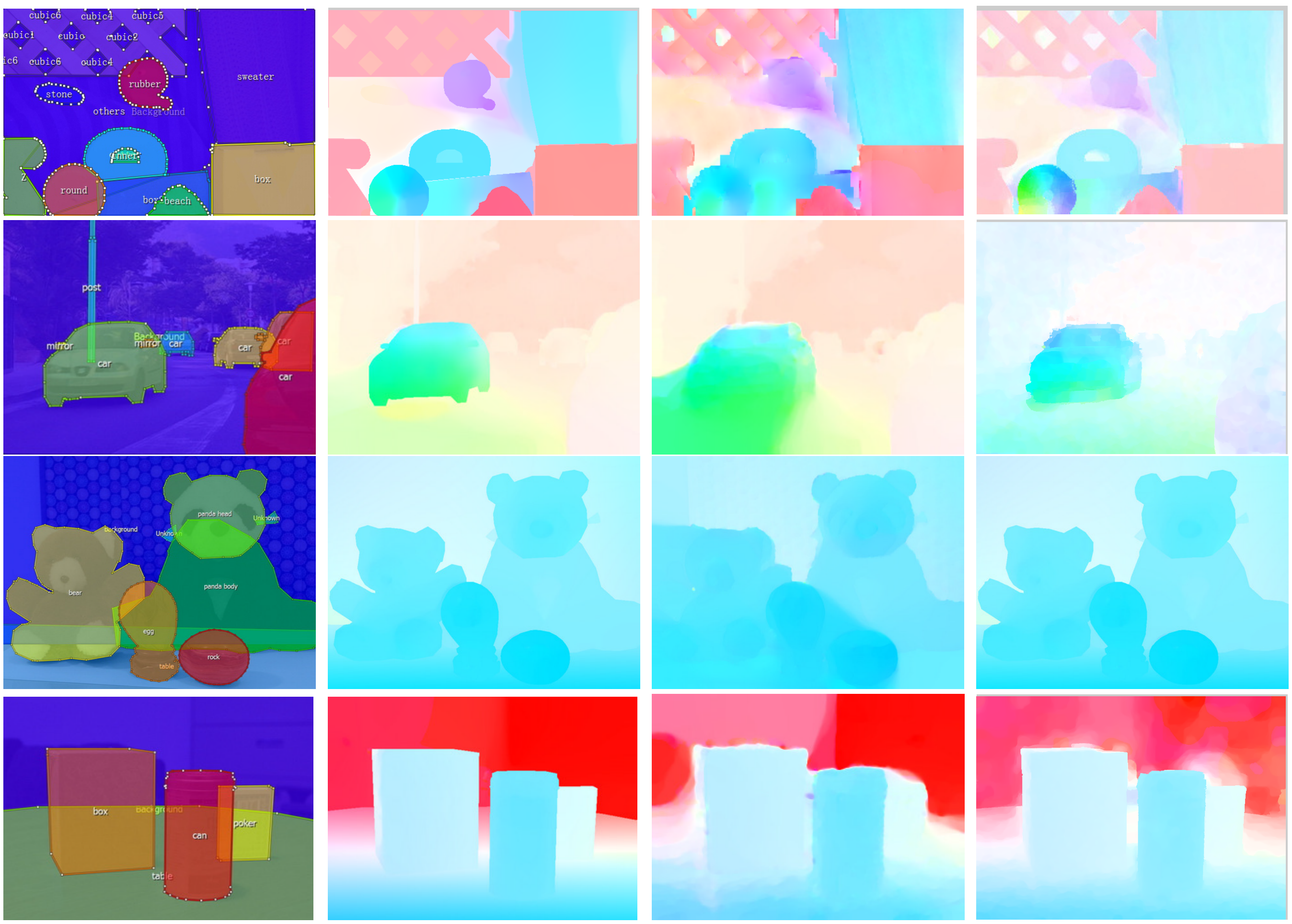

Figure 1: *The groundtruth data we get using Liu's method.*

BRUHN, A., WEICKERT, J., KOHLBERGER, T., AND SCHNÖRR, C. 2006. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision 70*, 257–277.

CHUANG, Y.-Y., AGARWALA, A., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2002. Video matting of complex scenes. *ACM Transactions on Graphics 21*, 243–248.

HORN, B. K. P., AND SCHUNCK, B. G. 1981. Determining optical flow. *Artificial Intelligence 17*, 185–203.

JIA, J., AND MATSUSHITA, Y. 2010. Motion detail preserving optical flow estimation. In *Proc. of CVPR*.

KÜHNE, G., RICHTER, S., AND BEIER, M. 2001. Motion-based segmentation and contour-based classification of video objects. In *Proceedings of the ACM international conference on Multimedia*, 41–50.

LIU, C. 2009. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology.

LUCAS, B. D., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence*, 674–679.

MEER, P., TORDOFF, B., TORR, P. H. S., DAVIDSON, C., SYNTHESIS, I., HUFFEL, S. V., AND WATSON, G. A., 2004. Robust techniques for computer vision.

SUN, D., ROTH, S., LEWIS, J. P., AND BLACK, M. J. 2008. Learning optical flow. In *Proc. of ECCV*.

SUN, D., ROTH, S., AND BLACK, M. J. 2010. Secrets of optical flow estimation and their principles. In *Proc. of CVPR*.

WEDEL, A., CREMERS, D., POCK, T., AND BISCHOF, H. 2009. Structure- and motion-adaptive regularization for high accuracy optic flow. In *Proc. of ICCV*.

WEDEL, A., POCK, T., ZACH, C., BISCHOF, H., AND CREMERS, D. 2009. *An Improved Algorithm for TV-L1 Optical Flow*. 23–45.

WERLBERGER, M., TROBIN, W., POCK, T., WEDEL, A., CREMERS, D., AND BISCHOF, H. 2009. Anisotropic huber-l1 optical flow. In *Proc. of British Machine Vision Conference (BMVC)*.

ZHANG, J., LI, L., ZHANG, Y., YANG, G., CAO, X., AND SUN, J. 2011. Video dehazing with spatial and temporal coherence. *Visual Computer 27*, 749–757.

ZHAO, W., AND SAWHNEY, H. 2002. Is super-resolution with optical flow feasible? In *Proc. of ECCV*.