Accurate Optical Flow Estimation with Sphere Representation

Xinglong Guo, Xinyan Su, Yan Han

National Key Laboratory for Electronic Measurement Technology, North University of China, Taiyuan, 030051, China

Abstract

This paper presents an accurate optical flow estimation approach by reformulating the conventional image brightness constraints and representing the flow in sphere coordinate. A more sophisticated representation for optical flow is derived to substitute current flow modeling. The new model provides better understanding of the behavior and limitations of conventional methods in terms of brightness constraint, and the sphere representation leads to modified algorithms that outperform traditional flow approaches. This seemingly small change in representation provides more direct access to the inner characteristics of a flow field. We compared to the conventional flow methods on standard data sets and show that we achieve more accurate and promising results.

Keywords: Optical flow, Brightness Constraints, Sphere Representation, Energy Minimization

1 Introduction

The estimation of accurate optical flow in image sequence is still an open and challenging problem in computer vision. Optical flow has broad applications, such as in motion estimation and video compression [16], object detection and tracking [5], robot navigation and visual odometry [10], micro air vehicles controlling [6], etc. Therefore, optical flow estimation is an extensive field in computer vision community.

A critical but difficult problem for optical flow estimation, obviously, is constructing correspondences. Correspondences could be in totally different forms, e.g., point correspondences, line correspondences, curve correspondences, even region correspondences. Sometimes, we can easily get some geometrical primitives from images, but sometimes not. According to our best understanding, there are two major methodologies: “dense” approach, and “sparse” or “feature-based” approach. The dense approach tries to build correspondences pixel by pixel, while feature-based approach tries to associate different image features. These two ideas result in totally different taste of motion and structure analysis. The method proposed in this paper belong the “dense” approach.

Horn and Schunck [7] proposed the pioneer work for “dense” flow estimation, which using the flow components in a Cartesian coordinate system. Unfortunately, however, designing objective functions for optical flow is often considered an art and relies primarily
on intuition. This paper is trying to take optical flow estimation to a slightly different direction. We ask whether the traditional brightness constraint is right with a time-step of 1 whereas spatial step can be any other values. In particular, we shall demonstrate in this paper that correct approach is to model the spatial-temporal volume in a uniform manner. This approach immediately allows us to explain the mechanism of well-know flow algorithms and also make their assumptions into more accurate sense. Meanwhile, during the optimization, we propose represent image velocity of the new model in terms of their sphere coordinate instead of their Cartesian ones. As advocated by Marr [8], we show that this not only “makes certain [flow] properties explicit, but it also “greatly affects how easy it is to do certain things with it. We show the experimental advantages of the proposed approach using Middlebury database [1].

2 Related work

The optical flow literature is far too extensive and diverse to allow an exhaustive review here. As mentioned above, modern optical flow research began in the early 1980s when Horn and Schunck [7] combined the brightness constancy assumption with a regularizer that expresses a hypothetical piecewise smoothness behavior, i.e., the assumption that the flows derivatives are zero almost everywhere. The two assumptions (Brightness Constancy & Piecewise Smoothness) were combined to form a single energy functional

\[ E(u, v) = \int \int_{\Omega} (I(x + u, y + v, t + 1) - I(x, y, t))^2 + \alpha(\|\nabla u\|^2 + \|\nabla v\|^2) dxdy \]  

(1)

where \( \Omega \) is the domain of image function as \( I : \Omega \rightarrow R \), \( I(x, y, t) \) is the image intensity at pixel \( (x, y) \in \Omega \) at time \( t \) and \( \alpha \) controls the strength of regularization (e.g., smoothness of the flow). With this functional defined, the goal of the computational algorithm is to find the vector field \((u, v)\) that minimizes it, for example by applying the Euler-Lagrange equations or alternative methods [3, 14, 15]. Over the years, optical flow algorithms have also incorporated various enhancements to improve accuracy such as coarse-to-fine strategies to deal with local minima [4, 9], median filtering to reduce the noise in the flow [11, 14], robust penalty functions to handle outliers [2], and texture decomposition to minimize effects of lightning variations [13, 14].

Somewhat surprisingly, an algorithm based on the basic energy function in Eqn. (1), combined with modern minimization approaches and other enhancements, can estimate the optical flow of traditional motion sequences (e.g., as in the Middlebury benchmark) fairly accurately [11]. Recent advances in flow estimation have been based on increasingly sophisticated computational approaches[4]. In contrast this paper focuses on the choice of data representation and the computational model simultaneously to investigate the inner mechanism of the flow estimation. The reformulated model and sphere representation proposed here can be incorporated into any computational framework that uses the brightness constraint.

3 Brightness Constraint Reformulation

Optical flow has been investigated over last two decades, both our understanding of the problem and its algorithmic implementation has become increasing sophisticated.
basis of the data term used by most algorithms is **Brightness Constancy**, the assumption is that when a pixel flows from one image \( I_t \) to another image \( I_{t+1} \), its intensity or color does not change. This assumption combines a number of assumptions about the reluctance properties of the scene (e.g., that it is Lambertian). The illumination in the scene (e.g., that it is uniform \([12]\) ) and about the image formation process in the camera (e.g., that there is no sketch). As defined in Eqn. (1), Brightness Constancy can be written as:

\[
I(x, y, t) = I(x + u, y + v, t + 1)
\]  

(2)

Linearizing Eqn.(2) by applying a first-order Taylor expansion to the right-hand side yields the approximation:

\[
I(x, y, t) = I(x, y, t) + u \frac{\partial I}{\partial x} + v \frac{\partial I}{\partial y} + \frac{\partial I}{\partial t}
\]  

(3)

which simplifies to the optical flow constraint as:

\[
I_x u + I_y v + I_t = 0
\]  

(4)

where \( I_x, I_y \) and \( I_t \) are derivatives in the \( x, y \) and \( t \) dimensions respectively. Since the flow at a point consists of two values \( u \) and \( v \), a single brightness constraint is insufficient, which makes flow estimation ill posed. Therefore, flow is estimated by imposing additional assumptions of smoothness on the flow field, which will be explained in next section.

There is a significant issue with using the brightness constraint representation in Eqn.(4), in which the derivation of the brightness constraint itself is based on an incorrect model where the temporal dimension is treated differently from the spatial dimensions which introduces undesirable biases. Perhaps this derives from the early methods which assumed that only two images (e.g., \( I_t, I_{t+1} \)) were available, i.e. with a time-step of 1. We shall demonstrate in this paper that the correct approach is to model the spatio-temporal volume in a uniform and continuous manner and introduce the specific discretization of the spatio-temporal image data only as an algorithmic detail. This approach immediately allows us to explain the behavior of well-known flow algorithms and also recast their assumptions into more accurate versions.

By treating the spatio-temporal dimensions in a uniform framework, a key insight that arises is that the correct representation for estimating image flow is not the two-dimensional vector field, but rather its homogeneous counterpart, i.e. normalized volume-flow. We denote the three-dimensional volume-flow at a point as \( F = [U, V, W]^T \), where \( U, V \) and \( W \) are the displacements in the \( x, y \) and \( t \) dimensions respectively. The two-dimensional optical flow is the projection of the volume flow vector \( F \) onto the \( x-y \) image plane and is denoted as \((u, v)\), where \( u = \frac{U}{W} \) and \( v = \frac{V}{W} \). It will be noted that normalized volume-flow \( f \) is given by \( f = \frac{F}{||F||} \) and is also projectively equivalent to the optical flow \((u, v)\), i.e. \( f \propto [u, v, 1]^T \). Using image brightness conservation, we have \( I(x+U, y+V, t+W) = I(x, y, t) \). By a Taylor series expansion around the point \((x, y, t)\) we have \( I(x, y, t) + U \frac{\partial I}{\partial x} + V \frac{\partial I}{\partial y} + W \frac{\partial I}{\partial t} = I(x, y, t) \) leading to

\[
U \frac{\partial I}{\partial x} + V \frac{\partial I}{\partial y} + W \frac{\partial I}{\partial t} = 0
\]  

(5)

which is a brightness constraint equation in three-dimensions and can be simply expressed as \( F^T \nabla I = 0 \), where \( \nabla I = [I_x, I_y, I_t] \) represent image gradient in spatial-temporal domain.
It will be immediately observed here that we have an unknown scale factor for $F$, i.e.
$F^T \nabla I = a F^T \nabla I = 0$, implying that we can only derive $F$ up to a scale factor $a$. Hence we fix the scale by using the normalized volume-flow vector as $f = \frac{F}{\|F\|}$, which is indicated geometrically on the surface of the unit sphere.

### 4 Sphere Representation and Optimization

According to the derivation in last section, the energy functional of the reformulated brightness constraints in terms of 3D flow sense is updated as

$$E(U, V, W) = \iint_{\Omega} (I_x U + I_y V + I_t W)^2 + \alpha(|\nabla U|^2 + |\nabla V|^2 + |\nabla W|^2) dxdy$$  \hspace{1cm} (6)

As we mentioned before, flow $F = [U, V, W]$ is defined on the surface of the unit sphere, so that the smoothness constraints in Cartesian coordinate as described in Eqn.(6) is ill-posed since $U, V, W$ may drift away from the surface of the unit sphere during the numerical optimization (e.g., Gradient Descent algorithm). Therefore, instead of using Cartesian coordinate representation for flow $F$, it’s better to use sphere coordinate to make sure the estimated flow exactly on the surface of the unit sphere, as $F = [U, V, W]^T = [r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta]^T$, and we have $r \equiv 1$ thanks to the unit sphere. Thus, flow $F$ can be further simplified as $F = [\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta]^T$, where only two variables ($\theta$ and $\phi$) should be estimated. Thus, with sphere representation, Eqn.(6) becomes:

$$E(\theta, \phi) = \iint_{\Omega} (I_x \sin \theta \cos \phi + I_y \sin \theta \sin \phi + I_t \cos \theta)^2 + \alpha(|\nabla \theta|^2 + |\nabla \phi|^2) dxdy$$ \hspace{1cm} (7)

As always, to estimate the optical flow one should find the minimizer $(\theta, \phi)$ of the energy functional in Eqn.(7). The Euler-Lagrange conditions for the minimization of Eqn.(7) as:

$$\nabla I^T F \nabla I^T D_\theta F = \alpha \Delta \theta$$ \hspace{1cm} (8)

$$\nabla I^T F \nabla I^T D_\phi F = \alpha \Delta \phi$$ \hspace{1cm} (9)

where $\nabla I = [I_x, I_y, I_t]^T$, $D_\theta F$ and $D_\phi F$ represent the divergence of $F$ w.r.t $\theta$ and $\phi$ respectively. In order to solve the Euler-Lagrange equations numerically, gradient descent is employed, where gradient descent is a generic algorithm for minimizing a smooth energy function. Now the gradient descent algorithm for our energy functional is

$$[\theta_0, \phi_0]^T = [\text{guess}, \text{guess}]^T$$ \hspace{1cm} (10)

$$[\theta_{k+1}, \phi_{k+1}]^T = [\theta_k, \phi_k]^T - \epsilon \nabla E(\theta, \phi)$$ \hspace{1cm} (11)

where $\nabla E(\theta, \phi)$ is defined in Eqn.(8) and Eqn.(9), and $\epsilon > 0$ is a small value to represent the step size. Since the negative gradient direction $-\nabla E(\theta, \phi)$ decreases the energy functional $E(\theta, \phi)$ in the steepest possible way, and the energy functional in Eqn.(7) is convex, therefore, the gradient decent algorithm will reach a global minimum.

### 5 Experimental results

In order to explore the potential of the suggested framework, an algorithm based on the reformulated brightness constraints and new sphere representation as defined in Eqn.(7)
Figure 1. Experimental results on the Middlebury dataset, the 1st row shows the one of original images, the 2nd and 3rd row respectively demonstrate the proposed algorithm and traditional approach, where the performance evaluation is illustrated in Fig.2.

Figure 2. Performance evaluation of Experiment in Fig.1, which shows that the proposed algorithm outperforms the traditional approach.
Figure 3. Image sequences that without ground truth data from Middlebury database, the 1st row shows the one of original images, the 2nd and 3rd row respectively demonstrate the proposed algorithm and traditional approach.
was implemented and compared to the traditional approach as characterized in Eqn.(1). As we mentioned before, we focused on the comparison by using Middlebury database [1], which is the most popular and focuses on various kind of scenes. Since our goal was to explore what are the possible benefits of the proposed method with reformulated brightness constraints and sphere representation only, so for fair comparison we don’t use any other additional optimization process, such as median filtering heuristic and texture decomposition as mentioned in [11].

The results of the initial eight Middlebury images with ground truth are displayed in Fig.1, and the corresponding performance evaluation are illustrated in Fig.2, in which we define three most common used measure of performance[1] for optical flow are the Average Angular Error (AAE), Standard Deviation of Angular Error (SDAE) and Average End Point Error (AEPE). More specifically, they are defined are:

\[
AE = \cos^{-1}\left(\frac{1.0 + u \times u_{GT} + v \times v_{GT}}{\sqrt{1.0 + u^2 + v^2} \sqrt{1.0 + u_{GT}^2 + v_{GT}^2}}\right)
\]

\[
AAE = \frac{1}{|\Omega|} \int_{\Omega} AE \, dxdy
\]

\[
SDAE = \sqrt{\frac{1}{|\Omega| - 1} \int_{\Omega} (AE - AAE)^2 dxdy}
\]

\[
AEE = \frac{1}{|\Omega|} \int_{\Omega} \sqrt{(u - u_{GT})^2 + (v - v_{GT})^2} \, dxdy
\]

where \((u_{GT}, v_{GT})\) is the ground truth flow, and \(|\Omega|\) indicates the total number of the pixels in the image. As can be seen from Fig.2, which suggested the proposed algorithm outperforms the traditional algorithm in all the sequences, we can attribute the better results to the reformulated brightness constraint and sphere representation.

In addition, we have also test all the other image sequence (without ground truth provided) in Middlebury database to show the benefits of the proposed algorithm in various scenes, the results are shown in Fig.3, which demonstrated that the proposed algorithm performs better perceptually (e.g., flow smoothness, since no statistics can be provided due to lack of ground truth data). We further observe that performance analysis based on sphere coordinate representation for the 3D flow field computation become even more reasonable and meaningful due to the better understanding of the image formation mechanism (e.g., noise, error, etc.). Therefore, the proposed method leads to better estimation algorithms.

6 Conclusion and Future Work

In this paper we investigate accurate flow estimation using sphere representation of motion as well as re-formulate the brightness constraint. We show that flow components seems to be more meaningful when sphere representation is employed. We then using gradient descent algorithm to optimize the flow energy functional to achieve the robust numerical solution. Lastly, we show how performance compares well with the traditional optical flows, and exceeds it on complex flows with Middlebury database. Our future work will focus on fully understanding the new model and new representation, e.g., how to handle the large motion, how to improve the performance with temporal regularity by adding more
image frames, and how to increase the speed to make it real-time. It’s our opinion that this research direction holds much promising for optical flow estimation in general.

Acknowledgements

Project supported by a grant from the Major State Basic Research Development Program of China (973 Program) (No. 2011CB311804).

References


