21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

A probabilistic formulation of the optical flow problem

Theodosios Gkamas[1], Giannis Chantas[2], Christophoros Nikou[3]

[1] LSIIT, UMR UDS-CNRS 7005, University of Strasbourg, France
[2] Aristotle University of Thessaloniki, Department of Informatics, Greece
[3] University of Ioannina, Department of Computer Science, Greece

Abstract

The Horn-Schunck (HS) optical flow method is widely employed to initialize many motion estimation algorithms. In this work, a variational Bayesian approach of the HS method is presented where the motion vectors are considered to be spatially varying Student's t-distributed unobserved random variables and the only observation available is the temporal image difference. The proposed model takes into account the residual resulting from the linearization of the brightness constancy constraint by Taylor series approximation, which is also assumed to be a spatially varying Student's t-distributed observation noise. To infer the model variables and parameters we recur to variational inference methodology leading to an expectation-maximization (EM) framework in a principled probabilistic framework where all of the model parameters are estimated automatically from the data.

1 Introduction

The estimation of optical flow is one of the fundamental problems in computer vision as it provides the motion of brightness patterns in an image sequence. A large family of optical flow methods are the global or variational techniques, relying on an energy minimization framework, with their main representative being the Horn-Schunck method [7], which optimizes a cost function using both brightness constancy and global flow smoothness and has also led to many variants of the basic idea. However, the spatial smoothness of the flow field assumed in the above techniques results in many cases to blurred flow boundaries. To overcome this drawback, many researchers proposed various approaches. The related literature being abundant, we may name methods such as robust statistics [10], variational methodologies [5], the integration of spatial priors [13], the segmentation of the image pixels or the motion vectors [15] and learning from ground truth data [14]. Moreover, efforts to combine local and global adaptive techniques were also proposed such as the technique in [3].

A significant issue in the variational methods is the relative importance between the brightness constancy term and the smoothness term which is usually controlled by a parameter determined by the user remaining fixed during the whole process. This is the case not only for the early algorithm of Horn-Schunck [7] but also for the latest versions of this category of methods. Another shortcoming is the linearization of the brightness constancy constraint which results in omitting the higher order terms of the Taylor series expansion [7].

In this paper, we propose a probabilistic formulation of the optical flow problem by following the Bayesian paradigm. The proposed model has intrinsic properties addressing the above mentioned shortcoming. More specifically, we consider the motion vectors in the horizontal and vertical directions to be independent hidden random variables following a Student's *t*- distribution. This distribution may model, according to its degrees of freedom, flows following a dominant model (spatial smoothness) as well as flows presenting outliers (abrupt changes in the flow field or edges). Therefore, to account for flow edge preservation with simultaneous smoothing of flat flow regions, the parameter of the *t*-distribution is also considered to be spatially varying and its value depends on pixel location.

Furthermore, the proposed model takes into account the residual resulting from the linearization of the brightness constancy constraint. The higher order terms of the Taylor series approximation are also represented by a spatially varying Student's *t*-distributed observation noise. It turns out that the update solution for the motion field has a form analogous to the update equations of the Horn-Schunck method [7], with the involved quantities being automatically estimated from the two images due to the principled probabilistic modeling. Numerical results revealed that the method provides better accuracies not only with respect to standard optical flow algorithms [7, 8] which are used to initialize more sophisticated methods, but also to a recently proposed version of their joint combination [3].

2 A prior for the motion vectors

Let $\mathbf{I}(\mathbf{x})$ be the first image frame (target frame) containing the intensity values lexicographically and let also $\mathbf{J}(\mathbf{x})$ be the second image frame (source frame) where $\mathbf{x} = (x, y)$ represents the $2D$ coordinates of a pixel. The brightness constancy constraint at a given location is expressed by:

$$\frac{\partial \mathbf{I}}{\partial x}\mathbf{u}_x + \frac{\partial \mathbf{I}}{\partial y}\mathbf{u}_y + \frac{\partial \mathbf{I}}{\partial t} = 0, \tag{1}$$

where we have removed the independent variable representing the location $\mathbf{x}$ for simplicity. In (1), $\mathbf{u}_x$ and $\mathbf{u}_y$ are the motion vectors in the horizontal and vertical directions respectively, $\partial\mathbf{I}/\partial x$ and $\partial\mathbf{I}/\partial y$ are the spatial gradients of the target image and $\partial\mathbf{I}/\partial t$ is the temporal difference between the two images $(\mathbf{J}(\mathbf{x}) - \mathbf{I}(\mathbf{x}))$.

For convenience, we compactly represent the optical flow values at the i-th location by $\mathbf{u}_k(i)$, for $i = 1, \ldots, N$ where $k \in \{x, y\}$ and N is the number of image pixels. We now assume that $\mathbf{u}_k(i)$ are i.i.d. zero mean Student's t-distributed, with parameters λ_k and ν_k:

$$\mathbf{u}_k(i) \sim \mathcal{S}t\left(0, \lambda_k, \nu_k\right), \forall i = 1, ..., N, \forall k \in \{x, y\}. \tag{2}$$

The Student's-t distribution implies a two-level generative process [4]. More specifically, $\mathbf{a}_k(i)$, $k \in \{x, y\}$ are first drawn from two independent Gamma distributions: $\mathbf{a}_k(i) \sim Gamma\left(\nu_k/2, \nu_k/2\right)$. Then, $\mathbf{u}_k(i)$, $k \in \{x, y\}$ are generated from two zero-mean Normal distributions with precision $\lambda_k \mathbf{a}_k(i)\mathbf{Q}_i^T\mathbf{Q}_i$ according to $p(\mathbf{u}_k(i)|\mathbf{a}_k(i)) = \mathcal{N}(0, (\lambda_k \mathbf{a}_k(i)\mathbf{Q}_i^T\mathbf{Q}_i)^{-1})$, where $\mathbf{Q}_i$ is the matrix applying the Laplacian operator to the flow field at the i-th location. Based on the assumption that the flow field should be smooth, it is common to assume this type of prior privileging low frequency motion fields [11].

The probability density function in (2) may be written as $p\left(\mathbf{u}_k(i)\right) = \int_0^\infty p\left(\mathbf{u}_k(i)|\mathbf{a}_k(i)\right) p\left(\mathbf{a}_k(i)\right) d\mathbf{a}_k(i)$, where the variables $\mathbf{a}_k(i)$ are hidden because they are not apparent in (2) since they have been integrated out. As the *degrees of freedom* parameter $\nu_k \to \infty$, the pdf of $\mathbf{a}_k(i)$ has its mass concentrated around its mean. This in turn reduces the Student's-t pdf to a Normal distribution, because all $\mathbf{u}_k(i)$, $k \in \{x, y\}$ are drawn from the same normal distribution with precision λ_k, since $\mathbf{a}_k(i)$ = 1 in that case. On the other hand, when $\nu_k \to 0$ the prior becomes uninformative. In general, for small values of ν_k the probability mass of the Student's-t pdf is more "heavy tailed".

We assume that the horizontal and vertical motion fields are independent at each pixel location. This assumption makes subsequent calculations tractable and is common in Bayesian image analysis. By defining the $N \times N$ diagonal matrices $\mathbf{A}_k = \text{diag}[\mathbf{a}_k(1), \ldots, \mathbf{a}_k(N)]^T$, $k \in \{x, y\}$, the pdf of the horizontal and vertical motion fields may now be expressed by $p\left(\mathbf{u}_k|\mathbf{A}_k\right) = \mathcal{N}\left(\underline{\mathbf{0}}, \left(\lambda_k \mathbf{Q}^T\mathbf{A}_k\mathbf{Q}\right)^{-1}\right)$, where $\mathbf{Q}$ is the Laplacian operator applied to the whole image and $\underline{\mathbf{0}}$ is a $N \times 1$ vector of zeros. Then, the overall pdf of the motion field $\mathbf{u} = [\mathbf{u}_x, \mathbf{u}_y]^T$ is given by $p(\mathbf{u}) = p(\mathbf{u}_x|\mathbf{A}_x)p(\mathbf{u}_y|\mathbf{A}_y)$, or equivalently:

$$p\left(\mathbf{u}|\tilde{\mathbf{A}}\right) = \mathcal{N}\left(\underline{\mathbf{0}}, \left(\lambda\tilde{\mathbf{Q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{Q}}\right)^{-1}\right), \tag{3}$$

where the $2N \times 1$ vector $\lambda = [\lambda_x, \lambda_y]^T$, the $2N \times 2N$ matrix $\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_y \end{bmatrix}$, the $2N \times 2N$ matrix $\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$ and $\mathbf{0}$ is a zero matrix of size $N \times N$. Hence, following (3), the marginal distribution $p(\mathbf{u})$ has a closed form.

The optical flow equation (1) may be written in matrix-vector form as:

$$\begin{bmatrix} \mathbf{G}_x & \mathbf{G}_y \end{bmatrix} \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_y \end{bmatrix} + \mathbf{w} = \mathbf{d}. \tag{4}$$

where the block diagonal $N \times 2N$ matrix $\mathbf{G} = \begin{bmatrix} \mathbf{G}_x & \mathbf{G}_y \end{bmatrix}$, with $\mathbf{G}_x = \text{diag}\left[\frac{\partial \mathbf{I}(\mathbf{x}_1)}{\partial x}, \ldots, \frac{\partial \mathbf{I}(\mathbf{x}_N)}{\partial x}\right]^T$, $\mathbf{G}_y = \text{diag}\left[\frac{\partial \mathbf{I}(\mathbf{x}_1)}{\partial y}, \ldots, \frac{\partial \mathbf{I}(\mathbf{x}_N)}{\partial y}\right]^T$ contains the spatial derivatives in the horizontal and vertical directions lexicographically and the $N \times 1$ vector $\mathbf{d} = [\mathbf{I}(\mathbf{x}_1) - \mathbf{J}(\mathbf{x}_1), \ldots, \mathbf{I}(\mathbf{x}_N) - \mathbf{J}(\mathbf{x}_N)]^T$ contains the temporal image differences. In order to take into account higher order terms of the Taylor series expansion of the brightness constancy constraint, we add a noise term $\mathbf{w}$ to (1) which is assumed to have spatially varying Student's t-statistics $\mathbf{w} \sim \mathcal{N}\left(\underline{\mathbf{0}}, (\lambda_\eta \mathbf{B})^{-1}\right)$, where $\lambda_\eta \mathbf{B}$ is the noise precision matrix and $\mathbf{B} = \text{diag}[\mathbf{b}(1), \ldots, \mathbf{b}(N)]^T$. The t-distribution implies that $\mathbf{b}(i) \sim Gamma\left(\mu/2, \mu/2\right)$.

Following the optical flow matrix-vector formulation in (4) and the noise modeling, we come up with the probability of the temporal image differences given the motion vectors: $p(\mathbf{d}|\mathbf{u}) = \mathcal{N}\left(\mathbf{G}\mathbf{u}, (\lambda_\eta \mathbf{B})^{-1}\right)$.

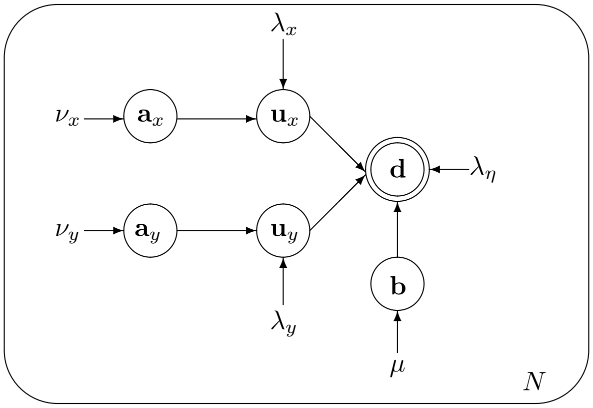

Figure 1. Graphical model for optical flow.

The above probabilistic formulation of the optical flow problem is represented by the graphical model of figure 1. As it may be observed, $\mathbf{d}$ is the vector containing the observations (temporal differences), denoted by the double circle, $\mathbf{u} = [\mathbf{u}_x, \mathbf{u}_y]^T$, $\mathbf{a}_x$, $\mathbf{a}_y$, $\mathbf{b}$, are the hidden variables of the model, denoted by the simple circles and λ_x, λ_y, λ_η, ν_x, ν_y and μ are the model's parameters. Notice that all of the variables and the observations are of dimension N except of the vector $\mathbf{u}$ collecting the horizontal and vertical motions. This shows the ill-posedness of the original optical flow problem where we seek $2N$ unknowns (vectors $\mathbf{u}_x$ and $\mathbf{u}_y$) with only N observations (vector $\mathbf{d}$).

3 Model inference

In the fully Bayesian framework, the complete data likelihood, including the hidden variables and the parameters of the model, is given by $p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) = p\left(\mathbf{d}|\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\mathbf{u}|, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) p\left(\tilde{\mathbf{A}}; \theta\right) p\left(\mathbf{b}; \theta\right)$ where $\theta = [\lambda_\eta, \lambda_x, \lambda_y, \mu, \nu_x, \nu_y]$ gathers the parameters of the model. Estimation of the model parameters could be obtained through maximization of the marginal distribution of the observations $p\left(\mathbf{d}; \theta\right)$:

$$\hat{\theta} = \arg\max_{\theta} \int\int\int p\left(\mathbf{d}, \mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}; \theta\right) \, d\mathbf{u} \, d\tilde{\mathbf{A}} \, d\mathbf{b}.$$

However, in the present case, this marginalization is not possible, since the posterior of the latent variables given the observations $p(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}|\mathbf{d})$ is not known explicitly and inference via the Expectation-Maximization (EM) algorithm may not be obtained. Thus, we resort to the variational methodology [2, 4] where we have to maximize a lower bound of $p\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right)$ by employing the *mean field approximation* [4]. Due to lack of space, the details of the derivation are given in [6]. The update equations for the motion vectors are given here.

Therefore, in the variational E-step of the algorithm the motion vectors are estimated by:

$$\mathbf{u}_x^{(t+1)} = \lambda_\eta^{(t)} \mathbf{R}_x^{(t)} \mathbf{B}^{(t)} \mathbf{G}_x \left(\mathbf{d} - \mathbf{G}_y \mathbf{u}_y^{(t)}\right) \quad (5)$$

$$\mathbf{u}_y^{(t+1)} = \lambda_\eta^{(t)} \mathbf{R}_y^{(t)} \mathbf{B}^{(t)} \mathbf{G}_y \left(\mathbf{d} - \mathbf{G}_x \mathbf{u}_x^{(t)}\right) \quad (6)$$

where t indicates the time step and the $N \times N$ matrix

$$\mathbf{R}_x^{(t)} = \left(\lambda_\eta^{(t)} \mathbf{G}_x^T \mathbf{B}^{(t)} \mathbf{G}_x + \lambda_x^{(t)} \mathbf{Q}^T \mathbf{A}_x^{(t)} \mathbf{Q}\right)^{-1} \quad (7)$$

and its counterpart $\mathbf{R}_y^{(t)}$ are computed using the Lanczos method [12]. Moreover, in the variational E-step, the expectations of the hidden random variables $\mathbf{a}_x(i)$ and $\mathbf{a}_y(i)$ and the noise $\mathbf{b}(i)$ are updated. In the variational M-step, the lower bound of $p\left(\mathbf{u}, \tilde{\mathbf{A}}, \mathbf{b}\right)$ is maximized with respect to the model parameters (noise precision λ_η, flow precisions λ_x, λ_y, degrees of freedom ν_x, ν_y, μ). The reader is referred to [6] for the expressions of the respective update equations.

Let us notice that as we can see from (5) and (6), there is a dependency between $\mathbf{u}_x^{(t+1)}$ and $\mathbf{u}_y^{(t)}$, as well as between $\mathbf{u}_y^{(t+1)}$ and $\mathbf{u}_x^{(t)}$. This is also the case in the standard Horn-Schunck method. The main contribution of the proposed approach is that all of the model parameters may be computed from the two images by the variational EM algorithm. In the standard Horn-Schunck methods and its variants these parameters are set heuristically. This also includes the noise precision λ_η which represents the error due to the linearization of the brightness constancy constraint which is generally omitted in standard optical flow estimation methods.

4 Experimental results

The method proposed herein is a principled Bayesian generalization of the Horn-Schunck (HS) method [7]. Therefore, our purpose is to examine its appropriateness to replace it in the initialization of more advanced optical flow schemes. We have also included the well-known and established rival algorithm of Lucas-Kanade (LK) [8]. These are the two methods widely used for initializing more sophisticated optical flow algorithms. Moreover, we have included in the comparison the algorithm proposed in [3], which combines the above two algorithms for feature tracking.

The proposed method was tested on image sequences including both synthetic and real scenes. A synthetic sequence included in our experiments consists of two textured triangles moving to different directions with different velocity magnitudes (*Triangles*). We have also applied our method to the *Yosemite* sequence as well as to the *Dimetrodon* sequence obtained from the Middlebury database [1].

To evaluate the performance of the method two performance measures were computed. The average angular error (AAE) [13], which is the most common measure of performance for optical flow and the average magnitude of difference error (AME) [9]. The latter measure normalizes the errors with respect to the ground truth and ignores normalized error vector norms smaller than a threshold T. We have employed $T = 0.35$ in our evaluation.

The numerical results are summarized in Table 1, where it may be observed that the method proposed in this paper provides better accuracy with regard to the other methods. More specifically, our algorithm largely outperforms the Lucas-Kanade method and is clearly better than the Horn-Schunck algorithm. Notice that the JLK algorithm is not very accurate as its behavior depends partially on a Lucas-Kanade scheme which fails in all cases (first table row). We conclude that JLK which combines the two approaches may perform better for sparse optical flow applied to features [3] but not for dense flow estimation.

Table 1. Comparative Results.

Average Angular Error (AAE)			
Method	*Triangles*	*Yosemite*	*Dimetrodon*
LK [8]	8.58°	11.65°	27.52°
HS [7]	5.57°	5.43°	8.50°
JLK [3]	6.95°	7.97°	33.14°
Proposed	**3.93°**	**4.45°**	**4.31°**

Average Magnitude Error (AME)			
Method	*Triangles*	*Yosemite*	*Dimetrodon*
LK [8]	0.17	0.26	0.56
HS [7]	0.13	0.14	0.49
JLK [3]	0.18	0.18	0.65
Proposed	**0.10**	**0.12**	**0.13**

The proposed algorithm takes on average less than a minute to converge o a standard PC running MATLAB, depending on the number of image pixels (e.g. it takes 80 seconds for the 584×388 sized *Dimetrodon* sequence). More than half of this time is due to the Lanczos method used for diagonalizing $\mathbf{R}_x$ in (7) (and respectively $\mathbf{R}_y$).

5 Conclusion

The optical flow estimation method proposed in this paper relies on a probabilistic formulation of the problem along with a variational Bayesian inference approach. The spatially varying Student's t-distribution of the motion vectors achieves selective application of smoothness leaving motion edges unaffected. Furthermore, any residuals of the linearization of the brightness constancy constraint are also modeled leading to better accuracy. A perspective of this study is to extensively evaluate the use of the algorithm as an initialization method for methods capturing large motions which incorporate more sophisticated constraints on the motion field, like the method in [5].

References

[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *ICCV*, pages 1–8, 2007. 4

[2] M. J. Beal. The variational Bayesian EM algorithm for incomplete data with application to scoring graphical model structures. *Bayesian Statistics*, 7:453–464, 2003. 3

[3] S. Birchfield and S. Pundlik. Joint tracking of features and edges. In *IEEE CVPR*, 2008. 1, 2, 3, 4

[4] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. 2, 3

[5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, volume 4, pages 25–36, 2004. 1, 4

[6] T. Gkamas, G. Chantas, and C. Nikou. Variational-Bayes optical flow. Technical Report 2012-1, University of Ioannina, Department of Computer Science, 2012. 3

[7] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. 1, 2, 3, 4

[8] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intellegence, (IJCAI)*, pages 674–679, 1981. 2, 3, 4

[9] B. McCane, K. Novins, D. Crannitch, and B. Galvin. On benchmarking optical flow. *Computer Vision and Image Understanding*, 84(1):126–143, 2001. 4

[10] E. Memin and P. Perez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703 –719, 1998. 1

[11] R. Molina. On the hierarchical Bayesian approach to image restoration. applications to astronomical images. *IEEE Trans. PAMI*, 16(11):1122–1128, 1994. 2

[12] C. Paige and M. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, 12:617–629, 1975. 3

[13] S. Roth and M. Black. On the spatial statistics of optical flow. *International Journal of Computer Vision*, 74(1):33–50, 2007. 1, 4

[14] D. Sun, S. Roth, J. P. Lewis, and M. J. Black. Learning optical flow. In *ECCV*, pages 83–97, 2008. 1

[15] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *CVPR*, pages 2464–2471, 2010. 1