

# Temporal Constraints in Large Optical Flow Estimation

Agustín Salgado and Javier Sánchez

Computer Science Department,  
University of Las Palmas de G.C.  
35017 Las Palmas de G.C., Spain  
{asalgado, jsanchez}@dis.ulpgc.es  
<http://serdis.dis.ulpgc.es/~jsanchez>

**Abstract.** The aim of this work is to propose a model for computing the optical flow in a sequence of images. We introduce a new temporal regularizer that is suitable for large displacements. We propose to decouple the spatial and temporal regularizations to avoid an incongruous formulation. For the spatial regularization we use the Nagel–Enkelmann operator and a newly designed temporal regularization. Our model is based on an energy functional that yields a partial differential equation (PDE). This PDE is embedded into a multipyramidal strategy to recover large displacements. A gradient descent technique is applied at each scale to reach the minimum.

## 1 Introduction

In this paper we consider the problem of estimating the optical flow assuming that the objects may undergo large displacements. The difference with respect to other related approaches is that we exploit the temporal dimension of the sequence. Most of the better known methods only deal with the problem of estimating the optical flow between two frames, ignoring that the sequence comprises several images which are all related. We show in this paper that it is possible to integrate the temporal information to improve the results.

We propose a variational technique in where an energy functional is minimized yielding a diffusion–reaction PDE. These kind of energy–based approaches have been largely used in optical flow estimation. Horn and Schunck [8], for instance, propose to minimize the so–called optical flow constraint equation (OFC) together with a smoothing term depending on the optical flow gradient. Later some authors have proposed several improvements to overcome the shortcomings of this method like in [3], [6], [11], [7].

In order to compute large displacements a common strategy is to use a multipyramidal decomposition like in [3] in where each scale is represented by decreasing size of images.

The first works on spatio–temporal methods are due to Nagel [10] and Black and Anandan [4]. More recently in [13] and [5] the authors propose similar continuous models for the spatio–temporal smoothness constraint.

In this work we propose a novel temporal regularizing term which is explicitly designed to support for large displacements. The previous mentioned spatio-temporal methods treat the spatial and temporal dimensions in the same way. We show here that when large displacements are present, it is not suitable to use temporal derivatives. It is more convenient to separate the temporal smoothing term. We also show that the major contributions of spatio-temporal smoothness are given by more stable and accurate results. This is clearly stated in the experimental results in where some comparisons with its corresponding spatial method is carried out on synthetic and real sequences.

In Sect. 2 we give an overview on related optical flow methods and a study on the generalization of spatial optical flows and justify the use of temporal regularizers. In Sect. 3 we derive the numerical scheme for the energy proposed in the previous section and in Sect. 4 we demonstrate the performance of our method by using synthetic and real sequences and compare with spatial methods. Finally in Sect. 5 the conclusions.

## 2 The Method

The optical flow,  $\mathbf{h}(\mathbf{x}) = (u(\mathbf{x}), v(\mathbf{x}))^T$ , is the apparent motion of pixels in a sequence of images. One of the first in introducing a variational formulation for the computation of the optical flow was Horn and Schunck [8]. They proposed the so-called optical flow constraint equation  $\frac{dI(\mathbf{x},t)}{dt} = \nabla I \cdot \mathbf{h} + I_t = 0$ , which states that the image intensity remains constant through the sequence –known as the Lambertian assumption–.

This equation is valid when the object displacements in the scene are continuous. This is not the case in our approach since the objects may undergo large displacements. A different formulation which is the corresponding to the optical flow constraint equation in the discontinuous case is  $I_1(\mathbf{x}) - I_2(\mathbf{x} + \mathbf{h}(\mathbf{x})) = 0$ . Other approaches different from the quadratic form are commonly used (see for instance [9] and [12]).

Typically, the data term is accompanied by a smoothness term depending on the gradient of the flow. Horn and Schunck [8], for instance, proposed to minimize the square norm of the optical flow gradient  $\|\nabla \mathbf{h}\|^2 = \|\nabla u\|^2 + \|\nabla v\|^2$  which provides smooth solutions and yields an isotropic diffusion equation at the PDE. This model has been improved during the last years and some authors has introduced some different approaches in order to respect the image or flow discontinuities. A well-known approach is that of Nagel–Enkelmann [11] that introduce an operator depending on the image gradient that enables anisotropic diffusion in order to respect the object contours. Other improvements are given by the methods explained in [2], [6] and [7].

Variational temporal regularizers have been introduced in some works like in [4] and [10]. More recently Weickert and Schnörr [13] introduced a method in where they used the continuous optical flow constraint equation as data term and a regularizing term of the form  $\mathcal{R}(|\nabla_3 u|^2 + |\nabla_3 v|^2)$ . In this case the temporal derivative is treated in the same manner as the spatial derivatives.

We separate the temporal and spatial derivatives. In the spatial domain we use the Nagel–Enkelmann operator and for the temporal domain the functional  $T(\mathbf{h}_i, \mathbf{h}_{i+1}) = \Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\mathbf{x} + \mathbf{h}_i)\|^2) + \Phi(\|\mathbf{h}_i - \mathbf{h}_{i-1}(\mathbf{x} + \mathbf{h}_{i-1}^*)\|^2)$  including the backward optical flows,  $\mathbf{h}^*$  as

$$\begin{aligned}
 E(\mathbf{h}) = & \sum_{i=1}^{N-1} \int_{\Omega} (I_i - I_{i+1}(\mathbf{x} + \mathbf{h}_i))^2 \, d\omega \\
 & + \alpha \sum_{i=1}^{N-1} \int_{\Omega} \text{trace}(\nabla \mathbf{h}_i^T \mathbf{D}(\nabla I_i) \nabla \mathbf{h}_i^T) \, d\omega \\
 & + \beta \sum_{i=1}^{N-2} \int_{\Omega} \Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\mathbf{x} + \mathbf{h}_i)\|^2) \, d\omega \\
 & + \beta \sum_{i=2}^{N-1} \int_{\Omega} \Phi(\|\mathbf{h}_i - \mathbf{h}_{i-1}(\mathbf{x} + \mathbf{h}_{i-1}^*)\|^2) \, d\omega \tag{1}
 \end{aligned}$$

where  $\Phi(x^2) = 1 - \gamma e^{-\frac{x^2}{\gamma}}$ . When displacements are very small,  $\mathbf{h}_i - \mathbf{h}_{i+1}(\mathbf{x} + \mathbf{h}_i)$  is an approximation of the temporal derivative.

With the last two integrals we are implicitly assuming a model for the object velocities. We enforce that the  $\mathbf{h}_i$  functions be similar in magnitude and direction, so this scheme is more suitable for objects that move with constant velocity in a permanent direction.

### 3 Minimizing the Energy

A functional variation of our energy leads to the Euler–Lagrange equations:

$$\begin{aligned}
 \mathbf{0} = & -(I_i(\mathbf{x}) - I_{i+1}(\mathbf{x} + \mathbf{h}_i)) \nabla I_{i+1}(\mathbf{x} + \mathbf{h}_i) \\
 & - \alpha (\mathbf{div}(\mathbf{D}(\nabla \mathbf{h}_i) \nabla u_i), \mathbf{div}(\mathbf{D}(\nabla \mathbf{h}_i) \nabla v_i))^T \\
 & + \beta \Phi'(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\mathbf{x} + \mathbf{h}_i)\|^2) \\
 & \cdot \left( (\mathbf{h}_i - \mathbf{h}_{i+1}(\mathbf{x} + \mathbf{h}_i))^T (\mathbf{Id} - \nabla \mathbf{h}_{i+1}^T(\mathbf{x} + \mathbf{h}_i)) \right) \\
 & + \beta \Phi'(\|\mathbf{h}_i - \mathbf{h}_{i-1}(\mathbf{x} + \mathbf{h}_{i-1}^*)\|^2) \\
 & \cdot (\mathbf{h}_i - \mathbf{h}_{i-1}(\mathbf{x} + \mathbf{h}_{i-1}^*)) \tag{2}
 \end{aligned}$$

We apply a gradient descent technique to reach the solution of the previous system of equations and embed it into a multi–pyramidal approach to deal with large displacements. We create a pyramid of scales for the whole sequence with different size of images and solve the previous system of equations at each level. Once we obtain a stable solution for a scale we use this as a first estimate for a finer scale.

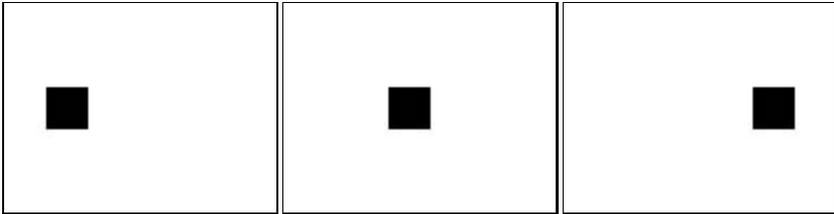
Therefore we have a number of scales  $s_1, s_2, \dots, s_n$ . At each scale we solve the previous system of equations for the whole set of unknowns  $\{u_i^s, v_i^s\}$  and then we use this as a first approximation for the following scale  $\{u_i^{s_1}, v_i^{s_1}\} \rightarrow \{u_i^{s_2}, v_i^{s_2}\} \rightarrow \dots \{u_i^{s_n}, v_i^{s_n}\}$ .

## 4 Experimental Results

The purpose of these experimental results is to show the performance of our novel method with respect to its spatial counterpart. We are also interested in comparing this new method with its simple form given by the forward temporal cost matching function  $\mathcal{T}(x) = \Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1} + \mathbf{h}_i\|^2)$ . In the sequel we name the spatial method as "Spatial", the simpler temporal method as "Temporal" and the complete temporal method as "Bi-Temporal" or "Bidirectional-Temporal" – Eq. (1) –.

### 4.1 Translating Square

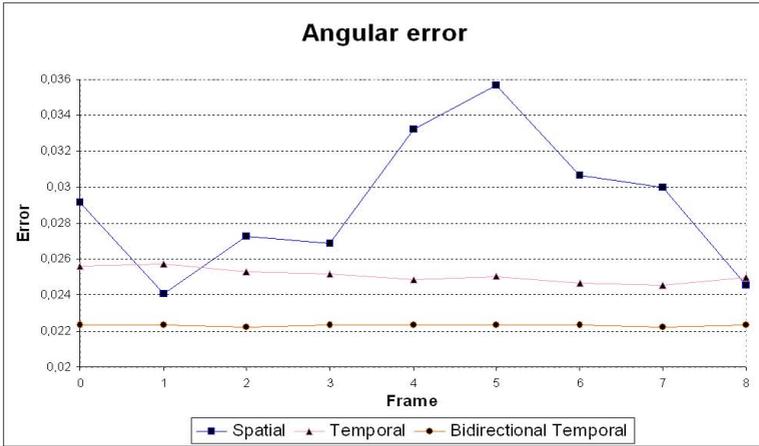
The first sequence consists of ten frames with a black square moving ten pixels horizontally forward with a constant velocity over a white background. In Fig. 1 we can see several frames of the sequence and in Fig. 2 we show the angular error for all the frames using the three methods. In table 1 we show the mean angular and euclidean errors.



**Fig. 1.** Frames 0, 4 and 9 of the square sequence

From table 1 and Fig. 2 we can see that the temporal methods are very stable: the euclidean and angular errors for all the frames are very similar. However, the "Spatial" method is clearly unstable.

The improvement of the "Temporal" method with respect to its spatial counterpart is about 13,79% of the angular error and in the case of the "Bi-Temporal" is about 24,14%. The euclidean error for the "Temporal" method is bigger than the spatial method. This is justifiable because the last optical flow has a big error with respect to the rest of frames. The simple temporal method does not improve the last solution because it does not receive information from other frames. Even more, when this last estimate is bad, it propagates the bad guesses to the rest of sequence. In the case of the "Bi-Temporal" method the improvement with respect to the spatial one is about 31,58%. This method is not affected by the last optical flow as in the previous case.



**Fig. 2.** Angular error for the square sequence

**Table 1.** Mean angular and euclidean errors for the square sequence:  $AE_\mu$  is the mean angular error and  $AE_\sigma$  its standard deviation;  $EE_\mu$  is the mean euclidean error and  $EE_\sigma$  its standard deviation

Method	$AE_\mu$	$AE_\sigma$	$EE_\mu$	$EE_\sigma$
Spatial	0,029°	3,6E-3°	0,0189	0,013
Temporal	0,025°	3,7E-4°	0,0246	3,8E-4
Bi-Temporal	0,022°	3,0E-5°	0,0132	1,9E-4

## 4.2 Marble Blocks

We used the Marble Block sequence – Fig. 3 – for the second test.

In Fig. 4 we can see the ground truth and the flows obtained with the spatial and temporal methods.

If we look at this image we observe that both solutions are similar but there are some differences: The floor is smoother in the temporal disparity map and the background is more respected. The right tower that disappears seems more continuous but, on the other hand, in the middle of the two left-most towers the disparity maps seem under-estimated. Globally, the temporal method seems to have better estimated the object discontinuities and the results inside each object are smoother.

In Fig. 5 we show the angular error for every frame of the sequence – for convenience we only show the central frames –. We may appreciate that the spatial method provides worse and less stable results. Nearly all the results for the temporal methods are improved. In all cases, the "Bi-Temporal" performs in a similar way as the "Temporal" but with a smaller magnitude.

As we could expect, the results obtained for this sequence are improved. This sequence is similar to the translating square in the sense that the objects are



Fig. 3. Frames 0, 10 and 20 of the Marble Blocks sequence\*

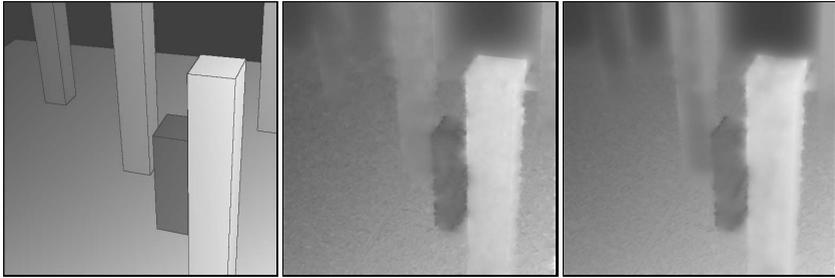


Fig. 4. On the left, the true disparity associated with frame 10; in the middle, the optical flow corresponding to the spatial method; and, on the right, the solution for the bidirectional temporal method

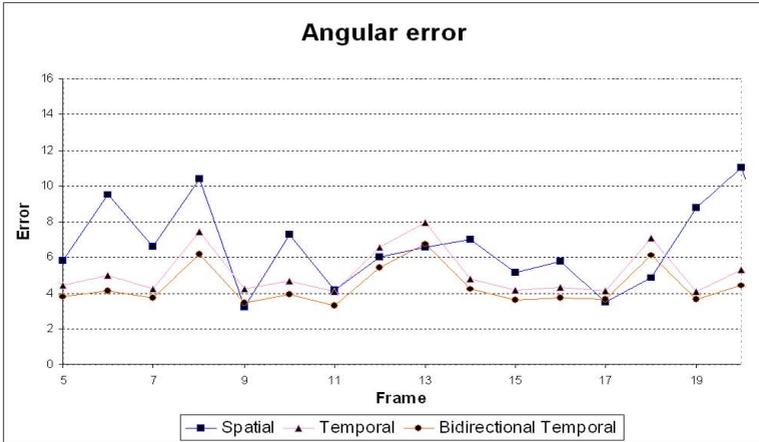


Fig. 5. Angular error for the Marble Block sequence

\* This sequence is composed of 30 frames and is copyright by H.-H. Nagel KOGS/IAKS, University of Karlsruhe, Germany, at [http://i21www.ira.uka.de/image\\_sequences/](http://i21www.ira.uka.de/image_sequences/).

**Table 2.** Mean angular and euclidean errors for the Marble Block sequence

Method	$AE_\mu$	$AE_\sigma$	$EE_\mu$	$EE_\sigma$
Spatial	6,695°	2,698°	0,2480	0,0963
Temporal	5,402°	1,327°	0,2081	0,0638
Bi-Temporal	4,731°	1,330°	0,1848	0,0661

moving with constant velocity in the same direction. In fact, the average motion of the sequence is 1,33 – with standard deviation 0,03 – except for frames 4, 8, 13, 18, 23 and 28 in which the average motion is 1,58 – with standard deviation 0,02–. This means that there is a shift on the magnitude of the velocities at these frames. If we look at Fig. 5 we observe maximum peaks that reduce the stability of the temporal methods. We suppose that this sequence was captured with different frame rates at these particular frames.

In table 2 we show the angular and euclidean errors and their standard deviations. The improvement of the "Temporal" method with respect to the "Spatial" one is about 19,31% for the angular error and 12,05% for the euclidean error. The improvement of the "Bi-Temporal" is about 29,33% for the angular and 25,48% for the euclidean error. In both cases the standard deviations are considerably reduced in a similar magnitude.

## 5 Conclusions

With this work we have proposed a novel variational method that includes a temporal smoothness term. We have created a new term that minimizes the difference of the optical flows that permits large displacements of the objects.

We have shown the necessity of including the backward flow in the energy functional. All the comparisons show that the bidirectional temporal method clearly outperforms the simpler temporal method.

As we can observe in the experimental results, the numerical studies show that this method provides more stable results. We have also shown that when translating motions are present on the scene then there is an increase of accuracy in all the frames.

To our knowledge, this is the first approach that considers temporal regularizations with large displacements.

## References

1. Alvarez, L., Deriche, R., Papadopoulos, T., Sánchez, J.: Symmetrical dense optical flow estimation with occlusions detection. International Journal of Computer Vision. Preprint (2006)
2. L. Alvarez, J. Weickert, and J. Sánchez, Reliable Estimation of Dense Optical Flow Fields with Large Displacements. International Journal of Computer Vision, 391 (2000) 41–56. An extended version maybe be found at Technical Report n°2 del Instituto Universitario de Ciencias y Tecnologías Cibernéticas

3. Anandan, P.: A Computational Framework and an Algorithm for the Measurement of Visual Motion. *International Journal of Computer Vision* 2, 283–310 (1989)
4. Black, M.J., Anandan, P.: Robust dynamic motion estimation over time. In: Proc. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 292–302 (1991)
5. Brox, T., Bruhn, A., Papenbergh, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J. (eds.) *ECCV 2004. Part IV. LNCS*, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
6. Cohen, I.: Nonlinear Variational Method for Optical Flow Computation. In: Proc. of the 8th Scandinavian Conference on Image Analysis, Norway (1993)
7. Deriche, R., Kornprobst, P., Aubert, G.: Optical flow estimation while preserving its discontinuities: a variational approach. In: Proc. Second Asian Conference on Computer Vision, vol. 2, pp. 290–295 (December 1995)
8. Horn, B., Schunck, B.: Determining Optical Flow. *Artificial Intelligence* 17, 185–203 (1981)
9. Mémin, E., Pérez, P.: Dense Estimation and Object-Based Segmentation of the Optical Flow with Robust Techniques. *IEEE Transactions on Image Processing* 7(5), 703–719 (1998)
10. Nagel, H.H.: Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. In: Faugeras, O. (ed.) *Computer Vision - ECCV 90. LNCS*, vol. 427, pp. 139–148. Springer, Heidelberg (1990)
11. Nagel, H.H., Enkelmann, W.: An Investigation of Smoothness Constraints for the Estimation of Displacements Vector Fields from Image Sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* 8, 565–593 (1986)
12. Wells, W.M., Viola, P., Atsumi, H., Nakajima, S., Kinikis, R.: Multimodal Volume Registration by Maximization of Mutual Information. *Medical Image Analysis* 1, 35–51 (1996)
13. Weickert, J., Schnörr, C.: Variational Optic Flow Computation with a Spatio-Temporal Smoothness Constraint. *Journal of Mathematical Imaging and Vision* 14, 245–255 (2001)