Illumination-robust variational optical flow based on cross-correlation*

József Molnár¹ and Dmitry Chetverikov²

¹ Computer and Automation Research Institute Hungarian Academy of Sciences, Budapest, Hungary

 ² Computer and Automation Research Institute csetverikov@sztaki.hu
 and Eötvös Loránd University, Budapest, Hungary

Abstract

We address the problem of variational optical flow for outdoor video processing applications that need fast operation and robustness to drastic variations in illumination. Recently, a solution has been proposed [12] based on the photometric invariants of the dichromatic reflection model [15]. However, this solution is only applicable to colour videos up to the validity of the dichromatic model. We propose a fast, illumination-robust variational scheme based on cross-correlation and applicable to both colour and greyscale sequences. We derive an explicit linearised algorithm for cross-correlation based variational optical flow and test the algorithm on challenging video data.

1 Introduction

Outdoor applications of machine vision such as vision-based intelligent vehicles [3], surveillance and traffic monitoring [9] require fast and robust solutions capable of handling situations when illumination and visibility may change suddenly and drastically. We are currently involved in a project whose purpose is traffic monitoring by mobile cameras. Traffic monitoring, vehicle counting and event detection are usually performed by video cameras mounted over roads in fixed positions. Traffic flow is measured in these fixed points at different times, and statistical models are used to estimate traffic between the points. An alternative solution is a mobile camera mounted on a vehicle participating in the traffic. In this case, traffic data is measured in varying points at different times. The two methods are complementary, as the mobile solution provides direct flow measurements between the fixed observation points, allowing to validate and improve the statistical models and to analyse the reasons of congestion.

In our project, we process greyscale videos provided by a calibrated stereo rig mounted on a regular bus carrying passengers to countryside settlements. Typical traffic images acquired by the mobile cameras are shown in Fig. 2 and Fig. 3. Obtaining correct optical flow for such video data is difficult for a number of reasons:

• Each camera continously adjusts itself so as to optimally use the current intensity range. This can cause global intensity changes between consecutive frames.

^{*}The authors thank the reviewers for valuable comments.



Horn-Schunck

proposed

Figure 1: Top row: Two frames from the greyscale foggy sequence [16]. Bottom row: Optical flow magnitudes.

- Although the two cameras of a stereo rig are synchronised, the intensity adjustment is done separately. This can cause global intensity difference between the two stereo images.
- Dynamic weather-related factors such as fog, sunshine, and clouds affect different parts of image in different ways. This can cause local intensity changes between consecutive frames. Intensity can also change locally due to shadow, shading, and shiny surfaces of cars.

In this paper, we address the problem of estimating the optical flow for each camera separately. Traditional optical flow methods based on the brightness constancy assumption fail when applied to video data strongly contaminated by the above factors. Fig. 1 illustrates the aim of our study. The upper row of the figure shows two frames from a foggy sequence from the Karlsruhe traffic video database [16]. The sequence is characterised by low and permanently changing visibility due to the intrinsic dynamics of the fog. The bottom row shows the results of optical flow estimation by the classical Horn-Schunck method [8] and the correlation-based method we propose in this paper.

Among numerous techniques used for motion estimation, the variational optical flow calculation methods are currently the most accurate. The variational problem for optical flow calculation is usually formulated as finding the displacement function $\mathbf{u}(\mathbf{x})$ that minimises a functional of the form

$$\mathcal{F}(\mathbf{u}) = \int_{\Omega} \left(E(\mathbf{x}, \mathbf{u}) + \lambda S(\boldsymbol{\nabla}_{\mathbf{x}} \otimes \mathbf{u}) \right) d\mathbf{x}, \tag{1}$$

where $E(\mathbf{x}, \mathbf{u}), \mathbf{x} = (x, y), \mathbf{u} = (u, v)$, is a scalar function describing optical constraints, S a scalar function of the dyadic product $\nabla_{\mathbf{x}} \otimes \mathbf{u}$ accounting for the smoothness of the flow, Ω the image domain, and λ a parameter.

Denote by $u_{\alpha}, v_{\beta}, \alpha, \beta \in \{x, y\}$, the spatial derivatives of the optical flow. Often, the smoothness

(regularisation) term is selected as

$$S = \frac{1}{2} \sum_{\alpha} \left(u_{\alpha}^2 + v_{\alpha}^2 \right),\tag{2}$$

then the solution of Eq. (1) is given by the Euler-Lagrange equations

$$\nabla_{\mathbf{u}} E = \lambda \,\Delta \mathbf{u},\tag{3}$$

where the Laplacian Δ is applied to each component of u. These equations are solved numerically. Most methods use an iterative solver that improves the optical flow estimate obtained in a previous step as $\mathbf{u} \to \mathbf{u}'$, repeating the procedure until a steady state is reached.

The classical Horn-Schunck [8] data term (energy function) is

$$E_{\rm HS} = \frac{1}{2} \left(I_t + uI_x + vI_y \right)^2 = \frac{1}{2} \left(I_t + \mathbf{u} \boldsymbol{\nabla}_{\mathbf{x}} I \right)^2, \tag{4}$$

where $I(\mathbf{x}, t)$ is the image brightness. The equation $I_t + \mathbf{u}\nabla_{\mathbf{x}}I = 0$ is a first-order Taylor approximation of brightness constancy between two consecutive images in a sequence: $I(\mathbf{x}+\mathbf{u}, t+1) = I(\mathbf{x}, t)$, thus minimising E_{HS} approximates the brightness constancy assumption.

At each iteration of the Horn-Schunck method, the updated displacement vector \mathbf{u}' is calculated as

$$\mathbf{u}' = \mathbf{A}^{-1} \big(\lambda \bar{\mathbf{u}} - I_t \boldsymbol{\nabla}_{\mathbf{x}} I \big), \tag{5}$$

where the central sum

$$\bar{\mathbf{u}}(x,y) = \mathbf{u}(x-1,y) + \mathbf{u}(x,y-1) + \mathbf{u}(x+1,y) + \mathbf{u}(x,y+1)$$
(6)

and

$$A_{\alpha\beta} = I_{\alpha}I_{\beta} + 4\lambda\delta_{\alpha\beta},\tag{7}$$

where $\alpha, \beta \in \{x, y\}$ and $\delta_{\alpha\beta}$ is the Kronecker delta. Since $A_{\alpha\beta}$ and $I_t \nabla_x I$ do not depend on u, they can be pre-computed resulting in a very efficient numerical scheme. For colour images, the energy term, the matrix **A** and the vector $I_t \nabla_x I$ are modified by taking the sum over the three channels; except for this, the same algorithm is applied.

In an attempt to cope with changing illumination, Kim at al. [10] modify the optical flow constraint by adding a linear term describing direct illumination $p(\mathbf{x})$ and ambient light $q(\mathbf{x})$:

$$I_t + \mathbf{u} \nabla_{\mathbf{x}} I + pI + q = 0.$$

However, this introduces two additional variables resulting in a highly under-determined equation. The method may tend to attribute intensity changes to illumination variations rather than to motion. When the modified optical flow constraint is used in the variational framework, the smoothness term needs setting two additional weights for p and q.

Another possibility is to include in the data term a function of image gradient which is known to be less sensitive to illumination variations. Brox et al. [4] avoid linearisation and use, for two consecutive frames I_0 and I_1 , the energy term

$$E = \sqrt{(I_1 - I_0)^2 + \gamma (\boldsymbol{\nabla}_{\mathbf{x}} I_1 - \boldsymbol{\nabla}_{\mathbf{x}} I_0)^2 + \varepsilon^2},$$

where γ is a parameter, while $\sqrt{x^2 + \varepsilon^2}$, $\varepsilon \ll 1$, is a differentiable version of |x|. Using the L^1 norm instead of L^2 makes the method more robust to noise and outliers. Zach et al. [18] also use L^1 , but linearise I_1 and apply a different approach to circumvent the non-differentiability of |x|.

Any differentiable function $z(I_x/I_y)$ is invariant under any differentiable transformation of intensity I' = f(I). In particular, the orientation of the gradient vector and the components of the unit gradient vector are functions of the ratio I_x/I_y . In principle, they can be used to obtain illumination-insensitive optical flow. However, this ratio is very noise-sensitive in image areas of low variation.

The above approaches are applicable to both greyscale and colour data. For colour data, Mileva et al. [12] propose to use the photometric invariants of the dichromatic reflection model [15] in order to make optical flow less sensitive to illumination changes, shadow and shading. They substitute the original RGB values by the normalised RGB or the angles ϕ , θ of the spherical (conical) transformation. The normalised RGB values are defined as $\frac{1}{N}(R, G, B)^T$, where N is either the arithmetic mean R + G + B or the geometric mean $\sqrt[3]{RGB}$. The angles of the spherical transformation are given by

$$\theta = \arctan\left(\frac{G}{R}\right)$$
$$\phi = \arcsin\left(\frac{\sqrt{R^2 + G^2}}{\sqrt{R^2 + G^2 + B^2}}\right)$$

The method is only applicable to colour data and under the assumption that the dichromatic reflection model [15] is valid. In particular, it will not work for colour light. We need a robust method applicable to greyscale images.

Of the discussed approaches, only the method by Mileva et al. [12] has been tested on data severely contaminated by changing illumination conditions. Optical flow methods can be tested on the Middlebury database [2] which currently does not contain outdoor data with drastic illumination changes we need for our purposes. In the rest of this paper, we present the proposed method and results of its application to challenging data containing natural and artificial illumination effects.

2 Cross-correlation based variational optical flow

There are two basic options to achieve robustness against varying illumination: using illuminationinvariant features and applying robust data metric. The two options can be combined: for example, Mileva et al. [12] use photometric invariants in a variational scheme with the robustified L^1 norm.

The method we propose in this paper uses normalised cross-correlation in a small window W. We apply cross-correlation to image intensity; however, it can be applied to any other image feature, such as colour, or photometric invariants. It is well-known that cross-correlation is a robust tool for comparing signal and image data. Cross-correlation is very efficient in block matching for numerous applications such as tracking, video coding or Particle Image Velocimetry.

Addressing the general problem of multimodal image matching, Hermosillo et al. [7] consider different statistical dissimilarity measures, including those based on correlation ratio and cross correlation. They derive the Euler-Lagrange equations for the statistical criteria and propose a number of sophisticated variational methods involving local probability density estimation using Parzen windows. The regularisation functional introduced by Alvarez et al. [1] is used. The resulting non-linear minimisation problem is solved by gradient descent. The method has been applied to elastic image matching [7], motion compensation [17], and 3D scene flow estimation [14]. It is very general, but too complex and slow for our purposes. In this section, we give an explicit, linearised iterative scheme for variational optical flow based on cross-correlation, which is robust, simple and fast.

Consider two consecutive frames, I_0 and I_1 . We use the standard smoothness term (2) but substitute the Horn-Schunck data term (4) by

$$E_{cor} = -\int_{\Omega} \frac{\int_{W} I_0(\mathbf{x} + \mathbf{x}') I_1(\mathbf{x} + \mathbf{x}' + \mathbf{u}(\mathbf{x} + \mathbf{x}')) d\mathbf{x}'}{\sqrt{\int_{W} I_0^2(\mathbf{x} + \mathbf{x}') d\mathbf{x}'} \cdot \int_{W} I_1^2(\mathbf{x} + \mathbf{x}' + \mathbf{u}(\mathbf{x} + \mathbf{x}')) d\mathbf{x}'}} d\mathbf{x},$$
(8)

where $\mathbf{x}' = (x', y')$ are local coordinates in the window W and the cross-correlation is negated since we minimise the functional.

Assuming a small window, after linearisation we obtain the following 'first-order' Euler-Lagrange equations¹:

$$\frac{1}{\sqrt{\int_{W} I_0^2 d\mathbf{x} \cdot \int_{W} I_1^2 d\mathbf{x}}} \left(\frac{\int_{W} I_0 I_1 d\mathbf{x}}{\int_{W} I_1^2 d\mathbf{x}} \int_{W} I_1 \nabla I_1 d\mathbf{x} - \int_{W} I_0 \nabla I_1 d\mathbf{x} \right) = \lambda \, \Delta \mathbf{u},\tag{9}$$

where for simplicity the local integration coordinates are changed to x and ∇ is ∇_x . A mathematically correct derivation of these equations is quite involved; it will be given in a forthcoming journal paper.

The above equations are nonlinear. Linearising Δu in a usual way leads to an iterative solution which is slow and imprecise. Fortunately, we can simplify the equations by further linearisation based on the small-size domain of the window integrals. This is done in three steps:

- Apply the first-order Taylor approximation $I_0 \approx I_1 \mathbf{u} \nabla I_1 I_{1t}$.
- Assume that **u** is approximately constant within W. Combining this with the above approximation, simplify the integrals like

$$\int_{W} I_0 I_1 d\mathbf{x} \approx \int_{W} I_1^2 d\mathbf{x} - \mathbf{u} \int_{W} \nabla I_1 d\mathbf{x} - \int_{W} I_1 I_{1t} d\mathbf{x}.$$

• In Eq. (9), do not linearise $\sqrt{\int_W I_0^2 d\mathbf{x}}$ but assume that it is constant within iteration at a given level of the Gaussian pyramid. When switching to another level, re-calculate the values.

The last step can be done because $\sqrt{\int_W I_0^2 d\mathbf{x}}$ essentially only modifies the weight of the smoothness term. Assuming that its order of magnitude is similar to the other integrals and observing that $\lambda \ll 1$, we conclude that this quantity has much less influence on the result.

After these steps, we obtain an iterative numerical solution similar to Eq. (5):

$$\mathbf{u}' = \mathbf{A}^{-1}\mathbf{b}(\bar{\mathbf{u}}),\tag{10}$$

¹The complete Euler-Lagrange equations contain an infinite series of integro-differential terms.

where

$$A_{\alpha\beta} = \sum_{\mathbf{x}\in W} I_1^2 \cdot \sum_{\mathbf{x}\in W} I_{1\alpha} I_{1\beta} - \sum_{\mathbf{x}\in W} I_1 I_{1\alpha} \cdot \sum_{\mathbf{x}\in W} I_1 I_{1\beta} + 4\lambda \delta_{\alpha\beta} \sum_{\mathbf{x}\in W} I_1^2 \cdot \sqrt{\sum_{\mathbf{x}\in W} I_1^2 \cdot \sum_{\mathbf{x}\in W} I_0^2}, \quad (11)$$

$$\mathbf{b}(\bar{\mathbf{u}}) = \lambda \sum_{\mathbf{x}\in W} I_1^2 \cdot \sqrt{\sum_{\mathbf{x}\in W} I_1^2 \cdot \sum_{\mathbf{x}\in W} I_0^2 \cdot \bar{\mathbf{u}}} + \sum_{\mathbf{x}\in W} I_1 I_{1t} \cdot \sum_{\mathbf{x}\in W} I_1 \nabla I_1 - \sum_{\mathbf{x}\in W} I_1^2 \cdot \sum_{\mathbf{x}\in W} I_{1t} \nabla I_1.$$
(12)

Here $\alpha, \beta \in \{x, y\}$ as before, the integrals are replaced by sums, and $\bar{\mathbf{u}}$ is given by Eq. (6).

Comparing the above expressions to Eq. (5) and Eq. (7), we observe that the original method operates with pointwise quantities, while the proposed method needs sums in small windows. Data arrays with such sums can be pre-calculated in an easy and fast way using techniques similar to running filters: when a sliding window moves to the next position, the previous sum is updated by adding the entering data and subtracting the exiting data. Since most of the computation time is consumed by the iterations, in practice the difference in processing speed between the Horn-Schunck method and our method is negligible.

Standard techniques to cope with large displacements, including scale-space or Gaussian pyramids and image warping, are applicable to the proposed method. When switching to a higher resolution level, the parameters of the iteration are re-calculated after the warping.

3 Test results

Systematic testing of the proposed method is in progress. We are comparing our algorithm to the state-of-the-art algorithms using our own data as well as the standard test data and benchmarks such as the Middlebury database [2]. We have hours of calibrated stereo video recorded by a rig mounted on a passenger bus travelling between countryside settlements. This data contains numerous examples of difficult lighting conditions that cause severe problems to the standard optical flow techniques. At the same time, we add artificial effects to the standard test data with ground truth available for the original images.

For test data without drastic illumination changes, such the Middlebury database, we do not expect our method to outperform the best current techniques in the precision of optical flow, although we do expect it to be competitive in that as well. Tab. 1 presents preliminary results comparing the proposed method to some state-of-the-art algorithms on the well-known synthetic Yosemite sequence with clouds illustrated in Fig. 4. Average angular (AAE) errors are given for the precise algorithms by Papenberg et al. [13] (2D regularisation), Bruhn et al. [5], Mémin-Pérez [11], and Alvarez et al. [1]. The figures are cited from [13] where quantitative results for other, less precise, methods are also available.

Papenberg (2006)	Bruhn (2005)	Mémin (2002)	Alvarez (2000)	proposed
2.44°	4.17°	4.69°	5.53°	4.36°

 Table 1: Average angular errors for the Yosemite sequence with clouds.

The cross-correlation based optical flow can handle severe illumination changes and other effects that influence visibility. Below, we show a few preliminary results demonstrating that cross-correlation is more robust than the traditional L^2 norm used by the Horn-Schunck and other algorithms. Since our



Horn-Schunck

proposed

Figure 2: Top row: Two consecutive frames of the greyscale road1 sequence. The brightness changes due to the self-adjustment of the camera. Middle row: image 1 mapped onto image 2 by the optical flow. Bottom row: Same part of the mapped images shown enlarged. Notice the distortions of the car and the house in the Horn-Schunck result.

algorithm and the Horn-Schunck procedure use the same smoothness term and similar iterations, the difference in results of the two methods is basically due to the difference in the data metric used.

The already mentioned Fig. 1 is a good illustration of robustness to fog which is a typical phenomenon in real-world road data. One can observe that the standard algorithm fails almost completely, while the proposed one is selective to the true motion of the cars despite the variations of the fog.

Images in Fig. 2 come from our own traffic data acquired by mobile cameras. The images are an example of the situation when the overall brightness suddenly changes due to the self-adjustment of the camera. We illustrate the difference between the two algorithms by mapping the first image onto the second one using the optical flow obtained. If the flow is correct, the result of the mapping is close to the second image. The proposed method is able to cope with the sudden brightness variation, while the Horn-Schunck result is deteriorated in the area of the car and the house. For better visibility, a part of the resulting images is displayed enlarged.

In another pair of images from our traffic data (see Fig. 3), we have added an artificial 'shadow' in the centre of the first image. The 'shadow' has an elliptical shape, and its density decreases from the centre outwards. Here again the standard algorithm is disturbed by the introduced effect, while our result does not contain major defects. Two parts of the resulting images are shown enlarged. Observe that the helmet is distorted by the standard algorithm, while our algorithm produces minor errors due to disocclusion.



Figure 3: Top row: Two consecutive frames of the greyscale road2 sequence with an artificial 'shadow' added in the centre of image 1. Middle row: image 1 mapped onto image 2 by the optical flow. Bottom row: Characteristic parts of the results shown enlarged.

Finally, Fig. 4 displays results for the synthetic Yosemite sequence where we again added a similar 'shadow' in the centre of the first image. The shadow is in the relatively dark part of the image, and it is barely visible. The bottom row of the figure shows the optical flows obtained by the two methods. The standard method is again visibly disturbed by the introduced effect, while our result is much more correct. However, the dark, poorly textured part of the sky poses certain problem to our algorithm, indicating that normalised cross-correlation, like most existing methods, may be sensitive to such areas. Typical average angular error of our method for the Yosemite sequence with the 'shadow' added is $7 - 8^{\circ}$. This is comparable to the results without the 'shadow' given in Tab. 1.

4 Conclusion

We have presented a novel algorithm for variational optical flow based on cross-correlation. The derivation of the linearised iterative numerical scheme assumes small size of the correlation window. We have used the first-order Euler-Lagrange equations, but higher order equations can also be derived and used. The obtained iterations are similar in structure to those of the classical Horn-Schunck method, but they use window sums rather than pixel-wise properties. Fortunately, this does not significantly affect the speed of the algorithm since the sums can be pre-calculated for subsequent iterations in an efficient way. The proposed method can be used with greyscale as well as colour images, and standard techniques to handle large displacements are applicable.

We have qualitatively demonstrated that cross-correlation is significantly more robust to drastic changes in illumination than the Horn-Schunck method. Currently, we are quantitatively comparing our



Figure 4: Top row: Two consecutive frames of the synthetic Yosemite sequence with an artificial 'shadow' added in the centre of image 1. Bottom row: Optical flow vectors.

method to the state-of-the-art robustified methods such as [4, 18] and, for colour data, [12]. We are also testing the mean-shifted version of the normalised cross-correlation which is easily obtained at no additional cost. At the same time, we are planning to test different data terms using the implicit non-linear variational optical flow algorithm [6]. The algorithm can accommodate a large variety of energy functions as it does not assume any particular analytical form of the function.

References

- [1] L. Alvarez, J. Weickert, and J. Sanchez. Reliable Estimation of Dense Optical Flow Fields with Large Displacements. *International Journal of Computer Vision*, 39:41–56, 2000.
- [2] S. Baker, S. Roth, D. Scharstein, M. Black, J. Lewis, and R. Szeliski. A database and evaluation methodology for optical flow. In *Proc. International Conf. on Computer Vision*, pages 1–8, 2007.
- [3] M. Bertozzi, A. Broggiand, and A. Fasciol. Vision-based intelligent vehicles: State of the art and perspectives. *Robotics and Autonomous Systems*, 31:1–16, 2000.
- [4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proc. European Conf. on Computer Vision*, volume IV, pages 25–36, 2004.
- [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*, 61(3):211–231, 2005.

- [6] S. Fazekas and D. Chetverikov. An implicit non-linear scheme for variational optical flow calculation. In 7th Conference of Hungarian Association for Image Processing and Pattern Recognition, 2009.
- [7] G. Hermosillo, C. Chefd'hotel, and O. Faugeras. Variational Methods for Multimodal Image Matching. *International Journal of Computer Vision*, 50:329–343, 2002.
- [8] B. Horn and B. Schunck. Determining optical flow. Artificial Intelligence, 17:185–203, 1981.
- [9] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis. A survey of video processing techniques for traffic applications. *Image and Vision Computing*, 21:359–381, 2003.
- [10] Y.-H. Kim, A.M. Martinez, and A.C. Kak. Robust motion estimation under varying illumination. *Image and Vision Computing*, 23:365–375, 2004.
- [11] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46:129–155, 2002.
- [12] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-Robust Variational Optical Flow with Photometric Invariants. In *Lecture Notes in Computer Science*, volume 4713, pages 152–162. Springer, 2007.
- [13] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67:141–158, 2006.
- [14] J.-P. Pons, R. Keriven, O. Faugerasand, and G. Hermosillo. Variational Stereovision and 3D Scene Flow Estimation with Statistical Similarity Measures. In *Proc. International Conf. on Computer Vision*, volume 1, pages 597–602, 2003.
- [15] S.A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10(4):210–218, 1985.
- [16] University of Karlsruhe, Institute of Algorithms and Cognitive Systems. Image Sequence Server. i21www.ira.uka.de/image_sequences/, 1998.
- [17] G. Hermosillo Valadez. Method and System for Motion Compensation in a Temporal Sequence of Images. US Patent Application Publication, 2005. Pub. No.: US 2005/0265611 A1.
- [18] C. Zach, T. Pock, and H. Bischof. A Duality Based Approach for Realtime TV-L1 Optical Flow. In *Pattern Recognition (Proc. DAGM)*, pages 214–223, Heidelberg, Germany, 2007.