Computer Science Department of The University of Auckland
CITR at Tamaki Campus (http://www.citr.auckland.ac.nz)

CITR-TR-102 October 2001

Experience with Optical Flow in Colour Video Image Sequences

John Barron[1] and Reinhard Klette[2]

Abstract

The paper studies optical flow methods on colour frames captured by a digital video camera. The paper reviews related work, specifies some new colour optical flow constraints and reports on experimental evaluations for one image sequence.

[1] Department of Computer Science, University of Western Ontario, London, Ontario, Canada, N6A 5B7, barron@csd.uwo.ca

[2] Center for Image Technology and Robotics Tamaki Campus, The University of Auckland, Auckland, New Zealand. r.klette@auckland.ac.nz

Experience with Optical Flow in Colour Video Image Sequences

John Barron* and Reinhard Klette†

Abstract

The paper studies optical flow methods on colour frames captured by a digital video camera. The paper reviews related work, specifies some new colour optical flow constraints and reports on experimental evaluations for one image sequence.

1 Introduction

Consider colour images $\mathbf{C}_1$, $\mathbf{C}_2$, ... $\mathbf{C}_n$ captured as n consecutive images of a video sequence where the camera is moving. An important problem consists in measuring image displacements between pairs of consecutive images using optical flow. Assume that one scene point $\mathbf{P} = (X, Y, Z)$ is projected onto image $\mathbf{C}_t(x, y)$ at time t and onto $\mathbf{C}_{t+1}(x + u_d, y + v_d)$ at time $t + \delta t$. The vector $\mathbf{u}_d = (u_d, v_d)^T$ is the *image displacement* of the projected scene point $\mathbf{P} = (X, Y, Z)$, and it will be estimated by calculating *optical flow* $\mathbf{u}_e = (u_e, v_e)^T$ using the given colour value distributions in $\mathbf{C}_t$ and $\mathbf{C}_{t+1}$, and possibly further image data of the captured sequence. The ideal solution would be $\mathbf{u}_d = \mathbf{u}_e$.

2 Review

The literature reporting on optical flow from colour images is quite sparse, despite the fact that multispectral approaches have been suggested already for about twenty years. Markandey and Flinchbaugh [6] have proposed a multispectral approach for optical flow computation. Their two-sensor proposal is based on solving a system of two linear equations having both optical flow components as unknowns. The equations are the standard optical flow constraints

$$\frac{\partial E}{\partial x}u + \frac{\partial E}{\partial y}v + \frac{\partial E}{\partial t} = 0, \tag{1}$$

where E is one image channel, formulated for each channel separately. Their experiments use colour TV camera data and a combination of infrared and visible imagery. In each experiment the data was smoothed by convolution with a Gaussian of $\sigma = 2.0$ before calculation of the derivatives and the calculated vector fields were post-processed using 10 iterations of median filtering, followed by 100 iterations of the vector field smoothing technique described by Schunck [8]. Their experiments have shown that "*vector field smoothing played a significant role in the computation of optical flow*" for two-channel imagery. The condition number of the integration matrix for the system of linear equations is used for estimating numerical stability.

Golland and Bruckstein [1] follow the same algebraic method. They compare a straightforward 3-channel approach using RGB data with two 2-channel methods, the first based on normalized RGB values and the second based on a special hue-saturation definition. The standard optical flow constraint may be applied to each one of the RGB quantities, providing an overdetermined system of linear equations:

$$\frac{\partial R}{\partial x}u + \frac{\partial R}{\partial y}v + \frac{\partial R}{\partial t} = 0\,, \tag{2}$$

$$\frac{\partial G}{\partial x}u + \frac{\partial G}{\partial y}v + \frac{\partial G}{\partial t} = 0\,, \tag{3}$$

$$\frac{\partial B}{\partial x}u + \frac{\partial B}{\partial y}v + \frac{\partial B}{\partial t} = 0\,. \tag{4}$$

The solution for the optical flow may be found by using a pseudo-inverse computation:

$$\mathbf{u} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} \tag{5}$$

where:

$$\mathbf{A} = \begin{pmatrix} R_x & R_y \\ G_x & G_y \\ B_x & B_y \end{pmatrix}\,, \quad \mathbf{b} = \begin{pmatrix} -R_t \\ -G_t \\ -B_t \end{pmatrix} \tag{6}$$

The matrix $\mathbf{A}$ must be non-singular. The smallest eigenvalue of $\mathbf{A}^T\mathbf{A}$ or the condition number of $\mathbf{A}^T\mathbf{A}$ can be used to measure numerical stability, i.e. if the smallest eigenvalue is below a threshold or the condition number is above a threshold, then we set $\mathbf{u}$

*Department of Computer Science, University of Western Ontario, London, Ontario, Canada, N6A 5B7, barron@csd.uwo.ca

†CITR, University of Auckland, Tamaki Campus, Building 731 Auckland, New Zealand, r.klette@auckland.ac.nz

Figure 1. (Left) The original video 720×576 colour image and (Right) the even interpolated colour image. The odd interpolated colour image looks about the same but is translated by about (-3,0) with respect to the even image.

to be undefined at this image location. We implement Golland and Bruckstein's method with weights for each channel in a standard least squares framework. If we assume a constant flow vector in a $(2n+1)\times(2n+1)$ then $\mathbf{A}^T\mathbf{A}$ becomes the least squares Lucas and Kanade optical flow method [4, 9]. We use $n = 1$ for the Lucas and Kanade results reported here. If all weights were 0 except for the Y channel (with value 1) then we would have the standard Lucas and Kanade calculation [4, 9]. If the R, G and B weights are 1 with all other weights 0 we have colour Lucas and Kanade. Note that Y=0.299R+0.587G+0.114B, so if we changed the non-zero weights to these values we would obtain exactly the same results as for the Y images. Lastly, if $n = 1$, we have the standard Golland and Bruckstein method.

We recast the standard Horn and Schunck regularization [2, 9, 10] as the minimisation of:

$$\sum_i^n w_i \left[E_{xi}u + E_{yi}v + E_{ti}\right]^2 + \alpha^2(u_x^2 + u_y^2 + v_x^2 + v_y^2).$$

The weights w_i indicate the attached importance of each channel of image information, e.g., $w_i = 0$ means channel i has absolutely no importance. If all w_i are 0 except the one corresponding to the Y channel (with value 1) and $\beta = 0$ then this minimisation is exactly Horn and Schunck on a grayvalue image [2, 9].

Golland and Bruckstein [1] have reported that the RGB-method "*provides sufficient quality for velocity estimates when the object undergoes translations in a plane parallel to the image plane, but if more complex kinds of motion are involved, this method, like any other using the brightness conservation assumption, produces estimates with significant errors.*" Furthermore, their methods "*provide much more precise information about the object motion.*" However, these statements have been deducted from a very small number of experiments, and more detailed studies might be worthwhile. In general, optical flow computed for a single image point has proved unsuccessful, local neighbourhoods under some parametric constraint (such as constant velocity) or global regularization of a functional is necessary to obtain useful flow.

Ohta [7] derives optical flow equations in a least squares framework on channel derivatives in colour images. Let E_{xi} and E_{yi} denote the x and y derivatives in colour channel i. For example E_{x5} could be the x derivatives in the hue image. Ohta assumes that there are $n \geq 2$ optical flow constraints for the different channels. For $n = 2$, it follows

$$u = \frac{E_{y1}E_{t2} - E_{t1}E_{y2}}{E_{x1}E_{y2} - E_{y1}E_{x2}} \quad \text{and} \quad v = \frac{E_{t1}E_{x2} - E_{x1}E_{t2}}{E_{x1}E_{y2} - E_{y1}E_{x2}}.$$

Both equations are solvable if $E_{x1}E_{y2} - E_{y1}E_{x2} \neq 0$. Letting the error in the i-th equation be e_i, taking the sum of all squared errors for all n equations, calculating its partial derivatives with respect to u and v, and letting these formulae be equal to zero, we obtain [7]

$$u = \frac{\sum E_{xi}E_{yi} \sum E_{yi}E_{ti} - \sum E_{yi}^2 \sum E_{xi}E_{ti}}{\sum E_{xi}^2 \sum E_{yi}^2 - \left(\sum E_{xi}E_{yi}\right)^2} \quad (7)$$

$$v = \frac{\sum E_{xi}E_{yi} \sum E_{xi}E_{ti} - \sum E_{xi}^2 \sum E_{yi}E_{ti}}{\sum E_{xi}^2 \sum E_{yi}^2 - \left(\sum E_{xi}E_{yi}\right)^2}, \quad (8)$$

Figure 2. (Left) The Horn and Schunck flow for the Y image with $\alpha = 3$ and 100 iterations and post-thresholding of $||\nabla I||_2 > 5.0$ and (Right) the Lucas and Kanade for for the Y image with smallest eigenvalue thresholding of $\tau > 5.0$.

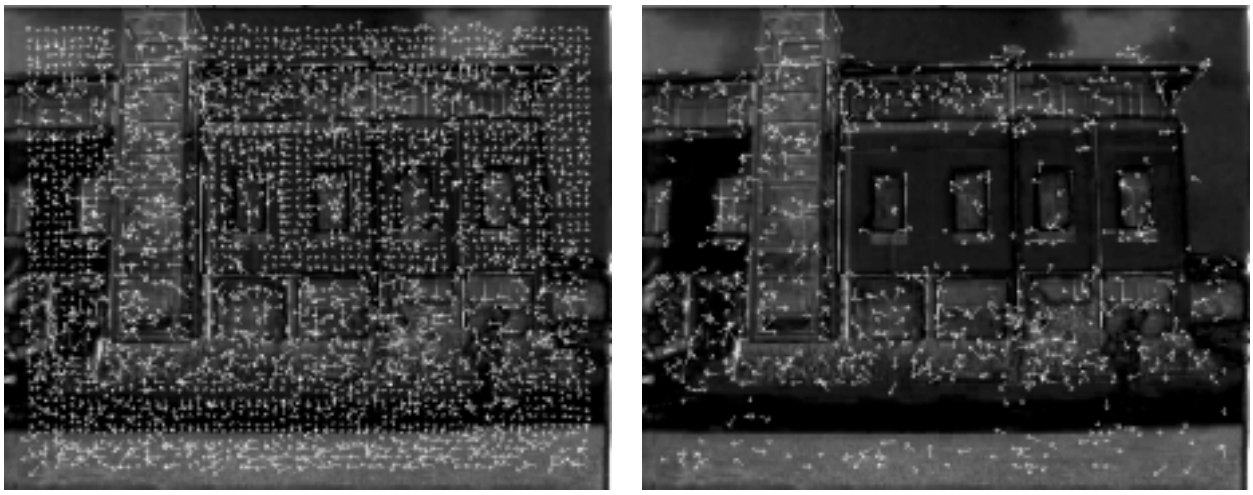

Figure 3. (Left) The Horn and Schunck flow for the saturation image with $\alpha = 3$ and 100 iterations and post-thresholding of $||\nabla I||_2 > 5.0$ and (Right) the Lucas and Kanade for for the Y image with smallest eigenvalue thresholding of $\tau > 5.0$.

where summations are from 1 to n. This approach and the generalized inverse method [1, 6] assume constant weights for all equations. Our implementation allows us to set arbitrary weights for each of the nine RGB, HSI and YIQ colour channels. Note that if all weights are zero except for the Y channel weight being 1, then in a $(2n+1) \times (2n+1)$ we again have a standard Lucas and Kanade [4, 9] least squares calculation.

3 New Approach

In a new approach, we recast the standard Horn and Schunck regularization as the minimisation of:

$$\sum_{i}^{n} w_i \left[E_{xi}u + E_{yi}v + E_{ti}\right]^2 + \beta^2 f_d(u,v). \qquad (9)$$

Again, as for the Horn and Schunck regularization, the weights w_i indicate the attached importance of each channel of image information and can be set so as to include/exclude any colour channel. $f_d(u,v)$ is a directional constraint that depends on one knowing whether

Figure 4. (Left) Standard Golland and Bruckstein flow and (Right) least squares Golland and Bruckstein flow, both with post-thresholding of $||\nabla I||_2 > 5.0$.

Figure 5. (Left) Colour Ohta flow and (Right) saturation/Y Ohta flow, both with post-thresholding of $||\nabla I||_2 > 5.0$.

the camera is panning or zooming. Because the 3D translational direction is known, the direction of flow can be computed everywhere using the standard optical flow equations [3] and this can be used to constrain the minimisation. The directional constraint can be expressed as:

$$f_d(u,v) = (\mathbf{u} \cdot \mathbf{u}_d)^2 - ||\mathbf{u}||_2^2 ||\mathbf{u}_d||_2^2 = 0, \qquad (10)$$

where, because due to sensor motion is only translation and can be written as [3] $\mathbf{u}_d$ can be written as:

$$\mathbf{u}_d = \begin{pmatrix} u_d \\ v_d \end{pmatrix} = \mu \cdot ||(x,y,f)||_2 \cdot \mathbf{A}_{(x,y)} \hat{\mathbf{m}}, \qquad (11)$$

where (x,y) is an image location, μ is the relative depth, i.e. $\mu = \frac{||\mathbf{m}||}{||(X,Y,Z)||_2}$, $||\mathbf{m}||$ is the (unknown) sensor translation magnitude and (X,Y,Z) are the 3D coordinates of some 3D moving point projected onto (x,y). Note that since we do not know μ, $\mathbf{u}_d$ is only in the direction of the true optical flow vector and scaled by the μ value used. We can write $A_{(x,y)}$, a 2×3 matrix, as:

$$A_{(x,y)} = \begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}, \qquad (12)$$

where f is the effective focal length (determined empirically to be about 40mm).

Figure 6. (Left) Colour Horn and Schunck flow with $\alpha = 3$ and (Right) colour Barron and Klette flow with $\beta = 1.0$, both with post-thresholding of $||\nabla I||_2 > 5.0$ and 100 iterations.

Figure 7. (Left) Barron and Klette flow for the Y and saturation images and (Right) Barron and Klette flow for the Y image only, both with post-thresholding of $||\nabla I||_2 > 5.0$ and 100 iterations and $\beta = 1.0$.

4 Experiments

This paper reports on an experimental evaluation of these 4 approaches [Golland and Bruckstein, Ohta, and Barron and Klette with Horn and Schunck and Lucas and Kanade serving as a standards] for optical flow in colour video images using various values for the w_i's and $\alpha = 3$ and $\beta = 1$. We show that sometimes optical flow results improve for non-zero weights (multiple colour image channels). Our α and β values produced the best results. Our investigation consists of a a qualitative evaluation of these algorithms on real colour video image sequences.

Figure 1a shows an original interlaced video image. There is motion of about 3 pixels to the right between the even and odd lines of the image. We generate 2 images from this single image. Figure 1b show the image generated by linearly interpolating the even lines, another second image is generated by linearly interpolating the odd lines. Note the the "jaggedness" of parts of the original image, for example the window sills, the speed bump or the roof eves, are smoothed out. The images were pre-filtered with a 2D Gaussian with standard deviation $\sigma = 2$ and Horn and Schunck 2-frame differentiation was used.

Figures 2a and 2b show the Horn and Schunck and Lucas and Kanade flows for the Y images. We see

that Horn and Schunck is considerably more dense than Lucas and Kanade but with some obvious errors. Both flows are right to left in the correct accepted direction. When we began this study we searched for a different colour image, other than the R, G, B, I (in HSI) or Y images, to aid in our optical flow calculation. H, I (in YIQ) and Q did not indicate any structure that could be tracked with motion, leaving only S images, which appeared to have trackable texture in areas of uniform colour. Figure 3a and 3b shows the flows for Horn and Schunck and Lucas and Kanade, which are quite poor compared to those in Figures 2a and 2b. Generally, the flow was right to left, with many erroneous velocities.

Figures 4a and 4b show the flows computed using Golland and Bruckstein's method for neighbourhood sizes $n = 0$ (standard) and $n = 1$ (least squares). Only the RGB images were used and all were weighted with 1. We see that the least squares solution is better than the standard solution; obviously larger least squares integration neighbourhoods improved the solution.

Figures 5a and 5b show the flows for Ohta's method on colour (RGB) images and on S/Y images (flow shown on the S image). Both flows are poor with obvious outliers. Figure 6a shows the flow for the RGB images (not as good as the flow for the Y image shown earlier) while Figure 6b show the Barron and Klette flow using the RGB images. Both are dense with generally correct flow but obvious outliers. Figures 7a and 7b shows the flows for the Y/S and Y alone images for Barron and Klette's method. From these two results, it is obvious that removing the S image from consideration significantly improved the flow.

5 Concluding Remarks

Surprisingly, colour seem to add little to the computation of optical flow for 2 frames. The use of the Y image alone (the standard black and white TV image) seemed to be the best. Our directional constraint proved quite useful in removing direction outliers. So far we only made a qualitative analysis of the computed flow fields. Future work includes an evaluation of computed optical flow based on a reliability analysis of the involved calculations [7], the numerical stability [1, 6], and a quantitative analysis of the flow for synthetic images (i.e. a real image displaced by known amount) [9, 10]. We will also analyze more image sequences, including zooming sequences. To achieve more accurate differentiation (and, hence, more accurate flow) we need to use more images [9]. The panning and zooming motions of the camera satisfy the inherent smoothness assumptions needed for this. We anticipate using colour flow measurements in a calculation of normalized dominant plane gradient [5] from the flow in a robust estimation framework. Such gradients could be used to stabilise video images and to describe secondary structures.

References

[1] P. Golland and A. M. Bruckstein. Motion from colour. Center for Intelligent Systems TR #9513, Computer Science Department, Technion, Haifa, August 1997.

[2] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, **17**, 1981, pp. 185–204.

[3] H.-C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. R. Soc. Lond.*, **B 208**, 1980, pp. 385–397.

[4] B. D. Lucas and T. Kanade. An iterative image-registration technique with an application to stereo vision. DARPA *Image Understanding Workshop*, 1981, pp. 121–130 (see also *IJCAI'81*, pp. 674–679).

[5] K. Kawamoto and R. Klette. Dominant plane estimation. CITR-TR-88, CITR Tamaki, The University of Auckland, Auckland, April 2001.

[6] V. Markandey, B. E. Flinchbaugh. Multispectral constraints for optical flow computation. Proc. 3rd *ICCV*, Osaka, Japan, December 1990, pp. 38–41.

[7] N. Ohta. Optical flow detection by colour images. *NEC Research and Development*, (97), 1990, pp. 78–84 (preliminary version: N. Ohta. Optical flow detection by colour images. *IEEE Int. Conf. on Image Processing*, Pan Pacific Singapore, pages 801 –805, 1989.)

[8] B. G. Schunck. Image flow: fundamentals and algorithms. In *Motion Understanding - Robot and Human Vision* (W.N. Martin and J.K. Aggarwal, eds.), Kluwer, Amsterdam 1988.

[9] J. L. Barron, D. J. Fleet and S. S. Beauchemin. Performance of optical flow techniques. *Int. Journal of Computer Vision*, **12**, 1994, pp. 43–77.

[10] P. Handschack and R. Klette. Quantitative comparisons of differential methods for measuring of image velocity. In Proc. *Aspects of Visual Form Processing*, Capri, Italy, World Scientific, 1994, pp. 241 - 250.