

# Optic Flow Experiments for Slow Motion

Kun Ju, Simon Hermann, and Reinhard Klette

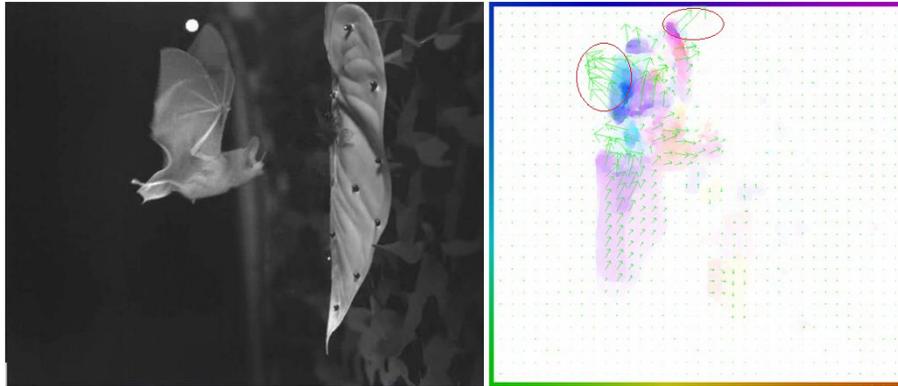
The *.enpeda.* Project, The University of Auckland  
Tamaki Innovation Campus, Auckland, New Zealand  
kju002@aucklanduni.ac.nz

**Abstract.** Prediction error analysis is a known method for evaluating the performance of optic flow algorithms. This paper applies this method to long image sequences with a special focus on opportunities when using metronome sequences. Two top-performing algorithms are discussed, known as TV- $L_1$  and BBPW.

**Keywords:** Optic flow, performance evaluation, prediction error, slow motion, TV  $L_1$ , BBPW, metronome sequence

## 1 Introduction

Optic flow is the instantaneous speed of a moving object's pixel, measured by visible displacements in the imaging plane. Optic flow research is about finding relationships between changes in image intensities in time and object movements in a 3-dimensional (3D) environment. In general, there are four events which can contribute to observable optical flow: camera movement, movement of objects in the



**Fig. 1.** *Left:* a flying bat (data courtesy by Bob Fisher, Edinburgh) recorded at 1000 Hz. *Right:* optic flow results when applying TV- $L_1$ . The coloured frame shows the used colour key; intensity corresponds to length of motion vectors. Both ellipses highlight local regions where wings move intensively (and actually into opposite directions).

scene, movements of both, or temporal changes in intensity due to illumination changes or simply due to noise. Optical flow methods are divided in [Ye 2002] into categories of matching-based, frequency domain-based, and gradient-based techniques. In this classification scheme, TV- $L_1$  [Papenberg et al. 2006] and BBPW [Brox et al. 2004] are gradient based, and both lead to good results on [OFWMC]. See Fig. 1 for the used colour key for visualizing results.

The original TV- $L^1$  algorithm applies an approximation of the  $L_1$ -norm in a total variation approach using an error function which combines a data and a smoothness term. Compared to the  $L_2$ -norm, the  $L_1$ -norm is more robust against outliers and preserves discontinuities. In this paper we use the improved TV- $L^1$  algorithm from [Wedel et al. 2008] which incorporates a median filter into the optimization scheme to increase convergence speed and to improve the consistency of the flow. It also proposes to work on the textural part of the input images to improve robustness against illumination changes.

The BBPW algorithm follows the Horn-Schunck algorithm but also minimizes an  $L_1$  data term in order to improve flow results, especially with regards to large displacements by using image pyramids with arbitrary scaling factors. It applies image warping and a robust penalizer for allowing also flow discontinuities. The implementation as used in this paper was done for [Vaudrey et al. 2008] and follows [Brox et al. 2004]. We use 50 levels in the pyramid with a scaling factor of 0.95.

The paper is structured as follows: Section 2 recalls prediction analysis and specifies our evaluation approach. Section 3 introduces the used sequences. Section 4 informs about experimental results. Section 5 concludes.

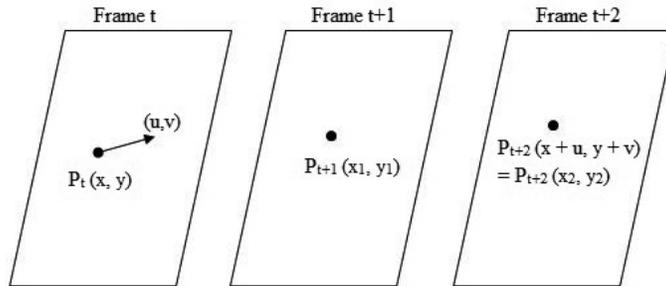
## 2 Prediction Error Analysis

Performance evaluation in computer vision is often designed under the assumption that ‘ground truth’ is available for measuring deviations from the ‘golden standard’.

Prediction errors [Szeliski 1999] can be utilized for analysing the performance of optic flow algorithms on real-world sequences, as illustrated there for very short sequences. Flow is calculated between images at times  $t$  and  $t + 2$ , and a virtual image calculated ‘half-way’ for time  $t + 1$ ; this is then compared with the actually recorded image at time  $t + 1$ . See Fig. 2.

The idea also applies to stereo vision: depth is calculated from cameras  $L$  and  $R$ , and, using camera calibration data, the intensity of camera  $L$  is then mapped into the view of a third camera  $T$  using the calculated depth or disparity values. This stereo evaluation approach was followed at some depth in [Morales and Klette 2009] using long trinocular image sequences recorded from a mobile platform.

However, it is very difficult to obtain long video sequences where recorded motion supports the application of prediction error analysis, due to relatively slow cameras (or relatively fast motion in the recorded scene). For example, 30 Hz is still an insufficient recording speed of cameras for common motions in the



**Fig. 2.** Illustration for prediction error analysis for evaluating optic flow methods. Point  $(x_1, y_1)$  is half-way between  $(x, y)$  and  $(x_2, y_2) = (x, y) + (u, v)$ , and the calculated intensity is the mean of  $P_t(x, y)$  and  $P_{t+2}(x_2, y_2)$ .

real world. The website [EISATS] provides a few long synthetic sequences with motion ground truth, but even for those synthetic sequences, the frame rate and the rendered object motions do not support prediction error analysis.

Consequently, we have recorded sequences showing very slow motion in the scene, and we also use (from other sources) sequences recorded at very high speed. Because mean brightness and brightness variations in individual images in a sequence vary due to some conditions (e.g. exposure or lighting), we can either use some kind of input data normalization (e.g. mapping images onto a constant mean and a constant standard deviation [Klette and Zamperoni 1996], applying a ‘simple’ edge operator [Guan and Klette 2008], or mapping images onto their residuals with respect to some blurring [Vaudrey and Klette 2009]), or we can compare the virtual and recorded image with a measure that does not take such brightness variations into account.

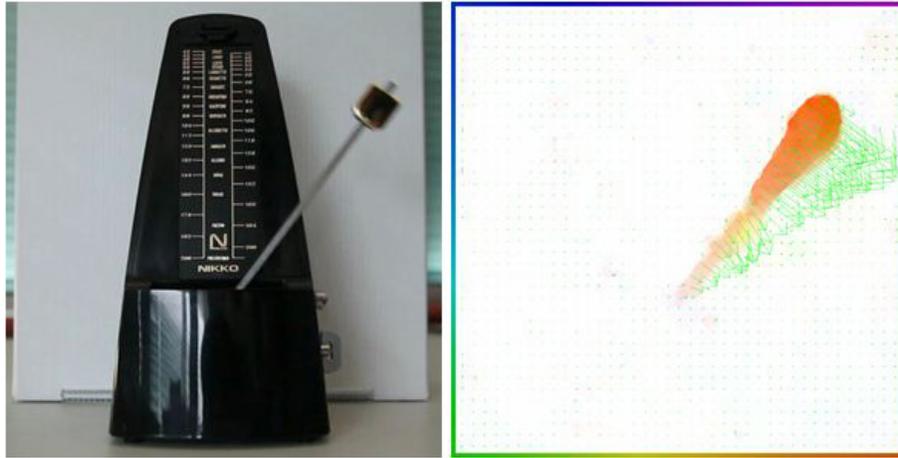
For this paper we decided for the latter approach. We report results for the original images and use the *normalized cross-correlation* (NCC) as the measure for detecting differences between virtual and recorded image:

$$N(t) = \frac{1}{|\Omega_{t+1}|} \sum_{(x,y) \in \Omega_{t+1}} \frac{[P_{t+1}(x, y) - \mu_{t+1}][P_{t+1}^v(x, y) - \mu_{t+1}^v]}{\sigma_{t+1}\sigma_{t+1}^v} \quad (1)$$

where  $\mu_{t+1}$  and  $\mu_{t+1}^v$  denote the means,  $\sigma_{t+1}$  and  $\sigma_{t+1}^v$  the standard deviations of recorded image  $P_{t+1}$  and predicted (or *virtual*) image  $P_{t+1}^v$ , and  $\Omega_{t+1}$  the set of all pixel positions used for frame  $P_{t+1}$  (i.e. all positions  $(x_1, y_1)$  obtained in the way as sketched in Fig. 2), and also contained in a generated mask (see beginning of Section 4). The NCC measure has been used, for example, for motion analysis in [Islam et al. 2009], and for stereo analysis in [Morales and Klette 2009].

### 3 Slow Motion Sequences

Image data for a flying bat was recorded (at Edinburgh university) at 1000 Hz. For the test discussed further below, 800 consecutive frames (i.e. less than one



**Fig. 3.** *Left:* Frame 2 of the first metronome sequence. *Right:* TV- $L_1$ -result when comparing Frame 2 with Frame 3.

second) have been selected for applying optical flow algorithms at frames  $P_t$  when  $t$  is even, thus generating 400 NCC values. See Fig. 1 for an input frame and a result using TV- $L_1$ .

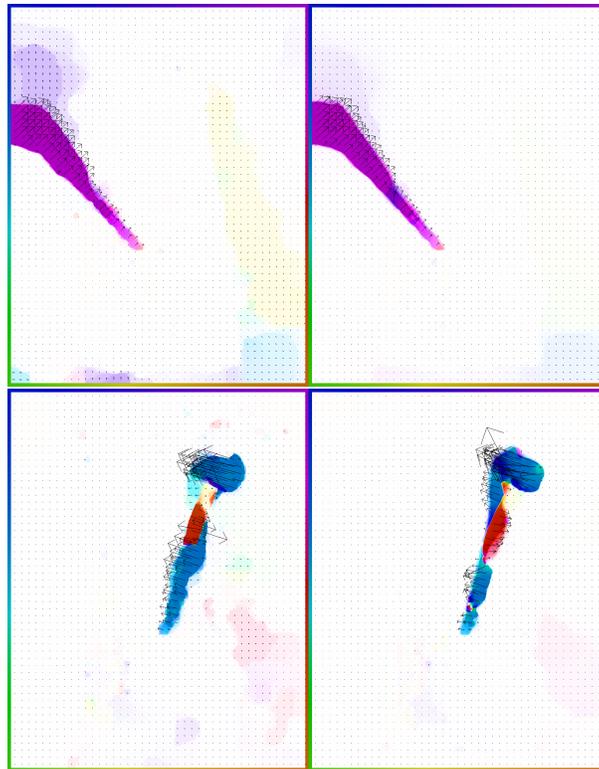


**Fig. 4.** Mask generated for Frame 57 of the used metronome sequence, prior to the application of the NCC measure. Note that the white region on the right and along the bottom is due to ‘incorrectly’ detected motion.

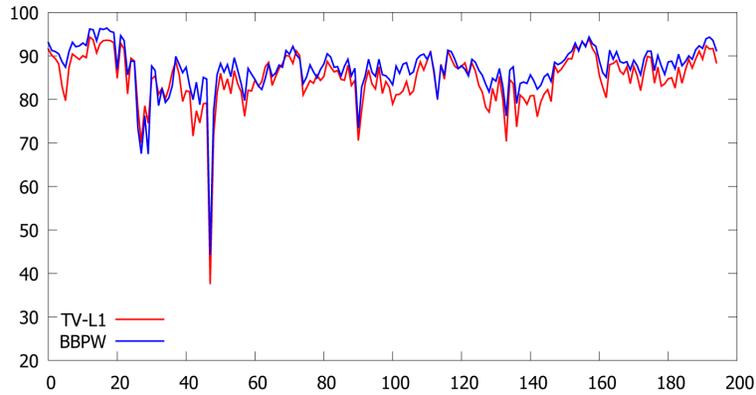
A metronome is known for its regular movements. It can be adjusted to a speed of 40 to 210 beats per minute. We recorded image sequences for 40 beats per minute because of our interest in slow motion. Metronome sequences were recorded with a Canon EOS 5D Mark II which is a high-performance digital SLR camera with a full-frame CMOS sensor with 21.10 effective megapixels. Videos were recorded in full  $1920 \times 1080$  pixel HD mode at 30 Hz. Videos were recorded for different lighting (e.g. lighting was changed by turning on and off regularly the office ceiling lights). Below we illustrate results for a sequence of 200 frames (i.e. 6.6 s) recorded for constant lighting, and also for a sequence with 800 frames where lighting changed regularly. See Fig. 3 for an input frame and a result using TV- $L_1$ .

## 4 Experiments and Discussion

Colour-coded optic flow for both TV- $L_1$  and BBPW is shown in Fig. 5. For prediction error analysis, we calculate the flow from  $t$  to  $t + 2$ , scale the flow



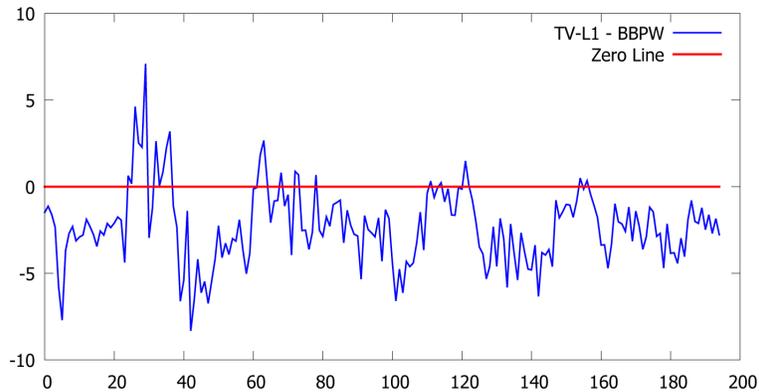
**Fig. 5.** Results for TV- $L_1$  (*left*) and BBPW (*right*) for Frame 57 (*top*) and Frame 111(*bottom*).



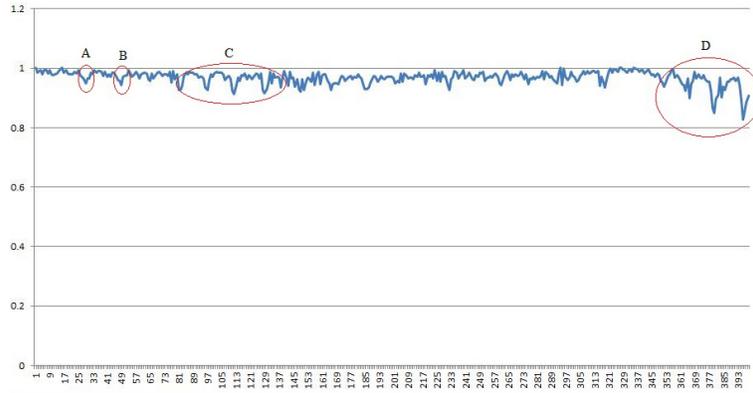
**Fig. 6.** NCC graphs for TV- $L_1$  (red line) and BBPW (blue line) in the constrained case (i.e. using a mask) for a metronome sequence with constant lighting. There is a high correlation between both curves, and BBPW appears to be superior. See Fig. 7 for a better illustration of this superiority.

vectors by 0.5 for generating a virtual image for time  $t + 1$  and compare with the original frame at  $t + 1$  using the NCC measure.

We apply two different selection criteria for the set  $\Omega_{t+1}$  in Equation (1). In the *unconstrained case* we use all the pixels that are identified by endpoints of halfway flow vectors in the virtual image. In the *constrained case* we calculate the NCC values only within regions of motion. Therefore we generate a mask as follows. At every pixel where the magnitude  $M = \sqrt{u^2 + v^2}$  of the flow extends 0.5 pixel, we validate the mask for a squared neighbourhood which has the size



**Fig. 7.** Differences of NCC values: those for TV- $L_1$  minus those for BBPW in the constrained case. The figure shows that BBPW's NCC values are in majority above those of TV- $L_1$ .

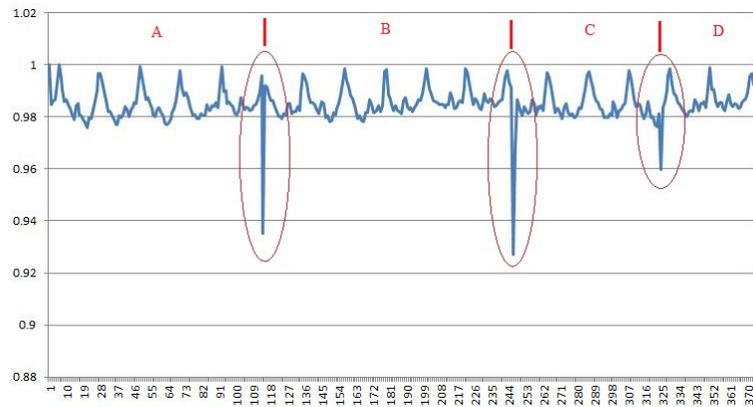


**Fig. 8.** NCC values for a flying bat sequence of 800 frames. Unconstrained case and use of TV- $L_1$ .

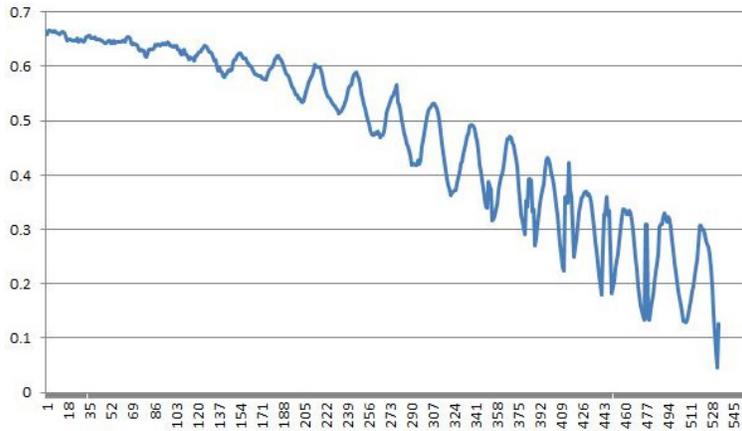
$M \times M$  (i.e. at least  $1 \times 1$ ). An example of a generated mask image is shown in Fig. 4.

At first we state a general observation that both techniques performed well to the expectations, as raised by their standing on [OFWMC], with a slight superiority of BBPW over TV- $L_1$ . See Figs. 6 and 7.

As a second result, the analysis of video data of a flying bat shows that optic flow is potentially a way for quantifying complex movements if the recording speed is sufficiently high. In Fig. 1, right, see, for example, the two ellipses highlighting particular motion patterns. When the bat is trying to generate a



**Fig. 9.** Unconstrained case and TV- $L_1$  when regularly changing the lighting for the metronome. NCC values decrease exactly at those frames where lighting changes and remain at the same level, no matter whether to office was darker or brighter for a given period A, B, C or D.



**Fig. 10.** Mean (over a whole metronome sequence) NCC values for the constrained case (i.e. using a mask) where the  $x$ -axis shows now the distance from the rotation centre of the pendulum.

lift, the wings are stretching rapidly; therefore the motion vectors at the edge of the wings are bigger than at other wing pixels.

In Fig. 8 we have encircled four parts with a particular ‘NCC failure pattern’. Such cases can be discussed as particularly ‘difficult motion patterns’ of the flying bat, such as no motion between two frames (but the prediction takes two frames where motion occurred), acceleration in movement, or changes in direction of movement.

Similar observations can be made for the metronome when the pendulum is close to the turning point: in such cases there may be a mismatch between the implicitly used motion model of prediction error analysis and the actual motion. This is our third statement.

Next, changes in lighting affect the performance of both optic flow methods, as already studied in [Vaudrey and Klette 2009] for other image data. Figure 9 shows how the NCC values decrease exactly at frames when the light was switched on or off.

Finally, the metronome sequence is also valuable for understanding the abilities of optic flow methods for estimating motion vectors of increasing lengths. See Fig. 10. These mean NCC values have been generated for pendulum positions including two turning points. Not surprising, the values are decreasing with distance. However, there are some periodic patterns in this histogram which require further analysis to be understood properly.

## 5 Conclusion

This paper introduces metronome image sequences for testing optic flow algorithms. Metronome sequences will be made publicly available at [EISATS]. The paper also discusses the use of image data recorded at very high speed. The experiments show how these data can be used for designing experiments aiming at particular studies.

**Acknowledgement.** The authors thank Andreas Wedel for providing his implementation of TV-L<sub>1</sub> as reported in [Wedel et al. 2008], and James Milburn for his sources for BBPW as used for [Vaudrey et al. 2008]. The bat sequence was provided by Bob Fisher (Edinburgh).

## References

- [Brox et al. 2004] Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In *ECCV*, LNCS 3024, pages 25–36.
- [EISATS] Enpeda image sequence analysis test site (EISATS). Retrieved 16 March 2011 at <http://www.mi.auckland.ac.nz/EISATS>.
- [Guan and Klette 2008] Guan, S. and Klette, R. (2008) Belief-propagation on edge images for stereo analysis of image sequences. In *Robot Vision*, LNCS 4931, pages 291–302.
- [Islam et al. 2009] Islam, M. Z., Oh, C.-M., and Lee, C.-W. (2009) Video based moving object tracking by particle filter. *Int. J. Signal Processing Image Processing Pattern Analysis*, **2**:119–132.
- [Klette and Zamperoni 1996] Klette, R. and Zamperoni, P. (1996). *Handbook of Image Processing Operators*. Wiley, Chichester.
- [Morales and Klette 2009] Morales, S., and Klette, R. (2009) A third eye for performance evaluation in stereo sequence analysis. In *CAIP*, LNCS 5702, pages 1078–1086.
- [OFWMC] Optic Flow Website at Middlebury College. Retrieved 16 March 2011 at <http://vision.middlebury.edu/flow/>.
- [Papenberg et al. 2006] Papenberg, N., Bruhn, A., Brox, T., Didas, S., and Weickert, J. (2006). Highly accurate optic flow computation with theoretically justified warping. *Int. J. Computer Vision*, **67**:141158.
- [Szeliski 1999] Szeliski, R. (1999). Prediction error as a quality metric for motion and stereo. In *ICCV*, pages 781788.
- [Vaudrey and Klette 2009] Vaudrey, T. and Klette, R. (2009). Residual images remove illumination artifacts for correspondence algorithms! In *DAGM*, LNCS 5748, pages 472–481.
- [Vaudrey et al. 2008] Vaudrey, T., Rabe, C., Klette, R., and Milburn, J. (2008). Differences between stereo and motion behaviour on synthetic and real-world stereo sequences. In *IVCNZ*, IEEE identifier 10.1109/IVCNZ.2008.4762133.
- [Wedel et al. 2008] Wedel, A., Pock, T., Zach, C., Bischof, H., and Cremers, D. (2008). An improved algorithm for TV-L1 optical flow. In *Statistical Geometrical Approaches Visual Motion Analysis*, LNCS 5604, pages 23–45.
- [Ye 2002] Ye, M. (2002) Robust visual motion analysis: Piecewise-smooth optical flow and motion-based detection and tracking. PhD thesis, Univ. Washington.