BENEFITS OF TEMPORAL OVERSAMPLING IN OPTICAL FLOW ESTIMATION

*SukHwan Lim** — Imaging Systems Lab, Hewlett-Packard Labs, Palo Alto, CA 94304

John Apostolopoulos — Mobile & Media Systems Lab, Hewlett-Packard Labs, Palo Alto, CA 94304

Abbas El Gamal — Information Systems Lab, Stanford University, Stanford, CA 94305

ABSTRACT

Recently it has been shown that optical flow estimation (OFE) accuracy can benefit from temporal oversampling especially for large displacements between frames. In this paper, we show that temporal oversampling also benefits OFE in the case of complex scenes with small displacements but high spatial bandwidth. Using synthetic test sequences and a high-speed real video sequence, it is shown that temporal oversampling improves the performance of OFE by reducing motion aliasing not only for areas with large displacements but also for areas with small displacements but with high spatial frequencies. We also demonstrate that the minimum frame rate necessary to achieve good OFE performance for the tested sequences is largely determined by the minimum frame rate necessary to prevent motion aliasing.

1. INTRODUCTION

Accurate optical flow estimation (OFE) is very important for many video processing and computer vision applications [1, 2]. It is well known that motion aliasing or temporal aliasing adversely affects the accuracy of OFE and many researchers pointed out that large systematic errors arise when displacements are large [3, 4, 5, 6]. Multi-resolution algorithms applied spatially help overcome these problems as the spatial subsampling in effect reduces the motion between frames by the subsampling factor [4, 7]. Although temporal aliasing can be caused by large displacements, it can also be caused by high spatial frequencies with low-to-moderate displacements. In this case, the spatial low-pass filtering used as part of the coarse-to-fine estimation can partially overcome the aliasing in the high spatial frequencies [5, 6, 7]. Note that these approaches adapt the spatial sampling, but do not modify the temporal sampling: the frame rate for processing is given by the frame rate of the desired output OFE which is also equal to the frame rate of the video acquisition. Recently, high-speed cameras are becoming economical and temporally oversampled OFE estimation has become practical and may provide a cleaner and more accurate approach for overcoming temporal aliasing. For example, it has been shown that several still image and video processing applications benefit from high frame rate video capture by using the additional information about illumination and motion obtained from temporally *oversampling* the video to enhance image quality or improve the performance of video processing applications [8, 9, 10].

*The work was done while the author was at Stanford University. It was supported partly under the DARPA Grant No. N66001-02-1-8940 and under Programmable Digital Camera Program by Agilent, Canon, HP, Interval Research and Kodak. Special thanks to Ali Ercan for assisting to capture high frame rate video sequences.

In our previous work [9, 11], we proposed an OFE algorithm that computes the OFE for a temporally oversampled video sequence and uses this to produce motion field estimates at the desired standard frame rate (see Figure 1). The scene is first captured at a high frame rate that is OV times the standard frame rate. The Lucas-Kanade algorithm is used to obtain high accuracy optical flow estimates between consecutive pairs of high-speed frames. The next step is to integrate optical flow temporally without losing the accuracy gained by using the high frame rate sequences. To do this, the optical flow estimates between consecutive high-speed frames are accumulated along the motion trajectories and then refined to provide an improved estimate of the optical flow between each pair of standard-speed frames. The method was tested on synthetically generated video sequences and the results showed significant improvements in optical flow estimation (OFE) accuracy especially for large displacements. Thus, the use of temporal oversampling appears to be a natural approach to overcome the large motion between frames. This paper extends our previous work [9] and shows that temporal oversampling also helps the case for complex scenes with low velocities but high spatial bandwidth.

Additional benefits of using temporal oversampling for optical flow estimation (OFE) include: (i) the brightness constancy assumption [1, 2] made implicitly or explicitly in most OFE algorithms becomes more valid as frame rate increases, and (ii) the accuracy of temporal gradient estimation increases. From our experiments in Section 3 the most important benefit of using temporal oversampling is avoiding motion aliasing, which suggests that the temporal oversampling factor should be high enough to avoid motion aliasing in order to perform accurate optical flow estimation.

This paper is organized as follows. In Section 2 we review 3-D spatio-temporal sampling theory and compute the minimum cap-

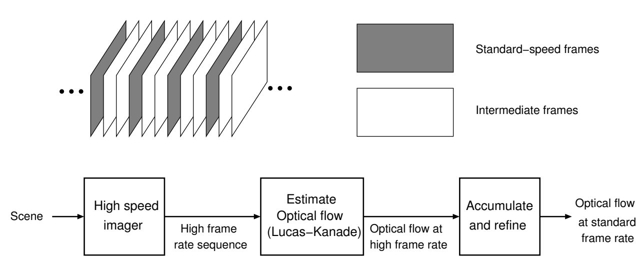


Fig. 1. Proposed method for estimating optical flow at the standard frame rate by processing a high frame rate sequence ($OV = 3$).

ture frame rate to avoid motion aliasing and accurately estimate the temporal gradient. In Section 3 we present optical flow estimation results obtained using sinusoidal test sequences, a synthetically generated natural image sequence and real video sequences. We confirm that temporal oversampling benefits OFE for complex scenes with low velocities but high spatial bandwidth and that the minimum frame rate needed to achieve high accuracy is largely influenced by the frame rate necessary to avoid motion aliasing.

2. TEMPORAL SAMPLING AND MOTION ALIASING

Motion aliasing can produce large errors even with the best optical flow estimators. Perhaps the most well known example is that of the wagon wheel in a Western movie (shot at 24 frames/s) which appears to be moving in the opposite direction from what physically makes sense given the wagon's motion. This section briefly reviews 3-D spatio-temporal sampling theory and shows that motion aliasing not only occurs for large displacements but also for high spatial frequency as well.

Let us consider the problem of how fast we should sample an original continuous video signal along the temporal dimension such that it can be perfectly recovered from its samples. Assume that an ideal low-pass filter with rectangular support in the 3-D frequency domain is used for reconstruction, although in certain ideal cases, a sub-Nyquist sampled signal can also be reconstructed by an ideal motion-compensated reconstruction filter assuming the replicated spectra do not overlap [2]. To recover the original continuous spatio-temporal video signal from its temporally sampled version, the temporal sampling frequency (or frame rate) f_s must be greater than $2B_t$ in order to avoid aliasing in the temporal direction. If we assume global motion with constant velocity v_x and v_y (in pixels per standard-speed frame) and spatially bandlimited image with B_x and B_y as the horizontal and vertical spatial bandwidths (in cycles per pixel), the minimum temporal sampling frequency $f_{\mathrm{s,Nyq}}$ to avoid motion aliasing is given by

$$f_{\mathrm{s,Nyq}} = 2B_t = 2B_x v_x + 2B_y v_y, \qquad (1)$$

where $f_{\mathrm{s,Nyq}}$ is in cycles per standard-speed frame. To clarify this unit, $f_{\mathrm{s,Nyq}} = 2$ cycles per standard-speed frame corresponds to 60 frames/s when the standard frame rate is 30 frames/s. Note that the temporal sampling frequency in cycles per standard-speed frame is the oversampling factor OV. Moreover, since OV is an integer in our framework to ensure that each standard-speed frame corresponds to a captured high-speed frame (see Figure 1), the minimum oversampling factor to avoid motion aliasing, OV_{theo}, is

$$OV_{\mathrm{theo}} = \lceil f_{\mathrm{s,Nyq}} \rceil = \lceil 2B_x v_x + 2B_y v_y \rceil.$$

To illustrate this relationship consider the simple case of a sequence with only global motion in the horizontal direction (i.e., with $v_y = 0$). Figure 2 plots $OV_{\mathrm{theo}} = \lceil 2B_x v_x \rceil$ versus horizontal velocity and spatial bandwidth for this case.

Although sampling just above the Nyquist rate is sufficient to avoid motion aliasing, it may not be enough for some OFE methods such as gradient-based methods. In gradient-based OFE methods, 2-tap filters are often used to calculate the temporal gradient in order to reduce the number of frame buffers, despite their poor performance for high frequencies. Gradient estimators using a small number of taps suffer from poor accuracy when high frequency content is present [12]. Specifically, a 2-tap temporal gradient estimator performs poorly for temporal frequencies $f_t > \frac{1}{3}$ [12].

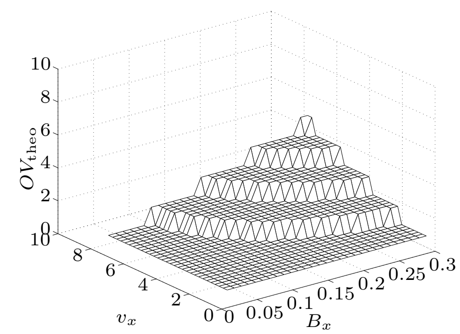


Fig. 2. Minimum OV to avoid motion aliasing, as a function of horizontal velocity v_x and horizontal spatial bandwidth B_x ($OV \geq 1$).

One solution is to sample at a rate 50% higher than that of the Nyquist temporal sampling rate, which ensures that $f_t \leq \frac{1}{3}$.

3. SIMULATION AND EXPERIMENTAL RESULTS

This section examines optical flow estimation results obtained using sinusoidal test sequences, a synthetically generated natural image sequence and a real video sequence captured from a high-speed camera.

While it is desirable to perform the analysis on natural video sequences acquired from the real-world, it is difficult to make quantitative assessments of the optical flow estimation accuracy because: (i) the true motion is generally not known in a natural sequence, (ii) it is difficult to understand the behavior of the proposed method with respect to local spatial frequency, since each local region in a natural sequence is likely to have different spatial frequency content and the Lucas-Kanade method estimates optical flow by performing spatially local operations, and (iii) typical figures of merit, such as average angular error and average magnitude error, are averaged out across the frame. To overcome these problems we use sinusoidal test sequences, which enable us to vary the spatial frequency and velocity in a controlled manner. This provides useful information on how to select the minimum high-speed frame rate for a natural scene.

We considered a family of 2-D sinusoidal sequences with equal horizontal and vertical frequencies $f_x = f_y$ moving only in the horizontal direction at speed v_x (i.e., $v_y = 0$). For each f_x and v_x, we generated a sequence with $OV = 1$ and performed optical flow estimation using the proposed method. We then incremented OV by 1 and repeated the simulation. We noticed that the average error drops rapidly beyond a certain value of OV and that it remained relatively constant for OVs higher than that value. Based on this observation we defined the minimum oversampling ratio OV_{exp} as the OV value at which the magnitude error drops below a certain threshold. In particular, we chose the threshold to be 0.1 pixels/frame. Once we found the minimum value of OV, we repeated the experiment for different spatial frequencies and velocities. The results are plotted in Figure 3.

Recall the discussion in Section 2 (including Figure 2) on the minimum oversampling factor as a function of spatial bandwidth and velocity needed to avoid motion aliasing. Note the similarity between the theoretical results in Figure 2 and their experimen-

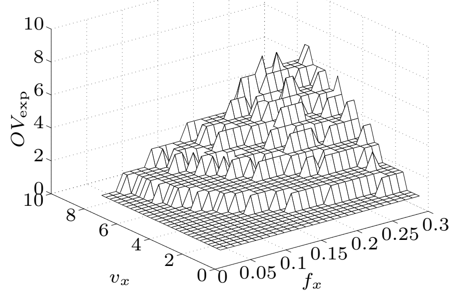


Fig. 3. Minimum OV as a function of horizontal velocity v_x and horizontal spatial frequency f_x.

tal counterpart in Figure 3. This is further illustrated by the plot of their difference and its histogram in Figure 4. This similarity shows that reduction in motion aliasing is one of the most important benefits of using high frame rate sequences. Note from the figure that oversampling is required, not only for the case with large displacements but also for high spatial frequency cases. The difference in Figure 4 can be further reduced by sampling at a higher rate than $\lceil f_{\mathrm{s,Nyq}} \rceil$ to better approximate brightness constancy and improve the estimation of temporal gradients. In our implementation, since we used a 2-tap temporal gradient estimator, we need to sample at least 50% faster than the Nyquist temporal sampling rate as discussed in Section 2. Choosing an OV curve that is 1.55 times the Nyquist rate (i.e., $\lceil 1.55 f_{\mathrm{s,Nyq}} \rceil$), in Figure 5 we plot the difference between the OV_{exp} curve in Figure 3 and the new OV curve. Note the reduction in the difference achieved by the increase in frame rate.

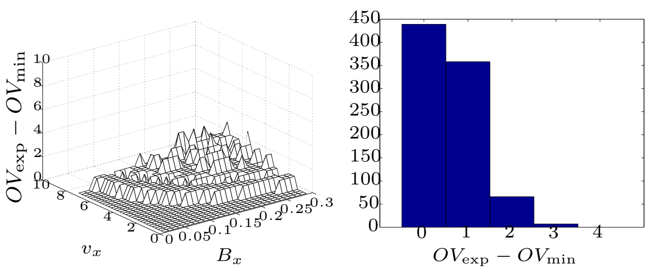


Fig. 4. Difference between empirical minimum OV for good optical flow estimation performance and OV corresponding to the Nyquist rate.

To investigate the effect of motion aliasing, we also estimated the energy in the image that leads to motion aliasing. A set of sequences with different capture frame rates but undergoing the same global motion with constant velocity of $v_x = v_y = 5$ pixels per standard-speed frame, was generated by warping the natural image shown in Figure 6 (a). The sequences were generated using a realistic image sensor model that includes motion blur, temporal noise and fixed pattern noise [9]. The temporal bandwidth of the sequences can be estimated as $B_t = 5B_x + 5B_y$ cycles per standard-speed frame. Thus, motion aliasing occurs for spatial frequencies $\{f_x, f_y\}$ that satisfy the constraint $f_x + f_y > OV/10$.

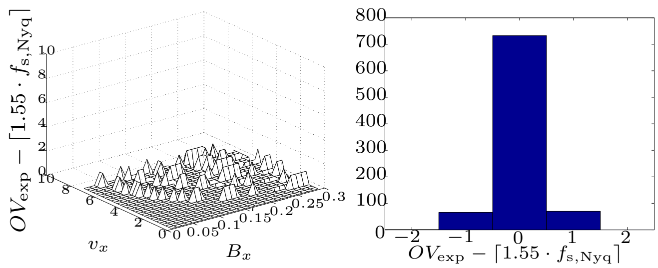
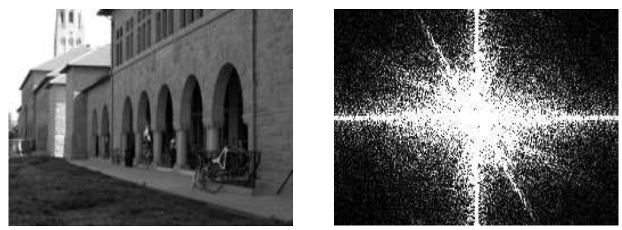


Fig. 5. Difference between empirical minimum OV for good optical flow estimation performance and OV corresponding to 1.55 times the Nyquist rate.

By computing the 2D-DFT of the first frame (shown in 6 (b)) and using this constraint, we calculated the energy in the sequence that is motion aliased for different OVs.



(a) **(b)**

Fig. 6. (a) The first frame of the sequence and (b) the magnitude of its 2-D Fourier transform.

Figure 7 plots the average angular error and the energy that is motion aliased while varying OV. Each point corresponds to an OV value and it is clear that the performance of the proposed OFE method is largely influenced by the presence of motion aliasing. As OV is increased, the motion aliased energy decreases and the average angular error of the optical flow estimation decreases as well. This confirms that motion aliasing significantly affects the performance of OFE and that a key benefit of using high frame rate sequences is the reduction of motion aliasing. Also, this example

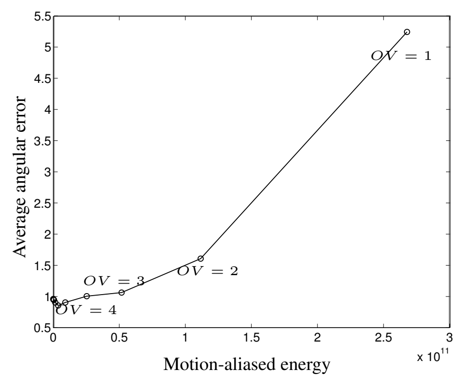
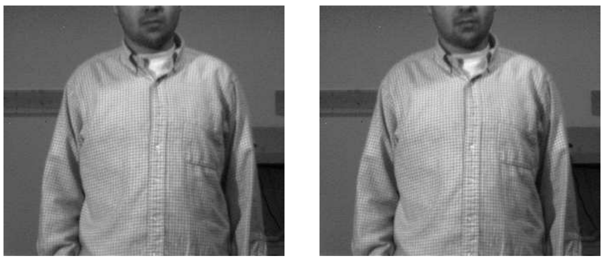


Fig. 7. Average angular error and energy in the image that leads to motion aliasing as a function of OV.

shows that with initial estimates of velocities, we can predict the amount of energy in the image that will be aliased. This can be used to identify the necessary frame rate to achieve high accuracy OFE for a specific scene.

We also show qualitative results obtained from real video sequences captured at a high frame rate. We used an experimental high speed imaging system [13], which is based on the Digital Pixel Sensor (DPS) chip and can operate at frame rates of up to 1400 frames/s. The experiment was carried out to examine the effect of aliasing in optical flow estimation, specifically when high spatial frequency is present with moderate displacements.



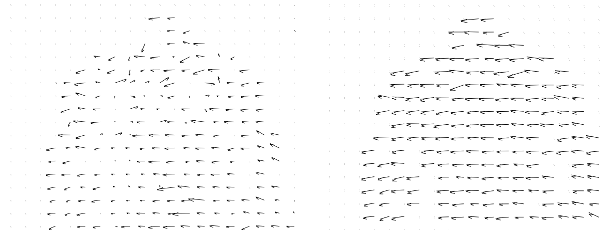
(a) Frame 9 (b) Frame 13

Fig. 8. Two frames from a real video sequence captured at 120 frames/s where the shirt has high spatial frequency content.

Figure 8 shows frames 9 and 13 of the real video sequence. The high-speed video sequence was captured at 120 frames/s ($OV = 4$) when a person was moving horizontally from right to left. The horizontal displacements between frame 9 and 13 were under 2 pixels. To estimate the optical flow between frames 9 and 13, the method described in [9] used all the frames between frames 9 and 13, while the standard Lucas-Kanade method just used frames 9 and 13. The resulting optical flows are shown in Figure 9. The effect of aliasing can be seen in some areas in Figure 9 (a) where the optical flow estimates point in the *opposite direction* of the true motion. It shows that the standard frame rate of, e.g., 30 frames/s is not sufficient to avoid motion aliasing and thus incorrect optical flow estimates. Note that the shirt has high spatial frequency although the displacements were small enough to be within general acceptable range of the standard Lucas-Kanade method. This illustrates that capturing sequences at a high frame rate not only helps when velocities are large but also for complex images with low velocities but high spatial bandwidths.

4. SUMMARY AND DISCUSSION

In this paper, we showed that temporal oversampling leads to reduction of motion aliasing and increased accuracy for a practical OFE algorithm. Temporal oversampling not only benefits optical flow estimation for large displacements but also for small displacements with high spatial frequency. Using synthetic video sequences we showed that oversampled OFE leads to reduced image energy that is motion aliased, resulting in smaller OFE errors. In addition, if 2-tap temporal gradient filters are used, we argue that the capture frame rate should be 50% higher than that needed in theory to avoid motion aliasing. Using a real high-speed video sequence, we demonstrated that oversampled OFE can lead to accurate motion estimates in contrast to standard frame rate OFE which can lead to estimates in the wrong direction. Therefore, this work demonstrates that temporal oversampling is a promising approach for improving OFE performance even when only standard frame rate optical flow is needed for subsequent video processing.



(a) Standard Lucas-Kanade (b) Proposed method [9]

Fig. 9. Optical flow estimation results for a real video sequence.

5. REFERENCES

[1] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Performance of Optical Flow Techniques," *International Journal of Computer Vision*, vol. 12(1), pp. 43–77, February 1994.

[2] A. M. Tekalp, *Digital Video Processing*, Prentice-Hall, New Jersey, USA, 1995.

[3] R. Battiti, E. Amaldi, and C. Koch, "Computing Optical Flow Across Multiple Scales: An Adaptive Coarse-to-Fine Strategy," *International Journal of Computer Vision*, vol. 6(2), pp. 133–145, 1991.

[4] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, no. 3, pp. 283–310, January 1989.

[5] J. Weber and J. Malik, "Robust Computation of Optical Flow in Multi-Scale Differential Framework," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 67–81, 1995.

[6] W.J. Christmas, "Spatial filtering requirements for gradient optical flow measurement," in *Proceedings of the Ninth British Machine Vision Conference*, September 1998, vol. 1, pp. 185–194.

[7] Eero Simoncelli, *Ph.D. Thesis: Distributed Analysis and Representation of Visual Motion*, MIT, Cambridge, MA, USA, January 1993.

[8] X. Liu and A. El Gamal, "Simultaneous Image Formation and Motion Blur Restoration via Multiple Capture," *IEEE International conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1841–1844, May 2001.

[9] S.H. Lim and A. El Gamal, "Optical Flow Estimation Using High Frame Rate Sequences," *Proceedings of the 2001 International Conference on Image Processing*, vol. 2, pp. 925–928, October 2001.

[10] D. Handoko, S. Kawahito, Y. Takokoro, M. Kumahara, and A. Matsuzawa, "On Sensor Motion Vector Estimation with Iterative Block Matching and Non-Destructive Image Sensing," *IEICE Transactions on Electronics*, vol. E82-C, no. 9, pp. 1755–1763, September 1999.

[11] Suk Hwan Lim, *Ph.D. Thesis: Video Processing Applications of High Speed CMOS Image Sensors*, Stanford University, Stanford, CA, USA, April 2003.

[12] E.P. Simoncelli, "Design of Multi-dimensional Derivative Filters," *IEEE International Conference on Image Processing*, vol. 1, pp. 790–794, 1994.

[13] A. Ercan, F. Xiao, X. Liu, S.H. Lim, A. El Gamal, and B. Wandell, "Experimental High Speed CMOS Image Sensor System and Applications," in *Proceedings of the First IEEE International Conference on Sensors*, June 2002.