Filtering Requirements for Gradient-Based Optical Flow Measurement

W. J. Christmas

Abstract—The accuracy of gradient-based optical flow algorithms depends on the ability to measure intensity gradients accurately. We will show how the temporal gradient can be compromised by temporal aliasing arising from motion and how appropriate post-sampling spatial filtering improves the situation. We will also demonstrate a benefit of using higher-order gradient estimators.

Index Terms—Aliasing, gradient method, motion measurement, optical flow.

I. INTRODUCTION

In all forms of optical flow measurement where there is significant motion ($\gg 1$ pixel/frame), if accurate measurements are required, some initial estimate of the upper limit of motion values is needed, creating a bootstrap problem. Hierarchical methods [1], [2], [4], [8] can alleviate this problem; however, the penalty is additional computational complexity. Thus for search-based methods, a limit has to be placed on the size of the search region, while for gradient-based methods [3], [8]–[10], and some other optimization methods [7], some spatial preprocessing is required. In what follows, we will show that the effect of this filtering is to reduce the effects of temporal aliasing (briefly discussed in [1]). This observation will enable us to quantify the amount of filtering needed to eliminate the aliasing entirely and hence improve the accuracy of the optical flow.

Also, the difference methods traditionally used for estimating spatial and temporal derivatives of sampled images introduce errors at higher frequencies. We will show that the effect of these errors can be reduced 1) if higher-order estimators are used and 2) as a side-effect of the antialiasing spatial filter itself.

II. TEMPORAL ALIASING AND HOW TO AVOID IT

Consider for simplicity a one-dimensional (1-D) time-varying image $h(x, t)$ of a scene under constant illumination moving with (unknown) velocity $v$

$$h(x, t) = h(x - vt).$$

Hence, if $H(\phi)$ is the 1-D Fourier transform of the stationary image $h(x)$, the transform $H(\phi, f)$ of $h(x, t)$ is

$$H(\phi, f) = H(\phi)b(\phi + f)$$

where $b(\phi + f)$ is the 1-D Dirac delta function embedded in two-dimensional (2-D) frequency space. Thus, $H(\phi, f)$ is nonzero only on the line $f = -\phi$ [Fig. 1(a)]. Note that each spectrum excurses by the same amount in the $\phi$-direction, whereas the excursions in the $f$-direction increase in proportion to the velocity $v$. The temporal spectrum at a particular position $x_0$ is given by the 1-D Fourier transform $H_{x_0}(f)$ of $h(x_0, t)$

$$H_{x_0}(f) = \frac{1}{v} e^{i2\pi f x_0/v} H \left( -\frac{f}{v} \right)$$

(1)

corresponding to a projection of the 2-D spectrum onto the $f$-axis.

If the image is sampled in the time and spatial domains with sampling frequencies $f_s$ and $\phi_s$, respectively, the spectrum of $H(\phi, f)$ in Fig. 1(a) will be replicated in the $f$- and $\phi$-directions at intervals of $f_s$ and $\phi_s$, respectively, [Fig. 1(b)]. From this, we can see that the replications of the temporal frequencies may overlap and interfere, depending on the extent of the high-frequency spatial content. This temporal aliasing will potentially affect any temporal filtering operations, in particular the calculation of the temporal image gradient $\dot{h}$. The problem will only happen for values of $v$ greater than one pixel/frame, and will clearly get worse with increasing $v$.

In the case of image sequences, temporal aliasing cannot be removed by a pre-sampling filter, since temporal sampling usually happens at the moment of acquisition. However, we can see from Fig. 1(b) that it is possible to remove the temporal aliasing by an appropriate spatial filter [Fig. 1(c)]. When the replicated spectra are projected onto the $f$-axis, we can see that they no longer overlap. The minimum spatial cutoff frequency to achieve this is $\phi_s/2v$.

To generalize the above analysis for 2-D image sequences, we note that, perpendicular to the motion direction, there is no temporal aliasing. The ideal spatial filter for 2-D image sequences would therefore be a 1-D lowpass filter, with a cutoff frequency of $\phi_s/2v$, acting in the direction of motion. However, since in practice we know neither the magnitude or direction of the motion in advance, we assume some upper limit $v_{\max}$ for the likely motion magnitude, and use a 2-D filter with cutoff frequency of $\phi_s/2v_{\max}$ along each axis. The response should be as close to zero as possible throughout the stop-band; in the passband, the accuracy of the filter response is not critical. Since the motion measurement is performed on the filtered image, the practical penalty for such filtering can be severe: in particular inaccuracies can be expected near object boundaries.

III. EXPERIMENTING WITH A SYNTHETIC SEQUENCE

We will next describe some results obtained using a synthetic image sequence. An image was generated from a set of independently randomly generated pixels, uniformly distributed. This image contains a significant and predictable amount of high-frequency information—the effects of both the antialias filter and the use of better derivative estimators show up more at higher spatio-temporal frequencies. The motion was synthesized by shifting the image by four pixels/frame horizontally. A small amount of noise was added—uniformly-distributed over a range of $\pm 5$ grey level values. To compute the optical flow, a simple least-mean-square regression algorithm [1], [5] was used to “solve” the motion constraint equation (2) over a small square neighborhood of each pixel

$$v \cdot \nabla h + \dot{h} = \epsilon$$

(2)

where $v$ is the image motion, $h$ is the image intensity field and $\epsilon$ is proportional to the noise.

1Note however that, for many cameras, some temporal filtering is effected by the temporal aperture of the camera, which partially alleviates the temporal aliasing problem.
A. Using an “Ideal” Filter

We initially used a 1-D equi-ripple design [6], with a maximum passband ripple of 3 dB and maximum stopband ripple of $-100$ dB, applied to the image successively in $x$- and $y$-directions. The transition region width was set equal to the passband width ($\approx \phi_s/4v_u$).

A square neighborhood of width $2v_u + 1$ pixels was used (three pixels if no filtering). We experimented with three different orders of derivative estimators, which were designed to correspond to the ideal at low frequencies; the effect of the higher order estimators was to improve the high frequency response. Table I shows the mean of the estimated velocity in the horizontal direction, together with the standard deviation of the error.

The mean velocity estimate in the vertical direction was always small ($<0.1$ pixel/frame), as might be expected. We interpreted the trends in Table I as follows.

**Estimator Bias:** For additive noise, if signal and noise derivatives are uncorrelated, and spatial and temporal noise derivatives are uncorrelated, we can show that the least-mean-square estimator is a biased estimator for $x$; in fact, $\tilde{v} \approx v/(1 + 1/r_{sn})$, where $r_{sn}$ is the signal-to-noise power ratio. In this context, the power refers to the square of the modulus of the spatial gradient. Thus, if there are alias components present that behave as random noise, we would expect the values of $\tilde{v}$ to be progressively biased toward zero as the filter cutoff frequency is increased beyond $\phi_s/2v_u$. This was certainly borne out in Table I: the bias increases dramatically for $v_u < 4$.

**Standard Deviation:** We would also expect the standard deviation to decrease with better filtering, reflecting a better signal-to-noise ratio. Table I in general bore this out. The exception was for the case of no filtering: this was presumably because the severe bias of $\tilde{v}$ toward zero scaled the standard deviation correspondingly.

**Derivative Estimator Order:** It is apparent from its frequency response that a simple first-order derivative estimator (i.e., central or forward difference of two pixels) will severely underestimate the correct value at higher frequencies. For $|\tilde{v}| > 3$ this will affect the temporal derivatives more than the spatial ones, and vice versa [Fig. 1(b)]. Increasing the order of the derivative estimator should therefore improve the accuracy; this was borne out in Table I. Decreasing the filter cutoff frequency beyond what is required for antialiasing should have a similar effect, as this would reduce the contribution of the higher frequency parts of the differentiator estimator response. In the table, we can see that there was indeed some significant improvement as $v_u$ increases beyond the ground truth motion value of four pixels/frame.

B. Comparison with Other Filters

We compared the effect of using the lowpass filter design suggested by Section II with a Gaussian filter, using $v_u$ as the “standard deviation” of the filter. The results were similar to those obtained using the corresponding equi-ripple filter. However, generous filter orders were used in both cases, so more experimentation with the filter order is needed, particularly for real-time applications. We also tried simple averaging filters with a range of widths from two to 16 pixels. None of them gave a mean velocity estimate of more than 25% of the correct value.

### IV. REAL IMAGE SEQUENCES

By comparison with the synthetic sequence, real image sequences have unknown spectral content, but which are generally weighted toward lower frequencies, so that less of the signal energy falls in the frequency band that suffers from aliasing. Also motion ground truth is not available. We therefore can only make qualitative comparisons between the effects of different filters. The sequence used here was from

![Fig. 1. Spatio-temporal spectra for moving 1-D image.](image1)

![Fig. 2. Sample image from sequence (before filtering).](image2)
Fig. 3. Motion vectors for different amounts of filtering using equi-ripple filter.

(a) none

(b) \( v_u = 4 \)

(c) \( v_u = 8 \)

(d) \( v_u = 12 \)

the “foreman” image sequence (Fig. 2). Since the motion is caused by camera movement, it is locally uniform, which is appropriate for the simple motion algorithm that was used. The maximum motion component was estimated to be around eight pixels/frame. The algorithm was run using the three filter types, with a second-order differencing scheme. The best results were obtained with the equi-ripple filter, for which the results for \( v_u = 0, 4, 8, \) and \( 12 \) are shown in Fig. 3. The arrows indicate the motion vectors to scale, and the circles the standard deviation of the regression fit over the patch. From these results we can see that, provided the amount of filtering is related to the actual motion (i.e., \( v_u > |v| \)), reasonable measurements can be made. The results using corresponding Gaussian filters were almost as good, while those using simple averaging filters were substantially worse.

V. CONCLUSIONS

Image sequences that contain significant motion can suffer from severe temporal aliasing, which in turn can significantly affect the accuracy of optical flow calculation. This aliasing cannot generally be removed by the usual temporal antialias filtering techniques. However, assuming constant illumination and a smoothly varying motion field, the aliasing can instead be removed by a postsampling spatial filter. The optimum amount of filtering needed is proportional to the motion magnitude. Since the motion magnitude is not known in advance, an upper limit has to be estimated. If this estimate is too high, more filtering is required, and hence more error at motion boundaries; on the other hand, if it is too low, the temporal aliasing will not be eliminated. While the exact shape of the filter appears not to be critical in practice, we found that better performance was obtained from designs that conform more closely to the ideal identified in this paper.

We also found that useful improvements to the optical flow accuracy can be obtained by increasing the order of the derivative estimators.

REFERENCES


