A Spatiotemporal/Spatiotemporal-Frequency Interpretation of Apparent Motion Reversal

Todd R. Reed
Department of Electrical and Computer Engineering
University of California
Davis, California 95616
U.S.A.

Abstract

Temporal aliasing artifacts are common in both computer generated and natural motion sequences. One of the most striking manifestations of temporal aliasing is the apparent reversal of motion commonly referred to as the "wagon wheel effect." In this paper, we examine temporal aliasing from the standpoint of joint spatiotemporal/spatiotemporal-frequency representations. We show that apparent motion reversal can be explained using these representations, and demonstrate that a motion estimation algorithm based on such a representation (the 3-D Gabor transform) can accurately predict this illusion.

1 Introduction

In temporal aliasing, components of the image sequence with high temporal frequency appear at lower frequencies due to an insufficient temporal sampling rate (commonly referred to as the frame rate). These components affect the visual quality of the sequence in two respects. The first, most noticeable at low frame rates, is a disruption in the smoothness of perceived motion (a breakdown of the apparent motion illusion underlying all motion picture display methods). Although predicting the sampling rate at which this effect will become visible is not trivial, the effect itself is intuitive - the motion appears nonsmooth because too few samples are presented. The second, which may be seen at comparatively high frame rates, is a distortion of the direction and speed of the perceived motion. This distortion is common in current film and video, and manifests in connection with objects that exhibit both significant high-spatial-frequency components (e.g., the spokes of a wheel) and that move with relatively high speed (the wheel is turning relatively rapidly). In the classical manifestation of this effect, a rotating wheel is seen to reverse its direction of rotation as the rate of rotation increases - the "wagon wheel effect." This effect is not so intuitive, and to the author's knowledge the connection between aliasing and apparent motion reversal, while observed, has not been explained in the literature.

The selection of an appropriate temporal sampling rate for a given application is of significant practical importance. A broad range of rates are in use, spanning from the very low rates (10 frames per second or less) used in current visual communications systems, through the moderate 24 - 30 frame per second rates used in animation, film, and standard television, to 60 frames per second in high definition television. Much higher sampling rates are used in scientific applications (hundreds and even thousands of frames per second), and are becoming increasingly common and economically viable as technology improves. For sequences involving high degrees of motion (particularly when detailed spatial structure must also be represented), temporal aliasing is the most important phenomenon in determining the temporal sampling rate[1], providing an additional motivation to understand it fully.

Motion is most intuitively a spatiotemporal phenomenon. However, it has been shown that it can also be characterized in the frequency domain via Fourier analysis. [Watson and Ahumada, 1983] have used this approach to investigate the relationship between frame rate, the bandwidth of the human visual system, and the perceived smoothness of motion for a moving line stimulus. Their results clearly demonstrate the value of frequency domain analysis for understanding aspects of visual motion perception.

In this paper, we consider the manner in which aliasing affects sequences at comparatively high frame rates, where motion is generally perceived as smooth, yet distortions of speed and direction may be visible. A method is demonstrated by which motion estimates consistent with those perceived visually are obtained, in cases both with and without aliasing, for sequences consisting of regions with different motions (a task that cannot be undertaken with conventional Fourier analysis). To do so, we utilize a generalization of frequency domain motion analysis, based on a joint spatiotemporal/spatiotemporal-frequency repre-

[1]Note that we distinguish here between sampling and screen update rates, the later being driven primarily by the perception of wide area flicker. Screen updates need not be unique samples, a fact reflected, e.g., in current film to video conversion practice.

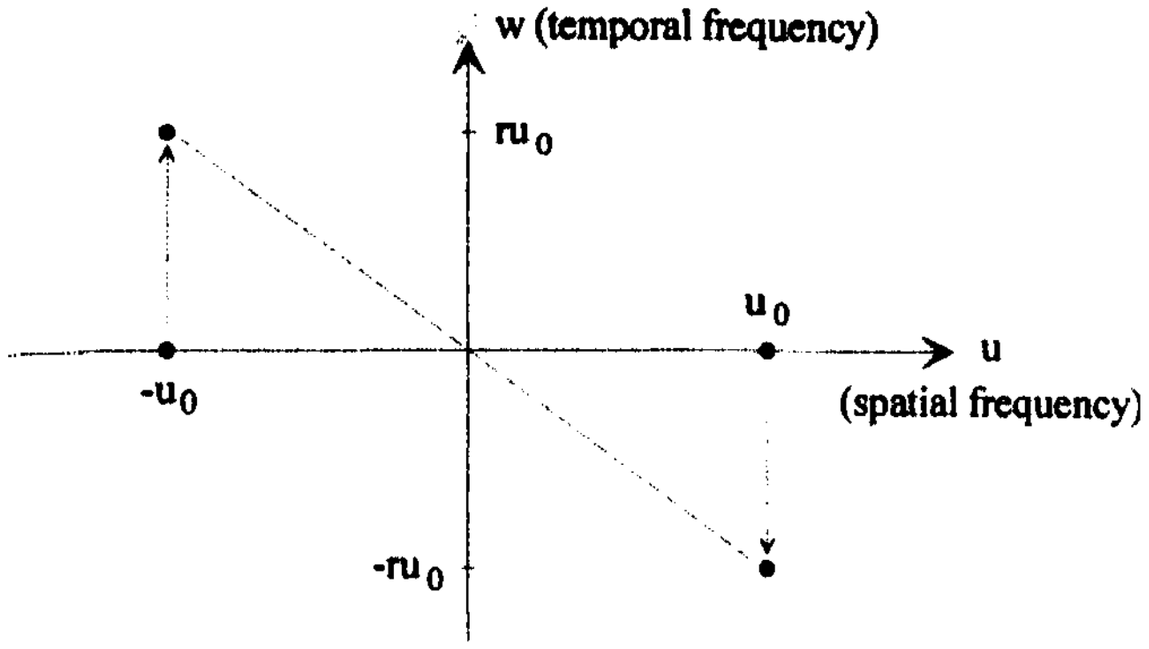

Figure 1: The spectrum of a static sinusoid with frequency u_0 and the same sinusoid moving with velocity r.

sentation (the 3-D Gabor transform).

2 Motion and Aliasing in the Frequency Domain

The frequency domain characterization of motion has been studied for some time, and for certain situations is well understood. In the case of an initially static image undergoing translation with constant velocity, motion is particularly straightforward to describe in the frequency domain. The spectrum of the sequence lies in an oblique plane, the slope of which indicates the velocity of the motion.

Consider a simple 1-D continuous sinusoidal "image" with spatial frequency u_0, moving with velocity r: $f(x,t) = sin(u_0(x - rt))$. When $r = 0$ (the image is static), the Fourier transform of the sequence consists of two components at the spatial frequencies $\pm u_0$. As r increases, the temporal frequency coordinates of the two components change from zero to $\mp ru_0$, respectively (see Figure 1). The slope of the line connecting the two components is $-r$.

As is well known from elementary sampling theory, if the sinusoid is spatially and temporally sampled, the spectrum described above is scaled and replicated on centers determined by the sampling rates. We will assume the spatial sampling rate is sufficient, and focus on the temporal sampling. As illustrated in Figure 2, for a given temporal sampling rate ω_s, the sequence can be either oversampled, critically sampled, or undersarnpled depending on whether r is less than, equal to, or greater than $\omega_s/2u_0$. Temporal aliasing occurs when the components of the replicated spectra occur at frequencies less than or equal to $\omega_s/2$ (the later two cases). Again this is well known, but sampling theory does not predict the visual impact of this aliasing.

Consider these two cases from the standpoint of the apparent motion represented (the slopes of the line or lines connecting the components of the spectra with temporal frequency less than $\omega_s/2$), however. In the critically sampled case there are in fact two lines, with slopes of $\pm r$, indicating simultaneously motions of the same speed but opposite directions. In this case, one would expect the sequence to appear essentially static (although a "jitter" or "flashing" might be seen). In the undersarnpled case there is only a single line, with slope depending on r and the degree of undersampling. For moderate undesampling the sign of the slope is positive indicating motion in the direction opposite to that of the original sequence. This is the basis of the "wagon wheel" illusion.

The above simple example might lead one to expect that temporal aliasing and its effects on perceived motion can be completely characterized using Fourier techniques. However, in sequences of practical interest (which may include multiple objects in independent motion) this is not the case. Aliasing is a local phenomenon, and the artifacts associated with aliasing are restricted to regions of the sequence which are insufficiently sampled. In practice, these regions typically correspond to objects or surfaces exhibiting high spatial frequencies, that are also moving at relatively high velocities. Although aliasing is certainly reflected in the Fourier transforms of sequences of this type, the connection between the aliased spectral components (or indeed any of the spectral components) and the spatiotemporal locations of the associated regions or objects cannot in general be made. For this reason, the Fourier transform cannot be used for the analysis of motion in complex sequences.

To identify the locations and motions of objects, frequency analysis localized to the neighborhoods of the objects is required. Windowed Fourier analysis has been proposed for such cases [Gafni and Zeevi, 1979]. However, the accuracy of a motion analysis method of this type is highly dependent on the resolution of the underlying transform, in both the spatiotemporal and spatiotemporal-frequency domains. It is known that the windowed Fourier transform does not perform particularly well in this regard. Filter bank-based approaches to this problem have also been proposed, e.g. [Fleet and Jepson, 1990], [Heeger, 1987]. A shortcoming of these approaches is the lossy nature of the proposed filter-banks, which can introduce a bias in the motion estimates.

There are a variety of alternative methods for local frequency analysis beyond the windowed Fourier transform. Examples include the Wigner distribution (a bilinear local frequency representation) and the Gabor transform (which is linear). Because they can provide a large degree of spatiotemporal locality and spatiotemporal-frequency resolution simultaneously (within the bounds of uncertainty), the use of these techniques is particularly promising for frequency-based motion analysis with multiple motions. The use of the Wigner distribution for this task was examined in [Jacobson and Wechsler, 1987]. However, the bilinear nature of the Wigner distribution (and the attendant cross terms produced by multicomponent signals) can make motion analysis difficult, in practice. A motion analysis technique based on the Gabor transform has recently been demonstrated (Reed, 1997). We will briefly described this approach in

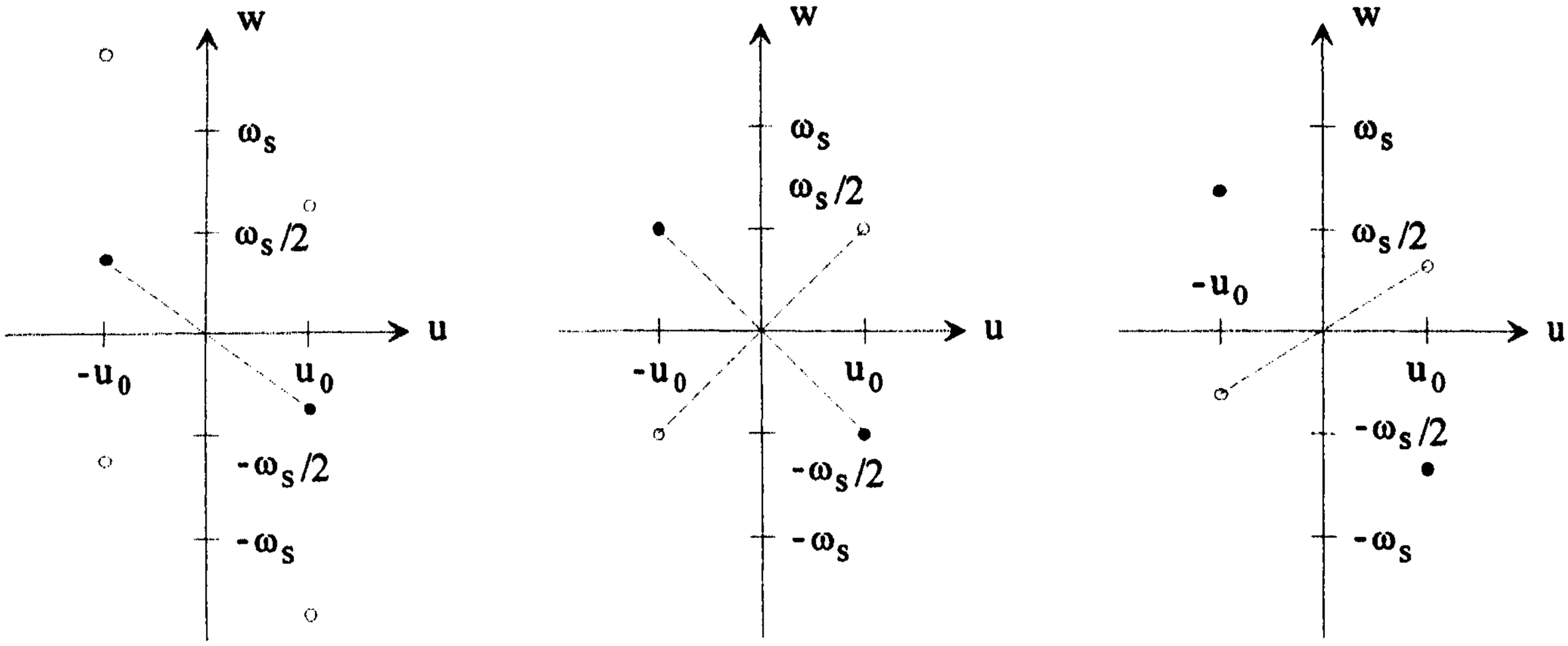

Figure 2: For a sinusoidal signal of spatial frequency Uo sampled temporally with frequencω_s, the spectra resulting from (left) oversampling, (middle) critical sampling, and (right) undersampling. The original spectral components are represented by filled circles, the replicated components by unfilled circles.

the following section, and use it to investigate apparent motion reversal in Section 4.

3 Motion Analysis using the 3-D Gabor Transform

The Gabor representation was first introduced for time-frequency analysis [Gabor, 1946], In this representation, the signal of interest is expressed as a weighted sum of basis functions formed by the products of shifted (usually Gaussian) windows and complex exponentials. The relative popularity of this representation has been due in part to the good spatial and spectral localization properties of the Gabor functions. It has also been demonstrated [Marcelja, 1980], [Daugman, 1985], [Webster and De Valois, 1985], [Field and Tolhurst, 1986], [Jones and Palmer, 1987] that the 2-D Gabor functions agree reasonably well with receptive-field profiles measured for simple cells in the cat striate cortex.

An image sequence can be considered a 3-dimensional (spatiotemporal) volume of data. Extending the Gabor representation to 3-D, this volume can be represented as the weighted sum of 3-D Gabor functions of the form

$$g(x,y,t) = \hat{g}(x,y,t) \cdot e^{j(u_0(x-x_0)+v_0(y-y_0)+w_0(t-t_0))}, \quad (1)$$

where

$$\hat{g}(x,y,t) = \frac{1}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_t} \cdot e^{-[((x-x_0)/\sigma_x)^2+((y-y_0)/\sigma_y)^2+((t-t_0)/\sigma t)^2]/2}$$

is a 3-D Gaussian function, σ_x, σ_y and σ_t determine the scale of the Gaussian along the respective axes, (x_0, y_0, t_0) is the center of the function in the spatiotemporal domain, and (u_0, v_0, w_0) is the center of support in the spatiotemporal-frequency domain.

A complete basis can be formed using the Gabor functions, resulting in an invertible transform (the Gabor transform). In the discrete case, for an image sequence with spatial dimensions *N by M* and *P* frames in length, *N.M.P* basis functions are required. The sequence can then be expressed at each discrete point (x_m, y_n, t_p) as

$$f(x_m,y_n,t_p) = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1}\sum_{l=0}^{L-1}\sum_{q=0}^{Q-1}\sum_{r=0}^{R-1}\sum_{s=0}^{S-1} c_{x_q,y_r,t_s,u_j,v_k,w_l} \cdot g_{x_q,y_r,t_s,u_j,v_k,w_l}(x_m,y_n,t_p) \quad (2)$$

where $J \cdot K \cdot L \cdot P \cdot Q \cdot R = N \cdot M \cdot P$ for completeness, $g_{x_q,y_r,t_s,u_j,v_k,w_l}(x_m,y_n,t_p)$ denotes the Gabor basis function with spatiotemporal and spatiotemporal-frequency centers of (x_q, y_r, t_s) and (u_j, v_k, w_l) respectively, and $c_{x_q,y_r,t_s,u_j,v_k,w_l}$ is the associated coefficient, which is generally complex. There is substantial freedom in selecting the locations of these basis functions, while maintaining completeness. Largely for computational reasons, in this work the functions are centered on a regular cubic grid.

Because the Gabor functions are not orthogonal, the Gabor transform coefficients cannot be calculated by simply computing the inner products of the basis functions and the signal to be transformed (or, equivalently, by convolving with the basis functions and subsampling).

Due to substantial interest over the past several years in the computation of this transform, there are a number of alternative methods available. The method used in this work is a 3-D extension of the algorithm first reported in [Ebrahimi *et* al., 1990].

From the 3-D Gabor transform of an image sequence, the motion parameters can be estimated at each spatiotemporal location by fitting the surface representing the spectral signature of the motion to the local spectrum. In the case of uniform translational motion, the slope of the planar spectrum is sought, yielding the motion vector $\mathbf{r}$. There are a number of ways in which this can be done.

A straightforward approach to estimating the slope of the local spectra, used in the examples which follow, is to form vectors of the u, v, and w coordinates of the basis functions that have significant energy (magnitudes exceeding a threshold) for each point in the sequence at which basis functions are centered. The motion vector and the coordinate vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ at each point are related as

$$\mathbf{w} = -(r_x\mathbf{u} + r_y\mathbf{v}) = -\mathbf{A}\mathbf{r} \tag{3}$$

where $\mathbf{A} = (\mathbf{u}|\mathbf{v})$. An LMS estimate of the motion vector at a given point can then be found using the pseudoinverse of A:

$$\mathbf{r}_{est} = -(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w} \tag{4}$$

4 Results

To examine the effects of temporal aliasing, we will first use a simple test sequence in which each frame includes two fields, consisting of both horizontal and vertical sinusoids with frequencies of $\pi/4$ and $\pi/2$ radians/pixel respectively, one field above the other. The sinusoids are scaled and offset so that all pixel intensities fall in the range 0 to 255. The sequence is 24 frames in length, with 256-by-256 pixels per frame. As the sequence progresses, the top and bottom fields move to the right at the same rate. We will consider three rates of translation: 1, 2, and 3 pixels/frame. The first three frames from the 3 pixel/frame sequence are shown in Figure 3. The top field is oversampled for all three velocities. The bottom field is oversampled, critically sampled, and undersampled, respectively.

We next compute the Gabor transform of the sequences, using a complete basis on 8 pixel centers in space-time, with a 3.5 pixel offset in each dimension, spaced $\pi/4$ apart in spatiotemporal- frequency, and with $\sigma_x = \sigma_y = \sigma_t = 1/8$. Computing the slope of the plane which best fits the local spectra for each sequence, the motion estimates shown in Figure 4 result for the point in time between frames 12 and 13 of the sequences. Note that the estimates are located between frames because the basis functions used in the transform are centered between frames. The arrows in the figure are scaled to the maximum velocity for each case. Similar results are obtained between frames 4 and 5, and frames 20 and 21.

As shown to the left in Figure 4, for a translation of 1 pixel/frame where both fields are oversampled, the motion estimates for the two fields are correct and identical (with the exception of some edge effects). For 2 pixels/frame (Figure 4, center), the estimate for the upper field (which is still oversampled) remains correct, while the motion estimate for the lower field is zero. In this case, the lower field is critically sampled. Visually, this field appears to "flash*' or "jitter", but not to translate, which is consistent with the motion estimate. In the third sequence (Figure 4, right), the motion estimate for the (oversampled) upper field is correct (3 pixels/frame to the right). The estimate for the lower field, which is now undersampled, is 1 pixel/frame in the reverse direction, exhibiting the apparent motion reversal discussed above. Visually, the lower field appears to move to the left, consistent with the estimate.

We next consider the sequence shown in Figure 5, 24 frames in length, with 256-by-256 pixels per frame. The background is static, consisting of a sinusoidal "plaid" field with horizontal and vertical frequencies of 7r/4 radians/pixel. Superimposed on this background are two blocks, 80 pixels square, starting in the upper and lower left of the frame. The first has spatial frequency components identical to the background, while the second has horizontal and vertical frequency components of 7r/2. The sinusoidal plaids are each scaled and offset, so that all pixel intensities fall in the range 0 to 255. As the sequence progresses, the two blocks move to the right at 3 pixels/frame.

The 3-D Gabor transform of the sequence was computed as in the previous example. Computing the LMS estimate described in equation 4 the motion estimates shown in Figure 6 result, for the points in time between frames 4 and 5, 12 and 13, and 20 and 21,respectively.

This example illustrates two points. Viewed as objects, the surface characteristics of the blocks lead to correct motion estimates for the interior of the upper block (consistent with the object motion), but incorrect (reversed in direction and reduced in speed) for the interior of the lower block. This is exactly as found for the two fields moving at 3 pixels/frame in the previous example. However, as is clear by examining Figure 6 from left to right, the motion of the blocks themselves remains correctly represented. They both move from left to right, retaining the same relative position. This is the same behavior observed in the classical "wagon wheel" illusion, where the wheel appears to reverse its direction of rotation and to rotate at a lower rate, but as a whole continues to move in the proper direction at the proper speed.

5 Conclusions

In this paper we have investigated the mechanism underlying apparent motion reversal in image sequences which exhibit temporal aliasing. We have demon-

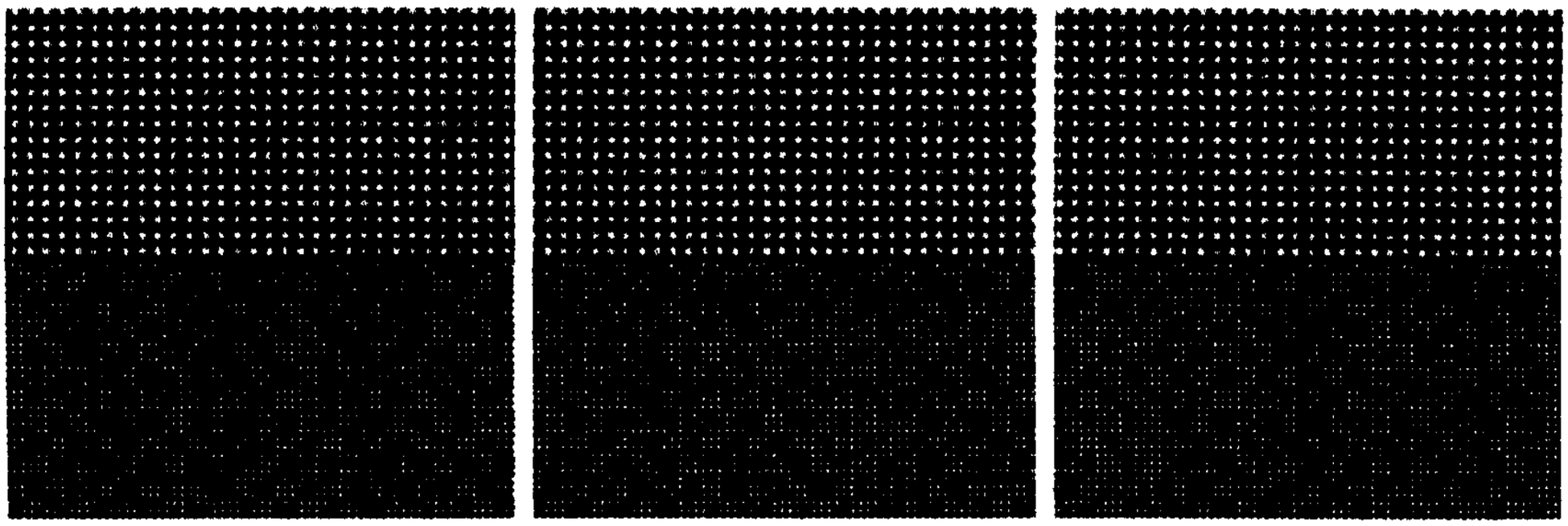

Figure 3: The first three frames from the first test sequence.

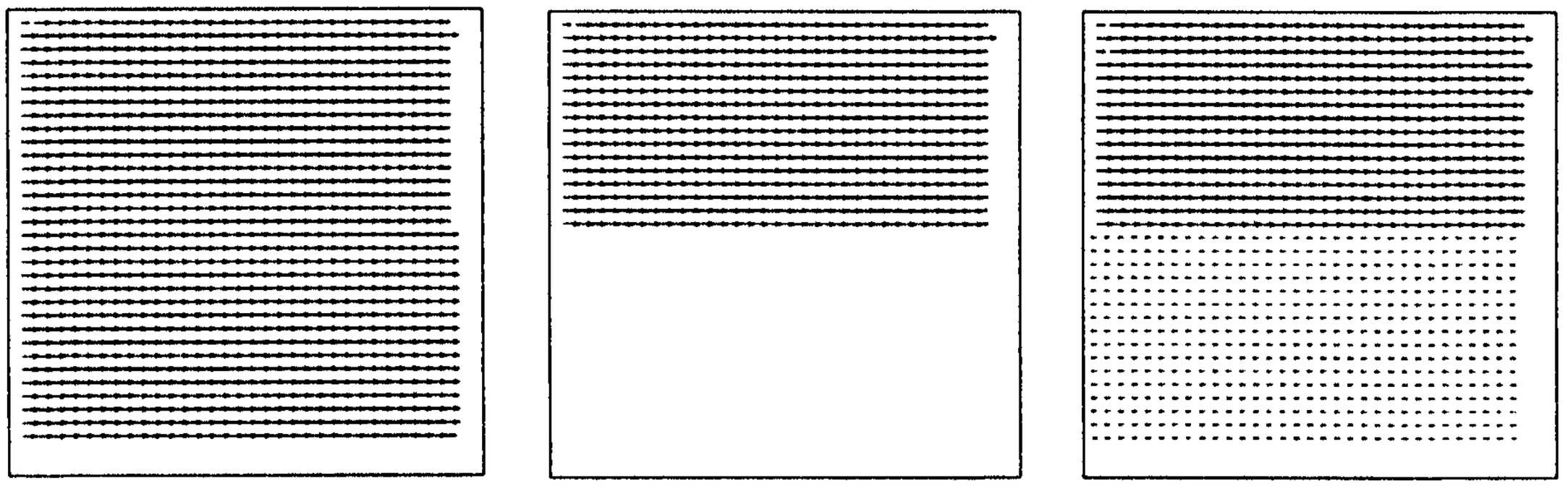

Figure 4: The optical flow fields for translations of 1, 2 and 3 pixels/frame.

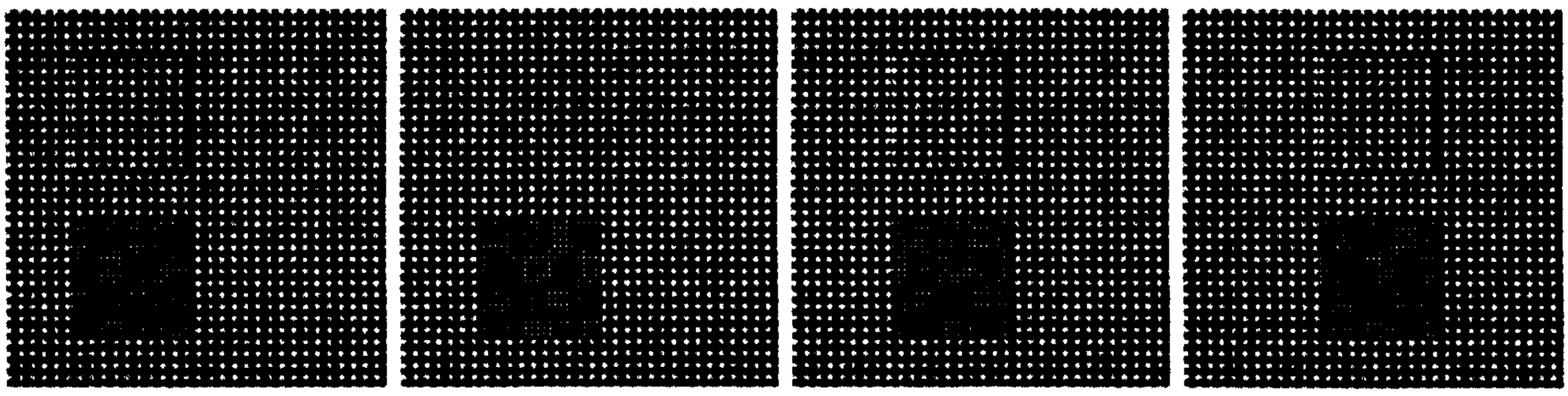

Figure 5: From left to right: frames 5, 9, 13 and 21 (of 24) from the second test sequence.

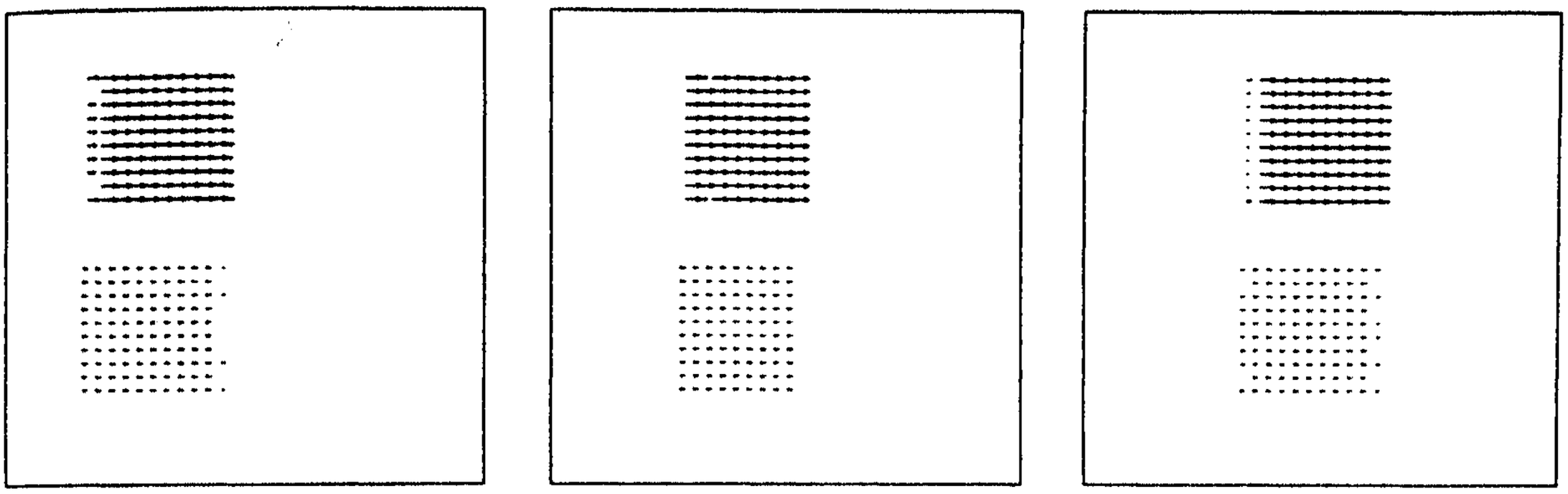

Figure 6: The optical flow fields between frames 4 and 5, 12 and 13, and 20 and 21.

strated analytically and experimentally that this phenomenon can be both understood and predicted using spatiotemporal/spatiotemporal-frequency representations. Using a motion estimation procedure based on one such representation, the 3-D Gabor transform, motion estimates consistent with those perceived visually were obtained in cases with and without aliasing. Finally, it was shown that overall object motion information is preserved using a procedure of this type, even when the surface properties of the object induce apparent motion reversal over the object surface. This is just as observed in the classical "wagon wheel" illusion.

Acknowledgments

This work was supported in part by the Computer Vision Laboratory, Department of Electrical Engineering, Linkoping University, Sweden.

References

[Daugman, 1985] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A,* 2(7) :1160-1169, July 1985.

[Ebrahimi *et al.*, 1990] T. Ebrahimi, T. R. Reed, and M. Kunt. Video coding using a pyramidal Gabor expansion. In *Proceedings of VCIP '90,* pages 489-502, Lausanne, Switzerland, October 2-4 1990.

[Field and Tolhurst, 1986] D. J. Field and D. J. Tolhurst. The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London,* 228(1253):379~400, September 1986.

[Fleet and Jepson, 1990] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *Int. J. of Comp. Vis.,* 5(1):77-104, 1990.

[Gabor, 1946] D. Gabor. Theory of communication. *Proceedings of the Institute of Electrical Engineers,* 93(26):429-457, 1946.

[Gafni and Zeevi, 1979] H. Gafni and Y. Y. Zeevi. A model for processing of movement in the visual system. *Biological Cybernetics,* 32:165-173, 1979.

[Heeger, 1987] D. J. Heeger. Model for the extraction of image flow. *J. Opt. Soc. Am. A,* 4(8): 1455-1471, 1987.

[Jacobson and Wechsler, 1987] L. Jacobson and H. Wechsler. Derivation of optical flow using a spatiotemporal-frequency approach *Computer vision$_f$ Graphics, and Image Processing,* 38:29-65, 1987.

[Jones and Palmer, 1987] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology,* 58(6): 1233-1258, December 1987.

[Marcelja, 1980] S. Marcelja. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America,* 70(11):1297-1300, November 1980.

[Reed, 1997] T. R. Reed. The analysis of motion in natural scenes using a spatiotemporal/spatiotemporal-frequency representation. In *Proceedings of the IEEE Int'l Conf. on Image Processing,* pages 1-93-1-96, Santa Barbara, California, October 26-29 1997.

[Watson and Ahumada, 1983] A. B. Watson and A. J. Ahumada. A look at motion in the frequency domain. In *SIGGRAPH/SIGART Interdisciplinary Workshop MOTION: Representation and Perception,* pages 1-10, Toronto, Canada, April 4-6 1983.

[Webster and De Valois, 1985] M. A. Webster and R. L. De Valois. Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America,* 2(7):1124-1132, July 1985.