CHAPTER 5

Fragmentary Motion Sensing

5.1. Introduction

It has been shown in the previous chapter that, by careful attention to details of spatially interactive coupling, it is possible to devise a preprocessor which is robust against exact sampling position for sensing of fragmentary profiles. However, this work presupposed that we were dealing with a single 'snapshot' image in time. In practice this is rarely the case. More often there is some small motion of parts of an image or the entire image with respect to the sight line and sampling matrix. Such motion is potentially able to provide (and does in practical human vision) a powerful means of discriminating scene information from noise and of sensing local motion in a scene. The potential of dynamic noise discrimination for an essentially stable imaging system will receive limited attention in Chapters 6 and 8. The purpose of *this* chapter is to discuss the problems associated with the sensing of components of motion of fragmentary profiles automatically, to present a theory for *direct* measurement of local motion of such fragmentary profiles and to demonstrate its power. More general motion analysis based on such fragmentary data will be dealt with in Chapter 9.

5.2. The problem of automatic sensing of motion

In principle it appears trivial to sense local motion in an image. It has been shown by several workers that simple subtraction, point by point, of the illuminance distribution in an image at two instants of time will suffice to *sense* motion (or, strictly, the component of motion orthogonal to an edge – the component *parallel* to an edge is normally lost due to the 'aperture problem' (see Chapter 9). Resultants will only be obtained in regions close to discontinuities in the luminance distribution which have moved (provided that the overall scene luminance is not changing with time) [e.g. 5.1]. If such an operation is carried out on two sharply imaged 'snapshots', then either a light or a dark band is obtained at each region where there has been local motion. However, whilst this can signal *existence* of motion, the information about the motion orthogonal to edges is coded in rather unsatisfactory terms. That is, the *amount* of motion is coded as the *widths* of the bands (which themselves must then be measured if objective data are required). If there is a *large* amount of *local* motion in a simple and otherwise static scene, it may well be possible to arrange some means of measuring the motion (as an

additional computation involving edge finding and then bandwidth determination). However, if there is a significant general motion, then the auxiliary task of computing local motion via such edge finding and bandwidth determination is very cumbersome (and is also essentially limited to a one pixel resolution). Yet again, for *considerable* motion in *complex* scenes, there will be a serious danger of confusion. It would therefore be better if one could sense local motion, including very small motion, more directly, at the same time as sensing any mean global motion. It is believed that our interpretation of the spatial preprocessing arrangement utilised by the human visual system is admirably suited to achieve this end.

5.3. Principles of motion from form

As already discussed in Chapter 2, the human visual system is believed to have two parallel spatio-temporal processing systems [e.g. 5.2, 5.3]. One of these systems senses form at high spatial resolution, subsequently deriving motion from form (the X ganglion or sustained system). The other primarily senses energy transients at lower resolution, thence sensing motion and subsequently form from motion at a more global level (the Y ganglion or transient system). However, study of available psychophysical and neurophysiological data has led us to the conclusion that, in the majority of everyday tasks, the sensing of form *and* local motion are handled by the sustained, high resolution system. The transient system we then believe to provide primarily the general awareness of global information flow for use by the balance mechanisms, etc.. Our attention has therefore, thus far, been given primarily to the sensing of local motion from the *form* channel (which is the channel previously studied in detail both at a probabalistic level in ORACLE and at a definitive level in VISIVE (see Chapter 3). As discussed in Chapter 3, for this channel the principle stages of processing are firstly a gross blurring of the image, then a very local inhibitory interactive processing similar to a Laplacian operation and, finally, a one-dimensional local integration and orthogonal differencing for location of fragmentary lines and edges. An important property of such processing, which is of great relevance to sensing of local motion is the provision, in the second difference outputs provided by the Laplacian type of operation on the blurred image, of a local approximately ramp change of signal. This ramp region typically extends over 2 or 3 pixels in the image, in the vicinity of any local luminance discontinuity in the original scene. In Chapter 4 the ramp again was utilised as one method of extracting vernier positional information (to about 0.1 pixels) from 'snapshots', for improved form discrimination. If such vernier positional information is recorded as a time sequence then, in principle, simple subtraction of vernier positional data for matched features on successive time samples will yield, directly, the *total* local motion with a high sensitivity [5.4]. In other words, the single frame vernier positional and orientation data permit a form of paired-frame matching with a potential for sensing local displacements with sub-pixel accuracy. Alternatively, and potentially more powerfully and directly, in the presence of motion the ramp zone around a local luminance discontinuity may be considered in terms of a

temporal ramp crossing a given pixel (Fig. 5.1). Thus, if sufficient time samples are taken whilst the ramp is crossing a pixel (i.e. during a motion of some 2 or 3 pixel spacings) high fidelity information is available about the component of fragmentary motion orthogonal to the local profile similar to that already demonstrated to be both available and powerful from *spatial* distribution on a 'snapshot'. It will be shown in Chapter 9 that, armed with distributed data on the orthogonal component of motion, pixel by pixel, it is relatively straightforward to reconstruct the complete optical flow field. It will then be shown, in Chapter 13, how the absolute flow of salient features such as corners may be obtained directly from the same orthogonal components of motion.

By inference, if the blurring and spatial sampling relationship assumed for the eye in VISIVE is near optimal for information transfer (cf. Chapter 3), equally, with the same blur, a time sampling such that the orthogonal component of motion is one pixel/time sampling interval will also be near optimal. However, this does not mean that, in order to utilise the induced temporal data, one must have a variety of sampling intervals to suit various motion rates, but rather that for a fixed sampling interval the motion response is a distributed bandpass filter in temporal frequency. This is in fact what has been observed to be the case for the human visual system. In that case the bandpass filter for the sustained system provides measurable performance extending from a very low threshold, of around 0.25

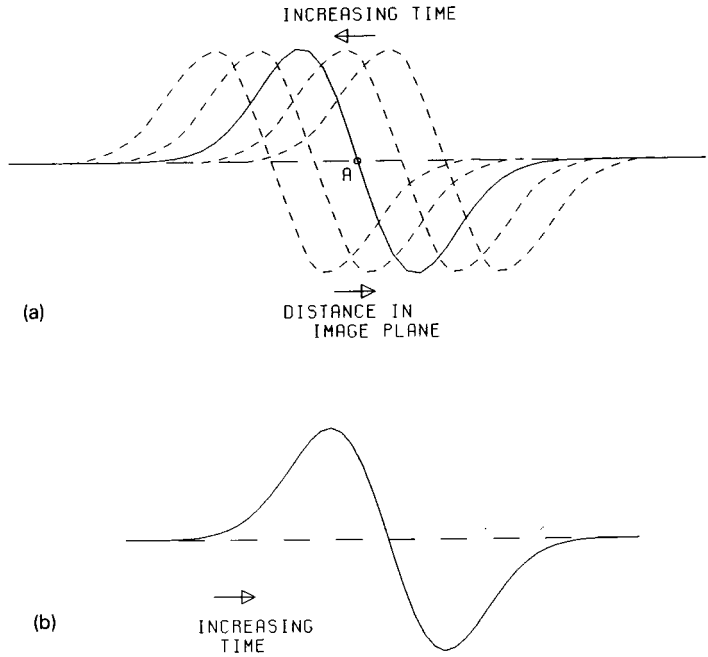

Fig. 5.1. Illustrating the automatic generation of temporal second difference functions by transit of spatial second difference functions past a given point. a) Progressive development of the spatial second difference function with time. b) Temporal second difference function at point A due to (a).

mrads/sec, up to something like 1 degree/sec. [5.5]. At higher rates the transient system mentioned earlier takes over [5.5].

5.4. Bases of practical implementation

In the previous Section, two alternative approaches to high fidelity local motion sensing have been discussed, in principle. What now of the practicalities? The practical aspects of the two methods will be discussed separately below.

5.4.1 Local motion from vernier position differencing

The problem of motion sensing by temporal differencing of vernier position data using VISIVE was the subject of a substantial research study during 1983/4. A version of VISIVE was at that time developed which could generate and store a series of vernier position and vernier orientation matrices for a sequence of time samples. Attempts were then made to generate local fragmentary velocity matrices from paired differences of such vernier data [5.4]. The only major problem, which made that program very cumbersome, was essentially the so-called 'correspondence problem' which other workers have experienced, particularly with relation to stereo association [e.g. 5.6]. This problem is essentially such that, if the local vernier profile fragments from two successive time samples of the same piece of profile do not fall on the same pixel, one must find suitable logic to relate *equatable* fragments on *different* pixels. Using positional information alone, as commonly employed in conventional feature matching techniques, this can be very unreliable. However, it was shown in Chapter 4 that one can obtain, from optimally blurred 2nd difference signals, not only vernier position, but also what we have called vernier orientation (to better than 1 degree). By utilising both the vernier position *and* vernier orientation data available from VISIVE, with an appropriate tolerance on orientation, one can minimise the correspondence problem (by greatly restricting permissible directions for comparisons), but at the expense of very cumbersome computation. The details of this practical implementation will not be reported further here. They were the subject of a technical report [5.4], but are not believed to be of great import, bearing in mind subsequent developments to be described. Suffice it to say that, within the constraints of tolerances related to the correspondence problem and noise, a system was implemented which could sense local motion of between 0.1 pixels/time interval and 2 pixels/time interval. Such a system could, if necessary, be extended to handle larger motions (at reduced absolute accuracy) by input image scaling.

5.4.2. 'Direct' sensing of local motion

The growing awareness of the correspondence problem, and the cumbersome processing necessary to overcome it, led us to look again, in parallel with the study just discussed, at the visual preprocessing. After all, human beings, and even more so many lower animals, are instantly and readily aware of local motion. It therefore

would seem reasonable to suppose that such sensing should be simple and direct, not cumbersome. The rethink, together with a reappraisal of known psychophysical data on real and apparent motion perception, led to the realisation that one may consider the second difference data, which is the basic transformed data available at the optic nerve, to contain not only *spatial* ramps but also *temporal* ramps in the vicinity of any local profile. [Note that in the following discussion, to be in line with conventional discussion related to second difference functions the concept of 'zero crossings' is used. Such zero crossings should be considered to be equatable to scene luminance discontinuities. A method of addressing *true* profile locations for 'difficult' scenes, where the profile does not coincide with the zero-crossing, will be discussed later – Section 5.10]. The rate of change of second difference signal with time, as sensed at a given pixel in the vicinity of a zero crossing, essentially yields a direct measure of the component of motion orthogonal to the local profile (Fig. 5.1). In fact the blurring of the image, plus time sampling, immediately provides the d'Alembertian function discussed, for instance, by Buxton and Buxton in their important paper on optical flow [5.7]. More importantly, if one takes two time samples, such that motion effects are contained well within the ramp function (i.e. motion of less than perhaps two pixels *in any direction*), one can, by taking both sum and difference of two time samples, obtain both a modified second difference signal for form extraction *and* a measure of the orthogonal component of motion.

Let us first look at an approximate concept. A more rigorous treatment, limited only by the assumption of a Gaussian form of blur function, will then be developed in Section 5.5. Consider Fig. 5.2. Let the means of two time samples be I_{AM} and I_{CM}. Then

$$I_{AM} = (I_{A1} + I_{A2})/2$$

$$I_{CM} = (I_{C1} + I_{C2})/2$$

For *single* frame analysis, the local profile strength is usually given from $(I_{A1} - I_{C1})$ or $(I_{A2} - I_{C2})$ (Chapter 3). Now let the motion be sufficiently small to

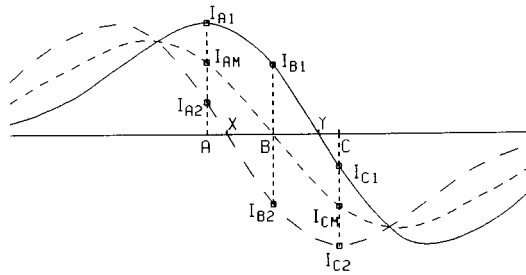

Fig. 5.2. Illustrating the concept and method of extracting motion *and* normal form information from pairs of time samples of second difference functions in the vicinity of a zero crossing. ——— time sample 1, — — — time sample 2, ------ mean second difference function. [*N.B.* The zero crossing at B is shown at a time midway between the two samples for simplicity. The analysis of motion is the same for B within ±0.5 pixels of the true zero crossing]

retain an adequate ramp in the mean of the two second difference functions. Under these conditions the strength may be approximated by $(I_{AM} - I_{CM})$. In such a case, from similar triangles,

$$(I_{B1} - I_{B2})/(X - Y) \simeq (I_{AM} - I_{CM})/(A - C)$$

But $(A - C) \equiv 2$ pixels. Therefore

$$(X - Y) \simeq 2(I_{B1} - I_{B2})/(I_{AM} - I_{CM})$$

where $(I_{B1} - I_{B2})$ is a measure of the maximum gradient in the time domain derived from the *difference* of the two time samples.

In the foregoing computation the sum (or mean) will yield single, direct values for vernier position and vernier orientation which are for the *mid-point* in time *between* the two time samples, whilst the difference is also a direct measure. Thus one immediately circumvents the correspondence problem (or perhaps I should say that the correspondence problem *does not exist*!). Also the strength of the form signal may be used to normalise the motion signals, in order to make them independent of local edge contrast and yield direct measures of *absolute* rate of motion. It has to be admitted that, as motion rate increases (with a fixed spatial sampling, fixed image blur and a fixed time interval between samples), the mean second difference ramps become distorted (and eventually break down – see Section 5.5 for details). But this is known to be a characteristic of human vision [5.5] and is probably a small penalty to pay for a very simple and efficient, high sensitivity, local motion sensing facility. In any case, the eventual breakdown can be offset to some extent by playing off spatial resolution and local motion sensitivity against maximum motion which can be sensed. This aspect is discussed in Chapter 9.

The alternative motion sensing technique was initially implemented on a version of VISIVE, simply by adding a two glimpse store at the second difference level (i.e. optic nerve simulation), an average and difference routine at the second difference level and a loop to run the one-dimensional local integration and orthogonal differencing routine, representing area 17 of the striate cortex, twice, once using mean second difference and once using differences of second differences. Practical applications showed much promise. Methods of overcoming the 'aperture problem', owing to measurement of only the components of motion orthogonal to local profile orientation, are discussed and developed in Chapters 9 and 13.

5.4.3. Motion of hot-spots

In Sections 5.4.1 and 5.4.2 we have been concerned with sensing local motion of profiles of fully resolved objects. What of the sensing of motion of hot-spots? One way of accomplishing this would be to sense the vernier position of the centre of the (blurred) hot-spot image on two successive time samples and then subtract them. This can be done in a manner similar to sensing vernier orientation, but the method is indirect and also suffers from some measure of correspondence prob-

lem. A better and more direct method is to acknowledge that a hot-spot image exhibits a well-defined zero crossing circle of finite diameter at the second difference level. Thus the vernier motion of a hot-spot may be characterised, directly, by essentially the same method as that suggested in Section 5.4.2. In this case the leading and trailing portions of the moving zero crossing circle define both the magnitude and direction of motion (normalised appropriately from the mean second difference signal), whilst the *existence* of the hot-spot is still available from either the mean first or second difference matrices.

5.5. More rigorous theory

The overall blur function of VISIVE, may be considered to be approximately represented by a 2-D Gaussian with sigma of about 1.7 pixels. This blur function is the composite of the optical blur before the Laplacian, a 3×3 weighting function and the sum of a set of weighted 5×1 local integrators at 30 degree differences of orientation, all referred to a hexagonal sampling matrix. Since the basic Laplacian differencing function is of minimal size (one pixel centre inhibited by a single ring of 6 surrounding pixels – Chapter 3), one may assume essentially that the summed bar array represents the pure second difference of a Gaussian for a point source, whilst the slope of the second difference function near a zero crossing represents the pure *third* difference of a Gaussian for such a point source. Now, for an edge, the blur function is the integral of a 1-D Gaussian (an erf function [5.9]). Thus the *second* difference function at the summed bar array for a fragmentary edge is actually represented by a *first* difference of a 1-D Gaussian with a sigma of 1.7. We are now in a position to define mathematically the approximate forms of both the sum and difference of paired frame data as approximated by simple triangles in Fig. 5.2. The general law for the first derivative of a Gaussian is $y = \mathrm{d}(\exp(-x^2/2\sigma^2))/\mathrm{d}x$ or $y = -x/\sigma^2 * \exp(-x^2/2\sigma^2)$. Then the average of two frames near a zero crossing with displacement δx between frames will be

$$y = -(1/2\sigma^2)\left[(x - \delta x/2)\exp\left(-(x - \delta x/2)^2/2\sigma^2\right) + (x + \delta x/2)\exp\left(-(x + \delta x/2)^2/2\sigma^2\right)\right] \tag{5.1}$$

Now the gradient near the zero crossing is estimated in VISIVE by taking the difference between points separated by 2 pixels. Thus on average the strength is given as

$$\begin{aligned}\delta y = -(1/2\sigma^2)\Big[&(x - \delta x/2 - 1)\exp\left(-(x - \delta x/2 - 1)^2/2\sigma^2\right) \\ &+ (x + \delta x/2 - 1)\exp\left(-(x + \delta x/2 - 1)^2/2\sigma^2\right) \\ &- (x - \delta x/2 + 1)\exp\left(-(x - \delta x/2 + 1)^2/2\sigma^2\right) \\ &- (x + \delta x/2 + 1)\exp\left(-(x + \delta x/2 + 1)^2/2\sigma^2\right)\Big]\end{aligned} \tag{5.2}$$

For approximate symmetry about zero (i.e. near the zero crossing of the mean)

$$\begin{aligned}\delta y \simeq -(1/2\sigma^2)\big[&-(1+\delta x/2)\exp\big(-(1+\delta x/2)^2/2\sigma^2\big)\\&-(1-\delta x/2)\exp\big(-(1-\delta x/2)^2/2\sigma^2\big)\\&-(1-\delta x/2)\exp\big(-(1-\delta x/2)^2/2\sigma^2\big)\\&-(1+\delta x/2)\exp\big(-(1+\delta x/2)^2/2\sigma^2\big)\big]\\ \simeq -(1/\sigma^2)\big[&-(1-\delta x/2)\exp\big(-(1-\delta x/2)^2/2\sigma^2\big)\\&-(1+\delta x/2)\exp\big(-(1+\delta x/2)^2/2\sigma^2\big)\big]\end{aligned} \tag{5.3}$$

At the zero crossing of the mean the *difference* between the outputs at the two time samples will be

$$\delta z = -(2(\delta x/2)/\sigma^2)\exp\big(-(\delta x/2)^2/2\sigma^2\big) = -(\delta x/\sigma^2)\exp(-\delta x^2/8\sigma^2) \tag{5.4}$$

where we can measure δz.

Now both δy and δz (our measurable functions) contain complex relationships of δx (our unknown). However, as δx tends to zero we find that δy tends to $2/\sigma^2\exp(-1/2\sigma^2)$, whilst $\exp(-\delta x^2/8\sigma^2)$ tends to unity. Thus for small motions one has the approximate law

$$\delta z/\delta y \simeq -(\delta x/\sigma^2)/\big[(2/\sigma^2)\exp(-1/2\sigma^2)\big] \simeq \delta x/\big[2\exp(-1/2\sigma^2)\big] \tag{5.5}$$

whence $\delta x \simeq \delta z * 2 * \exp(-1/2\sigma^2)/\delta y$. Finally, with $\sigma = 1.7$,

$$\delta x \simeq 1.68 * \delta z/\delta y \tag{5.6}$$

This is essentially the simple computation carried out in the flow version of VISIVE.

The values of δx computed from equation 5.6 are obviously subject to a progressive error as δx increases, since for such situations the assumption that $\exp(-\delta x^2/8\sigma^2) = 1$ is increasingly in error. A simple computational way of reducing the error is to compute an *estimate* of δx from equation 5.6 and then to

TABLE 5.1.

Theoretical errors in VISIVE fragmentary motion analysis for a noise free edge with mean position crossing a pixel centre.

True	Equation 5.6		Equation 5.7	
δx (pixels)	δx (pixels)	% error	$\delta x'$ (pixels)	% error
0	0	0	0	0
0.25	0.251	0.3	0.250	0.042
0.50	0.508	1.68	0.503	0.56
0.75	0.780	4.07	0.760	1.36
1.00	1.076	7.6	1.023	2.35
1.25	1.41	12.6	1.29	3.32
1.50	1.79	19.3	1.56	3.87
1.75	2.25	28.5	1.81	3.25
2.0	2.82	41.2	2.00	0.00
2.50	4.65	85.9	1.82	−26.9
3.00	9.25	208.4	0.23	−92.4

compute a second estimate $\delta x'$ where

$$\delta x' = (1.68\delta z/\delta y)/\exp(-\delta x^2/8\sigma^2) \tag{5.7}$$

i.e. using the first estimate to compute an approximate correction factor.

5.6. Error studies

Using equations 5.1 to 5.7 we have carried out an analysis of errors for sensing motion of ideal, straight, noise free edges using VISIVE. Potential errors fall into 2 categories – those due to the existence of the product term $\delta x \cdot \exp(-\delta x^2/8\sigma^2)$ in equation 5.4 and those due to the fact that mean edges do not always fall exactly across a pixel centre.

Errors were computed from both equation 5.6 and equation 5.7 for a centred mean and with δx up to 3 pixels/temporal sampling interval. These are shown in Table 5.1. It can be seen that the simple computations normally used in VISIVE are not quite adequate to provide an accuracy of better than 10% for motions of up to 1 pixel/temporal sampling interval. However, inclusion of the secondary computation of equation 5.7, where deemed necessary, can result in errors of less than 4% for up to 2 pixels/temporal sampling interval.

When a mean profile position is other than across a pixel centre there are deviations in the measured values of both δy and δz from those assumed in the simple computation. There is in this case no simple way of compensating for the deviations. However, although the progressive deviations in both δy and δz with increasing asymmetry are rather large, it so happens that they are both in the same sense. Thus, when introduced into equations such as 5.6 and 5.7, the overall effect

is compensatory and therefore much smaller than either. A number of computations have been carried out, for permutations of δx and the magnitude of asymmetry, from which it appears that the component of error due to asymmetry may be taken to be less than 0.1 pixels/temporal sampling interval for δx's up to 2 pixels/temporal sampling interval.

5.7. Discussion

It has been shown that the composite effect of the total sequence of processes contained in the optical flow version of VISIVE may be readily represented by simple approximate mathematical functions. By doing so it rapidly follows that one can provide an approximate measure of the orthogonal component of local motion from an extremely simple mathematical function, whilst a relatively simple single iteration can provide a first order correction, should such be necessary. The present analysis and practical results suggest that errors of less than 0.1 pixels or 10% of the motion, for orthogonal motion between samples of less than 1.5 pixels, and less than 15% up to motion between samples of not more than 2 pixels are achievable by the simple, direct method. It is doubtful, therefore, whether for most practical situations it is worth the additional computation to improve on such accuracy. For orthogonal motion between samples of more than 2 pixels, the errors increase rapidly. It is therefore necessary, for such situations, to consider flow coupled with progressive shrink (see Chapters 9 and 11), which will rapidly permit motion analysis up to many pixels between time samples for adequately coarse image details.

5.8. Practical local motion estimation

From the theory for paired frame analysis discussed in Section 5.5 it was deduced that the errors in predicted local motion by direct computation are progressive, becoming very large for between frame local motions in excess of 1.5 to 2 pixels (see Fig. 5.3). It was, however, shown to be possible to compensate for such theoretical errors with a between frame motion of 2 pixels by a single extra computational step (equation 5.7) – viz.

$$\delta x = \delta x_{\text{app}} / \exp\left(-\delta x_{\text{app}}^2 / (8 * \sigma^2)\right) \tag{5.8}$$

where δx_{app} is the computed estimate of between-frame motion and σ is the characteristic standard deviation of the composite Gaussian image blur function (in VISIVE $\simeq 1.7$ pixels).

Applying this correction, which has a very small effect for small values of δx_{app}, results in a theoretical progressive error curve as shown dotted in Fig. 5.3, producing a residual theoretical error of not more than 3% for motion between

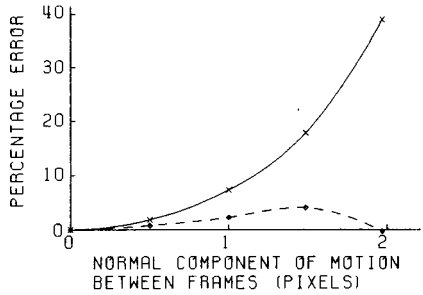

Fig. 5.3. Theoretical errors in prediction of apparent motion orthogonal to local profile orientation by VISIVE, as a function of magnitude of motion between frames. ——— Direct computation (as equation 5.6) ------ First order correction to balance error for a motion of 2 pixels between frames (as equation 5.7)

frames less than 2 pixels, but producing gross overcorrection for larger motions. Such an error function is considered by us to be very satisfactory for potential use in conjunction with the concept of progressive shrink (Chapter 11), where motion between samples significantly in excess of 2 pixels has to be handled.

In order to test the *practical* performance capabilities, basic test images were produced similar to those used for testing the vernier capabilities in the previous chapter. Many of the images were duplicated with displacements of 0.5, 1.0, 1.5 and 2.0 pixels either horizontal, vertical or at 45 degrees. Pairs of displaced images were then used to investigate both the positional and orientation vernier capabili-

TABLE 5.2.

Statistics of orientation differences for large disc images, as computed from paired frame analysis for various motion conditions.

Motion (pixels)			Mean orientation difference (degrees)	S.D. of orientation differences (degrees)
X	*Y*	radial		
1.5	1.5	0	3.22	1.41
0.75	0.75	0	3.21	1.23
0	1.5	0	3.26	1.30
0	1.0	0	3.27	1.28
0	0.5	0	3.27	1.31
1.5	0	0	3.22	1.39
1.0	0	0	3.23	1.24
0.5	0	0	3.23	1.37
1.0	1.0	0	3.20	1.29
0.5	0.5	0	3.27	1.36
0	0	1.0	3.16	1.38
1.0	1.0	0.5	3.21	1.31

ties for averages of two frames, as claimed to be acceptable for form extraction in the presence of local motion, and the accuracies of local velocity extraction.

5.9. Paired-frame analysis

For any two frames the motion sensing version of VISIVE has the theoretical ability to extract, pixel by pixel, the vernier position and orientation of profile fragments from the mean 2nd difference, plus the local motion orthogonal to the local fragmentary profile from the difference of 2nd differences. Similar analyses to those in Chapter 4 were therefore carried out on various paired samples of discs and squares, representing various motions between samples ranging from 0.25 to

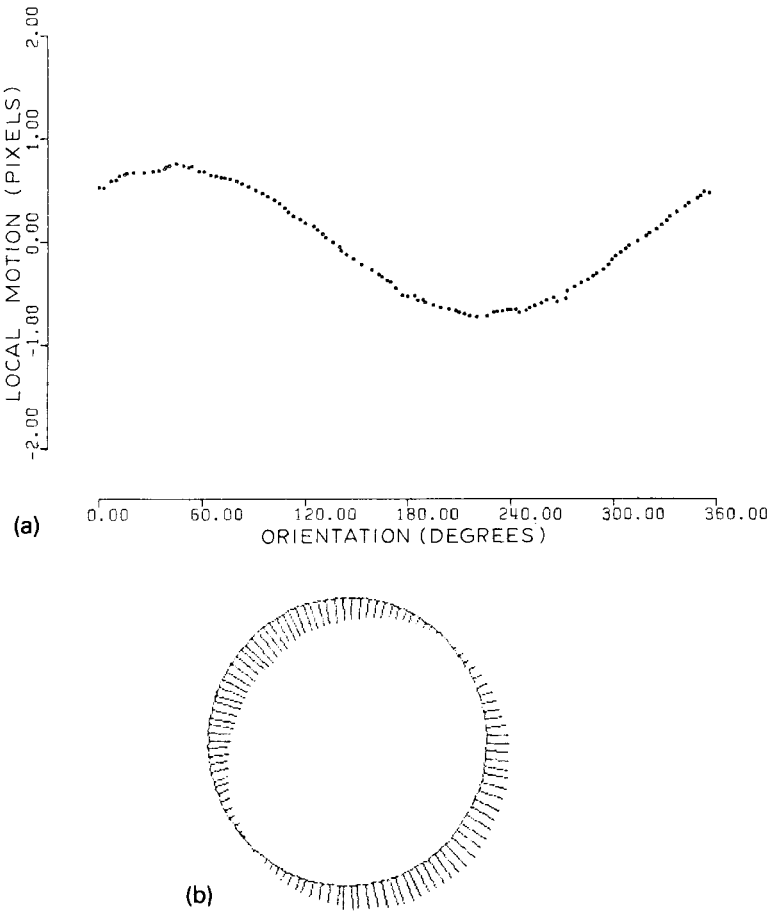

Fig. 5.4. Apparent motion outputs for a disc stimulus of 32 pixels diameter moving by 0.7 pixels between samples (using first order correction as Section 5.5). a) Scatter plot of fragmentary motion versus computed orientation. b) Motion vector and position/orientation dipole diagram.

2.8 pixels in various directions. In addition, for these paired images, the pixel by pixel apparent motion orthogonal to the fragmentary profile orientations was measured and recorded. For general visualisation purposes the high resolution graphical representation described in Chapter 4, Section 4.4 was extended, such that T's could represent the mean sub-pixel position of the fragmentary profiles, the local orientation and the local component of fragmentary motion orthogonal to the local profile.

5.9.1. Spatial factors

Table 5.2 shows the statistics of orientation differences for a number of paired frames of disc images. It will be seen that, by and large, they are very similar to

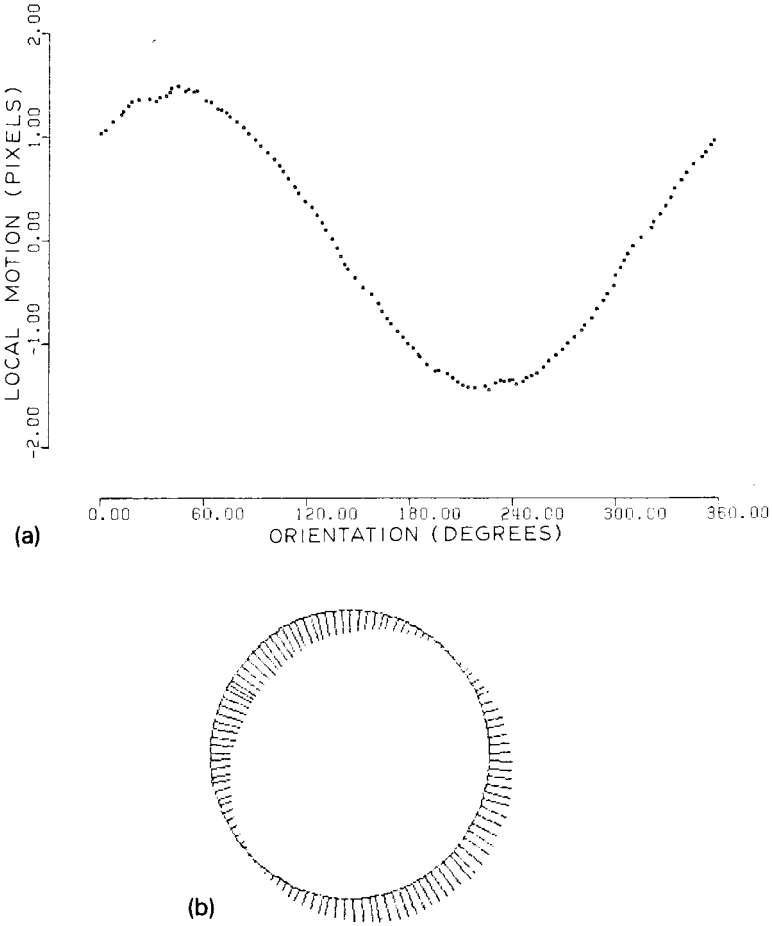

Fig. 5.5. Apparent motion outputs for a disc stimulus of 32 pixels diameter moving by 1.4 pixels between samples (using first order correction as Section 5.5). a) Scatter plot of fragmentary motion versus computed orientation. b) Motion vector and position/orientation dipole diagram.

those for single frame analysis (Chapter 4, Table 4.2), confirming that paired frame analysis is entirely adequate for sensing highly accurate orientations, as are necessary for successful full optical flow analysis (see Chapter 9). Also it is of importance to note that, for total motion of up to at least 2.0 pixels between samples, the orientation differences from paired-frame analysis are at least as stable as from single frame analysis.

The orientations obtained from paired frames, for central zones of edges of squares aligned with primary and secondary axes, were exact, just as for the single frames in Chapter 4. For 10 degree and 20 degree tilted squares, the mean orientations computed from paired frame analysis, for central zones of edges, were

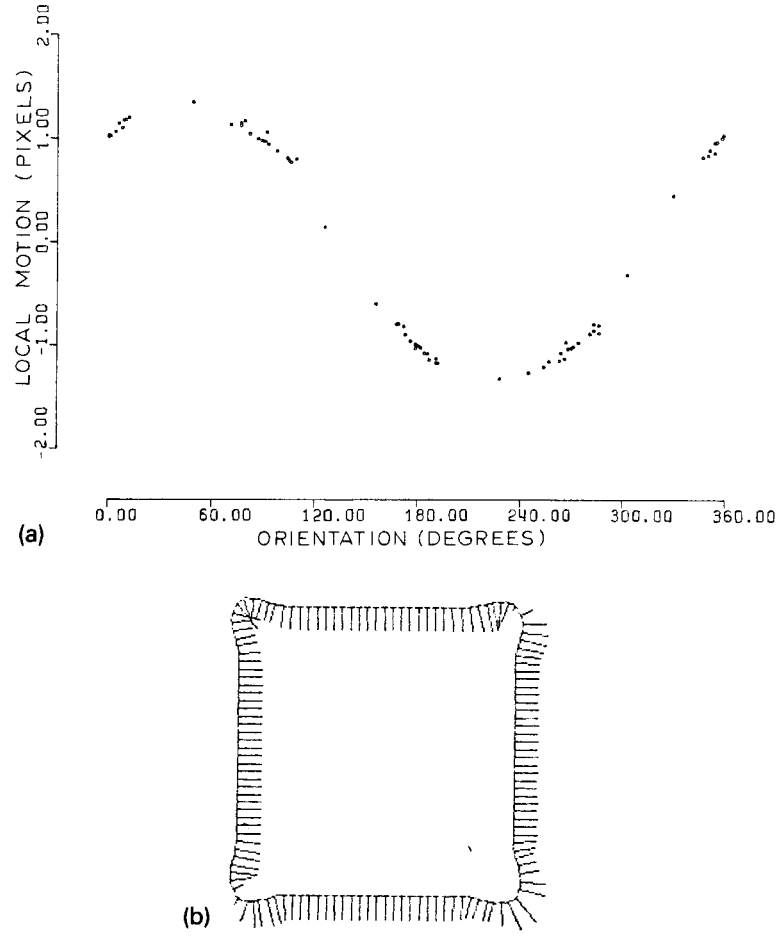

Fig. 5.6. Apparent motion outputs for a square stimulus of 32.25 pixels side length moving by 1.4 pixels between samples towards the bottom righthand corner (using first order correction as Section 5.5). a) Scatter plot of fragmentary motion versus computed orientation. b) Motion vector and position/orientation dipole diagram.

very similar to those for single frame analysis. There was, however, a noticeable *reduction* in pixel to pixel fluctuation, suggesting that there are significant *advantages* in carrying out form analysis from paired frame means in the presence of motion!

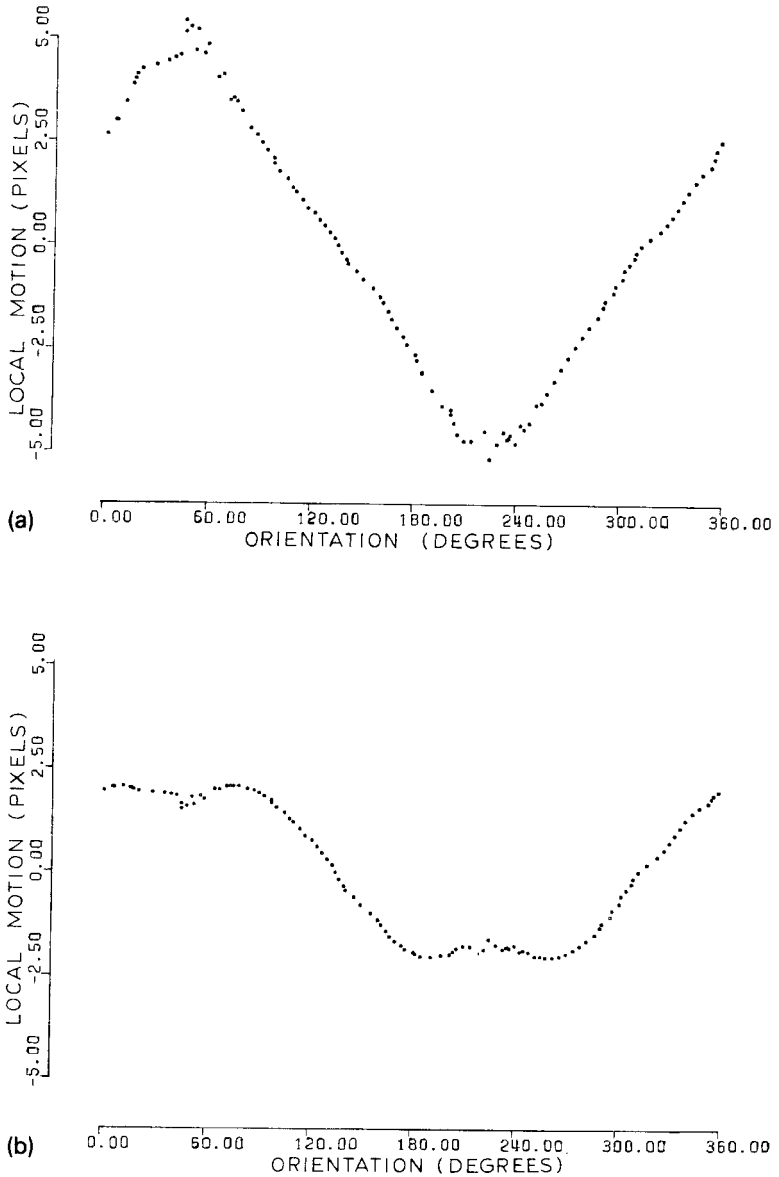

Fig. 5.7. Apparent motion scatter plots for a disc stimulus with 2.8 pixels movement between frames. a) Direct computation, b) With first order correction.

Vernier positional information obtained from paired frame analysis was also very similar to that from individual frames.

5.9.2. *Local motion*

Apparent local translational motion, for any single figure, inevitably results in a variation of measured (orthogonal component of) motion from a positive maximum, through zero, to a negative maximum as one goes round the profile of the object. Thus for any stimulus the ideal progressive apparent motion function around the profile should be a sinusoid with peak values $\pm v$, where v is the input true translation. That this is subjectively well represented for inter-frame motions up to 2 pixels can be seen from Figs. 5.4 to 5.6, where measured apparent fragmentary motion is plotted against measured local vernier orientation of profile fragments. For discs there are, of course, roughly regular samples at all angles (Figs. 5.4 and 5.5), whereas for squares the samples are clustered (Fig. 5.6). The rapid increase of errors when the inter-frame motion is greater than 2 pixels is well illustrated in Fig. 5.7, where a translational motion of 2.8 pixels between frames was studied. It will be seen that, for a simple, direct computation, the peak regions of the sinusoid are greatly extended and rather noisy (Fig. 5.7a), whilst for computations involving 1st order correction, to compensate for errors at a displacement of 2 pixels between samples, there is a pronounced over correction for large motion components (Fig. 5.7b). It is thus important to avoid analyses of motion where components significantly exceed 2 pixels/temporal sampling interval, and rather to employ a progressive blurring and resampling such as progressive shrink (Chapter 11) *before* carrying out motion analysis in such circumstances.

TABLE 5.3.

Summary of the amplitude and phase errors for best fit sinusoids to fragmentary flow data from basic 32 pixel diameter disc stimuli subjected to various motions between frames.

amplitude (pixels)			phase (degrees)		
true	best fit	error	true	best fit	error
0.50	0.508	0.008	0.0	−0.430	−0.430
0.50	0.51	0.01	90.0	90.040	0.040
0.707	0.720	0.013	45.0	45.246	0.246
1.00	1.014	0.014	0.0	0.548	0.548
1.00	1.017	0.017	90.0	89.971	−0.029
1.061	1.079	0.018	45.0	44.742	−0.258
1.414	1.432	0.018	45.0	44.777	−0.023
1.50	1.518	0.018	0.0	0.413	0.413
1.50	1.527	0.027	90.0	89.971	−0.029
2.121	2.049	−0.072	45.0	44.966	−0.034
2.828	2.224	−0.604	45.0	44.897	−0.103

To provide some statistics of the accuracy of local motion computation, a short program was written which could carry out a regression analysis for a best fit sinusoid to a given set of fragmentary flow data, according to the theory given in Appendix 5A. An error analysis about this best fit sinusoid was then performed. The summarised errors on amplitude and phase of the best fit sinusoids, for several disc stimuli, are presented in Table 5.3, where it will be seen that, in general, they are very small, except for the case where the motion between samples was 2.8 pixels. Residual errors of individual points from the best fit sinusoids were next explored visually. It was found that, again except for motion between samples of more than 2 pixels, the errors were substantially random (e.g. Figs. 5.8–5.11). In

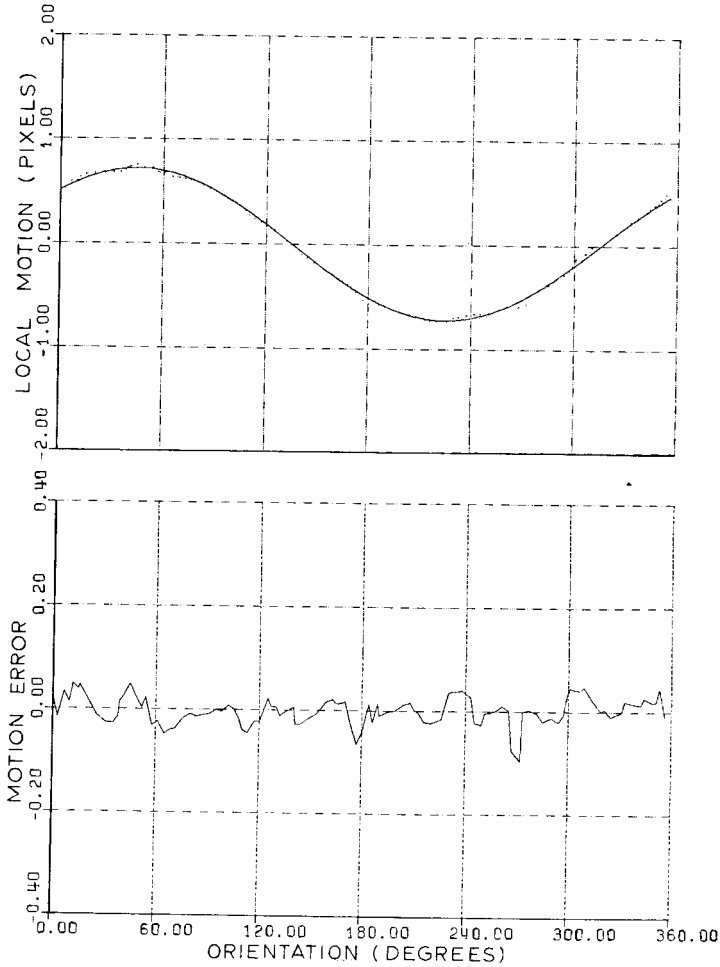

Fig. 5.8. Illustrating the computed best fit sinusoid to the data of Fig. 5.4 and the residual errors from this best fit sinusoid. (Errors are presented as fractions of the amplitude of the sinusoid).

the cases where the motion was in excess of 2 pixels between samples, the evident flattening of the peak regions resulted in a progressively more pronounced cyclic error component (Figs. 5.10 and 5.11). From the visual inspection it was considered satisfactory to determine errors as a root mean square (RMS) error from the best fit sinusoid. The results of such an analysis for various motions are summarised in Table 5.4.

To permit further statistics of local motion to be obtained, a large disc was paired with a disc of radius 1 pixel greater, and with coincident centres. Such a pair of stimuli should provide a set of local motions which are all constant at 1 pixel, representing pure growth. The result of the exercise is shown diagramatically

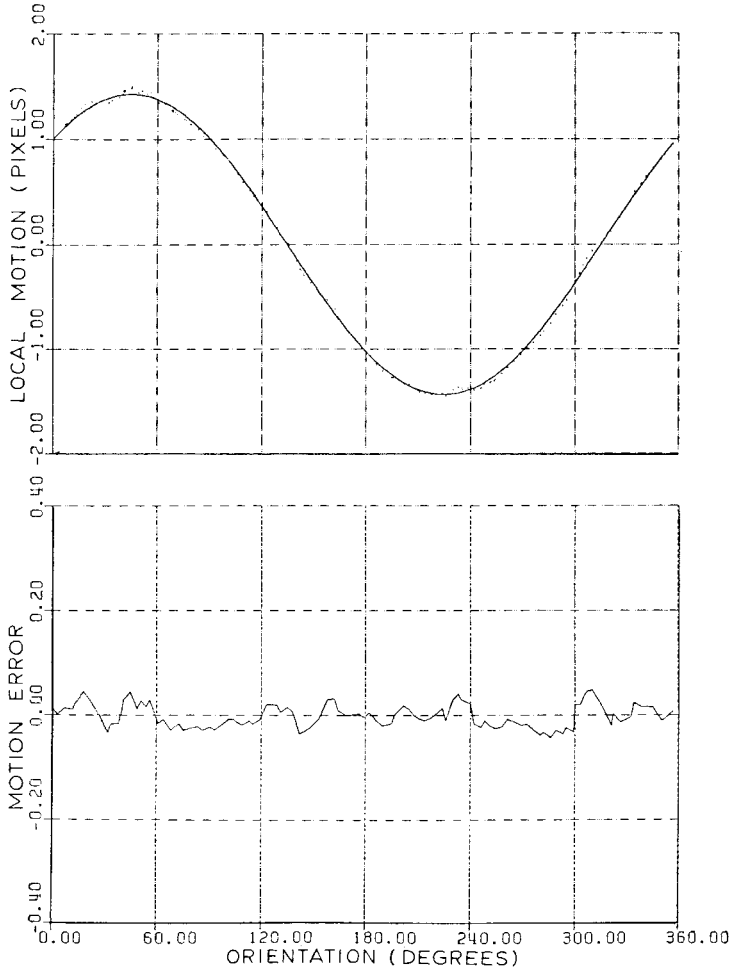

Fig. 5.9. Illustrating the computed best fit sinusoid to the data of Fig. 5.5 and the residual errors from this best fit sinusoid. (Errors are presented as fractions of the amplitude of the sinusoid).

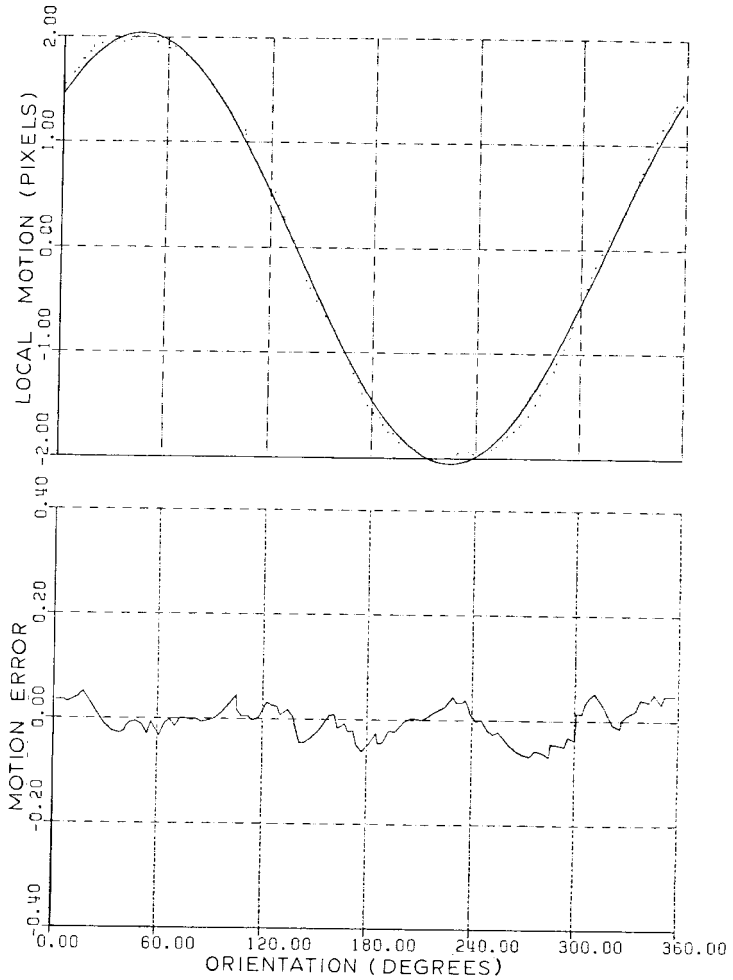

Fig. 5.10. Illustrating the computed best fit sinusoid to the data similar to Fig. 5.5 but with between frame motion of 2.1 pixels, together with the residual errors from this best fit sinusoid. (Errors are presented as fractions of the amplitude of the sinusoid).

in Fig. 5.12, whilst the statistics taken over all data points showed a mean local motion of 1.006 pixels, with a standard deviation of 2.4E-2 pixels. This is considered to be a very satisfactory result, and a powerful demonstration of the practical capability of VISIVE for local motion measurement.

Finally a large disc was paired with one of radius 0.5 pixels larger and with centres displaced by 1.4 pixels. This is a simulation of translational motion in 3 dimensional space, and should result in an apparent motion versus vernier orientation plot which is a sinusoid with a mean which is displaced from zero by 0.5 pixels due to the growth (see Chapter 9 for more detail on this growth theory). The resultant output is shown in Fig. 5.13, where it will be seen that a well defined,

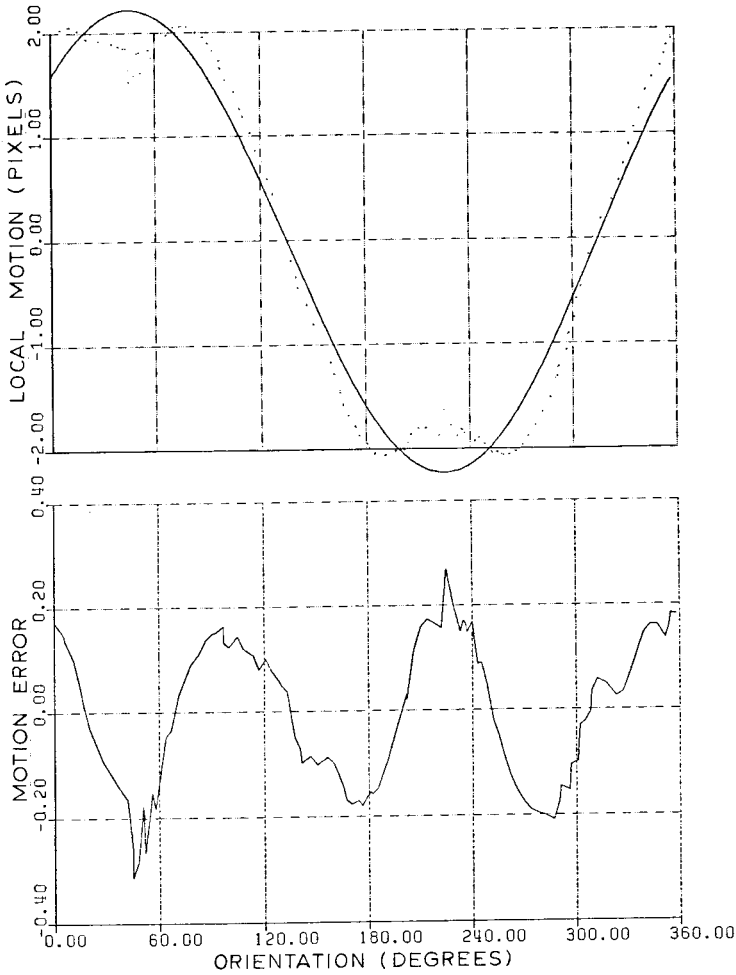

Fig. 5.11. Illustrating the computed best fit sinusoid to the data of Fig. 5.7b and the residual errors from this best fit sinusoid. (Errors are presented as fractions of the amplitude of the sinusoid).

displaced sinusoid is indeed generated. Statistical analysis of this data yielded errors in amplitude and phase of 0.007 pixels and 3.4E-2 degrees respectively, whilst the error on growth was computed to be -0.003 pixels between frames.

5.10. Cueing by first difference data

The *concept* of sensing the component of local motion normal to the local profile, from the spatio-temporal activity around zero crossings of second differences of energy, has been shown to be very simple, yet powerful. Practical application to complex images must, however, be subject to similar problems to

TABLE 5.4.

Summary of the RMS errors from best fit sinusoids for 32 pixel diameter disc stimuli as a function of between frame motion.

Amplitude (pixels)	phase (degrees)	RMS error from best fit sinusoid (fraction of amplitude)
0.50	0	1.89E-2
0.50	90	3.22E-2
0.707	45	2.69E-2
1.00	0	1.36E-2
1.00	90	3.14E-2
1.061	45	2.33E-2
1.41	45	2.16E-2
1.50	0	1.30E-2
1.50	90	2.92E-2
2.12	45	3.17E-2

those already highlighted for edge sensing (Chapter 4) – that is, gross distortions and spurious data points. Additionally, when comparing pairs of frames, the fact that zero crossings may occur other than at the centre of the characteristic ramp of second difference trends may lead to gross deviations from the simple theory. What one *really* requires is to be able to analyse the paired second difference functions *at the centre* of the characteristic ramp, whether this coincides with a zero crossing or not. Now it so happens that, because the simple analyses proposed for local motion analysis involve spatial and temporal *differences*, it does not matter to the computations whether they are carried out at a zero crossing or not. At the same time, from differential algebra, the location of the centre of the characteristic ramp in the second difference function will always coincide with the *peak* of the equivalent *first* difference function. But we have shown, in Chapters 3 and 4, that we can obtain both first and second difference distributions from the *same* initial analysis of a single image frame. By definition, the same is true for initial analysis of the *mean* image from a pair of frames in a time sequence. Hence, in analysing a pair of frames for motion data, it is readily possible to derive the *location* of centres of characteristic second difference ramps from *first* difference edge sensing. Then this may be used for cueing the pixels at which to carry out local motion analysis, using the algorithms described in this Chapter.

5.11. Conclusions

A theoretical appraisal of the potential errors in measurement of magnitude and orientation of local fragmentary motion by means of VISIVE has been presented, together with a study of practical errors of both local motion and vernier position and orientation from paired frame analysis. As a result of the theoretical study it was possible to refine some of the algorithms within VISIVE without

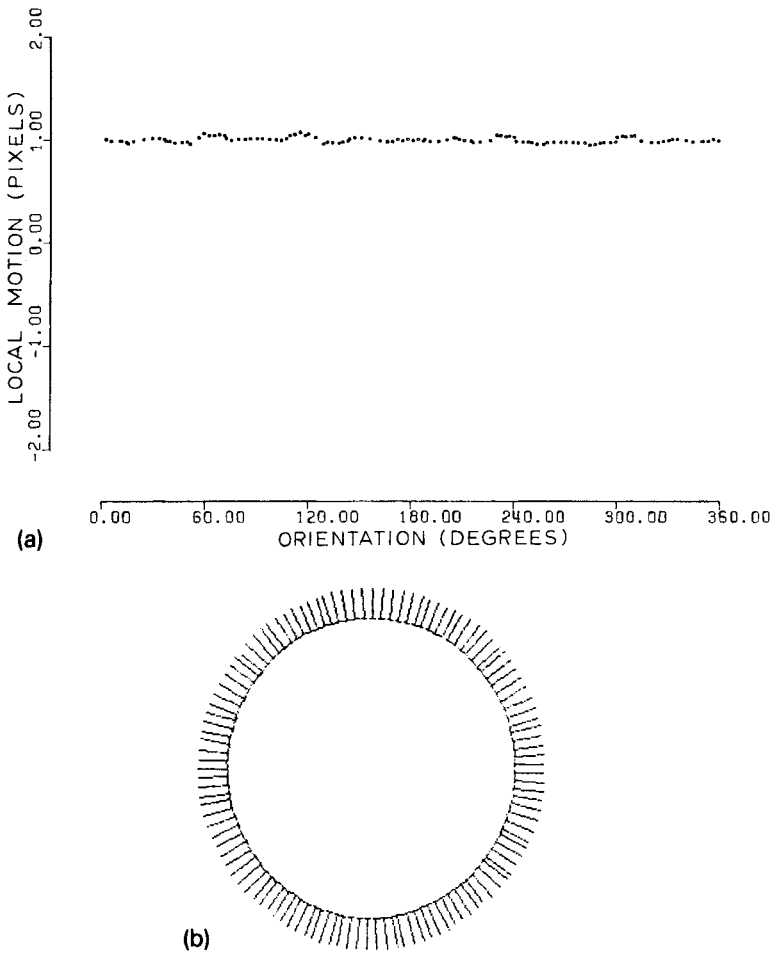

Fig. 5.12. Apparent motion outputs for a disc stimulus growing in radius by 1 pixel between time samples and with no translational motion. a) Scatter plot of fragmentary motion versus computed orientation. b) Motion vector and position/orientation dipole diagram.

increasing computational complexity, and thereby to reduce theoretical errors to a very low level. The practical study largely confirmed the capability of VISIVE to produce practical results with errors of basically the levels predicted from theory. Of particular interest is the finding that form analysis from paired frames with local motion is rather more stable than single frame analysis and appears to lose nothing, provided that motion between frames is less than 2 pixels. It has been noted that addressing problems characteristic of zero-crossing analysis in complex scenes can be readily circumvented by cueing locations for motion analysis from edge sensing using *first* differences. Further practical studies to explore the effects of noise on the presently observed capabilities are presented in Chapter 6. Theory

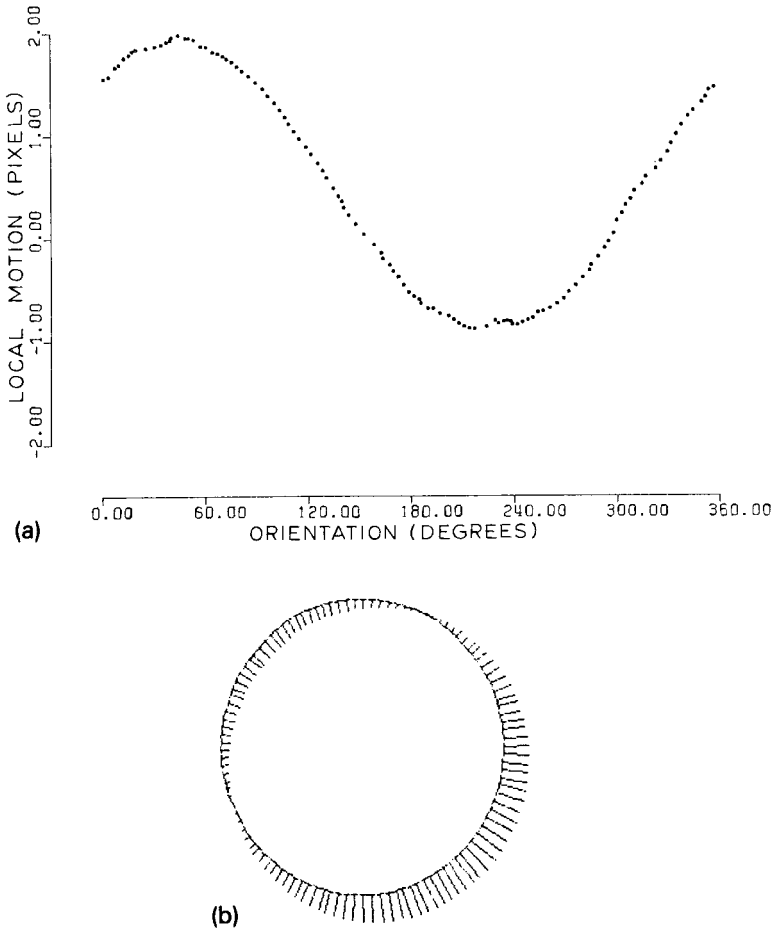

Fig. 5.13. Apparent motion outputs for a disc stimulus growing in radius by 0.5 pixels between time samples and translating by 1.4 pixels between frames. a) Scatter plot of fragmentary motion versus computed orientation. b) Motion vector and position/orientation dipole diagram.

and practical studies for a variety of optical flow analyses in real scenes are presented in Chapter 9.

Appendix 5A. Least squares fitting of a sine wave to sampled sinusoidal data

Suppose we have a large number M of data points, roughly equally spread over a full period of a sinusoid (i.e. $0 < \theta_i \leq 360°$ where θ_i is the ith sample angle). Then the data can be described by

$$Y_i = A \cos(\theta_i + \phi) + B + N_i \tag{5.A1}$$

where A and ϕ are the amplitude and phase of the underlying sine wave, B is the mean of the sine wave (which may in general be non-zero) and N_i is the noise (assumed random) on the ith sample. For a least squares fit we require to minimise

$$\sigma^2 = \sum [Y_i - B - A \cos(\theta_i + \phi)]^2 \tag{5.A2}$$

In turn this requires that

$$\partial\sigma^2/\partial A = -\sum [2 \cos(\theta_i + \phi)[Y_i - B - A \cos(\theta_i + \phi)]] = 0 \tag{5.A3}$$

$$\partial\sigma^2/\partial\phi = \sum [2A \sin(\theta_i + \phi)[Y_i - B - A \cos(\theta_i + \phi)]] = 0 \tag{5.A4}$$

$$\partial\sigma^2/\partial B = -\sum [Y_i - B - A \cos(\theta_i + \phi)] = 0 \tag{5.A5}$$

Therefore

$$\sum [(Y_i - B) \cos(\theta_i + \phi)] = A \sum \cos^2(\theta_i + \phi) \tag{5.A6}$$

$$\sum [(Y_i - B) \sin(\theta_i + \phi)] = A \sum [\cos(\theta_i + \phi) \sin(\theta_i + \phi)] \tag{5.A7}$$

$$\sum (Y_i - B) = A \sum \cos(\theta_i + \phi) \tag{5.A8}$$

But for a large number of roughly equally spaced samples as assumed;

$$\sum \cos^2(\theta_i + \phi) \to 0.5M$$

$$\sum [\cos(\theta_i + \phi) \sin(\theta_i + \phi)] \to 0$$

and

$$\sum \cos(\theta_i + \phi) \to 0$$

Therefore, from equation 5.A8,

$$\sum Y_i = \sum B = M * B \tag{5.A9}$$

Meanwhile equation 5.A7 becomes

$$\sum (Y_i - B) \sin(\theta_i + \phi) = 0$$

or

$$\cos \phi \sum [(Y_i - B) \sin \theta_i] + \sin \phi \sum [(Y_i - B) \cos \theta_i] = 0 \tag{5.A10}$$

where B is known from equation 5.A8. On the face of it one can solve this equation for tan ϕ. However, this produces an ambiguity in ϕ. But from equation 5.A6

$$\sum (Y_i - B) \cos(\theta_i + \phi) = A * M/2$$

or

$$\cos \phi \sum [(Y_i - B) \cos \theta_i] - \sin \phi \sum [(Y_i - B) \sin \theta_i] = A * M/2 \qquad (5.A11)$$

In equations 5.A10 and 5.A11 let $\Sigma[(Y_i - B) \cos \theta_i] = P$ and let $\Sigma[(Y_i - B) \sin \theta_i] = Q$. Then

$$Q \cos \phi + P \sin \phi = 0 \qquad (5.A12)$$

$$P \cos \phi - Q \sin \phi = A * M/2 \qquad (5.A13)$$

Multiplying 5.A12 by sin ϕ and 5.A13 by cos ϕ,

$$Q \cos \phi \sin \phi + P \sin^2 \phi = 0$$

$$-Q \cos \phi \sin \phi + P \cos^2 \phi = (A * M/2) \cos \phi$$

By summation, since $(\sin^2 \phi + \cos^2 \phi) = 1$,

$$P = (A * M/2) \cos \phi \quad \text{or} \quad A \cos \phi = 2P/M \qquad (5.A14)$$

Similarly, by multiplying 5.A12 by cos ϕ and 5.A13 by sin ϕ and subtracting,

$$A \sin \phi = -2Q/M \qquad (5.A15)$$

Then, squaring 5.A14 and 5.A15 and summing,

$$A = (2/M)\text{SQRT}(P^2 + Q^2) \qquad (5.A16)$$

Taking A as the positive root of equation 5.A16 (since amplitude is a scalar quantity) and substituting in equation 5.A14

$$\cos \phi = 2P/(M * A) \qquad (5.A17)$$

yielding an unambiguous value for ϕ.

References

[5.1] Gonzalez R.C. (1987), 'Digital Image Processing', Addison-Wesley.

[5.2] Holliday I.E. and Ruddock K.H. (1983), 'Two spatio-temporal filters in human vision', Biol. Cybern., *47*, 173.

[5.3] Kulikowski J.J. (1979), 'Neural stages in visual signal processing', in 'Search and the Human Observer' (Eds. J.N. Clare and M.A. Sinclair), Taylor and Francis, London.

[5.4] Puzey N.J. (1984), 'A correspondence method of solution for optical flow using VISIVE', B.Ae.D. SRC Report JS10019.

[5.5] King-Smith P.E. (1978), 'Visual sensitivity to moving stimuli: data and theory', in 'Visual Psychophysics: its Physiological Basis' (Eds. J.C. Armington, J.J. Krauskopf and B.R. Wooten), Academic Press, New York.

[5.6] Ballard D.H. and Brown C.M. (1982), 'Computer Vision', Prentice Hall.

[5.7] Buxton B.F. and Buxton H. (1983), 'Monocular depth perception from optical flow by space time signal processing', Proc. R. Soc. B., *218*, 27.

[5.8] Overington I. (1983), 'Computer simulation of preperceptual processing of form', Proc. of the SPIE, Vol. 397, page 73.

[5.9] Abrahamowitz M. and Stegun I.A. (1972), 'Handbook of Mathematical Functions', Dover Publications Inc., New York.