Estimating Motion In Noisy, Textured Images: Optical Flow In Medical Ultrasound

David H. Cooper, Jim Graham.
Department of Medical Biophysics
University of Manchester
Oxford Rd., Manchester M13 9PT
England, UK

Abstract

Motion is often an important feature in medical images and accurate motion estimation may have direct diagnostic benefits. There is a plethora of methods in the literature using tracking and optical flow which may be used for motion estimation. In addressing the particular application of estimating peristaltic motion in trans−abdominal ultrasound images we have adopted Heeger's method for optical flow. Heeger's formulation as an integral rather than differential approach offers significant advantages in this case. There are however a number of disadvantages of this method, primarily the absence of a confidence measure associated with the velocity estimate, and the inability to distinguish full from normal flow (the aperture problem). We present a principled approach to estimating confidence and demonstrate its application in our problem domain. In so doing we address the issue of establishing ground truth for motion in medical images.

1. Introduction

Traditionally, investigation of patients with suspected inflammatory bowel disease has been by small bowel barium radiology. This procedure results in one of the highest exposures to radiation dose. The National Radiation Protection Board estimates the number of tumours generated ultimately by exposure to this level of dose to be up to 20 out of 100,000 of the population examined. It is estimated that 50−60% of barium investigations currently performed could be avoided. We are interested in the use of transabdominal ultrasound as the *initial* radiological investigation in patients with suspected or recurrent bowel disease. The appearance of inflammatory bowel disease in ultrasound imaging is well recognised. The clinical features are thickening of the bowel wall and reduced or absent peristalsis. When peristalsis is near normal it can be difficult to decide whether a loop of bowel is normal or not. There are many equivocal cases. We seek to improve the the sensitivity of transabdominal ultrasound by attempting to measure peristaltic motion in ultrasound images[1].

In this paper we consider optical flow methods appropriate to this end, and stress the role that measures of confidence in velocity estimates play in maximising robustness and accuracy.

1.1 Image Properties and Optic Flow

In ultrasound images normal and equivocal cases exhibit amorphous but slightly textured appearance with little or no structure. When bowel structure is apparent it shows

BMVC 1996 doi:10.5244/C.10.9

highly variable, constantly changing shape. A typical frame from an ultrasound sequence is shown in Fig. 1.

Fig. 1. A typical trans−abdominal ultrasound scan of the small bowel. The well−defined dark area is a fluid−filled cyst.

The key features of these sequences are (a) poor or absent structure (b) textured signal (c) obvious motion (d) speckle. The nature of the images is such that estimation of motion by attempting to locate structure is exceedingly difficult. We have therefore chosen to investigate the application of optic flow methods to obtain a velocity field. We seek a method that will be successful on textured signals and can show some robustness to noise. The speckle caused by interference of the sound waves is systematic, correlated noise. Modelling speckle is the subject of research [12] but because of the complexity of the conducting media (skin, muscle, blood, fat, water, etc) the patterns are extremely complex and attempting to model them is beyond the scope of this project.

1.2 Ground Truth and Measures of Confidence

The systematic evaluation of flow methods requires that ground truth is known so that behaviour can be measured. It is often difficult (sometimes impossible) to obtain truth values for real data. In the interpretation of medical images ground truth is usually established by asking a number of "experts" familiar with the application to annotate real data with truth labels so that machine analysis can be compared to human analysis and that intra−expert variation can be accommodated to some extent. This is difficult where motion is involved as ultimately the velocity of individual pixels has to be specified.

We have used synthetic image sequences, for which ground truth is known, to assess the accuracy of velocity estimates. However, our experience elsewhere [7,8] indicates that the performance of algorithms on synthetic data does not necessarily match performance on real data. We have used visualisation techniques to assess accuracy subjectively but in the absence of ground truth it is essential that we have some measure of confidence associated with a velocity estimate. In this paper we concentrate on these problems and develop some possibilities for use with the method due to D. Heeger [5], which we believe may be appropriate for our application. We give our reasons below.

1.3 Optic Flow Methodologies

The literature on optical flow determination is large. An introduction to the topic is given in Horn [9] and more detailed discussions can be found in many of the references eg [2,11]. There are three main groups: differential, matching and spectral methods.

We have discarded differential methods for our application because of the difficulty of obtaining derivatives on textured ultrasound data [10]. We have discarded matching methods because of the lack of features which can possibly be tracked and the multiple responses that correlation methods give on textured data [2].

Spectral Methods

Spectral methods are based on certain properties of the Fourier transform (FT) of moving signals. Suppose a 2D signal *f(x,y)* has Fourier transform $F(\omega_x, \omega_y)$. The signal *f(x,y)* moving with constant translational velocity $\mathbf{v} = (u,v)$ is $f(x-ut, y-vt)$ and this has FT $F(\omega_x, \omega_y)\,\delta(\omega_t + u\omega_x + v\omega_y)$ where $\delta(s)$ is the delta function [3]. The spectrum of the moving signal therefore resides entirely within a plane

$$u\omega_x + v\omega_y + \omega_t = 0 \qquad \text{.....(1)}$$

To estimate the velocity *(u,v)* it is sufficient to estimate the orientation of this plane. The method due to Heeger [4,5] estimates the plane orientation by sampling the local power spectrum using Gabor filters tuned to various spatio−temporal frequencies and comparing their actual output with a predicted output. It is possible to estimate the velocity $\mathbf{v}$ by minimising the differences between the predictions and the measured data. The sampling is performed using integrals implemented as convolutions and so has the advantage that it can work on moving textures.

2. Heeger's Method

The power spectrum of a signal can be sampled by convolution of the signal in the spatial domain. This can be achieved (eg in 1D) by convolving a signal with functions $\sin(ax)\,f(x)$ and $\cos(ax)\,f(x)$ where $f(x)$ has Fourier transform $F(\omega_x)$, which is the sampling filter. The *cos* (even) function samples the real part of the spectrum and the *sin* (odd) function samples the imaginary part. Squaring and adding the outputs of these two filters results in the power sampled by the filter *F*. When *f* is a Gaussian, these so−called *quadrature pairs* are Gabor filters. In 3D, An odd (sine) 3D Gabor filter has the form

$$g(x,y,t) = \left[\frac{1}{\sqrt{2}\,\pi^{3/2}\sigma_x\sigma_y\sigma_t}\exp - \left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{t^2}{2\sigma_t^2}\right)\right] \sin(2\pi\omega_{x0}x + 2\pi\omega_{y0}y + 2\pi\omega_{t0}t)$$

This filter has a power response which is the sum of two gaussians centred at $(\omega_{x0}, \omega_{y0}, \omega_{t0})$ and $(-\omega_{x0}, -\omega_{y0}, -\omega_{t0})$ thus:

$$\begin{aligned} G(\omega_x,\omega_y,\omega_t) = {} & (1/4)\exp\{ - 4\pi^2[\sigma_x^2(\omega_x - \omega_{x0})^2 + \sigma_y^2(\omega_y - \omega_{y0})^2 + \sigma_t^2(\omega_t - \omega_{t0})^2]\} \\ & + (1/4)\exp\{ - 4\pi^2[\sigma_x^2(\omega_x + \omega_{x0})^2 + \sigma_y^2(\omega_y + \omega_{y0})^2 + \sigma_t^2(\omega_t + \omega_{t0})^2]\} \end{aligned}$$

The method makes the assumption that the moving signal is a random texture with uniform spectral content. Under this assumption all of the power is uniformly distributed along the plane $u\omega_x + v\omega_y + \omega_t = 0$. The predicted power sample is thus proportional to the integral of $G(\omega_x, \omega_y, \omega_t)$ subject to the constraint $u\omega_x + v\omega_y + \omega_t = 0$, that is:

$$P(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(\omega_x,\ \omega_y,\ -u\omega_x - v\omega_y)\ d\omega_x d\omega_y \qquad \text{.....(2)}$$

This integral can be solved analytically [4]. This predicted energy is a function *P(u,v)* depending on the unknown flow *(u,v)*. An error function *E(u,v)* is formed which

measures the squared sum of differences between the normalised predictions and the normalised measured energies. The location of the minimum of this function gives the flow estimate. Heeger suggests normalising each spatially−oriented filter group separately to accommodate possible differences in the relative components of spatial orientations in the signal. We have found this to be neccesary because real signals do not possess uniform spectral content.

The spatial organisation of the sampling pattern as suggested by Heeger is shown in Fig. 2. This pattern is duplicated at $\omega_t = 0.25, 0.00, -0.25$ Hz, making 12 samples in all. The filters are centred at a radial frequency $\omega_r = 0.25$ Hz and the filter standard deviations are $\sigma_{\omega_x} = \sigma_{\omega_y} = 0.0398$ Hz and $\sigma_{\omega_t} = 0.0159$ Hz. This corresponds to spatial standard deviations of 4 pixels and a temporal deviation of 1 pixel respectively. Heeger does not give any particular reason for choosing these values.

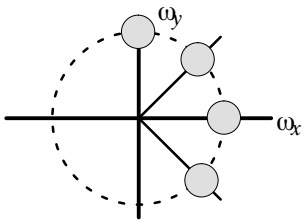


Fig. 2. The spatial sampling pattern suggested by Heeger, as if viewed along the ω_t axis. This arrangement is duplicated at temporal positions −0.25, 0, 0.25 Hz.

2.1 Advantages and Problems with Heeger's Method

Heeger has demonstrated that the method works well on textured signals. It does this because it estimates integrals rather than derivatives. This benefit is potentially so useful that we have invested some time in overcoming the disadvantages. They are (a) Heeger does not provide a suitable measure of confidence, (b) the method cannot readily distinguish full from normal flow (c) real signals do not possess the uniform spectral properties that are assumed in order to calculate the power predictions (d) speed: sampling followed by an optimisation at every point isn't exactly fast. Barron [2] has reported problems with the minimisation of $E(u,v)$. We have used a method due to Powell [6] which works well for sum−of−squares and sum−of−absolute deviation error functions. In the remainder of this paper we describe approaches to overcoming some of these difficulties.

3. Accuracy

3.1 Synthetic vs Real Data

Heeger's method makes an assumption that the data is a moving random texture, so it makes sense to verify that the method works using synthetic random data. Section 4.3 describes the evaluation of a confidence measure using synthetic data. In general, however, we are not particularly convinced of the usefulness of synthetic data for predicting behaviour with real data. We have found that methods which perform optimally on synthetic data are not necessarily optimal on real data [7,8].

We would like to be able to assess accuracy on our real data but in the absence of ground truth we are reduced to using subjective visualisation. To reassure ourselves that the method will work on real data we have implemented Heeger's method for (x,t) sequences. Fig. 3. shows an (x,t) sequence taken from an echocardiac cycle. At each time frame a streamline "marker" is updated by the calculated flow displacement, and the streamline thus formed is superimposed in the original (x,t) sequence. The velocity

data has been used regardless of whether it is deemed reliable or not. Where structure is visible the velocity estimates are clearly fairly accurate. Note the random−like displacements in regions of low contrast where noise dominates. This example serves to demonstrate the importance of rejecting estimates that are unreliable.

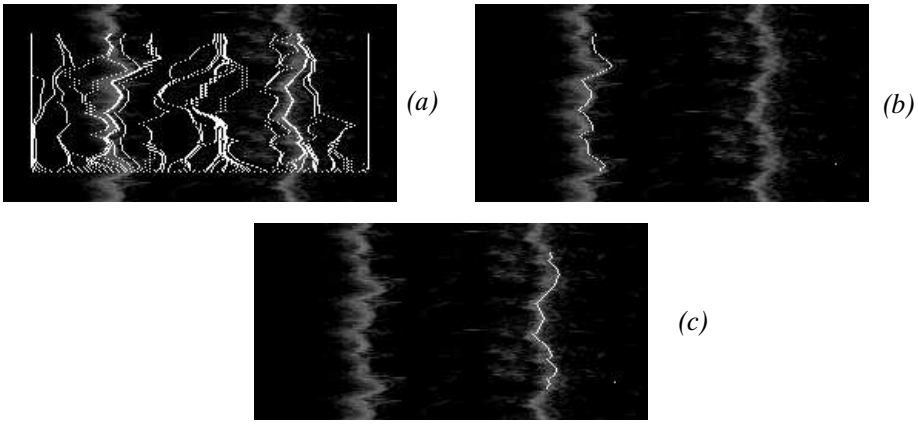


Fig. 3. Streamlines calculated from Heeger estimations for an (x,t) slice of echocardiac data. The horizontal axis is distance, x. the vertical axis is time, t. (a) all streamlines arising from every 4th pixel. (b), (c) two streamlines in regions of signal showing the calculated flow near the moving heart wall.

We have performed similar visualisation of streamlines for the *(x,y,t)* case on our real data sequences. Subjective assessment indicates comparable performance to the *(x,t)* case. Unfortunately, these visualisations are difficult to represent in a paper because projections do not show any well−defined structure against which to view the streamlines.

4. Confidence Estimates

4.1 Heeger's Uncertainty Measure

In [5] Heeger develops an uncertainty measure U based on a Fisher information matrix of the form

$$U = J^T(\hat{v})V^{-1}J(\hat{v})$$

where J is the Jacobian matrix of partial derivatives evaluated at the velocity estimate $\hat{v}$ corresponding to the mapping R: $\{p_1 .. p_N\} => \boldsymbol{v}$, V is a diagonal matrix of variances of the least−square residuals, and $\{p_i\}$ are the power samples. His results using test sequences of synthetic random textures indicate an almost linear relationship between U and $|\boldsymbol{v}|$ and between the absolute error in $|\boldsymbol{v}|$ and $|\boldsymbol{v}|$ itself. Thus U or $|\boldsymbol{v}|$ itself could be used to reject uncertain estimates according to this analysis.

The presence of J in this measure suggests that U is responding to the sensitivity of the method. If this is the case then such a measure is of little value in the case of real data because it is not known whether the assumption of a random patch, ie uniform spectral response, has been violated. We would prefer a goodness−of−fit measure independent of sensitivity which will allow us to reject all estimates of $\boldsymbol{v}$ (even with small $|\boldsymbol{v}|$) which are invalid because the constraints of the model have been violated.

4.2 The Sensitivity of Heeger's Method

The angles θ,ϕ subtended by the plane of power and the ω_{xt}, ω_{yt} planes are related to the velocity components by the relations $u=-tan(\theta)$, $v=-tan(\phi)$. Any method which estimates the plane angles θ,ϕ to derive $\boldsymbol{v}$. If $\theta > \pi/4$, errors in the angle estimates are magnified by a factor of $(1+|u|^2)$. Minimisation of the error surface $E(u,v)$ represents a function which maps the N (=12) power samples $\{p_i\}$ to (u,v) space. The partial derivatives $\frac{\partial u}{\partial p_i}, \frac{\partial v}{\partial p_i}$ express the sensitivity of the estimate as a function of the power samples. It turns out that the larger the estimate of $\boldsymbol{v}$ the larger these derivatives are. Typical value ranges for the partial derivatives using a moving random patch are:

```
|v| in ( 0.0 ..  0.5)        pd's in (0.001 .. 0.01)
|v| in ( 0.5 ..  5.0)        pd's in (0.01 ..  3.0)
|v| in (10.0 .. 50.0)        pd's in (0.01 .. 60.0)
```

Thus it seems that any solution $|\boldsymbol{v}| >> 1$ is unstable, and therefore potentially unreliable. In practice, image pyramiding is used to reduce displacements and the estimate with the smallest value of $|\boldsymbol{v}|$ is chosen. However, an estimate with $|\boldsymbol{v}| < 1$ is not necessarily reliable.

4.3 Goodness−of−Fit using Chi−Square Tests

To estimate velocities we minimise a quantity of the form

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - P(u,v)}{\sigma_i} \right)^2 \qquad \text{.....(3)}$$

where the y_i are the power estimates and the $P(u,v)$ is the power prediction. The value of χ^2 for any instance provides a goodness−of−fit measure. In problems where the function P is linear in the model parameters **v,** and the distribution of the residual differences are Gaussian with variance σ_i^2 then the distribution of χ^2 can be calculated analytically and is the chi−square distribution for $(N-M)$ degrees of freedom, where M is the number of model parameters. The probability that a value of χ^2 exceeds a given value can easily be calculated (a $P-value$) and if this probability is small for the χ^2at the minimum then there are grounds for rejecting the estimate as a bad fit to the model.

In our case P is not linear in (u,v). According to Press et al [6] it is common and not too wrong to assume that the chi−square distribution holds for models which are strictly non−linear. Since P is differentiable near to the minimum χ^2 then it is locally linear there. Inspection of the $P(u,v)$ surfaces verifies that this is the case. See Fig. 4.

The application of this test assumes that the σ_i^2 are known. We have estimated the σ_i^2 for the set of filter residuals by collecting statistics using an image of random numbers moving at velocity $\boldsymbol{v} = (0,1)$ pixels/frame. There is a technical problem here in that the σ_i^2 values are needed to estimate $\boldsymbol{v}$ before the residuals can be calculated. We assume a value of $\sigma_i^2=1$ in order to do this. Although this does not provide an independent assessment of these parameters, the final result is better than no estimate at all. Fig 5.(a) shows one frame from such a sequence. Fig 5.(b) is a scattergram showing the correlation between the theoretical P−value and Heeger estimates of $|\boldsymbol{v}|$ for a slice from the same test sequence. This result is encouraging because the high estimates all exhibit tiny P−values. This example illustrates a small problem, however. This random image will (by definition) contain regions where the assumption of uniform spec-

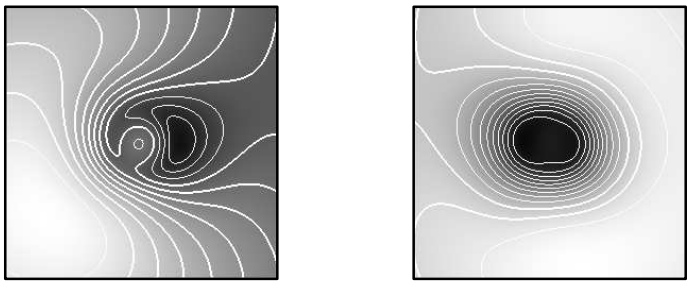


Fig. 4. Examples of E(u,v) surfaces for cases where $\mathbf{v}=(0.57, 0.02)$ *and* $(-0.37, 0.04)$*. The axis range of these images is* $-2 <= u,v <= 2$ *pixels/frame.*

tral response is violated. It is therefore not surprising that there is a spread of v−estimates. However, this also means that the estimates of the σ_i^2 will be biased because the test data does not exactly correspond to the model everywhere.

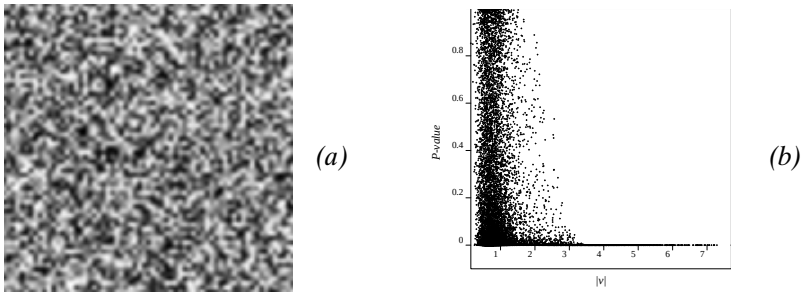


Fig. 5. (a) A frame from the random sequence used to estimate residual deviations. (b) the scatterplot of theoretical P−values versus $|\mathbf{v}|$ *for Heeger estimates on the same sequence. The ground truth for this sequence is* $\mathbf{v}=(0,1)$

The statistical variance that we are measuring is not an error in a measuring process, as is usually the case. In our case the power−measuring process is deterministic. The variation is due to the statistical variance in the image data, which we are assuming is Gaussian for the P−test to be valid. Examples of these distributions for three of the Gabor filter residuals are given in Fig. 6.

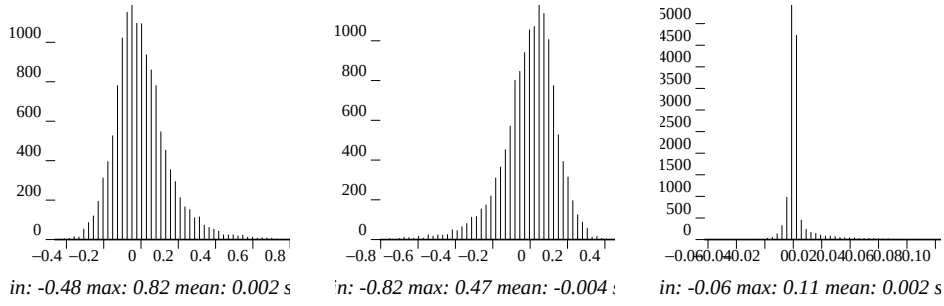


Fig. 6. The distributions of the residuals the three Gabor filters tuned to frequencies $(w_x,0,-w_t)$*,* $(w_x,0,0)$ *and* $(w_x,0,w_t)$ *for* $w_x,w_t=0.15$ *Hz.*

The response of a given Gabor filter varies as the velocity of a 2D signal changes (they are tuned to points in $(\omega_x, \omega_y, \omega_t)$ space) and so the variance statistics are really a function $\sigma_i(u,v)$ of velocity. We have estimated these functions by sampling velocity space

over the range $|\boldsymbol{v}| <= 1$ and using bilinear interpolation to evaluate P−values. Fig 7. shows a scattergram showing the correlation between the theoretical P−value and Heeger $|\boldsymbol{v}|$ estimates for a slice from a real sequence using the table of values for $\sigma_i(u, v)$. Fig 7.(a) shows the data from all of the velocity range, and Fig. 7.(b) has the velocity range truncated to $|\boldsymbol{v}| < 2$ and the P−value range truncated to .001. The behaviour is similar to the synthetic case and illustrates the tiny P−values associated with the large |v| estimates. It also illustrates that there are small $|\boldsymbol{v}|$ estimates with poor fit values.

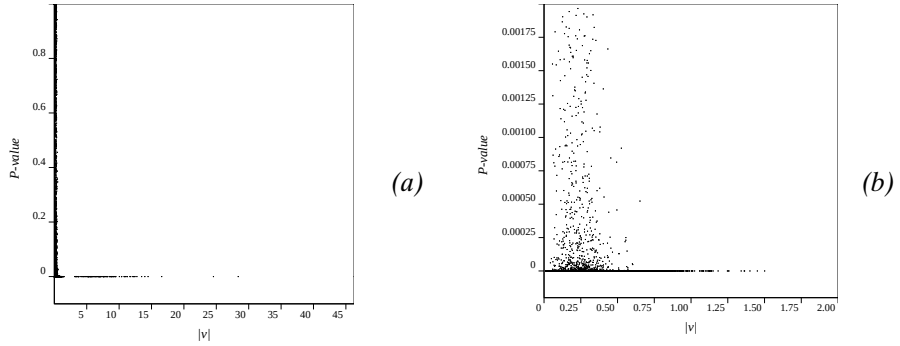


Fig. 7. Scatter plots of theoretical P−values and estimated $|\boldsymbol{v}|$ for a slice of real ultrasound data (a) all data; the range of $|v|$ in these plots is determined by the data, showing the max estimate of about 45 pixels/frame. (b) truncated data

Correlation with Errors

We have applied the P−test to the synthetic texture sequence to establish whether low P−values and errors are related as we would wish. The result is illustrated in Fig. 8. All the large errors are associated with tiny P−values. There are cases also where small errors are associated with poor fits. These cases suggest that poor fits give rise to a range of random velocity estimates, and that rejection using P−values makes more sense than rejection by assuming that all small $|\boldsymbol{v}|$ estimates are good.

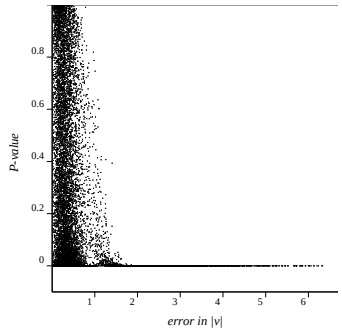
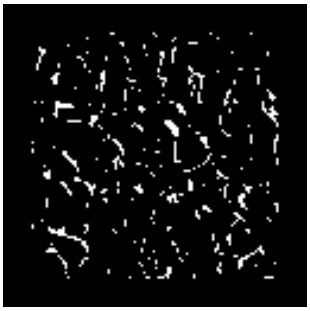


Fig. 8. A scatterplot of P−value against error in $|v|$ for the synthetic texture sequence moving with velocity (0,1) pixels/frame.

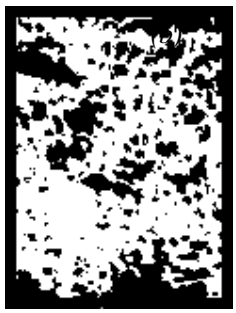
Rejection

It is not obvious how to set thresholds on P−values. It is common to accept quite low P−values (around P=0.001) when it is known that some of the assumptions are violated, but that it is wise to identify sources of error leading to the values. The distribu-

tions illustrated in Fig. 6. do not appear to be particularly Gaussian (although we have not performed any tests to date), and so a low value for rejection is possibly justified. A threshold of P<0.0001 is clearly large enough to reject the unlikely large $|\boldsymbol{v}|$ estimates, and results in densities of accepted data as shown in Fig. 9. For real data about 37% of the data is accepted. For the test data the figure is 93%. This latter figure is probably reasonable given that the random image will contain neighbourhoods where the aperture problem holds. The important point about this test is that it identifies estimates for which we believe full flow conditions exist *and* we can have confidence in the estimate.



(a)



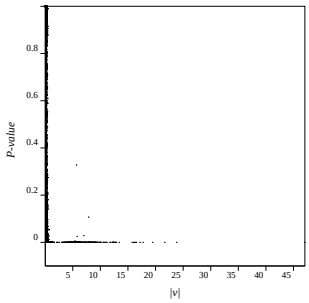
(b)

Fig. 9. Density maps for goodness−of−fit test, white=rejected, black=accepted (a) texture sequence, 93% accepted (b) real; data, 37% accepted

4.4 Addressing the Statistical Assumption

The results for real data have used statistics generated from a random image as depicted in Fig. 5. This data is not accurate for the ultrasound sequences. A better statistical model would be generated if we took some typical frames of real data and computed the statistics of the residues using this data. If we do this, but still assume that the power predictions are analytically derived from a uniform random texture, we obtain the result in Fig. 10. A marked increase in the acceptance level (66%) of the velocity estimates is observed. The correspondence of extremely large $|\boldsymbol{v}|$ values and tiny P−values is still apparent.

(a)



(b)

Fig. 10. (a) Density maps for a slice of real data using statistics computed from some other slices in the sequence (b) a scatterplot of P−value and $|v|$ for this slice.

5. Conclusions

We consider Heeger's method to be useful for ultrasound sequences of the small bowel because it does not use gradient estimation and is potentially successful on textured sequences. We have developed a standard statistical goodness−of−fit test for Heeger's method which gives us a reliability measure that we can use to identify those

velocity measurements which correspond to data that conform more closely to the underlying assumptions of the method. We have shown using synthetic data that this test possesses the characteristics needed to exclude erroneous velocity estimates and that behaviour of the measure with respect to large velocity estimates is comparable for the synthetic and real cases. There is a potential improvement in the usefulness of the test if we use statistics generated from real data sequences rather than assuming that the data is a moving random signal.

REFERENCES

1. **Shrimpton, P.C., Wall, B.F., Jones, D.G., Fisher, F.S., Hiller, M.C., Kendall, G.M., Harrison, R.M.** Doses to patients from routine X−ray examinations in England. *Brit. J. Radiol. 1986, 59, 749−758*

2. **Barron, J., Fleet, D.J., Beauchemin, S.S.** Performance of Optical Flow Techniques. *RPL−TR−9107 Queens University, Kingston, Ontario (1993)*

3. **Bracewell, R.N.,** The Fourier Transform and its Applications (2nd Ed. Revised) *McGraw−Hill (1986) ISBN 0−07−066454−4*

4. **Heeger, D.J.** Model for the extraction of image flow. *J. Opt. Soc. Am. 1987 4(8) 1455−1471*

5. **Heeger, D.J.** Optical Flow using Spatio−temporal Filters *Int. J. Computer Vision (1988) 1 279−302*

6. **Press, W.H. et al** Numerical Recipes in C (2nd Edition) *Cambridge University Press (1992)*

7. **Graham, J, Taylor, C.J.** Boundary Cue Operators for Model−Based Image Processing. *Proc. Alvey Vision Conf. AVC88 Manchester, UK. (1988) 65−72*

8. **Hunnerup, P.B., Nielsen, J.P.** Optic Flow Measurements In Ultrasound Images *MSc Thesis: Department of Medical BioPhysics, University of Manchester, UK (1994)*

9. **Horn, B.K.P., Schunk, B.G.** Determining Optical Flow. *Artificial Intelligence 1981 17, 185−203*

10. **Mailloux, G.E., Bleau, A., Bertrand, M., Petitclerc, R.** Computer analysis of heart motion from two−dimensional echocardiograms. *IEEE Trans. Biomedical Engineering 1987 34(5) 356−364*

11. **Singh, A., Allen, P.** Image Flow Computation: An Estimation−Theoretic Framework and a Unified Perspective. *CVGIP: Image Understanding (1992) 56,2 152−177*

12. **Oosterveld, B.J. et al** Texture of B−mode Echograms: 3D Simulations and experiments of the Effects of Diffraction and Scatterer Density. *Ultrasonic Imaging 7, 142−160 (1985)*