OBJECT LOCALISATION IN FETAL ULTRASOUND IMAGES USING INVARIANT FEATURES

Christopher P. Bridge and J. Alison Noble

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK

ABSTRACT

We address the task of object localisation in 2D fetal ultrasound images, where invariance to factors such as image contrast and object orientation is desirable. We build on recent methods for rotation-invariant detection and combine them with oriented measures of image structure derived from the monogenic signal. We test our approach on images containing the fetal heart. Our results suggest that although raw intensity features can achieve robust approximate detection, the structural measures can achieve better localisation.

Index Terms— Ultrasound, detection, monogenic signal, rotation invariant, histogram of gradients, feature symmetry/asymmetry, fetal heart

1. INTRODUCTION

In this paper, we address the task of automated object detection in ultrasound (US) images. The localisation of anatomical structures ('objects') in US images and videos may be useful as a first step in clinical procedures, such as automated biometry.

However, automated analysis of fetal US imagery is a challenging problem. A major obstacle is the variable quality of the images: there may be low contrast levels, speckle distortion, and acoustic shadowing and enhancement artefacts. A further complication, particular to fetal imaging, is the variable orientation of the fetal anatomy relative to the probe. In this work, we aim to design an algorithm with a high degree of robustness to many of these factors.

Previous successful object detection algorithms in the computer vision literature have generally focussed on *sliding window* approaches. Typically these use intensity features [1] or histograms of oriented gradients [2], which are generally found to have increased robustness to contrast changes.

Other work has found that the low-level *monogenic signal* representation provides a description of local US image structure with robustness to speckle and contrast variations. This has been used, for example, for boundary detection [3] and to provide a short-list of candidates for later detection [4], but to our knowledge has not been integrated fully into detectors.

Most popular object detection algorithms do not directly address the problem of unknown orientation, but instead resort to running the detector on multiple rotated versions of the image. In this work we build on recent advances [5] to provide low-level rotation invariance to detectors. Ultrasound images are not strictly invariant under rotations, but the rotation-dependent variation is not very pronounced for soft tissue.

2. DEFINING ORIENTED FEATURES

We follow a number of previous works in using the *monogenic* signal [6] to provide a degree of robustness to contrast variation and speckle distortion in US images. The monogenic signal is found using an even-symmetric bandpass image filter (in our case a log-Gabor filter) and two odd-symmetric filters formed from its *Riesz transform*. We denote the even part of the filter response by $q_e(\mathbf{x})$, and we express the two odd parts as a single complex number $q_o(\mathbf{x})$. The monogenic odd filter (MGOF) response itself forms one oriented feature.

Two useful measures that may be derived from the monogenic signal [3] are *feature symmetry*, $S(\mathbf{x})$, and *feature asymmetry*, $A(\mathbf{x})$. These quantities capture the intuition that highly symmetric regions give large responses to the even filter and low responses to the odd filter, and *vice versa* for asymmetric areas (boundaries). The standard definitions [7] are:

$$S(\mathbf{x}) = \frac{\lfloor |q_e(\mathbf{x})| - |q_o(\mathbf{x})| - T \rfloor}{P(\mathbf{x}) + \epsilon}$$
(1)

$$A(\mathbf{x}) = \frac{\lfloor |q_o(\mathbf{x})| - |q_e(\mathbf{x})| - T \rfloor}{P(\mathbf{x}) + \epsilon}$$
(2)

where T is a thresholding value (we empirically set T = 0.18), ϵ is a small constant to avoid division by zero, and $\lfloor \cdot \rfloor$ denotes the zeroing of negative values. Normalisation by the local amplitude, $P(\mathbf{x}) = \sqrt{|q_e(\mathbf{x})|^2 + |q_o(\mathbf{x})|^2}$ gives a value in the range 0 to 1 in each case. However, these definitions lose the directional information since they use only the magnitude of the filter responses. It is preferable to retain this information, which leads us to define *signed feature symmetry*, $\hat{S}(\mathbf{x})$, and *oriented feature asymmetry*, $\hat{A}(\mathbf{x})$ as follows:

$$\hat{S}(\mathbf{x}) = S(\mathbf{x}) \cdot \operatorname{sgn}\left(q_e(\mathbf{x})\right) \tag{3}$$

$$\hat{A}(\mathbf{x}) = A(\mathbf{x}) \cdot \frac{q_o(\mathbf{x})}{|q_o(\mathbf{x})|}$$
(4)



Fig. 1: Feature symmetry and asymmetry measures superimposed upon an image of the fetal heart. Colourless areas have symmetry/asymmetry values of zero due to thresholding.

where $sgn(\cdot)$ is the *signum* function. See Figure 1.

Signed feature symmetry (SFS) retains the sign of the even filter response, and hence lies in the range $-1 \leq \hat{S}(\mathbf{x}) \leq 1$ where negative values correspond to *troughs* in the image and positive values to *peaks*. For later description, we find it advantageous to split the SFS images into positive and negative (peak and trough) images and concatenate the descriptors from each image separately.

Oriented feature asymmetry (OFA) retains the complex orientation of the odd filter response by multiplying by a unit complex number whose argument is the local orientation. It therefore gives a complex-valued representation capturing the magnitude of the asymmetry *and* its direction in the image.

3. ROTATION INVARIANT DESCRIPTION

In order to achieve the desired rotation invariance we draw on the recent method of Liu et al. [5]. The method contains two distinct developments that together give overall rotation invariance: Fourier orientation histograms and rotation invariant basis functions. The original formulation was developed for a gradient representation of the image, but we apply the same methodology here to a number of different scalar and directional (2D vector-valued) image representations.

3.1. Orientation Histogram Representation Using Fourier Coefficients

Standard orientation histogram methods (e.g. [2]) have poor rotation behaviour as a consequence of discretising the orientation space. Since an orientation histogram is a periodic function of orientation, an alternative representation is a truncated (smoothed) Fourier series representation. With such a representation, rotation of the input image leads to a phase shift of the Fourier series coefficients. Only a few coefficients are needed to give a discriminative yet robust description.

A weighted Fourier histogram may be formed by considering each pixel's orientation in turn as a magnitudeweighted Dirac δ -distribution in an orientation histogram, and then representing it by the Fourier series expansion of the δ distribution. Using a complex notation for a general 2D vector representation, v, (e.g. gradient) the m^{th} -order coefficient is given by:

$$c_m(\mathbf{x}) = \|v(\mathbf{x})\|e^{-\mathbf{i}m \arg(v(\mathbf{x}))}, \qquad m = 0, 1, \dots, M$$
 (5)

where $\mathbf{i} = \sqrt{-1}$ and M is the largest order used. This expansion generates a group of coefficients $\{c_m\}$ for each pixel. Spatial aggregation can then take place for each coefficient order m separately, as if each forms one image channel.

3.2. Rotation Invariant Basis Functions

In order to create a rotation invariant description of a circular image window, it is integrated against a number of different circular *basis functions*. In the most general form, the basis functions may be expressed in polar coordinates, (r, θ) , as:

$$u_{j,k}(r,\theta) = p_j(r)e^{\mathbf{i}k\theta} \quad \text{for} \quad 0 \le r < R, \quad 0 \le \theta < 2\pi$$
(6)

Each basis function is the product of an arbitrary radial profile, p(r), and a Fourier series basis of order k in orientation. A descriptor is built up using a number of basis functions $u_{j,k}(r,\theta)$ with J different profiles, $\{p_j(r)\}_{j=0,1,\ldots,J-1}$, and rotation orders $k = -K, \ldots, K$.

To build a description for windows centred on each pixel in an image, each basis function, $u_{j,k}(\mathbf{x})$, is convolved ('*') with either the input image (or other scalar image representation such as feature symmetry), or the orientation histogram coefficient images, $c_m(\mathbf{x})$, from directional features (such as gradient). We can incorporate both cases into the same framework by considering a scalar representation to be an orientation histogram with a coefficient at m = 0 only. The result is 'images' of features, $D_{j,k,m}(\mathbf{x})$, where

$$D_{j,k,m}(\mathbf{x}) = u_{j,k}(\mathbf{x}) * c_m(\mathbf{x})$$
(7)

Because the angular basis is the Fourier basis, rotating the image by ϕ results in a simple phase shift of the resulting complex feature by $\hat{k}\phi$, where the *feature rotation order* $\hat{k} = k-m$. If $D_{j,k,m}$ describes a window, then $D_{j,k,m}e^{i\hat{k}\phi}$ describes the rotated version of the same window. Accordingly, the magnitude of the feature, $|D_{j,k,m}|$, is analytically rotationally invariant (small discrepancies of course arise due to sampling effects). Note that for m = 0, only bases with $k \ge 0$ are used to avoid redundancy in the feature set.



Fig. 2: Set of profiles and basis functions with J = 3, K = 4 (only $k \ge 0$ displayed). The saturation and hue represent the complex magnitude and argument respectively.

Following the original work [5], we use a set of overlapping 'triangular' profiles. Specifically, if R is the radius of the support of the largest basis function in the set, and J is the number of different profiles that make up the set, we use the set of profiles centred on radii $a_j = j\frac{R}{J}$:

$$p_j(r) = \left\lfloor 1 - \frac{|r - a_j|}{\frac{R}{J}} \right\rfloor \tag{8}$$

for $0 \le r < R$, j = 0, 1, ..., J - 1 where $\lfloor \cdot \rfloor$ represents the zeroing of negative values (see Figure 2 for an example set).

In practice, it is computationally more efficient to implement the filtering (7) as a multiplication in the Fourier domain using fast Fourier transforms of the image and basis functions.

3.3. Full Description Methodology

The full detection procedure is illustrated in Figure 3. First we form the image representation using the relevant method (§2). Then, if it is a directional representation, we form its Fourier histogram, $c_m(\mathbf{x})$, using (5). The images of features, $D_{j,k,m}(\mathbf{x})$, are then formed by Fourier domain filtering according to (7).

We then form the descriptor vector using all the rotation invariant quantities in the complex features. This includes the purely real features for which k = m = 0, the real and imaginary parts of features where $\hat{k} = 0$ and $k \neq 0$, and the magnitude of all other features $\hat{k} \neq 0$. We discard features with large rotation order ($|\hat{k}| > 4$). We then normalise this descriptor vector using the mean and standard deviation of the descriptors in the training set, and pass to a support vector machine (SVM).

4. EXPERIMENTS

To test our approach, we address the problem of detecting the fetal heart in 2D videos taken of the fetal abdomen. Our experiments were conducted on a dataset of 63 videos of fetuses at a range of gestational ages (16 to 40 weeks) giving a range of appearances and image qualities. We manually labelled the position of the heart, defined using the centre and radius of a

circular window fully encompassing all the chambers, in 10 random frames from each video, giving a total of 630 frames. One 'background' window of the same size was also randomly selected in each of the chosen frames. We split the dataset into seven partitions of 90 frames each, ensuring that a single video never contributed frames to both the training and test sets. We performed a seven-fold cross-validation on this dataset using these partitions. In each cross-validation iteration, the SVM parameters were trained using 6 of the 7 partitions, i.e. one positive and one negative window from each of 540 frames.

Evaluation was performed by passing each detector over the 90 images in its unseen test partition. The window with the largest SVM output score was chosen to be the heart detection. This was compared to the ground truth label, and if the distance between the two labels' centres was less than one-third of the heart radius, the detection was considered correct.

We evaluated algorithms based upon a number of different image representations including intensity (INT), gradient (GRAD, as used in the original work [5]), signed feature symmetry and oriented asymmetry (SFS and OFA), and monogenic odd filters (MGOF). We also combined representations by simply concatenating the descriptors before classification. For each image representation we used a small number of different sets of the parameters J, K, M. The centre-wavelength of the monogenic bandpass filter was set to $\lambda_0 = 75$ pixels in all cases, as this was empirically found to give good results.

We have not yet investigated methods for detecting objects at different scales. Therefore in these experiments the size of the basis functions, R, was set to the manual labels. This is not too unreasonable in practice, since the size of the heart is relatively predictable given the gestational age, and the physical dimensions of the pixels is usually known. It is necessary to normalise the features in proportion to basis function area.

5. RESULTS

Table 1a shows the classification accuracy (as defined above) of the various detectors. In addition, we show the mean normalised localisation error in Table 1b, where normalisation is by the manually-labelled heart radius. We see that the simple intensity detectors perform well in terms of classification accuracy (up to 87%), but that their results are significantly worse for localisation error (≥ 0.3 compared to as low as 0.21 for monogenic features). This may be because the phase-based features are better able to exploit the fine image structure to determine the location more precisely, but are more easily confused by background examples displaying similar structure.

It is interesting to note that gradient-based detectors (as used in previous work [5]) perform poorly relative to other features, including intensity. One possible reason is that the speckle artefact creates confounding gradients in the image.

We see from the results that using a large J generally improves classification performance, suggesting that separating image structure found at different radii aids detection. By con-



Fig. 3: Full detection process for a directional representation, Fourier histogram expansion is omitted for scalar representations.

Par	amet	ers	Image Representation								
J	K	M	INT	GRAI) INT+	SFS	MGO	F OFA	SFS+	SFS+	
				GRAD					OFA	MGOF	
3	4	4	0.70	0.79	0.80	0.81	0.83	0.76	0.78	0.80	
4	4	4	0.82	0.78	0.81	0.81	0.81	0.77	0.84	0.84	
5	4	4	0.86	0.78	0.80	0.83	0.82	0.79	0.83	0.87	
6	4	4	0.87	0.80	0.82	0.84	0.84	0.80	0.83	0.88	
4	2	4	0.80	0.80	0.82	0.79	0.83	0.81	0.82	0.88	
4	3	4	0.81	0.79	0.82	0.82	0.84	0.80	0.84	0.83	
4	5	4	0.82	0.78	0.81	0.81	0.75	0.78	0.84	0.84	
4	4	2	-	0.80	0.80	-	0.88	0.82	0.81	0.87	
4	4	3	-	0.82	0.81	-	0.87	0.80	0.80	0.88	
4	4	5	-	0.79	0.80	-	0.80	0.77	0.82	0.85	

	(a) Classification accuracies											
Parameters				Image Representation								
	J	K	M	INT	GRAD	INT+	SFS	MGOF	OFA	SFS+	SFS+	
					GRAD					OFA	MGOF	
	3	4	4	0.46	0.42	0.35	0.28	0.29	0.34	0.33	0.33	
	4	4	4	0.36	0.38	0.34	0.28	0.28	0.32	0.28	0.26	
	5	4	4	0.31	0.35	0.34	0.26	0.26	0.30	0.26	0.23	
	6	4	4	0.30	0.36	0.32	0.26	0.26	0.30	0.24	<u>0.21</u>	
	4	2	4	0.37	0.35	0.33	0.32	0.26	0.24	0.29	0.26	
	4	3	4	0.35	0.39	0.34	0.28	0.25	0.28	0.29	0.27	
	4	5	4	0.36	0.41	0.36	0.28	0.36	0.30	0.25	0.27	
	4	4	2	-	0.34	0.32	-	0.21	0.28	0.26	0.21	
	4	4	3	-	0.37	0.38	-	0.21	0.31	0.32	0.23	
	4	4	5	-	0.38	0.35	-	0.32	0.35	0.30	0.24	

(b) Ground truth distances normalised by heart radius

Table 1: Tables of experimental results for a number of detectors.

 Bold values are the best for the relevant image representation, underlined are best overall.

trast, in many cases using large K and M gives poorer results, suggesting that the higher order features contain little useful information and may confound the learning process. The best values for these parameters are likely to vary for structures of differing complexity. In this case the MGOF seemed a particularly effective feature with relatively low order histogram parameters (e.g. J = 4, K = 4, M = 2) as it achieved both a high classification accuracy of 88% and low average localisation error of 0.21.

Our MATLAB implementation took between 0.4 s per image (for raw intensity based detectors and low J, K, M), to just over 5 s (with phase-based features and large J, K, M) for images with side lengths in the range 300 to 600 pixels on a 3.40 GHz Intel i7 desktop PC. By comparison, previous work using spatial-domain convolutions reported 18 s per frame [5].

6. CONCLUSIONS

We have shown that quick, approximate, rotation-invariant detection of anatomical objects in images can be achieved using rotation invariant basis functions on the raw intensity images. More accurate localisation can be achieved through the use of oriented measures derived from the monogenic signal, as these measures are better able to describe the underlying image structure.

7. ACKNOWLEDGEMENTS

The authors are grateful to Christina Aye and Paul Leeson of the Oxford Cardiovascular Clinical Research Facility, John Radcliffe Hospital for the data used for experiments. CPB acknowledges the support of the UK Engineering and Physical Sciences Research Council Doctoral Training Award.

8. REFERENCES

- P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, 2001, vol. 1, pp. 511–518.
- [2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in CVPR, 2005, vol. 1, pp. 886–893.
- [3] K. Rajpoot, V. Grau, and J.A. Noble, "Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardiography images," in *ISBI*, 2009, pp. 783–786.
- [4] B. Rahmatullah, A.T. Papageorghiou, and J.A. Noble, "Integration of local and global features for anatomical object detection in ultrasound," in *MICCAI*, vol. 7512, pp. 402– 409. Springer Berlin Heidelberg, 2012.
- [5] K. Liu, H. Skibbe, et al., "Rotation-invariant HOG descriptors using Fourier analysis in polar and spherical coordinates," *IJCV*, vol. 106, no. 3, pp. 342–364, 2014.
- [6] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136– 3144, 2001.
- [7] P. Kovesi, "Symmetry and asymmetry from local phase," in *Tenth Australian Joint Conference on Artificial Intelli*gence, 1997, vol. 190, pp. 2–4.