Machine learning for the automatic localisation of foetal body parts in cine-MRI scans

Christopher Bowles[a], Niamh Nowlan[c], Tayyib T. A. Hayat[b], Christina Malamateniou[b], Mary Rutherford[b], Joseph V. Hajnal[b], Daniel Rueckert[a], and Bernhard Kainz[a]

[a]Biomedical Image Analysis Group, Department of Computing,
Imperial College London, London, UK;
[b]Department Biomedical Engineering Division of Imaging Sciences,
King's College London, London, UK;
[c]Department of Bioengineering, Faculty of Engineering,
Imperial College London, London, UK

ABSTRACT

Being able to automate the location of individual foetal body parts has the potential to dramatically reduce the work required to analyse time resolved foetal Magnetic Resonance Imaging (cine-MRI) scans, for example, for use in the automatic evaluation of the foetal development. Currently, manual preprocessing of every scan is required to locate body parts before analysis can be performed, leading to a significant time overhead. With the volume of scans becoming available set to increase as cine-MRI scans become more prevalent in clinical practice, this stage of manual preprocessing is a bottleneck, limiting the data available for further analysis. Any tools which can automate this process will therefore save many hours of research time and increase the rate of new discoveries in what is a key area in understanding early human development. Here we present a series of techniques which can be applied to foetal cine-MRI scans in order to first locate and then differentiate between individual body parts. A novel approach to maternal movement suppression and segmentation using Fourier transforms is put forward as a preprocessing step, allowing for easy extraction of short movements of individual foetal body parts via the clustering of optical flow vector fields. These body part movements are compared to a labelled database and probabilistically classified before being spatially and temporally combined to give a final estimate for the location of each body part.

Keywords: Foetal body part identification, Cine-MRI, Machine Learning, Motion Suppression, Segmentation

1. INTRODUCTION

Type and degree of foetal movement is thought to be a potential indicator of a number of musculoskeletal and neurological disorders.[1,2] Time resolved 2D foetal cine-MRI scans can provide an extremely valuable tool to be used to analyse foetal movements in the hope of ultimately identifying potential biomarkers for identifying foetal pathologies. However, locating a particular foetal body part can be a difficult and costly time sink when when exploring such biomarkers. Localisation and segmentation is often an unwelcome requirement which can take a disproportionately large amount of time, potentially requiring hours of tedious manual processing. As foetal cine-MRI scans become more common and data sets become larger, the greater the requirements for tools to automatically perform this preprocessing will become. Without such tools, it is easy to see a scenario in which only a fraction of available data can be used for investigations, simply because of the manual work which would be required to process more data. The work presented here describes a flexible framework to automatically locate individual body parts from a foetal cine-MRI scan, which can be adapted and used to reduce the burden of manual preprocessing in a wide range of foetal cine-MRI investigations. A key stage in the framework is a novel technique to suppress the effects of maternal breathing in the scan. This technique has potential applications beyond foetal imaging, in areas in which breathing or near periodic movement is present and undesirable.

Further author information: (Send correspondence to Christopher Bowles)
[a] [christopher.bowles10|t.hayat|d.rueckert|b.kainz]@imperial.ac.uk
[b] [christina.malamateniou|mary.rutherford|jo.hajnal]@kcl.ac.uk
[c] n.nowlan@imperial.ac.uk

2. METHOD

2.1 Preprocessing

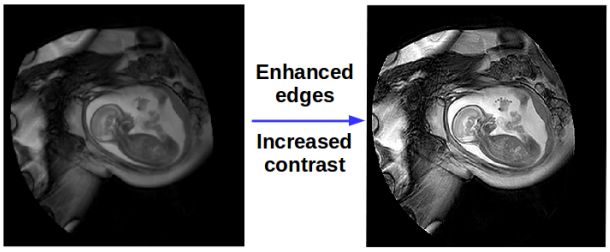


Figure 1: Sequence preprocessing

Prior to processing, each image sequence undergoes two operations in order to improve the performance of later procedures. Edges are sharpened using unsharp masking and pixel intensities are scaled such that 1% of pixels are saturated at the highest and lowest intensities as a way of normalising scans with respect to each other (Figure 1). Both these processes result in easier differentiation between the structures present in an image due to clearer edges and enhanced contrast and will improve the performance of intensity based processes such as optical flow (see Sections 2.5 and 2.6). In addition, image normalisation allows for the development of more robust methods.

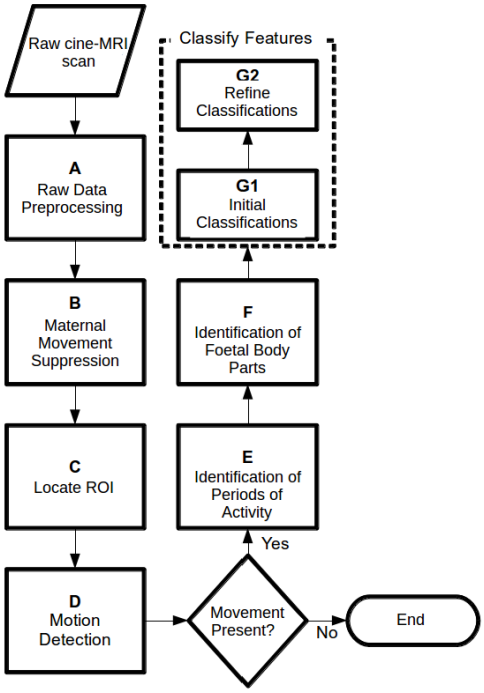


Figure 2: An overview of the different processes involved in the proposed method to identify individual foetal body parts

2.2 Maternal Motion Suppression

By treating the intensity value of each pixel over the duration of the scan as a signal and taking the Fourier transform it is possible to identify the frequency corresponding to the maternal breathing. Removing this frequency, along with higher frequency harmonics, results in the variations at this frequency to be removed and replaced by the mean signal intensity. This has the visual effect of blurring any periodic movement at this frequency in the image sequence. However, simply removing these frequencies from each pixel intensity signal results in the introduction of artefacts in the parts of the sequence which do not exhibit movements at this frequency, such as parts corresponding to foetal movements. These artefacts are characterised by the ghosting of movements before and after the true movement. It is therefore desirable to vary the application of the frequency filter in time and space according to whether these artefacts would be introduced. The details of this procedure are outlined in the rest of this section.

Once the maternal breathing frequency and its harmonics have been identified through the combined Fourier transforms of each pixel's intensity signal (Figure 3), a convolution kernel can be found which will remove these frequencies from a signal. The simplest way of creating such a kernel is to define a smoothly varying function over the combined Fourier domain of the signals, with a value of 0 at the identified frequency and its harmonics, and 1 everywhere else. The inverse Fourier transform of this function will yield the convolution kernel (Figure 4) which can be applied in the signal domain to remove these frequencies.

Convolving the entire signal for every pixel with this kernel will result in the artefacts mentioned previously, it is therefore necessary to selectively apply the convolution to only parts of the signals corresponding to the more periodic maternal breathing, as opposed to more random foetal movement. To do this it is necessary to differentiate between *constructive convolution* and *destructive convolution* (Figure 5).

Figure 4: The typical shape of a convolution kernel designed to remove a frequency from a signal

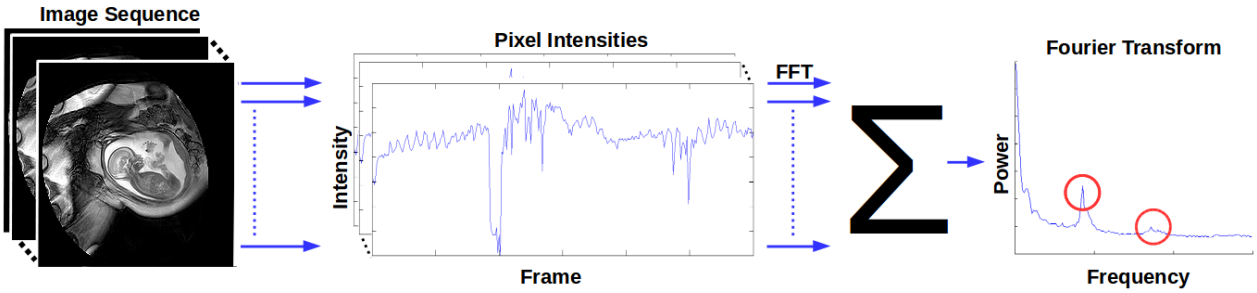
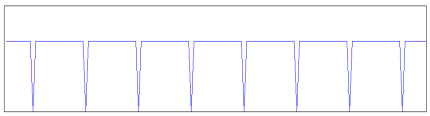


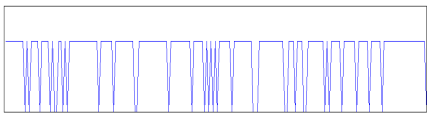
Figure 3: Periodic frequencies are identified through the combined Fourier transforms of each pixel intensity signal

Destructive convolution occurs when the application of a convolution kernel to a signal results in the loss of structure which the signal may have previously had. A purely periodic signal being convolved with a kernel designed to remove exactly the corresponding frequency results in destructive convolution, as the periodicity is destroyed and replaced by the mean of the signal.

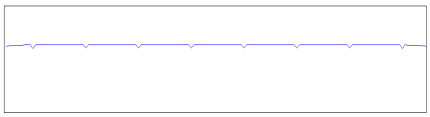
Constructive convolution occurs when convolution results in adding features to a signal which were not previously present. For example a convolution of a delta function with a kernel will result in the delta function being replaced by a copy of the kernel, leading to additional structures either side of where the delta was in the signal, which were not previously present. This effect is the reason for ghosting artefacts.



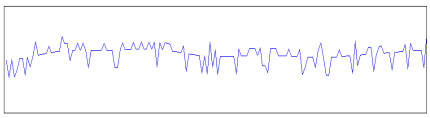
(a) A periodic signal



(b) A random signal



(c) Results of convolving the periodic signal in Figure 5a with the kernel in Figure 4. *Destructive convolution.*



(d) Results of convolving the random signal in Figure 5b with the kernel in Figure 4. *Constructive convolution.*

Figure 5

It is possible to gauge whether convolution at a particular point in a signal, S, is constructive or destructive by defining two new kernels, K_1 and K_2, and convolving the point in the signal with each separately. These kernels correspond to the peaks present in the original kernel and can be seen in Figure 6. In the case of convolution, the resulting signals would be similar in magnitude, with opposite signs, whereas if constructive case, then the resulting signals would have be entirely uncorrelated. This allows a measure to be defined as to whether the convolution of any given point in a signal with a given kernel results in constructive

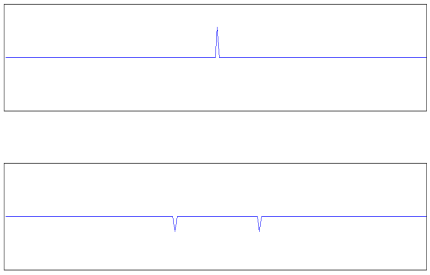


Figure 6: Two kernels derived from Figure 4 which can be used to differentiate between constructive and destructive convolution.

or destructive convolution.

By finding an optimal threshold, t, it is now possible to define a binary function, $C(x)$, which indicates whether or not the application of the convolution at a point x in the signal will result in constructive or destructive convolution (Equation 1). By applying convolution only in areas of destructive convolution, the introduction of unwanted ghosting artefacts is prevented, whilst still suppressing movement at the maternal breathing frequency.

$$C(x) = \begin{cases} 1 & M(x) \leq t \\ 0 & M(x) > t \end{cases} \tag{1}$$

$$M = |S * K_1| - |S * K_2| \tag{2}$$

2.3 Locating the Region of Interest (ROI)

Locating areas which consistently contain high frequency movement components as identified using the combined Fourier transform found previously proves to provide a reasonable estimate for the location of the foetus. This is due to the pseudo-random foetal movements contributing more to the higher frequency components. The image sequence can be reconstructed from the Fourier transforms of the individual pixel signals simply by taking the inverse Fourier transform. If all the frequencies apart from the highest are zeroed prior to reconstruction, the resulting image sequence will identify the areas which contain these high frequency movements. Treating locations which are consistently identified as a binary image and utilising heuristics such as that the ROI is likely to be towards the centre of the scan, morphological operations can be applied to automatically create a mask which can be used to crop each frame in the scan (Figure 7).

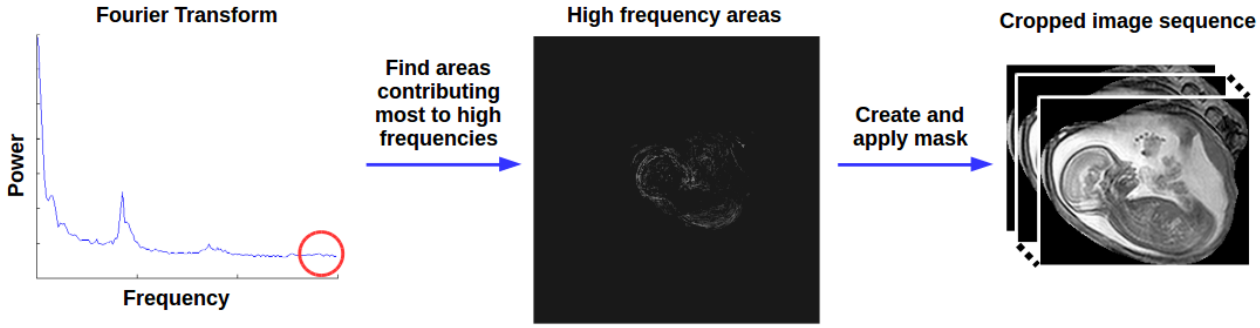


Figure 7: Localization of the foetal ROI.

2.4 Detecting Movement

The entire system of processes presented here relies upon there being a sufficient amount of foetal movement in order to identify the location of potential body parts and to differentiate between them. Image sequences in which the foetus does not exhibit any movement cannot be processed in this way and should therefore be filtered out. One method to detect the presence of foetal movement in a scan is to divide the scan into short sections and create an average image from the frames in each of these sections. In cases where there is little or no foetal movement, these images are likely to be very similar to each other, as repetitive maternal movement is averaged out. If there is foetal movement, some of the images will be significantly different to the others.

Comparing all images with each other and computing the sum of squared differences results a simple robust measure correlated with the amount of foetal movement present in a scan. By finding an optimal threshold on this value, it is possible to classify scans as either containing movement or not.

2.5 Identifying Periods of Foetal Activity

Foetal movement is often characterised by short bursts of activity between periods of inactivity. Identifying these shorter bursts of activity allows for each period to be processed independently. The proposed method uses optical flow[3] to create a series of motion vector fields representing the estimated movement between each consecutive frame. The sum of the magnitudes of the vector fields are used as a measure of the total amount of movement present at different times throughout the scan, allowing for periods of increased activity to be identified through the application of a smoothing filter followed by a threshold (Figure 8).

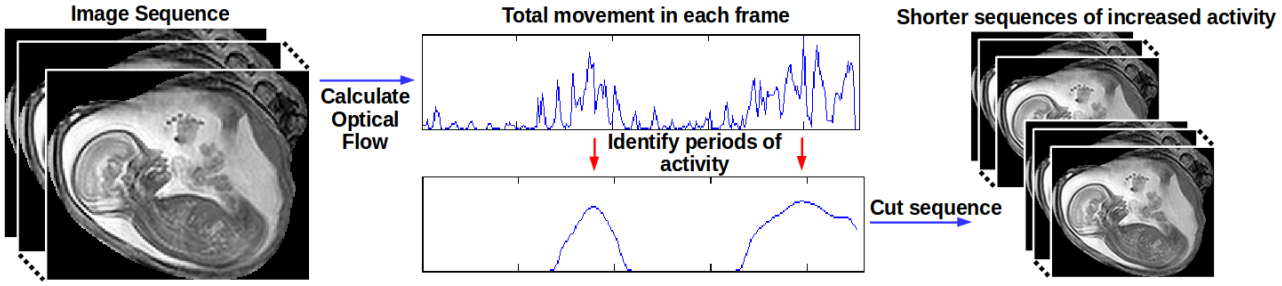


Figure 8: Partitioning of an image sequence into two periods of foetal activity

2.6 Identifying Foetal Body Parts and Deriving Movement Signatures

K-Means clustering[4] can be performed on each period of activity to locate four clusters corresponding to areas most likely to contain a body part. Clustering is applied to the optical flow vector fields. Each pixel is treated as a separate observation, with their displacement between each successive pair of frames considered as the feature vector. An additional two feature dimensions are added to reflect the pixel's spatial location, in order to promote spatially continuous regions. Increasing the relative weighting between spatial location and displacement vector features increases the likelihood of spatially continuous regions being identified, at the cost of smaller body parts being missed. The entire process serves to group nearby areas which consistently move at the same time and in the same direction as each other throughout the course of the period of activity. The choice to locate four clusters was made as empirical observations suggest that foetal movements can be broadly divided into head, body, arm and leg movements. The four clusters located therefore tend to correspond to these body parts. Cropping each frame around these four areas results in segments covering one of these body parts being identified (Figure 9). Each segment is further divided temporally into a number of individual movement signatures (MS), using the technique described Section 2.5. Each MS represents a single movement of an as yet unknown foetal body part. Depending on the length of scan and degree of foetal activity present, each scan can contribute between 10 and 100 MSs, each corresponding to a single movement of a single body part.*

2.7 Classifying Movement Signatures

A combination of four K-Nearest-Neighbour classifiers identifies each MS as Head, Body, Arm, Leg or Maternal Tissue. The maternal tissue class is required to accept MSs containing maternal tissue movements which were erroneously identified through clustering. These are usually the result of a global shift in the scan field of view, or of unsuppressed breathing. Each classification takes the form of a probability distribution over these five classes. The classifiers are trained on different descriptors derived from 3000 manually labelled MSs extracted from 53 scans.

A number of descriptors were evaluated, with a combination of the following four proving to provide the best classification rate.

- A modified variation of motion histograms.[5] The descriptor was given some rotational invariance by aligning the 0 degree direction to the most common direction present in the histogram.

*A sample head MS: `http://y2u.be/gMhFoKCSb_U`

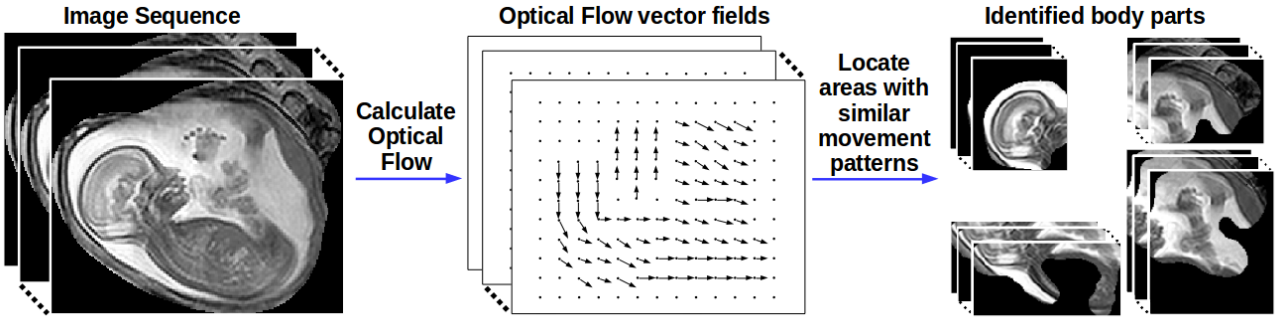


Figure 9: Spatial division into movement signatures (MSs)

- A vector of texture measurements derived from the first frame of each MS. Entropy, contrast, correlation, energy and homogeneity can be calculated and normalised between 0 and 1, and can form a vector to discriminate between MS.
- A modified version of the Histogram of Oriented Gradients (HOG)[6] technique was used by weighting the contribution of each edge by the amount of movement observed at that edge. This has the effect of making edges corresponding to objects which move more have a greater impact on descriptor. A normalised histogram with 16 bins was chosen, and a degree of rotation invariance was added by aligning the direction corresponding to an angle of 0 degrees to the most highly represented edge direction.
- A Fourier Descriptor of Movement Information is derived as a measure of the pattern of movement in an MS. The descriptor consists of a vector containing the magnitudes of 8 radial segments of the Fourier transform of the sum of the optical flow vector field over every frame.

2.8 Refining Classifications

The initial classification assigns each MS a probability distribution over the five possible classes. These probabilities are combined with other MSs extracted from the same areas to create probability maps corresponding to how likely a location is to contain a particular body part. Areas with a high probability of containing maternal tissue are identified and any MS extracted from one of these locations is concretely classified as maternal tissue, whilst the remaining MSs are reclassified as negative examples of maternal tissue. New probability maps are formed for the remaining body parts and the process is repeated. Figure 10 shows how new probability maps are formed as the process is repeated for each body part in turn.

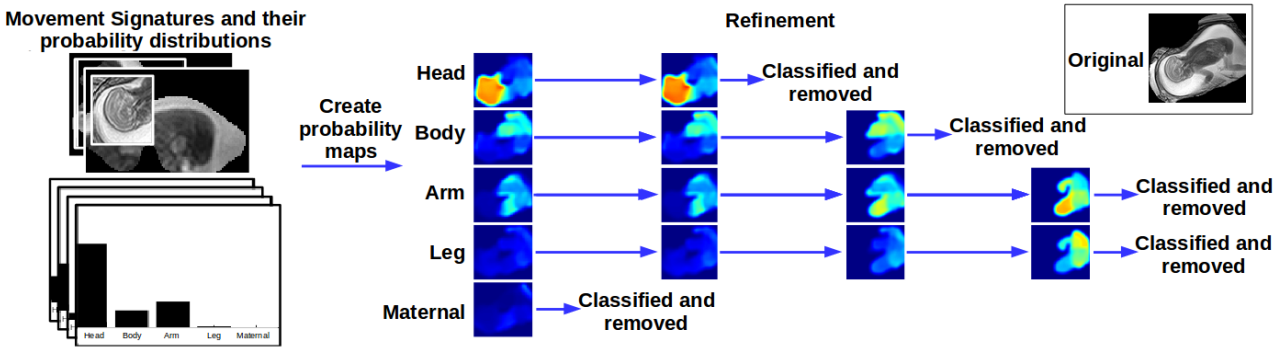


Figure 10: Classification refinement

3. RESULTS

The effects of maternal motion suppression using both complete filtering and space/time variant filtering on a sample cine-MRI scan has been provided and can be viewed online.[†]

We have evaluated our method on 53 scans of moving foetuses.[7] Locating the ROI is successful in 81% of scans, with success being defined as no loss of foetal tissue, and a clear reduction in maternal tissue.

The detection of movement in a scan using the method described was evaluated by introducing 33 scans of still foetuses, reporting an 80% accuracy, with a 95% sensitivity towards scans containing movement.

The process of dividing each scan into a shorter periods of significant foetal movement was analysed by treating it as the classification problem corresponding to the differentiation between periods of movement and of no movement. This resulted in an accuracy of 83%, with a sensitivity of 92% and a specificity of 75%.

81% of body parts which are visible in the scan are located through clustering, with the area containing the head being located in 100%, legs in 75%, arms in 68% and body in 79% of scans.

At least one body part was successfully identified in 74% of scans with an average of 1.3 body parts located per scan. The results of movement signature classification after refinement are summarised in the confusion matrix in Table 1, while the results of classification using each descriptor both individually and combined are shown in Table 2. All evaluation was performed using leave-one-out cross validation, ensuring that no MS from the scan being processed is present in the training set.

	Head	Body	Arm	Leg	Maternal
Head	**894**	61	21	59	24
Body	224	**180**	29	93	54
Arm	90	85	**90**	55	3
Leg	139	176	21	**108**	42
Maternal	42	109	8	29	**156**

Table 1: Confusion matrix of the classification results of movement signatures extracted from 53 scans.

	Motion Histograms	Weighted HOG	Texture Measures	Fourier Descriptor	Combined
Head	0.49	0.45	0.57	0.49	0.64
Body	0.22	0.29	0.27	0.35	0.29
Arm	0.27	0.17	0.4	0.2	0.53
Leg	0.21	0.27	0.26	0.24	0.31
Maternal	0.23	0.23	0.56	0.46	0.56

Table 2: The precision of classification using the four descriptors individually, and the combined together.

4. DISCUSSION

Whilst providing a good estimate of performance, measures derived from the confusion matrix are not necessarily reflective of the whole process. Ambiguous labelling when a MS contains multiple body parts is a frequent cause of misclassification. This is especially a problem between examples of legs and bodies as MSs of these tend to include areas of each other. The confusion matrix indicates that these are indeed often misclassified as each other. This is likely responsible for the poorer reported performance when it comes to locating these body parts.

By evaluating the individual stages of the process it is possible to better analyse the performance of the system as a whole. It is important to note that all the techniques used require a certain degree of foetal movement, and the performance is often directly related to the extent of the movement. However, during foetal MRI foetal

[†]Effects of maternal motion suppression: `http://y2u.be/tiqD5dfcIuc`

motion is usually frequent and intense. While this is a problem for conventional MRI image analysis our proposed method makes use of this otherwise potential source of artefacts.

The method used for movement detection revolves around the setting of a threshold which can be tuned to directly balance sensitivity versus specificity. The value chosen when testing was selected to maximise sensitivity while still retaining a specificity greater than 60%. Unless a very large number of scans are being processed, it is likely that a manual filtering of scans to remove those with no movement would be feasible, in which case this step can be avoided. In the case where manual filtering is not an option, it was decided that sensitivity was more important than specificity. This is because false positives will simply lead to failure to detect body parts in cases where the lack of movement would render this impossible anyway, whereas false negatives will result in scans being ignored in which body parts could have been successfully identified.

The method used to divide a scan into periods of foetal activity also involves the setting of a threshold which can be used to control sensitivity versus specificity. Sensitivity is again the more important of the two measures, as incorrectly removing periods of activity results in the loss of potentially useful data. However, a high specificity is also important, as continuing to process periods of no movement is likely to result in erroneous classifications.

The identification of body parts using clustering of optical flow vector fields appears to be a particularly promising method yielding good results, especially if the detection of the head is of significant importance. The head lends itself well to detection using clustering. The inhomogeneity, rigidity and relatively short movements of the head creates a very clear optical flow pattern, which is readily extracted. MSs containing heads therefore tend to be very precise, with little overlap of other body parts. It is important to consider also that while only body parts visible in the scan were used as ground truth in order to obtain the results, it is not necessarily the case that these body parts exhibited any movement during the scan. Since optical flow can inherently not track stationary objects, performance on scans containing a full range of movement is likely to be higher.

By examining Table 2, it can be seen that each descriptor has a high precision when locating the head. This is likely a result of the location of heads being very consistently and accurately found using clustering, leading to higher quality MSs to learn from. It is also useful to note how some descriptors are more able to discriminate between certain body parts than others. Except from in the case of locating the body, the result of combining classifiers is better than the best performing single classifier for a given body part. The precision of locating bodies and legs is significantly poorer than the other body parts, however this could be a result of ambiguous MSs as mentioned previously.

5. CONCLUSION

Despite large variations in the data set, the good results for the individual stages of the process suggest that the proposed method has the potential to automate the process of locating individual body parts prior to further analysis[‡]. The correlation between level of movement and performance of each stage has been clear throughout. Scans containing extensive movement can be analysed and body parts consistently and accurately located. In addition, the individual processing stages can be used effectively in isolation in different systems. The movement suppression technique is currently being used in other clinical applications.

The results of classification, to identify which body parts have been located, suggest that more work could be done to consistently differentiate between body parts. This could perhaps be achieved by the use of more sophisticated features, or simply by having a larger data set to train on. The extraction of non overlapping MSs would lead to less ambiguity and an improvement in learning specific body parts. This would be of particular benefit for the identification of bodies and legs. Despite these limitations when it comes to differentiating between multiple body parts, the system does perform well at consistently locating the head. A modified version of the system which aims to only locate the head is currently being investigated for use in identifying possible biomarkers based upon head movements.

[‡]Crude segmentation and colour based labelling of head, body and limbs: `http://y2u.be/uBh4ngNJSvc`

ACKNOWLEDGMENTS

Bernhard Kainz is supported by an European Union FP7 Marie-Curie Intra-European Fellowship (FP7-PEOPLE-2012-IEF F.A.U.S.T. 325661).

REFERENCES

[1] Giorgi, M., Carrieo, A., and Shefelbine, S. J., "Mechanobiological simulations of prenatal joint morphogenesis," *Journal of Biomechanics* **47**, 989–995 (2014).

[2] Nowlan, N., "Biomechanics of foetal movement.," *European cells & materials* **29**, 1 (2015).

[3] Beauchemin, S. S. and Barron, J. L., "The computation of optical flow," *ACM Computing Surveys* **27(3)**, 433–466 (1995).

[4] Kanungo, T., Mount, D., Netanyahu, N., Piatko, C., Silverman, R., and Wu, A., "An efficient k-means clustering algorithm: analysis and implementation," *Pattern Analysis and Machine Intelligence* **24(7)**, 881–892 (2002).

[5] Davis, J. W., "Hierarchical motion history images for recognizing human motion," in [*Detection and Recognition of Events in Video*], 39–46 (2001).

[6] Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection," in [*Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*], **1**, 886–893, IEEE (2005).

[7] Hayat, T., Nihat, A., Martinez-Biarge, M., McGuinness, A., Allsop, J., Hajnal, J., and Rutherford, M., "Optimization and initial experience of a multisection balanced steady-state free precession cine sequence for the assessment of fetal behavior in utero.," *American Journal of Neuroradiology* **32(2)**, 331–338 (2011).