

# Localisation of the Brain in Fetal MRI Using Bundled SIFT Features

Kevin Keraudren<sup>1</sup>, Vanessa Kyriakopoulou<sup>2</sup>, Mary Rutherford<sup>2</sup>,  
Joseph V. Hajnal<sup>2</sup>, and Daniel Rueckert<sup>1</sup>

<sup>1</sup> Biomedical Image Analysis Group, Imperial College London

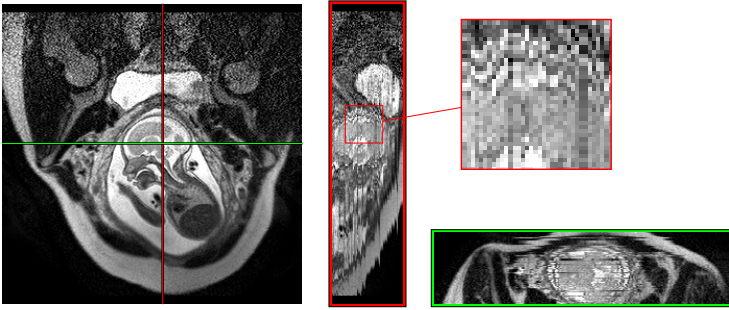
<sup>2</sup> Centre for the Developing Brain & Department Biomedical Engineering Division of  
Imaging Sciences, King's College London

**Abstract.** Fetal MRI is a rapidly emerging diagnostic imaging tool. Its main focus is currently on brain imaging, but there is a huge potential for whole body studies. We propose a method for accurate and robust localisation of the fetal brain in MRI when the image data is acquired as a stack of 2D slices misaligned due to fetal motion. We first detect possible brain locations in 2D images with a Bag-of-Words model using SIFT features aggregated within Maximally Stable Extremal Regions (called bundled SIFT), followed by a robust fitting of an axis-aligned 3D box to the selected regions. We rely on prior knowledge of the fetal brain development to define size and shape constraints. In a cross-validation experiment, we obtained a median error distance of 5.7mm from the ground truth and no missed detection on a database of 59 fetuses. This 2D approach thus allows a robust detection even in the presence of substantial fetal motion.

## 1 Introduction

Fetal Magnetic Resonance Imaging (MRI) has had great successes in the last years with the development of motion correction methods providing high quality isotropic volumes of the brain [7,11], thus enabling a better understanding of the fetal brain development. Such reconstruction methods typically rely on data acquired as stacks of 2D slices of real-time MRI, freezing in-plane motion. In order to reduce the scan time while avoiding slice cross-talk artefacts, contiguous slices are not acquired sequentially but in an interleaved manner. Slices are quite often misaligned due to fetal motion and form an inconsistent 3D volume (Figure 1). Motion correction is the registration of 2D slices of the fetal brain to an ideal 3D volume. Cropping a box around the brain is a prerequisite to exclude maternal tissues that can make the registration fail. We thus propose a method to automatically find a precise bounding box around the brain in order to speed-up the preprocessing steps of the motion correction procedure.

**Related Work.** Regression Forests perform the task of organ detection by learning a regression between image features, such as 3D Haar-like features, and the offset to the corners of the bounding boxes of organs [10]. By treating the



**Fig. 1.** Example scan with the native 2D slices in sagittal orientation. A zoomed patch of the coronal section highlights interleaving artefacts due to fetal motion.

localisation problem as a regression problem, it implies some rough alignment between part of the training database and the new data. An adult patient usually lies inside the scanner in a standard orientation, but this is not the case of a fetus. Moreover, the fetus is surrounded by maternal tissues instead of air, thus hindering a straightforward application of such a method.

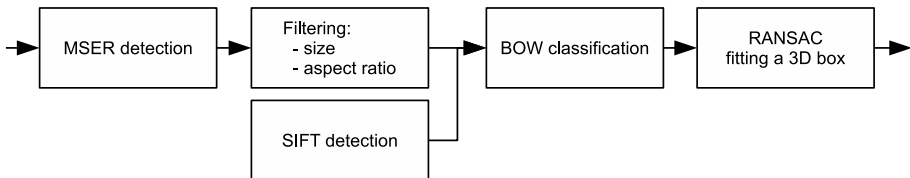
Marginal Space Learning (MSL) is a more general framework that can be described as an optimized sliding window search with a hierarchy of coarse-to-fine boosting-based organ detectors [15]. Learning is performed on affinely registered datasets, and the search covers all locations, scales and orientations. Steerable features apply a rotated grid over the image to sample features in order to perform local rotation and scaling without transforming the whole volume, thus making the search more computationally feasible. MSL has been successfully applied to various problems, such as the localisation of the four heart chambers in adult CT scans [15], semantic browsing of the fetal body in 2D ultrasound [2], or automatic fetal face detection in 3D ultrasound [4].

To the best of our knowledge, only two fully automated localisation methods of the fetal brain in MRI have been proposed so far in the literature: Anquez et al. [1] proposed to start from detecting the eyes with 3D template matching, followed by a segmentation of the brain using a 2D graph-cut segmentation of the mid-sagittal slice rotated at several angles. The best matching segmentation is selected and used to initialise a 3D graph-cut segmentation. Ison et al. [6] proposed a more general method based on a two-stage Random Forest classifier: the first stage distinguishes maternal tissues from fetal head tissues, while the second stage classifies fetal head tissues into 6 classes. The last step uses a Markov Random Field appearance model to establish an orientation of the brain based on the centroids of the previously identified fetal tissues.

**Overview.** In this paper, we chose to focus on the task of finding an axis-aligned bounding box for the fetal brain. This is in contrast with [6] which aimed at finding an oriented bounding box, and [1] which aimed at segmenting the skull bone

content. This relaxation of the problem enables us to focus on positioning a tight bounding box on the brain while dealing with 2D slices misaligned or corrupted due to fetal motion. Although acquisitions are carried out in conventional orientations, there is unpredictability in the positioning of the fetus. Hence, similarly to Marginal Space Learning, we decided to decompose the search space by performing a 2D detection process before accumulating the votes in 3D, and we removed the scale component by inferring it from the gestational age.

In a first stage, Maximally Stable Extremal Regions (MSER [9]) are extracted from the 2D slices and approximated by ellipses. Then, the regions whose size and aspect ratio conform with prior knowledge of the fetal development are classified into brain or not-brain based on histograms of the SIFT features found within the fitted ellipses (bundled SIFT [14]), following a Bag-of-Words model (BOW [3]). Finally, a RANdom SAMple Consensus procedure (RANSAC [5]) is performed to find a best fitting 3D cube whose dimensions are inferred from prior knowledge. This localisation pipeline is summarised in Figure 2. In the remainder of this paper, we will describe in more details our proposed method and evaluate it in a 10-fold cross validation, comparing it to a sliding window BOW using 2D or 3D SIFT features.



**Fig. 2.** Proposed pipeline for localising the fetal brain

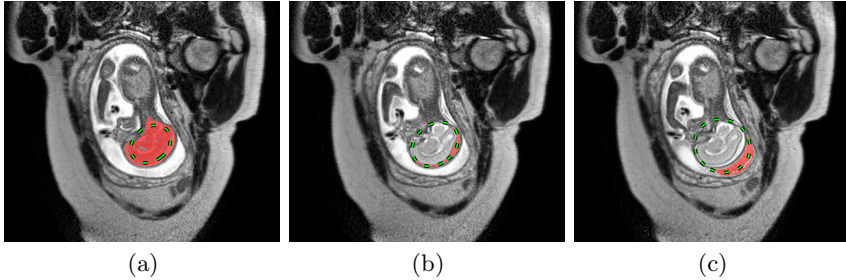
## 2 Method

### 2.1 Detection and Selection of MSER Regions

Maximally Stable Extremal Regions (MSER), introduced by Matas et al. [9], are a common feature detection method in computer vision. MSER regions can be defined as sets of connected pixels stable over a large range of intensity thresholds. Such regions can be characterised by homogeneous intensity distributions and high intensity differences at their boundary. Each region is defined by a seed pixel and a lower and upper intensity thresholds, the whole region resulting from a floodfill operation. In the case of fetal T2 MR images, MSER regions are well suited to the task of selecting candidate regions for the brain as the skull content appears much brighter than the surrounding bone and skin tissues.

In its first stages, our localisation pipeline proceeds slice by slice, working on 2D images, and starts by detecting candidate MSER regions  $\mathcal{R}_i$  which are sets of connected pixels. An ellipse  $\mathbf{E}_i$  is then fitted to each  $\mathcal{R}_i$  with a least-squares

minimization. As the shape of the brain is well approximated by an ellipsoid, the ellipses  $\mathbf{E}_i$  have the ability to recover the shape of the brain even if the detected  $\mathcal{R}_i$  only corresponds to a segment of cerebrospinal fluid (CSF, Figure 3.b), or amniotic fluid surrounding the fetal head (Figure 3.c). The ellipses  $\mathbf{E}_i$  are then filtered by size and aspect ratio using the gestational age combined with prior knowledge of the fetal development, namely the *occipitofrontal diameter*, *OFD*, and the *biparietal diameter*, *BPD* [12]. This prior knowledge is obtained from 2D ultrasound studies [12] and is discussed in more details in Section 3.



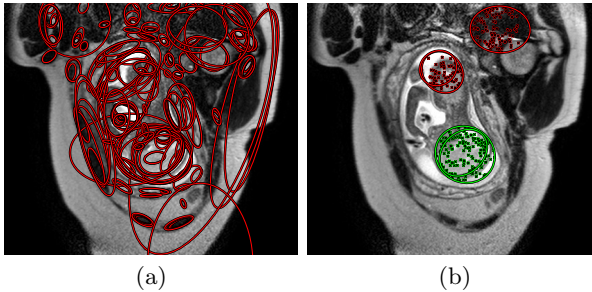
**Fig. 3.** Example MSER regions  $\mathcal{R}_i$  (red overlay) and their fitted ellipses  $\mathbf{E}_i$  (green dashes): (a) skull bone content, (b) cerebrospinal fluid, (c) amniotic fluid

## 2.2 Classification of Selected MSER Regions Using Bundled SIFT Features

MSER regions alone do not provide a descriptor that can be used for classification. As a versatile rotation invariant classification framework, we chose to use a Bag-of-Words model by computing for each region  $\mathcal{R}_i$  the histogram of the SIFT features (*Scale-Invariant Feature Transform* [8]) within the ellipse  $\mathbf{E}_i$ , similarly to the bundled SIFT features of Wu et al. [14]. In the Bag-of-Words method for image classification presented by Csurka et al. [3], SIFT features are extracted from training images and a set of words  $\mathcal{V}$  (*visual vocabulary*) is obtained through k-means clustering, each word being the centroid of a cluster. Each SIFT feature  $f_{\text{SIFT}}$  can then be associated to its nearest neighbour  $f_{\text{NN}} \in \mathcal{V}$  to build a histogram of words. A Support Vector Machine (SVM) classifier is then used to assign a class to this histogram.

SIFT features are the association of a keypoint corresponding to a local extrema in a scale-space Gaussian pyramid (*blob*) and a descriptor built from histograms of gradient orientations. The blob detection process provides scale invariance to the descriptor, whereas rotation invariance is obtained from the main gradient orientation over the blob. Rotation invariance helps accommodate the unknown orientation of the fetal brain, while scale invariance attenuates the variations due to gestational age.

In the **learning stage** of our pipeline, all SIFT features are extracted from the 2D slices of all training scans and clustered using a k-means algorithm. The  $N$  cluster centers form the vocabulary  $\mathcal{V}$ . Then, MSER regions  $\mathcal{R}_i$  are detected and selected (Section 2.1). The ellipses  $\mathbf{E}_i$  centered on the brain are kept as positive examples, while those further than  $OFD/2$  are kept as negative examples. For each  $\mathbf{E}_i$ , a histogram of bundled SIFT features is computed. After an  $L_2$  normalisation, the histograms are used to train an SVM classifier. As in [3], a linear kernel SVM has been used. In the **detection stage**, for each MSER region selected, a normalised histogram of SIFT features is computed, and the previously trained SVM is used to classify regions as brain or not-brain. Figure 4.a shows an example slice with all the candidate MSER regions, while Figure 4.b shows the result of the selection and classification processes.



**Fig. 4.** (a) All detected MSER regions. (b) MSER regions are filtered based on their size and aspect ratio, then classified according to their histograms of SIFT features (green for brain regions, red for not-brain)

### 2.3 RANSAC Fitting of a Cube

The 2D detection works reliably in mid-brain slices but not in peripheral slices of the brain. Estimating a 3D box is thus important for a reliable estimate of the entire brain region. As the set of candidate regions classified as brain by the SVM classifier may still contain outliers, we perform a RANSAC procedure [5] to find a best fitting position for a 3D axis-aligned bounding cube. We thus randomly select a small set of ellipses  $\mathbf{E}_i$  and use their centroid to define a bounding cube of width  $OFD$ . For a region  $\mathcal{R}_i$  to be considered an *inlier*, it must be completely included in this cube and the center of the ellipse  $\mathbf{E}_i$  must be at a small distance from the cube center. The cube position is then refined by taking the centroid of these inliers. This process is repeated a predefined number of times, typically 1000 times, and the cube with the largest number of inliers is selected.

## 3 Experiments

**Implementation.** To gain prior knowledge on the expected size of the brain knowing the gestational age, we used a 2D ultrasound study from Snijders and

Nicolaides [12] performed on 1040 normal singleton pregnancies. The 5<sup>th</sup> and 95<sup>th</sup> centile values of *OFD* and *BPD* have been used to define the acceptable size and aspect ratio for the selected MSER regions (Section 2.1). The size of the detected bounding box in the robust fitting procedure (Section 2.3) has been set to the 95<sup>th</sup> centile of *OFD* in order to contain the whole fetal brain. The size *N* of the BOW vocabulary (Section 2.2) has been set to 400. The automated detection process takes less than a minute on a normal PC.

**Data.** Fetal MRI was performed on a 1.5T MRI system, T2-weighted dynamic ssTSE images were obtained with the following scanning parameters: TR 15000ms, TE 160ms, slice thickness of 2.5mm, slice overlap of 1.5mm, flip angle 90°. The in-plane resolution varies between 0.94 and 1.25mm, with the number of interleaved packets ranging from 4 to 6. Our database includes 59 healthy fetuses whose gestational age range from 22 to 39 weeks, with 5 fetuses scanned twice and 1 three times, amounting to a total of 66 datasets. Each dataset consists of 3 to 8 scans acquired in three standard orthogonal planes, representing a total of 117 sagittal, 113 coronal and 228 transverse scans. We performed a 10-fold cross validation, with 39 training subjects and 20 testing subjects for each fold. For the ground truth, bounding boxes have been tightly drawn around the brain.

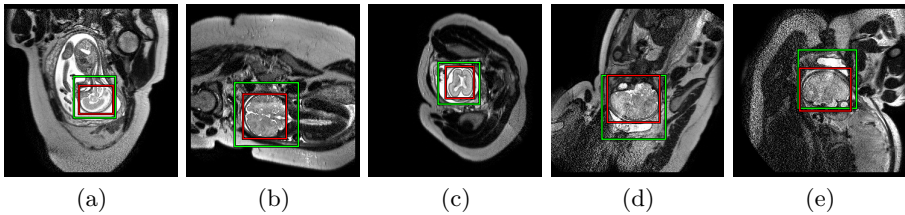
**Results.** The brain localisation results are summarised in Table 1, whereas examples of detected bounding boxes and their corresponding ground truth are shown in Figure 5 and in the supplementary material.<sup>1</sup> We compared our method against sliding a window of fixed size *OFD* with a Random Forest classifier on histograms of 2D or 3D SIFT features, using the extension to 3D of SIFT proposed in [13]. For each stack of slices, we measured the distance between the center of the ground truth bounding box and the detected bounding box. Similarly to [6], we defined a correct detection as 70% of the brain being included in the detected box. 2D SIFT features performed better than 3D SIFT features, which can be attributed to the inconsistency of the 3D data, whereas the bundled SIFT drastically improved the localisation accuracy with an error below 5.7mm in more than 50% of cases. This greater precision comes from applying a classifier only at specific locations (MSER regions) instead of all pixels, thus resulting in a more localised probability map. This localisation of the center of the brain shows improved results compared to [6] who reported a median error of 10mm in the detection of 6 landmarks corresponding to centroids of fetal head tissues. However, contrary to [6], the orientation of the brain is not determined. Our method is more general than [1] as it does not rely on localising the eyes. There has been no false detection or missed detection with bundled SIFT, with a worst case error of 25mm presented in Figure 5.e. In 85% of cases, the detected bounding box contains entirely the ground truth bounding box.

The prior knowledge gained from the gestational age plays an important role in disregarding most of the detected MSER regions. Indeed, on average during

<sup>1</sup> [www.doc.ic.ac.uk/~kpk09/MICCAI-451.mp4](http://www.doc.ic.ac.uk/~kpk09/MICCAI-451.mp4)

**Table 1.** Detection results averaged over the cross validation, all orientations combined

Centiles	Error (mm)		
	<i>2D SIFT</i>	<i>3D SIFT</i>	<i>Bundled SIFT</i>
<i>25<sup>th</sup></i>	10.9	14.8	<b>4.0</b>
<i>50<sup>th</sup></i>	15.5	20.8	<b>5.7</b>
<i>75<sup>th</sup></i>	20.5	30.4	<b>8.4</b>
<b>Detection</b>	98%	85%	<b>100%</b>
<b>Complete brain</b>	38%	23%	<b>85%</b>

**Fig. 5.** Examples of detected bounding boxes (green) and ground truth (red). a), b) and c) are respectively sagittal, coronal and transverse acquisitions of different subjects. d) and e) show detections with the largest error, which can be attributed to the presence of the maternal bladder near the fetal head.

the 10-fold cross validation, 98% of the candidate regions are disregarded in the first stage of the pipeline. As a comparison, the SVM classifier further disregards 41% of the remaining candidates, and the RANSAC procedure removes 18% outliers. No distinction has been made between sagittal, transverse or coronal acquisitions, which is another advantage of the proposed method.

When using normal growth charts to define a realistic range of sizes of the brain, our goal is to remove improbable detections while still allowing a large variation in brain size. From standard growth charts, extending to include 99.6% of subjects corresponds to a 1 week error. Simulating extreme growth restriction by adding a 5 weeks error in all gestational ages, the detection rate is still 91%.

## 4 Conclusion

We presented a novel automatic localisation method for the fetal brain in MRI. Proceeding slice by slice, MSER regions are first detected and filtered by size and aspect ratio before being classified using histograms of SIFT features. An expected size of the brain is inferred from the gestational age and prior knowledge of the fetal development. Finally, a 3D bounding cube is fitted to the selected regions with a RANSAC procedure. The method is not specific to the scan orientation, with a median distance error of 5.7mm from the ground truth. Further work will be required to integrate this brain localisation method in a motion correction pipeline, such as automatically masking out maternal tissues.

## References

1. Anquez, J., Angelini, E., Bloch, I.: Automatic Segmentation of Head Structures on Fetal MRI. In: ISBI, pp. 109–112. IEEE (2009)
2. Carneiro, G., Georgescu, B., Good, S., Comaniciu, D.: Detection and Measurement of Fetal Anatomies from Ultrasound Images using a Constrained Probabilistic Boosting Tree. *IEEE Transactions on Medical Imaging* 27(9), 1342–1355 (2008)
3. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual Categorization With Bags of Keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, vol. 1, p. 22 (2004)
4. Feng, S., Zhou, S., Good, S., Comaniciu, D.: Automatic Fetal Face Detection from Ultrasound Volumes via Learning 3D and 2D Information. In: CVPR, pp. 2488–2495. IEEE (2009)
5. Fischler, M., Bolles, R.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM* 24(6), 381–395 (1981)
6. Ison, M., Donner, R., Dittrich, E., Kasprian, G., Prayer, D., Langs, G.: Fully Automated Brain Extraction and Orientation in Raw Fetal MRI. In: Workshop on Paediatric and Perinatal Imaging, MICCAI, pp. 17–24. Springer (2012)
7. Jiang, S., Xue, H., Glover, A., Rutherford, M., Rueckert, D., Hajnal, J.: MRI of Moving Subjects using Multislice Snapshot Images with Volume Reconstruction (SVR): Application to Fetal, Neonatal, and Adult Brain Studies. *IEEE Transactions on Medical Imaging* 26(7), 967–980 (2007)
8. Lowe, D.: Object Recognition from Local Scale-invariant Features. In: ICCV, vol. 2, pp. 1150–1157. IEEE (1999)
9. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In: BMVC, pp. 384–393 (2002)
10. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast Multiple Organ Detection and Localization in Whole-body MR Dixon Sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011)
11. Rousseau, F., Glenn, O., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D., Barkovich, J., Studholme, C., et al.: Registration-based Approach for Reconstruction of High Resolution In Utero Fetal MR Brain Images. *Academic Radiology* 13(9), 1072–1081 (2006)
12. Snijders, R., Nicolaidis, K.: Fetal Biometry at 14–40 Weeks’ Gestation. *Ultrasound in Obstetrics & Gynecology* 4(1), 34–48 (2003)
13. Toews, M., Wells III, W.M.: Efficient and Robust Model-to-Image Alignment using 3D Scale-Invariant Features. *Medical Image Analysis* (2012)
14. Wu, Z., Ke, Q., Isard, M., Sun, J.: Bundling Features for Large Scale Partial-duplicate Web Image Search. In: CVPR, pp. 25–32. IEEE (2009)
15. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber Heart Modeling and Automatic Segmentation for 3D Cardiac CT Volumes using Marginal Space Learning and Steerable Features. *IEEE Transactions on Medical Imaging* 27(11), 1668–1681 (2008)