PVR: Patch-to-Volume Reconstruction for Large Area Motion Correction of Fetal MRI

Amir Alansary, Martin Rajchl, Steven G. McDonagh, Maria Murgasova, Mellisa Damodaram, David F.A. Lloyd, Alice Davidson, Mary Rutherford, Joseph V. Hajnal, Daniel Rueckert, and Bernhard Kainz

Abstract—In this paper we present a novel method for the correction of motion artifacts that are present in fetal Magnetic Resonance Imaging (MRI) scans of the whole uterus. Contrary to current slice-to-volume registration (SVR) methods, requiring an inflexible anatomical enclosure of a single investigated organ, the proposed patch-to-volume reconstruction (PVR) approach is able to reconstruct a large field of view of non-rigidly deforming structures. It relaxes rigid motion assumptions by introducing a specific amount of redundant information that is exploited with parallelized patch-wise optimization, super-resolution, and automatic outlier rejection. We further describe and provide an efficient parallel implementation of PVR allowing its execution within reasonable time on commercially available graphics processing units (GPU), enabling its use in the clinical practice. We evaluate PVR’s computational overhead compared to standard methods and observe improved reconstruction accuracy in the presence of affine motion artifacts compared to conventional SVR in synthetic experiments. Furthermore, we have evaluated our method qualitatively and quantitatively on real fetal MRI data subject to maternal breathing and sudden fetal movements. We evaluate peak-signal-to-noise ratio (PSNR), structural similarity index (SSIM), and cross correlation (CC) with respect to the originally acquired data and provide a method for visual inspection of reconstruction uncertainty. We further evaluate the distance error for selected anatomical landmarks in the fetal head, as well as calculating the mean and maximum displacements resulting from automatic non-rigid registration to a motion-free ground truth image. These experiments demonstrate a successful application of PVR motion compensation to the whole fetal body, uterus and placenta.

Index Terms—Motion Correction, Fetal Magnetic Resonance Imaging, GPU acceleration, Image Reconstruction, Super-Resolution

I. INTRODUCTION

THE advent of single shot fast spin echo (ssFSE) T2-weighted sequences has enabled spin echo image formation principles [1] in magnetic resonance imaging (MRI) to play an essential role in fetal diagnosis [2] and research [3]. In particular, cases for which ultrasound (US) fails to acquire conclusive image data benefit from fetal MRI [4], [5]. Recent advances in motion compensation for fetal MRI [6] facilitate advanced image-based diagnostics and lead to novel insights about the human development.

Fetal MRI enables an ability to distinguish between individual fetal structures such as brain, lung, kidney and liver, as well as pregnancy structures such as the placenta, umbilical cord and amniotic sac. It provides improved visualization and structural information of the fetal anatomy, which enables to study abnormalities during pregnancy such as neuro-developmental disorders [7], placental pathologies [8], fetuses with congenital lung masses [9], and conjoined twins [10]. MRI is considered to be safe after the first trimester [11] for 1.5T [12] and 3T [13] without the use of contrast agents, which may have teratogenic effects. Furthermore, this technology paves the way for researchers and clinicians to analyze correlations between childhood development and prenatal abnormalities.

During image acquisition the fetus is not sedated and moves freely as well as the mother breathes normally. As a result, movements are likely to corrupt the scans, hiding pathology and causing overlap between different anatomical regions. In order to limit these artifacts, fast scanning sequences such as ssFSE [14] allow for the rapid acquisition of single slices at high in-plane resolution in a large field of view and good tissue contrast of the uterus. However, when acquiring a 3D volume through a stack of slices, inter-slice artifacts in the out-of-plane views are highly likely. Consequently, this restricts reliable diagnostics to individual slices in the current clinical practice.

Fig. 1 depicts a typical example of motion related artifacts in a fetal ssFSE scan. The observed motion (c.f. Fig. 1 b & c) is of unpredictable nature and consists of a combination of maternal respiration, bowel and fetal movements.

Slice-to-volume registration (SVR) combined with super-resolution image reconstruction techniques [15] can be applied to compensate motion between single slices by reconstructing a high-resolution (HR) image from multiple, overlapping low-resolution (LR) images, as shown in Fig. 2. To provide a sufficiently high number of samples for such an approach, multiple stacks of 2D-slices need to be acquired, ideally in orthogonal orientations. A simple LR → HR reconstruction model [15] can be formalized as:

\[ Y_i = W_i X + n_i \quad \text{for } 1 \leq i \leq N, \]

where \( Y_i \) denotes the \( i \)-th LR image of total \( N \) images, and \( X \) is the HR image. The matrix \( W_i \) combines motion, subsampling and degradation effects: \( W_i = DBT_i \), where \( D \) is the sub-sampling matrix, \( B \) is the blurring matrix, and \( T_i \) is the transformation matrix of observation \( i \). The noise of observation \( i \) is represented by \( n_i \). LR images can be considered as a down-sampled, motion corrupted, blurred, and noisy version of the HR image. The resulting reconstruction can be divided into two main parts: motion correction (estimating \( W_i \)) and super-resolution (estimating \( X \)). Image registration can be used to estimate motion, interpolation to obtain a uniformly spaced HR image, and regularized super resolution with automatic...
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2017.2737081, IEEE Transactions on Medical Imaging

Practical limitations: SVR methods have been successfully employed to address these problems in fetal MR and are typically applied to small regions and organs with rigid body characteristics that are identified by manual annotations [16], [17], [18], [19] or less precise, automated segmentations [20], [21], [22]. Such approaches are prohibitive to whole body and uterus reconstruction because of the assumption of rigid motion in the 2D to 3D registration step of SVR. As a result, different areas in each slice that are likely to move in different directions will break this assumption, e.g., the head and thorax. Further, an extension of 2D-3D registration to include non-rigid deformations is only well defined with each slice and not well-constrained in 3D. Current SVR approaches will fail in the presence of non-rigid deformations and unpredictable organ shapes. This restricts the application of SVR to regions that are manually or automatically annotated. Thus, most of the previous SVR methods have been limited to the fetal brain as the main region of interest for fetal reconstruction due to the high incidence of neuro-developmental disability in premature infants. Only recently, [23], [24] proposed a motion corrected 3D reconstruction of fetal thoracic structures from prenatal MRI. Moreover, SVR is computationally expensive due to the exponential increase of computation with the size of the target area. This leads to prohibitive post-processing times in the clinical practice. Parallelized implementations [25] can address run-time problems, however, methodologically SVR is still restricted to small, rigid body areas.

Reconstruction of large-scale anatomy: MRI has further been shown to be useful for the evaluation of the whole uterus and structures like the placenta. During both normal and high-risk pregnancies, the whole uterine appearance and the condition of the placenta are considered to be an indicator for fetal health after birth [26]. Placental functions affect the birth weight as it controls the transmission of nutrients from the maternal to the fetal circulation [27]. However, the whole fetal body and secondary uterine parts can be inherently inconsistent. Different fetal body parts can move independently from the uterus. This makes the application of SVR and 2D-3D registration to the full uterus impossible in the presence of fetal motion and maternal respiration.

Besides, multiple births is a case where classical SVR pipelines, based on preprocessing steps to identify consistent rigid regions, will likely fail. The presence of multiple instances of the same fetal structure is usually not considered in previous methods. Therefore, a fully automatic motion correction method for the whole uterus, as it is presented in this paper, is very desirable and will enable the application of standard 3D image analysis techniques, e.g., [28], [29].

Related Work: Most motion compensation approaches for fetal MRI are based on SVR techniques that aim to obtain a motion-free and high resolution volume of a fetal target region. Registration of individual 2D slices with a higher resolution 3D volume [30] is the core approach of these algorithms. SVR methods assume that all acquired images are centered at a specific organ (e.g., brain, thorax) and cover three orthogonal image directions. Fig. 3 shows the core elements of SVR and the contribution of previous frameworks from the literature.

The first SVR-based reconstruction framework for fetal MRI was introduced by Rousseau et al. [16]. It includes steps to correct 2D slice misalignments, intensity inhomogeneity distortions, and reconstructs an isotropic HR fetal brain from sets of LR images. Motion correction is done by applying a global 3D rigid alignment between the LR images using one image as a reference to define the global coordinate system. Then every slice is aligned to the initial reconstructed HR volume. Normalized mutual information is maximized using the gradient ascent method for both registration steps. A narrow Gaussian kernel is applied as a point spread function (PSF) for volume reconstruction and empty voxels are filled using the mean of the surrounding voxels. The image contrast is corrected using one LR image as a reference.
Jiang et al. [31] introduced the acquisition of many thin slices to provide sufficient sampling of the region of interest. Cross correlation is used as a cost function for the SVR steps assuming that the data have consistent contrast properties. After that, multilevel B-splines are applied to the volumetric reconstruction for data interpolation, which has the advantage of reducing blurring of the reconstructed image supported by including the thin slices.

Kim et al. [17] proposed a method for slice intersection motion correction (SIMC) of multi-slice MRI for 3D fetal brain image formation. The method is based on slice-to-slice registration using spatially weighted mean square intensity differences (MSD) of the signal between slices as an energy, assuming that the MRI contrasts are identical. Maternal tissues are excluded from the energy computations using a windowing function of a parametric ellipsoid model. Similar to [16], temporally adjacent slices are grouped together then divided into half iteratively. The splitting process is performed using discrete cosine basis functions.

Gholipour et al. [18] were the first to introduce a mathematical model for super-resolution (SR) volume reconstruction from slice acquisitions of fetal brains. The main difference to previous methods is that it includes knowledge of the slice acquisition model and the SR reconstruction is performed based on maximum likelihood and a robust M-estimation minimization for an error norm function. A Tikhonov regularization term is added to the cost function in order to enforce a solution when the number of acquired samples is not high enough for solving the reconstruction problem.

Rousseau et al. [32], [33] proposed to use a variational regularization including an approximation of Total Variation (TV) to better preserve edges. Later, Tourbier et al. [34] introduced an adaptive regularization by applying novel fast convex optimization techniques to design an efficient optimization algorithm for the super-resolution problem using edge-preserving TV regularization.

Murgasova et al. [19] were able to reconstruct the fetal brain using intensity matching and complete outlier removal. The main steps of their reconstruction method are: (i) 3D registration of the acquired stacks using a template stack; (ii) extracting region of interest (the fetal head) from all the stacks; (iii) intensity matching and bias correction between the slices based on an EM framework, where the differential bias fields and slice-dependent scaling factors are estimated during the reconstruction; (iv) motion correction using [16] based on the normalized cross correlation as a similarity measure and an approximated 3D Gaussian PSF similar to [31]. A posterior probability is used to define the inlier and outlier voxels within the EM framework in order to remove the motion-corrupted artifacts and misaligned data. Blurring in reconstructed images is reduced by integrating edge-preserving regularization based on anisotropic diffusion within the SR reconstruction framework.

Kainz et al. [25] developed a fast multi-GPU accelerated implementation for the method presented in [19], which is based on 2D-3D registration, SR with automatic outlier rejection and an optional intensity bias correction. They extended the reconstruction framework by automatically selecting the stack with least motion as the reference stack and using a fully flexible and accurate PSF instead of approximated functions. Using a multi-GPU framework enabled the SR reconstruction process to be approximately five to ten times faster than using a multi-CPU framework.

To our knowledge, modeling non-rigid transformations with multiple rigid transformations and without an initial registration target has only been preliminarily explored for fetal MRI in [35]. Other works regarding non-linear 3D-3D registration between an outlier-free, regularly sampled source and target volume, constrain parts of an image to move independently and rigidly (e.g., bones through manual segmentation and landmark annotation), while allowing other parts to deform [36]. In 3D-3D settings globally continuous 3D deformation fields, which should not be assumed for 2D-3D fetal motion compensation problems, can be approximated by fusing multiple local rigid transformations as shown in [37], [38].

**Contributions:** In this paper we propose and evaluate a new paradigm for motion correction based on SVR and flexible subdivision of the input space into overlapping, highly redundant and partly rigid image patches [35], thus solving the motion compensation problem for large field of view reconstructions. We split the input into small overlapping areas and find these, which contain rigid components. This allows to iteratively learn their consistency compared to a global reconstruction optimization volume. Corrupted and inconsistent data is automatically identified and excluded using robust statistics. Our approach facilitates the automatic reconstruction of whole collections of motion corrupted stacks without the need of corresponding image segmentations. By treating rigid image patches as piecewise constant segments of organs further allows limited correction of non-rigid tissue motion. We also evaluate the non-rigid deformations that may result from PVR using anatomical landmarks annotated in the fetal head and non-rigid registration. The presented patch-to-volume reconstruction (PVR) finds automatically rigidly connected areas,
which can be used as segmentation prior for further refinement using conventional SVR in small regions of interest. In contrast to [25], we further introduce a multi-scale patch approach and thoroughly evaluate the reconstruction quality of the whole uterus including the fetal brain and placenta. We test the breaking points of SVR and variations of PVR on synthetically motion corrupted brain phantom data. The presented approach is the only currently available method that is able to reconstruct fetal organs and detailed 3D volumes of secondary, non-rigidly moving structures such as the placenta.

II. METHOD

SVR-based motion compensation methods make use of the assumption that rigid regions, e.g., brain and thorax, of 2D input slices deforms rigidly, where a global 3D volume is reconstructed by iteratively registering these 2D input slices. We propose to increase the granularity of the input data by using 2D image patches of arbitrary shape instead of whole slices for SVR reconstruction. We explore square patches and dilated superpixels [39] for the definition of the patch shape. Superpixels provide a method to define semantically meaningful regions while reducing the required data redundancy and computational overhead.

PVR relies on the fact that certain regions of the scanned anatomy are rigid and can be reconstructed with SVR algorithms. However, unlike SVR, it is fully automatic and provides a full field of view reconstruction. Data consistency is obtained by oversampling a region of interest at different scan orientations. Robust statistics can be used to identify misregistered or heavily corrupted data [18], [19]. Fig. 4 depicts a schematic overview of the proposed PVR framework.

Input data and initialization: A template stack is either randomly or automatically chosen from available input stacks by detecting the stack with fewest motion artifacts [25]. Global intensity matching is applied to normalize intensity values of all input images followed by global 3D-3D alignments to spatially initialize the reconstruction target. Input data can be represented as stacks of 2D images (patches) consisting of

\[
Y = \{y_s | s \in S\}, \quad (2)
\]

where \(y_s\) is a patch of arbitrary 2D-shape and indexed by the location \(s\). \(S\) is the set of all locations in all \(p\) stacks, \(S = \{s_1, s_2, \ldots, s_M\}\), and \(M\) is total number of patches.

Patch extraction: In the simplest, naîve case the shape of \(y_s\) is square defined via its edge length \(a\) and stride \(\omega\). This definition is generally applicable to any kind of over-sampled motion corrupted data. If \(a\) and \(\omega\) are fixed, no prior knowledge about the data is assumed. However, ideally each \(y_s\) corresponds to a meaningful subregion of the volume in which motion can be characterized as rigid. Typically, square patches are overlapping to provide redundant representations of the same locations. The computational expense of such an approach increases with patch number and size \(a\). Additional consideration must include the inherent trade-off between \(a\) and the assumption of it containing rigid motion.

An alternative to naîve shape definitions of \(y_s\) is to find correlation between voxel locations and their neighbors, which can be found by unsupervised image segmentation techniques such as superpixels (SP) [39]. These techniques allow similarly-sized segments to be obtained from local intensity information (see Fig. 5) instead of employing dense sampling of overlapping patches, enabling the image reconstruction with fewer but more useful data blocks. Further, reducing the total amount of required data blocks for reconstruction lowers the computational overhead, positively impacting the overall runtime. Additionally, larger rigid areas require less computational effort for image registration and super-resolution, and more importantly less dependency on inherent image data parameters (e.g., voxel spacing, organ size, subject size).

While there are several techniques for generating SP in the literature [39], [40], [41], a fast and efficient SP approach is desirable for the clinical practice. Simple linear iterative clustering (SLIC) [39] obtains regular SP based on minimizing the distance \(D\) between the centroids of SP with an initial step size \(a\). \(D\) is defined as:

\[
D = \sqrt{d_c^2 + \left(\frac{d_s}{\alpha}\right)^2 t^2}, \quad (3)
\]

where \(d_c\) and \(d_s\) are the intensity and spatial Euclidean distances that are controlled by the adaptive compactness parameter \(t\) for each SP. In our previous work [35], we use a defined number of SP centers \(N_{SP} \approx k\sqrt{(n/2)}\), where \(n\) is the number of voxels in an image of size \(d_x, d_y\) and \(k\) a constant factor. Here, we use a regular grid for initialization with a constant distance \(a\) between SP centroids so that \(N_{SP} = \frac{d_x}{a} + \frac{d_y}{a}\) and dilate by \(\gamma\%\). Thus, instead of providing a target number of SP, \(N_{SP}\) is defined by the image size and \(a\) controls the relative size of the SP.

Multi-scale patches: Although larger patch regions are less likely to include rigidly connected regions, they may perform better during 2D-3D registration due to the additional contextual information of each patch. In contrast, smaller patch sizes are more likely to represent rigidly deformed regions, but provide less contextual information, potentially affecting the 2D-3D registration. A good trade-off between the size of the patch region and the likelihood of rigid motion needs to be found. Here, we propose the use of multi-scale patches for reconstruction to exploit the advantages of different patch sizes. We represent input data as stacks of 2D patches:

\[
Y_i = \{y_{si} | s \in S_i\}, \quad (4)
\]

where, instead of using the same \(Y\) as a unique input, a different scale of \(Y_i\) is used for each iteration \(i\) at \(\gamma\%\) of its original size. Similar to Eq. 2, \(S_i\) is the set of all locations in all \(p\) stacks but with different size for each iteration \(i\). This is done by re-calling the patch extraction module with a different patch size (see Fig. 4). Additionally, to increase contextual information for estimating the transformations, we compute overlapping \(y_{si}\) superpixel patches and dilate each by \(\gamma\) pixels using a flat structuring element \(b\) with a fixed neighborhood (26 pixels in our case), hence \(y_{si} = y_{si} \oplus b\). Smaller \(\gamma\) values result in a faster reconstruction. Ideally \(\gamma\) is > 50% of \(a\), ensuring that every pixel is covered by multiple samples.

Patch to volume registration: An HR-image \(X\) is reconstructed from a number of motion corrupted patches \(y_s\) using
Fig. 4. A schematic and modular overview of the proposed patch-to-volume reconstruction (PVR) framework. The key parts are 3D-3D registration, patch extraction, 2D-3D registration, super-resolution, and EM-based outlier removal. Core contributions of PVR are written in red and marked with asterisk.

Fig. 5. An illustrative figure showing both square patches and superpixels methods for the patch extraction step. A 2D superpixel shows more flexibility than a square patch in extracting rigid regions or similar voxels. In practice, superpixels are dilated with few pixels to include some contextual information in order to increase the accuracy of the patch to volume registration step.

2D-3D registration-based super-resolution similar to [19], [25], where an accurate PSF calculation is used to generate a gradually improving approximation of $X$ and further employed to initialize the 2D-3D registration and computation of robust statistics. In [19], [25], the PSF is a sinc function for the in-plane and is the slice profile for the through-plane, measured for the employed MRI sequence (ssFSE), according to [31].

We employ an implementation of the PSF function by applying a Taylor series for a better approximation of small values close to zero. The Taylor series approximation of the sinc function is defined as $\text{sinc}(R) = 1 - \frac{R^2}{\pi} + \frac{R^4}{6\pi^2} - \frac{R^6}{120\pi^4} + \cdots$. The proposed approximate PSF achieves qualitative improvement of the reconstructed image compared to the sinc implementation. An example from the first iteration of a fetal brain reconstruction is shown in Fig. 6. This is expected since Taylor series approximation is common for computations requiring a high level of accuracy close to the machine precision limits of floating point computations [42].

During the optimization process, individual 2D patches are continuously rigidly registered to the current 3D reconstruction of $X$ and re-integrated into $X$ using iterative super-resolution with gradient descent optimization. Any similarity metric can be used as a cost function for the registration step such as mutual information [16], [19], cross correlation (CC) [31], [25], or mean square intensity differences [17], [18]. Choosing the best similarity metric for reconstruction depends on the input data. CC has been found to be effective for input data with similar intensity distribution [43]. In our experiments, we employ CC as the similarity metric for 2D-3D registration, after rescaling the intensities between the input stacks.

Super-resolution: Given that the position and orientation of each patch $y_s$ relative to $X$ is known at any point, patches $y_s^*$ can be simulated from sampling the current approximation of $X$ using the PSF. $X$ can subsequently be iteratively improved by minimizing the error between $y_s$ and $y_s^*$. Gradient decent is used for the optimization. The optimization has to
be regularized to prevent amplification of registration error, noise and under-sampling artifacts. Adaptive regularization is applied to reduce smoothing effects. Therefore the objective function is

$$\sum_i P(v_i) (y_{s_i} - y_{s_i}^*)^2 + \lambda \sum_i \sum_d \phi \left( \frac{x_i + d - x_i}{\delta |d|} \right),$$  

(5)

with $\phi(t) = 2\sqrt{1 + t^2} - 2$, voxel locations $x \in X$, $d$ a vector between the index of a voxel and one of its 26 neighbors, and $\delta$ a parameter that controls when the gradient between neighboring voxels is considered to be an edge. For our data a gradient magnitude between 100 and 400 yields good results. The regularization term represents edge preserving filtering [44] and is used similarly as presented in [25]. $\lambda$ is decreased after each iteration. Following the experimental findings in [19] a suitable starting value for $\lambda$ is 0.8$\delta^2$, halved after every iteration.

**EM-Based outlier removal:** Correctly registered patches $\hat{y}_s$ should provide a higher contribution to the final reconstruction, presenting a low error $e$ when compared to the original image data. [18] initially introduced an approach to account for outliers during super-resolution based on Huber function statistics. Similar to [19], we employ the expectation maximization algorithm for outlier removal by classifying $\hat{y}_s$ and the included pixels into an inlier and outlier class. A zero-mean Gaussian distribution $G_e(c)$ with variance $\sigma^2$ is used for the inliers and a uniform distribution with constant density

$$m = \frac{1}{\max(e) - \min(e)}$$

(6)

for the outliers. This makes use of available, highly redundant information (i.e., overlapping $\hat{y}_s$), to find partly matching patches and to depreciate or fully reject erroneous voxels. We aim to maximize the log-likelihood for each patch

$$y_s \log P(Y, \Phi) = \sum \log P(v|c, \sigma)$$

(7)

to be part of a region of rigid motion. $\Phi$ is the current estimate of the reconstructed volume $X$, the variance $\sigma^2$ of the errors $e$, and the proportion of correctly matched voxels $c$. The posterior probability for a pixel $\in \hat{y}_s$ being identified as inlier is

$$p = \frac{G_e(c)}{G_e(c) + m(1-c)}.$$  

(8)

We perform the updates of $c$ and $\sigma^2$ according to [19]:

$$\hat{p} = \sqrt{\frac{\sum p^2}{N}},$$

(9)

where $N$ is the number of pixels in $\hat{y}_s$. We further define an inlier and outlier probability for each $\hat{y}_s$ and exclude it from processing if classified as an outlier (e.g., if it contains structures moving in opposite directions during scanning, such as the fetal head and thorax). Only if information in $\hat{y}_s$ is consistent with the originally acquired data, will the registered patch contribute to the SR reconstruction of $X$.

**Identification of rigid regions and SVR refinement:** The rigidity of regions is measured by keeping track of the probability $p$ of each pixel of every $\hat{y}_s$. This allows the identification of locations best fitting the rigid 2D-3D registration constraints. Candidate regions, that contain rigid motion components, can be identified by combining $p$ and $\hat{p}$ into a 3D volume utilizing the same PSF as used in the reconstruction. This can further initialize the rigid SVR reconstruction and visualize the data uncertainty during reconstruction.

### III. IMPLEMENTATION

**Parallelization:** The high data redundancy required for the proposed approach makes conventional single threaded implementation practically not feasible. Computational complexity of PVR is exponentially higher than SVR, depending on the employed patch overlap. For optimal performance we implemented our approach via General-Purpose Programming on Graphics Processing Units (GPGPU) using the Compute Unified Device Architecture (CUDA, NVIDIA, Santa Clara, CA) language [45], [46]. CUDA is a highly evolved single instruction multiple data (SIMD) programming language, which allows a large part of the proposed framework to be mapped onto GPU hardware. Currently, CUDA is the only high-level general purpose GPU language that provides bi-directional texture access via surfaces in a kernel, which is essential for the efficient implementation of certain parts of our framework. In this section we discuss the key implementation details.

We use a modular design to allow experimentation with the separate components of the algorithm. An overview of this design is shown in Fig. 7. The modules are encapsulated in a CUDA library, which can be used independently from the instantiating framework. We employ the successor of IRTK for interfacing with medical image data.

PVR is parallelized on three levels:

**I. Patch-level:** Individual patches are mapped to blocks of a CUDA computing grid and the contained voxels are mapped to individual threads. Depending on the used GPU hardware, patch processing can also be mapped directly to the computing grid, such that each thread works on a complete patch (limited by the employed patch size). The resulting thread divergence provides opportunities for advanced GPU scheduling strategies [47] and for a direct translation of optimization strategies for image registration, for example patch-wise gradient descent.

**II. Voxel-level:** For the parallelization of PSF-based super-resolution and robust statistics we follow a similar three-fold procedure definition as used in [25]. The voxels within each patch are processed using kernel level parallelization and parallel pixel-volume, volume-pixel, and volume-volume procedures are applied.

**III. Patch-batch:** PVR scales to multiple GPUs through distributing independent subsets of patches over the desired number of devices. Synchronization is done through averaging of the resulting sub reconstruction volumes on the master GPU. Initial 3D-3D registration is performed on a single master GPU, which allows optimal coalesced memory access.


**Availability of source code:** We provide the source code of a C++/CUDA implementation of the proposed method, including parallelization strategies, in a publicly available software repository\(^2\). The source code for the implementation of PVR is licensed under MIT license.

**IV. Evaluation & Experiments**

**Evaluation of adult brain MRI reconstruction:** We evaluate the performance and limitations of PVR in terms of accuracy and robustness with synthetic deformations of adult brain data. Similar to [18], an isotropic 1 mm\(^3\) T2-weighted adult brain phantom with no noise obtained from the Brainweb database [48] is used for this experiment. Three experiments with different deformations are performed, namely:

I. **Rigid Transformation:** Synthetic rigid motion artifacts are generated by translating the axial, sagittal and coronal images with \(\pm d_x, mm\) in the \(x\)-direction.

II. **Bulk Transformation:** Synthetic bulk motion artifacts are generated by rotating the upper half of the original image with \(\pm \theta_z\) around the \(z\)-axis.

III. **Non-rigid deformation:** Synthetic non-rigid motion artifacts are generated by skewing (shearing) the axial, sagittal, coronal images using:

\[
T_{xyz} = \begin{bmatrix}
1 & S_{xy} & S_{xz} & 0 \\
S_{yx} & 1 & S_{yz} & 0 \\
S_{zx} & S_{zy} & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix},
\]

where we use one combined skewing value \(S_{xyz} = \tan(\pm \theta_{xyz})\) in the \(xyz\)-direction.

Following image deformation, a motion-corrupted 3D stack is constructed by sampling 2D images from both motion-free and motion-corrupted stacks in an interleaved manner similar to fetal MRI acquisition [49]. Where for the bulk motion experiment a 3D stack is constructed by only rotating the upper half of the brain. Three stacks are used for the reconstructions where each stack is sampled with a voxel size of 1.25x1.25x2.5 mm\(^3\). We use standard axial, sagittal, and coronal orientations as shown in Fig. 8. An HR image with isotropic voxel size 1.25 mm\(^3\) is reconstructed using SVR [25], square patch- and superpixel-based PVR.

**Evaluation of fetal organ MRI reconstructions:** Evaluating the quality of reconstructed fetal MRI is challenging due to the absence of motion-free ground truth data. For this purpose, we introduce a novel approach for the evaluation problem based on the originally acquired slice images. Assuming that 2D in-plane patches extracted from the original stacks contain no motion artifacts, we use them as gold standard and compare with corresponding simulated patches from the reconstructed volume. Evaluation metrics (see Sec. IV-A) are computed between the reconstructed input stacks and the final motion corrected image and averaged over the whole volume. The fetal brain is typically used to assess the quality of reconstruction as it moves rigidly, fulfilling the rigid motion assumption for SVR-based methods in the 2D-3D registration step. However, soft tissue organs such as the placenta deform non-rigidly. For this reason, we additionally chose to reconstruct the placenta and whole uterus as challenging test cases for PVR and SVR.

**Evaluation metrics:** We employ the following metrics for measuring the quality of the reconstructed image:

- **Cross-correlation (CC):** To measure the similarity between the intensities of input \(I(i,j)\) and reconstructed image \(I'(i,j)\) at the location \((i,j)\), which is defined as:

\[
CC = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{(I(i,j) - I_n)(I'(i,j) - I_n)}{\sigma_I \sigma_{I'}}
\]

where \(N\) and \(M\) are the dimensions of a 2D slice.

- The peak signal-to-noise ratio (PSNR) is used to measure the error introduced by motion and is based on the mean squared error (MSE) between the original 2D in-plane patch and the reconstructed image. PSNR is defined as:

\[
PSNR = 10 \log \frac{P_{\text{max}}^2}{MSE}
\]

where \(P_{\text{max}}\) is the maximum intensity in the original image. An improved reconstruction quality usually results in higher PSNR. However, PSNR does not reflect well subjective human perception of image quality as it is mainly based on estimating absolute errors between individual pixels.

- **Structural similarity index (SSIM):** Accounts for image degradation as perceived changes in structural information [50]. It measures the structural similarity by comparing normalized local patterns of pixel intensities, which is similar to the human visual system’s abilities to extract information based on structure. The SSIM is defined as:

\[
SSIM = \frac{(2\mu_I \mu_{I'} + c_1)(2\sigma_{I'} + c_2)}{(\mu_I^2 + \mu_{I'}^2 + c_1)(\sigma_{I'}^2 + \sigma_{I'}^2 + c_2)}
\]

where \(\mu_I, \mu_{I'}, \sigma_{I'}^2\) and \(\sigma_{I'}^2\) are the average and variance of the intensities of the original 2D in-plane slice and the reconstructed slice respectively. \(\sigma_{II'}\) is the covariance of \(I\) and \(I'\), \(c_1\) and \(c_2\) are defined as \(c_1 = k_1L^2\) and \(c_2 = k_2L^2\) in order to balance the division with weak denominator, where \(L\) is the dynamic range of the intensities in image \(I\). Similar to [50], \(k_1\) and \(k_2\) are equal to 0.01 and 0.03 respectively.

Structural dissimilarity (DSSIM) heat maps are calculated in order to visualize the dissimilarities between original and

---

\(^2\)https://github.com/bkainz/fetalReconstruction
reconstructed images. DSSIM is calculated as a distance metric derived from SSIM:

\[
DSSIM = \frac{(1 - SSIM)}{2}
\]  

(13)

**B. Evaluation of local deformations**

Compared to SVR, PVR relaxes rigidity constraints and may therefore introduce deformation artifacts like shrinking of anatomical structures. Hence, in order to assess the geometric integrity of the 3D image reconstruction in terms of mm we conduct two experiments:

**Evaluation with anatomical landmarks:** We use a set of 3D anatomical landmarks that can be clearly defined in 3D reconstructions of real fetal MRI head scans to assess the accuracy of non-rigid deformations resulting from PVR compared to SVR. In particular, we have selected 10 clinically relevant, anatomical landmarks in the eyes and sub-cortical region of the fetal brain, which define stable anatomical structures of the corpus callosum and the lateral ventricles. A clinical expert has chosen a set of 11 fetal MRI scans with the best reconstruction quality of the fetal head using SVR. We have repeated the reconstruction of this set from the raw scan data using PVR with patches of size 32 and stride 16 pixels. Two different observers have annotated the selected anatomical landmarks independently across the 11 subjects. The accuracy is assessed by calculating the average 3D Euclidean distance errors between PVR and SVR (mean annotation points of the two observers), measured for each landmark for all 11 subjects.

**Evaluation of displacement:** In this experiment we seek to evaluate the quality of PVR in cases of non-rigid motion between the scanned images using an approximated displacement error measured on the simulated skewing motion experiment of the brain phantom. The reconstructed images are first rigidly and then non-rigidly (b-spline [51]) registered to the motion-free ground truth image. The transformations resulting from non-rigid registration are used to transform a set of points sampled uniformly on a 3D grid (5% of the image size) inside the brain area. To ensure a fair comparison, we exclude reconstructed images from high motion skewing angles (\(\omega > 1.5^\circ\)), which may affect registration to the reference image if the reconstructed image is blurred or distorted.

**V. Results**

**A. Global image quality**

**Reconstruction of adult brain MRI:** Experiments on adult brain MR data using the Brainweb database [48] includes introducing synthetic non-rigid motion artifacts as described in Sec. IV-A. Example slices of standard planes of original and corrupted data are depicted in Fig. 8. Comparative experimental results of SVR and PVR reconstruction methods are shown in Fig. 9 for PSNR, SSIM and CC. For all metrics, PVR shows an improved performance over SVR, particularly in the presence of deformations with higher skewing angles. Further, we observe that superpixel-based PVR achieves performance similar to PVR using arbitrary square patches, while requiring a lower number of input patches.

**Reconstruction of fetal organs:** Exemplary PVR and SVR reconstructions under motion introduced by kicking of the fetus are shown in Fig. 10. PVR reconstruction results show an improved visual appearance and less blurring in the region with severe motion artifacts (arrow). An example of a challenging clinical case with a kidney malformation in one of twin fetuses, is shown in Fig. 11. Our clinical partners confirmed that such complications are easier to examine and to quantify after PVR-based reconstruction.

Comparative experiments of PVR variants were carried out on 32 fetal MR scans at gestational ages of approximately 20 weeks with voxel size around 1.00x1.00x2.50 mm\(^3\) and 1.25x1.25x1.25 mm\(^3\), constituting challenging image corruption samples. The ssFSE T2-weighted sequences data have been acquired on a Philips Achieva 1.5T, the mother lying 20\(^\circ\) tilt on the left side to avoid pressure on the inferior vena cava and aligned to the main axes of the fetus. Table I (a) & (b) show numerical results of evaluating individual stacks before reconstruction (baseline), and the final reconstructed image from 3-6 stacks using square patches (\(a = 32\) and \(\omega = 16\)), superpixels (\(a = 16, \gamma = 60\%)\) and multi-scale superpixels (with initial \(a = 16, \gamma = 60\%\) and \(\times 1.5\) scaling factor for every new iteration). Statistical testing between baseline and PVR variants was carried out using paired T-Tests and differences between using fixed or multi-scale and between square patches or superpixels were assessed via Two-factor ANOVA with repeated measures. In Table I (a) & (b) the names of PVR variants are marked in bold if statistically significant differences have been found during analysis, i.e., FS and MS and/or Square Patches and Superpixel pairs are bold if the results between them differ significantly.

The evaluation of the reconstruction quality of a whole 3D image into a single-valued metric may not properly reflect the performance differences, as it is based on averaging values of all the pixels of all the input stacks. Furthermore, Table I indicates significant differences between variants of PVR but these differences have only minimal qualitative effect on reconstruction accuracy. Therefore, Fig. 12 evaluates the reconstruction quality of PVR additionally using dissimilarity heat maps based on the measured DSSIM (see Sec. IV-A). This approach allows further qualitative evaluation and allows for uncertainty visualization of PVR reconstructions.

**Fetal brain reconstruction at late gestation:** Examination of the brain is one of the primary uses for fetal MRI. While PVR’s primary aim is to reconstruct large fields of view and non-rigid motion, we also compare how PVR performs under the conditions under which SVR is usually applied (accurate delineation of a brain mask followed by rigid full slice to volume registration). This experiment evaluates if PVR can efficiently substitute SVR in pipelines where only the reconstruction of a single, masked rigidly moving organ is required, i.e., for compensation of pure bulk motion.

To compare PVR’s performance, we collected a second dataset comprising of ten mature fetuses (GA > 33 weeks) with little motion, expert brain segmentation, and validated state-of-the-art SVR reconstructions [25]. PVR is applied only in the delineated brain area and compared to clinical SVR reconstructions in terms of PSNR, SSIM, CC and compu-
Fig. 8. Strong synthetic non-rigid motion artifacts caused by skewing an adult brain phantom with an angle of ($\theta_{xyz}^x = \pm 1$). Rows: MRI in standard orientations: coronal, axial, and sagittal. Columns: original scan (1x1x1 mm$^3$), and sampled (1.25x1.25x2.5 mm$^3$) axially, coronally and sagittally, and SVR, PVR-square patches and PVR-superpixel reconstructed isotropic images (1.25x1.25x1.25 mm$^3$).

Fig. 9. Comparative reconstruction performance of SVR and PVR methods on synthetically corrupted Brainweb [48] data using rigid translational transformations (1st column), bulk transformations (2nd column) and non-rigid skewing deformations (3rd column). Left to right: PSNR, SSIM and CC.

Average (a) PSNR and (b) SSIM results ($N = 32$) for the input stack (baseline) and PVR variants with fixed (FS) and multi-scale (MS) variants of square patches and superpixels. All mean differences of PVR against baseline are statistically significant ($p < 0.05$). Names of all statistically significantly different PVR variants are stated in **bold.**

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>(a) PSNR</th>
<th>(b) SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>Brain</td>
<td>Placenta</td>
</tr>
<tr>
<td>FS</td>
<td>16.97 ± 0.77</td>
<td>19.90 ± 4.40</td>
</tr>
<tr>
<td>MS</td>
<td>27.03 ± 1.36</td>
<td>26.35 ± 1.54</td>
</tr>
</tbody>
</table>

| **Baseline** | Brain | Placenta | Uterus |
| FS | 0.51 ± 0.01 | 0.48 ± 0.04 | 0.48 ± 0.04 | 0.51 ± 0.01 | 0.50 ± 0.03 | 0.47 ± 0.04 |
Multi-scale reconstruction: We evaluate different scale parameters for the multi-scale superpixel reconstruction to further assess the influence of the number of scales on the reconstruction quality. Similar to the phantom brain experiment, we reconstruct a brain image from 3-stacks deformed with skewing angle $\theta_{xyz} = \pm 1^\circ$. We use the ground truth image for evaluating quantitatively each reconstructed image with five different scales as shown in Table II. Multi-scale (MS) reconstruction is configured to increase the scale of the patch by a factor of $\times 1.5$, four times starting from $a = 16$ to 40 (2-iteration per each scale). Fixed-scale (FS) reconstruction is configured to fix the size of the patch for every reconstruction iteration. Figure 13 and Table II show that MS-PVR achieves similar performance as FS-PVR, while larger scale patches result in a lower number of input patches and overhead pixels.

B. Local deformations

Landmarks: The results for each landmark are summarized in Table III. The inter-observer error is $0.92 \pm 0.41 \text{ mm}$. The average distance between all landmarks in PVR reconstructions and the mean landmark location in SVR reconstructions between the two observers is $1.13 \pm 0.87 \text{ mm}$. In order to provide an indication of the scale of rigid motion being addressed in the fetal brain landmark images, the estimated displacement distance of all 3D landmarks during imaging is measured by calculating the distance between each annotated landmark and its inverse-transformed point from the correspondent slice of an input stack, assuming rigid slice motion for SVR. The maximum of the calculated displacements equals to $79.85 \text{ mm}$. The total average equals to $13.58 \pm 15.23 \text{ mm}$ for all landmarks, see Table III. The $p$-values from a paired T-test show that the difference between the error introduced between observers and PVR can be considered to be not statistically significant ($p = 0.0166$). Overall, this evaluation demonstrates that PVR does not introduce notable structural distortion compared to rigid SVR.

Displacement: The mean and maximum displacement of 9 reconstructed images are calculated for SVR and PVR with different parameters. The results in Table IV shows that PVR with patch size of 64 achieves the smallest mean error, while this error increases with the decrease of the patch size.

C. Performance Analysis

We further evaluate the computational performance of each PVR variant. Measuring the overall runtime is not meaningful because this would be highly machine specific and would include data transfer overhead and non optimized functions. The
TABLE III

<table>
<thead>
<tr>
<th>Landmarks</th>
<th>Slice-to-volume motion [mm]</th>
<th>Inter-observer error [mm]</th>
<th>PVR-SVR distance [mm]</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left center of the eye</td>
<td>15.10 ± 15.00</td>
<td>0.95 ± 0.29</td>
<td>1.00 ± 0.43</td>
<td>0.7983</td>
</tr>
<tr>
<td>Right center of the eye</td>
<td>11.87 ± 11.01</td>
<td>0.84 ± 0.40</td>
<td>0.96 ± 0.41</td>
<td>0.6381</td>
</tr>
<tr>
<td>Genus of corpus callosum outer aspect</td>
<td>13.86 ± 15.16</td>
<td>0.69 ± 0.36</td>
<td>0.81 ± 0.24</td>
<td>0.7670</td>
</tr>
<tr>
<td>Spectrum of corpus callosum outer aspect</td>
<td>13.02 ± 15.06</td>
<td>1.05 ± 0.56</td>
<td>1.11 ± 0.75</td>
<td>0.6091</td>
</tr>
<tr>
<td>Left Frontal horn of lateral ventricle</td>
<td>14.93 ± 16.83</td>
<td>1.05 ± 0.40</td>
<td>1.35 ± 1.00</td>
<td>0.2986</td>
</tr>
<tr>
<td>Right Frontal horn of lateral ventricle</td>
<td>17.04 ± 18.61</td>
<td>0.99 ± 0.22</td>
<td>1.25 ± 0.89</td>
<td>0.5111</td>
</tr>
<tr>
<td>Left occipital horn of lateral ventricle</td>
<td>11.40 ± 14.50</td>
<td>1.04 ± 0.34</td>
<td>0.98 ± 0.48</td>
<td>0.7483</td>
</tr>
<tr>
<td>Right occipital horn of lateral ventricle</td>
<td>11.41 ± 13.20</td>
<td>1.05 ± 0.38</td>
<td>1.05 ± 0.36</td>
<td>0.9906</td>
</tr>
<tr>
<td>Left temporal horn of lateral ventricle</td>
<td>12.89 ± 13.64</td>
<td>0.60 ± 0.36</td>
<td>1.28 ± 1.13</td>
<td>0.6278</td>
</tr>
<tr>
<td>Right temporal horn of lateral ventricle</td>
<td>11.85 ± 12.76</td>
<td>0.80 ± 0.46</td>
<td>1.47 ± 1.33</td>
<td>0.1396</td>
</tr>
<tr>
<td>All landmarks</td>
<td>13.58 ± 15.23</td>
<td>0.92 ± 0.41</td>
<td>1.13 ± 0.87</td>
<td>0.0565</td>
</tr>
</tbody>
</table>

TABLE IV

<table>
<thead>
<tr>
<th>Displacement</th>
<th>SVR</th>
<th>PVR-Superpixel 16</th>
<th>PVR-Patch 64</th>
<th>PVR-Patch 32</th>
<th>PVR-Patch 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.85 ± 1.55</td>
<td>1.70 ± 1.10</td>
<td>1.58 ± 1.03</td>
<td>1.67 ± 1.07</td>
<td>2.40 ± 1.71</td>
</tr>
<tr>
<td>Max</td>
<td>4.91 ± 4.26</td>
<td>4.86 ± 3.54</td>
<td>4.87 ± 3.92</td>
<td>5.02 ± 3.92</td>
<td>5.72 ± 4.09</td>
</tr>
</tbody>
</table>

VI. DISCUSSION

PVR surpasses reconstruction performance of the state-of-the-art SVR method for cases with considerable non-rigid deformations and inconsistent bulk motion. We have evaluated different variants of PVR using fixed-size and multi-size square patches and fixed-size and multi-size superpixels. ANOVA analysis has shown significant differences between these approaches for different areas of the uterus. However, evaluation of motion compensation methods is difficult especially due to the lack of ground truth in fetal MRI. Mapping the reconstruction quality of a whole 3D volume into a single-valued metric may not properly reflect qualitative differences, if based on averaging all measured values of all input stacks.
Therefore, we have additionally performed extensive qualitative analyses and present examples and evaluation based on structural dissimilarity (DSSIM) heat maps.

While an SVR reconstruction is constrained by using whole orthogonal slices assuming rigidity, PVR may introduce geometric deformations due to the relaxation of the rigidity assumption to smaller patches instead of whole slices. For this reason, we evaluated local deformations by selecting anatomical landmarks from the fetal head and compare PVR with SVR, which shows that the difference is not statistically significant for individual landmarks. We also employ dense, non-rigid registration to further validate the geometric accuracy of the reconstruction. Reconstructed images from SVR and PVR with different variants are non-rigidly aligned to a motion-free brain phantom. The mean and maximum displacements of the deformations are then calculated to measure the accuracy. Both experiments evaluate local deformations, although the latter may include errors introduced by the imperfect non-rigid registration process. Maximum displacement errors are mainly observed in feature-poor regions like the amniotic fluid, where automatic non-rigid registration does not perform well. Despite showing strong evidence that PVR works well for the whole uterus, it should be noted that our experiments to assess the reconstruction quality on brain anatomy do not necessarily extend to the achieved reconstruction accuracy of other areas (e.g., moving limbs). Focused validation of individual regions will be required for other, domain specific applications.

In addition to reconstruction and motion correction of the whole uterus, we have demonstrated that the proposed method works for multiple births cases with multiple fetal instances present in the volumetric image. These cases are more likely to have complications and to undergo MRI during pregnancy but would require extensive manual effort to be successfully reconstructed with state-of-the-art methods.

Although our method is able to reconstruct the whole uterus automatically, small parts like limbs that move rapidly between the acquisition of individual slices are more difficult to recover. This is especially problematic for very young fetuses that have more space to move inside the womb. In cases of rapid limb movements (>2 cm between individual slices) PVR is not able to find structural consensus between overlapping patches and blurry image regions will be reconstructed. This is a general problem of automatic intensity-based optimization methods and methods that are able to understand the semantic content of each patch will likely be required for future improvements.

PVR introduces a considerable computational overhead to the reconstruction stage of fetal MR image processing pipelines. We have evaluated the amount of necessary additional redundant information to give a general idea about the expected runtime of different PVR variants. Patches based on multi-scale superpixels are significantly more efficient than a naïve implementation of overlapping square patches, while maintaining a similar reconstruction accuracy. Quantitatively, square patches perform slightly better for the brain, which is most likely due to the rigid nature of the enclosing skull. Superpixel-based patches achieve better results for regions that are likely affected by non-rigid movements like the placenta and the whole uterus.

**VII. Conclusion**

We have introduced and evaluated the concept of patch-to-volume reconstruction (PVR) in order to compensate non-rigid motion artifacts from fetal MRI scans without requiring a defined region of interest. PVR splits the 3D input image into overlapping square patches and superpixels and employs automatic EM-based outlier rejection to find consistent data. Our method is able to automatically reconstruct whole collections of motion corrupted stacks without the need for image segmentation and manual identification of rigid regions. We have shown that PVR can reconstruct the whole uterus, selected fetal organs, and secondary, non-rigidly moving pregnancy structures such as the placenta.

PVR’s reconstruction quality and computational performance has been evaluated quantitatively and qualitatively on an adult phantom T2-weighted brain with synthetic non-rigid motion artifacts, as well as on the whole uterus from motion corrupted fetal MRI data including fetal brain, placenta and cases with multiple births.

**Acknowledgments:** We thank the volunteers, the radiographers J. Allsop and M. Fox, MITK (CC public license from IXICO Ltd.), the NIHR Biomedical Research Centre, Welcome Trust/EPSRC IEH award 102431, ERC 319456, NVIDIA (GPU donations), and EPSRC EP/N024494/1. Data access only in line with the informed consent of the participants, subject to approval by the project ethics board and under a formal Data Sharing Agreement. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. A. Alansy is supported by an Imperial College President’s PhD Scholarship.

**References**


