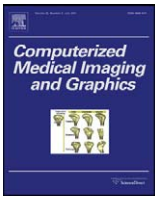
Computerized Medical Imaging and Graphics 35 (2011) 373–382

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag



Fusion of multi-planar images for improved three-dimensional object reconstruction

Xiaohui Yuan[a,*], Xiaojing Yuan[b]

[a] Department of Computer Science and Engineering, University of North Texas, P.O. Box 311366, Denton, TX 76210, USA
[b] Department of Engineering Technology, University of Houston, Houston, TX 77004, USA

ARTICLE INFO

Article history:
Received 18 April 2010
Received in revised form
16 November 2010
Accepted 24 November 2010

Keywords:
3D reconstruction
Magnetic Resonance Imaging (MRI)
Image fusion
Wavelet

ABSTRACT

Due to the scan time limitation, our MRI studies of the human tongue can acquire only a limited number of contiguous two-dimensional (2D) slices to form a volumetric data set in a given series. An interpolated three-dimensional (3D) reconstruction using images acquired in a single plane presents artifacts. To address this issue, we developed a wavelet-based bidirectional linear fusion method that uses slices acquired from sagittal and coronal planes to estimate the unknown values of the inter-slice voxels. We use an interpolation method to estimate the voxel value based on neighboring fiducial voxels in the bounding slices. This interpolation is followed by a wavelet fusion to recover image details by integrating prominent coefficients from the interpolated images. Our method was evaluated using 2D MR images and 3D phantoms. Experiments demonstrated that our method reduces interpolation artifacts and greatly improves the 3D reconstruction accuracy. The advantage of our method casts new light on MR imaging and image processing and permits us to achieve high resolution and short acquisition time simultaneously.

1. Introduction

To test the hypothesis that the *in vivo* human tongue is a muscular hydrostat with a constant volume regardless of task demands, we have embarked on a comprehensive study whereby serial 2D MR slices are acquired for the whole tongue in static, task-induced postures from the sagittal and the coronal planes. A precise image-based tongue volume measurement is, therefore, crucial to this project. It is our contention that a volume estimation based on 2D segmentation is inherently inaccurate because of the incomplete three-dimensional information. Several past volumetric studies [1–4] have demonstrated that 3D segmentation methods produce more accurate volume estimation than conventional 2D methods. The accuracy of 3D-based methods, however, depends on the reconstructed volumetric images.

In our study, the requirement of breath-holding with maximum effort tongue press limits our MR imaging time to roughly 20 s in a 3 Tesla (3 T) scanner. Because of the stringent timing, a three-dimensional (3D) MR scan is infeasible. Within this time window, a maximum of 19 contiguous slices of the tongue can be acquired in the coronal plane at a slice thickness of 4 mm. The small number of slices causes the reconstruction to have a staircase artifact using images from any single imaging plane. With rigorous experimental control of head position and movement during scanning, we ensure that the two sets of slices acquired separately are aligned in the 3D space. Fig. 1(b) and (c) illustrates a mid-coronal and a mid-sagittal MR images of the tongue, respectively. By assembling all the sagittal and coronal images in a 3D space (a sketch is illustrated in Fig. 1(a)), our goal is to improve the 3D reconstruction by fusing inter-weaved, multi-planar slices, as depicted in Fig. 1(d).

The value of a voxel can be estimated by a weighted sum of the neighboring voxels. A number of interpolation methods have been reported ([5]; see [6] for a review). In linear interpolation, for example, the value of a voxel with no value is approximated by the adjacent voxels and their distances to the target. Despite its insufficiency in high-order continuity, this method has been widely used because of its simplicity and efficiency. Cubic interpolation [7] and B-spline interpolation have been developed [6] to address high-order discontinuity. However, when applied to our intended 3D reconstruction from two sets of MR slices in perpendicular planes, these methods do not account for multi-planar image integration.

Methods to combine multi-planar slices for a more accurate 3D reconstruction have also been investigated. Kuwahara [8] described a contour-based reconstruction method for modeling a 3D heart. By tracing the heart boundary in six images of three acquisition planes (two images per plane), contours were obtained. A spline was used to connect the intersection points. The contours are then merged to form a denser 3D mesh. The need of manual tracing limits its application, and the accuracy of the 3D model is bounded by hand-tracing consistency and precision.

---

* Corresponding author. Tel.: +1 940 565 4256; fax: +1 940 565 2799.
E-mail address: xiaohui.yuan@unt.edu (X. Yuan).

0895-6111/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2010.11.013

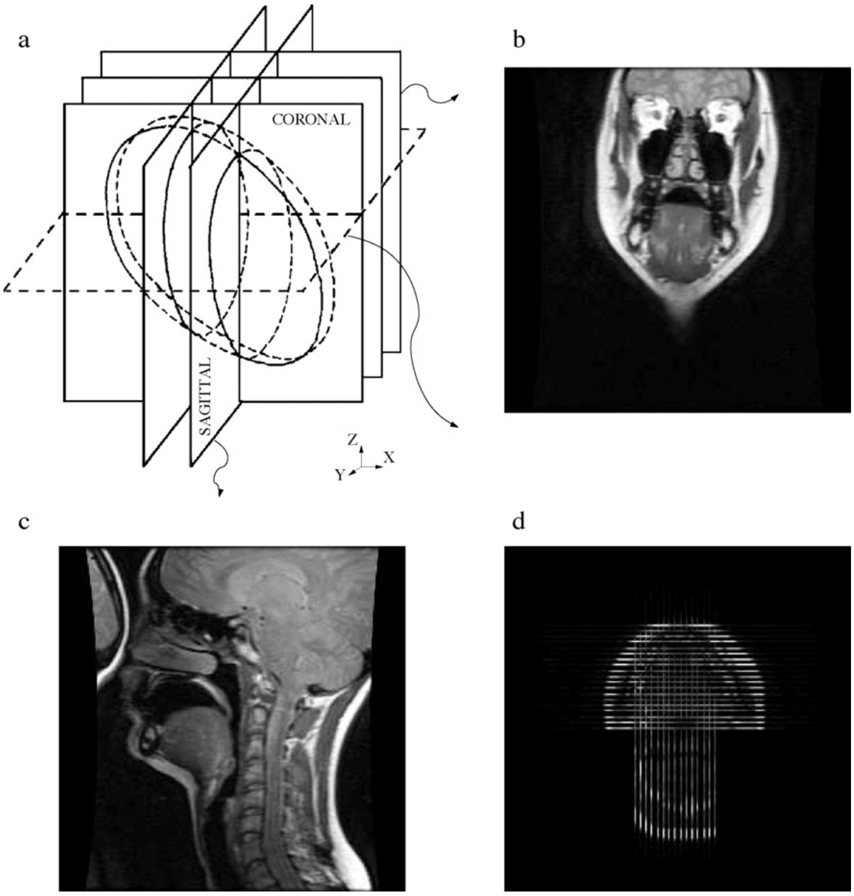


Fig. 1. The slice view shows the images in both sagittal and coronal planes and a cross-cut in the transverse plane. (a) The coronal and sagittal planes. The transverse plane is illustrated with a dashed line. (b) An example of a coronal scan. (c) An example of a sagittal scan. (d) A cross-section in the transverse plane that shows inter-waved sagittal and coronal scans.

Goshtasby and Turner [9] proposed a distance-weighted fusion method to reconstruct a 3D model using short-axis and long-axis cardiac images. The voxel value is approximated with a weighted sum of voxels within a neighborhood, and the weights are proportional to the inverse distance to the target:

$$I(x, y, z) = \frac{\sum_{j=1}^{N} I_j \left(r^2 + \sum_{p \in \{x,y,z\}} (p - p_j)^2\right)^{-1}}{\sum_{j=1}^{N} \left(r^2 + \sum_{p \in \{x,y,z\}} (p - p_j)^2\right)^{-1}},$$

where $I(x, y, z)$ is the value for the target voxel and I_j represents the known voxel values in a given window. A smooth factor r is used to gain better visual results. For an effective interpolation, the window size w needs to satisfy $w = 2dr^2 - 1$, where d is the number of new voxels between two image slices.

To achieve a more accurate reconstruction, Werahera et al. [10] employed the shape of rotund objects. Assuming a hemispherical shape for the prostate glands, their algorithm assigns unknown voxel values by linear interpolation and extrapolation. Similar methods were investigated in [11] and later in [12]. Based on its fine-tuned shape model, this method can achieve high accuracy in 3D reconstruction. However, underestimation is likely in rebuilding 3D models of small objects. Improvement was achieved by employing object deformation information derived from slice registration [13,14]. Penney et al. applied non-rigid registration to derive pixel association between a pair of sample images and used the associated pixels to estimate the missing values following linear interpolation [13]. Frakes et al. developed similar idea but used a hybrid motion estimation technique, i.e., control grid interpolation, to get the displacement field. Cubic interpolation is then applied [14].

In this paper, we describe a novel wavelet-based fusion method for 3D reconstruction and will address two important issues:

1. how to estimate unknown voxel values by combining complementary information from sparse, multi-planar MR images, and
2. how to improve reconstruction accuracy and preserve crucial image features (e.g., object boundaries) in a volume composite.

The ensuing sections of this paper are organized as follows: Section 2 presents our fusion-based 3D reconstruction algorithm. It begins with a description of our bidirectional linear interpolation method that estimates unknown values using information from both sagittal and coronal MR images. Wavelets-based fusion is then used to integrate features from interpolation results for reconstruction. Section 3 demonstrates our method using 2D tongue MR image, synthetic 3D phantom, and multi-planar, bidirectional human tongue MR scans. Our method was compared with five other interpolation methods. Section 4 concludes this paper with a summary and a brief discussion of our future work.

2. Methodology

2.1. 3D reconstruction via fusion

In MR imaging, slices are typically equally spaced and parallel to each other. An illustration of the imaging planes and their correspondence in the Cartesian coordinates is shown in Fig. 1(a). When the slices of a sagittal image set are mapped into a 3D space, the resolution in the Y–Z plane is the image resolution (or pixel size), whereas the resolution in the X direction is determined by the number of slices acquired (or slice thickness). Likewise, coronal images have a high resolution in X–Z plane, but low resolution in the Y direction.

Let $I_j(X_j, y, z)$ and $I_k(x, Y_k, z)$ denote sagittal and coronal slices, respectively. The capital letters indicate the dimensions that are constant to the image plane. By assembling I_j and I_k in a 3D space, we obtain high resolution in the X, Y, and Z directions in the sampled slices. These images are hereafter referred to as the *fiducial slices*, and the pixels/voxels in the fiducial slices are *fiducial pixels/voxels*.

Let the slice space be λ, $\lambda = |x_{j+1} - x_j|$ or $\lambda = |y_{k+1} - y_k|$. The reassembled 3D volume based on λ is re-sliced in the transverse plane, as illustrated in Fig. 1(a) with the dashed lines. Each transverse slice cuts through all the sample slices and is denoted with $I_l(x, y, Z_l)$, which I_l is a 2D grid-like image and contains fiducial pixels from the intersecting image slices but the rest pixels are voids as shown in Fig. 1(d). We assume that every voxel value is a function of its neighboring fiducial pixels in the transverse image:

$$I(x, y, z) = f(I_j(X_j, y, Z_l), I_k(x, Y_k, Z_l)), \tag{1}$$

where $x \notin x_j$ and $y \notin y_k$. Our task is, then, to estimate the 2D transverse images $I_l(x, y, Z_l)$ using the fiducial pixels in $I_j(X_j, y, Z_l)$ and $I_k(x, Y_k, Z_l)$. The 3D volume is hence created by assembling all the reconstructed transverse images.

Our method consists of two steps: (1) construction of transverse images via interpolation; and (2) a Wavelet-based Bidirectional Linear Fusion (WBLF) to integrate complementary information for an improved 3D reconstruction. In the first step, we use our bidirectional linear interpolation algorithm to estimate the values of the unknown voxels based on fiducial pixels in the sagittal and the coronal slices and construct a transverse image. Using the same sagittal and coronal slices separately, we then generated two transverse images with linear interpolation. In the second step, a wavelet transform is used to decompose three transverse images into subbands. A fused coefficient matrix is constructed by taking the most significant coefficients according to their strength in directional edges. The final reconstruction is achieved with an inverse wavelet transform.

2.2. Bidirectional linear interpolation

We modified the conventional bilinear interpolation to combine information from more than one source. As shown in Fig. 2, a transversal image consists of fiducial pixels (in gray) and voids (in white). To determine the value of I_0 (i.e., target pixel), four pixels are identified in the bounding fiducial slice that have the minimum distance to the target pixel, i.e.,

$$I_f = \{I_j | \min(\mathcal{D}(I_0, I_j), \forall I_j \in \{I^s, I^c\})\}, \tag{2}$$

where $\mathcal{D}(\cdot)$ denotes the Euclidean distance.

The four closest pixels are referred to as the fiducial neighbors of I_0 and are denoted by $I_n = \{I_1, I_2, I_3, I_4\}$. Two pairs are formed from the four neighbors, each of which consists of one pixel from each bounding slice. Let $\mathcal{D}_a$, $\mathcal{D}_a = \mathcal{D}(I_1, I_2) + \mathcal{D}(I_3, I_4)$ and $\mathcal{D}_b$, $\mathcal{D}_b = \mathcal{D}(I_1, I_4) + \mathcal{D}(I_2, I_3)$ denote two ways of pairing. Note that the farther the two fiducial pixels are apart, the more uncertain the interpolated value. Hence, our chosen pairing shall minimize the sum of the distance between every two paired pixels:

$$\mathcal{D}^* = \min(\mathcal{D}_a, \mathcal{D}_b), \tag{3}$$

where $\mathcal{D}^*$ is the total distance of the chosen pairing.

The value for I_0 is computed as follows:

$$I_0 = I_{12} + \mathcal{D}_{0\_12} \frac{I_{34} - I_{12}}{\mathcal{D}_{0\_12} + \mathcal{D}_{0\_34}}, \tag{4}$$

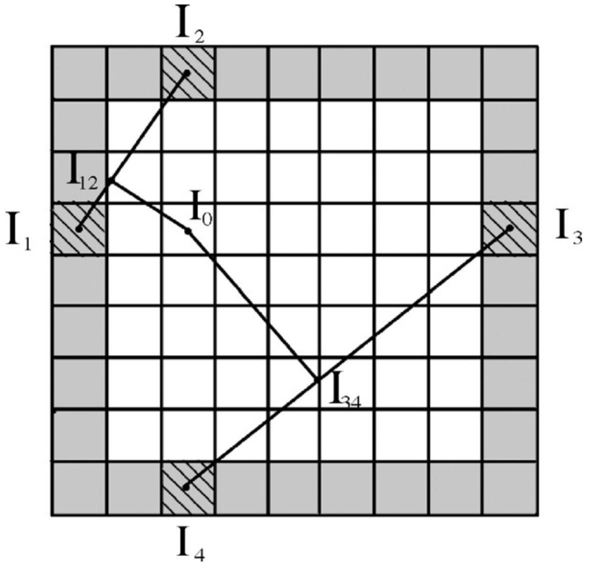


Fig. 2. A 2D scenario of the bidirectional linear interpolation. Fiducial pixels are shaded in gray and the voids are in white. Given the target pixel, I_0, four fiducial pixels are identified as I_1, I_2, I_3, and I_4.

where $\mathcal{D}_{0\_12}$ and $\mathcal{D}_{0\_34}$ are the distances of I_0 to the line segments of $\overline{I_1 I_2}$ and $\overline{I_3 I_4}$, respectively, and are proportional to the products of the X and Y displacements:

$$\mathcal{D}_{0\_12} = \mathcal{D}_{02} \frac{\mathcal{D}_{10}}{\mathcal{D}_{12}}, \tag{5}$$

$$\mathcal{D}_{0\_34} = \mathcal{D}_{04} \frac{\mathcal{D}_{30}}{\mathcal{D}_{34}}, \tag{6}$$

and

$$I_{12} = I_2 + \frac{\mathcal{D}_{02}^2}{\mathcal{D}_{12}^2}(I_1 - I_2), \tag{7}$$

$$I_{34} = I_4 + \frac{\mathcal{D}_{04}^2}{\mathcal{D}_{34}^2}(I_3 - I_4). \tag{8}$$

An advantage of bidirectional interpolation is that the diagonal features are mostly recovered. However, this method also suffers from sparse fiducial slices. An example is illustrated in Fig. 3. By removing evenly spaced 224 rows and 224 columns from a 256×256 image, we created a synthetic bidirectional sample, as illustrated in Fig. 3(a). The zoom-in view in Fig. 3(c) reveals the interpolation-induced discontinuity. The region enclosed by box 1 depicts continuous reconstruction with a clear boundary, whereas the interpolated region in box 2 exhibits artificial intensity variations. This difference is the consequence of having a sparse sampling space. The large difference makes it difficult to recover the pixel intensity.

2.3. Wavelet-based fusion

The values of two adjacent pixels in an image are very similar unless an edge exists. It is problematic, however, to recover missing values when the sample is sparse, i.e., the slice thickness is great. Fig. 4 shows a synthetic 2D image sampled in the X (horizontal) direction. The ground truth, illustrated in Fig. 4(a), has an 8-bit gray scale. The sampled fiducial pixels are shown in Fig. 4(b) as shaded squares, and the rest are voids, i.e., values that need to be estimated. Using linear interpolation, we construct an image using fiducial pixels, as shown in Fig. 4(c). It is clear that the vertical edges are successfully recovered. Similarly, if we sample in the Y (vertical)

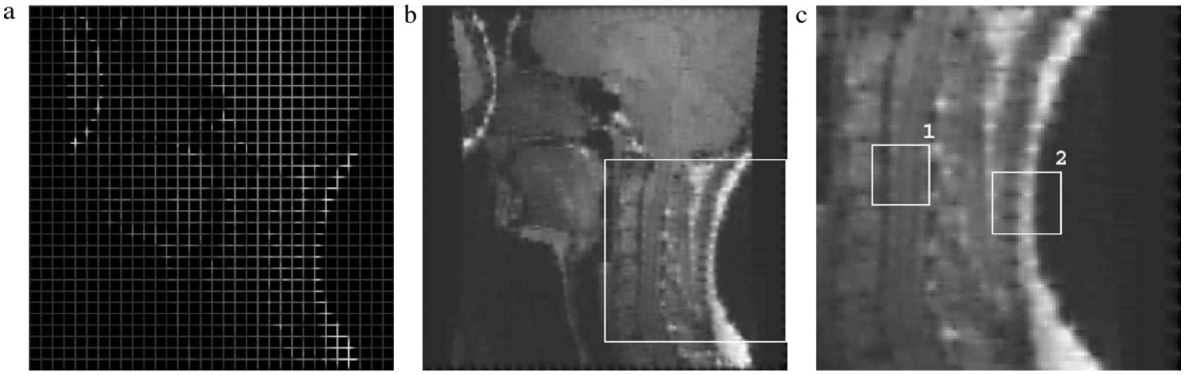


Fig. 3. Reconstruction using bidirectional linear interpolation. (a) The sample image. (b) The reconstructed image. (c) The zoom-in view of the boxed region in (b).

direction, continuous horizontal edges can also be recovered using linear interpolation.

The image features recovered with linear interpolation are complementary to those from bidirectional interpolation. To aggregate image features, we developed a method to fuse the interpolation results by combining the prominent wavelet coefficients for improved 3D reconstruction.

The wavelet transformation of a function *f*(*x*) at scale 2^j is a convolution of a filter function ψ to signal f and is computed as follows:

$$W_{2^j}f(x) \triangleq w_j(x) = f * \psi_{2^j}(x), \tag{9}$$

where $w_j(x)$ denotes the wavelet coefficients at index x. A wavelet function $\psi(x)$ is expended by a dilation parameter 2^j,

$$\psi_{2^j}(x) = \frac{1}{2^j}\psi\left(\frac{x}{2^j}\right).$$

By choosing $\phi(x, y)$ to be a 2D differentiable function to satisfy $\int_{x,y}\phi(x, y) = 1$ and $\phi(x, y)|_{x,y=\infty} = 0$, we have wavelets that consist of two components in the horizontal and vertical directions as follows:

$$\psi^x(x, y) = \frac{\partial\phi(x, y)}{\partial x}, \tag{10}$$

$$\psi^y(x, y) = \frac{\partial\phi(x, y)}{\partial y}. \tag{11}$$

Extending Eq. (9) to 2D and replacing the wavelets with Eqs. (10) and (11), the dyadic wavelet transformation of a function *f*(*x*, *y*) ∈ $L^2(R^2)$ can be written as follows:

$$W^x_{2^j}f(x, y) = f * \left(2^j\frac{\partial\phi_{2^j}}{\partial x}\right)(x, y), \tag{12}$$

$$W^x_{2^j}f(x, y) = 2^j\frac{\partial}{\partial x}(f * \phi_{2^j})(x, y), \tag{13}$$

$$W^y_{2^j}f(x, y) = f * \left(2^j\frac{\partial\phi_{2^j}}{\partial y}\right)(x, y), \tag{14}$$

$$W^y_{2^j}f(x, y) = 2^j\frac{\partial}{\partial y}(f * \phi_{2^j})(x, y). \tag{15}$$

The subband coefficients are partial derivatives of the convolution of image f with filter ϕ. That is, each subband is equivalent to an object edge map at various decomposition scales [15,16]. This transformation, therefore, resembles the Canny edge detector in that each subband consists of gradient values at a given scale [15,17].

As shown in Section 2.2, interpolation using information from different direction recovers complimentary image features. In a 2D scenario, interpolation from rows recovers most vertical features; interpolation from columns recovers most horizontal features; and interpolation with our bidirectional linear method preserves diagonal features. This property matches that of wavelet subbands and hence motivates our development of a fusion method. As shown in Fig. 5, three interpolated images are generated using bidirectional linear method (I_J), interpolation with row fiducial pixels (I_R), and interpolation with column fiducial pixels (I_C). By applying wavelet decomposition, each interpolated image generates 4 subbands at a given scale j, denoted with $W^{xx}_{2^j}$, $W^{yy}_{2^j}$, $W^{xy}_{2^j}$, and ϕ_{2^j}, where $W^{xx}_{2^j}$ consists of vertical derivatives, $W^{yy}_{2^j}$ of horizontal derivatives, and $W^{xy}_{2^j}$ of diagonal derivatives. Since I_R recovers mostly the vertical image features and I_C contains mostly the horizontal features, we construct the fused subbands $\tilde{W}^{xx}_{2^j}$ by selecting the coefficient with the greatest magnitude, i.e.,

$$\tilde{W}^{xx}_{2^j}(x, y) = W^{xx}_{2^j,Q}(x, y),$$

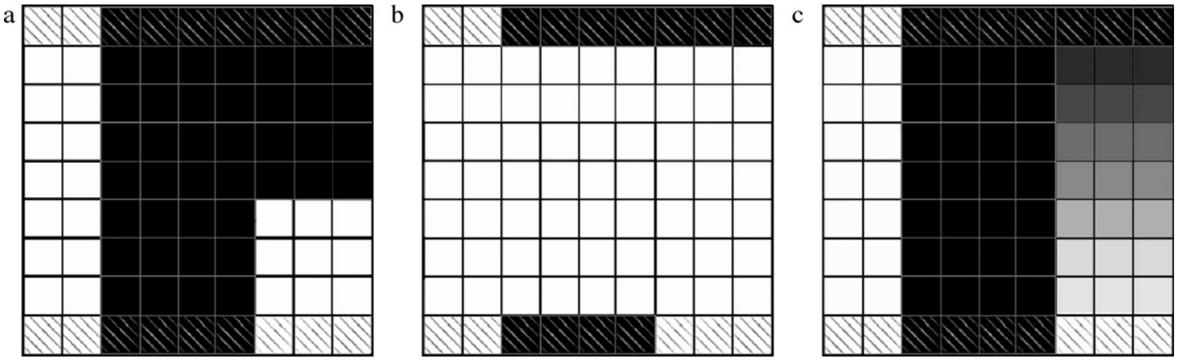


Fig. 4. (a) Original image. The shaded squares are the sampled locations. (b) Sampled image. (c) Linear interpolation result.

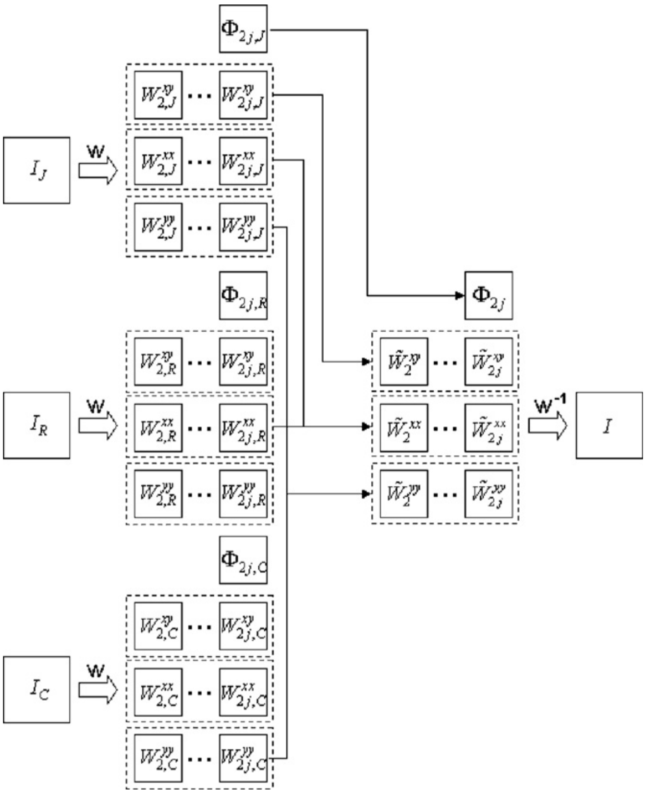


Fig. 5. Wavelet-based fusion. The block arrows represent wavelet transformations, and the dashed lines represent the subband selection and merging process. Three interpolation results are decomposed into subbands by using dyadic wavelet transformation. Based on our fusion rules, subbands are selected and a merged wavelets coefficient composite is obtained. The fusion result is synthesized with the inverse wavelet transform.

where $|W^{xx}_{2^j,Q}| = \max(|W^{xx}_{2^j,Q}|)$ and $Q \in \{J, R\}$.

Similarly, for the fused subbands $\tilde{W}^{yy}_{2^j}$, we select coefficients from the decomposition of I_C and I_J following the greatest magnitude rule:

$$\tilde{W}^{yy}_{2^j}(x, y) = W^{yy}_{2^j,Q}(x, y),$$

where $|W^{yy}_{2^j,Q}| = \max(|W^{yy}_{2^j,Q}|)$ and $Q \in \{J, C\}$.

However, since only subbands from I_J preserve diagonal image features, $\tilde{W}^{xy}_{2^j}$ carries the coefficients from the $W^{xy}_{2^j}$ of I_J, i.e.,

$$\tilde{W}^{xy}_{2^j}(x, y) = W^{xy}_{2^j,J}(x, y).$$

The fusion of the three interpolations is achieved by aggregating coefficients according to their maximum image feature. The fusion rule is based on coefficient magnitude, because a greater coefficient represents a stronger edge as follows:

$$\begin{aligned}
&\tilde{\phi}_{2^j}(x, y) = \phi_{2^j,J}(x, y),\\
&\tilde{W}^{xx}_{2^j}(x, y) = W^{xx}_{2^j,Q}(x, y), \quad |W^{xx}_{2^j,Q}| = \max(|W^{xx}_{2^j,Q}|) \quad \text{and} \quad Q = \{J, R\},\\
&\tilde{W}^{yy}_{2^j}(x, y) = W^{yy}_{2^j,Q}(x, y), \quad |W^{yy}_{2^j,Q}| = \max(|W^{yy}_{2^j,Q}|) \quad \text{and} \quad Q = \{J, C\},\\
&\tilde{W}^{xy}_{2^j}(x, y) = W^{xy}_{2^j,J}(x, y),
\end{aligned} \tag{16}$$

where $\tilde{W}$ and $\tilde{\phi}$ denote a wavelet subband of the fused composite. An inverse wavelet transform reconstructs the image from fused wavelet subbands.

In the 3D scenario, we assemble coronal and sagittal slices to form a 3D volumetric object with voids to be estimated. In a transversal plane that cuts across the 3D model, coronal slices provide the horizontal fiducial pixels and sagittal slices provides the vertical fiducial pixels. Hence, the reconstruction in 3D is converted into reconstruction of transversal slices in 2D. Reassembling the interpolated transversal slices gives the final 3D model.

Table 1
The RMSE of the 2D image reconstruction. S: sample spacing and MD: methods.

S	MD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	WBLF	3.96	3.65	3.56	3.58	3.73	3.70	3.87	3.89	3.88	3.97	3.86	4.01	3.76	3.62	3.96
	Gosh	4.44	4.26	4.18	4.24	4.44	4.36	4.61	4.62	4.71	4.81	4.69	4.89	4.56	4.39	4.72
	RI	17.1	18.6	16.9	18.8	17.2	18.2	17.1	17.4	17.5	19.0	20.9	20.8	16.2	20.4	18.1
	LI	28.4	26.9	25.9	25.9	27.1	25.9	27.4	27.9	28.5	28.6	28.1	29.0	28.2	28.2	30.5
	CI	28.6	27.0	26.1	26.1	27.3	26.1	27.6	28.1	28.6	28.8	28.3	29.3	28.5	28.4	30.7
	SI	28.7	27.1	26.2	26.2	27.4	26.2	27.7	28.2	28.8	29.0	28.5	29.4	28.6	28.6	30.9
5	WBLF	5.17	4.87	4.85	4.77	5.06	4.91	5.28	5.25	5.40	5.47	5.32	5.50	5.14	5.00	5.34
	Gosh	6.01	5.71	5.62	5.57	5.94	5.76	6.08	6.11	6.37	6.42	6.25	6.52	6.10	6.04	6.42
	RI	20.0	16.3	20.9	18.9	21.4	15.8	22.3	18.2	18.2	18.1	22.6	21.4	21.4	18.2	22.3
	LI	30.5	28.8	28.0	27.9	29.1	27.9	29.4	30.0	30.6	30.8	30.3	31.1	30.3	30.5	33.0
	CI	30.8	29.1	28.3	28.3	29.4	28.2	29.8	30.3	31.0	31.2	30.6	31.6	30.7	31.0	33.4
	SI	31.0	29.2	28.4	28.4	29.6	28.3	29.9	30.5	31.1	31.4	30.8	31.7	30.8	31.1	33.6
6	WBLF	6.12	5.75	5.73	5.77	6.10	5.94	6.31	6.53	6.67	6.69	6.42	6.64	6.08	5.90	6.37
	Gosh	7.14	6.71	6.65	6.70	7.12	6.95	7.36	7.60	7.81	7.78	7.44	7.83	7.23	7.06	7.48
	RI	17.4	20.6	19.4	20.3	21.2	16.1	18.7	19.4	22.3	17.8	18.9	21.1	19.5	20.2	24.0
	LI	30.1	28.4	27.5	27.6	28.8	27.5	29.0	29.6	30.3	30.4	29.8	30.7	29.8	29.9	32.2
	CI	30.4	28.7	27.7	27.8	29.1	27.8	29.3	29.9	30.6	30.7	30.1	31.1	30.1	30.2	32.5
	SI	31.0	28.8	27.9	28.0	29.2	27.9	29.5	30.1	30.8	30.9	30.3	31.3	30.3	30.4	32.7
7	WBLF	7.37	6.82	6.72	6.63	6.86	6.72	7.19	7.48	7.66	7.81	7.63	7.69	7.36	7.27	7.48
	Gosh	8.49	7.99	7.82	7.78	8.10	7.86	8.46	8.78	8.93	9.00	8.71	8.94	8.75	8.58	8.90
	RI	22.0	21.4	20.6	20.2	19.9	21.4	17.2	17.5	20.7	20.6	21.1	17.9	21.0	22.4	20.1
	LI	30.7	29.1	28.3	28.2	29.4	28.2	29.8	30.5	31.0	30.9	30.3	31.1	30.5	30.3	33.1
	CI	31.1	29.4	28.7	28.5	29.8	28.6	30.2	30.8	31.4	31.3	30.7	31.5	30.9	30.7	33.5
	SI	31.3	29.6	28.8	28.7	30.0	28.8	30.4	31.0	31.6	31.5	30.8	31.7	31.1	30.9	33.7
8	WBLF	8.23	7.66	7.52	7.47	7.81	7.68	8.21	8.37	8.63	8.59	8.34	8.59	8.16	7.84	8.35
	Gosh	9.33	8.76	8.57	8.51	8.97	8.77	9.40	9.50	9.75	9.82	9.49	9.83	9.40	9.20	9.74
	RI	21.1	20.6	18.9	21.0	21.6	16.2	16.8	19.6	23.3	22.2	20.9	18.5	22.8	22.9	23.7
	LI	30.7	28.8	28.0	28.1	29.3	28.1	29.7	30.3	31.1	31.0	30.4	31.1	30.1	30.2	32.6
	CI	31.0	29.1	28.3	28.4	29.6	28.4	30.0	30.6	31.5	31.4	30.8	31.5	30.4	30.6	32.9
	SI	31.2	29.3	28.5	28.7	29.9	28.6	30.2	30.9	31.8	31.6	31.1	31.7	30.7	30.8	33.2

3. Experimental results and discussion

For the performance evaluation, we applied our method to 2D MR images and synthetic 3D phantoms. Because our method reconstructs the image of a 3D volume by fusing fiducial pixels to form contiguous 2D transverse images, experiments with 2D images are close simulations of the 3D scenarios and permit evaluation of reconstruction accuracy. In the experiments with 3D phantoms, we segmented the reconstructed 3D images and compared the estimated volume against the ground truth. In these experiments, we compared our method (WBLF) with the conventional linear interpolation (LI), cubic interpolation (CI), spline interpolation (SI), the fusion method (Gosh) by Goshtasby and Turner [9] and the registration-based interpolation method (RI) by Penney et al. [13]. Last, we applied our method to the multi-planar, bidirectional MR scans of the human tongue.

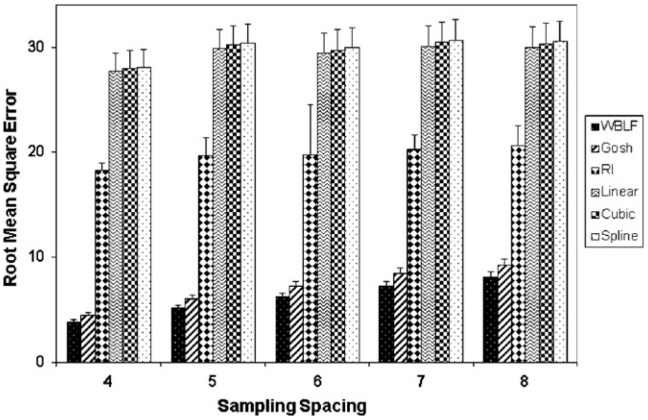


Fig. 6. Average RMSE given various sampling spaces.

3.1. Performance analysis

3.1.1. 2D image experiments

Our experimental 2D image set consists of 15 contiguous sagittal MR slices. Prior to the reconstruction, a histogram stretching was used to map the original images from 16-bit gray scale to 8-bit. The rescaled images were used as our references.

Given a sampling space λ, a row-sample image was generated by extracting every λth row in the reference image. For example,

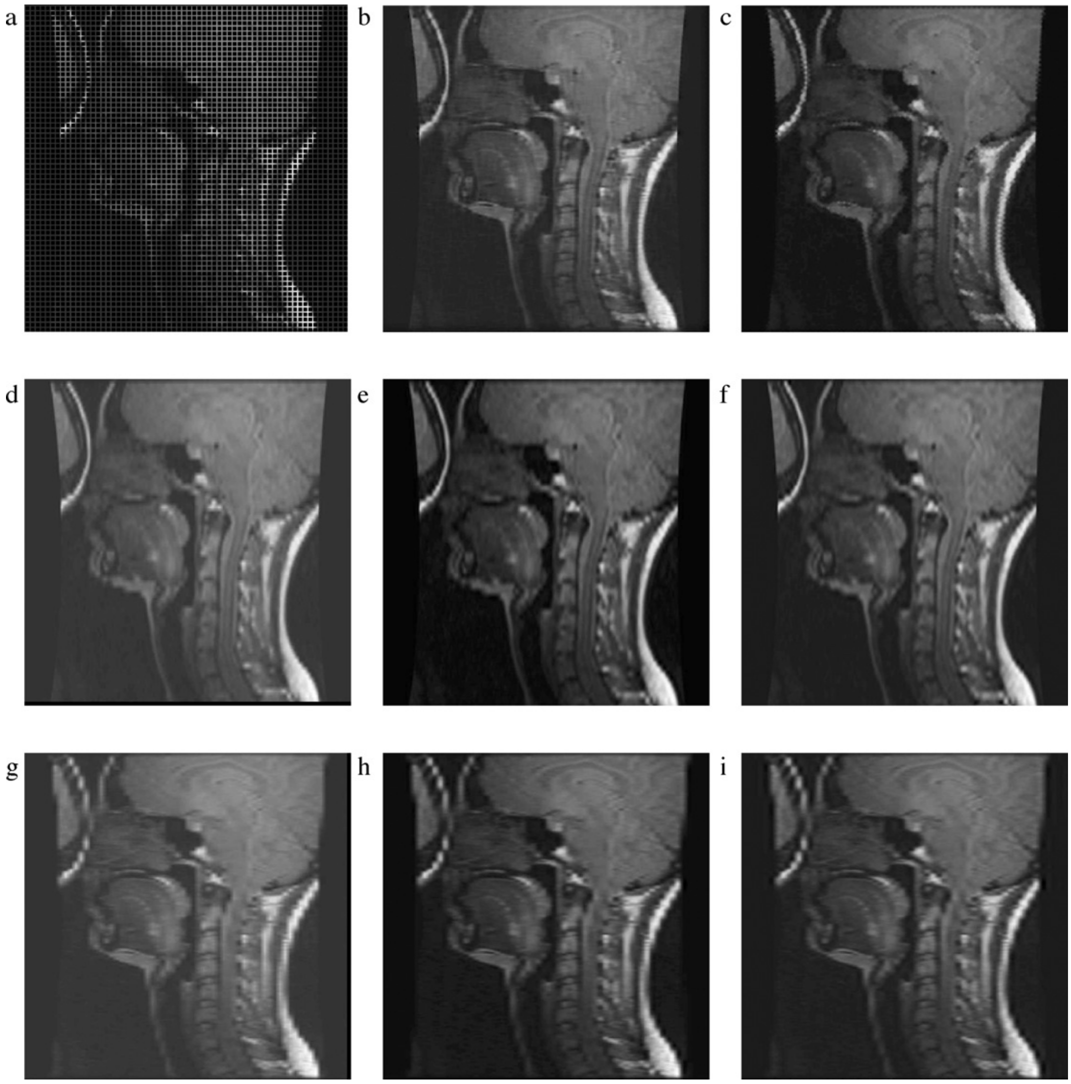


Fig. 7. 2D image reconstruction at sampling space 4. (a) Original image. (b) WBLF reconstruction result. (c) Distance-weighted fusion result. (d) Linear interpolation using row samples. (e) Cubic interpolation using row samples. (f) Spline interpolation using row samples. (g) Linear interpolation using column samples. (h) Cubic interpolation using column samples. (i) Spline interpolation using column samples.

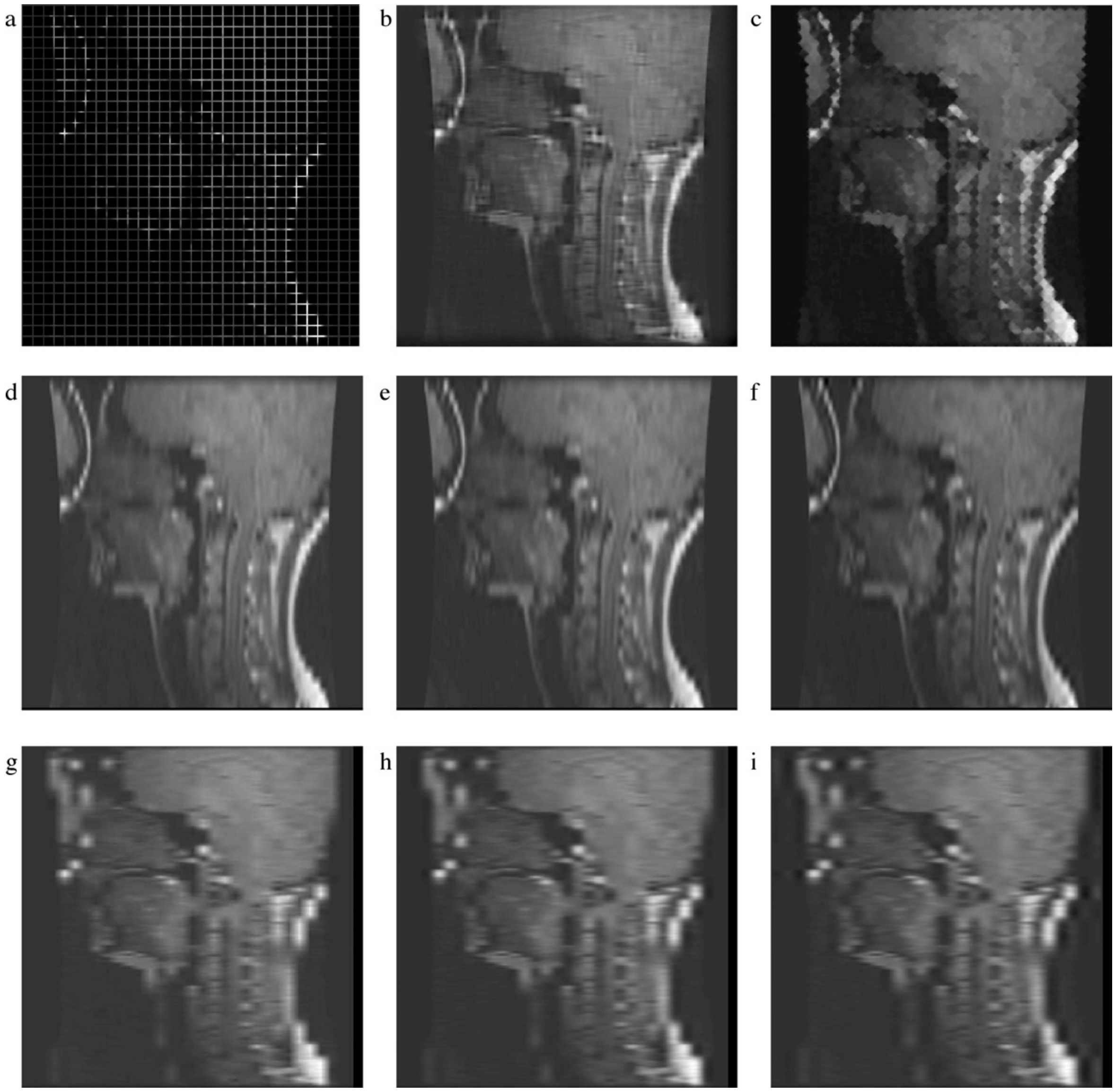


Fig. 8. 2D image reconstruction at sampling space 8. (a) Original image. (b) WBLF reconstruction result. (c) Distance-weighted fusion result. (d) Linear interpolation using row samples. (e) Cubic interpolation using row samples. (f) Spline interpolation using row samples. (g) Linear interpolation using column samples. (h) Cubic interpolation using column samples. (i) Spline interpolation using column samples.

for a 256×256 image with a sampling space of 8, a row-sample image is a 32×256 matrix that consists of rows $\{8, 16, 24, \ldots, 256\}$ from the reference image. Similarly column-sample images were created.

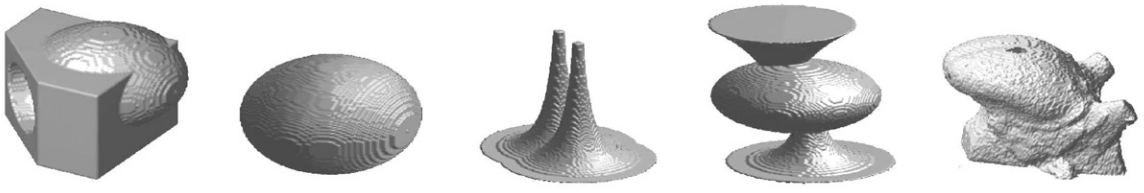
We used the root mean square error (RMSE) with respect to the reference images as a metric for reconstruction accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{1}^{N}(\hat{I} - I)^2},$$

where $\hat{I}$ and I denote the reconstructed and the reference images, respectively, and N is the total number of pixels.

Table 1 lists the RMSE of the methods at sampling spaces 4, 5, 6, 7, and 8. Fifteen experiments (in columns) were conducted and each row shows the RMSE of a method at a given sampling space. The reconstruction error increases as the sampling space increases. The mean row-sample-based and column-sample-based reconstruction RMSE values for the linear, cubic and spline interpolations were very close because the interpolation was unidirectional (i.e., performed in one dimension). The reconstruction error of these methods is the greatest. Registration-based interpolation (RI) resulted in smaller errors than the conventional interpolation methods. However, its results are sensitive to the registration accuracy and depict greater variance (as shown in Fig. 6). It is evident that fusion-based methods (Gosh and WBLF) are able to achieve a more precise reconstruction by integrating samples of two directions and our WBLF method (WBLF) consistently produced the smallest RMSE values. This result is further demonstrated in Fig. 6, which illustrates the average RMSE values of the five reconstruction methods as a function of sampling space. A two-way ANOVA ($\alpha = 0.05$) revealed significant improvement of our method in all sampling spaces ($p < 0.0001$ in each case). The Bonferroni post-tests revealed that our method had significantly lower RMSE scores at any sampling spacing than the conventional interpolation methods. It is clear that our method resulted in the smallest reconstruction errors in 2D scenario.

Figs. 7 and 8 illustrate the 2D reconstructed images at sampling spaces 4 and 8. The row-sample interpolation, (d)–(f) in Figs. 7 and 8, and column-sample interpolation, (g)–(i), preserved



(1) (2) (3) (4) (5)

Fig. 9. 3D phantoms used in our experiments.

vertical and horizontal details, respectively. In both cases, our method produced the most visually favorable reconstructions with elimination of local discontinuities, as shown in Fig. 7(b) and 8(b) in comparison to Fig. 3(c). The reconstructed images in Figs. 7 and 8 using LI, CI, and SI methods depict much difference; whereas the quantitative differences of the average RMSE error between sampling spaces 4 and 8 are 2.19, 2.34, and 2.45 for LI, CI, and SI, respectively. The statistics are computed by combining the errors of reconstructions using row samples and column samples, and represent a summary of 15 cases. Comparing against their average error magnitudes, which are 27.77, 27.96, and 28.08 at space 4 and 29.96, 30.3, and 30.53 at space 8 for LI, CI, and SI, respectively, the visual and quantitative results agree.

3.1.2. 3D volume estimation

We adopted the phantom generation approach used in [9] and created 3D phantoms that have distinct surfaces. Voxels within the phantoms have an intensity value of 200, whereas voxels outside the phantoms have values of 100. The 3D matrix size is 128 × 128 × 128. Five phantoms were created, including a digitized human tongue model, as illustrated in Fig. 9. These phantoms represent a variety of objects of different geometric characteristics.

Segmentation of the object was based on thresholding. The threshold was determined by choosing the medium value in the histogram of the reconstructed volume.[1] In our experiments, the value 150 was used as threshold.

We compared our WBLF method with the distance-weighted fusion method, registration-based interpolation, and linear and cubic interpolations. In addition, we performed volume estimation based on segmentation of 2D slices. That is, the 3D volume was approximated with the product of the area of the segmented 2D objects and the sampling space. For each method, we calculated the reconstructed volume error ratio (VER) based on the estimated volume, V^*; and the true phantom volume, V:

$$\mathrm{VER} = \frac{|V - V^*|}{V}. \tag{17}$$

Table 2 presents the error ratios at various sampling spaces with the best performance highlighted with bold face font. The volume is presented with the total number of voxels and listed in the second column. When the slice spacing was small, the methods achieve small VER values in volume estimation. Out of 25 experimental sets, our WBLF-based reconstruction resulted in best performances in 18 cases. When the object is in simple, regular shape, registration-based reconstruction method produced superior model (see phantom 2 for an example), had the smallest VER values. For phantoms 2, 3, and 4, the VER of WBLF at sampling space 8 was smaller than that of the other methods at sampling space 4. Statistical analysis ($\alpha = 0.05$) revealed a significant improvement in 3D reconstruction using our method ($p < 0.0001$).

In the experiments with phantom 2, both registration-based interpolation and the 2D segmentation produced exceptional better results. As shown in Fig. 9, phantom 2 is an ellipsoid with gradual change of its surface. Such a property favors registration-based interpolation method. However, WBLF-based method produced very competitive accuracy in volume estimation. Both registration-based interpolation and the 2D segmentation-based method encountered enormous difficulties when the object's shape was irregular or when the shape of the cross section changed sharply, for instance, phantom 4 and 5 in Fig. 9. In contrast, our method constantly achieved superior precision.

Fig. 10 illustrates the average volume estimation error ratios of the phantoms by sampling space as a function of reconstruction method. Our WBLF method had the smallest mean VER values and standard deviations in all cases and significantly better performance than the 2D method when sampling space is large, i.e. at 7 or 8, based on the Bonferroni post-tests. Note that VER values for the 2D segmentation-based method increased significantly with the increase in sampling space, and that the variances were the greatest. No significant difference was found between the bilinear and the bicubic interpolation methods, although the latter had a slight advantage at a larger sampling space. In contrast to WBLF-based method, registration-based interpolation is impacted by the registration error and depicted greater variance as the sampling spacing increases.

3.2. 3D reconstruction of the human tongue

Our *in vivo* human MR images were acquired in a GE Excite 3 T scanner using a fast spin-echo pulse sequence and the following parameters: $TR = 1916$ ms, $TE = 9.2$ ms, 4 mm slice thickness, zero slice spacing, 240 × 240 mm field of view (FOV), and 0.8 phase FOV. Fifteen sagittal slices and nineteen coronal slices were acquired for every subject. They were assembled in a 3D cube based on the physical slice locations recorded in the image header files. To achieve isometric 3D reconstruction, we needed to recover 8 slices between every two fiducial slices.

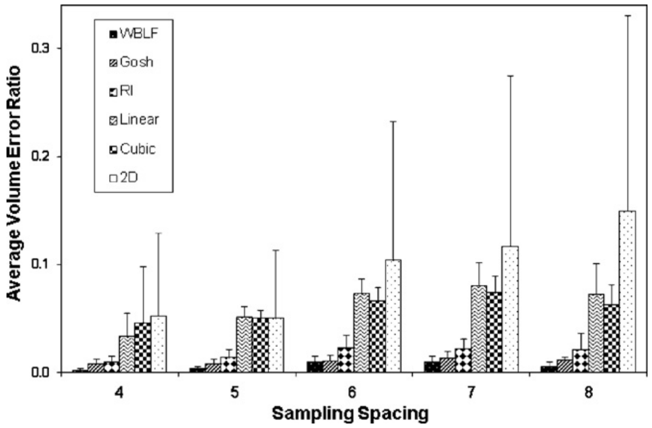


Fig. 10. The volume error ratio. Vertical bars denote the variance.

---

[1] Due to interpolation, the voxel values span in a range.

Table 2
The volume error ratio using different reconstruction methods. The volume is measured in the number of voxels. Pha: phantom. SS: sampling space.

Pha.	Voxel	SS	WBLF	Gosh	RI	LI	CI	2D
1	178,122	4	0.0051	0.0017	0.0073	0.0081	0.0076	0.0260
		5	0.0024	0.0058	0.0152	0.0386	0.0386	0.0250
		6	0.0059	0.0042	0.0268	0.0593	0.0591	0.0600
		7	0.0040	0.0093	0.0271	0.0692	0.0693	0.0570
		8	0.0091	0.0130	0.0235	0.0421	0.0426	0.0710
2	205,093	4	0.0006	0.0085	0.0039	0.0247	0.0240	0.0070
		5	0.0049	0.0060	0.0043	0.0493	0.0490	0.0140
		6	0.0136	0.0109	0.0092	0.0721	0.0721	0.0090
		7	0.0158	0.0142	0.0120	0.0893	0.0893	0.0120
		8	0.0026	0.0105	0.0054	0.0441	0.0405	0.0410
3	112,876	4	0.0002	0.0149	0.0057	0.0657	0.1370	0.1890
		5	0.0028	0.0157	0.0066	0.0485	0.0546	0.1620
		6	0.0132	0.0176	0.0106	0.0829	0.0578	0.3280
		7	0.0106	0.0201	0.0114	0.1138	0.0907	0.3970
		8	0.0018	0.0155	0.0156	0.1050	0.0750	0.4700
4	342,055	4	0.0006	0.0079	0.0092	0.0230	0.0149	0.0200
		5	0.0048	0.0060	0.0195	0.0659	0.0580	0.0250
		6	0.0125	0.0116	0.0388	0.0896	0.0849	0.0910
		7	0.0124	0.0167	0.0323	0.0696	0.0589	0.0800
		8	0.0006	0.0085	0.0367	0.0907	0.0781	0.1090
5	624,527	4	0.0001	0.0051	0.0182	0.0420	0.0419	0.0183
		5	0.0015	0.0057	0.0219	0.0513	0.0493	0.0227
		6	0.0016	0.0059	0.0250	0.0594	0.0571	0.0292
		7	0.0015	0.0037	0.0251	0.0596	0.0589	0.0345
		8	0.0019	0.0064	0.0341	0.0800	0.0759	0.0541

Fig. 11 illustrates the re-sliced transverse slices of a representative volumetric reconstruction. The top two rows are the reconstructed images using our method in a zoomed view of the tongue. In contrast, the bottom two rows show the results of the conventional linear interpolation method without fusion, applied to the same original images with the same slice locations. Clearly, by fusing bidirectional MR slices, the tongue boundary is greatly improved and the staircase artifact is eliminated. In addi-

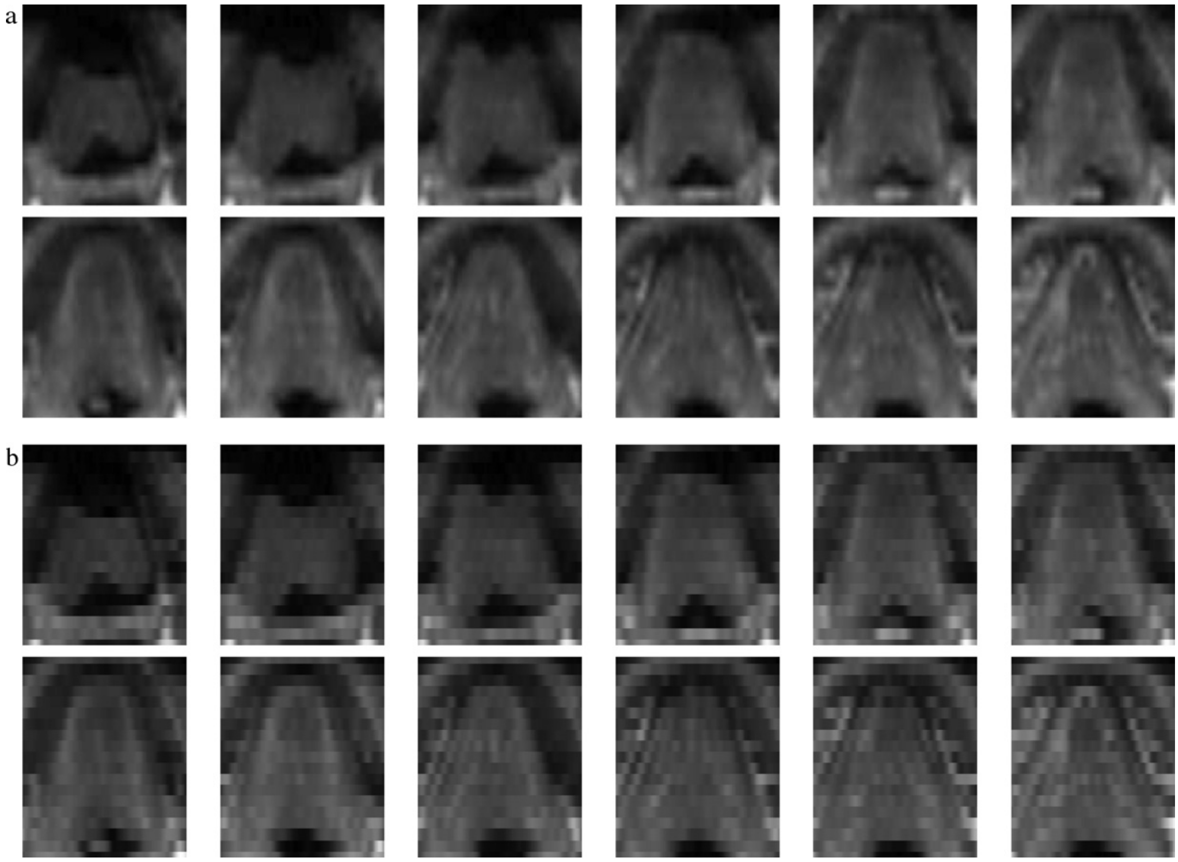


Fig. 11. Re-sliced transverse images of the reconstructed 3D tongue. The top two rows are the results of fusion. The bottom two rows are the results produced with conventional linear interpolation.

tion, the tongue body appears more homogeneous, which can potentially aid segmentation for more accurate delineation of the object.

4. Conclusions

In this paper, we present a wavelet-based fusion method to integrate multi-planar MR images and improve the accuracy of 3D reconstruction. Our method consists of an interpolation step followed by a wavelet-based fusion. First, we use our bidirectional linear interpolation algorithm to estimate the values of the unknown voxels based on fiducial pixels in the sagittal and the coronal slices and construct a transverse image. Using the same sagittal and coronal slices separately, we then generated two transverse images with linear interpolation. In the second step, a wavelet transform is used to decompose three transverse images into subbands. A fused coefficient matrix is constructed by taking the most significant coefficients. The final reconstruction is achieved with an inverse wavelet transform.

Experiments were conducted using 2D images and 3D phantoms. Our experimental results demonstrated that our fusion-based reconstruction method resulted in the smallest reconstruction error consistently in both 2D and 3D scenarios. In 3D volume estimation, both registration-based interpolation method and segmentation-based method performed well with regular-shaped objects. However, our method resulted in competitive performance and achieved the best results with other complicated objects. Further, by preserving crucial object edges without compromising the local spatial continuity, our wavelet-based fusion method successfully suppressed interpolation artifacts and produced improved 3D images.

The complexity of our algorithm for 3D reconstruction is $O(n^3 - ((2\lambda - 1)/\lambda^2)n^2)$, where λ is the sampling space and n is the dimension of the reconstructed 3D image.

The advantage of our method casts new light on MR imaging and image processing, as it is especially useful when high resolution and short acquisition time cannot be achieved simultaneously. We believe that accurate isotropic reconstruction can improve image segmentation performance. Our future work will focus on minimization of reconstruction errors induced by image noise and misregistration and the development of 3D segmentation algorithms using the fused 3D volume images.

References

[1] Clarke LP, Velthuizen RP, Camacho MA, Heine JJ, Vaidyanathan M, Hall LO, et al. MRI segmentation: methods and applications. Magnetic Resonance Imaging 1995;13(3):343–68.

[2] Dickens MM, Gleason SS, Sari-Sarraf H. Volumetric segmentation via 3D active shape models. In: Proceedings of the IEEE southwest symposium on image analysis and interpretation. 2002. p. 248–52.

[3] Kang Y, Engelke K, Kalender WA. A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. IEEE Transactions on Medical Imaging 2003;22(5):586–98.

[4] Yao J, Summers RM. Three-dimensional colonic polyp segmentation using dynamic deformable surface. In: Proceedings of SPIE, medical imaging, vol. 5369. 2004. p. 280–9.

[5] Grevera GJ, Udupa JK. An objective comparison of 3-d image interpolation methods. IEEE Transactions on Medical Imaging August 1998;17(4):642–52.

[6] Lichtenbelt B, Crane R, Naqvi S, Crane R, Naqvi S. Introduction to volume rendering. Prentice Hall; 1998.

[7] Keys RG. Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing 1981;29(6):1153–60.

[8] Kuwahara M. 3-D heart image reconstruction from MRI data. Computerized Medical Imaging and Graphics 1991;15(4):241–6.

[9] Goshtasby AA, Turner DA. Fusion of short-axis and long-axis cardiac MR images. In: Proceedings of the 1996 workshop on mathematical methods in biomedical image analysis. 1996.

[10] Werahera PN, Miller GJ, Taylor GD, Brubaker T, Daneshgari F, Crawford ED. A 3-D reconstruction algorithm for interpolation and extrapolation of planar cross sectional data. IEEE Transactions on Medical Imaging December 1995;14(4):765–71.

[11] Raya SP, Udupa JK. Shape-based interpolation of multidimensional objects. IEEE Transactions on Medical Imaging 1990;9:32–42.

[12] Grevera GJ, Udupa JK. Shape-based interpolation of multidimensional grey-level images. IEEE Transactions on Medical Imaging December 1996;15(6):881–92.

[13] Penney GP, Schnabel JA, Rueckert D, Viergever MA, Niessen WJ. Registration-based interpolation. IEEE Transactions on Medical Imaging July 2004;23(7):922–6.

[14] Frakes DH, an LPD, Pekkan dK, Kitajima HD, Sundareswaran K, Yoganathan AP, et al. A new method for registration-based medical image interpolation. IEEE Transactions on Medical Imaging March 2008;27(3):370–7.

[15] Bao P, Zhang L. Noise reduction for magnetic resonance images via adaptive multiscale products thresholding. IEEE Transactions on Medical Imaging 2003;22(9):1089–99.

[16] Mallat S, Zhong S. Characterization of signals from multiscale edges. IEEE Transactions on Pattern Analysis and Machine Intelligence 1992;14(7):710–32.

[17] Niedermeier A, Romaneeben E, Lehner S. Detection of coastlines in SAR images using wavelet methods. IEEE Transactions on Geoscience and Remote Sensing 2000;38(5):2270–81.

Xiaohui Yuan received a B.S. degree in electrical engineering from Hefei University of Technology, China in 1996 and a Ph.D. degree in computer science from Tulane University, U.S. in 2004. After graduation, he worked at the National Institutes of Health for two years on medical image analysis. He is an assistant professor in the Computer Science and Engineering Department at the University of North Texas. His research interests include data mining, machine learning, image processing, and pattern recognition, and he has published more than 50 technical papers. He is a member of IEEE and SPIE.

Xiaojing Yuan has years' experiences in embedding intelligence into sensors and actuators to deal with uncertainties. She authored and co-authored more than 50 technical papers; has one patent; and filed a permanent patient application in USA in 2010. She joined the University of Houston in 2004 after getting her Ph.D. in mechanical engineering from Tulane University in 2003 and working as postdoctoral researcher in Imaging Understanding Lab in Tulane University. Her research interests include pattern recognition, biomedical imaging analysis, data-centric data analysis and modeling, and intelligent wireless sensor networks. She is a member of IEEE, ASEE, and ISA.