

# Reconstruction of High-Resolution Tongue Volumes From MRI

Jonghye Woo\*, *Member, IEEE*, Emi Z. Murano, Maureen Stone, and Jerry L. Prince, *Fellow, IEEE*

**Abstract**—Magnetic resonance images of the tongue have been used in both clinical studies and scientific research to reveal tongue structure. In order to extract different features of the tongue and its relation to the vocal tract, it is beneficial to acquire three orthogonal image volumes—e.g., axial, sagittal, and coronal volumes. In order to maintain both low noise and high visual detail and minimize the blurred effect due to involuntary motion artifacts, each set of images is acquired with an in-plane resolution that is much better than the through-plane resolution. As a result, any one dataset, by itself, is not ideal for automatic volumetric analyses such as segmentation, registration, and atlas building or even for visualization when oblique slices are required. This paper presents a method of superresolution volume reconstruction of the tongue that generates an isotropic image volume using the three orthogonal image volumes. The method uses preprocessing steps that include registration and intensity matching and a data combination approach with the edge-preserving property carried out by Markov random field optimization. The performance of the proposed method was demonstrated on 15 clinical datasets, preserving anatomical details and yielding superior results when compared with different reconstruction methods as visually and quantitatively assessed.

**Index Terms**—Human tongue, magnetic resonance imaging (MRI), superresolution volume reconstruction.

## I. INTRODUCTION

THE mortality rate of oral cancer including tongue cancer is not considered to be high, but its morbidity in terms of speech, mastication, and swallowing problems is significant and seriously affects the quality of life of these cancer patients. Characterizing the relationship between structure and function in the tongue is becoming a core requirement for both clinical diagnosis and scientific studies in the tongue/speech research community [1], [2]. In recent years, medical imaging, especially magnetic resonance imaging (MRI), has played an important role in this effort. MRI is a noninvasive technology that has been extensively used over past two decades to an-

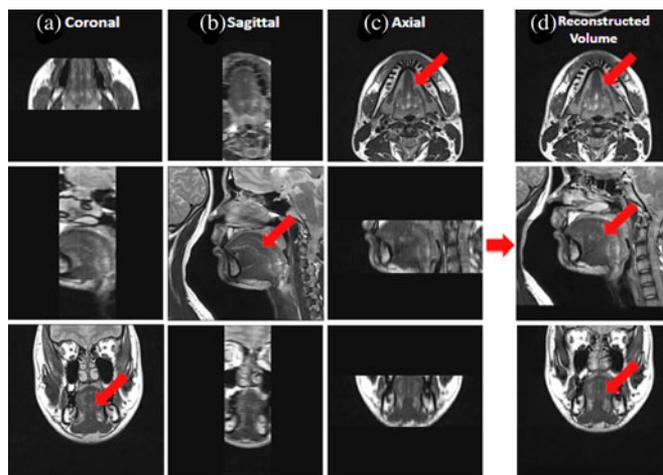


Fig. 1. Tongue images are acquired in three orthogonal volumes with FOV encompassing the tongue and surrounding structures. (a) Coronal, (b) sagittal, and (c) axial volumes are illustrated. The final superresolution volume using the proposed method is shown in (d). The red arrows indicate the tongue region.

alyze tongue structure and function ranging from studies of the vocal tract [3]–[6] to studies on tongue muscle deformation [1], [7]–[9]. For instance, high-resolution MRI provides exquisite depiction of muscle anatomy while cine MRI offers temporal information about its surface motion. With the growth in the number of images that can be acquired, research studies now involve 3-D high-resolution images taken from different individuals, time points, and MRI modalities. Therefore, the requirement for automated methods that carry out image analysis of the acquired tongue image data is expected to grow rapidly.

Time limitations of current MRI acquisition protocols make it difficult to acquire a single high-resolution 3-D structural image of the tongue and vocal tract. It is almost certain that tongue motion—particularly the gross motion of swallowing—will spoil every attempt to acquire such data. The 3-D acquisition takes at least 4–5 min and holding the tongue steady for more than 2–3 min will be likely to induce the involuntary motion. Therefore, in our current acquisition protocol, three orthogonal volumes with axial, sagittal, and coronal orientations are acquired one after the other. Motion may occur between these scans but the subject will return the tongue to a resting position for each scan. To speed up each acquisition, the fields of view (FOVs) of each acquired orientation are limited to encompass only the tongue itself, as shown in Fig. 1(a)–(c). Also, in order to rapidly acquire each stack of images with contiguous images (no gaps between the images), the through-plane resolution is worse than the in-plane resolution. In our case, the images have

Manuscript received May 8, 2012; revised August 31, 2012; accepted September 5, 2012. Date of publication September 27, 2012; date of current version November 22, 2012. This paper was presented in part at the 2012 SPIE Medical Imaging Conference. This work was supported by the National Institutes of Health under Grant R01CA133015 and Grant R00DC009279. *Asterisk indicates corresponding author.*

\*J. Woo is with the University of Maryland, Baltimore, MD 21201 USA, and also with Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: jschant@gmail.com).

E. Z. Murano and J. L. Prince are with the Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: emurano1@jhmi.edu; prince@jhu.edu).

M. Stone is with the University of Maryland, Baltimore, MD 21201 USA (e-mail: stone@umaryland.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBME.2012.2218246

an in-plane resolution of  $0.94 \text{ mm} \times 0.94 \text{ mm}$  but are 3 mm thick.

Because the slices are relatively thick and the regions of interest are tongue regions (e.g., tongue muscles) where certain muscle bundles are short and thin, none of the acquired image stacks are ideal for 3-D volumetric analyses such as segmentation, registration, and atlas building or even for visualization when oblique views are required. In particular, since each volume has poorer resolution in the slice-selection direction than in the in-plane direction, it is difficult to observe tongue muscles clearly in any one of the volumes. Therefore, reconstruction of a single high-resolution volumetric tongue MR image from the available orthogonal image stacks will improve our ability to visualize and analyze the tongue in living subjects.

In this study, we develop a fully automated and accurate superresolution volume reconstruction method from three orthogonal image stacks of the same subject by extending our preliminary approach [10]. We present a refined preprocessing and reconstruction algorithm design and provide validations on both tongue and brain data. This application is quite special in the sense that the function of the tongue involves a highly complex coordination of its different muscles to produce the precise movements necessary for proper speech, mastication, and deglutition. In addition, the tongue is a muscular hydrostat, with no bones or joints, thereby requiring an architecture of 3-D orthogonal muscles to move and deform it when producing speech [11]. Because of the nonrigid motion (e.g., tongue motion or swallowing), different FOV, and different intensity distributions between scans, our problem posed unique challenges, requiring specialized preprocessing. The challenges motivate us to use a number of preprocessing steps including motion correction, intensity normalization, etc., followed by a region-based maximum a posteriori (MAP) Markov random field (MRF) approach. The region-based approach allows us to reconstruct the tongue at high resolution, because it resides in the intersection of the image FOVs, as well as other regions around the tongue which are also used in the analysis of speech production. In order to preserve important anatomical features such as the subtle boundaries of muscle, we use edge-preserving regularization. To our knowledge, this is the first attempt at superresolution reconstruction applied to *in vivo* tongue high-resolution MR images for both normal and diseased subjects. The resulting superresolution volume improves both the signal-to-noise ratio (SNR) and resolution over the source images, thereby approximating an original high-resolution volume whose acquisition would have taken too long for the subject to refrain from swallowing (or making other motions).

Improvements in visualizing the internal structure of the tongue may provide better ways to determine tongue cancer boundaries relative to the tongue muscles leading to better tumor staging and planning for cancer surgery. For example, the impact of surgery on speech can be better understood by visualizing the muscles that will be affected by surgery and knowing their role in speech production. For scientific study of the role of muscles, this approach provides an alternative to the visualization of tongue muscle topography in cadaveric anatomy. In a cadaveric specimen, the lack of a rigid framework and the mus-

cle tone provided by the cocontraction of interdigitated extrinsic and intrinsic muscles lead to a loss of the *in vivo* topography. In summary, our reconstruction method yields a high-resolution, volumetric image of the *in vivo* tongue that can be ideal for both clinical and scientific purposes.

The structure of this paper is as follows. Related work is reviewed in Section II. Section III shows the proposed method using a region-based MAP-MRF approach. Section IV explains the validation method, followed by the experimental results of the proposed method on numerical simulations and *in vivo* MR images of the tongue in Section V. Section VI provides a discussion, and finally, Section VII concludes this paper.

## II. RELATED WORK

Superresolution image reconstruction from multiple low-resolution images has been an area of active research since the seminal work by Tsai and Huang [12]. It is an important problem in a variety of fields such as video restoration, surveillance, remote sensing, and medical imaging [13]–[15]. A superresolution reconstruction can overcome the limitations of the imaging system and enhance the performance of a variety of postprocessing tasks such as image segmentation, registration, atlas building, and motion analysis. If one knows the point spread function (PSF) of the imaging/camera system and also observes enough samples of shifted images, then resolution can be improved and noise reduced over the individual observations up to the limits posed by Shannon [16], [17]. It remains a difficult problem in practice, however, due to insufficient numbers of observed images, inexact registration, and inaccurate PSF and noise models.

Superresolution methods can be categorized into nonuniform interpolation, frequency domain, and spatial domain approaches [13]–[15], [18]. The nonuniform interpolation approach is the most intuitive method consisting of successive steps including 1) registration; 2) nonuniform interpolation; and 3) deblurring process. The advantages of this approach are that it is computationally inexpensive and makes real-time application possible. The disadvantages of this approach are that degradation models are limited and the optimality of the whole reconstruction is not guaranteed [14].

The frequency-domain approach utilizes aliasing in low-resolution images to reconstruct a high-resolution image [14]. Tsai and Huang [12] first formulated the multiframe superresolution reconstruction approach in the frequency domain. This approach is based on fundamental principles including 1) the shifting property of the Fourier transform; 2) the aliasing relationship between the continuous Fourier transform and the discrete Fourier transform (DFT); and 3) the band-limited property of the original scene. Frequency-domain approaches are generally intuitive and computationally efficient; however, they cannot easily incorporate spatial knowledge such as spatially varying degradation and poor image registration [18], [19] and they are also very sensitive to modeling errors [20].

The spatial-domain approach synthesizes high-resolution information either by learning from training sets (i.e., example-based approaches) to incorporate models involving

spatiallyvarying physical phenomena or by performing regularized reconstruction (i.e., model-based approaches). Example-based approaches capture the co-occurrence between low-resolution and high-resolution patches and synthesize high-resolution images by combining the high-resolution patches [21]–[30]. Model-based approaches incorporate forward models that specify how the high-resolution true image is transformed by the physics of the imaging system to generate low-resolution images including global or local motion, motion blur, varying PSF, compression artifacts, etc. This approach provides flexibility in modeling physical phenomenon whereas the computational complexity may be demanding [18]. Examples of the model-based approach include the projection onto convex sets (POCS) approach [14] and the maximum likelihood (ML)-POCS hybrid reconstruction approach [14]. The POCS approach simultaneously solves the restoration and interpolation problem to obtain the high-resolution image. The ML-POCS hybrid reconstruction approach finds high-resolution estimates by minimizing the ML cost functional with constraints. A superresolution reconstruction is found by solving an inverse problem yielding a high-resolution image that is likely to be the source of the acquired images under the modeled imaging conditions [31]. The classical solution of a linear inverse problem is analytically obtained through the Moore–Penrose pseudoinverse of the observation model [14]. In practice, however, the computation is demanding, and therefore, iterative solutions such as steepest descent [32] and expectation maximization [33] have been adopted to solve the problem [31]. In this study, we adopt a model-based spatial-domain approach because it allows us to incorporate *a priori* knowledge using constrained reconstruction algorithms to enhance the resolution through a variety of postprocessing techniques. A comprehensive survey on superresolution reconstruction can be found in [14] and [15].

Several studies have succeeded in reconstructing a high-resolution volume from a set of multislice MR images [15], [34]–[40]. Important factors in any MRI experiment include how to balance interdependent variables such as image resolution, SNR, and acquisition time. Longer acquisition time provides higher spatial resolution and SNR compared with shorter acquisition time and vice versa [34]. Superresolution reconstruction methods in MRI were first proposed by Peled and Yeshurun [39] and Greenspan *et al.* [40], adapting algorithms from computer vision community. A recent study has shown that superresolution volume reconstruction provides better tradeoffs between resolution, SNR, and acquisition time than direct high-resolution acquisition, which is especially useful when SNR and time constraints are taken into account [34]. In this study, image acquisition is limited by the time constraints of speech, and therefore, the factors were balanced to optimize the superresolution reconstruction.

In brain imaging, it is common to take multiple scans of the same subject from different orientations, where the acquired images typically have better in-plane than through-plane resolution. In particular, model-based reconstruction approaches have been actively used. Bai *et al.* [41] proposed a MAP superresolution method to reconstruct a high-resolution volume from two orthogonal scans of the same subject. This strategy is at

the heart of the method we propose here. In addition, Gholipour *et al.* [35] investigated a MAP superresolution method with image priors based on image gradients. Both of these methods used orthogonal stacks. Shilling *et al.* [37] proposed a superresolution reconstruction approach using rotated slice stacks based on POCS formalism for image-guided brain surgery. These methods did not address subject motion.

In addition, several researchers have developed methods for fetal brain imaging where due to uncontrolled motion it is common to acquire multiple orthogonal 2-D multiplanar acquisitions with anisotropic voxel sizes [18], [42], [43]. To yield a single high-resolution registered volume image of the fetal brain from this kind of data, Rousseau [29] investigated so-called *brain hallucination* by incorporating a high-resolution brain image of a different subject (an atlas) in order to synthesize likely high-resolution texture patches from a low-resolution image of a subject. Although this method is promising, we maintain that it is better to reconstruct higher resolution images from actual imaging data whenever possible rather than relying on examples that may or may not be representative of the object actually being imaged. This is especially true in the imaging of abnormal anatomy, such as the tongue cancer patients who are the primary target of our scientific studies on the tongue. Rousseau *et al.* [18] incorporated a registration method to correct motion and the final reconstruction process was based on a local neighborhood approach with a Gaussian kernel. Jiang *et al.* [43] proposed motion correction using registration and B-spline-based scattered interpolation to reconstruct the final 3-D fetal brain. The methods in [18] and [43] addressed the subject motion but employed scatter data interpolation, which ignores the physics of imaging. Gholipour *et al.* [31] proposed maximum likelihood (M-estimation) error norm to reconstruct fetal MR images using a model of subject motion. This work addressed both interslice motion and anisotropic slice acquisition using slice-to-volume registration and M-estimation solution, respectively. Our study also incorporated a model of subject and tongue motion. In recent work, Rousseau *et al.* [42] used an edge-preserving regularization method to reconstruct high-resolution images from the fetal MRI images and investigated the impact of the number of low-resolution images used. These approaches incorporate registration to align the observed data to a single anatomical model; this is a strategy we also use herein (although we use deformable registration due to the nature of the human tongue).

A superresolution reconstruction technique was successfully adopted to reconstruct diffusion-weighted images (DWI) from multiple anisotropic orthogonal DWI scans [44]. The problem of patient motion was addressed by aligning the volumes both in space and q-space. This advanced the state-of-the-art superresolution MRI application. In addition, a MAP formulation was used to tackle the reconstruction problem. In this study, we propose similarly to investigate a superresolution reconstruction approach for tongue MR images for the first time, aiming to advance the state-of-the-art superresolution MRI application. Inspired by the DWI application and recent developments in brain MR imaging [41], [42], our approach uses an MR image acquisition model, region-based subject/tongue motion correction, intensity matching between scans, and a region-based MAP

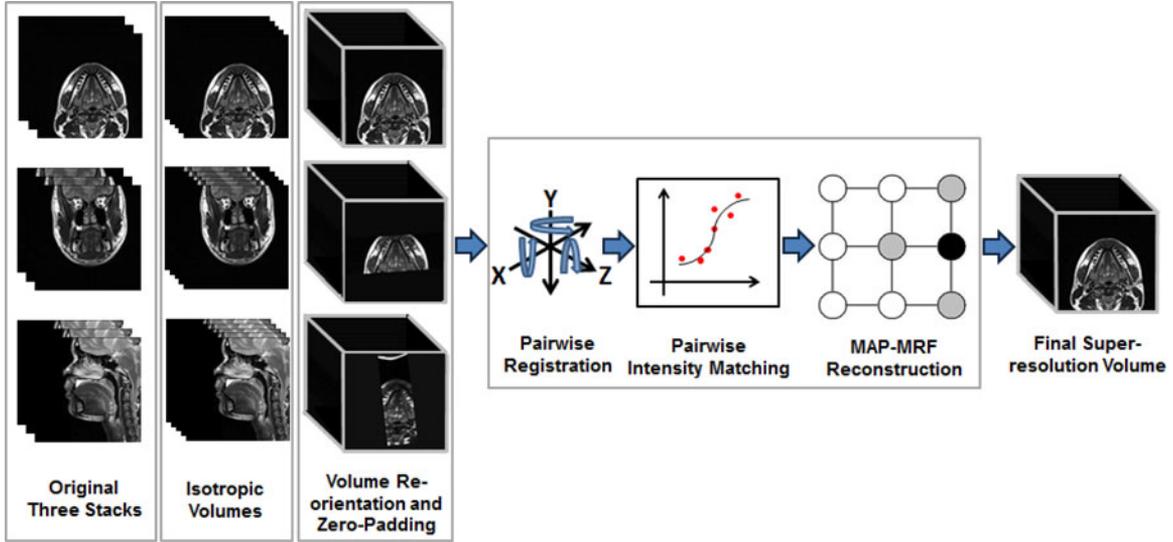


Fig. 2. Flowchart of the proposed method.

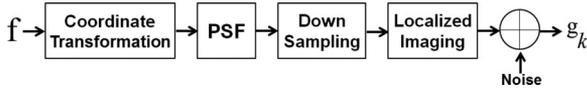


Fig. 3. Model of the imaging system.

formulation with an edge-preserving regularization to preserve crisp muscle information.

### III. METHOD

We view superresolution volume reconstruction as an inverse problem that combines denoising, deblurring, and up-sampling in order to recover a high-quality volume from several low-resolution volumes. In contrast to other superresolution frameworks, our application poses unique challenges in that multislice tongue MR data are affected by nonrigid motion (e.g., tongue motion or swallowing) between acquisitions and each volume is acquired with a different FOV. To efficiently tackle this problem, our proposed method comprises several steps: 1) preprocessing to address different orientation, size, and resolution; 2) registration to correct subject motion between acquisitions; 3) intensity normalization between volumes; and 4) a region-based MAP-MRF reconstruction. A flowchart of the proposed method is shown in Fig. 2.

#### A. Imaging Model With Partial Overlap Region

In our problem, the imaging system involves a coordinate transformation, a nonuniform decimation, a PSF, noise, and a limited FOV, as illustrated in Fig. 3. Let the high-resolution volume  $\mathbf{f} : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$  be defined on the open and bounded domain  $\Omega$ .

The FOVs of the low-resolution volumes are not identical because the acquisition protocols are designed to be “tight” around the tongue in each orientation (see Fig. 4). Let  $\Lambda_1$ ,  $\Lambda_2$ , and  $\Lambda_3$  be the regions formed on the image domain by the axial, coronal, and sagittal volumes, respectively. In particular,

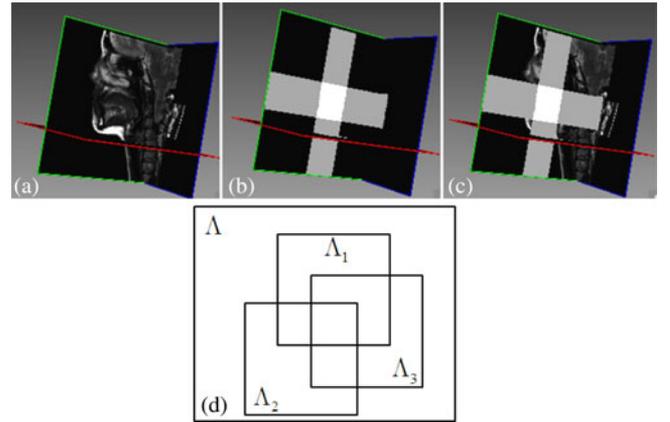


Fig. 4. (a) One representative final superresolution image; (b) regions defined from each low-resolution volume; (c) volume with regions defined from each low-resolution volume; and (d) schematic drawing of each region.

$\Lambda_k \subset \Omega$  is a subset of pixels at which the value of each low-resolution volume is observed acquisitions.

For example, as illustrated in Fig. 4(b) and (c), the white region represents  $S_{\Lambda_1 \cap \Lambda_2 \cap \Lambda_3}$ , the horizontal gray regions represent  $S_{\Lambda_1 \cap \Lambda_3}$ , and vertical gray regions represent  $S_{\Lambda_2 \cap \Lambda_3}$ , where  $S_\Lambda$  denotes the characteristic function of the domain of high-resolution volume, i.e.,

$$S_\Lambda(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \Lambda \\ 0, & \mathbf{x} \notin \Lambda. \end{cases} \quad (1)$$

The imaging model that incorporates the overlap regions (see Fig. 3) is then given by

$$\mathbf{g}_k(\mathbf{x}) = (\mathcal{D}_k \mathcal{H}_k \mathcal{W}_k \mathbf{f})(\mathbf{x}) + \mathbf{n}_k(\mathbf{x}), \quad \mathbf{x} \in \Lambda_k \quad (2)$$

where  $k$  denotes the  $k$ th observation,  $\mathbf{g}_k$  is one of the observed image volumes,  $\mathcal{W}_k$  is a coordinate transformation,  $\mathcal{D}_k$  is a downsampling operator,  $\mathcal{H}_k$  is an impulse response function (i.e., PSF) that blurs  $\mathbf{f}$ , and  $\mathbf{n}_k$  is a Gaussian noise with zero mean and variance  $\sigma_k^2$ . This model is combined as  $H_k = \mathcal{D}_k \mathcal{H}_k \mathcal{W}_k$ .

As is conventionally, we model the PSF as a Gaussian function [45]. The goal of this work is to reconstruct a single high-resolution volume  $\mathbf{f}$  from three orthogonal scans that approximates the original in-plane resolution in all three directions.

### B. Preprocessing

The three volumes to be combined into one superresolution volume have different slice positions, orientations, and volume sizes. Therefore, several preprocessing steps are applied prior to the reconstruction of the superresolution volume: 1) generation of an isotropic volume; 2) conversion of the orientation of each volume to the orientation of the target reference volume; 3) padding of zero values to yield the same volume sizes; 4) registration to correct subject motion between volumes; and 5) matching of intensities in the overlap region using spline regression. These steps are now summarized.

1) *Isotropic Volume Generation*: According to our imaging model, each low-resolution volume has its resolution degraded in only one dimension (slice-selection direction). In order to up-sample each volume in the slice-selection direction, a fifth-order B-spline interpolation is used.

2) *Orientation Conversion and Zero Padding*: The orientation of each volume is converted to that of the target reference volume. This step is required so that registration starts with the anatomy appearing in each volume roughly aligned. Without this step, the registration is likely to fail due to small overlap regions. In addition, although the previous steps produce an isotropic volume with the same orientation, the size of each volume is still different. For convenience in registration, each volume is padded with zero values in order to create digital volumes all having the same voxel dimensions. Without this step, the registration produces the same volume size as the target volume size (e.g.,  $256 \times 256 \times 20$  in the case of axial stack). Masks indicating the region of valid data—i.e.,  $\Lambda_1$ ,  $\Lambda_2$ , and  $\Lambda_3$ —are also created (see Fig. 4) for use in both registration and region-based reconstruction.

3) *Registration*: We use image registration to correct for subject motion between acquisitions. Accurate registration is of great importance in this application because small perturbations in alignment can lead to visible artifacts after applying the MAP-MRF reconstruction algorithm. This is because uncertainty in registration increases the variance of intensity values at each spatial location. To obtain accurate and robust registration results, we compute a global displacement estimate followed by a local deformation estimate. The global displacement estimate is characterized by an affine registration accounting for translation, rotation, and scaling (12 degrees of freedom in 3-D). We use mutual information (MI) [46] as the similarity measure for this step because orientation of acquisition can cause intensity differences even when the same pulse sequence on the same scanner is used. The algorithm we use is different from a conventional MI registration algorithm because of the FOV differences in the volumes.

We denote the images  $I_1 : \Omega_1 \subset \mathbb{R}^3 \rightarrow \mathbb{R}$  and  $I_2 : \Omega_2 \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ , defined on the open and bounded domains  $\Omega_1$  and  $\Omega_2$  as the template and target images, respectively. Mutual in-

formation  $\mathcal{M}$  can be computed using joint entropy  $\mathcal{H}$  as

$$\begin{aligned} \mathcal{H}(I_1(u_1(\mathbf{x})), I_2(\mathbf{x})) &= - \iint p(i_1, i_2) \log p(i_1, i_2) di_1 di_2 \\ \mathcal{M}(\mathbf{x}) &= \mathcal{H}(I_1(u_1(\mathbf{x}))) + \mathcal{H}(I_2(\mathbf{x})) \\ &\quad - \mathcal{H}(I_1(u_1(\mathbf{x})), I_2(\mathbf{x})) \\ &= \iint p_{u_1}(i_1, i_2) \log \frac{p_{u_1}(i_1, i_2)}{p_{I_1}(i_1)p_{I_2}(i_2)} di_1 di_2 \end{aligned} \quad (3)$$

where  $i_1 = I_1(u_1(\mathbf{x}))$ ,  $i_2 = I_2(\mathbf{x})$ , and  $p_{I_1}(i_1)$  and  $p_{I_2}(i_2)$  are marginal distributions. Here,  $p_{u_1}(i_1, i_2)$  denotes the joint distribution of  $I_1(u_1(\mathbf{x}))$  and  $I_2(\mathbf{x})$  in the overlap region  $V = u_1^{-1}(\Omega_1) \cap \Omega_2$  which can be computed using the Parzen window given by

$$p_{u_1}(i_1, i_2) = \frac{1}{|V|} \int_V \varphi\left(\frac{i_1 - I_1(u_1(\mathbf{x}))}{\rho}\right) \varphi\left(\frac{i_2 - I_2(\mathbf{x})}{\rho}\right) d\mathbf{x} \quad (4)$$

where  $\varphi$  is a Gaussian kernel whose width is controlled by  $\rho$ .

In this global registration step, we use the overlap regions of the two volumes to weight the evaluation of MI in order to emphasize influence on parameter estimation to regions that are known to be overlapping. Of note, two of the volumes were registered to the target reference volume. The optimal values for the 12 parameters that maximize the given energy functional are obtained by solving the associated Euler–Lagrange equations, and a gradient descent method [47] is performed. In addition, we use a coarse-to-fine multiresolution scheme for computational efficiency and robustness.

In order to achieve subvoxel accuracy, we estimate a local deformation using a modification of the diffeomorphic demons method [48]. Since diffeomorphic demons is an intensity-based registration method, we perform histogram matching between the two images to ensure that corresponding points in different volumes have similar intensity values for the local deformation model. The only modification that we make to diffeomorphic demons is to incorporate the overlap region in order to restrict the domain of registration (similar in concept to the global displacement estimation step).

4) *Intensity Matching*: Now that the volumes are registered, a more careful analysis of the intensity differences that exist between the images can be carried out via a spline-based intensity regression method that uses local intensity matching. We first revert to the original intensities (instead of the histogram-matched intensities) and recognize that if the registration is accurate then the overlap regions provide evidence of how the intensities of the different acquisitions correspond. In particular, we can observe intensity pairs at each voxel within the corresponding overlap regions and compute a regression that represents an estimate of the transformation required to make the intensities of one volume match those of another. In this step, intensity values of two volumes are matched to those of the target reference volume. We use 10–15 control points in a cubic spline fit to compute two such transformations that are then used to map the intensities of two volumes into the target volume.

### C. MAP-MRF Reconstruction

Once we have three preprocessed, orthogonal volumes, regularized reconstruction is carried out. Classical methods including Tikhonov regularization [49], Gaussian and Gibbs regularizations [50], and Laplacian regularization [19] have been used to achieve various image reconstruction objectives. But these methods often create oversmoothing of edges [20] in direct opposition to our superresolution goals. Edge-preserving regularization methods, including total variation (TV) [51], [52], bilateral TV [20], and half-quadratic regularization [53], have been developed to solve this problem.

In this study, we use the edge-preserving regularization method of Villain *et al.* [54] wherein MAP estimation with a half-quadratic approach is used to obtain a high-resolution volume  $\mathbf{f}$  given three orthogonal low-resolution images, i.e.,

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} P(\mathbf{f} | \mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3) \quad (5)$$

where  $\hat{\mathbf{f}}$  denotes the estimated solution. Applying the Bayes rule, this can be written as

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f}) P(\mathbf{f}) \quad (6)$$

where  $P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f})$  and  $P(\mathbf{f})$  represent likelihood and prior, respectively.

1) *Likelihood Model:* We assume that each observed image is corrupted by additive Gaussian noise, which is a good approximation since the MR images in the regions of interest (i.e., tongue region) have a high SNR. Accordingly, the likelihood is determined by the image formation process of (2) with additive noise, yielding

$$P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f}) = Q \exp(-D(\mathbf{f})) \quad (7)$$

where  $Q$  denotes a normalization factor and  $D(\mathbf{f})$  using the region-based approach [see Fig. 4(d)] is given by (8) as shown at the bottom of the page. Here, because the precise positions of these acquired volumes relative to the anatomy might be slightly wrong even after motion correction and to reduce the effects of “seams” between different regions, we use blurred versions of the masks—termed “softmasks”—which are given by

$$\mathcal{F}_k(\mathbf{x}) = G_\sigma(\mathbf{x}) * S_{\Lambda_k}(\mathbf{x}), \quad k = 1, 2, 3 \quad (9)$$

where  $*$  denotes the convolution operator and  $G_\sigma$  is a unit-height Gaussian kernel with standard deviation  $\sigma$ . In this study, we set  $\sigma = 2$  mm. The softmasks provide a probabilistic confidence about the intensity value of each scan, thereby allowing a soft transition between scans.

---

#### Algorithm 1 MAP-MRF Reconstruction

---

- 1: Initialization
  - 2: **for** each pixel  $s$  on the volume **do**
  - 3:   Determine overlapping volumes using regions defined in Fig. 4.
  - 4:   Calculate  $\mathbf{m}_s$  according to Eq. (18)
  - 5:   Update  $\mathbf{f}_s$  according to the over-relaxation scheme in Eq. (21)
  - 6:   For every clique  $c$  including  $s$ , update  $l$  according to Eq. (20)
  - 7: **end for**
  - 8: Iterate for loop until convergence
- 

2) *Prior Model:* The MRF is defined over a graph  $G = \langle V, E \rangle$  where  $V$  is the set of nodes representing the random variables  $s = \{s_v\}_{v \in V}$ ,  $E$  is the set of edges connecting the nodes, and the clique  $c$  is defined by the neighborhood system. The clique of MRF is defined by the local neighborhoods around pixels. Our prior model can be defined as

$$P(\mathbf{f}) = \frac{1}{W} \exp \left\{ - \sum_{c \in C} \varphi(u_c) \right\}. \quad (10)$$

Here,  $W$  is a normalization factor and  $\varphi(u) = \sqrt{1 + (u/\delta)^2}$ ,  $u_c = \Delta x_c / d_c$ , where  $\delta$  denotes a scaling factor and  $\Delta x_c$  and  $d_c$  denote the difference of the values and the distance of the two voxels in clique  $c$ , respectively (we set  $\delta = 0.8$  in our implementation).

Maximization of (6) is equivalent to minimizing the negative of the logarithm of the posterior probability, i.e.,

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} [-\log P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f}) - \log P(\mathbf{f})] \quad (11)$$

which leads to the following form:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} J(\mathbf{f}) = \arg \min_{\mathbf{f}} \left\{ D(\mathbf{f}) + \lambda \sum_{c \in C} \varphi(u_c) \right\} \quad (12)$$

where  $\lambda \in \mathbb{R}^+$  is a balancing parameter, controlling the smoothness of the reconstructed image (we set  $\lambda = 0.5$  in our implementation).

3) *Half-Quadratic Approach:* In order to efficiently minimize the objective function in (12), we adopt a half-quadratic regularization technique incorporating the region-based approach as in [54] and [55].

The summary of the MAP-MRF reconstruction algorithm is shown above.

---


$$D(\mathbf{f}) = \begin{cases} \sum_{k=1}^3 \mathcal{F}_1(\mathbf{x}) \mathcal{F}_2(\mathbf{x}) \mathcal{F}_3(\mathbf{x}) \frac{\|\mathbf{g}_k(\mathbf{x}) - H_k \mathbf{f}(\mathbf{x})\|^2}{2\sigma_k^2}, & \text{if } \mathbf{x} \in \Lambda_1 \cap \Lambda_2 \cap \Lambda_3 \\ \sum_{k \in \{i,j\}} \mathcal{F}_i(\mathbf{x}) \mathcal{F}_j(\mathbf{x}) \frac{\|\mathbf{g}_k(\mathbf{x}) - H_k \mathbf{f}(\mathbf{x})\|^2}{2\sigma_k^2}, & \text{if } \mathbf{x} \in \Lambda_i \cap \Lambda_j \\ \mathcal{F}_k(\mathbf{x}) \frac{\|\mathbf{g}_k(\mathbf{x}) - H_k \mathbf{f}(\mathbf{x})\|^2}{2\sigma_k^2}, & \text{if } \mathbf{x} \in \Lambda_k. \end{cases} \quad (8)$$

#### D. Implementation

The preprocessing steps including resampling, reorientation, global displacement (affine registration), local deformation model (diffeomorphic demons registration), and intensity matching were implemented using C++ and Insight Segmentation and Registration Toolkit (ITK) [56] and took  $18 \pm 2$  min. The MAP-MRF reconstruction step was also implemented using C++ and took about  $5 \pm 1$  min. The proposed method was performed on an Intel i7 CPU with a clock speed of 1.74 GHz and 8-GB memory. In addition, the ALGLIB library [57] was used to implement spline-based regression part. We used a multiresolution scheme for our global displacement model. The parameters are first estimated at the coarsest level, which are then used as the starting parameters for the subsequent level. Transformations at each level of the pyramid are accumulated and propagated. This scheme helps avoid local minima caused by partial overlap. In our implementation, we use two levels. Parameters used in the experiments were determined empirically to produce the best visual results. Once the parameters were determined, the same parameters were used in all experiments including tongue and brain datasets.

### IV. VALIDATION

#### A. Evaluation Using Tongue Data

1) *Tongue Data*: Fifteen high-resolution MR datasets were used in these experiments. They came from twelve normal speakers and three patients who had tongue cancer surgically resected (glossectomy). All MRI scanning was performed on a Siemens 3.0 T Tim Treo system using an eight-channel head and neck coil. In addition, T2-weighted Turbo Spin Echo sequence with echo train length of 12 and TE/TR of 62 ms/2500 ms was used. The FOV was 240 mm  $\times$  240 mm with a resolution of 256  $\times$  256. Each dataset contained a sagittal, coronal, and axial stack of images containing the tongue and surrounding structures. The image size for high-resolution MRI was 256  $\times$  256  $\times$   $z$  ( $z$  ranges from 10 to 24) with 0.94 mm  $\times$  0.94 mm in-plane resolution and 3 mm slice thickness. The datasets were acquired at rest position and the subjects were required to remain still from 1.5 to 3 min for each orientation.

2) *Quantitative Analysis*: The proposed method was first evaluated using 15 simulated datasets as described in Section IV-A. In order to quantitatively evaluate the performance of the proposed method in terms of accuracy of the reconstruction, the final superresolution volume using the proposed method was considered as our ground truth data. The ground truth was constructed with 256  $\times$  256  $\times$  256 voxels with a resolution of 0.94 mm  $\times$  0.94 mm  $\times$  0.94 mm. Three low-resolution volumes (i.e., axial, sagittal, and coronal volumes) were formed where the voxel dimensions was 0.94 mm  $\times$  0.94 mm in-plane and the slice thickness was 3.76 mm (subsampling by a factor 4). Prior to subsampling the volumes in slice-selection directions, Gaussian filtering with  $\sigma = 0.5$  (in-plane) and  $\sigma = 2$  (slice-selection direction) was applied in order to avoid an antialiasing effect.

In what follows, volume reconstruction was carried out in four ways. First, fifth-order B-spline interpolation was performed in each plane independently. Second, averaging of three up-sampled volumes was performed. Third, reconstruction from three up-sampled volumes using Tikhonov regularization [58] was performed. Finally, the proposed method using three up-sampled volumes was carried out. In this simulation study, preprocessing steps were not necessary except the up-sampling process using fifth-order B-spline interpolation as the volumes are already aligned and intensity values between volumes are identical. Tikhonov regularization is one of the most common approaches using quadratic functions, which takes the form

$$\hat{\mathbf{f}}_T = \arg \min_{\mathbf{f}} \left[ \sum_{k=1}^3 \frac{\|g_k - H_k \mathbf{f}\|_2^2}{\sigma_k^2} + \lambda_1 \|L\mathbf{f}\|_2^2 \right] \quad (13)$$

where  $\hat{\mathbf{f}}_T$  represents the reconstructed volume,  $\lambda_1$  denotes the balancing parameter, and we set  $L = I$  in this experiment.

As a quantitative measure, the peak signal-to-noise ratio (PSNR) is defined as

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}_X^2}{|\Omega_R|^{-1} \sum_{v \in \Omega_R} (X(v) - \hat{X}(v))^2} \right) \quad (14)$$

where  $\text{MAX}_X$  denotes the maximum possible voxel value of the volume,  $\Omega_R$  denotes a reference image domain,  $X$  denotes a reference volume, and  $\hat{X}$  represents a reconstructed volume.

3) *Noise Simulations*: To test robustness, Gaussian noise with zero mean and standard deviations of 50 (2% of intensity level) and 75 (3% of intensity level), respectively, were added to investigate the effect of noise on the reconstruction. The same quantitative analysis as described in Section IV-A2 was performed to evaluate the performance of different reconstruction methods.

#### B. Evaluation Using Brain Data

1) *Brain Data*: Three isotropic high-resolution brain MR datasets were used to objectively evaluate and compare the performance of different reconstruction methods. For each subject, two magnetization prepared rapid gradient echo (MP-RAGE) images were obtained using a 3.0T MR scanner (Intera, Phillips Medical Systems, The Netherlands). The MP-RAGE was acquired with the following parameters: 132 slices, axial orientation, 0.8 mm slice thickness,  $8^\circ$  flip angle, TE = 3.9 ms, TR = 8.43 ms, 256  $\times$  256 matrix, and 212  $\times$  212 cm FOV. The image size for these datasets was 288  $\times$  288  $\times$  225 with isotropic resolution (i.e., 0.8 mm  $\times$  0.8 mm  $\times$  0.8 mm).

2) *Quantitative Analysis*: A challenge in testing the proposed method is the lack of ground truth available in *in vivo* volumetric tongue MR data. In order to avoid potential bias in our experiments, we used two high-resolution brain datasets with isotropic resolution to evaluate the performance of different reconstruction methods. Three low-resolution volumes (i.e., axial, sagittal, and coronal volumes) were formed where the voxel dimensions were 0.5 mm  $\times$  0.5 mm in-plane and the slice thickness was 2.0 mm (subsampling by a factor 4). Prior to subsampling the volumes in slice-selection directions,

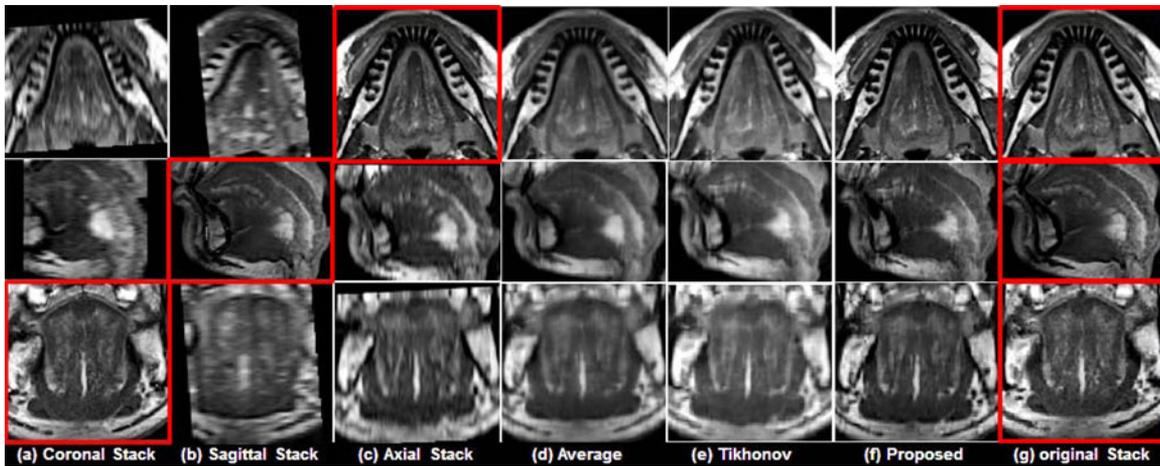


Fig. 5. Comparison of different reconstruction methods on a normal subject. The original coronal, sagittal, and axial volumes are shown in (a), (b), and (c), respectively. B-spline interpolation was used to yield higher pixel resolutions equal to that of the highest resolution orientations, which are indicated by the red boxes. Three different reconstruction methods including simple averaging, Tikhonov regularization, and the proposed method are shown in (d), (e), and (f), respectively. For reference, the highest resolution images from each original volume are repeated in (g).

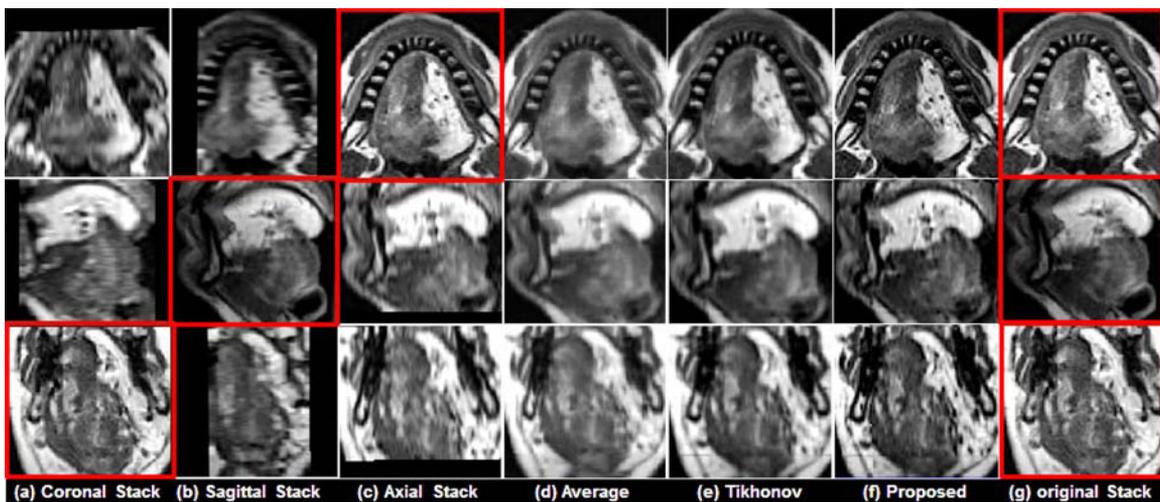


Fig. 6. Comparison of different reconstruction methods using a glossectomy patient. The original coronal, sagittal, and axial volumes are shown in (a), (b), and (c), respectively. B-spline interpolation was used to yield higher pixel resolutions equal to that of the highest resolution orientations, which are indicated by the red boxes. Three different reconstruction methods include (d) simple averaging, (e) Tikhonov regularization, and (f) the proposed method, respectively. For reference, the highest resolution images from each original volume are repeated in (g).

Gaussian filtering with  $\sigma = 0.5$  (in-plane) and  $\sigma = 2$  (slice-selection direction) was applied in order to avoid the antialiasing effect. The same quantitative analysis as described in Section IV-A2 was performed to evaluate the performance of different reconstruction methods.

## V. RESULTS

In Figs. 5 and 6, two representative results using a normal subject (see Fig. 5) and a glossectomy patient (see Fig. 6) with different reconstruction methods are demonstrated, respectively. The rows show slices of three orthogonal views including axial, sagittal, and coronal, respectively. The first three columns

(a)–(c) show three original scans after isotropic resampling (i.e.,  $0.94 \text{ mm} \times 0.94 \text{ mm} \times 0.94 \text{ mm}$ ) in the coronal, sagittal, and axial planes, respectively. Panel (d) shows the reconstruction using averaging, (e) shows reconstruction using Tikhonov regularization, and (f) shows the reconstruction using the proposed method. Panel (g) shows the original three orthogonal volumes. As shown in Figs. 5 and 6, the proposed method provided better muscle and fine anatomical detail than the other methods. The target reference volume into which the other volumes were registered was the axial volume in both cases. In our experiments, preprocessing steps (e.g., registration) were not validated in a quantitative manner. Instead, the quality of the results was confirmed visually. Notice that the reconstructed images and the

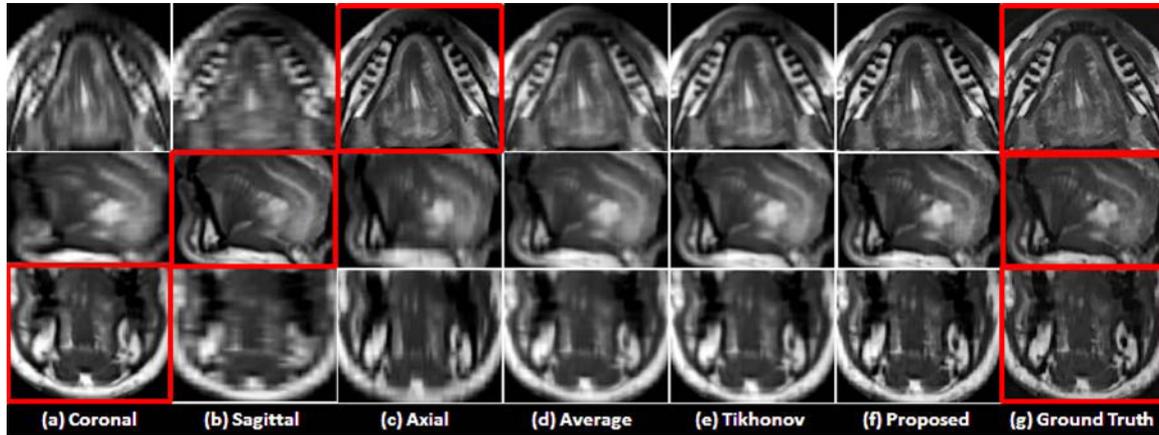


Fig. 7. Comparison of different reconstruction methods using a simulation study. Reconstructed volume are shown in (a), (b), and (c), respectively. B-spline interpolation was used to yield higher pixel resolutions equal to that of the highest resolution orientations, which are indicated by the red boxes. Three different reconstruction methods including simple averaging, Tikhonov regularization, and the proposed method are depicted in (d), (e), and (f), respectively. Ground truth data is illustrated in (g). The proposed method provides more detailed anatomical information than averaging and Tikhonov regularization methods as visually assessed.

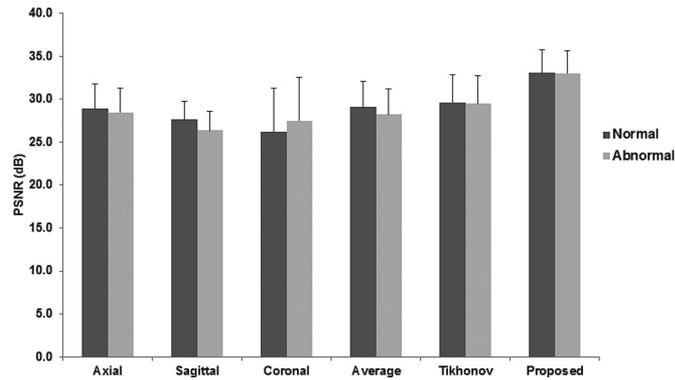


Fig. 8. PSNR was measured using 15 human datasets consisting of 12 normal and 3 abnormal subjects to compare the performance of the reconstruction methods quantitatively.

original images shown in the figures may not be exactly the same except in the axial slice, because the data are aligned to the axial stack.

#### A. Quantitative Analysis Using Tongue Data

Fig. 7 illustrates results using the simulated data. Fig. 7(a)–(c) shows the reconstructed results using the B-spline interpolation method [59] in which resolution was degraded in only one direction. Fig. 7(d)–(f) is reconstructed from three volumes [i.e., Fig. 7(a)–(c)] using averaging, Tikhonov regularization, and the proposed method, respectively, in which the proposed method provided a more visually detailed result than the other methods. Fig. 8 illustrates the PSNR results using the normal and patient datasets, respectively, in which the proposed method outperformed the other methods. Patients have more complex anatomy and scar tissue than controls due to forearm flap reconstruction. Therefore, it is more challenging to reconstruct but the results show that the proposed method was able to produce equally good results.

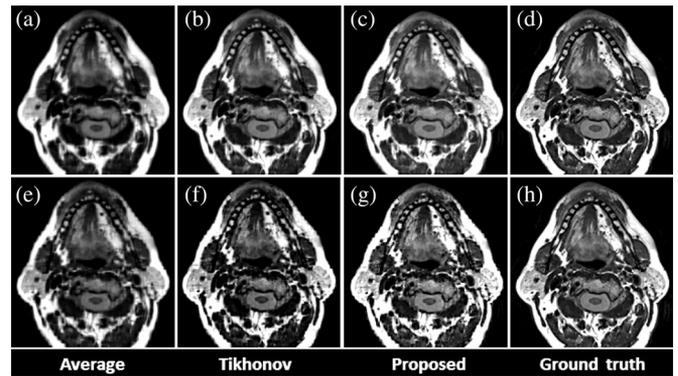


Fig. 9. Comparison of effect of Gaussian filtering that removes the antialiasing effect using a simulation study. The first row shows reconstructed volumes generated using Gaussian filtering that removes antialiasing effects including (a) averaging (PSNR = 24.36 dB), (b) Tikhonov regularization (PSNR = 27.74 dB), (c) the proposed method (PSNR = 28.74 dB), and (d) ground truth, respectively. The second row shows reconstructed volumes generated without using Gaussian filtering that removes antialiasing effects including (e) averaging (PSNR = 25.71 dB), (f) Tikhonov regularization (PSNR = 27.42 dB), (g) the proposed method (PSNR = 26.87 dB), and (h) ground truth, respectively. (d) and (h) are the same.

In addition, to compare the impact of antialiasing [60], volumes were reconstructed both with and without the use of Gaussian filtering. Results are shown in Fig. 9. It is observed that volumes reconstructed without using Gaussian filtering have step-like artifacts (see the second row). The averaging result in Fig. 9(e) without using Gaussian filtering had better PSNR as compared with the one using Gaussian filtering in Fig. 9(a) as the averaging is similar to the low-pass filtering, removing the step-like artifacts. The reconstructed volume using Tikhonov regularization [see Fig. 9(b)] provided almost similar results. However, the reconstructed volumes using the proposed method use edge-preserving property, and therefore, the aliasing artifacts are more prominent in Fig. 9(g) than in Fig. 9(c). The reconstructed volume [see Fig. 9(c)] using Gaussian

TABLE I  
PSNR (dB) FOR DIFFERENT RECONSTRUCTION METHODS ON SIMULATION DATA (MEAN  $\pm$  SD)

	Axial	Sagittal	Coronal	Averaging	Tikhonov regularization	Proposed method
No noise	28.8 $\pm$ 3.3	27.4 $\pm$ 3.2	26.5 $\pm$ 4.4	28.9 $\pm$ 3.6	29.3 $\pm$ 3.6	<b>33.1<math>\pm</math>3.1</b>
Gaussian Noise (m=0, $\sigma$ =50)	27.9 $\pm$ 2.8	26.7 $\pm$ 2.8	25.6 $\pm$ 3.8	28.2 $\pm$ 3.0	28.7 $\pm$ 3.6	<b>31.3<math>\pm</math>2.1</b>
Gaussian Noise (m=0, $\sigma$ =75)	26.8 $\pm$ 2.6	26.1 $\pm$ 2.7	25.1 $\pm$ 3.4	26.9 $\pm$ 2.6	28.5 $\pm$ 3.2	<b>29.7<math>\pm</math>3.7</b>

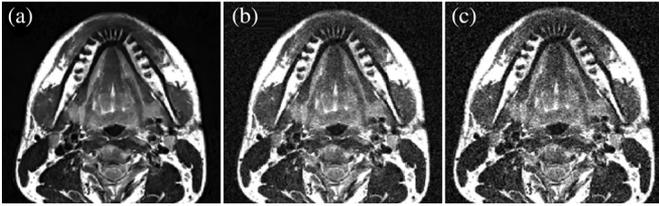


Fig. 10. To simulate the performance of the reconstruction in the presence of the noise, Gaussian noise images with two power levels were added to ground truth in (a): Gaussian noise with zero mean and standard deviations of (b) 50 and (c) 75, respectively.

filtering provided the best result as assessed both visually and quantitatively.

### B. Noise Simulations Using Tongue Data

Gaussian noise images with two power levels were added as depicted in Fig. 10. Table I summarizes the quantitative results after the reconstruction, showing that the proposed method is better than the other reconstruction methods. However, increasing the noise in our simulation degraded the proposed method (lower PSNR) while both averaging and Tikhonov regularization remained largely unaffected. This can be explained by the edge-preserving property of our method, which has a tendency to also preserve noise when it is present, whereas both averaging and Tikhonov regularization tend to suppress noise.

### C. Quantitative Analysis Using Brain Data

Fig. 11 illustrates reconstruction results using the brain data. Fig. 11(a)–(c) shows the reconstructed results using the B-spline interpolation method in which resolution was degraded in only one direction. Fig. 11(d)–(f) is reconstructed from three volumes [i.e., Fig. 11(a)–(c)] using averaging, Tikhonov regularization, and the proposed method, respectively. Fig. 11(g) shows the high-resolution ground truth data. Fig. 12 illustrates the quantitative measure (i.e., PSNR) of three datasets, where the proposed method provided better performance, confirming the proposed method is not biased with respect to the use of different datasets.

Fig. 13 shows the importance of using both registration and subsequent intensity matching steps. Fig. 13(a) and (b) illustrates that misalignments and difference in intensity distributions between scans were significant, respectively. The proposed approach corrects this problem as shown in Fig. 13(c). In addition, Fig. 14 illustrates that the reconstruction results with and without using a region-based reconstruction approach. Fig. 14(a) shows intensity mismatch compared to Fig. 14(b).

## VI. DISCUSSION

In our application, it was of great importance to preserve crisp muscle details and fine structures. Therefore, we treated them as edges and adopted an edge-preserving regularization as in [54]. The half-quadratic approach was carried out to minimize the objective function in the tongue MR images. In addition, we compared the results with a regularization scheme which does not use the edge-preserving property. Other edge-preserving regularization methods [61]–[64] are likely to give similar results.

In our experimental results, patients have more complex anatomy than controls. They have considerable scar tissue, their muscles are deformed around the scar and missing in the area of the resection, and they are asymmetrical. Therefore, it was possible that their datasets would be more difficult to align. However, it is seen from our experimental results that all were reconstructed with equally good success.

Noise simulations were also carried out to confirm the robustness of the proposed algorithm. Since the proposed algorithm tends to preserve noise due to the use of edge-preserving regularization, we tested it with a large amount of noise using Gaussian noise with zero mean and standard deviations of 50 and 75, respectively, the performance of the proposed method was better than other methods in our experiments.

Evaluation of the proposed study was a challenging task. In fact, the notion of accuracy is ill-posed in this application as precise validation is typically impossible due to the lack of ground truth available other than simulation and visual assessment. We considered the final reconstruction volume as our ground truth and simulated the reconstruction step. Moreover, we used the brain data to further evaluate the algorithm and avoid bias in our evaluation. The brain data allowed us to more definitively determine the quality of the algorithm because of its isotropic resolution. This leads to more confidence in the interpretation of the tongue data and is very important because the tongue itself is highly anisotropic. The tip is quite thin and it gets thicker toward the back. This is best seen in the sagittal orientation. On the other hand, only the axial plane allows us to see the lateral shape of the tip. In this paper, the axial orientation was chosen as the target. However, in a study where tongue tip thickness is the critical feature, the sagittal orientation might be chosen as the target.

The intensity matching results, while good, were not optimal because the regression was performed only on overlapped regions, and therefore, the full range of the tongue intensity may not have been covered. In our future work, we will make use of all three datasets to perform high-dimensional regression and devise a weighting scheme to incorporate any difference in

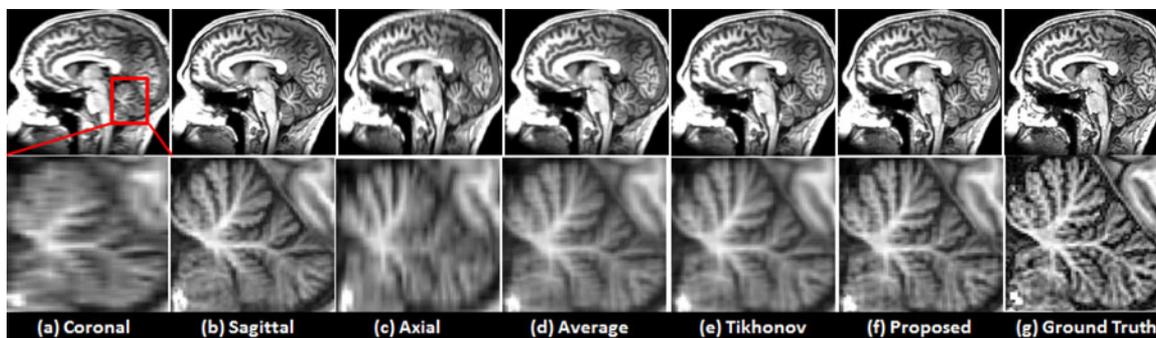


Fig. 11. Brain datasets were used to gauge the performance of different reconstruction methods. The images in the second row illustrate the enlarged region shown in the red box in the first column.

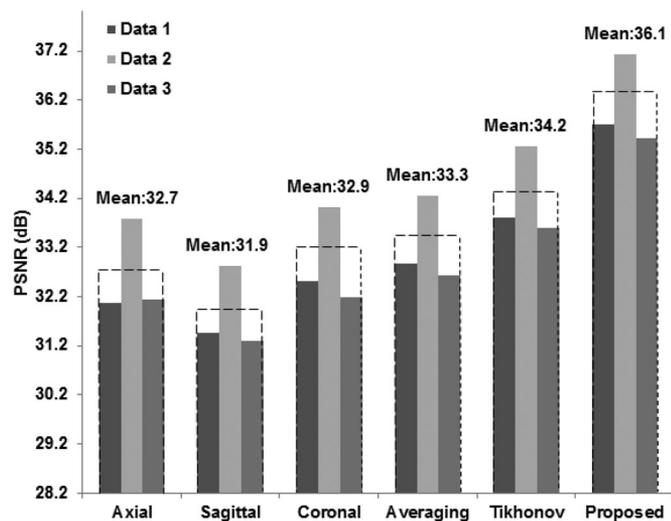


Fig. 12. PSNR (dB) for different reconstruction methods on brain data. Dotted lines represent the mean of the three datasets.

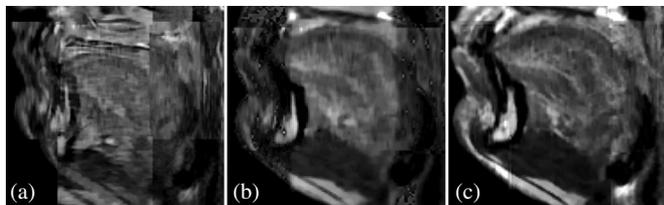


Fig. 13. One example showing the superresolution volume reconstruction (a) without using registration and intensity matching, (b) without using intensity matching after registration, and (c) the proposed method.

resolutions between the stacks. In addition, the performance of the reconstruction algorithm depends on the choice of the balancing parameter, which is a matter of modeling. In our future work, we will investigate the mechanism to find the best parameter in terms of the reconstruction performance.

The reconstruction quality heavily depends on the registration performance. In our study, we know the overlapped regions (i.e., tongue region) and explicitly used the region in our registration, which is an optimal way to deal with partial overlap. In addition, although affine registration may capture the differences of different stacks acquired at rest position, we further incorporated the diffeomorphic demons registration to accommodate subtle

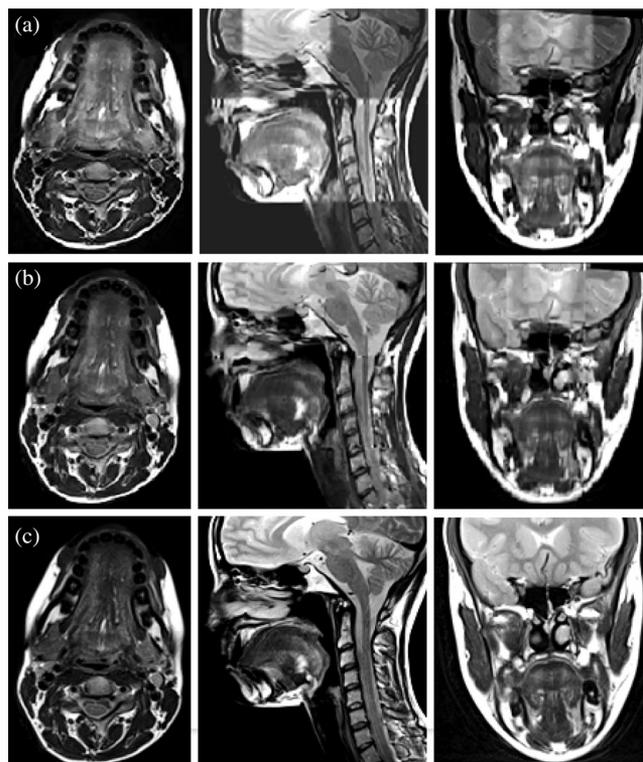


Fig. 14. Results of superresolution reconstruction with and without using the region-based approach. (a) MAP-MRF reconstruction without using the region-based approach, (b) the proposed method, and (c) original three volumes. Axial, sagittal, and coronal volumes are shown in the first, second, and third columns, respectively.

deformation in order to maximize the reconstruction quality. Diffeomorphic demons depends on accurate image intensities; however, these are only fully corrected after registration. Furthermore, intensities should only truly match if there were no differences in the direction of blurring between the images, which is obviously untrue.

In the clinical scenario, MRI has been used for structural imaging. The proposed superresolution volume reconstruction provides the groundwork for the full 3-D volumetric analyses such as image segmentation, registration, and atlas building. In addition, it overcomes the limitations of slice thickness problems. Therefore, this automated postprocessing technique can

$$\mathbf{m}_s = \begin{cases} \mathbf{f}_s + \frac{\sum_{k=1}^3 \mathcal{F}_1(\mathbf{x})\mathcal{F}_2(\mathbf{x})\mathcal{F}_3(\mathbf{x})([H_k^t g_k]_s - [H_k^t H_k \mathbf{f}]_s) + 2\sigma_n^2 \sum_{\{r,s\} \in C_2} \lambda l_{\{r,s\}} (\mathbf{f}_r - \mathbf{f}_s)}{[H_k^t H_k]_{s,s} + 2\sigma_n^2 \sum_{\{r,s\} \in C_2} \lambda l_{\{r,s\}}} \\ \mathbf{f}_s + \frac{\sum_{k \in \{i,j\}} \mathcal{F}_i(\mathbf{x})\mathcal{F}_j(\mathbf{x})([H_k^t g_k]_s - [H_k^t H_k \mathbf{f}]_s) + 2\sigma_n^2 \sum_{\{r,s\} \in C_2} \lambda l_{\{r,s\}} (\mathbf{f}_r - \mathbf{f}_s)}{[H_k^t H_k]_{s,s} + 2\sigma_n^2 \sum_{\{r,s\} \in C_2} \lambda l_{\{r,s\}}} \\ \mathbf{f}_s + \frac{\mathcal{F}_k(\mathbf{x})([H_k^t g_k]_s - [H_k^t H_k \mathbf{f}]_s) + 2\sigma_n^2 \sum_{\{r,s\} \in C_2} \lambda l_{\{r,s\}} (\mathbf{f}_r - \mathbf{f}_s)}{[H_k^t H_k]_{s,s} + 2\sigma_n^2 \sum_{\{r,s\} \in C_2} \lambda l_{\{r,s\}}} \end{cases} \quad (19)$$

be potentially used in any body part to improve 3-D resolution and decrease the noise in the images. In addition, because poorer through plane quality is acceptable, it could be used for patients that are unable to hold still for a long period of time, but require good image resolution for diagnoses and treatment.

To our knowledge, there have been no prior attempts to perform superresolution volume reconstruction in tongue MR imaging domain. We successfully implemented and evaluated interpolation, averaging, and MAP-MRF reconstruction using the Tikhonov regularization and compared with the proposed method. The proposed method provided the best performance and groundwork for further analyses.

## VII. CONCLUSION

In this study, a superresolution reconstruction technique based on the region-based MAP-MRF with an edge-preserving regularization and half-quadratic approach was developed for volumetric tongue MR images acquired from three orthogonal acquisitions. Experimental results show that the proposed method has superior performance to the interpolation-based method, simple averaging, and Tikhonov regularization. It also better preserved the anatomical details as quantitatively confirmed. The proposed method allows full 3-D high-resolution volumetric data, thereby potentially improving further image/motion and visual analyses.

## APPENDIX

### HALF-QUADRATIC APPROACH

The main idea in half-quadratic approach is to introduce and minimize a new objective function, which has the same minimum as the original non-quadratic objective function  $J(\mathbf{f})$ . By introducing an auxiliary variable  $l$ , the original objective function  $J(\mathbf{f})$  can be replaced by an augmented objective function  $K(\mathbf{f}, l)$ . We first define an auxiliary function  $\psi$  as follows:

$$\psi(l) = -\min_x (lx^2 - \varphi(x)). \quad (15)$$

This leads to the new prior model given by

$$\varphi(u) = \min_l (lu^2 + \psi(l)). \quad (16)$$

The augmented objective function can then be given by

$$K(\mathbf{f}, l) = (\mathbf{f}) + \lambda \sum_{c \in C_2} (lu_c^2 + \psi(l)) \quad (17)$$

which is a quadratic function whose minimization with respect to  $\mathbf{f}$  and  $l$  yields explicit and closed-form expressions,

respectively. Minimizing  $K(\mathbf{f}, l)$  with respect to  $\mathbf{f}$  leads to the following form:

$$\left. \frac{\partial K(\mathbf{f}, l)}{\partial \mathbf{f}_s} \right|_{\mathbf{f}_s = \mathbf{m}_s} = 0 \quad (18)$$

where its minimum value is achieved at  $\mathbf{m}_s$  [65]. Solving for  $\mathbf{m}_s$  yields an explicit expression of each condition in (8), i.e., (19) as shown at the top of the page, where  $\sigma_n$  denotes standard deviation of the final superresolution volume and the sums extend to the neighborhood of the currently visited voxel  $s$ . In addition, to minimize  $K(\mathbf{f}, l)$  with respect to  $l$ , we take the derivative, set it equal to zero, and solve for  $l$ , yielding

$$\hat{l} = \frac{\varphi'(u_c)}{2u_c}. \quad (20)$$

We also use an overrelaxation scheme to speed-up convergence as follows:

$$\mathbf{f}_s^{\text{new}} \leftarrow \max\{0, \mathbf{f}_s + \alpha(\mathbf{m}_s - \mathbf{f}_s)\} \quad (21)$$

where we set  $\alpha = 1.2$  in our implementation. The overall minimization of  $K(\mathbf{f}, l)$  is achieved by alternating between minimization with respect to  $\mathbf{f}$  and then with respect to  $l$  using (21) and (20), respectively.

## ACKNOWLEDGMENT

The authors would like to thank S. Ying, Department of Radiology, Johns Hopkins University, for sharing brain data in their simulation study. They would also like to thank the reviewers for their helpful comments.

## REFERENCES

- [1] M. Stone, E. Davis, A. Douglas, M. Aiver, R. Gullapalli, W. Levine, and A. Lundberg, "Modeling tongue surface contours from cine-MRI images," *J. speech, Lang., Hear. Res.*, vol. 44, no. 5, pp. 1026–1040, 2001.
- [2] R. Wilhelms-Tricarico, "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," *J. Acoust. Soc. Amer.*, vol. 97, pp. 3085–3098, 1995.
- [3] S. Narayanan, D. Byrd, and A. Kaun, "Geometry, kinematics, and acoustics of tamil liquid consonants," *J. Acoust. Soc. Amer.*, vol. 106, pp. 1993–2007, 1999.
- [4] S. Narayanan, A. Alwan, and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging," *J. Acoust. Soc. Amer.*, vol. 98, no. 3, pp. 1325–1347, 1995.
- [5] E. Bresch, Y. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging," *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 123–132, May 2008.
- [6] A. Lakshminarayanan, S. Lee, and M. McCutcheon, "MR imaging of the vocal tract during vowel production," *J. Magn. Reson. Imaging*, vol. 1, no. 1, pp. 71–76, 1991.
- [7] M. Stone, X. Liu, H. Chen, and J. Prince, "A preliminary application of principal components and cluster analysis to internal tongue deforma-

- tion patterns," *Comput. Methods Biomech. Biomed. Eng.*, vol. 13, no. 4, pp. 493–503, 2010.
- [8] V. Napadow, Q. Chen, V. Wedeen, and R. Gilbert, "Intramural mechanics of the human tongue in association with physiological deformations," *J. Biomech.*, vol. 32, no. 1, pp. 1–12, 1999.
- [9] S. Takano and K. Honda, "An MRI analysis of the extrinsic tongue muscles during vowel production," *Speech Commun.*, vol. 49, no. 1, pp. 49–58, 2007.
- [10] J. Woo, Y. Bai, S. Roy, E. Murano, M. Stone, and J. Prince, "Super-resolution reconstruction for tongue MR images," *Proc. SPIE*, vol. 8314, pp. 83140C–83140C-8, 2012.
- [11] R. Reichard, M. Stone, J. Woo, E. Murano, and J. Prince, "Motion of apical and laminal /s/ in normal and post-glossectomy speakers," *Acoust. Soc. Amer.*, p. 3346, 2012.
- [12] R. Tsai and T. Huang, "Multiframe image restoration and registration," *Adv. Comput. Vis. Imag. Process.*, vol. 1, no. 2, pp. 317–339, 1984.
- [13] S. Borman and R. L. Stevenson, "Super-resolution from image sequences—A review," in *Proc. Midwest Symp. Syst. Circuits*, 1998, pp. 374–378.
- [14] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
- [15] H. Greenspan, "Super-resolution in medical imaging," *Comput. J.*, vol. 52, no. 1, pp. 43–63, 2009.
- [16] M. Unger, T. Pock, M. Werlberger, and H. Bischof, "A convex approach for variational super-resolution," in *Proceedings of the 32nd DAGM Conference on Pattern Recognition*. Berlin/Heidelberg, Germany: Springer-Verlag, 2010, pp. 313–322.
- [17] C. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
- [18] F. Rousseau, O. Glenn, B. Iordanova, C. Rodriguez-Carranza, D. Vigneron, J. Barkovich, and C. Studholme, "Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images," *Acad. Radiol.*, vol. 13, no. 9, pp. 1072–1081, 2006.
- [19] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 479–490, Feb. 2007.
- [20] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
- [21] K. Kim and Y. Kwon, "Example-based learning for single-image super-resolution," *Pattern Recognit.*, pp. 456–465, 2008.
- [22] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 349–356.
- [23] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [24] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.
- [25] T. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," *Pattern Recognit. Lett.*, vol. 30, no. 5, pp. 494–502, 2009.
- [26] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar/Apr. 2002.
- [27] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.
- [28] C. Liu, H. Shum, and W. Freeman, "Face hallucination: Theory and practice," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 115–134, 2007.
- [29] F. Rousseau, "Brain hallucination," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 497–508.
- [30] J. Sun, N.-N. Zheng, H. Tao, and H.-Y. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, Jun. 2003, vol. 2, pp. 729–736.
- [31] A. Gholipour, J. Estroff, and S. Warfield, "Robust super-resolution volume reconstruction from slice acquisitions: Application to fetal brain MRI," *IEEE Trans. Med. Imag.*, vol. 29, no. 10, pp. 1739–1758, Oct. 2010.
- [32] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.
- [33] N. Woods, N. Galatsanos, and A. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 201–213, Jan. 2006.
- [34] E. Plenge, D. Poot, M. Bernsen, G. Kotek, G. Houston, P. Wielopolski, L. van der Weerd, W. Niessen, and E. Meijering, "Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?" *Magn. Res. Med.*, 2012. DOI: 10.1002/mrm.24187.
- [35] A. Gholipour, J. Estroff, M. Sahin, S. Prabhu, and S. Warfield, "Maximum a posteriori estimation of isotropic high-resolution volumetric MRI from orthogonal thick-slice scans," in *Proc. Int. Conf. Med. Imag. Comput. Comput.-Assist. Interv.*, 2010, pp. 109–116.
- [36] D. Poot, V. Van Meir, and J. Sijbers, "General and efficient super-resolution method for multi-slice MRI," in *Proc. Int. Conf. Med. Imag. Comput. Comput.-Assist. Interv.*, 2010, pp. 615–622.
- [37] R. Shilling, T. Robbie, T. Bailloleul, K. Mewes, R. Mersereau, and M. Brummer, "A super-resolution framework for 3-d high-resolution and high-contrast imaging using 2-d multislice MRI," *IEEE Trans. Med. Imag.*, vol. 28, no. 5, pp. 633–644, May 2009.
- [38] B. Scherrer, A. Gholipour, and S. Warfield, "Super-resolution in diffusion-weighted imaging," in *Proc. Int. Conf. Med. Imag. Comput. Comput.-Assist. Interv.*, 2011, pp. 124–132.
- [39] S. Peled and Y. Yeshurun, "Superresolution in MRI: Application to human white matter fiber tract visualization by diffusion tensor imaging," *Magn. Res. Med.*, vol. 45, no. 1, pp. 29–35, 2001.
- [40] H. Greenspan, G. Oz, N. Kiryati, and S. Peled, "MRI inter-slice reconstruction using super-resolution," *Magn. Res. Imag.*, vol. 20, no. 5, pp. 437–446, 2002.
- [41] Y. Bai, X. Han, and J. Prince, "Super-resolution reconstruction of MR brain images," in *Proc. 38th Annu. Conf. Inf. Sci. Syst.*, 2004, pp. 1358–1363.
- [42] F. Rousseau, K. Kim, C. Studholme, M. Koob, and J. Dietemann, "On super-resolution for fetal brain MRI," in *Proc. Med. Imag. Comput. Comput.-Assist. Interv.*, 2010, pp. 355–362.
- [43] S. Jiang, H. Xue, A. Glover, M. Rutherford, D. Rueckert, and J. Hajnal, "MRI of moving subjects using multislice snapshot images with volume reconstruction (SVR): Application to fetal, neonatal, and adult brain studies," *IEEE Trans. Med. Imag.*, vol. 26, no. 7, pp. 967–980, Jul. 2007.
- [44] B. Scherrer, A. Gholipour, and S. Warfield, "Super-resolution reconstruction to increase the spatial resolution of diffusion weighted images from orthogonal anisotropic acquisitions," *Med. Imag. Anal.*, 2012. DOI: <http://dx.doi.org/10.1016/j.media.2012.05.003>.
- [45] E. Carmi, S. Liu, N. Alon, A. Fiat, and D. Fiat, "Resolution enhancement in MRI," *Magn. Res. Imag.*, vol. 24, no. 2, pp. 133–154, 2006.
- [46] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.
- [47] G. Hermosillo, C. Chef'd'Hotel, and O. Faugeras, "Variational methods for multimodal image matching," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 329–343, 2002.
- [48] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.
- [49] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient super-resolution image reconstruction algorithm," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 573–583, Apr. 2001.
- [50] R. Hardie, K. Barnard, and E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.
- [51] M. Ng, H. Shen, E. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 74585, 2007.
- [52] T. Chan, M. Ng, A. Yau, and A. Yip, "Superresolution image reconstruction using fast inpainting algorithms," *Appl. Comput. Harmonic Anal.*, vol. 23, no. 1, pp. 3–24, 2007.
- [53] G. Chantas, N. Galatsanos, and N. Woods, "Super-resolution based on fast registration and maximum a posteriori reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1821–1830, Jul. 2007.
- [54] N. Villain, Y. Goussard, J. Idier, and M. Allain, "Three-dimensional edge-preserving image enhancement for computed tomography," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1275–1287, Oct. 2003.
- [55] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.
- [56] L. Ibáñez, W. Schroeder, L. Ng, and J. Cates, "The ITK Software Guide: The Insight Segmentation and Registration Toolkit," 1.4 ed: Kitware, Inc., 2003.

- [57] S. Bochkhanov and V. Bystritsky, (2008) "ALGLIB," [Online]. Available: [www.alglib.net](http://www.alglib.net)
- [58] A. Tikhonov, "Regularization of incorrectly posed problems," *Sov. Math. Dokl.*, vol. 4, no. 6, pp. 1624–1627, 1963.
- [59] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.
- [60] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli, "Super-resolution from unregistered and totally aliased signals using subspace methods," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3687–3703, Jul. 2007.
- [61] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. IEEE Int. Conf. Imag. Process.*, Nov. 1994, vol. 2, pp. 168–172.
- [62] S. Geman and D. E. McClure, "Bayesian image analysis: An application to single photon emission tomography," *Statist. Comput. Sec. Proc. Amer. Statist. Assoc.*, vol. 21, no. 2, pp. 12–18, 1985.
- [63] P. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.
- [64] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-d Bayesian reconstruction from poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 194–202, Jun. 1989.
- [65] S. Brette and J. Idier, "Optimized single site update algorithms for image deblurring," in *Proc. Int. Conf. Imag. Process.*, Sep. 1996, vol. 3, pp. 65–68.

Authors' photographs and biographies not available at the time of publication.