

Super-resolution Reconstruction for Tongue MR Images

Jonghye Woo^{1,2}, Ying Bai³, Snehashis Roy², Emi Z. Murano², Maureen Stone¹, Jerry L. Prince²

¹University of Maryland, Baltimore MD 21201

²Johns Hopkins University, Baltimore MD 21218

³HeartFlow Inc., Redwood City CA 94063

ABSTRACT

Magnetic resonance (MR) images of the tongue have been used in both clinical medicine and scientific research to reveal tongue structure and motion. In order to see different features of the tongue and its relation to the vocal tract it is beneficial to acquire three orthogonal image stacks—e.g., axial, sagittal and coronal volumes. In order to maintain both low noise and high visual detail, each set of images is typically acquired with in-plane resolution that is much better than the through-plane resolution. As a result, any one data set, by itself, is not ideal for automatic volumetric analyses such as segmentation and registration or even for visualization when oblique slices are required. This paper presents a method of super-resolution reconstruction of the tongue that generates an isotropic image volume using the three orthogonal image stacks. The method uses preprocessing steps that include intensity matching and registration and a data combination approach carried out by Markov random field optimization. The performance of the proposed method was demonstrated on five clinical datasets, yielding superior results when compared with conventional reconstruction methods.

1. 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. 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. 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 last two decades to analyze tongue structure and function ranging from studies of the vocal tract¹⁻⁴ to studies on tongue muscle deformation.⁵⁻⁸ 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 3D 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 3D 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. 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 Figs. 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 an in-plane resolution of 0.94 mm×0.94 mm but are 3 mm thick.

Because the slices are relatively thick, none of the acquired image stacks are ideal for 3D volumetric analyses such as segmentation, registration, and atlas building or even for visualization when oblique views are required. 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 work, we develop a fully automated and accurate super-resolution volume reconstruction method from three orthogonal image stacks of the same subject. Because of the motion between scans and limited FOV, our problem is somewhat

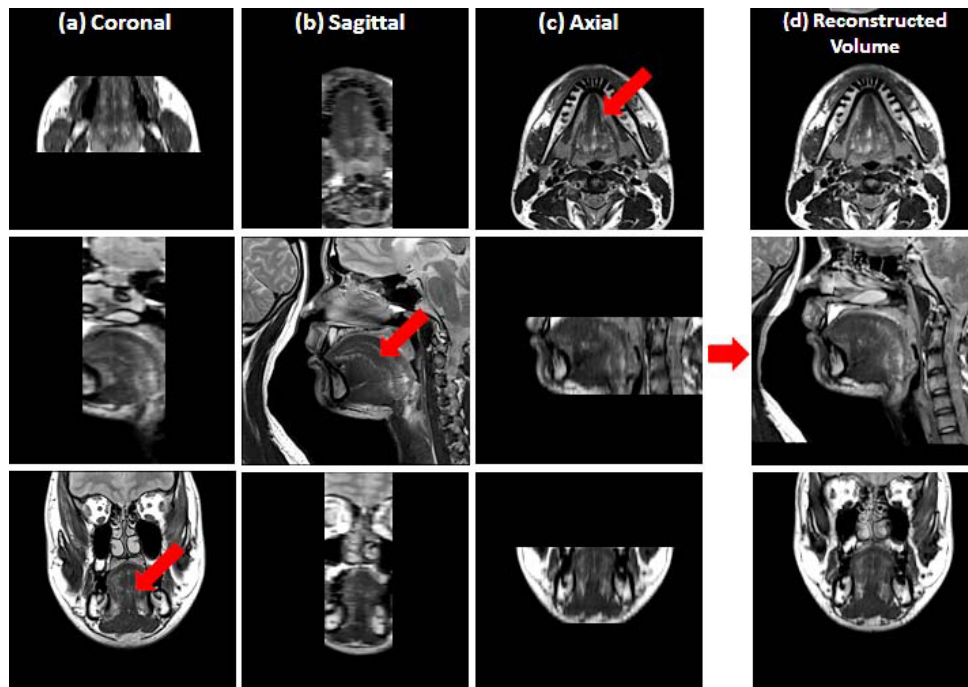


Figure 1. Tongue images are acquired in three orthogonal volumes with field of view encompassing the tongue and surrounding structures. (a) Coronal, (b) sagittal, and (c) axial volumes are illustrated. The final super-resolution volume using the proposed method is shown in (d). The red arrows indicate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

different than the conventional, well-studied super-resolution framework. We propose 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. 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 of super-resolution reconstruction applied to *in vivo* tongue high-resolution MR images. The resulting super-resolution 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). Our reconstruction allows full 3D volumetric data and improves further image/motion and visual analyses of the tongue.

2. RELATED WORK

To date there have been no reports describing super-resolution reconstruction of the tongue from low-resolution orthogonal MR images. However, some highly relevant work has been reported in other areas. For example, in brain imaging it is common to take multiple scans of the same subject from different orientations, and the acquired images typically have better in-plane than through-plane resolution. Bai et al.⁹ proposed an MAP super-resolution 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.

Several researchers have developed methods for fetal brain imaging where due to uncontrolled motion it is common to acquired multiple orthogonal 2D multiplanar acquisitions with anisotropic voxel sizes.^{10–12} To yield a single high-resolution registered volume image of the fetal brain from this kind of data, Rousseau et al.¹⁰ incorporated a registration method to correct motion and final reconstruction process was based on a local neighborhood approach with a Gaussian kernel. Jiang et al.¹² proposed motion correction using registration and B-spline based scattered interpolation to reconstruct the final 3D fetal brain. Gholipour et al.¹³ proposed maximum likelihood (M-estimation) error norm to reconstruct the fetal MR images. In recent work, Rousseau¹¹ used an edge-preserving regularization method to reconstruct high-resolution images from the fetal MRI images

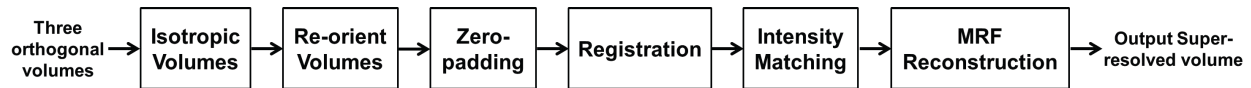


Figure 2. A flowchart of the proposed method.

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).

Example-based methods using non-local means have also been explored in the super-resolution of medical images. Rousseau,¹⁴ in particular, 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 these methods are 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.

3. METHOD

For our application, the imaging model that incorporates the overlap regions is then given by

$$\mathbf{g}_k = \Lambda_k(W_k S_k(h * \mathbf{f})) + \mathbf{n}_k, \quad k = 1, 2, 3 \quad (1)$$

where \mathbf{g}_k is one of the observed image volumes, Λ_k denotes a localized region (see Fig. 1), W_k is an intensity transformation caused by a coordinate transformation, S_k is a downsampling operator, h is a blurring operator, and \mathbf{n}_k is a Gaussian noise with zero mean and variance σ_k^2 . This model is combined as $H_k = \Lambda_k(W_k S_k(h * \mathbf{f}))$. 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. The goal of super-resolution is to estimate \mathbf{f} , which is a single high-resolution, isotropic image.

Our problem is somewhat different than the conventional super-resolution framework in that: (1) each volume is acquired with a different FOVs and (2) each observation has its resolution degraded in only one dimension. Our proposed method consists of multiple steps as follows: (1) preprocessing to address different FOVs and intensity differences, (2) registration to correct subject motion between volumes, and (3) a region-based MAP Markov random field (MRF) reconstruction. The flowchart of the proposed method is shown in Fig. 2.

3.1 Preprocessing

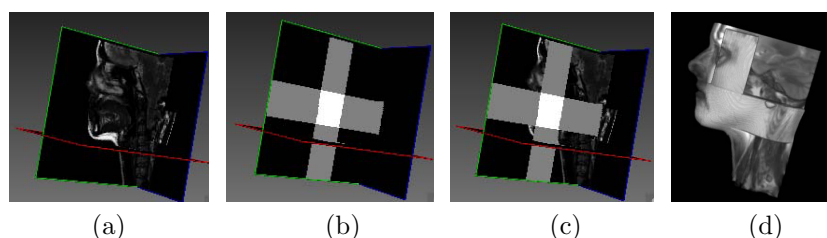


Figure 3. One representative final super-resolved image is shown in (a), regions defined from the masks are shown in (b), a volume with regions defined from the masks is illustrated in (c), and a 3D volume rendering is illustrated in (d).

The three volumes to be combined into one super-resolved volume have different slice positions, orientations, and volume sizes. Therefore, several preprocessing steps are applied prior to the reconstruction of the super-resolved volume: (i) generation of isotropic volume, (ii) conversion of the orientation of each volume to the orientation of the target reference volume, (iii) padding of zero values to yield the same volume sizes, (iv) registration to correct subject motion between volumes, and (v) matching of intensities in the overlap region using

piecewise spline regression, which normalizes the intensity values of a source volume based on the intensity values of a target volume. In addition, we generate masks from each volume to define overlap regions as shown in Fig. 3(b). We denote the regions of axial, coronal, and sagittal volumes by D_1 , D_2 , and D_3 , respectively. The masks are given by

$$M_k(x) = \begin{cases} 1, & x \in D_k \\ 0, & x \notin D_k \end{cases}, \quad k = 1, 2, 3, \quad (2)$$

where M_1 , M_2 and M_3 represent the characteristic functions of axial, coronal, and sagittal masks, respectively. In Fig. 3(b), the white region indicates the overlap region of all three volumes (i.e., $\Omega_1 = D_1 \cap D_2 \cap D_3$), the gray regions indicate the overlap regions of two volumes (i.e., $\Omega_2 = (D_1 \cap D_2) \cup (D_2 \cap D_3) \cup (D_1 \cap D_3)$), and the black regions indicate both background as well as regions where only one volume is available (i.e., Ω_3).

Because the precise positions of these acquired volumes relative to the anatomy might be slightly wrong due to patient motion, we use blurred versions of these masks—termed “softmasks”—which are given by

$$m_k = G_\sigma * M_k, \quad k = 1, 2, 3, \quad (3)$$

where $*$ denotes the convolution operator and G_σ is a unit-height Gaussian kernel with standard deviation σ . In this work, we set $\sigma=2$ mm.

3.1.1 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 disturbing artifacts within the MAP-MRF reconstruction algorithm. To obtain accurate and robust registration results, we compute a global displacement estimate followed by a local deformation models. The global displacement estimate is characterized by an affine registration accounting for translation, rotation, and scaling (12 degrees of freedom in 3D). We use mutual information (MI)¹⁵ 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 than a conventional MI registration algorithm because of the FOV differences in the volumes. In this global registration step, we use the overlap regions of two volumes to weight the evaluation of MI in order to emphasize influence on parameter estimation to regions that are known to be overlapping.

In order to achieve sub-voxel accuracy, we estimate a local deformation using the Demons method.¹⁶ We also incorporate the overlap region to restrict the domain of registration. Prior to registration, we perform histogram matching to standardize intensity distribution to ensure that corresponding points in different volumes have similar intensity values for the local deformation model.

3.2 MAP-MRF Reconstruction

To find optimal solution, we use a Bayesian framework whereby we can estimate the high quality image given the three volumes. We follow the work of Villain et al.¹⁷ in setting up an MAP-MRF estimation framework. Using Bayes rule and MAP estimator, one can write

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f}) P(\mathbf{f}), \quad (4)$$

where $P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f})$ and $P(\mathbf{f})$ denote likelihood and prior, respectively and $\hat{\mathbf{f}}$ is the estimated solution. In this work, we assume that the noise model is additive Gaussian noise and the likelihood $P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f})$ which depends on image formation process defined in Eqn. (1) and noise model can be expressed as

$$P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f}) = K \exp \left(\sum_{k=1}^3 - \frac{\|\mathbf{g}_k - H_k \mathbf{f}\|^2}{2\sigma_k^2} \right), \quad (5)$$

where \mathbf{g}_k is k th low resolution volume, \mathbf{f} is the high quality volume to be reconstructed, and K denotes a normalization factor. For a prior model, we use an MRF model because it can restore sharp discontinuities in the volumes. The Gibbs formulation is used as our prior model, i.e.,

$$P(\mathbf{f}) = \frac{1}{G} \exp \left\{ - \sum_{c \in C} V_c \right\}, \quad (6)$$

where V_c denotes the Gibbs potential¹⁷ defined on each set c of voxels which are mutual neighbors (called a clique) and G is a normalization factor. Maximization of the Eqn. (4) is equivalent to minimizing negative of logarithm:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} [-\log P(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 | \mathbf{f}) - \log P(\mathbf{f})], \quad (7)$$

which leads to

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} E = \arg \min_{\mathbf{f}} \left\{ \sum_{k=1}^3 \frac{\|\mathbf{g}_k - H_k \mathbf{f}\|^2}{2\sigma_k^2} + \lambda \sum_{c \in C} \varphi(u_c) \right\}. \quad (8)$$

Here, $\varphi(u) = \sqrt{1 + (u/\delta)^2}$, $u_c = \Delta x_c/d_c$ and λ is a balancing parameter, where δ denotes a scaling factor and Δx_c and d_c denote the difference of the values and the distance of the two voxels in clique c , respectively. We use half-quadratic regularization technique to solve the minimization problem in Eqn. (8).¹⁷ In addition, we incorporate the region based approach as there are partial overlaps from the three volumes to reconstruct the final volume as follows:

$$\hat{\mathbf{f}} = \begin{cases} \arg \min_{\mathbf{f}} \sum_{k=1}^3 m_1 \cdot m_2 \cdot m_3 \frac{\|\mathbf{g}_k - H_k \mathbf{f}\|^2}{2\sigma_k^2} + \lambda \sum_{c \in C} \varphi(u_c), & \text{if } x \in \Omega_1 \\ \arg \min_{\mathbf{f}} \sum_{k \in \{i,j\}} m_i \cdot m_j \frac{\|\mathbf{g}_k - H_k \mathbf{f}\|^2}{2\sigma_k^2} + \lambda \sum_{c \in C} \varphi(u_c), & \text{if } x \in \Omega_2 \text{ (determine } i, j \text{ by thresholding)} \\ \arg \min_{\mathbf{f}} \frac{\|\mathbf{g}_k - H_k \mathbf{f}\|^2}{2\sigma_k^2} + \lambda \sum_{c \in C} \varphi(u_c), & \text{if } x \in \Omega_3 \text{ (determine } k \text{ by thresholding).} \end{cases} \quad (9)$$

Here, softmasks are used as a weighting function in order to reduce the effects of “seams” between different regions.

4. VALIDATION

4.1 Clinical Tongue Data

Five high-resolution MR datasets, consisting of three normal and two patients who had tongue cancer surgically resected (glossectomy) native American English speakers, were used in our experiments. The image size and resolution for high-resolution MRI were $256 \times 256 \times z$ (z ranges from 10 to 24) and $0.94 \text{ mm} \times 0.94 \text{ mm} \times 0.94 \text{ mm}$, respectively. High-resolution MRI datasets were acquired at rest position using a head and neck coil. The subjects were required to remain still from 1.5 to 3 minutes for each plane.

4.2 Evaluation of the Algorithm

A challenge in testing the proposed method is the lack of the ground truth available in *in vivo* volumetric tongue MR data. The proposed method was first evaluated using five simulated datasets. In order to quantitatively evaluate the performance of the proposed method in terms of accuracy of the reconstruction, the final super-resolution 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 resolution of $0.94 \text{ mm} \times 0.94 \text{ mm} \times 0.94 \text{ mm}$. Three low-resolution volumes (i.e., axial, sagittal, and coronal volumes) were formed in which the voxel dimensions was $0.94 \text{ mm} \times 0.94 \text{ mm}$ in-plane and the slice thickness was 3.46 mm (subsampling by a factor 4). Prior to subsample 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 anti-aliasing effect.

In what follows, volume reconstruction was carried out in four ways: First, 5th-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 regularization¹⁸ was performed, and finally, the

proposed method using three up-sampled volumes were carried out. In this simulation study, preprocessing was not necessary except up-sampling process using 5th-order B-spline interpolation.

As a quantitative measure, the peak signal-to-noise ratio (PSNR) was used similar to the work,¹⁹ which is defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{D(X(v))}{|\Omega_R|^{-1} \sum_{v \in \Omega_R} (X(v) - \hat{X}(v))^2} \right), \tag{10}$$

where Ω_R denotes a reference image domain, X denotes a reference volume, \hat{X} represents a reconstructed volume, and $D(X(v))$ represents the dynamic range of a reference volume.

Second, the proposed method was also visually compared to the different reconstruction schemes using original low-resolution datasets including averaging and Tikhonov regularization after same preprocessing steps were applied.

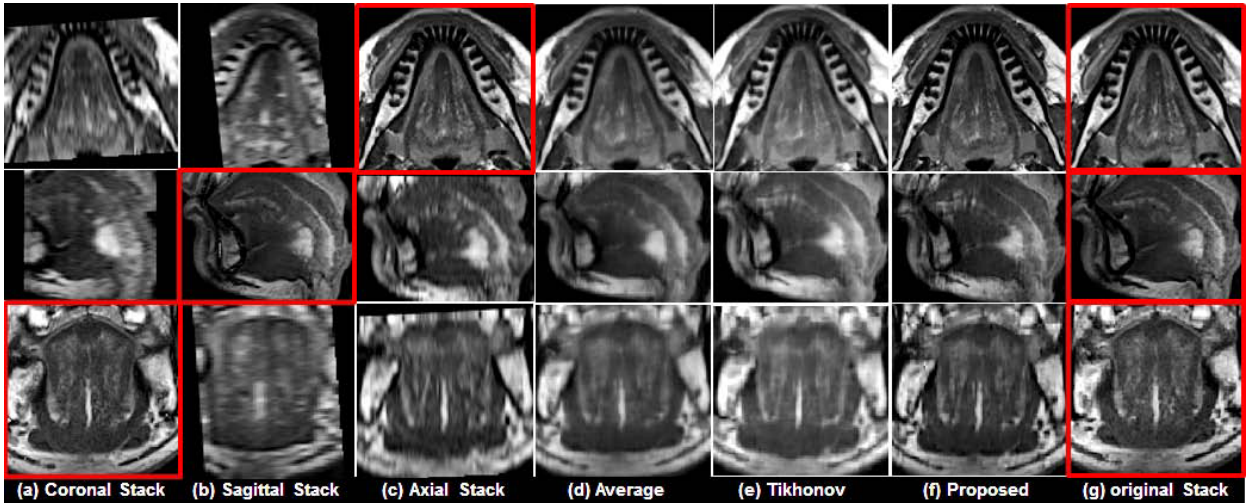


Figure 4. Comparison of different reconstruction methods using normal subject. Original low-resolution coronal, sagittal, and axial volumes are shown in (a), (b), and (c), respectively. Red boxes represent original volumes after isotropic volume upsampling using B-spline. Three different reconstruction methods including simple averaging, Tikhonov regularization, and the proposed method are presented in (d), (e), and (f), respectively. Original three volumes are illustrated in (g). Please notice that the proposed method provides detailed anatomical information compared to averaging and Tikhonov regularization methods as visually assessed.

5. RESULTS

Once we obtained the super-resolution volume, a simulation study was carried out in which we subsampled the reconstructed volume by a factor of 4 in each axis independently, generating three volumes degraded in only one dimension. In order to quantitatively measure the performance of the reconstruction, PSNR was used after reconstruction using different methods, showing that the proposed method achieved better than other reconstruction methods. Table 1 summarizes the quantitative results after the reconstruction.

Table 1. PSNR (dB) for differnt reconstruction methods on simulation data (Mean±SD)

Axial	Sagittal	Coronal	Averaging	Tikhonov	Proposed
28.9±3.8	27.1±3.7	26.5±4.6	29.1±4.0	30.2±4.4	34.1±3.2

In Figs. 4 and 5, two representative results using a normal subject (Fig. 4) and a glossectomy patient (Fig. 5) 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 mm×0.94 mm×0.94 mm) in the coronal, sagittal, and axial planes, respectively. (d) shows the reconstruction using averaging, (e) shows reconstruction using Tikhonov regularization,

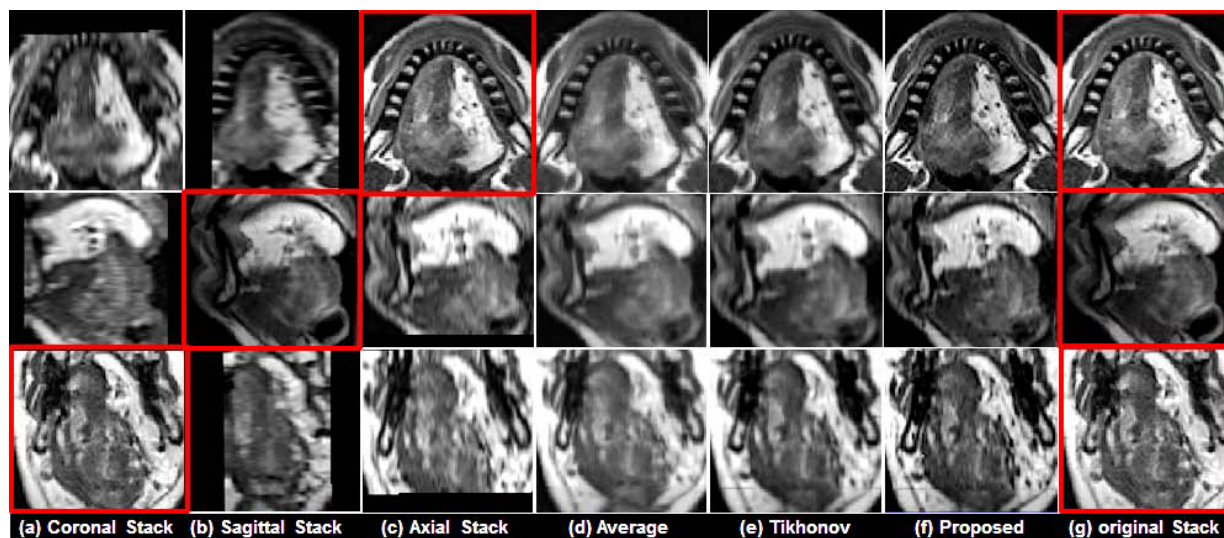


Figure 5. Comparison of different reconstruction methods using a glossectomy patient. Original low-resolution coronal, sagittal, and axial volumes are shown in (a), (b), and (c), respectively. Red boxes represent original volumes after isotropic volume upsampling using B-spline. Three different reconstruction methods include (d) simple averaging, (e) Tikhonov regularization, and (f) the proposed method, respectively. Original three volumes are illustrated in (g). Please notice that the proposed method provides detailed anatomical information compared to averaging and Tikhonov regularization methods as visually assessed.

and (f) shows the reconstruction using the proposed method. (g) shows the original three orthogonal volumes. As shown in the Figures, the proposed method provided sharp muscle and fine anatomical detail as compared with the other methods. The target reference volume into which the other volumes were registered was the axial volume in both cases and therefore the reconstructed images and the original images shown in the figures may not be exactly same except the axial slice.

6. CONCLUSION

In this work, we investigated a super-resolution technique to generate isotropic and high resolution images for the tongue MR images. Due to the limited FOVs of three orthogonal volumes, we proposed to use the region based MAP-MRF reconstruction method to obtain super-resolved volume with fine detail of the anatomical structures. The experimental results were demonstrated with superior performance when compared with the conventional reconstruction methods.

REFERENCES

- [1] Narayanan, S., Byrd, D., and Kaun, A., "Geometry, kinematics, and acoustics of tamil liquid consonants," *The Journal of the Acoustical Society of America* **106**, 1993–2007 (1999).
- [2] Narayanan, S., Alwan, A., and Haker, K., "An articulatory study of fricative consonants using magnetic resonance imaging," *The Journal of the Acoustical Society of America* **98**, 1325 (1995).
- [3] Bresch, E., Kim, Y., Nayak, K., Byrd, D., and Narayanan, S., "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging," *IEEE Signal Processing Magazine*, 123–132 (2008).
- [4] Lakshminarayanan, A., Lee, S., and McCutcheon, M., "MR imaging of the vocal tract during vowel production," *Journal of Magnetic Resonance Imaging* **1**(1), 71–76 (1991).
- [5] Stone, M., Davis, E., Douglas, A., NessAiver, M., Gullapalli, R., Levine, W., and Lundberg, A., "Modeling the motion of the internal tongue from tagged cine-MRI images," *The Journal of the Acoustical Society of America* **109**(6), 2974–82 (2001).
- [6] Stone, M., Liu, X., Chen, H., and Prince, J., "A preliminary application of principal components and cluster analysis to internal tongue deformation patterns," *Computer methods in biomechanics and biomedical engineering* **13**(4), 493–503 (2010).

- [7] Napadow, V., Chen, Q., Wedeen, V., and Gilbert, R., "Intramural mechanics of the human tongue in association with physiological deformations," *Journal of biomechanics* **32**(1), 1–12 (1999).
- [8] Takano, S. and Honda, K., "An MRI analysis of the extrinsic tongue muscles during vowel production," *Speech communication* **49**(1), 49–58 (2007).
- [9] Bai, Y., Han, X., and Prince, J., "Super-resolution reconstruction of MR brain images," in [*Proc. of 38th Annual Conference on Information Sciences and Systems (CISS04)*], (2004).
- [10] Rousseau, F., Glenn, O., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D., Barkovich, J., and Studholme, C., "Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images," *Academic radiology* **13**(9), 1072–1081 (2006).
- [11] Rousseau, F., Kim, K., Studholme, C., Koob, M., and Dietemann, J., "On super-resolution for fetal brain MRI," *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2010*, 355–362 (2010).
- [12] Jiang, S., Xue, H., Glover, A., Rutherford, M., Rueckert, D., and Hajnal, J., "MRI of moving subjects using multislice snapshot images with volume reconstruction (svr): application to fetal, neonatal, and adult brain studies," *IEEE Transactions on Medical Imaging* **26**(7), 967–980 (2007).
- [13] Gholipour, A., Estroff, J., and Warfield, S., "Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain mri," *IEEE Transactions on Medical Imaging* **29**(10), 1739–1758 (2010).
- [14] Rousseau, F., "Brain hallucination," *Computer Vision-ECCV 2008*, 497–508 (2008).
- [15] Viola, P. and Wells, W. M., "Alignment by maximization of mutual information," *International Journal of Computer Vision* **24**(2), 137–154 (1997).
- [16] Thirion, J., "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical image analysis* **2**(3), 243–260 (1998).
- [17] Villain, N., Goossard, Y., Idier, J., and Allain, M., "Three-dimensional edge-preserving image enhancement for computed tomography," *IEEE Transactions on Medical Imaging* **22**(10), 1275–1287 (2003).
- [18] Tikhonov, A., "Regularization of incorrectly posed problems," in [*Soviet Math. Dokl*], **4**(6), 1624–1627 (1963).
- [19] Rousseau, F., Kim, K., and Studholme, C., "A groupwise super-resolution approach: application to brain MRI," in [*IEEE International Symposium on Biomedical Imaging: From Nano to Macro*], 860–863 (2010).