Spatio-Temporal Resolution Enhancement of Vocal Tract MRI Sequences Based on the Wiener Filter

Ana L. D. Martins, Nelson D. A. Mascarenhas
Universidade Federal de So Carlos, Departamento de Computao
Via Washington Luís, Km 235, CP 676, CEP: 13.565-905, So Carlos, SP, Brazil
ana_martins@dc.ufscar.br; nelson@dc.ufscar.br

***Abstract*—Research on human speech production is highly dependent upon information about the position and movements of the speech articulators. Dynamic magnetic resonance imaging (MRI) has been the main tool to support this process. With this technique, image sequences can be acquired in the act of speech, which allows identifying shapes of the vocal tract in real time. However, the spatial and temporal resolution requirements are not known a priori and are expected to vary according to the speech task. Several available approaches enhance resolution by either changing the acquisition process of current devices, or by trading the acquisition devices themselves by more powerful ones. Both solutions involve additional hardware costs. In this paper, we propose an evolution of an approach to enhance spatio-temporal resolution of MRI image sequences of the vocal tract using only digital image processing techniques. On one hand, temporal resolution is increased by generating intermediate images according to the movement present in an observed sequence. On the other hand, spatial resolution is increased by applying a novel approach to super-resolution image reconstruction based on the Wiener filter. To evaluate the proposed approach, we processed a set of five simulated low resolution images in a sequence. Compared to available methods, results provide evidence of the effectiveness of the proposed method.**

***Keywords*-magnetic resonance imaging (MRI); human speech production research; spatio-temporal resolution enhancement; super-resolution image reconstruction.**

I. Introduction

Based on models of the human vocal tract shape and the acoustical processes involved in speech production, articulatory synthesis uses computational techniques for artificially synthesizing speech. Fig. 1 shows the contours of interest in speech production research: larynx, epiglottis, lips, pharyngeal wall, glottis, velum and hard palate. These anatomical components, which are called speech articulators, are controlled in order to change the shape of the vocal tract during the speech production process. Therefore, knowledge about speech articulators position and movements is essential for articulatory synthesis research.

Magnetic resonance imaging (MRI) is an emerging technique for studying speech production because it is an in vivo safe nonradioactive procedure which permits a good delineation of soft tissues. It is considered a powerful tool since it allows the three-dimensional visualization of the vocal tract during phonation. Since the initial application presented by Baer et al. [2], the use of MRI has produced very useful results to the speech production research [3]. Because

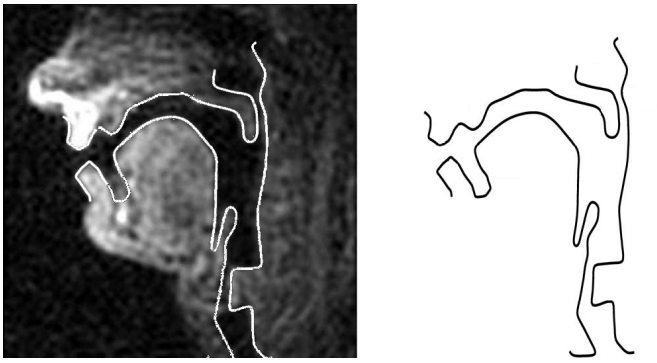


Fig. 1. Contours of interest to the speech production research. Adapted from the work of Bresch and Narayanan [1]

of the low spatio-temporal resolution of conventional MRI acquisition techniques, earlier studies were limited to static postures such as vowel sounds. Later, the development of the cine MRI technique allowed the imaging of dynamic vocal tract shaping. However, this method relies on the scanning of numerous exact repetitions of the same speech sequence to reconstruct the impression of articulatory movement in time. Therefore, this is not the most adequate imaging method for the study of continuous running speech. Real-time MRI refers to the direct capturing of moving image data with frame rates sufficient to capture the speech articulators movement. According to Bresch et al. [4], this became possible because of some advances in imaging technology, including: improvements in computer speed, parallel imaging, rapidly fluctuating gradients, novel k-space trajectories, etc.

However, fast acquisition of high-quality images, the detection of the main features in each image, and the analysis and modeling of the time-varying vocal tract shape still are great challenges in this subject [4]. Considering image quality, the temporal and spatial resolution requirements are expected to vary depending on the speech task and there is no prior information about these requirements. Moreover, even though MRI advances represent a significant improvement in the quality of information about changes in vocal tract shape over time, MRI capacity is still not close to the spatio-temporal resolution necessary for capturing the dynamic characteristics of tongue movement.

Several available approaches enhance resolution by either

empowering the acquisition process, or by trading the acquisition devices themselves by more powerful ones. In both cases, there are financial and hardware limitations. For instance, a long exposure time is required to obtain images with sufficient anatomic detail. In fact, the trade-off between image quality and frame rate limits the use of MRI in the study of some physiological events. Therefore, it is of great interest to increase the resolution of existing image sequences using only digital image processing techniques.

To the best of our knowledge, Martins et al. [5] were the first to develop a method for spatio-temporal resolution enhancement of vocal tract image sequences. They proposed a two stage approach based on a previous non-rigid image registration method. In the first stage of their approach, displacements and deformations, estimated by the image registration method proposed by Rueckert et al. [6], were used in a motion compensated interpolation procedure to generate intermediate images. In this way, the resulting images were coherent with the movement present in the observed sequence. Then, in the second stage, in order to increase spatial resolution, the displacements identified by the registration method were used in a super resolution image reconstruction (SRIR) approach that applied a maximum a posteriori probability (MAP) estimation based on a Markov random field (MRF) prior model. However, since the iterated conditional modes (ICM) algorithm was used to sequentially update high resolution pixel intensities, the approach turned out to be computationally costly.

In this paper, we propose an evolution of Martins et al.'s approach that generates higher quality images with reduced computational cost. The SRIR method used in the second stage of the previous proposal is replaced by a novel discrete Wiener filter based method. A separable covariance matrix of a first order Markov process [7] is constructed according to the distribution of the low resolution pixel areas along the high resolution grid. In this way, because of the characteristics of the Wiener filter, resulting images are reconstructed respecting the minimum mean squared error (MMSE) criterion in a single iteration.

Based on several observed vocal tract image sequences, equivalent sequences with higher spatio-temporal resolution were generated for a visual evaluation of the proposed approach in a real situation. Moreover, a numerical evaluation of the proposed method has been conducted by processing a simulated low resolution sequence with known deformations. Compared to the method proposed by Martins et al. [5], results provide evidence of the effectiveness of our method. The remainder of this paper is organized as follows. Section II provides a background for understanding the proposed method. Section III presents the proposed spatio-temporal resolution enhancement approach. Section IV evaluates our method in comparison to others. Section V discusses related work. Finally, Section VI concludes the paper.

II. Background

A. The Non-Rigid Image Reconstruction Method

Rueckert et al. [6] present a non-rigid image registration method based on free-form deformations (FFD) and cubic B-spline interpolations. This method is widely used in the medical context [8]. FFD was originally developed to model 3D deformable objects in computer graphics applications. It is defined by a discrete three-dimensional mesh of uniform spaced control points, each of which associated with a displacement vector. In contrast with nonparametric transformation, where a displacement vector is associated with every location, FFDs describe the displacements of a general location of the image by a set of vectors, so that the nearer have more influence in this location. The weights associated with each vector are defined by a weighting function, such as B-splines.

Following [6], deformations are modeled by a transformation φ, which combines global scene movement with local deformations. In the context of the vocal tract image sequences, it is not worth to consider the global scene motion, since deformations are highly localized around the jaw. Moreover, considering only midsaggital image sequences of the vocal tract, the transformation is bidimensional. Local deformations are modeled by the FFD, which is written as the three-dimensional tensor product of the uni-dimensional cubic B-splines. Considering the image volume $\Omega = \{0 \leq x \leq X, 0 \leq y \leq Y\}$, let Φ denote a $n_x \times n_y$ mesh of control points $\phi_{i,j}$ with uniform spacing δ. The local transformation is given by

$$\varphi(x,y) = \sum_{m=0}^{3}\sum_{n=0}^{3} B_m(u)B_n(v)\phi_{i+m,j+n}, \tag{1}$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$ e $v = y/n_y - \lfloor y/n_y \rfloor$. B_k represents the k-th basis function of the B-spline

$$\begin{aligned} B_0(u) &= (1-u^3)/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6. \end{aligned} \tag{2}$$

Note that this basis functions have limited support, which means that changes in one control point coordinates affects only its local neighborhood.

The control points act as parameters of the local transformation φ. Indeed, the kind of deformation that can be modeled by this transformation is highly dependent on the resolution of the mesh of control points. Control points with smaller spacings allow more localized deformations. However, the computational complexity is also defined by the resolution of the mesh. Therefore, in order to accomplish the best tradeoff between model flexibility and computational complexity, Rueckert et al. implemented a hierarchical multi-resolution approach in which the resolution of the mesh is increased in a coarse to fine way.

In order to regularize the transformation, guaranteeing its

smoothness, the following penalty term was used

$$C_{\text{smooth}} = \frac{1}{S}\int_0^X \int_0^Y \left[\left(\frac{\partial^2\varphi}{\partial x^2}\right)^2 + \left(\frac{\partial^2\varphi}{\partial y^2}\right)^2 + 2\left(\frac{\partial^2\varphi}{\partial x y}\right)^2\right] dx\, dy. \quad (3)$$

Moreover, to relate the reference image with the transformed one, the normalized mutual information is used as the similarity criterion

$$C_{\text{similarity}}(I_1, \varphi(I_2)) = \frac{H(I_1) + H(\varphi(I_2))}{H(I_1, \varphi(I_2))}, \quad (4)$$

where $H(I_1)$ and $H(\varphi(I_2))$ are the marginal entropies of I_1 and the transformed image $\varphi(I_2)$, and $H(I_1, \varphi(I_2))$ stands for their joint entropy. The optimal transformation is found minimizing the cost function

$$C(\Phi) = -C_{\text{similarity}(I_1, \varphi(I_2))} + \lambda C_{smooth}(\varphi), \quad (5)$$

where λ is the weighting parameter.

B. Super-Resolution Image Reconstruction

Super-resolution image reconstruction (SRIR) is a powerful methodology for increasing spatial resolution using only signal processing techniques. Lately, it has been a very active research area because, in cases that high spatial resolution images are required, the existing low resolution imaging systems can still be utilized. SRIR methods attempt to reconstruct a high resolution image from a set of low resolution observations of the same scene. The observed images must present subpixel displacements among them. It allows the existence of different information on each of the observations and the exceeding information is used to increase spatial resolution (Fig. 2). Images with this characteristic can be acquired from a single camera with several captures; from multiple cameras located in different positions; by scene motions or local objects movements; by vibrating imaging systems; using video frames, etc. Therefore, the SRIR methodology is valuable in several contexts. For instance, surveillance applications frequently need to do synthetic zooming of regions of interest (the face of a criminal or the license plate of a car) in low resolution video sequences. In the medical context, multiple acquisitions are usually possible, nevertheless the quality of these images is usually limited. Likewise, in satellite imaging applications, several images of the same area are commonly acquired and higher resolutions are often demanded.

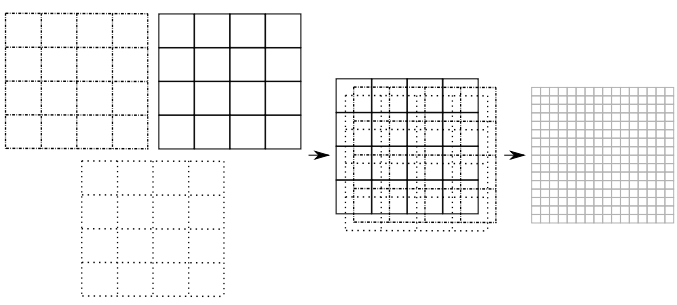


Fig. 2. Illustration of the SRIR process.

Tsai and Huang [9] were the first to discuss the SRIR problem. They adopted a frequency domain approach based on the shifting property of the Fourier transform, to model global translational scene motion. Recently, several algorithms were proposed, most of them in the spatial domain. In point of fact, despite the simplicity of frequency domain approaches, there are several disadvantages in this formulation. For instance, usually, it does not permit much flexibility with respect to the motion models. Spatial domain approaches are normally more flexible about motion models, degradation models and, mainly, the inclusion of a priori constraints. It is important to note that, similar to image restoration problems, the SRIR problem is considered ill-posed and regularized solutions using a priori constraints are usually required. Projection onto convex sets (POCS) based approaches impose prior knowledge by convex sets [10]. Nonetheless, despite the simplicity and flexibility of this approach, it demands high computational power and if the intersection of the sets is not a single point, there will be more than one solution. Therefore, the result depends on the initial estimation. Probabilistic reconstruction techniques are able to include prior knowledge in a more natural way and the Bayesian maximum a posteriori probability (MAP) estimation is the most promising method [11]. In this approach, all parameters and observable variables are considered unknown stochastic quantities and probability distributions are assigned to them based on subjective beliefs. In fact, the prior probability density function of the high resolution image is used to impose constraints to the solution. Markov Random Fields (MRF) prior models are considered the most flexible and realistic since they allow the inclusion of prior knowledge using only neighborhood relationships [12]. However, MAP-MRF approaches usually suffers from high computational burden. On the contrary, deterministic regularization methods are typically computationally simpler [13]. They use some desired information about the solution to stabilize it. Since images usually present limited high-frequency activity, smoothness is the most common constraint imposed.

To coherently analyze the SRIR problem, firstly it is essential to formulate the image formation model relating the desired high resolution image to the low resolution observations.

1) Image Formation Model: Consider $f[i,j]$, $0 \leq i,j \leq M$ an ideal undegraded image sampled at the Nyquist rate from the continuous scene of interest $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. In a real situation, the digital image is usually blurred by the optical system and also corrupted by noise. In this sense, following a lexicographic ordering, a low resolution degraded version $g_k[k,l]$, $0 \leq k,l \leq N$, $N \leq M$, of the high-resolution image f, can be modeled by

$$g_k = D_k f + n_k \quad (6)$$

where n_k stands for noise in the k-th low resolution observation, following an additive model. D_k models the sensor acquisition function. It implements the convolution with the sensor point spread function (PSF), followed by a sampling operator. According to Park et al. [14], most SRIR methods

in the literature model the sensor PSF as a spacial mean operator, assigning the mean of a high resolution block to the relative low resolution pixel. Moreover, in some proposals, this operator is also responsible for the sub-pixel displacements present among the observed images, as illustrated in Fig. 3. Considering d the scale factor, in practice, this operator is given by

$$D_k = \frac{1}{d^2} \begin{bmatrix} 11\ldots1 & & & 0 \\ & 11\ldots1 & & \\ & & \ddots & \\ 0 & & & 11\ldots1 \end{bmatrix}. \quad (7)$$

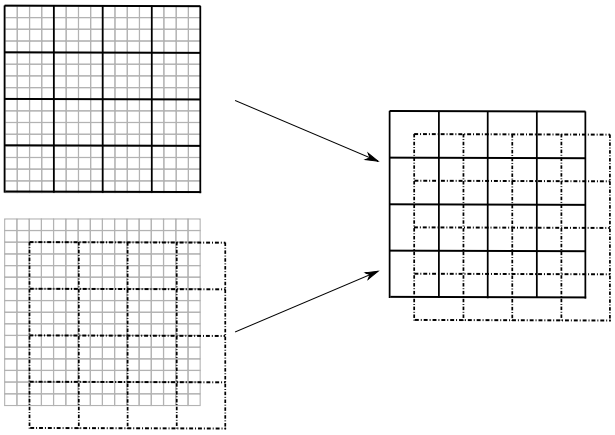


Fig. 3. Illustration of the downsampling operator causing the sub-pixel displacements among the observations.

The image formation model in Eq. (6) considers only one high resolution image. However, most SRIR approaches, which intend to reconstruct one high resolution image, are able to reconstruct a sequence of high resolution images following a sliding window approach [11]. For each high resolution image, a different low resolution image is considered as the reference one. Moreover, just a subset of the whole sequence is used in each step.

In order to estimate the sub-pixel displacements among the low resolution observations, the first step in an SRIR approach is to register all the observations considering one of them as a reference. In a sliding window approach, each reconstructed high resolution image corresponds to the low resolution reference in the considered subset of observations. Actually, given the sequence of low resolution observations g_k, $k = 1, \ldots, q$, a sequence f_k, $k = 1, \ldots, q$ will be reconstructed. In the reconstruction of f_k, g_k is considered as the reference, in the subset $g_{k-n}, \ldots, g_{k+n}$. Therefore, a subset of $2n + 1$ observations is being considered.

C. The Spatio-Temporal Resolution Enhancement Method

Martins et al. [5] proposed a two-step method to increase spatio-temporal resolution of human vocal tract MRI sequences. Considering a sequence of images, in order to increase temporal resolution, one could combine pixel values at the same spatial locations in adjacent images, to generate intermediate frames. However, this procedure blurs locations that moves and decreases resolution in small details of the images. In the first stage of their approach, displacements and deformations estimated by the non-rigid image registration method proposed by Rueckert et al. [6], were used in a motion compensated interpolation procedure to generate intermediate images. In this way, the generated images were coherent with the observed sequence motion.

Fig. 4 shows a sequence of vocal tract images, acquired with an MRI system operating at 1.5 Tesla (quantum gradients; 30mT/m amplitude; 0.24ms rise time; 125T/m/s Slew rate; 50cm FOV; 5 frames/s). The non-rigid registration algorithm was applied to each pair of images always selecting the first image as the reference one (the images of the meshes of control points are merely illustrative). Considering the correspondence between each control point in the sequence of identified meshes, intermediate meshes were generated by positioning control points in intermediate positions – linear interpolation of the corresponding control points coordinates in adjacent meshes. According to the intermediate mesh of control points, temporal resolution enhancement is performed by applying the transformation φ to both adjacent images, and then performing the mean of the transformed images as illustrated in Fig. 5. Considering that intermediate control points coordinates were found by linear interpolation of the corresponding control points coordinates in the adjacent images, the weights ω_1 and ω_2 are always 0.5.

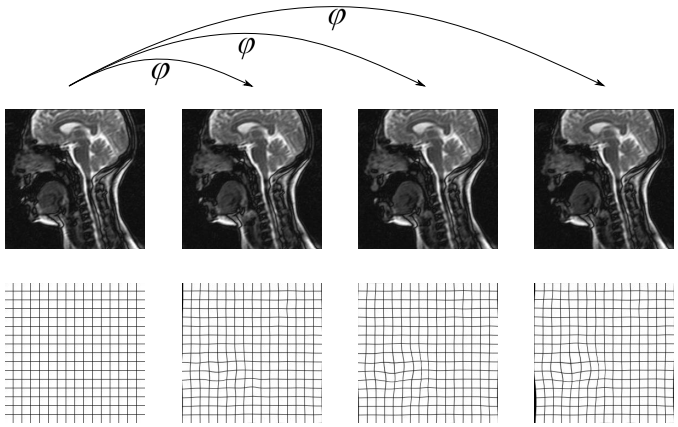


Fig. 4. Illustration of the registration of pairs of images.

In the second stage of the method proposed by Martins et al., in order to increase spatial resolution, the displacements identified by the registration method were used in a MAP-MRF SRIR procedure. Following a sliding window approach, each high resolution estimation f_n, $n = 1, \ldots, q$, was reconstructed considering a subset of the low resolution observations g_k, $k = n - 2, \ldots, n + 2$. Prior information was imposed by using the generalized isotropic multi-level logistic (GIMLL) MRF model to characterize each high resolution

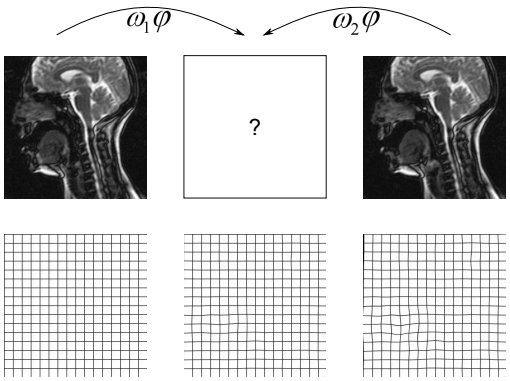


Fig. 5. Weighted sum of transformed neighboring images.

estimation. The iterated conditional modes (ICM) algorithm was used to sequentially update the high resolution pixel intensities f_n^i, $i = 0, \ldots, M^2$, by maximizing the posterior probability $P(f_n^i | g_{(n)}, f_n^{\eta_n^i})$, where η_n^i is the set of neighbors of pixel i in the high resolution image f_n, and $g_{(n)}$ indicate the subset of low resolution images in the sliding window $g_{(n)} = g_{n-2}, \ldots, g_{n+2}$.

III. The Proposed Method

A. Wiener Filter Based Super-Resolution Reconstruction

Mascarenhas et al. [7] present an statistical interpolator of the SPOT multispectral bands, under a Bayesian framework. The 20m resolution multispectral bands are interpolated to 10m resolution. The local linear estimation of the interpolated pixels was performed under the MMSE criterion, by using the orthogonality principle [15]. The authors assumed the separability of the correlation structure on the spatial and spectral domains, as well as the separability of the spatial correlation structure in the horizontal and vertical directions using the first order Markovian model in each direction. This is a common practice in the image processing literature [16].

Considering the ICM performance limitations, we replaced the second stage of the spatio-temporal resolution enhancement proposed by Martins et al. [5] by a discrete Wiener filter based SRIR method, similar to the statistical interpolator proposed by Mascarenhas et al. [7]. Considering the orthogonality principle and image formation model presented in Eq. (6), the estimation is given by

$$\hat{f} = E[f] + \Sigma_{fg}\Sigma_{gg}^{-1}\left(g - E[g]\right), \tag{8}$$

where $E[.]$ is the statistical expectation, g is the vector of low resolution pixel observations, Σ_{fg} the cross-correlation of f and g, and Σ_{gg} is the observations covariance matrix. Assuming separability on the horizontal and vertical directions, we define the prior covariance matrix Σ_{ff} using the first order Markovian model

$$r(x) = \sigma^2 \rho^{|x|}, |\rho| < 1, \forall x, \tag{9}$$

where σ^2 is the image variance. Thus, considering unit distances, Σ_{ff} will be defined as

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{N-1} \\ \rho & 1 & \rho & \ldots & \rho^{N-2} \\ \vdots & & \ddots & & \vdots \\ \rho^{N-1} & \ldots & & \rho & 1 \end{bmatrix}. \tag{10}$$

Considering the image formation model, Σ_{fg} and Σ_{gg} are given by

$$\Sigma_{gg} = D\Sigma_{ff}D^T + \Sigma_{nn}, \tag{11}$$

$$\Sigma_{fg} = D\Sigma_{ff}, \tag{12}$$

where Σ_{nn} is the noise covariance matrix.

The estimation is implemented for each high resolution pixel individually. For each low resolution pixel that is influenced by the current high resolution pixel, its eight neighbors are also used in the estimation of this high resolution pixel as illustrated in Fig. 6. In the following, we outline the proposed SRIR algorithm:

1) For each high resolution pixel f_i, assign to g the set of low resolution pixels influenced by it, together with their eight nearest neighbors;
2) Assign to f_i' the set of high resolution pixels that influences the low resolution pixels in g;
3) Define Σ_{ff} according to the spatial distribution of the high resolution pixels in f_i';
4) Define D_i, the downsampling operator relative to the low resolution pixels in g;
5) Define Σ_{gg} and Σ_{fg} (Equations (11) and (12));
6) Assign $E[f_i'] + \Sigma_{fg}\Sigma_{gg}^{-1}\left(g - E[g]\right)$ to f_i';
7) Extract f_i from f_i'.

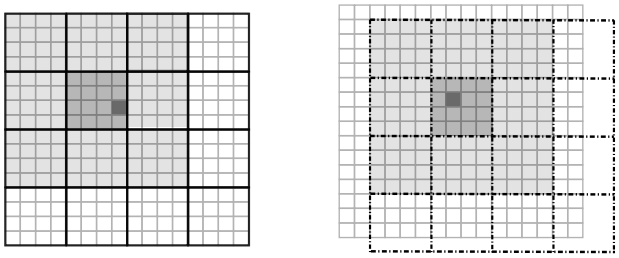


Fig. 6. Illustration of the low resolution pixel areas used in the covariance matrix construction.

There are computational savings associated with the estimation of a subset of high resolution pixels in each step of the algorithm. However, in this paper we only discuss the resulting images of the single pixel estimation.

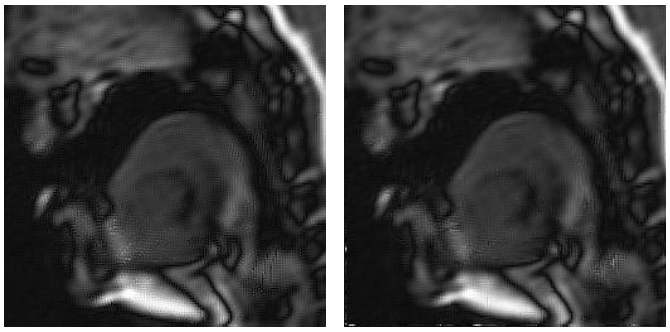
IV. Results

Considering a scale factor 2, Fig. 7 shows the details of the 256×256 images reconstructed by the GIMLL SRIR approach and by the proposed method. It is possible to note that, visually, they are very similar. However, the ICM algorithm

TABLE I
NMSE OF THE 5 SIMULATED IMAGES RECONSTRUCTED BY THE PROPOSED METHOD IN COMPARISON WITH THE BILINEAR INTERPOLATION OF THE LOW RESOLUTION SIMULATIONS AND THE IMAGES RECONSTRUCTED BY THE OTHER METHODS.

	Img01	Img02	Img03	Img04	Img05
Wiener	0.00071	0.00043	0.00031	0.00026	0.00025
GIMLL	0.00110	0.00066	0.00060	0.00056	0.00050
DAMRF	0.00110	0.00072	0.00063	0.00057	0.00051
TV	0.00330	0.00260	0.00240	0.00240	0.00240
Potts	0.02180	0.02160	0.02120	0.02140	0.02090
Bilinear	0.00800	0.01690	0.01610	0.01620	0.01650

converged in four iterations, performing the maximization of the local conditional probability for 65536 pixels in each iteration. On the other hand, the proposed method generated the high resolution image by performing the steps discussed in Section III only one time for each pixel.



(a) GIMLL (b) Wiener

Fig. 7. Images reconstructed by (a) the GIMLL SRIR approach and (b) by the proposed Wiener filter based approach.

Similar to Martins et al. [5], a numerical evaluation of the proposed method has been conducted by processing a set of five simulated low resolution images in a sequence. Transformations identified in a real sequence were used to simulate speech articulators' movement. These transformations were applied to one observed image and the simulated sequence was downsampled (scale factor of 2), generating a sequence of simulated low resolution images. Note that, in this experiment, the sub-pixel displacements are completely known. We used the same ICM-based methods considered by Martins et al. [5] in their evaluation. These methods are the following: the bilinear interpolation of the reference image, the GIMLL MRF proposed by Martins et al. [5], the discontinuity adaptive MRF (DAMRF) model proposed by Suresh and Rajagopalan [17], the Potts model discussed by Martins et al. [18], and the Total variation prior discussed by Li [19].

Fig. 8 presents the first high resolution image reconstructed by each of the methods. Note that the proposed method produced the best results.

The Normalized Mean Squared Error (NMSE) was used for this numerical evaluation. Table 1 shows the results. To have a visual idea of the data, Fig. 9 shows a boxplot of the results for the three best methods.

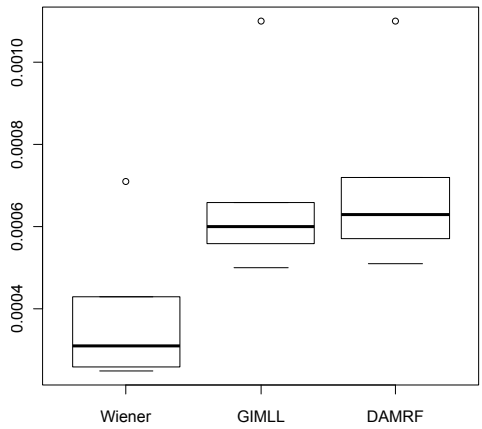


Fig. 9. Boxplot of the results for the three best methods.

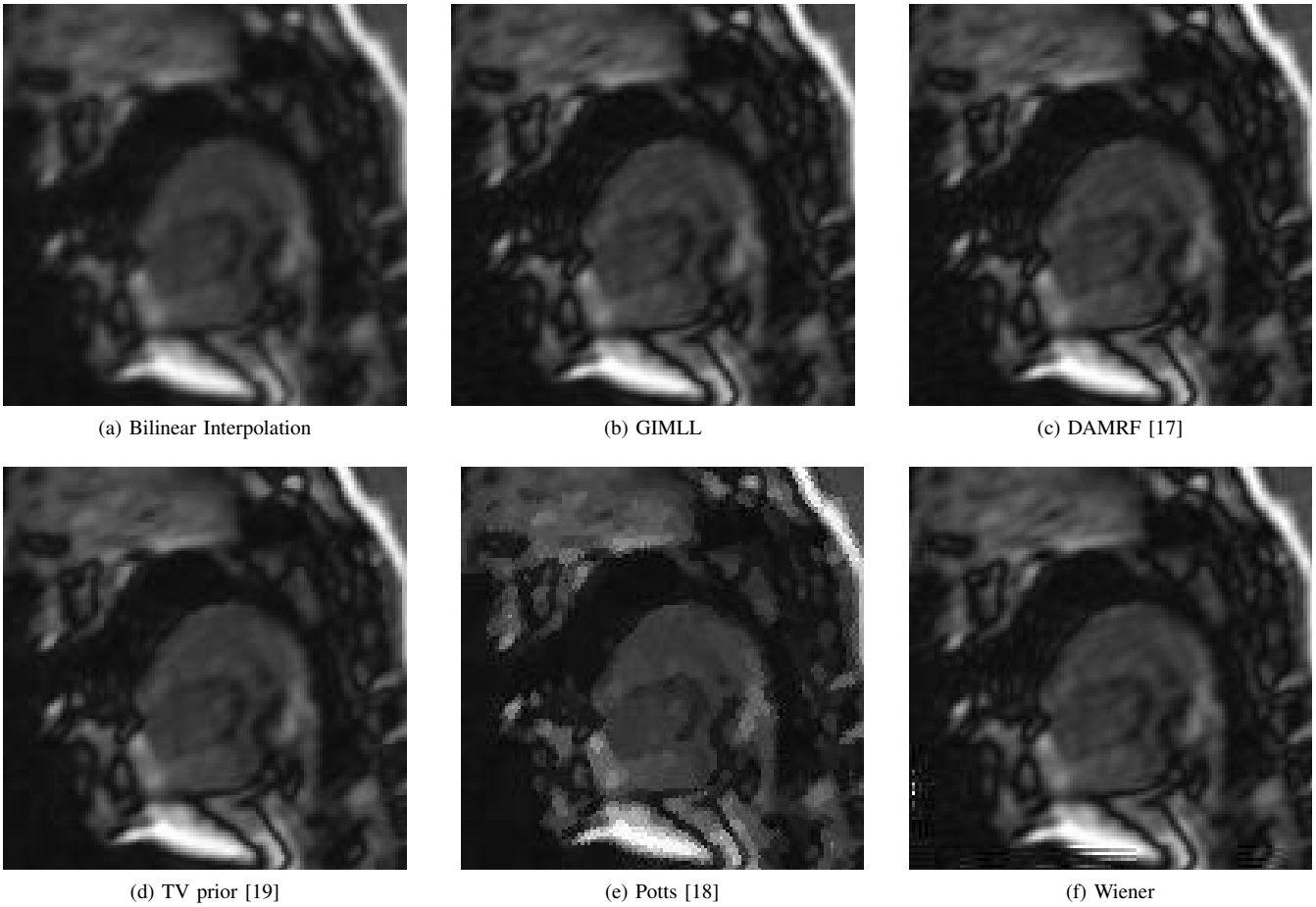
Note that the proposed method consistently outperforms the other methods. In order to check whether the differences among the NMSE means were statistically significant, we decided to run an analysis of variance (ANOVA) test. Since the distributions turned out to be non-normal (verified by running a Shapiro test), we used the Kruskal-Wallis non-parametric ANOVA method. At 95% confidence level, the test does indicate a difference among the means (χ^2 = 26.3157, df = 5, p-value = 0.0000775). A Wilcoxon signed rank paired test also shows a significant difference between the best two approaches (p-value = 0.01587). This is an evidence of the effectiveness of the Wiener based SRIR method in the reconstruction of high resolution images of the vocal tract.

V. RELATED WORK

To the best of our knowledge, Martins et al. [5] were the first researchers to develop a method for spatio-temporal resolution enhancement of vocal tract images used in the context of speech production. However, they adopted a MAP-MRF SRIR methodology and used the ICM algorithm to sequentially update high resolution pixel intensities. This algorithm is known to have a very fast convergence rate (it converges in five iterations at most). However, considering 128×128 low resolution images and a scale factor of 2, the SRIR methodology will generate a 256×256 high resolution image. Consequently, the ICM algorithm will maximize the local conditional probability for 65536 pixels in each iteration. Such method can be considered costly when compared to our Wiener based approach, since it requires only a single iteration.

Mascarenhas et al. [7] discuss an statistical interpolator of the SPOT multispectral bands, performing the local linear estimation of the interpolated pixels under the MMSE criterion. Similar to the approach presented in this paper, the separability of the spatial correlation structure and the first order Markovian model were also adopted. However, Mascarenhas et al. used the correlation model to characterize the observations, not the image to be reconstructed.

Hardie [13] proposes an SRIR algorithm that uses a type of adaptive Wiener filter, combining nonuniform interpolation



(a) Bilinear Interpolation (b) GIMLL (c) DAMRF [17]

(d) TV prior [19] (e) Potts [18] (f) Wiener

Fig. 8. Images reconstructed by (a) the GIMLL SRIR approach, (b) the DAMRF approach, (c) the TV prior, (d) the Potts method and finally (e) the proposed Wiener filter based approach.

and restoration into a single weighted sum operation. The proposed algorithm adopts an isotropic a priori model, while our method is based on the first order Markovian spatial correlation structure. Moreover, differently from our approach, Hardie focuses on scenes that remain static along the acquisition of frames, except for global relative motion between the scene and sensor. This is not the case in our context, since there are deformations localized around the jaw along the image sequences.

VI. Concluding Remarks

In this paper, we presented an evolution of the spatio-temporal resolution enhancement approach proposed by Martins et al. [5]. A novel Wiener filter based SRIR method is used as a more efficient alternative to enhance spatial resolution. We assumed the separability of the spatial correlation structure in the horizontal and vertical directions using the first order Markovian model in each direction. The performed experiments visually and numerically demonstrated the effectiveness of our approach in the context of the vocal tract MRI image sequences.

There are computational savings associated with estimating more than one high resolution pixel in each step of the algorithm. In this paper we only discuss the single pixel estimation. However, preliminary experiments indicate that the estimation of multiple high resolution pixels is feasible. In future work we plan to apply such procedure.

Acknowledgment

We would like to thank professors Antnio Teixeira and Augusto Silva from Instituto de Engenharia Eletrnica e Telemtica de Aveiro (IEETA) of Universidade de Aveiro, Portugal, for the vocal tract images used in this work. These images are part of the HERON Project - A Framework for Portuguese Articulatory Synthesis Research, POSI/PLP/57680/2004. Ana L. D. Martins is supported by FAPESP, Brazil, under grant number 2008/01348-2.

References

[1] E. Bresch and S. Narayanan, "Region Segmentation in the Frequency Domain Applied to Upper Airway Real-Time Magnetic Resonance Images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3, pp. 323–338, March 2009.

[2] T. Baer, J. Gore, L. C. Gracco, and P. W. Nye, "Analysis of vocal tract shape and dimensions using Magnetic Resonance Imaging: Vowels," *Journal of the Acoustic Society of America (JASA)*, vol. 90, no. 2, pp. 799–828, 1991.

[3] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *Journal of the Acoustical Society of America*, vol. 115, p. 17711776, 2004.

[4] E. Bresch, Y.-C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging [Exploratory DSP]," *Signal Processing Magazine, IEEE*, vol. 25, no. 3, pp. 123–132, May 2008.

[5] A. L. D. Martins, N. D. A. Mascarenhas, and C. A. T. Suazo, "Spatio-Temporal Resolution Enhancement of Vocal Tract MRI Sequences Based on Image Registration," *Integrated Computer-Aided Engineering*, vol. 18, no. 3, pp. 143–155, 2011.

[6] D. Rueckert, L. I. Sodona, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[7] N. D. A. Mascarenhas, G. J. F. Banon, and A. L. B. Candeias, "Multispectral image data fusion under Bayesian approach," *International Journal of Remote Sensing*, vol. 17, no. 8, pp. 1457–1471, 1996.

[8] D. Rueckert and P. Aljabar, "Nonrigid Registration of Medical Images: Theory, Methods, and Applications," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 113–119, July 2010.

[9] R. Y. Tsai and T. S. Huang, "Multi-frame image restoration and registration," *Advances in Computer Vision and Image Processing*, pp. 317–339, 1984.

[10] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of Optical Society America A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[11] A. K. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video*. San Rafael, CA: Morgan & Claypool, 2007.

[12] R. R. Schultz and R. L. Stevenson, "Extraction of highresolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.

[13] R. A. Hardie, "A fast image super-resolution algorithm using an adaptive Wiener flter," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2953–2964, 2007.

[14] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[15] A. Papoulis, *Probability, Random Variables ans Stochastic Processes*, 2nd ed. McGraw-Hill, 1984.

[16] W. K. Pratt, *Digital Image Processing: PIKS Scientific Inside*. Wiley-Interscience, 2007.

[17] K. V. Suresh and A. N. Rajagopalan, "Robust and computationally efficient superresolution algorithm," *Journal Optical Society of America A*, vol. 24, no. 4, pp. 984–992, April 2007.

[18] A. L. D. Martins, M. R. P. Homem, and N. D. A. Mascarenhas, "Super-resolution image reconstruction using the ICM algorithm," in *Proc. of the IEEE International Conference on Image Processing*. San Antonio, Texas: IEEE Computer Society, 2007, pp. IV 205–IV 208.

[19] S. Z. Li, *Markov Random Field Modeling in Image Analysis (Advances in Pattern Recognition)*, 3rd ed. Springer, March 2009.