ULTRASOUND TOMOGRAPHY IMAGE RECONSTRUCTION WITH LEARNED DICTIONARIES

Ivana Tošić [*], *Ivana Jovanović* [†], *Pascal Frossard* [*], *Martin Vetterli* [†‡] *and Neb Duric* [§]

[*] Signal Processing Laboratory (LTS4)
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[†] Audiovisual Communications Laboratory (LCAV)
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[‡] Department of Electrical Engineering and Computer Sciences
University of California at Berkeley (UCB), Berkeley, CA 94720, USA

[§] Karmanos Cancer Institute, 4100 John R. Street, WCCC, Detroit, USA

email : {ivana.tosic, ivana.jovanovic, pascal.frossard, martin.vetterli}@epfl.ch, duric@karmanos.org

ABSTRACT

We propose a new method for reconstruction of breast images from measurements obtained by ultrasound tomography (UT) scanners. Our solution for this inverse problem is based on sparse image representation in an overcomplete dictionary that is adapted to the properties of UT images. This dictionary is learned from high resolution MRI breast scans using an unsupervised dictionary learning method described in Ref. [1]. The proposed dictionary-based regularization method significantly improves the quality of reconstructed breast UT images. It outperforms the wavelet-based reconstruction and the $l_2+lowpass$ minimization algorithm, on both numerical and *in vivo* data. Our results demonstrate that the use of the learned dictionary improves the image accuracy for up to 4 dB with the exact measurement matrix and for 3.5 dB with the estimated measurement matrix over the wavelet-based reconstruction under the same setup.

Index Terms— ultrasound tomography, ray-based reconstruction, sparse representation, dictionary learning

1. INTRODUCTION

A variety of applications exploit the fact that transmission of sound is affected by the medium through which it passes. It provides information concerning the properties of the medium, including inhomogeneities in ultrasound speed and attenuation. Different methods have been developed over the past decades in order to recover and image these properties. An emerging application of ultrasound tomography (UT) is breast screening for cancer detection, which provides a much cheaper imaging solution than MRI. This study is motivated by the fact that MR and sound-speed images depict similar structures in the breast [2]. Although these two imaging modalities rely on totally different physical principles (magnetism versus acoustics) they trace similar structures because both water content (measured by MRI) and sound speed (measured by UT) increase with tissue density [3]. The high degree of spatial correlation of MR and sound speed images is therefore largely driven by similar sensitivity to changes in tissue density. However, UT imaging is a highly

ill-conditioned problem, which requires a proper regularization approach to assure a reliable and accurate reconstruction.

Assumption that an image has a sparse representation in an overcomplete dictionary can be used as an efficient regularization constraint for image reconstruction from ill-conditioned systems. Moreover, according to the compressed sensing theory it is possible to reconstruct a sparse signal from a reduced set of measurements [4]. These principles have been applied to MRI image reconstruction, assuming that the image is sparse in the wavelet domain [5]. However, MRI images are characterized by specific properties that substantially differ from natural images, so the wavelet-based representation might not be optimal. Indeed, the efficiency of sparse representations depends largely on the choice of the basis or the overcomplete dictionary. For UT or MRI images, an interesting approach to dictionary optimization is to learn it from a large database of signals from the same class. To the best of our knowledge, such learned dictionaries have never been applied to the reconstruction of UT images.

The contribution of this paper is a new sparsity-based method for reconstruction of UT breast images, using a learned dictionary of atoms adapted to the properties of UT images. We learn the dictionary from an MRI breast image database using the maximum likelihood dictionary learning method introduced in Ref. [1]. The UT image reconstruction problem is then formulated as a compressive sensing problem where the measurements represent the time of flight of the sound wave and the pixels correspond to the sound speed. The image is assumed to be sparse in the learned dictionary. We show that our approach outperforms the wavelet-based reconstruction for up to 4 dB with the exact measurement matrix, and for 3.5 dB with the estimated one. Our solution also permits the reconstruction from a smaller number of measurements, which can potentially increase the resolution of the UT images. Finally, we show that our reconstruction from *in vivo* measurements leads to better visual quality compared to state of the art approaches.

2. UT IMAGE RECONSTRUCTION BASED ON SPARSITY

The UT scanner that we consider consists of an array of 256 ultrasound emitters and 256 ultrasound receivers, uniformly spaced on a circle of radius 10 cm [6]. This circular setup surrounds a medium whose properties need to be imaged. In our case, we estimate the

speed of sound propagation inside this medium. The estimation of other parameters, such as the attenuation, can be performed similarly. The scanner works as follows. Each transmitter sends an ultrasound signal that is received by all receivers. The medium is divided into N tiles, each one representing a constant unknown sound speed $c_i, i = 1, ..., N$. The time that each signal takes to travel from an emitter to a receiver, hereafter referred to as time of flight t_{of}, is a function of the sound speed values c along the corresponding travel ray $\Gamma(c)$:

$$t_{of} = \int_{\Gamma(c)} \frac{dl}{c}. \tag{1}$$

Moreover, due to the inhomogeneities in the sound speed the rays of sound propagation are refracted, i.e., bent. Hence, the ray itself depends on the sound field via the Snell's law, which results in a non-linear relationship between the time of flights and the sound field values. An iterative scheme that alternates between finding the rays and setting up a linear system of equations to update the sound field for the next iteration needs to be applied. The stability, convergence and robustness of this iterative scheme are essential to the final reconstruction. Recently, a method to obtain bent rays for the UT setup has been proposed [7]. The stabilization is achieved by imposing the lowpass constraint on the intermediate sound speed solution, with the lowpass bandwidth determined from the number of available measurements and an oversampling factor. The output of the algorithm gives an estimate of the bent rays, which can be used to linearize the system of equations in Eq. (1) w.r.t. the 'slowness' $\mathbf{s} = [1/c_1 \, ...1/c_N]^\mathsf{T}$. The resulting linear system is:

$$\mathbf{t_{of}} = \mathbf{A} \cdot \mathbf{s}, \tag{2}$$

where $\mathbf{t_{of}}$ is the vector of measured t_{of} for each ray and $\mathbf{A}$ is an estimated bent ray measurement matrix. Additionally, we can linearize the equation above w.r.t. the sound speed using the first degree Taylor expansion in c_0, which results in:

$$2c_0 \mathbf{A} \cdot \mathbf{1} - c_0^2 \mathbf{t_{of}} = \mathbf{A} \cdot \mathbf{c}, \tag{3}$$

where $\mathbf{c} = [c_1 \, ...c_N]^\mathsf{T}$ and $\mathbf{1}$ is the vector of all ones. Since the sound speed reflects the tissue properties, the value of a pixel in a UT image corresponds to the sound speed value in the corresponding tile. Therefore, the solution of the system (3) is a ultrasound speed map of the scanned medium. However, this system turns to be very ill-conditioned. Different regularization methods have been developed over the past decades as a means to get stable and accurate image reconstruction from tomographic data. A common idea behind most of the methods is to use some *a priori* knowledge about the image model. For example, it may be known that in a certain basis or dictionary $\mathbf{\Phi}$ the images have a sparse representation. For a UT image $\mathbf{c}$, represented as $\mathbf{c} = \mathbf{\Phi a}$, this means that the l_0-norm of the coefficient vector $\mathbf{a}$ is small. However, finding the smallest l_0-norm solution to the signal representation problem is NP hard. Nevertheless, one can use convex relaxation and find the minimal l_1-norm solution by using convex optimization. Moreover, due to its sparsity, a given signal can be reconstructed from a number of measurements or samples that is below the Nyquist sampling rate. This concept is known as Compressed Sensing [4] and it is shown to be very useful for tomographic sampling in general [8]. Formally, the minimal l_1-norm signal reconstruction from measurements $\mathbf{y}$ is found as the solution of the following convex optimization problem:

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{A\Phi a}\|_2^2 + \lambda\|\mathbf{a}\|_1, \text{ where } \mathbf{y} = \mathbf{Ac} = \mathbf{A\Phi a} + \eta, \tag{4}$$

and $\mathbf{\Phi}$ is a matrix whose columns represent the dictionary atoms and η is the noise term. The vector $\mathbf{y}$ represents the vector of measurements, which in our UT problem formulation is $\mathbf{y} = 2c_0 \mathbf{A} \cdot \mathbf{1} - c_0^2 \mathbf{t_{of}}$.

Note that there is a trade-off parameter λ that can be chosen w.r.t. the amount of the desired sparsity in the final solution. The implicit assumption here is that our image $\mathbf{c}$ is sparse in a dictionary given by $\mathbf{\Phi}$. However, the choice of this dictionary is not obvious, especially for the UT images with statistics that significantly differ from natural images. In this paper, we choose to learn a dictionary in which our images have sparse representation.

3. DICTIONARY LEARNING

One of the first solutions to an unsupervised dictionary learning problem for natural images has been introduced by Olshausen and Field [1]. Their method is based on maximum likelihood estimation of the dictionary elements, given the sparse linear image model[1] $\mathbf{c} = \mathbf{\Phi a} + \eta$. The goal of the maximum likelihood learning is to find the overcomplete dictionary $\mathbf{\Phi}^*$ such that:

$$\mathbf{\Phi}^* = \arg\max_{\mathbf{\Phi}} \langle \max_{\mathbf{a}} \log P(\mathbf{c}|\mathbf{\Phi})\rangle, \text{ where } \mathbf{c} = \mathbf{\Phi a} + \eta. \tag{5}$$

When the coefficient vector $\mathbf{a}$ has a sparse structure, the prior distribution on the coefficients $P(\mathbf{a})$ is highly peaked at zero and heavy tailed. It can thus be modeled with a zero-mean Laplace distribution. In this case, we can formulate the optimization problem in Eq. (5) with the following energy minimization problem:

$$\mathbf{\Phi}^* = \arg\min_{\mathbf{\Phi}} \langle \min_{\mathbf{a}} E(\mathbf{c}, \mathbf{a}|\mathbf{\Phi})\rangle. \tag{6}$$

The energy function is:

$$E(\mathbf{c}, \mathbf{a}|\mathbf{\Phi}) = -\log P(\mathbf{c}|\mathbf{a}, \mathbf{\Phi})P(\mathbf{a}) = \|\mathbf{c} - \mathbf{\Phi a}\|_2^2 + \lambda\|\mathbf{a}\|_1, \tag{7}$$

where $\lambda = 2\sigma_N^2/\delta$, δ controls the steepness of the Laplace distribution and σ_N^2 is the variance of the noise η. A local minimum of the casted optimization problem can be found by iterating between two steps. In the first step, $\mathbf{\Phi}$ is kept constant and the energy function is minimized with respect to the coefficient vector $\mathbf{a}$. The second step keeps the obtained coefficients $\mathbf{a}$ constant, while performing the gradient descent on $\mathbf{\Phi}$ to minimize the energy $E(\mathbf{c}, \mathbf{a}|\mathbf{\Phi})$. Therefore, the algorithm iterates between the sparse decomposition and the dictionary learning steps until convergence, where the learned dictionary is determined. The dictionary is then applied to reconstruct any UT image using the CS based method described in Sec. 2.

4. EXPERIMENTAL RESULTS

In this section, we present the experimental results of dictionary learning and UT image reconstruction with the learned dictionary, and compare the reconstructions with standard wavelet bases.

To obtain the dictionary adapted to the statistical properties of medical breast images (MRI and UT), we have applied the maximum likelihood dictionary learning method on the CURE [6] database of MRI breast scans. The motivation behind using MRI instead of UT images for learning is that while they capture similar tissue structure, MRI provides images with higher resolution and less artifacts. Our learning included 41 3D-scans from 14 subjects, where each scan had different number of slices (from 36 to 160). Since we target the reconstruction of 2D images, each slice has been treated independently. We have learned a dictionary of 20 atom shapes shown in Fig. 1, where each shape can be shifted anywhere in the image.

We first show the denoising performance of sparse reconstructions using the obtained dictionary and compare to the wavelet basis

---

[1]We use here the notation corresponding to UT imaging.

(Daubechies wavelets of order 5), on a test MRI image in Fig. 2(a). The test image was not included in the dictionary training set. Sparse reconstruction has been evaluated with the conjugate gradient minimization of the objective function in Eq. (7). Let us denote the sparse reconstruction with the learned dictionary as LDS, and with the wavelets as WS. Fig. 2(b) shows a noisy image of SNR=23 dB, while Fig. 2(c) and (d) show the reconstructions using respectively LDS and WS. We can see that LDS gives a PSNR improvement of 2 dB compared to WS. Moreover, LDS leads to a better visual quality, without the ringing artifacts inherent to orthogonal wavelets. This shows the effectiveness of the sparse representation using the dictionary learned on MRI images.
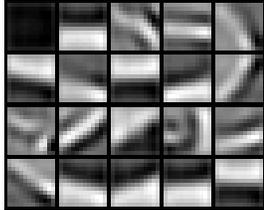
Fig. 1. Atoms in the learned dictionary. The whole dictionary Φ is formed by shifts of these atoms at all pixel positions.
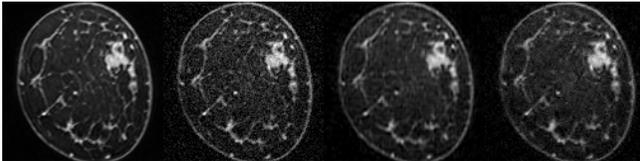
(a) original (c) noisy, 23dB (c) LDS, 28.1dB (d) WS, 26.2dB

Fig. 2. 128x128 pixel MRI image reconstruction with AWGN noise (CURE 191 patient), $SNR = 23dB, \lambda = 0.1$.

We further demonstrate the application of the compressive sensing framework for UT image reconstruction. In this case LDS and WS refer to the optimization problem in Eq. (4). For evaluation purpose, we have formed a phantom UT image that mimics breast structure seen in the MRI breast scan of the CURE 191 patient. Fig. 3(a) shows the phantom that is constructed by adapting the MRI scan to UT image properties (i.e., contrast and range adjustment and low pass filtering). We have simulated the screening of the phantom with the ultrasound wave propagation model for two cases: with the correct bent rays in the measurement matrix $\mathbf{A}$, and with the estimated bent rays in $\mathbf{A}$. In the case of correct rays, we have collected the total of 13448 measurements. However, due to the sparsity of the images in the learned dictionary and wavelet domains, not all measurements are necessary to reconstruct the MRI image with a certain quality. This is shown in Fig. 3(d), where we have plotted the PSNR of the reconstructed UT phantom with LDS and WS, for different number of measurements. The reconstructions with 50% of the measurements are given in Fig. 3(b) and (c) for LDS and WT, respectively. We can see that LDS outperforms WS, both quantitatively and visually, especially in the area of the salient image information (i.e., tumor). When the bent rays are estimated with a certain error, the reconstructed image quality decreases for both approaches. LDS and WS reconstructions of the phantom are shown in Fig. 4(b) and (c),

respectively. We have also shown in Fig. 4(a) the reconstruction result using the algorithm that evaluates image pixels directly from the measurements $\mathbf{y}$ and matrix $\mathbf{A}$. It performs l_2 minimization using the conjugate gradient (CG) and low pass filtering (LP) [8]. Again, we can see that LDS outperforms other methods in PSNR. More importantly, LDS leads to a cleaner and smoother reconstruction, while still preserving the tissue structure of the phantom.

Finally, we have applied the proposed UT image reconstruction method on *in vivo* dataset. CURE 191 patient was screened with the UT scanner prototype [6], collecting the total of 30314 measurements. Since we do not have the ground truth for the *in vivo* measurements, we show only the visual results of CG+LP, LDS and WS reconstructions in Fig. 5(a), (b) and (c), respectively. Since the algorithms based on sparsity do not necessitate all measurements, we use only 50% measurements for LDS and WS. Even with one half of the measurements, LDS gives more visually pleasant reconstruction than CG+LP, while keeping the important image features. WS also preserves the image structure, but it introduces high frequency noise that is usually not present in UT tissue images.

5. CONCLUSION

We have proposed a new UT image reconstruction method based on the image sparsity prior. In order to correctly represent the tissue images, we have learned an overcomplete dictionary from a large set of MRI breast tissue scans. We have found that the use of the learned dictionary in the image reconstruction outperforms the wavelet-based method, for both numerical and *in vivo* studies. This highlights the benefits of using learned dictionaries for representation and reconstruction of MRI and UT images, which are characterized by significantly different statistical properties than natural images. It also opens many perspectives for future work, particularly in 3D ultrasound tomography image reconstruction.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[2] B. Ranger, P. Littrup, N. Duric, C. Li, J. Lupinacci, L. Myc, O. Rama and L. Bey-Knight, "Breast imaging with acoustic tomography: a comparative study with MRI." *SPIE Medical Imaging*, 2009.

[3] T.D. Mast, "Empirical relationships between acoustic parameters in human soft tissues." *Acoustics Research Letters Online*, vol. 1, no. 2, pp. 37–42, 2000.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 – 1306, 2006.

[5] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182 – 1195, 2007.

[6] N. Duric, P. Littrup, L. Poulo, A. Babkin, E. Holsapple, O. Rama, and C. Glide, "Detection of breast cancer with ultrasound tomography: First results with the computed ultrasound risk evaluation (CURE) prototype," *Medical Physics*, vol. 2, no. 34, pp. 773–785, 2007.

[7] A. Hormati, I. Jovanovic, O. Roy and M. Vetterli, "Robust ray-based reconstruction in ultrasound tomography," *SPIE Medical Imaging*, 2010.

[8] I. Jovanovic, *Inverse Problems in Acoustic Tomography: Theory and Applications*, Ph.D. thesis, EPFL, Lausanne, Switzerland, 2008.

(a) original   **(b) PSNR=28.99 dB**   **(c) PSNR=26.42 dB**   **(d)**
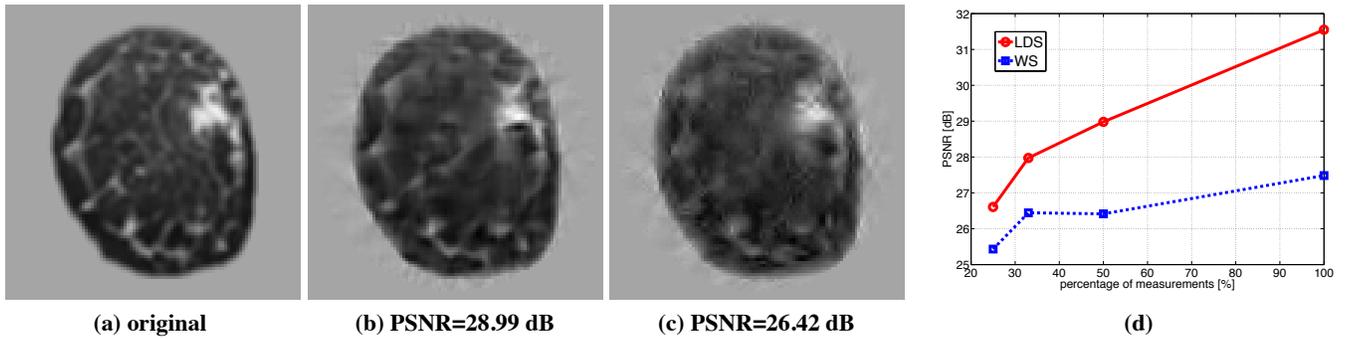
Fig. 3. Reconstruction of a numerical phantom that mimics breast structure seen from an MRI image of CURE 191 patient, using the correct bent rays in the measurement matrix: a) Original phantom b) LDS, 50% of the measurements; c) WS, 50% of the measurements, d) PSNR versus the percentage of measurements.

(a) PSNR=27.52 dB   **(b) PSNR=27.93 dB**   **(c) PSNR=24.45 dB**
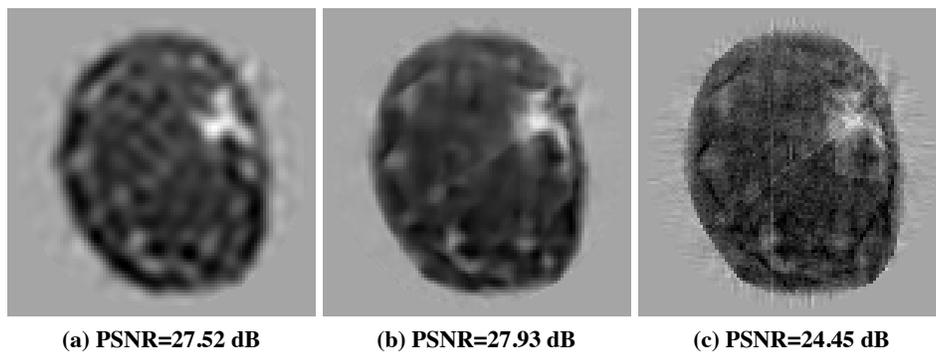
Fig. 4. Reconstruction of a numerical phantom that mimics breast structure seen from an MRI image of CURE 191 patient, using an estimated bent ray as the measurement matrix: a) CG+LP; c) LDS; d) WS.

(a)   **(b)**   **(c)**
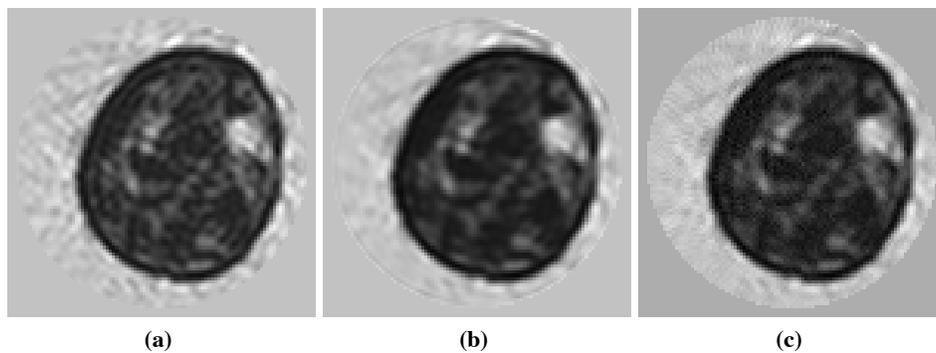
Fig. 5. Reconstruction from in vivo ultrasound tomography measurements for the CURE 191 patient: a) CG+LP; b) LDS, 50% of the measurements; c) WS, 50% of the measurements.