FUSION OF ULTRASOUND HARMONIC IMAGING WITH CLUTTER REMOVAL USING SPARSE SIGNAL SEPARATION

Javier S. Turek Jeremias Sulam Michael Elad Irad Yavneh

Department of Computer Science, Technion, Israel Institute of Technology, Haifa 32000, Israel
{javiert,jsulam,elad,irad}@cs.technion.ac.il

ABSTRACT
In ultrasound, second harmonic imaging is usually preferred due to the higher clutter artifacts and speckle noise common in the first harmonic image. Typical ultrasound use either one or the other image, applying corresponding filters for each case. In this work we propose a method based on a joint sparsity model that fuses the first and second harmonic images while performing clutter mitigation and noise reduction. Our approach, Fused Morphological Component Analysis (FMCA), uses two adaptive dictionaries for characterizing the clutter components in each image, and a common dictionary for the tissue representation. Our results indicate that the obtained images contain less clutter artifacts, less speckle noise and as such enjoy of the benefits of both harmonic input images.

Index Terms—Artifact reduction, signal separation, harmonic imaging, image fusion, morphological component analysis

1. INTRODUCTION
Ultrasound imaging is one of the most important medical imaging modalities, performed by acquiring the reflected ultrasound waves from the inner tissues. Although this modality is able to provide portable, online imaging with no radiation, it also suffers from several limitations, such as low resolution, low signal to noise ratio due to the multiplicative speckle noise, and other artifacts. In echocardiography in particular, the use of tissue harmonic imaging has been shown to improve the image quality [1]. Acquiring the second or higher harmonics instead of the reflected signals in the fundamental frequency enables suppressed side and grating lobes and minimal harmonic content in reverberant echoes. These contribute to a cleaner image with less artifacts and higher resolution. However, this comes at the cost of lower penetration depth and intensity [1, 2].

An important class of artifact in ultrasound imaging is that of clutter, which appears mainly as a quasi-static cloud of echo signals [3]. This problem is most severe in patients for whom the acquisition conditions are difficult [4]. Such artifacts obscure tissue areas and the resulting image has poor contrast and reduced readability. Therefore, it can mislead to corrupt diagnostic information like in myocardium strain evaluation [5], tracking techniques for functioning diagnosis [6], or visualization of cardiac abnormalities [7]. Cluttering is most prominent in fundamental frequency imaging, requiring aggressive filtering techniques, while it is usually reduced using harmonic imaging. Nevertheless, in some cases clutter is still present in the second harmonic image [8], requiring additional processing.

Some typical schemes for clutter mitigation involve linear filtering and transformation onto unitary bases such as the Discrete Fourier Transform [9] and Wavelets [10]. More effective methods are based on adaptive bases learned from the echo data, such as Principal Component Analysis (PCA) [11]. A different approach is that of Morphological Component Analysis (MCA), as suggested in [12] and [13], where the separation of the signals is done by means of an adaptive redundant non-orthonormal basis or dictionary.

Image fusion in medical imaging is an approach that seeks to combine the benefits of different medical imaging modalities to enhance salient features [14]. In ultrasound imaging in particular, one can consider fusion of several harmonic imaging bands. Indeed, the work reported in [15] considers such fusion directed towards superresolution.

In the last decade, sparsity related ideas have had a growing impact on image processing applications [16]. In this work we propose a method of fusing the first and second harmonic images while performing clutter mitigation based on joint sparse representations. We propose a model that enforces joint sparsity on the tissue component of both fundamental and second harmonic images, while allowing the removal of clutter from both of them simultaneously. As described later, this yields a cleaner image in terms of clutter and speckle noise removal, while providing a fused image which enjoys the benefits of both ultrasound harmonics.
2. SPARSE REPRESENTATIONS OF SIGNALS

A signal \( t \in \mathbb{C}^n \) is said to have a sparse representation over a known overcomplete dictionary \( D \in \mathbb{C}^{n \times m} \) if there exists a sparse vector \( x \in \mathbb{C}^m \), such that \( t = Dx \) and \( \|x\|_0 = k \ll n \). Here, \( \| \cdot \|_0 \) denotes the \( \ell_0 \) quasi-norm, which is the number of non-zero elements in a vector. In practice, an observed signal \( s \) is obtained by measuring a signal of interest \( t \) contaminated with noise \( v \), which is often assumed to be additive WGN with standard deviation \( \sigma \), i.e., \( s = t + v \). The objective is to recover the signal \( t \) from the noisy observation \( s \). Assuming a sparse representation prior on the signal \( t \), one can redefine the observed signal as \( s = Dx + v \). Then, \( x \) is computed by solving the following optimization problem:

\[
\min_x \|s - Dx\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq k. \tag{1}
\]

where \( k \) is the maximum number of non-zeros allowed in the representation. Once the sparse vector \( \hat{x} \) has been obtained, the estimate of the clean is computed by multiplying the dictionary \( D \) by \( \hat{x} \). Several methods exist to compute an approximate solution \( \hat{x} \) to Problem (1), one being the greedy pursuit algorithm called Orthogonal Matching Pursuit (OMP) [17]. An alternative is to relax the \( \ell_0 \)-norm with the \( \ell_1 \)-norm, obtaining the convex Basis Pursuit problem [18] which can be solved using standard optimization algorithms. The OMP has proven to be an affective compromise between accuracy and computational cost [19] and it will be used in our work when solving Problem (1).

The choice of the dictionary \( D \) is of great importance in obtaining the sparse vector \( x \). Though computationally efficient, dictionaries that are mathematically pre-defined are limited in terms of approximation performance [19]. Alternatively, the dictionary can be learned adaptively from the data to yield better results. Amongst these, the K-SVD [20] algorithm is a commonly used method for dictionary learning, which has proven to be effective in several image processing applications [16].

3. HARMONIC IMAGING FUSION WITH SPARSE SIGNAL SEPARATION

The echo waves correspond to a linear phenomena. Hence, and following our previous work in [12, 13], we assume that the acquired signal is the linear combination of echoes reflected by tissue, reverberation artifacts (clutter) and measurement noise:

\[
s = t + c + v, \tag{2}
\]

where \( t \) is the tissue signal and \( c \) represents the clutter artifacts. In order to construct the training signals we divide the data into small overlapping patches and apply the separation and fusion scheme to these small patches. We take each sample signal \( s^i \) as a two-dimensional patch in the axial and temporal dimensions\(^1\) from the observed signal \( s \), enabling the study of the temporal characteristics (static/moving) of a patch. That is, a signal \( s^i \) is a vectorized version of a two dimensional patch of \( M \) elements across the axial direction and \( N \) elements (frames) in the temporal direction. The size \( N \) influences the amount of motion that is captured by the patches\(^2\). The assumptions above about the signal model remain true also for each patch, and hence Eq. (4) is applicable to the patches \( s^i \):

\[
s^i = t^i + c^i + v^i. \tag{3}
\]

Assuming that the tissue and the clutter signals have sparse representations \( x^t_i \) and \( x^c_i \), under dictionaries \( D_t \) and \( D_c \), respectively, we may rewrite (3) as

\[
s^i = D_t x^t_i + D_c x^c_i + v^i = [D_t | D_c] \begin{bmatrix} x^t_i \\ x^c_i \end{bmatrix} + v^i = Dx^i + v^i, \tag{4}
\]

where \( D \) is a dictionary constructed by concatenating both dictionaries \( D_t \) and \( D_c \), and \( x^i \) represents the concatenation of the sparse representations of the tissue and clutter components. When a signal can be decomposed into two or more separate components having a sparse representation under different dictionaries, they can be separated through the scheme in Eq. (4), which is known as Morphological Component Analysis (MCA) [21]. Hence, the clutter component can be filtered by obtaining \( x^i \) and then removing the coefficients corresponding to the clutter as

\[
\hat{s}^i = s^i - D_c x^c_i = D_t x^t_i + v^i, \tag{5}
\]

where \( \hat{s}^i \) is the clutter-filtered version of the patch vector \( s^i \). Note that a different alternative would be to reconstruct \( \hat{s}^i \) considering just the tissue component. This way, however, some small and fast moving structures might be confused with noise and would be removed. Thus, we employ Eq. (5).

\(^1\)The echo data can be taken as a three dimensional element, by adding consecutive axial lines to every signal \( s^i \) and thus including information in the lateral direction.

\(^2\)For a given time lapse, the faster the frame rate the bigger the size \( N \) needs to be in order to capture the same amount of information.
While this signal model has proven useful in previous works [12, 13], we extend the model in Eq. (4) into a jointly sparse morphological analysis by considering signals \( s_1 \) and \( s_2 \) corresponding to the first and second harmonic images, respectively, and their corresponding aligned patches \( s_1^i \) and \( s_2^i \). Using the fact that the tissue component in both images has the same source [15], we can enforce these components to have the same sparse representation under the tissue dictionary \( D_t \). On the other hand, the clutter components will differ due to the different phenomena generating these artifacts in each frequency range [1]. This motivates us to propose the following signal model:

\[
\begin{bmatrix}
  s_1^i \\
  \cdots \\
  s_2^i
\end{bmatrix} = \begin{bmatrix}
  D_{c1} & 0 \\
  \cdots \\
  D_{c2}
\end{bmatrix} \begin{bmatrix}
  x_1^i \\
  \cdots \\
  x_2^i
\end{bmatrix} + v^i,
\]

where the expression in the left-hand side accounts for the concatenation of the signal vectors, while the expression in the left-hand side accounts for the real data instead. We follow a similar approach to that of [13], on the characteristics of the components, or may be trained on

The amount of clutter removed from an echo frame is measured by the Spatial Standard Deviation (SSD). The SSD is computed for each frame by averaging the CNR across frames.

In this section we demonstrate our method using two data sets of In-phase Quadrature (IQ) echo data acquired from a parasternal long axis view of two adult male volunteers. The sequences contain a full heart cycle (40–50 frames) and were acquired using a Vivid S6 (GE Medical Systems, Israel) ultrasound scanner transmitting at a frequency of 1.7 MHz. The fundamental and the second harmonic signals are received simultaneously at 1.7 and 3.4 MHz respectively, and separated later with a bandwidth of 0.7 MHz for the first harmonic and 1 MHz for the second harmonic. The images are of reduced quality due to limitations of the simultaneous acquisition of both harmonic signals. Clutter artifacts are present due to multi-path reverberations, mainly from the thoracic cage and sternum. Our implementation uses full-overlapping 15 × 15 patches, with clutter dictionaries of 248 atoms each, and a tissue dictionary of 900 atoms (before the cleaning stage in Eq. (8) with \( \epsilon = 0.2 \)). We used 10 iterations for K-SVD.

The amount of clutter removed from an echo frame is quantified by the Contrast-to-Noise-Ratio (CNR) [11] measure:

\[
\text{CNR} = 20 \log_{10} \left( \frac{\hat{\mu}_i - \mu_0}{\sigma_0} \right),
\]

where \( \hat{\mu}_i \) and \( \mu_0 \) are the mean envelope-detected quantities in regions with and without clutter artifacts (marked with magenta and blue rectangles in Fig. 2a, respectively), and \( \sigma_0 \) is the standard deviation in the clutter-empty region. Then, the mean CNR performance for an entire sequence is obtained by averaging the CNR across frames.

In order to quantify the reduction of speckle noise we measure the Spatial Standard Deviation (SSD) within a homogeneous region chosen in the interior of the ventricle where no tissue is to be expected. The SSD is computed for each frame and then averaged along the time direction.

In Fig. 2 we present the two harmonic images taken from dataset 1 ((a) and (b), where the clutter is marked with a red background).
Fig. 2: Examples of images from the first (a) and second (b) harmonics, their corresponding filtered versions \( \hat{s}_1 \) and \( \hat{s}_2 \) ((c) and (d), respectively), and the FMCA output image \( \hat{s}_f \) (e).

Table 1: Mean CNR results over two datasets [dB].

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orig. 1st Harm.</td>
<td>-2.84</td>
</tr>
<tr>
<td>Orig. 2nd Harm.</td>
<td>0.01</td>
</tr>
<tr>
<td>TA-MCA 1st Harm.</td>
<td>1.23</td>
</tr>
<tr>
<td>TA-MCA 2nd Harm.</td>
<td>2.46</td>
</tr>
<tr>
<td>TA-MCA Compounded</td>
<td>1.57</td>
</tr>
<tr>
<td>TA-MCA Improvement</td>
<td>3.87</td>
</tr>
<tr>
<td>FMCA 1st Harm.</td>
<td>1.87</td>
</tr>
<tr>
<td>FMCA 2nd Harm.</td>
<td>2.39</td>
</tr>
<tr>
<td>FMCA Compounded</td>
<td>2.24</td>
</tr>
<tr>
<td>FMCA Improvement</td>
<td>4.54</td>
</tr>
</tbody>
</table>

Table 2: Mean SSD results over two regions in each dataset.

<table>
<thead>
<tr>
<th>Region</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orig. 1st Harm.</td>
<td>53.37</td>
<td>44.20</td>
</tr>
<tr>
<td>Orig. 2nd Harm.</td>
<td>23.02</td>
<td>19.84</td>
</tr>
<tr>
<td>TA-MCA 1st Harm.</td>
<td>42.16</td>
<td>26.46</td>
</tr>
<tr>
<td>TA-MCA 2nd Harm.</td>
<td>19.86</td>
<td>13.55</td>
</tr>
<tr>
<td>TA-MCA Compounded</td>
<td>23.18</td>
<td>14.91</td>
</tr>
<tr>
<td>TA-MCA Improvement</td>
<td>23.25</td>
<td>13.99</td>
</tr>
<tr>
<td>FMCA 1st Harm.</td>
<td>18.05</td>
<td>12.43</td>
</tr>
<tr>
<td>FMCA 2nd Harm.</td>
<td>15.58</td>
<td>10.48</td>
</tr>
<tr>
<td>FMCA Compounded</td>
<td>15.58</td>
<td>10.48</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

We have presented a method that exploits the information shared by the first and second harmonic imaging by using a jointly sparse model. Our method makes use of dictionaries that allow to perform clutter filtering through morphological component analysis simultaneously in both harmonics and fuses the resulting images to reduce speckle noise while enhancing the tissue details. FMCA achieved better cluttering removal performance together with speckle noise reduction in the studied datasets. A broader validation would contribute to the understanding of the limitations of the method. Finally, a similar scheme may be beneficial in other medical imaging modalities.

6. ACKNOWLEDGMENTS

The authors would like to thank G.E. Medical Systems, Israel for their kind help and technical support.
7. REFERENCES


