OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author’s version published in:
http://oatao.univ-toulouse.fr/26432

Official URL
https://doi.org/10.1109/ICIP40778.2020.9191013


Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr
ULTRASOUND AND MAGNETIC RESONANCE IMAGE FUSION USING A PATCH-WISE POLYNOMIAL MODEL

O. El Mansouri(1),(2)    A. Basarab(2)    M. A. T. Figueiredo(3)    D. Kouamé (2)    J.-Y. Tounneret(1)

(1) University of Toulouse, IRIT/INP-ENSEEIHT/TéSA, 31071 Toulouse Cedex 7, France
(2) University of Toulouse, IRIT, CNRS UMR 5505, Université Paul Sabatier, Toulouse, France
(3) Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal

ABSTRACT
This paper introduces a novel algorithm for the fusion of magnetic resonance and ultrasound images, based on a patch-wise polynomial model relating the gray levels of the two imaging systems (called modalities). Starting from observation models adapted to each modality and exploiting a patch-wise polynomial model, the fusion problem is expressed as the minimization of a cost function including two data fidelity terms and two regularizations. This minimization is performed using a PALM-based algorithm, giving its ability to handle non-linear and possibly non-convex functions. The efficiency of the proposed method is evaluated on phantom data. The resulting fused image is shown to contain complementary information from both magnetic resonance (MR) and ultrasound (US) images, i.e., with a good contrast (as for the MR image) and a good spatial resolution (as for the US image).

Index Terms— Image fusion, magnetic resonance imaging, ultrasound imaging, super-resolution, despeckling, image enhancement, patch-based method.

1. INTRODUCTION
Magnetic resonance (MR) and ultrasound (US) images have been used intensively in many clinical diagnosis and guided surgery applications. While they both carry important information in assessing the condition of organs, they exploit different physical phenomena and thus have their own advantages and limitations. In particular, US imaging offers a good spatial resolution and high frame rate compared to MRI, at the cost of a very low signal to noise ratio (SNR), low contrast (depending on the central frequency of the probe), a presence of speckle noise and a reduced field of view. In contrast, MRI enables a wide field of view, with a good SNR, high contrast, but relatively low spatial resolution [1]. As a consequence of these complementary properties, MR and US images are commonly used jointly in various clinical applications. The objective of this paper is to propose a method to fuse the two images in a single image in order to improve the diagnosis capacity of each modality.

Image fusion refers to assembling all the important information from multiple images and including them in fewer images or into a single image. Its purpose is not only to reduce the amount of data but also to build enhanced images that are more comprehensible and informative for human and machine insight [2]. Fusion of medical images is becoming very common for the study of a given pathology [3–5], and generally allows for a better medical decision in clinical studies. Medical images that are commonly fused include CT scans and positron emission tomography [6], or gammagraphy and US images [7]. However, to the best of our knowledge, the fusion of MR and US images, which is the purpose of this work, has been less addressed in the existing literature.

In our previous work on MR and US image fusion [8], we introduced a new algorithm performing both super-resolution of the MR image and despeckling of the US image. That algorithm was based on a polynomial function relating the US and MR images, accounting for the discrepancy between these two modalities. The coefficients of this polynomial were pre-estimated from the observed images. This paper further improves the polynomial relation between the two images by estimating the polynomial coefficients patch-wise, thus allowing for a better matching between the two images to be fused. Note that a similar idea was used in [9] for MRI images.

The paper is organized as follows. Section 2 presents the observation models, the patch-based polynomial function relating the US and MR images, and the optimization problem considered to fuse these images. The algorithm proposed to solve the fusion problem is detailed in Section 3. Simulation results are presented in Section 4. Conclusions and perspectives are finally reported in Section 5.

2. MAGNETIC RESONANCE AND ULTRASOUND IMAGE FUSION

2.1. Observation models
Denote as \(y_{mr} \in \mathbb{R}^M\) and \(y_{us} \in \mathbb{R}^N\) the registered MR and US images, with \(M\) and \(N\) the number of pixels in each

The authors would like to thank Fabien Vidal for providing the ultrasound and magnetic resonance data, as well as for the fruitful discussions about the clinical pertinence of the proposed algorithm.
image\(^1\). This section introduces two observation models accounting for the low spatial resolution of MR images and the low SNR of US images. The low resolution of the MR image is modeled by a downsampling operation and a low pass filter [11], while an additive noise model is considered for the US B-mode image. Note that speckle is assumed to be a multiplicative noise, leading to additive perturbations when applying log-compression, which is classically considered before forming B-mode images. Furthermore, this work assumes that the speckle noise affecting B-mode images is distributed according to a log-Rayleigh distribution, as in [12, 13]. The two resulting observation models are

\[ y_{\text{us}} = x_{\text{us}} + n_{\text{us}} \]
\[ y_{\text{mr}} = SHx_{\text{mr}} + n_{\text{mr}}, \]

where \( y_{\text{us}} \in \mathbb{R}^N \) is the observed B-mode US image, \( x_{\text{us}} \in \mathbb{R}^N \) is the noiseless US image, \( n_{\text{us}} \in \mathbb{R}^N \) is the log-Rayleigh speckle noise, \( x_{\text{mr}} \in \mathbb{R}^N \) is the high-resolution MR image, \( y_{\text{mr}} \in \mathbb{R}^M \) is the observed (low-resolution) MR image, and \( n_{\text{mr}} \in \mathbb{R}^N \) is an additive Gaussian noise. The matrix \( H \in \mathbb{R}^{N \times N} \) is the blurring matrix and \( S \in \mathbb{R}^{M \times N} \) (with \( N = d^2M \)) is a decimation operator with decimation factor \( d \). Note that the decimation factor is such that \( x_{\text{us}} \) and \( x_{\text{mr}} \) have the same spatial sampling.

### 2.2. Patch-based polynomial model

The patch-based polynomial model proposed in this work (relating the gray levels of MR and US images) is motivated by the fact that US images highlight the interfaces between different anatomical structures with different acoustic impedances [14]. More precisely, the US image is expressed as a function of the MR image and its spatial gradient is computed in the direction of US wave propagation

\[ x_{\text{us}} = f(x_{\text{mr}}, \nabla x_{\text{mr}}^H u), \]

where \( f : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^N \) is unknown and \( \nabla x_{\text{mr}}^H u \in \mathbb{R}^N \) contains in its \( i \)th line the inner product between the \( i \)th local gradient \( x_{\text{mr}} \) and the US scan direction \( u \).

The function \( f \) was represented by a global polynomial in our previous work on image fusion [15], and in [14] for multimodal image registration. However, the relationship between MR and US images may depend on tissue acoustic and magnetic properties, and thus may change from one image region to another. Thus, considering a global polynomial model may lead to inaccurate gray level matching in specific image regions. To overcome this issue, this paper introduces a more general patch-based polynomial model, fitting independently low-order polynomial functions to each overlapping patch extracted from MR and US images. This patch-based polynomial model is defined as

\[ P_p x_{\text{us}} = f_p(P_p x_{\text{mr}}, P_p \nabla x_{\text{mr}}^H u), \]  

where \( P_p \in \mathbb{R}^{N \times N} \) is a binary operator that extracts the \( p \)th patch of size \( n \) from an image of size \( N \). In the following, \( N_p \) will denote the total number of patches. Replacing \( f_p \) by a polynomial function, the relation between patches from the US and MR images becomes

\[ P_p x_{\text{us}} = \sum_{l+k \leq d_p} c_{l,k,p} (P_p x_{\text{mr}})^l \odot (P_p \nabla x_{\text{mr}}^H u)^k, \]

where \( p = 1, ..., N_p \) is the patch number, \( d_p \) and \( c_{l,k,p} \) are the order and the coefficients of the polynomial function \( f_p \) corresponding to patch \( \# p \), \( \odot \) is the Hadamard product (element by element multiplication) and the power operations applied to vectors are element-wise. In this paper, the final function \( f \) is obtained by averaging patch-wise polynomials, since each pixel of the image is contained in several overlapping patches. More precisely, the transformation of the \( i \)th pixel denoted as \( f_i : \mathbb{R}^N \rightarrow \mathbb{R} \) is the average of all the polynomials associated with the patches containing this pixel.

### 2.3. Cost function

Using the observation models in (1), the relationship between MR and US images defined in (3) and (4), and the ideas proposed in [15], this paper formulates image fusion as the following optimization problem:

\[
\hat{x} = \arg \min_x \frac{1}{2} \left[ \frac{\|y_{\text{us}} - SHx\|^2}{\text{MRI data fidelity}} + \tau_1 \|\nabla x\|^2 + \tau_2 \|\nabla f(x, \nabla x^H u)\|^2 \right] + \tau_3 \sum_{i=1}^{N} \left[ \exp(y_{\text{us},i} - f_i(x, \nabla x^H u)) - \lambda (y_{\text{us},i} - f_i(x, \nabla x^H u)) \right],
\]

where \( \hat{x} \) is the fused image, \( y_{\text{us},i} \) is the \( i \)th pixel of \( y_{\text{us}} \) and \( \tau_1, \tau_2, \tau_3 \) are hyperparameters balancing the weights of the MR and US data fidelity terms and regularizations. Note that, following [15], total variation was used to regularize the solution, thus promoting piecewise constant fused images both in the US and MR domains.

### 3. OPTIMIZATION

#### 3.1. PALM algorithm for MR and US image fusion

The cost function in (5) is non-convex because of the presence of the polynomial functions \( f \) and \( f_i \). Therefore, we investigate a solution based on the proximal alternating linearized minimization (PALM) algorithm [16]. In order to fit the general form of this algorithm, we propose the following parametrization:
\[ l(x) = \frac{1}{2} \| y_{\text{mr}} - SHx \|^2 + \tau_1 \| \nabla x \|^2 , \]
\[ g(v) = \tau_2 \sum_i \left[ \exp(y_{\text{mr}} - v_i) - \gamma(y_{\text{mr}} - v_i) \right] + \tau_2 \| \nabla v \|^2 , \]
\[ H(x, v) = \tau_4 \| v - f(x, \nabla x^H u) \|^2 , \]
where
\[ v = f(x, \nabla x^H u) . \]

This parametrization allows (5) to be rewritten as
\[ \arg\min_{x,v} \quad l(x) + g(v) + H(x,v) , \tag{6} \]
where \( l \) and \( g \) are related to the MRI and US images, and \( H \) ensures the coupling between the two modalities (whose importance is controlled by the hyperparameter \( \tau_4 \)).

The PALM algorithm iteratively minimizes the cost function in (6) with respect to \( x \) and \( v \) (the reader is invited to consult [16] for more details about PALM). Note that this cost function depends on the coefficients \( c_{l,k,p} \) and degrees \( d_p \) of the different polynomials, which need to be estimated for each patch, as shown in the next subsection.

### 3.2. Estimation of the polynomial functions \( f_p \)

For a given degree \( d_p \), the polynomial function \( f_p \) relating patches \( P_p x_{\text{mr}} \) and \( P_p x_{\text{us}} \) is defined by \( (d_p + 1)(d_p + 2)/2 \) coefficients assembled in the vector \( c_{d,p} = \{ c_{k,l,p} \mid k + l \leq d_p \} \). To estimate these coefficients, we consider that the \( p \)-th observed MR and US patches are related according to
\[ P_p y_{\text{us}} = \sum_{k+l \leq 3} c_{k,l,p} P_p y_{\text{mr}}^l \odot (P_p \nabla y_{\text{mr}}^H u)^k + \epsilon_p , \tag{7} \]
or in a matrix form
\[ P_p y_{\text{us}} = A_{\text{mr,p}} c_{d,p} + \epsilon_p , \]
where \( A_{\text{mr,p}} \) is a matrix whose elements are \( P_p y_{\text{mr}}^l \odot (P_p \nabla y_{\text{mr}}^H u)^k \) for \( l + k \leq d_p \), and \( \epsilon_p \) is the measurement error. The least-squares estimator of \( c_{d,p} \) is defined by
\[ \hat{c}_{d,p} = A_{\text{mr,p}}^+ P_p y_{\text{us}} , \quad p = 1, \ldots, N_p , \]
where \( A_{\text{mr,p}}^+ = (A_{\text{mr,p}} A_{\text{mr,p}})^{-1} A_{\text{mr,p}}^T \) is the pseudo-inverse of the matrix \( A_{\text{mr,p}} \).

In order to estimate the polynomial degree of the \( p \)-th patch, we minimize the least square distance between \( P_p y_{\text{mr}} \) and \( P_p y_{\text{us}} \), i.e., solve the following problem
\[ \arg\min_{d_p} \quad \| P_p y_{\text{us}} - f_p(P_p y_{\text{mr}}, P_p \nabla y_{\text{mr}}^H u) \|^2 , \]
where we highlight that the polynomial degree \( d_p \) depends on the patch size. In the results provided in this paper, patches of size \( 30 \times 30 \) were extracted from images containing \( 600 \times 600 \) pixels, with an overlap of 25\%. The degree of the polynomial relating the patches was constrained to \( d_p \in \{1, \ldots, 3\} \).

## 4. RESULTS AND DISCUSSION

The proposed MR-US image fusion algorithm was validated on experimental phantom data. Figs. 1(a,b) show the observed MR and US images.

![Fig. 1. Original MR image (200 × 200 pixels) and US image (600 × 600 pixels) and fusion results: (a) observed MRI, (b) original US image, (c) noisy US image, (d) fused image using a global polynomial model [14], (e) fused image with the proposed path-based polynomial model.](image)

To mitigate the relatively good SNR obtained due to the phantom design, the US image was further degraded by log-Rayleigh noise as shown in Fig. 1(c). Figs. 1(a,b,c) highlight the differences in gray levels, spatial resolution, contrast, and noise between the two MR and US images. Three main structures can be observed in these images: a PVC phantom

---

\(^2\)More details about the experimental model design and image acquisition can be found in [17].
(bright structure in the MR image), a piece of beef meat (gray structure in the MR image), and the glue used to attach them, only visible in the US image. Figs. 1(c,d) show the fused images obtained with the algorithm in [15] and the new proposed approach. Both fused images gather information from MR and US images (with a small preference to the proposed method): they provide a good contrast between the PVC and the beef tissue (similar to MRI), a good spatial resolution (similar to US) allowing small structures such as the glue to be distinguished, and good SNR. Moreover, the image obtained after fusion seems to carry more information than MRI, especially in the beef tissue.

### Table 1. CNR results

<table>
<thead>
<tr>
<th></th>
<th>MRI</th>
<th>US</th>
<th>Fused image with [15]</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNR</td>
<td>48.76 dB</td>
<td>20.64 dB</td>
<td>37.73 dB</td>
<td>41.72 dB</td>
</tr>
</tbody>
</table>

In addition to visual inspection, the performance of the proposed patch-wise method was evaluated using two quantitative measures and compared to the global fusion method of [15]: 1) the contrast-to-noise ratio (CNR) [18] between the PVC and the beef meat, and 2) the slope between two neighboring structures as an indication of the spatial resolution [19]. As reported in Tables 1 and 2, the patch-wise approach offers a better compromise between MR and US images with a CNR close to that of the MRI and a slope close to that of the US image. Fig. 2 confirms these results, showing that the patch-wise fused image captures more details from the MRI than the global model-based fused image.

### Table 2. Slope values at the interface between different regions of interest in the MR, US and fused images, corresponding to the vertical profile in Fig. 2.

<table>
<thead>
<tr>
<th>Slope</th>
<th>MRI</th>
<th>US</th>
<th>Fused image with [15]</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>2.89</td>
<td>7.42</td>
<td>7.42</td>
<td>7.42</td>
</tr>
<tr>
<td>#2</td>
<td>-0.10</td>
<td>8.89</td>
<td>6.86</td>
<td>7.15</td>
</tr>
<tr>
<td>#3</td>
<td>3.57</td>
<td>5.47</td>
<td>4.61</td>
<td>5.24</td>
</tr>
<tr>
<td>#4</td>
<td>-1.35</td>
<td>-1.95</td>
<td>-2.05</td>
<td>-2.05</td>
</tr>
</tbody>
</table>

5. CONCLUSION

This paper studied a new approach for MR and US image fusion. The relation between MR and US images was modeled locally by low-order polynomial functions associated with the image patches. Interestingly, results obtained on a phantom show the advantage of using local polynomials associated with the image patches. A natural progression of this work is to combine the proposed fusion method with multimodal image registration in order to correct the registration errors and to further validate the algorithm on in vivo data.
6. REFERENCES


