Spine Image Fusion via Graph Cuts

Brandon Miles, Ismail Ben Ayed, *Member, IEEE,*, Max W. K. Law, Greg Garvin, Aaron Fenster, *Senior Member, IEEE*, Shuo Li

Abstract—This study investigates a novel CT/MR spine image fusion algorithm based on graph cuts. This algorithm allows physicians to visually assess corresponding soft tissue and bony detail on a single image eliminating mental alignment and correlation needed when both CT and MR images are required for diagnosis. We state the problem as a discrete multi-label optimization of an energy functional that balances the contributions of three competing terms: (1) a squared error, which encourages the solution to be similar to the MR input, with a preference to strong MR edges; (2) a squared error, which encourages the solution to be similar to the CT input, with a preference to strong CT edges; and (3) a prior, which favors smooth solutions by encouraging neighboring pixels to have similar fused-image values. We further introduce a transparency-labeling formulation, which significantly reduces the computational load. The proposed graph cut fusion guarantees nearly global solutions, while avoiding the pixelation artifacts that affect standard wavelet based methods. We report several quantitative evaluations/comparisons over 40 pairs of CT/MR images acquired from 20 patients, which demonstrate a very competitive performance in comparisons to the existing methods. We further discuss various case studies, and give a representative sample of the results.

Index Terms—Image Fusion, Graph Cuts, Medical Imaging, Spine.

I. INTRODUCTION

FOR spine diseases and injuries, it is common for a patient to receive both an MR and a CT scan because of their individual benefits. MR images depict useful soft-tissue details including the spinal discs, nerves, cerebral spinal fluid and spinal cord. Therefore, it is the primary modality to diagnose protruding and degenerated discs. CT images clearly depict bony structures, especially the bone cortex, allowing the assessment of damaged joints or osteophyte growth (bony spurs at the margins of a joint).

Radiologists currently display MR and CT images side by side, when both images are available. This does provide them with all the available image information, but its accessibility is limited to visual correlation between the two images. It can be difficult to determine whether narrowing of a spinal canal is caused by tissue or bone from clinical MR images hence, both CT and MR can be employed [1]. Using both CT and MR images, as opposed to relying on a single modality can benefit diagnosis and treatment of osteophytes and degenerate discs that impact bone and nerve structures. In addition both modalities can aid post operative follow up after spinal surgery

[2]. Here, both the CT and MR modalities provide complementary information. In order to properly visualize the related bone and soft tissue structures, the images must be mentally aligned and fused together. Detecting changes on unregistered, uncombined images is an error prone task [3] . Therefore, it is highly desirable to fuse these two modalities into a single image showing the clinically significant CT and MR details as well as their relative locations on a single image. This will remove the need for mental juxtaposition when examining multiple views. Our goal is to provide and validate such a system.

Here we present a novel method for image fusion of the spine, which preserves the bone structures and soft tissue detail in a single image. Spine image fusion has the potential to enable more effective and efficient evaluations of spine disorders, more so as the number of spine scans increases very rapidly. For instance, in the United States, there has been a 300% increase in lower spine MR scans in the period between 1994 and 2006 [4], and the percentage of adults who have suffered from back pain is 75% [5].

Multi-modality image fusion has been studied in other fields, with applications varying from multifocal [6] to geographical images [7]. In medicine, image fusion has been used for brain imaging [8], [9], MRI-SPECT fusion [10], epilepsy treatment planning [11], liver ablation [12] and digital subtraction angiography [13]. For the spine, registered and overlaid CT and MR spine images have been used for surgery planning [1], [14] and evaluation of bone implants [2]. A wavelet based approach to image fusion has been proposed by Li et al. [6]. The two input images were fused in the wavelet domain, and an inverse transformation was applied to produce the result. Other variations of this technique include additive wavelet decomposition [7], the contourlet transform [8], [15], the curvelet transform [16] and the complex wavelet transform [17], [18]. The wavelet or transform based methods can suffer from pixelation artifacts when two dissimilar images are fused. This is a result of the decimations involved in the wavelet transform as well as the translation dependence of standard wavelets. This may result in small details being distorted, leading to a significant loss in image quality. Pathologies that should have been visible may no longer be seen.

Variational fusion methods have also been investigated [19]–[23]. These methods consist of finding the optimum of an energy functional, often via standard continuous optimization techniques, e.g., gradient descent. For these variational approaches, the main difficulties come from the limitations of the optimizers. Gradient-descent procedures [19]–[23] yield suboptimal solutions and have a very high computational load.

Discrete optimization methods, which use graph cut algorithms have recently sparked a substantial research effort

in computer vision, and led to very efficient algorithms in image segmentation [24], [25], stereo vision [26] and image restoration [27]. For object recognition, graph cuts have been employed to segment planar surfaces from depth images [28]. Both depth and intensity values were combined as inputs to a graph cut based segmentation algorithm. Graph cuts have also been used to determine the optimal fusion rules for combining subbands of a beamlet transform [29]. Although this method utilizes a graph cut approach for combining subbands, it is still a transform based method and therefore is limited by the transformation performed. To the best of our knowledge, graph cut formulations have not been previously studied in the context of variational approaches to image fusion.

In this study, we state image fusion as a discrete multi-label optimization problem, which can be solved efficiently with graph cuts [26], [30], [31], via the well-known swap or alpha-expansion moves [26]. The proposed energy function [32] balances the contributions of three competing terms: (1) a squared error, which encourages the solution to be similar to the MR input, with preference to strong MR edges; (2) a squared error, which encourages the solution to be similar to the CT input, with preference to strong CT edges; and (3) a prior, favoring smooth solutions by encouraging neighboring pixels to have similar fused-image values. We further introduce a transparency-labeling formulation, which significantly reduces the computational load. The proposed graph-cut fusion guarantees nearly global solutions, while avoiding the pixelation artifacts that affect standard wavelet based methods. We report several quantitative evaluations/comparisons over 40 pairs of CT/MR images acquired from 20 patients. The results demonstrate very competitive performance in comparisons to existing variational and transform-based methods [6], [8], [19].

This work is a significant extension of a preliminary conference version [32]. The dataset was increased from 9 to 20 patients, resulting in validation on 40 image slices. A comparison to Piella's variational method and validation results based on the structural similarity information measure (SSIM) have also been added. Two more clinical case studies have also been included. In section II we present our formulation of image fusion as a graph cut labeling problem. This is followed by a description of the dataset, its registration, preprocessing and fusion in section III. We further discuss various case studies, and give a representative sample of the results in IV with a discussion following in V.

II. FORMULATION

A. Multi-label Formulation

We state image fusion as the following multi-label optimization problem:

$$\lambda^* = \min E(\lambda) \text{ with } E(\lambda) = \mathcal{D}(\lambda) + c_1 \mathcal{R}(\lambda) \qquad (1)$$

where:

- Variable λ is a labeling function that assigns each point in image domain Ω to a label l, which describes the intensity of the fused image at that point:

$$\lambda : p \in \Omega \rightarrow \lambda(p) \in \mathcal{L}, \qquad (2)$$

with $\mathcal{L} \subset \mathcal{I}$ denoting a closed finite set of integers (the possible output intensities).

- Data term $\mathcal{D}$ is defined as:

$$\mathcal{D}(\lambda) = \sum_{p \in \Omega} \mathcal{D}_p(\lambda(p))$$
$$= \sum_{l \in \mathcal{L}} \sum_{p \in R_l} \left[w_1 (l - u_1(p))^2 + w_2 (l - u_2(p))^2 \right]$$
$$(3)$$

where $u_1 : \Omega \rightarrow \mathbb{R}$ and $u_2 : \Omega \rightarrow \mathbb{R}$ denote the input images, and R_l is the l-label region defined by $\{p \in \Omega | \lambda(p) = l\}$. w_1 and w_2 are weights defined as follows:

$$s_1 = |\nabla u_1| * K \qquad s_2 = |\nabla u_2| * K$$
$$w_1 = \frac{s_1}{s_1 + s_2} \qquad w_2 = \frac{s_2}{s_1 + s_2} \qquad (4)$$

K is a kernel, for instance, a box filter. w_1 and w_2 bias the solution towards strong edges in u_1 and u_2, respectively. The data term balances the contributions of two competing terms:

1. A squared error which, encourages the solution to be similar to the first input u_1, with preference to strong edges in u_1; and

2. A squared error which, encourages the solution to be similar to the second input u_2, with preference to strong edges in u_2.

- Smoothness term $\mathcal{R}$ favors smooth solutions by encouraging neighboring pixels to have similar fused-image values:

$$\mathcal{R}(\lambda) = \sum_{\{p,q\} \in \mathcal{N}} r(\lambda(p), \lambda(q)) \qquad (5)$$

with $\mathcal{N}$ being a set containing all pairs of pixels p and q in a local neighborhood of p and $r(\lambda(p), \lambda(q))$ is defined by the truncated absolute value:

$$r(\lambda(p), \lambda(q)) = \min(c_2, |l_p - l_q|) \qquad (6)$$

with c_2 being a positive constant.

B. Alpha-blending Reformulation

The above formulation requires a one-to-one correspondence between the labels and pixel intensities. Therefore, the number of labels needed to express the output image is equal to the number of all possible pixel values. This may lead to a high computational load in the case of images with large dynamic ranges, as is common in medical imaging. To reduce the number of labels, we reformulate the data term as a transparency labeling. This is done by expressing the output image as a function of u_1 and u_2 via a transparency image α, with $\alpha(p) \in [0, 1] \; \forall \; p \in \Omega$:

$$u_\alpha = \alpha u_1 + (1 - \alpha) u_2 \qquad (7)$$

where u_α denotes the output fused image.

Based on this formulation, we rewrite the data term in (3) as follows:

$$\mathcal{D}(\lambda) = \sum_{p \in \Omega} D_p\left(\lambda(p)\right)$$
$$= \sum_{l \in \mathcal{L}_\alpha} \sum_{p \in R_l} \Big[w_1 \left(u_\alpha(p,l) - u_1(p)\right)^2 +$$
$$+ w_2 \left(u_\alpha(p,l) - u_2(p)\right)^2 \Big] \tag{8}$$

where

$$u_\alpha(p,l) = \frac{l}{N_l} u_1(p) + \left(1 - \frac{l}{N_l}\right) u_2(p) \qquad l \in \mathcal{L}_\alpha \tag{9}$$

with $\mathcal{L}_\alpha$ being a new (reduced) set of non-negative integer labels $\{0, 1, 2, ..., N_l\}$, parameterized by the user specified number of labels N_l.

C. Graph Cut Optimization

Our problem is amenable to efficient graph cut optimization [26], [30], [31]. An illustration of the multi-label graph cut problem is provided in Fig. 1. Exactly one label is given to each pixel in the image, with associated data and smoothness costs assigned to the links in the graph. To formulate this optimization let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a weighted graph, with $\mathcal{V}$ a set of nodes and $\mathcal{E}$ a set of weighted edges. $\mathcal{V}$ contains a node for each pixel in Ω and for each label in $\mathcal{L}_\alpha$. There is an edge $e_{\{p,q\}}$ between every pair of nodes p, q. A cut $\mathcal{C} \subset \mathcal{E}$ is a set of edges that separates all the label nodes from each other, thereby creating a sub-graph for each label. The minimum-cut problem consists of finding a cut $\mathcal{C}$ with the lowest cost. The cost of this minimum cut, denoted $|\mathcal{C}|$, equals the sum of the edge weights in $\mathcal{C}$. By properly setting the weights of the graph, one can use a series of swap moves from combinatorial optimization [30] to efficiently compute the minimum-cost cuts corresponding to a minimum of functional E.
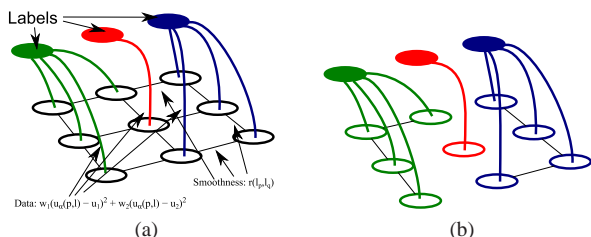

Fig. 1. An illustration of the graph-cut problem: a) A graph with 3 possible labels showing the data cost of assigning a label to a node and smoothness cost of assigning a labeling to adjacent pixel locations, b) the end result of the labeling of the graph.

A swap move starts with a labeled graph and determines for a given pair of labels, p and q, whether each node having a value in p, q should (1) retain its current label or (2) be updated to the other label in the pair. Each swap is accomplished globally in an exact manner by finding the minimum cut on a binary graph consisting of only two labels. This can be extended to the multi-label case by iterating over the set of all possible pairs of labels. The minimum cut is selected at each stage, with the final labeling corresponding to a

minimum of the energy functional. One can also use alpha-expansion moves [26] to optimize energy functions of the form E. It is well-known that alpha-expansion moves guarantee a solution that is within a constant factor of the global optimum [26]. However, experimentally, it is well established that swap moves outperform alpha expansions [26]. Therefore, in this work, we used swap moves with the edge weights defined in Table I; where $e_{\{l,p\}}$ denotes an edge between a label and a pixel, and $e_{\{p,q\}}$ an edge between two adjacent pixels.

TABLE I
WEIGHTS ASSIGNED TO THE EDGES OF THE GRAPH FOR MINIMIZING THE PROPOSED FUSION ENERGY

edge	weight	for
$\{l, p\}$	$w_1 \left(u_\alpha(p,l) - u_1\right)^2 + w_2 \left(u_\alpha(p,l) - u_2\right)^2$	$p \in \Omega, l \in \mathcal{L}_\alpha$
$\{p, q\}$	$r(l_p, l_q)$	$p, q \in \Omega$

III. METHODS

This retrospective study was approved by the Human Subjects Ethics Board of Western University, with the requirement for informed consent being waived. Twenty patient image sets were randomly selected with the criteria of patients having had both a lumbar MR and CT scan within a one-year time period. None of these patients had fractures, but other diseases such as degenerate / protruding discs, spinal stenosis and osteophytes were present. The images were acquired using either a Magnatom or Avanto Siemens 1.5T MR scanner (Seimens AG, Erlangen, Germany), with varying CT scanners depending on the location the CT images were obtained. The MR scans were acquired using a 3D T2 weighted pulse sequence, and the CT scans were acquired from either helical or axial slice CT images. No contrast was used in either scan. The lumbar spine was assumed to be rigid between scans, because the patients were scanned in a feet first prone position, resulting in very similar postures. We evaluated the proposed method over 40 pairs of CT/MR images acquired from these 20 patients. Twenty pairs were from the center sagittal slice, and 20 were from the left side of the patient through the nerve root bundle. T2-weighted 3D MR images were used because they clearly present the discs, nerve root bundle and cerebral spinal fluid. The 3D MR/CT images were then registered and preprocessed. Finally the images were fused as sets of 2D images because radiologists typically view 3D volumes as stacks of 2D images.

Validation was completed based first on visual results of the fusion, studying clarity of the detail presented in the fused image, and second via a statistical comparison of the clinically significant bone and tissue transferred to the fused images. Four clinical case studies were then examined to illustrate the potential clinical value of this technique. Our method was compared to four methods: (A) an averaging of the two images, and three methods from recent literature: (B) the discrete wavelet transform (DWT) [6] (C) the contourlet transform (CLT) [8] and (D) Piella's variational method [19]. These methods were implemented using the parameters listed in their papers. Table II contains a summary of these features.

TABLE II
IMPLEMENTATION DETAILS AND PARAMETERS FOR THE THREE METHODS FROM LITERATURE

Discreet Wavelet [6]	Contourlet Transform [8]	Piella Variational [19]
Subbands: 3 subbands **Wavelet:** Haar wavelet **Lowpass Rule:** pixel-wise averaging **Highpass Rule:** pixel-wise selection of the coefficient with the largest maximum value	**Subbands:** 4 subbands, with 4,8,16 directional subbands in levels 2,3,4 (lowest detail to highest) **Filters:** Lowpass 9-7 Filter, directional PKVA **Lowpass Rule:** local energy in a 3x3 window **Highpass Rule:** local contourlet contrast	**Parameters** $\eta = 0.1$, $\beta = 0.5$, $\gamma = 0.3$, $\delta t = 0.15$ **Kernel** w - Gaussian, $\sigma = 0.1$ **Polynomial** J' $n = 7$, $\alpha = 10$, $k = 0.25$

A. Registration and Pre-Processing

The input volumes were registered, using a rigid 3D versor based transform in ITK [33]. The optimizer used maximization of Mutual Information (MI) [34] to align soft tissue details present in both images (note the soft tissue details in the CT image are suitable for registration, but MR is better for diagnosis). For the purpose of aligning soft tissue each CT image was thresholded from -255 to 255 Houndsfield Units (HU) or -255 to 0 HU if needed. This kept many of the soft tissue details, but removed most of the bone detail. Both images were then scaled to an intensity range of 0 to 255 to be in the same range. The transform was initialized using two corresponding user-selected points, one from the CT and the other from the MR image. After this, MI was calculated from the voxels in both images, and the versor transform was iteratively updated based on MI of the two images at each step. Using the obtained optimal transform, the original MR image (without intensity scaling) was transformed and resampled to the voxel spacing of the CT image.

Manual points were selected in the 3D images for the target registration error (TRE) and the fiducial localization error (FLE) evaluation. The TRE is the mean post registration Euclidean distance between corresponding pairs of fiducials from the input images. The FLE is the root mean-squared difference in locations when selecting the same fiducial multiple times in an image [35]. The TRE used 17 points from two image pairs. For the FLE, 5 distinct points were defined on the CT image. On 5 separate days, corresponding points in the MR image were identified producing a total of 25 point sets. These errors were used to validate the registration.

After registration, the original CT images were thresholded at 0 HU, setting any negative values to 0 HU and leaving other values unchanged. This removed most of the soft-tissue details and was done because the MR presents the tissue detail with more clarity, so the CT tissue detail is undesirable for the fused image. For all 20 patients the MR images were found to have a maximum intensity of about 700, and the CT images were found to have a maximum intensity of about 1400. In view of this, the CT intensities were divided by two so that the MR and CT histograms would have similar intensity ranges prior to fusion. This was needed to eliminate bias in the fusion algorithm due to differing intensity ranges in the input images. MR and CT histograms are shown in Fig. 2 and Fig. 3 shows a flow chart describing the registration and pre-processing steps.
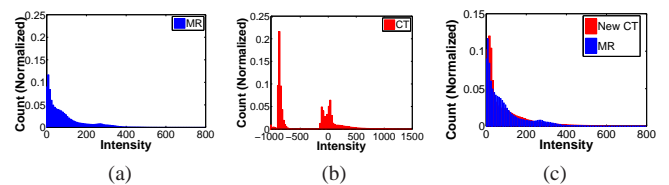

Fig. 2. Histograms of all the images, a) Histogram of all 40 MR images, b) Histogram of all 40 CT images, c) Histogram of all 40 CT images after Preprocessing (excluding pixels with an intensity of 0) and histogram of all 40 MR images.
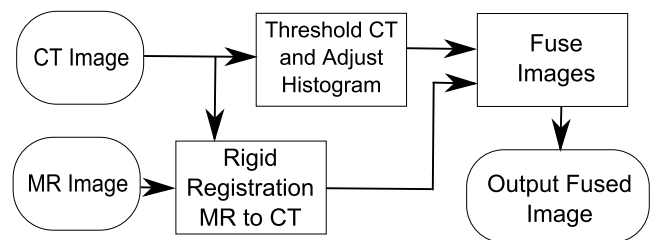

Fig. 3. Flow chart of the image fusion process: The MR is registered to the CT image. The CT is thresholded and its histogram is adjusted to match the MR. Finally the images are fused.

B. The Parameters of our Fusion Method

After preprocessing was completed, fusion could be carried out. For the purpose of these experiments, c_1 was set equal to $\frac{0.001(I_{max})^2}{c_2}$, c_2 equal to $0.40(N_l)$ and N_l equal to 20, with I_{max} being the maximum intensity value in both inputs. N_l and c_1 were tuned manually to balance image quality and speed of computation. c_2, was set empirically for smoothness. Constant c_1 was set in relation to the maximum value in the data term and in relation to c_2.

C. Statistical Evaluation of the Fused Images

We compared each of the MR and CT images to the fused result within: (1) the regions of soft tissues, and (2) the regions of bone structures. The soft-tissue details consisted of the discs, nerves, and cerebral spinal fluid from the MR image, and the bone details were from the CT image, with a specific focus on the bone cortex. The trabecular bone does not contact soft tissue and so was omitted. We created image masks of the tissue and bone details for each patient. The tissue masks were created by manual segmentations of the MR images, and the bone masks were obtained by thresholding the CT images at a user selected HU for each image and then manually correcting
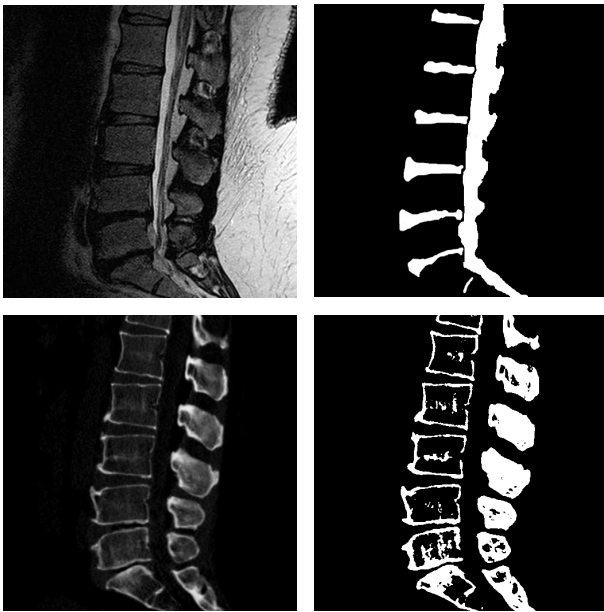
Fig. 4. Sample masks for validating the quality of the fusion algorithm. Top left: sample MR image, Top right: soft tissue Mask, Bottom Left: sample CT image, Bottom right: bone detail mask

any errors. Figure 4 shows sample masks of the tissue and bone detail.

We defined a fusion error as the mean absolute-value difference between the MR/CT images and the fused images in the tissue regions defined by the masks. For the MR images, we calculated the following two errors:

$$e_{\text{MR, Tissue}} = \frac{\sum_{M_{\text{Tissue}}} |I_{\text{MR}} - I_{\text{fused}}|}{\text{area of the tissue mask}} \quad (10)$$

$$e_{\text{MR, Bone}} = \frac{\sum_{M_{\text{Bone}}} |I_{\text{MR}} - I_{\text{fused}}|}{\text{area of the bone mask}} \quad (11)$$

where I_{MR} is the intensity of the MR image for a given pixel, I_{fused} is the intensity of the fused image at a given pixel and $M_{\text{Tissue}}, M_{\text{Bone}}$ are the non-zero domains of the two masks. Similarly, two additional errors, $e_{\text{CT, Bone}}$ and $e_{\text{CT, Tissue}}$ were defined for the CT images. Ideally, there should be no tissue differences between the MR images and the fused images in the tissue regions ($e_{\text{MR, Tissue}} = 0$) and no bone difference between the CT images and the fused images in the bone regions ($e_{\text{CT, Bone}} = 0$). The hypothesis we tested was that the error obtained for the MR images is lower than the one obtained for the CT images within soft-tissue regions, i.e. $e_{\text{MR, Tissue}} < e_{\text{CT, Tissue}}$, and higher within bone regions $e_{\text{MR, Bone}} > e_{\text{CT, Bone}}$.

Each of the four errors were calculated for each patient. Some of the data was found to be non-normal using a Shapiro-Wilks test [36], thus a non-parametric Wilcoxon test [37] was used to compare sets of errors. The tissue errors: $e_{\text{MR, Tissue}}$ and $e_{\text{CT, Tissue}}$ were compared to each other and the bone errors: $e_{\text{CT, Bone}}$ and $e_{\text{MR, Bone}}$, where also compared, in order to determine if there was a statistical significance difference between them for the 40 patient image sets. These calculations were performed for each of the five fusion methods using

version 20 of the SPSS statistical software (SPSS Inc., an IBM Company, Armonk, NY).

D. Additional Metrics for Evaluation

In addition to the above statistical tests, we have also examined the sensitivity and specificity or our algorithm along with the structural similarity in the masks [38]. For classification, we have defined true and false positives/negatives (TP,FP,TN,FN) per pixel as:

- $\text{TP}_{\text{tissue}}$ and TN_{bone} if $(e_{\text{MR, Tissue}} < e_{\text{CT, Tissue}})$
- $\text{FP}_{\text{tissue}}$ and FN_{bone} if $(e_{\text{MR, Tissue}} \geq e_{\text{CT, Tissue}})$
- $\text{TN}_{\text{tissue}}$ and TP_{bone} if $(e_{\text{MR, Bone}} > e_{\text{CT, Bone}})$
- $\text{FN}_{\text{tissue}}$ and FP_{bone} if $(e_{\text{MR, Bone}} \leq e_{\text{CT, Bone}})$

Sensitivity and specificity were calculated for each using the total number of TPs, FPs, TNs and FNs normalized over the image masks, which we denote by $\overline{n}\text{TP}, \overline{n}\text{FP}, \overline{n}\text{TN},$ and $\overline{n}\text{FN}$. Sensitivity and specificity are defined in (12). Since the tissue sensitivity is equal to the bone specificity and the bone sensitivity is equal to the tissue specificity, only the two sensitivity values have been reported:

$$\text{Sensitivity} = \frac{\overline{n}\text{TP}}{\overline{n}\text{TP} + \overline{n}\text{FN}}$$
$$\text{Specificity} = \frac{\overline{n}\text{TN}}{\overline{n}\text{TN} + \overline{n}\text{FP}} \quad (12)$$

The structural similarity metric [38] is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (13)$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}$ represent the means in the x and y images, the variances in the x and y images and the covariance of the two images respectively. This metric has been applied over a local window for pixels within the given masks, comparing the MR images to the fused images in the tissue mask, and the CT images to the fused images in the bone masks. The window was defined as an 11x11 Gaussian kernel with $\sigma = 1.5$. $C_1 = 0.01$ and $C_2 = 0.03$ are positive constants.

IV. RESULTS

In the following, we describe a representative sample of the fusion results, report several statistical evaluations, and discuss four clinical case studies based on our fusion method.

A. Registration Validation

The TRE [35] was found to be 1.9 ± 0.6 mm with a CT voxel spacing of 3 x 0.3 x 0.3 mm for the tested images. The FLE was found to be 0.8 ± 0.4 mm. This demonstrates that the registration accuracy is sub-voxel, since the TRE is greater than the FLE, but less than the diagonal size of the voxels.
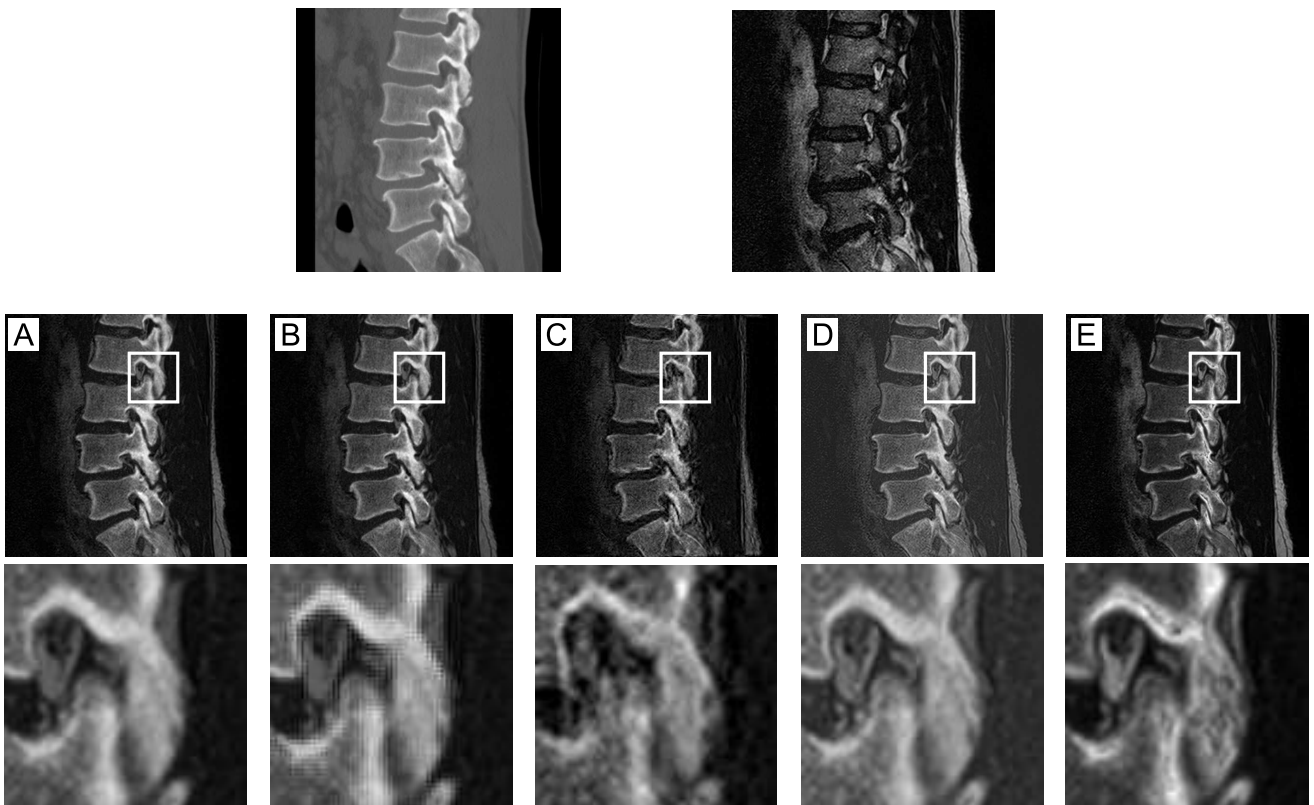
Fig. 5. Sample fused Images: top left - input CT, top right input MR, middle fused images: A) averaging method, B) discrete wavelet transform, C) contourlet, D) Piella's variational method, E) our proposed graph cut method, bottom row: magnified images within the region of interest indicated in the images above.

B. Sample Images

Figure 5 shows sample input images, including the registered CT and MR inputs and the results of the five fusion methods. It shows that the algorithms perform very differently in preserving the CT/MR details. As expected, the averaging method (A) loses many details, whereas the wavelet method (B) introduces block-structure artifacts because it does not account for shift invariance. The contourlet method (C) significantly blurs the MR details and adds noise to the CT detail, making it difficult to identify the nerve structures and bones. The variational method (D) preserves the details, but significantly reduces the intensity range of the solution. The graph cut result (E) depicts sharp MR and CT details, has a much larger dynamic range than Piella's method (D), and is artifact-free.

C. Evaluation of Masks

We have calculated the mean intensity in the tissue and bone masks, for all the MR and CT images. This provides a frame of reference for the fusion error calculations and to evaluate the effectiveness of the masks. These intensities are shown in Fig. 6, with separate graphs for the left sagittal slices through the nerve root bundle and mid sagittal slices, through the center of the subject. For the tissue mask, in the left slices the MR values are about 50, whereas the CT values are between 50 and 100, for the mid slices the CT values are about 50, whereas the MR values range between 100-250 for most patients. This demonstrates that the tissue masks perform

well at discriminating between tissue and bone for the mid slices (high MR, low CT), but have less differentiation power for the left slices. For the bone masks, on both the left and mid slices the intensities prior to fusion are between 200 and 300, with the MR values around 100 or less. This shows clear differentiation between bony and tissue detail in the masks.

D. Statistical Results

The results of the fusion errors for our graph-cut method are shown in Fig. 7. These show that $e_{\text{MR,Bone}}$ is greater than $e_{\text{CT,Bone}}$ error in all images as was hypothesized. For the tissue errors, $e_{\text{MR,Tissue}}$ is fairly constant at about 50 for all images, with $e_{\text{CT,Tissue}}$ being much lower on the left images, than the mid images, but higher than $e_{\text{MR,Tissue}}$ for the majority of subjects.

For all five methods, Table III reports the mean values of $e_{\text{MR,Tissue}}$, $e_{\text{CT,Tissue}}$, $e_{\text{CT,Bone}}$ and $e_{\text{MR,Bone}}$ over the 40 image sets. These were measured in pixel intensity. Table III also shows p-values for the pairwise and independent Wilcoxon tests comparing $e_{\text{MR,Tissue}}$ with $e_{\text{CT,Tissue}}$ and $e_{\text{CT,Bone}}$ with $e_{\text{MR,Bone}}$, again over the 40 sets of patient images.

As expected, $e_{\text{CT,Bone}} = e_{\text{MR,Bone}}$ and $e_{\text{MR,Tissue}} = e_{\text{CT,Tissue}}$ for the averaging method. For the wavelet/contourlet methods $e_{\text{MR,Tissue}}$ was slightly higher than $e_{\text{CT,Tissue}}$. This is the opposite of what is desired. On the contrary, for the proposed graph-cut method and Piella's method $e_{\text{MR,Tissue}}$ is lower than $e_{\text{CT,Tissue}}$. All the methods, except averaging, yielded $e_{\text{CT,Bone}} < e_{\text{MR,Bone}}$. Overall, the proposed graph cut method resulted in the lowest

Mean Intensity in Tissue Masks (Left Slices)

(a)

Mean Intensity in Tissue Masks (Mid Slices)

(b)

Mean Intensity in Bone Masks (Left Slices)

(c)

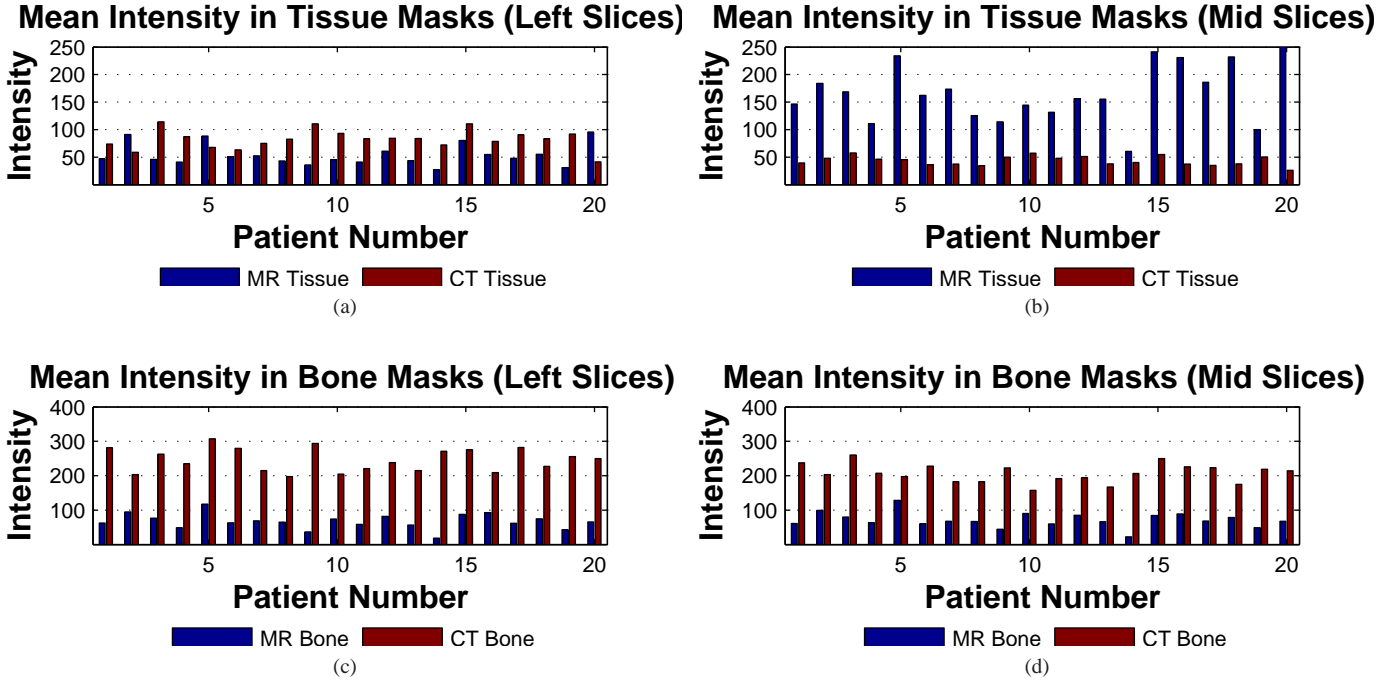## Mean Intensity in Bone Masks (Mid Slices)

(d)

Fig. 6. Masks Values before fusion, a) Tissue mask and left slices, b) Tissue mask and mid slices, c) Bone mask and left slices, d) Bone mask and mid slices

Tissue Fusion Error (Left Slices)

(a)

Tissue Fusion Error (Mid Slices)

(b)

Bone Fusion Error (Left Slices)

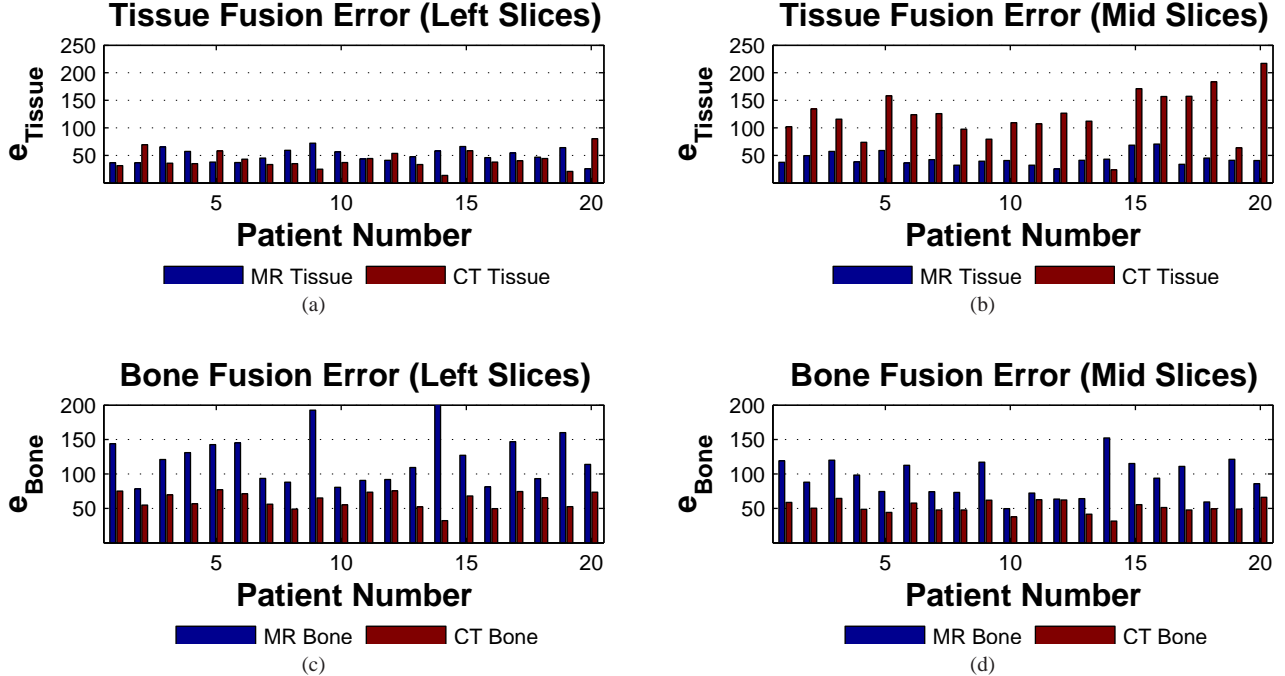

(c)

Bone Fusion Error (Mid Slices)

(d)

Fig. 7. Fusion Error for our graph cut method: a) e_{Tissue} for the left slices, b) e_{Tissue} for the mid slices c) e_{Bone} for the left slices and d) e_{Bone} for the mid slices

$e_{\text{CT,Bone}}$, and the lowest $e_{\text{MR,Tissue}}$, which corresponds well to our purpose. We obtained a mean $e_{\text{CT,Bone}}$ value of 57.0 based on the CT intensity dynamic range of 700. For the soft tissues $e_{\text{MR,Tissue}} = 46.6$. Note that Piella's method yielded the highest $e_{\text{MR,Tissue}}$. For bone regions, all the methods, except ours, yielded approximately the same $e_{\text{CT,Bone}}$.

With the exception of the averaging method $e_{\text{MR,Tissue}}$ was found to be pair-wise statistically different from $e_{\text{CT,Tissue}}$ and $e_{\text{CT,Bone}}$ was found to be pair-wise statistically different from $e_{\text{MR,Bone}}$. The independent Wilcoxon tests showed that only Piella's method and our graph cut method were statistically significantly different when analyzed as a group.

TABLE III
MEAN FUSION ERROR VALUES FOR THE 20 PATIENT DATA SETS AND
STATISTICAL-SIGNIFICANCE TEST RESULTS FOR THE FUSION METHODS.

Method	$e_{MR,Tissue}$	$e_{CT,Tissue}$	$P_{Pairwise}$	$P_{Independent}$
Averaging	64.2 ± 26.0	64.2 ± 26.0	1	1
DWT [6]	64.9 ± 26.1	63.9 ± 25.9	< 0.001	0.707
Contourlet [8]	73.4 ± 28.6	64.2 ± 24.9	< 0.001	0.083
Piella [19]	152.8 ± 52.6	175.9 ± 54.0	0.002	0.006
Graph Cuts	46.6 ± 12.3	81.7 ± 52.6	0.006	0.020
Method	$e_{CT,Bone}$	$e_{MR,Bone}$	$P_{Pairwise}$	$P_{Independent}$
Averaging	82.5 ± 19.8	82.5 ± 19.8	1	1
DWT [6]	81.9 ± 19.7	83.7 ± 19.8	< 0.001	0.583
Contourlet [8]	82.6 ± 19.7	89.0 ± 18.6	< 0.001	0.121
Piella [19]	84.1 ± 40.6	188.7 ± 48.0	< 0.001	< 0.001
Graph Cuts	57.0 ± 11.9	108.0 ± 36.2	< 0.001	< 0.001

E. Additional Metrics

The images have also been analyzed for the sensitivity and specificity of the number of correctly fused pixels along with the structural similarity between the input images and the fused images within the mask regions (Table IV).

TABLE IV
MEAN SENSITIVITY FOR TISSUE AND BONE DETAILS ALONG WITH MEAN
SSIM INDEX MEASURES FOR ALL FIVE METHODS.

Method	Sensitivity Tissue	Sensitivity Bone	SSIM Tissue	SSIM Bone
Averaging	0 ± 0	0 ± 0	0.59 ± 0.32	0.08 ± 0.04
DWT [6]	0.51 ± 0.01	0.76 ± 0.17	0.54 ± 0.32	0.11 ± 0.06
Contourlet [8]	0.50 ± 0.01	0.75 ± 0.17	0.22 ± 0.30	0.12 ± 0.08
Piella [19]	0.86 ± 0.11	0.87 ± 0.08	0.32 ± 0.31	0.15 ± 0.11
Graph Cuts	0.63 ± 0.12	0.84 ± 0.09	0.52 ± 0.33	0.21 ± 0.12

These results show that the averaging method has 0 sensitivity. The DWT and contourlet methods performed lower than the other two methods, with Piella's methods having the highest sensitivity for both CT and MR and graph cuts having the second highest for both. In regards to the SSIM index, the averaging method performed best on the MR data, followed by the DWT and Graph Cut methods. Piella's methods and the the contourlet transformed did much poorer. For bone details the graph cut method did the best, followed by Piella's method, the contourlet method, the DWT and finally averaging.

F. Case Studies

Visual Inspection - Lumbar Spine, Joint and Disc Disease
We present the first clinical case study for our fusion technique in Fig. 8. In the first case, the patient had a protruding spinal disc and damaged facet joint. The disc can be seen in the MR image as a hypointense region, whereas the facet joint is visible in the CT image. There is significant osteoarthritis in the joint. The fused image clearly shows both of these pathologies in a single image, allowing for a better diagnosis.
Visual Inspection - Osteophyte Growth The second case study shows osteophyte growth (see Fig. 9), which is the formation of bony spurs at the margins of a joint. On the MR image alone, it is difficult to see the location of the osteophyte. The CT shows the osteophyte, but none of the surrounding soft tissue. The fused image shows both the formation of the bony spurs and the surrounding soft tissue on a single image.

Visual Inspection - Abnormal Vertebrae and Cord damage
The third case study shows spinal cord damage (Fig. 11) and an abnormal vertebrae on a single fused image. The cord damage is not visible on the CT image, while the abnormal vertebrae is difficult to see on the MR image. The fused image presents both.
Visual Inspection - Osseous Erosion Secondary to Pannus
In this patient with rheumatoid arthritis (Fig. 10), pannus is eroding the posterior aspect of the dens. On the MR, the chronic pannus is dark and cannot be distinguished from the underlying bony cortex. On the CT, the margins of the bone are well seen, but soft-tissue contrast is poor. The relationship of the pannus to the underlying bone is best seen on the fused image.

V. DISCUSSION

We have investigated a novel CT/MR spine image fusion algorithm based on graph cuts. We have successfully fused MR and CT images to create a single fused image, providing a new and effective combined modality for diagnosis. Images were registered, pre-processed and then fused. This has been tested on 40 sets of clinical images from 20 patients. The graph cut results show better performance than the averaging method and the three state-of-the-art methods from the literature. Our method successfully transfers bone detail and soft tissue detail to the resulting fused image, with only a 57.0 difference in intensity values for the bone details and 46.6 intensity different for the soft-tissue details, in a dynamic range of 700. Visual inspection confirms these results, with graph cuts showing the sharpest detail for both the bone and soft tissue details.

The statistical tests showed pairwise significance for the CT vs MR error in every method except averaging, however, only Piella's method and our graph cut method showed groupwise statistical significance between errors, which is a stronger test. This indicates that these two methods perform better than the others, in transferring bone detail and soft tissue detail to the fused image. With regard to the additional methods, Piella's method performed better than graph cuts in regards to the sensitivity test. However, the graph cut method outperformed Piella's method in terms of structural similarity, a test where Piella's method is expected to perform well. In view of all the numeric and visual results, the graph cut method can be concluded to outperform the existing state-of-the-art methods.

The subvoxel accuracy of the registration ensures that the fusion errors are a result of the fusion techniques and not misregistration. A deformable registration could have been employed; however, this was not needed since the registration accuracy validates the assumption of rigidity.

One note of interest is that $e_{MR,Tissue}$ and $e_{CT,Tissue}$ for the DWT are similar or lower than the averaging, contourlet and Piella's methods, even though the DWT is visually worse than Piella's method. The DWT also has a high SSIM value. There was pairwise statistical significance of the fusion errors for the DWT, but no groupwise statistical significance. This indicates its poor ability to discriminate between tissue and bone detail, which is essential for clinical use.
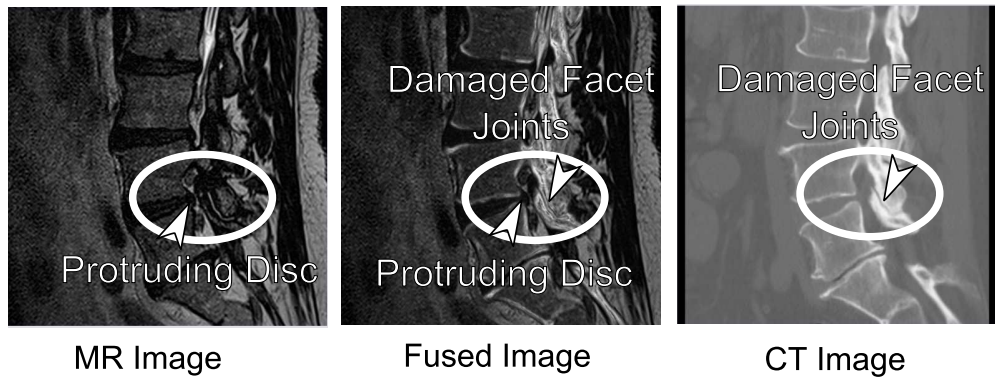
Fig. 8. Images of damaged spine: left - MR showing a protruding disc (hypointense region); middle - Fused image showing the disc, the spinal cord and the damaged facet joints; right - CT image showing damaged facet joints.
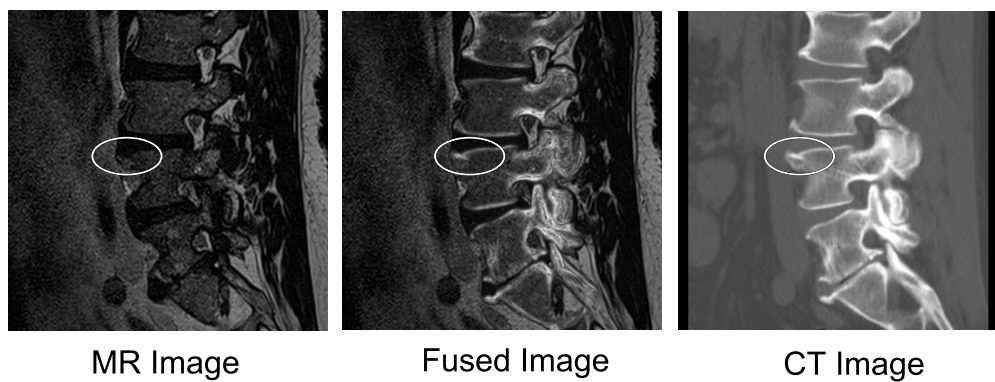

Fig. 9. Images of bony spur formation: left - in the MR, the osteophyte is not identifiable; middle - in the fused image, the osteophyte is clearly visible along with the surrounding soft tissue; right - the CT image shows the osteophyte, but not the soft tissue.
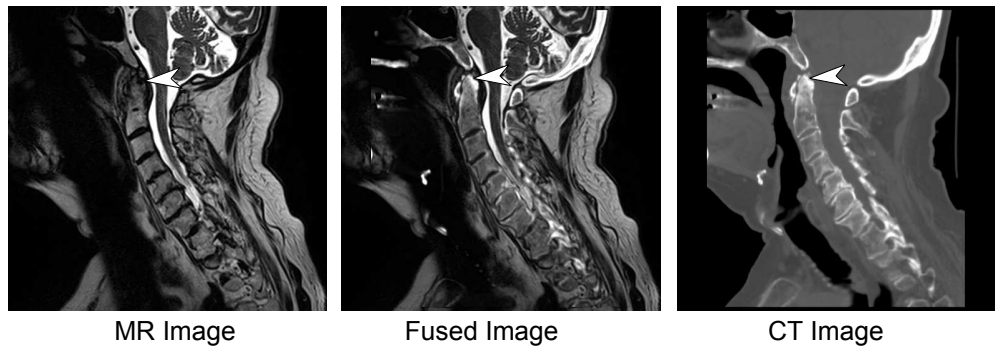

Fig. 10. The arrows show the pannus eroding the posterior aspect of the tip of the dens. The relationship between the pannus and the surface of the bone is best seen on the fused image.
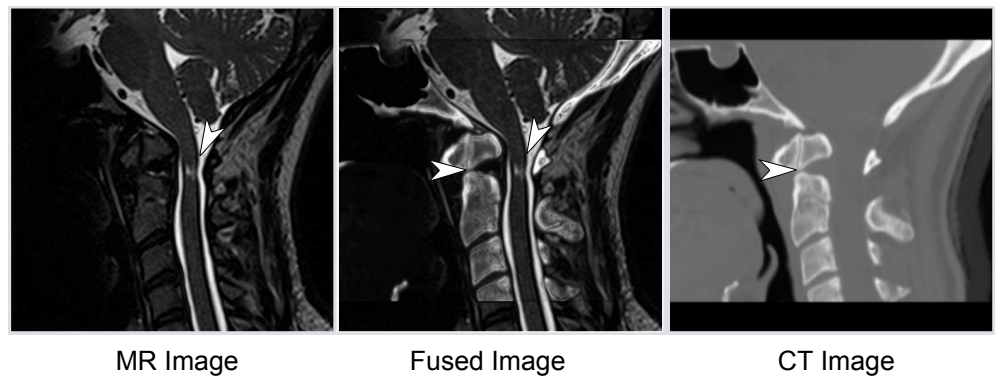

Fig. 11. Image of an abnormal vertebrae and cord damage: left - in the MR the cord damage is easily visible; center - the fused image: the cord damage and the abnormal vertebral body can be seen clearly; right - the CT image shows the abnormal vertebrae.

We have also shown the benefit of our fusion system on four clinical cases, where the fused image clearly shows both the bone and soft tissue detail on a single image. This highlights the pathology on a single image. Our method can successfully combine CT and MR images of the lumbar spine, while retaining the significant clinical detail. This eliminates the need for radiologists to mentally align and fuse two separate datasets, along with the associated potential for errors. Although we do not intend to have fused images replace CT and MR scans for clinical use, we do see this as a strong tool to add to the current practice and aid radiologists in completing more accurate and quicker diagnosis.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] Y. Hu, S. K. Mirzac, J. G. Jarvikb, P. J. Heagerty, and D. R. Haynor, "MR and CT image fusion of the cervical spine: a noninvasive alternative to CT-Myelography," in *Proceedings of SPIE*, vol. 5744. Bellingham, WA: SPIE, 2005.
[2] C. A. Karlo, I. Steurer-Dober, M. Leonardi, C. W. A. Pfirrmann, M. Zanetti, and J. Hodler, "MR/CT image fusion of the spine after spondylodesis: a feasibility study," *European Spine Journal*, vol. 19, pp. 1771–1775, 2010.
[3] R. Rensink, "Change blindness," *Annual Review of Psychology*, vol. 53, pp. 245–277, 2002.
[4] R. A. Deyo, S. K. Mirza, J. A. Turner, and B. I. Martin, "Overtreating Chronic Back Pain: Time to Back Off?" *Jour Am Board Fam Med*, vol. 22, no. 1, pp. 62–68, 2009.
[5] P. L. Owens, M. Woeltje, and R. Mutter, "Emergency Department Visits and Inpatient Stays Related to Back Problems," 2008. [Online]. Available: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb105.pdf
[6] H. Li, B. S. Manjunath, and S. K. Mitra., "Multisensor Image Fusion using the Wavelet Transform," *Graphical Models and Image Processing*, vol. 57, no. 3, pp. 235–245, 1995.
[7] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-Based Image Fusion with Additive Wavelet Decomposition," *IEEE Transactions On Geoscience And Remote Sensing*, vol. 37, no. 3, pp. 1204–1211, May 1999.
[8] L. Yang, B. L. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, pp. 203–211, 2008.
[9] G. K. Matsopoulos, S. Marshall, and J. N. H. Brunt, "Multiresolution morphological fusion of MR and CT images of the human brain," *IEEE Proceedings on Vision Image Signal Processing*, vol. 141, no. 3, pp. 137–142, 1994.
[10] A. Colin and J.-Y. Boire, "MRI-SPECT fusion for the synthesis of high resolution 3D functional brain images: a preliminary study," *Computer Methods and Programs in Biomedicine*, vol. 60, pp. 107–116, 1999.
[11] S. T. C. Wong, R. C. Knowlton, R. A. Hawkins, and K. D. Laxer, "Multimodal Image Fusion for Noninvasive Epilepsy Surgery Planning,," *IEEE Transactions on Computer Graphics and Applications*, vol. 16, no. 1, pp. 30–38, 1996.
[12] P. M. Mazurek, A. M. Silas, and J. M. Gemery, "Summation of CT Scans During Radiofrequency Ablation for Assessing Target Lesion Coverage," *American Journal of Roentgenology*, vol. 191, p. 790:792, September 2008.
[13] G. Zhang, Z. Cui, F. Li, and J. Wu, "DSA Image Fusion based on Dynamic Fuzzy Logic and Curvelet Entropy," *Journal of Multimedia*, vol. 4, no. 3, pp. 129–136, June 2009.
[14] M.-J. Sohn, D.-J. Lee, S. W. Yoon, H. R. Lee, and Y. J. Hwang, "The effective application of segmental image fusion in spinal radiosurgery for improved targeting of spinal tumours," *ACTA Neurochir*, vol. 151, pp. 231–238, 2009.
[15] Z. Youzhi, Q. Zheng, and Y. Jingyu, "Image Fusion Using a Hybrid Representation of Empirical Mode Decomposition and Contourlet Transform," in *Proceedings of Information Science and Technology (ICIST)*, 2011.
[16] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Information Fusion*, vol. 8, pp. 143–156, 2007.
[17] P. Hill, N. Canagarajah, and D. R. Bull, "Image Fusion Using Complex Wavelets," in *BMVC 2002*, 2002.
[18] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel - and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, pp. 119–130, 2007.
[19] G. Piella, "Image Fusion for Enhanced Visualization: A Variational Approach," *International Journal of Computer Vision*, vol. 83, no. 1, pp. 1–11, 2009.
[20] W.-W. Wang, P.-L. Shui, and X.-C. Feng, "Variational Models for Fusion and Denoising of Multifocus Images," *IEEE Signal Processing Letters*, vol. 15, pp. 65–68, 2008.
[21] W. Zhang and L. Yu, "SAR and Landsat ETM+ image fusion using variational model," in *2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*, 2010.
[22] C. Ballester, V. Caselles, L. Igual, and J. Verdera, "A Variational Model for P+XS Image Fusion," *International Journal of Computer Vision*, vol. 69, no. 1, p. 4358, 2006.
[23] C. Wang and Z.-F. Ye, "Perceptual Contrast-Based Image Fusion: A Variational Approach," *ACTA Automica Sinica*, vol. 33, no. 2, pp. 132–137, February 2007.
[24] Y. Boykov, V. Kolmogorov, D. Cremers, and A. Delong, "An integral solution to surface evolution PDEs via geo-cut," in *Proceedings of European Conference on Computer Vision*, vol. 3, 2006, pp. 409–422.
[25] M. B. Salah, A. Mitiche, and I. B. Ayed, "Multiregion Image Segmentation by Parametric Kernel Graph Cuts," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 545–557, 2011.
[26] Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, vol. 26, no. 9, pp. 1124–1137, September 2004.
[27] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.
[28] O. Khler, E. Rodner, and J. Denzler, "On fusion of range and intensity information using graph-cut for planar patch segmentation," *Int. J. Intelligent Systems Technologies and Applications*, vol. 5, no. 3/4, pp. 365–373, 2008.
[29] G. M. Zhang and Z. M. Cui, "A novel image fusion method using beamlet transform and graph cuts ()," *Key Engineering Materials*, vol. 467 - 469, pp. 1092–1096, 2011.
[30] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, vol. 23, no. 11, pp. 1222–1239, November 2001.
[31] V. Kolmogorov and R. Zabih, "What Energy Functions Can Be Minimized via Graph Cuts?" *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, vol. 26, no. 2, pp. 147–159, February 2004.
[32] B. Miles, M. W. K. Law, I. Ben-Ayed, G. Garvin, A. Fenster, and S. Li, "Pixel Level Image Fusion for Medical Imaging - an Energy Minimizing Approach," in *Proceedings of SPIE Medical Imaging*, 2012.
[33] Insight ToolKit. [Online]. Available: http://www.itk.org/
[34] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT Image Registration in the Chest Using Free-form Deformations," *IEEE Transactions on Medical Imaging*, vol. 22, no. 1, pp. 120–128, January 2003.
[35] J. M. Fitzpatrick, J. B. West, and C. R. Maurer, "Predicting Error in Rigid-Body Point-Based Registration," *IEEE Transactions on Medical Imaging*, vol. 17, no. 5, pp. 694–702, 1998.
[36] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, p. 591611, 1965.
[37] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, December 1945.
[38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.