http://dx.doi.org/10.1117/12.911613

Pixel Level Image Fusion for Medical Imaging - an Energy Minimizing Approach

Brandon Miles[a] Max W. K. Law[a,b] Ismail Ben-Ayed [b] Greg Garvin [c] Aaron Fenster [a,d] Shuo Li [a,b]

[a]University of Western Ontario, London Ontario
[b]GE HealthCare London Ontario
[c]Department of Radiology, St Joseph's Health Care, London Ontario
[d]Robarts Research Institute, London Ontario

ABSTRACT

In an attempt to improve the visualisation techniques for diagnosis and treatment of musculoskeletal injuries, we present a novel method for a pixel-wise fusion of CT and MR images. We focus on the spine and its related diseases including osteophyte growth, degenerate disc disease and spinal stenosis. This will have benefit to the 50-75% of people who suffer from back pain, which is the reason for 1.8% of all hospital stays in the United States.[1] Pre-registered CT and MR image pairs were used. Rigid registration was performed based on soft tissue correspondence. A pixel-wise image fusion algorithm was designed to combine CT and MR images into a single image. This was accomplished by minimizing an energy functional using a graph cut approach. The functional was formulated to balance the similarity between the resultant image and the CT image as well as between the resultant image and the MR image. Furthermore the variational smoothness of the resultant image was considered in the energy functional (to enforce natural transitions between pixels). The results have been validated based on the amount of significant detail preserved in the final fused image. Based on bone cortex and disc / spinal cord areas, 95% of the relevant MR detail and 85% of the relevant CT detail was preserved. This work has the potential to aid in patient diagnosis, surgery planning and execution along with post operative follow up.

Keywords: Image Fusion, Medical Imaging, Spine, Musculo Skeletal

1. INTRODUCTION

For musculo-skeletal diseases and injuries, a patient is commonly required to receive both a CT and an MR scan to reliably assess the damaged tissue. The MR image will provide good contrast and show important details of the soft tissues such as the nerves and the spinal discs. The CT image will show excellent details of the bone structures. In order to provide the best possible view of the patient, it is desirable to provide an image that combines the medically significant details from both the MR and the CT.

Our work is targeting the spine and offers a challenging use case because of the intricate connections between the bone tissue and soft tissue. Additionally, since up to of 75% of adults suffer from back pain, with many of those patients requiring hospitalisation (1.8% of all hospital stays in the United States),[1] there is a medical need for enhanced spine visualization. The applications of this work are first allowing better diagnosis with a single fused image, better surgery planning and execution, and finally post-operative follow up fusing a post-operative CT to the pre-operative MR image. This method will provides a new view of the patient combining two modalities into a single image.

The fusion process starts with pre-registered MR and CT images of the spine (rigid registration can be used to accomplish this). A new image can be created that incorporates the medically significant details from both the MR and CT images. This is then followed by applying an image fusion algorithm to combine two input images into a single output image.

Further author information: (Send correspondence to)
B.M. : E-mail: bmiles@uwo.ca, Telephone: (519)646-6000 ext:64626

Medical Imaging 2012: Computer-Aided Diagnosis, edited by Bram van Ginneken, Carol L. Novak,
Proc. of SPIE Vol. 8315, 831511 · © 2012 SPIE · CCC code: 1605-7422/12/$18 · doi: 10.1117/12.911613

Image fusion has been studied in other fields, and applications vary from multifocal images[2] to geographical.[3] In medicine, image fusion has been used for brain imaging,[4,5] MRI-SPECT fusion[6] and Epilepsy treatment planning.[7]

For applications regarding the spine, registration of CT and MR spine images have been used for surgery planning[8,9] and for evaluation of bone implants.[10] However, it should be noted that these works have focused on registration with very simple overlay methods and are not pixel-wise image fusion.

A wavelet based approach has been proposed by Li.[2] The two images are fused in the wavelet domain and an inverse transformation is applied to produce the resulting image. Variations of this technique include additive wavelet decomposition,[3] the contourlet transform[4] and the complex wavelet transform. Variational methods have also been employed to find the minimum of an energy functional.[11,12] There are drawbacks to these methods however. The wavelet based approaches are susceptible to pixelation effects due to shift invariance, especially with two very dissimilar input images. The variational methods do not guarantee a global solution and can get trapped in a local minimum.

We propose a pixel-wise graph cut based[13–15] image fusion method that will guarantee a global solution, while avoiding the pixelation artifacts that can be present with the use of wavelet based image fusion methods. To determine the accuracy of the fusion method, 20 sagittal spine image slices from 9 different patients were fused and compared to manual ground truths.

2. METHODS

2.1 Data set Preparation

Pre-registered CT and MR datasets where used as input data sets. They were rigidly registered based on corresponding soft tissue detail with a TRE of 3.7 +/- 0.8 mm. The CT images were also thresholded at 100 Houndsfield units, to preserve only the bone details. Finally the intensities were normalised to correspond with the dynamic range of the MR images.

2.2 Fusion Algorithm

Pixel level image fusion of the registered image pairs is achieved by minimizing an energy functional. The fusion algorithm aims to create a fused image that is:

1. Similar to the source CT image.
2. Similar to the source MR image.
3. Smooth (containing natural variations between pixels).

This is represented in the following energy minimisation equation from Wang.[12]

$$u(x) = \text{argmin}_{u\in\Omega} \left\{ \iint_\Omega \left[w_1(x) \left(u(x) - u_{CT}(x)\right)^2 + w_2(x) \left(u(x) - u_{MR}(x)\right)^2 \right] dx + 2\lambda \iint_\Omega |\nabla u(x)|\, dx \right\} \tag{1}$$

Where $u(x)$ is the desired output fusion image. The weights w_1 and w_2 are the convolution of the gradients of the inputs (u_{MR} and u_{CT}) with a 3x3 box filter applied as shown in Wang.[12] This is an excellent model for a fusion algorithm because the gradient magnitude of the weights implies favouring high gradients (edges) in the output image, while pixel similarities to the input images are also favoured. This means preserving both the pixel intensity information and the edge information in the final image, which will correspond to the the important details like edges of the bone, edges of the nerves and edges of the discs, while favouring similar intensity values in the final image. The smoothness term $2\lambda \iint_\Omega |\nabla u(x)|$ serves to enforce natural transitions in the image.

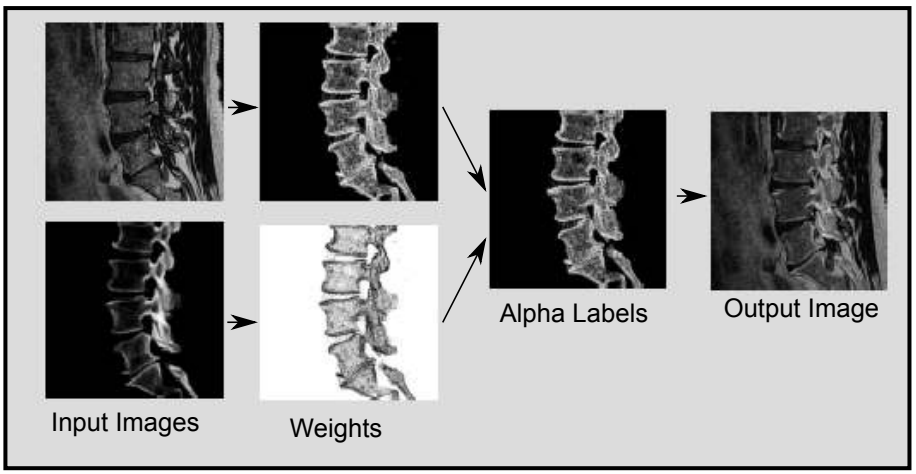

Figure 1. Diagram of the fusion process, left to right - registered input images, weighted images, alpha image, fused images

The above optimisation problem can be solved using graph cuts.[13–15] The Potts model for discontinuity preservation preserves both the edge discontinuities from the weights and the similarity in pixel levels from the similarity terms. The graph cut formulation is:

$$\begin{aligned} D(x) &= w_1(x)\left(u(x) - u_1(x)\right)^2 + w_2(x)\left(u(x) - u_2(x)\right)^2 \\ V(p_i, p_j) &= \min\ \ c_1, (l_i - l_j)^2 \end{aligned} \tag{2}$$

where $D(x)$ denotes the data cost and $V(p_i, p_j)$ denotes the smoothness cost. Also p_i and p_j are two adjacent nodes with l_i and l_j being values of their respective labels. The value c_1 is a constant to help preserve discontinuities. This will provide a global labeling that corresponds to the output intensities of the fused image at each pixel, solving the fusion problem.

Although the method (2) provides a high quality fused image with 256 labels, it is not feasible for high bit depth medical images such as DICOM images. The number of computational steps required grows polynomially with the number of labels used. This makes processing 16bit (high depth) CT or MR diagnostic images unfeasible because of the required computational expense.

The problem can be reformulated indirectly for targeting fusion of DICOM images. Pixel-wise image fusion can be considered as finding an alpha image, α, where:

$$u(x) = \alpha(x)u_{CT}(x) + (1 - \alpha(x))u_{MR}(x)\ \forall\ \alpha(x, y)\ \ \in [0, 1] \tag{3}$$

This alpha image dictates the percentage of a resulting pixel that will come from image u_{CT} and what percentage will come from image u_{MR}. The data and smoothness terms required for the graph cut optimisation can be formulated as shown in (4). The entire fusion process is detailed in figure 1.

$$\begin{aligned} D(x) &= w_1(x)\left(u_\alpha(x) - u_{CT}(x)\right)^2 + w_2(x)\left(u_\alpha - u_{MR}(x)\right)^2 \\ u_\alpha &= \alpha(x)u_{CT}(x) + (1 - \alpha(x))u_{MR}(x) \\ V(\alpha_i, \alpha_j) &= \min\left(c_1, |l_i - l_j|\right) \end{aligned} \tag{4}$$

2.3 Validation of Fused Images

Relevant anatomy from the input images were segmented, in order to validate the results of the image fusion. For the bone images, this included the bone cortex. For the MR images, this included the discs, spinal cord, cerebral spinal fluid and nerve tissues. The fused images were compared to the input MR images in areas of soft

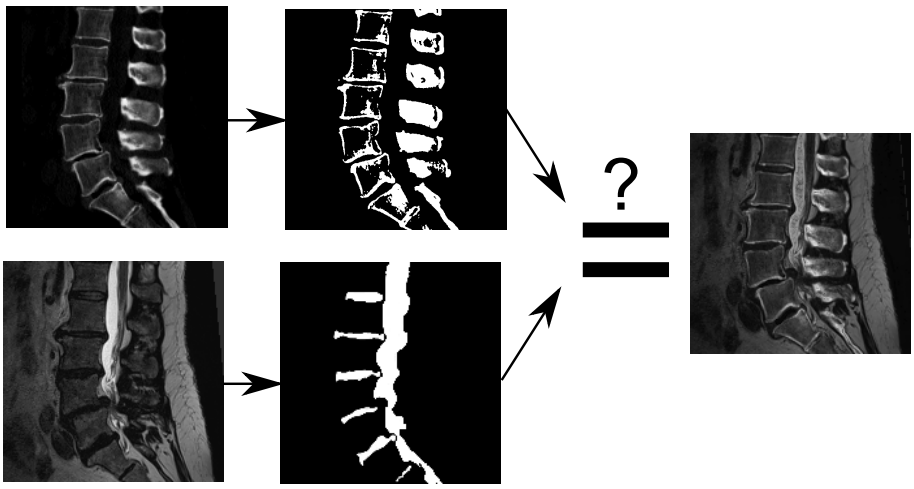

Figure 2. Sample masks for validating the quality of the fusion algorithm. Top, left and centre - CT input image and bone detail mask, Bottom left and middle - MR input image and soft tissue detail mask, Right - sample fused image

tissue, and then the input CT images in areas of bone detail, determining how much of the clinically relevant detail was successfully transferred from the input images to the fused images. This comparison was evaluated by a root-mean-squared (RMS) of the differences in the input and fused images is given by:

$$E_{\mathrm{RMS}} = \sqrt{\frac{\sum_{\Omega} \left(I_{\mathrm{input}} - I_{\mathrm{fused}}\right)^2}{\#\text{ of pixels in the image}}}$$

where I_{input} is the intensity of the input image for a given pixel, I_{fused} is the intensity of the fused image at a given pixel and Ω is the image domain.

To validate that this method was correct, the reverse was tested. The MR input image was compared with the fused image based on the bone details, and the CT input image was compared with the fused image based on the soft tissue detail. It could then be shown whether the results were statistically significant. Figure 2 illustrates this comparison process showing sample segmented images.

3. EXPERIMENTS

A total of 20 saggital lumbar spine slices from 9 patients have been used to validate the proposed graph cut method. A comparison was made with 2 methods found in the literature, namely: the discreet wavelet transform (DWT)[2] and the contourlet transform (CLT),[4] along with simply averaging the two input images. Sample images, statistical results and case studies are presented in this section.

3.1 Sample Images

Sample input images are shown in figure 3. Registered CT and MR images of the same subject have been provided and are also shonw in figure 3.

A few things can be noted based on figure 3. Although all methods do show both the MR and CT details, there are problems with some of the fusion schemes. The simple pixel-wise averaging method has a tendency to lose important details because it averages the modalities. The wavelet method suffers from pixelation artifacts (the output has block structures larger than individual pixels, as opposed to smooth edges) because it is not shift invariant. The contourlet has fewer artifacts, but it blurs the CT details. The graph cut method shows both sharp MR detail and sharp CT detail without artifacts.

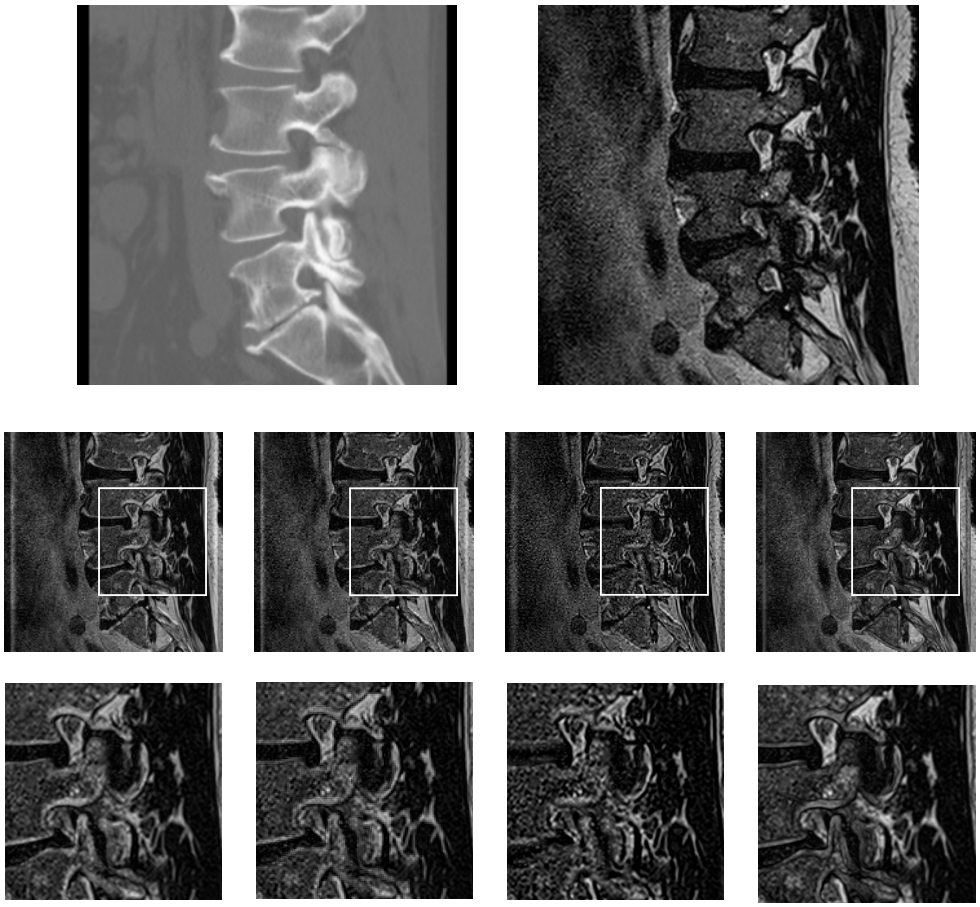

Figure 3. Sample fused Images: top left - input CT, top right input MR, middle: fused images from left to right, averaging method, discreet wavelet transform, contourlet, graph cut methods, bottom: close ups of the images.

3.2 Statistical Results

The results were validated by comparison with manually segmented tissue areas as detailed in section 2.3. The hypothesis was that the fused image would contain the soft tissue such as discs, nerves, cerebral spinal fluid and the spinal cord from the MR image, along with the bone details from the CT image. A comparison was made between the four different methods. Average differences in pixel intensity based on RMS error have been calculated. Additionally statistical significance has been tested using a T-test. The hypothesis is firstly that more detail will be transferred to the fused image, in the areas of soft tissue from the MR image, than from the CT and secondly that more detail will be transferred to the fused image in the areas of the bone cortex from the CT, image rather than the MR image. A pairwise single tailed T-test was used in this case.

The results are fairly consistent showing statistical significance for all methods except for averaging. Since the output of the average image will be halfway between the CT and MR images, the results are the same and the T-test will show no significance. For both the DWT and the CLT method although the results are significant, there is more soft tissue detail transferred from the CT image to the fused image than the from MR image to the fused image! This is the opposite of what is expected and means that these fusion algorithms are not providing the correct result.

The graph cut methods produces lower average errors for the MR data than all the other methods. It also produces lower average errors for the CT data than every method except the contourlet transform. However, as noted above, the contourlet transform is not producing expected results. For the graph cuts, there is an error of 89.9 pixels for the bone data, which means more than 85% of bone data is transferred to the fused image. For the soft tissue the error is 34.9 which means 95% of the soft tissue information is transferred to the fused images.

Table 1. Average sum of squares of differences and statistical significance. MR - Soft Tissue is the average pixel-wise intensity difference between the soft tissue areas (disc, spinal cord and nerve tissue) in the MR image and the fused image for all slices, The CT - Bone is the average pixel-wise intensity difference between the bone cortex areas in the CT image and the fused image for all slices. CT - Soft Tisse is the difference between soft tissue areas in the CT image and the fused image for all slices. MR - bone is the difference between the bone cortex areas in the MR image and the fused image for all slices. These numbers are based on dynamic range of 700. * See text, significance is shown, but in the wrong direction.

	MR - Soft Tissue	CT - Soft Tissue	P - Value	CT-Bone	MR-Bone	P - Value
Average	82.0	82.0	1	91.7	91.7	1
DWT*	83.5	81.4	< 0.0001	90.8	94.1	< 0.0001
CLT*	91.5	78.1	< 0.0001	84.9	105.2	< 0.0001
Graph Cuts	34.9	142.8	< 0.0001	89.8	101.5	0.003

3.3 Visual Inspection - Lumbar Spine, Joint and Disc Disease

The visual results of this method are shown in figure 4. The first case shows a patient with a protruding spinal disc and a damaged facet joint, see figure 4. The protruding disc can be seen in MR image, whereas the damage facet joint is visible in the CT image. There is significant osteoarthrytis in the joint. The fused image clearly shows both of these pathologies in a single image allowing for a better diagnosis.

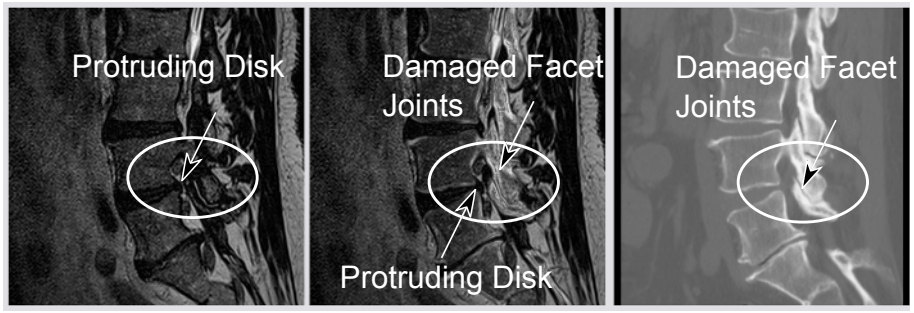

Figure 4. Images of damaged spine, left MR - showing bulging disk, middle - Fused image showing the disk, the spinal cord and the damaged facet joints, right - CT image with damaged facet joints.

3.4 Visual Inspection - Osteophyte Growth

The second case study shows osteophyte growth (see figure 5), this is the formation of boney 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 boney spurs and the surrounding soft tissue on a single image.

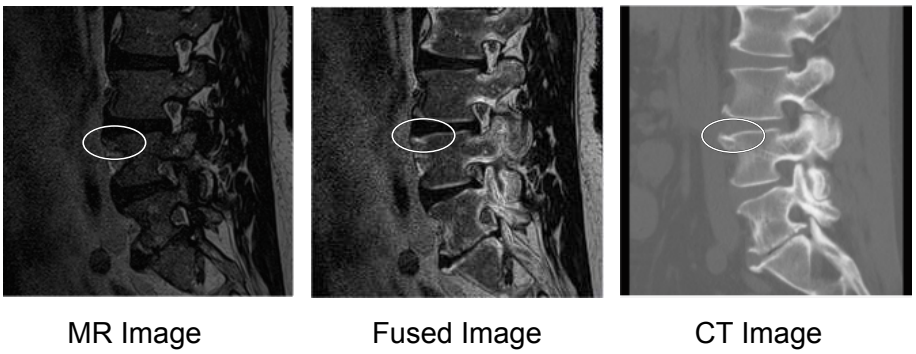

Figure 5. Images of boney spur formation, left MR - osteophyte not identifiable, middle Fused image - the osteophyte is clearly visible along with surrounding soft tissue, right CT image showing the osteophyte, but no soft tissue.

4. CONCLUSIONS

We have proposed a novel image fusion algorithm and applied it to images of the spine as a case study. We successfully fused MR and CT images to create a single new image providing a new and effective combined modality for the images to be assessed. This has been validated to show statistical significance and by visual inspection of the resulting images. The graph cut results are statistically significant with better performance than the other methods tested. The graph cut method successfully transfers 85% of the bone detail to the fused image and 95% of the soft tissue detail. This allows for better assessment of disease including disc disease and osteophyte growth.

In the future we plan to apply this method to other parts of the musculo-skeletal system including the upper and lower limbs. Interventional use of the proposed method is also an interesting research direction.

REFERENCES

[1] Owens, P. L., Woeltje, M., and Mutter, R., "Emergency Department Visists and Inpatient Stays Related to Back Problems," (2008). http://www.hcup-us.ahrq.gov/reports/statbriefs/sb105.pdf.

[2] Li, H., Manjunath, B. S., and Mitra., S. K., "Multisensor Image Fusion using the Wavelet Transform," *Graphical Models and Image Processing* **57**(3), 235–245 (1995).

[3] Núűez, J., Otazu, X., Fors, O., Prades, A., Palà, V., and Arbiol, R., "Multiresolution-Based Image Fusion with Additive Wavelet Decomposition," *IEEE Transactions On Geoscience And Remote Sensing* **37**, 1204–1211 (May 1999).

[4] Yang, L., B.L.Guo, and W.Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing* **72**, 203–211 (2008).

[5] Matsopoulos, G. K., Marshall, S., and Brunt, J. N. H., "Multiresolution morphological fusion of MR and CT images of the human brain," *IEEE Proceedings on Vision Image Signal Processing* **141**(3), 137–142 (1994).

[6] Colin, A. and Boire, J.-Y., "MRISPECT fusion for the synthesis of high resolution 3D functional brain images: a preliminary study," *Computer Methods and Programs in Biomedicine* **60**, 107–116 (1999).

[7] Wong, S. T. C., Knowlton, R. C., Hawkins, R. A., and Laxer, K. D., "Multimodal Image Fusion for Non-invasive Epilepsy Surgery Planning,," *IEEE Transactions on Computer Graphics and Applications* **16**(1), 30–38 (1996).

[8] Sohn, M.-J., Lee, D.-J., Yoon, S. W., Lee, H. R., and Hwang, Y. J., "The effective application of segmental image fusion in spinal radiosurgery for improved targeting of spinal tumours," *ACTA Neurochir* **151**, 231–238 (2009).

[9] Hua, Y., Mirzac, S. K., Jarvikb, J. G., Heagerty, P. J., and Haynor, D. R., "MR and CT image fusion of the cervical spine: a noninvasive alternative to CT-Myelography," in [*Proceedings of SPIE*], **5744**, SPIE (2005).

[10] Karlo, C. A., Steurer-Dober, I., Leonardi, M., Pfirrmann, C. W. A., Zanetti, M., and Hodler, J., "MR/CT image fusion of the spine after spondylodesis: a feasibility study," *European Spine Journal* **19**, 1771–1775 (2010).

[11] Piella, G., "Image Fusion for Enhanced Visualization: A Variational Approach," *International Journal of Computer Vision* **83**(1), 1–11 (2009).

[12] Wang, W.-W., Shui, P.-L., and Feng, X.-C., "Variational Models for Fusion and Denoising of Multifocus Images," *IEEE Signal Processing Letters* **15**, 65–68 (2008).

[13] Boykov, Y., Veksler, O., and Zabih, R., "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)* **23**, 1222–1239 (November 2001).

[14] Boykov, Y. and Kolmogorov, V., "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)* **26**, 1124–1137 (September 2004).

[15] Kolmogorov, V. and Zabih, R., "What Energy Functions Can Be Minimized via Graph Cuts?," *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)* **26**, 147–159 (February 2004).