Efficient Multi-Modal Least-Squares Alignment of Medical Images Using Quasi-Orientation Maps

A. Wong
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada

W. Bishop
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada

J. Orchard
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

Abstract - *In image registration, similarity metrics are used to determine the optimal alignment between two images. A common metric used for judging image similarity is the weighted sum of squared differences (SSD) cost function. Recently, it was demonstrated that the evaluation of the SSD cost function can be performed efficiently using the Fast Fourier Transform (FFT) to determine the optimal translation between two images based on pixel intensities. This paper extends this efficient approach by introducing the concept of quasi-orientation maps as features into the alignment framework. This feature-based method is invariant to intensity mappings, making it suitable for aligning medical images acquired with different modalities. Experimental results demonstrate overall multi-modal image alignment performance to be superior to that of previous work.*

Keywords: multi-modal registration, image alignment, medical imaging, orientation matching, least-squares

1 Introduction

Image registration is the process of determining the alignment between two images that have overlapping regions. This is a difficult task because the images may be acquired at different times and/or by different sensors. The difficulty is increased significantly in the case of multi-modal image registration, where the images are acquired using different imaging techniques. In such cases, images that share similar content may have very different intensity mappings. One area where the alignment of images of different modalities is important is in the field of medical imaging. Different medical imaging techniques such as x-rays, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) reveal different types of information about the body that can be used in the diagnosis of a disease. An example of this can be seen in Figure 1, which shows a slice from a cranial T1-weighted MRI scan and its corresponding CT scan. The bone structure of the head is clearly visible in the CT scan, as indicated by the bright areas in the image. The bone structure has a higher radiodensity than the surrounding tissue. This is not the case for the MRI scan. Conversely, tissue differences can be clearly distinguished from the MRI scan but not from the CT scan. Therefore, the alignment of images of different modalities allows for a better overall picture of a patient's condition.

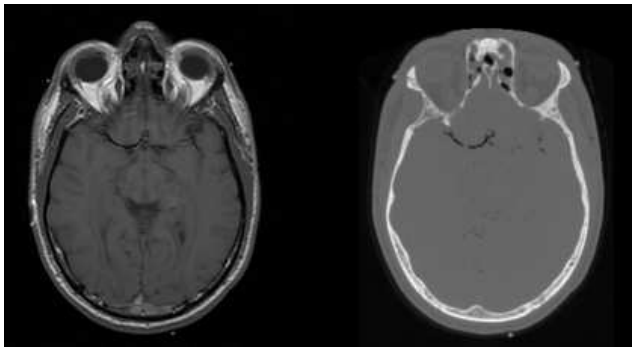

Figure 1: Left: T1-weighted MRI cranial slice
Right: Corresponding CT cranial slice

A number of different techniques have been proposed for the purpose of image registration. These methods are often classified based on: 1) the feature space used as the basis of comparison, and 2) the similarity metric used to compare the two images. Common feature spaces include intensity, edges, corners, regions, and contours. A number of different similarity metrics have been used in image registration. These include:

- cross-correlation [1]-[3]
- phase correlation [4]-[7]
- sum of squared differences (SSD) [8]-[10]
- mutual information [11]-[13]

Cross-correlation is commonly used for determining the translation alignment between two images because the correlation between the two images for all possible translations can be evaluated efficiently in the frequency domain by making use of the Fast Fourier Transform (FFT). Recently, it was demonstrated that the evaluation of a weighted SSD cost function can be accelerated through the use of the FFT [8][9]. This novel approach was shown to yield performance improvements ranging from 60 to over 500 times faster than the direct approach

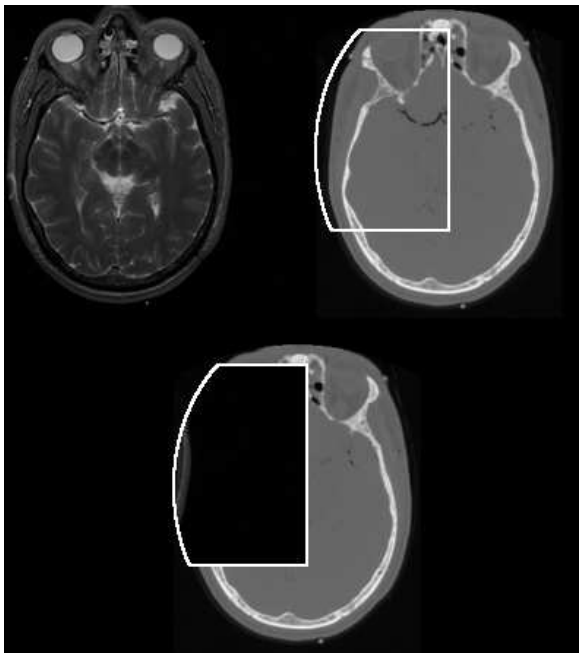

Figure 2: Top-Left: T2-weighted MRI cranial slice
Top-right: corresponding CT cranial slice
Bottom: alignment (SSD cost function with linear remapping)

of evaluating the SSD cost function for each individual shift [9]. Furthermore, the algorithm proposed in [9] also optimizes the contrast and brightness adjustment of the reference image to correspond with the other image by finding the best match over all linear intensity remappings. While this approach performs well for cases where the intensity mappings used for the two images are similar in nature due to the linear remapping, it is a purely intensity-based method and therefore does not explicitly take advantage of the structural characteristics of content in the images. Therefore, it is not suitable for the alignment of images with very different intensity mappings where no suitable brightness and contrast adjustment can be made to make the images similar on an intensity level. As an example, this is evident when aligning T2-weighted MRI slices with CT slices, as shown in Figure 2. Since the lower intensity regions in the CT slice correspond to the higher intensity regions in the MRI slice and the higher intensity regions in the CT slice correspond to the lower intensity regions in the MRI slice, and the dark background in the CT slice corresponds to the dark background in the MRI slice, no appropriate linear remapping exists between the two images. Thus, the correct image alignment cannot be achieved using the SSD cost function with a linear remapping. The goal of this paper is to extend this efficient alignment framework to address this issue by introducing the concept of quasi-orientation maps as a feature space over the use of pixel intensities to improve the level of invariance to intensity mappings.

The main contribution of this paper is an efficient feature-based SSD-type image alignment algorithm based on quasi-orientation maps. This method is highly invariant to intensity mappings and therefore suitable for aligning medical images acquired with different modalities. In this paper, the theory behind weighted SSD cost function evaluation using the FFT framework and the theory underlying quasi-orientation maps is presented and explained in Section 2 along with an outline of the proposed algorithm. The testing methods and test data are outlined in Section 3. Finally, experimental results comparing the proposed algorithm with previous research are discussed in Section 4, and conclusions are drawn based on the results in Section 5.

2 Theory

Before outlining the proposed alignment algorithm, it is important to discuss the theory behind the key components of the algorithm. The cost function evaluation is based on the technique introduced in [8] and [9]. Hence, the basic formulation of a weighted SSD cost evaluation using the FFT framework is presented to provide a context for this work. More importantly, the concept of quasi-orientation maps is described and explained in detail to justify the use of such a feature space over intensities for the purpose of comparing two images of different modalities.

2.1 Weighted SSD Cost Function Evaluation using the FFT

Given the 2-D images *f* and *g*, the similarity between the two images within a region of interest (ROI) in *g* can be found by evaluating the weighted SSD cost between the images,

$$SSD = \sum_{\chi} \left(f(\chi) - g(\chi) \right)^2 w(\chi), \qquad (1)$$

where $w(\chi)$ is a weighting function over $g(\chi)$ in the range [0,1] (in the case of a ROI, 0 indicates the regions outside the ROI and 1 indicates the regions inside the ROI), and χ = (*x*,*y*) which indicates the coordinate of a pixel. Using this cost function, a lower SSD cost indicates a higher image similarity. Likewise, the SSD cost between a shifted image *f* and an image *g* is

$$SSD(\delta) = \sum_{\chi} \left(f(\chi - \delta) - g(\chi) \right)^2 w(\chi), \qquad (2)$$

where δ = (*m*,*n*) represents the shift. Therefore, to find the shift of *f* that provides the global optimal alignment of images *f* and *g*, it is necessary to find the shift δ = (*m*,*n*) that results in the lowest SSD cost function, as given by

$$\delta = \arg_{\delta} \min \left[\sum_{\chi} \left(f(\chi - \delta) - g(\chi) \right)^2 w(\chi) \right]. \quad (3)$$

This is commonly known as the least squares problem. If done in a direct fashion, each shifted image of *f* needs to be compared to *g* using the SSD cost function. This evaluation process is computationally expensive and so a more efficient way to determine the shift associated with the minimum SSD cost is desired. As described in [8] and [9], Equation 2 can be expanded and expressed as

$$\begin{aligned} SSD(\delta) = & \sum_{\chi} \left(f^2(\chi - \delta) w(\chi) \right) \\ & -2 \sum_{\chi} \left(f(\chi - \delta) g(\chi) w(\chi) \right) + \sum_{\chi} \left(g^2(\chi) w(\chi) \right) \end{aligned}. \quad (4)$$

The last term of Equation 4 is irrelevant to the optimization problem and can be ignored, as it is independent of δ. The first two terms can be turned into convolutions. Therefore, the final weighted SSD cost function $PSSD(\delta)$ is expressed as

$$PSSD(\delta) = \left[\bar{f}^2(\chi) * w(\chi) \right]_{\delta} - 2 \left[\bar{f}(\chi) * \left(g(\chi) w(\chi) \right) \right]_{\delta}, \quad (5)$$

where $\bar{f}(\chi) = f(-\chi)$ and $*$ indicates a convolution. The bottleneck of this final cost function is the evaluation of the convolution operations. These can be evaluated very efficiently in the frequency domain due to the fact that convolution in the spatial domain become multiplication in the frequency domain. Therefore, the terms in Equation 5 can be evaluated for all possible values of δ by taking the Fourier transform of the convolution terms, performing multiplication in the frequency domain, and then taking the inverse Fourier transform of the result, which can be expressed as

$$\begin{aligned} PSSD(\delta) = & F^{-1} \left[F(\bar{f}^2(\chi)) F(w(\chi)) \right](\delta) - \\ & 2F^{-1} \left[F(\bar{f}(\chi)) F\left(g(\chi) w(\chi) \right) \right](\delta) \end{aligned}. \quad (6)$$

Therefore, the performance bottleneck becomes the computation of the FFTs and the inverse FFTs (IFFTs). For large image sizes, significant performance improvements can be achieved by evaluating the cost function in this manner since the FFT and the IFFT can be evaluated efficiently.

2.2 Quasi-Orientation Maps

The primary drawback for the use of pixel intensities as the feature space for multi-modal image registration is the fact that images acquired with different modalities can have significantly different intensity mappings. This makes it very difficult to evaluate similarity between images based on pixel intensities. Therefore, a feature space that is invariant to intensity mappings is highly desired. One such feature space is the gradient orientation, which can be calculated using the formula

$$O(x, y) = \tan^{-1} \left(\nabla_y / \nabla_x \right), \quad (7)$$

where ∇_y and ∇_x are the partial derivatives in the y direction and x direction, respectively. Since the gradient orientation is a measure of angle, it is invariant to the variations in image contrast and illumination. More importantly, this measure of gradient orientation provides a good representation of the structural characteristics of image content irrespective of the intensity mapping. This property can be visualized using a simple example. Imagine a straight line on a constant background being processed with two different intensity mappings to create two images. In the first intensity mapping, the straight line is mapped to black and the background is mapped to white. In the second intensity mapping, the straight line is mapped to white and the background is mapped to black. This scenario is analogous to the T2-weighted MRI to CT alignment scenario given in Section 1. It would be impossible to find a match between the two generated images using the SSD cost function if pixel intensity is used as a feature space directly. However, the gradient orientation computed using Equation 7 would be identical in the two images, making SSD cost evaluation suitable for finding image alignment.

A number of issues prevent the aforementioned gradient orientation measure from being used directly within a SSD cost evaluation framework. First, the gradient measure is highly sensitive to image noise. Any minor change in pixel intensities would result in the existence of a gradient direction regardless of the gradient magnitude. For example, a constant increase in pixel intensity in only one direction should yield a uniform gradient orientation map. However, the existence of noise would result in a gradient orientation map with many different gradient orientations. Second, the gradient measure does not yield a quantitative representation of an absence of gradient orientation (e.g., pixels in a uniform region) that can be used for evaluation in a SSD cost function. Finally, it is not possible to use just gradient orientation to distinguish uniform regions belonging to the foreground object and a uniform background, as is often the case with medical images where the background is represented as uniformly black. Therefore, the goal is to introduce the concept of quasi-orientation maps, a hybrid feature space that makes use of gradient orientation information while addressing these issues.

First, it is necessary to reduce the effect of noise on the effectiveness of the feature space for use in similarity evaluation. This can be accomplished by taking into account gradient orientation information only at locations where the gradient amplitude is above a reasonable threshold *t*. As gradient amplitude is a good indicator of structural feature significance (such as corners and edges), the gradient orientation at these locations provides a good representation of the structural characteristics of the image content while significantly reducing the effect of noise on regions with low structural feature significance.

Second, it is necessary to establish a quantifiable relationship between pixels with gradient orientations and those pixels that lack gradient orientations to permit evaluation using a SSD cost function. Let type-A pixels represent pixels with gradient orientations and type-B pixels represent pixels without gradient orientations (e.g., pixels in uniform regions). A way that a relationship can be established between type-A pixels and type-B pixels is through the use of an offset ε, which represents the distance between the two types. The value of ε should be chosen large enough such that the minimum difference between type-A pixels and type-B pixels is greater than the maximum difference between the gradient orientations in type-A pixels. For example, Equation 7 yields gradient orientations within the range of [-90, 90]. Since SSD is used as a cost function, the orientation maps for both *f* and *g* can be remapped to the range of [0, 180] without affecting the outcome. In this case, an offset of ε = 181 can be used such that type-B pixels are set to 0 and type-A pixels are in the range of [181, 361]. The minimum difference between type-A pixels and type-B pixels would then be 181, which is more than the maximum difference between the gradient orientations in type-A pixels (which would be 361-181=180). By using an offset, uniform areas will more likely be matched with other uniform areas rather than being ignored, while areas with similar structural characteristics will be matched to each other.

Finally, it is important to distinguish the foreground from the background so that uniform regions from the foreground objects are not matched to the uniform background. This can once again be accomplished through the use of an offset γ. Pixels belonging to the background are assigned a value such that the SSD between background pixels and foreground pixels are large enough to not result in a match. Expanding upon the previous example, the background pixels may be mapped to γ = –181. In practical scenarios, the distinction between foreground objects and background objects is necessarily known prior to image alignment. This is the case in [9], where the ROI is known for the reference image but not known for the other images involved in image registration. In such a case, an adaptive threshold algorithm such as Otsu's algorithm [14] may be used. Using the aforementioned techniques, quasi-orientation maps can be constructed for use as a feature space for the SSD-based image alignment framework.

2.3 Alignment Algorithm

Based on the theory presented, the alignment algorithm can be outlined as follows:

1) Given images *f* and *g*, construct quasi-orientation maps *f'* and *g'* in the following manner:
   a) Compute the gradient magnitude *G* and orientation *O* of the image using Equation 7.
   b) Compute a binary image *F* that separates the foreground objects from the background using an adaptive thresholding method such as Otsu's algorithm (where 1 indicates foreground objects and 0 indicates the background)
   c) Apply morphological close operator to binary image to fill in some of the holes in the image.
   d) Compute the quasi-orientation map *QO* using

$$QO(x,y) = \begin{cases} O(x,y) + \varepsilon & \text{if } G(x,y) > t \\ 0 & \text{if } G(x,y) \le t \ \& \ F(x,y) = 1 \\ \gamma & \text{if } G(x,y) \le t \ \& \ F(x,y) = 0 \end{cases} \quad (8)$$

2) Using *f'* and *g'*, perform weighted SSD cost evaluation using FFT as described in Section 2.1 to determine the optimal shift between *f* and *g*.

3 Testing Methods

The proposed algorithm was implemented in MATLAB based on code that was used in [9]. In the current implementation, the optimal translation between two images is computed based on the lowest SSD cost within a selected region of interest (ROI). The proposed algorithm was tested using three sets of images derived from the Visible Male dataset of the National Library of Medicine's Visible Human Project. All test images are 8-bit grayscale images 256×256 in size. A description of each test set is given below.

- **THORAX:** A set of 5 images of a coronal slice through the thorax. These include: T1-, T2-, PD-weighted MRI, CT, and a grayscale photo.
- **HEAD:** A set of 4 images of an axial slice through the head. These include: T1-, T2-, PD-weighted MRI, and CT.

Table 1: Image Alignment Performance

Test Set	DIRECT		LINEAR		QUASI	
	Mean Alignment Error[1]	% Error < 3 Pixels[2]	Mean Alignment Error[1]	% Error < 3 Pixels[2]	Mean Alignment Error[1]	% Error < 3 Pixels[2]
THORAX	58.68	29.50	49.68	39.50	12.59	63.00
HEAD	30.29	26.16	54.67	40.83	7.17	83.33
PELVIS	22.86	70.00	4.79	94.17	3.28	95.83
Overall	**44.06**	**34.84**	**44.17**	**48.55**	**9.39**	**74.60**

1: The mean alignment errors are computed over 400 test alignments for THORAX, 240 for HEAD, and 120 for PELVIS.
2: The % error < 3 pixels is computed based on 400 test alignments for THORAX, 240 for HEAD, and 120 for PELVIS.

- **PELVIS:** A set of 3 images of a coronal slice through the pelvis and upper thigh. These include: T1-, T2-, PD-weighted MRI.

Each image in a test set is aligned with every other image in the same test set over a set of 20 randomly generated ROIs using one of the three alignment methods:

1) SSD evaluation using pixel intensities directly (DIRECT)
2) SSD evaluation using pixel intensities with linear remapping as described in [9] (LINEAR)
3) SSD evaluation using quasi-orientation maps (QUASI)

To judge the performance of the alignment methods, the mean alignment error and the percentage of image alignments with alignment errors less than or equal to 3 pixels (1.17% of image dimension) were computed. The alignment error is calculated as the Cartesian distance from the estimated shift to the gold standard shift.

4 Experimental Results

The experimental results are shown in Table 1. It can be observed that the proposed QUASI method showed noticeably better performance over the DIRECT method in all test cases. Furthermore, the QUASI method performed noticeably better than the LINEAR approach in the THORAX and HEAD test sets and slightly better results in the PELVIS test set. An example of the alignment achieved using the three methods for each test set is shown in Figures 3, 4, and 5. The QUASI method performed particularly well in aligning T2-weighted MRI with CT when compared to the DIRECT and LINEAR methods, as shown in Figure 6. The intensity mappings of T2-weighted MRI and CT are too different for the two intensity-based methods to find a match. However, since the structural characteristics of the content remain the same despite the intensity mappings, the proposed QUASI method is able to find a match. These results illustrate the advantage of the QUASI method in aligning images where the intensity mappings are very different from each other, as is often the case with medical images acquired using different modalities.

5 Conclusions and Future Work

In this paper, we have introduced a new method for efficient least squares image alignment based on the concept of quasi-orientation maps. Experimental results show that overall alignment accuracy is noticeably improved over previous work. It is our belief that this method can be successfully implemented for multi-modal medical image registration purposes. Future work includes investigating the effectiveness of quasi-orientation maps as a feature space for other cost functions, as well as extending the algorithm to perform non-rigid image registration.

Acknowledgments

The authors would like to thank the National Library of Medicine for the Visual Human Project data.

References

[1] A. Cideciyan, "Registration of ocular fundus images: an algorithm using cross-correlation of triple invariant image descriptors," in *IEEE Engineering in Medicine and Biology Magazine*, 1995, Vol. 14, No. 1, pp. 52-58.

[2] M. Solaiyappan and S. Gupta, "Predictive registration of cardiac MR perfusion images Using geometric invariants," in *Proc. ISMRM*, 2000, Vol. 1, p. 37.

[3] L. Fonesca and M. Costa, "Automatic registration of satellite images", in *Proc. X Brazilian Symposium on Computer Graphics and Image Processing*, 1997, pp. 219-226.

[4] E. Castro, and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, Vol. 9, No. 5, pp. 700–703.

[5] B. Reddy and B. Chatterji, "An FFT-based technique for translation, rotation and scale invariant image registration," in *IEEE Transactions on Image Processing*, 1996, Vol. 5, No. 8, pp. 1266–1271.

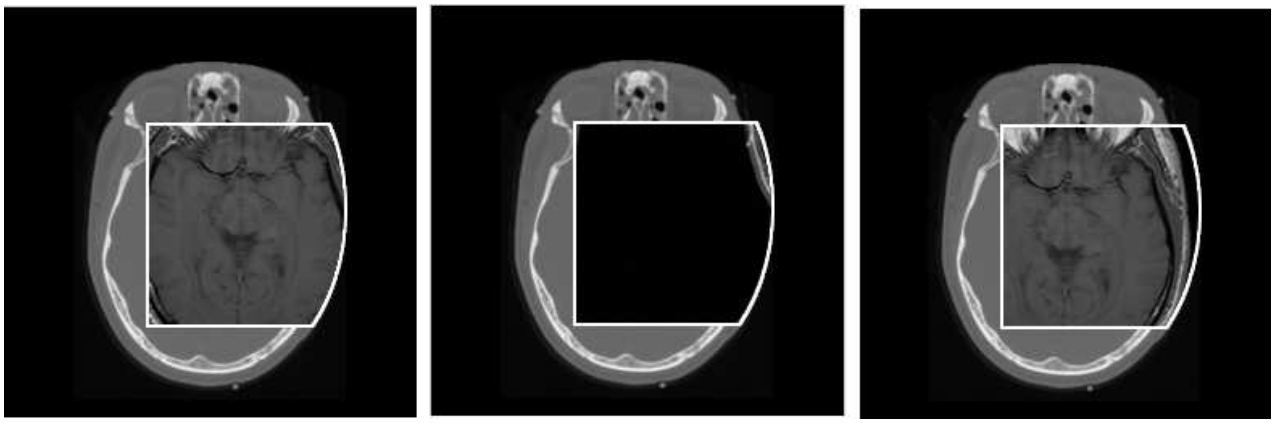

Figure 3: Alignment of T1-weighted MRI with CT from HEAD test set using three methods
Left: DIRECT; Center: LINEAR; Right: QUASI
Only QUASI method resulted in a correct alignment.

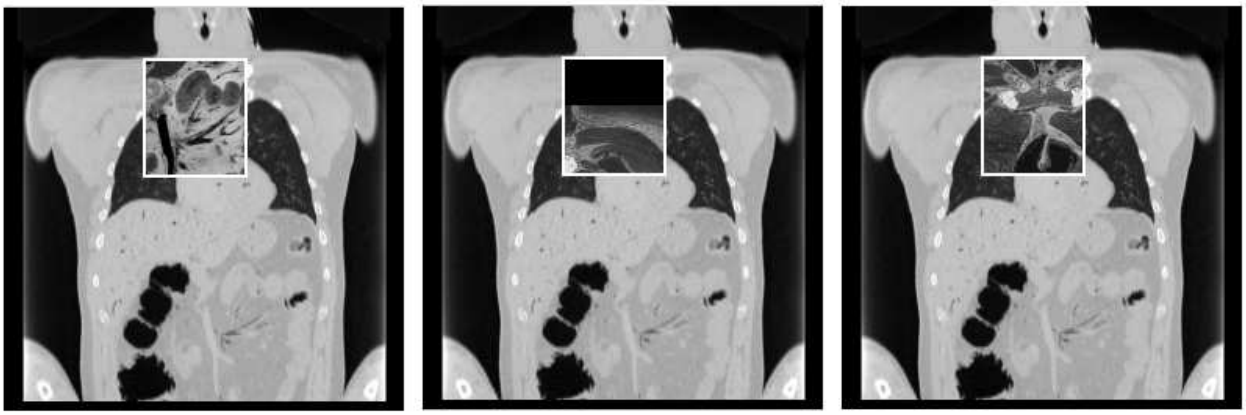

Figure 4: Alignment of grayscale photo with CT from THORAX test set using three methods
Left: DIRECT; Center: LINEAR; Right: QUASI
Only the QUASI method resulted in a correct alignment.

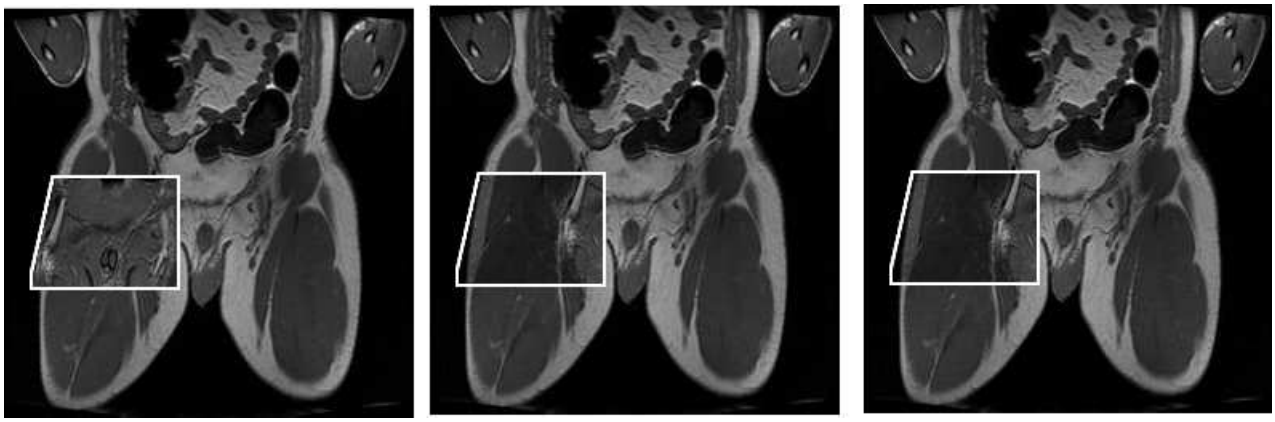

Figure 5: Alignment of PD-weighted MRI with T1-weighted MRI from PELVIS test set using three methods
Left: DIRECT; Center: LINEAR; Right: QUASI
Both the LINEAR and QUASI methods resulted in a correct alignment, while the DIRECT method resulted in an incorrect alignment.

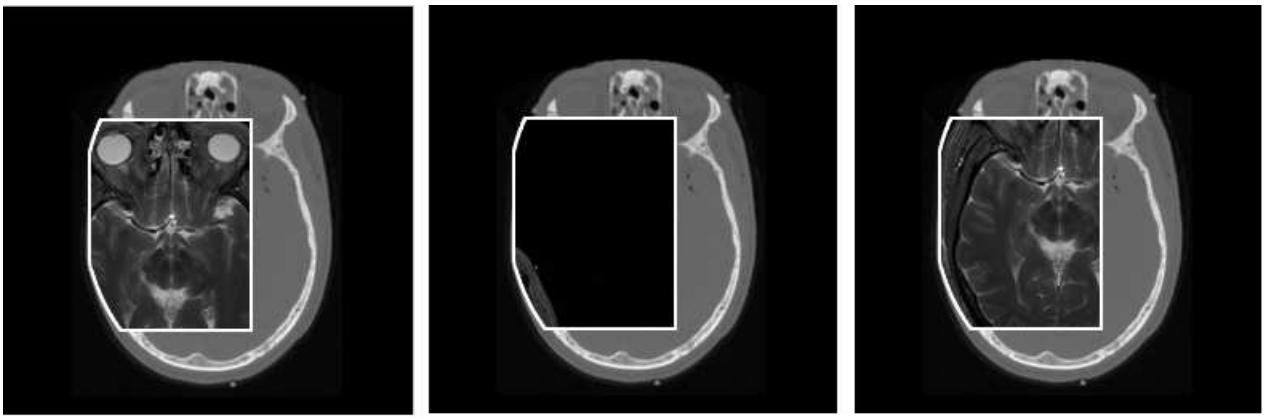

Figure 6: Alignment of T2-weighted MRI with CT from HEAD test set using three methods
Left: DIRECT; Center: LINEAR; Right: QUASI
Only the QUASI method resulted in a correct alignment.

[6] L. Heng, E. Hwa, and H. Kok, "High accuracy registration of translated and rotated images using hierarchical method," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, Vol. 6, pp. 2211-2214.

[7] A. Averbuch and Y. Keller, "FFT based image registration," in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002, Vol. 4, pp. 3608-3611.

[8] A. Fitch, A. Kadyrov, W. Christmas, and J. Kittler, "Fast robust correlation," in *IEEE Trans. Image Processing*, 2005, Vol. 14, No. 8, pp. 1063–1073.

[9] J. Orchard, "Efficient global weighted least-squares translation registration in the frequency domain," in *Image Analysis and Recognition (ICIAR)*, 2005, pp. 116–124.

[10] B. Lucas and T. Kanade, "An interative image registration technique with an application to stereo vision," in *Proc. Imaging Understanding Workshop*, 1981, pp. 121-130.

[11] W. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volumn registration by maximization of mutual information", in *Medical Image Analysis*, Oxford University Press, Vol. 1, No. 1, pp. 35-51.

[12] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multi-modality medical image registration by maximization of mutual information," in *IEEE Transactions on Medical Imaging*, 1997, Vol. 16, No. 2, pp. 187-198.

[13] R. Shekhar and V. Zagrodsky, "Mutual information-based rigid and nonrigid registration of ultrasound volumes," in *IEEE Transactions on Medical Imaging*, 2002, Vol. 21, No. 1, pp. 9-22.

[14] N. Otsu, "A threshold selection method from gray level histograms," in *IEEE Trans. Systems, Man, and Cybernetics*, 1979, Vol. 9, pp. 62-66.