Simultaneous Multi-modal Registration of Multiple Images based on Multi-Dimensional Joint Phase Moment Distributions

Alexander Wong
Systems Design Engineering, University of Waterloo, Canada
a28wong@engmail.uwaterloo.ca

Abstract

In this paper, a novel method for simultaneously registering multiple images acquired from different imaging modalities is presented. The optimal alignment is computed as the set of transformations that minimize the dispersion of the multi-dimensional joint phase moment distribution. Dispersion is measured as the cumulative quadratic orthogonal distance between samples from the joint phase moment distribution and the corresponding multi-dimensional fitted hyperline. The proposed method is designed to be computationally efficient, robust to signal non-homogeneities and noise, and maintains internal consistency amongst all images. Experimental results using real-world medical and remote sensing images show that the proposed method achieves a high level of registration accuracy when simultaneously registering multiple multimodal images.

1 Introduction

A problem of great interest in the field of visual pattern recognition is image registration, where images of the same scene captured under different conditions are aligned with each other. Image registration is important to a wide range of applications, such as environment change analysis, superresolution [1], building extraction [2], and computer-assisted surgery. Much of recent research in image registration has focused on multimodal image registration, where images captured using different imaging modalities are registered together.

Multimodal image registration is a very challenging problem for many reasons. Images captured using different imaging modalities are represented using different intensity mappings, making it difficult to compare images in a direct manner based on intensity values. This situation is further complicated by the presence of signal non-homogeneities, noise, and geometric distortions. Furthermore, there are many cases where it is necessary to register multiple images from different imaging modalities to obtain a more complete picture of the scene (e.g., T1/T2/CT registration, LANDSAT 4/5/7 inter-band registration). Hence, it is important to maintain internal consistency between all the images being registered, which is a difficult task to accomplish in the case of multimodal registration given the different information being captured. Therefore, an automatic method that can address all of the above issues is highly desired.

Many methods have been proposed for the purpose of multimodal registration. Such methods can be generally divided into entropy-based methods [3, 4] and feature-based methods [5, 6]. Current methods have mainly focused on dealing with issues such as signal non-homogeneities, noise, and geometric distortions. However, current methods register images in a pair-wise manner, where only two images of different modalities are registered at a time. To the best of our knowledge, there are currently no multimodal image registration methods that register multiple images from different modalities simultaneously. This is very important for avoiding error accumulation during the registration process.

The main contribution of this paper is a simultaneous approach to the problem of registering multiple images from different modalities. In this paper, the theory behind the proposed method is described in Section 2. The proposed method is described in Section 3, and experimental results using multi-spectral LANDSAT images and T1/T2/PD/CT images are presented in Section 4.

2 Theory

Prior to explaining the proposed method, it is important to first explain the theory behind the proposed method. Suppose we wish to register n different images $f_1, f_2, ..., f_n$ acquired using n different image modalities. The optimal set of transformations $\hat{\mathrm{T}}_1, \hat{\mathrm{T}}_2, ..., \hat{\mathrm{T}}_\mathrm{n}$ that bring the n images into alignment can be formulated as the following optimization problem:

$$\left\{\hat{\mathrm{T}}_1, \hat{\mathrm{T}}_2, ..., \hat{\mathrm{T}}_\mathrm{n}\right\} = \underset{\mathrm{T}_1,...,\mathrm{T}_\mathrm{n}}{\arg\min}\left[C\left(f_1^{\mathrm{T}_1}, f_2^{\mathrm{T}_2}, ..., f_n^{\mathrm{T}_\mathrm{n}}\right)\right] \tag{1}$$

where $f_k^{\mathrm{T}_\mathrm{k}} = f_k\left(\mathrm{T}_\mathrm{k}(\underline{\mathrm{x}})\right)$ and C is the objective function that evaluates the dissimilarity between all images. It can be seen that registration accuracy is heavily dependent on the objective function used. In the context of multimodal image registration, the most widely used objective functions are those based on mutual information [3, 4], which measures the reduction in uncertainty of one image when another image is known. An objective function based on mutual information can be expressed as follows:

$$C(f_1^{\mathrm{T}_1}, f_2^{\mathrm{T}_2}) = -\left(H(f_1^{\mathrm{T}_1}) + H(f_2^{\mathrm{T}_2}) - H(f_1^{\mathrm{T}_1}, f_2^{\mathrm{T}_2})\right) \tag{2}$$

where $H(f_1^{\mathrm{T}_1})$ and $H(f_2^{\mathrm{T}_2})$ are the marginal intensity entropies and $H(f_1^{\mathrm{T}_1}, f_2^{\mathrm{T}_2})$ is the joint intensity entropy. In this case, minimizing the objective function maximizes mutual information, which in effect minimizes the joint intensity entropy. Methods that utilize entropy-based objective functions take advantage of the fact that correctly registered images are characterized by tightly packed joint distributions and that minimizing the joint intensity entropy effectively minimizes the dispersion of the joint distribution.

While entropy-based objective functions have proven to be effective in multimodal registration, there are several important drawbacks when taken into the context of registering multiple images of different image modalities. Entropy-based objective functions, including normalized mutual information, are highly under-constrained in terms of evaluating intensity relationships. Therefore, the convergence planes of such objective functions are highly non-monotonic with many local optima in non-ideal situations [7]. This becomes increasedly problematic as the number of dimensions (in this case, number of imaging modalities) increase, thus making it difficult to converge to the global optima. More importantly, entropy-based objective functions suffer from what is referred to as the "curse of dimensionality". As the number of dimensions increase, the number of samples required to reliably approximate entropy grows exponentially. Therefore, this makes it very difficult to extend entropy-based methods to high-dimensional multimodal registration problems as there are typically insufficient number of samples to estimate entropy properly. Intuitively, it would seem that an alternative approach to measuring the dispersion of joint distributions that is well-constrained and can be efficiently extensible to high-dimensional problems is much desired.

To construct a multimodal objective function that is more well-constrained than entropy-based functions, we propose that the multimodal images being registered are first transformed from their individual intensity feature spaces to a common feature space based on complex wavelet phase moments. This complex wavelet feature transform allows information acquired using different imaging modalities to be compared in a direct manner. Hence, a more well-constrained distance measure can to be integrated into the objective function to improve convergence monotonicity. The phase moment transform M can be computed as follows [8]:

$$\begin{aligned} M(\underline{x}) = \;& \tfrac{1}{2}\sum_{\theta} P(\underline{x},\theta)^2 + \\ & \tfrac{1}{2}\Bigg[4\left(\sum_{\theta}\left(P(\underline{x},\theta)\sin(\theta)\right)\left(P(\underline{x},\theta)\cos(\theta)\right)\right)^2 + \\ & \left(\sum_{\theta}\left[\left(P(\underline{x},\theta)\cos(\theta)\right)^2 - \left(P(\underline{x},\theta)\sin(\theta)\right)^2\right]\right)^2\Bigg]^{\frac{1}{2}} \end{aligned} \tag{3}$$

where $P(\underline{x}, \theta)$ is the phase coherence at orientation θ computed using the iterative phase coherence estimation scheme described in [9]. This phase moment transform is highly robust to signal non-homogeneities and noise.

To construct a multi-modal objective function that is efficient and easily extensible to high-dimension problems, it is important to first observe the behavior of the joint phase moment distribution as the images come into alignment. The joint phase moment distributions at various stages of alignment for a particular three-dimensional case are shown in Figure 1. It can be observed that as the images come into alignment, the samples in the joint phase moment distribution becomes tightly packed around a 3-dimensional hyperline. This is intuitive since in the ideal case where all images contain the same information, the samples in the joint distribution would lie entirely along the hyperline $x = y = z$ when the images are in alignment. This can be generalized to the n-dimensional case, where the optimal alignment between n images occurs when the samples in the joint distribution are maximally packed around a n-dimensional hyperline. Based on this theory, we propose that the dispersion of a n-dimensional joint phase moment distribution can be measured efficiently based on the orthogonal distances between samples in the joint distribution and the fitted n-dimensional hyperline that minimizes the error residual within the distribution.

For the case of n images, the proposed objective function is defined as the cumulative quadratic orthogonal distance between samples in the joint phase moment distribution J and the fitted n-dimensional hyperline $\underline{h}$:

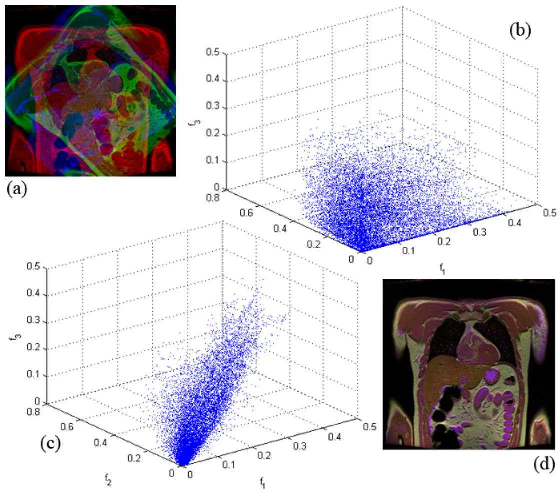

Figure 1. a) misaligned images, b) joint distribution of misaligned images, c) joint distribution of aligned images, d) aligned images

$$C(f_1^{\mathrm{T}_1}, f_2^{\mathrm{T}_2} ..., f_n^{\mathrm{T}_n}) = \sum_{i=1}^{N} \left\| \left(\underline{j}_i - \underline{\mu} \right) \times \underline{\varphi} \right\|_2 \quad (4)$$

where N is the number of samples in the joint distribution, $\underline{\mu}$ and $\underline{\varphi}$ are the centroid and direction cosines of the hyperline $\underline{h}$ respectively, $\times$ and $\|.\|_2$ indicates a cross product and a L2 (quadratic) norm respectively, and $\underline{j}_i$ is a vector representing the i^{th} sample from joint distribution J and can be defined as follows:

$$\underline{j}_i = [M_1 (\mathrm{T}_1(\underline{\mathrm{x}}_i)) \quad M_2 (\mathrm{T}_2(\underline{\mathrm{x}}_i)) \quad ... \quad M_n (\mathrm{T}_\mathrm{n}(\underline{\mathrm{x}}_i))] \quad (5)$$

where $\underline{\mathrm{x}}_i$ is the n-dimensional co-ordinate of the i^{th} sample. The fitted hyperline $\underline{h}$ that minimizes the error residual within the joint distribution J can be calculated by performing orthogonal regression on the joint distribution. The orthogonal regression method used can be described as follows. First, a matrix D is computed based on the difference between samples in the joint distribution J and the centroid of the samples $\underline{\mu}$ (which is also the centroid of the hyperline):

$$\mathrm{D} = \begin{bmatrix} \underline{j}_1 - \underline{\mu} \\ \underline{j}_2 - \underline{\mu} \\ \vdots \\ \underline{j}_n - \underline{\mu} \end{bmatrix} \quad (6)$$

Singular value decomposition (SVD) is then performed on D and the direction cosines of the hyperline $\underline{\varphi}$ is determined as the singular vector corresponding to the largest singular value. The fitted hyperline $\underline{h}$ can be represented by the direction cosines $\underline{\varphi}$ and the centroid $\underline{\mu}$.

3 Proposed Method

Based on the above theory, the proposed method can be described as follows. An iterative solver based on sequential quadratic programming (SQP) [10] is used to solve the optimization problem defined in Equation (1) using the objective function defined in Equation (4) The set of transformations are then re-estimated iteratively until convergence is reached to determine the optimal alignment between all images being registered. A multi-resolution scheme involving three different scales ($s = \frac{1}{4}, \frac{1}{2}, 1$) was used to improve convergence speed as well as avoid local optima.

4 Experimental Results

The effectiveness of the proposed method was evaluated using four different test sets. TEST1 consists of 256×256 T1/T2/CT images of an axial cranial slice from the NLM Visible Human project. TEST2 consists of 256×256 T1/T2/PD images of a coronal pelvic slice from the NLM Visible Human project. TEST3 and TEST4 consist of 761×748 images acquired using Landsat 7 ETM+ band 3 and Landsat 4-5 TM bands 4 and 5 at (Lat/Long: 46.0/-83) and (Lat/Long: 69.6/-92.7) respectively from the USGS project. Each test set was distorted using 20 different random affine transformations and were then registered using the proposed method. To judge the registration accuracy of the proposed method, the RMSE was determined based on 20 ground-truth control point triplets. As stated earlier, there are currently no methods that registers multiple images from different modalities simultaneously.

The registration accuracy results are shown in Table 1. The proposed method achieved low RMSE for all of the test sets. Sample registration results of TEST1, TEST2, and TEST3 achieved using the proposed method are shown in Figures 2 to 4 respectively. Based on visual inspection, the registration appears accurate in both cases. These results demonstrate the effectiveness of the proposed method for registering multiple images from different modalities while maintaining internal consistency amongst all the images.

5 Conclusions

In this paper, we introduced a novel method for simultaneous multimodal registration of multiple images. A new objective function was introduced based on the dispersion of joint phase moment distributions. The proposed method is highly efficient and robust to noise and signal non-homogeneities. Experimental results using real-world multimodal image sets indicate that high

Table 1. Registration accuracy

Test Set	RMSE[1] (pixels)		
	min	mean	max
TEST1	0.4458	0.7812	1.5356
TEST2	0.2793	0.4810	0.9625
TEST3	1.0449	2.2784	2.7399
TEST4	0.6372	1.1349	2.0156

1: The RMSE is computed over 20 random distortions.

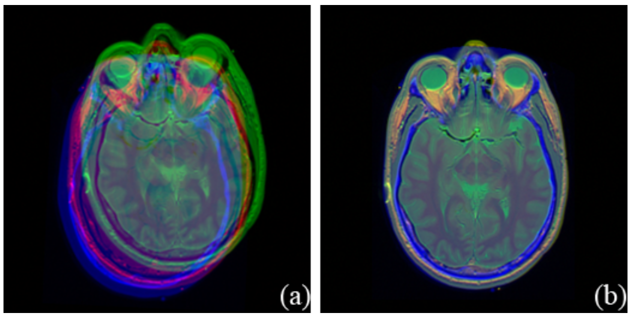

Figure 2. TEST1: a) overlay of T1/PD/CT (green/red/blue), b) registered images

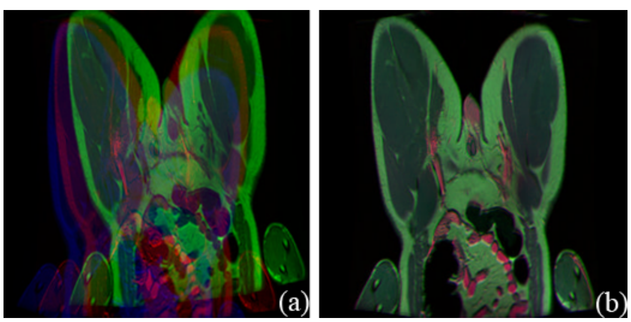

Figure 3. TEST2: a) overlay of PD/T1/T2 (green/red/blue), b) registered images

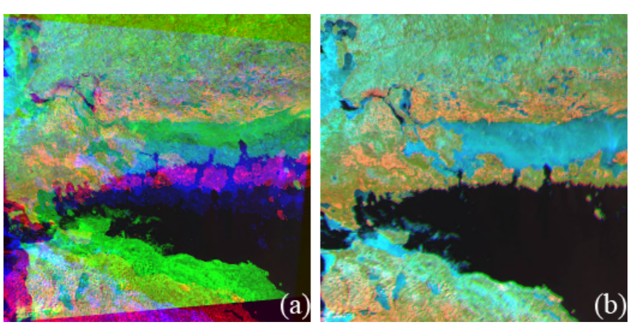

Figure 4. TEST3: a) overlay of Landsat 7 band 3 and Landsat 4/5 bands 4-5 (green/red/blue), b) registered images

registration accuracy can be achieved for situations involving multiple images. Future work involves investigating more robust orthogonal distance functions to reduce the effect of outlier samples on registration accuracy.

References

[1] J. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari, "Super-resolution in PET imaging," *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 137-147, 2006.

[2] L. Chen, T. Teo, J. Rau, J. Liu, and W. Hsu, "Building reconstruction from LIDAR data and aerial imagery," *Proc. IEEE International Geoscience and Remote Sensing Symposium*, pp. 2846-2849, 2005.

[3] P. Viola and W. Wells, "Alignment by Maximization of Mutual Information," *IJCV*, vol. 24, no. 2, pp. 137-154, 1997.

[4] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986-1004, 2003.

[5] E. Haber and J. Modersitzki, "Intensity gradient based registration and fusion of multi-modal images," *Proc. MICCAI*, pp. 726-733, 2006.

[6] A. Wong and D. Clausi, "ARRSI: Automatic Registration of Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, Part II, pp. 1483-1493, 2007.

[7] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration," *International Journal of Imaging Systems and Technology*, vol. 11, no. 1, pp. 71-80, 2000.

[8] P. Kovesi, "Phase congruency detects corners and edges," *Proc. Australian Pattern Recognition Society Conference*, pp. 309-318, 2003.

[9] A. Wong, "An iterative approach to improved local phase coherence estimation", *Proc. Canadian Conference on Computer and Robot Vision*, 2008.

[10] P. Boggs and J. Tolle, "Sequential quadratic programming," *Acta Numerica*, pp. 151, 1995.