A MULTI-MODAL AUTOMATIC IMAGE REGISTRATION TECHNIQUE BASED ON COMPLEX WAVELETS

Milad Ghantous, Somik Ghosh, Magdy Bayoumi
The Center for Advanced Computer Studies
University of Louisiana at Lafayette
{mmg4545, sxg5317, mab}@cacs.louisiana.edu

ABSTRACT

Image registration is considered one of the most fundamental and crucial pre-processing tasks in digital imaging. This paper describes a fast multimodal automatic image registration algorithm that handles the alignment of IR and visible images. A multi-resolution approach based on Dual Tree-complex wavelet transform is employed to speed up the process. At the coarsest level, an accurate registration estimate for higher levels is achieved, using edge detection and cross correlation. Mutual Information, on the other hand, is applied at higher levels as a matching criterion applied to the six orientation bands of the complex wavelet. The process is completely automatic, and was tested on several sets of synthetic and real data. Experimental results show that the proposed technique exhibits better accuracy than DWT-based algorithms for uni and multi-modal cases.

***Index Terms*—** Image Registration, Image Fusion, Multiresolution processing, Complex Wavelet Transform, Mutual Information.

1. INTRODUCTION

Image Registration is the process of geometrically aligning two or more images acquired from different viewpoints (MultiView registration), at different times (Mutli-Temporal registration), by different sensors (Multi-Modal registration), or a combination of two or more of the aforementioned. In MultiView analysis, the images may differ in translation, rotation, scaling, or more complex transformations mainly due to camera positions, while in multi-temporal analysis, images of the same scene may be acquired at different times or under different lighting conditions, and finally, in multimodal analysis, images are acquired by different types of imagers or sensors.

A plethora of proposed algorithms can be found for image registration which has gained, and is still receiving a lot of attention in the research community, due to its importance and necessity in many applications such as remote sensing, image mosaicing, image fusion (Surveillance, historical monument preservation), medicine (change detection, tumor growth monitoring) and computer vision (Target tracking). A detailed survey on conventional and newly proposed registration algorithms can be found in [1].

The process of image registration usually consists of four steps: **1. Feature detection**, also called Control Point (CP) selection such as lines, edges, corners, etc. **2. Feature matching**, in which, a match between the control points chosen in step 1 is established. **3. Mapping estimation,** which consists of estimating the best parameters responsible of registering the sensed image to the reference one. And finally, **4. Image resampling** consisting of transforming the sensed image using the optimal parameters found in the previous step.

In manual registration, the selection of CPs is usually performed by a human operator. Despite the extensive applications of this inherently simple method, it has proven to be inaccurate, time consuming, and unfeasible due to image complexity that makes it cumbersome or even impossible for the human eye to discern the appropriate control points. In addition, it fails to meet the real time execution requirements of modern applications**.** Therefore, researchers have focused on automating the feature detection, to align two or more images with no need for human intervention. However, one must keep in mind that no registration algorithm will work for all kinds of applications, and at the same time, it must not be too application specific.

Image registration is interpreted as the common bottleneck in the achieved accuracy of image fusion algorithms. This paper aims at developing a technique able to align infrared and visible images, which serves as a preprocessing step for an image fusion scheme published in a previous paper [2]. Thus, a fast MultiView/Multimodal automatic registration algorithm is proposed. The contribution of our work is twofold. First, employing Dual-Tree Complex Wavelet Transform (DT-CWT) as the pyramidal approach, not only offers a faster processing, but also better accuracy due to directional sensitivity and shift invariance. Second, a new metric joining edge detection, cross correlation, and mutual information is proposed to handle the multimodal nature of the problem while maintaining the applicability of the algorithm to different cases.

The remainder of this paper is organized as follows: Section 2 covers the related work, followed by the developed algorithm in section 3. Section 4 summarizes the simulation results, while section 5 concludes the paper.

2. RELATED WORK

Automatic registration has been extensively researched in the past 20 years; however, this section covers the main proposed schemes that employ multi-resolution processing, Mutual Information, or the combination of both. The idea of addressing the registration problem by applying coarse-to-fine resolution strategy has proven to be an elegant method to speed up the whole process while preserving, if not enhancing the accuracy of the algorithm. In [3], a mutli-resolution scheme based on Discrete Wavelet transform (DWT) is employed to register satellite images. Maximum Modulus Maxima is applied on the LH and HL frequency bands to extract edge points, and correlation is then applied for matching. The authors in [4] developed a parallel algorithm using the maxima of DWT coefficients for the feature space, and correlation for the search space. Despite their achieved performance, the above mentioned methods operate directly on gray intensity values and hence they are not suited for handling multi-sensor images. Mutual information methods on the other hand, originating with Viola and Wells [5], are able to register multimodal images since MI represents a measure of statistical dependency between the reference and the senses images rather than gray intensity values, which vary when different types of imagers are used, or under different lighting conditions. A multimodal brain image registration is developed and presented in [6]. It combines the sum of difference (SAD) and the mutual information (MI) into a matching criterion to enhance the registration accuracy. A mutli-resolution scheme is adopted making use of the LL band. Even though SAD is applied directly to the gray intensity values, the authors claim their algorithm work for multimodal images. [7] presents an automatic registration algorithm suited for airborne imagery, based on DWT and a maximization of mutual information (MMI) optimization. A similar technique is presented in [8]. The work developed in [9] on the other hand, explores a new hybrid metric based on mutual information and spatial information to register medical images.

3. PROPOSED ALGORITHM

The image registration problem can be stated as follows. Let $I_{REF}(x,y)$ and $I_{SRC}(x,y)$ be the reference and the source images respectively with $(x,y) \in \Phi \subset \Re^2$, where Φ represents a common region of interest between I_{REF} and I_{SRC}. Therefore, $\exists$ $T(x,y)$, where $T(x,y)$ is a geometric transformation, such that $I_{REF}(x,y) \approx I_{SRC}(T(x,y))$. One of the most common transformations found between images is a RST (Rotation/Scale/Translation) transformation. In that case,

$$T(x,y) = \begin{bmatrix} s\cos\alpha & s\sin\alpha & t_x \\ -s\sin\alpha & s\cos\alpha & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3.1)$$

Where **s** is the scaling factor, **α** is the rotation angle, and $(\mathbf{t_x},\mathbf{t_y})$ are the translation parameters in the x and y directions respectively. In other words, the problem of registration is reduced to find an optimal $v = [s\ \alpha\ t_x\ t_y]$, a vector of transformation parameters, that can be used in Eq. 3.1 to find the best corresponding match between I_{REF} and I_{SRC}.

The aim of this paper is to find the best geometric transformation able to align two images captured from visible and infrared imagers. A block diagram representing the proposed method is illustrated in figure 1.

3.1 Dual-Tree Complex Wavelet Transform (DT-CWT)

Due to the computational burden imposed by the search over the whole image to find the best geometric transformation, mutli-resolution schemes were adopted and used by researchers to speed up the process. The multi-resolution pyramidal approach allows us to exhaustively search over a small image at a coarse resolution to find an estimate of transformation parameters. Once found, the search space is narrowed, and an estimate of the higher resolution parameters is found. This is repeated until the highest level is reached, thus decreasing the amount of computations required compared with the search over the whole image.

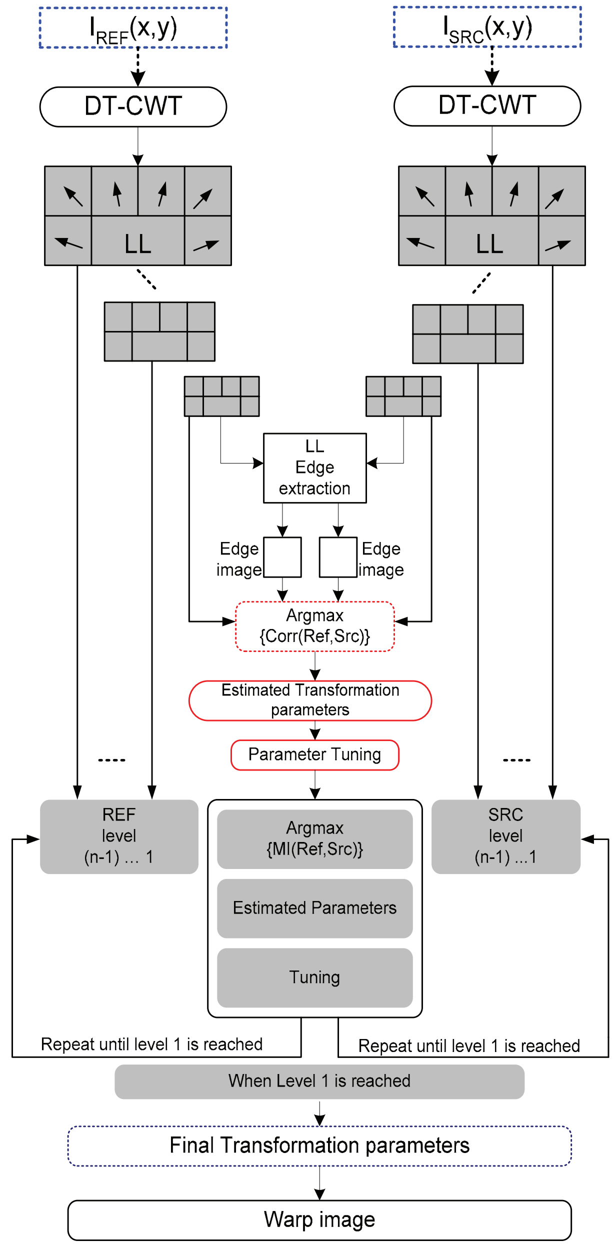


Figure 1. Proposed Algorithm Block Diagram

Discrete Wavelet Transform (DWT) [10] has been investigated and used to speed up the registration process. However, it suffers from several shortcomings such as shift sensitivity due to the sub-sampling at each level, poor directionality (three orientation bands: vertical, horizontal and diagonal), and lack of phase information. The Shift-Invariant DWT (SIDWT [11]) eliminates the shift sensitivity problem at the cost of an over-complete signal representation. On the other hand, the recently proposed Dual-Tree

Complex Wavelet Transform (DT-CWT [12]) not only addresses the over-completeness problem of the SIDWT, but is also characterized by a better directional sensitivity representing the image at six orientations at ±15º, ±45º, and ±75º.

3.2 Registration steps

The proposed algorithm aims to register two images acquired from different sensors (Infrared and visible spectrums in our case), and from different point of views, hence the difference in rotation, translation in both directions x and y, as well as possible scaling. The algorithm starts by decomposing the two input images, $I_{REF}(x,y)$ and $I_{SRC}(x,y)$ using the aforementioned DT-CWT(Near-Symmetric 13,19 tap filters, Q-Shift).
Let $D_{REF,l}(x,y)\{l=1,..,n\}$ and $D_{SRC,l}(x,y)\{l=1,..,n\}$ represent the decomposed images respectively, where l denotes the decomposition level and n is the total number of levels. Each decomposed image consists of a real part representing an approximation of the image and a complex part comprising six orientation bands (±15º, ±45º, and ±75º).
The algorithm is divided in two main parts: Registration of the lowest decomposition level ***n***, and the registration of higher levels $l = n-1, \ldots 1$. Starting at level n, the coarsest level of decomposition, a first estimate of the transformation vector v = [α t_x t_y] must be found. Scaling is omitted in this paper for simplicity. This step must be handled with extreme care since it constitutes the initial estimate upon which, higher levels of decomposition depend. For this reason, we choose Cross Correlation as a matching criterion due to its effectiveness and accuracy. This choice, however, suffers from two problems:

- Cross Correlation cannot handle multimodal images since it operates directly on intensity values.
- It is a computationally demanding task.

To overcome the situation, we propose to extract edge maps for the reference and source low passed images, M_{ref} and M_{src} respectively. Operating on edge maps instead of the image itself not only solves the correlation limitation (correlating edge information instead of intensity values), but also have reduced computational requirements since the majority of the map consists of zero values except for edge locations.
The search space is initially chosen to be [-ϴ, + ϴ] for the angle with accuracy Λ_α, $[-\tau_x/2, +\tau_x/2]$ for the translation in the X-direction, and $[-\tau_y/2, +\tau_y/2]$ for the translation in the Y-direction with accuracy Λ_t , where $\tau_{x,y}$ denotes the image dimensions at level n. A priori information gathered from camera locations and movements can be used to narrow the initial search space. An exhaustive search is then performed over the search space to determine the best initial transformation vector v, called v_{init}=[α_{init} $t_{x,init}$ $t_{y,init}$], according to equation 3.2.

$$v_{init} = \arg\max_{v} \rho_{M_{ref}, T(M_{src})} \tag{3.2}$$

Where ρ denotes the cross correlation, and T(.) the warped image using vector v.
The second part of the algorithm starts at decomposition level n-1. V_{init} is used as the center of the new search interval at this level as follows: [α_{init}-μ, α_{init}+ μ], [2*$t_{x,init}$- μ, 2* $t_{x,init}$+ μ], [2* $t_{y,init}$- μ, 2* $t_{y,init}$+ μ] become the new search intervals for rotation and translations with accuracies 2*Λ_α and 2*Λ_t respectively. μ is a variable bounded by two values: a minimum value to compensate for any erroneous estimate at level n, and a maximum value to narrow down the search interval to speed up the process. The matching criterion for levels $l \le$ (n-1) is the Mutual Information, an entropy based concept that measures the statistical dependence between 2 images A and B defined as:

$$I(A,B) = \sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)\,p(b)} \tag{3.3}$$

Where a and b denote the intensity of images A and B respectively, p(.) is the marginal distribution, and p(a,b) is the joint distribution. Mutual information is not only suitable for multimodal images but is also computationally light compared to cross correlation. The transformation vector V_l is found according to equation 3.4.

$$v_l = \arg\max_{v} \begin{bmatrix} I(R\{D_{ref,1}\}, T(R\{D_{src,1}\})) \\ + I(\|C\{D_{ref,1}\}\|, T(\|C\{D_{src,1}\}\|)) \end{bmatrix} \tag{3.4}$$

Where R{.} and C{.} denote the Real and Complex parts of the images. V_{n-1} becomes the center of the search interval at level (*n-2*) following the same reasoning presented above, and the process is repeated until the highest level of decomposition is reached. μ is divided by 2 in each iteration to narrow the search space, while Λ_α and Λ_t are doubled. A pseudo-code of the algorithm is presented in Algorithm 1.

Algorithm 1. Multi-resolution Registration

START
$D_{REF,l} \leftarrow$ DT-CWT(I_{REF},n) , $D_{SRC,l} \leftarrow$ DT-CWT(I_{SRC},n)
IF l = n **DO**
$M_{ref} \leftarrow$ EdgeMap(R{D_{REF},n}) , $M_{src} \leftarrow$ EdgeMap(R{D_{SRC},n})

$$v_{init} \leftarrow \arg\max_{v} \rho_{M_{ref}, T(M_{src})}$$

END IF
WHILE l >=1 **DO**
Adjust search interval according to V_{l-1}

$$v_l \leftarrow \arg\max_{v} \begin{bmatrix} I(R\{D_{ref,1}\}, T(R\{D_{src,1}\})) \\ + I(\|C\{D_{ref,1}\}\|, T(\|C\{D_{src,1}\}\|)) \end{bmatrix}$$

END WHILE
Warp image using V=[α_1 2*$t_{x,1}$ 2*$t_{y,1}$]
END

To further reduce the computational requirements, an alternative search method, consisting of splitting the search space into complementary sub-spaces is proposed in [4] and can be easily applied to our proposed method to further reduce the computational burden. It is however omitted due to lack of space.

4. EXPERIMENTS AND RESULTS

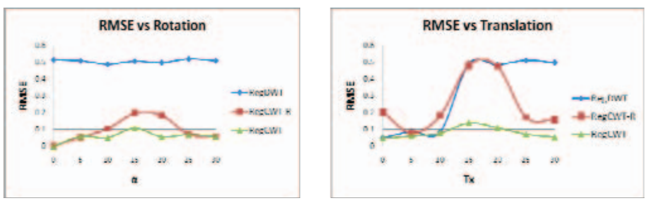
The proposed algorithm was developed and tested on several sets of uni-modal and multi-modal images. However, we limit the simulation results to two sets only due to the lack of space. For each set of images, three algorithms were implemented: (1) A Discrete Wavelet Transform employing correlation at the lowest level and MI at higher levels, dubbed RegDWT, (2) a Complex wavelet based algorithm making use of the real part only, RegCWT-R, and (3) our proposed algorithm, RegCWT. Λ_t, Λ_α, and μ are chosen to be 5 pixels, 4 degrees, and 10 respectively.

The first experiment consists of uni-modal registration that is well handled by the developed technique. Initially, the source image is warped manually. Figures 2.a, 2.b and 2.c represent the reference, the source, and the registered image respectively.

(a) (b) (c)

Fig.2 a. I_{REF} b.I_{SRC} c. Registered image

The plots in figures 3.a and 3.b represent the registration accuracy measured as Root Mean Square Error (RMSE), versus the rotation and the translation respectively. The registration accuracy of RegCWT-R is clearly higher than that achieved by RegDWT. RegCWT further improves the accuracy by an average of 20%.



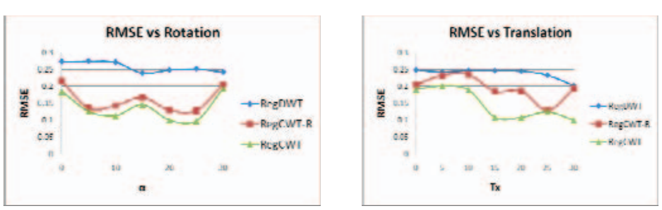
(a) (b)

Fig.3.Unimodal Registration Accuracy a. RMSE vs Rotation variation b. RMSE vs T_X variation

The effectiveness of the proposed scheme on multimodal images was investigated using two images captured from a visible and infrared cameras. Figure 4 represents the qualitative results of RegCWT. A quantitative evaluation is shown in figure 5. Clearly, RegCWT outperforms RegDWT and RedCWT-R by around 13 to 25% accuracy.

(a) (b) (c)

Fig.4 a. I_{REF} b.I_{SRC} c. Registered image



(a) (b)

Fig.5. Multimodal Registration Accuracy a. RMSE vs Rotation variation b. RMSE vs T_X variation

5. CONCLUSION

In this paper, a new technique for multimodal automatic image registration algorithm is presented. To speed up the processing, the algorithm is employed in a pyramidal fashion based on Dual tree complex wavelet transform. At the lowest level, edge maps are extracted and the matching is based on cross correlation measure. The search interval is then refined for higher levels employing Mutual Information as a matching criterion due to its ability to register multimodal images and its light computational load. The developed technique handles multi-modal as well as uni-modal cases and has shown to have superior accuracy when compared to its DWT counterpart.

6. REFERENCES

[1] B.Zitova and J.Flusser,"Image Registration methods:a survey" Image and Vision Computing, pp.977-1000, 2003

[2] M. Ghantous, S. Ghosh, and M. Bayoumi, "A Gradient Based Hybrid Image Fusion scheme Using Object Extraction", IEEE International Conference on Image Processing, pp. 1300-1303, 2008

[3] L. Fonseca, and M. Costa, "Automatic Registration of Satellite Images", in *Proc*. Brazilian Symposium on Computer Graphics and Image Processing, pp.219-226, 1997

[4] J. Le Moigne, W. Cambell, and R. Cromp, "An Automated Parallel Image Registration Technique Based on the Correlation of Wavelet Features", IEEE *Trans*. Geoscience and Remote Sensing, Vol.40, No.8, pp. 1849-1864, August 2002

[5] P. Viola and W.M. Wells, "Alignment by Maximization of Mutual Information", International Journal of Computer Vision 24, pp. 137-154, 1997

[6] J. Wu, and A. Chung, "Multimodal Brain Image Registration Based on Wavelet Transform Using SAD and MI", MIAR 2004, LNCS 3150m, pp. 270-277, 2004

[7] X.Fan, H. Rhody, and E. Saber, "Automatic Registration of Multi-Sensor Airborne Imagery", in *Proc*. IEEE Applied Imagery and Pattern Recognition workshop, pp. 80-86, 2005

[8] I. Zavorin, and J. Le Moigne, "Use of Multiresolution Wavelet Feature Pyramids for Automatic Registration of Multisensor Imagery", IEEE *Trans*. On Image Processing, Vol 14., No.6, pp.770-782, June 2005

[9] R. Xu, and Y. Chen, "Wavelet-Based Multiresolution Medical Image Registration Strategy Combining Mutual Information With Spatial Information", International Journal of Innovative Computing, Information and Control Vol.3, No.2, 2007

[10] S.Mallat, "A Theory for Multiresolution Signal Decomposition", IEEE PAMI, Vol.11, No.7, pp. 674-693, 1989

[11] O. Rockinger, "Image Sequence Fusion Using a Shift Invariant Wavelet Transform", IEEE *Trans*. Image Processing, vol.3, pp. 288-291, October 1997

[12] N. Kingsbury, "A Dual-Tree Complex Wavelet Transform with improved orthogonality and symmetry properties", In *Proc*. IEEE Conf. on Image Processing, 2000