International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:5, No:11, 2011

# Hidden State Probabilistic Modeling for Complex Wavelet Based Image Registration

F. C. Calnegru

***Abstract*—**This article presents a computationally tractable probabilistic model for the relation between the complex wavelet coefficients of two images of the same scene. The two images are acquisitioned at distinct moments of times, or from distinct viewpoints, or by distinct sensors. By means of the introduced probabilistic model, we argue that the similarity between the two images is controlled not by the values of the wavelet coefficients, which can be altered by many factors, but by the nature of the wavelet coefficients, that we model with the help of hidden state variables. We integrate this probabilistic framework in the construction of a new image registration algorithm. This algorithm has sub-pixel accuracy and is robust to noise and to other variations like local illumination changes. We present the performance of our algorithm on various image types.

***Keywords*—**Complex wavelet transform, image registration, modeling using hidden state variables, probabilistic similarity measure.

## I. Introduction

IMAGE registration is the process of geometrically overlapping two or more images of the same scene. This basic capability is needed in various image analysis applications. For example image registration is a critical component of remote sensing, medical, industrial image analysis systems, etc..[1]

The two images involved in the task of image registration are named reference image, and target image [2]. Image registration consists in finding a coordinate transformation from the target image to the reference image such that the two images become similar. The precise meaning of 'similar' depends on the specific registration problem to be solved [1].

Depending on the type of the transformation, that we look for, there are two types of registration: parametric, and non-parametric registration. In the case of the parametric registration, the transformation can be expanded in terms of some basis functions. In the case of non-parametric registration, the transformation is no longer restricted to a parametrizable set.

As the algorithm that we propose is from the category of parametric image registration, we will no further insist on the non-parametric image registration. Parametric image registration can be divided into: landmark based parametric image registration, principal axes-based registration and optimal parametric registration [1].

Landmark based parametric registration is a type of registration that relies on the features extracted at an initial stage in the process of registration. Those features can be line intersections [3], road crossings, inflection points of curves, corners [2], local extremes of wavelet transform [4], etc. The quality of landmark based registration is highly dependent on the performances of the feature detector that is being used. If the feature detector is not reliable enough, the registration will be low quality.

This is why sometimes is better to use a registration method that relies on features that can be automatically deduced from the image. Such intrinsic features are for example the principal axes [5]. Although principal axes registration is fast and necessitates very few parameters, it needs the moment matrix and the eigenvalue decomposition of two large matrixes, it is not suitable for multimodal registration and its results can be ambiguous [1].

The disadvantages of the landmark-based registration and principal axes registration led to the emergence of a more general and flexible class of parametric image registration. This category consists of optimal parametric registration algorithms.

The basic idea of optimal parametric registration is to define a distance (similarity) measure between the reference image and the target image, and then to find the parameters of the transformation that optimize this similarity measure. The most known similarity measures are the sum of squared differences, the correlation [2], and the mutual information [6]. The domain of this similarity measures is either the intensities space like in [7], or another feature space like the wavelet coefficients with a magnitude above a certain threshold [4], the energy map [8], and the wavelet coefficients from the first decomposition level [9].

The algorithm that we introduce is an optimal registration algorithm. The similarity measure that we use is a probability defined on the wavelet coefficients space. This makes our similarity measure more robust to noise than the similarity measures defined on the intensity space. The latter mentioned similarity functions are affected by the noise that usually corrupts the image intensities.

We chose to use for our algorithm the complex wavelet transform because unlike the discrete wavelet transform, it is almost shift invariant and almost rotationally invariant and because, it has a good directional selectivity [10]. We used the dual tree complex wavelet transform to construct the complex wavelet transform

This article is structured as following. In section II we present an introduction to the dual-tree complex wavelet transform. In section III we expose the probabilistic model underlying the registration algorithm. In section IV we present

F.C. Calnegru is with the Department of Computer Science University of Pitesti, 110040 Romania (e-mail: calnegru_florina@yahoo.com)

the registration algorithm and introduce different modalities to integrate it into real world registration systems. In section V we present experimental results on artificial data as well as on real data. In section VI we present our conclusions.

## II. Complex Wavelet Transform

Discrete wavelet transform (DWT) is a modality to project a signal onto an orthogonal wavelet basis. By using the DWT it is possible to obtain local information about a signal both in the spatial domain and in the frequency domain. For a 2-D signal the DWT coefficients are obtained by passing the signal through a cascade of orthogonal high pass and low pass filters. The original image is decomposed at any scale j, into 4 components: *HHj* (contains the diagonal details), *HLj* (contains the horizontal details), *LHj* (comprised of vertical details), and *LLj* (contains the approximation coefficients). For more details on DWT see for example [11].

Any signal f(x, y) can be reconstructed via the inverse discrete wavelet transform from its detail and approximation coefficients as in (1)

$$f(x,y)=\frac{1}{\sqrt{MN}}\sum_m\sum_n W_\varphi(j_0,m,n)\varphi_{j_0,m,n}(x,y) + \frac{1}{\sqrt{MN}}\sum_{k=1}^{3}\sum_{j=j_0}^{\infty}\sum_m\sum_n W_\psi^k(j,m,n)\psi^k_{j,m,n}(x,y) \quad (1)$$

In (1) $\varphi_{j,m,n}$ represents the scaling function scaled with a factor of $j$ and translated with $m$ and $n$, $\psi^k_{j,m,n}$ represents k-th mother wavelet function, scaled with a factor of $j$ and translated with $m$ on Ox, and $n$ on Oy, $W_\varphi(.,.)$ represents the aproximation coefficients, and $W^k_\psi(.,.)$ represents the detail coefficients.

For the 2-D DWT there are 3 mother wavelet functions: one that permits the extraction of horizontal details, one for the vertical details, and one for the diagonal details. So we can say that $W_\varphi(j,.,.)$ corresponds to $LL_j(.,.)$ and that, for example, $W_\psi^1(j,.,.)$ corresponds to $HL_j(.,.)$, $W_\psi^2(j,.,.)$ corresponds to $LH_j(.,.)$, and that $W_\psi^3(j,.,.)$ corresponds to $HH_j(.,.)$.

Unfortunately, the DWT has some major drawbacks that make it less appropriate for registration. Among those drawbacks, we mention poor directional selectivity, as the *HH* coefficients cannot differentiate between edges at 45 degrees and edges at 135 degrees and rotation and translation variance. Complex wavelet transform constitutes a remedy for these problems.

It is possible to observe that by taking in (1), instead of a real scaling function and real wavelet functions, a complex scaling function, and complex wavelet functions, for which the real and the imaginary part form a Hilbert pair, the drawbacks of the DWT are eliminated [12].

In our article, we employed dual tree complex wavelet transform to obtain the complex wavelet decomposition for our images. The dual tree complex wavelet transform uses 6 complex mother wavelets that distinguish spectral features oriented at {75˚, 45˚, 15˚, -75˚, -45˚, -15˚}. By projecting the image onto the 6 complex wavelet functions, we obtain 6 complex wavelet coefficients for each scale and translation.

To facilitate the presentation, from now on, every time we mention wavelet transform, we refer to the dual tree complex wavelet transform.

## III. Probabilistic Framework

The research behind this article is motivated by the desire to understand and model the relation between the corresponding complex wavelet coefficients of two images of the same scene, that are captured at different times, or from different viewpoints or with different sensors.

What happens with the complex wavelet coefficients of the two images, when we have previously registered them (by some means)? How does the relation between the coefficients differ, in the case in which we already registered the two images, from the case in which we haven't yet registered the images?

By considering that every wavelet coefficient can be in one of the two states: large or small, we claim that the probability, of having the magnitudes from one image in the same state as the magnitudes from the other image, reaches its maximum when the two images are registered (see Fig. 1).

We construct below the probabilistic framework that is needed to define the above-mentioned probability.

First we model the distributions of the wavelet coefficients magnitudes, and then, based on those distributions, we define, for the considered images, the probability that their complex wavelet coefficients are in the same states.

In the following, each set, consisting of all coefficients of a certain level, corresponding to the same mother wavelet [10], will be referred as a layer.

In order to model the distribution of the wavelet coefficient magnitudes we performed the following experiment. We considered several images, and compute different layers of coefficients. By selecting a certain threshold, we divided each layer into two classes, the class containing the coefficients with magnitudes larger than the threshold and it's complementary. (see Fig. 2 for such an example)

As Fig. 2 (c) and Fig. 2 (d) suggest, the distribution of coefficient magnitudes larger than a threshold, and the distribution of coefficient magnitudes smaller than that threshold can be assumed as being log-normal distributions. This yielded us to consider the following probabilistic model.

Let $S$ be a random variable modelling the state of each coefficient magnitudes [12], that is $S$ can take one of the two values *low* and *high*

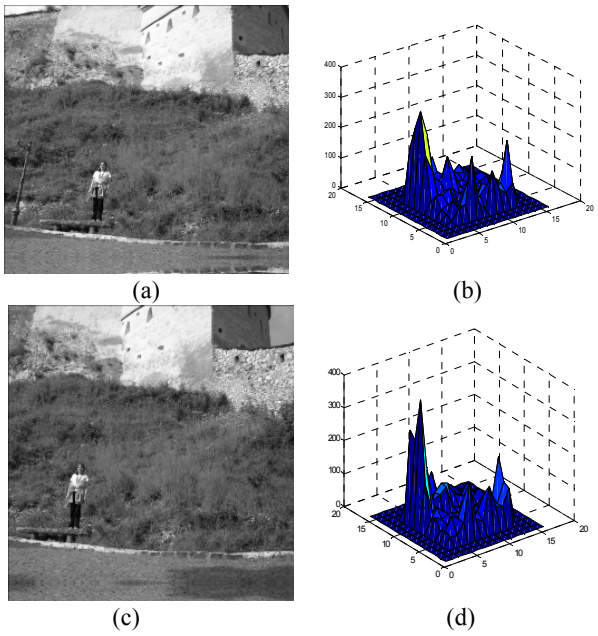



Fig. 1 Image (a) and (c) are two outdoor images not registered yet. Image (b) represents the fourth level wavelet coefficients, of image (a). Image (d) represents the fourth level of wavelet coefficients, of (c). The wavelet coefficients in (b) and (d) correspond to the same mother wavelet and they are computed on the overlapping region between images (a) and (c) after they were registered. Being given the law relevancy of the wavelet coefficients computed in the neighbourhood of the borders, in our tests we used exclusively the wavelet coefficients computed at a conventionally established distance from the image borders

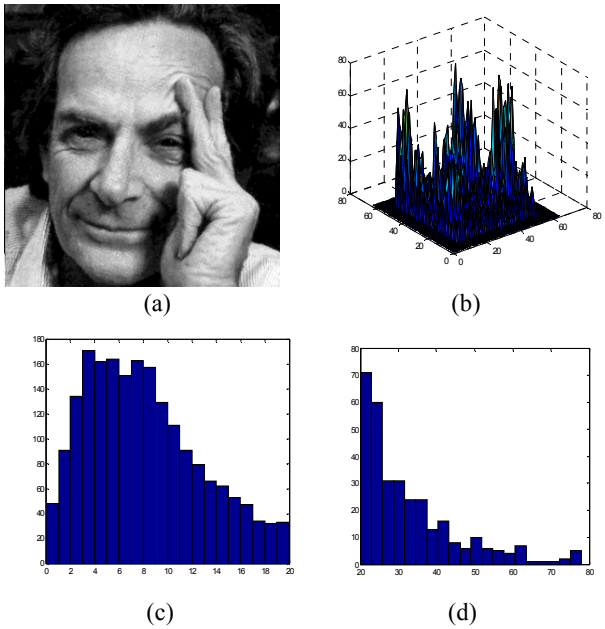



Fig. 2 Image (b) represents a layer of level 3 computed for the initial image (a). Image (c) represents the histogram of the magnitudes of the coefficients of (b) smaller than 20. Image (d) represents the histogram of the magnitudes of the coefficients of (b) larger than 20

Usually low coefficient magnitudes correspond to smooth regions, and high magnitudes correspond to regions that contain edges.

Since the magnitude values can be observed, while the values taken by $S$ can not be observed directly, we say that $S$ is a *hidden variable.*

We conjecture that the magnitude of every coefficient is conditioned by a different hidden state variable, therefore in case of a layer the number of hidden state variables equals the number of coefficients. Also we assume that there are no correlations among the hidden state variables corresponding to different wavelet coefficients. This assumption corresponds to the idea that behind each coefficient magnitude there is a switcher whose on/off state determines that certain values are more probable than others, and the switchers are independent.

From mathematical point of view, we consider that the magnitude of each coefficient $i$ from a layer is the realization of a variable $W_i$, that is conditioned by the hidden state variable $S_i$, associated to that coefficient. This means that in order to generate a particular realization of $W_i$, we have to generate a realization $s$ of $S_i$, and then draw an observation $w$ according to the density function $f_{W_i \mid S_i}(w \mid S_i = s)$. Since the hidden state variable $S_i$ can take only two values we get:

$$f_{W_i}(w) = f_{W_i|S_i}(w \mid S_i = \text{small})\, p_{S_i}(\text{small}) + f_{W_i|S_i}(w \mid S_i = \text{large})\, p_{S_i}(\text{large}) \tag{2}$$

In our model we take:

$$f_{W_i|S_i}(w \mid S_i = \text{small}) = \frac{1}{w\sigma_{i,small}\sqrt{2\pi}} e^{-\frac{\left(\ln w - \mu_{i,small}\right)^2}{2\sigma_{i,small}^2}} \tag{3}$$

$$f_{W_i|S_i}(w \mid S_i = \text{large}) = \frac{1}{w\sigma_{i,\text{large}}\sqrt{2\pi}} e^{-\frac{\left(\ln w - \mu_{i,\text{large}}\right)^2}{2\sigma_{i,\text{large}}^2}} \tag{4}$$

Besides we assume that the variables that model the magnitudes of the wavelet coefficients from a layer are identically distributed. Consequently we can drop the indexing of the collection of variables $W_i$, and we can say that for every coefficient pair $(i, j)$ from a layer the equalities (5) and (6) hold.

$$\sigma_{i,small} = \sigma_{j,small} = \sigma_{small} \quad \forall i,j$$
$$\sigma_{i,\text{large}} = \sigma_{j,\text{large}} = \sigma_{\text{large}} \quad \forall i,j \tag{5}$$

$$\mu_{i,\text{large}}=\mu_{j,\text{large}}=\mu_{\text{large}} \quad \forall i,j$$
$$\mu_{i,small}=\mu_{j,small}=\mu_{small} \quad \forall i,j \tag{6}$$

In the following, we assume that the variables, modelling the coefficient magnitudes from a layer, are independent.

In order to model the distribution of the magnitudes from a layer we need to estimate the parameters $\sigma_{small}$, $\sigma_{large}$, $\mu_{small}$, $\mu_{large}$. We look for the values of these parameters that maximize the expression (7).

$$\prod_i \left( \begin{array}{l} f_{W|S_i}(w_i \mid S_i=\text{small})\, p_{S_i}(\text{small}) + \\ f_{W|S_i}(w_i \mid S_i=\text{large})\, p_{S_i}(\text{large}) \end{array} \right) \tag{7}$$

If we would know the probability mass functions of all $S_i$, this method would be equivalent with the maximum likelihood method for estimating the parameters $\sigma_{small}$, $\sigma_{large,}$ $\mu_{small}$, $\mu_{large}$. Since we do not know those probability mass functions, we also have to estimate them and this complicates the problem. Luckily, we can simplify the estimation problem by means of the assumptions that we have made. According to those assumptions, when maximizing in relation with the probability mass functions, all the terms of the product from (7) are independent. This implies that the maximum of (7) can be reached if all the terms in (7) are maximum. We use this observation to first estimate the probability mass functions and then $\sigma_{small}$, $\sigma_{large,}$ $\mu_{small}$, $\mu_{large}$.

As the state variables are independent, the estimates, of the probability mass functions that maximize (7), are given in (8) and (9)

$$p_{S_i}(\text{small})= \begin{cases} 1, & f_{W|S_i}(w_i \mid S_i=\text{small}) > f_{W|S_i}(w_i \mid S_i=\text{large}) \\ 0, & otherwise \end{cases} \tag{8}$$

$$p_{S_i}(\text{large})=1-p_{S_i}(\text{small}) \tag{9}$$

(8) and (9) tell us that, in fact, the hidden states variables are not random variables, since they have deterministic values.

After the hidden variables of the layer are divided in the two classes, *large* and *small*, we can estimate the values of $\sigma_{small}$, $\sigma_{large,}$ $\mu_{small}$, $\mu_{large}$ using maximum likelihood for the observations also separated in the two categories, according to the values of their hidden state variables:

$$\mu_{\text{large}}=\frac{\sum_{w\in L^{\text{large}}} \ln w}{N^{\text{large}}} \tag{10}$$

$$\sigma_{\text{large}}=\sqrt{\frac{\sum_{w\in L^{\text{large}}} \left(\ln w-\mu_{\text{large}}\right)^2}{N^{\text{large}}}}, where \tag{11}$$
$$L^{\text{large}}=\{w_i \mid S_i=\text{large}\}, N^{\text{large}}=\left|L^{\text{large}}\right|$$

$\sigma_{small}$, $\mu_{small}$ are computed in the same way by replacing *large* with *small*.

The algorithm for estimating the parameters is:

*Begin Algorithm*

*Step 1:* Generate randomly a value *t* between the minimum value of the magnitudes from the layer and the maximum value of those magnitudes

*Step 2:* for all i do

$$S_i=\begin{cases} \text{large}, & \text{w}_i > t \\ \text{small}, & \text{otherwise} \end{cases}$$

*Step 3:* compute $\sigma_{small}$, $\sigma_{large,}$ $\mu_{small}$, $\mu_{large}$ using (10) and (11)

*Step 4:*

$$\sigma_{\text{small}}^{\text{new}}=\sigma_{\text{small}},\ \sigma_{\text{large}}^{\text{new}}=\sigma_{\text{large}},$$
$$\mu_{\text{small}}^{\text{new}}=\mu_{\text{small}},\ \mu_{\text{large}}^{\text{new}}=\mu_{\text{large}}$$

*Step 5:*

repeat

1. use (8) and (9) to compute for every i $p_{S_i}(\text{small})$, $p_{S_i}(\text{large})$
2. compute $\sigma_{\text{small}}$, $\sigma_{\text{large,}}$ $\mu_{\text{small}}$, $\mu_{\text{large}}$ using (10) and (11)
3. Assign

$$\sigma_{\text{small}}^{\text{old}}=\sigma_{\text{small}}^{\text{new}},\ \sigma_{\text{large}}^{\text{old}}=\sigma_{\text{large}}^{\text{new}},$$
$$\mu_{\text{small}}^{\text{old}}=\mu_{\text{small}}^{\text{new}},\ \mu_{\text{large}}^{\text{old}}=\mu_{\text{large}}^{\text{new}}$$
$$\sigma_{\text{small}}^{\text{new}}=\sigma_{\text{small}},\ \sigma_{\text{large}}^{\text{new}}=\sigma_{\text{large,}}$$
$$\mu_{\text{small}}^{\text{new}}=\mu_{\text{small}},\ \mu_{\text{large}}^{\text{new}}=\mu_{\text{large}}$$

$$\text{until} \left( \begin{array}{l} \left|\sigma_{\text{small}}^{\text{old}}-\sigma_{\text{small}}^{\text{new}}\right|<\varepsilon \ \&\ \left|\sigma_{\text{large}}^{\text{old}}-\sigma_{\text{large}}^{\text{new}}\right|<\varepsilon \ \& \\ \left|\mu_{\text{small}}^{\text{old}}-\mu_{\text{small}}^{\text{new}}\right|<\varepsilon \ \&\ \left|\mu_{\text{large}}^{\text{old}}-\mu_{\text{large}}^{\text{new}}\right|<\varepsilon \end{array} \right)$$

*End Algorithm*

We use the distributions of the complex wavelet coefficient magnitudes to define the probability that the correspondent magnitudes are in the same state.

Let us consider

$$L_j^t = \left\{ \begin{matrix} w \mid w \in \text{layer j of wavelet coefficient} \\ \text{magnitudes of the target image} \end{matrix} \right\}$$
$$L_j^r = \left\{ \begin{matrix} w \mid w \in \text{layer j of wavelet coefficient} \\ \text{magnitudes of the reference image} \end{matrix} \right\} \quad (12)$$

$$L_{\text{j,common}}^{\text{large}} = \left\{ \left( w_i^{j,t}, w_i^{j,r} \right) \left| \begin{matrix} w_i^{j,t} \in L_j^t, w_i^{j,r} \in L_j^r, \\ f_{W|S_i}(w_i^{j,t} \mid S_i = \text{large}) > f_{W|S_i}(w_i^{j,t} \mid S_i = \text{small}) \\ , f_{W|S_i}(w_i^{j,r} \mid S_i = \text{large}) > f_{W|S_i}(w_i^{j,r} \mid S_i = \text{small}) \end{matrix} \right. \right\} \quad (13)$$

$$L_{\text{j,common}}^{\text{small}} = \left\{ \left( w_i^{j,t}, w_i^{j,r} \right) \left| \begin{matrix} w_i^{j,t} \in L_j^t, w_i^{j,r} \in L_j^r, \\ f_{W|S_i}(w_i^{j,t} \mid S_i = \text{large}) \le f_{W|S_i}(w_i^{j,t} \mid S_i = \text{small}) \\ , f_{W|S_i}(w_i^{j,r} \mid S_i = \text{large}) \le f_{W|S_i}(w_i^{j,r} \mid S_i = \text{small}) \end{matrix} \right. \right\} \quad (14)$$

We denote:

$$N_{\text{j, common}} = \left| L_{\text{j,common}}^{\text{large}} \right| + \left| L_{\text{j,common}}^{\text{small}} \right| \quad (15)$$

$$N_j = \left| L_j^t \right| = \left| L_j^r \right| \quad (16)$$

We define in (17) the probability that the correspondent magnitudes, of the layer *j,* are in the same state.

$$p_j = \frac{N_{\text{j, common}}}{N_j} \quad (17)$$

Since all the layers are independent, we define in (18), the probability that the correspondent magnitudes of the two images are in the same state.

$$p = \prod_j p_j \quad (18)$$

## IV. The Registration Algorithm

If we parameterize the transformation that registers the two images by $\theta$, than the registration algorithm consists in searching for $\theta$ , that maximize the probability from (18).

In the implementation of our algorithm, we maximized (18) as in [13] by means of simulated annealing [7]. This helped us to avoid that the algorithm outputs a local maximum instead of a global maximum. We have used this algorithm to register images that differ by a similarity transform. For those images, the search space of the parameter is quite large. Fortunately, we observed that the value for the parameter $\theta_i$, for which the maximum of (18) is attained when the parameters $\theta_j, j \neq i$, are fixed, is somehow close to the value that this parameter has in the set θ when the global maximum of (18) is reached. This especially happens when the other $\theta_j, j \neq i$ are also somehow close of their values from the set $\theta$ for which the global maximum is attained.

This allowed us to find the optimizing parameters not by searching on the Cartesian product of the spaces of all parameters but by alternatively searching on the space of each parameter. This process necessitates several iterations, depending how far the solution is from the initial guess.

For high temperature the generation function, for our variant of the simulated annealing algorithm, is the product between a Gaussian centered in the current value, *x*, and the function h from (19). For smaller temperatures, the generation function is simply a Gaussian centered in *x*

$$h(x) = \begin{cases} 1, & r < 0.5 \\ -1, & r \ge 0.5 \end{cases} \quad (19)$$

where $r \in (0,1)$ is a uniform random number

In order to speed up the convergence, we have used an annealing scheduling function *s(.)*, that decreases faster for larger iteration indexes.

$$s(k) = \begin{cases} t_0 / k, & k < k_1 \\ t_0 / k^2, & k_1 \le k < k_2 \\ t_0 / k^3, & k \ge k_1 \end{cases} \quad (20)$$

where $k$ is the current iteration index, $t_0$ is the initial temperature and $k_1$, $k_2$ are positive integers.

Although the optimization idea is similar with the one used in [13], the similarity measure, from this article, and the probabilistic framework, on which we construct this similarity measure, are radically different from the ones presented in [13].

In practice our algorithm requires two stages: one of training, in which the values for all $\sigma_{small}$, $\sigma_{large}$, $\mu_{small}$, $\mu_{large}$ are obtained, and the other one for the actual optimization of (18).

Once the parameters $\sigma_{small}$, $\sigma_{large}$, $\mu_{small}$, $\mu_{large}$ are estimated for every layer, the speed of our algorithm depends of three factors. The first is the initial value for the set of parameters $\theta$. The second is the stopping criterion. We have used as the stopping criterion the value of $p$ as long as a number of iterations is not reached This means that the algorithm searches for a solution until a certain value of is reached or until a number of iterations is reached. The third speed factor is the number of parameters from the set $\theta$. If this number is large, than the algorithm is more time costly.

Like the algorithm from [13] our algorithm can be used in many situations: For example one can use it to estimate the vector $\theta$, starting from a random value for this vector. When the vector $\theta$ has a single parameter, for example the rotation angle, the speed is reasonable and the algorithm can be considered even for time dependent applications. The speed decreases with the increase in the number of parameters. If the registration task consists in finding a transformation with a large number of parameters, we recommend our algorithm when the time is not crucial, but instead the registration accuracy is. This recommendation is supported by the fact that our algorithm allows finding a solution with sub-pixel accuracy.

This algorithm can also be used for tuning the solution vector $\theta$. This requires of course that a less accurate solution was already found by a different method. For example as in [14], one can find a first alignment from low frequency components. In this situation, the speed of convergence to a highly accurate solution is reasonable for time dependent applications even when the number of parameters is large.

## V. Experimental Results

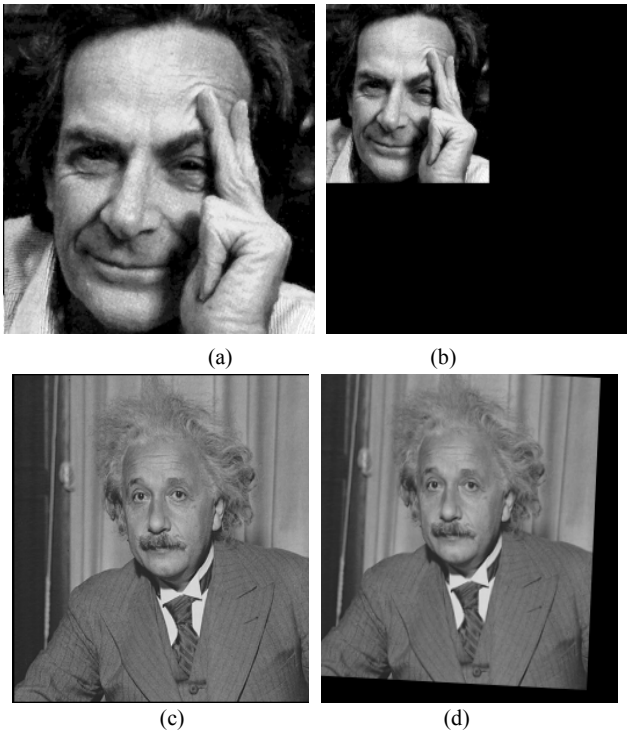
### A. Synthetic uni-modal data

In order to perform an experimental study on synthetic pairs, that contain images captured with the same imaging modality, we have used two sets of synthetic data

The first set of pairs contains pairs of images that are obtained one from the other by a single parameter transform. For every pair the value and the type of this parameter (i.e the parameter can be either the rotation angle, either, the scaling factor, either the translation on Ox or the translation on Oy) are randomly chosen. More specifically, the rotation angle is random in [-90˚,90˚], the translation parameters have random values in [0, size_of_the_image /2], the scaling factor is random in [0.1, 3]. We have produced 100 pairs of images. Fig. 3 (a) and 3 (b) shows an example of such a pair.

The second set of pairs contains pairs of images with images obtained one from another by a similarity transform. This consists of scaling, together with translation and rotation. We obtained 50 pairs. You can see in Fig. 3 (c) and(d) an example of such a pair. If we denote, for a pair of images, $(I_1, I_2)$, by $T\theta_t$ the real transform (the transform that is used to obtain $I_2$ from $I_1$) and by $T\theta_a$ the transform outputted by the algorithm, then the inaccuracy for this pair of images is given in (21).

$$\sum_{i=1}^{P} \frac{\left\| T\theta_t(x_i, y_i) - T\theta_a(x_i, y_i) \right\|}{P}, \text{ where } \left(x_i, y_i\right)_{i=\overline{1,P}} \text{ are random points in } I_1 \quad (21)$$



(a) (b) (c) (d)

Fig. 3 Example of artificial pairs of images.The image (b) is the scaled version of image (a).The image (d) is the image (c) transformed by a similarity transform

### B. Real uni-modal data

We have tested this algorithm on a real data set consisting of 50 outdoor image pairs. An example of pair of outdoor images used for testing the algorithm, is the one consisting in images 1 (a) and 1 (c). The algorithm has found, for every outdoor pair, a similarity transform that registered that pair. We compare it with the similarity transform inferred from our ground truth. The results for those synthetic sets are shown in Table I. The results computed by using (21) are shown in Table II

TABLE I
Results for Synthetic Data

| Type of the artificial transform | Percent of image pairs registered by our algorithm with an inaccuracy < 1 pixel |
|---|---|
| Single parameter | 100% |
| Multiple parameter | 100% |

TABLE II
Results for Outdoor Data.

| Image cathegory | Percent of image pairs registered by *our algorithm* with an inaccuracy < 5 pixels |
|---|---|
| Oudoor | 98% |

### C. Multi-modal data

We tested the capacity of our algorithm to register multi-modal images using images from the BrainWeb database [15]. These images contain 3% noise and 20% intensity non uniformity, in order to achieve realistic results. The ground truth alignment for the images is manually chosen by us. We

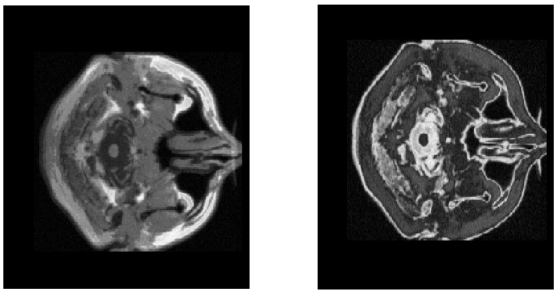
show in Fig. 4 examples from the image classes that we have used. We considered 50 pairs of T1 and T2 images, 50 pairs of T1 and PD images and 50 pairs of T2 and PD images. We obtained, by means of formula (21), the results summarized in TABLE III.

TABLE III
RESULTS FOR MULTI-MODAL BRAIN IMAGES

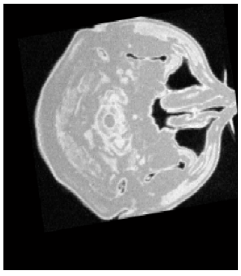
| Registered images | Percent of image pairs registered by our algorithm with an inaccuracy < 5 pixels |
|---|---|
| T1-T2 | 96% |
| T1-PD | 94% |
| T2-PD | 98% |

## IV. CONCLUSION

In this paper, we have presented a new mathematical model for the relation between the complex wavelet coefficients. The importance of this model derives from the fact that it provides a constraint, not on the values of the magnitudes of the wavelet coefficients, which are more prone to variations, but on their states, which are in exchange more stable to variations. In consequence, the registration algorithm, based on our model, is more robust to noise (and others variations). Other characteristics of the registration algorithm presented in this paper are: large applicability, the ellimination of any neccesity for preprocessing, and sub-pixel accuracy (this was proved only on the set of images artificially transformed, since the accuracy on real images (as it is outputed by the tests) depends on the limited precision of the humans that created the ground truth.



(a) (b)



(c)

Fig. 4 Example of brain images that we used in our tests. (a) is a T1 image, (b) is a T2 image and (c) is a proton density weighted image(PD)

## REFERENCES

[1] J. Modersitzki, *Numerical Methods for Image Registration*. New York, Oxford University Press, 2004, pp 1-74.
[2] B. Zitova, J. Flusser, Image registration methods: a survey. *Image and Vision Computing*, Vol. 21, No. 11, pp. 977-1000, 2003
[3] A. Goshtasby, *2-D and 3-D Image Registration for medical, remote sensing, and industrial applications.* New Jersey, John Wiley & Sons, Inc., Hoboken, 2005.
[4] J. L. Moigne, W. J. Campbell., R. F. Cromp, "An automated parallel image registration technique based on the correlation of wavelet features," *IEEE Trans. On Geoscience and Remote Sensing*, 40(8), pp. 1849–1864, 2002.
[5] N. M. Alpert, J. F Bradshaw, D. Kennedy, and J. A Correia,. "The principal axes transformation - A method for image registration,". *Journal of Nuclear Medicine* 31(10), pp. 1717–1722, 1990.
[6] P. Viola, W. M Wells III, "Alignment by maximization of mutual information,". in *International Conference on Computer Vision*, pp. 16–23, 1995.
[7] N. Ritter, R. Owens., J. Cooper., R.. H. Eikelboom., P. van Saarloos, "Registration of Stereo and Temporal Images of the Retina," *IEEE Trans. Medical Imaging*, Vol.. 18, No. 5, 1999.
[8] O. Pauly., N. Padoy., H. Poppert., L. Esposito., N. Navab, "Wavelet energy map: A robust support for multi-modal registration of medical images," in: *IEEE Conference on Computer Vision and Pattern Recognition*, pp.2184-2191, 2009.
[9] S. Li., J. Peng., J. T Kwok., J. Zhang, "Multimodal registration using the discrete wavelet frame transform," *Proc. of ICPR Conf*., pp. 877–880, 2006.
[10] I. W. Selesnick, R. G. Barniuk., N. G Kingsbury, "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Processing Magazine,* 2005.
[11] R. Gonzalez., R. Woods, *Digital Image Processing*. New Jersey, Prentice Hall Upper Saddle River, 2002, pp. 350-402.
[12] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc*., vol. 46, pp. 886–902, Apr. 1998.
[13] F. C. Calnegru. "A probabilistic framework for complex wavelet based image registration," *Lecture Notes in Computer Science*, 2011, vol 6978, pp 9-18, 2011
[14] Indian Institute of Information Technology, Allahabad, http://mtech.iiita.ac.in/A grade/Sukriti - Medical Image Registration using Next Generation Wavelets.pdf
[15] BrainWeb database http://www.bic.mni.mcgill.ca/brainweb/