Multi-modal Registration Using a Combined Similarity Measure

Juan Wachs[1], Helman Stern[1], Tom Burks[2], Victor Alchanatis[3]

[1] Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er-Sheva, Israel, 84105,
{juan, helman}@bgu.ac.il
[2]Agricultural and Biological Engineering, University of Florida, Gainesville, FL, 110570
TFBurks@ifas.ufl.edu
[3] Institute of Agricultural Engineering Volcani Center
Bet-Dagan, Israel, 50250,
victor@volcani.agri.gov.il

Abstract. In this paper we compare similarity measures used for multi-modal registration, and suggest an approach that combines those measures in a way that the registration parameters are weighted according to the strength of each measure. The measures used are: (1) cross correlation normalized, (2) correlation coefficient, (3) correlation coefficient normalized, (4) the Bhattacharyya coefficient, and (5) the mutual information index. The approach is tested on fruit tree registration using multiple sensors (RGB and infra-red). The combination method finds the optimal transformation parameters for each new pair of images to be registered. The method uses a convex linear combination of weighted similarity measures in its objective function. In the future, we plan to use this methodology for an on-tree fruit recognition system in the scope of robotic fruit picking

Keywords: Mutual information, multi-modal registration, similarity measures, sensor fusion

1 Introduction

Multi-modal image registration is a fundamental step preceding detection and recognition in image processing pipelines used by the pattern recognition community. This preprocessing stage concerns the comparison of two images –the base and sensed images- acquired from the same scenario at different times or with different sensors in a way that every point in one image has a corresponding point on the other images, in order to align the images. This procedure has a broad use in the medical field to obtain insights regarding bone structure (CT scans) or tissues softness (MR scans), or to see the evolution of a patient based on images obtained over the years, see [1][2][3] for reviews in this field. Other fields that relies on multi-modal image registration preprocessing are remote sensing [4],[5], surveillance [6] and to mention a few. Our main problem is the automatic registration of fruit trees images obtained

by multiple sensors. This is for the purpose of ultimately recognizing apples in a tree canopy using visual and thermal infrared cameras. In this paper, the initial problem of registering the fruit trees images by combining the images from different modalities is addressed. According to the position of the cameras with regard to the scene, it can be assumed that the transformation between images of different modalities are affine, which means rotations, translations, scaling and shearing are allowed. In this context, a standard image registration methodology called the correspondence problem includes the following steps [7]: a) Feature detection, b) Feature matching, and c) Transformation. Two approaches exist for feature detection step: a) Feature-based methods and b) Area-based methods. We focus in the area-based methods. The main representative in the area-based group was proposed by Viola and Wells [8] and is called mutual information methods, however a second group commonly used are the correlation-like methods [9].

A coarse-to-refined medical image registration method is presented in [10] by combining mutual information with shape information of the images. In [11] a new joint histogram estimation algorithm is proposed for computing mutual information to register multi-temporal remote sensing images, and is compared to correlation-like methods. Remotely sense image registration is addressed in [12] using the maximization of MI on a limited search space range obtained from a differential evolution strategy.

In this paper, an affine multi-modal registration method based on fusion and selection of SM is proposed. First, the sensed image is cropped to a "patch" and matched to areas in the base image. The best correspondence solution is achieved by maximizing the SM over the registration parameters search space. Solutions obtained by the SM are combined such that the mean squared error is minimized. We allow weighted use of all the similarity methods combined or the selection of the best method per affine parameter. Our approach is similar to [13] by adopting a combined methodology between powerful similarity measures, however we extend their work to include additional measures and to the affine registration scope.

The paper is organized as follows. A brief summary to mutual information and other similarity measures is given in section 2. In section 3, the proposed method is described in detail. Then we present experimental results in section 4, and finally conclusions in Section 5.

2 Mutual Information and Similarity Measures

Template matching methods are based on computing a SM between rectangular patches in an image pair. Corresponding patches between the two images are obtained when the maximum of the correlation is achieved. We will deploy three of these measures: cross correlation normalized (CC_1), correlation coefficient (CC_2), and correlation coefficient normalized (CC_3). The other two measures are histogram based: the Bhattacharyya coefficient (BC) and the Mutual Information index (MI). They rely on the joint/or and marginal histograms of the mutual areas. We denote V as a set of SM methods indexed as u = 1,2,3,4,5 for BC, MI, CC_1, CC_2 and CC_3, respectively.

2.1 Mutual Information (MI)

Let A, B be two random variables; let $p_A(a)$ and $p_B(b)$ be the marginal probability distributions; and let $p_{AB}(a,b)$ be the joint probability distribution

The degree of dependence between A and B can be obtained by their mutual information (MI):

$$I(A,B) = \sum_{a,b} p_{AB}(a,b) log \frac{p_{AB}(a,b)}{p_A(a).p_B(b)} \tag{1}$$

Given that H(A) and H(B) are the entropy of A and B, respectively, then their joint entropy is H(A,B); and H(A|B) and H(B|A) is the conditional entropy of A given B and B given A, respectively. Then, the entropy can be described by:

$$I(A,B) = H(A) + H(B) - H(A,B) \tag{2}$$
$$= H(A) - H(A|B) \tag{3}$$
$$= H(B) - H(B|A) \tag{4}$$

In terms of the marginal and joint probabilities distribution:

$$H(A) = -\sum_{a} p_A(a) log\ p_A(a) \tag{5}$$
$$H(A,B) = -\sum_{a,b} p_{AB}(a,b) log\ p_{AB}(a,b) \tag{6}$$
$$H(A|B) = -\sum_{a,b} p_{AB}(a|b) log\ p_{A|B}(a|b) \tag{7}$$

In the context of registration, A is the sensed and B the base (or reference) images, and *a* and *b* are the grayscale value of the pixels in A and B, then the marginal and joint distributions $p_A(a)$, $p_B(b)$ and $p_{AB}(a,b)$ can be obtained by the normalization of the marginal and joint histograms of the overlapping areas in A and B. I(A,B) is maximum when the overlapping areas in A and B are geometrically aligned.

2.2 Histogram comparison using the Bhattacharyya coefficient

In statistics, the Bhattacharyya distance measures the similarity of two discrete probability distributions. This attribute can be used to measure the similarity between two overlapping areas (A' $\subseteq$ A, B' $\subseteq$ B) in the sensed and base images.

BC(A',B') is maximum when the areas A' and B' are geometrically aligned.

$$BC(A',B') = \sum_{a \in A, b \in B} \sqrt{p_A(a) p_B(b)} \tag{8}$$

2.3 Correlation-like measures

Similarity measures (SM) are computed for pairs of overlapping areas between the sensed and base images. The maximum of these measures indicates corresponding areas. Let I_1 be the base image, I_2 be a patch in the sensed image with size (w ,h). Let i,j be coordinates in the overlapping area, then

The cross correlation normalized is:

$$CC_1(i,j) = \sum_{i',j' \in T} \frac{I_2(i',j') \cdot I_1(i+i', j+j')}{\sqrt{\sum_{i',j' \in T} I_2'(i,j)^2 \sum_{i',j' \in T} I_1'(i+i', j+j')^2}} \tag{9}$$

The cross correlation coefficient

$$CC_2(i,j) = \sum_{i',j' \in T} I_2'(i',j') \cdot I_1'(i+i', j+j')$$

where

$$I_2'(i'j') = I_2(i',j') - 1 \Big/ \left[(w \cdot h) \cdot \sum_{i'',j'' \in T} I_2(i'',j'') \right] \tag{10}$$

$$I_1'(i+i', j+j') = I_1(i+i', j+j') - 1 \Big/ \left[(w \cdot h) \sum_{i'',j''} I_1(i+i'', j+j'') \right]$$

The cross correlation coefficient normalized

$$CC_3(i,j) = I_2'(i',j') \cdot I_1'(i+i', j+j') \Big/ \sqrt{\sum_{i',j' \in T} I_2'(i,j)^2 \sum_{i',j' \in T} I_1'(i+i', j+j')^2} \tag{11}$$

Some of the limitations of the CC based methods are: a) they are not able to cope with pairs of images where the sensed (template) and the base images differ by more than slight rotation and scaling. b) they cannot represent the intensity dependence between images from different modalities. Nevertheless, their simplicity and low time complexity compared to the MI method makes them useful for real-time applications.

3 Methodology

3.1 Transformation

Given that we want to register two input images referred to as the, the base A and the sensed B images from different modalities, and assuming that the scene presented in A is totally contained in B we want to find a transformation based on rotations, translations and isometric scaling that transforms every point in A to a point in B. This means that there exists a geometric transformation T_α defined by the registration parameter α such that A(x,y) is related to B(x,y). The optimal parameter α* indicates that the images A and B are correctly geometrically aligned.

In this paper, we have restricted T_α to a 2D affine transformation. This is expressed by the registration parameter vector α by including a scaling factor s, a rotation angle θ (measured in degrees) and two translation distances t_x and t_y (measured in pixels). Hence $\alpha_j=[\alpha_{1j},\ldots, \alpha_{ij},\ldots, \alpha_{nj}]$ where α_{ij}=represents the ith parameter for the jth pair of images ($\alpha_{1j}=s$, $\alpha_{2j}=\theta$, $\alpha_{3j}=t_x$ and $\alpha_{4j}=t_y$ for the jth pair of images to be registered).

Transformation of the coordinates P_A and P_B from the sensed image A to the base image B is given by:

$$(P_B - C_B) = sR(\theta).(P_B - C_B) + t$$
$$R(\theta) = \begin{pmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{pmatrix} \quad t = \begin{pmatrix} t_X \\ t_Y \end{pmatrix} \tag{12}$$

Where C_A and C_B are the coordinates of the centers of the images, R(θ) is the rotation vector, and t is the translation vector.

Then, a *SM* based on measure m_μ tries to solve the registration problem by finding the optimal registration parameter α* such that the m_μ is maximized:

$$\alpha^* = \arg\max_{\alpha} \; m_\mu(A, B, \alpha) \tag{13}$$

Solving (13) does not always result in an optimum registration parameter, and the level of success in the registration depends on the capability of each SM to capture the relationship between the mutual areas in the images. However, it is possible that some SM's are more suitable for a specific registration parameters than others. In this vein, we try to find the best combination of SM's such that the absolute error between the optimal registration parameter and the observed one is minimized. The optimal registration parameter can be found in advance by manual registration.

3.2 Algorithm

Preprocessing. Initially, a contrast-limited adaptive histogram equalization algorithm is applied to the infra-red (IR) images to enhance their contrast. Secondly, we crop the sensed image to make a patch that fully overlaps the base image. We discard k=10% of each of the four sides around the IR images. This value was found empirically. A smaller value may leave areas not overlapping, and a higher value may leave out useful information and cause miss-registration.

Training. Given a training set of images $S=\{(a_1,b_1),\ldots\ldots,(a_m, b_m)\}$ where pair (a_j, b_j) represents the jth pair of sensed and base images that need to be registered. Let (w_a,h_a) and (w_b,h_b) be the with and height for the sensed and base images respectively. Let Λ and R be the parameter range for scaling and rotation respectively.

To register the images in the set S, follow the steps below:

1. For each sample pair of images $(a_j,b_j) \in S$. Set k=0.
    2. For each $s=\alpha_1 \in \Lambda$, scale a_j by factor s,
        3. For each $\theta= \alpha_2 \in \Phi$, rotate b_j by θ,
            4. For each coordinates $\alpha_3=x \in [0,\ldots,w_b-w_a]$ and $\alpha_4=y \in [0,\ldots,h_b-h_a]$

```
                5. Position the image a_j such that it left top corner coincides with x,y.
                6. Compute ρ_k=m_μ(a_j, b_j, α_k) for all μ=1,…,V.
                7. k=k+1,
            8. End
        9. End
    10. End
  11. Find α_jμ = arg max ρ
  12. Set Ω_jμ= α_jμ
13. End
```

Algorithm 1. Registration Algorithm

This algorithm results in a matrix Ω where the entries $\alpha_{ij\mu}$ are registration parameter i, for the pair of images j, using the SM m_μ. Given the true parameters obtained from manual registration, we denote the error of registration as: $e_{ij\mu} = (\alpha_{ij\mu} - \alpha_{ij\mu}*)$.

For a training set of size N, the root mean square error (RMS) for parameter i using m_μ is

$$RMS_{i\mu} = \frac{1}{N} \sum_{j=1,..,N} \sqrt{(\alpha_{ij\mu} - \alpha_{ij\mu}*)^2} \tag{14}$$

Therefore we can suggest a combined SM, CM with weight vector $\{w_{i1,...,}w_{ik},..,w_{i\mu}\}$ where the weight w_{ik} is associated with k^{th} SM for the registration parameter i, such that the RMS of the new combined method is minimized for each parameter of the registration vector. This is formulated in the following problem:

$$Min\, Z = RMS_{i\mu} = \frac{1}{N} \sum_{j=1,..,N} \sqrt{(\alpha_{ij\mu} \cdot w_{i\mu} - \alpha_{ij\mu}*)^2} \quad , \qquad \mu \in V,\, i \in U \tag{15}$$

s.t.:

$$\sum_{\mu \in V} w_{\mu i} = 1 \,, \quad i \in U \tag{16}$$

$$w_{\mu i} \geq 0 \quad , \quad \mu \in V,\, i \in U \tag{17}$$

The approach proposed above suggests that the each registration parameter can be corrected by some weight such that the RMS is minimized. This implies that indirectly we assign weights to each of the SM explored so they can 'collaborate' together towards an optimal registration. An alternative approach is the selective SM, noted by LM, based on constraining (17) to integers only, such that the registration parameters are only corrected by one SM, each time.

The equations (15-17) can be solved using different optimization methods; in this work we used a pattern search method for linearly constrained minimization [15].

Testing. For the testing set of images $S'=\{(a'_1,b'_1),\ldots\ldots,(a'_m, b'_m)\}$ where each pair (a'_j, b'_j) of images is to be registered. Repeat the registration algorithm 1, however after step 11 add the line: $\alpha_{j\mu} = \alpha_{j\mu} \,.\, w_{\mu j}$

The values of $w_{\mu j}$ can be floating point or integers according to the LM or CM, respectively. The testing performance is obtained using again the RMS measure over the number of testing samples.

4 Experiments

This section presents a comparison in the performances of five SM's in the context of 2D non-rigid registration using fruit tree images from different modalities: RGB and IR. The performance measures are: MI, BC, CC_1, CC_2 and CC_3 and two additional indicators introduced in this paper: a combined (CM) and the selective (LM) methods. The registration algorithm proposed was tested using 2 different datasets. Dataset 1 contains 28 pairs of color and infrared images captured from a 3-5 meters distance to the fruit trees, while the images in Dataset 2 were 19 pairs, obtained from a 8-10 meters distance. We use manual registration as a "ground truth" reference to validate the registration performance of the different SM for datasets 1 and 2. Table 1 lists the input parameters and the range of the registration parameters.

Table 1. Input Parameters

Parameter	Value	
w_b,h_b (RGB Width and Height)	2560, 1920	
w_a,h_a (IR Width and Height)	320, 240	
Scale range (Λ)	6.8→ 7.2; δ=0.05	5.2→ 5.6; δ =0.05
Angle range (Φ)	-3 → 3; δ=0.5	
Translation x range	W/α_1-w	
Translation y range	H/α_1-h	
Joint Histogram size (only for MI)	256 x 256	

The images of datasets 1 and 2 were registered using the five SM's: MI, BC, CC_1, CC_2 and CC_3. In each case the direct search optimization approach was used with the range of parameters in Table 1, and in the order (s, θ, t_x, t_y). Let $\alpha^*=\{s^*, \theta^*, t_x^*, t_y^*\}$ be the optimal parameters obtained from the manual registration and the registration parameters error be $\Delta\alpha = \{\Delta s, \Delta\theta, \Delta t_x, \Delta t_y\}$. Table 2 shows the MSE using both datasets (N=47) together for each SM. Note, that the best translations were obtained by the MI method, while cc_2 resulted in the best scale and rotation The minimum error is highlighted in bold letters). As a comparative example, the performances of each *SM* are illustrated in Figure 1, when registering the images (a_j,b_j) for j=1, for Dataset 1. Figure 1.(f) shows a close-up around the top right apple. The mutual information method was able to keep the whole shape of the apple.

For this example, the joint probability of the pair of images was plotted, see Figure 2. The axes x,y are the coordinates on the overlap area using the base image axes. The peak of the surface gives the solution for the MI method, where the second best is the cc_1 measure.

A second experiment was conducted for validation purposes using the k-fold cross validation method, where k=N. This time the new two measures were added to the evaluation, LM and CM and were compared to MI.

Table 2. Registration error obtained over Dataset 1+2

Measure	RMS			
	Δs	Δθ	Δtx	Δty
BC	0.605017	4.384621	62.97621	83.14283
MI	0.574475	3.770517	**25.68827**	**25.34178**
CC_1	0.59338	4.276189	59.82712	72.52225
CC_2	**0.522892**	**3.706814**	35.85383	42.03153
CC_3	0.524599	3.726948	36.03657	43.02969

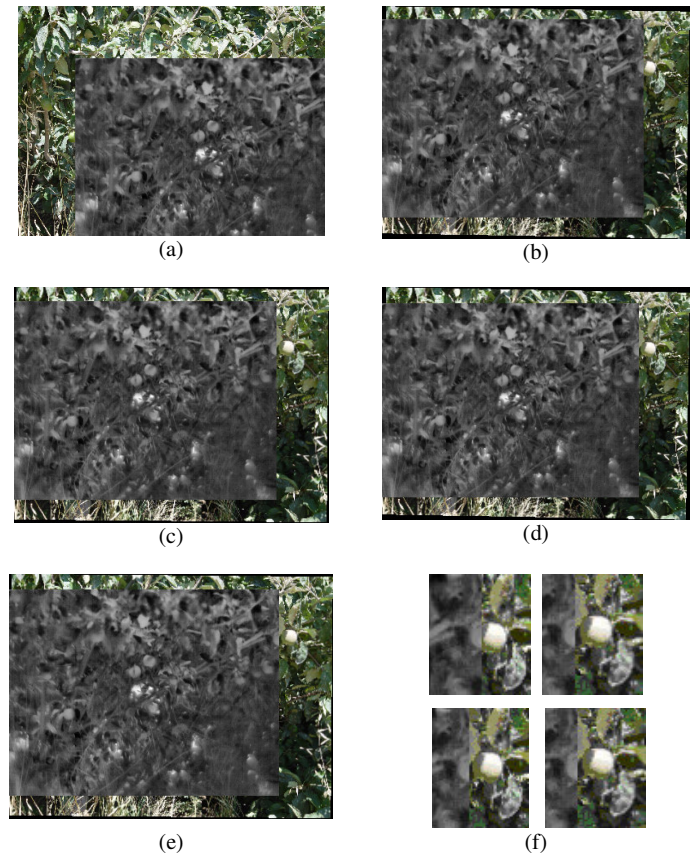

Figure 1. Comparison of the five SM for a pair of images: (a) BC, (b) MI, (c) CC_1, (d) CC_2, (e) CC_3, (f) zoom-in in the top right apple for the last 4 methods. The top left which corresponds to (b) resulted in the best registration.

The RMS obtained from all the sessions are presented in Table 3. The values of w_i were obtained from (15-18) using a pattern direct search method. The integer values of w_i were determined by selecting the w_i* that minimized the training error in the session, and later, the same value was used in the testing session.

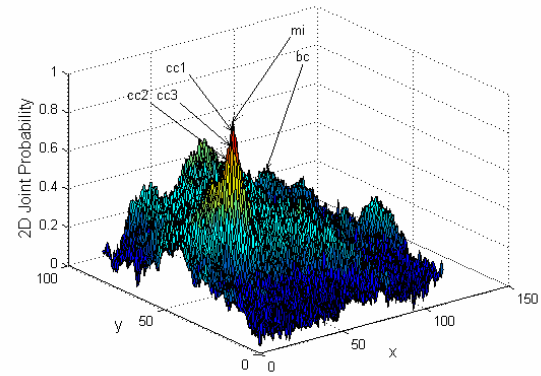

Figure 2. Mutual Information Search Function and best solutions for each method

Table 3. Registration error obtained over Dataset 1+2 using k-fold cross with k=47

		MI		CM		LM	
		Train	*Test*	*Train*	*Test*	*Train*	*Test*
RMS	Δs	0.574	**0.240**	**0.520**	0.252	0.522	0.255
	Δθ	3.762	**1.777**	3.758	1.807	**3.697**	2.042
	Δtx	25.681	**15.381**	**24.037**	16.719	25.681	15.381
	Δty	25.322	**13.013**	**24.479**	15.451	25.322	13.013

The results show that the training performance for the combined measure CM was better than MI and LM for training (except for the rotation parameter), with a significance <0.05 using a two tailored t-test. When testing, MI showed better results than the other SM however this was not statistically significant. This validation suggests that for the cohort of image types tested the transformation parameters found can not be generalized to register new images. It is recommended that new registration parameters be obtained for each new pair of images using the CM approach which finds the optimal SM weights.

4 Conclusions

In this work, we have presented a method of combining similarity measure of alternative image registration methods for multi-modal images registration. The measures used are: (1) cross correlation normalized, (2) correlation coefficient, (3) correlation coefficient normalized (CC_3), (4) the Bhattacharyya coefficient, and the (5) the mutual information index. The registration parameters found are weighted according to the strength of each measure to predict each type of parameter, the combined approach. An alternative approach is to find the best similarity measure per registration parameter.

We found that the combined approach performed statistically better than using each measure individually or using mutual information for the training sessions.

However, during testing it was found that the prediction capability of the combination and best similarity measure approaches are no better than that of using the mutual information measure. Since the training results were statistically better for the combination approach, which uses a convex linear combination of weighted similarity measures. In the future, we plan to use this methodology for on-tree fruit recognition system using multi-modal data, in the scope of robotic fruit picking.

Acknowledgments. This research was supported by Research Grant No US-3715-05 from BARD, The United States - Israel Binational Agricultural Research and Development Fund, and by the Paul Ivanier Center for Robotics Research & Production Management, Ben-Gurion University of the Negev.

References

1. Van den Elsen and Viergever M.A., Medical image matching - a review with classification, IEEE Engng. Med. Biol. vol. 12, pp.26-39, Mar. 1993.
2. Maintz, J.B.A., and Viergever, M.A. (1999). A Survey of medical image registration. In U. Spetzger, H.S. Stiehl, & J.M. Gilsbach (Eds.), Navigated Brain Surgery. pp. 117-136. Aachen: Verlag Mainz.
3. Lavallée, S. Registration for Computer Integrated Surgery: Methodology, state of the art. R.H. Taylor, S. Lavallée, G.C. Burdea, and R.W. Mosges, eds., Computer Integrated Surgery, Cambridge, MA: MIT Press, 1995.
4. Fan X., Rhody H. and Saber E. Automatic Registration of Multi-Sensor Airborne Imagery. AIPR workshop, Washington, D.C., Oct. 19-21, 2005.
5. Fransens R., Strecha C., Van Gool L. Multimodal and Multiband Image Registration using Mutual Information. Theory and Applications of Knowledge driven Image Information Mining, with focus on Earth Observation; EUSC, Madrid (Spain); March 17-18, 2004.
6. Krotosky, S.J.; Trivedi, M.M. Mutual Information Based Registration of Multimodal Stereo Videos for Person Tracking; Computer Vision and Image Understanding , vol. 106, no. 2-3, 2007.
7. Zitova B and Flusser, Jan. Image Registration methods: a survey. Image and Computing, 21, 2003, pp. 977-1000.
8. Viola P. and Wells WM. Alignment by maximization of mutual information, Intl. Journal of Computer Vision, 24 ,1997, pp. 137-154
9. Pratt WK. Digital Image Processing, 2nd ed., Wiley, New York, 1991.
10. Weiqing C. Zongying O., and Weiwei S, A Coarse-to-Refined Approach of Medical Image Registration Based on Combining Mutual Information and Shape Information. Intl. Conference on Neural Networks and Brain, 2005. ICNN&B '05. 2005. pp. 816- 820.
11. Hua-Mei Chen; Varshney, P.K.; Arora, M.K. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Trans. on Geoscience and Remote Sensing, Nov. 2003, vol. 41, no. 11, pp. 2445-2454.
12. De Falco, Della Cioppa A., Maisto D., Tarantino E. Differential Evolution for the Registration of Remotely Sensed Images. Soft Computing in Industrial Applications, Recent and Emerging Methods and Techniques. Advances in Soft Computing, vol. 39 Saad, A.; Avineri, E.; Dahal, K.; Sarfraz, M.; Roy, R. (Eds.) 2007.
13. Roche A., Malandain, G and Ayache, N. Unifying Maximum Likelihood Approaches in Medical Image Registration. Int. J. of Imaging Systems and Technology, 11:71,2000.
14. Audet, Charles and J.E. Dennis Jr. Analysis of Generalized Pattern Searches. SIAM Journal on Optimization, vol. 13, no. 3, pp. 889-903, 2003.