Multimodality Image Registration using Ordinary Procrustes Analysis and Entropy of Bivariate Normal Kernel Density

Wan-Hyun Cho[1], Sun-Worl Kim[1], Myung-Eun Lee[2], Soo-Hyung Kim[2], Soon-Young Park[3], and Chang-bu Jeong[4]

***Abstract*— We present a registration method for medical images based on shape information and voxel intensities. First, we segment volume images using the Markov random field and the Gibbs distribution. We extract the 3D feature points of the shape from the surface of the segmented object. Then, we conduct first registration using ordinary Procrustes analysis for two sets of 3D feature points. For the second registration, we define the new optimization measure of registration as the entropy of the bivariate normal kernel density for pairs of intensities given from the extracted feature points as well as the transformed feature points. The final registration for two volume images is carried out by finding the appropriate transformation parameter yielding the minimum value of this optimization measure. To evaluate the performance of the proposed registration method, we conduct various experiments comparing our method with existing ones such as the Mutual Information measure.**

I. INTRODUCTION

MEDICAL image registration is a newly emerged task in medical image processing is used to match two independently acquired images. To register medical images, the geometrical relationship between them is determined using various spatial alignment methods. Matching all of the geometric data available for a patient provides better diagnostic capability, better understanding of data, and improves surgical and therapy planning and evaluation. The imaging modalities are acquired by tomography modalities including CT (computed tomography), MRI (magnetic resonance imaging), X-ray, US (ultrasound), and PET (positron emission tomography).

Generally, registration algorithms in medical images can be broadly classified into three regions. These criteria can be landmark-based, segmentation-based, and intensity-based [1] [2] [3]. Landmark-based registration uses salient features selected by the user. These features are usually points but can also be lines or more complex structures such as corners. Since the number of identified features is sparse compared to the image content, landmark-based methods are fast to compute. The major drawback is that this method requires user interaction to locate the landmarks.

Segmentation-based methods attempt either rigid or deformable alignment for two binary structures obtained by segmentation [4]. The segmented structure of one image can be aligned either to a segmented structure on the second image or to the whole unsegmented second image. In the latter case, the criteria typically require that the boundary of the binary structure matches to the edges in the second images. Due to reduction of information, segmentation-based methods are faster than methods using full image content. On the other hand, one of the drawbacks of segmentation-based methods is that the performance of the registration relies on the accuracy of the segmentation pre-processing step.

Intensity-based methods operate directly on the image intensity. They are more flexible than landmark-based or segmentation-based methods as they use all of the available information without previous reduction of data either by the user or by a segmentation algorithm. These methods are typically automatic. However, using full image contents is computationally very expensive especially for 3D images and hence may not be suited to time-constrained applications. There have been many research efforts on the registration methods using image intensities. In many cases, registration optimum measures often use Mutual Information (MI) or the maximum likelihood (ML) which are based on the pixel intensity of the registered image [5][6][7][8][9]. The MI, originating from information theory, is a measure of statistical dependency between two data sets. It is particularly suitable for the registration of images from the same or different modalities. Also, with a ML approach to image registration we assume that the pixel values in two images for the registration are probabilistically related. When the likelihood has its maximum value, the two images are considered to be registered.

In this paper, we propose a new registration method combining the segmentation-based approach and the intensity-based approach. First, we segment two images using Markov random field model and Gibbs distribution. Next, we extract the feature points from the segmented images, and then we apply an ordinary Procrustes analysis (OPA) to two sets of extracted feature points to conduct the initial registration. Second, we define the new optimization measure of registration as the entropy of the bivariate normal kernel density (EBND) computed from the image intensities

[1]*Department of Statistics, Chonnam National University, Korea.* *whcho@chonnam.ac.kr, sunworl@gmail.com*
[2]*Department of Computer Science, Chonnam National University, Korea.* *melee@chonnam.ac.k, shkim@chonnam.ac.kr*
[3]*Department of Electronics Engineering, Mokpo National University, Korea.* *sypark@mokpo.ac.kr*
[4]*Department of Internet Software, Honam University, Korea.* *cbjeong@honam.ac.kr*

of corresponding voxels in both the reference and floating images. If the reference and floating images are geometrically aligned, this measure will take a minimum value. Finally, we conduct various experiments to assess the accuracy and precision of our method by comparing the EBND based registration with the MI based registration.

II. Segmentation using Markov random field model and Gibbs distribution

The Markov random field (MRF) is a class of statistical models that describe contextual constraints [10] [11]. It can be interpreted as a generalization of the Markov chain models, which describe temporal constraints. It also provides a convenient way to combine both the observed intensity and spatial information under a Bayesian framework.

To formally describe MRF modeling, we first consider a neighborhood system. We need to define a clique. Let S denote a lattice indexing the pixels in a given target region. Let s be the lattice point (or pixel). The neighboring system N of each lattice point s contains its neighbor points. It must be symmetric. A clique is a set of points, $c \in C$, which are all neighbors of each other. An 8-point neighborhood system N_s around center pixels is used throughout this study. Let $L = \{1,\ldots,k\}$ denote the label set indicating each group. The number k is 3 (GM, WM and CSF) in our case.

Let the random variable X denote the labeling process of S such that $x_s \in L$ is the value of X at pixel s [5]. Then the Markov property is expressed by

$$\begin{aligned} P(X_s = x_s \mid X_r = x_r,\, r \in S,\, r \neq s) \\ = P(X_s = x_s \mid X_r = x_r,\, r \in N_s). \end{aligned} \tag{1}$$

For practical use, a means to specify these conditional probabilities is required. This is provided by the equivalence of the MRF model and the Gibbs random field (GRF) model, since the GRF model can be specified in terms of clique potentials. That is, for modeling labeling compatibilities in an MRF, only cliques have to be considered. According to the Hammersley and Clifford theorem, the density of X is given by the following Gibbs density:

$$p(\mathbf{x}) = Z^{-1} \exp\{-\beta U(\mathbf{x})\}, \tag{2}$$

where Z is a normalizing constant known as the partition function and $U(\mathbf{x})$ is the energy function composed of the clique potentials V_c:

$$U(\mathbf{x}) = \sum_{c \in C} V_c(\mathbf{x}). \tag{3}$$

The potentials for a point clique containing more than one site are defined as

$$V(x_c) = \begin{cases} -\beta_c & \textbf{if all sites in c have the same label} \\ \beta_c & \textbf{otherwise} \end{cases} \tag{4}$$

This formula expresses only the prior distribution.

Let the observed image $\mathbf{y}$ be a realization of a random field $\mathbf{Y}$. Let $\mathbf{x}^*$ be the true unknown label of the observed pixels and let $\hat{\mathbf{x}}$ indicate an estimate of $\mathbf{x}^*$. The objective now is to find $\hat{\mathbf{x}}$ given $\mathbf{y}$. In a Bayesian framework, $p(\mathbf{x})$ can be viewed as the prior distribution for the true image $\mathbf{x}^*$. Then the posterior probability is, by Bayes' theorem [12][13], proportional to

$$p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{x}). \tag{5}$$

In general, the conditional densities $p(\mathbf{y} \mid \mathbf{x})$ of the observed image are generally modeled as a Gaussian distribution in image processing. But we use here an exponential distribution because the distribution shape of a brain image exhibits the similar appearance with an exponential distribution. The conditional density function of the observed intensity y given the class x_s is given as

$$p(y \mid x_s) = \frac{1}{b_s} \exp\left(-\frac{(y - a_s)}{b_s}\right), \quad a \le y \le \infty \tag{6}$$

where a_s and b_s are the distribution parameters of each class. Then, the estimator of unknown label variable $\hat{\mathbf{x}}$ can be obtained by computing the maximum a posterior (MAP) probability. That is, it is given as

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\max_{\mathbf{x}} \log(p(\mathbf{x} \mid \mathbf{y})) \\ &= \arg\max_{\mathbf{x}} \{-\log b - (\mathbf{y} - a)/b - \beta U(\mathbf{x})\}. \end{aligned} \tag{7}$$

Here, we note that the logarithm of the posterior density is being maximized and all of the constant terms, not affecting the maximizing, are removed. However, this maximization is a computationally huge task because the number of possible configurations for pixel labels is too many. In order to remedy this situation, we use a deterministic algorithm called iterated conditional modes (ICM) which maximizes local conditional probabilities sequentially. The ICM algorithm solves this maximization by sequentially minimizing the following equation at each pixel:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}_s \in L} \{\log b_s + (\mathbf{y} - a_s)/b_s + \beta U(x_s)\} \tag{8}$$

Finally, as we assign each pixel to a class with large posterior probability, we can segment the brain image.

III. The First Registration using An Ordinary Procrustes Analysis

We suppose that the configuration matrix $\mathbf{X}$ is the $(k \times m)$ matrix of Cartesian coordinates of the k feature

points or landmarks in m-dimensions extracted from the segmented boundary surface of floating volume images $\mathbf{F}$. The configuration matrix $\mathbf{Y}$ is also the matrix of Cartesian coordinates of the k feature points selected from the reference volume images $\mathbf{R}$. Then, we can center each matrix of feature points by multiply the centering matrix

$$C = \mathbf{I}_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^T, \tag{9}$$

where $\mathbf{I}_k$ is the $(k \times k)$ identity matrix, and $\mathbf{1}_k$ is the vector of ones.

In order to match the two centered configurations in shape we need to establish a measure of distance between the two matrices. We primarily concentrate on the full Procrusteres distance [14], which is defined as follows.

Definition 3.1 The full Procrustes distance between X and Y with centered matrix $\mathbf{X}_C = \mathbf{C}\mathbf{X}$ and $\mathbf{Y}_C = \mathbf{C}\mathbf{Y}$ is

$$D^2_{OPA}(\mathbf{X},\mathbf{Y}) = \left\|\mathbf{Y}_C - \beta\mathbf{X}_C\boldsymbol{\Gamma} - \mathbf{1}\boldsymbol{\alpha}^T\right\|^2, \tag{10}$$

where $\|\mathbf{X}_C\| = \{trace(\mathbf{X}_C^T\mathbf{X}_C)\}^{1/2}$ is the Euclidean norm, $\boldsymbol{\Gamma}$ is a rotation matrix, β is a scale parameter, and $\boldsymbol{\alpha}$ is an location vector.

In this case, we wish to match the two configurations as closely as possible up to a similarity transformation. To carry out this matching, we use least square techniques. It requires estimating all of the parameters minimizing the full Procrustes distance given in **Definition 3.1**.

Result 3.2 The ordinary Procrustes solution to the minimization of the full Procrustes distance is given by

$$\boldsymbol{\alpha} = \mathbf{0}, \ \hat{\boldsymbol{\Gamma}} = \mathbf{U}\mathbf{V}^T \tag{11}$$

where

$$\mathbf{Y}_C^T\mathbf{X}_C = \|\mathbf{X}_C\|\|\mathbf{Y}_C\|\mathbf{V}\Lambda\mathbf{U}^T, \tag{12}$$

and Λ is a $(m \times m)$ diagonal matrix of positive eigenvalues of $\mathbf{X}_C^T\mathbf{Y}_C\mathbf{Y}_C^T\mathbf{X}_C$. Furthermore,

$$\beta = \frac{trace(\mathbf{Y}_C^T\mathbf{X}_C\hat{\boldsymbol{\Gamma}})}{trace(\mathbf{X}_C^T\mathbf{X}_C)}. \tag{13}$$

Proof: We wish to minimize

$$D^2_{OPA} = \left\|\mathbf{Y}_C - \beta\mathbf{X}_C\boldsymbol{\Gamma} - \mathbf{1}_k\boldsymbol{\alpha}^T\right\| = trace(\|\mathbf{Y}_C\|^2 + \beta^2\|\mathbf{X}_C\|^2 - 2\beta\,\mathbf{Y}^T\mathbf{X}\boldsymbol{\Gamma}) + k\boldsymbol{\alpha}^T\boldsymbol{\alpha},$$

where $\mathbf{X}_C$ and $\mathbf{Y}_C$ are centered. So, we must have $\boldsymbol{\alpha} = \mathbf{0}$. If we define the centerd pre-shape of a configuration matrix $\mathbf{X}$ and $\mathbf{Y}$ as

$$\mathbf{Z}_\mathbf{X} = \frac{\mathbf{X}_C}{\|\mathbf{X}_C\|}, \quad \mathbf{Z}_\mathbf{Y} = \frac{\mathbf{Y}_C}{\|\mathbf{Y}_C\|},$$

then we need to minimize

$$trace(\|\mathbf{Y}_C\|^2 + \beta^2\|\mathbf{X}_C\|^2 - 2\beta\|\mathbf{Y}_C\|\|\mathbf{X}_C\|\mathbf{Z}_\mathbf{Y}^T\mathbf{Z}_\mathbf{X}\boldsymbol{\Gamma}).$$

Since the first two terms in the parenthesis of the trace are positive, to minimize the above trace becomes finding the maximum of $trace(\mathbf{Z}_\mathbf{Y}^T\mathbf{Z}_\mathbf{X}\boldsymbol{\Gamma})$ over the orthogonal matrix $\boldsymbol{\Gamma}$. First, we consider a singular value decomposition of $\mathbf{Z}_\mathbf{Y}^T\mathbf{Z}_\mathbf{X}$ given by

$$\mathbf{Z}_\mathbf{Y}^T\mathbf{Z}_\mathbf{X} = \mathbf{V}\Lambda\mathbf{U}^T, \tag{14}$$

where $\mathbf{V}$ and $\mathbf{U}$ are orthogonal matrices and Λ is a diagonal matrix with eigenvalues $\lambda_1, \cdots, \lambda_m$ of the matrix $\mathbf{Z}_\mathbf{Y}^T\mathbf{Z}_\mathbf{X}$. Hence, the above trace using this decomposition is equivalent to

$$\max_\mathbf{R} trace(\mathbf{R}\Lambda) = \sum_{i=1}^{m} r_{ii}\lambda_i \tag{15}$$

where $\{r_{11}, \cdots, r_{mm}\}$ are the diagonal elements of the orthogonal matrix $\mathbf{R}$. Now the set of diagonals of $\mathbf{R}$ is a compact convex set with extreme points $\{(\pm1, \cdots, \pm1)\}$ with an even number of minus signs. Hence, it is clear in our case that the maximum is achieved when we choose

$$r_{ii} = 1, \ i = 1, \cdots, m.$$

Thus, we have taken the minimization rotation matrix as $\hat{\boldsymbol{\Gamma}} = \mathbf{U}\mathbf{V}^T$ because it is obtained from the result that

$$trace(\mathbf{Z}_\mathbf{Y}^T\mathbf{Z}_\mathbf{X}\boldsymbol{\Gamma}) = trace(\mathbf{V}\Lambda\mathbf{U}^T\mathbf{U}\mathbf{V}^T) = trace(\Lambda). \tag{16}$$

Second, differentiating the full Procrustes distance with respect to β, we obtain:

$$\frac{\partial D^2_{OPA}}{\partial\beta} = 2\beta\,trace(\mathbf{X}_C^T\mathbf{X}_C) - 2\,trace(\|\mathbf{Y}_C\|\|\mathbf{X}_C\|\mathbf{Z}_\mathbf{Y}^T\mathbf{Z}_\mathbf{X}\boldsymbol{\Gamma}). \tag{17}$$

Hence we can estimate the scale parameter as the following value:

$$\hat{\beta} = \frac{trace(\mathbf{Y}_C^T\mathbf{X}_C\hat{\boldsymbol{\Gamma}})}{trace(\mathbf{X}_C^T\mathbf{X}_C)}. \tag{18}$$

Finally, we have only considered the rigid-body transformation in this paper. Therefore, the value of the scale

parameter can be taken as one. Then, the transformation of the Cartesian coordinates $\mathbf{X}$ to $\mathbf{Y}$ from the floating image $\mathbf{F}$ to reference image $\mathbf{R}$ is expressed by

$$(\mathbf{Y} - C_{\mathbf{Y}}) = (\mathbf{X} - \mathbf{C}_{\mathbf{X}})\boldsymbol{\Gamma} + \mathbf{1}_k \boldsymbol{\alpha}^T, \quad (19)$$

where $C_{\mathbf{X}}$ and $C_{\mathbf{Y}}$ are the Cartesian coordinates of the centers of the two image, $\boldsymbol{\Gamma}$ represents the rotation matrix and $\boldsymbol{\alpha}$ is the translation vector. Hence, if we use the conclusion of **Result 3.2**, we have that the initial value of the rigid transformation parameters is given:

$$\hat{\boldsymbol{\Gamma}} = \mathbf{U}\mathbf{V}^T \text{ and } \boldsymbol{\alpha} = \mathbf{0}. \quad (20)$$

Therefore, we have obtained that the initial transformation from Cartesian coordinates $\mathbf{X}$ to $\mathbf{Y}$ be given:

$$\mathbf{Y} = (\mathbf{X} - \mathbf{C}_{\mathbf{X}})\hat{\boldsymbol{\Gamma}} + C_{\mathbf{X}} - C_{\mathbf{Y}}. \quad (21)$$

IV. The Second Registration using the Entropy of the Bivariate Normal Kernel Density

To achieve the final registration for the two given images, we define the new optimal measure based on the entropy of the bivariate normal density function. Let $\mathbf{S}$ be the overlapping volume of the floating images $\mathbf{F}$ and the reference image $\mathbf{R}$. We suppose that the image intensity u_i $(i = 1, \cdots, n)$ of the position $\mathbf{s}$ in the floating image $\mathbf{F}$ corresponds to the intensity v_j $(j = 1, \cdots, m)$ at the transformed position $T_\alpha(\mathbf{s})$ in the reference image $\mathbf{R}$.

Here, we first define the bivariate normal kernel density computed from all pairs of two intensities (u_i, v_j). This is generally defined by the product of the univatriate kernel densities. It is given by the formula:

$$f_K(u,v) = \frac{1}{|\mathbf{S}|}\sum_{i,j}^{n,m}\left(\frac{1}{h_1}K\left(\frac{u - w_{ij1}}{h_1}\right)\cdot\frac{1}{h_2}K\left(\frac{v - w_{ij2}}{h_2}\right)\right),$$

where w_{ijk} is the k-th component of the (i,j)-th observation, and h_k is the bandwidth of the k-th component density.

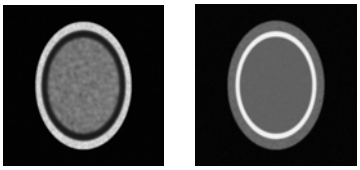
Next, we define the new optimal measure using the entropy of this density. We call this measure the entropy of bivariate normal density function (EBND). It is defined:

$$EBND = -\sum_u\sum_v f(u,v)\log f(u,v). \quad (22)$$

It can be used as the measure of registration representing how much the floating image aligns with the reference image. We note that the more the two images align well, the more the value of the EBND measure decreases. Hence, we can find an optimal transformation T_α of an image $\mathbf{F}$ with respect to an image $\mathbf{R}$ by searching the minimum value of the EBND measure used as the criterion of the image registration

V. Experimental Results

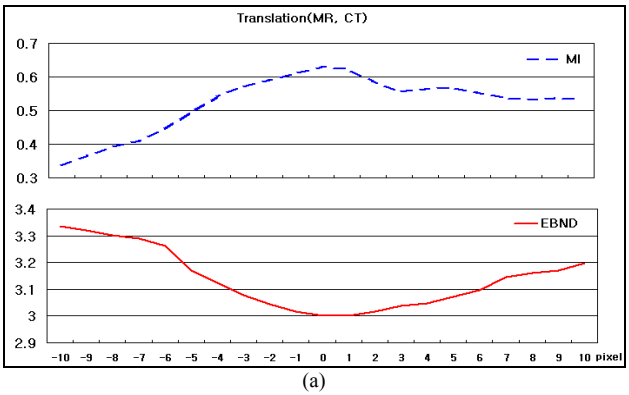
We evaluated the efficacy of our registration method on both synthetic and clinical datasets. First, we evaluated the performance of the proposed EBND measure for multimodality registration using synthetic CT and MR volume datasets. The synthetic MR datasets were 256x256x40, 8-bit grayscale volumes and also the synthetic CT datasets were 512x512x40, 8-bit grayscale volumes. Note that the general head volume images are composed with background, tissue, skull, and brain regions. Thus, we have constructed the synthetic volume images so that they have same structure as actual head images. Fig.1 shows the 10th slice of synthetic MR and CT volumes, respectively.



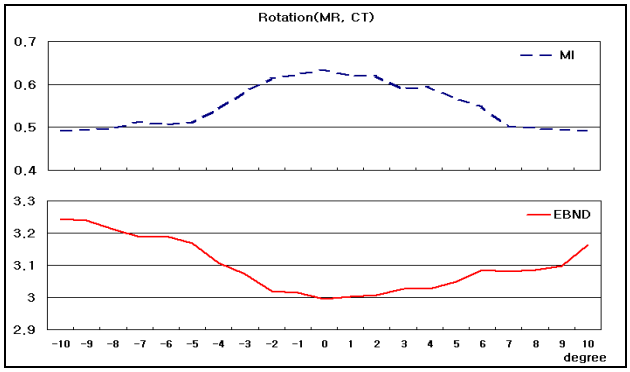
(a) (b)

Fig. 1. Synthetic datasets: (a) the 10th slice of MR synthetic volumes, and (b) the 10th slice of CT synthetic volumes

We investigated the precision of our method by comparing registration traces with those of an MI based registration. Fig. 2 displays the registration traces for the x-axis translation and rotation of CT floating volumes over MR reference volumes. As expected, we observed that the MI-based registration had a maximum value at the optimal location (0 pixel or 0 degree) but the EBND-based registration had a minimum value at the optimal location.



(a)



(b)

Fig. 2. MR to CT registration: (a) a translation in the range from -10 to +10 pixel, and (b) a rotation in the range from -10 to +10 degree.

Second, to compare the efficiency of our method with a MI based method for clinical datasets, we executed the rigid registration of unimodality and multimodality using the ITK Toolkit [15]. As for the unimodality registration, Table I provides the parameter values obtained from the experimental results for the MR volume images. The extracted parameters of the EBND and MI based registrations were comparable to those of the true registration parameters.

TABLE I
COMPARISON RESULTS BETWEEN THE MI AND EBND METHODS

Parameter \ Method		EBND	MI	Optimal value
Translation (pixel)	T_x	15.1185	14.863800	15
	T_y	0.2380	0.034089	0
	T_z	-0.9167	0.611958	0
Rotation (quaternion)	Real	0.9962	0.996146	0.996
	R_x	-0.0014	0.000686	0
	R_y	0.0018	0.000723	0
	R_z	-0.0862	-0.087706	-0.086

Fig. 3 also shows the final unimodality registration results of the two methods for the MR volume images. Two MR images are shown as a checkerboard where each block alternately displays data from each MR image. The checkerboard images before and after registration display visibly the effect of registration. From our results, we note that the MI and EBND methods provide a relatively good registration outcome.

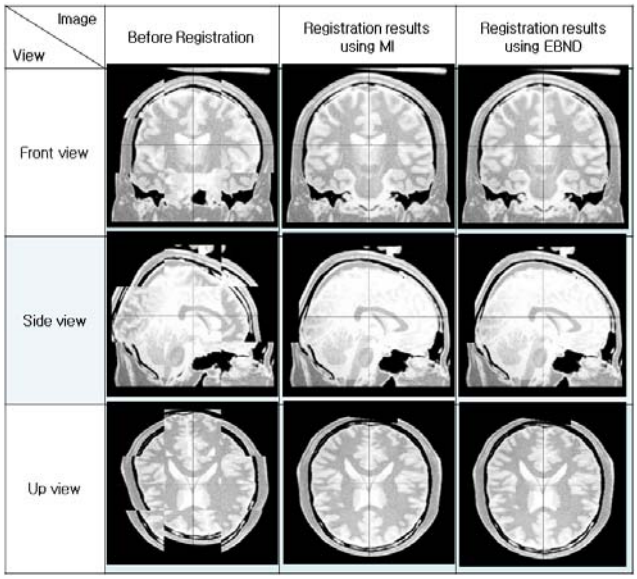


Fig.3. Registration results for the MI and EBND methods using a unimodality volume image

For multimodality registration, the MR volume images were used as the reference image and the CT volume images were used as the floating images. To assess the OPA for the initial registration, we applied the MRF model-based segmentation and Procrustes fit to the MR and CT images. Fig. 4 shows the process of an initial registration which includes segmentation, feature points extraction, and a Procrustes fit. Here, the feature points were extracted by sub-sampling the boundary into an equal spacing of 15 points. We noted that the shapes of the brain object were well preserved and the MR and CT images were roughly aligned.

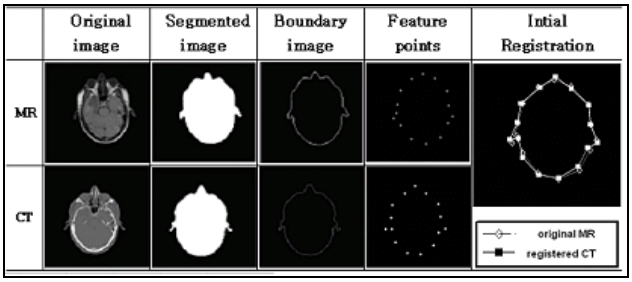


Fig. 4. The process of an initial registration applied to the MR and CT volume images.

Table II provides the parameter values obtained from the registration results for the two volume images. The initial registration using the OPA yielded the rough alignment and the final registration resulted in the more accurate alignment. Hence we can also expect that the registration parameter values given by EBND method are comparable to the MI method.

TABLE II
COMPARISON RESULTS OF THE TWO METHODS IN A MULTIMODALITY REGISTRATION

Parameter \ Method		Initial Registration		Final Registration	
		OPA	MI	OPA+EBND	MI
Translation (pixel)	T_x	-33.6902	0	-27.7393	-26.3844
	T_y	22.4839	0	24.5770	22.9486
	T_z	0.1063	0	11.8428	11.9285
Rotation (quaternion)	Real	0.9990	1	0.9985	0.9982
	R_x	0.0000	0	-0.026	-0.0207
	R_y	0.0000	0	-0.0044	0.0022
	R_z	0.0380	0	0.0465	0.0436

Fig. 5 shows the registration results for the MR and CT volumes using the MI and EBND registration methods. The CT and MR images are shown as a checkerboard. The bright color values represent the skull of the CT image and the contour in the less bright color represents the tissue of the MR image. After registration, the alignment can be observed from the overlap between the bright region representing the skull on the CT image and the corresponding dark region in the MR image. In each case, the registered floating images are in good agreement with the referenced images.

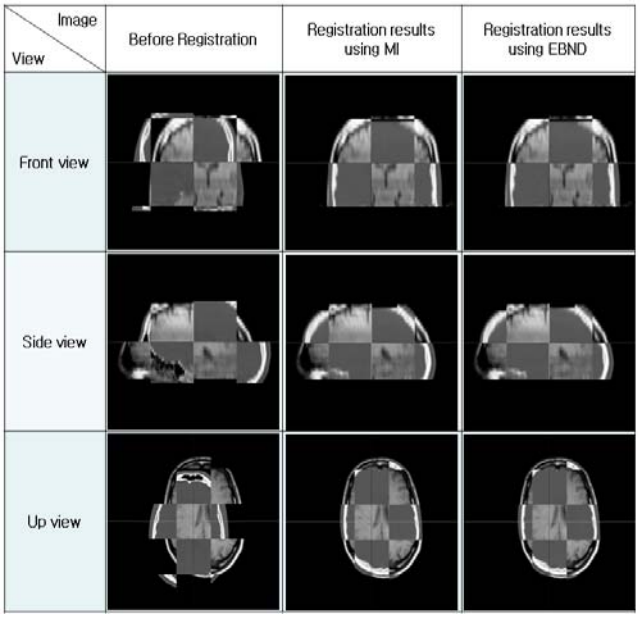


Fig. 5. Registration results for the MI and EBND methods using multimodality volume images

VI. CONCLUSIONS

We presented a new registration method using both a ordinary Procrustes analysis and the EBND function. The initial registration was conducted by using ordinary Procrustes analysis and the final registration was based on the proposed EBND measure. We defined this function by using the entropy of the joint density of the bivariate normal kernel given from two intensities where one of them maps with the other intensity. The precision of our measure was evaluated by comparing registration traces obtained from a MR image and transformed CT images. Experimental results showed that the proposed registration method was highly robust and accurately aligned the referenced volume images and the floating ones at the global minimum of the EBND measure.

ACKNOWLEDGEMENTS

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. R01-2007-000-20486-0).

REFERENCES

[1] M. Sonka and J. M. Fitzpatrick, Handbook of Medical Imaging: Volume 2. *Medical Image Processing and Analysis*, SPIE Press, Washington, 2000.

[2] T. S. Yoo, *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*, 2004.

[3] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration", *Medical Image Analysis, 2(1),* 1998, pp. 1-36.

[4] H. Lee and H. Hong, "Robust Surface Registration using a Gaussian weighted Distance Map in PET-CT Brain images", *Iberoamerican Congress on Pattern Recognition*, LNCS 3773, 2005, pp. 794-803.

[5] J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, "Mutual information based registration of medical images: a survey", *IEEE Transaction on Medical Imaging*, 20(Y), 2003, pp. 1-20.

[6] F. Mases , A. Collignon, D. Vandermeulen,, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information", *IEEE Transaction on Medical Imaging*, 16(2), 1997, pp. 187-198.

[7] Y. M. Zhu and S. M. Cochoff,, "Likelihood Maximization Approach to Image Registration", *IEEE Transaction on Image Processing*, 11(12), 2002, pp. 1417-1426.

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

[9] J. V. Hajnal, D. L.G. hill and D. J. Hawkes, "Medical Image Registration", *CRC Press*, New York, 2001.

[10] M. A. Wirth, J. Narhan and D. Gray, "Non-rigid mammogram registration using mutual information", SPIE Medical Imaging: Image Processing, Vol. 4684, 2002, pp. 562-573.

[11] Y. M. Zhu and S. M. Cochoff,, Likelihood Maximization Approach to Image Registration, *IEEE Transaction on Image Processing*, 11(12), 2002, pp. 1417-1426.

[12] R. Gan and A. C. S. Chung, "Distance-Intensity for Image Registration", LNCS, Vol. 3765, 2005, pp. 281-290.

[13] G. Wen, Z. Wang, S. Xia and D. Zhu, "Least-squares fitting of multiple M-dimensional points sets," Visual Computation, Vol. 22, 2006, pp. 387-398.

[14] P. Kelland and P. G. Tait, "Introduction to Quaternions," Adamant Media Corporation, 2005.

[15] L. Ibanez, W. Schroeder, L. Ng and J. Cates, "The ITK software Guide," Kitware, Inc.