RULE-BASED DECISION-MAKING FRAMEWORK FOR KNOWLEDGE-BASED ANATOMICAL LANDMARK LOCALIZATION (K-BALL)*

Mohammad-Reza Siadat[a], Hamid Soltanian-Zadeh[a,b], and Kost Elisevich[c]

[a]Radiology Image Analysis Lab., Henry Ford Health System, Detroit, MI 48202, USA
[b]Control and Intelligent Processing Center of Excellence, Department of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran
[c]Department of Neurosurgery, Henry Ford Health System, Detroit, MI 48202, USA

ABSTRACT

K-BALL is a general method for localization of anatomical phenomena of the same origin with natural discrepancies distributed over a reference space, e.g., human brain anatomical structures. In this paper, we focus on information analysis step (2nd step) of K-BALL during which landmarks extracted in its first step are evaluated. We provide a framework in which rules are automatically generated based on estimated and derived models. We show that the rules based on the derived models can improve the overall success rate of K-BALL. Each rule evaluates the extracted points by producing an intermediate confidence factor (*ICNF*). A total confidence factor is calculated using *ICNF*'s to facilitate the acceptance or rejection of a set of points as landmarks of interest. Using the rules merely based on the estimated models, simulation study produced an overall success rate of 91.8%. Using the rules based on both of the estimated and derived models, this rate increased to 92.5%.

1. INTRODUCTION

Anatomical Landmark localization is important [1] as it provides: 1) initial information for registration, 2) navigation and retrieval guidance through the image data [2], 3) initial models for segmentation [3], and 4) valuable (though rough) information about morphologic or volumetric features of the organs or structures of interest [2]. We have proposed a two-step knowledge-based method to localize the hippocampus in the human brain. The search for this structure passes through lateral landmarks of the lateral ventricles (LV), superior landmarks of the hippocampus (SH), inferior landmarks of the insular cortex (IC), inferior and lateral landmarks of hippocampus (IH, LH) [4]. The search takes place within certain areas and paths, which all together is called a statistical roadmap. We have used the localization results as an initial state of a 3D deformable model to segment the hippocampus. The steps involved in the proposed method are: 1) anatomical information extraction, and 2) information analysis. In this paper, we present a mathematical framework as well as some simulation results (using models estimated in real circumstances) of the second step when a certain rule-based system is used.

2. METHOD

The localization method requires an expert to define a rough roadmap passing through a set of high-contrast landmarks (milestones), and eventually reaching at the structure of interest. The expert is asked to mark the milestones as desired points and a few points around them as undesired points. We estimate Gaussian models for the marked points and we use them to determine the optimal search area for each desired landmark. The search areas estimated in this step are considered as segments of the statistical roadmap. Details of the first step of the K-BALL are described in our previous publication [4].

In the second step, we use the above statistical roadmap along with: 1) symmetry and 2) absolute statistical models to analyze the extracted information. The symmetry models indicate how similar the landmarks of interest are located relative to the interhemispheric plane. The absolute spatial models provide the distribution of landmarks of interest in a reference coordinates system.

2.1. Definitions

Definition I: An uncertain *roadmap*, $\boldsymbol{G} = (V, D)$, is a directed acyclic graph (DAG). V is a set of nodes, vertices or *milestones*. D is a set of search segments as ordered pairs or edges $(v_i, v_j) \in D$, where $v_i, v_j \in V$.

A vertex or milestone $v_i \in V$ consists of one desired landmark ($L_{i,0}$) and one or several undesired ones ($L_{i,j}, j \neq 0$). Each landmark is spatially characterized by an uncertain (e.g., Gaussian) distribution model except for the initial milestones ($v_i \in V_0$), which are considered deterministic. Each initial milestone consists of only one desired landmark ($L_{i,0}$). The initial points are defined based on a high contrast, robustly identifiable landmark. There is a deterministic relationship between the initial landmarks.

Definition II: A universe, $U = (\Omega^{(i)})_{i \in N_n}$, is a set of domains $\Omega^{(1)}, \ldots, \Omega^{(N_n)}$, each of which corresponds to a milestone of the roadmap $\boldsymbol{G}$ where there are n milestones (excluding the

*This work is partially supported by NIH Grant R01 EB002450.

initial ones). Therefore, in domain $\Omega^{(i)}$, there are one desired landmark, $L_{i,0}$, and n_i undesired landmarks, $L_{i,j}$, where $j \in \{1, \ldots, n_i\}$.

As *a priori* source of knowledge, we define a set of rules $R^{M_1},...,R^{M_m}$, where each of them deals with one or several milestones in ***G***. The modularization set *M* of N_n is a power set of milestones indices that determines what combinations of the milestones are in the rule base. For instance, when there is a rule that involves *i*-th and *j*-th milestones, then there exist an index set $M_k = \{i, j\} \in M$.

Definition III: $R(U, M) = \{R^{M_i} \mid M_i \in M\}$ is a rule base with regard to the universe *U* and the modularization *M*.

The total evidence consists of a set of observations $E^{N_1},...,E^{N_m}$ each of which corresponding to one or several milestones (initial ones excluded) of the ***G***. An evidence may look like $E^{N_k} = \left(f_{v_i,0}(\vec{x}), f_{v_i,1}(\vec{x}), ..., f_{v_i,n_i}(\vec{x})\right)$, which is produced by desired and undesired models of the corresponding milestone, v_i, for the extracted point, $\vec{x}$, where $f_{v_i,j}(.)$ is the distribution function of the *j*-th landmark of the *i*-th milestone. Note that similar to modularization, a partition, *N*, is a power set of milestones that determines the framework in which our observations take place.

Definition IV: $E(U, N) = \left\{E^{N_i} \mid N_i \in N\right\}$ is an evidence system with regard to the universe *U* and the partition *N*.

2.2. Rule Base Design

The uncertainty models of the milestones enable us to evaluate the reliability and accuracy of our observations of the milestones. Our observations are the results of the searches performed during the information extraction (first phase) of the K-BALL.

2.2.1. Rules Structure and Implication Mechanism

In this paper, we only consider bipolar rules that either accept or deny an observation. If an extracted point resides inside the desired iso-contour (an iso-hyper-contour in *n*-D) of its uncertain model, the corresponding rule verifies the observation, i.e., it produces intermediate confidence factor (*ICNF*) = 1. Otherwise, the rule rejects it, i.e., *ICNF* = 0. A weighted linear combination of the *ICNF*'s finally produces a total confidence factor (*CNF*) based on which we decide to accept or reject a set of observations,

$$CNF = \frac{1}{NofR} \sum_{i=1}^{NofR} \beta_i \times ICNF_i \,, \qquad (1)$$

where *NofR* is number of rules.

2.2.2. Rule Categories

We define three categories of rules based on: (*i*) absolute locations of milestones, (*ii*) relative locations with regard to other milestones, and (*iii*) general symmetry of the brain relative to the interhemispheric plane. The subset M_{abs} of the modularization *M* defined as: $M_{abs} = \{\{i\} \mid i \in \aleph \; and \; v_i \notin V_0\}$ determines the absolute rules. For the relative and symmetric rules, we limit ourselves to the rules where only two milestones are involved. M_{rel} defined in (2) provides the framework for the relative rules. When $Path(v_i, v_j) = 1$, there is a path from milestone v_i to v_j and when $Path(v_i, v_j) = 0$, there is no such path. Formula (2) implies that the milestones from different hemisphere of the brain are not included in relative rules. M_{sym} defined in (3) provides the framework for the symmetric rules, where *S* is the set of all milestone pairs of the same structures that are located in different hemispheres of the brain, e.g., hippocampus. The modularization is the union of the above three sets: $M = M_{abs} \bigcup M_{rel} \bigcup M_{sym}$.

2.2.3. Estimated and Derived Rules

The uncertainty models used in K-BALL are twofold: estimated and derived. Therefore, there are two types of rules: the ones designed based on estimated models (*est* rules) and the ones designed based on the derived models (*drv* rules). Let's consider a case in which there exist estimated models of the relative distributions of the consecutive milestones. As one may notice, M_{rel} allows defining relative rules for non-consecutive milestones, too. In order to construct the rules for non-consecutive milestones, we compute the relative uncertainty models between them. Let's consider v_i and v_j in a simple case with $Path(v_i, v_j) = 1$ and only one intermediate milestone, v_k, in between. When the uncertain models are jointly Gaussians, we directly convolve v_i and v_j models to compute the model of v_j from the v_i standpoint, i.e., $\mu_{L_{j,m}}^{v_i} = \mu_{L_{k,0}}^{v_i} + \mu_{L_{j,m}}^{v_k}$ and $C_{L_{j,m}}^{v_i} = C_{L_{k,0}}^{v_i} + C_{L_{j,m}}^{v_k}$, where μ and *C* are mean vector and covariance matrix, respectively. The question of whether or not the derived models would bring any useful information to the rule-based system will be addressed in the following paragraph.

In general the Gaussian density function is $f_{v_i}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |D|}} \exp\left(-\frac{1}{2}\tilde{\mathbf{x}}^T C^{-1} \tilde{\mathbf{x}}\right)$, therefore, the iso-subspace (hyper-ellipse) is: $\left(-\frac{1}{2}\tilde{\mathbf{x}}^T C^{-1} \tilde{\mathbf{x}}\right) = -2\ln\left(\sqrt{(2\pi)^n |C|}\, z_0\right)$, where $|C|$ is the determinant of *C*. In 2D the iso-contour (ellipse) is: $\sigma_y^2 \tilde{x}^2 + \sigma_x^2 \tilde{y}^2 - 2\sigma_{xy} \tilde{x}\, \tilde{y} + 2|C| \ln\left(2\pi\sqrt{|C|}\, z_0\right) = 0$.

Also, note that throughout this paper the iso-contours are set at a level that 90% of the total volume under the distribution resides inside them (except for the **S** that will be discussed later).

$$M_{rel} = \left\{\{i, j\} \mid i, j \in \aleph \; and \; v_i, v_j \in V \; and \; \exists v_q \in V; Path(v_q, v_i) = 1 \wedge Path(v_q, v_j) = 1\right\} \quad (2)$$

$$M_{sym} = \left\{\{i, j\} \mid i, j \in \aleph \; and \; v_i, v_j \in V \; and \; \exists v_q \in V_0; Path(v_q, v_i) = 1 \wedge Path(v_q, v_j) = 1 \; and \; (v_i, v_j) \in S\right\} \quad (3)$$

We intersect the above iso-contour with an arbitrary line $\tilde{y} = m\tilde{x}$ passing through the mean location of the distribution. This will provide us with a sense as how the iso-contour of the derived models $\left(L_{j,m}^{v_i}\right)$ varies as its covariance matrix deviates from its estimated model $\left(L_{j,m}^{v_k}\right)$, incorporating the vagueness of the intermediate model $\left(L_{k,0}^{v_i}\right)$. Let **P** and **Q** and **R** in Fig. 1(a) be the iso-contours of $L_{k,0}^{v_i}$, $L_{j,0}^{v_k}$, $L_{j,0}^{v_i}$, respectively, which are all jointly Gaussian distributions. Computing the following system of equations produces the intersecting points:

$$\begin{cases} \sigma_y^2\tilde{x}^2 + \sigma_x^2\tilde{y}^2 - 2\sigma_{xy}\tilde{x}\,\tilde{y} + 2|C|\ln\left(2\pi\sqrt{|C|}z_0\right) = 0 \\ \tilde{y} = m\tilde{x} \end{cases}$$

where $\tilde{x} = x - \mu_x$, $\tilde{y} = y - \mu_y$. Direct substitution of $\tilde{y}$ in the first equation yields:

$$\tilde{x} = \pm\left(\frac{-2\left|\sigma_x^2\sigma_y^2 - \sigma_{xy}^2\right|\ln\left(2\pi\sqrt{\left|\sigma_x^2\sigma_y^2 - \sigma_{xy}^2\right|}z_0\right)}{\sigma_y^2 + m^2\sigma_x^2 - 2m\sigma_{xy}}\right)^{\frac{1}{2}}$$

when m = tan(θ) and θ = π/2 then $\tilde{x} = 0$ therefore, $\tilde{y} = \pm\left(\frac{-2|C|\ln\left(2\pi\sqrt{|C|}z_0\right)}{\sigma_x^2}\right)^{\frac{1}{2}}$. Using these formulae and changing $0 < \theta < \pi$, we calculate the distance between **P**, **Q**, **R**, etc. The distance between **Q** and **R** is shown in Fig. 1(b) by a dot curve. The distance between **Q** and a contour computed by **P**+**Q** is depicted in Fig. 1(b) as a solid curve. Note that **R** is not even close to **P**+**Q**. Theoretically, the latter contour may not be even an ellipse and in general it cannot be an iso-contour of any Gaussian distribution. With p, a realization of $L_{k,0}^{v_i}$, as shown in Fig. 1(a), we expect the iso-contour **T** to represent **Q**. Consequently, a point like q can be accepted by the corresponding relative consecutive (*est*) rule. On the other hand, the exact same point will be rejected by *drv* rule represented by **R**. Even if we increase the sensitivity of the latter rule; producing the larger iso-contour **S** that encompasses **T**, it may not still be a match for **P**+**Q**. Fig. 1(b) clearly shows the discrepancies between the above two contours; dashdot curve vs. solid curve, respectively. The immediate consequence is that the set of *est* rules do not generate the same results as that of neither *drv* rules nor the combination of two (*est*+*drv*). Therefore, producing rules based on derived models can be justified. The above findings can be formally summarized by the following observations and theorem.

Observation I: $\exists\; p_1 \prec L_{k,0}^{v_i}$, $p_2 \prec L_{j,0}^{v_k}$, and $p_3 \prec L_{j,0}^{v_i}$ such that $f_P(p_1) < 0$ and $f_Q(p_2 \mid p_1) = f_T(p_2) < 0$ but $f_R(p_3) > 0$, depending on the choice of threshold for iso-contour per each model. Where $\prec$ indicates that p is a realization of L, and $f_Z(\cdot)$ is the distribution based on which iso-contour **Z** has been computed.

Observation II: $\exists\; p_1 \prec L_{k,0}^{v_i}$, $p_2 \prec L_{j,0}^{v_k}$, and $p_3 \prec L_{j,0}^{v_i}$ such that $f_P(p_1) < 0$ and $f_Q(p_2 \mid p_1) = f_T(p_2) > 0$ but $f_R(p_3) < 0$, depending on the choice of threshold for iso-contour per each model with $\prec$ and $f_Z(\cdot)$ as before.

This insight leads to the following theorem that constitutes design of the rules using derived models.

Theorem: $R(U, M_{rel}^{drv})$ does not produce identical results as $R(U, M_{rel}^{est})$ for a given observation E^{N_o}. Note that M_{rel}^{est} and M_{rel}^{drv} are the *est* and *drv* subsets of the rule base, respectively.

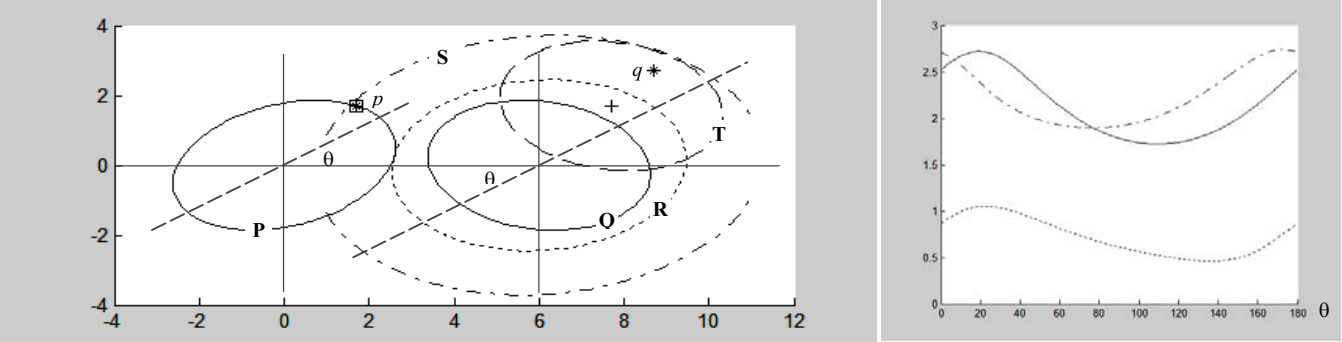

Fig. 1. a) Iso-contour representations of the estimated and derived models. Note that **P**, **R**, and **S** are in reference coordinates system (v_i) while **Q** and **T** are in coordinate system of v_k. b) Distances between **Q** and **R** (dot), **Q** and **P**+**Q** (solid), **Q** and **S** (dashdot).

The above theorem shows that it is possible for the *drv* rules to be useful in the sense that they may support some observations, which are not supported by *est* rules and vice versa. The following algorithm generates the relative subset of the Modularization, M_{rel}.

1. $M'_{rel} = \left\{\{i, j\} \mid i, j \in \aleph \; and \; v_i, v_j \in V \; and \; (v_i, v_j) \in D\right\}$, $M_{rel} = M'_{rel}$, $BfS = \phi$.
2. $\forall (v_i, v_j) \in M'_{rel}$ and $(v_j, v_{j+1}) \in D$, if $\forall L_{j+1,k} \in v_{j+1}, \int_{Iso(L_{j+1,0}) \cap Iso(L_{j+1,k})} f^{v_i}_{j+1,0}(\mathbf{x}) d\mathbf{x} < olth$ (*olth* is an overlap threshold) $\wedge \int_{Iso(L_{j+1,0}) \cap Iso(L_{j+1,k})} f^{v_i}_{j+1,k}(\mathbf{x}) d\mathbf{x} < olth$ then $BfS = BfS \bigcup (v_i, v_{j+1})$.
3. If $BfS = \phi$ terminate the algorithm, otherwise $M_{rel} = M_{rel} \bigcup BfS$, $M'_{rel} = BfS$ and $BfS = \phi$ and go to the second step.

3. SIMULATION AND EXPERIMENTAL RESULTS

We have simulated random vector $\boldsymbol{x}$, using white noise $\boldsymbol{w}$, as shown in block diagram of Fig. 2, by calculating $L = V\Lambda^{1/2}$ where V is the orthogonal matrix comprised of eigenvectors of the covariance matrix of $\boldsymbol{x}$ and Λ is the diagonal matrix of the corresponding eigenvalues, and $\boldsymbol{m_x}$ is the mean value of $\boldsymbol{x}$.

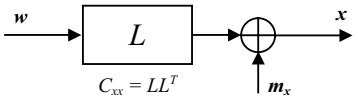

Fig. 2. Generating colored noise $\boldsymbol{x}$ from white noise $\boldsymbol{w}$.

We used the simulation of the relative models estimated for hippocampus localization to optimize the weight vector, $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{\beta}} \left(\sum_{i=1}^{NofS} (CNF_i - GT_i)^2 \right) \quad (4)$$

We generated 3000 patterns out of which 1000 were from desired models and the rest were from undesired models. GT_i is 1 for the patterns generated by only desired models and 0 for the ones generated by undesired models. We have applied (4) to the relative rules of the estimated models both on the right and left side and the result of this experiment was the following weighting vectors, which is valid for both sides using the *est* rules and *est+drv* rules, respectively:

$\boldsymbol{\beta}_{est}$ = [0.3280 0.2307 0.2170 0.0813 0.1430]
$\boldsymbol{\beta}_{drv}$ = [0.2177 0.1186 0.0876 0.0470 0.0706 -0.0216 0.1076 0.3725]

where M^{est}_{rel} = {{iPnt,LV},{LV,SH},{SH,IC},{SH,IH}, {IC,LH}}, M^{drv}_{rel} = {{LV,IC},{LV,IH},{SH,LH}} and iPnt is the roadmap starting point. The rule corresponding to the index set {iPnt,SH} was excluded from M^{drv}_{rel} since it produced a low specification of 0.6 (less than the choice of *olth* = 0.7) in spite of its relatively high sensitivity (0.92). As shown in Fig. 3(left), the specificity of the proposed rule base system is increased when using both *est* and *drv* rules. The sensitivity of the information analysis step is almost the same when switching between *est* rules and *est+drv* rules. Note that this is a preliminary study and we did not implement the algorithm to estimate the optimized set of *est+drv*. Therefore, there is a good chance to get even better results in case we thoroughly follow the proposed theory.

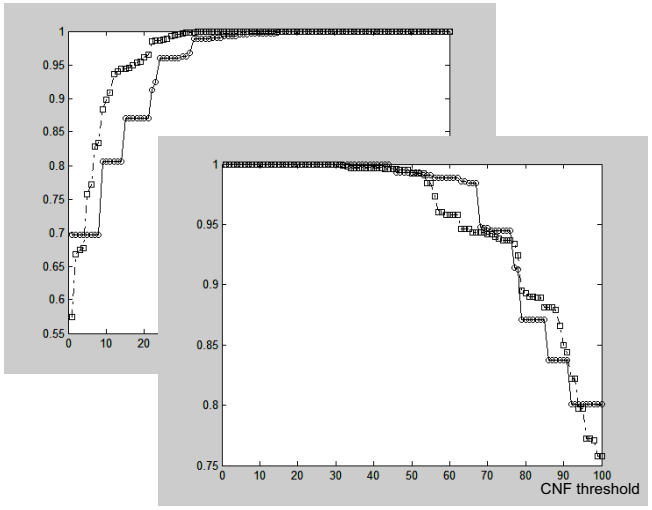

Fig. 3. (left) Specificity, and (right) sensitivity. Curves with squares and circles are produced by *est* and *est+drv* rules, respectively.

We have applied the rule-based method on T1-weighted brain MRI of 10 epileptic patients to localize the hippocampus. The method made no false alarms and the overall success rate of the K_BALL was 83.3%. Two perpendicular views of the MRI data with overlaid localization results are shown in Fig. 4.

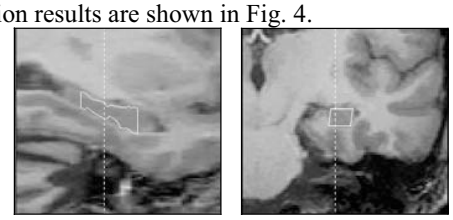

Fig. 4. Sagittal (left) and coronal (right) views of T1-weighted MRI with cross sections of hippocampus initial models overlaid.

4. REFERENCES

[1] L. Verard, P. Allain, J.M. Travere, J. C. Baron, D. Bloyet, "Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis," *IEEE Trans. on Medical Imaging*, vol. 16, no. 5, 1997.

[2] M.-R. Siadat, H. Soltanian-Zadeh, F. Fotouhi, K. Elisevich, "Content-based image database system for epilepsy," *J. of Computer Methods and Programs in Biomedicine*, Elsevier Publisher, in press.

[3] A. Ghanei, H. Soltanian-Zadeh, "A discrete curvature-based deformable surface model with application to segmentation of volume images," *IEEE trans. on Information Technology in Biomedicine*, vol. 6, no. 4, 2002.

[4] M.-R. Siadat, H. Soltanian-Zadeh, F. Fotouhi, K. Elisevich, "Bayesian landmark identification in medical images," *SPIE*, vol. 5370, pp. 628-639, 2004.