Context-Dependent Classification of Multi-Echo MRI Using Bayes Compound Decision Model

Junchul Chun[1], Soo Il Kwon[2]

Purpose : This paper introduces a computationally inexpensive context-dependent classification of multi-echo MRI with Bayes compound decision model. In order to produce accurate region segmentation especially in homogeneous area and along boundaries of the regions, we propose a classification method that uses contextual information of local neighborhood system in the image.

Material and Methods : The performance of the context free classifier over a statistically heterogeneous image can be improved if the local stationary regions in the image are disassociated from each other through the mechanism of the interaction parameters defined at the local neighborhood level. In order to improve the classification accuracy, we use the contextual information which resolves ambiguities in the class assignment of a pattern based on the labels of the neighboring patterns in classifying the image. Since the data immediately surrounding a given pixel is intimately associated with this given pixel, then if the true nature of the surrounding pixel is known this can be used to extract the true nature of the given pixel. The proposed context-dependent compound decision model uses the compound Bayes decision rule with the contextual information. As for the contextual information in the model, the directional transition probabilities estimated from the local neighborhood system are used for the interaction parameters.

Results : The context-dependent classification paradigm with compound Bayesian model for multi-echo MR images is developed. Compared to context free classification which does not consider contextual information, context-dependent classifier show improved classification results especially in homogeneous and along boundaries of regions since contextual information is used during the classification.

Conclusion : We introduce a new paradigm to classify multi-echo MRI using clustering analysis and Bayesian compound decision model to improve the classification results.

Introduction

Many studies on the classification of multi channel M-RI have been introduced to segment region of interest in the image(1,2,3). Conventional context-free medical image classification techniques classify each pixel independently and do not use spatial information between any pair of pixels during the classification procedure. In an image, however, the intensity-level of a pixel is depen-

JKSMRM 3:179-187(1999)

[1]Department of Computer Science, Kyonggi University
[2]Department of Medical Physics, Kyonggi University

Received; May 27,1999, acceepted; June 22, 1999

Address reprint requests to : Junchul Chun, Ph.D., Department of Computer Science, Kyonggi University, Suwon Korea
Tel. 82-331-249-9668/9660, Fax. 82-331-253-1165, Email: jcchun@kuic.kyonggi.ac.kr

dent on those of neighboring pixels unless the image is simply uncorrelated random noise. Therefore, the use of contextual information in classifying image data has long been desired to increase the efficiency and accuracy of the classification(4). The role of context in context-dependent classification is to resolve ambiguities in the class assignment of a pattern based on the labels of the neighboring patterns. Since the data immediately surrounding a given pixel is intimately associated with this given pixel, then if the true nature (estimated class) of the surrounding pixel is known, this can be used to extract the true nature of the given pixel. Welch and Salter used compound decision theory to introduce contextual information into decision scheme and laid the basic background for contextual pattern classification(5). The previous works show the Markov-Gibbs model can be incorporated with Bayes decision rule and contextual information for the image classification(6,7).

In this paper, we present a computationally inexpensive context-dependent Bayes classification with compound decision model. Compared to context free classification, the unsupervised context-dependent classification performs better in homogeneous area and along boundaries as well as images corrupted with noise.

Theory and Method

1) The Aspects of Multi-Echo MR Image Model

Let S be a set of sites over which an MR image is defined. In multi-echo MR images, at each location $s \in S$, the three channel observation (gray level) $ys = (ys^1, ys^2, ys^3)$ can be constructed. Each component of ys, $ys^1 = SD$, $ys^2 = T1$, and $ys^3 = T2$, is called spin density (proton density), spin-lattice relaxation time and spin-spin relaxation time, respectively. The spin density is measured by imposing a strong magnetic field on the object (whose MR image is to be obtained) and then exposing the object radio waves. The radio frequency response from the spinning hydrogen nuclei results in an SD image being formed. The relaxation time T1 is the time required for the hydrogen nuclei to return their equilibrium position in line with the magnetic field. T2 is the time required to dephase the processing magnetic moments of the protons to point when there is no detectable signal. Usually it is not possible to acquire all three of these signatures simultaneously. However, it is possible, using specific sequences of RF pulses, to acquire approximations to some of the signatures from the same slice almost simultaneously. One such imaging methodology is called the Spin-Echo protocol. In our case, the imaging protocol results in images which are predominantly SD and T2.

2) Clustering Analysis Algorithm

The task of clustering analysis is to uncover a reasonable categorization of the data set. Those clustering strategies of importance to us are parametric techniques. If the class conditional densities happen to be Gaussian, these parametric technique is optimal(9,10). In our work, we make an assumption that the clusters in the distributions are approximately normally distributed and a statistical analysis is done to validate goodness of fit to a Gaussian model.

A well-known clustering algorithm is the k-means algorithm, which can be used for estimating the parameters of underlying distributions when the number of classes is given. The main idea of the algorithm is the minimization of a certain criterion functions, which are usually functions of the derivations between all patterns from their respective cluster centers. With an arbitrary chosen initial cluster configuration, the algorithm changes the cluster members iteratively to obtain a better configuration. In general, the sum of squared Euclidean distances is adopted as a respective criterion to the algorithm. Due to its computational simplicity, the algorithm is straightforward to understand. However, the major weakness of the optimization problem is that the k-means algorithm has no guarantee that the iterative process will converge and it may result in a solution which corresponds to a local minimum.

The other approach is the agglomerative algorithm which attempts to find an intrinsic number of classes in the image. With a large number of classes, the algorithm merges them until some minimum number of classes is reached. In this work we employed an agglomerative algorithm in which square neighborhoods of the image are used as initial cluster center. The idea behind the agglomerative clustering algorithm is as follows:- First, the MR image is subdivided into $n \times n$ square for two dimensional images. Each square may be of any size greater than 1×1 but we limit its size to a maximum of 3×3. For each neighborhood, the mean vectors and covariance matrices are computed. Then initial clusters are chosen from spatial neighborhoods which occur in homogeneous regions of the image. In other words, where the intensity variation is less than a specific level is selected. When an arbitrary number of clusters are found, then they are merged to provide a number of final clusters.

The underlying philosophy used here to determine a homogeneous neighborhood is to find areas of low variance which represent homogeneous areas of the image. We opt to use the trace of the covariance matrix, which is the sum of the variance of the sampling neighborhoods for each multivariate image, given by

$$T_{var} = \frac{1}{n^2} = \sum_{l=0}^{1} \sum_{i=x}^{x+n} \sum_{i=y}^{y+n} (g^l_{i,j} - M^l)^2 \qquad [1]$$

where (x,y) is the upper-left location of the n×n neighborhood and M is the mean of the neighborhood. Only those neighborhoods with the variance less than a threshold are accepted as the initial clusters. The next step in the process is to merge the clusters into some meaningful number of clusters which we believe represents the number of classes actually present. The merging process iteratively merges two clusters based on some criteria.

One of the criteria of optimality used to evaluate feature vectors is the Bayes risk. In the real world, however, obtaining the a posteriori distribution is a difficult problem. Thus, finding more practical criteria is desirable, and we consider here two types of such criteria frequently used in practice. One is a family of functions of scatter matrices based on a metric such as Euclidean distance. This criteria is conceptually simple but it does not relate to the Bayes error directly. The other is a family of criteria which give upper bounds of the Bayes error such as Bhattacharyya distance. The criteria are used to measure the class separability between distributions. The Euclidean distance between the means of the class distributions is a simple way to measure class separability. Once the class centers M_i of the class ω_i are calculated, the Euclidean distance between ω_i and ω_j is defined as

$$E_{ij} = \sqrt{(M_i - M_j)^T (M_i - M_j)} \qquad [2]$$

The Euclidean distance is easily computable and understandable. However, a large distance between the means does not simply mean that two distributions do not overlap because the covariance between the class is not used in the calculation.

The Bhattacharyya distance uses both the class means and the covariances between classes to measure class separability. The Bhattacharrya distance between two normal distributions is written by

$$B_{ij} = \frac{1}{8} = (M_i - M_j)^T \left(\frac{\Sigma_i + \Sigma_j}{8}\right)^{-1} + \frac{1}{2} \text{ in } \frac{\frac{\Sigma_i + \Sigma_j}{8}}{\sqrt{|\Sigma_i||\Sigma_j|}} \qquad [3]$$

where Σ_i and Σ_j are covariance matrices of ω_i and ω_j, respectively.

It is important to note that the above distance function uses both distribution parameters (mean, covariance) in its calculation. The equation shows that the first term of the Bhattacharyya distance is related to the class separability due to a difference in means, while the second term is related to the class separability due to difference between the covariance matrices.

Then the estimation of the mean vectors (M_m) and covariance matrices (Σ_m) for the merged clusters is required and expressed by the following formulas provided the clusters being merged are Gaussian(8):-

$$M_m = \frac{n_p M_p + n_q M_q}{n_p + n_q}$$

$$\Sigma_m = \frac{n_p(\Sigma_p + M_p M_p^T) + n_q(\Sigma_q + M_q M_q^T)}{n_p + n_q} - M_m M_m^T, \qquad [4]$$

where Σ_p and Σ_q are the predecessor covariance matrices. The clustering process is continued until some predetermined number of classes are found or the smallest distance between clusters is greater than a threshold value.

3) Statistical Analysis of Clustering Model

Parametric classification including Bayes' s method are optimal if the class conditional densities are Gaussian. Therefore, statistical verification of normality of the given data set to justify deploying Bayes' s decision rule in the image classification is necessary. The Chi-square (x^2) test, which is used to test whether there are significant variation in the probability function, is sufficient when the dimensionality of the data set is low. However, the number of cells increases exponentially with the dimensionality and thus the x^2 test is impractical for a high dimensional data set. One technique adapted to high dimensional data set is the Kolmogorov-Smirnov(KS) test(10). Based on available samples, a statistic is computed which may be used to accept or reject the Gaussian hypothesis. The KS statistic, which tests whether there is a significant difference in the Cumulative Distribution Function, is desirable for our case.

Given a distribution of X with the sample mean vector M and the sample covariance matrix Σ, the empirical distribution of X is expresses by the following quadric form

$$\zeta = \frac{1}{n-1} (X-M)^{T} \Sigma^{-1} (X-M). \quad [5]$$

The density function of ζ has the form of gamma distribution, $\Gamma(a, \lambda)$. Specifically, the density function ζ has the form of gamma distribution $\Gamma(n/2,2)$ where n is the dimension of the given data set provided M and Σ are known. The test can be extended when M and Σ are not known, but estimated by the sample mean and sample covariance. When X is normal, ζ, becomes beta distribution, which is the distribution on [0,1] parameterized by the positive parameters p and q, is represented by

$$B(\zeta) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} \zeta^{p-1}(1-\zeta)^{q-1} \quad 0 \leq \zeta \leq 1 \quad [6]$$

where $p = \frac{n}{2}$, $q = \frac{N-n-1}{2}$.

For the Kolmogorov-Smirnov test the computed distribution of ζ and the theoretical distribution are used as input values. The cumulative distributions of both the theoretical and empirical distribution is used for the test. The KS statistic is defined as follow:

Definition : Let x_i, $i = 1 \cdots, N$ be the sample of a cumulative distribution(CDF) $S_N(x)$ and let $F^*(x)$ be the empirical CDF. Then the Kolmogorov-Smirnov statistic(d) is

$$d = \max| S_N(x) - F^*(x) |. \quad [7]$$

The Kolmogorov-Smirnov statistic d is the maximum difference between the two cumulative distribution functions. Under the significance level of probability for the null hypothesis that the data sets are drawn from the same distribution, the KS statistic is useful for determining the equivalence of two distribution.

4) Bayesian Compound Model for Context-Dependent Classification

Parametric classification refers to the development of statically defined discriminant functions in which the underlying probability functions are assumed known. Assume that homogeneous region of the images is Gaussian, and it is defined in the N-dimensional image histogram spaces; then the distribution of a variable x is describes by its mean μ and its covariance ϕ. According to Bayes's rule:

$$P(S_i|x) = \frac{P(S_i|x)P(S_i)}{P(x)} \quad [8]$$

If probability density functions of the classes, S_i, are assumed to be normal, then the N variate normal distribution becomes:

$$P(x|S_i) = \frac{1}{(2\pi)^{N/2}|\phi_i|^{1/2}} \exp\{-\frac{1}{2}(x-\mu_i)^{t}[\phi_i]^{-1}(x-\mu_i)\} \quad [9]$$

where:

μ_i is the mean vector of class i

ϕ_i is the covariance matrix of class i

The context free classifier assign x to class S_i if $p(x \mid S_i) \geq p(x \mid S_j)$ for all $i \neq j$. In practices, in order to avoid exponential calculation during the actual classification we evaluate $\ln P(S_i \mid x)$.

Suppose the statistical decision problem is repeated n times and there is no relationship among the n repetitions; then the n decision problems having identical generic structure constitute a compound decision problem(3). Given a set $\Omega = \{\omega_1, \cdots, \omega_n\}$ of states of nature and a corresponding set $G = (g_1, \cdots, g_n)$ of vector-valued random variables, the compound decision rule will minimize the compound Bayes loss

$$\sum_{\omega_s=1}^{C} L(\omega, \omega^*)\, P(g_n \mid \omega)\, P(\omega) \quad [10]$$

where G is the set of pattern vectors for all pixels in the image frame and C is the number of classes. The sig-

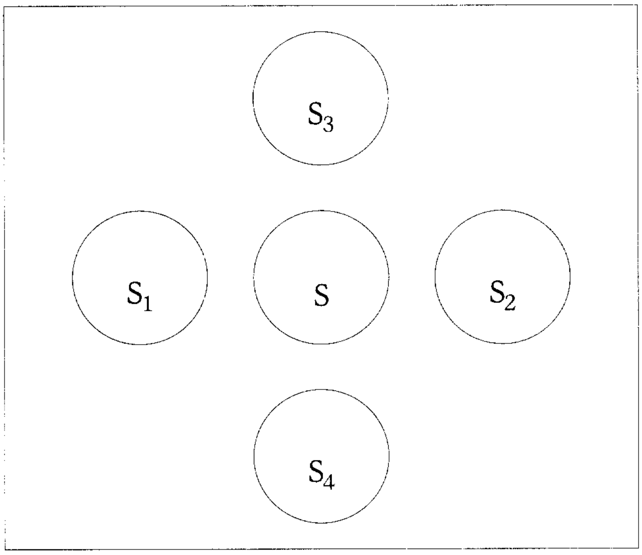

Fig. 1. Labeling of neighbors of pixel *s*

nificant difference between the compound decision rule and the simple decision rule is that the first uses all the pattern vectors to estimate the nature ω_s, while the second uses the vector from the pixel s only.

In practical situations, the direct calculation of the compound Bayes loss is impossible. For example, suppose the size of an image frame is 256×256 so that the G is a 256^2 dimensional vector of vectors. Thus, the density function $P(G \mid \omega_s)$ defined on this vector space cannot be evaluated with any ease since the dimensionality of the configuration space is enormous. For the evaluation of density function some simplified assumptions which are based on the observation that two adjacent pixels in the image are unconditionally correlated and the degree of correlation depends on the distance between pixels is used. This observation provides us with the neighborhood system, where only the neighboring

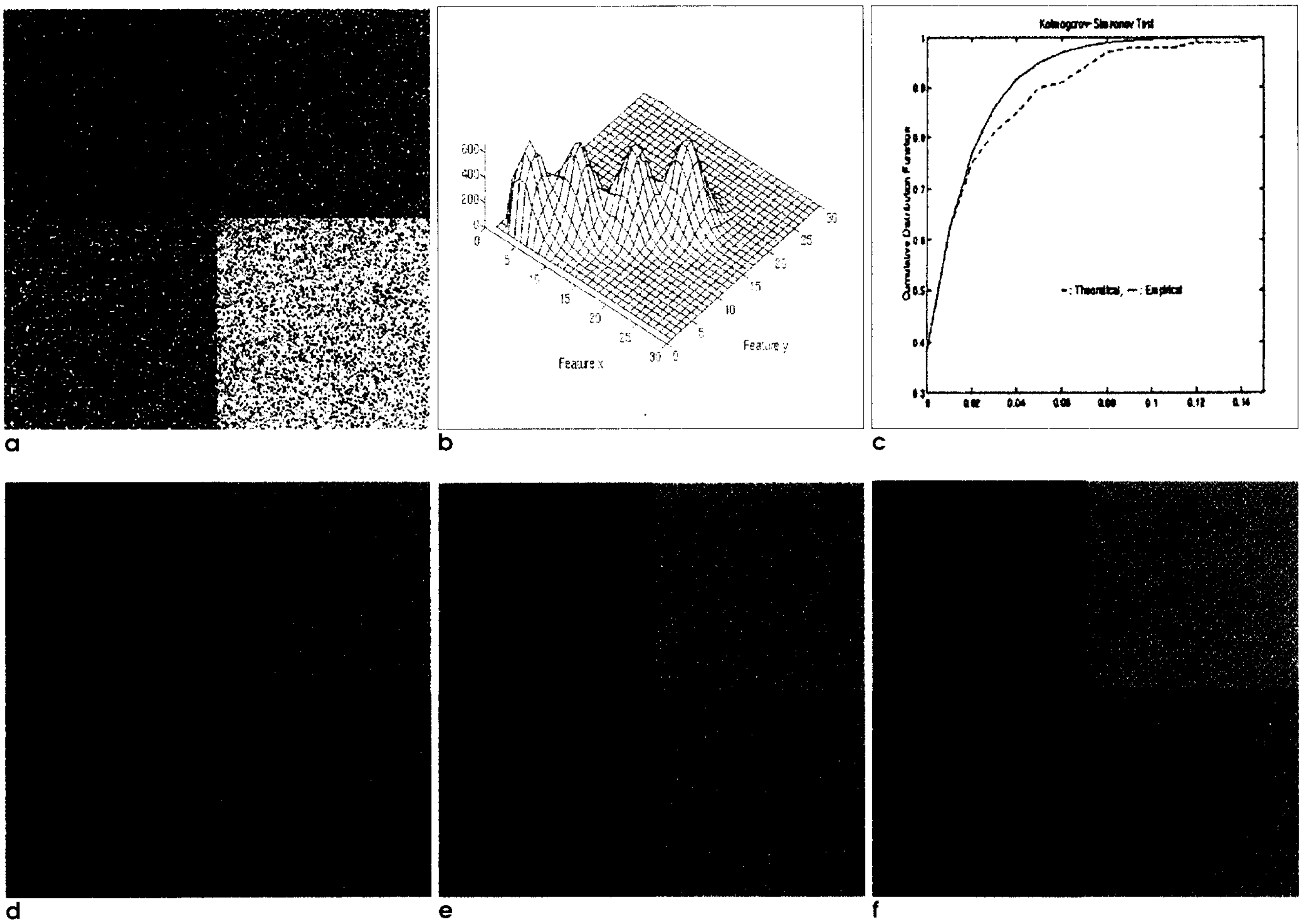

Fig. 2. (**a**) Bivariate test image (**b**) 3D histogram of distributions in test image (**c**) KS test for the distribution of quadrant 1 (**d**) Result of minimum distance classification (**e**) Result of maximum likelihood classification (**f**) Result of Bayes compound classification

pattern vectors of a certain pixel s are thought to contain information about the unknown nature of ω_s. The first assumption is that the contextual relations between non-adjacent cells are negligible while defining adjacent cells to be any two pixels which are physically adjacent. Therefore, in a 4-pixel neighborhood system(Fig. 1) the pixels s_i, $(i=1,2,3,4)$ interact directly with only with center pixel s_i not with each other. The second assumption is that the pattern g_s of pixel ω_s is a function of only of ω_s, the true class of the pixel.

Based on such assumption, the context-dependent compound decision model uses the compound Bayes decision rule with the contextual information. As for the contextual information in the model, the directional transition probabilities are used. The decision rule for a four neighborhood system becomes $g_s \in \omega_k$ if the following conditions are satisfied for all $k \neq l$

$$P(g_s \mid \omega_k)\, P(\omega_k) \prod_{i=1}^{4} \sum_{j=1}^{C} P(g_{s_i} \mid \omega_j)\, p(\omega_j \mid \omega_k) > P(g_s \mid \omega_l)\, P(\omega_l) \prod_{i=1}^{4} \sum_{j=1}^{C} P(g_{s_i} \mid \omega_j)\, p(\omega_j \mid \omega_l). \qquad [11]$$

The four multipliers in the product term represent the contextual distributions from the four adjacent neighbor pixels. The estimation of the transition probability $P(\omega_{s_i}=c_j \mid \omega_s=c_i)$ would be $n_{c_ic_j}/n_{c_i}$ where n_{c_i} is the number of occurrences of class c_i and $n_{c_ic_j}$ is the number of occurrence of ω_{s_i} having class c_j when ω_s has class c_i. The compound decision model provides contextual information in the form of directional probabilities during the classification. However, usually the true classes of neighbors are not available; only the neighboring patterns are available. Thus the step of classifying the neighbors is necessary before classification can proceed.

Experimental Results

Classifications for the synthetic image and actual MRI have been done for the evaluation of the proposed ap-

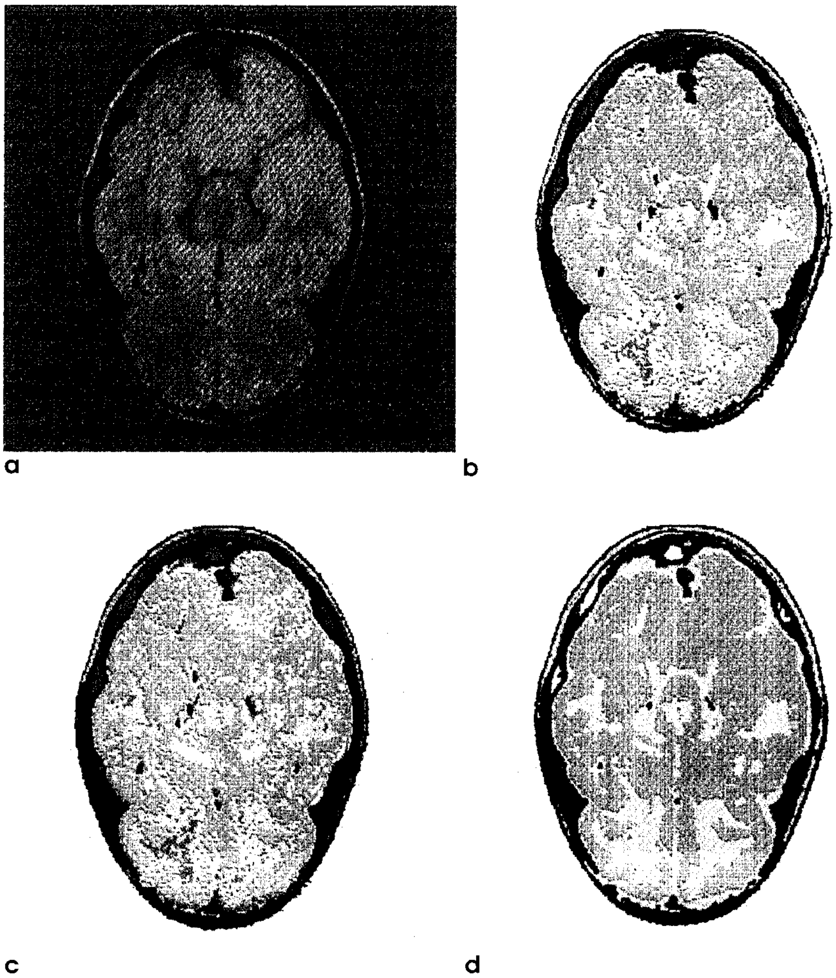

Fig. 3. (**a**) Sample MR image (**b**) Result of minimum distance classification (**c**) Result of maximum likelihood classification (**d**) Result of Bayes compound classification

proach. In image classification, it is often necessary to generate a simulated data set before applying the proposed classification algorithm to the multi-echo MR images. In our case here, the sample images to be tested should be normally distributed samples according to given expected mean vectors(M) and covariance matrices(Σ). Therefore, given means $\Sigma_1 = (60,50)$, $M_2 = (130,120)$, $M_3 = (220, 190)$, $M_4 = (290, 270)$ and the covariance $\sigma_{11} = \sigma_{22} = \sqrt{1000}$, a bivariate synthetic image, which is composed of four distributions, is generated (Fig. 2(a)). The natural clustering and 3 dimensional histogram of the synthetic image is composed of four distributions with overlaps as shown in Fig. 2 (b). The next step we need is to test the normality of each distribution of the image because Bayes theory is optimal for the normally distributed sample data. The Kolmogrov-Simirnov statistic is used to test the hypothesis that the pixel populations from which the individual clusters are derived are normally distributed in the bivariate synthetic image parameter space. For the test, 100 random samples are collected from quadrant 1 and the theoretical and empirical distributions are evaluated. Fig. 2(c) explains that the sample distribution is normally distributed. As a result, the class map of the synthetic image (Fig. 2(f)) using context-dependent classification based on Bayes compound model shows much clear result rather than using context-free classification (Fig. 2(d)-(e)) especially in homogeneous region and regions along the boundaries.

To actual muti-echo MR images, each classification method is applied. The mid brain images (Fig. 3(a), Fig. 4(a)) which are composed of SD and T2 image are used for the test. Both the minimum distance classifier and maximum likelihood classifier are adopted for context-free classification. The results of minimum distance classification and maximum likelihood classification are illustrated in Fig. 3(b),4(b) and Fig. 3(b),4(c), respectively. Classification results(Fig. 3(d), 4(d)) using context-de-

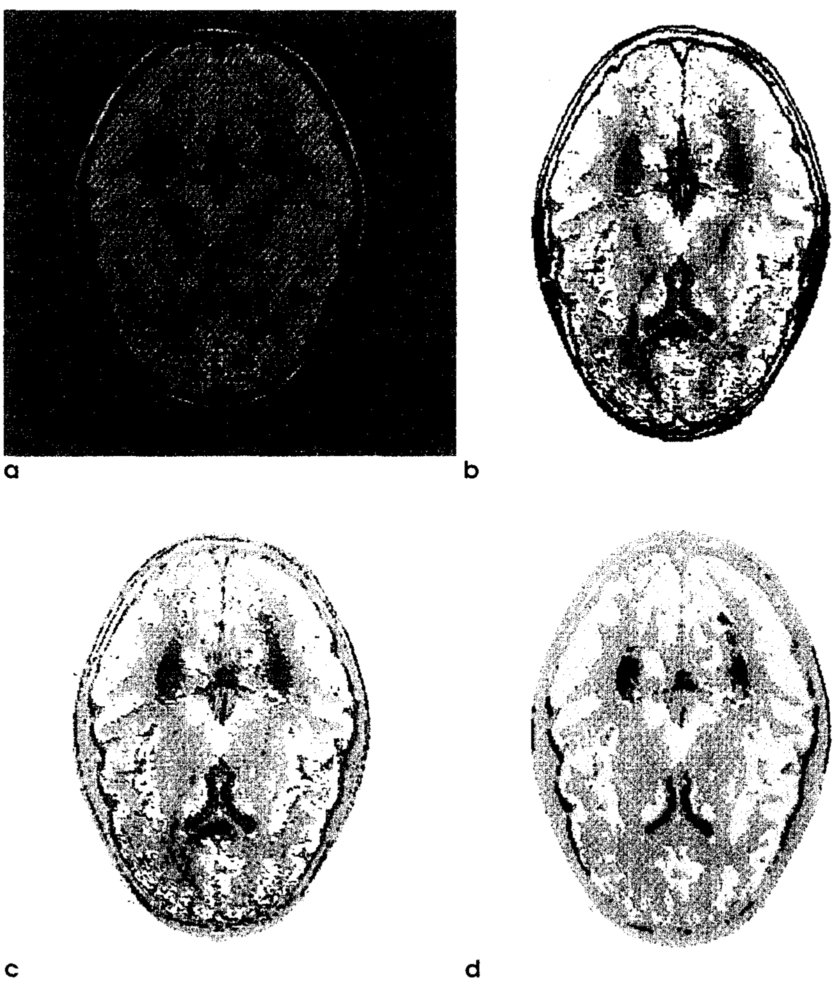

Fig. 4. (**a**) Sample MR image (**b**) Result of minimum distance classification (**c**) Result of maximum likelihood classification (**d**) Result of Bayes compound classification

pendent compound Bayes classification show better results rather than the results using context-free classification by reducing numbers of misclassifications of each class.

Conclusion

This paper introduces a context-dependent classifier using Bayes compound decision model for multisignature magnetic resonance images. For this purpose, we first use clustering analysis for the target images to analyze the distributions of the image. From the analysis of the image, we extract the parameters of distributions, i.e., means and covariance vectors of each class which are basically used for context-free classification. To increase the performance of the classifier over statistically heterogeneous image, we adopt Bayes compound decision model for context-dependent classification. The significant difference between the compound decision rule and simple decision rule is that the compound decision rule uses all the pattern vectors to estimate configuration, while the simple rule uses the vector from the pixel itself. However, usually the true class of neighbors are not available and only the neighboring patterns are available. Thus, the contextual information can be achieved from the context-free classification. The compound decision model provides contextual information in the form of directional transitional probability during the classification. The strength of correlation among the pixels in the neighborhood system is explained by the directional transitional probability.

As illustrated in the experimental results, the context-dependent classification using Bayes compound decision model provides better performance in the classification rather then the results using only context free classification. Especially, the numbers of misclassification in the homogeneous region and along the boundaries are reduced when the context-dependent classification is applied.

References

1. P. Santago and H. D. Gage. Quantification of MR Brain Images by Mixture Density and Partial Volume Modeling. IEEE Trans. on Medical Imaging, 1993; 12: 566-574.
2. H. S. Choi, D. R. Haynor, and Y. Kim. Partial Volume Tissue Classification of Multichannel Magnetic Resonance Images-A Mixel Model. IEEE Trans. on Medical Imaging, 1991; 10: 395-407.
3. T. Taxt and A. Lundervold. Multispectral Analysis of the Brain Using Magnetic Resonance Imaging. IEEE Trans. on Medical Imaging, 1994; 13: 470-481.
4. K. S. Fu and T. S. Yu, Statical Pattern Classification using Contextual information, Research Studies Press, New York, NY, 1980.
5. J. R. Welch and K. G. Salter, A Context Algorithm for Pattern Recognition and Image Interpretation, IEEE Trans. on Systems, amn, and Cybernetics, 1971; 1: 24-30.
6. J. C Chun and S. I Kwon, Classification of a Volumetric MRI Using Gibbs Distributions and a Line Model, Journal of the Korean Society of Magnetic Resonance in Medicine, 1998; 2: 58-66.
7. Geman, S. and D. Geman, Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images, IEEE Trans. on Pattern Anal. Mac. Intel, 1984; 6: 721-741.
8. J.B. Perkins, I. R. Greenshields, F. Dimario, and G. Rambsy, Unsupervised Classification of Multi-Echo Magnetic Resonance Images of Pediatric Brain with Implicit Spatial and Statical Hypothesis Validation, in Proc, SPIE, Medical Imaging, 1993; 1692.
9. T. N. Pappas, An adaptive clustering algorithm for image segmentation, IEEE Trans. Signal Process., 1992: 40, 901-914.
10. K Fukunaga, Introduction to Statical Pattern Recognition, Academic Press, New York, NY 1990.

대한자기공명의과학회지 3:179-187(1999)

Bayes의 복합 의사결정모델을 이용한 다중에코 자기공명영상의 context-dependent 분류

[1]경기대학교 정보과학부 영상처리연구실
[2]경기대학교 의학물리학과

전 준 철[1] · 권 수 일[2]

목적 : 본 논문은 Bayes의 복합 의사결정모델을 이용한 효과적인 다중에코 자기공명영상의 분류방법을 소개한다. 동질성을 갖는 영역 혹은 경계선부위 등 영역을 명확히 분할하기 위하여 영상 내 국소 부위 이웃시스템상의 주변정보(contextual information)를 이용한 분류 방법을 제시한다.

대상 및 방법 : 통계학적으로 이질적 성분들로 구성된 영상을 대상으로 한 주변정보를 이용한 분류 결과는 영상내의 국소적으로 정적인 영역들을 이웃화소 시스템 내에서 정의되는 상호작용 인자의 메커니즘에 의해 분리함으로서 개선시킬 수 있다. 영상의 분류과정에서 분류결과의 정확도를 향상시키기 위하여 분류대상 화소의 주변화소에 대한 분류 패턴을 이용한다면 일반적으로 발생하는 분류의 모호성을 제거한다. 그러한 이유는 특정 화소와 인접한 주변의 데이터는 본질적으로 특정 화소와 상관관계를 내재하고 있으며, 만일 주변데이터의 특성을 파악할 수 있다면, 대상화소의 성질을 결정하는데 도움을 얻을 수 있다. 본 논문에서는 분류 대상화소의 주변정보와 Bayes의 복합 의사결정모델을 이용한 context-dependent 분류 방법을 제시한다. 이 모델에서 주변 정보는 국소 부위 이웃시스템으로부터 전이확률(transition probability)을 추출하여 화소간의 상관관계의 강도를 결정하는 상호인자 값으로 사용한다.

결과 : 본 논문에서는 다중에코 자기공명영상의 분류를 위하여 Bayes의 복합 의사결정모델을 이용한 분류 방법을 제안하였다. 주변 데이터를 고려하지 않는 context-free 분류 방법에 비하여 특히 동질성을 갖는 영역 혹은 경계선 부위 등에서의 분류 결과가 우수하게 나타났으며, 이는 주변정보를 이용한 결과이다.

결론 : 본 논문에서는 클러스터링 분석과 복합 의사결정 Bayes 모델을 이용하여 다중에코 자기공명영상의 분류 결과를 향상시키기 위한 새로운 방법을 소개하였다.

통신저자 : 전준철, 경기도 수원시 팔달구 이의동 산 94-6, 경기대학교 정보과학부 영상처리연구실
Tel. 82-331-249-9668/9660 Fax. 82-331-253-1165 Email: jcchun@kuic.kyonggi.ac.kr