DEPARTMENT OF STATISTICS
University of Wisconsin
Medical Science Center
1300 University Ave.
Madison, WI 53706

TECHNICAL REPORT NO. 1113

November 18, 2005

Parametrization and Classifiication of Closed Anatomical Curves

Shubing Wang[1] and Moo K. Chung[1,2]

[1] Department of Statistics
[1] Department of Biostatistics and Medical Informatics
1300 University Ave.
Madison, WI 53706

ABSTRACT

We present a streamlined mathematical framework for modeling and classifying closed curves from medical images. The method of gradient vector flow (GVF) snakes is used to extract object boundaries in the magnetic resonance images of the human brain. The curvature based arc-length parametrization is used to parameterize the extracted boundary. A linear curve registration is then performed to align the orientation of the curves via the least-squares method. Afterwards, a finite Fourier series expansion is used to get a smooth functional representation of the parametrization. Various classification and regression tree techniques are applied to determine the best classification techniques in separating the corpus callosum (CC) boundary curves obtained from 12 autistic and 15 normal control subjects. The goal of this paper is to develop a methodological framework for classifying CC curves into different clinical groups, and compare the performance of different classification methods.

1. INTRODUCTION

The modern magnetic resonance imaging (MRI) technique provides extremely detailed anatomical information about the corpus callosum (CC) shape of the human brain in 1mm scale. Although various methods such as Talairach space normalization [3], factor analysis [14] and voxel-based morphometry [2] are proposed for modeling and quantifying the CC shape, there has not been an attempt to classify the corpus callosum shape itself. In this paper, we present a streamlined image processing and analysis framework for segmentation, parametrization, and classification of the CC curves.

The CC boundary curves were extracted using snakes [8, 17] in the midsagittal section of MRIs. Traditional snakes have problems associated with small capture range and poor convergence to boundary concavity. To address these problems, new external force for snakes called the *gradient vector flow* (GVF) fields was introduced [17]. The GVF snakes have advantages over the traditional snakes with their larger capture range and ability to move into the concavity. Hence, we used the GVF snakes in extracting the CC boundaries that have many concave regions (Figure 1).

The extracted CC boundaries were parameterized by arc-length, which is an often used parametrization in computer vision [6, 12]. A new algorithm for computing the arc-length parametrization using curvature is provided. This algorithm reduces the estimation error in computing the arc-length from finite points. Afterwards, the parameterized CC curves were aligned linearly by minimizing the sum of squared errors between two curves. The average shape of the curves are then calculated by just averaging the curve over the corresponding parameters.

As a way to reduce the dimension of the CC curves, the finite Fourier expansion is used to represent the coordinate functions. Then the coefficients of the Fourier expansion is used as predictors in classifying the CC curves into two groups: normal controls and autism. Although the spherical harmonic (SPHARM) representation of the coordinates of 2D anatomical surfaces such as the hippocampus has been proposed [4, 16], the Fourier representation of the 1D CC curves has not been proposed previously.

Classical methods of curve and image classification include functional data analysis [7], support vector machines (SVM) [15, 16] and mixture model and EM algorithm [13]. In this study, we mainly applied regression tree based classification techniques such as QUEST [11], CRUISE [9], GUIDE [10], and to determine if it is possible to differentiate autism purely based on the shape of CC curves. These regression tree based methods have been widely used in statistical literatures since there is no implicit statistical assumption about predictor and response variables [1]; however, they have not been used in classifying curves or images.

2. CLOSED CURVE MODELING

2.1. GVF snakes

A traditional active snake is a 2D curve $\boldsymbol{x}(t) = (x(t), y(t))^\tau$ ($t \in [0, 1]$) that moves within an image and converges to the desired boundary and minimizes the energy functional

$$E = \int_0^1 \frac{1}{2}(\alpha|\boldsymbol{x}'(t)|^2 + \beta|\boldsymbol{x}''(t)|^2) + E_{\text{ext}}(\boldsymbol{x}(t))dt,$$

where α and β are weight parameters that control the snake's tension and rigidity, E_{ext} defines the external energy. For vector norm $|\cdot|$, we used the length of the vector. Given image intensity function $I(x, y)$, the typical external energy is given as

$$E_{\text{ext}} = -|\nabla(G_\sigma * I(x, y))|^2, \quad (1)$$

where ∇ is the gradient operator and G_σ is a 2D Gaussian kernel with standard deviation σ. A snake that minimizes E must satisfy a force balance equation:

$$\mathbf{F}_{\text{int}} + \mathbf{F}_{\text{ext}} = 0,$$

where $\mathbf{F}_{\text{int}} = \alpha\boldsymbol{x}''(t) - \beta\boldsymbol{x}''''(t)$ and $\mathbf{F}_{\text{ext}} = -\nabla E_{\text{ext}}$. The GVF snake introduces a new external force $\mathbf{F}_{\text{ext}} = (u, v)^\tau$, which is the solution to the Euler equation

$$\begin{cases} \mu\Delta^2 u - (u - f_x)(f_x^2 + f_y^2) = 0 \\ \mu\Delta^2 v - (v - f_y)(f_x^2 + f_y^2) = 0 \end{cases} \quad (2)$$

where μ is the regularization parameter and $f(x, y)$ is the edge map [17]. The major advantages of the GVF snakes are their large capture range and their capability to converge to the concavities of boundaries. Based on this new external force, we minimized the energy functional and obtained the CC boundaries (Figure 1).

Fig. 1. Representative results of the exacted CC curves using the GVF snakes.

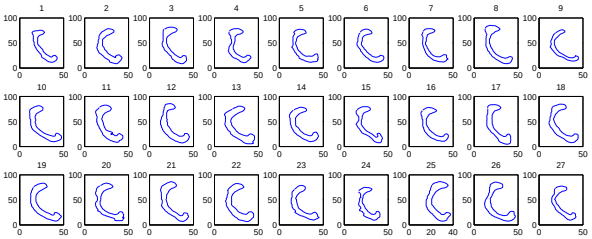

Fig. 2. 27 corpus callosum boundary curves used in the study.

2.2. Curvature based arc-length parametrization

For given points $\{p_i = (x_i, y_i)^\tau\}_{i=0}^n$ ($p_n = p_0$) that form a closed CC curve, our goal is to find an arc-length parametrization $\gamma(s) = (x(s), y(s))^\tau, s \in [0, L]$ such that

$$\gamma(s_i) = (x_i, y_i)^\tau, \quad i = 0, 1, \cdots, n, \tag{3}$$

where L is the total arc length of the CC curve. The curvature function is defined as

$$k(s) = |\gamma'(s)| = \lim_{\Delta s \to 0} |\gamma(s + \Delta s) - \gamma(s)|/\Delta s$$

which implies $\Delta s \approx |\gamma(s + \Delta s) - \gamma(s)|/k(s)$ for small Δs. For convenience, we denote $p_{-1} = p_{n-1}$ and $p_{n+1} = p_1$. s_i's in (3) is calculated by

$$s_0 = 0, \; s_{i+1} = s_i + \frac{|p_{i+1} - p_i|}{k(s_i)},$$

where $k(s_i)$ is calculated by the following formula [5]

$$k(s_i) = d_i \cdot \frac{4A(p_{i-1}, p_i, p_{i+1})}{|p_{i-1} - p_i| \cdot |p_i - p_{i+1}|} \tag{4}$$

where $A(p_{i-1}, p_i, p_{i+1})$ is area of the triangle with vertices p_{i-1}, p_i and p_{i+1}, and $d_i = 1$, if the triangle is inside the curve; -1 otherwise. Then $k(s)$ for arbitrary s is estimated by linear interpolation

$$k(s) \quad = \tfrac{s_{i+1}-s}{s_{i+1}-s_i} k(s_i) + \tfrac{s-s_i}{s_{i+1}-s_i} k(s_{i+1}), \quad s \in [s_i, s_{i+1}].$$

Since each CC curve has different arc-length, we need further reparameterize γ by mapping the parameter space $[0, L]$ onto a unit circle. Consider a linear mapping $\theta = 2\pi \frac{s}{L}$ which maps $[0, L]$ to $[0, 2\pi]$. Then our final parametrization is given by $\varrho(\theta) = \gamma(\frac{\theta}{2\pi} L)$.

2.3. Curve alignment

After the parametrization, a linear curve alignment is necessary to factor out the orientation and translational difference. We do not need to factor out the size difference since our parametrization basically maps a CC curve onto the unit circle effectively normalizing the global size.

Suppose we have parameterized curves ϱ_1 and ϱ_2 using the previous section. For n control points for each of curves ϱ_1 and ϱ_2,

$$\hat{\varrho}_1 = \{\varrho_1(\frac{2\pi i}{n})\}_{i=0}^n, \quad \hat{\varrho}_2 = \{\varrho_2(\frac{2\pi i}{n})\}_{i=0}^n,$$

the corresponding coordinates in the Cartesian coordinate system is given by $(x_{1,i}, y_{1,i})$ and $(x_{2,i}, y_{2,i})$. Then we define the warping map $A : L^2([0, 2\pi]) \to L^2([0, 2\pi])$:

$$A_{(\theta,a,b,t)}((x_{1,i}, y_{1,i})) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_{1,i+t} - a \\ y_{1,i+t} - b \end{pmatrix}$$

with convention $x_{n+k} = x_k, y_{n+k} = y_k$. Then the linear warping between $\hat{\varrho}_1$ and $\hat{\varrho}_2$ is given by minimizing the discrepancy

$$A_{(\theta^*,a^*,b^*,t^*)} = \arg \min_{(\theta,a,b,t)} \sum_{i=1}^n \|A_{(\theta,a,b,t)}(\hat{\varrho}_1) - \hat{\varrho}_2\|_2$$

where $\|\cdot\|_2$ is the l^2-norm.

Figure 2 shows the all 27 CC curves that were linearly aligned. 15 curves are from the control group and the remaining 12 are from the high functioning autistic group. Figure 3 shows the averages shapes of the two groups after the curve alignment.

2.4. Fourier representation

The natural basis on a unit circle is given by

$$f_{2n-1} = \cos(nx), \quad f_{2n} = \sin(nx), \quad n = 1, 2, \cdots.$$

with a constant term $f_0 = 1/\sqrt{2}$. Then $\{f_i\}_{i=0}^\infty$ forms a complete orthonormal basis set in Hilbert space $L^2([0, 2\pi])$, the space of square integrable functions on the unit circle. The inner product is defined as $\langle f, g\rangle = \frac{1}{\pi}\int_0^{2\pi} f(x)g(x)dx$. Since $L^2([0, 2\pi])$ is separable, for $u \in L^2([0, 2\pi])$, we have the Fourier expansion

$$u = \sum_{i=0}^\infty \langle u, f_i\rangle f_i.$$

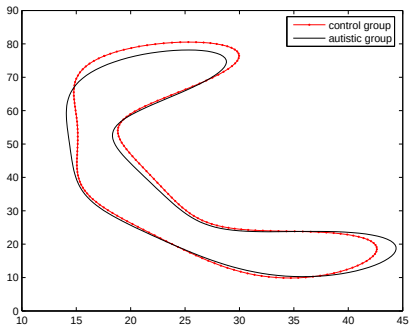

Fig. 3. The average shapes of control and autistic groups. The red curve is the control subject and the black curve is the autistic subjects.

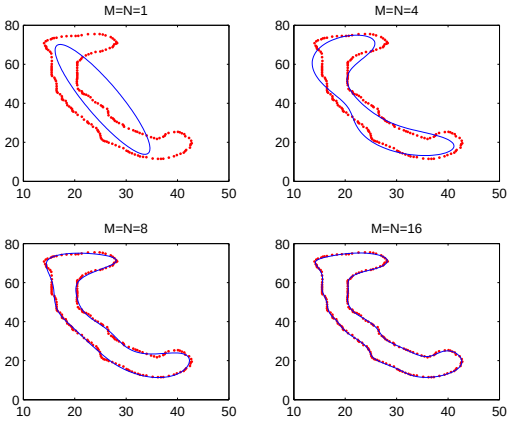

Fig. 4. The finite Fourier expansions of closed curve with different (M, N)'s.

Based on the convergence of Fourier expansion, we approximate the CC curve $\varrho(s) = (x(s), y(s))$ by

$$\varrho \approx (\sum_{i=0}^{M} \langle x, f_i \rangle f_i, \sum_{j=0}^{N} \langle y, f_j \rangle f_j).$$

Figure 4 shows the Fourier series representation with increasing degrees of M and N. The Fourier coefficients $\langle x, f_i \rangle, \langle y, f_j \rangle$ fully characterize the shape variation of the CC curves and will be used as the predictors for classification in Section 3.

3. CLASSIFICATION OF CURVES

First, we applied different classification techniques to simulated curves. We generated two closed curves ϱ_1 and ϱ_2 (Figure 5). Two shape of two curves are different in the par to the bottom arc. 100 curves with Gaussian white noise $N(\mu, \sigma^2)$ are generated from ϱ_1 and another 100 curves with $N(\mu, \sigma^2)$

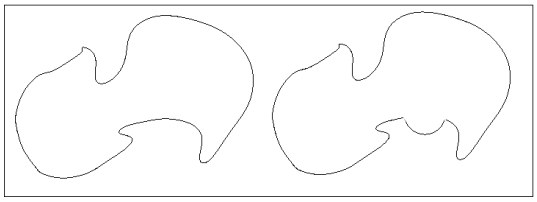

Fig. 5. Curve ϱ_1 (left) is generated by a smooth spline while curve ϱ_2 (right) is generated from ϱ_1 by adding a small convex part.

Methods \ $D(\sigma)$	0.1	0.2	0.4	0.8	1.6
LDA	0	0	0	0.08	0.25
CRUISE	0	0	0	0.06	0.34
GUIDE	0	0	0	0.11	0.40
QUEST	0	0.02	0.06	0.11	0.37

Table 1. The classification results of simulated data. Each entry is the misclassification rate for a particular classification algorithm at different $D(\sigma)$ level.

noise are generated from ϱ_2. We computed the amount of noise with respect to signal using measure

$$D(\sigma) = \frac{\sigma}{\|\varrho_1 - \varrho_2\|_\infty},$$

where $\|\cdot\|$ is L^∞-norm. $D(\sigma)$ represent the relative amount of noise difference between two curves.

After finite Fourier expansion is calculated for each curve, we use half of curves in each group as the training set and the other half as the test set. The classical methods applied to the simulated curves are linear discriminant analysis (LDA). We have also applied the regression tree techniques including QUEST, which is a fast, unbiased binary decision tree method [11], CRUISE, which stands for Classification Rule with Unbiased Interaction Selection and Estimation [9], GUIDE, a generalized unbiased regression tree algorithm [10]. The classification results for the simulated curves are shown in Table 1. As expected, the difficulty of classification increased with increasing $D(\sigma)$ as shown in Table 1. LDA performed best since LDA assumes Gaussian noise which is exactly the case for our simulated data.

For the actual CC curves, we picked 8 of the control group and 6 of the autistic group as the training set, the rest of the data as the test set. The misclassification rates for the actual data are 0.25 (LDA), 0.22 (CRUISE), 0.15 (GUIDE) and 0.37(QUEST).

4. CONCLUSION

In this paper, we presented a streamlined framework for the 2D corpus callosum boundary curve extraction, parametrization and classification. For the CC boundary extraction, the

GVF snake is used. For parameterizing the CC curve, we used a new curvature-based arc-length parametrization. Afterwards, as a scheme for data reduction, the Fourier representation is used to characterize the CC curves using only a finite number of Fourier coefficients. Then these coefficients are used in classifying both the simulated and the actual autism data set. We conclude that among many regression tree based techniques, GUIDE provides the smallest misclassification error in differentiating the CC curves between the two groups.

5. ACKNOWLEDGMENT

We wish to thank Kim Dalton and Richard Davidson of the Waisman laboratory for brain imaging and behavior, University of Wisconsin-Madison for providing the imaging data illustrated in this study.

6. REFERENCES

[1] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1984.

[2] Chung, M.K., Dalton, K.M., Alexander, A.L. and Davidson, R.J,. Less white matter concentration in autism: 2D Voxel-based morphometry. *NeuroImage.* 23:242-251, 2004.

[3] Davatzikos, C., Vaillant, M., Resnick, S.M., Prince, J.L., Letovsky, S., Bryan, R.N. A computerized approach for morphological analysis of the corpus callosum. *J. Comp. Assist. Tomog.*, 20(1):88-97, 1996.

[4] Gerig, G., Styner, M., Jones, D., Weinberger, D. and Lieberman, J., Shape analysis of brain ventricles using SPHARM, IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), 2001.

[5] Gonzalez, O. and Maddocks, J.H., Global curvature, thickness and the ideal shapes of knots, *The Proceedings of the National Academy of Sciences, USA*, 96, 4769-4773, 1999.

[6] Guenter, B. and Rarent, R., Motion control: computing the arc length of parametric curves, *IEEE Comp. Graphics and Applications*, Vol, 10, No. 3, 72-78, 1990.

[7] Hall, P. and Poskitt, D.S., A functional data-analytic approach to signal discrimination, *Technometrics*, 43, 1-9, 2001.

[8] Kass, K., Witkin, A. and Terzoploulos, D., Snakes: Active Contour Models, *Int. J. Computer Version*, Vol. 1(4), 321-331, 1987.

[9] Kim, H. and Loh, W.Y. , Classification trees with bivariate linear discriminant node models, *Journal of Comp. and Graphical Stat.* Vol. 12, 512-530, 2003.

[10] Loh, W.Y., Regression trees with unbiased variable selection and interaction detection, *Statistica Sinica*, Vol. 12, 361-386, 2002.

[11] Loh, W.Y. and Shih, Y.S., Split selection methods for classification trees, *Statistica Sinica*, Vol. 7, 815-840, 1997.

[12] Madi, M., Close-form expressions for the approximation of archlength parameterization for Bézier curves, *Int. J. Appl. Math. Comput. Sci.*, Vol. 14, No. 1, 33-41, 2004.

[13] Miller, D.J. and Browning J., A mixture model and EM-Based algorithm for class discovery, robust Classification, and outlier rejection in mixed labeled/unlabeled data sets, *IEEE Trans. on Pattern Ana. and Machine Intel.*, Vol. 25 No.11:1468-1483, 2003.

[14] Perterson, B.S., Felnelgle, P.A., Stalb, L.H. and Gore, J.C., Automated measurement of matent morphological features in the human corpus callosum, *Human Brain Mapping*, 12:232-245, 2001.

[15] Pontil, M. and Verri, A., Support vector machines for 3D object recoginition, *IEEE Trans. on Pattern Ana. and Machine Intel.*, Vol. 20, No. 6, 637-646, 1998.

[16] Shen, L., Ford, J., Makedon, F. and Saykin, A., A surface-based approach for classification of 3D neuroanatomical structures. *Intelligent Data Analysis*, 8(5). 2004.

[17] Xu, C. and Prince, L., Snakes, shapes and gradient vector flow, *IEEE Trans. on Image Proc.*, 7(3), 359-369, 1998.