行政院國家科學委員會專題研究計畫成果報告

自我相似性，維度簡化及
醫學影像分析上統計方式的資料發掘(3/3)

Self-Similarity, Dimension Reduction, and Statistically Data Mining for Medical Image Analysis

計畫編號：NSC 90-2118-M-009-017
執行期限：90 年 8 月 1 日至 91 年 7 月 31 日

主持人：國立交通大學統計學研究所盧鴻興副教授

一、中文摘要

自我相似性在資料的結構發掘上是一重要關鍵因素。在自我相似性的多重解析表示法中，傅立葉轉換，碎形，Gabor 分析和小波理論是當代的技術。我們將應用及比較這些技術在影像分析中特徵向量的尋找上，目的是以較先進的數學理論及統計工具來模擬及分析人類的知覺反應。然而，具有高維度的資料在影像分析處理中十分不易。為了以特徵向量來表示這些巨量資料，在沒有損失訊息的原則之下，維度簡化對保有資料的幾何結構及對計算上的可行性上就顯得特別重要。而主成分分析、獨立成分分析，切片逆回歸、分類迴歸樹等這些統計方法將與自我相似性的技術相整合，應用於影像的叢聚、分類、分割和形態辨別。完備的分析架構及理論方法是此階段研究的目標。最後，在醫學影像分析的領域中，這些先進的技術將會結合專家的知識來進行統計方式的資料發掘。這也就是說醫學的專業知識將轉換成統計檢定及決策，引導出客觀及自動的決策。以這些決策結果與醫學專家的診斷結果相比較，來找出有效的判別方法。在這長期的計劃中，最終的目的是研究出新與有效的統計方法來幫助醫學影像分析及診斷。實證的醫學影像將包括臺大醫院的超音波影像或榮民總醫院的功能性磁振造影影像。

關鍵詞：傅立葉轉換、碎形、Gabor 分析、小波、多重解析、主成分分析、獨立成分分析、切片逆回歸、叢聚分析、分類迴歸樹、巨量資料、統計檢定及決策、維度簡化、醫學影像。

Abstract

Self-similarity is a key ingredient in discovering the structure of functions. Fourier transform, fractals, Gabor analysis and wavelets are state-of-art techniques in the multiresolution representation of self-similarity. These will be studied and compared in the findings of feature vectors for image analysis. The goal is to mimic and analyze human perception with advanced mathematical and statistical tools. For the represented feature vectors of large volume data in high dimension that occur in image analysis, dimension reduction without loss of intrinsic information is important to preserve the geometric structure and make computation feasible. Principal component analysis (PCA), independent component analysis (ICA), sliced inverse regression (SIR), classification and regression trees will be integrated with self-similarity techniques in the clustering, classification, segmentation and pattern recognition of images. A general framework and methodology are aimed at this stage. Finally, these new techniques will be combined with expertise information to statistical data mining in medical images. That is, the medical expertise will be converted into statistical tests and decisions. These developments lead to objective and automatic decisions, which will be compared with the diagnosis results by medical expertise to find out the effectiveness. At the end, new and effective

statistical methods for image analysis will be proposed and investigated to assist the medical diagnosis of medical images in this long-term project. The medical images will include the ultrasound images at the National Taiwan University Hospital or the functional MRI images at the Veterans General Hospital-Taipei.

Keywords: Fourier transform, fractals, Gabor analysis, wavelets, multiresolution analysis, principal component analysis, independent component analysis, sliced inverse regression, classification and regression trees, large volume data, statistical tests and decisions.

二、緣由與目的

Texture segmentation is the first key step and the next step is to perform the further image analysis, such as pattern recognition and diagnosis. As the computation power increases dramatically, the size of data accumulates extremely fast. For large data sets, the analysis methodology becomes challenging. The data analysis methods have to take into account all perspectives of effectiveness, efficiency, computational complexity and cost. Currently, the community of data mining and knowledge discovery is excited in identifying the patterns from data by machine learning, such as in Fayyad *et al.* (1996). However, the advanced statistical methods are not integrated successfully into this arising area yet as pointed out by Friedman (1997). More interactions between the fields of statistics and data mining are needed to cross-fertilize both areas. Hence, this project will go through the studies of medical image to find out effective and efficient methods of data mining by the advanced statistical methods. Specifically, we will use the techniques of multiresolution analysis for self-similarity, dimension reduction, clustering, classification, hypothesis testing and decision rules to mimic human vision and medical expertise. The medical images will include the ultrasound images at the National Taiwan University Hospital or the functional MRI images at the Veterans General Hospital-Taipei. The ultimate goal is to advance the statistical methodologies and establish a bridge to the communities of statistics, data mining, as well as medical images.

三、結果與討論

本三年期計畫到目前為止已有 10 篇論文被國際期刊接受發表，其發表狀態與部分摘要如下。

1. “An Early Vision Based Snake Model for Ultrasound Image Segmentation," *Ultrasound in Medicine and Biology*, 26, 2, 273-285, 2000.

2. “Bayesian Wavelet Shrinkage for Nonparametric Mixed-Effects Models," *Statistica Sinica*, 10, 4, 1021-1040, 2000.

3. “Extended Gauss-Markov Theorem for Nonparametric Mixed-Effects Models," *Journal of Multivariate Analysis*, 76, 2, 249-266, 2001.

4. “An Adaptive Snake Model for Ultrasound Image Segmentation: Modified Trimmed Mean Filter, Ramp Integration and Adaptive Weighting Parameters,” *Ultrasonic Imaging*, 22, 214-236, 2001.

5. “A Textural Approach Based on Gabor Functions for Texture Edge Detection in Ultrasound Images,” *Ultrasound in Medicine and Biology*, 27, 4, 515-534, 2001.

6. “Adaptive Symmetric Mean Filter: A New Noise Reduction Approach Based on the Slope Facet Model,” *Applied Optics*, 40, 29, 5192-5205, 2001.

7. “A Dual Snake Model of High Penetrability for Ultrasound Image Boundary Extraction,” *Ultrasound in Medicine and Biology* 27, 12, 1651-1665, 2001.

8. “A Discrete Region Competition Approach

Incorporating Weak Edge Enhancement for Ultrasound Image Segmentation," to appear in *Pattern Recognition Letters*.

Abstract:
Ultrasound images are difficult for analysis due to their complex textures and speckle noises. Taking into account these two characters, in this paper, we present a new region-based approach for ultrasound image segmentation. The proposed approach is composed of two primary algorithms, namely, discrete region competition and weak edge enhancement. The discrete region competition features four techniques, i.e., region competition, statistics modeling, early vision modeling, and discrete concept. To prevent a region from flooding out of the desired area, weak edges located on the slowly varying slope are enhanced according to their position on the slope and the length of the slope. The discrete region competition incorporating weak edge enhancement has been verified on clinical ultrasound images and promising results have been achieved.

9. "Cell-Based Dual Snake Model: A New Approach to Extracting Highly Winding Boundaries in The Ultrasound Images," to appear in *Ultrasound in Medicine and Biology*.

Abstract:
Two common deficiencies of most conventional deformable models are the need to place the initial contour very close to the desired boundary and the incapability to capture a highly winding boundary for sonographic boundary extraction. To remedy these two deficiencies, a new deformable model, namely, the cell-based dual snake model, is proposed in this paper. The basic idea is to apply the dual snake model in the cell-based deformation manner. While the dual snake model provides an effective mechanism to allow a distant initial contour, the cell-based deformation makes it possible to catch the winding characteristics of the desired boundary. The performance of the proposed cell-based dual snake model has been evaluated on the synthetic images with the simulated speckles and the clinical ultrasound images. The experimental results show that the mean distances from the derived to the desired boundary points are 0.9 ± 0.42 pixels and 1.29 ± 0.39 pixels for the synthetic and the clinical ultrasound images, respectively.

10. "Generalized Cross-Validation for Wavelet Shrinkage in Nonparametric Mixed-Effects Models," to appear in *Journal of computational and Graphical Statistics*.

Abstract:
A nonlinear wavelet shrinkage estimator was proposed in Huang and Lu (2000). Such an estimator combined the asymptotic equivalence to the best linear unbiased prediction and the Bayesian estimation in nonparametric mixed-effects models. In this article a data-driven GCV method is proposed to select hyperparameters. The proposed GCV method has low computational cost and can be applied to one or higher dimensional data. It can be used for selecting hyperparameters for either level independent or level dependent shrinkage. It can also be used for selecting the primary resolution level and the number of vanishing moments of wavelet basis. The strong consistency of the GCV method is proved.

四、計畫成果自評

由上述的報告中，可以發現我們的研究內容與原計畫相符，達成預期的目標。我們將進一步將完成的技術報告投稿到學術期刊發表，並進一步將這些技術應用到實際的影像分析方面，提供更正確和有效的統計分析。因此，本計畫的研究除了在學術上分析方法的突破，也同時具備應用的價值。

五、參考文獻

[1] A. Antoniadis and G. Oppenheim (1995) *Wavelets and Statistics*, Springer-Verlag, New York.

[2] I. Daubechies (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia.

[3] M. Farge, J.C.R. Hunt, and J. C. Vassilicos (1993) *Wavelets, Fractals, and Fourier Transforms*, Oxford University Press.
[4] U. M. Fayyad *et al.* (1996) *Advances in Knowledge Discovery and Data Mining*, AAAI Press / The MIT Press, Cambridge, MA.
[5] J. H. Friedman (1997) “Data Mining and Statistics: What's the Connection?” Technical Report, Department of Statistics, Stanford University, CA.
[6] K.-C. Li (1991) ``Sliced inverse regression for dimension reduction, (with discussions),'' *J. Amer. Statist. Assoc.*, 86, 316-342.
[7] S. Mallat (1998) *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA.
[8] Y. Meyer (1993) *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, PA.
[9] R. T. Ogden (1997) *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhauser, Boston, MA.
[10] L. Prasad and S. S. Iyengar (1997) *Wavelet Analysis with Applications to Image Processing*, CRC Press, New York, NY.
[11] G. Strang and T. Nguyen (1996) *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA.

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※　　　　**自我相似性，維度簡化及**　　　　※
※　**醫學影像分析上統計方式的資料發掘(3/3)**　※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：■個別型計畫　　□整合型計畫
計畫編號：NSC　90-2118-M-009-017
執行期間：90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：國立交通大學統計學研究所盧鴻興副教授
共同主持人：
計畫參與人員：

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學統計學研究所

中　華　民　國　91 年　10 月　30 日