

國立交通大學

統計學研究所

碩士論文

使用空間相關性的高斯混合模型

對 PET/CT 影像作分割

Segmentation of PET/CT Images Using Spatial
Dependence in Gaussian Mixture Model

研究生：陳亮勳

指導教授：盧鴻興 教授

中華民國九十九年六月

使用空間相關性的高斯混合模型

對 PET/CT 影像作分割

Segmentation of PET/CT Images Using Spatial Dependence in
Gaussian Mixture Model

研 究 生： 陳亮勳

Student：Liang-Xun Chen

指導教授： 盧鴻興 教授

Advisor：Dr. Henry Horng-Shing Lu



Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Master
in
Statistics
June 2010
Hsinchu, Taiwan

中 華 民 國 九 十 九 年 六 月

使用空間相關性的高斯混合模型 對 PET/CT 影像作分割

學生： 陳亮勳

指導教授： 盧鴻興 教授

國立交通大學理學院
統計學研究所

摘 要

PET 影像中，高亮度的區塊常被視為疑似腫瘤產生的地方。若能將 PET 影像中的腫瘤部位精確的分割出來，將會對醫生的診療有很大的幫助。近年來，由於 PET/CT 的發明，它結合了 PET 和 CT 的優點，能將腫瘤細胞的活動狀況及位置融合在同一張影像中，使醫生對腫瘤診斷有更進一步的發展。而本研究在分割 PET 影像的同時也加進了 CT 影像的資訊，目的是希望能將腫瘤細胞更精確的分割出來。

我們使用 Gaussian mixture model (GMM) 對 PET 和 CT 的融合影像作分割。此外，我們考慮了 PET 和 CT 的相關性，使用一個二維的 GMM 對 PET/CT 影像作分割。我們還在 GMM 中加入空間相關性，將影像的中每一個像素都考慮它們的鄰近點，然後使用一個多維的 GMM 模型去配適。這些方法的結果均較單對 PET 影像作分割的結果為佳。

Segmentation of PET/CT Images Using Spatial Dependence in Gaussian Mixture Model

Student : Liang-Xun Chen

Advisor : Dr. Henry Horng-Shing Lu

Institute of Statistics

National Chiao Tung University

Hsinchu, Taiwan

ABSTRACT

The specific brightened regions of PET images are the suspected regions of tumor. Segmenting the region of tumor on PET images will be very helpful to doctors. In recent years, the invention of Positron emission tomography/computed tomography (PET/CT) has allowed combination of the advantages of PET and CT: namely, that the activities and location of tumor cells can be merged in one image. This merged images provides significant progress for doctors diagnosing tumors. In this study, the information of CT is used when segmenting the PET images. The aim is to enhance accuracy of segmenting the regions of tumor.

The fusion images of PET and CT are segmented by Gaussian mixture model (GMM). In addition, a two-dimensional GMM is used to fit the PET and CT image data by considering the correlation of PET and CT. The spatial dependence is also considered in a GMM. Our approach is to consider points surrounding each pixel to fit a multi-dimensional GMM to the data. These methods all have better performance than the result of only implementing segmenting PET images in this study.

誌 謝

感謝盧鴻興老師對我的耐心指導，才能使本論文順利完成！跟著老師的這段日子，讓我學到了真正做研究的精神與態度，也開始學會從不同的角度思考、解決問題。接著，還要特別感謝陳泰賓學長，總是在我遇到瓶頸的時候適時的幫助我，學長提供的建議及勉勵的話，讓我受益良多。

在統計所的這兩年，謝謝郭姐一直都這麼照顧我們，郭姐就像個媽媽一樣支持、鼓勵我們，為我們處理一切大小事物，才能讓我們能無後顧之憂的學習。

最後，謝謝家人們對我的支持，以及曾在我求學路上幫助過我的人。還有研究所的同學們，因為有了你們的陪伴，才能讓我在最後的學生生涯充滿著歡笑，也因為有你們的鼓勵與幫助，才能讓我順利完成學業。即將離開交大，各自邁向新的旅程了，我會想念你們和這裡所有的一切，祝大家未來都能夠一切順利！

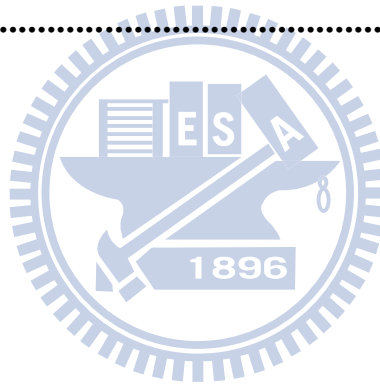
陳 亮 勳 謹誌於

國立交通大學統計所研究所

中華民國九十九年七月

Contents

| | | |
|------------------|--|-----------|
| Chapter 1 | Introduction | 1 |
| Chapter 2 | Methodologies | 1 |
| 2.1 | Gaussian Mixture Model (GMM) | 3 |
| 2.2 | Using Spatial Dependence in GMM | 6 |
| 2.3 | Determine the Initial Parameters of GMM..... | 8 |
| 2.4 | Bayesian Information Criterion for Model Selection | 10 |
| 2.5 | The Method of Fused PET and CT Images | 10 |
| 2.6 | F-measure..... | 11 |
| 2.7 | Procedure..... | 12 |
| Chapter 3 | Phantom Study..... | 14 |
| Chapter 4 | Conclusions | 23 |
| Reference | | 24 |



Chapter 1 Introduction

Positron emission tomography (PET) is a type of medical imaging technique utilizing nuclear technology. Due to the different ratio of nuclear medicine absorption in human cells, PET can provide high contrast images. Doctors can detect the metabolic activity in a human body and diagnose the abnormal region with PET images. Because cancer cells need to consume a great amount of glucose as energy for mitosis, PET images can produce extraordinarily brightened regions at the suspected tumor region. Therefore, PET is a critical tool for doctors to diagnose cancer.

Segmenting the region of tumor on PET images will be very helpful to doctors. However, because of the higher noise property associated with PET images, it is difficult to segment PET images accurately. In recent years, the invention of Positron emission tomography /computed tomography (PET/CT) has allowed combination of the advantages of PET and CT: namely, that the activities and location of tumor cells can be merged in one image. This merged image provides significant progress for doctors diagnosing tumors. In this study, the information of CT will be used when segmenting the PET images. The aim is to enhance the accuracy of segmenting regions of high activity.

Besides segmenting PET images, we also segment fusion images of PET and CT. Due to the different variances in different clusters, a Gaussian mixture model (GMM) can be used to classify the image data (McLachlan and Basford, 1988, Hsiao, Rangarajan and Gindi, 1998). In addition, a two-dimensional GMM is used to fit the PET and CT image data by considering the correlation of PET and CT. The spatial dependence is also considered in a GMM. Our approach is to consider points surrounding each pixel to fit a multi-dimensional GMM to the data.

The image data, based on a GMM, is generated by the weighed sum of several Gaussian distributions. The unknown parameters of the GMM can be estimated by the Expectation-

maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977, Wu, 1983). The initial number of clusters and initial values of EM algorithm can be determined by kernel density estimation (KDE). The issue of model selection can be solved by Bayesian information criterion (BIC) (Schwarz, 1978).

The torso phantom data were provided by Dr. Tai-Been Chen from I-Shou University. It is designed for comparing the accuracy of different methods in this study. Because the exact location of the highest activity regions is known, the F-measure (van Rijsbergen, 1979) can be used to compare the accuracy. These methods using information of CT are all better than the result of only implementing segmenting PET images in this study.



Chapter 2 Methodologies

2.1 Gaussian Mixture Model (GMM)

GMM is an extension of a single Gaussian distribution. It can flexibly fit the asymmetric and multi-modal data. Since the variances of different clusters may be different, GMM's can be used to classify the image data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, $\mathbf{x}_i \in \mathbf{R}^d$ (McLachlan and Basford, 1988, Hsiao, Rangarajan and Gindi, 1998). Suppose \mathbf{x}_i is generated by a mixture of K different Gaussian distributions and the prior weight of k th distribution is π_k , the GMM model can expressed as

$$p(\mathbf{x}_i | \Theta) = \sum_{k=1}^K \pi_k f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (2.1.1)$$

with

$$\sum_{k=1}^K \pi_k = 1 \text{ and } 0 \leq \pi_k \leq 1 \quad (2.1.2)$$

where

$$f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k)}{2}\right) \quad (2.1.3)$$

represents the density function of the k th d -dimension Gaussian distribution with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. The overall parameters are $\Theta = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_K, \pi_1, \pi_2, \dots, \pi_K)$. The maximum likelihood estimator (MLE) can be use to estimate the parameters of GMM. The log likelihood function is

$$\begin{aligned} \log L(\Theta) &= \log \prod_{i=1}^N p(\mathbf{x}_i | \Theta) \\ &= \sum_{i=1}^N \log \left(\sum_{k=1}^K \pi_k f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \end{aligned} \quad (2.1.4)$$

Next, we need to solve the partial differential equations to find Θ such that the log likelihood function is maximized,

$$\frac{\partial \log L(\Theta)}{\partial \Theta} = 0 \quad (2.1.5)$$

However, it is difficult to solve the above equations directly. Alternatively, the EM algorithm can be used to estimate the parameters (Dempster, Laird, and Rubin, 1977, Wu, 1983). First, we denote the unobserved vector $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iK})$, where Z_{ik} represents the indicator function as follows,

$$Z_{ik} = \begin{cases} 1, & \text{if } \mathbf{x}_i \text{ comes from } k\text{th normal distribution} \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, $\{(\mathbf{x}_1, \mathbf{Z}_1), \dots, (\mathbf{x}_N, \mathbf{Z}_N)\}$ form the complete data for EM algorithm. The joint density function of \mathbf{x}_i and \mathbf{Z}_i is

$$p(\mathbf{x}_i, \mathbf{Z}_i) = \sum_{k=1}^K Z_{ik} \pi_k f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \quad (2.1.6)$$

The log likelihood function becomes

$$\begin{aligned} \log L(\Theta) &= \log \prod_{i=1}^N p(\mathbf{x}_i, \mathbf{Z}_i) \\ &= \sum_{i=1}^N \log \left(\sum_{k=1}^K Z_{ik} \pi_k f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\ &= \sum_{i=1}^N \sum_{k=1}^K Z_{ik} \log(\pi_k f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) \\ &= \sum_{i=1}^N \sum_{k=1}^K Z_{ik} (\log \pi_k + \log f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)). \end{aligned} \quad (2.1.7)$$

Then, we start the Expectation step in the EM algorithm when given initial parameters $\Theta^{(old)}$ and \mathbf{X} .

$$\begin{aligned} Q(\Theta; \Theta^{(old)}) &= E(\log L(\Theta) | \mathbf{X}, \Theta^{(old)}) \\ &= E \left(\sum_{i=1}^N \sum_{k=1}^K Z_{ik} (\log \pi_k + \log f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) \middle| \mathbf{X}, \Theta^{(old)} \right) \\ &= \sum_{i=1}^N \sum_{k=1}^K E(Z_{ik} | \mathbf{X}, \Theta^{(old)}) (\log \pi_k + \log f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) \end{aligned} \quad (2.1.8)$$

Given \mathbf{X} and $\Theta^{(old)}$, the expectation of Z_{ik} can be evaluated as follows:

$$\begin{aligned}
E(Z_{ik} | X, \Theta^{(old)}) &= 1 \cdot p(Z_{ik} = 1 | \mathbf{x}_i, \Theta^{(old)}) + 0 \cdot p(Z_{ik} = 0 | \mathbf{x}_i, \Theta^{(old)}) \\
&= \frac{\pi_k^{(old)} f_k(\mathbf{x}_i | \boldsymbol{\mu}_k^{(old)}, \boldsymbol{\Sigma}_k^{(old)})}{\sum_{j=1}^K \pi_j^{(old)} f_j(\mathbf{x}_i | \boldsymbol{\mu}_j^{(old)}, \boldsymbol{\Sigma}_j^{(old)})}
\end{aligned} \tag{2.1.9}$$

Hence, the equation can be obtained,

$$Q(\Theta; \Theta^{(old)}) = \sum_{i=1}^N \sum_{k=1}^K \frac{\pi_k^{(old)} f_k(\mathbf{x}_i | \boldsymbol{\mu}_k^{(old)}, \boldsymbol{\Sigma}_k^{(old)})}{\sum_{j=1}^K \pi_j^{(old)} f_j(\mathbf{x}_i | \boldsymbol{\mu}_j^{(old)}, \boldsymbol{\Sigma}_j^{(old)})} (\log \pi_k + \log f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) \tag{2.1.10}$$

Next, we start the Maximization step in the EM algorithm. It is a step to find what Θ maximizes (2.1.10). The solutions are used to update $\Theta^{(old)}$.

On the constraint of (2.1.2), we can use the Lagrange multipliers to solve the following function:

$$Q(\Theta; \Theta^{(old)}; \lambda) = Q(\Theta; \Theta^{(old)}) + \lambda \left(\sum_{k=1}^K \pi_k - 1 \right). \tag{2.1.12}$$

We can find the solution on the constraint of (2.1.2), as follows:

$$\pi_k^{(new)} = \frac{1}{N} \sum_{i=1}^N p(k | \mathbf{x}_i, \Theta^{(old)}), \tag{2.1.13}$$

$$\boldsymbol{\mu}_k^{(new)} = \frac{\sum_{i=1}^N \mathbf{x}_i p(k | \mathbf{x}_i, \Theta^{(old)})}{\sum_{i=1}^N p(k | \mathbf{x}_i, \Theta^{(old)})}, \tag{2.1.14}$$

$$\boldsymbol{\Sigma}_k^{(new)} = \frac{\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T p(k | \mathbf{x}_i, \Theta^{(old)})}{\sum_{i=1}^N p(k | \mathbf{x}_i, \Theta^{(old)})}, \tag{2.1.15}$$

where

$$p(k | \mathbf{x}_i, \Theta^{(old)}) = \frac{\pi_k^{(old)} f_k(\mathbf{x}_i | \boldsymbol{\mu}_k^{(old)}, \boldsymbol{\Sigma}_k^{(old)})}{\sum_{j=1}^K \pi_j^{(old)} f_j(\mathbf{x}_i | \boldsymbol{\mu}_j^{(old)}, \boldsymbol{\Sigma}_j^{(old)})}, \tag{2.1.16}$$

represents the posterior probability when given \mathbf{x}_i and Θ . After obtaining the estimated parameters of GMM, we can use this posterior probability to classify the data. If the conditional probability of the data point classified to the specific cluster larger than the others, it is classified to the cluster. A summary of the GMM algorithm follows below:

- Step 1. Determine the numbers of clusters K , and initialize the values, $\Theta^{(old)}$.
- Step 2. Update $\Theta^{(new)}$ by using (2.1.13), (2.1.14) and (2.1.15).
- Step 3. Repeat Step 3 until $\log L(\Theta^{(new)}) - \log L(\Theta^{(old)}) < \varepsilon$.
- Step 4. Using posterior probability (2.1.16) to classify data.

2.2 Using Spatial Dependence in GMM

Besides using one-dimensional GMMs to segment PET images and fusion images of PET and CT, a two-dimensional GMM is also used to fit the PET and CT image data by considering the correlation of PET and CT. It is shown as follows:

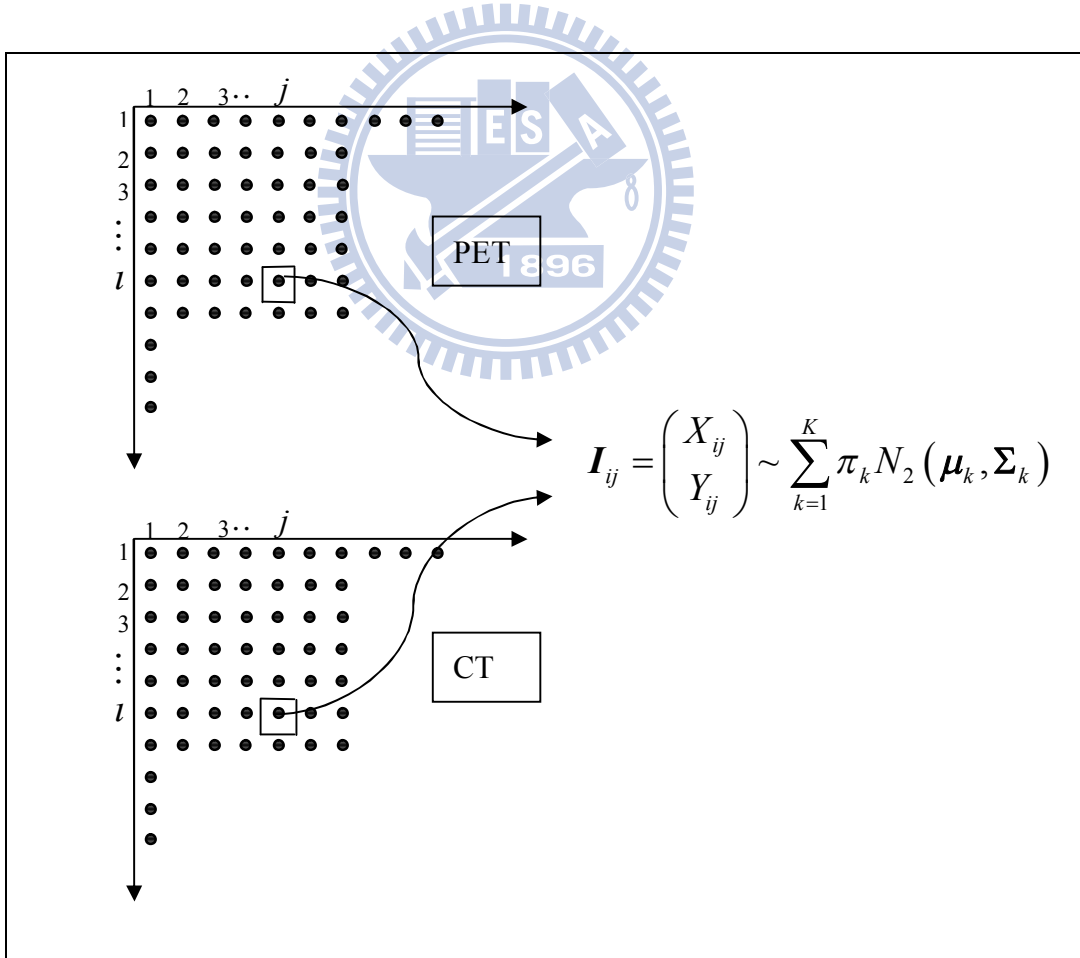


Figure 2.3.1: Using two-dimensional GMM to fit the PET and CT image data.

Next, the 1st and 2nd order spatial dependence are considered in GMM. The 1st order spatial dependence is to add the four adjacent points of each pixel on PET and CT images. A ten-dimensional GMM can be used to fit the data (PET and CT). The 2nd order spatial dependence is to add the eight points surrounding each pixel on PET and CT images. An eighteen-dimensional GMM can be used to fit the data (PET and CT). They are shown as follows:

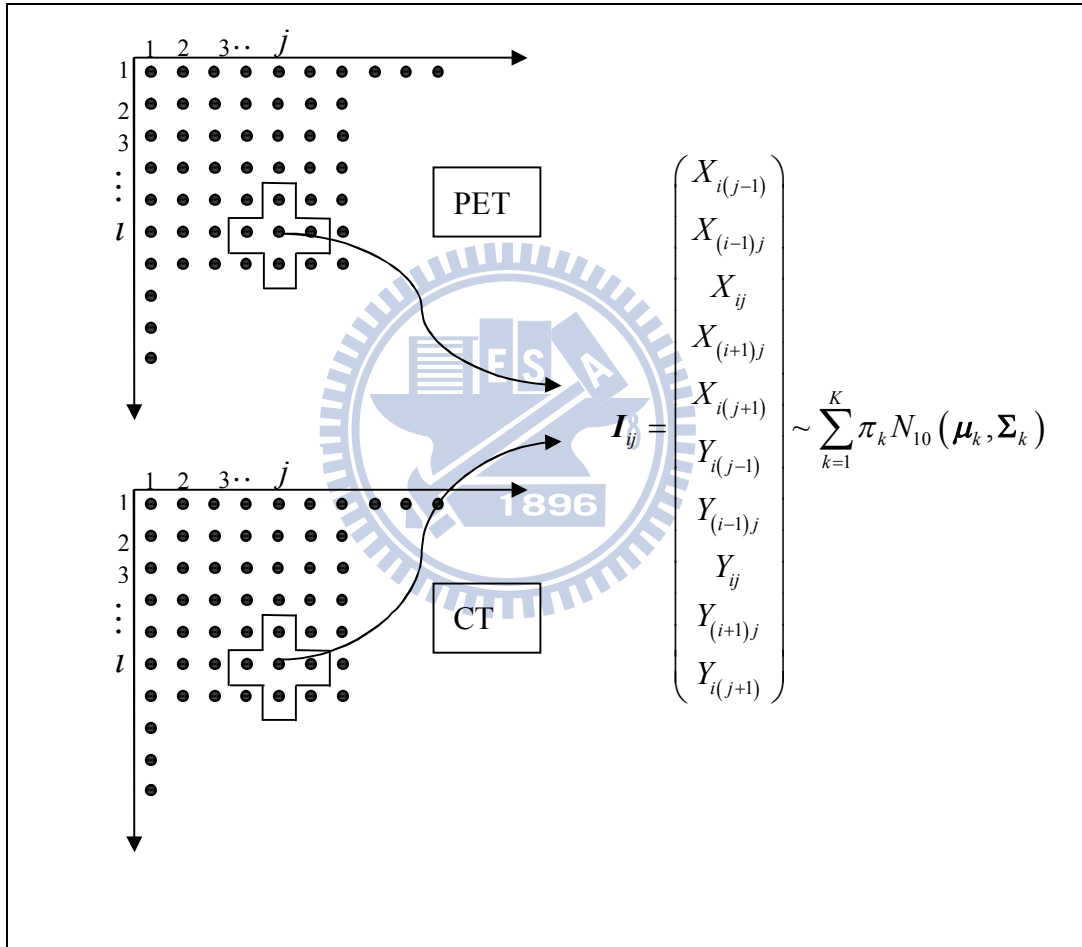


Figure 2.3. 2: The 1st order spatial dependence in GMM.

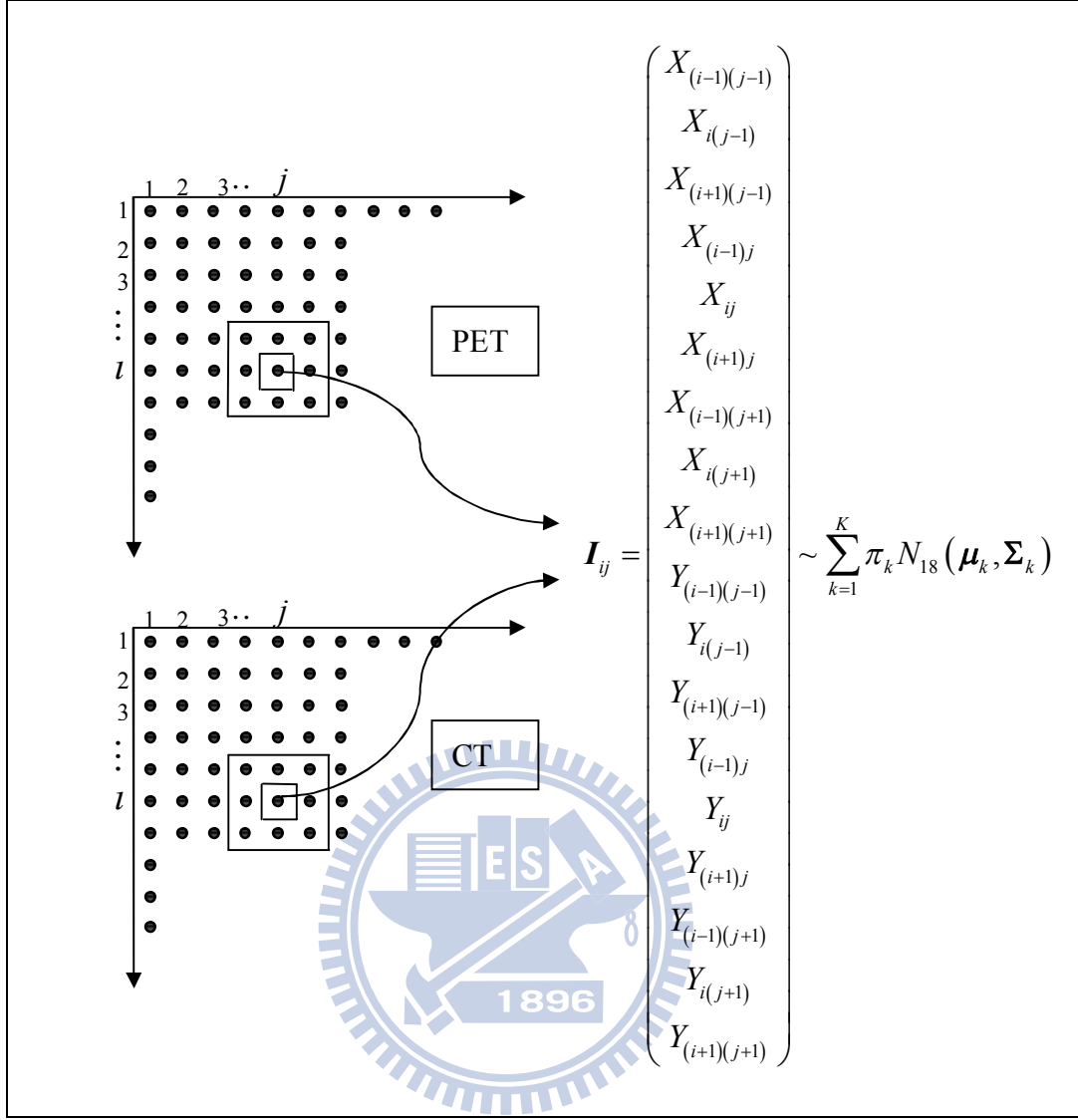


Figure 2.3.3: The 2nd order spatial dependence in GMM.

2.3 Determine the Initial Parameters of GMM

The different initial values and number of clusters affect the clustering results in a GMM significantly. We will discuss how to select initial values of the GMM in this section.

Kernel density estimation (KDE) is a non-parametric approach to estimate the density function of data (Sheather and Jones, 1991). For a one-dimensional GMM, KDE can be applied to solve the issue of selecting initial values and the initial number of clusters. The K high peaks can be observed in a kernel density curve. By the concept of mixture distribution, The K high peaks correspond to K different structure of distribution. Hence, the initial

number of clusters K can be determined. The means and standard deviations of GMM can be determined by high peaks and low peaks. It is based on the 68-95-99.7 rule from a Gaussian distribution that about 99.7% data lie within the interval $[\mu - 3\sigma, \mu + 3\sigma]$. The following is a basic example that shows a procedure to select the initial values of the GMM.

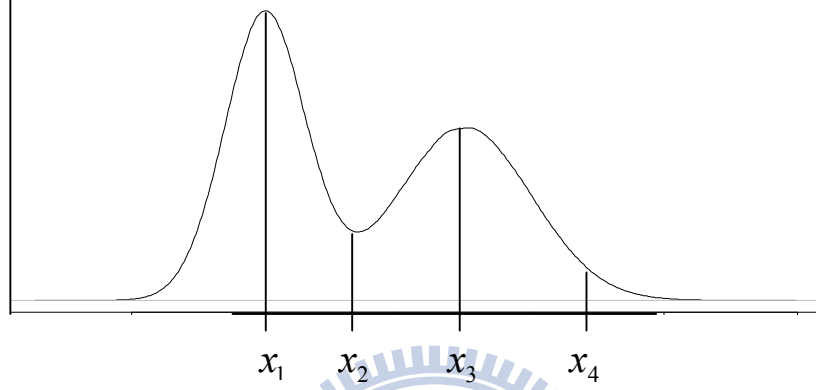


Figure 2.3.1: The mixture model of two Gaussian distributions.

From the Figure 2.3.1, the number of clusters $k=2$ can be determined by the density curve. The initial values of GMM are $\mu_1^{(0)} = x_1$, $\mu_2^{(0)} = x_3$, $\sigma_1^{(0)} = \frac{x_2 - x_1}{3}$ and $\sigma_2^{(0)} = \frac{x_4 - x_3}{3}$. In addition, the initial clustering result can be obtained by the cutoff point x_2 .

For a multi-dimensional GMM, we can reduce the data to one-dimension by averaging them. The initial clustering result can be obtained by the cutoff points (low peaks) of the kernel density curve. And we use maximum likelihood estimator (MLE) of Multivariate Gaussian distribution to estimate the parameters of the multi-dimensional GMM. For $k = 1, 2, \dots, K$,

$$\pi_k^{(0)} = \frac{\sum_{xy=1}^N I_{\{I_{xy} \in G_k\}}}{N},$$

$$\mu_k^{(0)} = \frac{\sum_{I_{xy} \in G_k} I_{xy}}{N},$$

$$\Sigma_k^{(0)} = \frac{\sum_{I_{xy} \in G_k} (I_{xy} - \mu_k)(I_{xy} - \mu_k)^T}{N},$$

where G_k represents the k th group, I_{xy} is d -dimensional vector.

In this study, Gaussian kernel functions are used to estimate the density function of image data. The bandwidth of KDE is determined by the program code of "density" in software R (Silverman, 1986).

2.4 Bayesian Information Criterion for Model Selection

Bayesian information criterion (BIC) is a criterion for model selection developed by Schwarz (1978). It is similar to Akaike information criterion (AIC) (Akaike, 1974). The difference between BIC and AIC is that the penalty of the BIC for additional parameters is stronger than the penalty of the AIC. They are defined as follows:

$$AIC = -2\log(L) + 2p$$

$$BIC = -2\log(L) + p\log(N)$$

where p is the number of free parameters, L is maximized value of likelihood function for the estimated model, and N is the number of data. The values of AIC and BIC are the same when the sample size $N = 7.389$. When the sample size is greater than 7.389, the penalty term in BIC is greater than the AIC. Since the size of the image data are larger than 7.389, we will use BIC to select model. Given several estimated models, the model with the lower value of BIC is preferred. Hence, the number of clusters for GMM can be determined by the model with the minimum value of BIC.

2.5 The Method of Fused PET and CT Images

The advantage of PET is that it is easier to detect abnormal regions by high and low

intensity on images. CT, however, has a better outlining effect. The profiles of organs can be observed very clearly on CT images. We can combine the advantages of PET and CT. And we take weighed sum of PET and CT images to form one image as follow.

$$\text{Fusion} = \alpha \times \text{PET} + (1 - \alpha) \times \text{CT}, \alpha \in [0, 1].$$

In this study, let $\alpha = 0.5$. That is, we take the average of PET and CT images to form fusion images. In addition, before we fused PET and CT images, we normalized the images between 0 and 1.

2.6 F-measure

In a statistical classification task, F-measure is a measure of testing accuracy (van Rijsbergen, 1979). It combines the precision (or we called Predicted Positive Value , PPV) and recall (or we called True Positive Rate, TPR) to form a single measure. Precision could be seen a measure of exactness and recall could be seen a measure of completeness. They are defined as following.

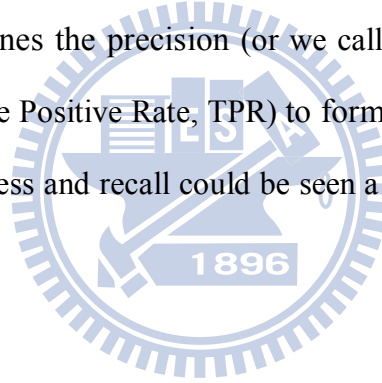


Table 2.6.1: The 2×2 contingency table of the real and estimated result.

| | | Real result | |
|------------------|-------|---------------------|---------------------|
| | | True | False |
| Estimated result | True | True Positive (TP) | False Positive (FP) |
| | False | False Negative (FN) | True Negative(TN) |

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$F = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{(\text{Precision} + \text{Recall})}.$$

The F-measure could be interpreted as the harmonic mean of precision and recall. It has the best performance when the value is 1 and worst performance when the value is 0.

2.7 Procedure

In this study, we have five procedures to segment the images shown as follows. And we will compare the accuracy of these methods by phantom study on Chapter 3.

Procedure 1: PET + GMM

Use one-dimensional GMM to segment PET images only.

Procedure 2: Fusion + GMM

Use one-dimensional GMM to segment fusion images of PET and CT.

Procedure 3: GMM (PET & CT)

Use two-dimensional GMM to segment PET/CT images.

Procedure 4: 1st order spatial dependence in GMM (PET & CT)

Use the 1st order spatial dependence in GMM to segment PET/CT images.

Procedure 5: 2nd order spatial dependence in GMM (PET & CT)

Use the 2nd order spatial dependence in GMM to segment PET/CT images.

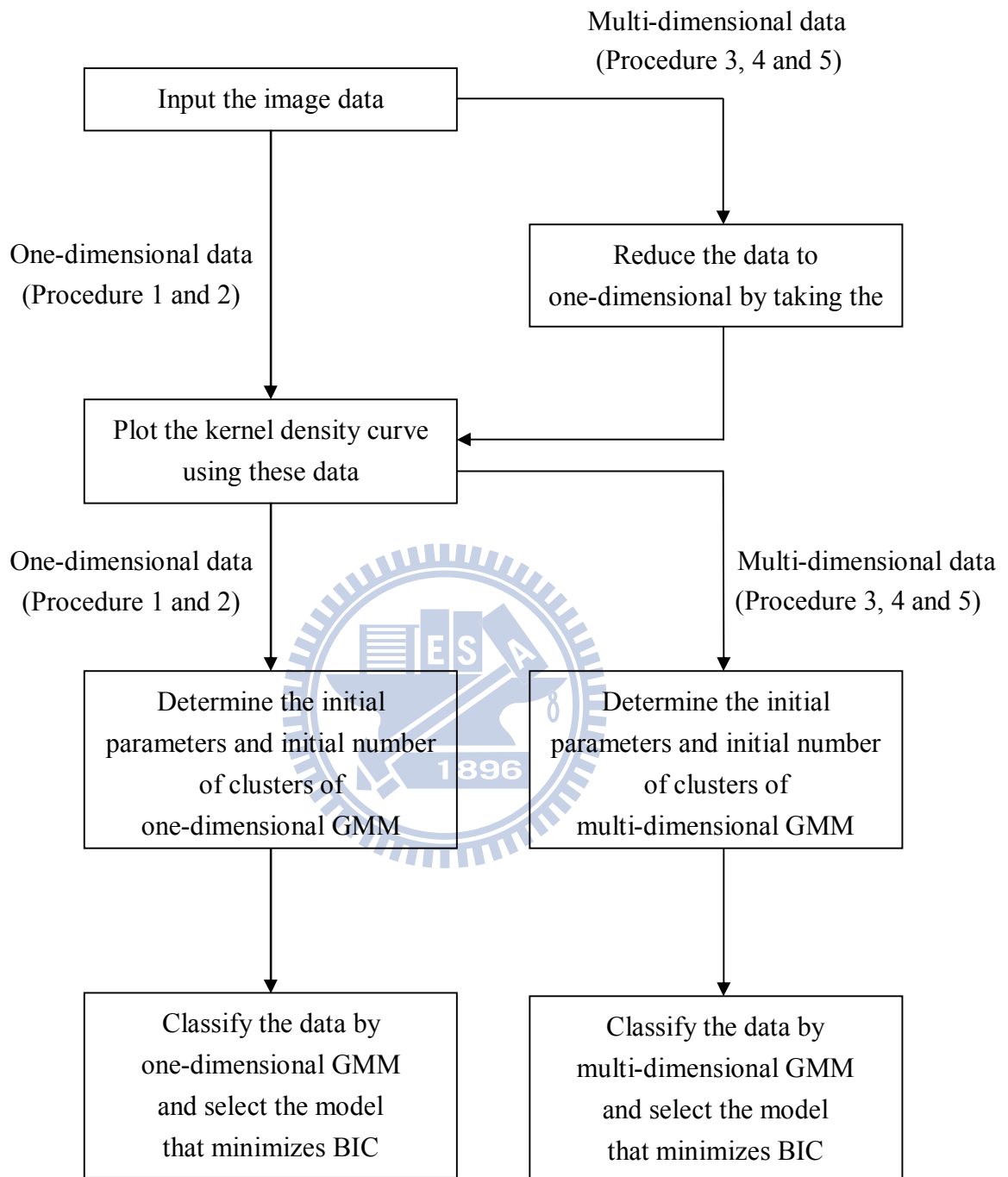


Figure 2.7.1: The procedure of segmenting images.

Chapter 3 Phantom Study

The torso phantom data were provided by Dr. Tai-Been Chen from I-Shou University. It is designed for comparing the accuracy of different segmenting methods in this study. The information of phantom container is shown below. The region of the simulated heart container is injected a large amount of medicine. This makes the image appear in high intensity on PET images.

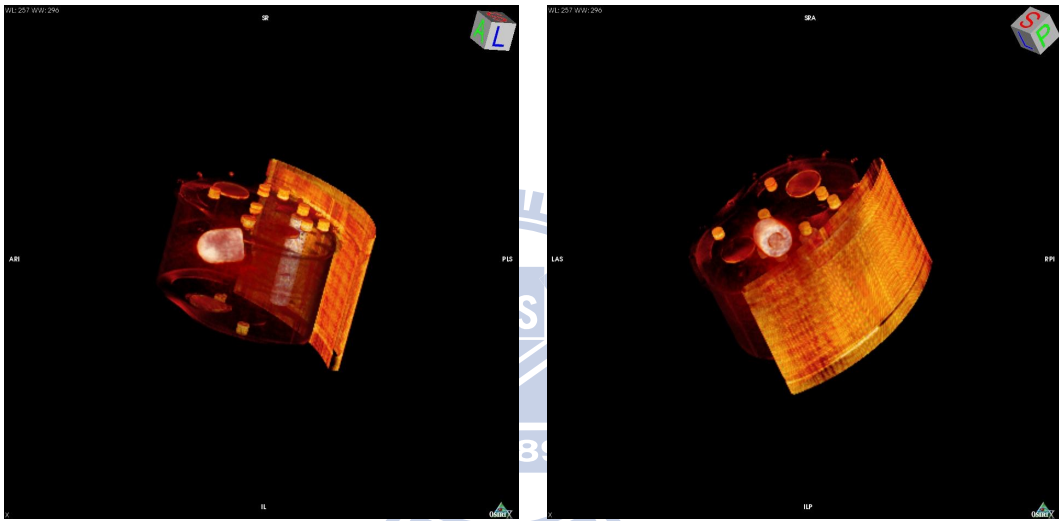


Figure 3.1: The information of phantom container.

PET and CT both have 114 slices where pixel size is 256×256 in the phantom experiment. The PET images which contain high intensity identify the area with high activity in the simulator. Different segmenting methods are used to segment the region of the highest activity in the simulator. The aim of this phantom study is to compare the accuracy of the highest region by these different segmenting methods. The following is the 43th slice of PET and CT images.

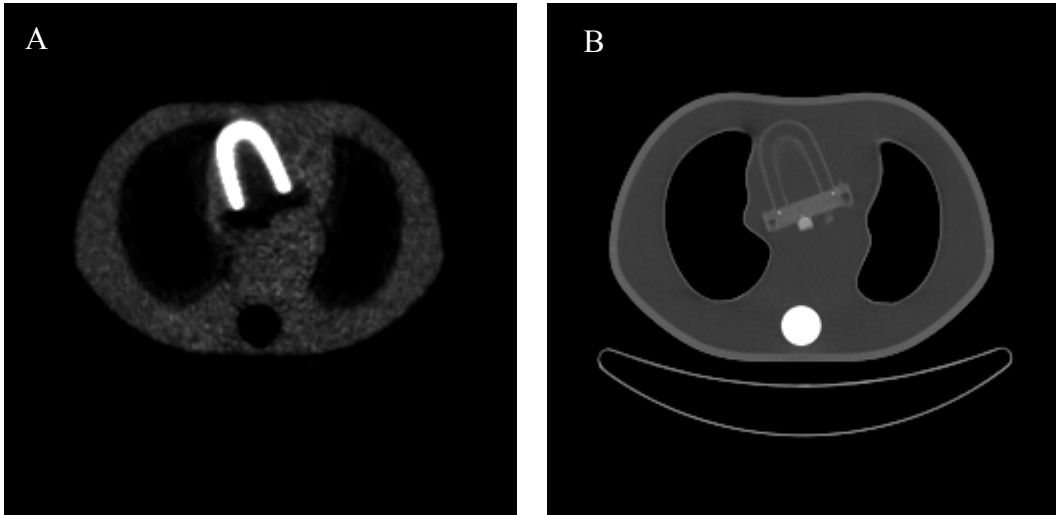


Figure 3.2: (A) The 43th slice of 114 PET images. (B) The 43th slice of 114 CT images.

The region of high activity can be observed in the PET image. From CT image, the profile and organ can be observed very clearly. Therefore, we can find the boundary of the simulated heart container from CT image. Because the region of the boundary of simulated heart container is injected a large amount of medicine, the true highest activity region can be selected by CT images' good outlining effect. It is shown as follow:

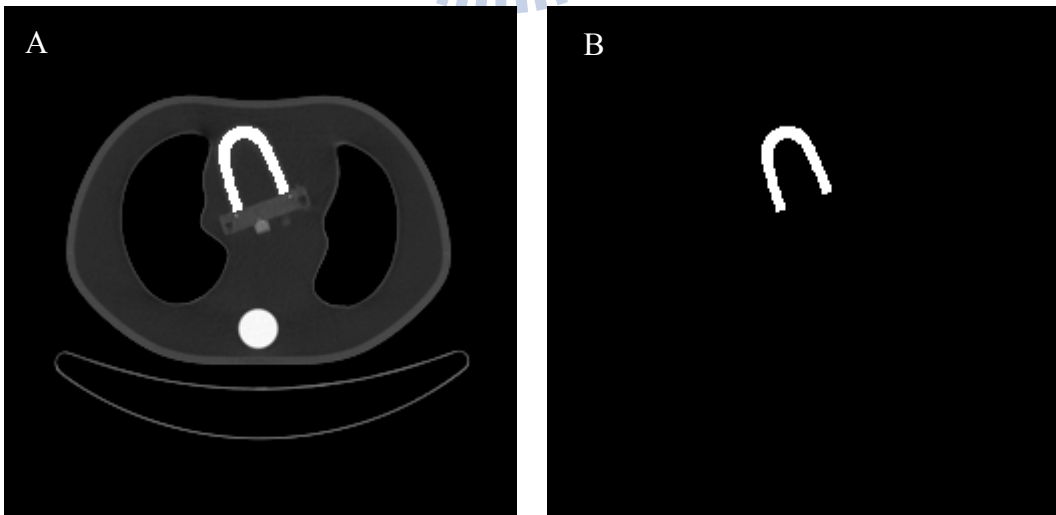


Figure 3.3: (A) The boundary of simulated heart container can be found from CT images. (B) The region of the true highest activity of 43th slice of CT image.

After obtaining the true highest activity region, we can compare the accuracy by F-measure (van Rijsbergen, 1979). First, the GMM is used to segment PET images. The following is kernel density curve of PET image data.

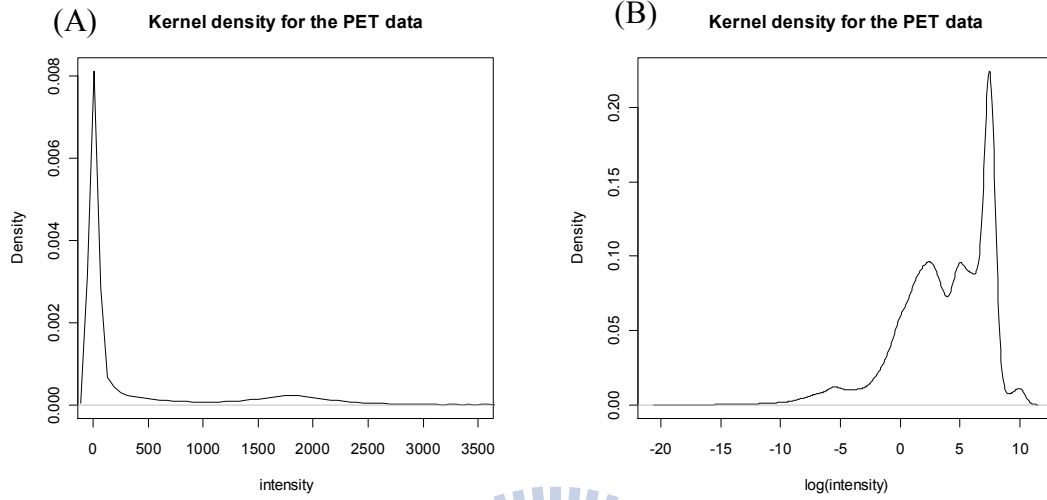


Figure 3.4: (A) The kernel density curve of PET image (B) Take log of data and plot density curve. These peaks are more obvious.

By the kernel density curve from Figure 3.4, the initial values of GMM can be determined. And the model with minimum value of BIC can be selected as $k = 6$ (the number of clusters) from Figure 3.5 shown as follow.

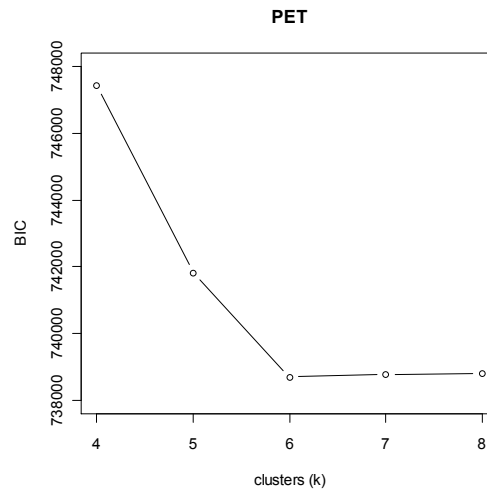


Figure 3.5: The BIC values of different clusters from GMM (PET).

The results of segmenting PET image by GMM shown as following. It obviously overestimates the real highest region.

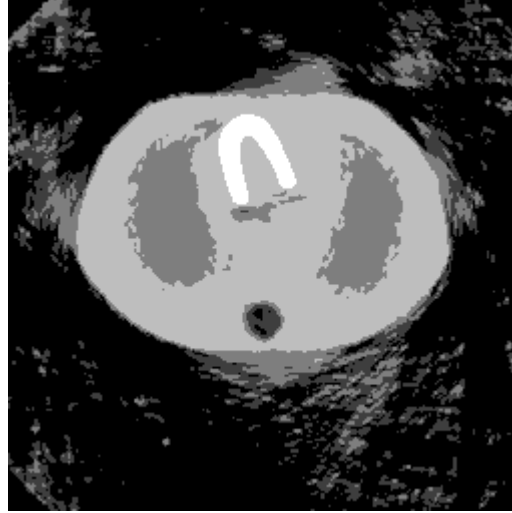


Figure 3.6: $k = 6$, The result of segmenting PET image by GMM with KDE.

Next, the fusion image of PET and CT is segmented by GMM. The following is kernel density curve of fusion image data.

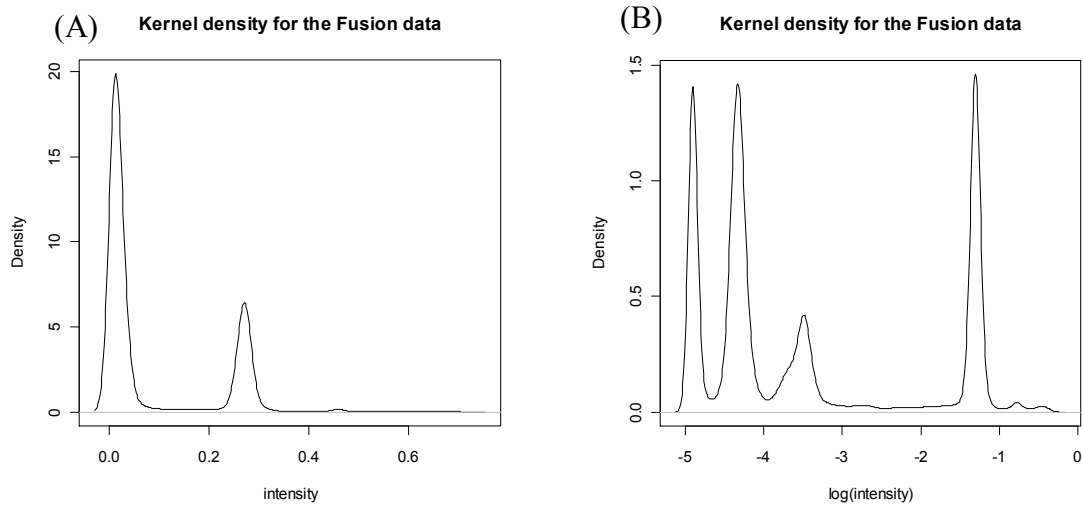


Figure 3.7: (A) The kernel density curve of fusion image of PET and CT. (B) Take log of data and plot density curve. These peaks are more obvious.

The initial values of GMM can be determined by the kernel density curve from Figure 3.7. And the model with minimum value of BIC can be selected as $k = 8$ from Figure 3.8 shown as follow.

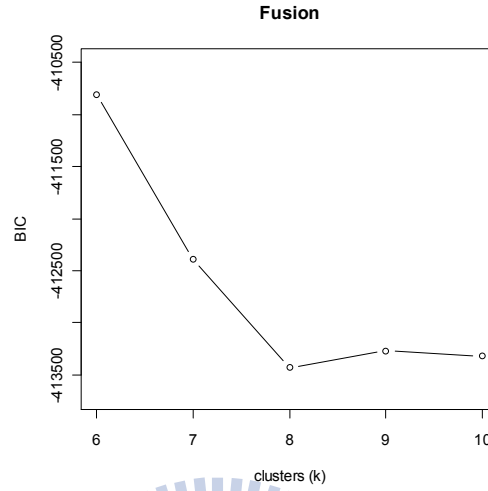


Figure 3.8: The BIC values of different clusters for GMM (Fusion). k is the number of clusters.

The result of segmenting fusion image by GMM is shown as following. It is better than the result of only implementing segmenting PET image at the highest region.



Figure 3.9: $k = 8$, The result of segmenting fusion image by GMM with KDE.

Finally, the two-dimensional GMM is fitted to the PET and CT data. And the 1st order and the 2nd order spatial dependence are used to in GMM as shown Figure 2.3.2 and Figure 2.3.3 on Chapter 2. The followings are the BIC values of different clusters of these three methods.

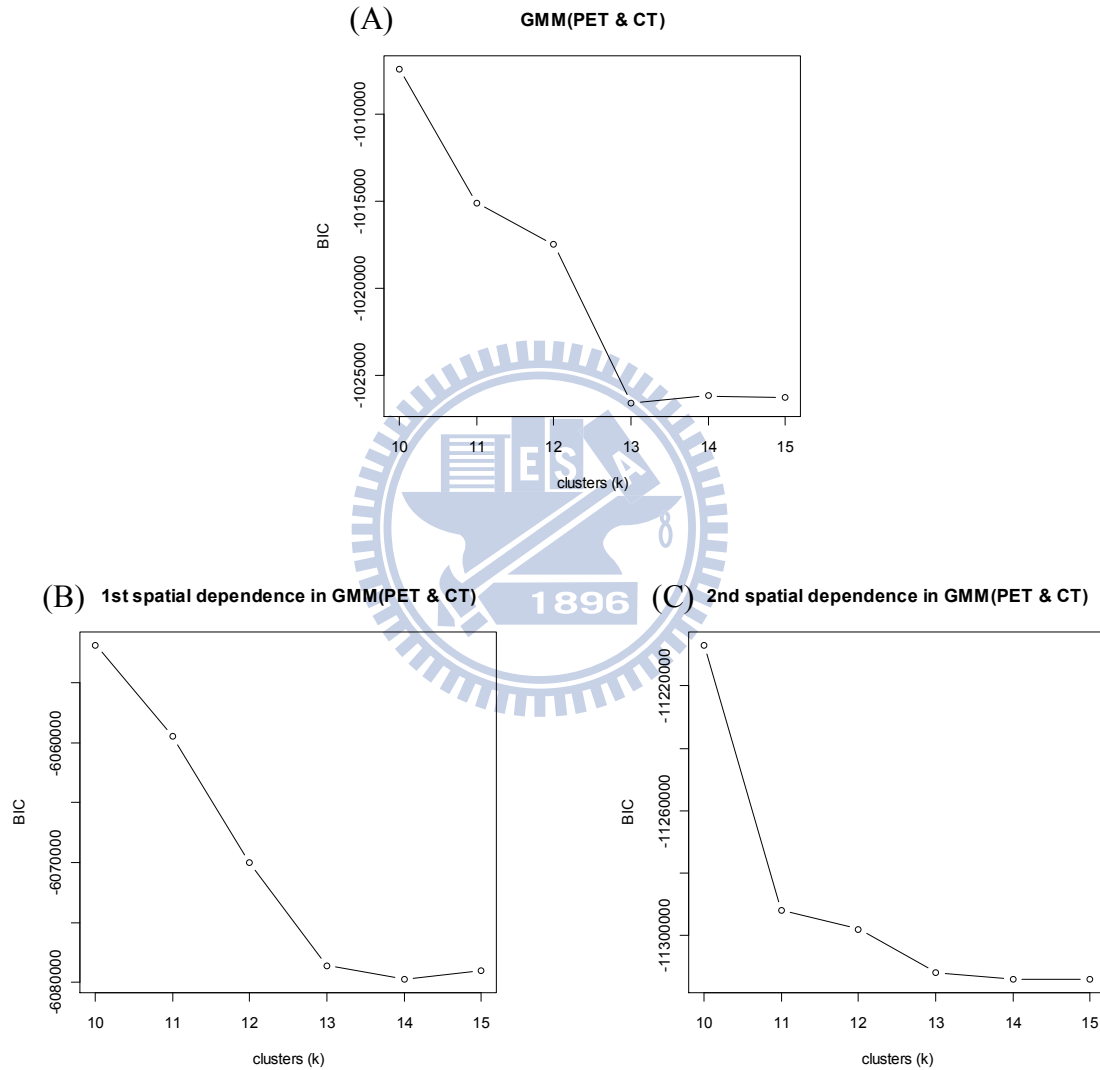


Figure 3.10: The BIC values of different clusters. (A) GMM(PET & CT). (B) 1st spatial dependence in GMM(PET & CT). (C) 2nd spatial dependence in GMM(PET & CT).

From Figure 3.10, the cluster sizes can be selected $k=13$ for GMM(PET & CT), and $k=14$ for the 1st and 2nd spatial dependence GMM. These clustering results and accuracy are shown as following.

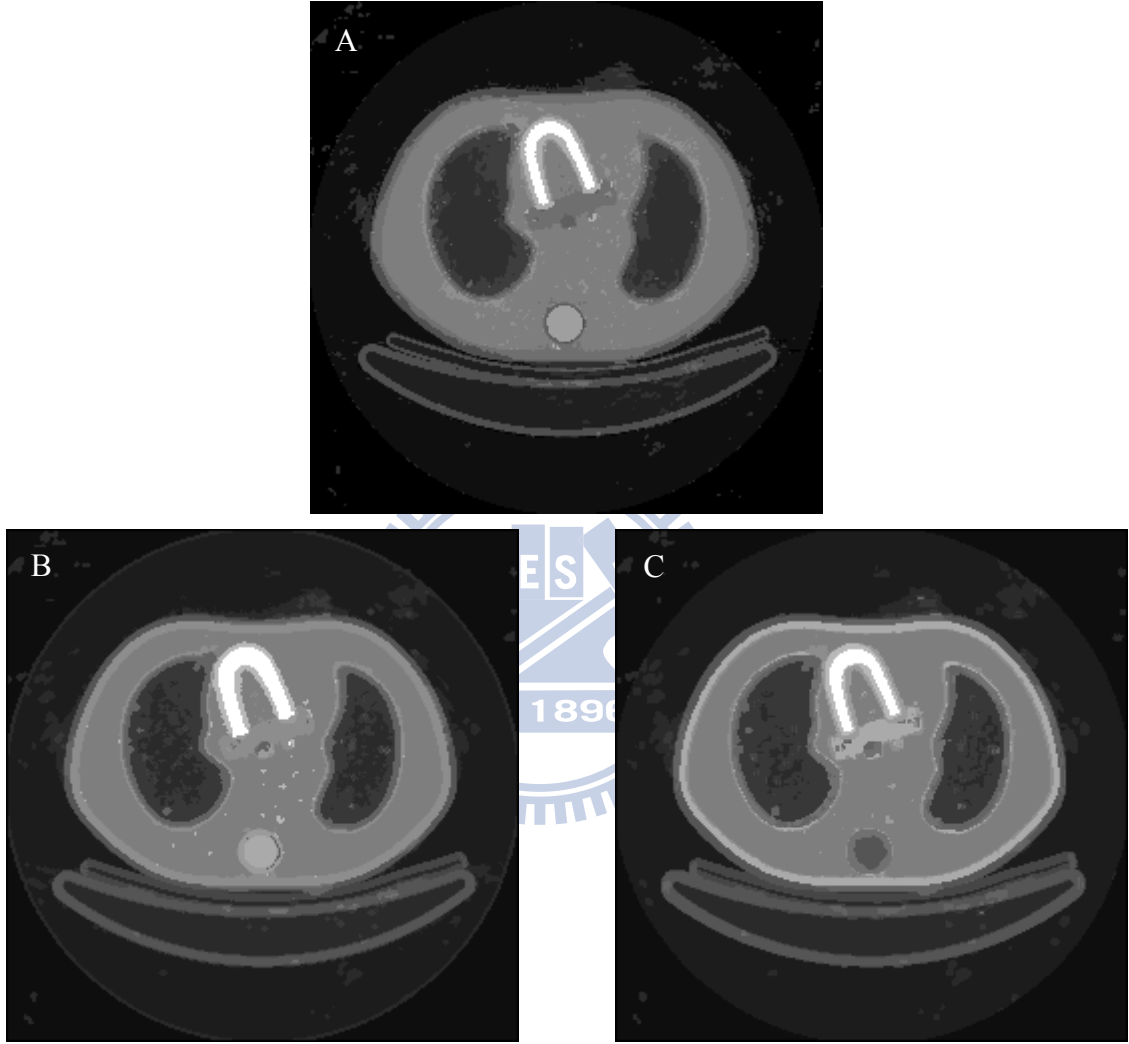


Figure 3.11: (A) $k=13$, the result of GMM (PET & CT). (B) $k=14$, the result of 1st spatial dependence GMM(PET & CT). (C) $k=14$, the result of 2nd spatial dependence GMM (PET & CT).

Table 3.1: Comparison for the accuracy of the highest activity region of different methods by F-measure.

| | Recall (TRR) | Precision (PPV) | F-measure |
|--------------------------------|--------------|-----------------|-----------|
| PET + GMM | 1.0000 | 0.5256 | 0.6891 |
| Fusion + GMM | 0.9682 | 0.9346 | 0.9511 |
| GMM (PET&CT) | 0.8962 | 0.9976 | 0.9442 |
| 1st order spatial GMM (PET&CT) | 1.0000 | 0.7307 | 0.8444 |
| 2nt order spatial GMM (PET&CT) | 0.9322 | 0.9692 | 0.9503 |

The comparison from Table 3.1 shows that the results of Fusion + GMM, GMM (PET&CT), 1st order spatial GMM (PET&CT) and 2nd order spatial GMM (PET&CT) all have higher accuracy than PET + GMM. Although GMM (PET&CT), the 1st and 2nd order spatial GMM (PET&CT) have slightly lower F-measure values than Fusion + GMM, they provided more information for the data. We can find the correlation of PET and CT on the regions of interesting. The correlation of the result of PET and CT could be used to detect the association pattern between the pixels of these two images. The correlation is positive as the area of a PET image has high (or low) isotope radioactivity and the area of a CT image has high (or low) X-ray absorption overlap very much. The correlation is negative as the area of a PET image with high (or low) isotope radioactivity and the area of a CT image with low (or high) X-ray absorption overlap very much. In addition, the correlations of neighbor points on PET and CT images also can be obtained by using spatial dependence in GMM. The following is some correlations of PET and CT on the result of using 2nd order spatial dependence in GMM.

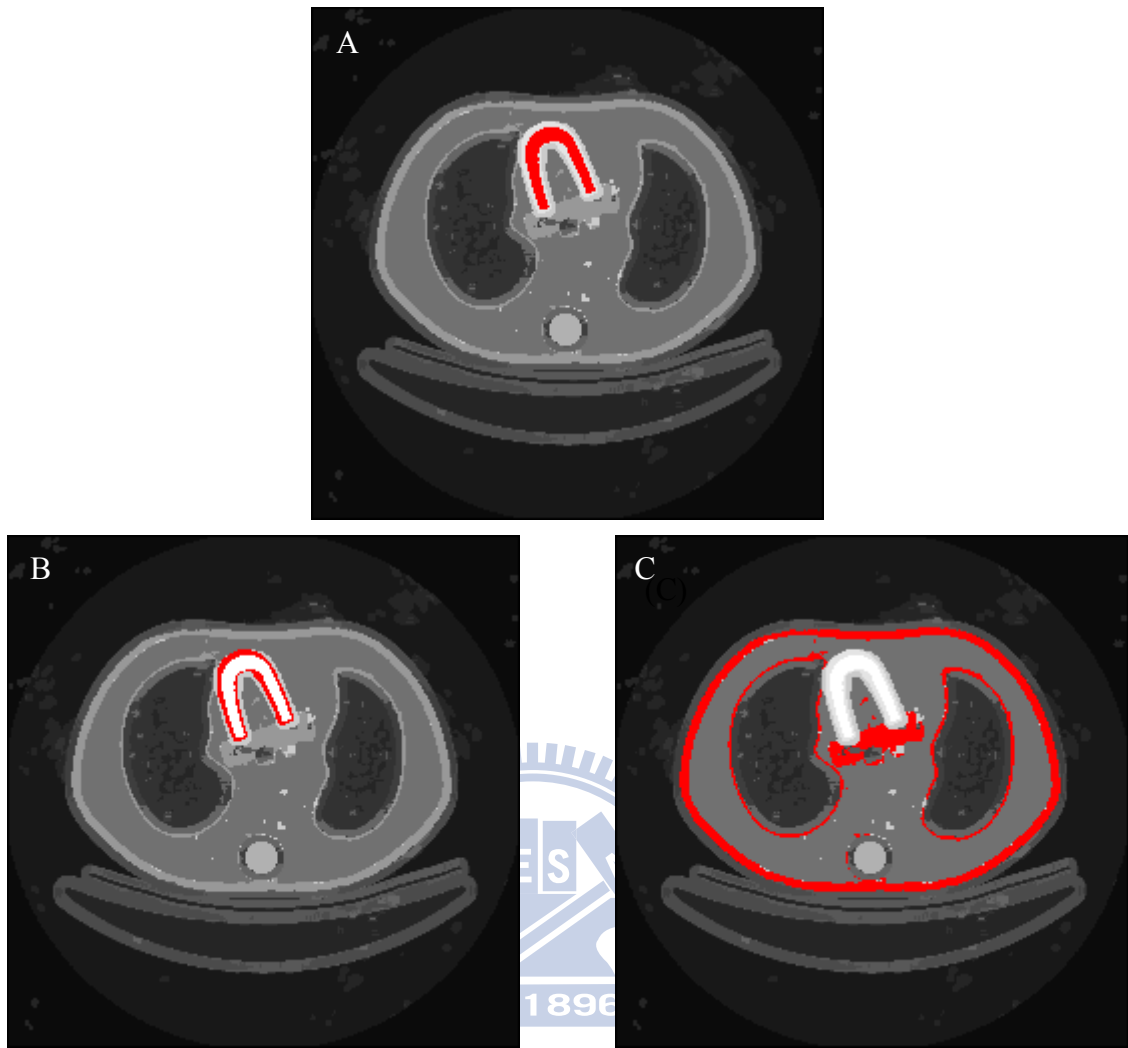


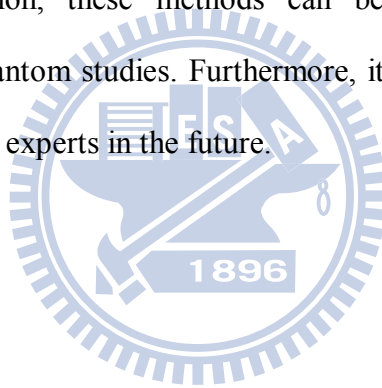
Figure 3.12: The red regions show different correlations of PET and CT on the result of using 2nd order spatial dependence in GMM. (A) The correlation is -0.7575. (B) The correlation is 0.3704. (C) The correlation is -0.7540.

From Figure 3.12, the correlations of (A) and (C) are negative. It represents that the area of the PET image with high (or low) isotope radioactivity and the area of the CT image with low (or high) X-ray absorption overlap very much.

Chapter 4 Conclusions

The information of CT images is used when segmenting PET images in this study. Besides segmenting fusion images of PET/CT and taking two-dimensional GMM fit the PET and CT data, we also used spatial dependence in GMM. These clustering results on the highest activity region all have better performance than the result of segmenting PET images only in Chapter 3. Besides the correlation of PET and CT, we also can find the correlations between neighbor points of PET and CT on the regions of interesting by using spatial dependence in GMM. Therefore, the model using spatial dependence is preferred in this phantom experiment since it provides more information for the data.

For further investigation, these methods can be tested for the stability of the performances with more phantom studies. Furthermore, it also could be applied in empirical study and judged by medical experts in the future.



Reference

- [1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, **19**, 716-723.
- [2] Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B*, **39**, 1-38.
- [3] Engel, J., Herrmann, E., Gasser, T. (1994). An iterative bandwidth selector for kernel estimation of densities and their derivatives. *Journal of Nonparametric Statistics*, **4**, 21-34.
- [4] Hsiao, I.T., Rangarajan, A., Gindi, G. (1998). Joint-Map Reconstruction/Segmentation for transmission Tomography Using Mixture-Model as Priors. *Proc. IEEE Nuclear Science Symposium and Medical Image Conference*, **3**, 1689-1693.
- [5] Chiang H.Y. (2004). Segmentation of Dynamic MicroPET Images by K-mean and Mixture Methods. Master's thesis, Institute of Statistics National Chiao Tung University.
- [6] McLachlan, G.J., Basford, K.E. (1988). *Mixture Model: Inference and Applications to Clustering*. Marcel Dekker, New York.
- [7] Ye M.C. (2009). Segmentation of PET/CT images by Flexible Mixture Models and Comparison with K-means and Gaussian Mixture Models. Master's thesis, Institute of Statistics National Chiao Tung University.
- [8] Olinger, J.M., Fessler, J.A. (1997). Positron emission tomography. *IEEE Signal Processing Magazine*, **14**, 43-55.
- [9] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*. **6**, 461-464.
- [10] Sheather, S.J., Jones, M.C. (1991). A reliable data-based bandwidth selection method for

kernel density estimation. *Journal of Royal Statistics Society. B*, **53**, 683-690.

- [11] Silverman, B. W. (1986) *Density Estimation*. Chapman and Hall, London.
- [12] Chen T.B. (2007). Statistical Applications of Maximized Likelihood Estimates with the Expectation-Maximization Algorithm for Reconstruction and Segmentation of MicroPET and Spotted Microarray Images. Ph.D. Dissertation, Institute of Statistics National Chiao Tung University.
- [13] Townsend, D. W., Carney, J.P., Yap, J.T, Hall, N. C. (2004). PET/CT Today and Tomorrow. *The Journal of Nuclear Medicine*, **45**, 4S-14S.
- [14] Van Rijsbergen, C. J. (1979). *Information Retrieval*. London. 2nd Edition. Butterworth. London, England.
- [15] Vardi, Y., Shepp, La., Kaufman, L. (1985). A statistical model for positron emission tomography. *Journal of the American Statistics Association*, **80**, 8-20.
- [16] Wong, K.P., Feng D., Meikle, S.R., Fulham, M.J. (2002). Segmentation of Dynamic PET Images Using Cluster Analysis. *IEEE Transactions on Nuclear Science*, **49**, 200-207.