

基因微陣列影像分析之研究

薛 義 誠*

楊 勝 琦**

郭 建 明**

摘要

生物晶片 - 基因微陣列是一種新興且能夠同時偵測分析千萬點基因是否有表現的工具。基因微陣列的影像擷取與分析辨識是進行後續實驗分析的第一個步驟，也是不可或缺的重要步驟。分析一張含有萬點基因表現點的高密度基因微陣列影像時，如何將不同形狀或是歪斜位置的基因表現點從基因微陣列影像中正確的切割出來，對於往後資料分析的精確度會直接影響到實驗的最後結果。因此本研究之著重在於藉由理想的影像邊緣偵測演算法（Canny Edge Detection Algorithm）從一張基因微陣列影像中正確的切割出每一個基因表現點的位置。我們並以史丹佛大學基因微陣列資料庫中的基因微陣列影像做為本研究的基因表現點切割測試資料，以本研究建立自初步基因表現點方格劃分、偵測基因表現點邊緣、動態修正基因表現點方格功能，測試找尋及分析比較基因微陣列影像上面的基因表現點，證明可提供一個良好的基因微陣列影像分析工具，可提高後續資料分析的精確度。

關鍵詞：生物晶片、基因微陣列、基因微陣列影像、基因表現點切割

* 國立中央大學 資訊管理所教授

**國立中央大學 資訊管理所研究生

The Microarray Image Process and Analysis

Yih-Chearng Shiue*

Sheng-Chi Yang**

Gen-Ming Guo**

Abstract

Biochip – Microarray is a new tool applied to examine expression levels for thousands of genes simultaneously. In the operation of the microarray experiment data analysis, both image acquisition and analysis are the first step and they are one of the most important processes. When biologists analyze a high density microarray image with thousands of gene spots, segmenting the gene spot accurately from irregular shapes and crooked position may affect the final result of bio-experiment greatly. In this paper, we are focused on using an ideal edge detection algorithm – Canny edge detection algorithm in order to segment every spot accurately. The Microarray images obtained from the Stanford Microarray Database were taken as the testing samples. Through the developed system with grid、spot edge detection and dynamic gridding functions in this study, the system is verified and proved to be able to provide a good method for each spot's segmentation in a biochip image.

Key Words : Biochip、Biochip Image、Microarray、Spot Segmentation

* Professor, Department of Information Management, National Central University

** Graduate Student, Department of Information Management, National Central University

壹、緒論

基因為一切生命現象的基礎，回顧過去幾十年來應用於篩選差異表現基因的方法，例如差異性篩選法（Sargent, 1987）、蛋白質序列推衍法（Kishimoto, 1987）、免疫篩選法（Hogervorst, 1990）、刪減篩選法、演化保存區域篩選法（Suzuki, 1990）及差異性顯示法（Liang, 1992；Liang, 1993）。這些方法或可單獨進行或可結合其他實驗步驟以使篩選基因更加有效率，但仍須耗費大量人力、物力及時間。近年由於人類基因組計劃（Human Genome Project）的效應，迅速且多量的基因功能分析技術便應運而生，因而帶動了全球生物晶片的蓬勃發展，幾項新發展且具前瞻性的生物晶片技術，如基因微陣列（Microarray）、基因晶片（DNA chip）、蛋白質晶片（Protein array）、組織晶片（Tissue array）及實驗室晶片（Lab-on-a-chip）等，正加速瞭解許多基因的功能及其交互作用與調控。

而所有基因微陣列影像分析的研究探討幾乎都將焦點放置於怎麼去統計分析所得到的基因微陣列影像資料，例如如何利用評估每張片子上面基因表現點和其背景的差異而得到每個基因表現點的顏色濃度對比等等時，一個所必須思考的問題就是在分析一張含有萬點表現點以上的基因微陣列影像時，如何將不同形狀或是歪斜位置的基因表現點從基因微陣列影像中正確的切割出來，對於往後資料分析的精確度的影響是值得注意的（Carlos, 2004），而這個步驟的品質往往都讓人給忽略了。對於從事生物晶片-基因微陣列實驗者而言，在基因微陣列實驗步驟中，經由雜交後所掃描出來的基因微陣列影像中充滿著許多雜訊（Noise）以及呈現不規則分佈的基因表現點（Spot shape）暈開的部分，而所呈現出來的大小（Spot size）、基因表現點位置（Spot position）也都會因為當初實驗流程中品質控制的不好而讓後續的實驗分析有所誤差，甚至是整張基因微陣列影像無法辨認，尤其是在高密度點的基因微陣列中。

以前在分析基因微陣列影像時，往往都是將焦點放在如何去偵測基因表現點的背景訊號以及雜訊的擷取，而不是將每一個影像中的基因表現點分割出

來。而每張影像裡的每一個像素點（Pixel）的值是代表著此點顏色濃度的基本單位，而每一個基因表現點影像裡的濃度變化對於其資料分析的量化有著直接的影響（Brown, 2001），所以將每一個基因表現點正確的切割出來對於分析基因表現點裡的不同螢光劑濃度的資料精確度能較正確的提高。Hirata (2001)等學者曾利用數學的架構提出一個切割基因微陣列影像上面基因表現點位置的演算法，然而，其對於影像圖片上面基因表現點的半徑大小以及基因表現點之間的相對位置必須靠手動的方法修正才能夠正確的將基因表現點的位置切割出來。Liew (2003)等人也會提出一個方法來切割出基因微陣列影像上的基因表現點，但是其方法對於辨識基因微陣列影像上的基因表現點有著許多的限制，像是影像上面的表現點的大小必須適當，而且此方法是靠偵測表現點本身所帶有的顏色濃度來找出表現點的正確位址。由於此項限制，所以只能辨識出帶有非常高的濃度的顏色變化的基因表現點的外形並加以切割。所以，在許多數學統計或是電腦影像圖學研究的方法論中，已經有許多關於微陣列影像上基因表現點切割的相關研究及方法論，其中也有些改良的方法論來適用於微陣列影片的分析，但大部分都只能夠切割出近似的基因表現點外形位置，但是不能精確的切割出基因表現點位置。

在本論文研究中，我們致力於開發一個將電腦影像處理技術應用於生物晶片基因微陣列影像分析的系統，利用一個理想的影像邊緣偵測的演算法（Canny Edge Detection Algorithm）來用於基因微陣列影像上基因表現點位置切割的找尋，此方法首先設計一個邊緣濾波器來平滑化影像上的雜訊，然後計算影像中具有高梯度值的區域再向前追蹤並運用兩個門檻值過濾判斷每一個表現點的區域邊緣。最後在本研究的基因微陣列影像分析流程，可將每張微陣列影像上面的基因表現點的位置正確的切割出來。我們所要強調的地方有兩點：一、初步劃分方格內的基因表現點的輪廓並透過演算法標示出來。二、基因微陣列影像中基因表現點所在的方格位置動態調整的切割出來。要達到上述的目的，需要用到一些傳統的影像處理技巧，如影像平滑處理，色彩轉換空間理論等。配合

本研究的分析流程和影像處理的演算法設計出一個針對基因微陣列影像內基因表現點偵測的系統，以對於分析基因微陣列影像的貢獻能有不錯的表現。

貳、系統設計與實作

以下將介紹本研究所設計的系統的影像分析的四個流程，一、色彩模式轉換、二、初步基因表現點方格劃分、三、邊緣偵測演算法，及四、動態修正基因表現點方格切割；並說明每一個流程所採取的方法及演算法。在基因微陣列的影像的分析中，通常會嘗試各種方法來偵測影像裡所包含的上萬個基因表現點，例如利用顏色，濃度變化或是統計分析來粹取特徵，透過不同的分析流程和演算法來分析經由雜交實驗後所產生的基因微陣列影像。在本篇論文中，我們採取灰階濃度的變化以及邊緣偵測的演算法將基因表現點標示出來，以提供基因表現點的訊號截取而達到其目地。

一、色彩模式轉換

一張原始尚未經過處理的基因微陣列影像圖片主要是由兩種色彩組成：使用 Cy3 所呈現綠色反應的螢光染劑以及使用 Cy5 所呈現紅色反應的螢光染劑。而做為計算基因微陣列影像上面每一點像素的色飽和度則是由這兩種顏色所混合組成，所以一張基因微陣列影像圖片所呈現的各種不同色光基本上都是由三種基本色光，紅（Red）、藍（Blue）、綠（Green）所混合組成，也就是所謂的 RGB 三原色。

在本研究中，為了讓後續的基因表現點方格劃分及基因表現點邊緣偵測的處理中有效的避免因為顏色濃度的變化而影響到其準確度。必須先對基因微陣列影像作色彩模式轉換。YIQ 色彩模式為用在工業色彩電視廣播中，其目的是提高轉移效率並保持和單色電視台的相容性，YIQ 為 RGB 的一個紀錄，Y 對應於灰階亮度，而 I、Q 是兩個彩色分量。YIQ 彩色模式的優點是亮度（Y）和彩色資訊（I 和 Q）分離，不會因為彩度而干擾到灰階的亮度。因而本研究將其 RGB 色彩模式轉換成 YIQ 色彩模式，利用 Y 值得到所謂的灰階影像，而 RGB

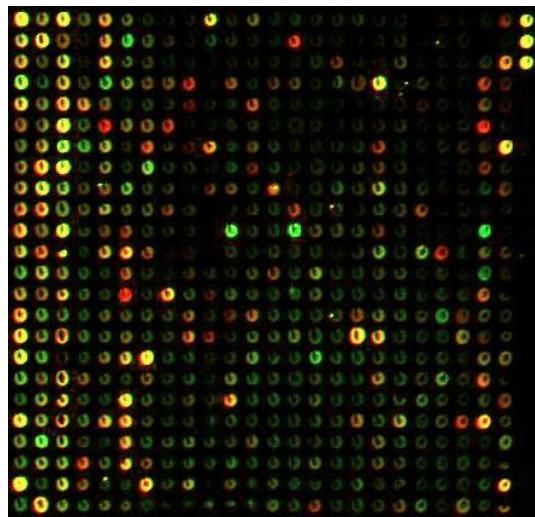
和 NTSC 協會(National Television System Committee)制定的 YIQ 色彩模式(鐘國亮, 2002) 存在着下列算式轉換：

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.528 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

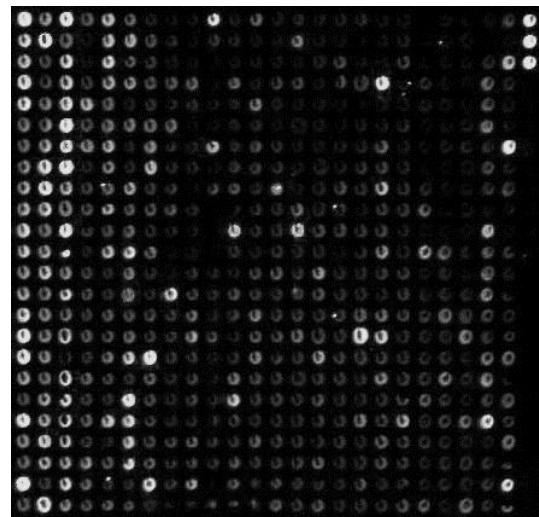
根據上述的關係，可以將 RGB 的色彩影像轉換成 YIQ 後以 Y 值來表示灰階影像(Gray Image)，而其轉換的公式如下：

$$Y(x, y) = 0.299 * P_R(x, y) + 0.587 * P_G(x, y) + 0.114 * P_B(x, y) \quad (2)$$

其中 $Y(x, y)$ 代表在 $P(x, y)$ 點轉換後所對應的灰階值。 $P_R(x, y)$ 、 $P_G(x, y)$ 、 $P_B(x, y)$ 分別代表著此點個別的 R、G、B 值，而在基因微陣列影像中， $P_B(x, y)$ 的值為 0，而圖 1 即為我們所轉換的結果。



RGB 彩色微陣列影像



Y 值灰階微陣列影像

圖 1 利用 YIQ 色彩模式將 RGB 影像轉成 Y 值灰階影像

二、初步基因表現點方格劃分

初步的基因表現點方格劃分的想法是先在一張基因微陣列影像中把所有的表現點約略的位置先劃分出來。此步驟假想基因微陣列上面所處的每一個基因表現點的相對位置都相同且整齊的排列在雜交過後的微陣列影片上，那麼所有的基因表現點皆擁有著相同空間位置交錯於基因微陣列影片上面。因此利用交錯的直線所形成的方格將基因表現點的範圍約略的包圍起來，初步的基因表現點方格的劃分操作可以先將每一個表現點近似的邊界切割出來，這一個步驟提供了我們一個初步的基因表現點的外形，以便在下一個步驟中可以更準確的找出每一個基因表現點的邊界。

要達到初步切出基因表現點方格並將每一個表現點約略的包圍起來，必須先在基因微陣列影像上面尋找出切割直線所要開始的切割點。本研究利用電腦影像處裡中空間域影像特徵表示的概念（連國珍, 2001），而所謂比較輸入基因微陣列影像的灰度值即可偵測出我們在影像中所要尋找的像素點：空間域影像的性質主要表現在每一個像素點灰度的變化，灰度的分佈以及相鄰點灰度的總合。所以，影像內的動態分佈的變化是空間影像中的任何一條線的灰度變化，而此灰度變化為一維的信號分佈。在此先將單一基因表現點放大從它的一維灰度變化看起，如圖 2 及表 1 所示：

圖 2 單一基因表現點放大

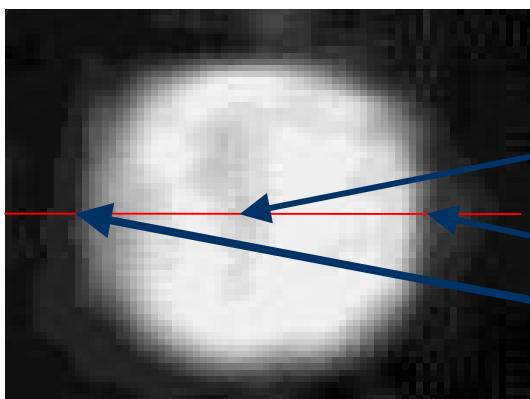


表 1 一維的灰度變化分佈

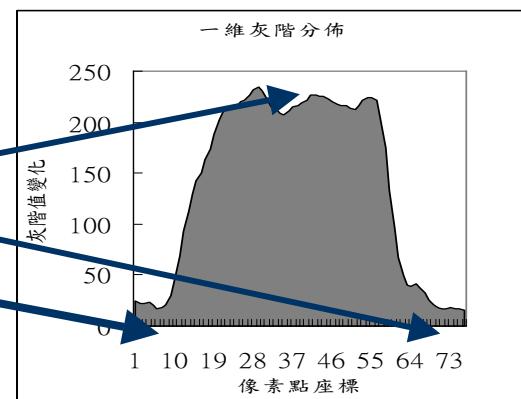


圖 2 為單一基因表現點放大的圖片，沿著圖中的紅線動態的掃描其所經過每一個像素點所表示的灰階值，並且將其紀錄下來繪成表 1 得到的一維的灰階分佈。藉由統計著其灰階分佈的變化發現其形成一個往上攀升而後下降的山峰圖，對照著圖 2 發現灰階濃度最高的中心像素座標點對應到表 1 為其峰頂，而基因表現點兩側的黑色濃度較低的座標點則對應到表 1 的谷底。再看下圖 3 其兩個相鄰的基因表現點所對應到的表 2 的一維灰度變化的分佈圖：

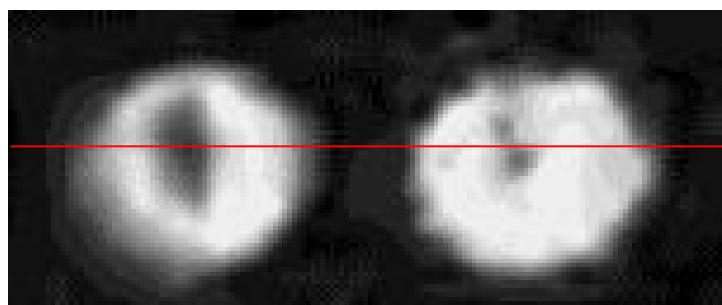
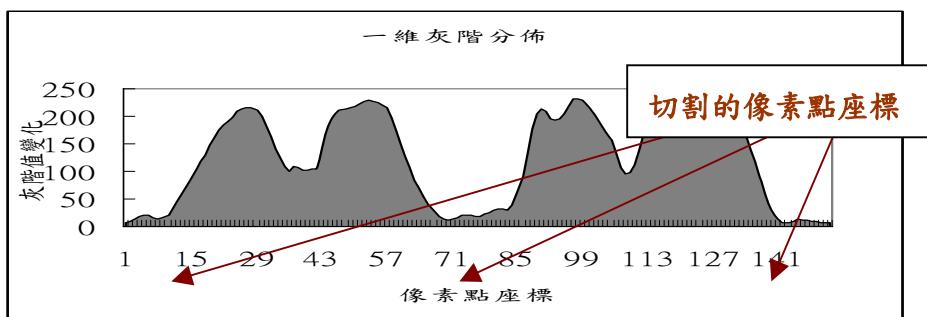


圖 3 兩個相鄰的基因表現點

表 2 兩個相鄰的基因表現點一維灰階分佈



再一次由圖 3 的兩個相鄰的基因表現點及其對應表 2 之一維灰階分佈可以看起來相鄰的兩個基因表現點就像是兩個相鄰的兩山峰，山峰與山峰中間的波谷對應著圖 3 中兩基因表現點相鄰的中心位置。所以根據上述的論述，可將每一個最低點的峰谷的像素點座標當作切割像素點（Grid point），將其串聯起來作直線的劃分，便能做出最初的基因表現點的方格劃分。

所以本研究分別由橫向和縱向的欲切割的基因維陣列影像動態的掃描一遍，找尋所有的最低點山谷所代表的切割像素點，而這些切割像素點分別代表著其縱軸與橫軸的平均最低的灰階值，最後將所有找到的切割像素點交錯的連接起來，這些交錯的直線便形成了基因維陣列影像的初步基因表現點方格的劃分。圖 4 所示為初步的基因表現點方格劃分的結果，在此結果中可以看到，初步的基因表現點的方格切割約略的將基因微陣列影像中每一個基因表現點的位置劃分出來，當知道了這些基因表現點約略的位置後便分別開始對方格內的基因表現點做邊緣偵測，將其基因表現點的輪廓給標示出來。

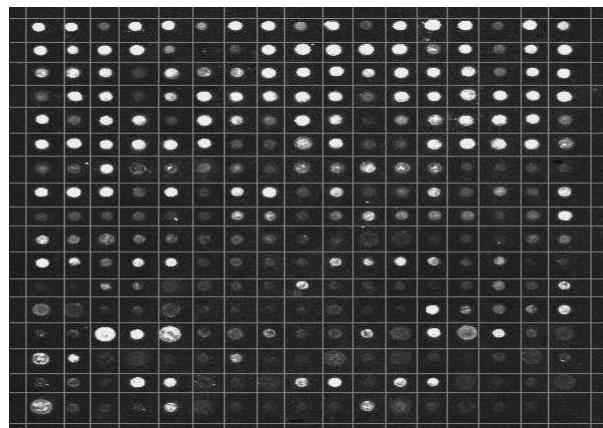


圖 4. 初步的基因表現點方格切割結果

三、邊緣偵測演算法將基因表現點的輪廓標示出來

經由上一步驟的初步的基因表現點方格劃分，將每一個基因表現點在整張基因微陣列影像中的約略位置已經劃歸於方格框框內後，為了要更正確定位出方格內部所屬的基因表現點輪廓與位置，一個定位基因表現點輪廓的演算法是所被需要的。在本研究中，採用一種最佳化的邊緣偵測演算法將基因表現點的輪廓給定位出來。首先將整張影像用高斯平滑法來做旋積，然後用 Sobel 運算子的梯度運算將被平滑過後的影像中明顯的區域給突顯出來，那麼邊緣便會被標示於明顯區域的突出點，於是此演算法便開始沿著這些最高點追蹤並且忽略

那些不是位於最高點的點，且在追蹤的過程中依據著兩個門檻值： T_H 和 T_L 且 $T_H > T_L$ ，從大於 T_H 值的點開始持續追蹤直到遇到了一個低於 T_L 的點後便停止，這兩個門檻值的限定是確保不必要的邊緣不會擾亂到邊緣偵測的結果。這種邊緣追蹤的模式是一種叫做 Canny 邊緣偵測的模式(Canny, 1986)，關於 canny 邊緣偵測模式的演算法步驟（其中包括了高斯平滑法、Sobel operator）將在後面的小節中作說明。

(一) 高斯平滑法

高斯平滑法 (Gaussian Smoothing) 是用一個類似二維矩陣的遮罩 (mask) 來做旋積運算將影像平滑化，本研究首先用算式 3 高斯 2-D 算式以及他所呈現的分佈圖（圖 5）來帶出我們所要訂定的遮罩：

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

σ 為分佈的標準誤差值，我們通常都假設此分佈包含著通過零點的中心 ($x=0$ 的中心線)，而其 2-D 分佈圖如圖 5 所示。一張影像是由 $n \times n$ 個像素點所展現而成的，所以在開始要做旋積之前，必須先找出一個適合的高斯方程式來展現本研究所需的遮罩，而此概念是由算式 3 及圖 5 所衍生出來的一個叫做點的延展 (Point-Spread) 的觀念來產生本研究旋積所需的遮罩。理論上，高斯分佈是無窮的，所以需要一個無限大的旋積遮罩才能完全的處理，但事實上是不可能的，綜合以上，我們截斷理論上無限大的旋積遮罩成一個 5×5 遮罩再運用點的延展的概念來對影像作旋積，圖 6 為我們所選取的 5×5 的遮罩：

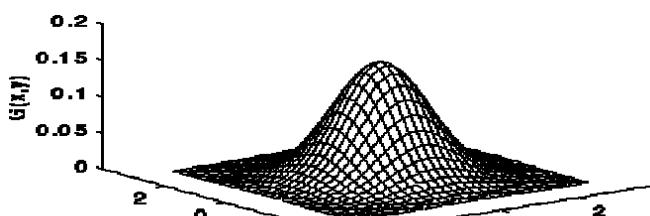


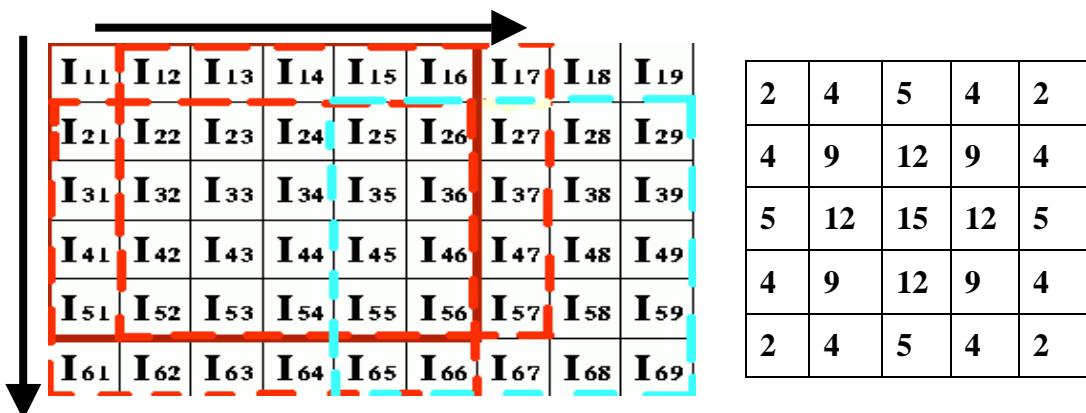
圖 5 2-D 在 mean $(0,0)$ 與 $\sigma=1$ 時的高斯分佈

2	4	5	4	2
4	9	12	9	4
5	12	15	12	5
4	9	12	9	4
2	4	5	4	2

$\frac{1}{115}$

圖 6 5x5 的遮罩且 $\sigma=1.4$

根據高斯方程式將遮罩決定後，便可展開對影像作旋積的處理來達到平滑的效果，分別從 X 軸和 Y 軸開始對影像作旋積平滑，在此以一個例子（如圖 7）來說旋積的運作：

圖 7 一個 9x6 的影像（左）和本研究所訂的 5x5 遮罩（右）， I_n 顯示每一個像素點的值

旋積平滑開始於將遮罩放置於輸入影像的左上角，之後分別由 X 軸和 Y 軸開始作旋積，而遮罩滑過的每個像素點的值將會經過重新運算而得到一個新的值重新存入影像中，舉例來說， I_{25} 的值將會被輸出為：

$$O_{25} = I_{25} * 2 + I_{26} * 4 + I_{27} * 5 + I_{28} * 4 + I_{29} * 2 + \dots + I_{65} * 2 + I_{66} * 4 + I_{67} * 5 + I_{68} * 4 + I_{69} * 2 \quad (4)$$

而對於一張 $M \times N$ 大小的基因微陣列影像與本研究所訂定的遮罩而言，可利用下列的公式作高斯平滑法的旋積：

$$O(x,y) = \sum_{k=1}^5 \sum_{l=1}^5 I(x+k-1, y+l-1) K(k,l) \quad (5)$$

(二) Sobel 運算子

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

圖 8 Sobel 運算子，包含 G_x 和 G_y

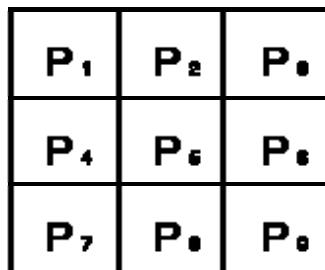


圖 9 3×3 影像區域

G_x 及 G_y 分別負責檢知 X 與 Y 方向的邊緣變化，由圖 8， ΣG_x 與 ΣG_y 皆為零，所以如果一個點其相鄰的八個點都是相同值，表示沒有邊緣變化，其值也會是零。如果左右的值不同， G_x 輸出結果就不會是零，但 G_y 仍會是零，直到上下的值不同為止。對於每一點的 G_x 與 G_y 計算結果會累加起來存於另一張圖上，以確保兩個方向的邊緣變化都會被找到。基於此運算子，本研究得到一個梯度值：

$$|G| = |(P1 + 2 \times P2 + P3) - (P7 + 2 \times P8 + P9)| + |(P3 + 2 \times P6 + P9) - (P1 + 2 \times P4 + P7)| \quad (6)$$

重複上述計算過程直到所有可能位置都執行過後，就得到一個和原圖大小相同的梯度圖，接著開始確認哪些像素是邊緣。最簡單的方法就是取一個門檻值，也就是說當梯度值大於這個門檻值就判定它為邊緣，但是在基因微陣列的影像中，存在著因為雜交和清洗時遺留在影像上面的細微雜訊，為了避免這些細微雜訊的誤導，而使用了兩個門檻值來過濾而得到更正確的基因表現點的邊緣。

(三) Canny 邊緣偵測模式

關於 Canny 邊緣偵測模式運作，當輸入一張基因微陣列影像後，先利用高斯平滑法將影像平滑化，然後利用 Sobel 運算子的梯度運算將可能是邊緣的區域標示出來，之後沿著這些區域開始追蹤每一個像素點強度，並利用兩個門檻值做為過濾的標準，如果此像素點的強度低於第一個 T_L 門檻值，便將之設為零（不設為邊緣），如果強度大於第二個 T_H 門檻值的時候，我們便將他設為我們所要尋找的邊，但是如果強度介於第一個門檻值和第二個門檻值的中間時，除非有另外一個像素點和此點相鄰且其門檻值大於或等於 T_H 時，就將此像素點也設為零。此模式總共有五個程序：(1) 使用 5×5 遮罩對影像作高斯平滑化（步驟一）、(2) 使用 Sobel 運算子找出影像梯度（步驟二）、(3) 找出檢測邊緣的方位梯度（步驟三）、(4) 挑選邊緣追蹤點到邊緣的指派（步驟四&步驟五）、(5) 利用兩個門檻值判定出最後的邊緣（步驟六）。以下為 canny 邊緣偵測模式的步驟說明：

在步驟一中，利用本研究決定的 5×5 遮罩對基因微陣列影像做平滑，步驟一主要是將影像平滑，在開始定位和偵測任何影像的邊緣時，可以先將影像中一些明顯的雜訊給過濾掉。在本研究中，使用圖 6 的 5×5 遮罩來作為過濾器，一但遮罩決定後，運用高斯平滑法對影像做旋積，當然遮罩相對於整張影像而言是非常小的，所以我們一邊滑動遮罩，一邊對於所掃過的像素點做運算而得到新的值，改變整張影像而得到平滑效果。圖 10 為所示的一個案例。

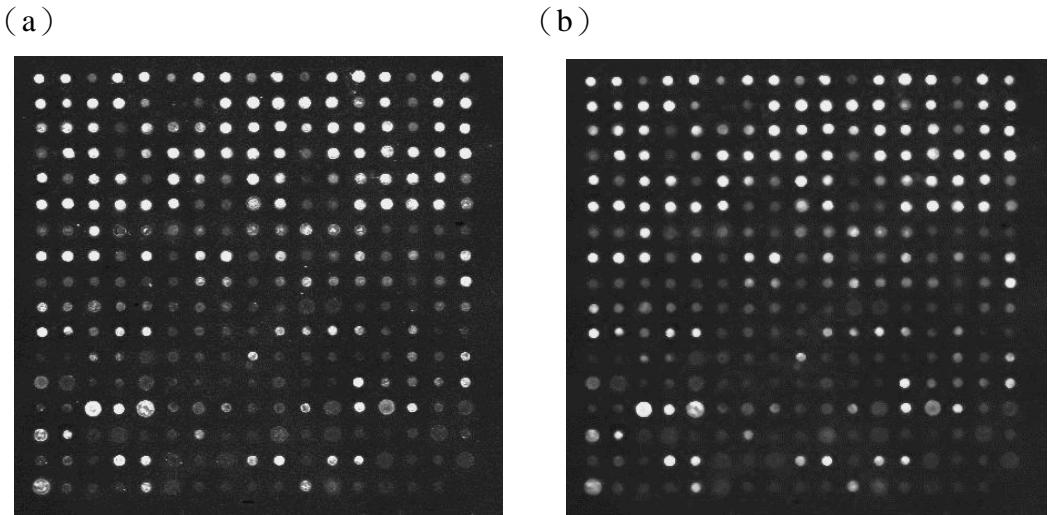


圖 10 高斯平滑過濾雜訊 (a) 原始影像 (b) 平滑後影像
(案例中將基因表現點方格線拿掉以清楚其平滑結果)

在步驟二中，利用高斯平滑法將影像過濾後，接著對影像作梯度運算來找出邊緣的強度變化。並運用 Sobel 運算子對影像展現一個二維的梯度測量，並使用圖 8 的兩個 3×3 遮罩 Gx 及 Gy 分別對於 X 軸方向（橫軸）以及 Y 軸方向（縱軸）作旋積，分別檢測 X 和 Y 方向的邊緣變化，以確保兩個方向的邊緣變化都會被找到，重複上述計算過程直到所有可能位置都執行過後，會得到 Gx 和 Gy 兩個不同方向所執行的梯度運算的結果，之後再利用式子 7 將兩個結果合併便得到完整的整個影像梯度。

$$|G| = |Gx| + |Gy| \quad (7)$$

在步驟三中，一但步驟二中的邊緣梯度被確定後，步驟三開始決定邊緣的方位。公式 8 為邊緣方位計算的公式，取 Gx 和 Gy 的正切（tangent）值便可以得到每一個像素點的方位值。在計算每一個點的方位值的同時，需考慮到一個例外的狀況，就是在求 Gx 通過零點的正切時其值會因為是無限大而產生錯誤，必須依 Gy 值來決定設為 90° 或是 0° ，當 Gy 也為零時，其方位設為 0° ，不為零則設為 90° 。

$$\theta = \tan^{-1} \left(\frac{Gy}{Gx} \right) \quad (8)$$

在步驟四中，一但邊緣的方位都建立後，接著下一步就是把找到的邊緣的方位都相關聯起來可以依照關聯的方向來做邊緣追蹤。要如何達到這步驟，在此以 5×5 個像素點排成的影像作說明，從圖 11 的 a 看起，對 a 這像素點來說，環顧四週有四種可能的方位，依序為 0° (水平方位)、 45° (沿著正相的對角線)、 90° (垂直面) 以及 135° (沿著負相的對角線)。所以現在邊緣的方向可依據相關聯的方位由此被這四個方位來定位 (例如：如果發現一個方位為 3° ，那麼把他設為 0° 水平方位)。

x	x	x	x	x
x	x	x	x	x
x	x	a	x	x
x	x	x	x	x
x	x	x	x	x

圖 11 5×5 個像素點排成的影像

再從圖 12 (Canny, 1986) 的一個半圓分成五個方位區域來看會更清楚，任何一個邊緣的方位落入 ($0\sim 22.5$ & $157.5\sim 180$) 的區域都會被設為 0° ，任何落入 ($22.5\sim 67.5$) 區域的邊緣方位將會被設為 45° ，任何落入 ($67.5\sim 112.5$) 區域的邊緣方位將會被設為 90° ，而最後，任何落入 ($112.5\sim 157.5$) 區域的邊緣方位將會被設為 135° 。



圖 12. 半圓的五個方位

在步驟五中，在邊緣方位確定後，便開始選擇從哪一點開始追蹤，在追蹤之前必須決定開始的追蹤點，再沿著這些確定的邊緣方位作追蹤並利用門檻值將那些不被考慮為邊緣的像素值設為 0 並忽略之。

在步驟六中，最後，要確保不必要的雜訊邊緣不會擾亂到邊緣偵測的結果，藉由所訂定的兩個門檻值 T_H 和 T_L 來決定（在此本研究定 $T_H = 120$, $T_L = 40$ ）。任何像素點的強度值比 T_H 大的時候，就認為此像素點是構成所要找的邊緣，並將之標示出來，且任何與其相關的像素點而其的像素值雖然小於 T_H 但是大於 T_L ，則亦先認定是構成邊緣的像素點，並將之標示出來，在延展到該點的相鄰像素點偵測，如此下去一直找到一個像素點的值介於 T_H 和 T_L 之間且其和其相關的像素點值小於 T_L 時，那麼此邊緣的延展找尋才會停止。

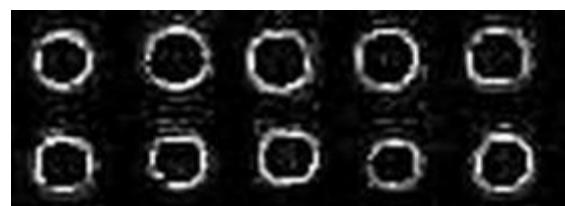


圖 13 Canny 邊緣偵測模式所找出的基因表現點邊緣

四、動態修正基因表現點方格切割

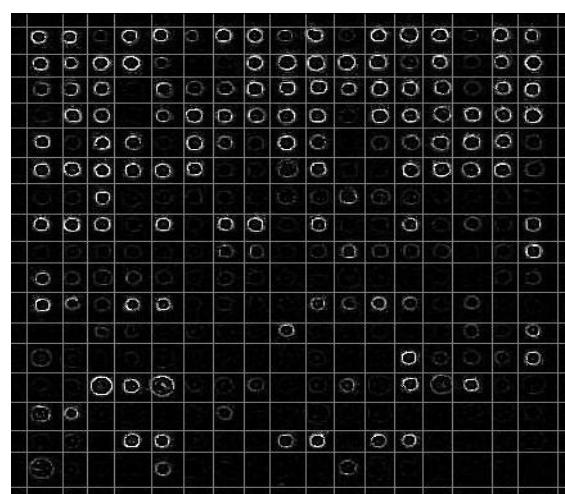


圖 14 基因表現點邊緣偵測+初步基因表現點切割

在做完基因表現點邊緣的偵測後，最後便開始動態調整基因表現點所在的地方。如圖 14 所示，在開始的基因表現點方格的初步劃分中可以看的出來，有些方格的格線雖然將所偵測出來的表現點劃分出來了，但是有些表現點的邊緣超出了所劃分出來的地方，因此，便需開始動態調整初步的方格劃分而將正確的基因表現點所在位置給切割出來。

藉由本研究的概念來實踐調整的方法，亦即藉由格線的動態移動直到和表現點的邊緣有交集的時候便停止，藉由上下的格線的動態調整，劃分出基因表現點位置，以提高基因表現點資料分析的正確性（如圖 15 所示）。

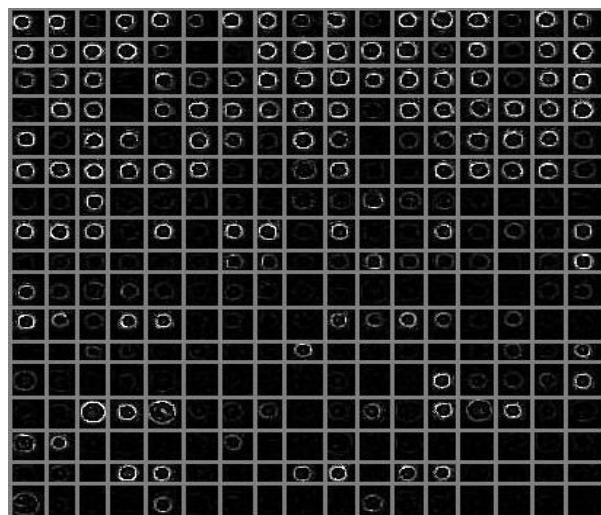


圖 15. 動態的調整基因表現點方格結果

參、實証與分析

一、系統介面瀏覽

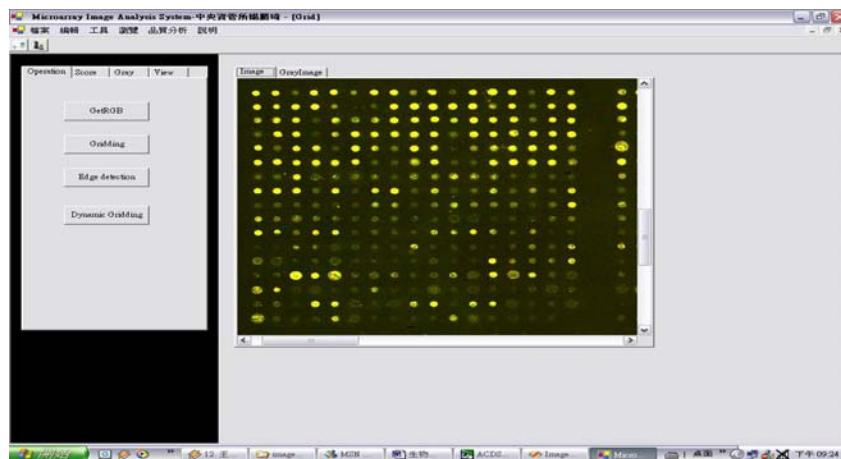


圖 16 基因微陣列影像分析系統介面

本研究使用 Microsoft .Net 為開發系統的平台，開發「電腦影像處理技術應用於生物晶片基因微陣列影像分析」的系統，利用影像邊緣偵測模式的演算法來用於基因微陣列影像上基因表現點位置切割的找尋。首先先建立一個簡單方便的操作介面，當開啓左上角工具列上的“檔案”選項將所要分析的基因微陣列影像載入系統中的子畫面。如圖 16 所示，左邊半部為主要的分析操作介面，分別透過基因表現點 RGB 的攝取，初步的基因表現點方格切割，基因表現點邊緣偵測以及最後的動態調整基因表現點方格的功能得到本研究所要的最後結果（如圖 17）。

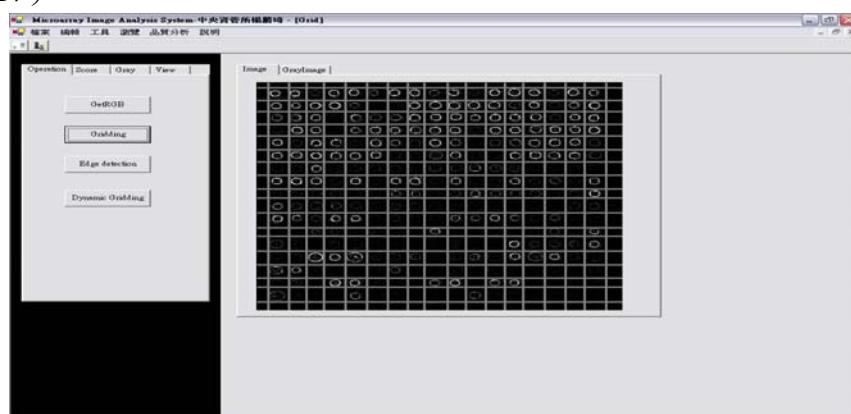


圖 17. 基因微陣列影像分析結果畫面

二、實驗結果

本研究測試的基因微陣列影像係由 Stanford 大學的基因微陣列資料庫 (Stanford, 2004) 中取得。每一張完整的基因微陣列影像包含著數個區塊 (Sector)，而每一個區塊包含著我們所關心的基因表現點。先對每一個區塊作手動的切割之後再對切出來的區塊放入本研究系統分析。

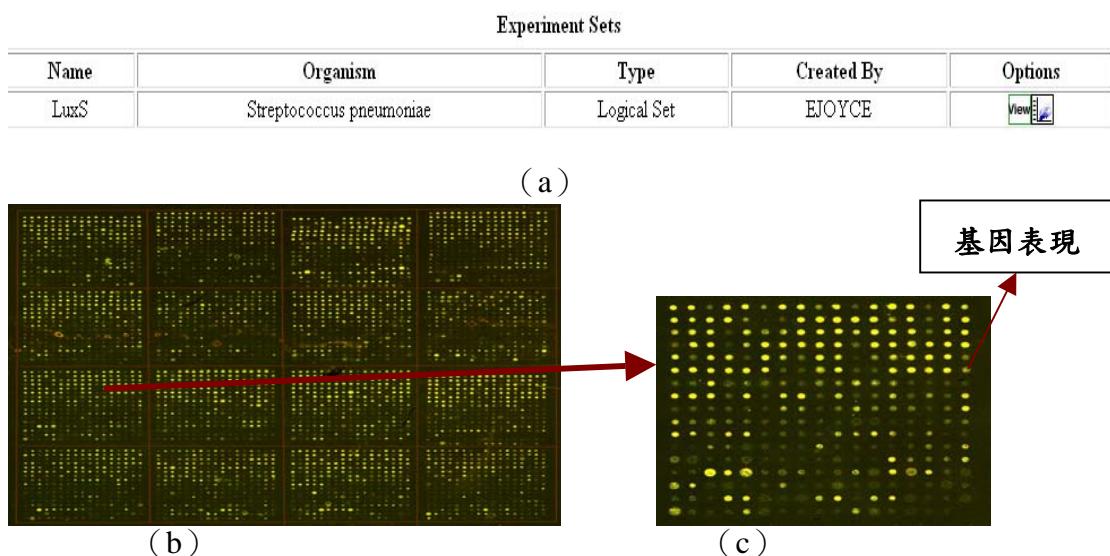


圖 18 從史丹佛大學微陣列影像資料庫中所測試的基因微陣列影像；(a) 此影像實驗名稱；(b) 此影像所包含的每一個區塊；(c) 單一區塊所包含的基因表現點。

圖 18 顯示測試的基因微陣列影像中的一個用來說明本研究測試的結果。圖 19 顯示本研究一個完整的實驗測試結果，當原始的一張影像輸入本研究的系統之後，輸出最後分析過後的影像，顯示出被偵測出來的基因表現點的邊緣以及被切割出來的方格。從圖 19 來看經由雜交過後所呈現的基因微陣列的影像，沒有表現的基因表現點將被過濾忽略掉，而剩下有表現的基因表現點將會被本研究系統偵測出來。

原始輸入影像 (a) 與基因表現點偵測後影像 (b) 對照; (c) (d) 為局部放大

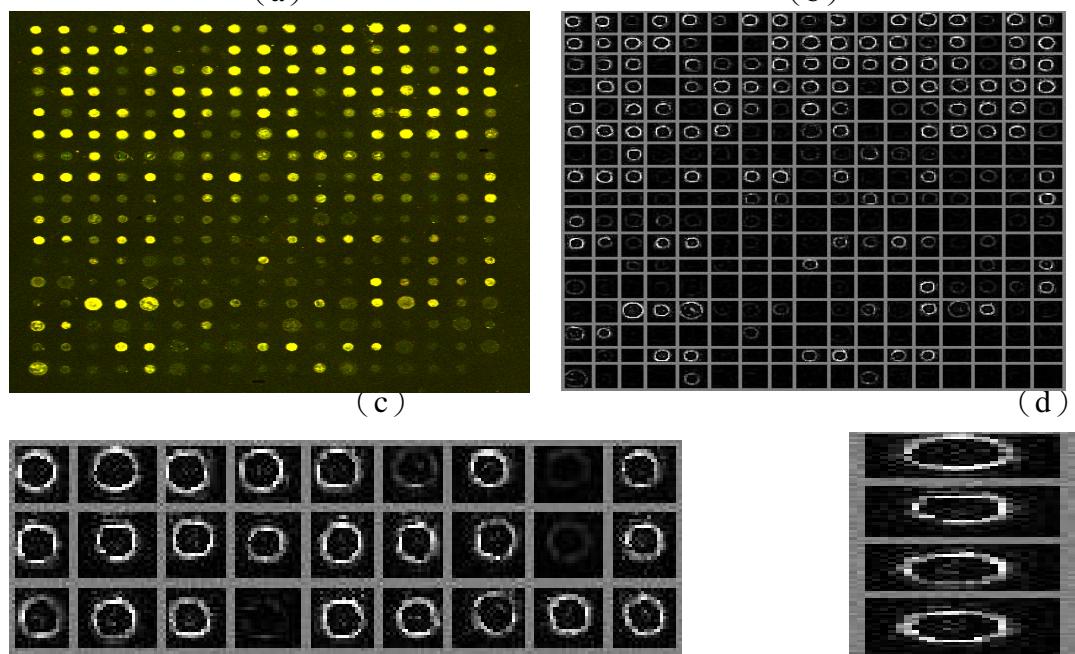


圖 19 實驗測試一

另外一個測試的結果，圖 20 為另一個區塊的原始基因微陣列影像，而最後輸出的是本研究偵測出來的基因表現點結果。

原始輸入影像 (a) 與基因表現點偵測後影像 (b) 對照; (c) 為局部放大

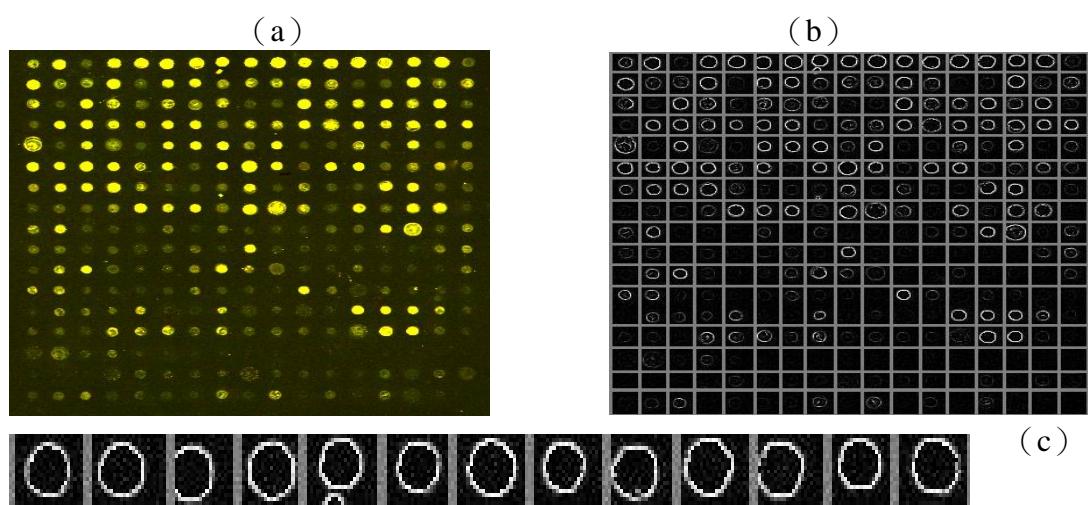


圖 20 實驗測試二

三、實驗比較

為了說明本研究所提供的方法對於基因表現點的偵測與切割能夠提高其正確率，而以史丹佛的基因微陣列資料庫（Stanford, 2004）中已經切割好的基因表現點位置來作比較。圖 21 (a) 為從其資料庫中所取出的單一個切割好的基因表現點，可從圖中看出整個切割的方格並沒有把基因表現點給包含進去。而圖 21 (b) (c) 為利用本研究方法所作出的單一基因表現點切割，以邊緣偵測演算法找出基因表現點的範圍，而將正確的基因表現點大小給切割出來了。

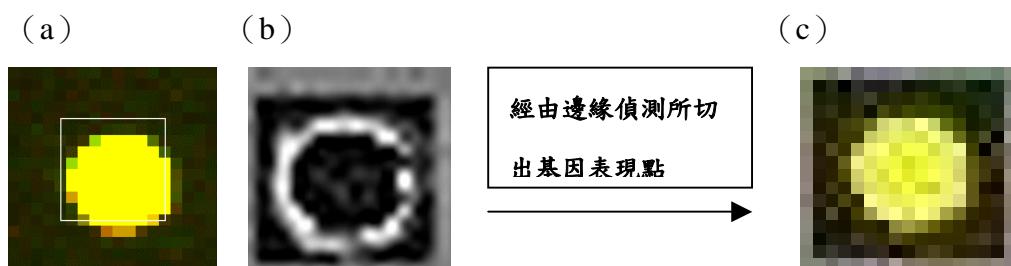


圖 21 本研究與史丹佛大學資料庫比較；(a) Stanford 所切出來的單一基因表現點位置；(b) (c) 為本研究所提供的方法所切出來的基因表現點

肆、結論

在本研究中，致力於開發一個「電腦影像處理技術應用於生物晶片基因微陣列影像分析」的系統利用理想的影像邊緣偵測的演算法（Canny Edge Detection Algorithm）於基因微陣列影像之基因表現點位置切割的找尋，以實現另一個在基因微陣列影像之基因表現點邊緣的找尋和切割。

本研究所強調的兩個重點：一、初步劃分方格內的基因表現點的輪廓並透過演算法標示；二、基因微陣列影像中基因表現點所在的方格位置動態調整的切割。經過實驗分析後皆得到良好的結果，並與 Stanford 的基因微陣列資料庫中切割好的基因表現點位置比較結果顯示本研究所建立的系統的確能完整切出基因表現點的位置；並證明不需要使用者額外的手動調整，便能完整的截取基因表現點訊號且能完整的偵測出每個基因表現點的濃度變化，而提供一個良好的基因微陣列影像分析工具來提高後續資料分析的精確度。

誌謝

感謝國科會計劃之編號 NSC 92-2213-E-008-021 贊助本研究之經費。

參考文獻

- 連國珍 (2001)：數位影像處理，儒林圖書有限公司，pp 4-2。
- 鍾國亮 (2002)：影像處理與電腦視覺，東華書局。
- Brown C. S., Goodwin P. C. and Sorger P. K., (2001) "Image metrics in the statistical analysis of DNA microarray data," Proc. Natl Acad. Sci. USA, Vol. 98, No. 16 pp. 8944-8949.
- Caldas Carlos and Brenton D. James, (2004) "Microarray segmentation methods significantly influence data precision," Nucleic Acids Research, Vol. 32, No. 5, pp. e50.
- Canny J., (1986) "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, Issue 6, pp. 679-698.
- Hirata R., Hashimoto R. F. and Dantas D. O., (2001) "Microarray gridding by mathematical morphology," International Brazilian Symposium on Computer Graphics, pp. 112-119
- Hogervorst F., Kuikman I., Von dem Borne AE and Sonnenberg A, (1990) "Cloning and sequence analysis of beta-4 cDNA : an integrin subunit that contains a unique 118 kd cytoplasmic domain," EMBO, Vol. 9, pp. 765-770.
- Kishimoto T. K., Connor K. O', Lee A. and Roberts T. M. (1987) "Springer : Cloning of the beta subunit of the leukocyte adhesion proteins : homology to an extracellular matrix receptor defines a novel supergene family," Cell, Vol. 48, pp. 681-690.
- Liang P. and Pardee A. B., (1992) "Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction," Science, Vol. 257, No. 5072, pp. 967-71.

- Liang P., Averboukh L and Pardee A. B., (1993) "Distribution and cloning of eukaryotic mRNAs by means of differential display," Nucleic Acids Res., Vol. 21, No. 14, pp 269-75.
- Liew A. W., Yan H. and Yang M., (2003) "Robust adaptive segmentation of DNA microarray images," Pattern Recognition, Vol. 36, pp 1251-1254.
- Sargent T. D., (1987) "Isolation of differentially expressed genes," Methods Enzymol, Vol. 152, pp. 423-32.
- Stanford MicroArray DataBase, (2004), <http://genome-www5.stanford.edu>.
- Suzuki S, Huagn ZS and Tanihara H, (1990) "Cloning of an integrin b subunit exhibiting high homology with integrin b3 subunit," PNAS, Vol. 87, pp. 5354-5358.