

微陣列基因晶片影像之網格定序

王清林 王清德 江孟霖、史珮妮、羅健銘
國立勤益科技大學 國立勤益科技大學 國立勤益科技大學資訊管理系
資訊管理系 資訊管理系
clwang@ncut.edu.tw ctwang@ncut.edu.tw s39532021@student.ncut.edu.tw

摘要

二十一世紀各界學者完成人類染色體組定序後，更進一步利用生物晶片科技進行基因功能研究。微陣列基因影像擷取與定位辨識是進行基因功能研究後續實驗分析之重要的第一個步驟。本研究旨在透過影像處理技術針對微陣列基因晶片影像提出一基因表現點網格自動定序方法，可針對點針不完全或失序殘破之表現點作網格繪製；另由建構得之基因表現點邊緣輪廓計算其色彩濃度，作為病理推估計算及區分等級，可提供醫學使用上縮短檢測之時效性。經實驗顯示本方法確實能完整的繪製網格並截取基因表現點訊號，不需要手動調整及指定相關參數，即可依其濃度變化區分等級，提供一個良好的影像分析工具，提高後續資料分析的精準度及相關應用範疇。

關鍵詞：生物晶片、微陣列基因、基因表現點

Abstract

From all walks of life scholar finish human chromosome group make after the preface 21st century, is it utilize biochip science and technology carry on gene function research to go still one step further. It is the first step of carrying on follow-up experimental analysis that the microarray gene image picks and fetches and orients distinguishing, it is also indispensable important step, and how to display some positions and do the accurate localization to different forms, crooked gene, have direct influence on follow-up analysis, this method displays some edge picture element and clicks and takes a sample to calculate completely with the gene on the image, improve some needle complete and out-of-sequence net estimate, it makes to be can effective and automatic completion net make preface and draw out perfect

localization click. Another calculate its thickness of color widely by building the edge that must construct, is it can do with gene thick tolerance value ordered to display pathology is it is it regard as to estimate to push away, prescription that measure on offer medical science use to make. Analysed that compares its gene to behave and click by this research method, prove its experimental result, can improve the perfect degree of application categories calculated with pathology of follow-up analysis.

Keywords: biochips, microarrays, Spot expression

1. 緒論

生物晶片的概念為 1991 年 Fodor 等人為首提出，其概念如同電腦晶片一般將各種不同類別之功能整合一晶片運作，而生物晶片藉此利用這種概念把核酸和蛋白質等固定於固相載體的微型檢測元件，且它的技術發展極為迅速目前有基因晶片(DNA microarray)、蛋白質晶片、縮微實驗室晶片等。現今生物晶片技術中，微陣列晶片是生物晶片發展與應用中較為成熟的技術之一，其製作方式為將晶片上的核酸探針與受測之液體中的基因做雜交的動作，假使某一段基因有雜交地反應會呈現不同螢光色，而雙螢光標記是以綠色(Cy3)與紅色(Cy5)，這兩種顏色重疊一起產生其他顏色，而每個像素點的值是代表著此點顏色濃度的基本單位，每一個基因表現點影像中色彩濃度的變化對於其資料分析的量化有著直接的影響[2]，故每一個基因表現點若能正確地切割出來對於分析基因表現時，可提高後續資料分析時的精確度[3]。Hirata[6]等學者曾利用數學的架構提出一個切割微陣列基因影像中其基因表現點位置的演算法，其方法必須靠手動的方法來判斷半徑大小以及基因表現點之間的相對位置，才能夠正確地將基因表現點的位置作切

割。Liew[8]等人也曾提出一個方法來切割出微陣列基因影像上的基因表現點，但是其方法對於辨識基因表現點必須利用適當的顏色濃度來找出表現點的位置。因此，只能辨識出帶有高衰度之濃度的色彩變化並對基因表現點的位置加以切割。薛義誠[1]等人利用 Canny 影像邊緣偵測的演算法，以兩個門檻值判斷邊緣，雖可以正確偵測邊緣但針對不同影像中其平滑係數及門檻值仍需不斷嘗試最佳參數，而在過於破碎及殘缺之基因表現點偵測中仍不能有效建構邊緣。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基因晶片影像之基因表現點切割的相關研究及方法論，雖可利用邊緣偵測法切割出基因表現點位置，卻仍不能解決破碎及殘缺基因表現點之偵測。

本研究提出一晶片影像之基因表現點網格定序法，將微陣列基因晶片影像轉換為 HSI 色彩空間，為色調、飽和度、強度(Hue、Saturation、Intensity, HSI)三張影像。而當中最為清晰明顯的強度空間之影像作為實驗基底，對此強度影像以索柏運算子(Sobel operator)建構出二值化的邊緣影像，並利用標記連通成分(Label connected component)將二值化影像之物件進行標記並篩選掉不構成圓的物件，將影像內的圓進行三點圓上重心估圓，並取其眾數作為標準半徑值 R_1 ，對每個表現點圓之誤差值最小的四組基因表現點之圓來作為定位，從平滑化之直方圖中找出定位圓所在之行、列個數，依 R_1 半徑值與定位圓的相對位置將網格依序繪製，以完成網格定序。本研究另議一基因表現點之色彩濃度計算功能，將最後定序得之基因表現點，依色彩濃度進行病理推估，可直接檢視所有基因表現點之色彩濃度是否有異。

現今醫療科技日漸進步，微陣列晶片細胞生化學為近代基因研究與病理推測所盛行，但在此方面的相關研究與疾病上的診斷等之影像技術卻仍有改善的空間，本研究希望藉由影像處理技術，針對以下缺點加以改善：(1)製作基因晶片過程中，微陣列晶片影像的樣本液在點片時超出其基因表現點之範圍。(2)沖洗樣本液時造成基因的剝落，導致在最後拍攝成影像資料時，影像內的基因表現點不完整甚至破碎的情形產生。

本研究提出晶片影像之網格自動定序方法，希望將影像中的基因能更有效地標記定位，使其進行更進一步的基因分析處理時，能取得有效的樣本數據，在疾病上的診斷此技術

亦可協助醫生判斷與分析病狀。另此技術在分析與檢驗時，仍需依循著基因樣本一一做比對，此步驟將會浪費相當多的時間在比對基因上，因此本研究除了將微陣列晶片做切割定位之外，並提議一基因表現點之色彩濃度計算方法，希望藉以提高微陣列晶片的檢驗的時效與成果。

2. 相關技術

本章回顧本研究所使用的影像處理相關技術，首先在第二節會介紹影像色彩模型的相關特色與使用方式；接著在第三節會對於物件輪廓偵測階段所使用的邊緣偵測演算法做概述與探討；而在第四節介紹本研究中用於標記二值影像之標記連通成分的原理與技術。

2.1 HSI 色彩空間

HSI 色彩空間[10]由色調(Hue)、飽和度(Saturation)、強度(Intensity)所組成，其色調和飽和度為色彩上表示顏色的類別與深淺，人眼的視覺對強度的敏感遠勝於顏色濃淡程度，其空間影像之色彩物件比 RGB 影像具有更強的色彩描述且符合人眼的視覺特性。

HSI 色彩模型(如圖 2-1)，將色彩資訊中帶有強度成分的資訊與色調和飽和度做分離。色調 H 的範圍定義在 0 到 2π 角度之間，紅色則位在 0° 、綠色位在 $2\pi/3$ 、藍色位在 $4\pi/3$ ；當飽和度 S 為 1 時色調表示為完全飽和，飽和度 S 為 0 時色調表示為完全不飽和；而強度 I 代表色彩的強度其範圍為 $[0, 1]$ 。HSI 色彩空間所需的色調、飽和度、強度其值可以由 RGB 的色彩立方體中取得，可以將任何 RGB 點轉換為 HSI 色彩模型中的點，其轉換公式於第三章中再行介紹。

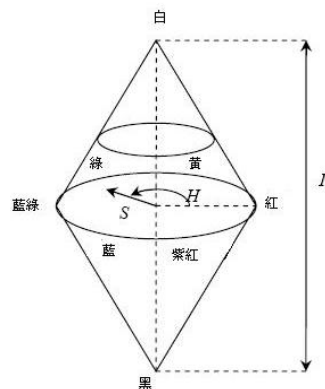


圖 2-1 HSI 色彩模型

2.2 邊緣偵測演算法

目前全世界有許多的邊緣偵測方法，較常被使用的為 Gradient [5]、Canny[4,7]、Sobel[12] 以及 Laplacian[11]，而這幾個方法的共通點是針對影像利用邊緣偵測濾波器來偵測影像的輪廓，並將影像中的雜訊剔除掉，區分出灰階影像裡高頻與低頻的部份，取得高頻用以將影像邊緣與鄰近像素之間的差異加以放大讓其更為明顯，讓邊緣偵測能夠更準確地被執行。

現在增強影像邊緣輪廓常用的方法為 Gradient filter，運用微分運算來達到效果。先假設 $f(x, y)$ 為影像 f 在座標點 (x, y) 上的灰階值，在 (x, y) 座標處的梯度向量 $\nabla f(x, y)$ 如式 2-1 求得：

$$\nabla f(x, y) = \left(\frac{\partial f(x, y)}{\partial x}, \frac{\partial f(x, y)}{\partial y} \right) = (f_x, f_y) \quad (2-1)$$

其中 $f_x = f(x+1, y) - f(x, y)$ ，

$f_y = f(x, y+1) - f(x, y)$ 。

該向量的梯度大小為 $|\nabla f| = \sqrt{f_x^2 + f_y^2}$

，其向量的梯度方向定義為 $\theta = \tan^{-1} \left(\frac{f_y}{f_x} \right)$ 。

當 $|\nabla f|$ 越大時，就代表此像素為邊緣輪廓的可能性就越高，同時也與邊緣連線越呈垂直。

2.2.1 Canny edge detection

Canny 輪廓檢測法相同於 Gradient filter，首先求出每個像素的梯度 $|\nabla f| = \sqrt{f_x^2 + f_y^2}$ 與梯度方向 $\theta = \tan^{-1} \left(\frac{f_y}{f_x} \right)$ ，此處 f_x 與 f_y 分別為對像素值在 x 與 y 方向上作一階導數。若梯度方向上像素的 $|\nabla f|$ 大於某被給定的門檻值，則視該此像素為邊緣輪廓點。也可設定兩個門檻值 T_h 與 T_l ，若梯度方向上像素的梯度大於 T_h 時，視此像素為邊緣輪廓點；而此像素連接出的其它像素，如其梯度大於 T_l 時，也視該連接出的像素為邊緣輪廓點。

2.2.2 Sobel edge detection

Sobel 跟其他邊緣偵測法最不同的地方在於它除了邊緣強化以外，還多了使邊緣平滑的好處。先計算出遮罩的中心所在像素點的梯度，如圖 Sobel 之區塊遮罩如圖 2-2：

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

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)$$

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

$$G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

圖 2-2 Sobel 之區塊遮罩

經由 Sobel 運算遮罩的處理後，將遮罩中心移到下一個像素點，重複上述運算可得到第二個像素的梯度值。當對整張影像的像素位置都執行這個過程後，可得就得到一張與原來影像同樣大小的梯度圖，會平滑影像並讓對應到像素其邊緣梯度的部位特別突出。其中，若中央像素點與其相鄰的八個點 ΣG_x 與 ΣG_y 皆都是相同值，即表示沒有邊緣通過，其梯度值也會是零。若左、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，重複上述計算過程直到最後一個像素值可得到一個梯度值：

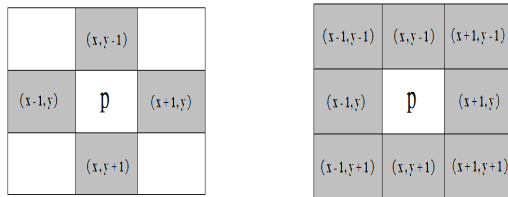
$$g = [G_x^2 + G_y^2]^{1/2}$$

對此梯度值取一門檻值 T ，當 $g \geq T$ 時，則該像素值則判定為邊緣像素。

2.3 標記連通成分

標記連通成分 (connected component) [10]

可將影像中像素的周圍鄰近點以互相連結的方式來組成物件 (object)，其主要連結的方式與路徑之鄰接性有關，故標記連通成份的特點取決於所選擇鄰接的形式。最為普遍的鄰接形式為 4 鄰接 (4-adjacent) 圖 2-3 (a) 與 8 鄰接 (8-adjacent) 圖 2-3 (b)。因此，利用這兩種的鄰接形式可將二值影像之像素群組化為物件，擁有物件群組後可再進一步去分析其形狀、大小與位置等。



(a) 像素 p 的 4 近鄰

(b) 像素 p 的 8 近鄰

圖 2-3 標記連通成份之近鄰

當像素 p 之座標在 (x, y) 時，其上下左右各有兩個水平與垂直的近鄰，座標分別為 $(x, y-1)$ 、 $(x, y+1)$ 、 $(x-1, y)$ 與 $(x+1, y)$ 時，表示為 p 的 4 近鄰 (4-neighbors) 如圖 2-3(a)，以符號 $N4(p)$ 來表示；像素 p 的四個對角其近鄰座標為 $(x-1, y-1)$ 、 $(x+1, y-1)$ 、 $(x-1, y+1)$ 與 $(x+1, y+1)$ ，以符號 $ND(p)$ 表示；將 $N4(p)$ 與 $ND(p)$ 兩者相互做聯集為 p 的 8 近鄰 (8-neighbors)，如圖 2-3 (b)，以符號 $N8(p)$ 表示。

3. 晶片影像之網格定序法

本研究所提出的微陣列基因晶片影像之網格定序共分成三個主要階段：3.1 節影像強化轉換、3.2 節物件輪廓偵測與 3.3 節網格定序三個階段，並於 3.4 節計算網格內基因表現點之色彩濃度及分類用以協助病理判斷。而在影像強化轉換階段中以色彩空間轉換並選擇最佳的空間模式作為基底，讓影像上之特徵得以被適當偵測；經由轉換強化後之影像，在物件輪廓偵測階段，由邊緣偵測演算法進行二值化邊界的運算，並藉著標記連通成分讓邊緣像素完成標記，再進一步依影像內的邊緣像素進行三點亂數並以重心公式來估計圓心及半徑，篩選出 50 組趨近於圓之基因表現點物件。在網格定序階段，依得之每個基因表現點物件其半徑的眾數來找出該影像具代表性之半徑 R_1 值，再依半徑 R_1 值計算平均絕對誤差值進行定

位點的評估，將誤差值最小的四組邊緣像素物件為定位基準，以完成網格之推估。本章節將對此三個階段做詳細描述。

3.1 影像強化轉換階段

微陣列基因影像圖片之基因表現點主要由使用綠色反應 (Cy3) 與紅色反應 (Cy5) 之螢光染劑所混合組成，為能計算由這兩種色彩所混合組成的影像上其每一點像素之色彩飽和度，因此，必須將微陣列基因的 RGB 影像作轉換。在本研究中，為了讓後續基因表現點的邊緣偵測處理與網格定位能有效利用基因表現點作邊緣取樣，其影像在拍攝時，可能因所使用燈光的強度不同，而造成在對同一影像所攝取影像之色彩亮度對比造成差異。為減少影像拍攝時使用燈光強度不同對後續物件輪廓偵測之影響，必須依微陣列基因影像之特性轉換至適當色彩空間。影像強化轉換階段包含以下二步驟：

Step1: 影像轉換與評選

將影像轉換至 HSI 色彩空間，評選一轉換後的灰階影像作為實驗基底。

Step2: 影像正規化

針對經過 HSI 色彩空間轉換之強度灰階影像 (I) 作正規化至 0 到 255 之間。

3.1.1 影像轉換與評選

HSI (Hue-Saturation-Intensity) 色彩空間紀錄影像的色調 (Hue)、飽和度 (Saturation)、強度 (Intensity)，對影像之色彩物件比起 RGB 影像有更強的色彩描述，其色彩頻率轉換方式較符合本研究探討之實驗影像，故將影像經 HSI 色彩轉換表示，將色彩資訊中帶有強度成分的資訊與色調和飽和度做分離。此色彩模型將色調 H 的範圍定義在 0 到 2π 角度之間，紅色位在 0° 、綠色位在 $2\pi/3$ 、藍色位在 $4\pi/3$ ；當飽和度 S 為 1 時色調表示為完全飽和，飽和度 S 為 0 時色調表示為完全不飽和；而強度 I 代表色彩的明度其範圍為 $[0, 1]$ 。HSI 彩色模型的優點在其強度與彩色資訊 (I 和 HS) 作分離，不會因為色調或飽和度而干擾到灰階的強度。根據上述的關係，可以將 RGB 的色彩影像轉換成 HSI 色彩空間後可以 H 、 S 、 I 值來表示三種灰階影像，而其轉換的公式如下：

$$\text{色調為, } H = \begin{cases} \theta & \text{若 } B \leq G \\ 360 - \theta & \text{若 } B > G \end{cases}$$

$$\theta = \cos^{-1} \left\{ \frac{\sqrt{(R-G) + (R-B)}}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\}$$

$$\text{飽和度為, } S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)]$$

$$\text{強度為, } I = \frac{1}{3}(R+G+B)$$

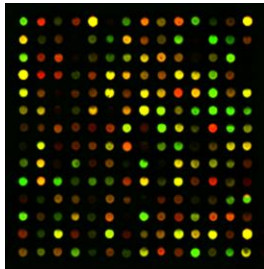


圖 3-1 RGB 彩色微陣列基因影像

本研究將其 RGB 色彩空間轉換成 HSI 色彩空間，可獲得 H 值、S 值及 I 值灰階影像，H 值灰階影像如圖 3-2，表示色調頻率域之色彩表示，因背景與基因表現點之顏色度對比度過低及雜訊過高，故不選用。另 S 值灰階影像如圖 3-3，物件的色彩濃度不夠故不夠清楚且對比度也明顯過高，亦不選用。而圖 3-4 為經轉換後之 I 值灰階影像，其對比度高、且能清楚表達每顆基因表現點，最能表示其微陣列基因影像，故選為實驗基底影像。

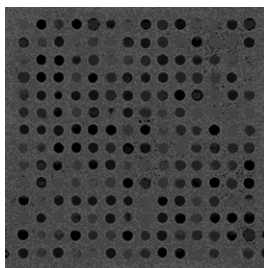


圖 3-2 H 值灰階影像

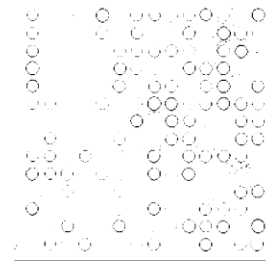


圖 3-3 S 值灰階影像

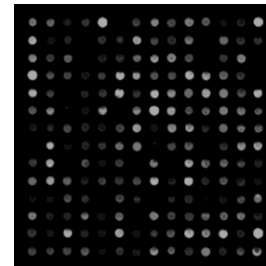


圖 3-4 I 值灰階影像

3.1.2 影像正規化

經 HSI 色彩空間轉換得之 I 值灰階影像，其影像中整體對比度可能過低，或像素值明顯皆處於低頻率之分佈，不利後續物件輪廓偵測階段繼續處理，故將影像依式 3-1 作正規化。設輸入影像 p，其 P_M 與 P_m 分別為所有像素中最大值與最小值，表示該 P 影像像素值中之組距，每一像素 $p(x,y)$ 減去影像最中之最小值 P_m ，再乘上 255，則被正規化更新為 $O(x,y)$ 。

$$O(x, y) = \left(\frac{p(x, y) - P_m}{P_M - P_m} \right) \times 255 \quad (3-1)$$

正規化後得之影像 O，如圖 3-5，經公式計算之每點像素值皆會被更新為 0 至 255 之間，使影像之所有像素值皆能分佈在 0 至 255 之間，使原圖 3-4 之影像中能見度較低之像素值，經公式強化後能拉開對比度、提高亮度，有利於後續物件輪廓偵測階段使用邊緣偵測法來建構邊緣。

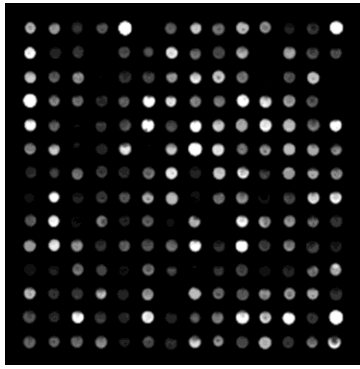


圖 3-5 正規化後之 I 值灰階影像

3.2 物件輪廓偵測階段

在不同的微陣列基因晶片影像中，其基因表現點物件之輪廓理應為一完整的同心圓，並擁有一致的半徑及周長，而邊緣附近像素值的顏色值，通常會與其鄰近像素值的顏色值有很大的差異。但在本研究中所發現問題在於因雜交和清洗時造成基因表現點的破碎、殘缺，以及點針狀況不理想導致基因附著太少，雖經影像強化後仍不足已辨識，文獻[1]的方法亦是如此，故不能只依靠邊緣偵測法進行物件偵測，需再依偵測得之邊緣像素進行估算。因此，本研究利用 Sobel 邊緣偵測建構邊緣像素，並利用建構之邊緣像素物件，選取 50 組物件來作為估計樣本，利用這 50 組趨近於完整之基因表現點在網格定序階段再進行篩選。本階段包含以下三步驟：

Step1: Sobel 邊緣偵測

使用 Sobel 邊緣偵測法對強化後影像進行邊緣像素之建構。

Step2: 標記連通並紀錄標記欄位

利用標記連通將建構得之邊緣像素進行標記，並將標記所得之邊緣像素座標值、邊緣像素個數紀錄至標記欄位。

Step3: 篩選估計樣本

依標記欄位中所紀錄的邊緣像素個數來篩選出 50 組物件，作為網格定序階段定位點估計的樣本。

3.2.1 Sobel 邊緣偵測

經由 Sobel 檢測器遮罩計算出之像素差異值為梯度，當影像中物件邊界輪廓較明顯處像素的梯度就越大，否則梯度就越小，依靠梯度

值去進行邊緣偵測。目前已有很多種計算邊緣梯度的方法被提出，經過第二章第二節提及之各項邊緣偵測法，存在著所用之邊緣其梯度仍不足已偵測出邊緣或是使用兩階門檻值造成基因表現點過多細線的情形。且其不同的邊緣檢測器對梯度的計算，針對不同影像亦必須擁有十分準確的門檻值，才能完整地將邊緣正確的偵測，如拉普拉斯檢測器易造成梯度太小，Canny 邊緣偵測法又過甚造成細線情形，如圖 3-7。

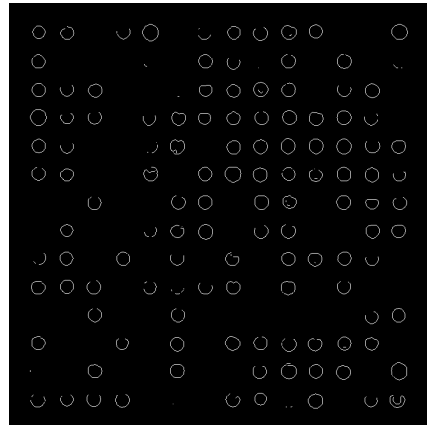


圖 3-6 Sobel 邊緣偵測之二值影像

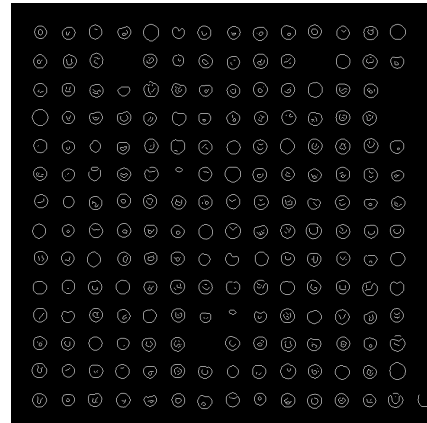


圖 3-7 Canny 邊緣偵測之二值影像

分析各類計算方法的限制與其優缺點後，發現 Sobel 之檢測運算子，能適度的偵測縱向與橫向之邊緣並平滑邊緣像素，應用 Sobel 之檢測運算子計算邊緣，如圖 3-6，再對偵測後得到的物件邊緣進行取樣，能進一步做更精準的整體網格估計，亦可解決邊緣偵測法無法對像素值差異太小的基因表現點做邊緣建構的問題。

3.2.2 進行標記連通並紀錄標記欄位

利用文獻提到之標記連通成分，利用 4 鄰接與 8 鄰接等連通特性，在不同標記連通成分

的像素裡頭會被指派從 1 到標記連通成分的總數,如當標記(Flag)為 1 的時候像素為第一個連通成分;當標記為 2 的時候像素為第二個連通成分,按照這樣的方法推到最後,將背景的像素標記為 0,輸出標記矩陣。最後,將標記矩陣裡之像素座標值依標記的不同,依序建構基因標記欄位(SpotFlag Table),使每一個標記(Flag)的邊緣像素座標值與邊緣像素個數加以歸納,如圖 3-8:

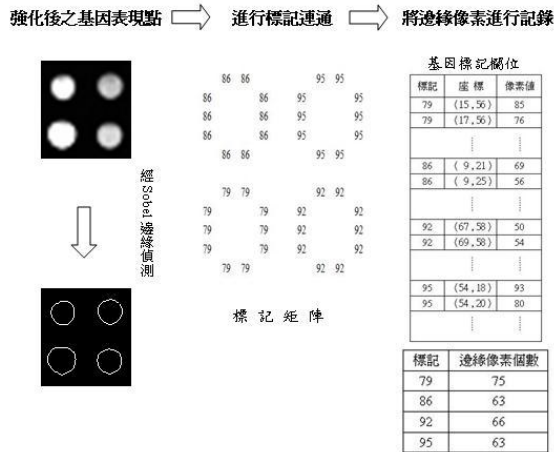


圖 3-8 邊緣偵測與物件標記流程

3.2.3 篩選估計樣本

雖獲得影像上基因表現點的邊緣與標記,得以計算推估每一表現點應有的圓心及半徑,但邊緣像素個數過少的物件,將會降低本研究經由樣本來估計整體網格的準確度。因此,在圖 3-9 中利用基因標記欄位之物件邊緣個數繪製邊緣像素次數分配圖,其 X 軸為物件之邊緣像素個數, Y 軸為該物件數。而圖 3-9 中,邊緣像素個數位於 0 於 10 之間為破碎邊緣物件,代表一個基因表現點經邊緣偵測後得到的邊緣並不完全,其邊緣像素個數只有 10 個像素值不到,而這類的破碎邊緣物件將不利後續網格估計,實驗過程將其視為干擾值,將透過標準差遞增篩選趨近於 50 組圓之物件。本研究選定 50 組樣本作為系統參數,其原因為在統計取樣的估計上,學理常將大樣本視為 ≥ 30 ,而母體樣本的基因表現點個數又因實驗影像不同可給定不同參數,在此先預設估計樣本為 50 組。扣除干擾值之後,該圖為左偏之鐘形分配,故計算出所有物件邊緣個數之平均數後,以標準差逐漸遞增過濾掉破碎邊緣物

件、扣除邊緣物件個數過小,以篩選出 50 組趨近於完整圓之基因表現點邊緣物件,作為下節網格定序中之估計樣本。

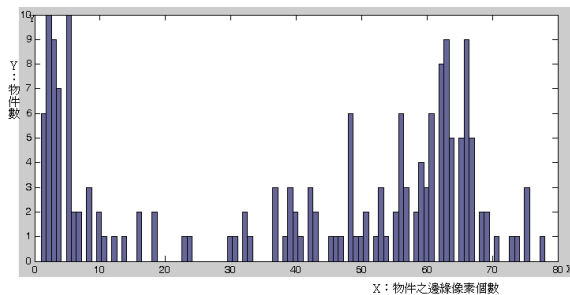


圖 3-9 邊緣像素次數分配圖

3.3 網格定序階段

本研究在網格定序階段中,將從上節獲得的 50 組估計樣本中找出最具代表性的四組定位點來作為定位,再從這四組定位點中計算其座標值,對應到其行、列兩張直方圖中,決定其定位點在網格繪製的基準。而如何取得最佳代表性的定位點,即是這個階段最重要的問題。本研究將計算 R_1 值與平均絕對誤差值 D 來作為定位點之評選,以計算出誤差值最小之四組物件作為定位點。而繪製直方圖的過程中,需先對其作平滑化並計算其斜率變化趨勢,以利定位點座標值判定。本階段包含以下六步驟:

Step1:基因表現點之圓心估計

以亂數選取邊緣像素三點,並依圓公式估計每一組樣本之圓心座標值及半徑。

Step2:基因表現點之半徑估計

依上述獲得之半徑值,再利用 250 組數據取眾數並決定影像標準半徑 R_1 值。

Step3:計算平均絕對誤差值

計算平均絕對誤差值 D ,依標準半徑 R_1 值來對每一物件邊緣像素作相減,評估每一組樣本其誤差值大小。

Step4:取出誤差值最小之物件

取出 4 組誤差值最小的物件,作為定位基準點。

Step5:判別定位點所在之行列數

判別定位點所在之行、列數,需先針對直方圖做平滑化,並計算直方圖之斜率變化趨

勢，得以判斷四組定位點的所在座標值。

Step6:基因表現點定位

進行基因表現點定位，計算其之相對位置，自動完成網絡繪製。

3.3.1 基因表現點之圓心估計

為得以完整的依估計標記樣本來估出該基因表現點的圓面積，本研究將依照該標記欄位紀錄的邊緣像素座標值以亂數來決定三點座標值，用以估計基因表現點之圓心及其半徑值。但是，以亂數決定座標值的過程中，可能導致產生的三點座標過近，無法正確估出該物件應該擁有的圓。因此，本研究在亂數決定三點座標值來估計圓之前，必須先取得一 R_0 半徑估計值，使產生的三點之間的距離均大於 R_0 值，以確保三點距離夠遠，亦即此三點邊緣像素具代表性，可用以估計該基因表現點之圓心座標值。本研究以基因表現點邊緣像素取得該邊緣像素之第一點、中間點及最末點三點座標值，如圖 3-10 中之 a、b、c 三點進行 R_0 值的估算。

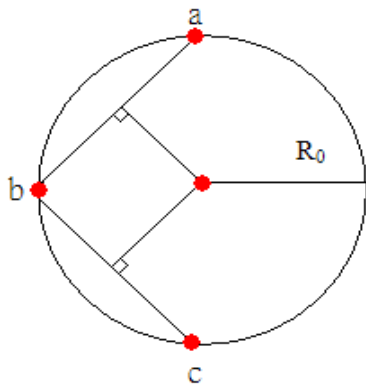


圖 3-10 以圓樣本取得 R_0

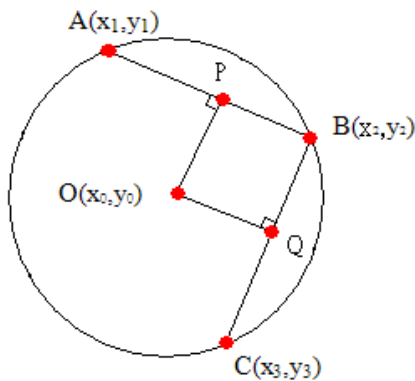


圖 3-11 以亂數決定物件圓上三座標點估圓

求得 R_0 估計值後，設一亂數種子取得物件

圓上三點 $A(x_1, y_1)$ 、 $B(x_2, y_2)$ 、 $C(x_3, y_3)$ ，如圖 3-11，而 \overline{AB} 、 \overline{BC} 、 \overline{AC} 線段之間距離均需大於 R_0 半徑值，即為 $\overline{AB} > R_0$ 、 $\overline{BC} > R_0$ 、 $\overline{AC} > R_0$ 。為求得圓心 O 點令座標 (x_0, y_0) ，將 A、B、C 三點座標代入圓公式 3-2。得結果如式 3-3，解 3-3 式求得 (x_0, y_0) 及 r 值。

$$(x - X_0)^2 + (y - Y_0)^2 = r^2 \quad (3-2)$$

$$\begin{cases} (x_1 - X_0)^2 + (y_1 - Y_0)^2 = r^2 \\ (x_2 - X_0)^2 + (y_2 - Y_0)^2 = r^2 \\ (x_3 - X_0)^2 + (y_3 - Y_0)^2 = r^2 \end{cases} \quad (3-3)$$

3.3.2 基因表現點之半徑估計

依所有 50 組物件，每組物件隨機亂數亂數 50 組樣本共可得到 250 組樣本供估計半徑值，估算結果如表 3-1。本研究以這 250 組平均數四捨五入至小數兩位數，取眾數為標準 R_i 半徑值，作為本實驗影像之基因表現點之標準半徑。

表 3-1 250 組樣本估計之圓半徑值

物件\組別	1	2	3	4	5	...	50
1	13.21	11.1	11.01	11.1	10.93	...	10.7
2	13.04	11.24	10.92	10.69	10.64	...	11.29
3	13.27	11.2	11.15	11	10.85	...	11.42
4	13.04	10.7	10.95	11.04	11.18	...	10.85
5	13.03	11.11	11.19	11.44	10.93	...	11.24
6	13.21	10.98	11.16	11.45	11.26	...	11.13
7	13.08	10.56	11.4	11.21	10.97	...	11.03
8	12.99	10.76	11.23	10.98	10.96	...	10.42
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	13.17	11.59	10.98	11.1	11.37	...	10.97

3.3.3 計算平均絕對誤差值

接著本研究由這 50 組物件估計樣本中取

出 4 組最趨近於圓的基因表現點物件作為定位點，而為了評量該物件是否趨近於圓，依式 3-4 計算，針對圓之物件其邊緣像素座標值做估算；設邊緣像素座標個數為 n ，邊緣像素座標至圓心座標 (x_0, y_0) 之距離為半徑 r ，依 $r_1 \dots r_n$ ，減去 R 值取絕對值，再算出該樣本物件之平均絕對誤差值，為式 3-4。

$$D = \frac{1}{n} \sum_{i=1}^n |r_i - R| \quad (3-4)$$

3.3.4 取出 4 組誤差值最小的物件

依上述 3-4 公式計算得之平均絕對誤差值，依序紀錄於表 3-2 中，本研究選擇誤差值最小之四組標記作為定位點，可以在後續繪製網格的步驟中作為基準。

表 3-2 標記誤差表

物件編號	物件標記	圓心座標 (x, y)		平均絕對誤差值
1	133	513.9923	616.4938	0.261994
2	145	561.7915	427.8186	0.266935
3	54	281.9989	571.1008	0.27291
4	7	50.22979	239.5956	0.27819
5	129	513.7758	48.14223	0.281379
6	86	376.5085	242.0009	0.315222
7	74	328.8121	333.8902	0.323041
8	73	328.7335	286.5226	0.331819
9	28	144.9536	146.9753	0.341723
⋮	⋮	⋮	⋮	⋮
50	156	607.8483	288.0497	0.346426

3.3.5 判別定位點所在之行列數

擁有誤差值最小之 4 組作為網格定序的定位點，為了能準確地展開網格的繪製，並自動偵測出該圖之行列個數，不需再手動輸入行列

數及間距。而一個定位點分別在垂直方向與水平方向有對應之行列數，因此，針對影像之像素的量值描繪垂直與水平兩張直方圖，再行對應至定位點之座標值。如圖 3-12 為垂直方向的直方圖，並依定位點之座標值對應到直方圖中，判斷定位點坐落的行個數，水平方向亦同。其中，進行判斷定位點的過程中，需對直方圖中之峰谷值進行平滑化，且避免該數值序列之基因表現點殘破所造成行列數個數誤判，故再對直方圖計算其斜率變化趨勢繪製成圖，如圖 3-13 為斜率變化趨勢圖，以供判別出正確之峰值及谷值。

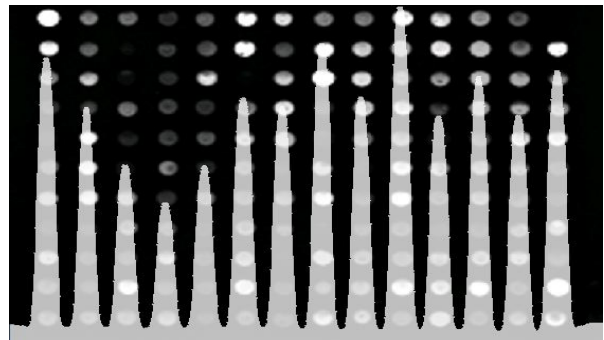


圖 3-12 經平滑化之垂直方向直方圖

在趨勢圖中，若斜率遞增值愈來愈小直到小於 0，而其量值轉變為遞減值的過程中，則在通過 0 的該點即代表該點為峰頂；若該點遞減斜率值進行標記直至其量值大於 0 為止，在通過 0 的該點即為谷底結束標記。圖 3-13 為垂直方向直方圖之斜率變化趨勢圖，可針對列之座標值進行註記後，即可判斷定位點所在之座標值所對應到的列個數，另水平方向亦同。

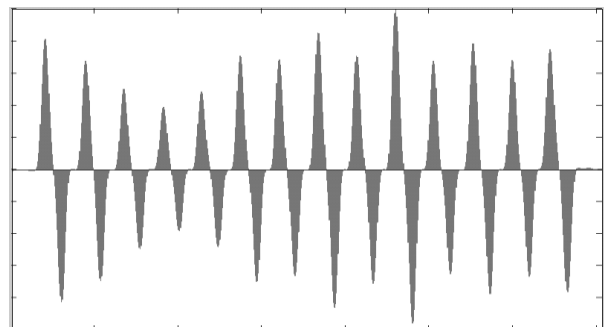


圖 3-13 垂直方向直方圖之斜率變化趨勢圖

3.3.6 基因表現點定位

依照四個定位點之座標位置進行相對位置計算，如圖 3-14。依行、列個數分別推估各點圓心座標位置依序排列，再依標準 R_1 半徑值及圓心代入圓公式 3-5，即可計算出該網格上圓之座標點；而 (X_0, Y_0) 為該點推估出之圓心座標，其 θ 角度為 0 到 2π ，輸出圓上各點座標 (X, Y) ，直到將整張影像網格繪製完成。

$$(x, y) = P(x_0 + R_1 * \cos \theta, y_0 + R_1 * \sin \theta) \quad (3-5)$$

其中， $\theta = 0 \leq \theta \leq 2\pi$

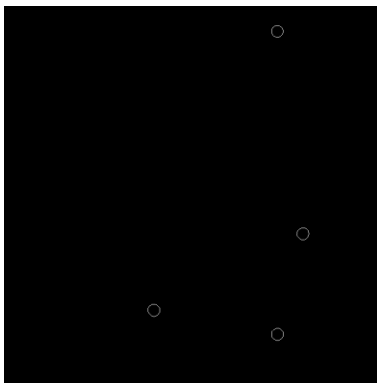


圖 3-14 具代表性之四組定位點

3.4 基因表現點之色彩濃度及分級

根據物件所標記的網格內所有座標值，每一點皆對應到微陣列基因晶片原圖的座標值，依序取出 R、G、B 色彩值。在文獻[2]中提及微陣列基因晶片在製作過程中，其基因表現點使用綠色反應 (Cy3) 與紅色反應 (Cy5) 之螢光染劑所混合組成。本研究將針對紅色、綠色兩個色彩值進行分析，將網格內所有座標的紅色值 (R) 及綠色值 (G) 進行排序，並篩選掉干擾的黑色量值，取中位數為該物件之代表值，如表 3-3。若該基因表現點之代表值 $R > G$ ，即代表其基因表現點擁有明顯的紅色之色彩濃度；若 $R < G$ ，則代表其基因表現點擁有明顯的綠色之色彩濃度量值。在此即可做簡單的病理判斷，學理上紅色量值較高的基因表現點，則代表其基因中病人擁有的量高於正常人的範圍，若綠色量值較高的基因，則代表該基因正常人擁有的量高於病人擁有的範圍，在這兩種情況下其基因表現點則表現其基因皆有狀況，一情況為其基因高於正常人範圍，另一情況又為病人所缺少之基因；而 R 與 G 兩者之色彩

濃度值皆高者，亦代表在兩種情況下皆無顯著差異。此外，可再將所有基因表現點上之紅色值與綠色值作排序，可依醫學病理上之需求區分等級，若濃度深度愈高則其分級之等級愈高，基因附著量相當高；若深度相較之下不高，基因表現點附著量則相對的低，亦代表分級之等級則較低。

表 3-3 色彩濃度數值表

Label	Red value	Green value
1	84	217
2	213	41
3	41	45
4	123	29
5	254	254
6	13	6
7	52	106
8	241	98
9	96	77
10	235	84
⋮	⋮	⋮

4. 實驗結果

本章節主要的目的在於以實驗的方式，來探討本研究所提議的微陣列基因晶片之網格定序之成效。本研究共採 6 張不同之微陣列基因晶片影像當作測試影像。本研究第一個實驗目的旨在完成網格繪製及定序，使用 Sobel 邊緣偵測法建構邊緣像素，並利用 50 組估計樣本再進行估計誤差值，以完成網格的定位與推估圖 4-1 為第六張測試影像，顯示出依本研究之結果，從圖 (b) 中可將基因表現點中完整的物件圈選標記邊界，再從局部放大圖 (c) 中看出圈選之表現點像素之範圍。

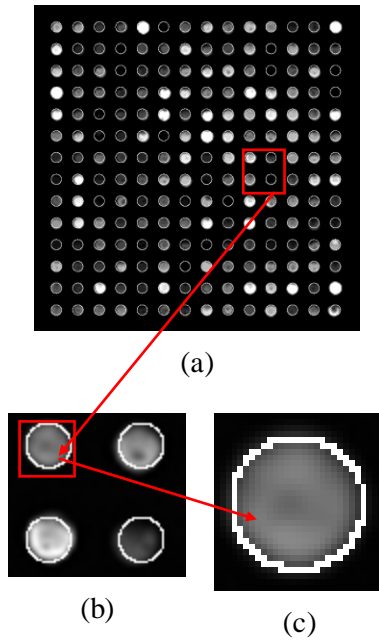


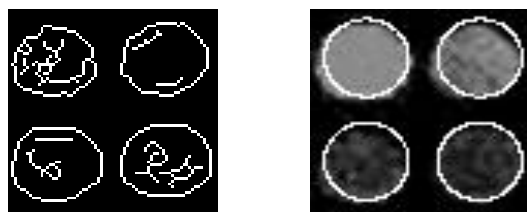
圖 4-1 第六張測試影像網格定位結果

將相同之實驗影像依[1]之研究方法作對照，該方法以 Canny 邊緣偵測法再針對其不同影像在做高斯模糊的同時將會造成參數上的一致性，另在門檻值設定時會造成圖 4-2 (a) 其干擾之邊緣，在第一張測試影像中，圖 4-3 (a) 的結果中其干擾情形將更加嚴重，而本研究之方法仍可估計出適當邊緣來圈選其表現點像素圖(b)。



(a) 薛義誠[1]之方法 (b) 本方法

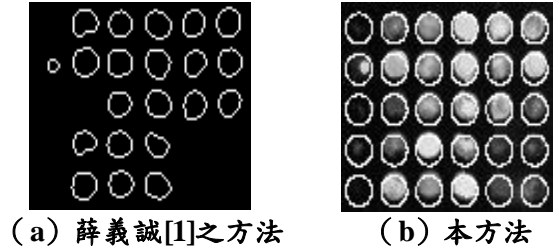
圖 4-2 第六張測試影像局部切割比較



(a) 薛義誠[1]之方法 (b) 本方法

圖 4-3 第一張測試影像局部切割比較

在第三張測試影像圖 4-4，該區塊(a)之基因表現點破碎情況嚴重，導致邊緣偵測之方法皆無法偵測得當，而本研究研究之方法仍可利用其它邊緣物件去建構出應有邊緣之位置(b)。



(a) 薛義誠[1]之方法 (b) 本方法
圖 4-4 第三張測試影像局部切割比較

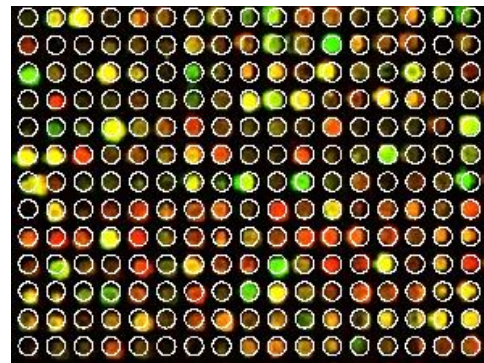


圖 4-5 第三張測試影像網格定位結果比較

在第三張測試影像中，圖 4-5 裡許多基因表現點已破碎、偏移，在其它方法論中針對該類型影像在定位或繪製邊緣時皆有其困難度，但在本研究研究方法之實驗結果中，仍可有效的將微陣列基因晶片影像完成定位並繪製網格。

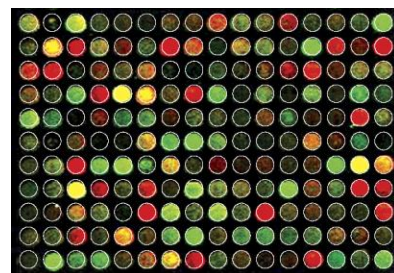


圖 4-6 第一張測試影像網格定位結果

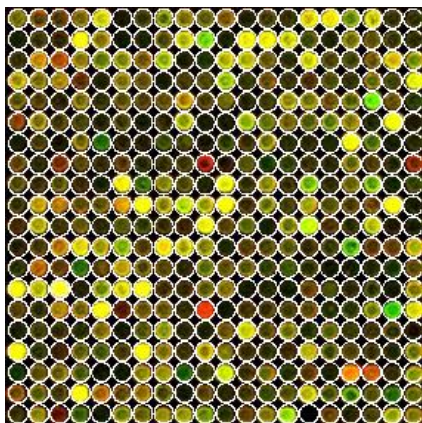


圖4-7第四張測試影像網格定位結果

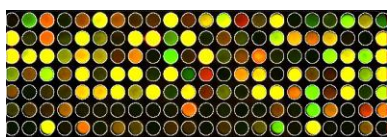


圖4-8第五張測試影像網格定位結果

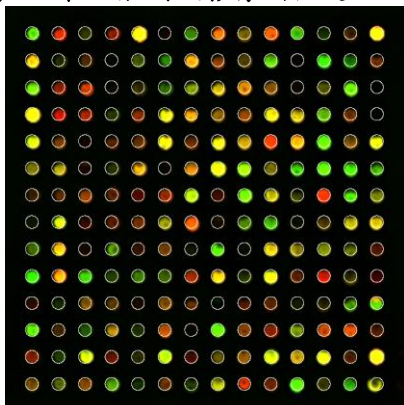


圖4-9第六張測試影像網格定位結果

本研究第二個實驗結果旨在探討依色彩濃度來協助進行病理分析，分別依中位數當作參考指標，將各基因表現點之色彩濃度數值區分成五個等級，依據第六張測試影像中如圖4-9，其基因表現點之色彩濃度紅色值達到第五個等級者(R_5)將以白色標記，如圖4-10。圖4-11為基因表現點之色彩濃度綠色值達到第五個等級者(G_5)。圖4-12為其基因表現點之色彩濃度兩者交疊之黃色值達到第五個等級(R_5G_5)，可有效辨別出基因表現點之色彩濃度所呈現不同之病理症狀，可供學界加快對照基因與其功能時之時效性及便利性。

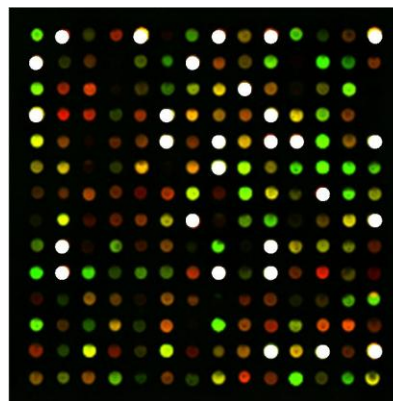


圖4-10第六張測試影像之 R_5

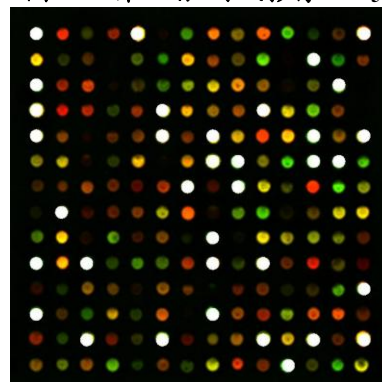


圖4-11第六張測試影像之 G_5

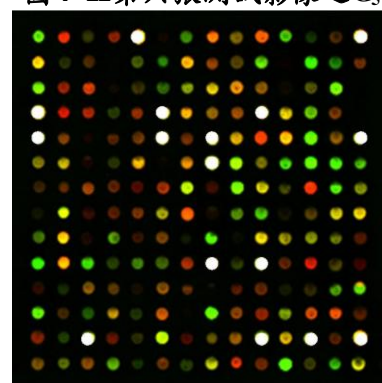


圖4-12第六張測試影像之 R_5G_5

5. 結論

本研究致力於開發一個以影像處理技術達成微陣列基因晶片網格定序的自動化系統，利用有效的邊緣偵測與網格推估之演算法，於微陣列基因影像之基因表現點位置能有效的繪製網格與定位，以實現在基因微陣列影像之基因表現點之正確判讀。本研究達成所強調的兩個重點：一、自動繪製定位之網格將基因表現點之正確輪廓位置透過演算法標示；二、基因微陣列影像中基因表現點所在圈選之色彩濃度像素值並區分等級，有利於後續分析與應用。

經過實驗分析後皆得到良好的結果，並與

薛義誠等人的基因微陣列邊緣切割演算法[1]比較結果顯示本研究所建立的演算法，的確能完整定位基因表現點的位置；本方法不需要使用者額外的手動調整，亦不必指定相關參數，即可準確繪製網格來圈選基因表現點，即便能完整的截取基因表現點訊號，來提高後續資料分析的精確度與應用。而繪製所得之網格範圍能完整的偵測出每一個基因表現點的濃度變化，並可依濃度變化區分等級作對應，可縮短檢驗之辨別所需時效，提供一個良好的基因微陣列影像分析工具來協助檢驗之應用範疇。

參考文獻

- [1] 薛義誠，楊勝琦，郭建明，”基因微陣列影像分析之研究”，*康寧學報*，7：261—283，2005。
- [2] Brown C. S., Goodwin P. C. and Sorger P. K., “Image metrics in the statistical analysis of DNA microarray data,” *Proc. Natl Acad. Sci. USA*, Vol. 98, No. 16 pp. 8944-8949, 2001.
- [3] Caldas Carlos and Brenton D. James, “Microarray segmentation methods significantly influence data precision,” *Nucleic Acids Research*, Vol. 32, No. 5, pp. 50, 2004.
- [4] Canny J., “A Computational Approach to Edge Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, Issue 6, pp. 679-698, 1986.
- [5] E. Davies, *Machine vision: theory, algorithms and practicalities*, Academic Press, 1990.
- [6] Hirata R., Hashimoto R. F. and Dantas D. O., “Microarray gridding by mathematical morphology,” *International Brazilian Symposium on Computer Graphics*, pp. 112-119, 2001.
- [7] J. F. Canny, “A computational approach to edge detection,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, pp. 679-698, 1986.
- [8] Liew A. W., Yan H. and Yang M., “Robust adaptive segmentation of DNA microarray images,” *Pattern Recognition*, Vol. 36, pp. 1251-1254, 2003.
- [9] L. Ding and A. Goshtasby, “On the canny edge detector,” *Pattern Recognition*, Vol. 34, No. 3, pp.721-725, 2001.
- [10] Rafael C. Gonzalez , Richard E . Woods , Steven L. Eddins , *Digital image processing using MATLAB*, Prentice-Hall, 2005.
- [11] R. F. Gonzalez and P. Wintz, *Digital image processing*, 3rd Editor, Addison-Wesley, 1992.
- [12] R. Gonzalez and R. Woods, *Digital image processing*, Prentice- Hall, 2002.