

國立交通大學

電控工程研究所

碩士論文

在具雜訊的正常血液抹片中進行白血球分
類計數



Differential Count of White Blood Cell in Noisy
Normal Blood Smear

研究生：洪裕筆

指導教授：林昇甫 博士

中華民國一百年七月

在具雜訊的正常血液抹片中進行白血球分類計數

Differential Count of White Blood Cell in Noisy

Normal Blood Smear

研究生：洪裕筆

Student: Yu-Bi Hong

指導教授：林昇甫 博士

Advisor: Dr. Sheng-Fuu Lin

國立交通大學



Submitted to Institute of Electrical Control Engineering

College of Electrical Engineering

National Chiao-Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master

in

Electrical Control Engineering

June 2011

Hsinchu, Taiwan, R.O.C

中華民國一百年七月

在具雜訊的正常血液抹片中進行白血球分類計數

學生：洪裕筆

指導教授：林昇甫 博士

國立交通大學電控工程研究所 碩士班

摘要

對醫學檢驗單位而言，血液抹片人工鏡檢是不可廢除的一項重要檢驗依據，但是人工鏡檢確實是一個耗時耗力的過程，除了血液抹片本身製作上的優劣之外，還得要考慮到不同醫檢人員因為疲勞或者是標準不同而有不一樣的結果，因此，以數位影像分析的技術來協助這項工作的進行，將能夠減少人力的消耗，讓血液抹片鏡檢更有效率。

本論文所研究的對象為具有雜訊的正常血液抹片，雜訊主要來自於製作不良的血液抹片，例如抹片太薄、抹片太厚、染色不均，以及細胞破裂等等因素，在同一個人的血液抹片影像裡面，可能會因為這些原因而在後續的分析中產生不一致的結果，因此，針對這類雜訊進行處理有其必要性，如此一來實驗結果才能具有讓人信服的依據。

本論文的貢獻有三，第一，透過二值化以及區域成長法，能將上述所定義的雜訊排除，進而找出本論文所要尋找的白血球細胞核區域；第二，在找到白血球細胞核區域之後，將可以完成多顆白血球影像定位，以利後續分析；第三，利用距離轉換(distance transform)以及平均值移動演算法(mean shift)，可以找出白血球細胞核分葉特徵，配合形狀特徵以及紋理特徵將能夠使判斷更為精確。

Differential Count of White Blood Cell in Noisy Normal Blood Smear

Student: Yu-Bi Hong

Advisor: Dr. Sheng-Fuu Lin

Institute of Electrical Control Engineering

National Chiao Tung University

ABSTRACT

For the medical examination unit, the artificial blood smear examination is an important test for abolition, but the process is indeed a time-consuming examination process. In addition to making blood smear on their own merits, but also have to take into account the different medical laboratory personnel due to fatigue or different standards then have different results. Therefore, to digital image analysis technology, assisting in this work, will be able to reduce human consumption, making microscopic examination of blood smears more efficient.

The object of study in this paper as having normal blood smear noise, noise mainly from the production of bad blood smears, for example, thin smear, thick smear, stain unevenly, and cell rupture, among other factors, blood smears in the same individual images which may be because of these reasons and in the subsequent analysis produced inconsistent results, therefore, were necessary for the processing of such noise, this way in order to have convincing results basis.

There are three contributions of this paper, first, through the binarization and region growing method, able to rule out noise as defined above, and then find out in this paper to find the white blood cell nucleus area; second, find the white blood cells in the nucleus zone, more satellites will be able to complete the positioning of white blood cell imaging, to facilitate subsequent analysis; Third, the use of distance transform (distance transform) and moving average algorithm (mean shift), you can find leaf characteristics of white blood cell nuclei, with the characteristic shape and texture features will be able to make more accurate judgments.

誌 謝

首先要以最誠摯的謝意，感謝我的指導教授林昇甫博士，感謝他在研究上的啟迪與指導，以及細心校閱本論文的謬誤之處。同時也要感謝本實驗室的學長士哲、晉嘉、國育、啟耀、逸章、俊偉、長安、煒清、世雄、家興，還有同甘共苦的銘達、勁成、則銘以及學弟們，由於他們熱心的協助與討論，使得研究過程中的困境能夠迎刃而解。

最後要感謝我親愛的父母，感謝他們的養育之恩以及在我求學生涯中給我最大的鼓勵與支持，使我得以在精神與生活上無後顧之憂，順利完成學業。



目錄

中文摘要.....	i
英文摘要.....	ii
誌謝.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	vii
第一章 緒論.....	1
1.1 白血球計數之介紹.....	1
1.2 影像雜訊.....	2
1.3 研究動機.....	3
1.4 相關研究之探討.....	3
1.5 論文主體與貢獻.....	5
第二章 相關知識及理論.....	5
2.1 色彩模型轉換.....	5
2.1.1 RGB 色彩模型.....	6
2.1.2 RGB 色彩模型轉換成灰階色彩模型.....	6
2.1.3 RGB 色彩模型轉換成 HSV 色彩模型.....	8
2.2 全域二值化與區域二值化.....	9
2.2.1 最大類間方差二值化法.....	10
2.2.2 尼布蘭克二值化法.....	12
2.3 高斯混合模型.....	13
2.3.1 高斯分布.....	14
2.3.2 高斯混合模型.....	15
2.4 灰階共生矩陣.....	17
2.5 平均值移動演算法.....	20
2.6 支持向量機.....	22
第三章 系統流程.....	24
3.1 整體系統架構.....	24
3.2 RGB 直方圖分析與修正.....	27
3.3 建立紅血球模型.....	29
3.3.1 影像二值化處理.....	29
3.3.2 二值影像邊緣提取.....	31
3.3.3 高原度區塊面積分析.....	32

3.3.4 高原度區塊半徑分析.....	34
3.4 白血球細胞核影像處理與特徵抽取.....	36
3.4.1 色彩空間轉換.....	37
3.4.2 區域成長法復原細胞核區.....	39
3.4.3 細胞核區重新描述及多重細胞核定位.....	43
3.4.4 擷取白血球細胞核特徵.....	48
3.5 分類決策樹.....	54
第四章 實驗結果與分析.....	58
4.1 實驗設備與方法.....	58
4.2 實驗結果分析與討論.....	61
4.2.1 白血球細胞核影像切割.....	61
4.2.2 多重細胞核影像切割及定位.....	64
4.2.3 白血球細胞核分類.....	69
第五章 結論.....	72
參考文獻.....	73



表目錄

表 4.1	訓練及驗證資料。.....	61
表 4.2	白血球細胞核分割計數。.....	63
表 4.3	多重細胞分割定位結果。.....	69
表 4.4	根據本論文所作的分類實驗結果。.....	70
表 4.5	根據參考文獻[15]所作的分類實驗結果。.....	70



圖目錄

圖 1.1	五種類白血球，分別為 (a) 嗜中性粒細胞，(b) 嗜酸性粒細胞，(c) 嗜鹼性粒細胞，(d) 單核細胞，(e) 淋巴細胞。	2
圖 2.1	RGB 色彩模型。	6
圖 2.2	10 灰階。	7
圖 2.3	256 灰階。	7
圖 2.4	HSV 圓錐模型。	8
圖 2.5	最大類間方差二值化的結果。	12
圖 2.6	尼布蘭克二值化結果。	13
圖 2.7	三種不同的高斯分布。	14
圖 2.8	高斯分布的資料量比例。	15
圖 2.9	血液抹片影像之飽和度影像。(a)為原圖 (b)為(a)的立體化	18
圖 2.10	平均值移動演算法示意圖。	22
圖 2.11	支持向量機示意圖。	22
圖 3.1	白血球細胞核影像。(a)血液抹片影像 (b)紅血球影像 (c)白血球細胞核影像。	24
圖 3.2	血液抹片中的雜訊。	25
圖 3.3	血液抹片中的嗜鹼性顆粒。	25
圖 3.4	主要辨識流程。	26
圖 3.5	前處理流程。	27
圖 3.6	RGB 三通道分析。	27
圖 3.7	RGB 三通道未做修正。	28
圖 3.8	經過 RGB 三通道的分析與修正，降低了雜訊。	28
圖 3.9	紅血球模型建立。	29
圖 3.10	二值化範例。(a)RGB 轉灰階圖 (b)最大類間方差法 (c)尼布蘭克二值化法 (d)由(c)和(b)作 AND 運算。	30
圖 3.11	二值影像邊緣提取。(a)血液抹片二值化影像 (b)填補影像中的空洞 (c)則是由(b)作邊緣提取所得的邊緣影像。	31
圖 3.12	正圓度分析。	32
圖 3.13	根據正圓度圈選出可能為紅血球的區塊。	33
圖 3.14	分析高圓度區塊面積直方圖。	34
圖 3.15	高圓度區塊近似半徑直方圖。	35
圖 3.16	白血球細胞核影像處理流程圖。	36
圖 3.17	色彩空間轉換及面積過濾流程圖。	37
圖 3.18	白血球細胞核影像飽和度切割。(a)原圖 (b)來自(a)的飽和度影像 (c)飽和度臨界值(0.75)切割 (d)濾除小區塊之結果	39
圖 3.19	由(a)到(t)依序為區域成長法復原白血球細胞核區塊過程。	43

圖 3.20	複雜血液抹片影像。	44
圖 3.21	複雜血液抹片影像-飽和度。	44
圖 3.22	複雜血液抹片影像-細胞核。	45
圖 3.23	無法連通為一個完整區塊的細胞核。	45
圖 3.24	一個色彩代表一個細胞核。	46
圖 3.25	三組離散細胞核之長短軸分析。	47
圖 3.26	重新描述完整之白血球細胞核影像。	47
圖 3.27	分葉數計算流程圖。	50
圖 3.28	細胞核分葉特徵抽取。(a)為經過區域成長法以及結合邊界資訊 之後的白血球細胞核影像 (b)白血球細胞核像素距離邊緣最近影像 (c)距離影像以三維模型表示 (d)~(i)距離影像以多重臨界值切割的結 果 (l)白血球細胞核分葉總數。	53
圖 3.29	支持向量機二元樹。	55
圖 3.30	五種類白血球。分別為(a)嗜中性粒細胞 (b)嗜酸性粒細胞 (c)嗜鹼 性粒細胞 (d)單核細胞 (e)淋巴細胞	56
圖 4.1	實驗資料製備。(a)劉氏染色劑 AB 劑 (b)染色後血液抹片。	58
圖 4.2	實驗拍攝平台。(a)血液抹片拍攝平台 (b)XY 軸微動平台 (c)LED 背 光板。	59
圖 4.3	掃描血液抹片路徑示意圖	59
圖 4.4	驗證用血液影像。(a)(b)無重疊無干擾影像 (c)(d)紅血球重疊影像 (e)(f)染色不均影像 (g)(h)曝光不足影像。	62
圖 4.5	多重細胞核定位及切割。(a)原圖 (b)細胞核切割且尚未重新描述之 影像 (c)重新描述後影像。	64
圖 4.6	多重細胞核定位及切割。(a)原圖 (b)細胞核切割且尚未重新描述之 影像 (c)重新描述後影像。	66
圖 4.7	多重細胞核定位及切割。(a)原圖 (b)細胞核切割且尚未重新描述之 影像 (c)重新描述後影像。	67

第一章 緒論

本論文所要研究的對象為血液抹片，為了能在廣泛的案例中皆可進行分析，系統抵抗雜訊的能力將會是研究重點，而在此所指涉的雜訊主要是指因不同的製作技術而讓同一人的血液抹片有不同的分析結果，導致結果不具一致性以及正確性，這將是本論文所要解決的主要問題。

本章共分成四節，第一節先概略介紹目前血球技術上有哪些做法以及優缺點，第二節則是闡述本論文的研究動機，第三節則是介紹目前相關研究領域有哪些做法，第四節則是本論文之貢獻。

1.1 白血球計數之介紹

白血球細胞分類計數(white blood cell differential count)，是被廣泛使用的血液常規檢查項目之一，目的在於測定動物體內各類型白血球細胞的含量，提供臨床診斷或者疾病篩檢的線索；事實上，幾乎所有疾病都和白血球有密切關係，透過白血球分類計數來計算不同白血球間的數量比例，將可以協助診斷與篩檢及病。

在目前的醫學應用技術上，主要分成人工鏡檢法和自動血球計算儀兩種方式，人工鏡檢法是把外周血液塗抹於抹片上，進行細胞化學染色，再將抹片置於光學顯微鏡底下，由專業的醫檢人員依據大小、細胞核結構、核質比，以及細胞內顆粒等等細胞型態學上的相對差異，將白血球分成五類：嗜中性粒細胞(Neutrophil)、嗜酸性粒細胞(Eosinophil)、嗜鹼性粒細胞(Basophil)、單核細胞(Monocyte)、淋巴細胞(Lymphocyte)。

而自動血球計算儀在分類時的依據則依廠牌不同而有所不同，主要是利用血球在稀釋液中流動，讓血球一個一個通過小孔道，並且在此孔道中安裝檢測器，油檢測器所得結果，將血球進行分類，主流作法有光學式和電子式兩種，其中，

光學式透過將雷射光照向通過孔道的細胞，由光學散射來分析細胞的大小、容積，以及內部密度等資訊；而電子式則是檢測血球通過孔道時，孔道阻抗的變化量，或者是在孔道中心部位加裝電極，並且施以高頻電流，如此一來將可在血球通過孔道時檢測出脈衝或者是電量的變化，依據這些資訊則可以對血球細胞做分類。

雖然自動化血球計算儀相較起人工鏡檢有許多的優勢，但是在血球計算儀的分類依據內卻欠缺了一個血球分類的重要因素：細胞型態學，因此血球計算儀只能對常規血液進行分類，對於不典型，或者是異常型態的細胞則無法檢出，只能對使用者提出警告，再由使用者來做人工檢查，因此自動化血球計算儀仍然無法完全取代人工鏡檢。

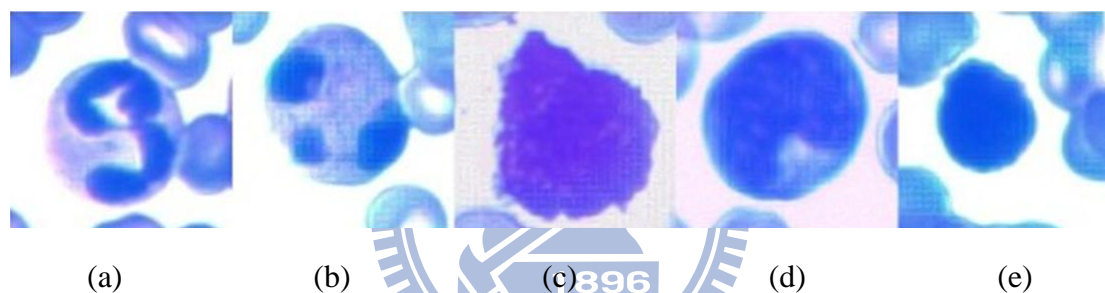


圖 1.1 五種類白血球，分別為 (a) 嗜中性粒細胞，(b) 嗜酸性粒細胞，(c) 嗜鹼性粒細胞，(d) 單核細胞，(e) 淋巴細胞。

1.2 影像雜訊


對於影像雜訊較為廣泛的說法，是指因為外來干擾而導致影像品質劣化，這個劣化現象通常可以用某種數學模型來表達，例如胡椒鹽雜訊(impulse noise)、高斯雜訊(Gaussian noise)因此，在確定雜訊內容為哪種類型的模型之後，通常都可以提出一個較為可行的減低雜訊方案來解決掉雜訊問題。

上述關於雜訊的討論，主要著重在「影像退化」，所以試著要找出一些方法來讓影像品質恢復到一定水準，這類型的問題通常是指某些無用的訊號干擾了我們真實想要擷取的訊號，而這些無用的訊號就被稱為雜訊；有別於此，另一類雜

訊的定義並非是指無用的訊號，而是在某種情況下，某些訊號可以被視為是雜訊，例如，當 X 光穿透人體時，不可避免地會產生折射，使得許多不該曝光的部分曝光了，因此這些部分將被視為是雜訊；事實上，這些曝光的部分和之前所提到的胡椒鹽雜訊以及高斯雜訊有很大的不同，此處所指的雜訊除了無法用一個數學模型來表達，這些被視為雜訊的對象，也是「真的」訊號，只是因為在研究過程中鎖定的對象不同，所以為了避免研究結果受到干擾，而需要被排除，實際上，

而在血液抹片中，則因抹片製作技術上的差異，而使得染色情況不甚一致，除了影響了數位影像的判讀之外，更嚴重的話可能也會干擾醫檢人員的判斷，因此，本論文所想要解決的雜訊問題，就是「在同一隊像的血液組織內卻呈現出不一樣染色情況」。

1.3 研究動機



自動化血球計算儀大量使用了光學以及電學理論，但是卻缺少了血球型態學的資訊，在人類視覺無法觸及的環境下，自動血球計算儀的功能也就因此而受限了；倘若能夠引入數位影像分析的方法，則因為數位影像分析非常具有直觀性，除了可以單獨鑑別異常細胞之外，也可以學習人工鏡檢法的經驗與病理學家的專業知識，充分發揮計算機視覺與病理學結合的優勢，因此如果發展一個能自動識別白血球細胞的系統，將具有重要的意義。

1.4 相關研究之探討

近年來，以數位影像分析血液抹片以計算白血球細胞數量的研究已經獲得了許多成果，而在這個研究方向又分成影像切割，以及血球識別，首先討論影像切割的部分，參考文獻[8]透過設定臨界值將白血球細胞核自灰階影像中切割出來，再以主動輪廓來搜尋白血球細胞質的邊界，而參考文獻[9]為了克服染色過深區塊，以及為了產生主動輪廓所需要的初始輪廓，進一步使用了型態學，但是卻因

此造成了細胞核區域的失真，標示出來的細胞核區塊與肉眼所視細胞核區域有很大的差別，然而，較為特別的是，[9]已經開始注意到當把問題提高到較為普遍的狀況時，得要開始處理本論文所提到的染色問題；由於[8]以及[9]都採用主動輪廓作為解決白血球細胞影像切割的方法，因此都產生了幾個問題，第一，主動輪廓受到初始輪廓的影響太大，如何找到適當的初始輪廓就是一個很大的問題；第二，主動輪廓所耗費的時間太久，通常需要經過長時間的迭代次數方能找到適當的邊界；第三，有些細胞的細胞質含量極少，在數位影像上會看到幾乎只有細胞核，倘若以[8]、[9]的做法來做細胞核切割，主動輪廓很有可能會發散，而無法找到最佳解，因此主動輪廓可能不是個恰當的想法。

參考文獻[10]對於影像切割則沒有執著在整個白血球細胞影像的切割，而是著重在白血球細胞核切割，因為白血球細胞的分類上比較明顯的特徵還是來自於白血球細胞核；[10]將每個像素的 RGB 三通道數值視為一個向量，將影像進行史密特正交化(Gram-Schmidt orthogonalization)，藉此來分割出白血球細胞核，但是這樣的做法忽略了染色差異，不同染色技術以及不同操作人員都可能會對血液抹片有不一樣的影響，以 RGB 通道數值作為標準，並不是很可靠的作法。

參考文獻[11]則想要用交離變換的方式來定位白血球細胞核，但是交離變換只能找出特定的像素結構，在白血球五種類別中，除了淋巴球以外，其餘五種類別的白血球都具備有兩種以上的變體，只以特定結構來分析白血球細胞，只能對淋巴球有一定的效果，對於其他類型的白血球則沒有明確的規則可以遵守。

因此，若想要在血液抹片上利用影像切割來識別白血球，有一個最重要的問題得要解決，也就是「對於同一個對象的血液抹片分析，是否都能產生一致的結果？意即分析系統的結果將具有可重現性？」，在這個問題框架之下，首要解決的就是影像切割，這也將是本論文將要嘗試解決的問題之一。

1.5 論文主體與貢獻

在進行血液抹片分析的時候，系統的流程主要可以分成三個步驟：影像切割、特徵抽取，以及分類。

因此本論文將在第二章介紹相關技術以及原理，在第三章則介紹本論文將提出的系統之流程，第四章則是與近年相關領域論文結果之比較與分析，第五章則是結論與未來展望。

而在本論文的研究中，將會把重點集中在影像切割，以及特徵抽取，而在分類的部分將會採用前人所作的支持向量機二元樹(support vector machine binary tree)，本論文所達成的貢獻有三，如下所示：

1. 當血液抹片影像出現本論文所指涉的雜訊時，本論文所提出的方法可以在很大程度上排除掉這類雜訊。
2. 可以對多顆白血球進行同時定位以及分析，不需要限定影像只能存在單一白血球細胞。
3. 分析出白血球細胞核分葉數目，提供粒細胞以及非粒細胞明確的鑑別資訊。

第二章 相關技術與原理

本章節將會介紹與本論文有關的知識與理論，將從影像處理以及圖型識別的角度來做介紹，在 2.1 節介紹關於本論文所使用的色彩模型轉換，在 2.2 節則會介紹影像二值化，包含全域二值化以及區域二值化，在 2.3 節則會提到高斯混合模型(Gaussian mixture model)，在本論文中常以高斯分佈或者是高斯混合模型來建立影像模型，或者是作物體切割與選取；2.4 節介紹灰階共生矩陣；2.5 節則會介紹平均值移動演算法(mean shift algorithm)，本論文使用平均值移動演算法來對未知分群數目的資料作分析，以取得分群數目；2.6 節則是介紹本論文用來做細胞核分類的主要演算法，也就是支持向量機(support vector machine)的原理。

2.1 色彩模型轉換

本論文中所使用到的色彩模型包括 RGB、HSV，還有灰階(gray-level)影像，針對不同的處理階段，需要不一樣的色彩模型輔助。

在本論文中，色彩模型的轉換，主要目的在於輔助影像切割，使影像切割的結果更接近實驗需求；在進行影像切割的時候，通常假設影像包含了感興趣的區域與不感興趣的區域，兩者通常可以由某種特性來做明顯的區分，例如：強度(intensity)、飽和度(saturation)，以及色調(hue)等等，根據不同需求而採用不同的色彩模型，而本論文基於兩個目的，第一個目的是紅血球模型建立，需要將 RGB 色彩模型轉換成灰階色彩模型，以獲得像素強度資訊，可利用於後續的邊緣提取；另一個目的則是白血球細胞核搜尋，主要參考的特性為細胞核區塊的飽和度遠高於背景以及其他血液中的血球，因此將 RGB 色彩模型轉換成 HSV 色彩模型，以取得飽和度影像。

2.1.1 RGB 色彩模型

在 RGB 色彩模型，每種色彩是由紅、綠、藍三個主要頻譜成分來顯示，倘若把 RGB 色彩模型以可視化模型來表示，如圖 2.1，則可以發現紅、綠、藍三種分量構成了一個立方體，且三種分量互相正交，一張以 RGB 色彩模型表示的影像，得要用上三個數值來描述同一個像素；用來表示像素(pixels)所需的位元數稱為「像素深度」，一張以 RGB 色彩模型表達的影像，其紅、綠、藍影像皆為 8 位元影像，則每個 RGB 像素需用 24 位元來表達。

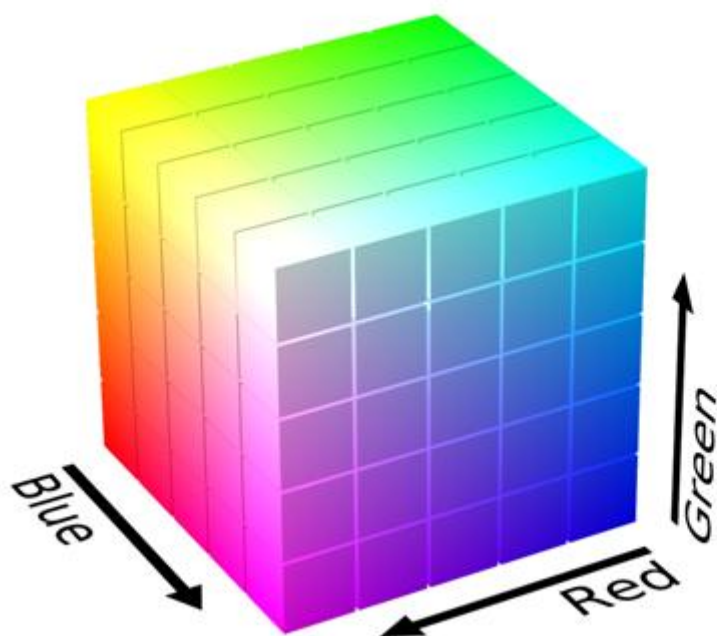


圖 2.1 RGB 色彩模型

2.1.2 RGB 色彩模型轉換成灰階色彩模型

灰階色彩模型，有別於 RGB 色彩模型得要用三個數值來描述單一像素，灰階色彩模型只有單一數值，灰階模型所呈現出來的影像，代表著影像的強度 (intensity) 資訊，最弱灰階為黑色，最強灰階為白色，灰階模型則是在這兩個灰階限制下，計算出從最弱灰階到最強灰階中的變遷灰階，圖 2.2 將最弱到最強灰

階之間額外區分出 8 個灰階值。



圖 2.2 10 灰階

而在計算機影像處理領域中，最常用的灰階模型具有 256 強度，這是因為電腦常用的 1 位元組等於是 8 位元，而 8 個位元就代表最低可顯示的十進位數字為 0，最高可顯示的十進位數字為 255，共有 256 個數值，此外也就代表著一個像素可以用 1 位元組來代表，另一方面也因為 256 灰階也大約是一般人眼所能分辨的灰階數目，所以才會使用此數值作為常用灰階數目。



圖 2.3 256 灰階

在做 RGB 模型轉灰階模型的時候，並不是將 RGB 三個成分相加再除以 3，而得要考慮人眼對於綠色的感受度較高，對於藍色的感受度較低，因此在考慮這個因素之後，轉換出來的灰階影像才會與人眼觀察影像所感受到的明暗變化類似，詳細轉換公式如下：

$$Gray = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (2.1)$$

2.1.3 RGB 色彩模型轉換成 HSV 色彩模型

HSV 色彩空間是從 RGB 色彩空間經過計算而得的一種表達色彩的方式，H 代表色相(hue)、S 代表(saturation)，V 則是代表著明度(brightness)；比起 RGB，HSV 可以更為適當地聯繫人類的感知顏色，因此，HSV 常常被用於計算機視覺或者是影像處理相關領域。

倘若將 HSV 模型以一個可視覺化的方式表達，將會是一個圓錐體，色相代表著繞圓錐體中心軸的角度，而飽和度則為圓錐體的某橫截面上的一點到橫截面圓心的距離，明度則被表示為從圓錐體的橫截面積的圓心，到圓錐體頂點的距離，如圖 2 所示，此可視化模型很適當地表達的 HSV 色彩空間的概念。

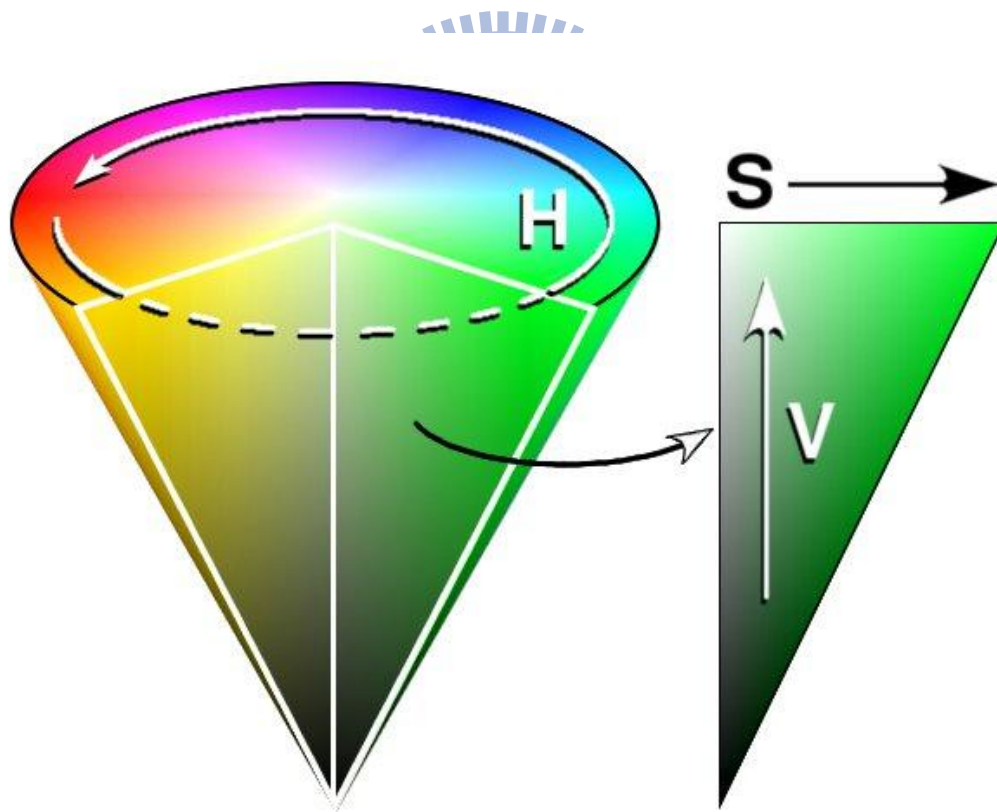


圖 2.4 HSV 圓錐模型

為了將 RGB 色彩空間轉換到 HSV 色彩空間，得要先將 RGB 三通道數值進

行正規化(normalize)。

$$r = \frac{R}{256}, \quad g = \frac{G}{256}, \quad b = \frac{B}{256} \quad (2.2)$$

$$\max = \max(r, g, b), \quad \min = \min(r, g, b) \quad (2.3)$$

$$h = \begin{cases} 0^\circ & \text{if } \max = \min \\ 60^\circ \times \frac{g-b}{\max-\min} + 0^\circ & \text{if } \max = r, g \geq b \\ 60^\circ \times \frac{g-b}{\max-\min} + 360^\circ & \text{if } \max = r, g < b \\ 60^\circ \times \frac{b-r}{\max-\min} + 120^\circ & \text{if } \max = g \\ 60^\circ \times \frac{r-g}{\max-\min} + 240^\circ & \text{if } \max = b \end{cases} \quad (2.4)$$

$$s = \begin{cases} 0 & \text{if } \max = 0 \\ \frac{\max-\min}{\max} = 1 - \frac{\min}{\max} & \text{otherwise} \end{cases} \quad (2.5)$$

$$v = \max \quad (2.6)$$

$h \in [0, 360]$ 是代表色相的色相角；而 $s \in [0, 1]$ 與 $v \in [0, 1]$ ，則分別代表著飽和度(saturation)與明度(brightness)。

2.2 全域二值化與區域二值化

影像的二值化處理就是將影像上的點的灰階設定為最弱色階或者是最高色階，也就是讓整張影像從原本的灰階影轉換成黑白兩色，在 256 個灰階等級的灰階模型中，黑色代表灰階值為 0，白色代表灰階值為 255，通過適當的臨界直選取，將會從灰階影像獲得可以反映影像整體與局部特徵的二值化影像。

在數位影像處理中，二值化處理佔有非常重要的地位，在實際應用的圖像處理系統中，以二值化影像處理實現與建構的系統非常多；要進行二值化影像處理與分析的時候，得要先將灰階影像轉換為二值影像，得到二值影像之後，在處理上將有一個很大的優勢，就在於影像的集合性質只與像素值是 0 或者是 255 有關，

不再涉及像素的灰階變遷，使處理變得更簡單，也使影像的數據處理量縮小。

常見的二值化方式是設定一個臨界值，倘若一像素之灰階大於此臨界值，則判斷此像素屬於某特定物體，將其灰階重新設定為最高灰階；否則該像素將被排除在物體區域外，灰階值將被設定為最低灰階，以表示此像素屬於背景或者其他物體區域，類似此類作法被稱為全域二值化法；相對於全域二值化，也存在著區域二值化方法，可以進一步考慮到影像的細部特性。

在本論文中採用了兩種二值化方式，第一種為最大類間方差法(Otsu)二值化，另一種則為尼布蘭克(Niblack)二值化，前者為全域二值化，考慮整張影像的特性之後，設定一個臨界值來將整張影像進行二值化處理；後者則為區域二值化，則是先建立一個遮罩，將此遮罩滑動過整張影像，每次所處理的範圍限定在此遮罩範圍內，依照某種判斷準則，將遮罩內影像作二值化處理，此一過程持續到整張影像都被處理過為止，如此一來可以透過限定處理範圍取得更為細部的資訊；本節分成兩部分，第一部分先介紹最大類間方差二值化法，接著再介紹尼布蘭克二值化法。

2.2.1 最大類間方差二值化法

最大類間方差二值化法，是在 1975 年由大津(Nobuyuki Otsu)所提出的二值化方法；將灰階影像轉換二值化影像是影像處理中一個很基本的處理過程，最大類間方差法包含了反覆計算所有可能的臨界值，並且計算出這些臨界值兩側的像素灰階分布，目的是找出一個臨界值，可使類內方差值(within class variance)最小，則此臨界值即為最佳的分割值，圖 2.5 即為以最大類間方差法對血液抹片影像進行全域二值化的結果。

先定義變數如下： N_p 代表整張影像的像素總數， N_{p_b} 代表背景部分各個灰階所佔的像素數目， P_b 代表了全背景像素灰階的集合， W_b 為背景之權重(waigt)， μ_b 為背景之均值(mean)， σ_b 為背景之方差(variance)； N_{p_f} 代表前景部分各個灰

階所佔的像素數目， P_f 代表了全前景像素灰階的集合， W_f 為前景之權重， μ_f 為前景之均值， σ_f 為前景之方差。

背景
$$W_b = (\sum P_b) / Np \quad (2.7)$$

$$\mu_b = [\sum_i P_b(i) \times Np_b(i)] / \sum Np_b \quad (2.8)$$

$$\sigma_b = [\sum_i Np_b(i) \times ((P_b(i) - \mu_b))^2] / \sum Np_b \quad (2.9)$$

前景
$$W_f = (\sum P_f) / Np \quad (2.10)$$

$$\mu_f = [\sum_i P_f(i) \times Np_f(i)] / \sum Np_f \quad (2.11)$$

$$\sigma_f = [\sum_i Np_f(i) \times ((P_f(i) - \mu_f))^2] / \sum Np_f \quad (2.12)$$

類內方差
$$\sigma_w^2 = W_b \sigma_b^2 + W_f \sigma_f^2 \quad (2.13)$$

類間方差
$$\sigma_w^2 = W_b W_f (\mu_b - \mu_f)^2 \quad (2.14)$$

原先定義的類內方差值為(2.13)，但是因為計算量太過龐大，因此採用另一個等價觀點來解決這個問題，當類間方差(between class variance)最大的時候，會發現類內方差最小，因此找到一臨界值使(2.14)最大值時，(2.13)則會是最小，詳細演算法如下：

最大類間方差二值化演算法

1. 計算直方圖，以及每個色階的機率分佈
2. 設定初始的 W_b 、 μ_b 、 W_f 以及 μ_f
3. 臨界值設定從最低色階變化到最高色階，每變換一次臨界值，就得要做下列操作：
 - a. 更新 W_b 、 μ_b 、 W_f 以及 μ_f

b. 計算類間方差 σ_w^2

4. 最佳的臨界值設定對應到最大的類間方差。

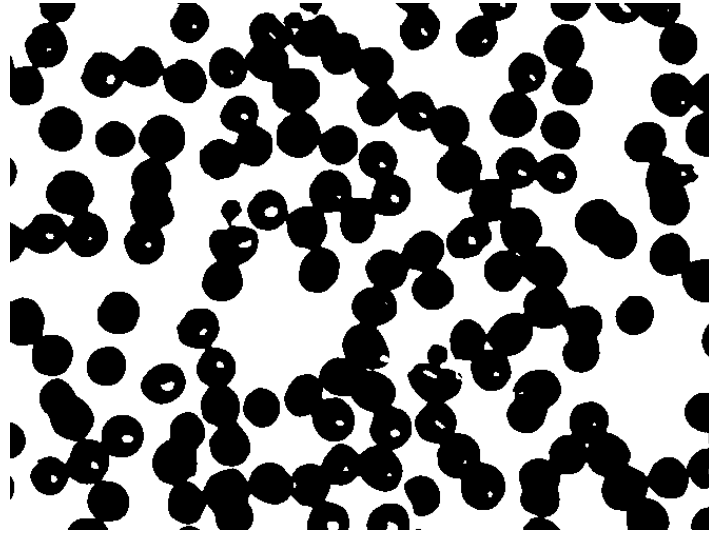


圖 2.5 最大類間方差二值化法的結果

2.2.2 尼布蘭克二值化法

區域二值化方式並不像是空間濾波(spatial filtering)將濾波系數與濾波遮罩所涵蓋區域內的影像像素之乘積作為其響應結果，而是在這個濾波遮罩涵蓋的範圍內，對此範圍分析其灰階分布，接著可以得到此遮罩涵蓋區域內的二值化影像；將此遮罩滑動過整張影像，並且每移動一次都要進行相同的分析與處理，透過這樣的方式，最終將會得到一個考慮到區域特性的二值化影像，圖 2.6 即為利用尼布蘭克二值化法對血液抹片影像進行區域二值化的結果。

尼布蘭克二值化方法則是利用統計學的知識來進行區域二值化，詳述如下：

$$T(x, y) = m(x, y) + k \cdot s(x, y) \quad (2.7)$$

先取一個寬度為 w 的矩形框，以 (x, y) 為這個框的中心，依此建立遮罩；在遮罩內的運算則是以統計框內數據為主， $T(x, y)$ 為區域臨界值， $m(x, y)$ 為平均值， $s(x, y)$ 則為均方差， k 則為一個自訂參數，一般設定為(-2)；計算出 $T(x, y)$ 之後，就可以對此遮罩內影像進行二值化處理，高於區域臨界值者判定為 255，低於區

域臨界值者判定為 0。

雖然尼布蘭克二值化方法，具有快速以及效果好的優點，並且相較於全域二值化來說，還能夠保留細部特性，但缺點是會在平滑區塊產生雜訊，這是需要重視的一個問題。

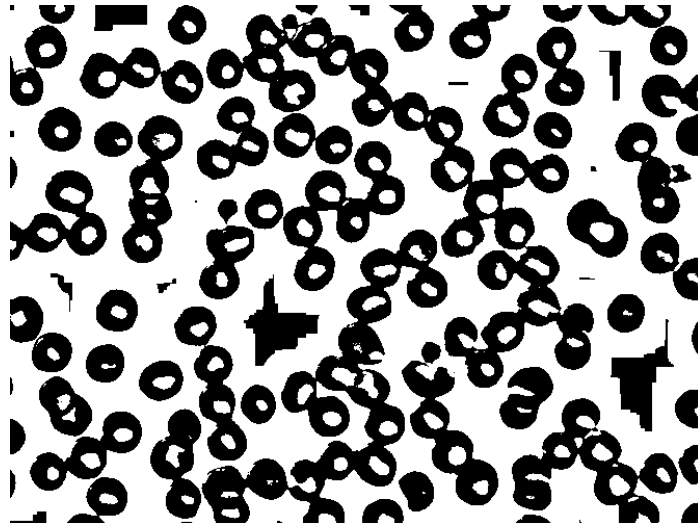


圖 2.6 尼布蘭克二值化結果

2.3 高斯混合模型

高斯分布(Gaussian distribution)，是在數學以及自然科學中的定量現象的一個簡單模型，儘管這些現象是未知的；目前在統計學以及多種統計測試中被廣泛地使用。

而高斯混合模型(Gaussian mixture model)，則是高斯分布的延伸，簡稱 GMM，他的特色在於能夠利用高斯分佈模型來平滑地近似某特定的機率分佈。

高斯混合模型利用高斯機率密度函數，也就是常態分布曲線來量化資料，將一個影像分解為若干基於高斯密度函數形成的模型，對影像建立高斯模型的原理以及過程，常用於前景與背景的分割；因為灰階直方圖反映的是影像中某個灰階值出現的次數，也可以看成是影像灰階機率密度的估測，如果影像所包含的目標區域與背景具有一定程度上的差異，那麼該影像的灰階直方圖將會出現明顯的雙

峰，以及存在兩者間明顯的波谷，其中一個波峰對應於目標，另一個波峰對應於背景；對於複雜的影像，通常在灰階直方圖的波形通常會顯現多重峰值，可以將波形看成是多個高斯分布的疊合，透過波峰與波谷的概念，可以選定某特定高斯函數作為切割複雜影像中的物體的依據。

在本論文中高斯混合模型被大量地使用，除了影像分割之外，也用於建立模型，包括紅血球模型，以及白血球細胞核相似度模型，或者是用來從複雜影像中挑選具有某些特性的區塊；本節將分成兩個部分，先介紹高斯分布，再介紹高斯混合模型。

2.3.1 高斯分布

高斯分布又稱為常態分布，在數學、物理以及工程領域都扮演著很重要的角色；對於隨機變量 X ，倘若 X 滿足期望值 μ 、方差為 σ^2 的高斯分布，則可以說， $X \sim N(\mu, \sigma^2)$ ，其機率密度函數為(2.8)；設定不同的參數值則可產生不同的高斯分布，期望值決定了高斯分布的中心所在位置，而均方差則決定了高斯分布圖形的陡峭平滑程度，可由圖 2.7 看出參數不同所導致的不同結果。

$$p(x) = e^{-[(x - \mu)^2 / 2\sigma^2]} / \sigma\sqrt{2\pi} \quad (2.13)$$

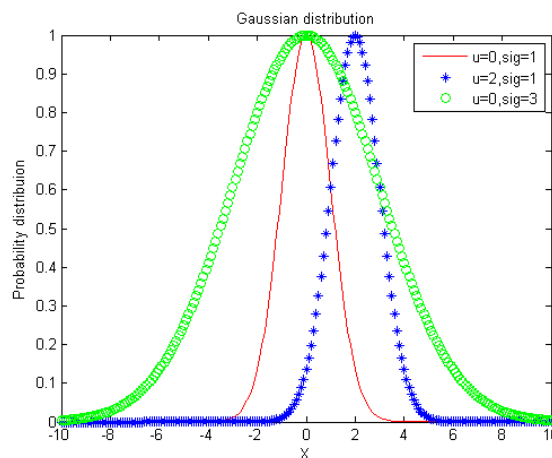


圖 2.7 三種不同的高斯分布

高斯分布具備有以下這些特性：

1. 密度函數對於平均值對稱，可由圖 2.8 看出這特性
2. 平均值是它的眾數 (statistical mode) 以及中位數 (median)
3. 函數曲線下 68.26% 的面積在平均值左右的一個標準差範圍內
4. 95.44% 的面積在平均值左右兩個標準差 2σ 的範圍內
5. 99.72% 的面積在平均值左右三個標準差 3σ 的範圍內

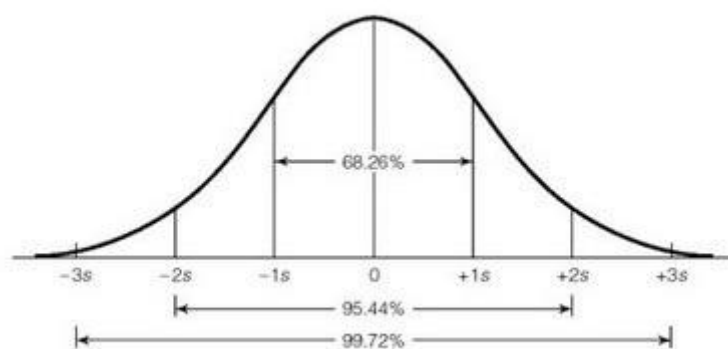


圖 2.8 高斯分布其資料量比例

2.3.2 高斯混合模型

對於觀測數據 $X = \{x_1, x_2, \dots, x_N\}$ 中，取其成員 x_i ，其高斯混合分布的機率密度函數為

$$P(x_i | \Theta) = \sum_{k=1}^K \pi_k p_k(x_i | \theta_k) \quad (2.14)$$

其中 π_k 是混合係數，代表每個混合模型成分的先驗機率 (prior possibility)，而 $\Theta = (\theta_1, \theta_2, \dots, \theta_K)$ ，則是各個混合成分的參數向量， $\theta_k = (\mu_k, \sigma_k^2)$ ，是高斯分布的參數，也就是期望值與均方差。

倘若已經知道分布的形式，當所要估計的參數是非隨機的未知常數，或者待測參數是隨機，但是先驗機率未知的情況，就通過使機率密度函數最大化來決定

參數的數值為何，這就是最大似然估計：

$$P(X | \Theta) = \prod_{i=1}^N P(x_i | \Theta) = \text{Like}(\Theta | X) \quad (2.15)$$

$$\Theta^* = \arg \max(\text{Like}(\Theta | X)) \quad (2.16)$$

對於高斯混合分佈而言，求(2.16)式的極大值並不是一件簡單的事，因為先驗知識的缺乏，也就是不知道第 k 個成分究竟生成了哪些 x_i 向量，也不知道哪一個 x_i 影響到哪一個 θ_k ，倘若對(2.15)式直接求解偏導數，則會得到多個根的超越方程式，因此，需要找一個方法來解決這個模糊性質，而最大期望值演算法(expectation maximization algorithm)就是一個可以克服先驗知識缺乏的最佳化演算法。

最大期望值演算法是一種由「不完整數據」中求解模型分佈參數的最大似然估計的方法[14]，而所謂的不完整數據通常指向兩種狀況，第一種是由於觀測過程本身的限制或者錯誤，造成了觀測數據有所遺漏的不完整數據，另一種則是因為參數的似然函數想要直接最佳化相當困難，但是引入了額外的參數後就比較容易最佳化，因此定義原始觀測數據再加上額外的參數所組成的數據稱之為「完整數據」，所以，相對於完整數據而言，原始數據自然就成為了「不完整數據了」。

引入了額外數據 Y 之後，定義 $Z = \{X, Y\}$ 為完整數據，則觀測數據 X 就是不完全數據，同時，前面所述的似然函數(2.15)式也被稱為「不完全數據的似然函數」，而「完全數據的似然函數」為：

$$L(\Theta | Z) = L(\Theta | X, Y) = P(X, Y | \Theta) \quad (2.17)$$

若將 X 和 Θ 視為固定，(2.17)式就是一個以隨機變數 Y 為自變量的函數，則這個完全數據的似然函數也可以被看成是一個隨機變數；雖然隨機變數想要求解最大值有點困難，但是隨機變數的期望值卻是一個明確的函數，因此最佳化就容易得多了，這也是最大期望值演算法的基本想法，以下介紹最大期望值演算法。

最大期望直演算法透過兩個基本步驟交替進行，來達到最佳化的目的，第一個步驟為計算期望值(expectation)，第二個步驟為最大化(maximization，透過最

大化在計算期望值步驟所求得的最大似然值來計算參數的值，在最大化步驟上所取的參數估計值將被用在下一次迭代中的計算期望值步驟中。

計算期望值步驟(E-Step)

$$Q(\Theta, \Theta^t) = E[\log P(X, Y | \Theta) | X, \Theta^t] \quad (2.18)$$

最大化步驟(M-Step)

$$\Theta^{t+1} = \arg \max Q(\Theta, \Theta^t) \quad (2.19)$$

基本步驟只是呈現一種概念，想要實現演算法，其關鍵在於尋找隱藏的數據 Y 的機率密度表示式。

在高斯混合模型中，引入預設的變數，也就是預設觀測數據與混合成分的對應關係的變數，將此變數作為隱藏數據[4]，就可以推導出預設變數的機率密度公式，從而推導出最大期望值演算法求解高斯模型的迭代公式：

$$a_k^t = \sum_{i=1}^N p(k | x_i, \Theta^t) \quad (2.20)$$

$$\pi_k^t = \frac{1}{N} a_k^t \quad (2.21)$$

$$\mu_k^{t+1} = \frac{1}{a_k^t} \sum_{i=1}^N x_i p(k | x_i, \Theta^t) \quad (2.22)$$

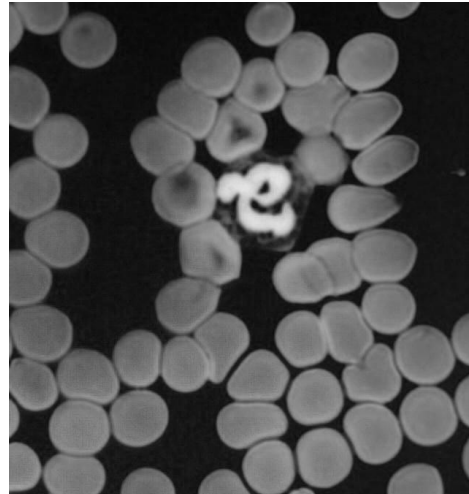
$$\sigma_k^{t+1} = \frac{1}{a_k^t} \sum_{i=1}^N p(k | x_i, \Theta^t) (x_i - \mu_k^{t+1})(x_i - \mu_k^{t+1})^T \quad (2.23)$$

其中， $p(k | x_i, \Theta^t) = \frac{\pi_k^t p_k(x_i | \Theta^t)}{\sum_{j=1}^K \pi_j^t p_j(x_i | \Theta^t)}$

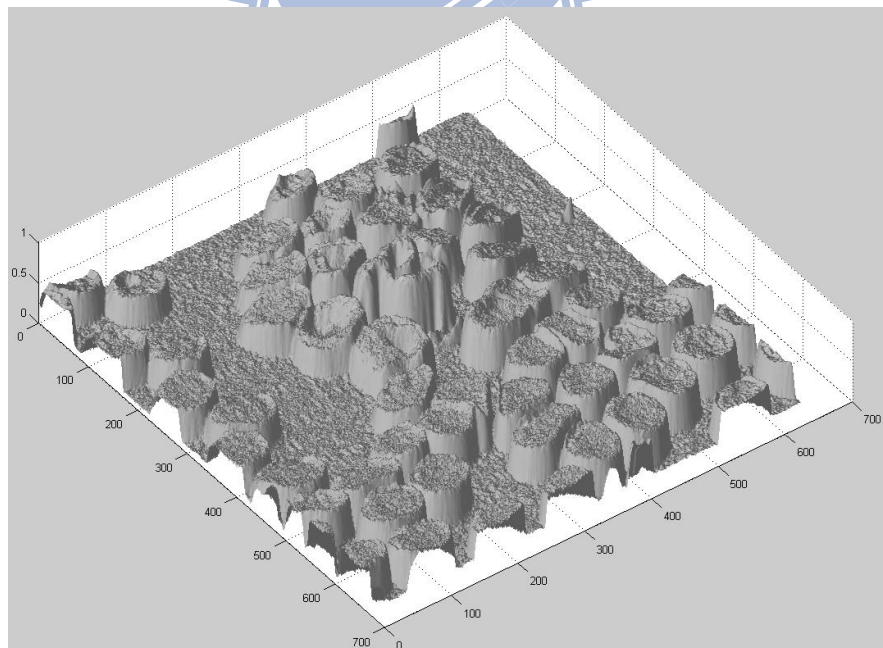
2.4 灰階共生矩陣

灰階共生矩陣(gray level co-occurrence matrix)，可以提供影像灰階的分佈特性，也可以提供具有相同灰階或者式不同灰階之間的位置分布特性，可以透過灰階共生矩陣來定義一組紋理特徵。

任何影像事實上都可以看成是三維空間中的一個曲面，如圖 2.9 所示；考慮到紋理是由於某灰階分佈在空間位置上反覆出現而形成，因此在上述三維空間中相隔某固定距離的兩個像素之間會存在某種灰階關係，這就是影像中灰階的空間相關特性，而灰階共生矩陣就是用來描述這個特性的常用方法。



(a)



(b)

圖 2.9 血液抹片影像之飽和度影像：(a)為原圖；(b)為(a)的立體化。

灰度共生矩陣是由影像中保持某固定距離的兩個像素之灰階值在經過統計之後而得；假定影像大小為 $N \times N$ ，任一點 (x, y) 與距離 (dx, dy) 的另一點 $(x + dx, y + dy)$ 的灰階值分別為 g_1 與 g_2 ，記錄為 (g_1, g_2) ； (x, y) 在曲面上移動，在掃瞄過整個曲面之後，統計出每一種 (g_1, g_2) 出現的次數，並且以 (g_1, g_2) 為兩座標軸，將次數記錄在相對應的座標點 (g_1, g_2) 上面，因此倘若原始影像的灰階值上限為 L ，則所得的矩陣大小為 $L \times L$ ，接下來再利用總次數對這個矩陣作正規化處理(normalization)，經過這樣的處理之後，兩像素之灰階同時發生的機率，就從空間座標，轉換成灰階對 (g_1, g_2) ，而最後所得的機率密度方陣即為灰度共生矩陣。

在灰度共生矩陣形成的過程中，距離並不能夠任意設定，對於較細緻的紋理得要設定較小的距離，對於較粗糙的紋理則選擇較大的距離。

定義灰階共生矩陣在 (g_1, g_2) 座標的機率值是 $p(g_1, g_2)$ ，因此可以由灰階共生矩陣定義出一組紋理特徵：

1. 能量(energy)：

$$E = \sum_{i,j} (p(i, j))^2 \quad (2.24)$$

能量是影像灰階分布均勻性的度量，當灰階共生矩陣的元素分布較集中於主對角線附近的時候，代表局部區域內影像灰階分布較為平均，因此紋理分布較粗糙的時候能量相對於紋理分佈細緻時的能量較大。

2. 對比度(contrast)：

$$Con = \sum_{i,j} (i - j)^2 p(i, j) \quad (2.25)$$

對比度反映了局部區域的變化情形，倘若局部變化很大，則對比度越大，可以將對比度理解成影像的清晰程度，也可以說是紋理的清晰程度。

3. 相關度(correlation)：

$$Cor = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (2.26)$$

相關度是用來描述灰階共生矩陣中的元素在行或者是列方向上的相似程度，提供了某個灰階延某方向的延伸長度，若延伸越長，則相關度越高。

4. 同質性(homogeneity)：

$$H = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (2.27)$$

同說明了紋理局部變化的程度，倘若影像紋理的不同區域間缺乏變化，則同質性將會很高，影像局部均勻。

5. 熵(entropy)：

$$Et = -\sum_{i,j} p(i,j) \times \log_2(p(i,j)) \quad (2.28)$$

代表了影像中所具有的信息量也可以說是影像中紋理的非均勻程度、複雜程度；當共生矩陣中所有元素具有最大的隨機性的時候，共生矩陣所有值均相同，熵較大，

2.5 平均值移動演算法



平均值移動演算法(mean shift algorithm)這個概念最早是由 Fukunag 及 Hostetler [6]所提出，平均值移動並非是指平均值的移動，平均值移動指的是一個向量，指的是偏移的均值向量。

平均值移動分群法是一個迭代的過程，先計算出目前的偏移均值，移動該點到即偏移均值，然後以此作為新的起始點，繼續移動，直到滿足一定條件時才結束這個迭代過程；1995年 Cheng 引入了權值的概念[7]，使得不同的取樣點對於偏移均值的貢獻也有所不同，大幅擴充了平均值移動分群法的功能，在圖型識別、影像切割，以及目標物追蹤等等領域，都具體的應用。

平均值移動演算法是一個非參數性的分群法，除了不需要事先知道分群數目之外，也不需限定分群的形狀，因此對於欲從未知群數的資料中挖掘出該資料可

能的分群數此一目標來說，平均值移動演算法是個非常恰當的方法，以下開始介紹平均值移動演算法的詳細內容：

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (2.29)$$

$$K(x) = c_{k,d} k(\|x\|^2) \quad (2.30)$$

有 n 個資料點 x_i ， $i=1,2,\dots,n$ ，分佈於 R^d ，而 $c_{k,d}$ 是個正規化常數，目的是讓 $K(2.30)$ 的積分等於 1，再由核函數 $K(2.30)$ 以及視窗半徑為 h ，可以得到多變量核密度估測 $f(2.29)$ 。

$$\begin{aligned} \nabla f(x) &= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x_i - x) g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \\ &= \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \left[\frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right] \right] \end{aligned} \quad (2.31)$$

其中 $g(s) = -k'(s)$ ，前一項正比 $G(x) = c_{g,d} g(\|x\|^2)$ 於在 x 的密度估測，而第二項則是偏移均值向量：

$$m_h(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad (2.32)$$

偏移均值向量永遠指向密度中最高的增加量，而接下來的平均值移動流程如下：

1. 透過 $m_h(x^t)$ 計算偏移均值向量
2. 移動遮罩視窗 $x^{t+1} = x^t + m_h(x^t)$
3. 重複 1 和 2 步驟，直到滿足收斂條件為止，例如： $\nabla f(x_i) = 0$

圖 2.10 為平均值移動演算法示意圖。

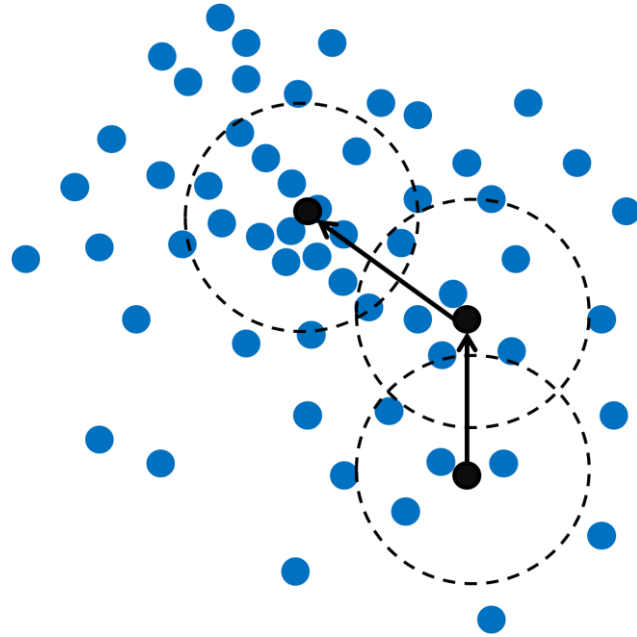


圖 2.10 平均值移動演算法示意圖

2.6 支持向量機

對一群在 R^d 空間中的資料，為了將該資料分成兩群，需要在該空間內找到一超平面(hyper-plane)來將該資料做區隔，如圖 2.11 所示：

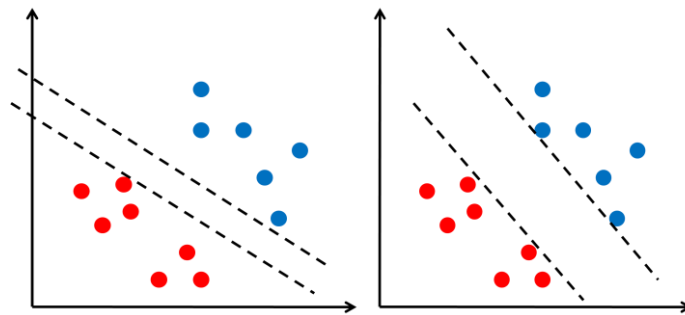


圖 2.11 支持向量機之範例圖

虛線部分為超平面所在位置，比較左右兩圖發現，在右圖的超平面間隔較大，此間隔稱為幾何間隔，間隔較大也就意味著，以右圖的超平面間隔寬度，進行分

類時較不容易發生錯誤分類的情形，因此右圖的超平面比起左圖的超平面要來得好。

一般在低維度的資料往往會因為資料的數量提高，漸漸地變成線性不可分的情況；雖然在低維度無法找到線性可分的超平面但是倘若思考將低維度的資料映射到高維度，就會發現低維度的資料有機會成為線性可分的資料，所以為了能夠最大化幾何間隔，現在得要找個方法將低維度的資料應設到高維度。

用來映射的函數即稱為「核函數」，透過核函數，可以將低維度資料朝高維度應設。

因此支持向量機的問題，可以視為是個優化問題：

$$\min : \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (2.33)$$

$$\text{subject to : } y_i [(wx_i) + b] \geq 1 - \xi_i \quad i = 1 \sim l; \xi_i \geq 0 \quad (2.34)$$

其中 ξ_i 在這裡是用來解決離群點的問題，又稱為鬆弛變數；而 C 則是懲罰因數，作用在於衡量對於離群點的重視程度。

第三章系統流程

有鑑於雜訊對於血液抹片的影響之劇，本論文提出一個有效率的系統，除了對雜訊有很高的寬容度之外，對於將細胞核自血液抹片中切割並且定位的能力，在經過實驗之後，確實展現出很不錯的效果。

本章共分成 4 節，先說明整體系統流程，再進一步說明血液抹片影像處理以及辨識策略；3.1 節介紹整體系統流程； 3.2 節說明在進入正式的影響處理分析前，所需進行的預處理；3.3 節說明系統架構中的紅血球模型建立，3.4 節介紹系統架構中的白血球細胞核影像處理及特徵抽取，3.5 節則是關於分類決策樹的建立。

3.1 整體系統架構

本論文將輸入的血液抹片影像先區分成兩個部分來處理，第一部分為紅血球影像，第二部分為白血球細胞核影像。

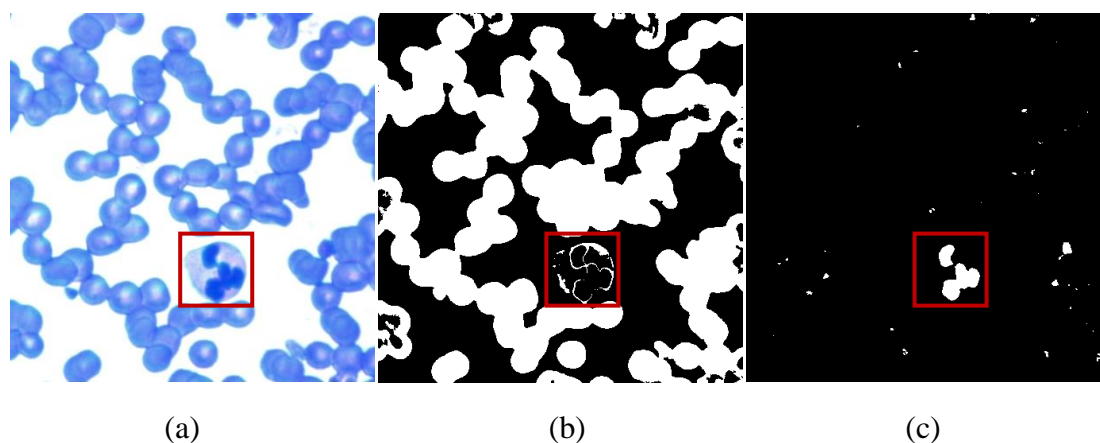


圖 3.1 白血球細胞核影像：(a)血液抹片影像；(b)紅血球影像；(c)白血球細胞核影像。

其中，紅血球影像將用來建立紅血球大小的高斯分布模型，此模型會是後續白血球細胞核分割及辨識的參考資料之一；而白血球細胞核影像的取得方式，則因為血液抹片影像雜訊過多而得要進行白血球細胞核的影像重建，而雜訊的來源主要是血小板(圖 3.2 橘色虛框處)，紅血球互相重疊(圖 3.2 橘色實框處)，或者是嗜鹼性顆粒在血液抹片製作過程中破碎(圖 3.3 中可見深色區塊處皆是)，此三種現象都導致白血球細胞核切割的困難，因此本論文將提出一個修正方法，將有效地排除掉這類雜訊的干擾，成功地將白血球細胞核自抹片影像中切割出來。

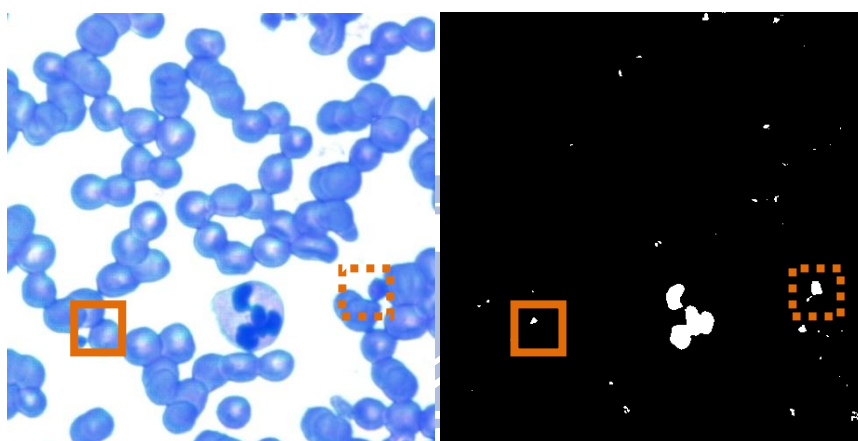


圖 3.2 血液抹片中的雜訊。

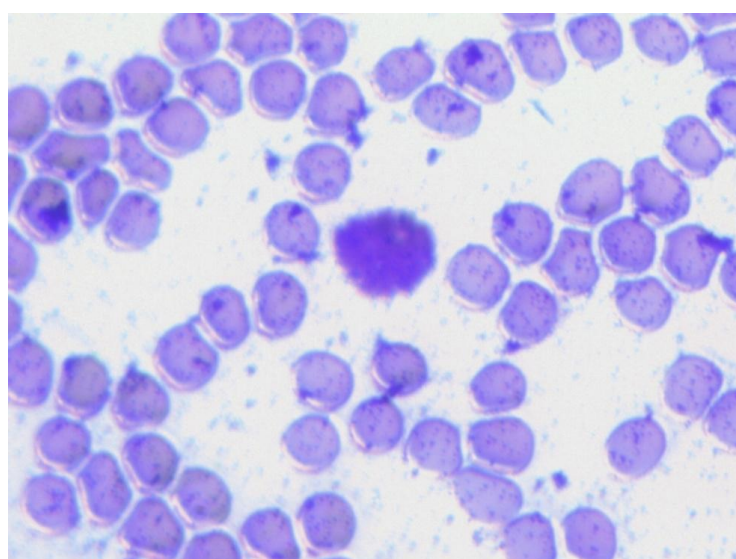


圖 3.3 血液抹片中的破碎嗜鹼性顆粒。

經過白血球細胞核切割後的影像，並未能夠完整描述白血球細胞核，主要原因在於粒細胞的分葉(lobes)，同一白血球細胞核，可能會因為分葉與分葉間的連接處飽和度過低，而無法透過前一步驟將白血球連結成一個完整的區塊，因此看似離散的細胞核區塊卻可能同屬於一個白血球細胞核，需要透過某種方式來將離散區塊重新描述為同一細胞核；此現象將於後面章節討論。

白血球細胞核切割完成之後，將進行特徵抽取，所參考的特徵主要分成形狀，以及紋理特徵，透過這兩類特徵來描述白血球細胞核，接著再由各類血球之間的差異，建立 SVM 決策樹，最後將可透過此決策樹來辨識此細胞核區塊為嗜中性粒細胞、嗜酸性粒細胞、嗜鹼性粒細胞、單核細胞，或者是淋巴細胞。

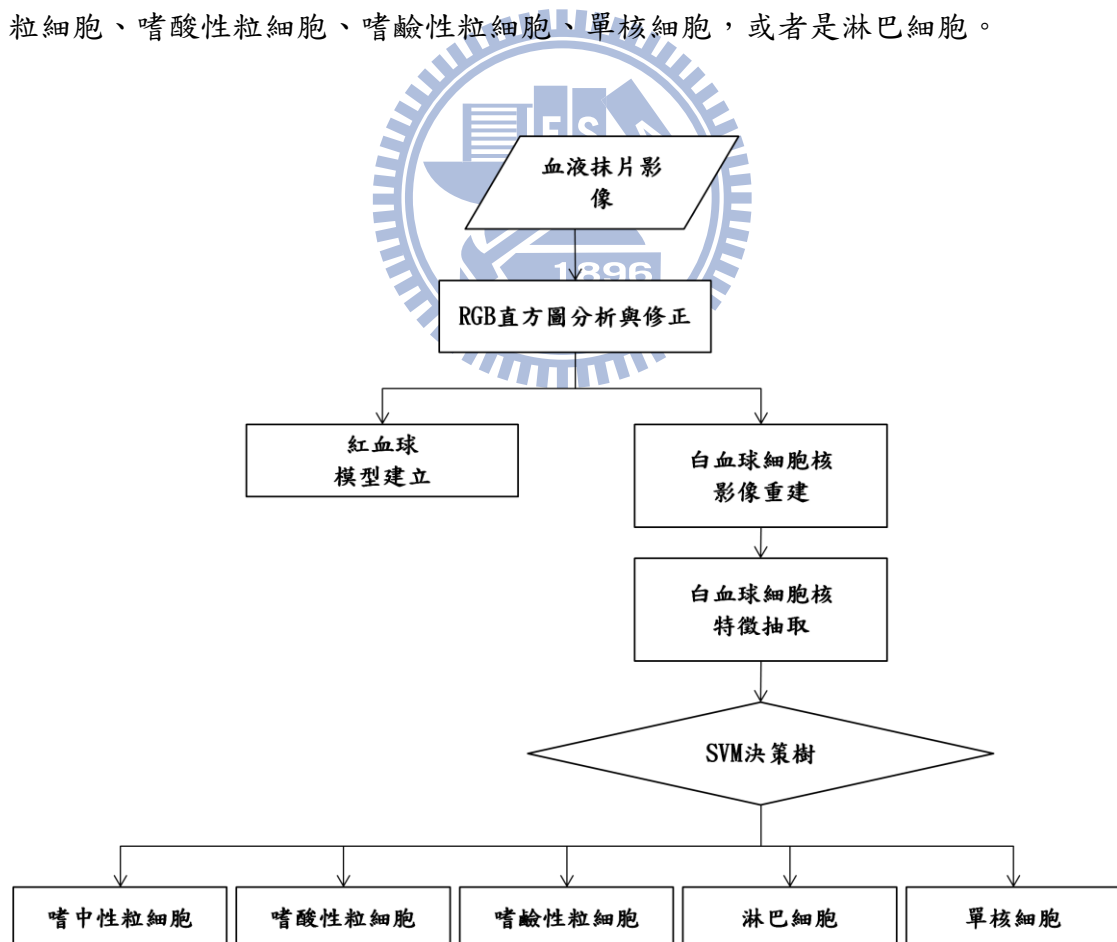


圖 3.4 主要辨識流程。

3.2 RGB 直方圖分析與修正

因為 RGB 三通道之直方圖分佈不等的影響，進行色彩空間轉換之效益可能無法到達預期中的目標，因此得要先對 RGB 三通道進行分析，倘若發現有不足的部分，可以在此步驟做修正。

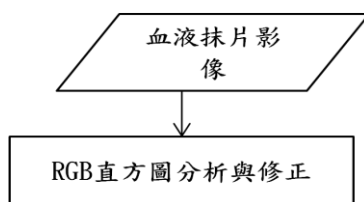


圖 3.5 前處理流程圖。

在進行二值化處理之前，得要先分析 RGB 三通道直方圖的分布是否會影響到後續的影像處理步驟；分析結果如圖 3.6 所示：

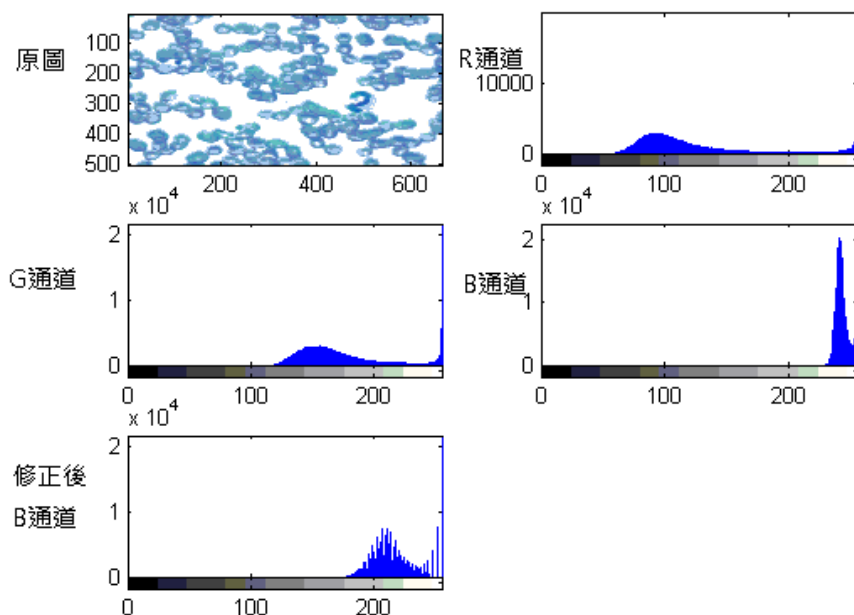


圖 3.6 RGB 三通道分析。

由分析結果可以得知，RGB 三通道中，以 B 通道的直方圖分布最不均勻，

這一現象對於紅血球模型的建立，影響並不大；但是對於白血球細胞核影像的分析則影響深遠，因為白血球細胞核影像的產生，是將原本 RGB 色彩空間轉換到 HSI 色彩空間，接著只取 S 通道影像來作進一步分析，依據 RGB 色彩空間轉換到 HIS 色彩空間的公式而論，原本 B 通道的集中分布將會對色彩飽和度分割造成不利的影響，以下兩張影像皆設定色彩飽和度為 0.75 作為臨界值，以此值為界線對飽和度影像進行二值化處理，圖 3.7 為尚未修正過後的 S 通道二值化影像，圖 3.8 則為修正過後的 S 通道二值化影像，由這兩張圖來比較，確實同一場景下修正前與修正後作比較，會發現修正後的影像雜訊相對減少了很多：

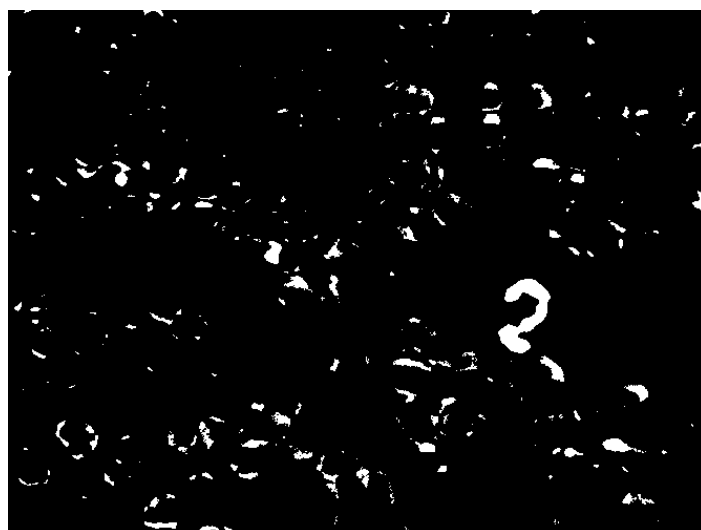


圖 3.7 RGB 三通道未作修正。

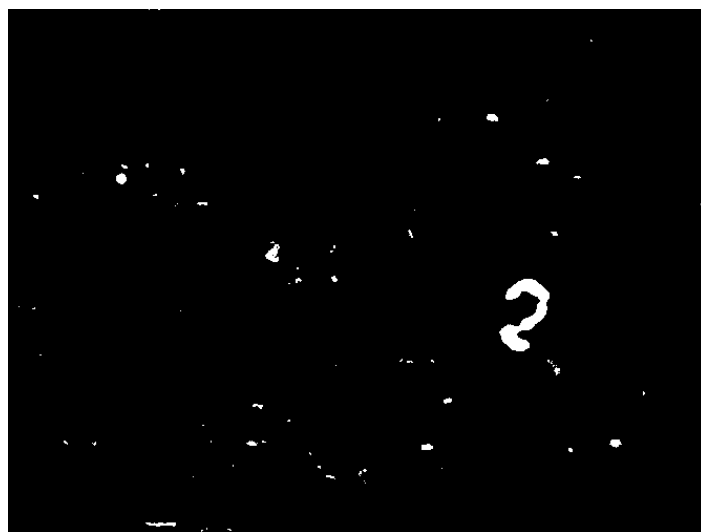


圖 3.8 經過 RGB 三通道的分析與修正，降低了雜訊。

3.3 建立紅血球模型

本論文並未嘗試去解決紅血球重疊問題，而是以正圓度以及區塊大小為線索，來找出可能為單一紅血球的區塊，以這些區塊來建立紅血球模型。

在本論文中所討論的紅血球模型，包含了紅血球的近似面積大小，以及紅血球的近似半徑，兩者都以像素(pixels)為單位。



圖 3.9 紅血球模型建立。

3.3.1 影像二值化處理

在經過 RGB 三通道重新修正之後，將色彩空間自 RGB 轉換到 GRAY 色彩

空間，見 3.10，接著分別進行最大類間方差法(Otsu)全域二值化(圖 3.10(b))及尼布蘭克(Niblack)區域二值化(圖 3.10(c))，再將全域二值化結果與區域二值化結果以 AND 運算作結合(圖 3.10(d))，在全域二值化的部分，因為背景與血球細胞影像之間灰階值有一定差距，因此對於背景轉換到血球之間模糊邊界非常敏感，導致二值化後的影像中，無法將血球與血球接觸的邊界做出清楚的畫分(圖 3.10(a)紅色實框)，因此，需要區域二值化方法來彌補這個缺點；但區域二值化也無法獨立完成一幅完美的二值化影像，其二值化的結果忽略了整張影像的資訊，而會產生錯誤的二值化區域(圖 3.10(b)紅色實框)，因此也需要全域二值化來協助彌補這缺失，如圖 3.10 所示，這就是為何無法單用一個全域臨界值或者是區域二值化方法來完成影像切割的原因，透過圖 3.10(d)可知兩者的缺失在結合之後都已經被彌補回來了。

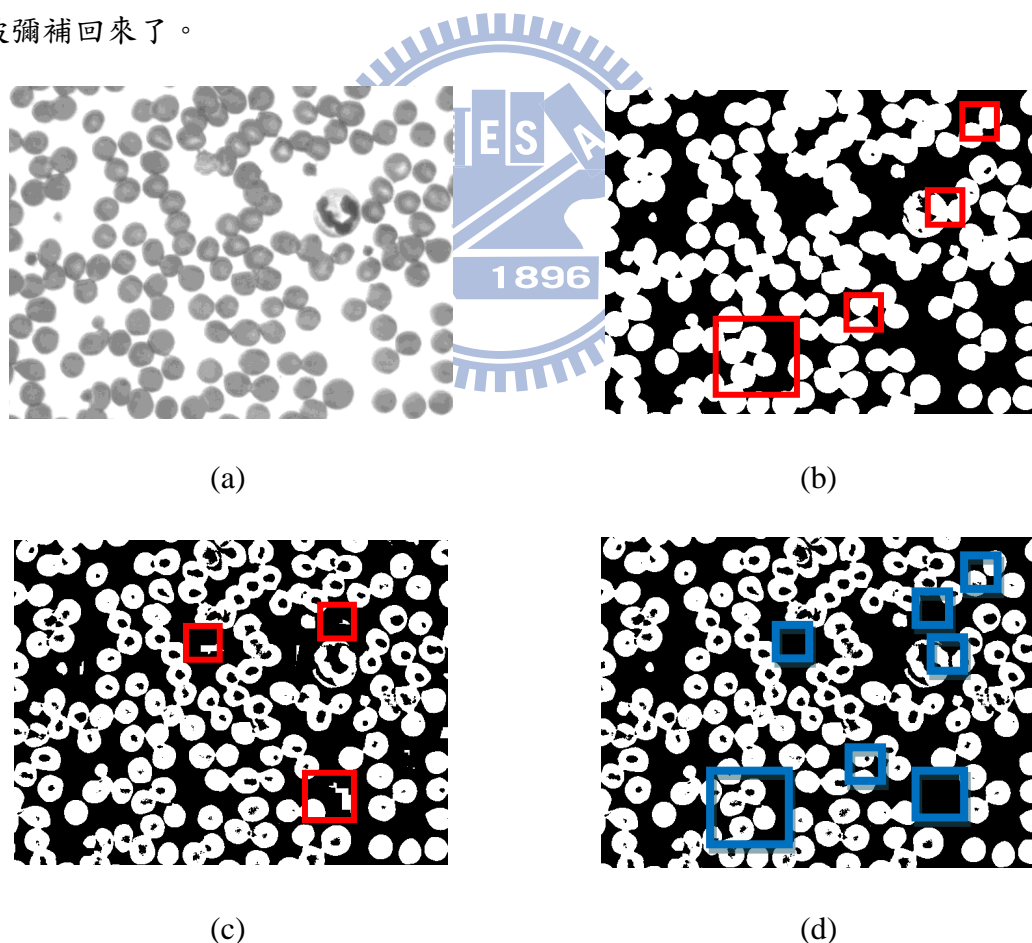


圖 3.10 二值化範例：(a)RGB 轉灰階圖；(b)最大類間方差法；(c)尼布蘭克二值化法；(d)由(c)和(d)作 AND 運算。

3.3.2 二值影像邊緣提取

當二值化處理結束之後，可以看到影像中仍然有許多空洞存在，此空洞可能是因為紅血球本身外厚內薄的原因，所以二值化處理的時候無法將紅血球影像形成一個完整無空洞的區塊；在本論文中，該空洞對於實驗後續分析並沒有任何助益，更有可能造成無謂的計算量上升，因此，將空洞填滿將對於實驗的進行有所幫助，填補空洞後的影像如圖 3.11(b)；再利用 Sobel 運算子對圖 3.11(b)計算邊界，可得邊界影像如圖 3.11(c)。

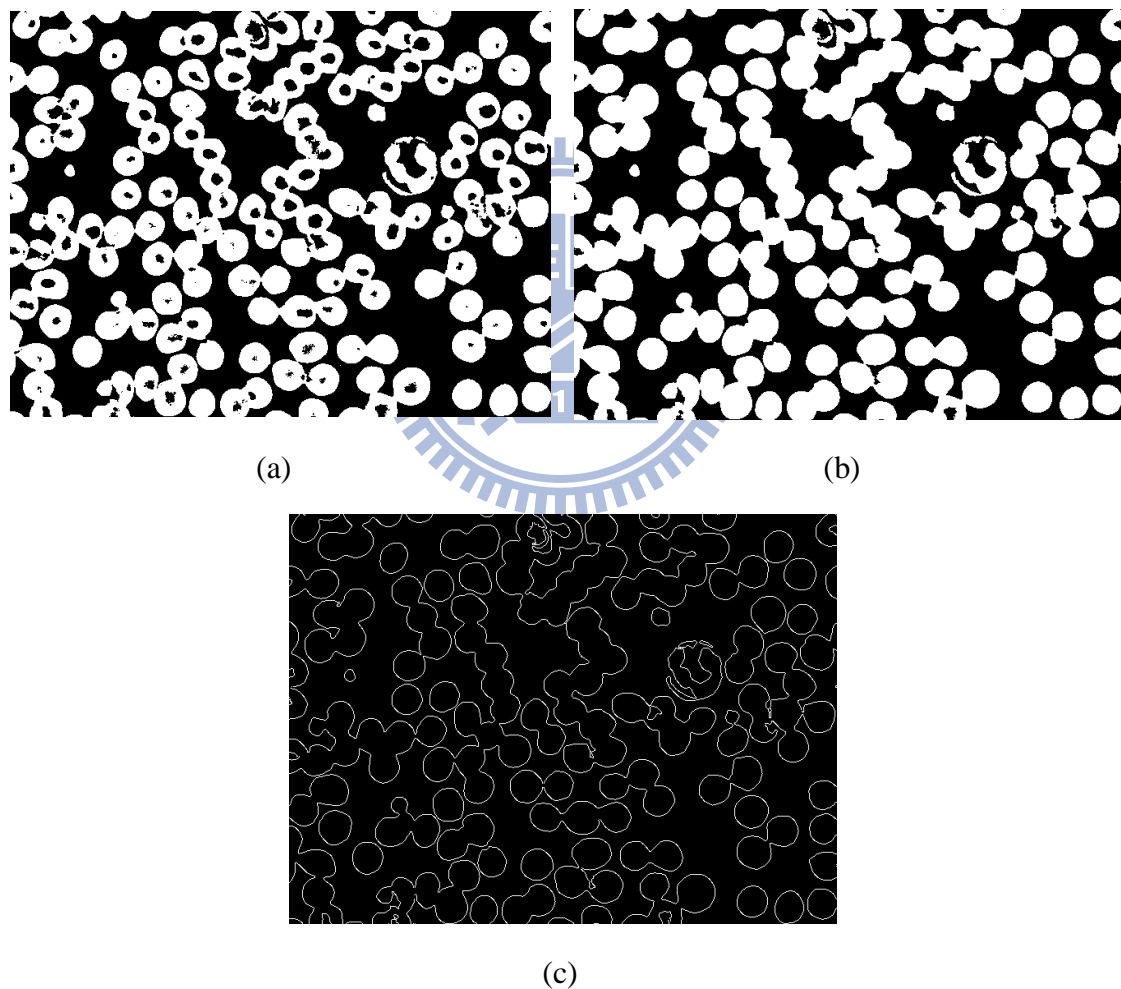


圖 3.11 二值影像邊緣提取：(a)血液抹片二值化影像；(b)填補影像中的空洞；(c)則是由(b) 作邊緣提取所得的邊緣影像。

3.3.3 高圓度區塊面積分析

參考 3.4.4，每個封閉邊界都能計算出圓度，再將所得的圓度利用高斯混成模型分成三類(圖 3.12)，預期會形成高、中，以及低圓度三個高斯分布模型；根據 2.3 節，高圓度高斯分布模型的資料將有 68.27% 落在平均值左右一個標準差的範圍內，因此設定平均值減去一個標準差為正圓度臨界值，高於此數值則判斷此封閉邊界近似於圓，很有可能是本節所想要尋找的紅血球區塊，由此臨界值，可以保留下約 84.135% 的高圓度區塊，但是即使滿足高圓度限制，也未必是紅血球區塊，另外還得要再考慮區塊面積大小的問題。

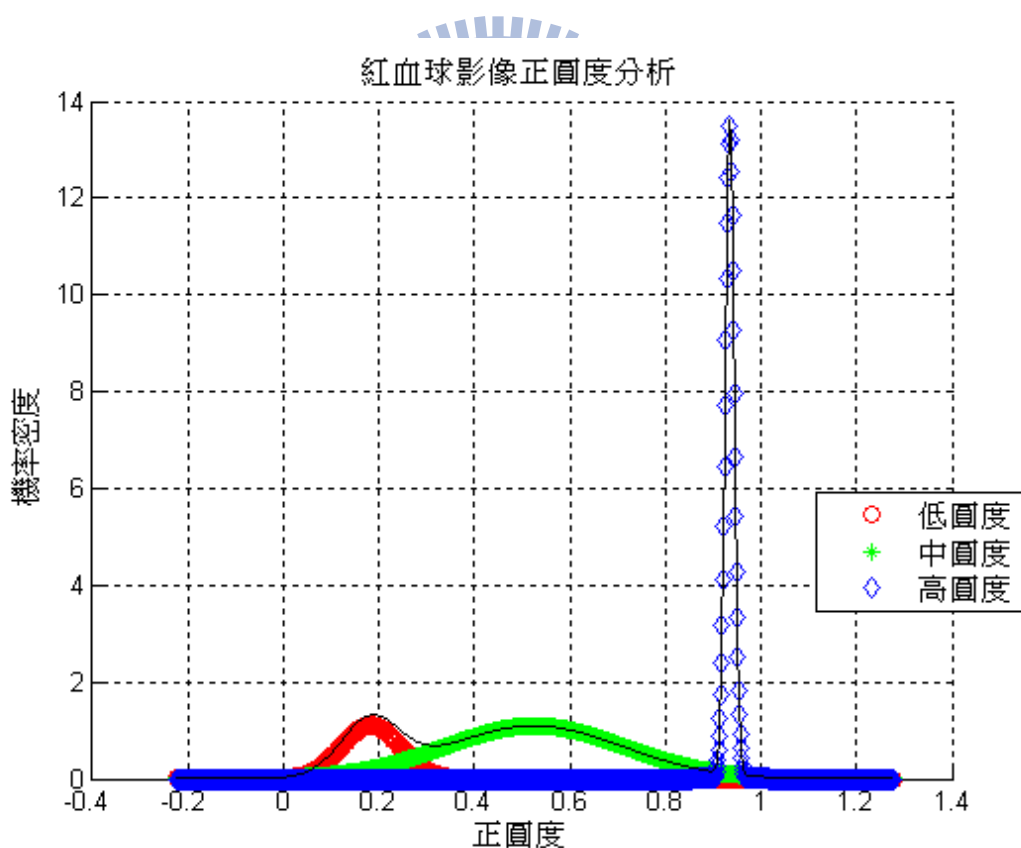


圖 3.12 正圓度分析。

圖 3.13 是參考上一步驟所選定的高圓度臨界值而標記的高圓度區塊，當區塊內包含有一個黑色圓圈，此區塊即為高圓度區塊，其圓度標記於該區塊的左下角；由此參考圖片來看，本論文只尋找單一旦圓度高的區塊，對於區塊交疊或者是相連的部分則不作處理。

圖 3.14 為高圓度區塊面積直方圖，這些疑似紅血球的高圓度區塊其面積並沒有固定範圍，受到血小板、血液抹片製作技術或者是血液來源本身的影響，所以差距可以到六、七倍之譜；雖然就面積大小來說有很大的差異，但也可以看出，此直方圖實際上具備集中趨勢，在健康的動物體內，紅血球的大小應該是落在一個固定的區間內，此區間雖然沒有明確界定，但是可以合理假設經過統計後，可以由直方圖得到一個高機率分佈的區間，而在實驗過程中也確實證明了這項假設。

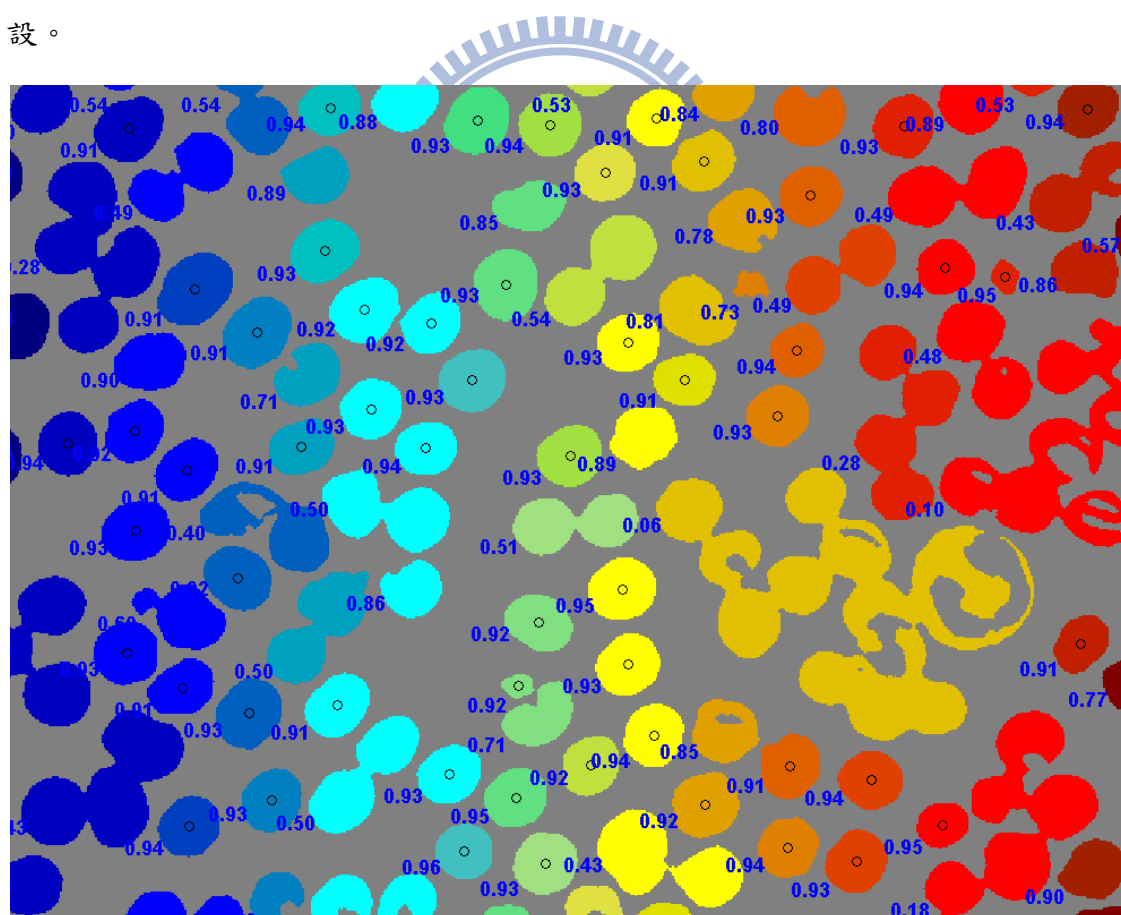


圖 3.13 根據正圓度圈選出可能為紅血球的區塊。

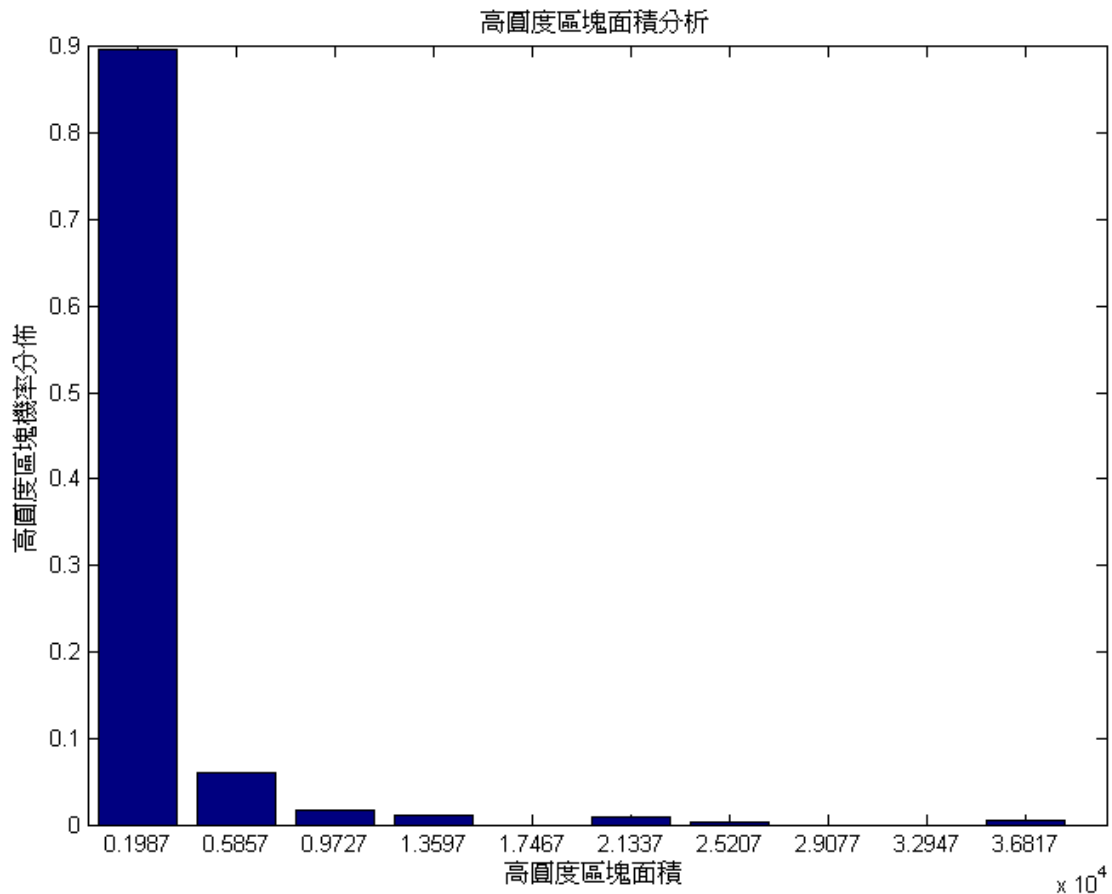


圖 3.14 分析高圓度區塊面積直方圖。

為了取得紅血球的近似大小，本論文將每個圖 3.14 中每個區間內的數量除以高圓度區塊的總數，因此可以得到機率密度分布，再由此機率密度分布作為權值，與每個區間的均值作加權運算，則可以得到紅血球的近似大小。

3.3.4 高圓度區塊半徑分析

本論文所討論的紅血球模型還包含了紅血球半徑(r)，此半徑的計算方式乃是由高圓度區塊面積計算而得；已知區塊具備高圓度，因此可以簡單地由圓形的面積計算公式：

$$Area = radius^2 \times \pi \quad (3.1)$$

$$radius = \sqrt{\frac{Area}{\pi}} \quad (3.2)$$

分析半徑大小的分布，可以得到一個關於半徑大小分布的直方圖，如圖 20 所示：

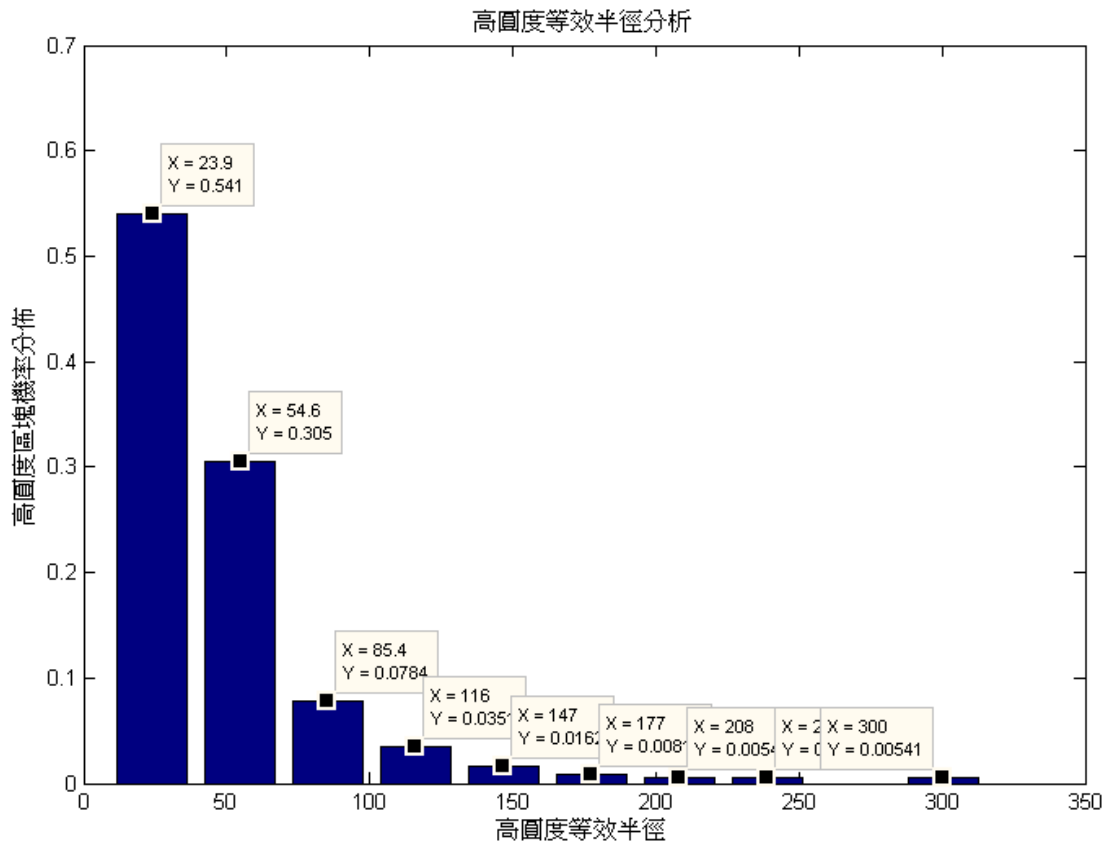


圖 3.15 高圓度區塊近似半徑直方圖。

依照 3.3.1 節高圓度區塊面積分析，在這裡可以對近似半徑直方圖進行同樣的處理，將每個區間內的數量除以總數量，計算得近似半徑落在該區間的機率密度分佈，再將此機率密度分佈作為權值，與每個區間的均值作運算，即可得到高圓度區塊的近似半徑。

至此，紅血球面積的近似大小以及紅血球半徑的近似大小已經獲得，紅血球模型即由此二參數來描述，在後續的白血球細胞核影像處理中，將會有重要的作用。

3.4 白血球細胞核影像處理與特徵抽取

白血球細胞大致上可分成細胞核與細胞質區塊，在本論文的實驗進行過程中發現，當血液抹片製作較差，導致染色不均，或者是數位顯微鏡曝光程度差異，都會影響到白血球細胞質邊界的清晰程度，因此，因為其易受人為因素影響的特性，所以在本章並不討論白血球細胞質。

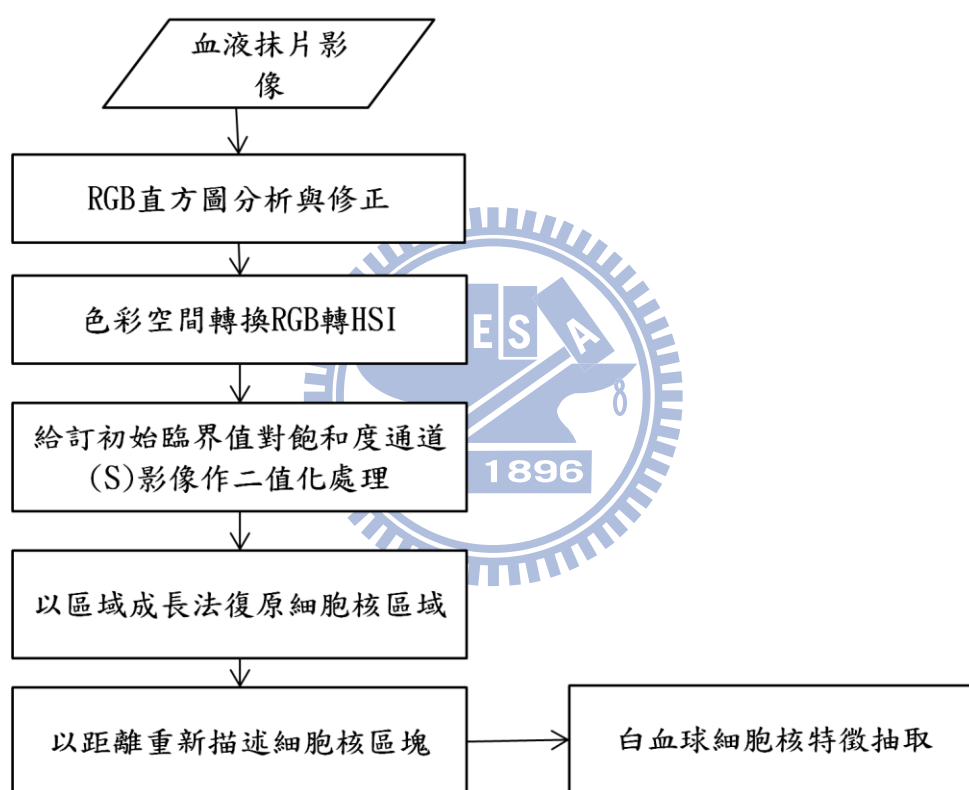


圖 3.16 白血球細胞核影像處理流程圖。

本節共分成四小節，3.4.1 介紹色彩空間轉換，3.4.2 介紹區域成長法復原細胞核區域，3.4.3 則因為 3.4.2 節所復原出來的細胞核區域並未必真實地描述了白血球細胞核區域，因為白血球細胞核分葉的現象，導致可能有兩三個區域同屬於同一細胞核，若直接以 3.4.2 節的結果來作為白血球細胞核，而不描述這些不同區塊間的關係，最後產生出來的白血球細胞核描述會是錯誤的描述；3.4.4 說明

在 3.5.3 節的處理之後，核區已經重建完成，於是可以從這些已經重新描述過的白血球細胞核區塊抽取出特徵，作為白血球分類的依據。

3.4.1 色彩空間轉換

本論文為了要將細胞核盡可能地與背景有更多的差異，因此需要透過色彩空間轉換，讓屬於白血球細胞核的特性更為顯著；觀察白血球細胞核影像之後，本論文決定採用 HSI 色彩空間，由此色彩空間特性，可以得到影像中色調、飽和度、以及亮度資訊；透過 HSI，本論文可以分析影像中色調，飽和度，以及亮度的變化，經過實驗發現，在血液抹片影像上，色調的變化因染色技術以及曝光參數有關，並不具有一致性，因此不能從色調來作分析；亮度資訊也和色調一樣，不具有一致性，只有飽和度在各種染色技術，都能突顯出白血球細胞核，即使曝光參數變化，也能有很高的容忍程度，因此由 RGB 轉換成 HSI 確實有其必要性。



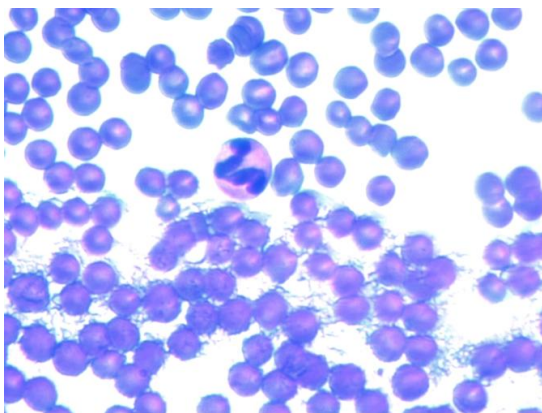
圖 3.17 色彩空間轉換及面積過濾流程圖。

根據第 3.2 節 RGB 直方圖分析與修正之後的結果，可以相信 RGB 三通道已經獲得適當的修正，使三通道的直方圖分布上都較為均勻，以此為前提，開始進行 RGB 色彩空間轉換到 HSV 色彩空間；在實驗過程中發現 H(hue)通道很容易受到曝光程度、白平衡，以及不同染色技術的影響，並不具有後續實驗分析的價值；而 I(intensity)通道影像也隨著曝光程度改變而有不同的變化，同樣地，I 通

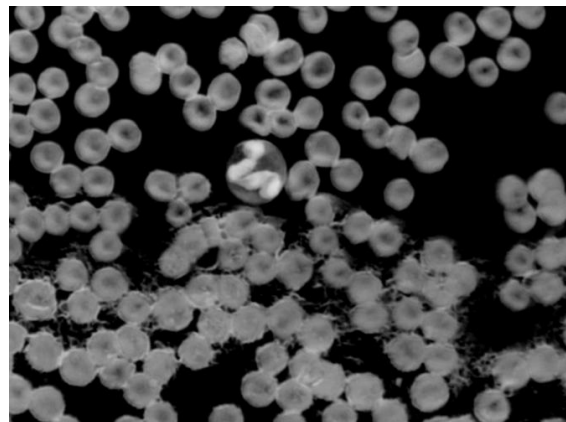
道也不能拿來作分析；再來看 S(saturation)通道影像，可以發現無論曝光程度，白平衡怎樣變化，白血球細胞核區塊的飽和度相較起影像中其他區域有很大的差異，可參考圖 3.18(b)，因此，可以設定臨界值，將白血球細胞核區塊自 S 通道影像中切割出來。

初始臨界值設定乃是根據經驗，在本論文的實驗過程中發現，白血球細胞核區塊飽和度幾乎都維持在 0.75 以上，因此可以先設定臨界值為 0.75 來對 S 通道影像作切割，圖 3.18(c)即為以色彩飽和度為 0.75 為臨界值，對 S 通道影像切割後的結果。

由圖 3.18(c)和圖 3.18(b)來做比較，在圖 3.18(c)中發現依然存在著一些會影響到後續實驗進行的雜訊，在此以紅色方框標記；由本例來說，雜訊是因為紅血球重疊而產生的陰影，其色彩飽和度也在 0.75 以上，所以也被錯誤歸類到白血球細胞核區塊；除了紅血球重疊這個因素之外，有時候也因為曝光不足、血小板，或者是嗜鹼性顆粒的關係，而出現錯誤歸類的區塊，為了克服這一點，本論文基於一個假設：「雜訊面積和白血球細胞核區塊面積相比較小」，因此依經驗設定一面積臨界值為 100，大於此數值者，才會被歸類為白血球細胞核區塊；此一假設並沒有理論支持，單純是實驗過程中的心得，而這個假設所導致的最大缺點，就在於可能會有白血球細胞核區塊也小於 100 像素，也被這個假設所濾除，如圖 3.18(d)所示，為了修正這個缺點，本論文將會在 3.5.2 節提出修正方法。



(a)



(b)



圖 3.18 白血球細胞核影像飽和度切割；(a)原圖；(b)來自(a)的飽和度影像；(c)飽和度臨界值(0.75)切割；(d)濾除過小區塊之結果。

3.4.2 區域成長法復原細胞核區

為了要尋找白血球細胞核區塊，並且定位白血球區塊，本論文在一個前提下進行後續的操作：「當雜訊完全去除掉的時候，剩下的區塊即為白血球細胞核區塊」，因此，如果單純只對整張飽和度影像選取一個臨界值作分割，從 3.5.1 節可以發現依然會存在本論文所不希望出現的錯誤區塊，想要從影像中找到一個全域臨界值來做二值化切割幾乎是不可能的事，因此，在本小節延續 3.5.1 節的結果，將探討的範圍從整張影像，限定到可能為白血球細胞核區塊的近鄰區域，一方面修正 3.5.1 節為了在雜訊與真實細胞核區塊之間所作的取捨，一方面也因為搜索細胞核區塊的範圍從整張影像，縮減到極小的範圍，將可大幅提高定位白血球區塊的準確度，但是即使復原了核區，在此章節仍未能定位白血球細胞核，理由將由 3.5.3 節回答。

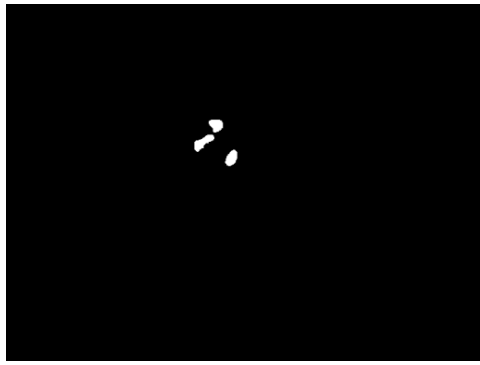
由圖 3.18(d)，可以看到三個離散的白血球細胞核區塊，再參照圖 3.18(a)，會發現這三個離散的白血球細胞核區塊實際上都同屬於一個細胞核，但是從這三個白血球細胞核區塊並沒有辦法得到此細胞核任何資訊；觀察圖 3.18(a)，可以

發現原本的細胞核區塊遠比這三個白血球細胞核區塊來得大，為了實驗的進行，有必要由現有的資訊來恢復完整的白血球細胞核。

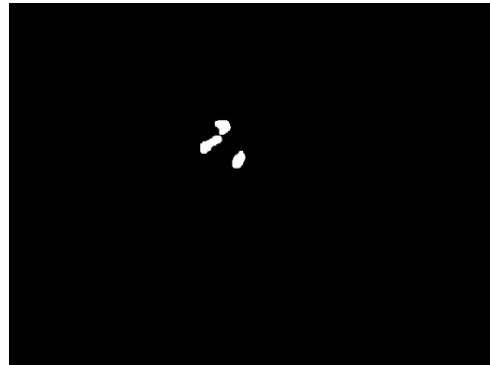
首先第一個考慮的要點，若要恢復完整細胞核區塊，得要找出有什麼資料是這些區塊的共通性質，以連結這些區塊；既然我們切割出白血球細胞核的依據是來自於飽和度，那麼很理所當然地，若兩像素同在同一細胞核區的範圍內，則兩像素間最大的相似性在於飽和度。

由現有的白血球細胞核區塊，分析這些區塊內的飽和度分布，計算其平均值與標準差，建立一個以現有區塊飽和度為基礎的高斯分佈模型，此高斯分布模型將被用來描述像素與像素間的相似性，在本論文中稱此高斯模型為「相似性模型」。

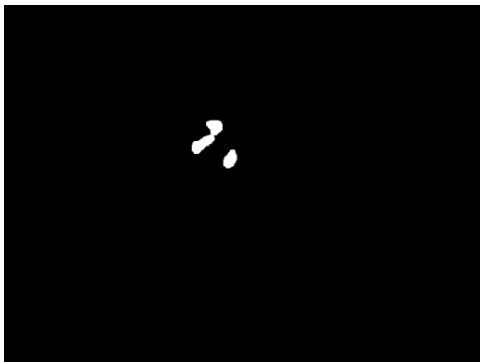
初始輪廓為圖 3.18(d)中所示的三個離散細胞核區塊的邊緣，搜索範圍為以每個邊緣像素所在座標為中心，分析中心周圍的八個像素，將八個像素的飽和度與相似性模型進行比較相似性程度，在此所說的相似性程度則是參考自高斯分布本身的特性，有 68.27%的面積落在平均值左右一個標準差範圍內，95.45%的面積落在平均值左右兩個標準差的範圍內，因此判斷該像素是否與目前的白血球細胞核區塊相似，代表著該像素的飽和值落在相似性模型的哪個位置，越遠離平均值，代表著越相似或者越不相似，在本論文中定義一個臨界值為平均值減去一個標準差，倘若該像素的飽和值大於此臨界值，則代表該像素和目前的核區很相似，因此，該像素將被歸類為核區的範圍之一；重複此一過程，直到舊有輪廓其周邊像素都已經被比較且更新過，此時新的白血球細胞核區塊已經產生，依據此新的白血球細胞核區塊重新計算其輪廓，重複此一流程，直到沒有新的像素被更新為止，此時白血球區塊即重建完成，詳細演進過程如圖 3.19 所示：



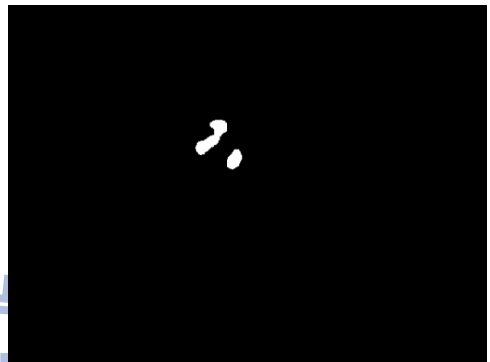
(a)



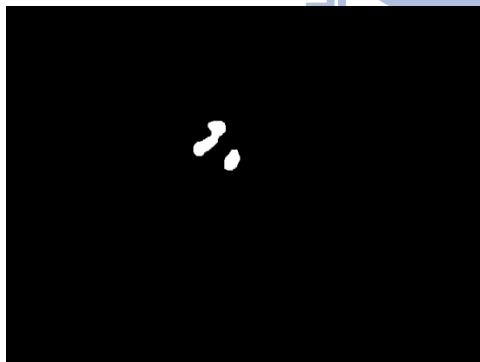
(b)



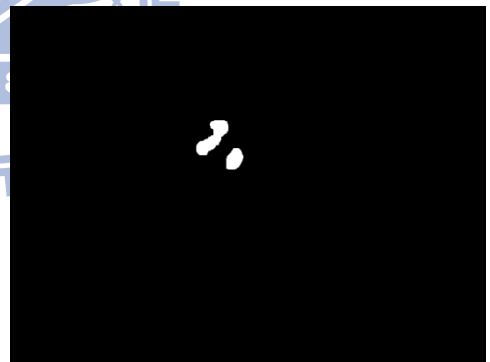
(c)



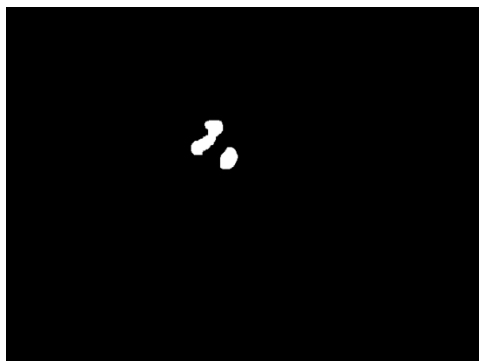
(d)



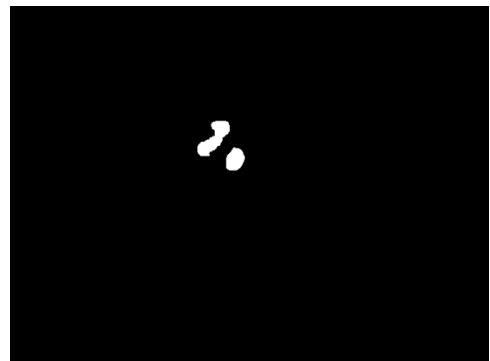
(e)



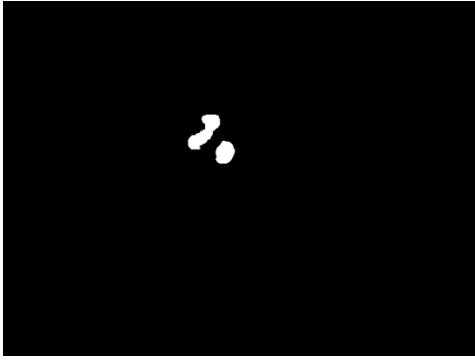
(f)



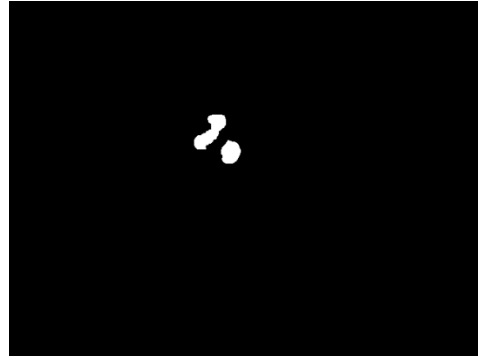
(g)



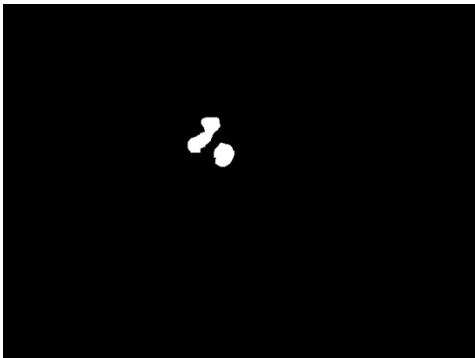
(h)



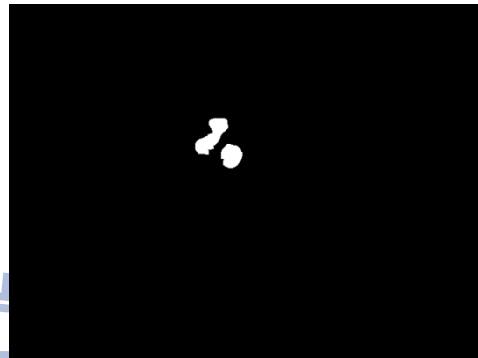
(i)



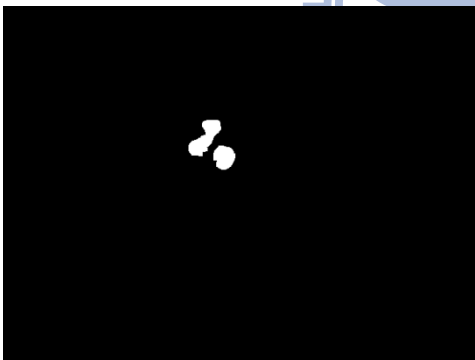
(j)



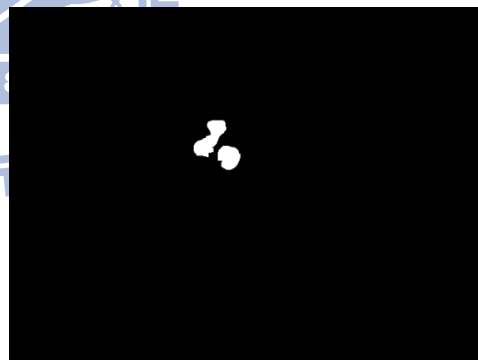
(k)



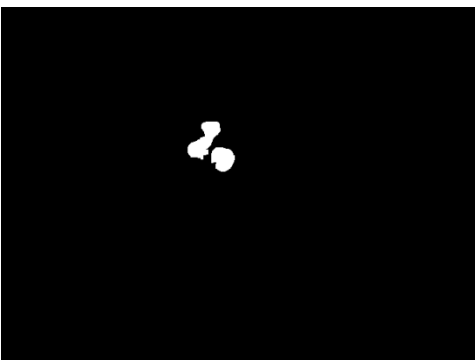
(l)



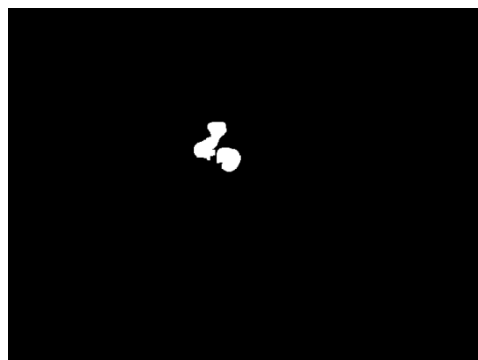
(m)



(n)



(o)



(p)

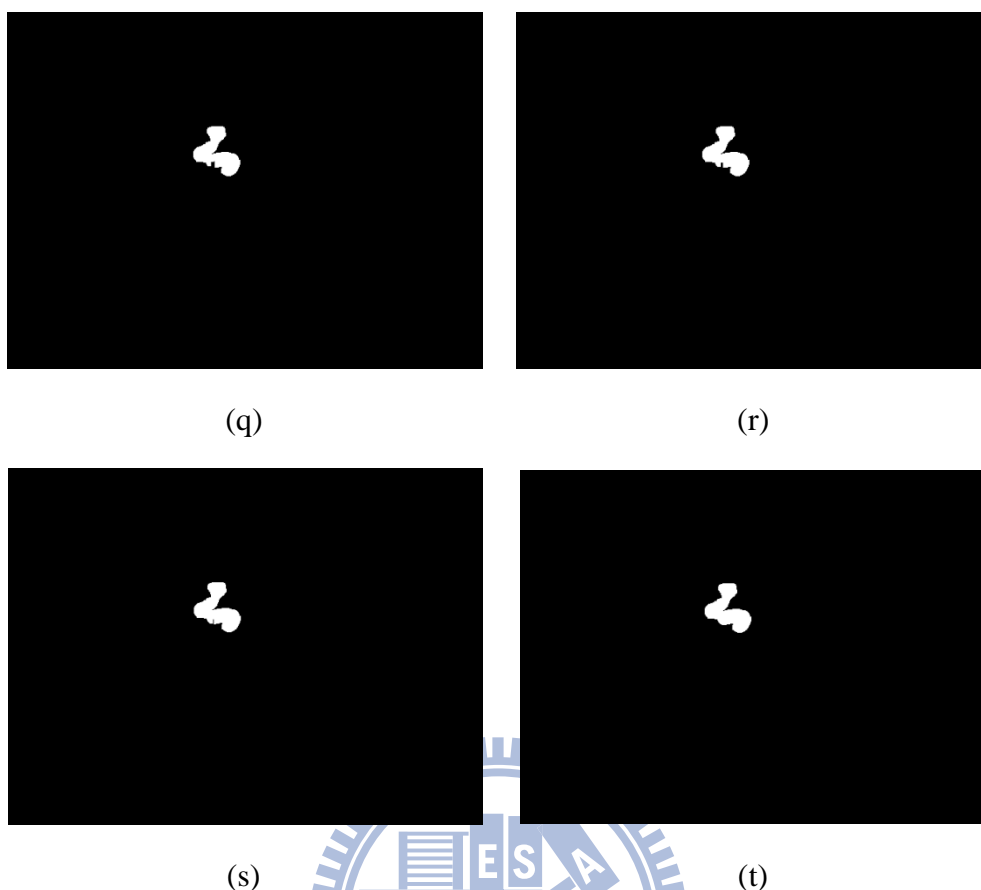


圖 3.19 由(a)到(t)依序為區域成長法復原白血球細胞核區塊過程。

3.4.3 細胞核區重新描述及多重細胞核定位

由 3.5.2 節所重新成長出來的細胞核區塊未必為完整連通區域，主要的因素為白血球細胞核的分葉問題，常見於嗜中性粒細胞，或者是嗜酸性粒細胞，因此在經過區域成長所處理過的影像仍然未必能定位出白血球細胞核，原因就在於此；當細胞核區經過重新描述之後，就可以將同屬於同一細胞核的區塊，但卻不相連通的區域，經過標記相同數字，而使這些區塊得以被認為是同一個類別；在接下來的討論中，將舉一例子，並且將參考 3.5.1、3.5.2，再到 3.5.3 這些章節內所述，依序對此例影像作處理，然後再來說明本章節所要呈現的細胞核區塊重新描述，以及多重細胞核定位。

事實上在實驗過程中以 500 倍放大倍率觀察血液抹片，同一個視野中很少出現三個以上的白血球細胞，因此為了更清楚地表明本論文的作法在複雜影像中依然能夠切割並且定位多個白血球區塊，本論文在此以多張包含白血球細胞，但是卻不同曝光值的影像，合成為同一張影像，即圖 3.20，接下來將以圖 3.20 為例，進行說明。

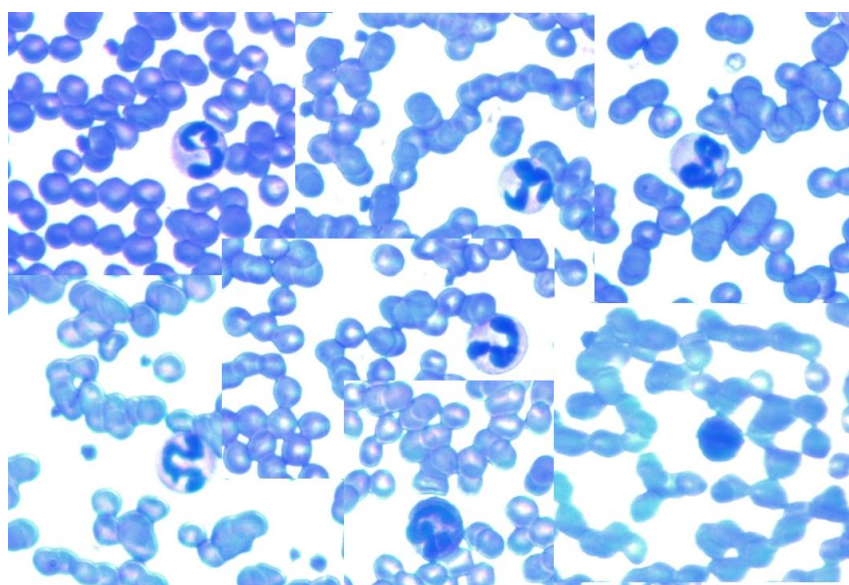


圖 3.20 複雜血液抹片影像。

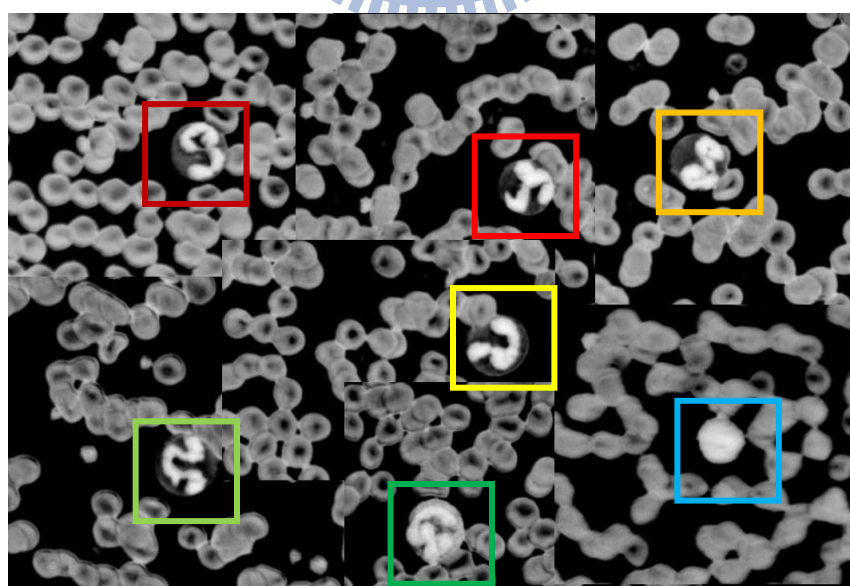


圖 3.21 複雜血液抹片影像-飽和度。

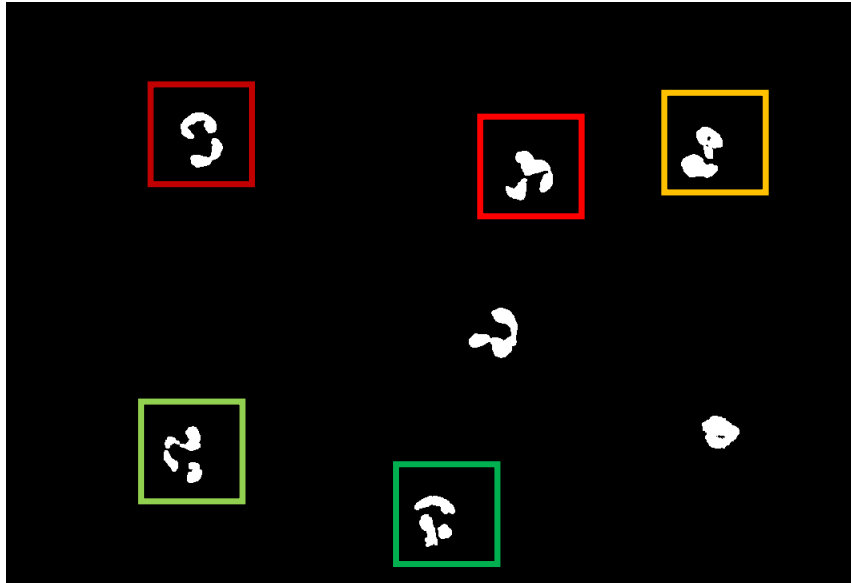


圖 3.22 複雜血液抹片影像-細胞核。

圖 3.21 為圖 3.20 的色彩飽和度影像，在圖 3.21 中依然可以發現白血球細胞核區域呈現高飽和度(白色區域)，在經過 3.5.1 所述飽和度臨界值設定以及濾除掉連通像素小於 100 的區塊後，所呈現的影像如圖 3.22 所示；注意到，雖然在原始影像中可以清楚辨認為同一細胞核的區塊，在經過二值化之後，卻被切割成數個區塊的組合，為了修復白血球細胞核影像，得要再進行 3.5.2 所使用的區域成長法處理圖 3.22 的影像，結果如圖 3.23 所示。



圖 3.23 無法連通成為一個完整連通區域的細胞核。

由圖 3.23 可以發現上有三個白血球細胞核無法形成一個完整的連通區域，這是因為白血球細胞核的分葉與分葉間的區域，其飽和度相較起同一細胞核的飽和度還不足以被 3.5.2 所提到的相似度模型判斷為是細胞核區域，但是明顯地，這些在方框內的細胞核區塊，確實同屬於同一細胞核，如何描述這些不連通區域為同一細胞核就是本小節所要講述的重點。

繼續關注圖 3.23，在圖 3.23 中，尚有三群離散白血球細胞核區塊沒有被描述為同一群，為了更清楚地看出這個事實，在這裡先將不同的連通區域標記上數字，再將同樣數字的區域顯示為同樣色彩，如圖 3.24 所示。

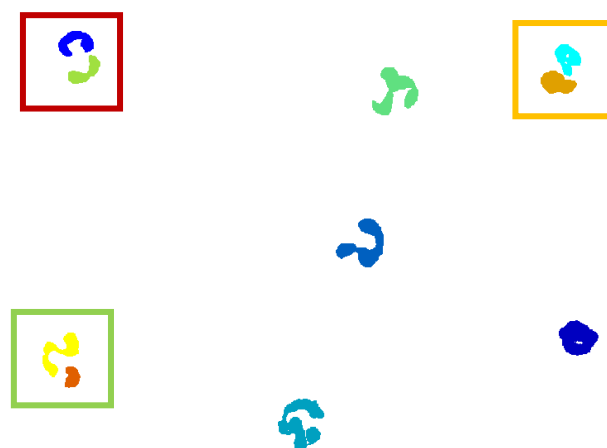


圖 3.24 一個色彩代表一個細胞核。

最直覺的想法是利用區塊與區塊間的距離，作為合併的依據；倘若兩區塊距離較遠，則不可能為同一細胞核區，反之，若兩區塊距離較近，則為同一細胞核區的機會很高，所以，如何定義遠近關係則是這裡所需要思考的問題。

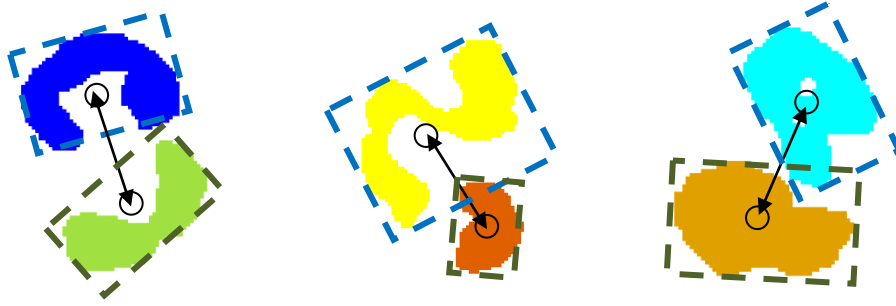


圖 3.25 三組離散細胞核之長短軸分析。

要分析區塊與區塊間的距離，得要先對單一區塊做些處理；對於單一區塊，得要先計算其凸殼(convex hull)之邊緣，再由此邊緣計算得長軸與短軸，前者之定義為凸殼邊緣上距離最遠的兩端點距離，後者的定義為垂直於長軸，並且在凸殼邊緣上距離最近的兩端點距離；由長軸與短軸可以計算出以長軸與短軸為長與寬的長方形的對角線長，以此對角線長度為線索，倘若兩區塊之中心距離大於兩區塊對角線長的最大值，則可判定這兩個區塊並非屬於同一白血球細胞核，倘若兩區塊之中心距離小於兩區塊對角線長之平均，則可判定這兩個區塊屬於同一白血球細胞核，將這兩個區塊通通標記為同一數字，以這樣的方式來重新定義區塊與區塊間的關係；最後結果如圖 3.26 所示，總計有七個白血球細胞核區，與圖 3.20 相比較，可以發現這是個正確的結果。

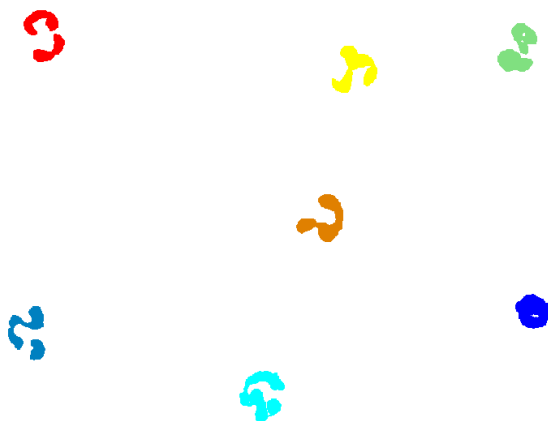


圖 3.26 重新描述完整之白血球細胞核影像。

然而在這個部分隱藏著一個缺點，也就是當所有細胞核區塊都非常接近的時候，可能會錯誤地將細胞核的零碎區塊分類到錯誤的細胞核類別中，但因為在本實驗中，這樣的情況尚未被觀察到，此外根據參考文獻[4]所述，在動物血液裡，白血球細胞凝集通常意味著病菌入侵或者是白血球不正常增生的現象，這部分並不在本論文探討的範圍裡，因此並不把這類型的資料列入考量。

3.4.4 擷取白血球細胞核特徵

白血球細胞核特徵可以分成兩部分來探討：形狀以及紋理，本小節先介紹本論文所使用的形狀特徵，再介紹本論文所使用的紋理特徵。

1. 與紅血球面積的相對比例：在本論文中，面積的單位為像素，計算白血球細胞核範圍內所占面積，並且與紅血球面積相比較，最後可以得知白血球細胞核所占面積相對於紅血球細胞所占面積的倍數。
2. 邊界曲率均方差：邊界曲率變化是評估白血球輪廓變化程度的依據，但是如果將整個白血球細胞核邊界都納入考量，那麼特徵向量的大小將會非常驚人，因此，根據最初的想法，使用此特徵的目的在於求得輪廓變化程度，所以只取邊界曲率均方差，均方差越大，則邊界曲率變化越大，均方差越小，則邊界曲率變化越小。
3. 邊界到中心的平均距離與均方差：因為邊界曲率很容易受到擾動，還需要一個更為穩固的特徵來協助描述邊界，本論文採用的特徵為邊界到中心的平均距離以及均方差；作法則是將邊界到中心距離作為高斯混合模型擬合的目標，以三個高斯模型來近似，這三個高斯模型分別代表遠、中、近距離模型，最後再以最遠距離模型的均值，將所有數據作正規化處理(normalization)，透過這三個高斯模型的平均值以及均方差，可以得到描述邊界的另一種看法，雖然無法完全取代邊界曲率對於整個邊界描述的重要性，但也能在邊界曲率受

到干擾的時候提供一個較為穩定的邊界描述。

4. 正圓度(roundness)：用來描述細胞核邊界近似圓形的程度，因為淋巴細胞與單核細胞之細胞核形狀變化較少，基本上很接近圓形，相較起淋巴細胞與單核細胞，其他三類粒細胞之細胞核變化就非常不規則，因此，用正圓度作為特徵，可以很清楚地分別出淋巴細胞、單核細胞與粒細胞，對於初期的粗分類是非常有幫助的；正圓度的定義如下：

$$\text{Roundness} = 4\pi \times \frac{\text{Area}}{\text{Perimeter}^2} \quad (2)$$

5. 堅實度(solidity)：堅實度的定義如下：

$$\text{Solidity} = \frac{\text{Area}}{\text{Convex Hull Area}} \quad (3)$$

堅實度代表著在目標區塊凸殼(Convex Hull)內的像素同時也在目標區塊中的比例；若白血球細胞核區塊其扭曲程度或者是正圓度太低，通常會導致其凸殼面積遠大於細胞核區塊面積，因此堅實度會大幅下降；堅實度的高低也可以做為判斷淋巴細胞、單核細胞以及粒細胞的一個依據。

6. 分葉數(lobes)：細胞核分葉，對於判斷細胞核為哪種類別是很重要的特徵，例如：嗜中性粒細胞通常有 3 到 4 分葉左右，而嗜酸性粒細胞則為 2 分葉，嗜鹼性粒細胞其分葉數不明顯，淋巴細胞與單核細胞之分葉數皆為 1 分葉。在白血球細胞核分類特徵中，分葉數應可以說是最重要的特徵，本論文在此嘗試用平均值移動演算法(mean-shift)來找出白血球細胞核的分葉數目，流程如圖 3.27，而分葉特徵抽取流程請見圖 3.28。

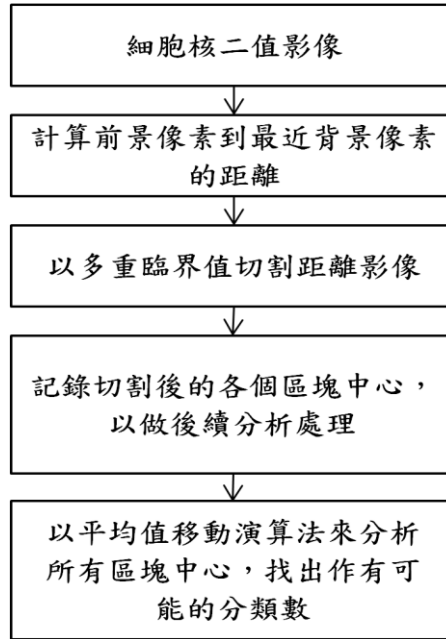
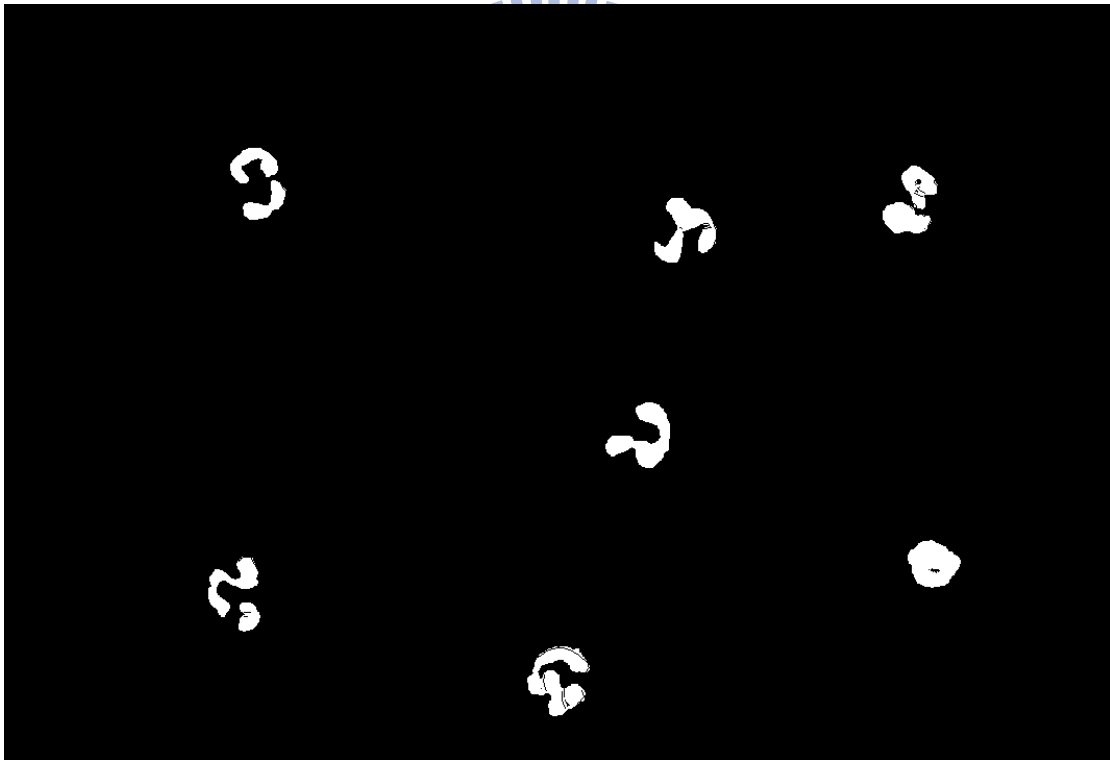
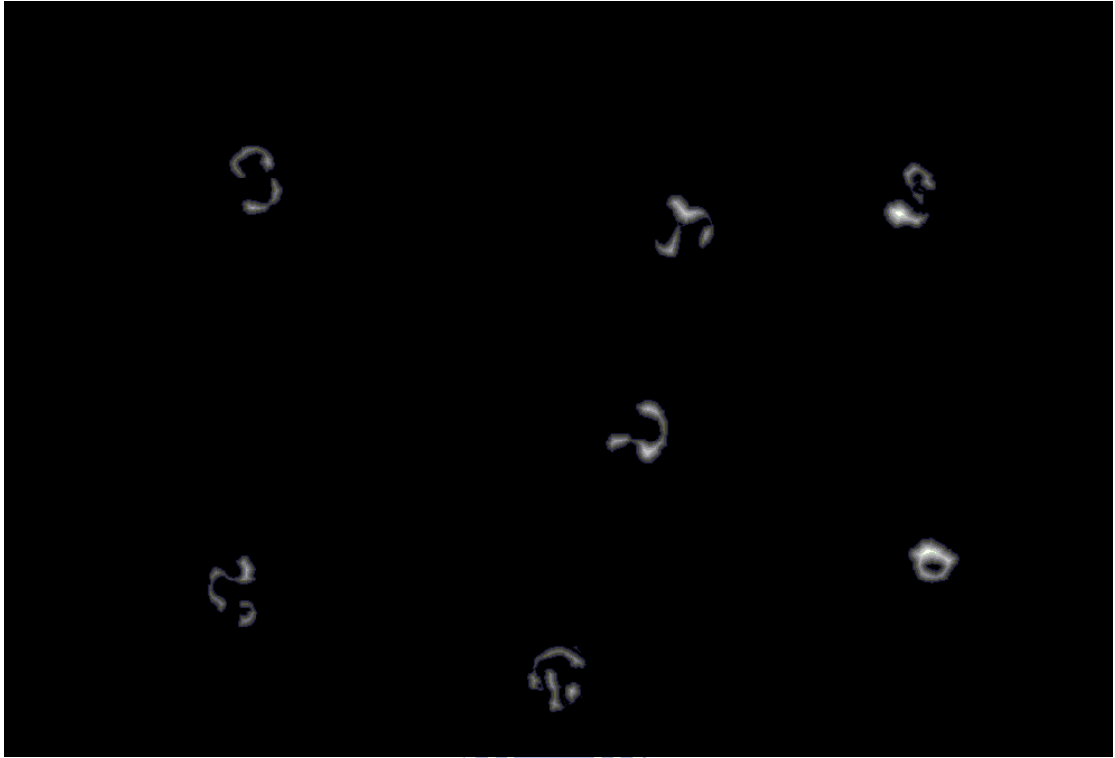


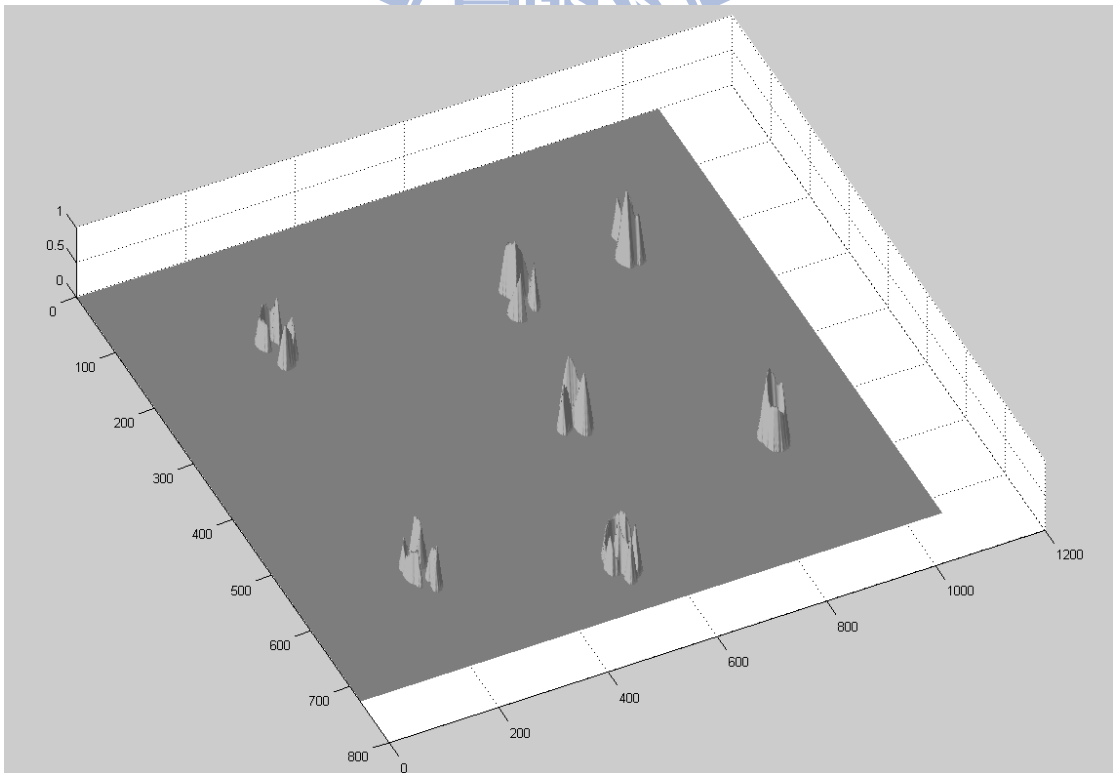
圖 3.27 分葉數計算流程圖



(a)



(b)



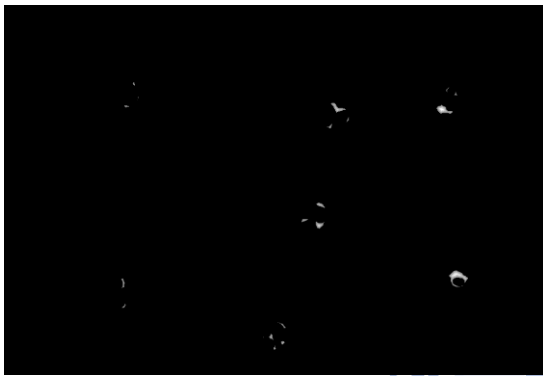
(c)



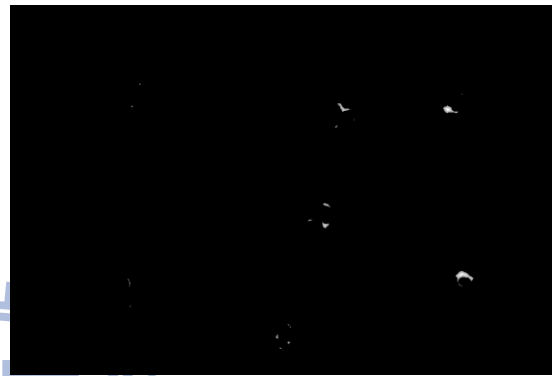
(d) 臨界值為 0.4



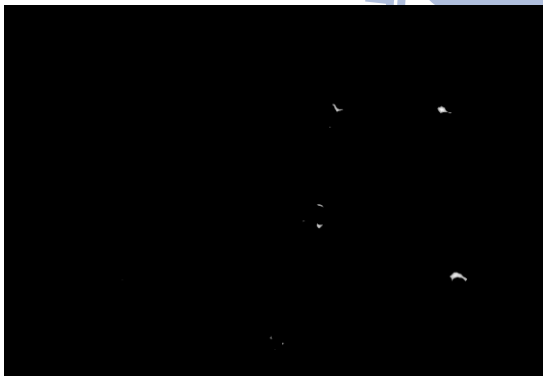
(e) 臨界值 0.45



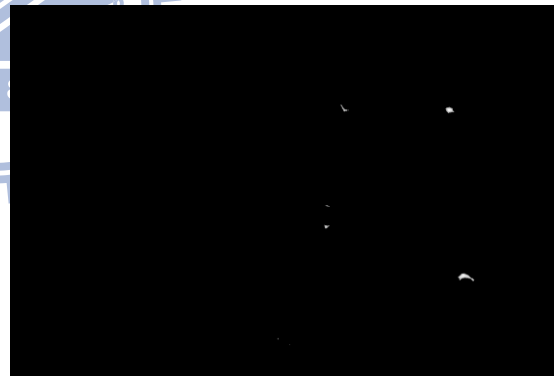
(f) 臨界值 0.5



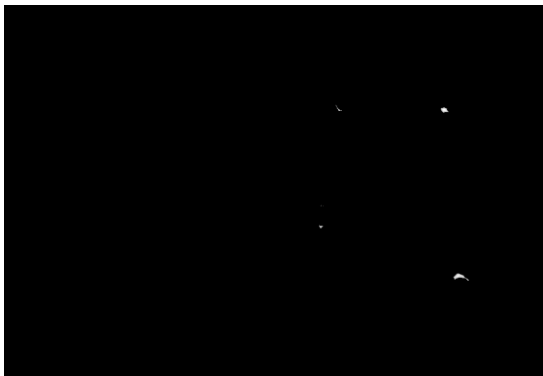
(g) 臨界值 0.55



(h) 臨界值 0.6



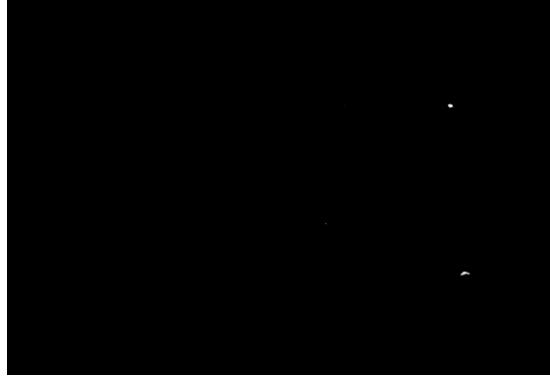
(i) 臨界值 0.65



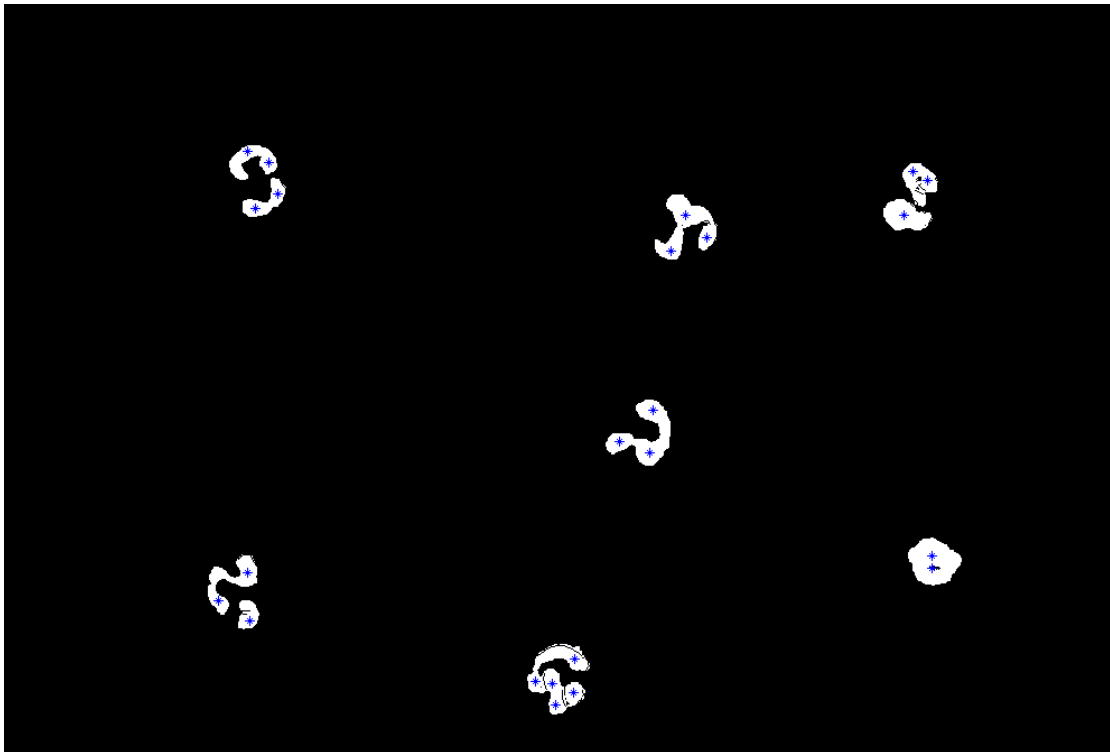
(j) 臨界值 0.7



(k) 臨界值 0.75



(I) 臨界值 0.8



(I) 白血球細胞核分葉數量

圖 3.28 細胞核分葉特徵抽取。(a)為經過區域成長及結合邊界資訊之後的白血球細胞核影像；(b) 白血球細胞核像素距離邊緣最近距離影像；(c) 距離影像以三維模型表示；(d)~(I) 距離影像以多重臨界值切割的結果；

(I) 白血球細胞核分葉數量。

由圖 3.28(b)可以看出透過計算細胞核影像距離最近的背景像素距離，可以得到白血球細胞核每個區段的寬窄資訊，通常在判斷分葉的時候，所依據

的是扭曲程度，以及寬窄；在細胞核分葉區，寬度通常會比分葉與分葉連接處的寬度要寬，根據這個線索，可以先把圖 3.28(a)距離影像中每個細胞核區像素值除以每個細胞核裡面最高距離值，把每個細胞核的距離影像進行正規化(normalization)，因此距離影像中每個細胞核核區的像素都呈現出和該核最大寬度相對比例；如此一來，將可以進一步分析分葉。

從多重臨界值切割以及圖 3.28(b)，可以知道在距離影像的所提供的相對距離資訊，裡面含有許多的區域極值，隨著臨界值變動可以看出這現象，因此為了不讓分葉計算受到區域極值影響，本論文在此以多重臨界值切割來累積數據，每一次臨界值切割所產生的區塊中心將被保留下來，當多重臨界值完成的時候，已經累積了足夠的資訊來分析分葉數。

由於我們並不知道一個白血球細胞應該分葉成幾葉，得要由累積下來的細胞核各區塊的中心，以平均值移動演算法來找出分葉數，結果如圖 3.28(l)。

7. 紋理特徵：詳細資訊請參考第 2.4 節灰度共生矩陣，以灰度共生矩陣提取出細胞核區塊的凸殼範圍內的能量(energy)、對比度(contract)、相關度(correlation)、熵(entropy)、以及同質性(homogeneity)，可以用這些特徵分析凸殼區塊內的亮度分佈特性，藉此描述細胞核凸殼範圍內的染色情形，一般來說，嗜中性粒細胞與嗜酸性粒細胞除了在細胞核分葉上有些差別之外，另一個差異在於嗜酸性粒細胞內含嗜酸性顆粒，因此相較之下，兩者在細胞質的區域之飽和度比較，嗜酸性粒細胞會比嗜中性粒細胞來得高，因此，透過灰階共生矩陣，可以將這個特性描述出來。

3.5 分類決策樹

本論文所採用的分類方式是以二元樹(binary tree)結合支持向量機(support vector machine) (圖 3.29)，將白血球細胞核分類，最終目的是要將輸入的特徵向

量分類分成五種分類，也就是嗜中性粒細胞、嗜酸性粒細胞、嗜鹼性粒細胞、淋巴細胞以及單核球(圖 3.30)。

本論文在分類器結構使用二元樹的原因在於某些白血球細胞核的形狀非常不固定，像是嗜中性粒細胞、嗜酸性粒細胞以及嗜鹼性粒細胞，而在本論文的特徵選取部分又大量使用了形狀特徵，因此藉由設計適當的二元樹節點，或許有助於分類表現，所以二元樹成為了本論文的最佳選擇。

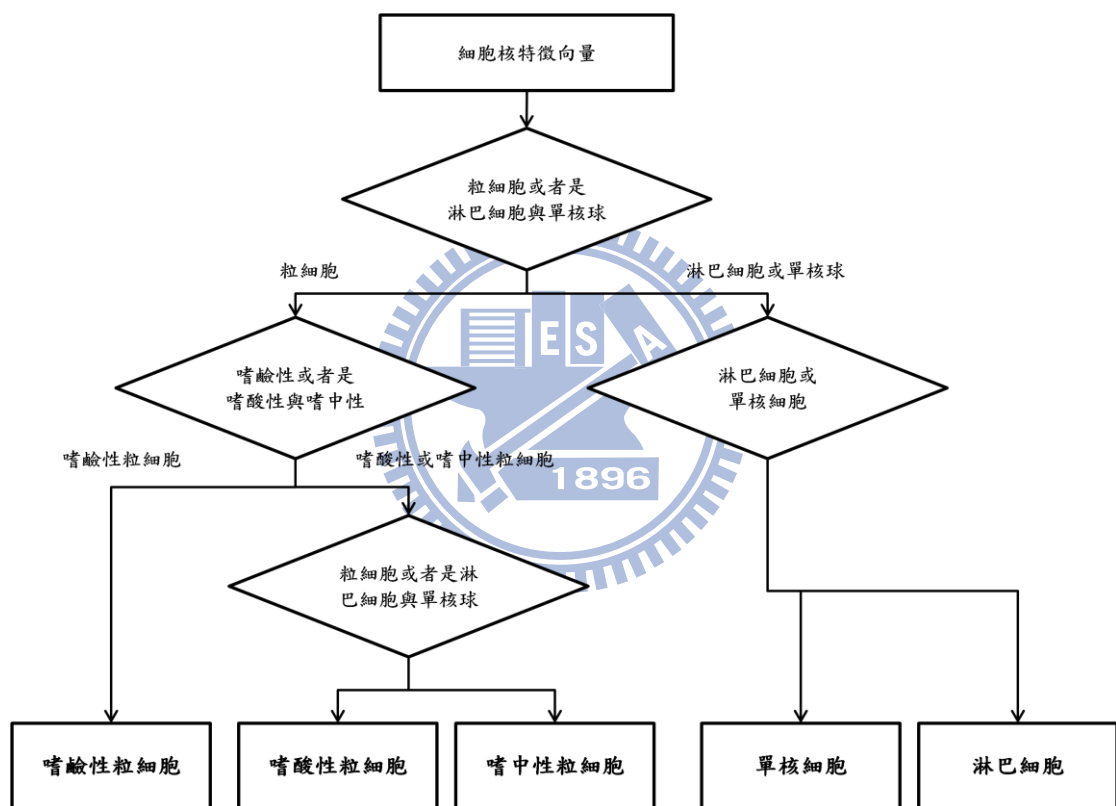


圖 3.29 支持向量機二元樹

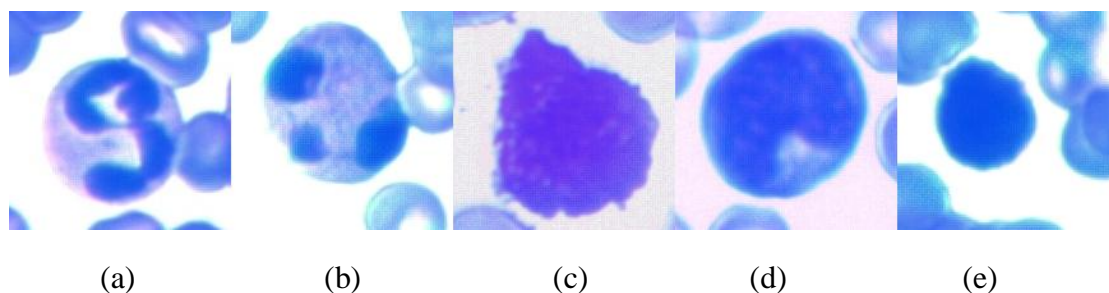


圖 3.30 五種類白血球，分別為 (a) 嗜中性粒細胞，(b) 嗜酸性粒細胞，(c)嗜鹼性粒細胞，(d) 單核細胞，(e) 淋巴細胞。

由二元分類樹的結構可以知道共有四個分類節點，依照不同的分類目的將分類器分別訓練，在根節點的部分將細胞核區分成粒細胞或者是淋巴細胞或者單核細胞兩大類別，根據所選取的形狀特徵，在輪廓、正圓度以及分葉數量上有很大的差異，因此在根節點將粒細胞以及淋巴細胞與單核細胞先做區分，將有利於後續的分類器工作。

順著根節點分類結果，右側子節點目的在於區分淋巴細胞與單核細胞，兩者最大的差別在於面積，單核細胞在血液中無論體積以及面積都是最大的，而淋巴細胞則與紅血球大小相差無幾，這一明顯特徵將有助於此分類器將兩者作明確的區分。

根節點左側子節點的目的是要將粒細胞進一步區分成兩大類，分別是嗜鹼性粒細胞，以及嗜中性與嗜酸性粒細胞；嗜鹼性粒細胞因為其細胞內含有大量嗜鹼性顆粒，因此細胞核常常會被嗜鹼性顆粒所遮蔽，倘若以本論文中的作法，以飽和度作為線索來切割細胞核，將會把嗜鹼性顆粒連帶細胞核一併切割出來，可以想見細胞核堅實度將會因為嗜鹼性顆粒的影響而變得非常高；圓度方面也因為細胞核邊界並沒有太大變動所以呈現高圓度；邊界到中心的距離則因為圓度偏高，所以變異數與其他兩類粒細胞相比會偏低；而在紋理的部分，因為嗜鹼性顆粒與細胞核互相遮蔽，所以在凸殼區域內同質性(homogeneity)、相關度(correlation)都會比嗜酸性及嗜中性粒細胞要高，透過這些特性或許可以將嗜鹼性及嗜中性與

嗜酸性做出區分。

最後一個分類器則是用來區別嗜中性粒細胞與嗜酸性粒細胞，兩者的區別主要在於嗜酸性粒細胞內含有嗜酸性顆粒，所導致的直接結果則是在紋理特徵，在細胞核凸殼區塊的部分，因為嗜酸性粒細胞的影響，所以就細胞核飽和度影像而論，嗜酸性粒細胞的對比度應該會比嗜中性粒細胞來得低，同質性以及相關度則會比嗜中性粒細胞來得高；因為嗜中性粒細胞的扭曲程度遠比嗜酸性粒細胞來得高，因此邊緣到中心的距離平均值還有均方差的部分，就遠、中以及近距離分佈而言，嗜中性粒細胞的變異程度會比嗜酸性粒細胞來得高，同樣地，因為細胞核扭曲程度的緣故，邊緣輪廓變異程度依然是嗜中性粒細胞高於嗜酸性粒細胞，透過這些特徵，可能可以將嗜中性粒細胞與嗜酸性粒細胞作出區分。



第四章 實驗結果與分析

本章節利用實驗來驗證本論文的貢獻，主要分為三個部分：白血球細胞核影像分割、多重白血球細胞核切割及定位，以及白血球細胞核類別識別，將會呈現出在本論文所指涉的雜訊範圍內，系統能夠有效地切割出白血球細胞核；而在切割出白血球細胞核範圍之後，將可利用分葉特徵，讓粒細胞以及非粒細胞能夠有明確的區別。

本章共分成兩節，第一節介紹實驗資料以及方法，第二節則是針對本論文各項貢獻進行比較以及實驗。

4.1 實驗設備與方法

本論文之研究資料主要是由新竹市立安心動物醫院所提供的犬隻血液抹片，玻片製備方式介紹如下：先對犬隻採集血樣後，塗抹於玻片上，待血液塗層乾燥後，再將玻片浸泡於特殊染色劑中進行染色，在此所採用的染色劑為「劉氏染色劑」(圖 4.1(a))，此染色計分成 A 劑與 B 劑，而 A 劑與 B 劑的用量為 1:2，分別會與血液塗層中的不同物質產生作用，將該物質染上不同顏色，先將玻片浸泡於劉氏染色劑 A 劑中，將血液塗層內的細胞直以及血紅素染成紅色，再將劉氏染色劑 B 劑倒進適才所使用的 A 劑中，將細胞核以及白血球內的嗜鹼性顆粒作用，染成藍紫色，玻片取出後，自然風乾，即製作完成(圖 4.1(b))。

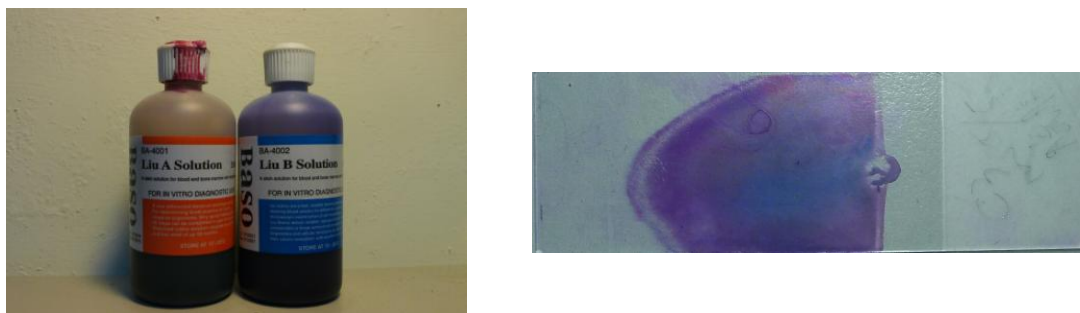
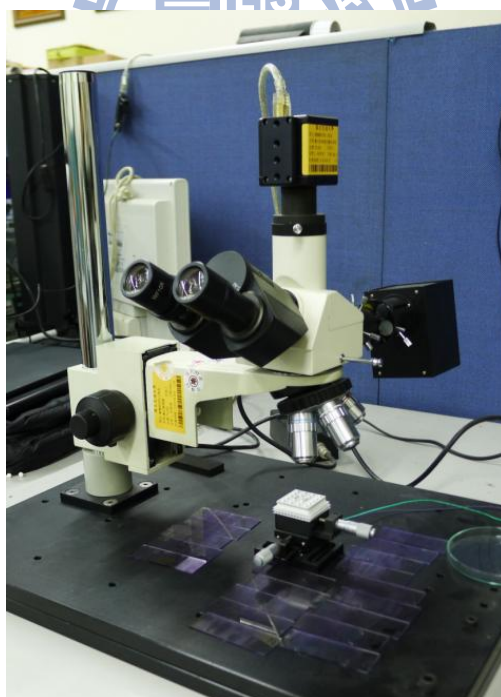


圖 4.1 實驗資料製備。(a) 劉氏染色劑 AB 劑 (b) 染色後血液抹片。

觀察血液抹片所使用的器材包括：數位顯微鏡、微動台以及背光源(圖4.2)，而在血液抹片上所要觀察的對象則以白血球為主，主要有五種分類：嗜中性粒細胞、嗜酸性粒細胞、嗜鹼性粒細胞、單核細胞以及淋巴細胞；數位顯微鏡之物鏡倍率為50倍，而目鏡倍率為10倍，所截取到的影像為500倍的放大倍率，再將載玻片置於背光源上，開始調整數位顯微鏡之曝光參數，注意曝光參數並無最佳參數值，而是以「能用肉眼分辨出細胞核與其他胞器的差別」此一準則來做為判斷標準，配合背光源的不同亮度來修正曝光參數，以取得較適當的曝光參數設定；使用微動平台來移動玻片位置，在顯微鏡下以圖4.3的移動方式掃描玻片，在掃描過程中，視野內若出現白血球，即進行影像擷取。

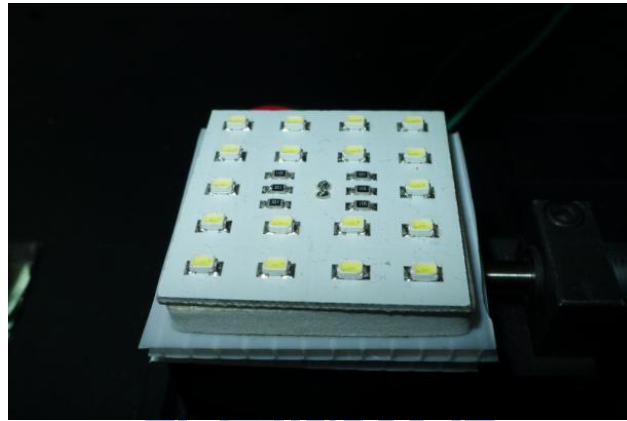
在影像取得之後，即可將影像輸入到本系統中，進行各項流程處理，從分析RGB 色彩分佈、二值化、飽和度臨界值、區域成長、最後則是特徵抽取與分類。



(a)



(b)



(c)

圖4.2實驗拍攝平台。(a) 血液抹片拍攝平台(b) XY軸微動平台 (c)LED背光板。



圖 4.3 掃描血液抹片路徑示意圖。

本論文共拍攝 400 張含有白血球的影像，經過獸醫師分類過後，可以將各類血球先以人工分類計數，如表 4-1，隨機使用其中 200 張來訓練支持向量機(support vector machine)分類器，再以另外 200 張來進行驗證，並將在 4.2 節進行實驗結果分析。

表 4.1 訓練及驗證資料

細胞種類	訓練資料	驗證資料
嗜中性粒細胞	156	156
嗜酸性粒細胞	29	29
嗜鹼性粒細胞	32	32
淋巴細胞	149	149
單核細胞	34	34

4.2 實驗結果分析與討論

本論文的貢獻在於白血球細胞核影像的切割，以及白血球細胞核分葉特徵抽取，在本章節將會利用實驗結果來驗證本論文之貢獻，實驗總共分為三部分，第一部分為白血球細胞核影像切割，第二部分為多重白血球分割定位，第三部分則為白血球細胞核影像識別。

4.2.1 白血球細胞核影像切割

作為對照，本論文引用參考文獻[2]以及幾個傳統方法來切割白血球細胞核影像，在參考文獻[2]中，使用了最大類間方差二值化以及尼布蘭克二值化法，再透過飽和度將影像分成白血球細胞核次影像，以及紅血球次影像，利用分水嶺演算法來分析這兩個次影像，並且計算出白血球個數以及紅血球個數；此外，本論文也將與在飽和度影像中只使用最大類間方差二值化法的白血球細胞核分割方式做比較，證實本論文所提出的方法比起近年所研究的細胞核影像分割法以更具有穩健性；而在影像分割這個階段，將分別使用染色正常以及紅血球較少重疊的血液影像、紅血球發生重疊的血液影像、染色技術不良之血液影像、以及曝光不足的血液影像以作為分割結果之比較標準。

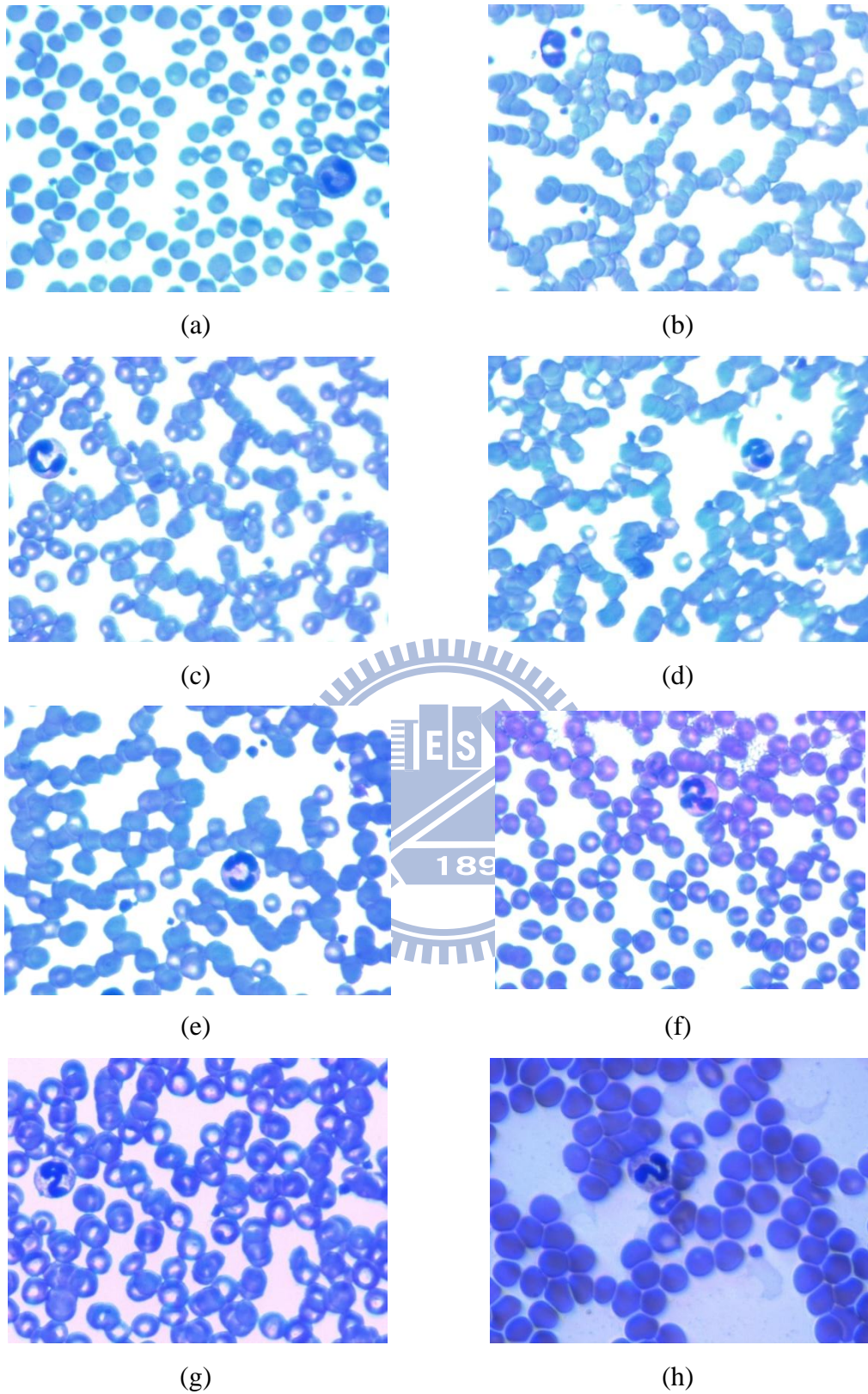


圖 4.4 驗證用血液影像。 (a)(b)無重疊無干擾影像 (c)(d)紅血球重疊影像(e)(f)染色不均影像 (g)(h)曝光不足影像。

表 4.2 白血球細胞核分割計數

計算影像中白血球細胞總數				
影像\方法	手算	本論文所提出的方法	[2]的方法	最大類間方差
(a) 無重疊無干擾	1	1	1	2
(b) 無重疊無干擾	1	1	1	4
(c) 紅血球重疊	1	1	1	5
(d) 紅血球重疊	1	1	2	7
(e) 染色不均	1	1	3	5
(f) 染色不均	1	1	1	5
(g) 曝光不足	1	2	2	9
(h) 曝光不足	1	1	2	15

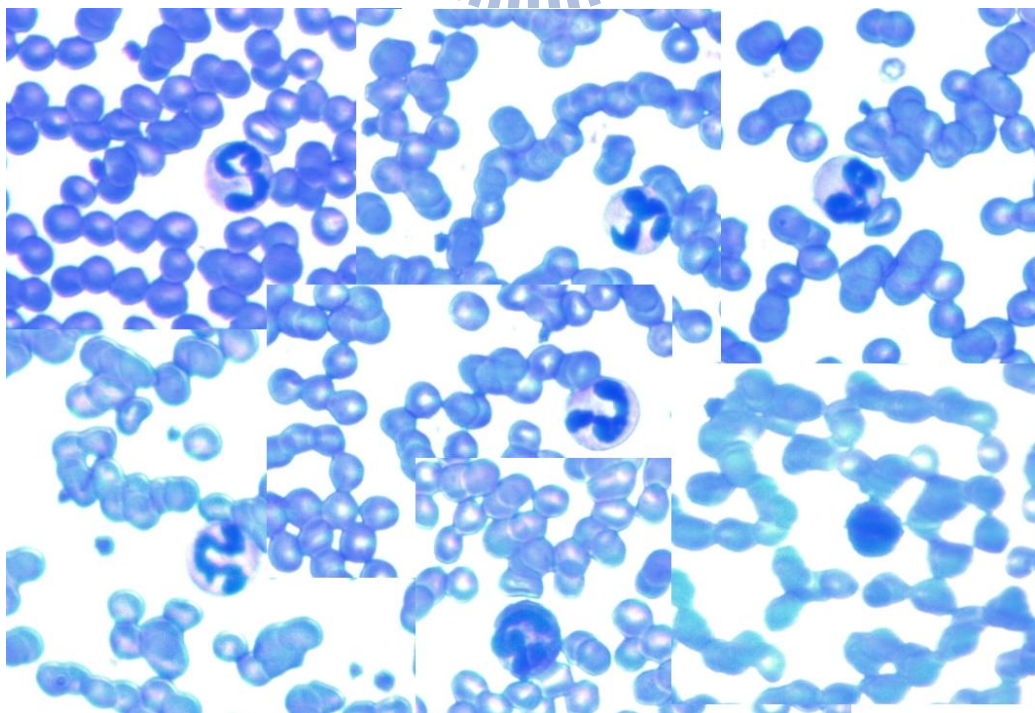
在這個影像分割的實驗中，造成分割上的困難的因素主要為血小板以及染色不均或者曝光不足，而透過紅血球細胞的大小估測模型，可以盡可能地排除掉血小板的干擾，因此在本論文的系統以及參考文獻[2]中，血小板的影響幾乎是微乎其微，主要構成影響的原因來自於染色不均，以及曝光不足，但是即使是這兩種情況，本論文的系統依舊有不錯的表現，原因在於本論文在影像輸入之後已經對其 RGB 三通道進行了分析與校正，因此當色彩空間自 RGB 轉換成 HSV 時，染色不均的影響已經被減弱了，如此一來 HSV 空間中的飽和度影像將能夠更貼近視覺上的感受，在後續以飽和度影像作分析時，將能夠更明確地將白血球細胞核與其他組織區分開來，相對於參考文獻[2]直接進行色彩空間轉換，而沒有考慮到在飽和度轉換公式上對於染色不均的影像會有不良的效果，本論文所提出的方法在染色不均的影像也有不錯的效果。

而相較於傳統的最大類間方差分割法，本論文所提出的方法以及參考文獻[2]的作法都有壓倒性的優勢。

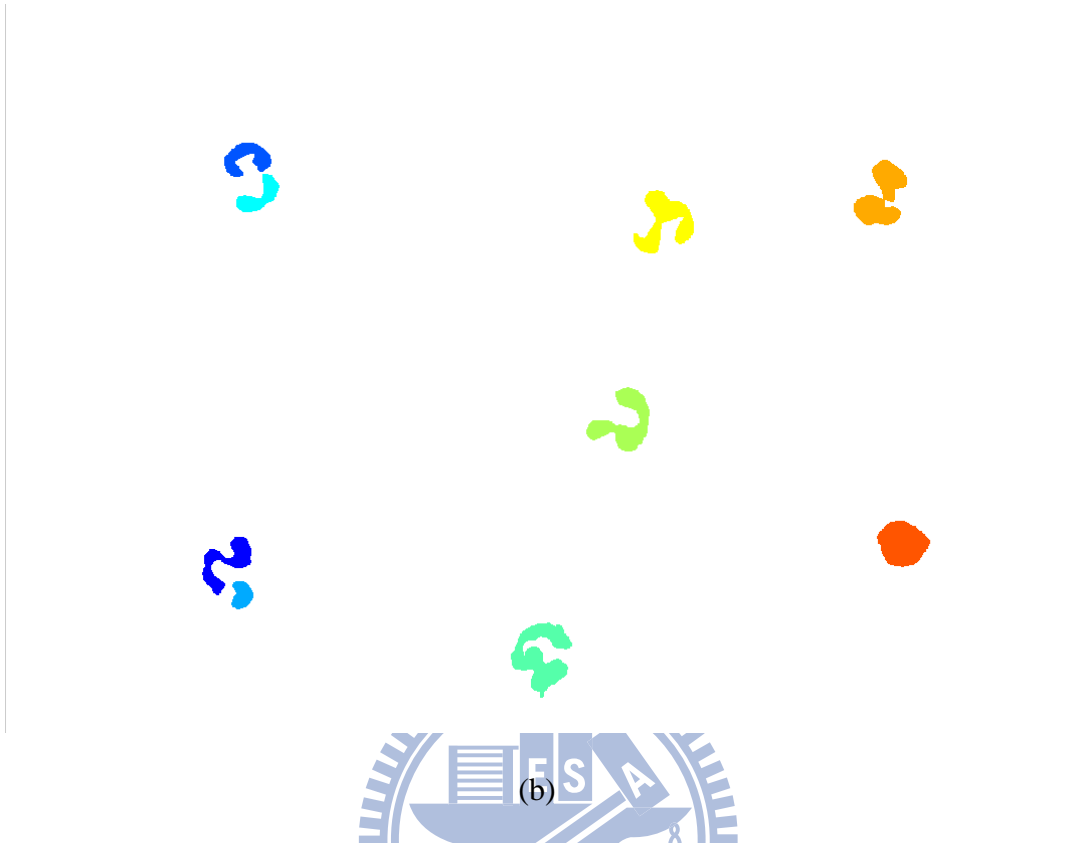
4.2.2 多重細胞核影像切割及定位

在 4.2.1 節，已經呈現出本論文在血液抹片影像中所產生的影像退化現象有不錯的穩健性，接下來將對於本論文的系統是否能完成多重細胞核影像切割及定位來作探討，在這個章節，將以四個例子(圖 4.5~圖 4.9)展示本系統之能力範圍，並非所有情況都能對目標區塊進行重新描述。

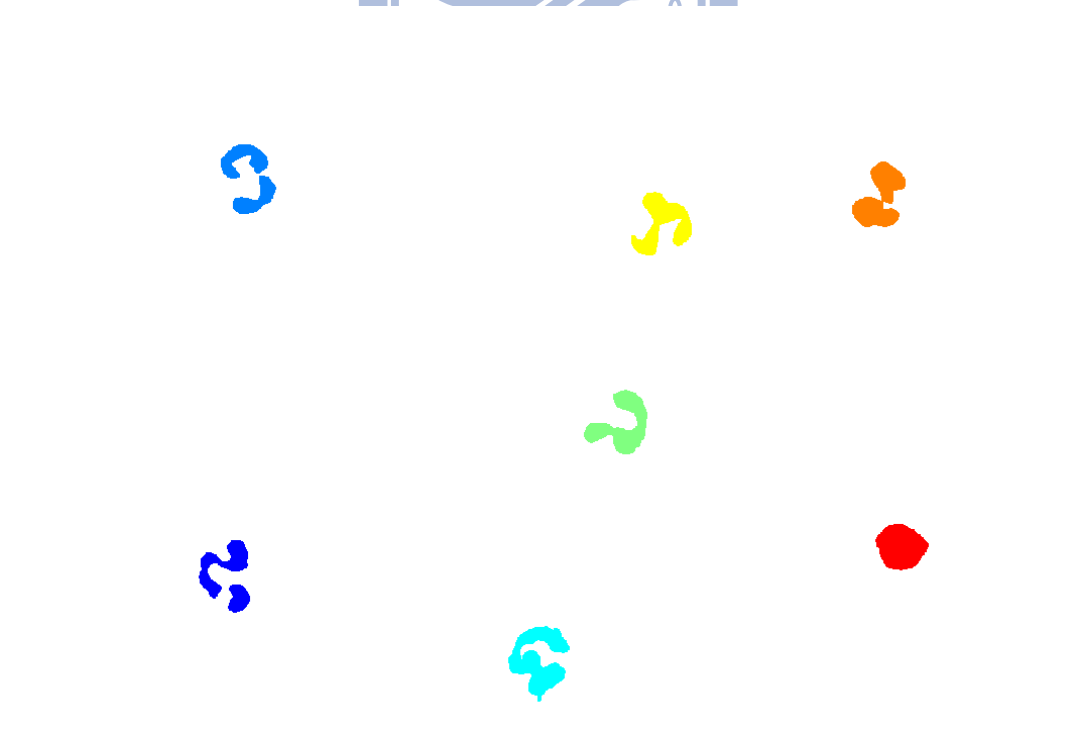
圖 4.5(a)並非取自真實血液影像，而是自行安排出來的影像，因為在真實血正常液抹片中，少有白血球凝集的現象發生，因此為了實驗確定本論文將可以達成多重細胞核影像切割及定位，在圖 4.5~圖 4.7 中，每張實驗用原影像皆是由實驗者自行剪貼自其餘血液抹片影像，試圖利用這些人工圖片來證實本論文的貢獻；在圖 4.5(a)中，白血球細胞核區塊呈現稀疏地錯落配置，在經過白血球細胞核分割之後，透過距離關係來重建各個區塊的關係，以獲得離散區塊是否為同一細胞核的資訊，由圖 4.5(c)比較圖 4.5(a)可知重新描述結果無誤。



(a)



(b)

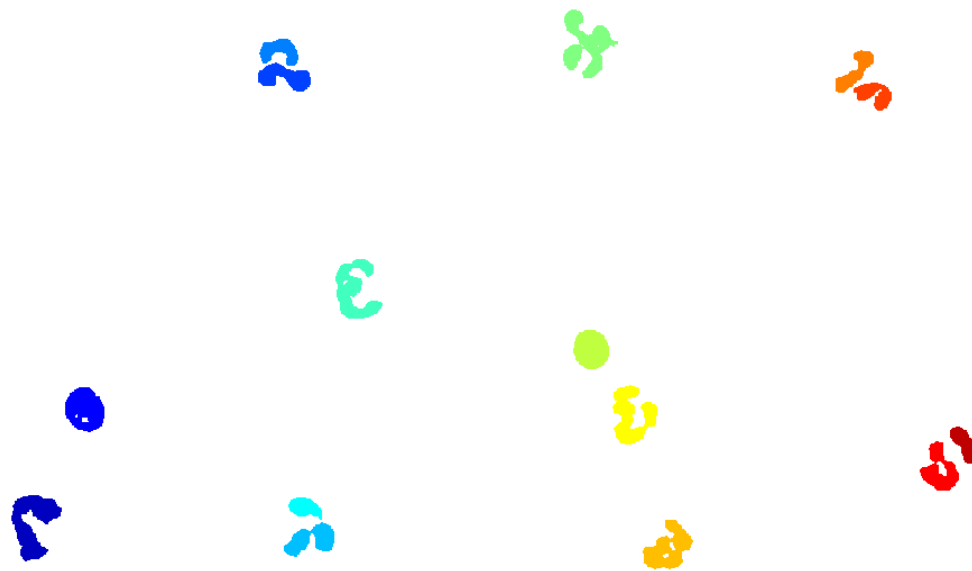
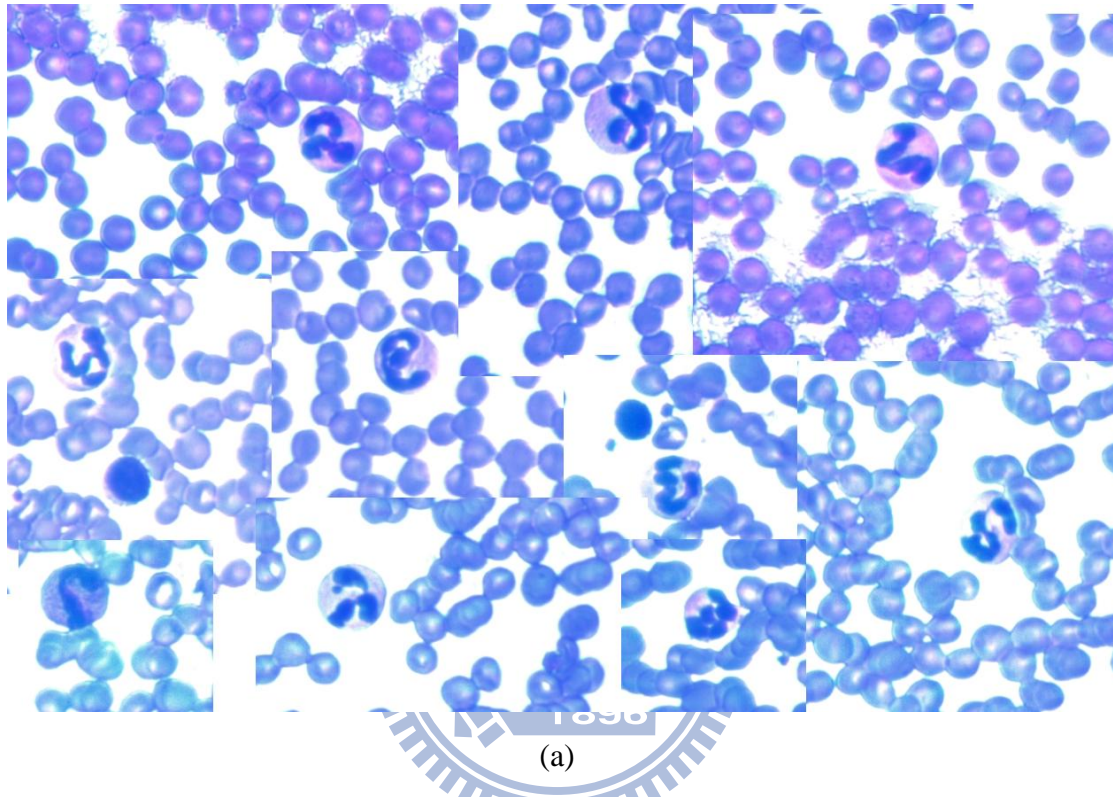


(c)

圖 4.5 多重細胞核定位及切割。(a)原圖 (b)細胞核切割且尚未重新描述之影

像 (c)重新描述後影像。

在圖 4.6 中，考慮加入更多的白血球細胞核，想要由此得知細胞核的數量多寡是否會影響系統的穩定性。



(b)

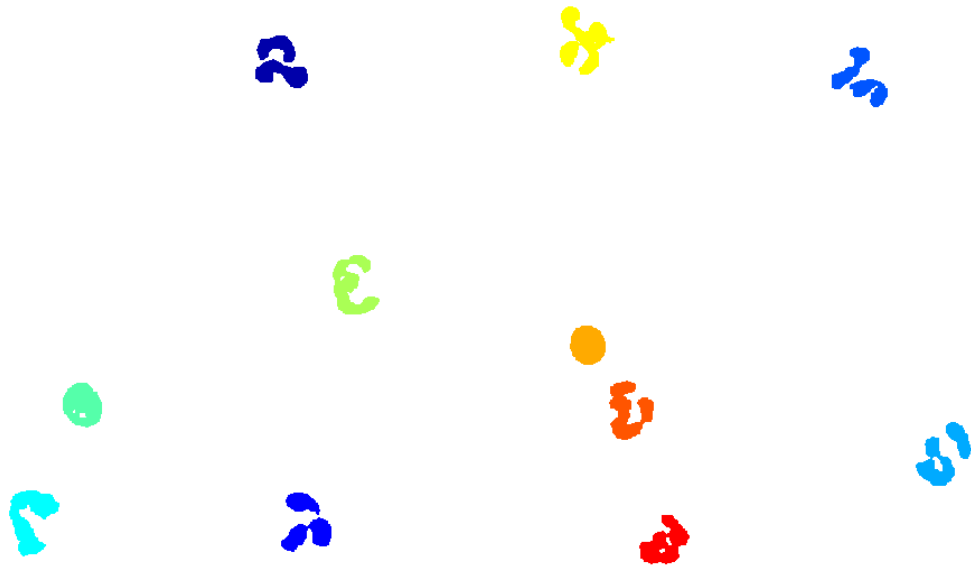
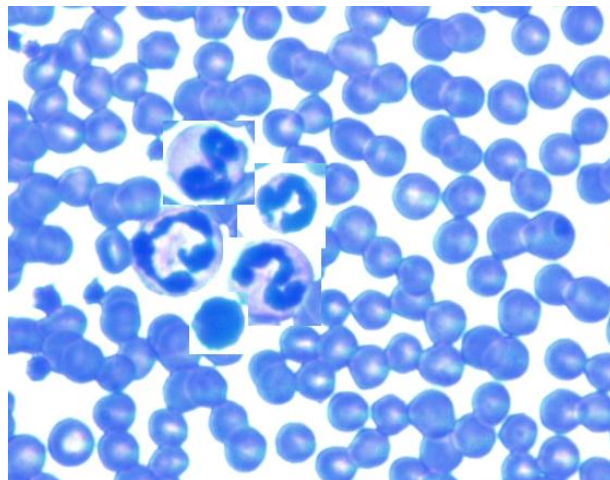


圖 4.6 多重細胞核定位及切割。(a)原圖 (b)細胞核切割且尚未重新描述之影像 (c)重新描述後影像。

由實驗可知本論文在重新描述白血球時，細胞核影像的多寡並不是影響本系統穩健性的原因，因為本系統在重新描述白血球時，所考慮的因素為距離關係，因此，白血球細胞核若太過接近可能會導致誤分類的情況發生，從圖 4.7 將可以發現這個現象。



(a)



(b)



(c)

圖 4.7 多重細胞核定位及切割。(a)原圖 (b)細胞核切割且尚未重新描述之影像 (c)重新描述後影像。

因此，可以知道本論文在細胞核影像太過接近的時候，無法正確地重新描述，主因即在於本論文用以重新描述的方法乃是以距離作為線索，來進行迭代，直到所有細胞核區塊之類別都不再變動為止，或者是超過迭代上限；倘若細胞核影像之距離太過接近，將無法把各個細胞核區塊做適當地分類，而在不斷迭代的結果

之後，將會達到迭代上限，或者是產生錯誤分類，因此本系統在多重白血球定位的能力上，受限於白血球細胞核聚集程度，但在正常血液中，白血球凝集現象不太可能發生，因此就正常的血液抹片而論，本論文所提出的方法仍具有不錯的效果。

表 4.3 多重細胞分割定位結果

影像\方法	手算	重新描述前	重新描述後
圖 4.5	7	9	7
圖 4.6	11	15	11
圖 4.7	5	9	7

4.2.3 白血球細胞核分類

在白血球細胞核影像分布較為稀疏的情況下，本論文將把各個白血球細胞核區塊抽取特徵並且進行分析，主要想要呈現的結果是在粒細胞以及非粒細胞間能夠有明顯區別而不會被誤分，要達成這個目的所使用的方法為抽取分葉特徵，以及利用分類器結構來達成目的，所採用的訓練集以及驗證集則由表 4.1 所標示的各類血球類別來擔任。

分類比較對象為人工辨識，以及參考文獻[15]的分類方法，該方法在分類部分與本論文所提出的方法最大的差異在於[15]使用了主動輪廓來做細胞核以及細胞質切割，為了減少實驗上的誤差，將跳過[15]的切割方法，直接以[15]的分類特徵以及分類法與本論文提出的分類特徵與分類法作比較，實驗結果將以三個表格來呈現，表 4.2 則為本論文所提出的方法實驗後的結果，表 4.3 為參考文獻[15]所提出的方法實驗後的結果。

表 4.1 訓練及驗證資料

細胞種類	訓練資料	驗證資料
嗜中性粒細胞	156	156
嗜酸性粒細胞	29	29
嗜鹼性粒細胞	32	32
淋巴細胞	149	149
單核細胞	34	34

表 4.4 根據本論文所作的分類實驗結果

輸入\輸出	嗜中性	嗜鹼性	嗜酸性	單核細胞	淋巴細胞	正確率
嗜中性	149	0	7	0	0	95.5%
嗜鹼性	0	30	2	0	0	93.75%
嗜酸性	3	0	26	0	0	89.65%
單核細胞	0	0	0	30	4	88.23%
淋巴細胞	0	0	0	8	141	94.63%

表 4.5 根據參考文獻[15]所作的分類實驗結果

輸入\輸出	嗜中性	嗜鹼性	嗜酸性	單核細胞	淋巴細胞	正確率
嗜中性	141	0	5	3	0	90.38%
嗜鹼性	0	21	1	0	10	65.6%
嗜酸性	5	0	24	0	0	82.76%
單核細胞	5	0	0	18	11	52.94%
淋巴細胞	0	0	0	15	134	89.93%

在分類實驗之後，來探討關於參考文獻[15]在分類上所提出的方法比較本論文的方法有何不同，為何會造成這樣的落差，討論的角度可以從分類所使用的特徵，以及分類器結構來討論。

在分類特徵的部分，[15]使用的特徵之維度為 26，其中 18 個為色彩特徵，8 個為形狀特徵，其中色彩的特徵的部分來自於細胞核與細胞質的 HSL 三通道數值的均值、標準差以及偏度(skewness)；另外 8 個形狀特徵則是凸度(convexity)、主軸比(principle axis ratio)、完備度(compactness)、圓變異度(circular variance)，以及橢圓變異度(elliptic variance)。

相較起本論文的特徵來說，在色彩特徵的選擇方式基本上都是為了衡量目標區域所呈現出來的「透光度」，因此得要考慮目標區域的色彩濃度，以及色調等等，雖然本論文採用的是紋理特徵，而[15]採用的是三通道的統計特徵，但目的都在於表達這一點。

但是在形狀特徵上則有很大的差異，本論文在抽取形狀特徵時，也考慮到不同細胞核的大小比例、細胞核分葉數目，以及輪廓曲率變化，因此就形狀特徵的選擇而論，比起[15]更能凸顯出細胞核的特色。

在分類器結構上，[15]為了能夠分類出白血球類別，因此建立 78 個節點的支持向量機二元樹，每個節點都只是為了區別出兩種類別，但是考慮到特徵只有 26 維度，過度的二元樹分支可能只會讓分類誤差累積到無法忽視的地步，因此，這個龐大的支持向量機二元樹或許是造成分類結果不佳的因素之一。

而在分類器結構上，本論文雖然也採用支持向量機二元樹，但只採用了 4 個節點，每個節點所參考的特徵向量隨著節點所要區分的類別而有差異，在根節點的部分，其功能在於區分出粒細胞以及非粒細胞，因為分葉特徵的作用，所以讓後續的節點不致於將粒細胞誤分類成淋巴細胞或者單核細胞，這項優點可以從表 4.2 了解，雖然分類結果仍未有滿意的表現，但是在本實驗過程中粒細胞以及非粒細胞的部分並不會有誤分的情況發生，因此，在這樣的架構下，分葉特徵確

實讓分類的結果更為穩定，而這也是本論文的貢獻之一。

第五章 結論

本論文在相關研究領域中所產生的貢獻在於克服了本論文所指涉的雜訊發生的情況下的白血球細胞核抽取，以及在抽取之後的分類特徵分析，這在檢驗醫學中所能產生的貢獻在於減輕人工鏡檢在經濟、人力以及時間上的負擔，倘若能在不改變現行檢驗器材的情況下(數位顯微鏡)，而能夠對血液進行計數分析，那麼就能夠將更多的經費、人力以及時間花費在更需要這些資源的檢驗項目，或者是其他單位部門。

本論文的研究目前僅止於帶雜訊的正常血液抹片，而在現行的全血計數分析中，事實上較常用自動化血液分析儀器，而非傳統人工鏡檢，只有遇到病變、異常，或者非典型的細胞核，才會需要人工進一步檢驗；本論文只是後期研究的開端，先對正常血液抹片進行白血球細胞核的分析，以了解相關研究實現方法的異同，以及針對一種白血球細胞核的異常(分葉數目多寡)進行了初期研究，在某些情況下，白血球細胞核的過度分葉，或者是分葉不足，都可能是暗示著某種疾病，或者能夠提供白血球細胞的成熟程度；在未來的研究中，研究資料將會是異常血液抹片，針對某些特定的血液疾病進行數位影像分析，以期能提供某些重要的診斷資訊。

此外，若把數位影像技術用來分析異常血液抹片，對於檢驗醫學也能產生很大的助益，因為在醫學檢驗上，針對各種不同的疾病需要不同的儀器或者是檢驗套組，例如寄生蟲偵測、菌血症偵測、愛滋病以及白血病等等，未來倘若能結合數位影像技術，以及特殊細胞化學染色法，則這些疾病或許都能夠通過同樣的方法來檢驗，如此一來，或許原本耗時費力的過程，在未來將會變成是一個簡單的流程。

參考文獻

- [1] 何敏夫, 血液學, 4th edition, 合記圖書出版社, 台北市內湖區安康路322-2號, Sep. 2010.
- [2] M. Habibzadeh, A. Krzyzak, and T. Fevens, "Counting of RBCs and WBCs in Noisy Normal Blood Smear Microscopic Images," in *Proceedings of SPIE on Medical Image*, Lake Buena Vista, Florida, USA, Feb. 2011, vol. 7963, pp. 79633I-79633I-11.
- [3] M. Bober, "MPEG-7 Visual Shape Descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, June 2011.
- [4] J. W. Harvey, *Atlas of Veterinary Hematology - Blood and Bone Marrow of Domestic Animals, 1st edition*, Saunders, Philadelphia, USA, Jan. 2001.
- [5] M. Partio, B. Cramariuc, M. Gabbouj, and A. Visa, "Rock Texture Retrieval Using Gray Level Co-occurrence Matrix," in *Proceedings of 5th Nordic Signal Processing Symposium*, Oct. 2002.
- [6] K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32-40, Jan. 1975.
- [7] Y. Cheng, "Mean Shift Mode Seeking and Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790-799, Aug. 1995.
- [8] J. Theerapattanakul, J. Plodpai and C. Pintavirooj, "An Efficient Method for Segmentation Step of Automated White Blood Cell Classifications," in *Proceedings of IEEE Region 10 Annual International Conference (TENCON 2004)*, Chiang Mai, Thailand, vol. 1, pp. 191-194, Nov. 2004.
- [9] F. Zamani and R. Safabakhsh, "An unsupervised GVF Snake Approach for White Blood Cell Segmentation based on Nucleus," *IEEE Conference on Signal Processing, The 8th International Conference*, vol 2, Nov. 2006.

[10] S. H. Rezatofghi, H. Soltanian-Zadeh, R. Sharifian and R.A. Zoroofi, “A New Approach to White Blood Cell Nucleus Segmentation Based on Gram-Schmidt Orthogonalization,” *IEEE International Conference on Digital Image Processing*, pp. 107-111, 2009.

[11] K. Cheng “White Blood Cell Detection Using a Novel Fuzzy Morphological Shared-Weight Neural Network,” *IEEE International Symposium Computer Science and Computational Technology* ,vol 2 , pp. 532-535 , Dec. 2008.

[12] E. Xie, T. M. McGinnity, and Q. X. Wu, “Automatic Extraction of Shape Features for Classification of Leukocytes,” *IEEE International Conference Artificial Intelligence and Computational Intelligence*, vol 3, pp. 220-224 , Oct. 2010.

[13] J. A. Bilmes, “A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models,” University of Berkeley, Tech. Rep. ICSI-TR-97-021, 1997.

[14] D. Comaniciu and P. Meer, “Mean Shift: a robust approach toward feature space analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 24, no. 5, pp. 603–619, 2002.

[15] H. Ramoser, V. Laurain, H. Bischof, and R. Ecker, “Leukocyte segmentation and classification in blood-smear images,” in 27th Annual Intern. Conf. of the Engineering in Medicine and Biology Society, September 2005, pp. 3371–3374.