

影像分類機制實作 - 以數位典藏管理系統為例

王喻正 李政宏 陳貴青 王祥安 邱志義

中央研究院 資訊科學研究所

{wyc, chli, ching64, sawang, cychiu}@iis.sinica.edu.tw

摘要

近年來，隨著創作及傳播影像資料的工具和系統有著顯著地進步，使得數位影像的產量呈現爆炸性地成長，對於存放大量數位影像資料的數位典藏系統而言，一個有效率的影像管理機制是必要的。在本文中，我們提出一個有效的影像分類技術，能有效地提高影像管理的效率，當典藏系統匯入大量影像資料的同時，此一技術可以被用來對大量影像自動進行分類；並且，透過此一分類機制，將可發展與改良影像自動註解之技術。本文提出的方法首先將影像分割成多個不重疊的區塊並且擷取影像的顏色與紋理特徵，並透過支撐向量法(Support Vector Machine, SVM)進行分類模組之訓練以及影像資料的分類。由實驗結果的驗證，本文提出的影像分類機制可以有效分類影像，有助於提升數位典藏管理系統對影像的管理能力。

關鍵詞：影像分類、支撐向量法

1. 前言

近幾年，在數位典藏單位中典藏的數位影像檔案的數量有著巨大的成長，使得管理這些影像檔案變得越來越困難，為了降低在數位影像管理上所花費的人力物力，一個自動化的管理工具是數位圖書館(數位典藏單位)所迫切需要的。此外，在數位影像管理相關的研究活動中，包含影像擷取(image retrieval)、數位權利管理(digital rights management)和影像註解(image annotation)，這些研究活動都與影像分類技術有很大的關聯性，若典藏單位有大量的影像需匯入資料庫，一個高效率的分類技術對於維護數位典藏單位之影像資料品質與正確性是非常必要的。例如：有一大量未嵌入影像資訊(metadata)的原始影像，影像分類技術可以用來將影像儲存於適當的分類集合中，甚至將這些影像加上分類註解。除此之外，此一分類技術也可以用來驗證其他透過人工分類的影像標籤正確性，並且對於難以透過人工決定所屬分類的影像，能夠確認其所屬的分類群組，以維持數位典藏單位中影像資料之品質。

『數位典藏國家型科技計畫』[1]在民國 91 年 1 月 1 日正式成立，是承襲行政院國家科學委員會『數位博物館計畫』、『國家典藏數位化計畫』、『國際數位圖書館合作計畫』三個計畫的經驗，依據國家整體發

展，重新規劃而成。在數位典藏計畫中，各典藏單位將其珍貴的歷史文物轉換成數位檔案進行儲存與典藏，產出了大量的數位內容。數位內容在形式上有別於傳統有形著作，必須面臨許多不可避免的問題與挑戰，例如資料如何被管理、保護。國家典藏的數位化，可以有效提升知識的累積、傳承與運用，是知識經濟的重要基礎環節，要如何享受國家典藏數位化後為我們帶來的便利，是一門相當重要的課題。

本研究將影像分類技術實作並與現有數位典藏管理系統進行整合，影像分類在此系統包括二個主要元件：離線影像訓練模組以及即時影像分類模組。在離線影像訓練模組中，每張影像被分割成 $n \times n$ 個不相重疊的區塊，每一個區塊的顏色和紋理分布直方圖(color/texture histogram)資訊被取出作為這個影像區塊的特徵值，並透過支撐向量法(Support Vector Machine, SVM)進行數個影像類型模組的訓練。而即時影像分類模組中，亦擷取與上述相同的特徵值，最後再透過訓練過之支撐向量法模組進行影像分類處理。

本篇論文將於第二節探討影像處理、檢索、圖形辨識以及機器學習之相關技術。第三節將詳細說明本文提出之架構、影像特徵的擷取以及支撐向量法的訓練與分類方法。在第四節將描述實作的系統與實驗結果。最後，第五節提出結論並且探討影像分類機制的未來研究方向。

2. 相關文獻探討

早從 1990 年初期，許多的內涵式影像檢索系統(content-based image retrieval, CBIR)已經被發展出來。其中，針對影像資料庫的內涵式影像檢索[6]已經被廣泛地應用於相關的研究領域，如影像處理(image processing)，資訊擷取(information retrieval)，圖形辨識(pattern recognition)和資料庫管理(database management)，大多數的內涵式影像檢索研究，致力於尋找能有效表示影像的特徵資訊。包含分布直方圖的方法已經被廣泛地使用於解決各種影像分類的問題；Szummer 和 Picard[10]，以及 Vailaya 等人[11]利用顏色分布直方圖資訊，將戶內和戶外的影像進行分類，Chapelle 等人[4]則以影像顏色分布直方圖為特徵資訊並且應用支撐向量法進行影像分類，包含影像物件類別的分類。這些研究說明了分布直方圖方式所花費的計算處理成本較低而且對於某些分類個案有很好的效

率。然而，一般分布直方圖表示法的最大缺點就是忽略了空間結構的資訊。因此，其他的特徵值像是紋理特徵值和形狀特徵值等，被提出來解決空間結構相關特徵值擷取之問題。

另外亦有許多研究探討將影像分割成多個區塊來擷取特徵值的方法，Gorkani 和 Picard[5]的方法首先劃分影像成 16 個相同大小且相互不重疊的區塊，接著計算每一個區塊的主方向(dominant orientation)值，並根據區塊的主要方向做為特徵值分類城市或鄉鎮場景。此外，Wang 等人[12]發展了一套分類演算法，將圖片(graph)或照片(photograph)進行分類，這個分類法將影像分割成數個區塊，而且將每個區塊根據其高頻區域的小波系數來將影像分類為圖片類別還是照片類別，若區塊被分類為照片區塊的百分比高於預先設定之門檻值，這張影像就會被標記為照片；反之則被標記為圖片。Yu 和 Wolf [14]針對室內和室外場景的分類提出隱藏式馬可夫模型(Hidden Markov Model, HMM)分類法，此方法則是針對向量量化(vector quantized, VQ)後的影像區塊進行顏色分布直方圖訓練。Li 和 Wang[7]提出一個以影像區塊的顏色和紋理特徵為基礎，並透過二維多解析隱藏式馬可夫模型訓練，以期獲得特定的影像類別。Yanai[13]提出一通用的影像分類系統，此系統並非使用人工所收集的影像集合，而是透過網際網路上自動收集的影像作為訓練資料。Murphy 等人[9]則提出四種圖片模型可以將圖片物件與影像區塊的

特徵值產生關聯，得以進行相連之場景以及物件辨識之處理。

3. 影像分類架構

影像分類架構在本系統包括二個部分：離線影像訓練模組以及即時影像分類模組。在離線影像訓練模組部分，本系統自典藏系統的影像資料庫中擷取影像資料特徵值，並且透過支撐向量法訓練，得到各分類之訓練模型，接下來，在即時影像分類模組部分，對未知影像擷取影像特徵值，再透過已訓練完成之支撐向量法分類器模型將未知的影像分類至合適的影像類別中，如圖 1 所示為本文提出之數位典藏管理系統影像訓練和影像分類的流程架構。

3.1 特徵擷取

在數位典藏計畫中，由於同種類別的典藏影像通常具有相似的特徵，包括文物位置、文物顏色和紋理等，例如，在考古資料中的影像通常在影像的中央會有一個古代的文物，而在漢代簡牘的影像資料中其彩色影像通常有著自然的木質顏色，且其紅外線影像則有彩色影像無法顯現的細微紋理。因此，本系統將以位置、顏色與紋理做為擷取之特徵資訊，下列為顏色和紋理特徵值之擷取過程。

圖像訓練模型

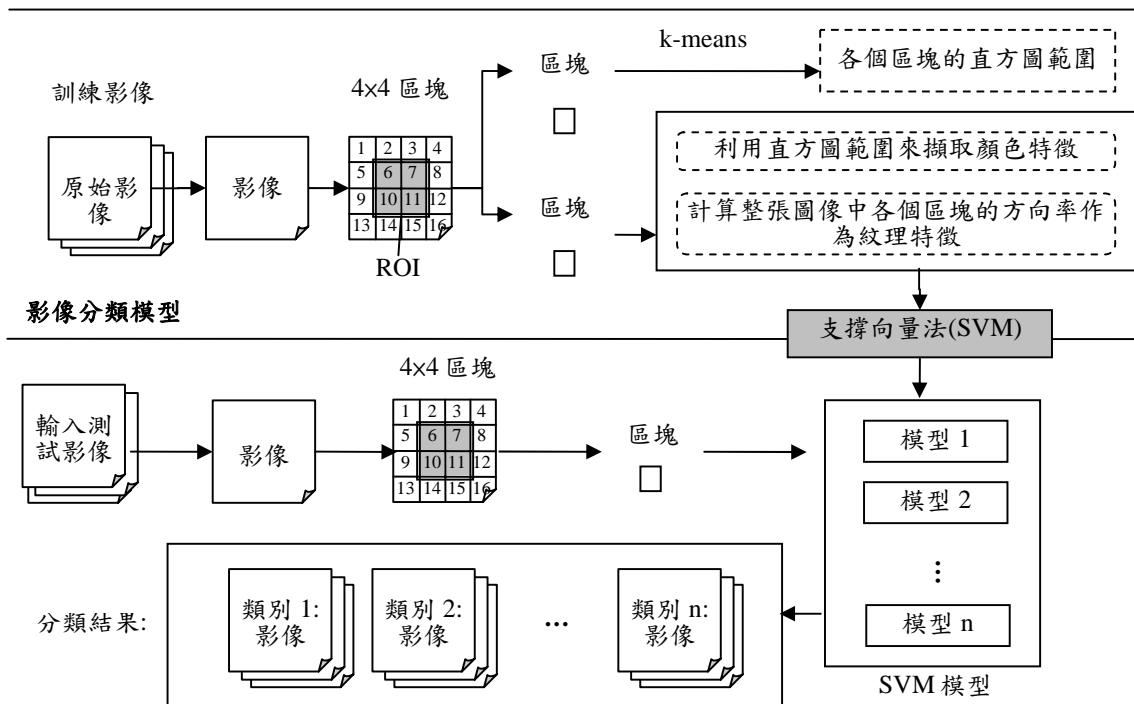


圖 1. 影像訓練和分類流程圖

為了獲得顏色的特徵值，第一步先將要訓練的影像進行影像大小的正規化處理，將訓練的影像大小轉成 96x20 像素大小的影像，並且將它劃分成 4x4 個不重疊的區塊(如圖 1 所示)。接著，計算每一個區塊的顏色分布直方圖向量矩陣，平均向量則經由計算十六個區塊的顏色分布直方圖的加總平均得到，接下來計算每一個區塊顏色分布直方圖的向量與平均向量的差異，最後，我們會得到十六個影像特徵值作為影像的顏色特徵表示。

在紋理特徵的擷取上，為了取得每一個區塊之方向性，首先對每個區塊採用了 Canny 的邊緣偵測演算法(Canny's edge detection algorithm)[2]，該演算法套用高斯遮罩(Gaussian Mask)的平滑效應消除影像中的雜訊，並且使用一組 3x3 的索貝爾迴旋遮罩(Sobel convolution mask)來估計 x 軸方向(Gx)和 y 軸方向(Gy)的梯度值(gradient)。接著，我們使用 Gx 和 Gy 的梯度值來計算像素的方向性，如下。

$$Direct = 180 - ((\tan^{-1}(Gy/Gx)) * (180/\pi)); \quad (1)$$

最後再將計算後的邊緣方向性進行量化處理，對於邊緣方向性的值(edge direction)會根據它所座落的範圍給定一個對應的值。邊緣方向性的值如果位於(0~22.5 & 157.5~180 度)的範圍內將被設為 0 度。任何邊緣方向性的值如果位於(22.5~67.5 度)的範圍內則設為 45 度；位於(67.5~112.5 度)設為 90 度；位於(112.5~157.5 度)設為 135 度。由於數位典藏系統中之影像，文物多處於影像中央，因此，本系統僅計算影像中四個我們所感興趣的區域(region-of-interest, ROI)的區塊(為圖 1 中的區塊 6, 7, 10 和 11)來決定這張影像的紋理特徵表示值。

3.2 支撐向量法的訓練和分類

支撐向量法是一種利用最小的架構風險策略的機器學習理論，此一理論在處理沒有交互相關領域知識的樣本分類問題上是非常有效率的[8]，該理論根據二元分類法，尋找超平面(hyper-plane)將物件劃分成兩類，並保證分類的最小誤差率。支撐向量法主要的好處是，它可以藉由訓練和分析來處理線性不能分離的資料來產生支撐向量資料，再從訓練資料中屏除較極端資料之後，將此支撐向量建立成一個標準模型(model)。接著，此系統將現有存在的測試資料依據建立好的標準模型，分類至適當的類別。

本文採用顏色和紋理特徵作為訓練和分類的依據，對此我們將訓練影像的特徵資訊以下列的正規表示法來描述。

[標籤 I] 索引 1]:[值 1] [索引 2]:[值 2] ...

例如，第一張影像的特徵向量可以被表示為：

1 1:0.268169 2:0.332564 3:0.246752 4:0.101631:...

這裡，[標籤]表示為影像在所屬總類的編號，[索引]則是影像特徵值的序列號。在這篇論文中，我們採用支援向量法的類神經(radial basis function)技術作為核心學習功能。

4. 實驗結果

本系統的影像資料庫建置於 Redhat Linux 7.3 系統上，其影像資料庫包含考古資料、佛教石刻造像拓本、漢代石刻畫像拓本、傅斯年善本古籍和漢代簡牘。而系統核心主要分成兩大部分：多媒體中心(the multimedia center web system, MMC)和影像分類模組(the image process daemon, IPD)，這兩個核心部分共享功能模組、資料庫及儲存媒體。多媒體中心負責檔案上傳時的接收、檔案的儲存、以及檔案的管理。而影像分類模組負責檔案上傳時的分類工作。多媒體中心透過呼叫和編譯應用伺服器的 JavaBeans 來執行 Servlet。影像分類模組使用功能模組層來聯繫資料庫及儲存媒體。分類模組則是用 Java-Servlet 和 SVM 工具[3]開發出來。圖 2 為部份數位典藏系管理系統的類別影像。

本實驗總共採用六種典藏資料類別，共 2716 張影像(# CI)、385 張訓練影像(# TRI)和 2331 張待測試的影像(# TI)，在訓練階段，本實驗在各個類別使用不同的交叉確認(Cross Validation)參數(# CV)來增加分類準確性((%) CVA)。最後的實驗結果透過正確率(recall rate)、準確率(precision rate)以及 F-Measure 來評估分類的成果。表 1 列出本實驗影像群集(clustering)的正確率，而表 2 則顯示經由實驗來評估我們所提出的分類技術所獲得準確率，正確率(Rec. (%))用 TI_R / TI_T 來表示，精確率(Pre. (%))定義為 $TI_R / (TI_R + TI_W)$ 來表示，而 F-Measure 為正確率和精確率的調和平均數， TI_R 、 TI_W 和 TI_T 分別為正確分類之測試影像個數、分類錯誤之測試影像個數和所有測試影像個數，所有的值都用百分比來表示，測試影像經由各個支援向量法分類器評估並得到一個決定係數(decision value)，並且根據分類器所得到最高的決定係數作為測試影像的類別配置依據，ADV 為在分類器中決定係數的平均值。我們得到一個平均的 F-Measure 為 86.67%，如表 2 所示。

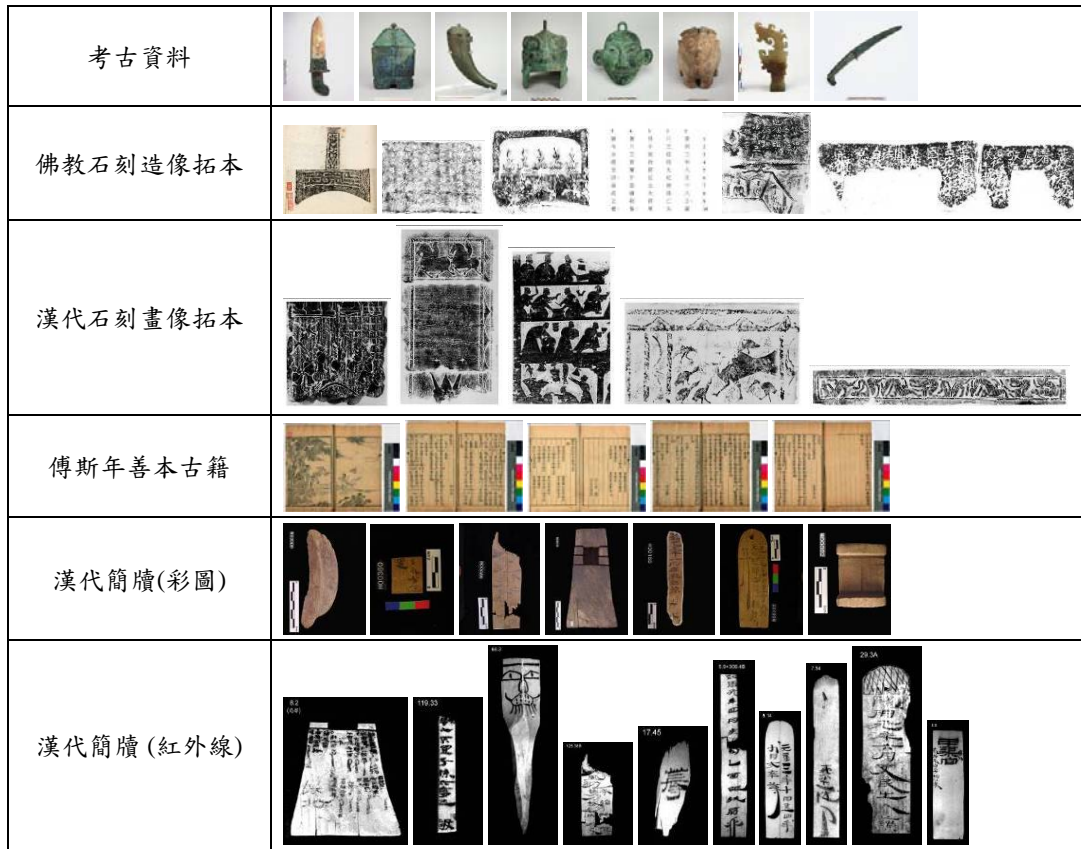


圖 2. 數位典藏管理系統各類別影像

表 1. 影像群集實驗結果

類別	考古資料	佛教石刻造像拓本	漢代石刻畫像拓本	傅斯年善本古籍	漢代簡牘(彩圖)	漢代簡牘(紅外線)	總和 / 平均
# CI	154	498	499	498	569	498	2716
# TRI	36	72	72	72	71	62	385
# TI	118	426	427	426	498	436	2331
# CV	5	50	40	5	5	15	
(%) CVA	98.96	90.65	89.87	100	98.96	94.55	

表 2. 影像分類的實驗結果

類別	考古資料	佛教石刻造像拓本	漢代石刻畫像拓本	傅斯年善本古籍	漢代簡牘(彩圖)	漢代簡牘(紅外線)	總和 / 平均
ADV	1.571	1.008	0.958	0.989	0.904	0.982	1.073
Rec. (%)	114/118 96.61	339/426 79.58	281/427 65.81	421/426 98.83	498/498 100.00	352/436 80.73	86.93
Pre. (%)	114/122 93.44	339/457 74.18	281/358 78.49	421/422 99.76	498/574 86.76	352/398 88.44	86.85
F (%)	95.00	76.78	71.59	99.29	92.91	84.41	86.67

圖 3 和圖 4 為在數位典藏管理系統中，考古資料和傅斯年善本古籍資料的分類結果。由於考古、傅斯年善本古籍和漢代簡牘(彩圖)各類別中的影像資料顏色調較相似，而佛教石刻造像拓本、漢代石刻畫像拓本和漢代簡牘(紅外線)由於各類別中之影像解析度不規

則以及有不同的紋理特徵。經由實驗發現，考古資料、傅斯年善本古籍和漢代簡牘(彩圖)的實驗結果相較於佛教石刻造像拓本、漢代石刻畫像拓本和漢代簡牘(紅外線)來得好。另外，透過表 2 的實驗結果可以明確地看出本文所提出之影像分類技術是極為有效的。



圖 3. 考古資料分類結果



圖 4. 傅斯年善本古籍分類結果

5. 結論與未來的研究方向

本文將影像分類技術結合於數位典藏管理系統中，影像被分割成 $n \times n$ 個不重疊的區塊，針對各個區塊擷取其顏色和紋理的分布直方圖資訊作為代表它們的特徵值，並利用這些特徵表示資訊來訓練並產生支撐向量法的分類模型。經由實驗顯示，本文所提出的技術對於處理數位典藏單位所提供之影像分類，有著顯著有效的分類結果，相信此一影像分類技術對於龐大的影像資料庫進行自動註釋和分類驗證是非常有效的。

在未來的研究課題上，本研究將延伸其它影像特徵，包括：形狀(shape)、空間架構(spatial layout)等，以及分類技術，包括：貝氏網路(Bayesian networks)、決策樹(decision tree)等方法。此外，未來將可建置一個多層式的語義註釋架構以促進影像管理和檢索。

6. 誌謝

- [1] 行政院國家科學委員會，數位典藏國家型科技計畫-技術研發分項計畫-總計畫：NSC95-2422-H-001-007。
- [2] 行政院國家科學委員會，數位典藏國家型科技計畫-技術研發分項計畫-典藏系統建置與相技術研發計畫：NSC95-2422-H-001-008。

7. 參考文獻

- [1] 數位典藏型國家科技計畫，<http://www.ndap.org.tw/>
- [2] Canny, J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6) (1986) 679-698
- [3] Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. (2001). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [4] Chapelle, O., Haffner, P., Vapnik, V.N.: Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5) (1999) 1055-1064
- [5] Gorkani, M.M., Picard, R.W.: Texture orientation for sorting photos "at a glance". In *Proceedings of the*

- 12th IAPR International Conference on Pattern Recognition,, (1994) 459-464
- [6] Kim, D.-H., Chung, C.-W.: QCluster: Relevance Feedback Using Adaptive Clustering for Content-based Image Retrieval. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, (2003) 599-610
- [7] Li, J., Wang, J.Z.: Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9) (2003) 1075-1088
- [8] Moghaddam, B., Yang, M.-H.: Gender Classification with Support Vector Machines. *IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG)*, (2000) 306-311
- [9] Murphy, K., Torralb, A., Freeman, W.T.: Using the forest to see the trees: a graphical model relating features, objects, and scenes. In *Advances in Neural Information Processing Systems 16*. Cambridge, MA:MIT Press, (2004)
- [10] Szummer, M., Picard, R.W.: Indoor-outdoor image classification. In *Proc. IEEE Int'l Workshop on Content-Based Access of Image and Video Databases*, (1998) 42-51
- [11] Vailaya, A., Figueiredo, M.A.T., Jain, A.K., Zhang, H.-J.: Image classification for content-based indexing. *IEEE Transaction on Image Processing*, Vol. 10, No. 1, (2001) 117-130
- [12] Wang, J.Z., Li, J., Wiederhold, G.: SIMPLiCity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9) (2001) 947-963
- [13] Yanai, K.: Generic Image Classification Using Visual Knowledge on the Web. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, (2003) 167-176
- [14] Yu, H.H., Wolf, W.H.: Scenic classification methods for image and video databases. In *Proc. SPIE Int'l Conf. on Digital Image Storage and archiving systems*, 2606 (1995) 363-371