

植基於主成份分析與支援向量機的影像語意內容分析系統

張傳育
國立雲林科技大學
電腦與通訊工程系
chuanyu@yuntech.edu.tw

王宏仁
國立雲林科技大學
工程科技研究所
g9210810@yuntech.edu.tw

李綺芳
國立雲林科技大學
資訊工程研究所
g9417707@yuntech.edu.tw

摘要

隨著數位相機的普及，數位影像的拍攝與取得變得極為容易，現行數位像機中幾乎都是以日期加上編號的方式來命名，造成日後影像檢索與管理上的困難。另外，由於數位影像、資料庫和網路技術的快速發展，對於大型資料庫而言，以內容(Content)為基礎的影像檢索(Image Retrieval)已成為一個重要的研究題目。另一方面，如何有效的管理大量的數位影像，特別是對於語意式的影像檢索系統而言，仍未獲致良好的解決方法。因此，我們將提出一個有效的影像語意內容分析技術，利用支援向量機(Support Vector Machine, SVM)結合低階的影像特徵以分析影像的內容，並計算影像中各種高階語意項目的成分，進而達到影像管理的目的。同時，由於影像特徵具有高維度，我們進一步使用主成份分析(Principal Component Analysis, PCA)來分析影像特徵，從中選取有效及有鑑別度的影像特徵，以適當地降低 SVM 的輸入特徵維度。實驗結果顯示，本論文提出的結合 PCA 與 SVM 方法對於語意為基礎的照片分析具有高的準確度，而且分析的結果比較接近人類視覺的判斷，也比傳統的半徑式基底函數類神經網路(Radial Basis Function, RBF)的分析準確度高。

關鍵詞：影像語意內容分析、支援向量機、主成份分析。

一、前言

隨著電腦儲存設備容量與日俱增、網路頻寬的增加，再加上數位相機、掃瞄器等影像設備的普及，影像取得與傳送更為方便，為我們帶來了大量的影像資料。面對這些龐大的影像資料，影像的檢索與管理已成為一個重要的課題。目前已經有學者提出以內容為基礎的影像檢索系統，例如：QBIC[1]和 Photobook[2]。影像檢索系統是藉由擷取輸入影像中的一些特徵值，與資料庫中事先計算好的相對特徵作相似度的比對，然後將最相似的幾張影像及相關資訊回傳給檢索的使用者。儘管這些系統的研究已經有不錯的成果，但是檢索的結果仍然未達到令人滿意的效能，其主要原因在於影像的顏色分佈、影像的物件形狀及影像的紋理結構等這些低階特徵和符合人類視覺感知之間存在的語意差距。在

影像分類技術中，已有使用語意標籤的方法[3]被提出使用，希望可以解決上述的問題。然而，對於一張影像而言，影像中可能包含相當多的物件，不同的物件有不同的語意類別，導致影像分類系統很難辨識這麼大量的物件語意。目前利用切割影像尋找物件並且根據使用者提供語意的影像分類方法已被提出，用以解決高階語意與低階特徵的問題，使低階特徵與高階人類視覺感知的語意得以溝通對應連接。因此，我們嘗試以區域為基礎的切割方法進行影像切割(Segmentation)，用以解決高階語意與低階特徵的問題，並採用物件的影像特徵作為代表。將影像切割成許多小區域，理想的切割方法所切割出來的每個小區域代表一個物件區域，這種切割方法跟人類視覺是類似的。

由於各種影像的特徵分佈，無法由一般的線性分類器達成，因此，我們使用監督式學習方式，利用支援向量機(Support Vector Machine, SVM)[5-6]經由最佳化超平面分類並於高維度空間俱有良好分類能力的特性，透過監督式學習機器，將各種影像的特徵建立模型。

在影像資料中，鄰近像素的相關性頗高，因此，影像所擷取出的低階影像特徵可能會存在一些多餘的資訊，而且特徵越多不見得分類效果越好，不良的特徵值會影響有效特徵的識別力，可能導致影像分類誤判，也會增加網路的學習時間。因此，我們採用主成份分析(Principal Component Analysis, PCA)[7-8]對所有特徵進行分析，挑選具鑑別度的特徵，以降低特徵維度。

在我們提出的影像語意內容分析系統架構中，包含兩個階段：訓練階段和測試階段，圖 1 展示影像語意內容分析系統流程。在訓練階段，首先對影像進行分割，並且計算每個區域的各種影像特徵與高階語意的對應，同時，使用 PCA 分析影像特徵，從中選取有效及有鑑別度的影像特徵，做為 SVM 的訓練樣本。在測試階段，類似於訓練階段，先對測試影像進行分割，計算每個區域的影像特徵，使用 PCA 分析影像特徵，將這些提供給 SVM 網路進行分類。此外，我們也將使用機率的方式描述影像中物件的高階語意，也就是說，我們將分析照片中包含哪些高階語意的物件，計算每一個高階語意物件占整張影像的機率及分析後的影像準確度。

在本論文的其他章節安排如下：在第二節中，將簡介影像分割、特徵擷取、PCA 等原理。我們使

用的 SVM 網路將在第三節中做介紹。第四節為實驗結果及討論，同時，也和傳統的半徑式基底函數類神經網路(RBF)做比較。最後，第五節為本論文的結論。

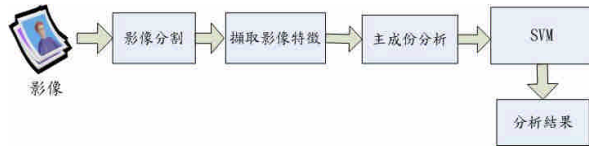


圖 1、影像語意內容分析系統流程

二、 相關理論

(一)影像分割

由於一張影像中可能包含多個物件，故影像切割所產生的區域資訊對影像分類具有相當大的影響。因此，我們使用 Deng 等學者[9]所提出的 *J*-image segmentation (JSEG) 非監督式彩色紋理切割演算法來對影像進行切割。由於要在同一時間去分析顏色以及其分佈狀況是很困難的，因此，此種方法的流程包含彩色空間量化和空間切割兩個階段，其中，空間切割是由計算數值 *J*、區域成長和區域合併三部分所組成，如此便具有比較容易實現的優點。此切割演算法的作法是將顏色從空間資訊中取出，並分成顏色以及紋理作個別處理，最後再作合併的動作。圖 2 展示影像經過 JSEG 切割演算法的範例。圖 3 展示使用 JSEG 演算法切割結果。從圖 3，我們可以知道，JSEG 切割演算法實現一個合理的切割結果，而且此結果近似於人類視覺的感知。因此，經過 JSEG 切割之後的區域，可以提供使用者進一步的語意類別指定，以方便作為後續處理之用。

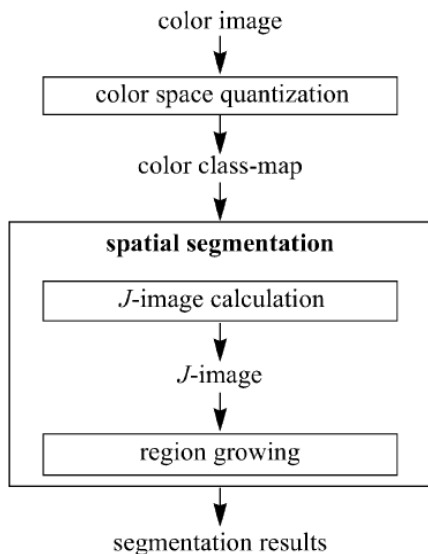


圖 2 JSEG 流程圖



(a) (b)
圖 3 使用 JSEG 演算法切割結果 (a)原始影像 (b)JSEG 切割後的結果

(二)特徵擷取

影像切割演算法後所產生的區域，對每個區域計算各種與人類視覺感知有高度相關的影像特徵(如：HSV 彩色直方圖、紋理和形狀特徵)並分析這些特徵對於影像內容分析的能力。

物件區域的顏色特徵採用 HSV 彩色空間重新量化成 64 個種類，因此，顏色特徵我們使用 64 個彩色特徵作為物件的顏色特徵。物件區域的紋理特徵使用 5 個灰階統計參數如方程式(1-5)，配合 4 個不同的角度(0、45、90 和 135 度)來表示。由於邊緣方向直方圖[10]具有位移和尺度不變的特性，可用來表示物件的形狀特徵。因此，每張影像可表示成一個由顏色特徵、紋理特徵及形狀特徵所組合而成的特徵向量(feature vector)，維度為 108。

$$\text{Contrast} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 P_{\delta}(i, j) \quad (1)$$

$$\text{Energy} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{\delta}(i, j)^2 \quad (2)$$

$$\text{Entropy} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{\delta}(i, j) \log(P_{\delta}(i, j)) \quad (3)$$

$$\text{Homogeneity} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P_{\delta}(i, j)}{1 + (i-j)^2} \quad (4)$$

$$\text{Max probability} = \text{Max}[P_{\delta}(i, j)] \quad (5)$$

其中 $P_{\delta}(i, j)$ 是在條件為 $\delta(r, \theta)$ 下的 co-occurrence matrix，而 r 代表像素距離， θ 是兩個像素 (i, j) 的角度。

(三)主成份分析

主成份分析(PCA)是一種多變量分析的方法，被廣泛地應用在影像處理、資料維度降低等。PCA 是在一個多維空間中找出一條軸線，使得所有的多維向量能夠在這條軸線上產生一個投影點，並使相鄰的投影點彼此擁有最大的變異數。降低維度的過程中，PCA 是將原始資料透過線性組合產生一組較原始資料少的新變數，這個新變數就被稱為主成份。

假設每張影像擷取的影像特徵有 N 個，形成原

始特徵向量 $Y = \{y_1, y_2, \dots, y_N\}$ ，首先計算原始特徵向量的共變異矩陣 C_x ：

$$C_x = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})^T \quad (6)$$

其中， \bar{y} 為 Y 的平均向量。

接著計算共變異矩陣的特徵值(Eigenvalue)，並依其數值大小排序為： $\lambda_1, \lambda_2, \dots, \lambda_N$ ，並求出相對應的單位特徵向量(Eigenvector)： E_1, E_2, \dots, E_N 。由於重要的資訊集中在較大特徵值所對應之特徵向量中，利用方程式(7)可以得到第 k 個主成份 x_k ：

$$x_k = E_k^T Y \quad k=1,2,3,\dots,N \quad (7)$$

主成份的個數是透過主成份貢獻度 P 來決定，其數學式如下：

$$P = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_M}{\lambda_1 + \lambda_2 + \dots + \lambda_N} \quad \text{其中 } N > M \quad (8)$$

原始的影響特徵參數經過 PCA 轉換後，得到新的特徵參數 $\{x_1, x_2, \dots, x_M\}$ 。這組特徵參數除了維度降低之外，其全域散佈矩陣也是最大，如此便可以讓我們用較少維度的特徵來表示每張影像，並當作 SVM 網路的輸入。

三、支援向量機

支援向量機是屬於監督式學習的學習機器，利用最小風險函數來逼近實際函數，並透過統計學習理論找到 VC 維(VC dimension)的底限並控制機器的推廣能力，圖 4 為 SVM 的示意圖。

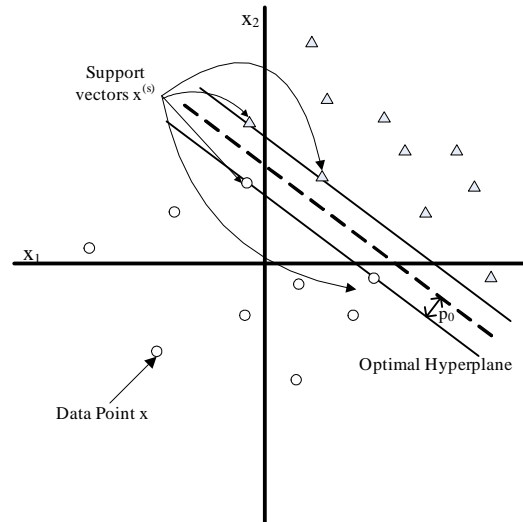


圖 4、SVM 示意圖

(一)最小化經驗風險函數

經由 PCA 所得到的輸入空間再經由核函數(kernel function) $\Phi(\mathbf{w})$ 將輸入空間映射至高維度空間，在高維度空間中就可將原本線性不可分割的資

料使用線性的方式一分為二，並藉由最小化經驗風險(empirical risk)函數來達成結構風險函數最小化(structural risk minimization, SRM)，經驗風險函數如下：

$$R_{emp} = \frac{1}{N} \sum_{i=1}^N |d_i - f(\mathbf{x}_i, \mathbf{w})| \quad (9)$$

其中 N 是訓練樣本數， d_i 為目標輸出， f 為預測函數。

(二)最佳化超平面

最佳化超平面的鑑別函數(discriminant function)如下式所示。

$$g(\mathbf{x}) = \mathbf{w}_o^T \mathbf{x} + b_o \quad (10)$$

其中 \mathbf{w}_o 及 b_o 分別為最佳化的權重向量及偏移值。

利用最佳化超平面進行分類之高維空間的超平面必須符合下列兩個限制條件：

$$\begin{cases} d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 & \text{for } i=1,2,\dots,N \\ \Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \end{cases} \quad (11)$$

此具有限制條件的最佳化問題稱為原始問題(primal problem)，我們藉由 Lagrangian 函數，可將問題表示成：

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i [d_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \quad (12)$$

其中 α_i 為 Lagrangian multipliers，此具限制條件最佳化問題的解於 Lagrangian 函數 $J(\mathbf{w}, b, \alpha)$ 的鞍點(saddle point)，得下列兩個最佳化條件

$$\mathbf{w} = \sum_{i=1}^N \alpha_i d_i \mathbf{x}_i \quad (13)$$

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (14)$$

由上面的條件，我們可以將 primal form 重新表示成 dual form 如下：

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j \quad (15)$$

此目標函數(objective function)的解有下列兩個限制

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (16)$$

$$\alpha_i \geq 0 \quad \text{for } i=1,2,\dots,N$$

我們可解得最佳化權重向量 \mathbf{w}_o 及偏移值 b 如下：

$$\mathbf{w}_o = \sum_{i=1}^N \alpha_{o,i} d_i \mathbf{x}_i \quad (17)$$

$$b_o = 1 - \mathbf{w}_o^T \mathbf{x}^{(s)} \quad \text{for } d^{(s)} = 1 \quad (18)$$






我們亦可將此最佳化超平面線性可分的問題轉移到線性不可分的資料樣本上，只需加上鬆弛項即可達成。

$$Accuracy = 1 - \left(\frac{\sum_{i=1}^m A_i}{S} \right) \quad (20)$$

四、 實驗結果及討論

在照片分析實驗中，我們選擇 Corel 影像資料庫中的 257 張影像作為 SVM 網路訓練影像，並且定義 8 種語意的類別[3]：建築物、雲、天空或水、草地、樹木、花、山和日出，8 種語意的類別影像數量分別為 52 張、38 張、32 張、39 張、29 張、32 張、32 張及 38 張。8 種語義類別與其區塊影像的顏色對照如表 1 所示。

表 1. 使用在區塊影像的顏色對照表

語意類別	顏色區塊
建築物	
雲	
花	
草地	
山	
天空/水	
日出	
樹木	

實驗中，每個物件特徵表示成一個由 64 個顏色特徵、20 個紋理特徵及 24 形狀特徵所組成維度 108 的特徵向量(feature vector)。然後，將每類特徵個別透過 PCA 運算，取出主成份貢獻度 $P=0.95$ 的主成份。最後我們得到一個維度為 26 的新特徵向量(feature vector)用以表示每個物件，並當作 SVM 網路的輸入，其中每個物件包含 12 個顏色特徵、11 個形狀特徵及 3 個紋理特徵。

照片分析考慮區域歸屬的語意類別以及面積的大小，因此，對於第 j 類語意類別，其分析機率定義如下：

$$P_j = \sum_{i=0}^k \frac{W_i^j}{S} \quad (19)$$

其中 S 代表影像的大小， k 為影像分割後的區域數目， W_i^j 是第 j 個語意類別的第 i 個區域的區域大小。

從照片分析中，可以得到每張照片所包含的每個類別機率，但是，若從影像分類的角度來看，光分析每個類別的機率是不夠直覺的。因此，我們進一步計算每張照片分析後的分類準確度，用來判斷系統的效能。準確度的定義如下：

其中， S 為影像的面積大小， A_i 為被誤判的區域面積， m 為被誤判的區域個數。

圖 5 和表 2 展示本系統以語意為基礎的照片分析結果，(a)欄為原始影像，(b)欄為 JSEG 切割後的影像，(c)欄為結合 PCA 與 RBF 的分析結果，(d)欄則為本論文提出之結合 PCA 與 SVM 法的分析結果。實驗結果中，圖 5(A-B)拍攝的主題都是山，觀察圖 5(Ac-Bc)發現，在 RBF 法的分析結果中，影像中有些屬於山的區域被誤分為雲和樹木，使得圖 5B 的整體辨識率僅有 45.8%。反觀，本論文提出的結合 PCA 與 SVM 法的分析結果如圖 5(Ad 及 Bd)所示，大部分的區域都被正確的辨識，山的類別機率分別為 57.38% 及 61.05%。雖然圖 5(Ad)有些區域被誤分為雲，但該區域的山被雲霧所遮蔽，即便是肉眼也很難正確的分辨，因此，分類的結果頗符合人眼視覺感受。圖 5(C)為庭院影像，圖 5(Cc)顯示 RBF 法的分類結果，其中有大片的草地被誤分為建築物及山，造成整體辨識率僅有 46.38%。圖 5(Cd)顯示 SVM 法的分類結果，其中天空/水有 24.12% 的機率、樹木有 15.05% 的機率、建築物和草地各有 29.52% 的機率和 31.31% 的機率，整體辨識率高達 99.76%。圖 5(D)是一張風景影像，影像中包含天空/水、草地、山及雲，在 SVM 法的分析的結果中，可以完全正確的分成四個類別，天空/水的類別機率為 37.16%，雲的類別機率為 22.68%，山和草地的類別機率分別為 12.31% 及 27.86%，整體辨識率為 100%。圖 5(E)為日出影像，在使用 RBF 法的分析結果中發現，整張影像被錯分為天空或水，而 SVM 法則可正確的將它分為日出。圖 5(F)是一張森林影像，在這張影像中，我們可以發現，不論是 RBF 或 SVM 法都只有少部分的小區域分類錯誤，而這些小區域通常是人眼所忽略的區域。另外，從表 2 中可知，本論文提出的結合 PCA 與 SVM 法在圖 5 中各影像的分類準確度介於 93.30%~100% 之間，其中，(A)、(D)和(E)的準確度為 100%。反觀以 PCA 與 RBF 法的分析結果的分類準確度介於 0%~92.5% 之間，起伏劇烈，代表 RBF 法的穩定性欠佳。從圖 5 和表 2 的分析比較結果可知，除了圖 5(F)的分析結果較接近外，其餘的部分都有蠻大的差異。因此，本系統的分析結果確實比結合 PCA 與 RBF 的表現佳，而且，本系統的平均準確度相當高，也相當符合人眼視覺感官的認知。由上述的實驗分析結果可知，即使是影像的分類結果仍存在些微的錯誤，但這些錯誤都是肉眼難以分辨的區域，因此，分析的結果都相當接近人類視覺的判斷。

五、 結論

本論文提出的影像語意內容分析系統，利用

SVM網路結合低階的影像特徵分析影像的內容，計算影像中各種高階語意項目的成分，達到影像語意內容分析的目的。我們採用以物件為基礎的特徵擷取方式，而不是整張影像的特徵，並進一步利用PCA分析影像特徵，選取有效及具有鑑別度的影像特徵，降低影像特徵維度。從實驗結果顯示，SVM網路分析的結果確實相當接近人類視覺的判斷。而且PCA確實可大幅減少資料維度，進而加速SVM網路的訓練，及提升分類能力。

致謝

本論文的研究經費由國科會提供，NSC95-2221-E-224-059。

參考文獻

- [1] C. Faloutsos, R. Barder, M. Flickner, J. Hafner, W. Niblack, D. Petkovic and W. Equitz, "Efficient and effective querying by image content," *J. Intell. Inform. Systems*, vol. 3, pp. 231-262, 1994.
- [2] A. Pentland, R.W. Picard and S. Sclaroff, "Photobook: content-based manipulation of image database," *SPIE Storage and retrieval for image and video database II*, vol. 2185, pp. 34-47, 1994.
- [3] Yixin Chen, James Z. Wang and Robert Krovetz, "CLUE: Cluster-based retrieval of image by unsupervised learning," *IEEE Trans. Image Processing*, vol. 14, pp. 1187-1201, August 2005.
- [4] Y. Liu, D. Zhang, G. Lu and W.Y. Ma, "Region-based image retrieval with high-level semantic color names," *IEEE Proc. Multimedia Modeling Conference*, pp. 180-187, 2005.
- [5] Jianping Fan, Yuli Gao, Hangzai Luo and Guangyou Xu, "Statistical modeling and conceptualization of natural images," *Pattern Recognition*, Vol.38, pp. 865-885, June 2005.
- [6] C.C. Chang, and C.J. Lin, "LIBSVM: a library for support vector Machines" Soft ware available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [7] J. Latin, D. Carroll and P.E. Green, *Analyzing Multivariate Data*, Duxbury, 2003.
- [8] A.C. Rencher, *Multivariate Statistical Inference and Applications*, John Wiley and Sons, 1998.
- [9] Y. Deng, B.S. Manjunath and H. Shin, "Unsupervised segmentation of color-texture regions in image and video," *IEEE Tran. Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800-810, Aug. 2001.
- [10] A.K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognit.*, vol. 29, pp. 1233-1244, 1996.

表 2 語意為基礎的照片分析結果(單位：%)

	圖(A)		圖(B)		圖(C)		圖(D)		圖(E)		圖(F)	
	RBF	SVM	RBF	SVM	RBF	SVM	RBF	SVM	RBF	SVM	RBF	SVM
建築物	0.00	0.00	5.88	0.00	24.79	29.52	0.00	0.00	0.00	0.00	2.06	1.79
雲	0.00	7.09	0.00	7.66	0.00	0.00	0.00	22.68	0.00	0.00	0.00	2.28
花	0.00	0.00	0.00	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
草地	0.00	0.00	46.31	0.00	1.55	31.31	27.86	27.86	0.00	0.00	2.39	2.05
山	54.34	57.38	1.99	61.05	6.05	0.00	0.00	12.31	0.00	0.00	0.85	0.00
天空或水	42.62	35.53	31.28	31.28	24.12	24.12	37.79	37.16	100.0	0.00	1.28	2.06
日出	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.18	0.39
樹木	3.04	0.00	14.55	0.00	43.50	15.05	34.36	0.00	0.00	0.00	93.24	91.43
準確度	91.37	100.00	45.80	93.78	46.38	99.76	67.97	100.00	0.00	100.00	92.50	93.30

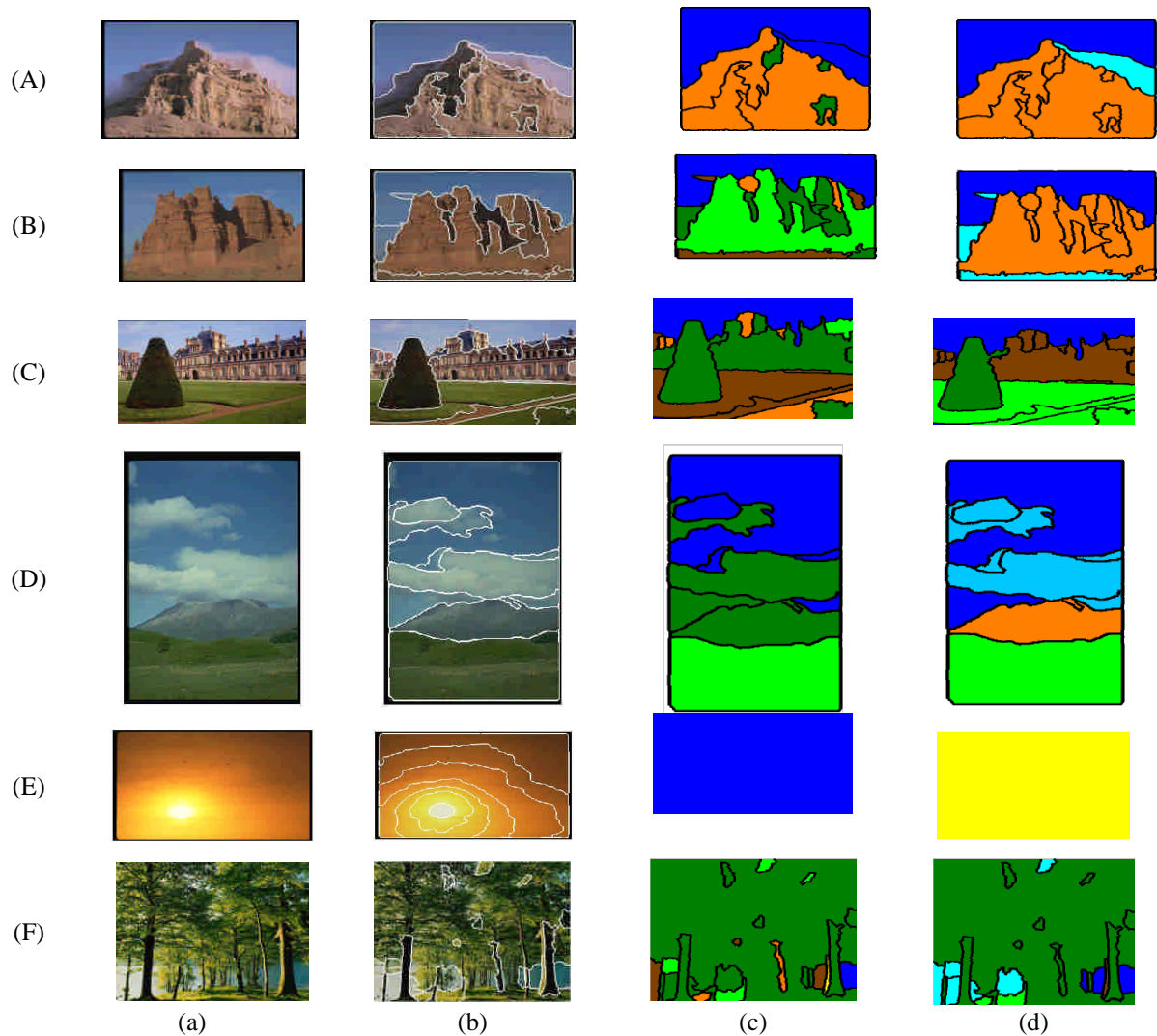


圖 5 語意為基礎的照片分析結果 (a)原始影像，(b)切割後的圖片，(c) 使用 PCA+RBF 分類後所對應的區塊影像，(d)使用 PCA+SVM 分類後所對應的區塊影像。