

國立交通大學

資訊科學與工程研究所

碩士論文

影像中建築物偵測之研究

The Study of Building Detection in Image

研究生：陳昭翰

指導教授：傅心家 教授

中華民國九十七年七月

影像中建築物偵測之研究
The Study of Building Detection in Image

研究生：陳昭翰

Student : chao-han chen

指導教授：傅心家

Advisor : Prof. Hsin-Chia Fu

國立交通大學
資訊科學與工程研究所
碩士論文



Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年七月

影像中建築物偵測之研究

研究生:陳昭翰

指導教授:傅心家 教授

國立交通大學資訊科學與工程研究所

摘要

影像中偵測建築物在影像語意分析上是一個重要的任務。影像語意分析能夠利用一些低階的特徵例如：顏色、紋理、形狀...等來表示，在影像檢索與影像瀏覽上是一項有用的技術，而建築物就是其中一種重要的語意。本篇論文中提出一個以紋理特徵為基礎的建築物偵測方式。方法首先會使用較小區塊中的紋理特徵找出影像中有可能為建築物的區域，再利用較大範圍區域中之紋理特徵來判斷是否為建築物，來找出建築物的區域，研究並延伸至建築物影像分類與特定類型建築物偵測。我以科立爾影像資料庫中的建築物影像來做實驗，測試在本方法下建築物區域偵測結果的召回率與正確率，召回率方面最高能達到 75.70%，而正確率可以最高能達到 80.32%，結果顯示紋理特徵在建築物偵測有不錯的效能。

The Study of Building Detection in Image

Student: chao-han chen

Advisor: Prof. Hsin-Chia Fu

Institute of Computer Science and Engineering
National Chiao Tung University

Abstract

Building detection in images is an important task in image semantic analysis. The image semantic analysis can be inferred from lower-level features such as color, texture, shape, etc. which is a very useful technique for the image retrieval and browsing. And building is one kind of the important image semantic. In this paper, we propose a method for building detection based on texture feature. First, we used the texture feature in smaller blocks to find the candidate regions, and then used texture feature in bigger range of regions to determine the correct building regions. The study also extend to building image classification and specific category building detection. The experiments are conducted on the Corel image database. We tested the region's recall and precision under our method. The best recall rate is 75.70% and the best precision rate is 80.32%. The experimental results show that texture feature is efficient in the detection of building.

誌謝

謝謝傅老師在我研究所兩年的生涯給予我的指導和照顧，並幫助我的論文找到研究方向，學習到做研究的方法與態度，才得以完成此篇論文。同時，感謝實驗室博士後研究以及博士班學長，永煜、柏伸、政龍、岳宏、士賢，還有同學逸凡、佳蓁，學弟坤隆、威人平常在生活上及學業上的建議與指教。感謝大學同學以及朋友在生活上的鼓勵。最後，感謝爸爸、媽媽、姊姊、妹妹一直在背後支持我，給我無憂無慮的生活，讓我可以專注在學業上，才得以順利完成學業。



目錄

摘要.....	i
Abstract.....	ii
誌謝.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	vii
第一章 緒論.....	1
1.1 研究背景.....	1
1.2 研究動機.....	1
1.3 研究目的.....	2
1.4 章節介紹.....	2
第二章 建築物偵測之相關研究探討.....	3
第三章 建築物區域偵測與建築物影像分類.....	5
3.1 建築物區域之偵測.....	6
3.1.1 建築物區塊偵測.....	7
3.1.2 建築物區域偵測之方法.....	12
3.2 建築物影像分類.....	23
3.3 特定類型建築物偵測.....	24
第四章 實驗結果.....	28
4.1 建築物區域偵測實驗結果與分析.....	28
4.2 建築物影像分類實驗結果與分析.....	32
4.3 特定類型建築物偵測實驗結果與分析.....	35
第五章 結論與未來展望.....	36
5.1 結論.....	36

5.2 未來展望.....	36
參考文獻.....	37



表目錄

表 3.1	不同尺度的遮罩視窗大小.....	16
表 4.1	不重疊區塊之區域偵測召回率與精確率.....	29
表 4.2	重疊區塊之區域偵測召回率與精確率.....	30
表 4.3	不重疊區塊之建築物影像分類結果.....	32
表 4.4	重疊區塊之建築物影像分類結果.....	33
表 4.5	特定類型建築物偵測之召回率與精確率.....	35



圖目錄

圖 3.1	建築物區域偵測流程圖.....	6
圖 3.2	建築物區塊偵測.....	7
圖 3.3	取得訓練建築物區塊模型之資料.....	8
圖 3.4	支持向量機之概念圖.....	10
圖 3.5	核心函式資料投影.....	11
圖 3.6	建築物區域偵測.....	12
圖 3.7	JSEG 彩色影像切割流程.....	13
圖 3.8	J值與點分佈的關係，+，o，* 代表不同色彩等級的點.....	15
圖 3.9	JSEG 視窗遮罩.....	15
圖 3.10	篩選出建築物區域.....	18
圖 3.11	建築物區域偵測.....	19
圖 3.12	建築物區域偵測結果(一).....	20
圖 3.13	建築物區域偵測結果(二).....	21
圖 3.14	建築物區域偵測結果(三).....	22
圖 3.15	區域中之建築物區塊.....	23
圖 3.16	都市方形小窗戶建築物之訓練資料與偵測結果.....	24
圖 3.17	古羅馬式建築物(柱子)之訓練資料與偵測結果.....	25
圖 3.18	哥德式建築物之訓練資料與偵測結果.....	26
圖 3.19	都市建築物之訓練資料與偵測結果.....	27
圖 4.1	不重疊區塊之區域偵測召回率與精確率曲線圖.....	29
圖 4.2	重疊區塊之區域偵測召回率與精確率曲線圖.....	30
圖 4.3	區塊重疊與不重疊下召回率之比較曲線圖.....	31
圖 4.4	區塊重疊與不重疊下精確率之比較曲線圖.....	31

圖 4.5 不重疊區塊之建築物影像分類結果曲線圖..... 32

圖 4.6 重疊區塊之建築物影像分類結果曲線圖..... 33

圖 4.7 區塊重疊與不重疊下影像分類結果(建築物)之比較曲線圖..... 34

圖 4.8 區塊重疊與不重疊下影像分類結果(非建築物)之比較曲線圖..... 34



第一章

緒論

1.1 研究背景

目標物偵測(object detection)是影像分析中一個重要的問題，用於將影像中有興趣的物體偵測出來，並進一步對影像做分析，目前常見的目標物有汽車、人臉、貓、狗、飛機、建築物…等物體，建築物的偵測是近幾年一項熱門的領域，常應用於影像檢索，衛星空照圖分析…等應用，建築物的偵測大致上分為對航照圖和對地面拍攝影像作偵測這兩種，前者目前已有蠻多系統完成，例如：Carnegie Mellon University 的 build system[6][7][8]以及 Southern California ASCENDER2[9]這些系統大多採用屋頂存在角點與直線統計等方法分析的假設，採用邊緣檢測及形狀的分析，而後者通常用於影像檢索，場景分析，例如：[2]利用多尺度隨機場的方法在自然影像中進行偵測，[3]對視覺上明顯的建築物作偵測，將天空與建築物分隔後再對所選出之候選區域進行偵測，[5]對建築物作偵測，然後進行場景分析，建築物的偵測通常有兩個重要問題，第一為如何將建築物與背景區分，第二為如何描繪出所偵測的建築物。

1.2 研究動機

目標物偵測這項技術目前已行之有年，不過目前較成熟的技術大概只有人臉偵測，行人偵測，汽車偵測…等，也廣泛的運用在許多應用上，而建築物的偵測在這個領域也有其重要性，例如應用於影像檢索系統上，建築物就是一個相當重要的物件，若能有效的在影像中找到建築物，則進一步可以取出這張影像的語義，使用者對感興趣影像的取得會更加的方便。

1.3 研究目的

目標物偵測的目的就是從影像中找出感到興趣的物件，若能夠利用機器自動對大量的影像作偵測，這對於許多應用上都有相當大的幫助，例如圖片的搜尋系統，若使用者希望能夠找到存在建築物的圖片，利用這項技術先將圖片中建築物找出並紀錄下此圖為建築物的圖片，則使用者能更方便的找到想要的圖片，然而建築物的偵測受限於不同風格的建築物和拍攝角度，大小比例…等往往不容易找到共同的特徵，且建築物可能會受到遮蔽物和複雜背景的影響，使得偵測的難度提高，另外影像本身的解析度也可能影響偵測的準確率，所以本研究希望能找到一個穩定且運算量小的方法，有效的找出影像中的建築物使得使用者能更方便且有效的運用存在影像中的資訊。

1.4 章節介紹

在以下章節中，第二章首先介紹建築物偵測與其相關研究，並討論不同方法的優缺點；第三章介紹本論文如何以紋理特徵為基礎來偵測建築物，並且將研究延伸至建築物影像分類與特定建築物偵測；第四章是以第三章所提出建築物偵測的方法來驗證其效能；第五章是結論及對未來的展望。

第二章

建築物偵測之相關研究探討

目前建築物偵測的應用，通常應用於影像檢索、場景分析…等，也有一些相關的論文，而大致上分成兩類方法：

第一類是先從影像中找出可能的候選區域，然後在由這些候選區域中去偵測出哪些為建築物而哪些為非建築物，例如[4]先利用小波轉換的技術，對水平邊緣與垂直邊緣的區域作強化，進而將天空與建築物作區隔，在由建築物與天空的對比，找出人眼覺得明顯的區域作為候選區域，最後使用支持向量機分類器來判斷這些候選區域是否為建築物，建築物模型是以 SOBEL 運算子對建築物取出邊緣的特徵，再利用支持向量機訓練出建築物的模型；其優點是能利用小波的特性快速的將建築物突顯出來，不過卻只能對顯著的建築物進行偵測，且 SOBEL 運算子對於建築物的分類結果不是太好。

而第二類是由整張圖來偵測，利用一些建築物之邊緣、角、顯著點的特徵將建築物的區域給強化出來，例如[3]利用相鄰區域對比、顯著點、邊緣出現的密度、邊緣並排的出現這些特徵來找出建築物所在的區域，並對每種特徵訂門檻值，當特徵值超過門檻值，則認定某區域為建築物，此方法是對建築物的面做偵測，所以比較不會有拍攝角度不同的問題，但執行速度相對較慢，且門檻值的決定以及是否認定為建築物的條件也是一個問題。

第一種類型的方法比較常被使用，因為他在處理速度上相對比較快，而且候選區域中建築物與非建築物之間的差異較容易掌控，所需的特徵也相對較容易選出，而第二種類型的方法通常需要使用到較多且較複雜的特徵，對於影像上解析度的要求也相對的較為嚴格。

而在影像索引(image retrieval)以及影像瀏覽(image browsing)方面也有相關的研究主題，希望能夠利用一些低階的特徵如顏色(color)，形狀(shape)，紋理(texture)……等，再一堆影像中找到符合自己語義觀念(semantic concept)的影像，例如[18]分別利用了紋理和邊緣的特徵來將影像的語意取出，判斷是否有建築物的存在，紋理特徵是取局部(local)的特徵，而邊緣特徵則是取總體(global)的特徵，紋理的作法為首先將整個影像切割為3x2或2x3的區塊，然後取出區塊中紋理的特徵，接著利用支持向量機(SVM)分類器來判斷每個區塊是否為建築物區塊，而判斷是否為建築物影像則使用了一個簡單的投票機制，當被判定為建築物區塊的數量比非建築物區塊多時則此影像被認定為建築物影像，而邊緣特徵的則是取整張影像的邊緣直方圖(histogram)，總共使用了72個bin來表示，然後再利用支持向量機(SVM)分類器來判斷是否為建築物影像，不過此方法對於建築物區域較小的影像較難判斷其為建築物影像。



第三章

建築物區域之偵測與建築物影像分類

本章敘述從影像中找出建築物區域並將建築物影像分類出來的方法，方法首先將影像中所有可能為部分建築物的小區塊找出，然後再將相似的區塊合併為一候選區域，最後從這些候選區域中篩選出為建築物的區域，接著可利用上述 2 階段的結果來判斷此影像是否為建築物影像。於本章 3.1 節介紹如何從影像找出建築物區域；3.2 節介紹如何判斷影像是否為建築物影像。



3.1 建築物區域之偵測

本節敘述如何在影像中找出建築物的區域，圖 3-1 是建築物區域偵測流程圖。在 3.1.1 小節介紹建築物區塊的偵測，在 3.1.2 小節介紹建築物區域之偵測方法。

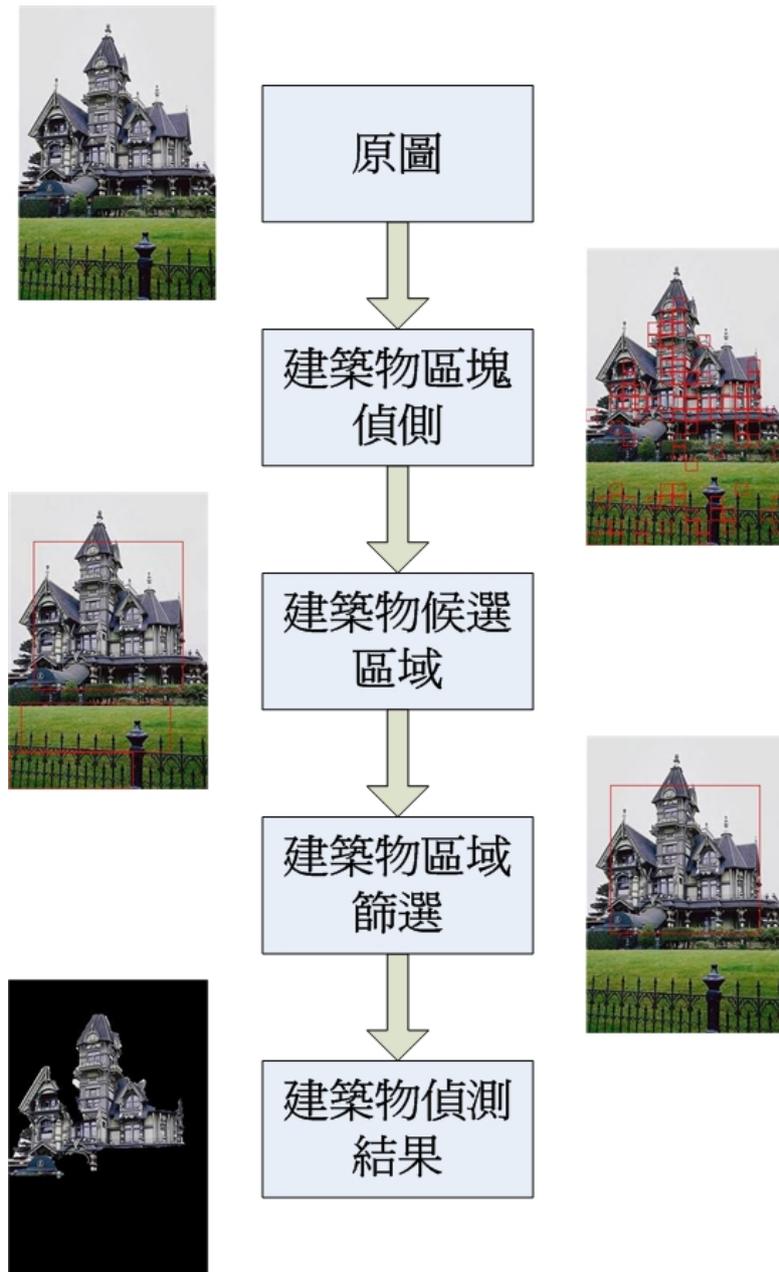


圖 3.1 建築物區域偵測流程圖

3.1.1 建築物區塊偵測

首先將原圖切成相同尺寸的區塊，每塊區塊以紋理(texture)為特徵取出 48 維的特徵向量，並以事先由支持向量機(support vector machine) 訓練的建築物區塊模型來判斷哪些區塊為建築物區塊。

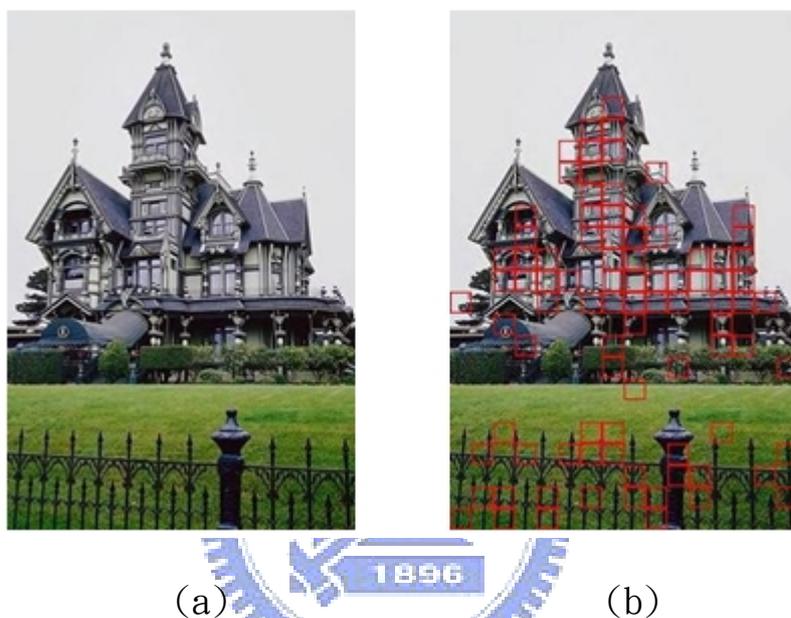


圖 3.2 建築物區塊偵測(a)原影像，(b)建築物區塊偵測結果。

建立建築物區塊模型方法及步驟：

步驟 1. 取得訓練建築物區塊模型所需資料

選擇一些不同類型的建築物影像，接著以人工從影像中切出建築物與非建築物的區域，再將這些區域切成大小相同的區塊，用來作為訓練建築物區塊模型所需之資料。

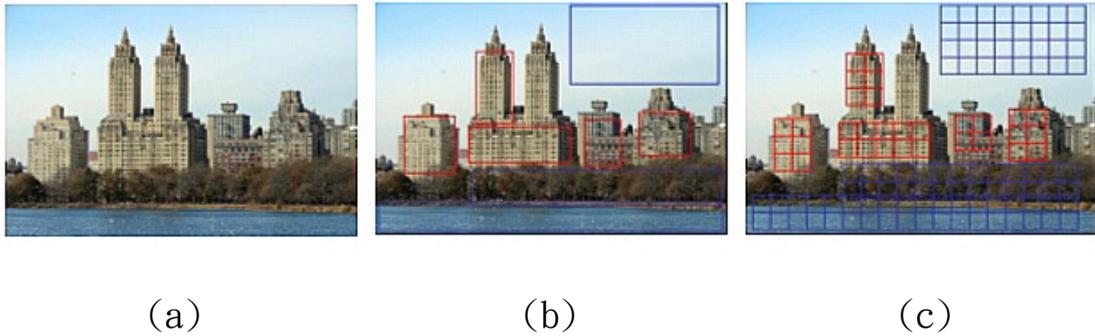


圖 3.3 取得訓練建築物區塊模型之資料 (a)建築物影像，(b)人工畫出之建築物與非建築物區域，(c)切割出訓練建築物區塊模型所需資料。

步驟 2. 取出判斷為建築物或非建築物區塊之特徵

特徵的選擇是非常重要的，所選擇的特徵必須是能夠很有效的分辨不同的類別資料，本研究中我選擇以紋理(texture)來作為判斷是否為建築物的特徵，紋理指的是具有齊次(Homogeneous)或非齊次的圖案，是否具有齊次性則由影像中的色彩與影像亮度來決定。紋理包含了物件表面重要的結構資訊，以及他們和當時環境的關係，利用適當的紋理描述子(Texture Descriptor)來描述影像中的紋理特徵在特徵的擷取上有很大的幫助。

我所使用的紋理特徵擷取方法為賈柏過濾器(Gabor filter)，賈柏紋理(Gabor texture)已經被用在許多內容語意(semantic content)與分類(classification)方面的研究[14][15]，賈柏過濾器(Gabor filter)已被證明為空間(space)域與頻率(frequency)域間轉換的理想方法，並且有考慮到方向(orientation)與尺度(scale)的因素。

二維的賈柏函數(Gabor function)定義如下：

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi_j W_x\right] \quad (1)$$

$g(x, y)$ 為母賈柏小波函數(mother Gabor wavelet)，藉由適當的擴大(dilations)和旋轉(rotations)動作可得到一組濾波器組(filter bank)，其生成函數如下：

$$\begin{aligned}
g_{mn}(x, y) &= a^{-m} g(x', y') \\
x &= a^{-m} (x \cos \theta + y \sin \theta) \\
y &= a^{-m} (-x \sin \theta + y \cos \theta)
\end{aligned} \tag{2}$$

其中 $a > 1$, $\theta = n\pi/K$, $n=0, 1, \dots, K-1$, and $m=0, 1, \dots, S-1$. K 和 S 為所選擇的方向(orientations)和大小(scales)的個數。

此濾波器的詳細設計策略敘述可參考[16]。

影像 $I(x, y)$ 的賈柏轉換(Gabor transform)定義如下:

$$W_{mn}(x, y) = \int I(x, y) g_{mn}^*(x - x_1, y - y_1) dx_1 dy_1 \tag{3}$$

其中*表示完全共軛(complete conjugate)。

而紋理特徵(texture feature)的取法為取 $W_{mn}(x, y)$ 所有係數之均值

(mean) μ_{mn} 與標準差(std.) σ_{mn} ，例如本實驗是以6種方向(orientation)與4種尺度(scale)組合之濾波器組來實作，則所取的賈柏紋理特徵表示為:

$$f = [\mu_0, \sigma_0, \mu_1, \sigma_1, \dots, \mu_{23}, \sigma_{23}] \tag{4}$$

特徵相量(feature vectors)之維度為48維。

步驟 3. 是否為建築物區塊之判斷

而是否為建築物之判斷，使用的是廣為大家所使用的分類器支持向量機(support vector machine)，支持向量機這種學習的技術是由V. Vapnik於1985年提出，SVM不像一般傳統學習技術以訓練資料的分類誤差最小為目標，而是使測試資料的分類誤差在一個機率上界以下，此機器學習技術能夠將錯誤率的上限最小化。支持向量機在圖形分類(pattern classification)問題上不需要加入領域知識，且擁有不錯的效能。

支持向量機(SVM)介紹:

支持向量機的概念就是從一群 R^d 空間的資料中找出一個超平面 (hyper-plane) 來將這群資料分成2大群, 同一群的資料則會在超平面的同一側, 並且希望資料離此平面越遠越好, 如此才能更明確的分辨出某點屬於哪一群, 否則計算上容易因為精確度問題而產生誤差, 如下圖3.4所示,

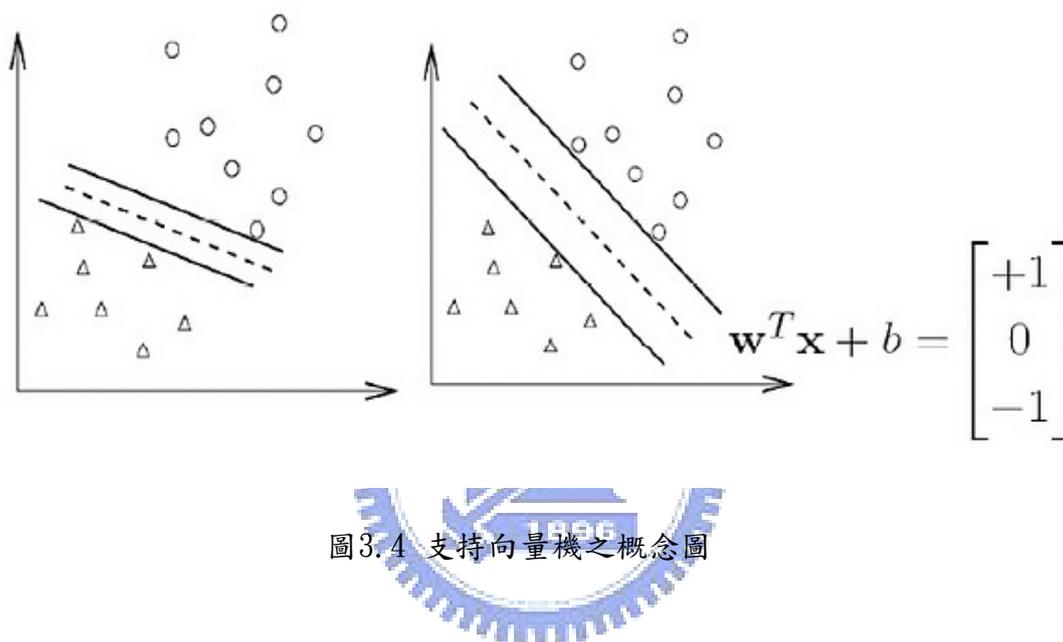


圖3.4 支持向量機之概念圖

右圖的邊界(margin)比起左圖來的大, 所以我們認為右圖的結果較好, 接下來以數學式來描述以上問題, 假設有一堆點集合 $\{x_i, y_i\}, i=1 \dots n$ 且 $x_i \in R^d, y_i \in \{+1, -1\}$, 我們希望找到一條直線 $f(x) = w^T x - b$, 使的 $y_i = -1$ 的點落在 $f(x) < 0$ 這一邊, 而 $y_i = +1$ 的點則落在 $f(x) > 0$ 這邊, 所以我們要解出下式之 w, b

$$y_i (w^T x_i - b) - 1 \geq 0, i = 1 \dots n \quad (5)$$

邊界(margin)大小為 $2/\|W\|$, 所以 $\|W\|$ 越小則邊界越大, 所以有了下面目標函式:

$$\text{最小化 } \frac{1}{2} \|W\|^2, \quad y_i (w^T x_i - b) - 1 \geq 0, i = 1 \dots n \quad (6)$$

我們利用Lagrange Multiplier Method將上面式子轉為 L :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1] \quad (7)$$

我们去最大化上式，使用二次規方法(quadratic programming method)來解決，中間的解法省略，最後可以得到此超平面的判斷式(decision function)如下：

$$f(x) = \text{sgn}\left(\sum_{i=1}^n \alpha_i^0 y_i x_i \cdot x + b^0\right) \quad (8)$$

其中x為class +1 與 class -1的支持向量(support vector)。

但如果兩個類別是無法線性分割(non-linear separable)的話，我們則必須透過核心函式(kernel function)將資料投影到更高維度的空間，如下圖所示，

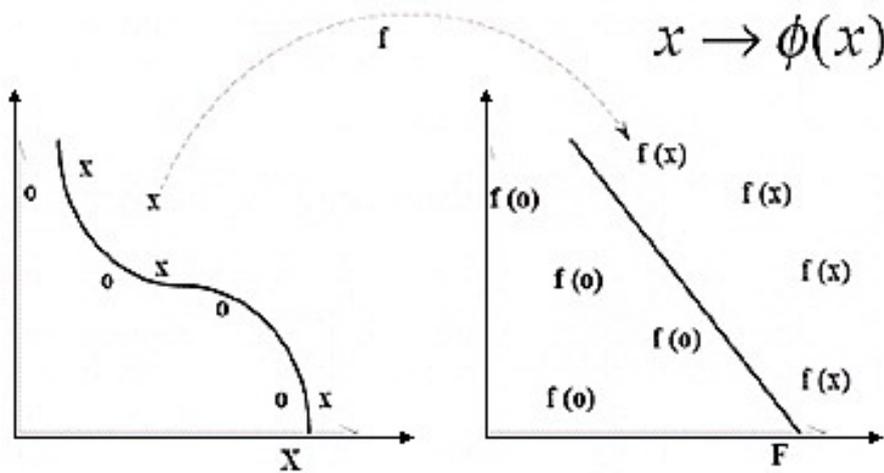


圖3.5 將資料投影到更高維度的空間，使資料可線性分割。

而 ϕ 即核心函式，有三種常見的核心函數如下：

$$\text{Polynomial: } (x \cdot y + 1)^d \quad (9)$$

$$\text{Gaussian RBF: } \exp\left(-\frac{1}{2\sigma^2} \|x - y\|^2\right) \quad (10)$$

$$\text{Sigmoid: } \tanh(k(x \cdot y) - \mu) \quad (11)$$

而到底要如何選擇核心函式來使用其實沒有特定的方法，通常就是經由實驗來選擇結果最好的那組核心函式來使用。

3.1.2 建築物區域偵測之方法

步驟 1. 找出建築物的候選區域

首先將影像中偵測出來的區塊依據顏色、紋理特徵來分群，將相似的區塊分為同一群，並刪除較小的群，再以最小矩形將同一群中的區塊框為一區域，則這些區域即建築物的候選區域，而分群的方法是使用影像切割技術 JSEG[12]先對影像做切割，而位於相同切割區域(segmentation region)的區塊就視為同一群，動作如圖 3.6。

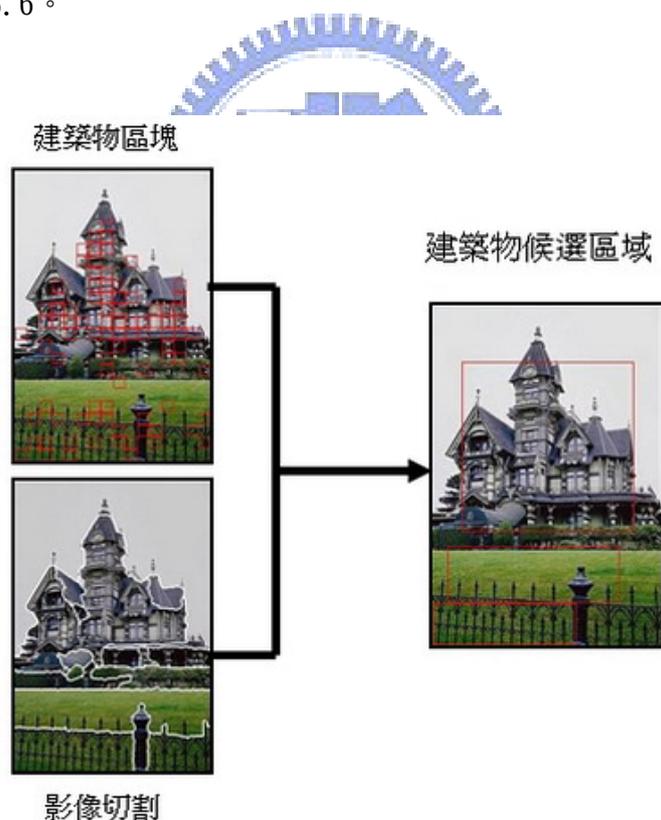


圖3.6 建築物候選區域之選取

JSEG影像切割技術介紹：

JSEG[12]為一種區域擴張與融合的影像切割方法，利用顏色與紋理為特徵來進行分割，大致上分成2個主要的階段色彩空間量化(color space quantization)與空間切割(spatial segmentation)，方法概要流程圖如下圖3.7

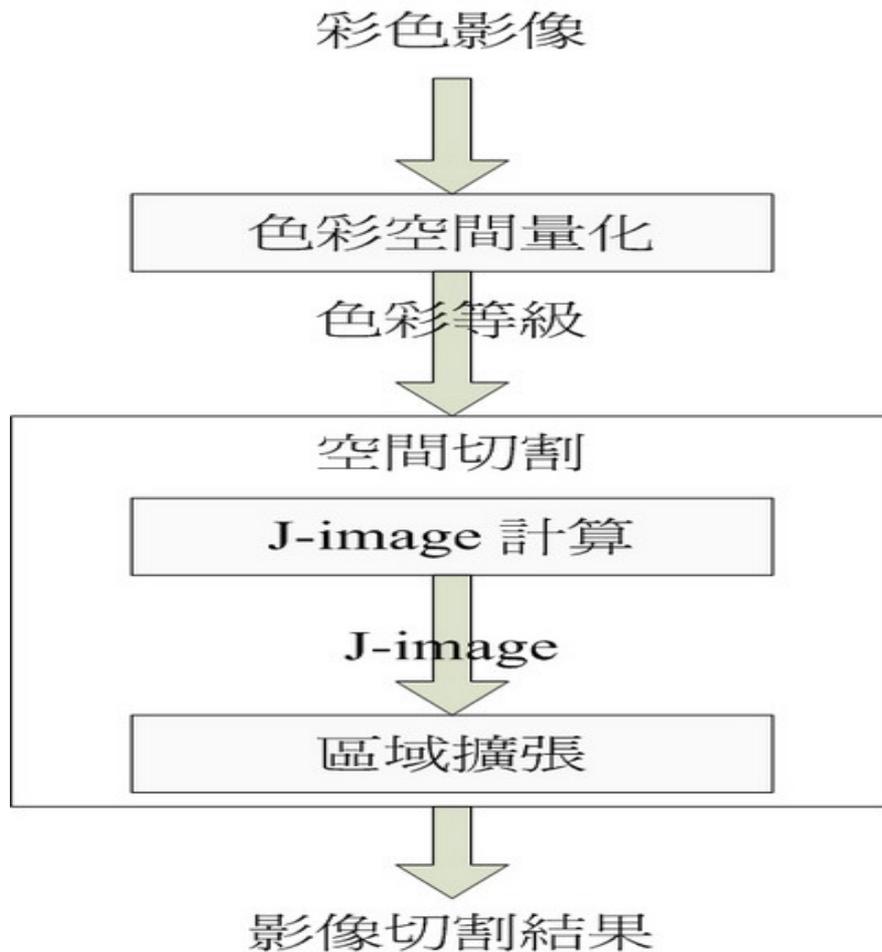


圖3.7 JSEG彩色影像切割流程圖

1. 色彩空間量化

首先將顏色量化成不同大小的色彩等級，而量化後的值將被影像區域所使用，因此影像中每個像素(pixel)的值將被量化的結果所取代，量化的步驟如下：

- (1) 利用一個非線性的演算法 Peer Grouping filter[17]，將影響模糊並達到去雜訊的效果。

(2) 經過 Peer Grouping filtering 後，每個像素值具有一個權重值，越平滑的區域有越高的權重值。

(3) 使用一種變形的 General Lloyd Algorithm(GLA)[t5]的量化方法，所使用的色彩空間為 CIE LUV 色彩空間，其顏色差異距離公式定義如下：

$$D = \sum_i D_i = \sum_i \sum_n v(n) \|x(n) - c_i\|^2, x(n) \in C_i \quad (12)$$

其中 c_i 為群集 C_i 之中心點， $x(n)$, $v(n)$ 為像素 n 的值與權重，而 D_i 為群集 C_i 中的差異性，而中心點的更新方法如下：

$$c_i = \frac{\sum v(n)x(n)}{\sum v(n)}, x(n) \in C_i \quad (13)$$

(4) 進行合併階段，當2個相似的群集其距離低於一個預設的門檻值(threshold)時，將這兩個群集合併。

最後量化後的結果稱為色彩等級分配圖(class map)，而分配圖中每個點即為原影像所對應的量化值。



2. 計算J值

計算J值部份是使用經過色彩量化的值來計算，而不會使用到原來的像素值，而我們可以把一張經過量化的影像認為是一種特殊的紋理(texture)影像來處理，我們令 Z 為色彩等級分配圖所有點之集合， $z = (x, y), z \in Z$, m 為平均值，

$$m = \frac{1}{N} \sum_{z \in Z} z \quad (14)$$

假設 Z 被分成 C 類色彩等級， $Z_i, i = 1, \dots, C$ ， Z_i 中 N_i 點的平均為 m_i ，

$$m_i = \frac{1}{N_i} \sum_{z \in Z_i} z \quad (15)$$

且令

$$S_T = \sum_{z \in Z} \|z - m\|^2 \quad (16)$$

和

$$S_w = \sum_{i=1}^C S_i = \sum_{i=1}^C \sum_{z \in Z_i} \|z - m_i\|^2 \quad (17)$$

J值的定義如下：

$$J = (S_T - S_w) / S_w \quad (18)$$

當影像中相鄰區域的像素越相近，則越容易被分割成同一區域，而J值也就會越大，如下圖3.8(a)相同類形型的點都群聚在一起，較容易分割區遇，所以J值較大，而圖3.8(b)不同色彩等級的點非常的分散，因此J值很小。

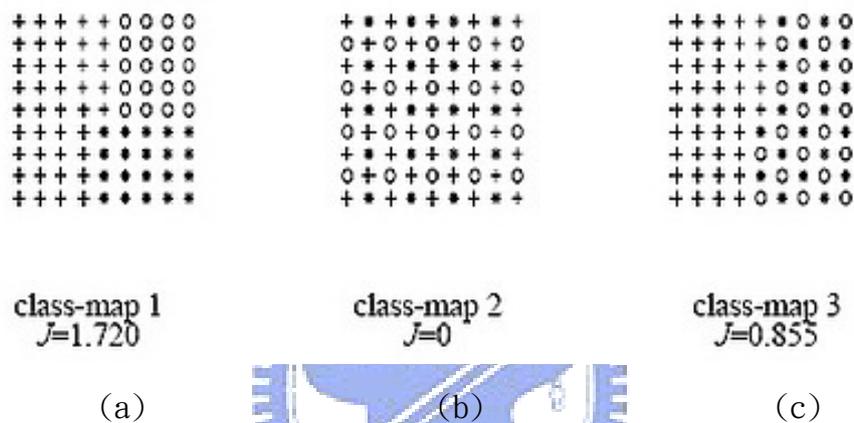


圖3.8 J值與點分佈的關係，+，o，* 代表不同色彩等級的點

取J值方法，會以一個遮罩視窗(window)對影像做處理，從中取出一個J值，使用時會先以較大的尺度開始做運算，然後再依序縮小尺度，如圖3.9(a)為一基本的遮罩視窗而圖3.9(b)為一尺度為2遮罩，其處理方式為隔一行與一列做運算。

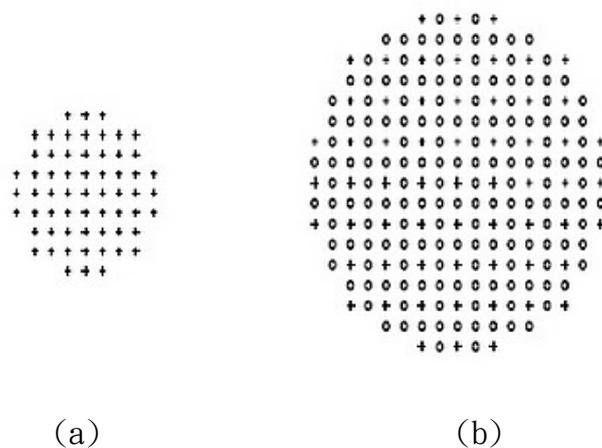


圖3.9 (a)基本視窗遮罩(b)尺度為2之視窗遮罩，只有標記+的點有做運算。

3. 區域成長(region growing)

區域成長包含種子決定與種子成長兩個階段，首先在影像中決定種子的分布與個數，然後再經由種子的成長來產生新的區域，而種子的個數會影響最後分割區域的個數。

決定種子步驟：

- (1) 計算每一個區域的J值，並計算所有J值的平均值 μ_j 與標準差 σ_j 。
- (2) 設定一門檻值 T_j ， $T_j = \mu_j + a\sigma_j$ ，其中a為一預設值，且此值會影響種子的個數。J值小於此門檻值的點會被當做候選的種子，最後使用4-連通(4-connectivity)的方式將做連接來取得候選種子的區域。
- (3) 如果候選種子區域大於表3.1中最小種子個數，則此點被視為種子。

表3.1 不同尺度的遮罩視窗大小

尺度	視窗	取樣 (1/像素)	區域大小 (像素)	最小種子 (像素)
1	9x9	1/(1x1)	64x64	32
2	17x17	1/(2x2)	128x128	128
3	33x33	1/(4x4)	256x256	512
4	65x65	1/(8x8)	512x512	2048

區域成長步驟：

- (1) 移除種子中的洞。
- (2) 計算其他未切割區域中的點其J值，並算出平均，將J值小於平均的點連接組成一區域，如果此區域鄰近某一個種子點，則將此區域合併至種子的區域。
- (3) 重複步驟(2)，使用更小的視窗尺度做運算直到做到尺度1。
- (4) 持續作完所有尺度的區域成長後，將剩下的點儲存到緩衝區，每次從中取出最小J值的點並將他分配給最鄰近的種子，直到所有點都被分配完為止。

4. 區域合併(region merging)

經過區域成長後，會產生一個初步的影像切割，此時的切割會有太過破碎的情況產生，所以要進一步將這些破碎的區域做合併，我們將顏色較相近的區域合併，每個區域以顏色直方圖(histogram)來表示顏色特徵，並計算區域間的相似性，計算距離的公式如下：

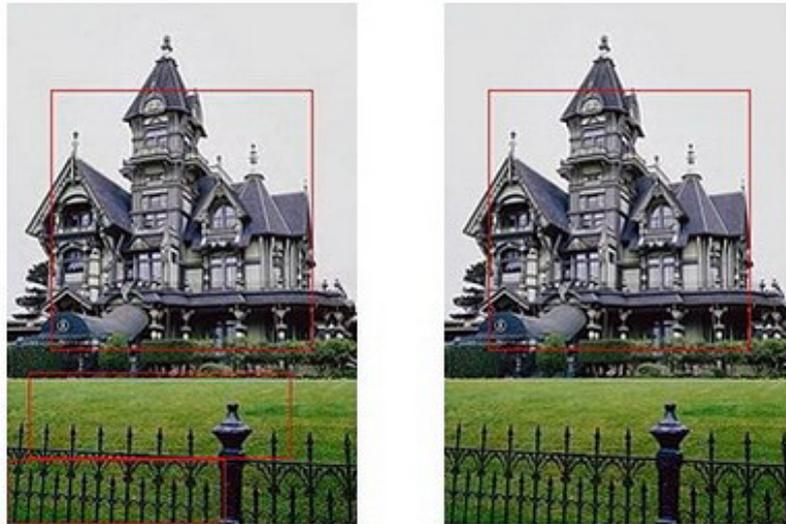
$$D(i, j) = \|H_i - H_j\| \quad (19)$$

其中H表示顏色直方圖的向量，而所採用的色彩空間為 CIE LUV。

合併區域前，先將任兩個區域的色彩直方圖距離全部算出，然後從最小直方圖距離的區域開始合併，合併後會再重新計算距離，在從更新的距離中取出最小距離的區域來合併，直到所有區域之間的距離都大於門檻值為止，合併後的結果即最後影像切割的結果。

步驟2. 由候選區域中篩選出建築物區域

首先將候選區域影像縮放到固定的大小，然後利用事先以支持向量機(support vector machine)所訓練的建築物區域模型來判斷哪些候選區域為建築物區域，建築物區域模型所選用的特徵與建築物區塊模型一樣為賈柏紋理特徵，測試發現利用此特徵來分辨建築物與非建築物有不錯的效果，如圖3.10(b)最上面那塊候選區域被認定為建築物區域。



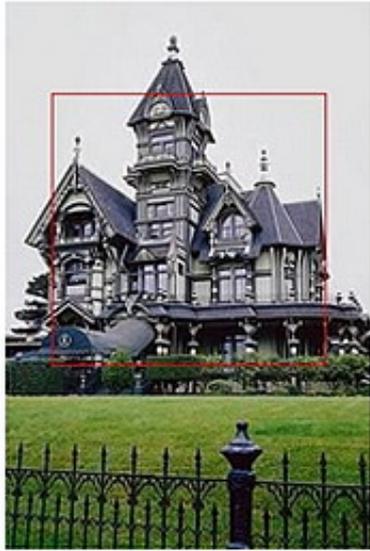
(a)

(b)

圖3.10 篩選出建築物區域(a)建築物候選區域，(b)建築物區域篩選結果

步驟3. 建築物區域偵測結果

因為在做影像切割有時會將一棟建築物切成2個以上的區域，因此最後會有2個以上的方塊來表示同一棟偵測出之建築物，視覺上不適於表達偵測結果，因此我想以更貼近建築物輪廓的方式來表示偵測的結果，利用剛剛影像切割的結果，選出建築物區域裡擁有最多建築物區塊的切割區域(segmentation region)來表示所偵測出的建築物，如圖3.11(b)即最後偵測結果表示圖。



(a)



(b)

圖3.11 建築物區域偵測(a)建築物區域，(b)建築物偵測結果

下圖列出一些不同類型建築物區域偵測的結果，左邊為原影像而右邊為偵測的結果。



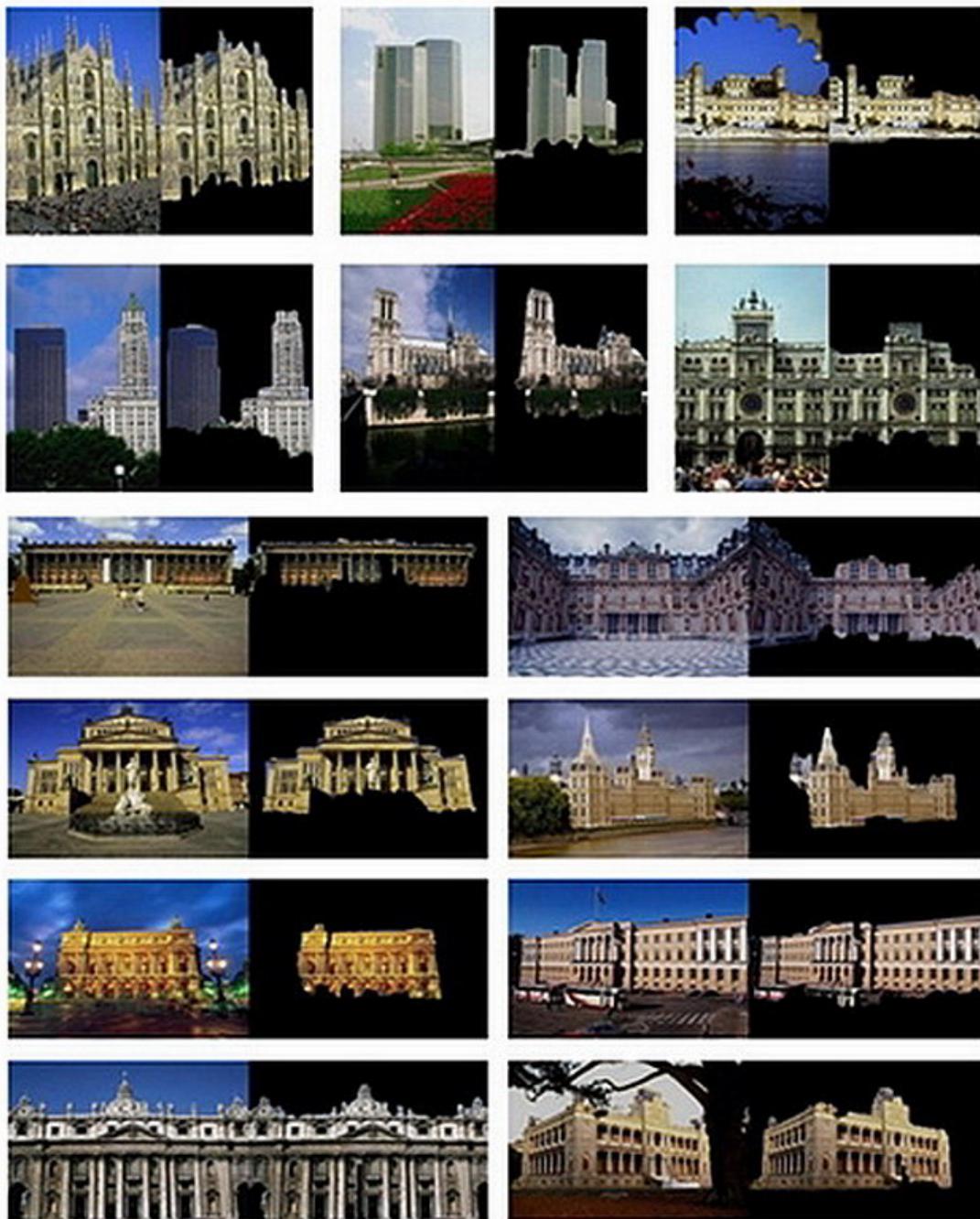


圖 3.12 建築物區域偵測結果(一)

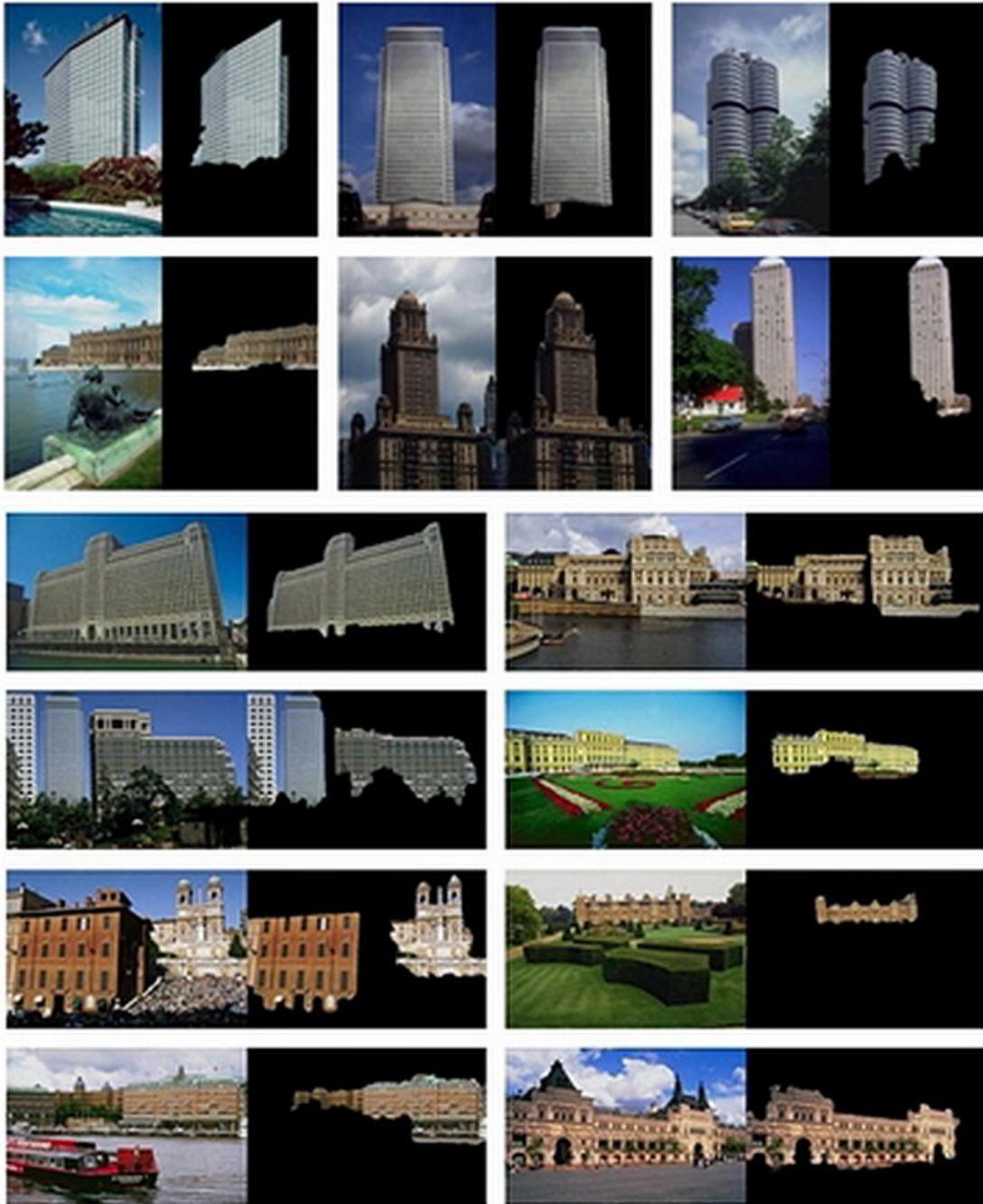


圖 3.13 建築物區域偵測結果(二)

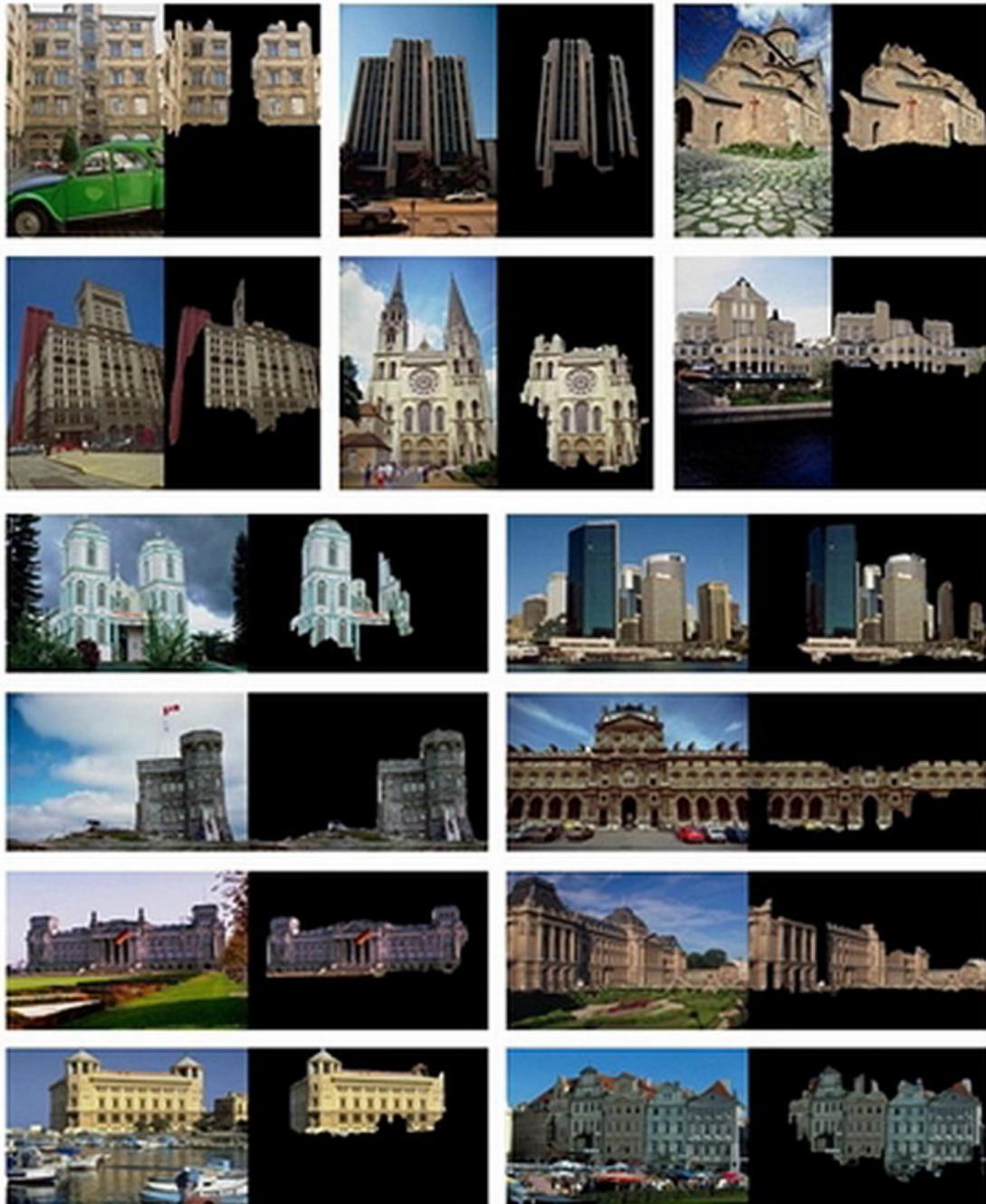


圖 3.14 建築物區域偵測結果(三)

3.2 建築物影像分類

本節我將介紹如何從建築物偵測動作中判斷此影像是否為建築物影像之方法，利用此方法可從一堆影像中有效的挑出建築物影像，方法首先會使用到建築物偵測的中間結果，由建築物偵測的第 2 步驟可將所有候選區域分成建築物區域與非建築物區域 2 種，我們來計算這 2 種區域裡所包含的建築物區塊各別有多少，然後比較哪種區域所含的建築物區塊較多，建築物區域擁有較多的區塊則判定此影像為建築物影像，反之則判定為非建築物區域，例如下圖 3.15 中建築物區域有 56 個區塊而非建築物區域有 24 個區塊，所以此圖被認定為建築物影像。

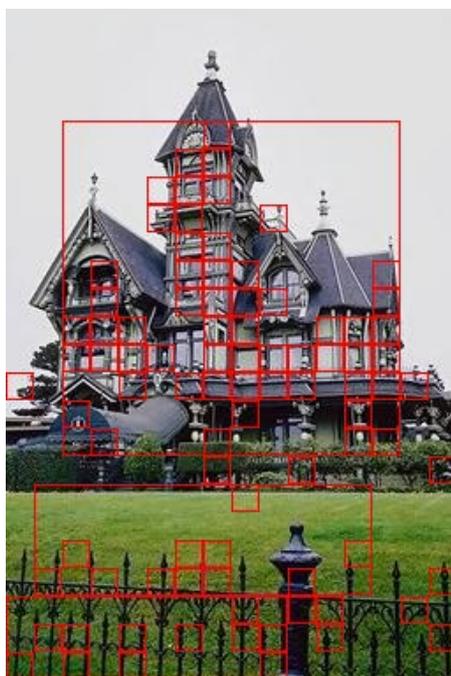


圖 3.15 區域中之建築物區塊

此方法的想法是認為在建築物影像中通常會有較多的建築物區塊被辨出，若建築物區塊有絕佳的偵測結果，則區塊數量差異應該是非常明顯的，而在非建築物影像中即使有建築物區塊被偵測出，也會根據偵測第 2 步驟的結果將他判定為非建築物所屬區塊，所以此方法之正確性與建築物區塊及區域偵測結果有非常緊密之關聯。

3.3 特定類型建築物偵測

本節我要針對特定類型的建築物來偵測，這個動作能更精準的取出影像中的語意，讓影像有更明確的分類，我的作法仍是使用紋理這個特徵來分類，作法為在訓練特定建築物模型時，支持向量機中的正面資料為特定建築物的區塊，而反面資料為所有其他類型建築物的區塊，執行建築物區塊偵測時只會在之前所取出的建築物區域中偵測，當執行區塊偵測完畢後，計算所偵測出來的區塊數佔所有建築物區域中區塊數之比例，當大過一定比例就認定為此種類型建築物的影像，我找了幾種類型建築物做測試，如下：

1. 都市方形小窗戶的建築物

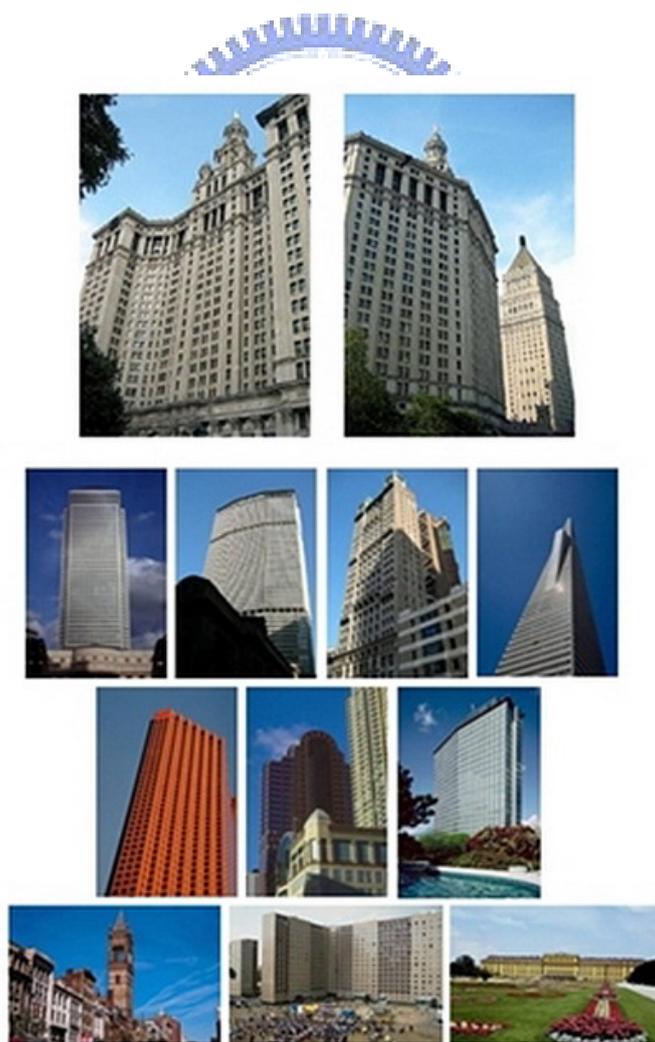


圖 3.16 都市方形小窗戶建築物（上）訓練資料（下）偵測結果

2. 古羅馬式建築物(柱子)



圖 3.17 古羅馬式建築物(柱子) (上) 訓練資料 (下) 偵測結果

3. 哥德式建築物



圖 3.18 哥德式建築物 (上) 訓練資料 (下) 偵測結果

4. 都市建築物



圖 3.19 都市建築物（上）訓練資料（下）偵測結果

第四章

實驗結果

在這章中我針對第三章所提的方法，設計實驗來評估此方法的在偵測建築物區域與建築物分類的效能。對於實驗的平台，在硬體方面使用了 Intel Pentium-4 1.8Ghz core dual 中央處理器的個人電腦，搭配 1 Gigabytes 主記憶體，作業系統為 Microsoft Windows XP 專業版，程式的執行使用 MATLAB 7.1。

4.1 建築物區域偵測實驗結果與分析

本節我將測試在不同大小的區塊以及區塊是否有重疊的情況下對偵測結果的影響與本方法之偵測效能，我以偵測結果與正確結果(ground-truth)之間重疊面積的多寡來做為偵測效能之依據，來估計召回率(recall rate)與正確率(precision rate)，召回率與正確率之算法如下：

$$\text{召回率} = \frac{\text{偵測結果與正確結果之重疊面積(像素)}}{\text{正確結果之面積(像素)}}$$

$$\text{正確率} = \frac{\text{偵測結果與正確結果之重疊面積(像素)}}{\text{偵測結果之面積(像素)}}$$

實驗資料來源為 COREL image database，取 200 張訓練資料，測試資料取 550 張建築物影像與 1000 張非建築物影像來做測試，而影像大小為 384x256 或 256x384。

首先以沒有重疊的區塊來測試，區塊大小由 8x8 到 32x32 共 6 種，

表 4.1 不重疊區塊之區域偵測召回率與正確率

區塊大小	區域召回率	區域正確率	時間 (秒/張)
8x8	67.90%	76.86%	62
12x12	70.73%	79.78%	35
16x16	73.99%	80.32%	24
20x20	68.32%	77.66%	18
24x24	64.82%	75.50%	15
32x32	61.68%	76.75%	12

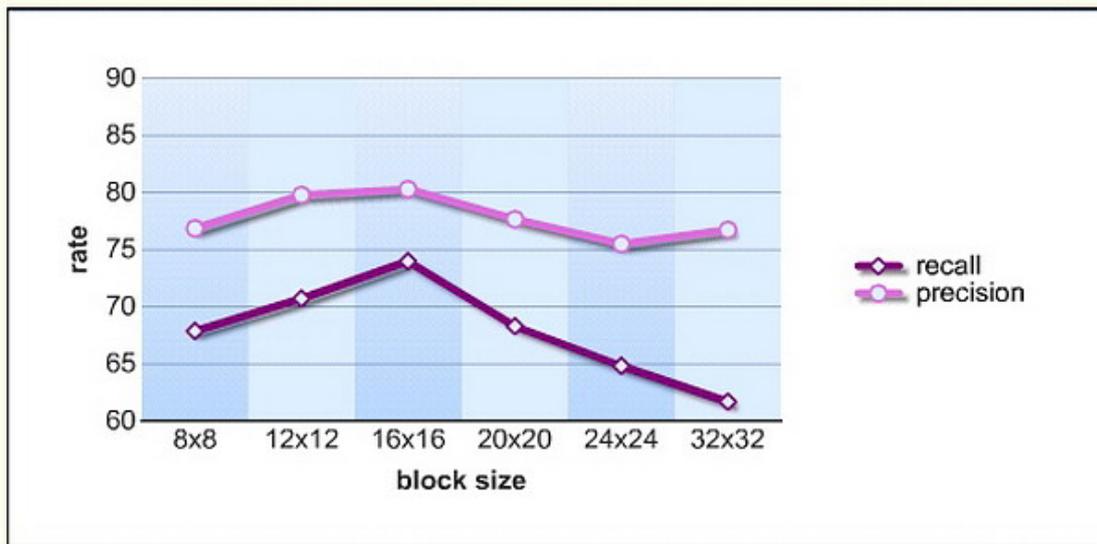


圖 4.1 不重疊區塊之區域偵測召回率與正確率曲線圖

接下來使用重疊(50%)的區塊來測試，區塊大小由 8x8 到 64x64 共 10 種，

表 4.2 重疊區塊之區域偵測召回率與正確率

區塊大小	區域召回率	區域正確率	時間 (秒/張)
8x8	70.38%	79.04%	173
12x12	71.56%	79.43%	80
16x16	75.75%	78.88%	51
20x20	74.77%	77.72%	37
24x24	75.70%	77.24%	28
32x32	74.75%	77.60%	22
40x40	70.90%	74.19%	17
48x48	67.27%	72.48%	15
56x56	64.11%	73.37%	14
64x64	64.25%	74.21%	12

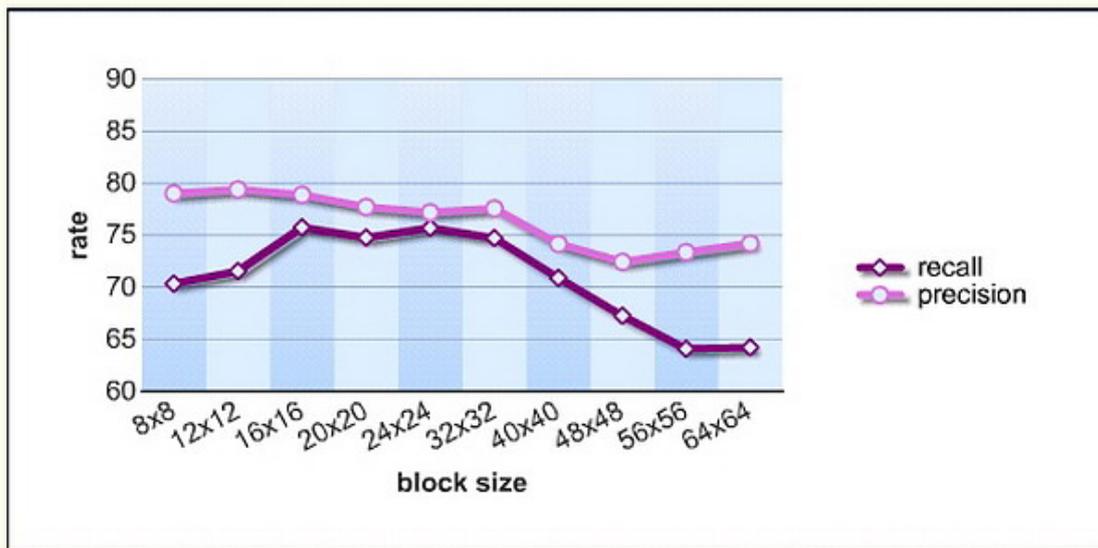


圖 4.2 重疊區塊之區域偵測召回率與正確率曲線圖

實驗結果比較與分析：

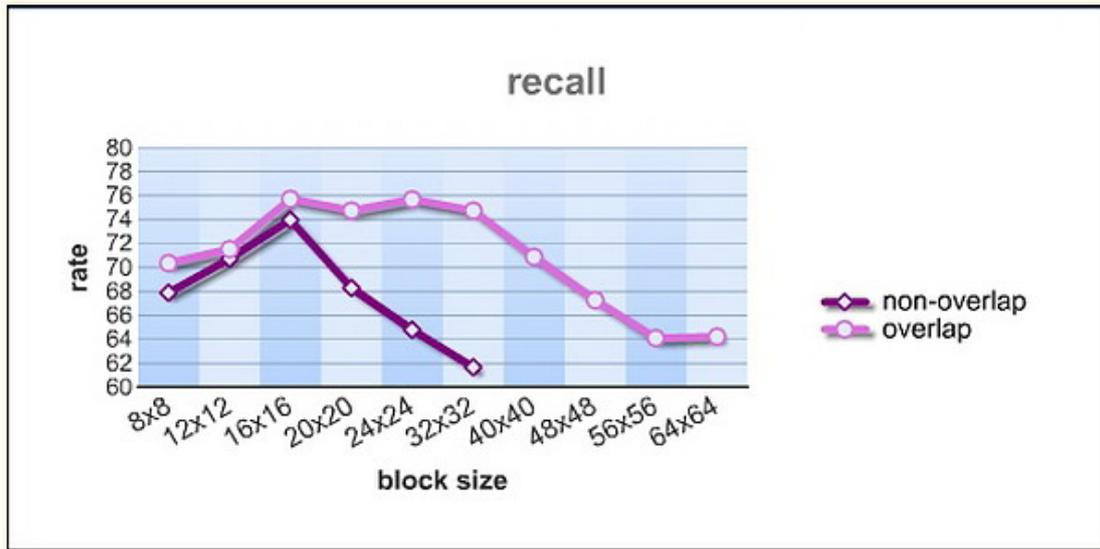


圖 4.3 區塊重疊與不重疊下召回率之比較曲線圖

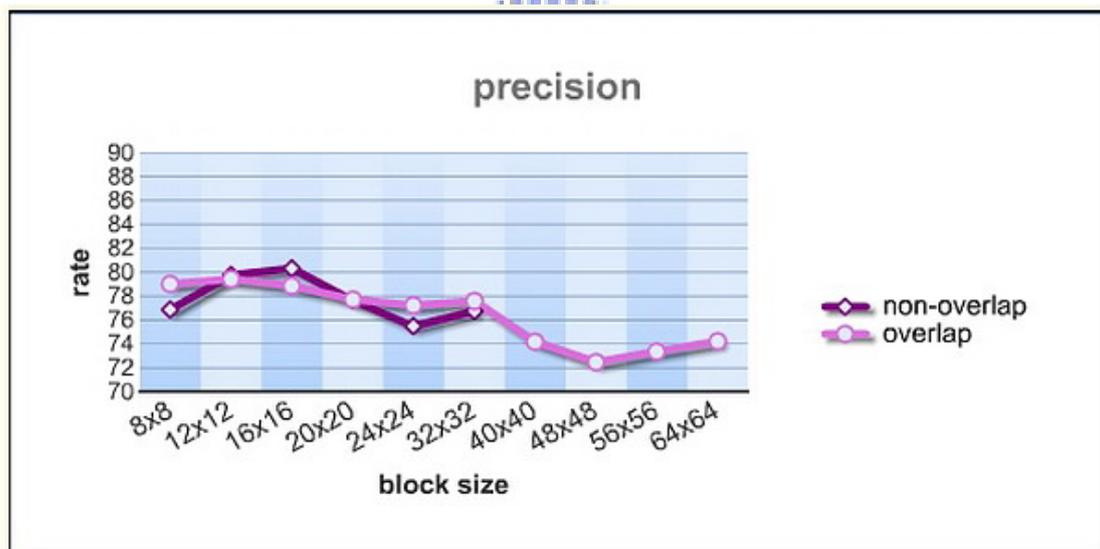


圖 4.4 區塊重疊與不重疊下正確率之比較曲線圖

由上圖來看，召回率在區塊重疊下明顯比不重疊要好，且區塊大小 16x16 到 32x32 之間有較佳的結果，另外區塊較大其結果也明顯下降，而正確率方面區塊重疊下最大值比起不重疊的最大值稍微的低，但曲線較穩定直到區塊大小超過 32x32 才有明顯的下降。原因在於區塊太小建築物的結構特徵較不明顯，而太大則會抓到太多非建築的區域而使結果變差了。

4.2 建築物影像分類實驗結果與分析

本節我要來測試 3.2 節介紹的建築物影像分類方法，同樣還是測試在不同大小的區塊以及區塊是否有重疊的情況下的分類結果之準確率 (accuracy rate)，實驗使用 corel image 550 張建築物影像以及 1000 張非建築物影像來測試，結果如下：

首先以沒有重疊的區塊來測試，區塊大小由 8x8 到 32x32 共 6 種，

表 4.3 不重疊區塊之建築物影像分類結果

區塊大小	分類 (建築物影像)	分類 (非建築物影像)
8x8	80.07%	96.60%
12x12	81.35%	96.20%
16x16	82.44%	96.60%
20x20	78.98%	95.70%
24x24	79.98%	96.30%
32x32	77.69%	96.10%

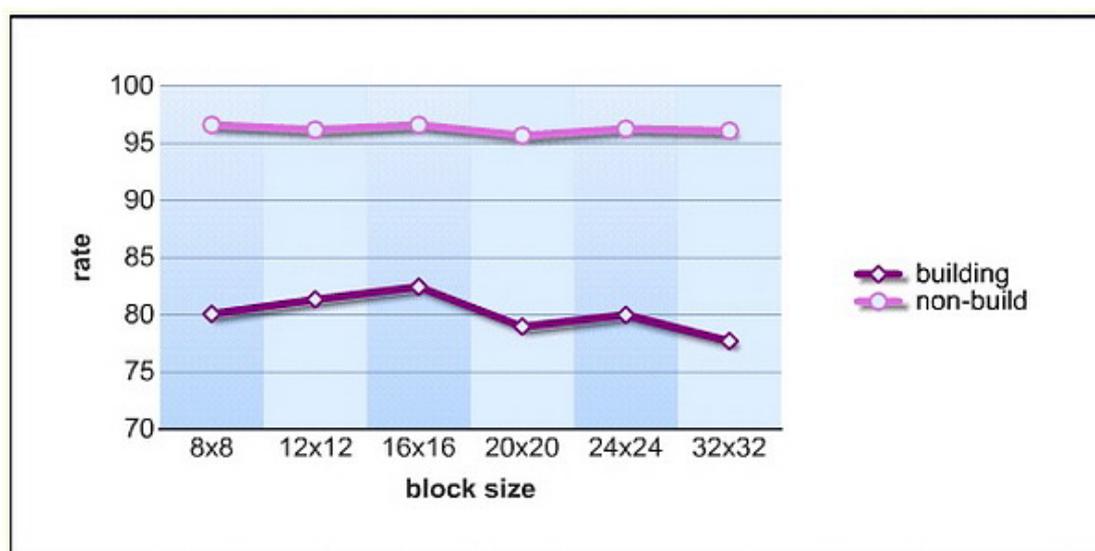


圖 4.5 不重疊區塊之建築物影像分類結果曲線圖

接下來使用重疊(50%)的區塊來測試，區塊大小由 8x8 到 64x64 共 10 種，

表 4.4 重疊區塊之建築物影像分類結果

區塊大小	分類 (建築物影像)	分類 (非建築物影像)
8x8	82.99%	97.10%
12x12	82.26%	97.70%
16x16	82.63%	97.50%
20x20	83.54%	97.50%
24x24	83.54%	97.20%
32x32	83.54%	97.00%
40x40	81.90%	96.40%
48x48	79.89%	96.30%
56x56	81.53%	96.20%
64x64	80.99%	96.40%

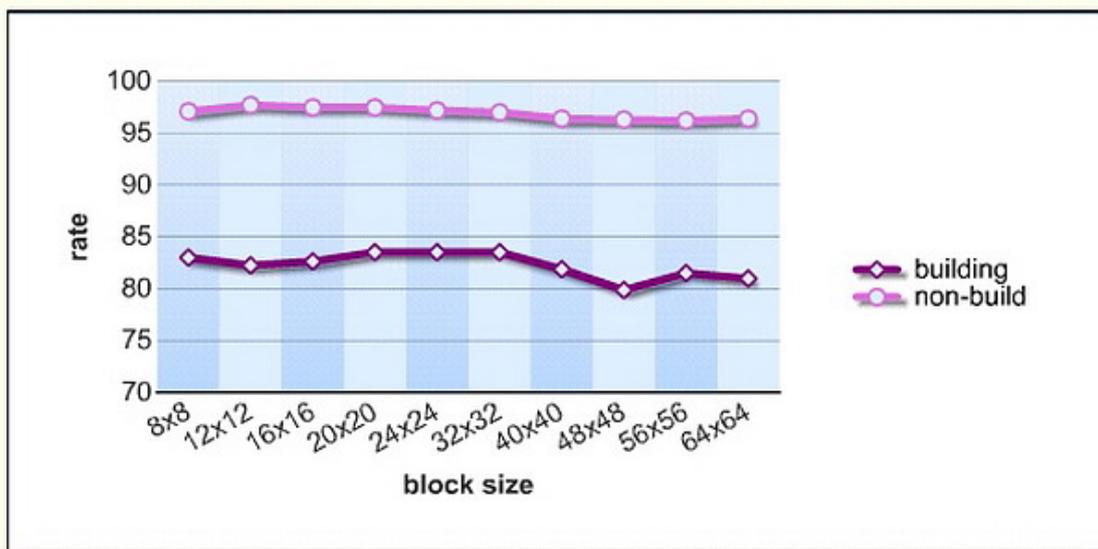


圖 4-6 重疊區塊之建築物影像分類結果曲線圖

實驗結果比較與分析：



圖 4.7 區塊重疊與不重疊下影像分類結果(建築物)之比較曲線圖

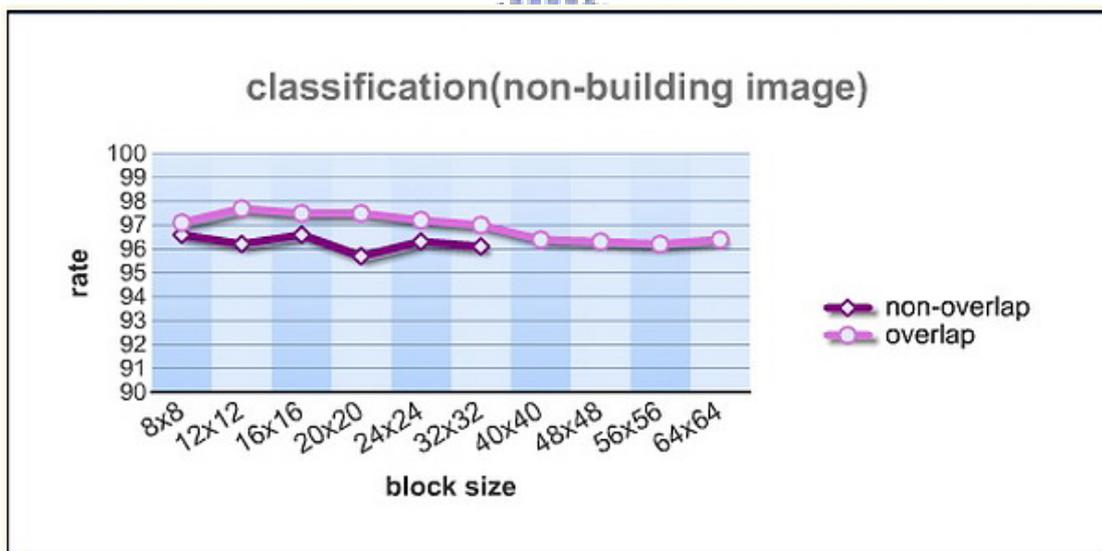


圖 4.8 區塊重疊與不重疊下影像分類結果(非建築物)之比較曲線圖

由上圖來看，區塊重疊確實在建築物影像分類上有較好的結果，原因在於影像上相同位置會被偵測 2 次，因此建築物區塊被辨識出來的機會增高了，且偵測出的區塊也更貼近建築物的輪廓，而在不同區塊大小偵測下非建築物影像的判斷正確率差異不大且精確，而建築物影像則是在區塊大小 16x16~32x32 重疊與 16x16 不重疊下結果比較好，而為什麼非建築物影像的結果會比建築物影像的結

果要好，原因是區域的分類器在將非建築物區域判斷為非建築物區域下比將建築物區域判斷為建築物區域來的準確。

4.3 特定類型建築物偵測實驗結果與分析

本節我要來測試 3.3 節介紹的特定建築物偵測方法，使用大小為 32x32 的重疊區塊來測試，並記錄實驗結果的召回率與正確率，實驗使用 corel image 550 張建築物影像，結果如下：

表 4.5 特定類型建築物偵測之召回率與正確率

建築物類型	召回率	正確率
都市方形小窗戶建築物	85.7%	60%
古羅馬式建築物(柱子)	60%	54%
哥德式建築物	50%	46.2%
都市建築物	52.8%	48.6%

都市方形小窗戶建築物有較明顯與規律的紋理特徵，因此有較好的實驗結果，而古羅馬式建築物與哥德式建築物的紋理特徵較不規律，因此結果大概只在五成左右。

第五章

結論與未來展望

5.1 結論

本文提出了以紋理特徵為基礎的建築物偵測方法來進行不同類型建築物的偵測，方法會先使用較小區塊中的紋理特徵找出影像中有可能為建築物的區域，再利用較大範圍區域中之紋理特徵來判斷是否為建築物，來找出建築物的區域，因此偵測結果較不容易受到建築物區域大小影響，在判斷影像是否為建築物影像時，則是會受到建築物區域偵的結果所影響。

我們利用本方法來實作建築物的偵測，實驗證明了本方法在建築物區域偵測與建築物影像分類有不錯的結果。



5.2 未來展望

在本論文的研究與實驗中，發現有幾個細節是我們可以繼續改進的重點，在此說明如下：

1. 本研究中建築物區塊與建築物區域的判斷已有不錯的結果，但可以再找出其它有效的特徵合併使用，來增加正確率。
2. 特定類型建築物偵測目前是使用紋理的特徵，但有些類型偵測的結果較好有些較差，所以希望未來能找到其他更有效的特徵，且不同建築物類型使用不同的特徵。

參考文獻

- [1] A. Iqbal and J.K. Aggarwal, "Applying perceptual grouping to content-based image retrieval: Building images", In Proc. IEEE Int. Conf. CVPR, 1:42-48, 1999.
- [2] S. Kumar and M. Hebert, "Man-made Structure Detection in Natural Images using a Causal Multiscale Random Field", In Proc. IEEE Int. Conf. on CVPR, 2003
- [3] Hoang-Hon Trinh Dae-Nyeon Kim and Kang-Huyn Jo "Urban building detection by visual and geometrical features, 2004
- [4] Q. Yanyun, Z. nanning, salient building detection in natural image using SVM IEEE Intl Conf, 2005
- [5] Yqing Song and Aidong Zhang, analyzing scenery images by monotonic tree, ACM Multimedia System Journal, 2002
- [6] D.M. McKeown, "Toward Automatic Cartographic Feature Extraction, Mapping and Spatial Modelling for Navigation, L.F. Pau, ed., NATO ASI series, vol. F65, pp. 149-180, 1990.
- [7] R.B. Irvin and D.M. McKeown, "Methods for Exploiting the Relationship Between Buildings and Their Shadows in Aerial Imagery," IEEE Trans. Systems, Man, and Cybernetics, vol. 19, no. 6, pp. 1,564-1,575, Nov. 1989.
- [8] J.C. McGlone and J.A. Shufelt, "Projective and Object Space Geometry for Monocular Building Extraction," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 54-61, June 1994.
- [9] C. Lin, R. Nevatia, "Building Detection and Description from a Single Intensity Image", Computer Vision and Image Understanding

- [10] Christopher J. C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition". *Data Mining and Knowledge Discovery* 2:121 - 167, 1998
- [11] Jianbo Shi, J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. PAMI*, no. 22, vol. 8, pp. 888-905, 2000.
- [12] Y. Deng, B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. PAMI*, vol. 23, no. 8, pp. 800-810, 2001
- [13] D. Comaniciu, P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. PAMI*, vol. 24, no. 5, pp. 603-619, 2002.
- [14] Y. Rubner and C. Tomasi. "Texture-based image retrieval without segmentation. In *ICCV' 99*, Corfu, Greece, Sep. 1999
- [15] A. Vailaya, M. Figueiredo, A. K. Jain, and H. J. Zhang. content-based hierarchical classification of vacation images. In *IEEE conf. on Multimedia Computing and System*, vol.1 1999
- [16] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol.18 no.8, Aug. 1996
- [17] Y. Deng, C. Kenney, M. S. Moore, and B. S. Manjunath, "Peer Group Filtering and Perceptual Color Image Quantization" *IEEE Proc. International Symposium on Circuits and Systems*, vol. 4, 1999
- [18] YAN-NI WANG, LONG-BIN CHEN, BAO-G HU, "I semantic extraction of the building images using support vector machines" *Machine Learning and Cybernetics*, 2002. Proceedings. 2002 International Conference