

國立交通大學

多媒體工程研究所

碩士論文

自動圖像註解於圖像檢索系統之研究

The Study of Automated Image Annotation
for Image Retrieval Systems

研究生：周逸凡

指導教授：傅心家 教授

中華民國九十七年七月

自動圖像註解於圖像檢索系統之研究
The Study of Automated Image Annotation
for Image Retrieval Systems

研究生：周逸凡

Student：Yi-Fan Chou

指導教授：傅心家

Advisor：Prof. Hsin-Chia Fu

國立交通大學
多媒體工程研究所
碩士論文



Submitted to Institute of Multimedia 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

自動圖像註解於圖像檢索系統之研究

研究生：周逸凡

指導教授：傅心家 教授

國立交通大學多媒體工程研究所

摘要

圖像自動產生註解文字是一個極具挑戰性且困難的工作，因為其必須找出圖像內容與語意間對應的關係。本篇論文提出一個以圖像區塊為基礎產生自動圖像註解的方式。首先對圖像進行圖像分割找出其可能的物件區塊，並對圖像區塊取出特徵，以此特徵在資料庫中與其他帶有標記字詞資訊的圖像區塊做視覺相似度計算，再將與其視覺相似度較大的圖像區塊所帶的標記字詞設定到該圖像區塊上，並依據平均相似度產生各字詞所對應的權重，最後綜合圖像中各圖像區塊的標記字詞及權重資訊產生圖像註解字詞。我們以 Corel 圖像資料庫的圖像進行自動圖像註解實驗，驗證所提方法的可行性與正確率，實驗顯示當系統給出 1 個自動註解字詞時，正確率可以達到 40% 左右，而在給定 15 個自動註解字詞時，註解字詞的召回率可以達到 50% 左右。論文最後更進一步將所提的方法整合至一個既有的以內容為基礎的圖像檢索系統中，提供一個多模式的圖像搜尋與檢索機制，並融入 web 2.0 的概念，讓圖像提供者也能為圖像區塊加入人工標記字詞，以增進系統的自動圖像註解能力。

The Study of Automated Image Annotation for Image Retrieval Systems

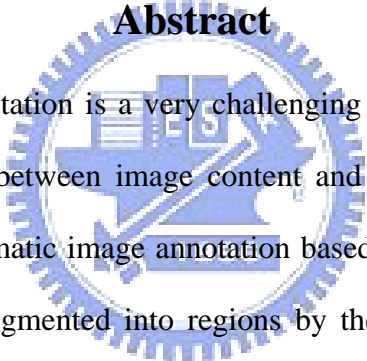
Student: Yi-Fan Chou

Advisor: Prof. Hsin-Chia Fu

Institute of Multimedia Engineering

National Chiao Tung University

Abstract



Automatic image annotation is a very challenging and difficult task, because it needs to find the relation between image content and semantic words. This paper presents a method for automatic image annotation based on region information of an image. First, images are segmented into regions by their low-level visual features. Second, visual features are extracted from regions and compute the visual similarity with other regions which have tags information in database. Then, tags are assigned to regions which have higher similarity. And weight of each tag is computed. Finally, we merge the information from regions of an image to generate the annotation words. In the experiment, we use the Corel images to verify the accuracy and feasibility of the proposed method. Moreover, the proposed method is integrated into an existed content-based image retrieval system to provide a multimodal approach for image query. We add the web 2.0 concept to make the image provider be able to add tags to regions of the image. By this way, we can maintain or improve the system ability of automatic image annotation.

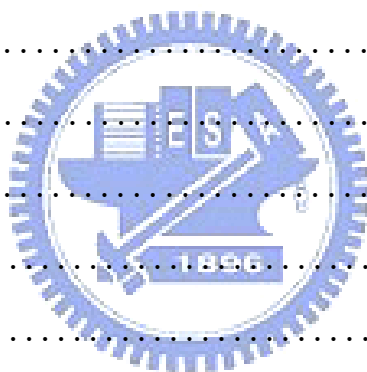
誌謝

終於完成這本論了！首先要感謝傅心家老師的指導和照顧，承蒙老師給予學生的鼓勵與支持並幫助我找到研究方向，使得本論文能順利完成。在此謹向老師致上最誠摯之謝忱與敬意。同時，感謝實驗室博士後研究以及博士班學長，永煜、岳宏、柏伸、政龍、士賢，同學昭翰、佳蓁，還有學弟坤隆、威人平常在生活及學業上的建議與指教，也讓我的研究生生涯多采多姿充滿歡樂。特別感謝永煜學長在論文上的極大幫助，讓我認識資訊檢索方面的相關知識也幫助我解決困難並修改論文，讓論文更為完善。感謝平時一起運動的室友信龍在生活上的照顧，讓我可以有健康的身體為課業打拼。感謝我最心愛的未來老婆馨俐，在生活中不斷給我鼓勵與支持，在我退縮的時候幫我打氣，成為我得以進入交大就讀的推手，也感謝俐爸、俐媽、俐姊們、友千、成家對我的關心與幫忙，讓我倍感窩心。最後，感謝從小一直支持我的爸爸、脾氣超好的媽媽、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 前言.....	1
1.1 研究動機.....	1
1.2 研究目標.....	2
1.3 研究方向.....	2
1.4 章節介紹.....	3
第二章 相關研究.....	4
2.1 圖像檢索.....	4
2.1.1 以文字為基礎的圖像檢索(TBIR).....	4
2.1.2 以內容為基礎的圖像檢索(CBIR).....	5
2.2 自動圖像註解.....	5
2.2.1 文字-圖像聯合模型.....	6
2.2.2 監督式分類方法.....	6
2.2.3 web 2.0 標記分類.....	7
第三章 以圖像區塊為基礎自動圖像註解.....	8
3.1 圖像分割與圖像區塊特徵擷取.....	10
3.2 圖像區塊標記字詞及權重建構.....	11
3.2.1 圖像區塊候選標記字詞之產生.....	12



3.2.2	圖像區塊標記字詞權重計算.....	13
3.3	綜合圖像區塊資訊產生圖像註解字詞.....	15
3.3.1	根據最大權重產生圖像註解字詞.....	15
3.3.2	根據最大出現頻率標記字詞產生圖像註解字詞.....	16
3.3.3	根據權重與圖像區塊在圖像中所佔比例產生圖像註解字詞...	17
第四章	實驗結果.....	18
4.1	圖像區塊自動標記字詞實驗與結果分析.....	19
4.1.1	圖像區塊標記字詞與權重建構實驗與結果分析.....	19
4.1.2	標記字詞及權重建構次數對圖像區塊標記正確性之影響析...	21
4.2	圖像自動註解實驗與分析.....	21
4.2.1	圖像區塊標記權重建構對自動圖像註解之影響.....	22
4.2.2	圖像區塊大小限制對自動圖像註解之效能影響.....	27
4.2.3	有正確人工標記字詞提供對自動圖像註解效能影響.....	30
4.2.4	使用不同方法綜合圖像區塊資訊產生圖像註解.....	31
4.3	自動圖像註解範例.....	32
第五章	結論與未來展望.....	35
5.1	結論.....	35
5.2	未來展望.....	36
參考文獻	37

表目錄

表 4-1：標記權重建構對圖像區塊自動正確標記之影響.....	20
表 4-2：資料庫標記權重建構次數對圖像區塊標記正確性之影響.....	21
表 4-3：使用不同圖像區塊大小限制產生自動圖像註解所需平均時間.....	29



圖目錄

圖 1-1： 圖像註解字詞.....	2
圖 3-1： 單張圖像自動註解流程圖.....	9
圖 3-2： 圖像處理流程.....	10
圖 3-3： 圖像模糊化範例.....	10
圖 3-4： 圖像區塊標記字詞及權重概念圖.....	12
圖 3-5： 候選標記字詞產生流程.....	13
圖 3-6： 標記字詞設定流程.....	14
圖 3-7： 標記字詞權重產生流程.....	15
圖 3-8： 圖像區塊標記字詞及權重建構示意圖.....	16
圖 4-1： 人工標記字詞示意圖.....	22
圖 4-2： 自動圖像註解正確率(標記字詞及權重建構之影響).....	23
圖 4-3： 自動圖像註解召回率(標記字詞及權重建構之影響).....	23
圖 4-4： 自動圖像註解正確率(資料庫未經標記字詞及權重建構).....	25
圖 4-5： 自動圖像註解召回率(資料庫未經標記字詞及權重建構).....	25
圖 4-6： 自動圖像註解正確率(資料庫經過標記字詞及權重建構).....	26
圖 4-7： 自動圖像註解召回率(資料庫經過標記字詞及權重建構).....	26
圖 4-8： 自動圖像註解正確率(限制圖像區塊大小).....	28
圖 4-9： 圖像大小限制 300 像素點時自動註解正確率與召回率.....	29
圖 4-10： 自動圖像註解正確率(持續提供人工標記字詞).....	30
圖 4-11： 自動圖像註解召回率(持續提供人工標記字詞).....	30
圖 4-12： 自動圖像註解正確率(使用不同註解字詞產生方法).....	31
圖 4-13： 自動圖像註解召回率(使用不同註解字詞產生方法).....	32

圖 4-14：自動圖像註解較為成功範例..... 33


圖 4-15：自動圖像註解無法找到適當的註解字詞的範例..... 34



第 1 章

前言

1.1 研究動機



俗話說「一圖勝千文」，說明了圖像所能帶給人的感受往往可以勝過千言萬語，而隨著網際網路的蓬勃發展，網路中的圖像日以百萬計地增加，如何從眾多的圖像中找出需要的圖像，便成了一個重要的研究領域；在現今的圖像檢索系統中，大部分均須以賴以文字為基礎(text-based)的檢索方式來找出想要的圖像，然而在此種檢索方式下，圖像的描述文字便成了影響檢索結果的一個重要因素，使用人工對圖像分類、描述的方式雖然可以有不錯的分類效果，但此種方式耗時費力，且不同的人看同一張圖像往往會有不同的解讀，會因為人的主觀因素使得檢索系統漸漸變得缺乏客觀性。因此，如果有一個系統能夠對圖像自動產生與其相關的字詞，將可以大量降低人力的介入，並且可以對圖像搜尋引擎從網路中搜尋到的圖像提供其相關的字詞讓以文字為基礎的檢索使用，所以我們希望能發展一個自動圖像註解系統。

1.2 研究目標

在本研究中我們希望所建置的系統能夠對使用者提供的圖像，以客觀的方式自動產生與該圖像相關的註解字詞(annotation words)，如圖 1-1，並且將本研究所提的方法應用在以內容為基礎(content-based)的圖像檢索系統中，提供一個多模式的圖像檢索系統。



圖像		
註解字詞	mountain tree snow sky forest river	tiger grass water

圖 1-1：圖像註解字詞。

1.3 研究方向

本論文的研究方向在於使用以圖像區塊為基礎的方式，藉由圖像區塊之間的視覺相似度比對，將圖像區塊所擁有的標記字詞自動設定到與其視覺相似度較高的圖像區塊上來達到產生自動註解字詞的目的，此外我們希望能將所提的方法整合到已有的以內容為基礎的圖像檢索系統中使其成為多模式的圖像檢索系統，並加入以 web 2.0 為基礎的圖像標記字詞建構機制，藉由使用者檢索時提供的範例圖像相關字詞或是對範例圖像設定標記字詞，來增進系統的自動圖像註解正確率，以及增加系統中的字詞種類，並提供一個標記字詞驗證機制，對使用者所提供的標記字詞做把關，降低不適當標記字詞對圖像檢索結果的影響。

1.4 章節介紹

在以下章節中，第二章首先介紹圖像檢索、自動圖像註解相關研究；第三章介紹本研究如何以圖像區塊為基礎進行圖像區塊標記字詞及權重建構，以及如何綜合圖像區塊的資訊產生自動圖像註解；第四章是第三章所提出的自動圖像註解方法實驗結果，以驗證其可行性與效能；第五章是結論及對未來的展望。



第 2 章

相關研究

2.1 圖像檢索

如何從眾多圖像中找出需要的圖像是一個被研究多年的題目，而其方法大致可以分為兩大類：(1)以文字為基礎的圖像檢索(Text-Based Image Retrieval, TBIR)；(2)以內容為基礎的圖像檢索(Content-Based Image Retrieval, CBIR)。

2.1.1 以文字為基礎的圖像檢索(TBIR)

在以文字為基礎的圖像檢索中，檢索系統主要根據使用者輸入的關鍵字詞在資料庫中與圖像的描述文字做比對，找出符合的圖像，此種方式通常可以很快地找出檢索結果，因為系統的主要工作只是在資料庫中做搜尋，然而檢索結果是否符合所需則需倚賴資料庫中的圖像描述文字。

在早期的系統中，圖像的描述文字通常以分享者所提供的資訊或是檔案名稱為主或是由人工進行標記分類，然而在這個網際網路資料量倍增的時代，此種方式太過耗時費力；Shen 等人【1】曾提出將圖像所在的網頁中的文字作為圖像的潛在註解文字，用以自動產生圖像的描述文字供以文字為基礎的圖像檢索使用；Srihari 等人【2】提出藉由圖像週遭的敘述文字透過分析後取出適合該圖像的索

引字詞作為圖像的描述文字，然而這些方式均需要在網頁圖像週遭有相關文字資訊的情況下，才能夠順利產生圖像的描述文字。

2.1.2 以內容為基礎的圖像檢索(CBIR)

儲存體、硬體成本的降低以及網際網路的普及促進了以內容為基礎的圖像檢索朝著許多不同的方向迅速發展，Smeulders 等人【3】將其應用大致分為三種廣泛的型態：(1)目標搜尋(target search)－主要利用樣式比對(pattern matching)及物件辨識技術來達到搜尋的目的，在不同場景及大量物件型式的情況下，以內容為基礎的檢索方式面臨一大挑戰；(2)分類搜尋(category search)－植基於物件辨識與樣式模型統計(statistical pattern recognition)等問題上，大量的類別及缺乏特徵擷取時的明確相位為其困難處；(3)關聯性搜尋(search by association)－以未具體指定的目標物為搜尋目標，其結果往往會受限於電腦視覺領域上所使用的特徵集合與相似度函數，亦須解決語意上的困難問題。

以內容為基礎的圖像檢索主要目的就是希望在缺乏元資料(meta-data)的情況下，客觀地找出圖像中可能擁有的概念(concept)；徐【4】利用複合式高斯混合模型進行圖像檢索，首先將圖像進行圖像分割，對於分割後的圖像讓使用者選擇其感興趣的部份，即欲檢索的圖像區塊(region of interest)做為視覺關鍵字在資料庫中進行相似度計算；然而在以內容為基礎的圖像檢索方式中，由於需要做較高維度的相似度計算比較，速度上往往較以文字為基礎的圖像檢索需要更多的時間。

2.2 自動圖像註解(Automated Image Annotation)

如何對一張圖像自動給予適當的圖像註解是一個被認為很具挑戰性且困難的問題，而其亦與以內容為基礎的圖像檢索有著很大的相關性【5】，若圖像與字詞間的自動對應能夠達到一定的可靠程度，則在此自動對應的輔助之下，以內容為基礎的圖像搜尋在語意方面將變得更具有意義【6】。

2.2.1 文字-圖像聯合模型(Joint Word-Picture Modeling Approach)

自動圖像註解就像是將一個圖像分割區塊集合轉換到一個文字集合中【7】；在此種方法下，通常先將所有圖像分割後，再將所有分割後的區塊(region)分群成多個小群(blobs)，然後由這些小群來描述每一張圖像，對於訓練資料(training data)集合的每張圖像可以看成是由幾個小群與文字所組成，只要文字與小群之間的聯合機率經過學習之後，自動圖像註解便成為了在小群與文字之間尋找近似值(likelihood)的問題。

Jeon 等人【8】用跨媒體相關模型(Cross-Media relevance models)來達到自動圖像註解的目的，他們假設每一張圖像 I 存在一些潛在的機率分布 $P(\cdot|I)$ ，即 I 的相關模型(relevance model)， I 經過分割分群後可以用小群 $\{b_1, \dots, b_m\}$ 來描述，再用 $P(w|I)$ 來估測字典中每一個字詞 w 在圖像 I 中的機率來選出適當的註解字詞，但直接計算 $P(w|I)$ 並不容易，故經由 $P(w|I) \approx P(w|b_1, \dots, b_m)$ 來求其機率，小群本身並不包含文字，然而可以利用訓練資料集合來估測字詞 w 與小群 b_1, \dots, b_m 在同張圖像的聯合機率 $P(w|b_1, \dots, b_m) = \sum_{J \in T} P(J)P(w, b_1, \dots, b_m | J)$ ，進而找出 $P(w|I)$ 較大的字詞作為圖像 I 的註解字詞。

2.2.2 監督式分類方法(Supervised Categorization Approach)

此方法將圖像註解的問題視為一個監督式分類的問題，它將每一個不同的註解字詞或是語意分類視為一個獨立的集合，並創造一個不同的圖像分類模型給每一個字詞或類別。

Vailaya 等人【9】提出階層式的分類方法來達到圖像分類的目的，藉由訓練多個貝氏分類器，將圖像先做大範圍的分類，進而再做較小範圍的分類，舉例來說，在最上層的分類器先將圖像分為 indoor 與 outdoor 兩個分類，在 outdoor 的類別裡再將其分為 city 或是 landscape，最後再進一步將 landscape 的類別分為 sunset、forest 與 mountain 等子類別；然而，此種方法須作不同階層分類類別的判斷，會侷限於初始設定的類別個數。

Li 等人【10】提出的 ALIP(Automatic Linguistic Indexing of Pictures)系統是用此方法的一個最具代表性的系統，其假設訓練資料圖像均已經過適當的分類，對於每一個分類集合訓練一個 2 維度多解析式的隱馬可夫模型(Hidden Markov Model)，當查詢圖像進入系統後，用各個已儲存事先訓練好的分類模型計算其近似值，並依照這些近似值作為註解字詞排序依據。

2.2.3 web 2.0 標記分類

Tim O'Reilly【11】在其文章中提到了 web 2.0 的關鍵原則：“用戶越多，服務越好”，藉由網際網路作為平台並結合集體智慧來達到各方面的應用，而其中一種被人們稱為“分眾分類”(有別於傳統分類法，folksonomy)的概念更獲得了廣泛的關注，“分眾分類”是一種使用用戶自由選擇的關鍵字對網站進行協作分類的方式，而這些關鍵字一般稱為標籤(tags)，標籤化的分類法有別於傳統的分類法，其運用類似人類大腦所使用的多重關聯式的分類法，例如對一張小狗的照片而言，其可能被加上「小狗」或是「可愛」這樣的標籤，而搜尋時允許系統依照用戶行為所產生的自然方式進行檢索。

在著名的網路書籤網站「美味書籤」(del.icio.us)【12】中，便是藉由收集不同使用者對每個不同的網址所用的標記字詞資訊，來產生每個網址對應的標記字詞及分類，當下一次有使用者在書籤中加入該網址時，便將該網址出現頻率較大的標記字詞提供給使用者作為建議標記字詞，如此長久下來，可以將較熱門的網址對應到一些合適的標記字詞，做不同的分類，例如 www.google.com 會對應到「入口網站」、「信箱」等。

第 3 章

以圖像區塊為基礎自動圖像註解

本章敘述所提的自動圖像註解方法，本方法主要針對單張圖像中各圖像區塊進行標記字詞與標記字詞權重計算，再綜合各區塊的標記字詞及權重產生該張圖像的註解字詞排序，圖 3-1 是對單張圖像進行自動圖像註解的流程圖，首先對新進圖像做圖像分割，切割出可能的物件區塊，針對較大的區塊在資料庫中做以內容為基礎的相似度比對，對於比對後相似度較大的結果統計其各個標記字詞的出現次數，出現次數較多者視為候選標記字詞並計算其平均相似度作為權重，並將權重較大的候選標記字詞設定給該圖像區塊，最後綜合該圖像各圖像區塊的標記字詞與權重產生該張圖像註解字詞。本章 3.1 節介紹圖像分割與圖像區塊特徵擷取；3.2 節介紹圖像區塊標記字詞及權重之建構方式；3.3 節介紹如何綜合圖像各區塊資訊產生圖像註解字詞。

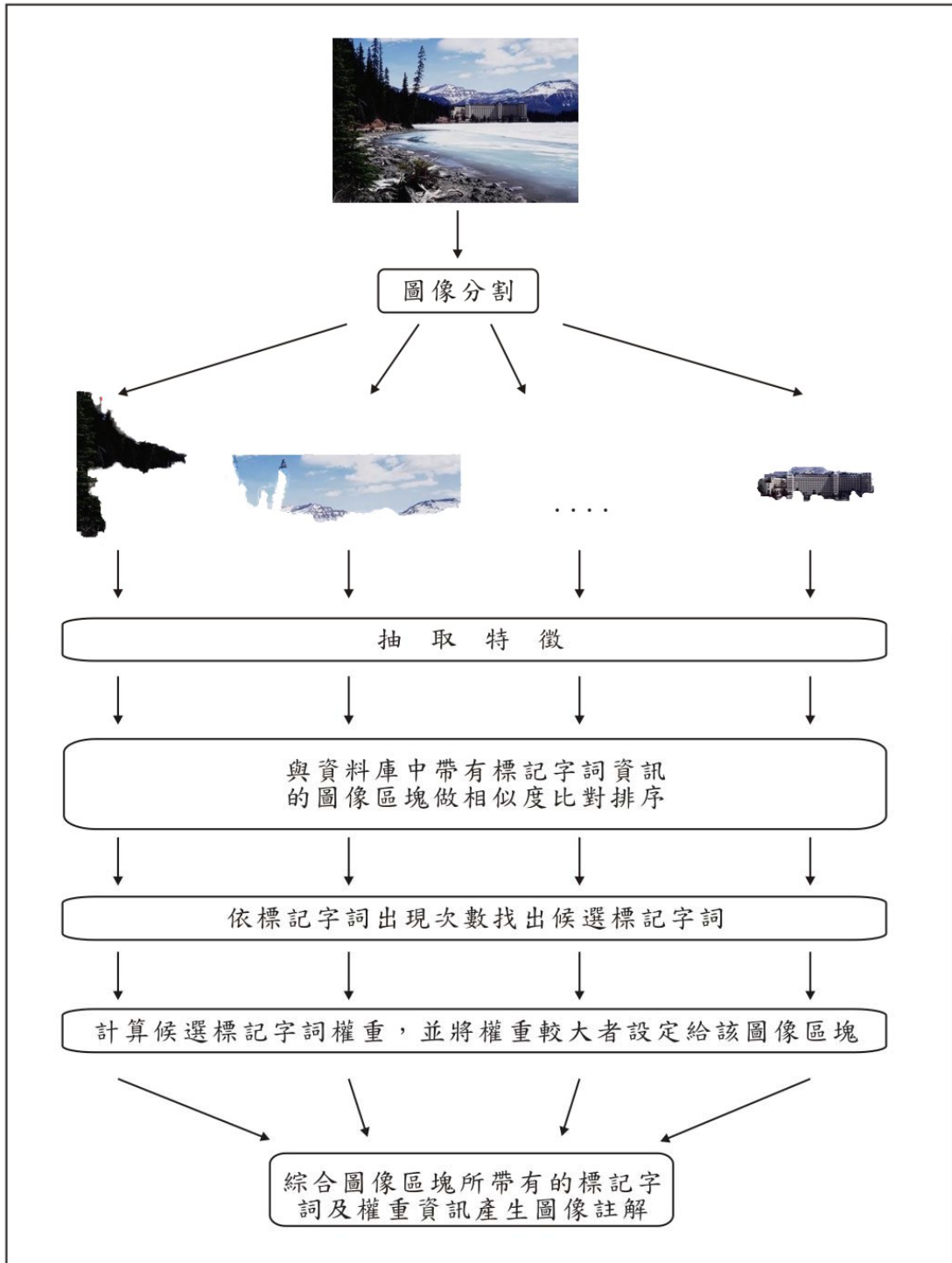


圖 3-1：單張圖像自動註解流程圖。

3.1 圖像分割與圖像區塊特徵擷取

本節說明對圖像所做的前處理與特徵擷取，首先藉由視覺相似度的比較將一張圖像分割成許多視覺相似區塊，再對各個視覺相似區塊取出一組特徵向量作為其描述，用於與其它圖像區塊間的相似度計算。

圖像處理流程如圖 3-2，本方法希望能將圖像中具有視覺相似性的物件區塊取出，為了降低雜點對圖像分割的影響，先將圖像模糊化，再將視覺上相似度較高的點分成幾個區域，流程圖中各步驟詳述如下：

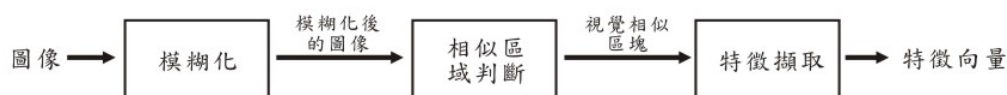


圖 3-2：圖像處理流程。

步驟一：圖像模糊化

高斯低通濾波器(Gaussian low-pass filter)【13】是很常用來做圖像模糊化的一種方法，然而此方法也可能會將圖像中很明顯的邊緣或邊界模糊化，為了避免此種現象發生，使用非線性擴散(anisotropic diffusion)【14】方法做處理，其原理為對圖像中一個座標點做模糊化時，先偵測其附近是否有顯著的邊界存在，再對邊界內的區域進行模糊化，如此只會將區域內部顏色較相近的部份模糊化而保留原有清楚的邊緣或邊界，圖 3-3 為圖像模糊化的範例，圖 3-3(a)為原圖像，(b)與(c)分別為使用高斯低通濾波方法與非線性擴散方法模糊化後的圖像。

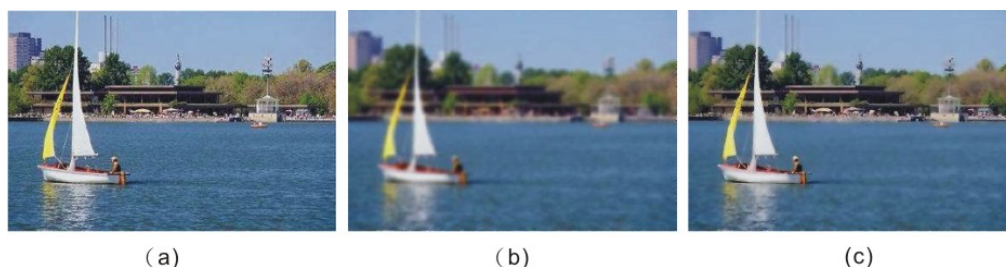


圖 3-3：圖像模糊化範例。(a)原圖像，(b)經高斯低通濾波模糊化圖像，(c)經非線性擴散模糊化圖像。

步驟二：相似區域判斷

為了使得人們視覺上看起來相近的顏色在顏色空間上的數值亦很接近，在計算相似區域時，我們採用 Lab 色彩系統來計算相似度，即對於圖像 I 中的每一個點用一個 3 維度的向量 $C_I(x)$ 表示，分別對應到 Lab 色彩空間的三個向度。

如此，對於一個視覺相似區域 A_i 可以從圖像 I 中的一個參考點 x_s 開始找尋其周圍附近視覺相似的點，直到 A_i 中的點其視覺上與其它點差異較大為止，即 $A_i = \{x: x \in I, \|C_I(x) - C_I(x_s)\| \leq \lambda\}$ ，其中 $\|\cdot\|$ 使用 Lab 的色彩距離計算公式，而 λ 為一個臨界值。

步驟三：視覺相似區塊特徵擷取

當找出視覺相似的區域後，我們對每個區域取出一組 6 維度的特徵向量

$D_{A_i} = [CM, CV]^T$ 作為對該圖像區塊的特徵描述，其中 CM 為 A_i 中所有點的顏色平均值，即 $CM = \frac{1}{|A_i|} \sum_{x \in A_i} C_I(x)$ ，而 CV 為 A_i 中所有像素點的顏色變異數，即 $CV = \frac{1}{|A_i|} \sum_{x \in A_i} (C_I(x) - CM)^2$ 。

圖像區塊與圖像區塊之間的視覺相似度計算方式可以參考【15】所提出的色彩相似度計算方式。

3.2 圖像區塊標記字詞及權重建構

在前一節中介紹完圖像分割與區塊特徵擷取後，由於本研究所提之自動圖像註解方法是根據圖像的各個區塊標記字詞及權重而得，本節將說明如何對圖像區塊進行標記字詞及權重之建構，在經過圖像區塊標記字詞及權重建構後每一個圖像區塊所帶的資訊如圖 3-4，目的是希望將沒有任何標記字詞的圖像區塊經過標記字詞及權重建構後，輸出的圖像區塊可以帶有多個不同標記字詞及權重資訊；而圖像區塊標記字詞及權重建構的另一個目的在於對資料庫中的訓練資料而言，希望能藉由此流程將少量的人工標記字詞經過相似度的比較，自動設定標記

字詞給相似度較大的圖像區塊，使得資料庫中可以產生大量帶有標記字詞資訊的圖像區塊。

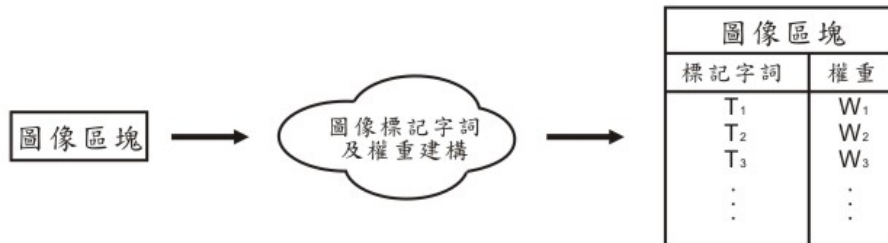


圖 3-4：圖像區塊標記字詞及權重概念圖。

根據 3.1 節所提到的圖像分割方法，一張圖像 I 經過分割處理後可以表示成 $I = \{A_1, \dots, A_n\}$ ，每一張圖像視其複雜度不同可能會有不同的 n 值，即由其各個區塊組合而成，而資料庫中的訓練資料相當於是由許多的圖像區塊組成，我們希望能對圖像區塊 R 標記出一組關係，使得 $R = \{T_1^R, T_2^R, \dots, T_k^R; w_1^R, w_2^R, \dots, w_k^R\}$ ，其中 $\{T_1^R, T_2^R, \dots, T_k^R\}$ 為系統找出可能適合該圖像區塊的標記字詞， $\{w_1^R, w_2^R, \dots, w_k^R\}$ 為這些標記字詞相對於該圖像區塊的權重值。在 3.2.1 節中說明圖像區塊候選標記字詞之產生方式；3.2.2 節說明如何對候選標記字詞計算其權重並將其設定到圖像區塊上。

3.2.1 圖像區塊候選標記字詞之產生

在對圖像區塊進行標記字詞設定前須先找出可能適合該圖像區塊的候選標記字詞，而候選標記字詞的產生流程如圖 3-5。首先對圖像區塊 R 與資料庫中其他已帶有標記字詞資訊的圖像區塊進行視覺相似度比對並將結果排序，在結果中統計所有視覺相似度與圖像區塊 R 較相近的圖像區塊所附帶的各個不同的標記字詞出現次數，若某一標記字詞出現次數高於某一事先設定好的臨界值 N ，則將其視為候選標記字詞，並進行該標記字詞權重之計算。

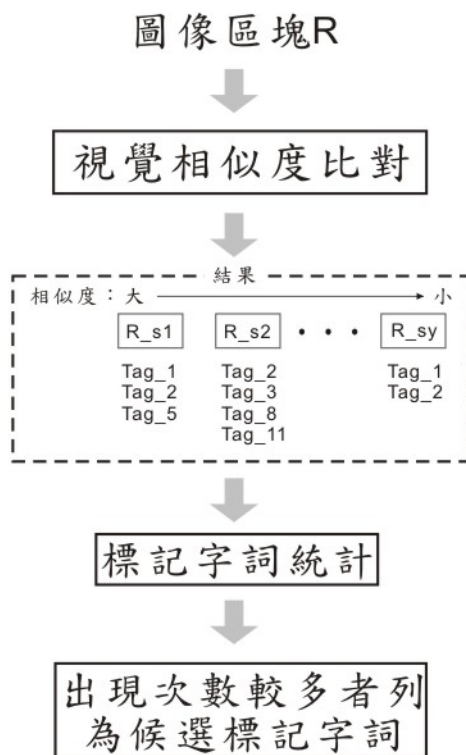


圖 3-5：候選標記字詞產生流程。

3.2.2 圖像區塊標記字詞權重計算

使用標記字詞權重值的效益有二：(1)當使用文字為基礎的檢索時，權重可作為檢索結果排序的依據，在產生圖像註解時，權重亦可作為該標記字詞出現的排序優先權；(2)在 Web 2.0 的圖像檢索系統中，當使用者錯用或故意誤設標記字詞給圖像時會因為得不到高的權重值而將其所給定的標記字詞消除，可降低人為因素造成的不良影響甚至錯誤的檢索結果。

將標記字詞設定到圖像區塊上的流程如圖 3-6，候選標記字詞輸入到權重產生器產生權重後，若其大於預先設定的臨界值，則將該後選字詞及權重設定到圖像區塊上，若小於預先設定的臨界值 W 則表示該圖像區塊與該候選標記字詞的相關性不高，將該標記字詞捨棄。

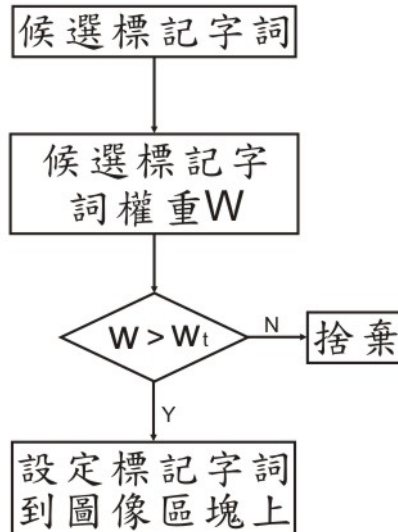


圖 3-6：標記字詞設定流程。

標記字詞之權重可視為該標記字詞對於該圖像區塊的相關程度，當候選標記字詞輸入到權重產生器之後，權重產生器先以候選標記字詞為關鍵字做以文字為基礎的檢索(TBIR)，找出資料庫中具有該標記字詞的圖像區塊，再對檢索出的結果與欲標記的圖像區塊做視覺相似度比對後並排序，並取其前 m 個平均值做為該標記字詞在該圖像區塊上的權重值，權重產生器的詳細過程如圖 3-7，其中 TBIR 的檢索結果若太少則將標記字詞暫存但不計算其權重，這是因為在 web 2.0 的標記字詞建構下，若使用者所提供的標記字詞在資料庫中未出現過，則其權重用此方法亦無法計算出，故將其暫時保留且給予一個人工設定標籤，待資料庫更新時若有多個該標記字詞，則可將其權重算出，如此可以讓整個系統的標記字詞隨著 web 2.0 的方式建構，而有越來越多種類的標記字詞產生。

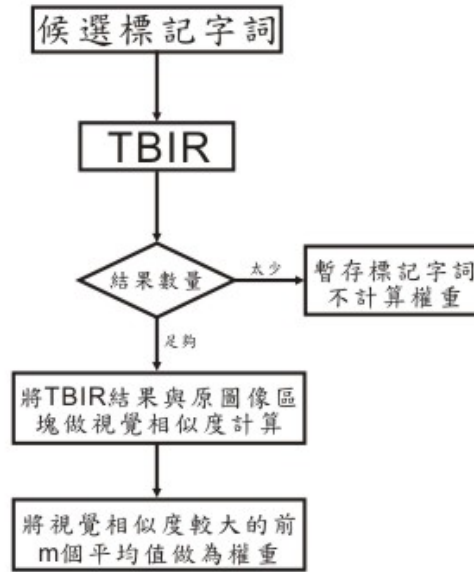


圖 3-7：標記字詞權重產生流程。

3.3 綜合圖像區塊資訊產生圖像註解字詞

在本節中我們設計了三種不同的方法利用圖像區塊的資訊來產生圖像註解；3.3.1 節的方法主要考慮圖像區塊上的標記字詞權重來產生註解字詞排序；3.3.2 節的方法忽略標記字詞權重而僅考慮標記字詞在圖像中的出現次數來作為產生圖像註解的依據；3.3.3 節則將圖像區塊在圖像中所佔的面積大小也加入成為產生圖像註解的考量因素之一；對於本節所提出的幾個方法，將在第四章中以實驗驗證其效能與可行性。

3.3.1 根據最大權重產生圖像註解字詞

當圖像 I 中的圖像區塊經過標記字詞與權重建構後，藉由這些帶有標記字詞的圖像區塊資訊來產生圖像 I 的註解字詞；首先找出在圖像 I 的所有圖像區塊中所出現的不同的標記字詞，並找出每一個不同的標記字詞所擁有的最大權重值，再利用此數值將各個不同的標記字詞排序，產生圖像的註解字詞，此數值可視為該標記字詞在圖像 I 上的近似值。

舉例如下，假設圖 3-8 為圖像 I 經過圖像區塊標記字詞及權重建構後的結果， $R_x^I, x \in \{A, B, \dots, F\}$ 為圖像 I 的圖像區塊資訊，即 $R_A^I = \{T1, T3, T6; w_{T1}^A, w_{T3}^A, w_{T6}^A\}$ ， $R_B^I = \{T1, T2, T4, T8; w_{T1}^B, w_{T2}^B, w_{T4}^B, w_{T8}^B\}$ ，以此類推，則圖像 I 的註解字詞為 $\{T1, T2, T3, T4, T5, T6, T8\}$ ，而其對應到一組權重值 $\{W_{T1}^I, W_{T2}^I, W_{T3}^I, W_{T4}^I, W_{T5}^I, W_{T6}^I, W_{T8}^I\}$ ，其中 $W_{T1}^I = \max\{w_{T1}^A, w_{T1}^B, w_{T1}^C, w_{T1}^E\}$ ， $W_{T2}^I = \max\{w_{T2}^B, w_{T2}^D\}$ ，以此類推，而 $W_{T_i}^I$ 值越大就表示 T_i 這個標記字詞適用於圖像 I 的可能性越高，在產生圖像註解字詞時，便可使用 $W_{T_i}^I$ 來決定其排序優先順序。

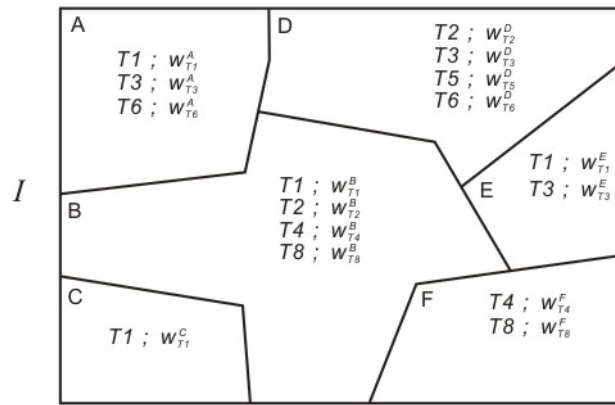


圖 3-8：圖像區塊標記字詞及權重建構示意圖。

然而在實作系統時，圖像分割的結果通常不會如圖 3-8 那樣的理想化，很有可能會產生許多零碎、較小且不具語意的圖像區塊，因此在對圖像做自動圖像註解時，我們另外對圖像區塊的大小做限制，令其必須大於一定值才進行圖像標記字詞及權重建構。

3.3.2 根據最大出現頻率標記字詞產生圖像註解字詞

此方法不將圖像區塊所帶的標記字詞權重值納入考慮，而直接將圖像 I 中所有圖像區塊所帶的不同的標記字詞統計其出現次數，再依照其出現次數作為產生圖像註解字詞的排序依據，若相同則比較該標記字詞所在的圖像區塊面積，以圖 3-8 為例，其所有不同的標記字詞 $\{T1, T2, T3, T4, T5, T6, T8\}$ 出現次數為 $\{4, 2, 3, 2, 1, 2, 2\}$ ，則最終產生的圖像註解字詞排序為 $\{T1, T3, T2, T4, T8, T6, T5\}$ ，此方法將在

第四章中以實驗驗證其可行性。

3.3.3 根據權重與圖像區塊在圖像中所佔比例產生圖像註解字詞

此方法與 3.3.1 節的方法類似，假設圖 3-8 為圖像 I 經過圖像區塊標記字詞及權重建構後的結果，即 $R_A^I = \{T1, T3, T6; w_{T1}^A, w_{T3}^A, w_{T6}^A\}$ ， $R_B^I = \{T1, T2, T4, T8; w_{T1}^B, w_{T2}^B, w_{T4}^B, w_{T8}^B\}$ ，以此類推，則圖像 I 的註解字詞為 $\{T1, T2, T3, T4, T5, T6, T8\}$ ，而其對應到一組權重值 $\{W_{T1}^I, W_{T2}^I, W_{T3}^I, W_{T4}^I, W_{T5}^I, W_{T6}^I, W_{T8}^I\}$ ，而 W_{Tt}^I 值越大就表示 Tt 這個標記字詞適用於圖像 I 的可能性越高，在產生圖像註解字詞時，便可使用 W_{Tt}^I 來決定其排序優先順序，在此我們定義 $Area_A$ 、 $Area_B$ 、...、 $Area_F$ 為圖像 I 各圖像

區塊在圖像 I 中所佔的面積比例，則 $W_{T1}^I = \frac{\sum_{i=A,B,C,E} w_{T1}^i * Area_i}{\sum_{i=A,B,C,E} Area_i}$ ，以此類推，如此對

於較大的圖像區塊而言，可以對自動圖像註解有比較多的影響力。



第 4 章

實驗結果

在本章中對於第 3 章所提自動圖像註解方法，加以實作，並設計實驗來評估此方法的效能。對於實驗的平台，在硬體方面使用 Intel Pentium-4 3.0GHz 時脈速率中央處理器的個人電腦，搭配有 1Gigabytes 主記憶體，作業系統為 Fedora Core 5 搭配 MySQL 資料庫；在圖像資料來源方面，使用 Corel 圖像資料庫中 28 個圖像類別集合，每個類別集合有 100 張圖像，並對這 2800 張圖像做 3.1.1 節所述的圖像分段，總共分成約 21 萬個圖像區塊，在實驗中我們以人工標記字詞作為標準答案，而設定人工標記字詞的標準如下：

- (1) 針對每張圖像中較顯著的圖像區塊設定一個或多個人工標記字詞；
- (2) 對於每張圖像設定 1~7 個不同的人工標記字詞，即使用 1~7 個不同的字詞描述一張圖像。

如此一共對約 16000 個圖像區塊設定了人工標記字詞，其中包含約 300 種不同的標記字詞，實驗時從每個類別集合中隨機選取 80 張圖像做為訓練資料，即已存在資料庫中的圖像，其餘做為測試資料。

4.1 圖像區塊自動標記字詞實驗與結果分析

由於本論文提出的方法是以圖像區塊上的標記字詞為基礎建構出圖像的註解字詞，因此在本節中先進行以圖像區塊為基礎的相關實驗，即以圖像區塊為單位進行實驗。在 4.1.1 節的實驗中，比較當資料庫經過標記字詞與權重建構與資料庫未經過標記字詞與權重建構兩種情況，對於新進圖像區塊使用兩種評比方法能比較其自動正確標記的效能差異；4.1.2 節的實驗中，我們將資料庫中的圖像區塊進行多次的圖像區塊標記字詞及權重建構流程，比較兩種評比方法所得結果的變化。

4.1.1 圖像區塊標記字詞與權重建構實驗與結果分析

本小節的實驗要比較當資料庫中的訓練資料經過與未經過標記字詞及權重建構流程，對於測試資料圖像區塊能否正確標記的效能影響；此實驗測試的方式是將測試資料中擁有人工標記字詞的圖像區塊輸入到 3.2 節中所提出的標記字詞與權重建構流程，而資料庫中的訓練資料分為未經過標記字詞及權重建構與經過一次標記字詞及權重建構兩種情況。

評比的方法有兩個，第一個方法使用召回率，即若圖像區塊的人工標記字為 T_{GT}^R ， $|T_{GT}^R|$ 為其個數，而 T_S^R 為圖像區塊經過系統設定的自動標記字詞， $|T_S^R|$ 為其個數，則召回率 = $\frac{|T_{GT}^R \cap T_S^R|}{|T_{GT}^R|}$ 。

第二個方法是若測試資料的圖像區塊經過自動標記字詞及權重建構後，其原本人工標記字詞必須出現在電腦自動標記字詞中且須具有電腦自動標記字詞中的最大權重值才算正確標記。

在實驗中我們使用在 3.2.1 節中所提到的不同的門檻值 N 進行實驗，表 4-1 中第一欄為所使用的門檻值 N ，第二欄是資料庫中未經過標記權建構的情況下使用前述兩種評比方式所得到的結果，第三欄是資料庫中的訓練資料經過前述兩種評比方式所得到的結果。

表 4-1：標記權重建構對圖像區塊自動正確標記之影響。

門檻值 N	未經標記權重建構		經過一次標記權重建構	
	評比方法一	評比方法二	評比方法一	評比方法二
N = 2	40.3%	22.3%	69.4%	21.8%
N = 3	39.1%	23.2%	65%	22.5%
N = 4	38.3%	23.4%	61.9%	22.8%
N = 5	37.7%	23.3%	58.4%	23%
N = 6	37.1%	23.2%	55.8%	23.3%

從表 4-1 的結果中可以發現在資料庫未經過標記及權重建構與經過標記權重建構兩種情況下，最大差別在於使用評比方法一對於圖像區塊標記字詞的召回率，由於本論文提出的圖像註解字詞建構方法是根據其圖像區塊上的標記字詞而產生，因此對於圖像區塊是否能被正確設定出適當的標記字詞這個中間過程會對圖像自動註解字詞的效能有很大的影響，若是在圖像區塊上無法標記出正確的標記字詞，相對地，在產生圖像註解時正確的字詞也不會出現。

在經過標記權重建構的情況下，評比方法一的結果會大幅提升，這是因為在沒有經過標記權重建構時資料庫中擁有人工標記字詞的圖像區塊較少，在視覺相似度的計算上造成較少的標記字詞能夠被找出，而經過標記權重建構後，可以達到藉由少量的人工標記字詞產生大量的圖像區塊標記字詞，如此在測試時會有較多的標記字詞能夠被列為候選標記字詞，使評比方法一的結果提升。

隨著門檻值 N 的提高，可以發現評比方法一的結果逐漸下降，這是因為較高的門檻值會造成較少的標記字詞能夠被列為候選標記字詞，而在這些較少的標記字詞中，要能夠命中人工標記字詞相對機會較小，但這並不代表門檻值 N 的設定越小越好，因為在產生圖像註解時同時會考量圖像區塊上的標記字詞及其權重值，當門檻值越小時，可能會有越多的不正確的標記字詞被設定到圖像區塊上，若其又擁有很大的權重值則反而降低自動圖像註解的效能。

從評比方法二的結果看來，由於其條件訂得較為嚴格，所以其正確率較評比方法一的結果低了許多，而隨著門檻值 N 的增加，在未經過標記權重建構的情況下，其正確率看不出顯著的變化，但在經過標記權重建構後，可以發現正確率有隨著門檻值的增加稍微提升。

4.1.2 標記字詞及權重建構次數對圖像區塊標記正確性之影響

本小節實驗圖像區塊標記字詞及權重建構次數對圖像區塊正確標記效能的影響，我們對資料庫中訓練資料的圖像區塊進行 1 至 3 次的標記字詞及權重建構流程，並以 4.1.1 節中的兩種評比方法測試其對測試資料的標記效能，使用 $N=3$ ，結果如表 4-2。

表 4-2：資料庫標記權重建構次數對圖像區塊標記正確性之影響。

	資料庫標記權重建構次數			
	0	1	2	3
方法一(召回率)	39.1%	65.4%	66.1%	66.1%
方法二	23.2%	22.5%	22.2%	22.3%

從表 4-2 中可以發現，當資料庫中的圖像區塊經過多次的圖像區塊標記字詞及權重建構後，使用兩種評比方法測試所得的結果都顯示在經過 1 至 3 次標記字詞及權重建構後，所得到的正確率可以趨近一個穩定的狀態。

4.2 圖像自動註解實驗與分析

本節將以圖像為主進行實驗設計，評估產生圖像自動註解字詞的效能，主要以正確率及召回率為評比標準，其計算方式如下，假設測試資料圖像 I 的註解字詞標準答案為 T_{GT} ， $|T_{GT}|$ 為其個數，而圖像自動註解系統設定給測試資料圖像 I 的

註解字詞為 T_S ， $|T_S|$ 為其個數，則 $\frac{|T_{GT} \cap T_S|}{|T_S|}$ 即為正確率，而 $\frac{|T_{GT} \cap T_S|}{|T_{GT}|}$ 為召回率，

在此定義測試資料圖像 I 的註解字詞標準答案 T_{GT} 為圖像 I 中所有圖像區塊上的不同的人工標記字詞為標準答案，舉例如圖 4-1，若 $\{A, B, C, D, E, F\}$ 為圖像 I 的圖像區塊，而 MT 為其圖像區塊的人工標記字詞，則將 $\{MT1, MT3, MT4, MT8\}$ 視為圖像 I 的註解字詞標準答案，而若系統產生的自動註解字詞為 $\{MT1, MT2, MT3, MT4, MT6\}$ ，則正確率為 0.6，召回率為 0.75。

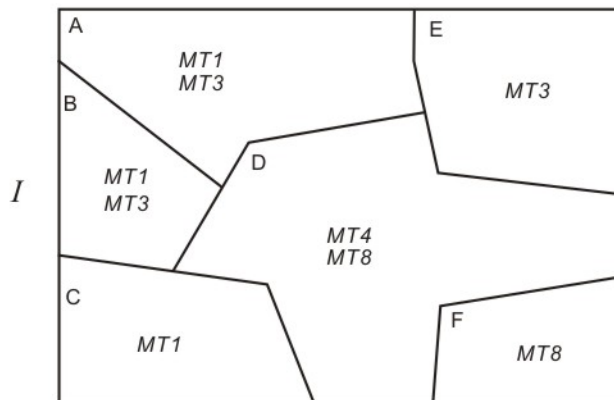


圖 4-1：人工標記字詞示意圖。

4.2.1 圖像區塊標記權重建構對自動圖像註解之影響

本小節測試圖像區塊標記權重建構對自動圖像註解正確性的影響，類似 4.1.1 節的方式，資料庫的訓練資料分為經過圖像區塊標記權重建構與未經過圖像區塊標記權重建構兩種情況，這個實驗中我們先不對圖像區塊大小做任何限制，意即將測試圖像 I 中的所有圖像區塊都進行圖像區塊標記字詞與權重建構，評比方式使用正確率與召回率；我們讓系統給出 1 至 15 個不同的自動註解字詞，在此我們使用 $N=3$ ，而自動圖像註解字詞使用 3.3.1 節中所提出的方法產生，將測試資料圖像單張輸入系統中進行自動圖像註解，完畢後將其移除換下一張測試資料圖像，圖 4-2、4-3 分別為資料庫未經過標記字詞及權重建構與資料庫經過標記字詞及權重建構的正確率與召回率比較，橫軸為系統給的註解字詞個數。

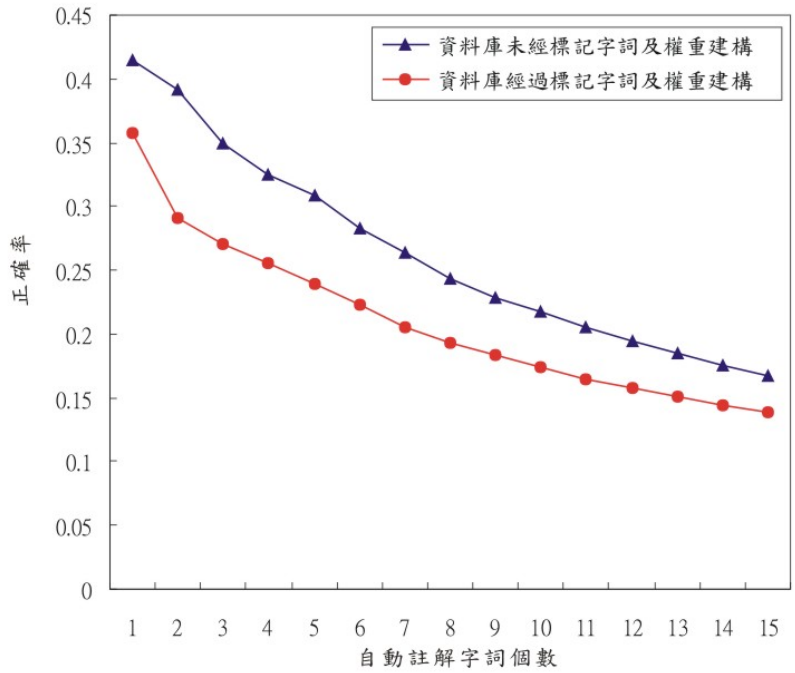


圖 4-2：自動圖像註解正確率(標記字詞及權重建構之影響)。

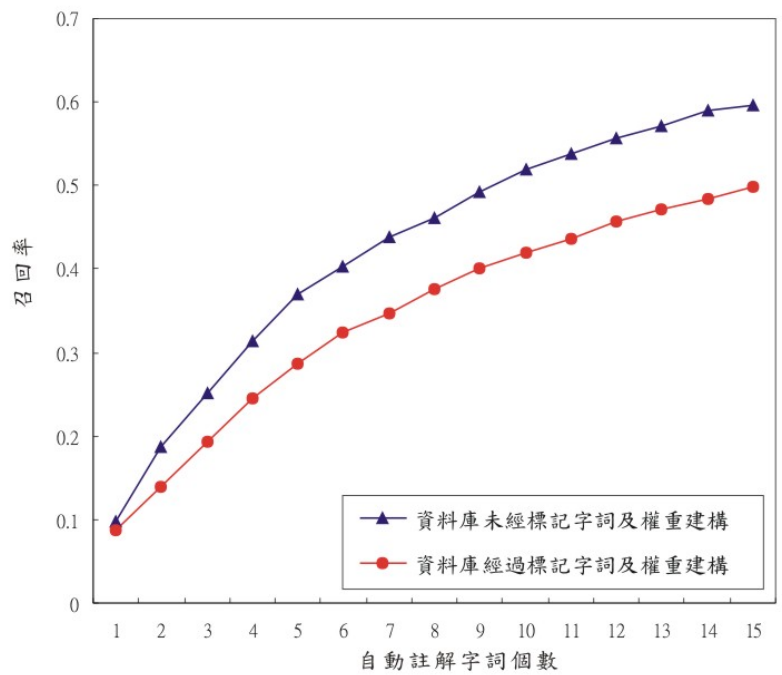


圖 4-3：自動圖像註解召回率(標記字詞及權重建構之影響)。

從圖 4-2 與 4-3 的結果可以發現隨著系統給的自動註解字詞個數增加時，正確率逐漸下降、召回率逐漸上升，這些結果都是在預期的範圍中，因為系統給的註解字詞排序在越前面代表其具有越大的權重值，即其適合該圖像的可能性越高，當個數增加時，會使得一些可能性比較不高的字詞被選出，所以正確率會下

降，而系統給的註解字詞個數越多時，圖像註解字詞的標準答案被找出的機率相對也較高。

圖 4-2 與 4-3 的結果亦可發現在資料庫未經過標記字詞及權重建構的情況下，正確率與召回率均高於資料庫經過標記字詞及權重建構的情況，這是因為在經過標記字詞及權重建構後，經由視覺相似度比較後系統自動設定一些標記字詞給圖像區塊，這個動作雖然可以達到藉由少量的人工標記字詞產生較大量的自動標記字詞，但同時也可能因為視覺相似度算出來較相近而設定了不適當的標記字詞給資料庫中的圖像區塊，造成測試資料計算註解字詞時受到不正確字詞的干擾，進而使得正確率與召回率降低，然而，在資料庫經過標記字詞及權重建構的情況下正確率與召回率曲線較為平順。

在以下實驗中，我們將測試資料以隨機順序輸入系統中進行整批測試，標記後不將先進入的測試資料移除，即隨著測試資料圖像一張一張進入資料庫後，先進入的圖像有可能影響後一張圖像的自動圖像註解結果，而資料庫中的圖像會越來越多，這是因為本論文所建置的系統是採用 web 2.0 的概念，使用者可提供圖像或字詞進行檢索，而圖像輸入到系統後，亦可作為之後輸入圖像的訓練資料；實驗同樣分為資料庫中的圖像區塊未經過標記字詞及權重建構與資料庫中的圖像區塊經過標記字詞及權重建構兩重情況，並將其結果與圖 4-2、圖 4-3 的結果比較，實驗結果如圖 4-4、圖 4-5、圖 4-6 與圖 4-7，圖 4-4 與圖 4-5 為資料庫圖像區塊未經過標記字詞及權重建構的情況下，將測試資料圖像單張測試與整批測試的結果正確率與召回率比較，圖 4-6 與圖 4-7 為資料庫圖像區塊經過標記字詞及權重建構的情況下，將測試資料圖像單張測試與整批測試的結果正確率與召回率比較。

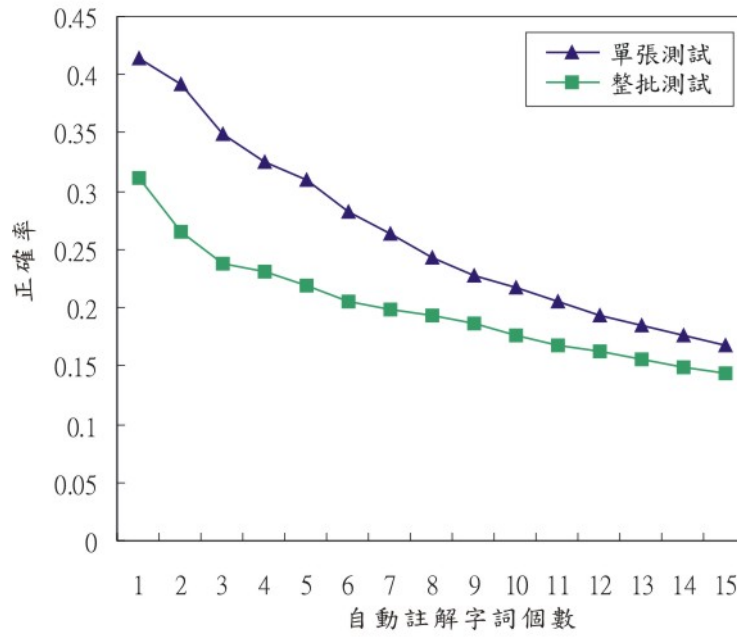


圖 4-4：自動圖像註解正確率(資料庫未經標記字詞及權重建構)。

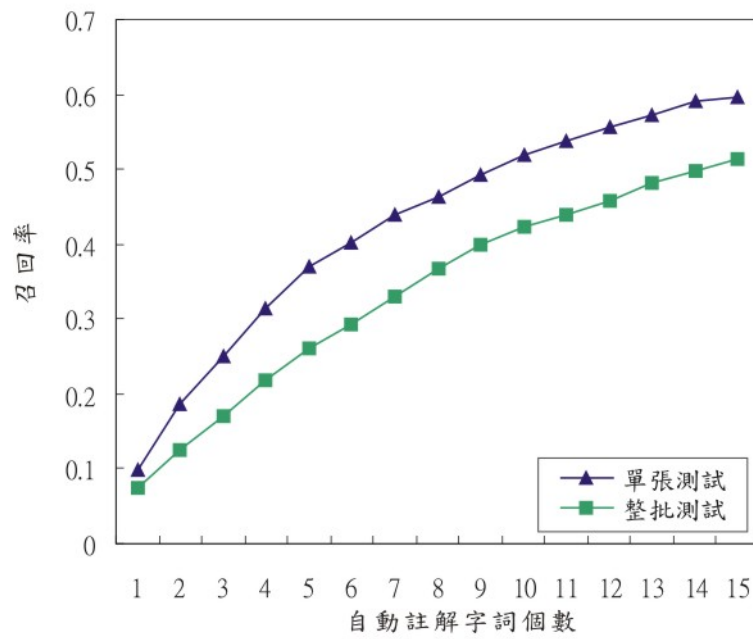


圖 4-5：自動圖像註解召回率(資料庫未經標記字詞及權重建構)。

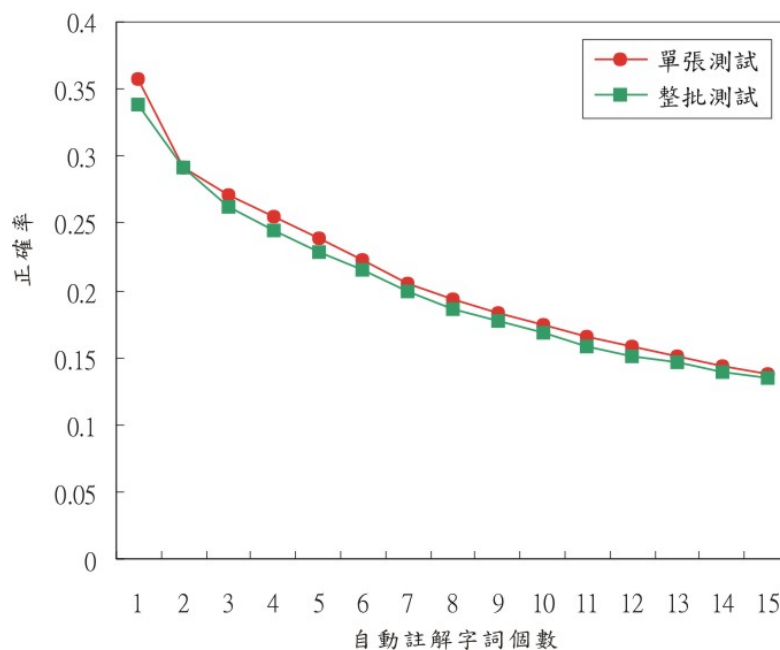


圖 4-6：自動圖像註解正確率(資料庫經過標記字詞及權重建構)。

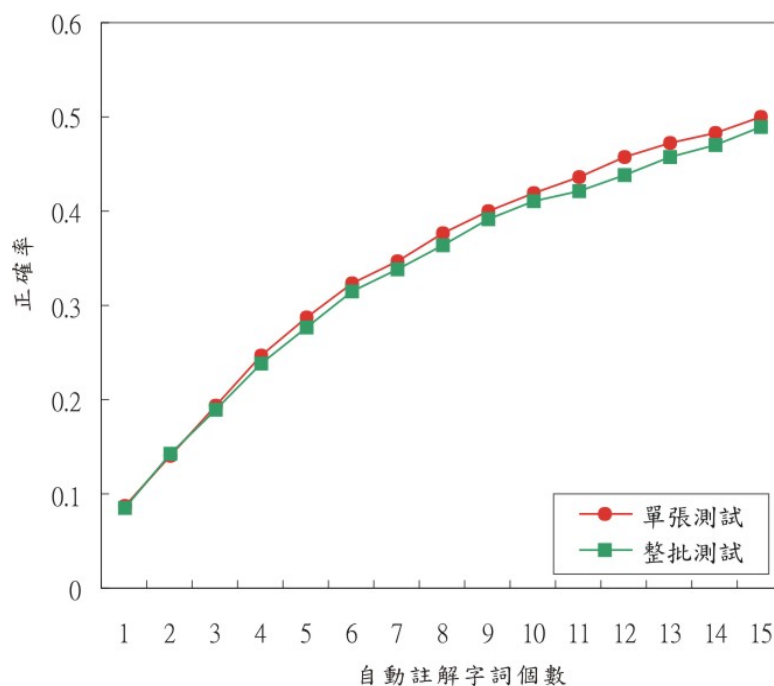


圖 4-7：自動圖像註解召回率(資料庫經過標記字詞及權重建構)。

從圖 4-4 與圖 4-5 的結果可以看出在資料庫未經過圖像區塊標記字詞及權重建構的情況下，整批測試所得到的結果不論在正確率與召回率與單張測試都產生很大幅度的下降，而在資料庫圖像區塊經過標記字詞及權重建構的情況下，圖 4-6 與圖 4-7 顯示整批測試的結果與單張測試則僅有些微的降低；整批測試對於

本系統而言相當於不斷對系統增加圖像而沒有增加人工標記字詞，因此正確率與召回率比起單張測試會降低是可以預期的，在資料庫圖像區塊未經過標記字詞及權重建構的情況下，若擁有某一標記字詞的圖像區塊數目不多，則很有可能被新加入的圖像區塊影響該標記字詞判斷的正確率，然而在資料庫圖像區塊經過標記字詞及權重建構後的情況下，擁有某一標記字詞的圖像區塊數目會有較多，比較不容易受到新進的圖像區塊影響其判斷正確率，且在 web 2.0 的建構下，若有人工標記字詞持續加入則可維持系統自動圖像註解的正確率與召回率，這點在 4.2.3 節的實驗中可以看出。

4.2.2 圖像區塊大小限制對自動圖像註解之效能影響

在 4.2.1 節的實驗中，我們在綜合圖像區塊資訊產生圖像自動註解字詞時並沒有對圖像區塊的大小做限制，然而在測試資料圖像 I 經過圖像分割後，有可能因為圖像的複雜度因素導致產生許多零碎、較小且不具語意的圖像區塊，這對圖像自動註解時有可能降低其效能，因此本小節將以實驗驗證圖像區塊大小限制對圖像自動註解效能的影響，在本實驗中分別將圖像區塊大小門檻值設定在 0、300、600、900、1200 與 1500 個像素點以上，即超過門檻值大小的圖像區塊才用來計算產生註解字詞，並使用 3.3.1 的方法產生自動註解字詞，資料庫中圖像區塊經過標記字詞及權重建構，採用 $N=3$ ，我們比較在不同的區塊大小限制條件下讓系統自動給出 1 至 5 個自動註解字詞時的正確率，結果如圖 4-8，橫軸為自動註解字詞個數。

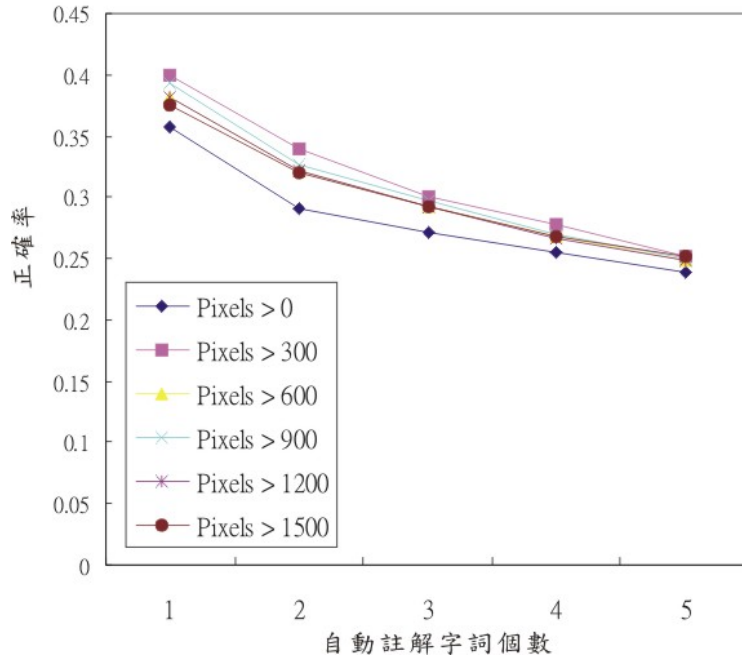


圖 4-8：自動圖像註解正確率(限制圖像區塊大小)。

由圖 4-8 可以發現當系統給的自動註解字詞個數為 1 個的時候，圖像區塊大小限制 300 個像素點以上與不限制圖像區塊大小自動註解正確率相差近 5 個百分比，且在自動註解字詞個數較少時，將圖像區塊大小門檻值設定在 300 個像素點以上可得到較佳的正確率，其原因可歸咎於一些較零碎、較小的圖像區塊，在人看起來覺得不具語意於其上，然而在計算時視覺相似度可能與某些圖像區塊非常相似，導致錯誤的標記字詞被設定到這些較零碎的小圖像區塊且具有很大的權重值，造成自動註解時這些錯誤的標記字詞擁有較高的優先權，降低自動圖像註解的正確率，然而若將圖像區塊大小限制設定太大，也可能忽略掉一些較小但卻帶有語意的圖像區塊資訊，像是一些圖像場景中距離較遠的物件就有可能較小，但可能帶有語意於其上，圖 4-9 為圖像區塊大小限制在 300 像素點以上時，系統給定 1~15 個自動註解字詞的正確率與召回率。

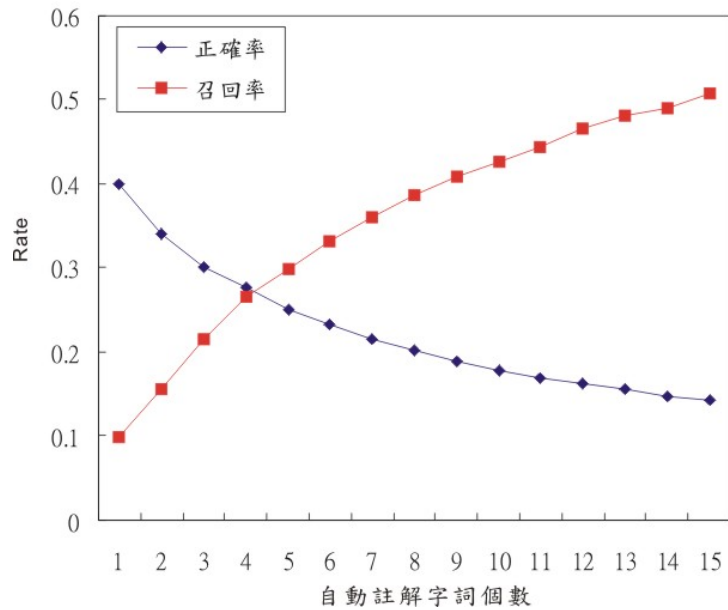


圖 4-9：圖像大小限制 300 像素點時自動註解正確率與召回率。

進行自動圖像註解時，限制所需考慮的圖像區塊大小除了可以增進自動圖像註解的正確率外，對於計算的時間也有很大的幫助，表 4-3 為在不同的圖像區塊大小限制下單張圖像經過圖像區塊標記字詞及權重建構後產生自動圖像註解的平均時間，當限制在大於 600 像素點時比起不限制圖像區塊大小可以節省近 7 倍的時間，因此，在進行自動圖像註解時對圖像區塊大小做適當的條件限制不但可以增進自動圖像註解的正確率且有節省運算時間的優點，這對於本論文所建置的 web 2.0 圖像自動註解及檢索系統是很重要的。

表 4-3：使用不同圖像區塊大小限制產生自動圖像註解所需平均時間。

圖像區塊大小限制(像素)	平均花費時間(秒)
> 0	7.1
> 300	1.5
> 600	1.0
> 900	0.85
> 1200	0.75
> 1500	0.64

4.2.3 有正確人工標記字詞提供對自動圖像註解效能影響

在本小節的實驗中，我們讓測試資料圖像以整批測試的方式輸入到系統中，而先輸入的測試圖像經過系統自動註解並計算其正確率與召回率後，我們將其圖像區塊的人工標記字詞設定到對應的圖像區塊上，然後再將下一張測試圖像輸入系統中進行自動註解，這個動作相當於對本論文所建置的系統持續增加圖像及人工標記字詞；在本實驗中我們將區塊大小限制在 300 像素點以上，系統自動給定 1~15 個自動註解字詞，觀察其正確率與召回率之變化，結果如圖 4-10 與圖 4-11，橫軸為系統給定的自動註解字詞個數。

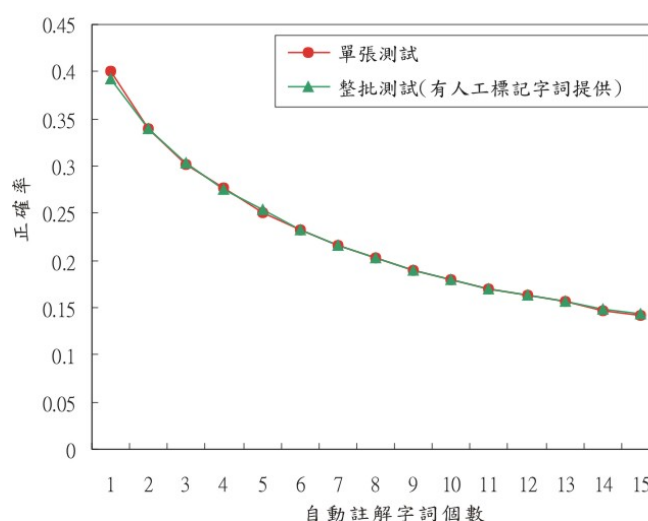


圖 4-10：自動圖像註解正確率(持續提供人工標記字詞)。

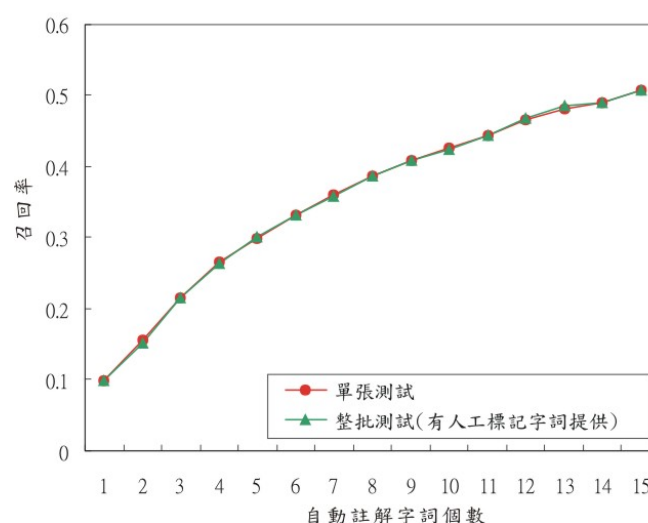


圖 4-11：自動圖像註解召回率(持續提供人工標記字詞)。

從圖 4-10 與圖 4-11 的結果可以看出在持續提供人工標記字詞給新進圖像區塊的情況下，可以維持自動圖像註解系統的正確率與召回率，因此本論文所提的自動圖像註解系統可藉由使用者的回饋資訊，維持穩定的自動圖像註解效能，並增加圖像數量。

4.2.4 使用不同方法綜合圖像區塊資訊產生圖像註解

在本小節中我們比較在 3.3 節中所提的三種綜合圖像區塊資訊產生圖像註解字詞的方法與隨機方式產生圖像註解字詞的效能，由於本論文所建置的系統希望能達到即時產生圖像註解字詞，所以將綜合圖像區塊資訊時的圖像區塊大小限制在 300 個像素點以上，以達到速度上的要求，在評比效能方面使用正確率與召回率，系統給定 1~15 個自動註解字詞，其結果正確率與召回率如圖 4-12、圖 4-13，橫軸為自動註解字詞個數。

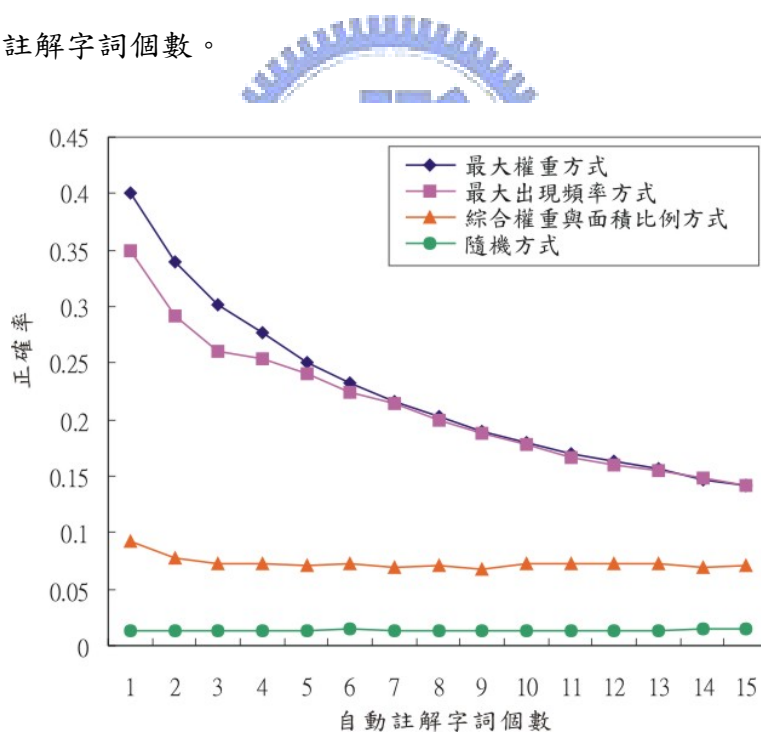


圖 4-12：自動圖像註解正確率(使用不同註解字詞產生方法)。

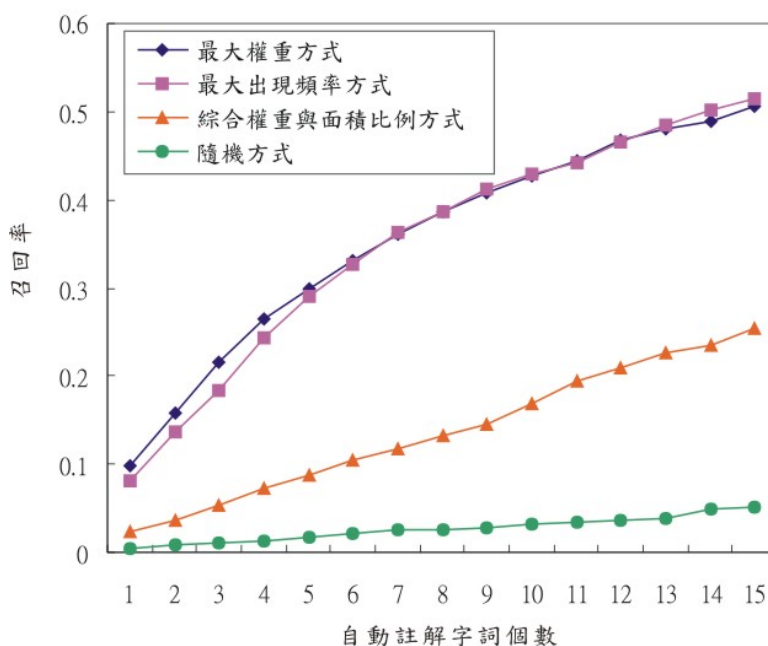


圖 4-13：自動圖像註解召回率(使用不同註解字詞產生方法)。

從圖 4-12 與圖 4-13 的結果來看，可以發現 3.3 節中所提的三種方法種，使用最大權重方式產生圖像註解字詞可以得到最佳的效果；而使用最大出現頻率方式產生的註解字詞在系統給定的自動註解字詞個數少時，不容易命中正確的註解字詞，這表示在不考慮標記字詞權重時，正確的註解字詞無法依據權重取得排序上的優先權，進而降低了自動圖像註解的效能；而從實驗結果看來，綜合標記字詞權重與圖像區塊面積比例產生圖像註解字詞的效果似乎不如預期，其正確率只能達到約 1 成左右，與前兩個方法效能相差甚遠，其原因可歸咎於有些較大的圖像區塊擁有不適當且權重不大的標記字詞，但卻因為其區塊所佔面積比例較大而加重了這些不適當標記字詞在產生圖像註解時的影響力，所以此方法不適合用於產生自動圖像註解。

4.3 自動圖像註解範例

本節展示一些用所提的方法產生的自動圖像註解範例，圖 4-14 與圖 4-15 分別為一些自動圖像註解比較成功的範例與自動圖像註解無法找到適當的註解字詞的範例，在自動註解字詞的欄位中，依照各個不同註解字詞權重排序由系統給

出前 15 個註解字詞，其中粗體字是代表其符合人工標記字詞，加註底線的字詞表示由人工進行評估後，覺得可以適用於該圖的註解字詞；因為人工標記字詞的設定方式是對圖像中較顯著的圖像區塊進行人工標記，但有可能會有些較不明顯的圖像區塊沒有被設定到正確的標記字詞，因此經由自動圖像註解產生的註解字詞有可能會適用於該圖像但卻不符合初始設定的人工標記字詞。

圖像			
人工標記字詞	people sand beach	building sky water tree harbor boats	tree grass people
自動註解字詞	beach ice rocks sand houses building field fox pillar stone ground grass water boats plants	water sky clouds mountain tree building ground snow boats fog wave rocks people shadows	tree <u>plants</u> grass stone rocks <u>forest</u> ruins <u>leaf</u> people horses water boats foals birds flowers
圖像			
人工標記字詞	grass tiger cat water	sky mountain tree clouds	forest fox tree grass shadows
自動註解字詞	<u>tree</u> water <u>plants</u> rocks grass leaf building stone field tiger sculpture deer <u>ground</u> houses ruins	sky mountain water clouds <u>fog</u> <u>snow</u> building tree ground boats wave rocks <u>shadows</u>	water mountain building rocks tree people fox houses tiger elk boats <u>branch</u> stone nest birds

圖 4-14：自動圖像註解較為成功範例。


圖像			
人工標記字詞	sunset sky frost	moose forest antlers	cottage roof houses lawn tree path
自動註解字詞	stone ice rocks field boats building ground sand fox grass houses beach water pillar plants	<u>tree</u> <u>grass</u> <u>plants</u> water rocks building stone field leaf ruins houses street tiger ground <u>deer</u>	<u>building</u> people ice water mountain sky clouds snow arctic fog wave horses mare stone frost
圖像			
人工標記字詞	road people		
自動註解字詞	<u>stone</u> ruins <u>tree</u> <u>grass</u> water field houses rocks mountain building street fox sand nest wood		

圖 4-15：自動圖像註解無法找到適當的註解字詞的範例。

第 5 章

結論與未來展望

5.1 結論



本論文提出了以圖像區塊為基礎的方式，藉由圖像區塊間的視覺相似度計算將人工標記字詞設定到未帶有標記字詞資訊的圖像區塊上，再綜合圖像區塊資訊來達到產生圖像註解字詞的目的，對於缺乏關鍵字資訊的圖像使用本方法可以用一個較客觀的角度自動找出其可能適合的關鍵字，用來描述該張圖像，並可用於以文字為基礎的圖像檢索。

我們利用所提的方法建置了一個以 web 2.0 為基礎的多模式圖像檢索系統；在以內容為基礎的圖像檢索系統中，通常需使用者提供範例圖像做為檢索依據，而對我們建置的系統而言，範例圖像的來源除了可以讓使用者提供外，亦可讓使用者以文字方式在系統中搜尋，同時並提供一個可以讓使用者新增人工標記字詞的機制；系統從網路中收集圖像時，所能得到的資訊通常僅有圖像本身與其檔案名稱，因此我們在以內容為基礎的圖像檢索系統中加入本論文所提方法，對收集來的圖像以客觀的方式產生描述字詞，供使用者使用文字檢索。

5.2 未來展望

在本論文的研究與實驗和應用中，發現有幾個方向是我們還可以繼續改進的重點，在此說明如下：

1. 在綜合圖像區塊資訊產生圖像註解字詞時，在所提方法中及實作的系統中僅使用標記字詞的權重大小作為註解字詞排序依據，然而如此所產生的字詞通常在語意上屬於較低階層級，如花、草等等，如果能綜合各不同標記字詞間的關係，則可能可以產生較抽象且具有較高階語意層級的註解字詞，如慶典、聚會等等。
2. 在本論文所提的架構中，並不侷限於所提的圖像分割方法，如果能有更準確的圖像分割方法能正確切割出圖像中的物件區塊，將有助於自動圖像註解正確率的提升。
3. 在本論文中為達到較佳的即時自動圖像註解速度，僅使用圖像的色彩資訊進行圖像前處理與特徵擷取，如果能夠使用其他更有分別性的特徵資訊則可以達到更加的自動圖像註解正確率。



參考文獻

- 【1】 Heng Tao Shen, Beng Chin Ooi, and Kian-Lee Tan, “Giving meanings to www images,” in MULTIMEDIA ’00: Proceedings of the eighth ACM international conference on Multimedia, New York, NY, USA, 2000, pp. 39-47, ACM.
- 【2】 Rohini K. Srihari, Zhongfei Zhang, and Aibing Rao, “Intelligent indexing and semantic retrieval of multimodal documents,” *Inf. Retr.*, vol. 2, no. 2-3, pp. 245-275, 2000.
- 【3】 A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of early years,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, No. 12, pp.1349-1380, 2000.
- 【4】 徐永煜, “The Study of Mixture Faussian Neural Networks,” 國立交通大學, 資訊工程研究所博士論文, 民國九十三年.
- 【5】 Jia Li and James Z. Wang, “Real-Time Computerized Annotation of Pictures,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, No. 6, pp.985-1002, June. 2008.
- 【6】 R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” *ACM Computing Surveys*, Vol. 40, No. 2, April, 2008.
- 【7】 P. Duygulu, K. Barnard, J.F.G. de Freitas, and D.A. Forsyth, “Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary,” *ECCV*, pp. IV:97-112, 2002.
- 【8】 J. Jeon, V. Lavrenko and R. Manmatha, “Automatic Image Annotation and Retrieval using Cross-Media Relevance Models,” *Proc. ACM SIGIR*, 2003.
- 【9】 Vailaya, A., Figueiredo, M. A. T., Jain, A. K., and Zhang, H.-J., “Image Classification for Content-Based Indexing,” *IEEE Trans. Image Processing*, Jan 2001.
- 【10】 Li, J., Wang, J.Z., “Automatic Linguistic indexing of pictures by a statistical modeling approach ,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2003.
- 【11】 Tim O’Reilly, “What Is Web 2.0,” <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, 2005.
- 【12】 Uche Ogbuji, “Real Web 2.0: Bookmarks? Tagging? Delicious!,” <http://www.ibm.com/developerworks/xml/library/wa-realweb1/index.html>, 2006.

- 【13】 John C. Russ, *The Image Processing Handbook*, Ron Powers, 3 edition, 1998.
- 【14】 Pietro Perona and Jitendra Malik, “Scale-space and edge detection using anisotropic diffusion ,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629-639, July 1990.
- 【15】 <http://www.brucelindbloom.com/>

