

行政院國家科學委員會專題研究計畫成果報告

智慧型多媒體資訊處理系統的研究(二)

The Study of Intelligent Multimedia Information Processing System(II)

計畫編號：NSC 89-2213-E-009-094

執行期限：88年8月1日至89年7月31日

主持人：傅心家 交通大學資訊工程系

一、中文摘要

隨著電腦速度的提昇以及網路的普及，如何在浩瀚的網路多媒體資訊中快速及正確找尋想要的多媒體資訊，已是一個非常重要的課題。目前已有的影像搜尋系統均著重於如何有效的在已知資料庫中搜尋靜態影像，其比對之特徵均以整個影像為基礎。而對於環球網際網路（WWW）上的動態影像的搜尋目前少有成熟的研究出現。本計畫提出了一個以人工智慧及類神經網路為基礎的智慧型多媒體資訊處理系統（Intelligent Multimedia Information Processing Systems, IMIPS）。

這個系統包括了(一) 網際網路視訊收集器，(二) 視訊切割及關鍵影像擷取器，(三) 影像分類及檢索器三部分。除了利用傳統的網際網路通訊影像處理及辨認等技術外，特別的是，我們為影像分類及檢索器提出了一個“EM base Multi-Instance Learning”演算法，使得搜尋工具不但可以有效的將搜尋到的影像分類，並可以將使用者個人搜尋喜好，以學習的方式來調適(Adaptive)，達到個人理想的搜尋工具。

目前實做完成的系統我們在網際網路搜尋到三千多個視訊檔。而在分類檢索方面，我們做了五種分類（animal、human、sky、star、fire）其平均正確率達到69%。

關鍵詞：智慧型多媒體資訊處理系統、類神經網路、動態影像搜尋引擎

Abstract

Retrieving multimedia information efficiently among the world wide web becomes more and more demand. Many images search engines have been developed, so far their retrieved image were based on the image database constructed by themselves, and the developed of video system are seldom to be announced. In this project, We intend to design an intelligent multimedia information processing system (IMIPS) based on neural networks.

The system include three modules: (1) world wide web video collector, (2) video segmentation and key frame selector, (3) image classifier and retriever. We propose an "EM base Multi-Instance Learning" algorithm to classify the images efficiently and adapt the search engine to a special user style.

The experimental results show that the correct rate for image retrieval can reach 69% for five classes problem.

Keywords: IMIPS, neural networks, video search engine

二、緣由與目的

網際網路資料及數位圖書館上之資訊查詢，目前多以文字搜尋的方式在進行[2][3]。然近年多媒體資料在網際網路上普遍置，對這類型資料（包括影像（Images）、

音訊(Audio)、視訊(Video)等)查詢需求也日益迫切[4]。目前這方面現有之查詢工具[1]也漸有所見。

在網路上看到的若干搜尋工具[1]所用的方法概言之是在影像、音訊、視訊檔中附加一段簡短的介紹文字及若干關鍵字(key words),而在搜尋時使用者仍需將所需搜尋的標的以關鍵字描述,而搜尋器將使用者之關鍵字與之比對[5],再將相符的網址及簡短文字及小圖(ICON)列出供使用者參考。這類工具確能提供相當程度的方便以及實用價值。但是我們知道,一張影像有時候很難用文字來描述,何況用關鍵字。因此有時候這些搜尋到的結果常常很難令使用者滿意。有鑑於此,本研究希望能讓使用者以圖像來描述他/她個人的搜尋概念,在網際網路上搜尋所需視訊資料。為此,我們將前述的原理及技術,實作在一部以 Pentium II 及 Windows NT 為核心的伺服器(Server)上建立了一套以網站的型態供使用者利用的智慧型多媒體資訊處理系統(Intelligent Multimedia Information Processing Systems, IMIPS),使用者可利用常用如 Netscape 或 IE 等瀏覽器(Browser)進入本站,來搜尋他/她們想要找的多媒體資訊。

IMIPS 包括了(一)網際網路視訊收集器,(二)視訊切割及關鍵影像擷取器,(三)影像分類及檢索器三部分。我們除了利用傳統的網際網路通訊影像處理及辨認等技術外,並利用人工智慧及神經網路技術來研製一快速及智慧型搜尋工具。特別的是,我們為了影像分類及檢索器設計了一個"EM based Multi-Instance Learning"演算法,使得我們的搜尋工具不但可以有效的將搜尋到的影像分類,並可以將使用者個人搜尋喜好,以學習的方式來調適(Adaptive),達

到個人理想的搜尋工具。

三、結果與討論

本年度我們完成了以網際網路為搜尋資料庫的智慧型多媒體資訊處理系統[5]。茲就此系統的組織架構及工作原理詳述如下。

(一) 網際網路視訊收集器

網際網路視訊收集器(Internet Robot)最重要的部份就是如何搜尋資料及如何管理資料,在這裡將介紹我們的漫游器是如何找尋資料,以及如何利用伺服器中的資料庫有效管理所需要的資訊。本系統的視訊收集器主要分成二個部份:漫游器(spider)及維護器(Maintainer)。Spider 的主要工作是負責尋找網路上還未拜訪過的網頁,而 Maintainer 則是負責監督已拜訪過的網頁,若拜訪過的某個網頁內容有變動,則對該網頁重新進行分析,以確保資料庫中的資料能即時更新。我們在 Spider 的設計中加入了參照記數器(Reference counter)來記錄網頁被參考的次數。使用參照記數器可以優先拜訪比較大的站或很多人常看的網頁,參照記數器愈大,代表該網頁被很多其他不同網頁所參考,不但容易連得上,也有較豐富的內容及額外資訊。在網頁的維護方面,我們給每一個網頁不同的更新週期。我們在每次更新時,加入統計更新次數的動作。在定期修改過幾次後,可以知道哪些網站更新次數很頻繁,哪些不常更動。依據這項統計資料,便可對不同的更新頻率,排以不同的更新週期。

(二) 視訊切割及關鍵影像擷取器

在本研究中,我們提出了一個以偵測『突然的場景切換』(abrupt scene change)及『特別的編輯效果』(例如影片中的淡入/淡出)來抽取關鍵片段(key frames)的演算

法，作為視訊摘要的技術。場景的切換方式可分為兩種：第一種稱為直接的場景切換 (Direct Scene Change)，這種場景切換在一般的視訊資料中最常發生。另一種則為淡入/淡出 (fade in/out)型的場景切換，這種場景切換技術多應用在 KTV 等影片編輯效果上。我們所提出的偵測場景切換的演算法主要是針對在 MPEG 位元串中的 DCT DC 係數與移動向量作一個統計的處理。我們先由一些 I-、P-、B-frames 得到候選的場景切換點，最後再由 I-frame 決定真正的場景切換的位置。

(三) 影像特徵擷取及分類器

在影像分類與檢索的範疇中，很多是使用顏色統計圖 (color histogram) 的方法 [7]。顏色統計圖是計算整張圖的顏色分佈，所以圖中某顏色出現的位置與方向並不會影響該張圖的顏色分佈，但是就因為沒有空間的關係，所以在識別上有不足之處。目前有很多的影像搜尋系統就是用整張圖或是使用者所點選區域的顏色分佈與紋理 (Texture) 來當檢索的依據，如 QBIC [8]、NETRA [9]。也有人把空間的關係加入特徵中，如在 [10] 中的方法，先建立許多模版，例如下雪的山會被訓練成「藍色色塊下面有白色色塊，再下面有棕色的色塊」。

上述的方法都必須在一張影像之中，精確地標示出使用者所想要的部分，除了造成使用上的不便，也容易因為標示錯誤而得到錯誤的結果。因此在 [11] 中採用了以多事例學習 (Multiple-Instance Learning) 為主要架構，並以互異密度 (Diverse Density) 作為學習的指標，用斜率陡降法 (Gradient Decent method) 來最大化互異密度的方法來作影像分類與檢索。這個方法的優點在於使用者可不用精確地標示出使用者於影像中所想要的部分，使用者只要標示出那些影

像是要的，那些是不要的，則系統會根據使用者的選擇學習找出使用者所要的影像。在本研究中，我們提出了以 EM 為基礎的多事例學習演算法 (EM base Multiple-Instance Learning Algorithm) 來作為訓練系統影像分類與檢索的演算法。

四、計畫成果自評

本系統實作了一個架構於網際網路上的視訊搜尋系統。本系統包含了三個子系統：網際網路漫遊器、視訊切割及關鍵影像擷取器和影像特徵擷取及分類器。

在網際網路漫遊器部分，我們設計出了一個智慧型網際網路視訊收集器。此網際網路漫遊器可自動學習及尋找網際網路上受歡迎的網頁，並可根據網頁更新的頻率，給予資料庫中不同網頁不同的更新週期。在視訊切割及關鍵影像擷取器部分，我們針對直接的場景切換與淡入/淡出型的場景切換，設計出不同的視訊切割演算法。在影像特徵擷取及分類器部分，我們設計了一個以 EM 為基礎的多事例學習演算法來作為影像特徵擷取及分類。在本系統中預先建立了五個樣板類別，分別為 Human, Star, Sky, Animal 跟 Fire。我們的實驗結果如下表所示。由結果可看出，本系統確能有效的將使用者心中的視訊從網際網路中找到。在使用者介面方面，由於本系統是架構於網際網路上的視訊搜尋系統，因此使用者可直接用網頁瀏覽器 (如 Internet Explorer 或 Netscape) 操作本系統。除此之外，使用者可利用使用者介面與系統互動，訓練本系統學習使用者所欲尋找的類別，進而使得系統所搜尋的視訊結果更符合使用者心目中的要求。由實驗結果可知本系統確能有效的從網際網路中尋找到使用者心中的視訊。

		Total Number of retrieved frames					
Correctness		10	20	30	40	50	Hit rate
Template name	[1] Human	10	20	30	39	48	96%
	Star	10	15	18	22	23	46%
	Sky	7	13	19	26	29	58%
	Animal	10	20	30	35	40	80%
	Fire	9	13	n/a	n/a	n/a	65%
Hit rate		92%	81%	81%	72%	70%	

五、參考文獻

- [1] AltaVista, Available <http://www.altavista.com>.
- [2] Wu, S. and U. Manber, "Fast Text Searching Allowing Errors," *Communications of ACM*, Vol. 35, Oct., pp. 83-91, 1992.
- [3] Manber U. and S. Wu, "Glimpse: A Tool for Search Through Entire File Systems," *USENIX 94 Winter Conference*, Jan., 1994.
- [4] 曾元顯。「多媒體資訊檢索技術之探討」。21世紀資訊科學與技術國際學術研討會。台北市：世界新聞傳播學院圖書資訊學系，民85年11月7 9日，頁281-298。
- [5] Natalie Oakes Sturr, "WAIS: an Internet Tool for Full-text Indexing," *Computers in Libraries* June, pp. 52-54, 1995
- [6] Martijn Koster, "World Wide Web Robots, Wanderers, and Spiders," <http://www.webcrawler.com/mak/projects/robots/robots.html>.
- [7] Rangachar Kasturi, Susan H. Strayer, "An Evaluation of Color Histogram Based Methods in Video Indexing," *Research Progress Report CSE-96-053*, October, 1995.
- [8] M. Flickner, and et al., "Query by image and video content: The QBIC System," *IEEE Computer*, 28:23-32, 1995.
- [9] W.Y.Ma, *NETRA: A Toolbox for Navigating Large Image Databases*, Ph.D. Dissertation, Dept. of Electrical and Computer Engineering, University of California at Santa Barbara, June 1997.
- [10] J. Smith and S. Chang, "VisualSEEK: a fully automated content-based image query system," In *Proc. ACM International Conference on Multimedia*, Morgan Kaufmann, 1996.
- [11] O. Maron and A. LakshmiRatan, "Multiple-Instance Learning for Natural Scene Classification," *Machine Learning: Proceedings of the 15th International Conference*, 1998.