

基於音訊與文字訊息之視訊摘要

孫惠民
長榮大學資訊管理學系
副教授
prince@mail.cjcu.edu.tw

黃志偉
長榮大學資訊管理學系
學生
joke-168@hotmail.com

摘要

本研究係以主觀視訊品質評估為基礎，探討視訊摘要(Video Abstraction)技術中的動態影像視訊摘錄(Video Skim)，以三種不同的擷取序列片段長度(Skimmed Segment)，分別以 2.5 秒、5 秒、與 10 秒，作為擷取序列片段的固定單位長度，以及三種不同的視訊摘錄總長度(Total Skimmed Length)，分別以原始影像檔的 5%、10%、與 15%，觀察受測者對上述不同組合，哪一組最能代表原始視訊內容。依據這個實驗結果應用於動態影像視訊摘錄(Video Skim)中的技術，以上述的最佳組合，提出摘錄演算法來自動擷取視訊片段，達到有效率內容存取使得摘錄視訊具有原始影像之代表性與建立出自動擷取視訊摘要之系統工具。

關鍵詞：視訊摘要、動態影像視訊摘錄、主觀評估法

Abstract

Based on the subjective measurement method, we study the video skim in the video abstraction field. We probe the video skimmed combination by the total skimmed length and the skimmed segment. The total skimmed length includes three levels which are 5%, 10%, and 15% for total length, while the skimmed segment includes three levels which are 2.5, 5, and 10 seconds for a separated unit. We will propose a video skimmed algorithm based on the optimal combination of above. Our goal is to skim a video segment such that it can represent the original video content.

Keywords: Video Abstraction, Video Skim, Subjective Measurement

1. 前言

近年來數位影像革命已經在各式各樣的電腦設備中快速的發展，例如：增強電腦處理的效能、更大容量且便宜的儲存記憶體、以及更加快速的網際網路。這個革命帶來了許多新的應用，因此，研究以準確性(Effectiveness)與執

行速度(Efficiency)的影像擷取、歸類、編入目錄與索引為目的，將可增加儲存影像的可用性。在這研究範圍中，如何達到迅速的瀏覽大量的影像資料以及如何達到有效率的內容存取和敘述是重要的議題。針對這個問題，使得視訊摘要(Video Abstraction)技術的出現，並在近幾年吸引了許多學者對於此研究的興趣。

視訊摘要(Video Abstraction)技術應用於 MPEG-4 技術中，將可以使 MPEG-4 的影音更加的減少影片原始的容量，並保留影片的重點，使影片更容易流通於現行的網際網路中，不只如此，還有助於視訊管理的應用。視訊摘要(Video Abstraction)技術目前有兩個主要探討的議題，一是靜態影像的關鍵畫格(Keyframes)，另一是動態影像的視訊摘錄(Video Skim)。靜態影像的關鍵畫格(Keyframes)由最初的視訊來源抽出最重要的圖片來收集組成視訊摘錄；動態影像的視訊摘錄(Video Skim)則由原始的影片抽出部分影像片段，包含聲音或文字所收集組成的視訊摘錄。

本研究係以主觀視訊品質評估為基礎，探討視訊摘要(Video Abstraction)技術中的動態影像視訊摘錄(Video Skim)，其研究目的為以下幾點：

1. 探討在不同的電影視訊內容(Video Content)類型下，以三種不同的擷取序列片段長度(Skimmed Segment)，分別以 2.5 秒、5 秒、與 10 秒，作為擷取序列片段的固定單位長度，以及三種不同的視訊摘錄總長度(Total Skimmed Length)，分別以原始影像檔的 5%、10%、與 15%，觀察受測者對上述不同組合，哪一組最能代表原始視訊內容。

2. 應用動態影像視訊摘錄(Video Skim)中的技術，提出的摘錄演算法來自動擷取視訊片段，達到有效率內容存取並據有原始影像之代表性。

2. 文獻探討

2.1 視訊摘要(VIDEO ABSTRACTION)技術

視訊摘要(Video Abstraction)技術為一種將原始影片擷取製作出摘要短片的一種機制，一段好的視訊摘要，將能使觀看者在最小時間裡，獲得關於擷取出的短片所要表達的最大資訊，視訊摘要(Video Abstraction)技術分為兩大類，一為靜態影像的關鍵畫格(Keyframes)為基礎之影像摘要；另一為動態影像的視訊摘錄(Video Skim)為基礎之影像摘要[1]。

靜態影像的關鍵畫格也可以稱為代表性畫格(Representative Frames)，或稱為 R 畫格(R-frames)，靜止的圖片摘錄(Still-image Abstracts)或是靜態的故事板(Static story-board)，由最初的影像來源抽出最重要的圖片來收集組成。依據[1]Keyframes 擷取技術的架構包含了擷取出 Keyframes 所需設置的大小(Number of R-frames)、基礎單位(Unit)、表述的範圍(Representation)以及相關技術(Mechanism)的方法。

動態影像的視訊摘錄(Video Skim)也可以稱為移動影像摘錄(Moving-image Abstract)，移動的分鏡腳本(Moving Storyboard)，或是摘要序列(Summary Sequence)。從原始的影片抽出部分影像片段包含聲音所收集組成的，這些片段再經由修剪和轉場效果結合在一起，可算是屬於影片本身的剪輯，且表示原始影片的時間將經由動態影像視訊摘錄後變的更為簡短的[1]。動態影像的視訊摘錄(Video Skim)被分類為兩個種類：一是以摘要為目的(Summary-oriented Skim)例如電影預告片，另一是以影片中最強調的片段為目的(Highlight-oriented Skim)例如運動影片中的十大好球[2]。依據[1]產生 Video Skim 擷取架構需包含五個步驟，程序(Process)、設置的大小(Skim length)、觀點(Perspective)、相關技術(Mechanism)、與特徵(Features)。

2.2 影片內容的影響

不確定性(Uncertainty)與模稜兩可性(Equivocality)被定義為影響資訊處理的兩大力量[2]。而影像摘要技術(Video Abstraction)所擷取出的影片應該著重降低影片的不確定性(Uncertainty)與模稜兩可性(Equivocality)，使影片能完整表達所提供的資訊量與豐富性。

不確定性代表明確的資訊缺乏，所以當資訊增加，不確定性便減少。Jianping Fan[3]等四位學者提議一種 Cost-Sensitive 的視訊分類技術處理概念不確定(Concept Uncertainty)超時的問題。

模稜兩可性表示資訊模糊，多樣性和衝突的發生關於組織的情況，當資訊充滿著不確定性、高模稜兩可性的時候，必須配合能夠傳遞豐富資訊的媒體，資訊才能有效地被處理。高模稜兩可性表示混亂和理解的不足。舉例：觀看低畫質模糊的影片，雖能經由模糊影像以及聲音來簡略理解影片所表達的資訊，但卻無法完整的理解影片中較細微的人物動作、表情所要傳達的資訊。

2.3 視訊品質測方法之相關研究

本研究所採取的主觀量測方法是參照 ITU-R Recommendation 500[4]所定義出來的標準，標準內容包含觀測者、觀測條件、測試場地的選擇、評估程序、與分析方法，主要的目的是要透過試測者進行觀看實驗測試進而視訊品質的好壞，而常見的主觀視訊品質測方法有以下四種，分別為雙重受測連續性品質評分法(Double Stimulus Continuous Quality Scale, DSCQS) [5][6][7][8]，雙重受測損壞評分法(Double Stimulus Impairment Scale, DSIS) [8]，單一受測連續性品質評估法(Single Stimulus Continuous Quality Evaluation, SSCQE) [6][7][8][9][10]，同步雙重受測連續性評估法(Simultaneous Double Stimulus for Continuous Evaluation, SDSCE) [4]。

DSCQS 方法是採用兩個顯像畫面來讓受測者進行視訊品質評估，在視訊序列裡有不同的視訊內容類別，並依照不同的品質參數進行處理，實驗過程讓受測者觀看多組簡短的視訊序列，包含原始的視訊序列與經過處理的視訊序列，兩者是以隨機方式播放，每一個視訊序列播放十秒鐘後會有一小段時間給予受測者進行品質評估(from Bad to Excellent)。

DSIS 與 DSCQS 所使用的方式類似，都是使用兩個顯像畫面播放，每個視訊序列播放時間以十秒鐘為主，兩者不同的差別在於 DSIS 是將原始視訊序列與經過處理過的視訊序列兩者以先後順序播放，並且不重複，撥放完一組顯像畫面後會有一小段時間給予受測者進行五等級品質評估，評估優劣從“very annoying”到“imperceptible”。

SSCQE 方式是採取單一顯像畫面進行實驗設計，每個顯像畫面播放時間大約二十分鐘到三十分鐘之間並且可以連續播放給予受測者觀看，在觀看期間受測者必須在規定的間隔時間內進行品質評估，品質評估方式與 DSCQS 相同(from Bad to Excellent)。

SDSCE 方式是採取雙顯像畫面進行實驗設計，原始的視訊序列與經過處理過的視訊序列兩者是相鄰，播放時間五-十分鐘，播放過程為同步但不隨機更換先後順序，受測者在播放完後進行品質評估 (from Bad to Excellent)。

在客觀的品質量測方法裡，常見的量測方法有訊號品質與雜訊的比例的計算，例如尖峰訊號雜訊比 (Peak Signal to Noise Ratio, PSNR)，或是使用原始視訊與處理過的視訊擷取特徵值後相互做比較，或是瑞士洛桑聯邦理工學院所發展的彩色移動影像品質規範，此方法考慮了 HVS 幾個特徵值作為參數影響：多頻道架構、顏色對比敏感性和遮罩等，然而其模式具有高度複雜性且無法時實做在即時環境。

Aeluri et al.[11]以客觀的視訊品質評估方法，分析移動特徵(從無移動到高移動)、編碼器、畫面播放率以及資料傳輸率這四個參數組合，其客觀的視訊品質評估則依平均平方差作為評估測量。

Yao et al.[12]以客觀的視訊品質評估方法，分析三個影響視覺感受的重要參數，分別為視覺遮罩錯誤(Visual Masking Error)、模糊的扭曲(Blurring Distortion)以及結構扭曲(Structural Distortion)，並以主觀的視訊品質評估方法，驗證其相關性。

3. 研究方法

3.1 研究架構

回顧視訊摘要相關文獻發現先前學者曾使用 2 秒、3 秒甚至 5 秒的視訊摘要擷取序列長度進行動態影像的視訊摘錄[13][14]，但皆未能證實視訊摘要擷取序列長度該擷取多長才是最適當的；另一方面，回顧先前學者進行視訊摘要總長度的選擇，通常以擷取原始視訊片段的 10 分之一為標準[15][16]，但同樣皆未提出證實選取多少視訊摘要總長度，才能保留足夠影片資訊使影片具代表性，使觀看者能透過動態影像的視訊摘錄快速理解視訊內容。本研究架構如圖 1 所示，基於探討在不同的電影視訊內容(Video Content)，變化視訊摘要擷取序列長度(Skimmed Segment)與視訊摘要總長度(Total Skimmed Length)下，觀察觀看者對其視訊摘錄代表性之影響。

3.1.1 電影視訊內容(Movie Video Content)

本研究所使用的電影視訊內容為劇情類影片，藉由劇情類影片有其故事性的特性，進行動態影像的視訊摘錄，將研究劇情電影經視訊摘錄後是否會影響原始視訊內容所要傳達給觀看者理解的故事內容，使其間接影響到原始電影所具有的代表性，例如剪輯後可能造成影片畫面播放的不連貫、音訊模糊或字幕詞意不清的情況發生，而讓觀看者對影片內容產生誤解，這便是降低電影所摘錄的代表性。

3.1.2 視訊摘要片段長度(Skimmed Segment)

視訊摘要片段長度意指進行影片剪輯時所設定間隔的剪輯秒數。第二章文獻探討中得知動態影像的視訊摘錄在視訊摘要擷取序列上並沒有一定的秒數規定，因此我們希望透過我們的演算法將電影視訊依不同的視訊摘要擷取序列長度分別為：2.5 秒、5 秒以及 10 秒，重新編碼後再組成新的視訊摘錄，探討不同的視訊摘要擷取序列是否會影響到電影視訊摘錄的代表性。

3.1.3 視訊摘要總長度(Total Skimmed Length)

視訊摘要總長度意指進行影片剪輯時所設定的影片擷取比率，舉例原為 60 分鐘的電影影片經由設定視訊摘要總長度為 10%，視訊摘錄後將會產生 6 分鐘的影片。在文獻探討中得知視訊摘要總長度常見為 10%，但如同上小節所述一樣，並沒有硬性規定一定的視訊摘要總長度，也未證實為何要取 10%最為適當，因此本研究將視訊摘要總長度分為：5%、10%以及 15%三種，藉由實驗變化不同的視訊摘要總長度證實是否會影響到電影視訊摘錄的代表性。

3.1.4 視訊摘錄代表性

視訊摘錄代表性意指視訊內容不確定性低與理解程度高，摘錄代表性越高表示觀看者越能理解原始影片所要傳得的訊息內容。在文獻探討影片內容的影響中，發現由媒體豐富理論所延伸出的不確定性(Uncertainty)構面，會影響到視訊內容所要傳達的資訊，不確定性高資訊便會減少，另外本研究提出視訊理解程度構面，理解程度越高表示越能讓觀看者了解原始電影所傳達的資訊，綜合兩構面本研究將在下一節以實驗設計方式證實不確定性與理解程度兩項構面足以代表視訊摘錄的代表性。

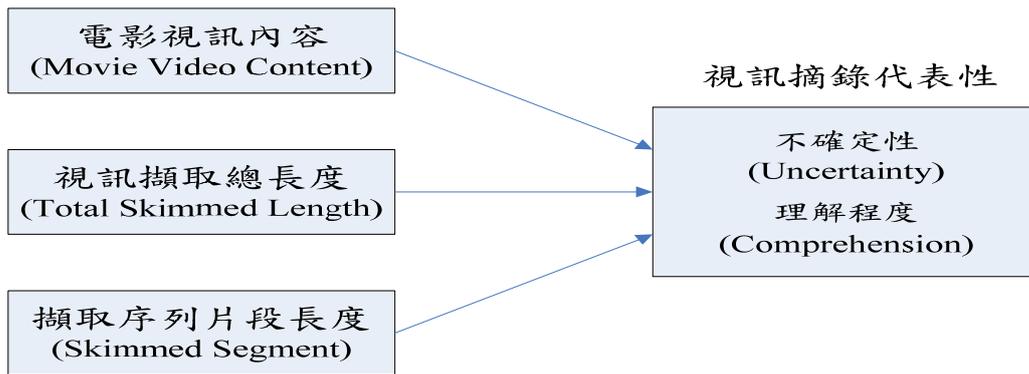


圖 1 研究擷取架構圖

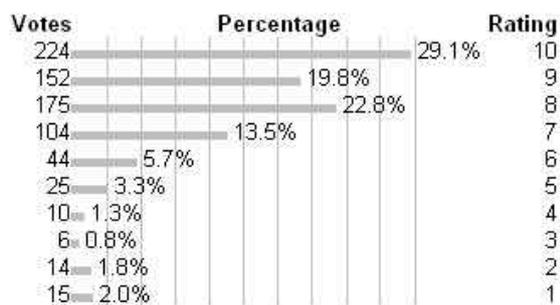
3.2 實驗設計與衡量

為了瞭解不同的電影視訊內容 (Video Content)，在變化視訊摘要擷取序列長度 (Skimmed Segment) 與視訊摘要總長度 (Total Skimmed Length) 下，觀察觀看者對視訊代表性之影響。本研究依據主觀的視訊評估方法制定以下實驗步驟：

3.2.1 實驗視訊選取

基於本實驗以電影視訊為研究前提，電影常透過影音光碟 (Video Compact Disc, VCD) 或數位多功能影音光碟 (Digital Versatile Disc, DVD) 的方式呈現給社會大眾觀看，而 DVD 比 VCD 又具有高密度、高畫質、高音質、高兼容性和高可靠性等特點，因此本研究決定以 DVD 電影光碟為研究格式。

本研究所挑選之劇情片電影「不能說的秘密」，為美國網路電影資料庫 (Internet Movie Database, IMDB) 由 769 位影迷與電影相關工作者投票評分為 7.7 分的電影 [18]，圖 2 為 769 位投票者所投票分數的條狀統計圖，由圖中可看出超過二分之一的投票者皆給予 8 分以上的投票分數，此外此部電影也榮獲第 44 屆金馬獎年度台灣傑出電影獎，因此本研究以此電影為研究範本。表 1 為此視訊樣本的內容描述。



Arithmetic mean = 8.1. Median = 8

圖 2 IMDB 評分者投票分數統計圖

表 1 視訊內容描述

影片名稱	內容概述
不能說的秘密 (The Secret)	描述高中生葉湘倫出身單親家庭，並在父親任教學校就讀。而在父親的耳濡目染下，他熱愛音樂並且琴藝過人。某日，班上來了一位同樣喜愛彈琴的新同學路小雨，投緣的兩人形影不離，情感也日漸加溫，然而小雨總是相當神祕，還常彈奏一首未曾問世，但優美動聽的曲子。而每當小倫想多了解小雨一些，她常欲言又止，只都推說是秘密。但有一天，在一場誤會發生後，小雨再也沒來上過課，思念小雨又一頭霧水的小倫決心要找出這個不能說的秘密。

3.2.2 實驗設計方法挑選

基於本研究為兩因子的重複測量研究，因此本實驗將使用「受測者間與受測者內的混合設計」進行實驗設計，針對視訊摘要擷取序列長度與視訊摘要總長度兩項因子進行混合設計，透過混合設計將本研究實驗分為兩部份進行測試，如表 2 與表 3 所示，表 2 實驗一針對視訊摘要擷取序列長度找出多長的擷取秒數才是最適當的摘錄選擇，將表 2 中橫向 A1、A2、A3 三組各別進行實驗；而表 3 實驗二針對視訊摘要總長度找出多長的擷取率才會是最適當的視訊摘錄選擇，將表 3 中縱向 B1、B2、B3 三組各別進行實驗。

表 2 視訊摘要擷取序列長度 Test Condition 表

實驗一			
視訊摘要 總長度 (Total Skimmed length)	視訊摘要擷取序列長度 (Skimmed Segment)		
	2.5sec	5sec	10sec
5%	A1	A1	A1
10%	A2	A2	A2
15%	A3	A3	A3

表 3 視訊摘要總長度 Test Condition 表

實驗二			
視訊摘要 總長度 (Total Skimmed length)	視訊摘要擷取序列長度 (Skimmed Segment)		
	2.5sec	5sec	10sec
5%	B1	B2	B3
10%	B1	B2	B3
15%	B1	B2	B3

3.2.3 進行動態影像的視訊摘錄

經過第一階段對實驗視訊選取後，第二階段即對所選取出的視訊依表 2 與表 3 實驗需求，對於不同視訊摘要擷取序列長度與視訊摘要總長度兩項因素進行 Video Skim，表 4 中可知經摘錄後共會產生 9 個不同組合的 Video Skim，在表中分別以 C1、C2、C3、C4、C5、C6、C7、C8、C9 表示，舉例說明 C1 為 2.5 秒視訊摘要擷取序列長度與 5% 視訊摘要總長度所摘錄出的 Video Skim，其它視訊摘錄樣本以此類推。

表 4 Video Skim 實驗樣本分配表

視訊摘要 總長度 (Total Skimmed)	視訊摘要擷取序列長度 (Skimmed Segment)		
	2.5sec	5sec	10sec

length)			
5%	C1	C2	C4
10%	C4	C5	C6
15%	C7	C8	C9

根據研究架構我們提出了下列假說：

假說 1：不同的視訊摘錄總長度(分別取原始視訊的 5%、10%及 15%)對於視訊摘錄代表性有顯著的差異。

假說 2：不同的擷取片段長度(分別取每一片段為 2.5 秒、5 秒、及 10 秒)對於視訊摘錄代表性有顯著的差異。

假說 3：擷取片段長度(分別取每一片段為 2.5 秒、5 秒、及 10 秒)與視訊摘錄總長度(分別取原始視訊的 5%、10%及 15%)，兩者之間對於視訊摘錄代表性無顯著的差異。

假說 4：不同的時間下所作的測試結果，對於視訊摘錄代表性無顯著的差異。

假說 5：不同的視訊摘錄組合(Video Skim C1~C9)，對於視訊摘錄代表性有顯著的差異。

本實驗設計出一套進行 Video Skim 的工具，系統介面如圖 3 所示，透過圖 3 動態影像視訊摘錄剪輯工具，使用者可挑選視訊摘要擷取序列長度與視訊摘要總長度依自己所需自動剪輯出符合本實驗設計 Video Skim 影片；另一方面使用者也可在視訊摘錄後依自己喜愛於下拉式選單選取 DVD 檔案進行修改 Video Skim 的動作，最後使用者可透過程式中開啟多媒體檔選項進行 Video Skim 的觀看如圖 4 所示。

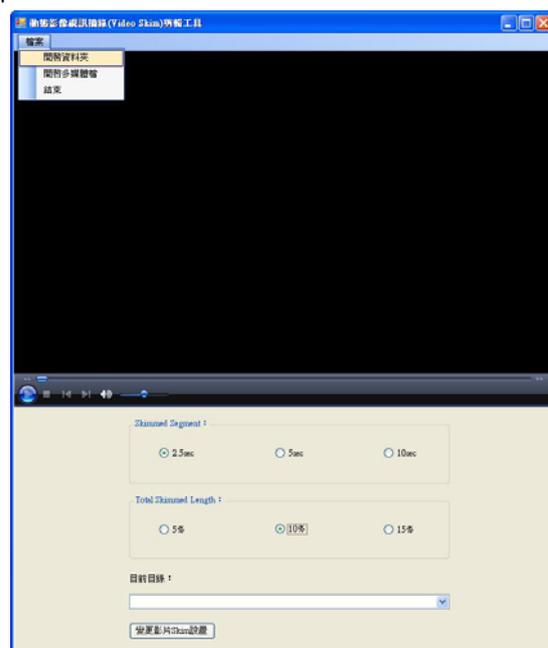


圖 3 Video Skim 剪輯工具圖



圖 4 多媒體檔案播放

動態影像視訊摘錄剪輯工具系統流程圖如圖 5 所示。以下簡述圖 5 中系統元件之功能：

1. 設定 SS 與 TSL 參數：負責設定使用者所挑選之視訊摘要擷取序列長度 (Skimmed Segment, SS) 與視訊摘要總長度 (Total Skimmed length, TSL)。
2. 開啟資料夾：選擇進行動態影像視訊摘錄的原始影片位置。
3. 選擇欲修改檔案：Video Skim 完成後，可透過系統介面中下拉式選單，選取欲修改視訊檔案位置。
4. 開啟多媒體檔：使用者可透過這個功能觀看擷取完成的 Video Skim；也可選取其它多媒體檔案進行觀看。



圖 5 動態影像視訊摘錄剪輯工具系統流程圖

3.2.4 實驗視訊摘錄演算法

本實驗摘錄演算法如圖 6 所示，首先演算法步驟一為輸入影音檔，步驟二為設定擷取片段長度與總長度大小，步驟三進行判斷每次皆先由片頭開始進行剪輯直到影片結束，且 GOP(Group of Picture) 累計之數量秒數不可大於 L，步驟四以固定週期的時間進行視訊摘錄，步驟五結束迴圈。舉例說明，假設 SS 為 5 秒、TSL 為 10%、L 為 100 分鐘，則套進演算法中 Skim SS length 應為 50 秒，表示演算法會以每隔 50 秒取當中的 5 秒為視訊摘錄的片斷。在演算法中以 1GOP 為單位，1GOP 約為

0.5 秒，Count 變數的加總方法是以判斷影音檔中視訊編碼為 0xBB 表示抓取到 1GOP 視訊串流。表 5 為摘錄演算法之參數表。

STEP1. 輸入影音檔 (Video Pack)
STEP2. 設定擷取片段長度 (SS) 與總長度 (TSL)
STEP3. while (Count * GOP < L)
STEP4. Skim SS length for a period of $\frac{SS}{TSL}$
STEP5. end while

圖 6 Video Skim 實驗摘錄演算法

表 5 摘錄演算法參數表

參數表

影音檔總長度	L 秒
擷取片段長度	SS 秒
擷取總長度的百分比	TSL 百分比
1GOP	0.5 秒
計算累積 GOP 之數量變數	Count

3.2.5 實驗環境描述

在測試環境(Test Environment), 整個觀測條件中, 本研究盡可能符合 ITU-R Rec. BT500[4]所描述的觀測條件, 並根據一般使用者需求對實驗室進行調整, 本實驗以一間無人且安靜的電腦實驗室中進行測試, 實驗設備電腦規格與屬性如表 6 所示。

表 6 實驗電腦設備規格

實驗電腦設備
作業系統: Windows XP Professional Version2002 Service Pack2
中央處理器(CPU): AMD Turion(tm)64 Mobile Technology 803MHz
記憶體: 2GB

實驗測試材料(Test Material)為 3.2.2 所挑選針對不同視訊摘要擷取序列長度與視訊摘要總長度所擷取出 9 個 Video Skim, 實驗中為瞭解多少視訊摘要擷取序列長度與多少視訊摘要總長度最為適合當做 Video Skim 的標準, 因此本研究將會使用在 3.2.2 實驗設計方法挑選這章節所說混合設計方法將實驗分為兩部分進行測試, 實驗一為針對視訊摘要擷取序列長度找出多長的擷取秒數才是最適當的摘錄選擇, 由表 2 中以橫向 A1、A2、A3 三組為實驗一分別進行實驗; 實驗二針對視訊摘要總長度找出多長的擷取率才會是最適視訊摘錄選擇, 表 3 中則以縱向 B1、B2、B3 三組為實驗二分別進行實驗。

針對所有實驗測試(Test Procedure), 本研究將依照以下程序進行。每次實驗將有 9 位受測者進入電腦實驗室中, 每人將有一台固定的電腦進行實驗, 在實驗進行前研究人員會先將實驗流程進行解說, 並進行一次簡單的測試做為示範, 已確定受測者瞭解實驗流程, 待確認受測者瞭解實驗流程後便開始進行實驗測試, 當受測者每看完一次視訊摘錄後, 畫面會停止顯像, 已提示受測者本段視訊摘錄已播放完畢, 受測者便可進行評分的動作, 受測者可以藉由滑鼠移動評分指標(如圖 7 所示)針對他們認定的確定性與理解程度七題問項(如表 7

所示)進行評分, 評分尺度使用李克特 7 點量表(Likert scale)越左邊代表越差越右邊表示越好, 評分完成後, 評分指標將會回到中心位置, 受測者可以進行觀看下一段視訊摘錄。

本實驗受測者(Subjects)以青少年為主, 實驗人數共 54 人, 實驗以隨機方式挑選長榮大學學生進行受測。



圖 7 Video Skim 影片代表性之實驗介面

表 7 實驗問項

確定性 3 題問項
1. 此「視訊」是否提供足夠資訊使您有效理解原影片內容
2. 視訊中「文字訊息」是否提供足夠資訊使您有效理解原影片內容
3. 「音訊」是否提供足夠資訊使您有效理解原影片內容
理解程度 3 題問項
1. 此「視訊」內容與原影片內容之相關程度
2. 視訊中「文字訊息」與原影片內容之相關程度
3. 「音訊」內容與原影片內容之相關程度
驗證確定性、理解程度與原始視訊比較之 1 題問項
1. 您所觀看的這段「視訊摘要」是否足以代表原始電影內容概況

3.3 分析方法

本研究因應驗證研究目的之需要, 因此利用 STATISTICA6.0 等統計軟體進行資料分析, 並採用下列資料分析方法:

3.3.1 一般統計分析(General Statistic Analysis)

利用受測者主觀的李克特 7 點量表(Likert scale)評分進行分析, 並統計不同的視訊摘要擷

取序列長度、視訊摘要總長度對視訊摘錄代表性之相關影響程度，針對受測者評估的分數進行觀察與說明。

3.3.2 信度分析(Reliability Analysis)

運用相關係數分析與 Cronbach' s α 係數之值來衡量整體實驗數據結果是否具有可信度與代表性，Cronbach' s α 係數越高表示此實驗數據可信度越高且越具代表性。

3.3.3 變異數分析(Analysis of Variance, ANOVA)

透過變異數分析將可以瞭解受測者對於本實驗因素視訊摘要擷取序列長度(Skimmed Segment)與視訊摘要總長度(Total Skimmed Length)的差異性是否具有顯著關係，由此便可探究差異來源。

4. 實驗分析

4.1 一般統計分析(General Statistic Analysis)

根據實驗設計所得主觀評估結果分析，針對在劇情視訊內容裡，不同擷取序列片段長度(Skimmed Segment)與不同的視訊摘錄總長度(Total Skimmed Length)，受測者對上述不同組合，哪一組最能代表原始視訊內容進行說明與觀察。圖 8 為依據 3.2.2 實驗設計方法挑選表 2 中所進行之實驗一，針對視訊摘要擷取序列找出多長的擷取秒數才是最適當的摘錄選擇，圖中 C1 至 C9 的說明可參考表 4 Video Skim 實驗樣本分配表，由圖 8 中可看出觀看者評分偏好落在 C3、C6、C9 這個區間內，並且給予較高的視訊代表性評分，觀察這三個類型發現它們皆具有擷取序列片段為 10 秒的特性，由此可知擷取序列片段長度(Skimmed Segment, SS)秒數越長越能使動態影像視訊摘錄越具代表性。

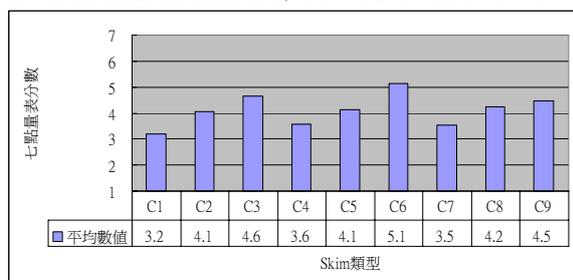


圖 8 實驗一 SS 視訊代表性之平均數

圖 9 為依據 3.2.2 實驗設計方法挑選表 3

中所進行之實驗二，針對視訊摘要總長度找出多長的擷取率才會是最適視訊摘錄選擇，由圖 9 中可看出觀看者評分偏好落在 C7、C8、C9 這個區間內，並且給予較高的視訊代表性評分，觀察這三個類型發現它們皆具有視訊摘要總長度為 15%的特性，由此可知視訊摘要總長度(Total Skimmed Length, TSL)越長越能使動態影像視訊摘錄越具代表性。

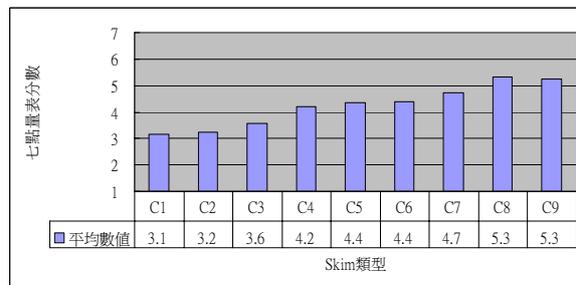


圖 9 實驗二 TSL 視訊代表性之平均數

圖 10 與圖 11 為本實驗為了驗證不確定性(Uncertainty)構面與理解程度(Comprehension)構面兩構面足以代表視訊摘要代表性，分別在實驗中對於不確定性與理解程度各設計三題問項共六題，與設計一題問項詢問“您所觀看的這段「視訊摘要」是否足以代表原始電影內容的概況”，用以證實兩者之間的關係。

圖 10 為實驗一中不確定性與理解程度兩構面共六題問項的平均數在圖中以確定性與理解程度值表示，而問項七的平均數則以原始視訊代表值表示，在此折線圖中可看到兩者之間的曲線相當接近，因此可以說明不確定性(Uncertainty)與理解程度(Comprehension)兩構面足以代表視訊摘要代表性。

圖 11 折線圖中也可看到兩者之間的曲線相當接近，因此可以說明不確定性(Uncertainty)與理解程度(Comprehension)兩構面足以代表視訊摘要代表性。

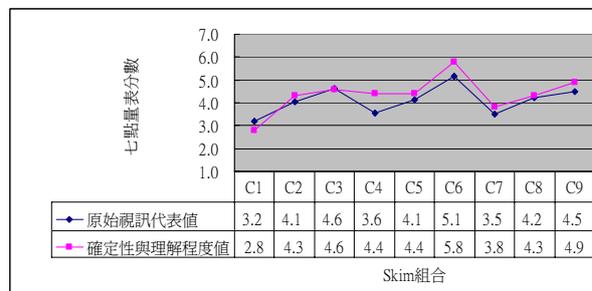


圖 10 實驗一確定性、理解程度與原始視訊代表性問項之平均數

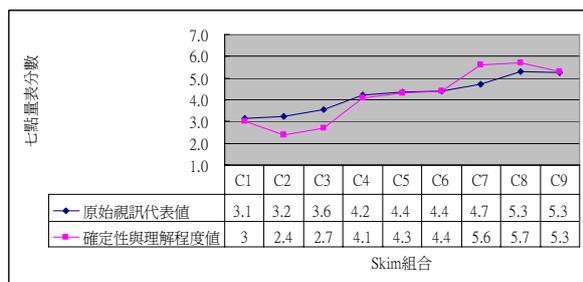


圖 11 實驗二確定性、理解程度與原始視訊代表性問項之平均數

4.2 信度分析(Reliability Analysis)

根據實驗設計所得評估量表數據結果進行整體信度分析。在信度分析方面，採取 Cronbach α 係數之值來估計本研究問卷整體及量表的信度，Cronbach α 係數為目前社會科學研究最常使用的衡量標準。

實驗一由 27 位受測者分三組 A1、A2、A3(可參考表 2)，分別實驗三次共 81 筆資料。實驗二由 27 位受測者分三組 B1、B2、B3(可參考表 3)，分別實驗三次共 81 筆資料。Nunnally[17]認為 Cronbach α 係數應在 0.70 或 0.70 以上，由表 8 中實驗一整體 Cronbach α 係數為 0.930550；實驗二整體 Cronbach α 係數為 0.937887，由此可知本實驗問卷與量表具有一定信度。

表 8 實驗整體信度分析

實驗序列	量表	Cronbach α
實驗一	視訊代表性	0.930550
實驗二	視訊代表性	0.937887

4.3 變異數分析(Analysis of Variance, ANOVA)

變異數分析是為檢定各水準下其結果的平均數是否有顯著差異。透過統計分析的變異數分析法，針對設定的研究假說進行檢定分析，目的是檢定各組平均數相等的假設是否有達到統計上的顯著差異，ANOVA 是探討一個計量的準則變數與一個或多個以上的預測變數之間的相依關係，預測變數為不同的視訊類型、擷取序列片段與視訊摘錄總長度，而準則變數為觀看者對視訊摘錄的代表性。

4.3.1 重複量測變異數分析

根據表 9 重複量測 ANOVA 分析之結果，得知視訊摘要總長度(Total Skimmed Length, TSL) P 值 < 0.001 ，表示視訊摘要總長度對於視訊摘錄代表性有顯著的影響。由上述可知假說 1 成立：不同的視訊摘錄總長度(分別取原始視訊的 5%、10%及 15%)對於視訊摘錄代表性有顯著的差異。

擷取序列片段長度(Skimmed Segment, SS) P 值 < 0.001 ，表示擷取序列片段對於視訊摘錄代表性有顯著的影響。由上述可知假說 2 成立：不同的擷取片段長度(分別取每一片段為 2.5 秒、5 秒、及 10 秒)對於視訊摘錄代表性有顯著的差異。

表 9 中 TSL*SS 其 P 值 > 0.05 ，代表視訊摘要總長度與擷取序列片段之配對對於視訊摘錄代表性並沒有顯著的影響。由此可知假說 3 成立：擷取片段長度(分別取每一片段為 2.5 秒、5 秒、及 10 秒)與視訊摘錄總長度(分別取原始視訊的 5%、10%及 15%)，兩者之間對於視訊摘錄代表性無顯著的差異。這個分析結果可以得知視訊摘要總長度與擷取序列片段對於視訊代表性皆屬同等重要。

TIME*TSL 其 P 值 < 0.05 表示不同的時間對於視訊摘要總長度有顯著的交互作用，有顯著的時間效應。TIME*SS 其 P 值 > 0.05 表示不同的時間對於擷取序列片段長度無顯著的交互作用，無顯著的時間效應。由上述可知假說 4 在 TSL 的實驗情況下成立而在 SS 的實驗情況下則不成立：不同的時間下所作的測試結果，對於視訊摘錄代表性無顯著的差異。推論假說 4 不成立的原因，因是實驗時間太過長久，導致受測者出現不耐煩而亂填問卷的情況發生。

表 9 本實驗重複量測 ANOVA 表

影響變數	視訊代表性
TSL	**
SS	**
TSL*SS	0.89492
TIME	0.46984
TIME * TSL	*
TIME * SS	0.09800
TIME * TSL*SS	0.88012

註：*表 $t > 2$ ， $P < 0.05$ ；**表 $t > 2$ ， $P < 0.001$

表 10 為試做九種 Video Skim 組合對視訊摘錄代表性之影響 ANOVA 分析，由表 10 中

Skim Type 其 P 值 <0.001 可看出本研究九種不同 Video Skim 組合對視訊摘錄代表性有顯著的影響。由上述可知假說 5 成立：不同的視訊摘錄組合(Video Skim C1~C9) ，對於視訊摘錄代表性有顯著的差異。

TIME * Skim Type 其 P 值 <0.05 得知時間對於 Video Skim 組合有顯著的交互作用，有顯著的時間效應。

表 10 九種 Skim 組合重複量測 ANOVA 表

影響變數	視訊代表性
Skim Type	**
TIME	0.46984
TIME * Skim Type	*

註：*表 $t > 2, P < 0.05$ ；**表 $t > 2, P < 0.001$

5. 結論

本研究係以主觀視訊品質評估為基礎，探討視訊摘要(Video Abstraction)技術中的動態影像視訊摘錄(Video Skim)，以三種不同的擷取序列片段長度(Skimmed Segment)，分別以 2.5 秒、5 秒、與 10 秒，作為擷取序列片段的固定單位長度，以及三種不同的視訊摘錄總長度(Total Skimmed Length)，分別以原始影像檔的 5%、10%、與 15%，觀察受測者對上述不同組合，哪一組最能代表原始視訊內容。

表 11 為本研究假說之成果表，表中可看出假說 1、假說 2、假說 3 與假說 5 經實驗分析後皆為成立，而假說 4 則在 TSL 的實驗情況下成立；SS 的實驗情況下則不成立。

表 11 研究假說成果表

實驗假說	結果
假說 1：不同的視訊摘錄總長度(分別取原始視訊的 5%、10%及 15%)對於視訊摘錄代表性有顯著的差異。	假說成立
假說 2：不同的擷取片段長度(分別取每一片段為 2.5 秒、5 秒、及 10 秒)對於視訊摘錄代表性有顯著的差異。	假說成立
假說 3：擷取片段長度(分別取每一片段為 2.5 秒、5 秒、及 10 秒)與	假說成立

視訊摘錄總長度(分別取原始視訊的 5%、10%及 15%)，兩者之間對於視訊摘錄代表性無顯著的差異。	
假說 4：不同的時間下所作的測試結果，對於視訊摘錄代表性無顯著的差異。	TSL 的實驗情況下成立 SS 的實驗情況下則不成立
假說 5：不同的視訊摘錄組合(Video Skim C1~C9)，對於視訊摘錄代表性有顯著的差異。	假說成立

本研究主要研究成果為以下幾點：

1. 透過實驗可知不同的擷取序列片段長度(Skimmed Segment)與視訊摘錄總長度(Total Skimmed Length)會影響視訊摘錄代表性。

2. 透過實驗可知不確定性與理解程度兩構面足以代表觀看者對於視訊摘錄的代表性。

3. 由實驗發現觀看者對於擷取序列片段長度(Skimmed Segment)與視訊摘錄總長度(Total Skimmed Length)所組成的動態影像視訊摘錄(Video Skim)，偏愛於 10 秒擷取序列片段與 15%視訊摘錄總長度的這個區間內，綜合實驗一與實驗二取分數較高的 Video Skim 類型，此區間包含 C6、C8、C9 三個類型。

4. 本實驗應用動態影像視訊摘錄(Video Skim)中的演算法，製作出自動電影摘錄動態影像視訊摘錄工具。

參考文獻

- [1] Venkatesh, S. and Truong, B. T., "Video abstraction: A systematic review and classification," *ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP)*, Vol. 3, No. 1, pp.1-37, 2007.
- [2] Ying, L., Shih, H. L., Jay, K. and Chia, H. Y., "Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques," *Signal Processing Magazine, IEEE*, Vol. 23, No. 2, pp.79-89, 2006.
- [3] Luo, H., Wu, L., Fan, J. and Xiao, J., "Semantic video classification and feature subset selection under context and concept uncertainty," *Digital Libraries 2004 Proceedings of the 2004 Joint ACM/IEEE Conference on*, pp.192-201, 2004.

- [4] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures." *ITU, Geneva, Switzerland*, pp.1-48, 2002.
- [5] Hands, D. S., "A Basic Multimedia Quality Model," *IEEE Transaction on Multimedia*, Vol. 6, No. 6, pp.806-816, 2004.
- [6] Pinson, M. and Wolf, S., "Comparing Subjective Video Quality Testing Methodologies," *SPIE Video Communications and Image Processing Conference*, Vol. 5150, pp.573-582, 2003.
- [7] Evain, J. P. and Alpert, T., "Subjective quality evaluation-The SSCQE and DSCQE methodologies," *ITU-R Recommendation BT.500-7*, No. 271, pp.12-20, 1997.
- [8] Winkler, S., "Issues in Vision Modeling for Perceptual Video Quality Assessment," *Signal Processing*, Vol. 78, No. 2, pp.231-252, 1999.
- [9] Pinson, M. and Wolf, S., "The Impact of Monitor Resolution and Type on Subjective Video Quality Testing," *NTIA Technical Memorandum TM-04-412*, pp.1-7, 2004.
- [10] Dufaux, F. and Winkler, S., "Video quality evaluation for mobile applications," *Proc. SPIE/IS&T Visual Communications and Image Processing Conference*, Vol. 5150, pp.593-603, pp.8-11, 2003.
- [11] Weeks, A., Aeluri, P. K., Richie, S. and Bojan, V., "Objective quality analysis of MPEG-1, MPEG-2 & Windows Media video," *Image Analysis and Interpretation, 2004. 6th IEEE Southwest Symposium on*, pp.221-225, 2004.
- [12] Ong, E., Etoh, M., Yao, S., Lin, W. and Lu, Z., "Objective quality assessment for compressed video," *Circuits and Systems, 2003.ISCAS '03.Proceedings of the 2003 International Symposium on*, Vol. 2, pp.688-691, 2003.
- [13] Michael, S. A. and Takeo, K., "Video skimming and characterization through the combination of image and language understanding," *Content-Based Access of Image and Video Database 1998 Proceedings 1998 IEEE International Workshop on*, pp.66-70, 1998.
- [14] Lee, S. H., Jay, K. and Chia, H. Y., "Video skimming based on story units via general tempo analysis," *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, Vol. 2, pp.1099-1102, 2004.
- [15] Sergio, B., Riccardo, L. and Pierangelo, M., "Hidden Markov Models for Video Skim Generation," *Image Analysis for Multimedia Interactive Services, Eighth International Workshop on*, pp.6, 2007.
- [16] Michael, G. C., Roy, T. C., Michael, A. S. and David, B. W., "Evolving video skims into useful multimedia abstractions," *Proceedings of the SIGCHI conference on Human factors in computing systems CHI '98*, pp.171-180, 1998.
- [17] Nunnally, J. C., *Psychometric Theory, 2nd ed, New York:McGraw-Hillock*, 1978.
- [18] <http://www.imdb.com/title/tt1037850/>