

根據空間事件之影片相似尋取

梁恩輝

淡江大學資訊管理系

副教授

ehliang@mail.tku.edu.tw

陳鈺薇

淡江大學資訊管理系

碩士班研究生

hs920232@hotmail.com

林詒慧

淡江大學校務資訊組

技士

YHLin@mail.tku.edu.tw

摘要

一個影片的內​​容可以視為一連串的影格，而每一個影格可以經由轉換以一個符號影像(symbolic image)來表示。根據符號影像中物件間空間關係之影片相似尋取是一個影片尋取的重要方法。在先前許多相關的研究中，並未直接明確的考慮到在影片中物件間之空間關係的變化。

在本論文中，我們定義在相鄰影格間，物件間的空間關係的變化稱為空間事件(Spatial Event)。當多個空間事件同時發生時，以字串的方式表示，稱為空間事件字串(Spatial Event String, 簡稱 SES)。我們提出計算兩 SES 間之相似度方法。將查詢的 SES 序列與參考影片之 SES 序列做比較，並根據 SES 間之相似度，我們提出在參考影片之 SES 序列中找出與查詢影片相似的片段的方法。因此，即可進行以空間事件為基礎的影片相似尋取。

關鍵詞：符號影像、影片相似尋取、空間事件

Abstract

A video can be viewed as a sequence of frames. Each frame can be transformed into a symbolic image. Similarity retrieval of video based on the spatial relationship between objects in the symbolic image is an important method. In the previous research, the change of the spatial relationship between objects in the video is not considered directly and precisely.

In this paper, the spatial event is defined as the change of the spatial relationship between objects from a frame to its next. When multiple events occur, the spatial events are represented as a string, named the spatial event string (SES). We propose a method to calculate the similarity between two SESs. Similarity retrieval of video can be performed by comparing the query SES sequence with the SES sequence of the reference video. We proposed a method to locate the segment in the reference video similar to the query SES sequence based on the similarity between SESs. Hence, Similarity retrieval of

video based on spatial events can be performed.

Keywords— symbolic image, video similarity retrieval, spatial event.

1. 緒論

隨著科技及網路的進步，每個人都可透過桌上型電腦、NB 及行動裝置(例如：智慧手機、平板電腦、iPhone...等)這些設備在網路上搜尋或是分享資訊。影片便是其中一個重要且有意義的資訊，因為影片包含豐富的資訊(影像、聲音及文字)，是很好的知識或訊息傳遞的資料類型。根據YouTube(線上影片社群)的統計顯示，使用者每天在YouTube 網站上觀賞二十億部影片，每天上傳的影片也有數十萬部之多。如何建置一個多媒體系統管理這些大量的影片並方便使用者搜尋影片是很重要的議題。

基於內容的影片相似尋取(Content-Based Video Retrieval)是一種處理影片資料較合適的解決方法之一。在之前的研究中，最重要的方法之一是根據存在於物件之間時間與空間關係來進行的。許多索引方法已經被提出來表示一個符號影像中物件之間的空間關係，例如：2D string[3]、2D G-string[2,7]、2D C-string[12]、2D C+-string[6]、2D Z-string[8]等。實際上，影片的內​​容是可以被視為一個影像序列[4,15]。許多由2D string的概念延伸出的索引方法已被提出來表示在一個影片中物件之間的時間與空間關係。如：2D B-string[14]、2D C-tree [5]、9DLT strings [1]、3D-list [13]、3D C-string [9,10]和3D Z-string [11]。一些基於這些符號影片模型的尋取方法也已被提出。在文獻[16]中提出了基於9DST-string的加權時間和空間關係的影片相似尋取方法。

利用在符號影片中物件之間的時間和空間關係的相似尋取是影片尋取的重要方法之一。在先前許多的研究中，對於影片相似尋取的匹配問題，是以查詢影片之影格序列和原始影片之影格序列間之序列匹配問題之處理方

式進行。然而，在這種情況下，是根據影像中物件之空間關係來計算查詢影片中之影像與參考影片中之影像的相似度，並以此做為影片查詢之基礎，並未直接明確考慮到物件間空間關係的變化，這種變化在本論文中定義為空間事件。

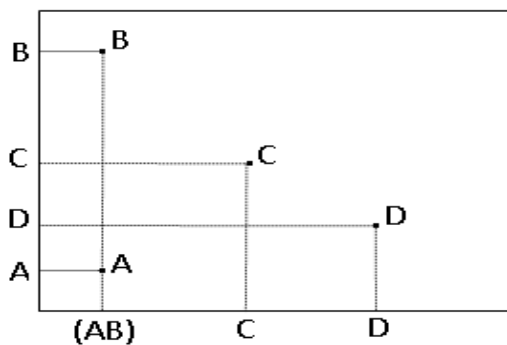
在影片中通常物件會移動，因此一物件由一影格至下一影格之移動而產生空間事件，而空間事件是影片中重要的資訊。因此，在本論文中我們提出如何根據空間事件來進行影片相似尋取的方法。

2. 空間事件字串及影片相似尋取

本節將介紹空間事件及序值，再以序值定義空間事件。接著我們將說明如何根據查詢影片中之空間事件，來進行影片之相似尋取，尋取參考影片中與查詢影片相似之片段。

影片中的物件已經過影像辨識技術，並取得每張影格中各物件之最小邊界矩形 (Minimum Boundary Rectangle, MBR) 之後，由於我們的關注在於物件的移動資訊，因此以 MBR 之中心點做為代表物件之參考點，並且投影至 X 軸及 Y 軸，形成兩個字串分別為 u-string 及 v-string，此為一種表示符號圖之方法，其中投影在同一點之物件會以 “(” 和 “)” 包圍起來，如圖 1 中，u-string 為 (AB)CD，v-string 為 ADCB。

2.1 序值(Rank)



u-string: (AB)CD
v-string: ADCB

圖 1、範例符號圖

在 u-string 及 v-string 中之物件將分別給予序值，計算物件序值時需考慮以下三種情況：(1) 單獨投影至一點的物件，(2) 括弧裡的第一

個物件，(3) 括弧裡非第一個物件的其他物件。以圖 1 的 Y 軸為例，物件 A、B、C、D 分別單獨投影至 Y 軸上，此為情形(1)，物件序值為出現在其之前物件個數加 1，A 為 Y 軸上第一個投影點，因此序值為 1，D 為第二個投影點，則為 A 的序值加 1，等於 2，C 和 B 以此類推，序值分別為 3 和 4。在 X 軸上，物件 A、B 投影至 X 軸上的同一點，我們用 (AB) 來表示，由情形(2)可知，A 之序值為 1，而由情形(3)可知，B 和 A 有相同之序值，因此 B 之序值也為 1，而 C 和 D，根據情形(1)，其序值分別為 3 和 4。物件在影格中的序值以下列方式來表示：

$\text{Rank}^x(A)_i$ 表示物件 A 在 frame i 中 X 軸上之序值。

$\text{Rank}^y(A)_i$ 表示物件 A 在 frame i 中 Y 軸上之序值。

2.2 空間事件及空間事件字串

前述提到，影片中每個畫面可視為一個靜態影像，每個影像中可能有多個物件。當物件在移動時，物件會改變位置，而物件間的空間關係就會有所變化。對於空間事件，我們做以下之定義。

定義一：兩個物件間的空间關係在連續的兩個畫面中發生變化時，我們將此稱之為一個空間事件(Spatial Event)。

我們利用物件間序值大小變化，定義出了在 X 軸上之六種空間事件，分別為事件 1 至事件 6，如表 1。在 Y 軸上，也可用類似的方法定義空間事件。若兩相鄰畫面的所有物件對之空間關係無發生任何相對變化時，則無任何空間事件發生。

我們以圖 2 中的物件 A 與物件 B 舉例說明，在 frame i 中，物件 A 與物件 B 投影至同一點，因此 $\text{Rank}^x(A)_i$ 和 $\text{Rank}^x(B)_i$ 皆等於 1；而在 frame i+1 時， $\text{Rank}^x(B)_{i+1}=1$ ， $\text{Rank}^x(A)_{i+1}=3$ ；兩物件在 X 軸方向的序值關係由 $\text{Rank}^x(A)_i = \text{Rank}^x(B)_i$ 變為 $\text{Rank}^x(A)_{i+1} > \text{Rank}^x(B)_{i+1}$ ，根據表二之定義，此時發生空間事件 6。也就是說，在 X 軸或者是 Y 軸上， f_i 中物件 A 的序值等於物件 B 的序值，且在 f_{i+1} 中物件 A 的序值大於物件 B 的序值，則物件對 AB 由 f_i 至 f_{i+1} 發生空間事件 6。一個空間事件包含了空間事件編號與物件對名稱，因此，此空間事件表示為 6AB。

表 1、空間事件定義表

空間事件	X 軸上序值大小關係變化
空間事件 1	$\text{Rank}^x(A)_i < \text{Rank}^x(B)_i$ 且 $\text{Rank}^x(A)_{i+1} > \text{Rank}^x(B)_{i+1}$
空間事件 2	$\text{Rank}^x(A)_i > \text{Rank}^x(B)_i$ 且 $\text{Rank}^x(A)_{i+1} < \text{Rank}^x(B)_{i+1}$
空間事件 3	$\text{Rank}^x(A)_i < \text{Rank}^x(B)_i$ 且 $\text{Rank}^x(A)_{i+1} = \text{Rank}^x(B)_{i+1}$
空間事件 4	$\text{Rank}^x(A)_i > \text{Rank}^x(B)_i$ 且 $\text{Rank}^x(A)_{i+1} = \text{Rank}^x(B)_{i+1}$
空間事件 5	$\text{Rank}^x(A)_i = \text{Rank}^x(B)_i$ 且 $\text{Rank}^x(A)_{i+1} < \text{Rank}^x(B)_{i+1}$
空間事件 6	$\text{Rank}^x(A)_i = \text{Rank}^x(B)_i$ 且 $\text{Rank}^x(A)_{i+1} > \text{Rank}^x(B)_{i+1}$

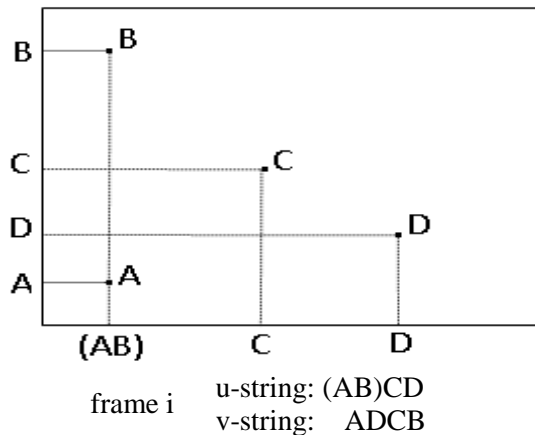


圖 2、兩相鄰畫面中物件的移動

而一個影格間，可能發生多個空間事件；因此，將影格間所有有發生空間事件的物件對用字串記錄下來，我們稱為空間事件字串 (Spatial Event String, SES)。

接下來，我們將以圖 3 之範例影片說明如何產生 SES。此範例影片中，有 6 個畫面 $f_1 \sim f_6$ ，各包含三個物件，分別為物件 A、物件 B、物件 C。其 X 軸與 Y 軸序值整理如表 2。

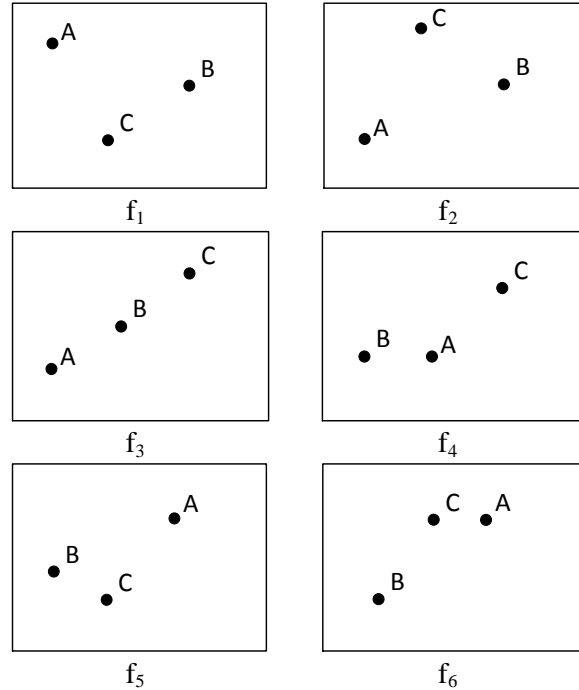


圖 3、範例影片

表 2、範例影片中物件的序值

(a) X 軸上之序值

	f_1	f_2	f_3	f_4	f_5	f_6
A	1	1	1	2	3	3
B	3	3	2	1	1	1
C	2	2	3	3	2	2

(b) Y 軸上之序值

	f_1	f_2	f_3	f_4	f_5	f_6
A	3	1	1	1	3	2
B	2	2	2	1	2	1
C	1	3	3	2	1	2

首先定義一影格間之空間事件字串表示法。若發生空間事件之兩相鄰畫面為 frame i 與 frame i+1，而在此影格間之 X 軸及 Y 軸上之空間事件字串，則分別為 $\text{SES}^x[i, i+1]$ 和 $\text{SES}^y[i, i+1]$ 。而此二者可整合為 $\text{SES}[i, i+1]$ ，也就是說 $\text{SES}[i, i+1] = (\text{SES}^x[i, i+1], \text{SES}^y[i, i+1])$ 。

假設在 frame i 和 frame i+1 之間，X 軸上發生了 K 個空間事件， $N_j P_j Q_j$ 代表第 j 個空間事件，其中 $1 \leq j \leq K$ ， $P_j Q_j$ 代表第 j 個物件對而 N_j 代表其空間事件編號。而在 Y 軸上發生了 L 個空間事件， $n_j R_j S_j$ 代表第 j 個空間事件，其中 $1 \leq j \leq L$ ， $R_j S_j$ 代表第 j 個物件對而 n_j 代表其空間事件編號。因此我們可將影格間之空間事件字串表示成：

$$\begin{aligned} & \text{SES}[i,i+1] \\ &= (\text{SES}^x[i,i+1], \text{SES}^y[i,i+1]) \\ &= (N_1P_1Q_1, N_2P_2Q_2 \dots, N_KP_KQ_K, \\ & \quad n_1R_1S_1, n_2R_2S_2 \dots, n_LR_LS_L) \end{aligned}$$

從表 2 中，根據其序值的改變，我們可得知該範例影片，在 X 軸， f_2 與 f_3 、 f_3 與 f_4 、 f_4 與 f_5 有發生空間事件；在 Y 軸， f_1 與 f_2 、 f_2 與 f_3 、 f_4 與 f_5 、 f_5 與 f_6 有發生空間事件。若該影格間的 X 軸或 Y 軸沒有發生空間事件，則用 \emptyset 來表示。

圖 3 之範例影片之 SES 序列如下：

$$\begin{aligned} \text{SES}[1,2] &= (\emptyset, 2AB2AC \ 2BC), \text{SES}[2,3] = (2BC, \\ & \emptyset), \text{SES}[3,4] = (1AB, 3AB), \text{SES}[4,5] = (1AC, \\ & 6AB1AC1BC), \text{SES}[5,6] = (\emptyset, 4AC2BC). \end{aligned}$$

2.3 空間事件字串間之相似度

接著說明 SES 間相似度的計算方法，我們對兩 SES 做比較。由於 SES 包含 SES^x 及 SES^y ，因此先分別計算兩 SES^x 之間與兩 SES^y 之間的相似度。兩 SES^x 所共有的空間事件個數即為此二 SES^x 之間的相似度。兩 SES^y 之間的相似度也可用相同的方法計算。最後將此二相似度加總，即為 SES 的相似度。

假設兩個 SES： S_1 和 S_2 ，如下所列：

$$\begin{aligned} S_1 &:= (3AB3AD2BC, 3BC1AD); \\ S_2 &:= (4AD1CE4CD3AB2BC3AC, \emptyset) \end{aligned}$$

在 X 軸上， S_1 與 S_2 共有的空間事件為 3AB、2BC，因此其相似度為 2；在 Y 軸上， S_1 與 S_2 沒有任何共有的空間事件，因此相似度為 0。而 S_1 與 S_2 之相似度為以上二相似度之總和，為 $2+0=2$ 。以 $\text{sim}(S_1, S_2)$ 代表此相似度。

2.4 影片相似尋取

我們提出的影片查詢方法，由包含 $m+1$ 個影格之查詢影片中，取得包含 m 個 SES 之查詢序列進行查詢，用 $U=u_1u_2u_3\dots u_m$ 來表示，其中 u_i 為一 SES，包含 SES^x 與 SES^y 。由於是以 SES 序列做相似比較，因此必須先將欲查詢之對象(又稱參考影片)進行處理，取得參考 SES 序列，以 $V=v_1v_2\dots v_r$ 表示。

首先利用前述計算兩 SES 相似度的方法取得 U 中的每個 u_i 和 V 中的每個 v_j 之間的相似度，再根據以下之定義二來定義 U 和 V 中某一段之相似度。

定義二：

對一查詢 SES 序列 $U=u_1u_2\dots u_m$ ，若有一影片片段之 SES 序列 $W=w_1w_2\dots w_z$ ，且 $m \leq z$ ，對 W 中每一子序列 $w_{j_1}w_{j_2}\dots w_{j_m}$ ， $1 \leq j_1 < j_2 < \dots < j_m \leq z$ ，計算 $\text{sim}(w_{j_1}, u_1) + \text{sim}(w_{j_2}, u_2) + \dots + \text{sim}(w_{j_m}, u_m)$ ，則此值即為 U 和此子序列之相似度，其中和 U 最大之相似度即為 U 和 W 之相似度。

此外，為了可以彈性地參考影片中不同範圍內尋取與查詢影片相似之片段，在查詢時我們可以設定查詢範圍之大小，如定義二中之 z ，可以以 $z = m$ 或 $z > m$ 兩種方式進行查詢。根據定義一找到與查詢影片相似度最高之片段。

3. 影片尋取範例

我們以圖 3 之範例為例，以此範例影片為參考影片，並提供一個查詢範例影片來尋取，該查詢範例影片為三張連續影格組成，其中包含三個物件：物件 A、物件 B 及物件 C。三個物件在影片中逐一移動，如圖 4 所示。

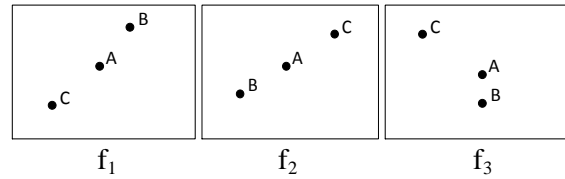


圖 4、查詢範例影片

首先根據物件在查詢影片及參考影片中各影格內之序值，分別產生查詢 SES 序列及參考 SES 序列如下：

查詢 SES 序列：

$$\begin{aligned} \text{SES}[1,2] &= (1AB2AC2BC, 1AB2AC2BC), \\ \text{SES}[2,3] &= (4AB1AC \ 1BC, \emptyset); \end{aligned}$$

參考 SES 序列：

$$\begin{aligned} \text{SES}[1,2] &= (\emptyset, 2AB2BC2AC), \\ \text{SES}[2,3] &= (2BC, \emptyset), \\ \text{SES}[3,4] &= (1AB, 3AB), \\ \text{SES}[4,5] &= (1AC, 6AB1AC1BC), \\ \text{SES}[5,6] &= (\emptyset, 2BC4AC). \end{aligned}$$

接著我們利用上節提到的方法將參考序列中之 SES 一一與查詢序列中每個 SES 計算其

相似度。以計算參考之 SES[1,2]與查詢之 SES[1,2]為例，其中之二 SES[1,2] 間之相似度為 0，但其中之二 SES[1,2]間之相似度為 1，因此此二 SES[1,2]之相似度為 1。其餘各查詢 SES 與每一個參考 SES 的相似度也可用同樣方法計算取得並列於表 3。

在此範例中，我們使用 $z = m$ 的尋取方法，其中 m 為查詢 SES 的長度，其值為 2，因此由參考 SES 序列中以兩個 SES 為單位依序取出四個子序列(如表 4 所示)。再一一檢查他們是否與查詢序列相似。

根據定義二，子序列 1 之相似度為 3，子序列 2 之相似度為 1，子序列 3 之相似度為 2，子序列 4 相似度為 0。在這四個子序列當中，與查詢 SES 序列相似度最高者為子序列 1 與子序列 3，因此子序列 1 及子序列 3 為該範例的最佳相似尋取結果，代表參考影片的影格 1 至影格 3 及影格 3 至影格 5。

表 3、查詢 SES 與參考 SES 之間的相似度

查詢之 SES 參考之 SES	SES[1,2]	SES[2,3]
SES[1,2]	2	0
SES[2,3]	1	0
SES[3,4]	1	0
SES[4,5]	0	1
SES[5,6]	1	0

表 4. 參考 SES 序列中長度為二之子序列

子序列 編號	參考影片 SES 子序列	相似度
1	SES[1, 2], SES[2, 3]	2
2	SES[2, 3], SES[3, 4]	1
3	SES[3, 4], SES[4, 5]	2
4	SES[4, 5], SES[5, 6]	0

4. 結論

根據符號影像中物件間空間關係之影片相似尋取是一個影片尋取的重要方法，但在先前相關許多影片相似尋取的研究中，並沒有直接明確的考慮到空間事件的因素。在本論文中，我們提出了一個以空間事件為基礎的影片相似尋取的方法，並用一個範例來說明其有效性。

在本論文中，我們定義物件間的空間關係變化稱為空間事件及在影格間之空間事件字

串(SES)，並提出計算兩 SES 間之相似度方法。影片的相似尋取可以根據一串查詢的 SES 進行，將查詢 SES 序列和參考影片之 SES 序列做比較，並根據 SES 間之相似度，提出了在參考影片之 SES 序列中找出與查詢影片相似的片段的方法。如此一來，即可進行以空間事件為基礎的影片相似尋取。

參考文獻

- [1] Chan, Y.K., and Chang, C.C., "Spatial similarity retrieval in video databases," *Journal of Visual Communication and Image Representation*, Vol. 12, No. 2, pp. 107-122, 2001.
- [2] Chang, S.K., Jungert, E., and Li Y., "Representation and retrieval of symbolic pictures using generalized 2D strings," in: *SPIE Proceedings on Visual Communications and Image Processing, Philadelphia*, pp. 1360-1372, 1989.
- [3] Chang, S.K., Shi, Q.Y., and Yan, C.W., "Iconic indexing by 2-D strings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 3, pp. 413-428, 1987.
- [4] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P., "Query by image and video content: the QBIC system," *IEEE Computer Society*, Vol. 28, No. 9, pp. 23-32, 1995.
- [5] Hsu, F.J., Lee, S.Y., and Lin, B.S., "Video data indexing by 2D C-trees," *Journal of Visual Languages & Computing*, Vol. 9, No. 4, pp. 375-397, 1998.
- [6] Huang, P.W., and Jean, Y.R., "Using 2D C+-string as spatial knowledge representation for image database systems," *Pattern Recognition*, Vol. 27, No. 9, pp. 1249-1257, 1994.
- [7] Jungert, E., "Extended symbolic projections as a knowledge structure for spatial reasoning," in: Proceedings of 4th BPRA Conference on *Pattern Recognition*, pp. 343-351, 1988.
- [8] Lee, A.J.T., and Chiu, H.P., "2D Z-string: a new spatial knowledge representation for image databases," *Pattern Recognition Letters*, Vol. 24, No. 16, pp. 3015-3026, 2003.

- [9] Lee, A.J.T., Chiu, H.P., and Yu, P., “3D C-string: a new spatio-temporal knowledge representation for video database systems,” *Pattern Recognition*, Vol. 35, No. 11, pp. 2521–2537, 2002.
- [10] Lee, A. J.T., Chiu, H.P., and Yu, P., “Similarity retrieval of videos by using 3D C-string knowledge representation,” *Journal of Visual Communication and Image Representation*, Vol. 16, No.6, pp. 749–773, 2005.
- [11] Lee, A. J.T., Yu, P., Chiu, H.P., and Hong, R.W., “3D Z-string: A new knowledge structure to represent spatio-temporal relations between objects in a video,” *Pattern Recognition Letters*, Vol. 26, No. 16, pp. 2500–2508, 2005.
- [12] Lee, S.Y., and Hsu, F.J., “2D C-String: a new spatial knowledge representation for image database systems,” *Pattern Recognition*, Vol. 23, No. 10, pp. 1077-1087, 1990.
- [13] Liu, C.C., and Chen, A.L.P., “3D-list: a data structure for efficient video query processing,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 14, No. 1, pp. 106–122, 2002.
- [14] Shearer, K., Venkatesh, S., and Kieronska, D., “Spatial indexing for video databases,” *Journal of Visual Communication and Image Representation*, Vol. 7, No. 4, pp. 325–335, 1997.
- [15] Yeung, M., Yeo, B.L., and Liu, B., “Extracting story units from long programs for video browsing and navigation,” in: *Proceedings of the Third IEEE International Conference on Multimedia Computing and Systems*, pp.296–305, 1996.
- [16] Yu, P., “Weighted similarity retrieval of video database,” *ICIC Express Letters*, Vol. 4, No. 5(B), pp. 2009–2014, 2010.