

双面影印

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

公告本

※ 申請案號：97120917

※ 申請日期：97.6.5

※IPC 分類：G06F 17/30(2006.01)

一、發明名稱：(中文/英文)

使用多個搜尋組合改良音訊/視訊指紋搜尋正確性之技術

IMPROVING AUDIO/VIDEO FINGERPRINT SEARCH ACCURACY USING MULTIPLE SEARCH
COMBINING**二、申請人：(共 1 人)**

姓名或名稱：(中文/英文)

杜比實驗室特許公司 / DOLBY LABORATORIES LICENSING CORPORATION

代表人：(中文/英文)

張莉妮 / CHEUNG, LILY S.

住居所或營業所地址：(中文/英文)

美國加州舊金山市·帕翠洛街 100 號

100 Potrero Avenue, San Francisco, CA 94103, U.S.A

國籍：(中文/英文)

美國 / U.S.A.

三、發明人：(共 2 人)

姓名：(中文/英文)

1. 江文字 / JIANG, WENYU

2. 包爾 克勞斯 / BAUER, CLAUS

國籍：(中文/英文)

1. 中國 / CHINA

2. 德國 / GERMANY

四、聲明事項：

主張專利法第二十二條第二項 第一款或 第二款規定之事實，其事實發生日期為：。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

1. 美國、 2007/06/06、 60/933,614

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

五、中文發明摘要：

一種用於識別音訊及/或視訊資訊之一片段的方法，包含以下步驟：獲得該片段內的多數個隔開的時間位置之每個上的一查詢指紋；對每個此查詢指紋在一資料庫內搜尋指紋以獲得一可能的匹配；對每個此查詢指紋，獲得該資料庫內的一被找出的指紋之一可能的匹配之一置信等級；以及組合可能的匹配之搜尋結果，其中每個可能的匹配結果被賦予一個別置信等級之權重。一置信等級可以是以下中的至少一者或二者之一函數：(1)一查詢指紋與一被找出的指紋之間的差值之一測量；以及(2)一查詢指紋之時間位置與一被找出的指紋之時間位置之間的相對時序關係。

六、英文發明摘要：

A method for use in identifying a segment of audio and/or video information comprises obtaining a query fingerprint at each of a plurality of spaced-apart time locations in said segment, searching fingerprints in a database for a potential match for each such query fingerprint, obtaining a confidence level of a potential match to a found fingerprint in the database for each such query fingerprint, and combining the results of searching for potential matches, wherein each potential match result is weighted by a respective confidence level. A confidence level may be a function of at least one or both of (1) a measure of difference between a query fingerprint and a found fingerprint and (2) the relative timing relationship between the time location of a query fingerprint and the time location of a found fingerprint.

七、指定代表圖：

(一)本案指定代表圖為：第 (3) 圖。

(二)本代表圖之元件符號簡單說明：

(無)

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

九、發明說明：

【發明所屬之技術領域】

發明領域

本發明是關於音訊及視訊指紋。本發明尤其是關於改良指紋搜尋正確性之精確度。

【先前技術】

發明背景

隨著數位內容之快速發展，已增加對正確地管理及定位內容之需求。一具有前景的應用是識別在一已知內容儲存庫內的音訊或視訊之摘錄。這可用於監測網際網路上的內容之非法下載/傳輸。其也可用於提供使用者正在收聽或觀看以及感興趣的內容之智慧存取，但是他們不知道該等內容之名稱或作者資訊。該使用者可(例如)將一行動電話靠近目前正在播放一歌曲的揚聲器旁，且讓行動電話經營者之軟體找出歌曲之名稱、其演唱者以及專輯資訊等。

此識別能力一般藉由首先產生特徵資訊(被稱為指紋)而實現，該特徵資訊被設計以唯一地識別音訊及視訊信號，接著在來自儲存庫/資料庫的指紋與來自該搜尋查詢內的摘錄的指紋之間執行某形式的型樣匹配搜尋。此一資料庫一般以一搜尋樹之形式實施，雖然其他資料結構也是可能的。一般而言，每個指紋對應音訊或視訊之某一片段。因此一兩秒音訊指紋將對應音訊之一兩秒片段。一指紋一般以一般被稱為簽章的特徵資訊之小的連續區塊實施。一兩秒指紋可(例如)由兩百個10毫秒(ms)長的簽章組成，其中

每個簽章自10ms的新音訊或視訊資訊計算出。因此，指紋之型樣匹配是一種用於比較對應簽章的程序。這在第1圖中被描述，顯示了自簽章產生一指紋。

為了在來自一查詢摘錄的指紋與來自該資料庫的指紋
5 間執行正確的型樣匹配，至關重要的是在比較期間其兩者之間的正確的時間對準。為了確保此，一般以每個可能的時間偏移開始的所有指紋都被加到該資料庫以保證其等中的至少一者具有足夠接近該查詢指紋的時間對準。若一簽章是10ms長，則一兩秒指紋在一兩秒滑動視窗上每隔10ms
10 被偏移，接著被加到該資料庫。這也在第1圖中被描述，且在此情況下，其在連續的指紋之間產生一99.5%重疊，但是此冗餘一般被需要以確保良好的搜尋效能。對於小於10ms(或者一般是每個簽章之期間)的任何其他時間未對準，一種良好設計的指紋產生方法應按照時序選擇較接近
15 的簽章以較佳地匹配來自該查詢的對應簽章。簡而言之，一指紋搜尋系統之目的是為了找到一查詢指紋之對應的配對指紋，且若此種未對準存在，則該配對應最小化其與該查詢指紋的時間未對準。

因為查詢摘錄可能經歷一些編輯或處理步驟(例如使
20 用一行動電話重新擷取自一揚聲器播放的聲音)，在被擷取的音訊/視訊信號中可能具有一些失真。因此，假設具有一配對，則產生的指紋也可能相對於該資料庫內的其等配對稍微改變。

該摘錄內的失真之可能性表示此一搜尋內的最佳匹配

一般不是一相同的匹配，而是一種最接近匹配。為了定義最接近匹配需要定義兩指紋之間的差值之一測量。例如，差值之一般被使用的測量是漢明距離，即來自該查詢摘錄的指紋與來自該資料庫的指紋之間的不同位元之數目。利用差值之測量的定義，最接近匹配之對應準則是與來自該查詢摘錄的指紋具有最小漢明距離之來自該資料庫的指紋。以一指紋內的位元之數目相除的兩指紋之間的漢明距離一般被稱為位元錯誤率(BER)。該BER是相對差值之測量的一例子。當來自該摘錄的指紋與其配對之間的BER小時，最小漢明距離準則運作良好。然而，當BER增加時，產生逐漸增加的最小漢明距離的搜尋結果未找到實際的配對。幸運地是，在大多數指紋搜尋應用中，只需要識別正確的音訊/視訊，而並不需要識別對應的片段。但是當BER進一步增加時，該搜尋結果甚至可能找到錯誤的音訊/視訊，而漏掉在該音訊/視訊段內的正確的片段。該BER取決於該查詢摘錄內的失真程度以及該指紋取出方法相對於此等失真之強健性。

此外，一摘錄不屬於該資料庫內的任何音訊/視訊是可能的。例如，該摘錄可以是音樂之一新的樂曲之錄音。因為沒有任何搜尋演算法可預先知道(沒有被告知)一摘錄是否屬於該資料庫，所以其最佳做法仍是應用最小漢明距離之相同準則，但是預期在此等情況下找出的最小漢明距離非常不同於(較佳的是高於)源於該資料庫的一摘錄之最小漢明距離，且使用某一臨界值決定該摘錄是否來自該資料庫。

因此，在應用任何臨界值之前(例如，按照BER)，在一個單一搜尋操作之後可能具有3個可能的結果(其中只有一個查詢指紋被用以搜尋該資料庫)：

1. 該摘錄屬於該資料庫，且該搜尋返回正確的音訊/視訊(找到正確的音訊/視訊是足夠的，此處並不需要找到正確的配對片段)。

2. 該摘錄屬於該資料庫，且該搜尋返回錯誤的音訊/視訊。

3. 該摘錄不屬於該資料庫，且因為該搜尋總是返回某一音訊/視訊，所以答案將總是錯誤的。

第2圖顯示了一個單一搜尋之三個不同的可能結果之BER分佈之一例子。該等結果中的每個將產生該BER之一對應的機率密度函數(PDF)分佈。對於一良好設計的指紋取出演算法，第一結果之BER一般應比第二及第三結果之BER小，如第2圖中所描述。

然而，若第二與第三結果之BER具有非常類似的PDF分佈，則難以區別屬於該資料庫但是具有一錯誤搜尋結果的一摘錄與不屬於該資料庫的一摘錄。此外，對於源於該資料庫的音訊/視訊，在施加一般的音訊/視訊失真(例如，編碼解碼器壓縮)之後，一典型的實施態樣內的搜尋結果是正確的(按照識別該正確音訊/視訊)的機率一般在90至99%的範圍內，取決於在施加任何BER臨界值之前的指紋期間及失真之類型。這是良好的，但是一較高等級的正确性肯定是被期望的，且在施加BER臨界值之後(比如第2圖中的

BER=0.2)，正確的搜尋結果之比率只是稍微減少，因為結果1之BER分佈之尾部被丟棄以避免錯誤地選擇太多結果2中的頭部分佈。這表示單獨微調該BER臨界值在單個搜尋中無法產生非常高的正確性(比如99.9%)。

5 【發明內容】

依據本發明之一實施例，係特地提出一種用於識別音訊及/或視訊資訊之一片段的方法，包含以下步驟：獲得該片段內的多數個隔開的時間位置之每個上的一查詢指紋；對每個此查詢指紋在一資料庫內搜尋指紋以獲得一可能的
10 匹配；對每個此查詢指紋，獲得該資料庫內的一被找出的指紋之一可能的匹配之一置信等級；以及組合可能的匹配之搜尋結果，其中每個可能的匹配結果被賦予一個別置信等級之權重。

圖式簡單說明

15 第1圖是一特定類型的指紋可藉以產生的方式之一示意性例子；

第2圖顯示了以一個單一指紋匹配搜尋之三個可能的結果中的每個可能產生的一機率密度函數(PDF)為位元錯誤率(BER)之函數的一理想的例子；

20 第3圖是利用關於三個搜尋的多數決的進行指紋匹配搜尋結果組合之一例子的一示意性概念；

第4a圖是關於三個搜尋的指紋匹配搜尋結果之一例子的示意性概念，其中時間位置對於三個可能的匹配中的兩者是一致的；

第4b圖是關於三個搜尋的指紋匹配搜尋結果之一例子的示意性概念，其中時間位置對於三個可能的匹配中的兩者是不一致的；

第5圖是一置信等級可藉以被映射為位元錯誤率(BER)之一函數的方式之一例子；

第6圖是用於理解該指紋查詢之時間與一資料庫內找到的指紋之時間之間的相對時間位置可被決定之方法的一理想化例子，在此例中使用線性迴歸；以及

第7圖是一置信等級可藉以被映射為該指紋查詢之時間與一資料庫內找出的指紋之時間之間的相對時間位置內的時間差值之一函數的方式之一例子。

【實施方式】

較佳實施例之詳細說明

依據本發明之層面，一種用於識別音訊及/或視訊資訊之一片段的方法包含以下步驟：在該片段內的每個隔開的時間位置獲得一查詢指紋、在一資料庫內搜尋指紋以對每個此查詢指紋獲得一可能的匹配、對每個此查詢指紋獲得該資料庫內的一被找出的指紋之一可能匹配的一置信等級，且合併可能匹配的搜尋結果，其中每個可能的匹配結果被賦予一個別置信等級之權重。

一置信等級可以是以下中的至少一者或兩者之一函數：(1)一查詢指紋與一被找出的指紋之間的差值之測量；以及(2)一查詢指紋之時間位置與一被找出的指紋之時間位置之間的相對時序關係。該差值之測量可以基於漢明距離

或者其相對差值變體(例如，位元錯誤率)，或者其可基於 L_p 範數，其中組成一指紋的一向量之每一維上的差值之絕對值的 p 次方之總和的 p 次根被獲得。該 L_p 範數測量可以是一 L_2 範數差值測量，其中差值之絕對值的平方和之一平方根被獲得。作為另一選擇，差值之測量可藉由自相關獲得，其中相關性之值表示匹配之接近度。

該置信等級可以是以下中的至少一者或二者之一函數中的每個之個別置信等級之一組合：(1)一查詢指紋與一被找出的指紋之間的差值之測量；以及(2)一查詢指紋之時間位置與一被找出的指紋之時間位置之間的相對時序關係。個別置信等級之組合可以是(例如)個別置信度之總和或者個別置信等級之乘積。

一置信等級可以是該查詢指紋之時間位置與該資料庫內的可能匹配的被找出的指紋之時間位置之間的至少相對時序關係的一函數，在此情況下，具有至少三個隔開的時間位置。

音訊及/或視訊資訊之片段內的隔開的時間位置可被隔開以最小化查詢指紋之間的相關性。

可以具有多組隔開的時間位置，關於這些組中的每個，可能的匹配之搜尋結果被組合。搜尋之結果可被組合進入每一組內，且此等組合之結果接著可被進一步組合。可選擇的方式是，搜尋之結果可在所有組之間被合併。

該資料庫內的指紋可能具有與該音訊及/或視訊資訊有關的相關元資料。此元資料可包括該音訊及/或視訊資訊

之至少一名稱。

一查詢指紋可以實質上相同的方式在該等隔開的時間位置中的每個上獲得。

一種用於組合來自多個搜尋的結果之簡單但是有用的方法是多數決。例如，如第3圖中所描述，若進行三個搜尋，其中兩者返回A則(piece)，一者返回B則，則我們推斷最後的搜尋結果應是A則。若我們獲得三個不同的答案(A、B及C則)，則沒有多數獲勝者，且答案是非決定性的。在此情況下，最佳地是推斷該摘錄不屬於該資料庫，因為若其實屬於該資料庫，則三個搜尋給出三個不同的結果是相對較不可能的。

這可對任何數目的搜尋被一般化，利用以下所示的偽碼所描述的多數決(其中獲得多於一半票數者獲勝)，或者實施者期望的任何投票方案(例如，具有大多數票數且沒有平手者獲勝)。一投票方案應被設計以利用機率分析最大化正確答案之數目且最小化錯誤答案之數目。例如，對於源於該資料庫的音訊/視訊，若一個單一搜尋具有98%的正確性，則只有在三個搜尋中的至少兩者是錯誤的時，則利用多數決組合三個搜尋會得到錯誤的結論

$3 \times 0.98 \times 0.02^2 + 0.02^3 \approx 1.2 \times 10^{-3}$ ，因此組合的正確性應幾乎為99.9%。此機率分析假設每個搜尋彼此獨立，即搜尋之間沒有相關性。

用於使用多數決的搜尋組合之示範性偽碼可表示如下：

$N=3$;

在該摘錄之時間 $t, t+d[1], \dots, t+d[N-1]$ 分別執行 N 個指紋搜尋；

若 ($>N/2$ 的搜尋返回相同的 A 則)

5 返回 A 則；

否則

返回“摘錄不屬於該資料庫”。

第3圖中的時間間隔 $d[1], d[2], \dots$, 被用以在不同的位置執行搜尋，且一般而言，它們應被選擇以最小化它們的對
10 應搜尋片段之間的重疊以減少搜尋結果之間的任何不被期望的相關性。在非決定性投票之情況下，該組合程序可由多個搜尋之額外疊代被進一步級聯，且只在新的疊代中考慮該等投票，或者利用該實施者選擇的一投票方案組合先前的投票中的一些或所有。

15 看似如多數決此類的簡單方案可給出非常良好的結果。以上例子建議一幾乎99.9%之正確性，但是當該搜尋內具有錯誤之強烈叢發或相關性時，實際的正確性可能只是比原來的98%好一些：若一搜尋給出錯誤的結果，則一附近時間位置上的搜尋也可能給出錯誤的結果，且一般是相
20 同的錯誤結果。這可能(例如)由於音訊/視訊之一些區域(例如，視訊中的高度動作片段或安靜的音訊信號)對失真(分別為低位元率壓縮或背景雜訊)較敏感而產生。這將會誤導一多數決策演算法認為其找到正確答案，但實際上沒有。除此之外，若一音訊/視訊不屬於該資料庫，則必須處理多於

一個搜尋可返回相同音訊/視訊之情況，因為其將產生一錯誤的多數獲勝者。

藉由使用本發明之層面，正確性可增加，甚至在存在此等叢發錯誤之情況下。這可(例如)藉由不僅組合來自每個
5 搜尋結果的音訊/視訊之名稱資訊，而且組合附屬資訊(例如，在被找出的音訊/視訊內的該被找到的片段之時間位置)以及該搜尋結果之BER而實現。例如，在第4(a)圖中，若三個搜尋中的兩者返回A則，且該兩搜尋片段之時間位置相差5秒，則該兩搜尋結果內的被找出的片段之時間位置也應相
10 差5秒(或者在某一錯誤邊限內)。若此未被滿足，如第4(b)圖，其是一錯誤的多數獲勝者幾乎是肯定的。這是一公平假設，因為若一搜尋返回錯誤的音訊/視訊，則其幾乎肯定是由於以下偶然事故：其恰巧是產生最小漢明距離的結果。因此，被找出的片段之時間位置應是相當隨機，而不是
15 確定性的。因此，兩個錯誤的搜尋結果返回相同的錯誤音訊/視訊及一對一致接近的時間位置之機率相當低。實際的機率可藉由分析正確及錯誤的搜尋結果之時間位置的PDF分佈而被估計。

為了一般化具有N個搜尋的比較，其表示該等搜尋之時間位置之間的差值對於該等查詢及搜尋結果應是相同或接近的。然而，為了使此比較更易於以軟體實施，可計算出原始音訊/視訊與一搜尋結果內的摘錄之時間位置之間的偏移(即，差值)，且此值在第4(a)及4(b)圖中被稱為 t_{offset} 。第一搜尋之偏移是 $t_{\text{offset},1}$ ，第二搜尋之偏移是 $t_{\text{offset},2}$ 。對於

返回相同音訊/視訊的每對搜尋結果，則可計算出它們的 t_{offset} 之間的差值。接著，對於獲得最小差值的每對搜尋結果，若此最小差值小於某一臨界值(比如0.2秒)，則其音訊/視訊名稱可被認為是最終的答案。這在以下用於使用附屬

5 資訊的搜尋組合之偽碼中被描述，其中 Min_t_offset_dist 是最小差值，且 Candidate 表示獲得此最小值的搜尋(對)。一具有0.2秒之一示範性值的臨界值 d_A 被用以決定該結果是否足以信賴。

用於使用 t_{offset} 附屬資訊進行搜尋組合之示範性偽碼可

10 表示如下：

```

N=3 ;
分別在該摘錄之時間  $t, t+d[1], \dots, t+d[N-1]$  執行 N 個指紋搜尋；
Min_t_offset_dist = min( $t_{\text{offset},i} - t_{\text{offset},j}$ , 其中
search[i].piece == search[j].piece,  $i < j$ );
15 Candidate = arg min ( $t_{\text{offset},i} - t_{\text{offset},j}$ , 其中
search[i].piece == search[j].piece,  $i < j$ );
d_A = 0.2;
若 ( $\text{Min\_t\_offset\_dist} < d\_A$  &&  $\text{Candidate}$  有  $>N/2$  票數)
返回 search[Candidate].piece;
20 否則
返回“摘錄不屬於該資料庫”；

```

在以上偽碼中，“search[i]”表示來自第 i 個搜尋的結果，以及“search[i].piece”是該則被找出的音訊/視訊之識別資訊，例如其名稱。一種用以使用在以下描述的附屬資訊

25 時序資訊之較一般的方法使用一時間測量 t_{distance} ，而不是 t_{offset} 。

對於一大的 N 數目的搜尋，若較多搜尋返回相同的音訊

/視訊(若結果是正確的), 則很清楚地, Min_t_offset_dist 之值將較小。例如, 若其餘都相等, 則五個樣本之間的最小值肯定比三個樣本之間的最小值小。因此, 我們可定義更多的臨界值, 例如除了 d_A 之外的 d_B 、 d_C , 取決於幾個

5 搜尋返回相同的音訊/視訊。這樣, 可基於該多數決資訊(多少搜尋返回相同的音訊/視訊)智慧地使用附屬資訊(在此例中是 t_{offset})且達成較高的正確性。

若查詢內容經歷某加速或減速(其有時是廣播節目之情形), 則該等查詢指紋之間的時間對來自該等搜尋結果的

10 指紋之間的時間不再相同。而是它們將相對於該等查詢指紋之間的時間慢慢地漂移開。為了處理此, 可確保該等查詢指紋之間的時間不會引起 t_{offset} 之偏差超過一臨界值(例如 d_A)。或者, 若具有足夠數目的票數, 則可觀察該等查詢指紋與來自該等搜尋結果的指紋之間的時序相關性, 其應

15 產生一線性趨勢, 其一斜率稍微高於或低於一, 取決於其是否經歷加速或減速, 假設沒有任何其他失真對來自該等搜尋結果的指紋之時序內的錯誤產生貢獻。接著, 我們可使用該線性趨勢(例如, 基於線性迴歸)之置信度之測量決定被組合的結果之置信度。此一線性迴歸之一例子在以下討

20 論的第6圖中被顯示。

類似於 t_{offset} 之概念, 若一搜尋結果內的BER非常低(其表示一非常接近的匹配), 則該搜尋結果之正確性應被高度信任。然而, 若BER高, 則結果較不可信。此外, 當多於一則音訊/視訊的附屬資訊被用於一邏輯AND方式時, 並不

限制只對它們使用固定的臨界值。例如，若該臨界值 d_A 一開始為0.2秒且BER臨界值為12%，但是一目前的搜尋之BER極低(例如0.1%)，則可將該臨界值 d_A 從0.2秒放寬至(比如)0.3秒，具有一強置信度。這可被稱為動態臨界。此

5 動態臨界之調整的確切量可由實施者決定，使用機率分析以確保自此動態臨界產生的錯誤匹配之數目是可忽略的。

若該結果在組合三個搜尋之後是非決定性的，則不是放棄且斷言該摘錄不屬於該資料庫，而是可執行較多搜尋且組合所有結果。例如，在第4(b)圖之情形下，可執行多於兩個

10 搜尋。若該摘錄實際屬於A則，但是第二搜尋不知未何返回錯誤的時間位置，則在有多於兩個搜尋之情況下，該兩搜尋中的一者非常可能也返回具有正確時間位置的A則。然而，當組合五個搜尋時，一投票方案之定義變得較複雜。在一簡化的實施態樣之例子中，可定義返回相同音訊/視訊且按照

15 時間位置是最一致的兩個搜尋(五個中的)作為獲勝者，以“min”及“argmin”運算符被用於以上偽碼之方式。

因此，依據本發明之層面，指紋搜尋正確性可藉由組合來自多數個搜尋的結果而被改良。較佳地，依據本發明之其他層面，此等結果被智慧地組合。

20 在以上所描述的本發明之層面中，一種基於搜尋結果內的可用資訊之投票方法被用以改良指紋搜尋正確性。主要可用的資訊是被搜尋的音訊/視訊(查詢摘錄)之(假定的)名稱。附屬可用資訊(可實質上改良組合該等結果之正確性)包括如 t_{offset} (如以上所描述的)或 $t_{distance}$ 以及一距離之測量

(例如，BER)此類的參數。該參數 t_{distance} 表示該查詢音訊/視訊與該(等)被找出的音訊/視訊之間的時間位置之線性：該等時間位置越接近一直線(可藉由線性迴歸獲得)，則關於該等結果之匹配應越可信任。該 t_{distance} 參數在以下結合第6
5 圖被進一步描述。

類似地，自漢明距離得到的BER表示被找出的音訊/片段與該查詢音訊/片段有多不同。BER越高，則它們彼此越不同，以及BER越低，則它們相對於彼此較少不同且較類似。若所有其他都相同，則具有一較低BER的匹配應該比
10 具有一較高BER的匹配較值得確信的。

因此，附屬資訊之該等值表現它們對應的搜尋結果之置信度。一種用以智慧地組合它們的方法(在此被描述)是在組合該等結果期間使用附屬資訊之個別值的置信等級作為權重。第5圖顯示了以BER之一函數的置信等級之一例子。
15 可稱此函數為 $f_B(\text{BER})$ 。

雖然在先前例子中，BER被映射到一置信等級，但是此差值之測量對於本發明之不是關鍵的且其他形式的差值測量也可被使用。一指紋可被視為數字之一向量。在漢明距離及BER之情形下，其被視為一二進制數字之一向量
20 (即，一二進制向量)。此一向量之每一維僅一位元，但是可設想除了二進制之外的其他數字之指紋向量，且每一維之大小甚至在不同的維之間可以是不同的，雖然其應是可預測的。因此，較一般地，差值之測量可以是一 L_p 範數測量，其中組成一指紋的一向量之每一維上的距離之絕對值之 p

次方之總和的 p 次根被獲得。該BER基於漢明距離，其是一
L1範數差值測量(每一維上的差值之絕對值的總和，每一維
是1位元)。也可有L2範數差值測量(其中差值之絕對值的平
方和之平方根被獲得)，雖然當每一維是1位元時，L2範數
5 類似L1範數。也可具有藉由自相關的差值測量，其中一較
高值表示較接近的匹配。

類似地，可測量一資料庫內的與一查詢指紋之一可能
匹配之時間位置離被假設的位置多遠，如(例如)藉由線性迴
歸所預測的。這在第6圖中被顯示，其中 $time_{query}$ 是查詢指
10 紋之時間位置，以及 $time_{found}$ 是被找出的音訊/視訊內的一可
能匹配的指紋之時間位置，且該兩個值對被顯示為第6圖中
的菱形。接著線性迴歸在該等菱形點上被執行。每一菱形
點與其產生的線性迴歸預測(直線)之偏差被稱為時間位置
距離或 $t_{distance}$ (被顯示為第6圖中的垂直線)。接著可將此距
15 離映射到一置信等級，如第7圖中所示。我們可稱此函數為
 $f_i(t_{distance})$ 。任何適合類型的線性迴歸可被使用。用以移除離
群值的各種技術可在執行線性迴歸之前被使用或者作為線
性迴歸程序之部分。

如以上所提到的，該 $t_{distance}$ 參數是該參數 t_{offset} 之較一般
20 的形式。一使用該 t_{offset} 參數之例子在第4a及4b圖中被顯
示。其測量該查詢摘錄內的一時間位置與一匹配內的對應
時間位置之間的偏移。若連續的指紋匹配是正確的，則其
應大約為一常數，但是當沒有加速或減速時，並不需要接
近零(因為查詢內容之不確定的絕對時間)值。該參數 $t_{distance}$

較一般，如第6圖中所示。其測量一樣本點(其X及Y座標分別是該查詢摘錄與該匹配內的時間位置)與藉由任何各種適合形式的線性迴歸所獲得直線之間的垂直距離。若其對應一良好匹配，則該 t_{distance} 參數被認為接近零。 t_{distance} 之概念甚至在具有加速或減速時也可用，因為其使用線性迴歸且其沒有假設線性迴歸之斜率為1，如同 t_{offset} 參數所執行的。也應注意到的是，因為 t_{offset} 參數假設-1之斜率，所以可能對僅兩個時間位置檢查 t_{offset} 之一致性。然而，當使用 t_{distance} 時，因為沒有假設斜率為1，所以至少需要三個時間位置以獲得一非顯然的線性迴歸(雖然若期望，可使用至少兩個時間位置以獲得一線性迴歸，接著丟棄遠離1之斜率的任何斜率)。

注意到在第7圖中，該置信等級比第5圖中降落得快得多。這是因為我們預期在典型的實施態樣中，BER之容限可能相當高，而 t_{distance} 之任何非顯然量是該結果較不可信賴得多的一指示。除此之外，因為兩個點總是產生完美通過它們的一直線，所以基於 t_{distance} 的置信等級之使用應只在具有三個或更多個點之情況下被考慮。

依據本發明之層面的用於組合置信等級且智慧地決定搜尋結果之結果的一程序之一例子在以下偽碼中被顯示。

用於組合置信等級且決定搜尋結果之結果的示範性偽碼可被表示如下：

```

For (X in set of unique pieces in the results) {
  For(i in where search[i].piece==X) {
    Confidence1=fB(search[i].BER);
  }
}

```

```

        If(the set where results[i].piece==X has>=3 points)
            Confidence2= ft(search[i].t_distance);
        Else
            Confidence2=0;
5         Total_score[X]+=Confidence1+Confidence2;
        }
        Pick X for which Total_score[X] is the highest;
            If (Total_score[X]>threshold)
                Return "the query excerpt is piece X";
10        Else
            Return "unknown query excerpt";

```

如該偽碼中所指示的，首先將BER及 $t_{distance}$ 映射到它們對應的置信等級，接著組合它們。在此例子中，該組合僅僅是一求和運算，雖然可使用其他運算，例如乘法(於是該等置信等級必須為非負)，只要該組合操作相對於每個置信等級是一單調非減少函數。若 $f_B()$ 總是返回1，且 $f_t()$ 總是返回0，且變數“threshold”具有值 $N/2$ (其中N是被執行的搜尋之數目)，則以該偽碼表示的程序被簡化為一簡單的多數決方案。實際上，應選擇該等函數 $f_B()$ 及 $f_t()$ 使得它們最大化最後輸出是正確的機率。

實施態樣

本發明可以硬體或軟體或其二者之一組合(例如，可規劃邏輯陣列)實施。除非被另外指明，否則被包括為本發明之部分的方法或演算法並不固有地與任何特定電腦或其他裝置相關。特別地，各種通用機器可與依據此處的教示撰寫的程式一起被使用，或者製造較專用裝置(例如，積體電路)以執行所需的方法步驟可能是較方便的。因此，本發明

可以在一或多個可規劃電腦系統上執行的一或多個電腦程式實施，該等可規劃電腦系統各自包含至少一處理器、至少一資料儲存系統(包括依電性及非依電性記憶體及/或儲存元件)、至少一輸入裝置或埠以及至少一輸出裝置或埠。

- 5 程式碼被施加給輸入資料以執行本文所描述的功能且產生輸出資訊。該輸出資訊以已知方式被施加給一或多個輸出裝置。

每個此程式可以任何期望的電腦語言(包括機器、組合或高階程序、邏輯或以物件為導向的程式語言)實施以與一
10 電腦系統進行通訊。在任何情況下，該語言可以是一編譯或解譯語言。

每個此電腦程式較佳地被儲存在或下載到一可由一般或通用可規劃電腦讀取的儲存媒體或裝置(例如，固態記憶體或媒體或者磁或光學媒體)以當該儲存媒體或裝置被該
15 電腦系統讀取時組配及操作該電腦以執行本文所描述的程序。本發明系統也可被認為需被實施為被組配一電腦程式的電腦可讀儲存媒體，其中被如此組配的儲存媒體使一電腦系統以一特定及預定方式操作以執行本文所描述的功能。

20 本發明之一些實施例已被描述。然而，將明白的是，在不背離本發明之精神及範圍下，可進行各種修改。例如，以上所描述的一些步驟可以與順序無關，因此可以不同於所描述的順序被執行。因此，其他實施例仍落於以下申請專利範圍之範圍內。

【圖式簡單說明】

第1圖是一特定類型的指紋可藉以產生的方式之一示意性例子；

第2圖顯示了以一個單一指紋匹配搜尋之三個可能的結果中的每個可能產生的一機率密度函數(PDF)為位元錯誤率(BER)之函數的一理想的例子；

第3圖是利用關於三個搜尋的多數決的進行指紋匹配搜尋結果組合之一例子的一示意性概念；

第4a圖是關於三個搜尋的指紋匹配搜尋結果之一例子的示意性概念，其中時間位置對於三個可能的匹配中的兩者是一致的；

第4b圖是關於三個搜尋的指紋匹配搜尋結果之一例子的示意性概念，其中時間位置對於三個可能的匹配中的兩者是不一致的；

第5圖是一置信等級可藉以被映射為位元錯誤率(BER)之一函數的方式之一例子；

第6圖是用於理解該指紋查詢之時間與一資料庫內找到的指紋之時間之間的相對時間位置可被決定之方法的一理想化例子，在此例中使用線性迴歸；以及

第7圖是一置信等級可藉以被映射為該指紋查詢之時間與一資料庫內找出的指紋之時間之間的相對時間位置內的時間差值之一函數的方式之一例子。

【主要元件符號說明】

(無)

十、申請專利範圍：

1. 一種用於識別一音訊及/或視訊資訊之片段的方法，包含以下步驟：

獲得該片段內在多個隔開的時間位置中各個位置處的一查詢指紋；

針對每個此查詢指紋搜尋在一指紋資料庫內之指紋以尋求一可能的匹配；

針對每個此查詢指紋，獲得對該資料庫內的一被找出指紋之一可能的匹配之一置信等級，其中該置信等級係該查詢指紋之時間位置與該被找出指紋之時間位置之間的相對時序關係之一函數；以及

組合對於可能的匹配之搜尋結果，以識別該音訊及/或視訊資訊之片段，或推斷該片段不屬於該資料庫，其中每個可能的匹配之結果係以一個別置信等級予以加權。

2. 如申請專利範圍第1項所述之方法，其中該置信等級亦是該查詢指紋與該被找出指紋之間的差異之一測量值的一函數。

3. 如申請專利範圍第2項所述之方法，其包含下列步驟：

若一搜尋所具查詢指紋與被找出指紋間之相對時序上的一最小差異小於一臨界值，則從該一搜尋識別該音訊及/或視訊之片段，其中該臨界值取決於識別出音訊及/或視訊之相同片段的搜尋數目而改變。

4. 如申請專利範圍第2項所述之方法，其中：

該查詢指紋之時間位置與該被找出指紋之時間位置之間的該相對時序關係為時間位置上之一差值，及

若該等查詢指紋與被找出指紋之間的時間上差值係一最小值且小於在查詢指紋與被找出指紋間之差異的該測量值為低時所產生的一動態臨界值，則查詢指紋即被視為識別出一音訊及/或視訊資訊之片段。

5. 如申請專利範圍第1至4項中任一項所述之方法，其中該等時間位置被隔開以使查詢指紋之間或之中的該等搜尋結果之相關性最小化。
6. 如申請專利範圍第5項所述之方法，其中有多組隔開的時間位置，且對於各該等多組之搜尋結果被組合。
7. 如申請專利範圍第6項所述之方法，其中該等搜尋結果在每一組內被組合，且此等組合之結果被進一步組合。
8. 如申請專利範圍第6項所述之方法，其中該等搜尋結果跨越所有組而組合。
9. 如申請專利範圍第6項所述之方法，其中有至少三個隔開的時間位置。
10. 如申請專利範圍第9項所述之方法，其包含使用該查詢指紋之時間位置與在該等時間位置之該可能匹配被找出指紋之時間位置之間的相對時序關係之線性趨勢的置信度之測量值，來決定被組合的該等搜尋結果之置信度。
11. 如申請專利範圍第1至4項中任一項所述之方法，其中：
該等搜尋結果係藉由一投票方案予以組合，

在非決定性投票之情況下，該組合係由多個搜尋之額外疊代予以級聯，及

投票僅在新的疊代中被考慮或者與一些或所有先前的投票予以組合。

- 5 12. 一種適於執行用於識別音訊及/或視訊資訊之片段的方法之裝置，該方法係如申請專利範圍第1至11項中任一項之方法，該裝置包含：

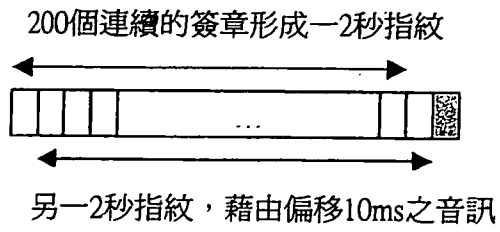
○ 用以獲得該片段內在多個隔開的時間位置中各個位置處的一查詢指紋之構件；

- 10 用以針對每個此查詢指紋搜尋在一指紋資料庫內之指紋以尋求一可能的匹配之構件；

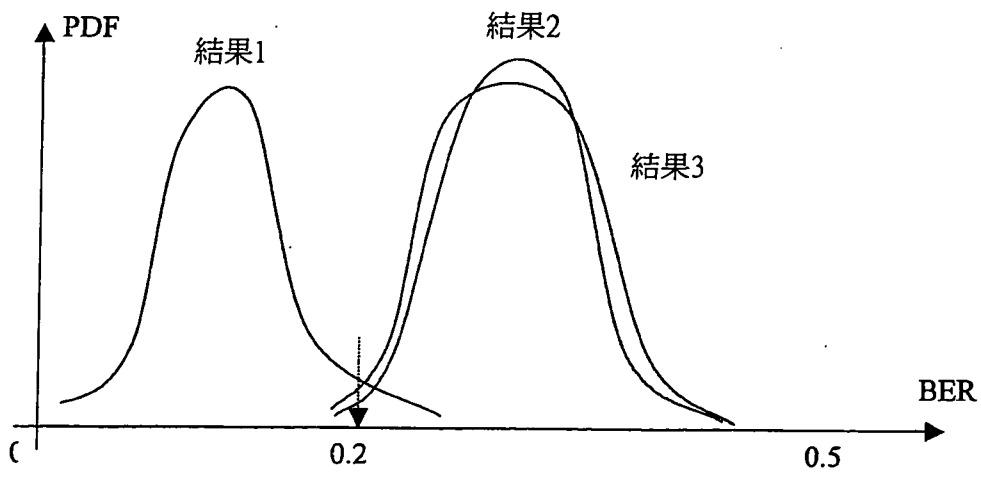
用以針對每個此查詢指紋獲得對該資料庫內的一被找出指紋之一可能的匹配之一置信等級之構件，其中該置信等級係該查詢指紋之時間位置與該被找出指紋之時間位置之間的相對時序關係之一函數；以及

- 15 ○ 用以組合對於可能的匹配之搜尋結果以識別該音訊及/或視訊資訊之片段或推斷該片段不屬於該資料庫之構件，其中每個可能的匹配之結果係以一個別置信等級予以加權。

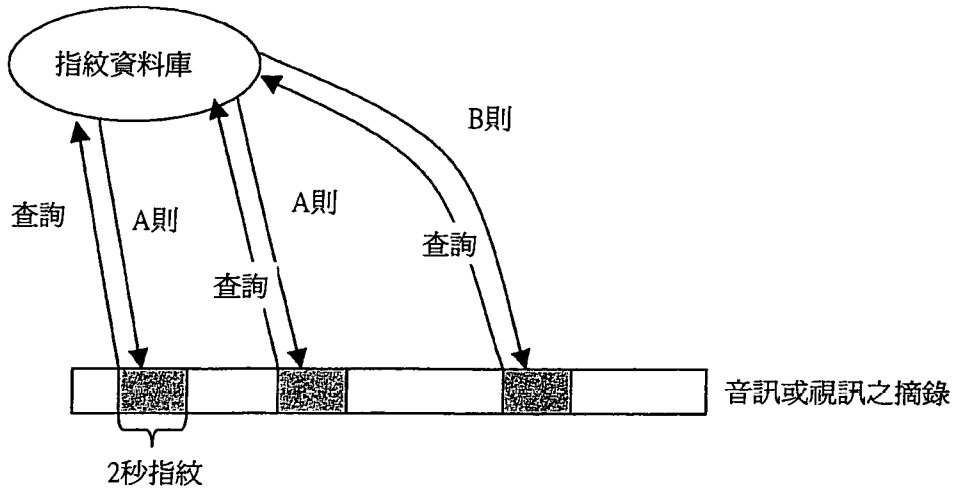
- 20 13. 一種儲存有電腦程式之電腦可讀媒體，該電腦程式係用以致使一電腦執行如申請專利範圍第1至11項中任一項之方法。



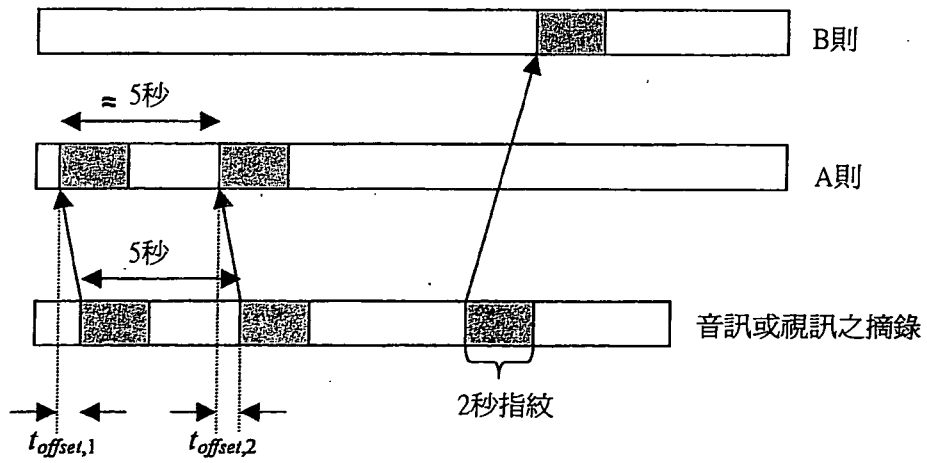
第 1 圖



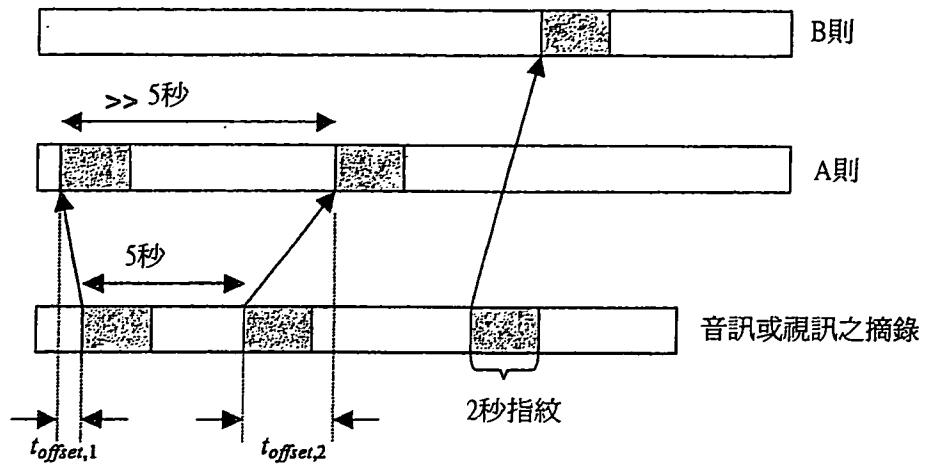
第 2 圖



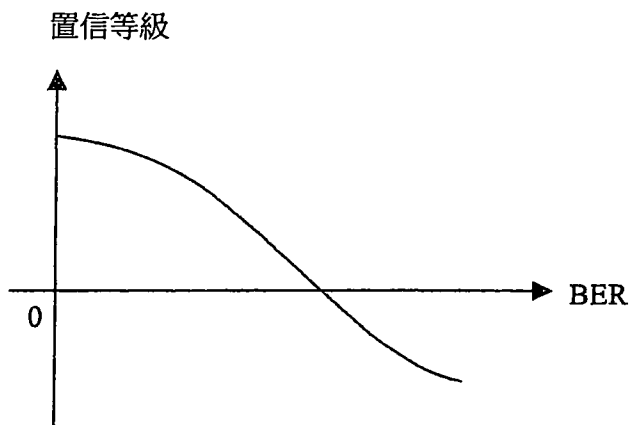
第 3 圖



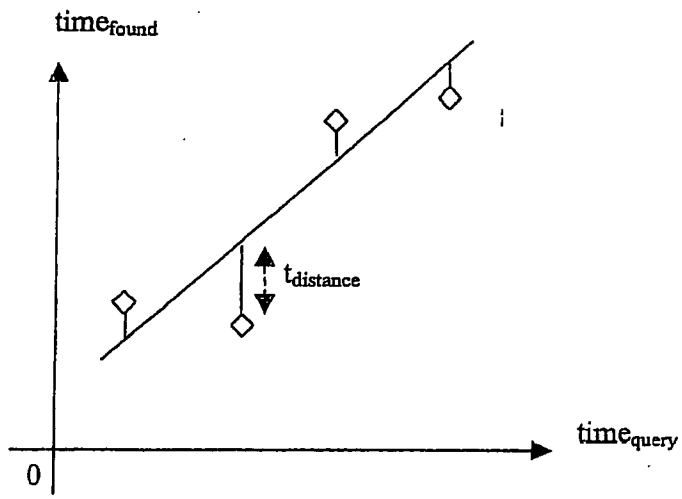
第 4a 圖



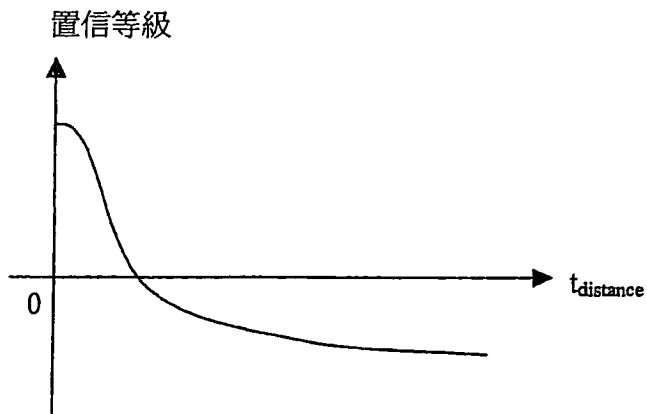
第 4b 圖



第 5 圖



第 6 圖



第 7 圖