



(19) 中華民國智慧財產局

(12) 發明說明書公告本

(11) 證書號數：TW I455064 B

(45) 公告日：中華民國 103 (2014) 年 10 月 01 日

(21) 申請案號：097147218 (22) 申請日：中華民國 97 (2008) 年 12 月 05 日
 (51) Int. Cl. : G06T5/00 (2006.01) G10L19/00 (2013.01)
 (30) 優先權：2007/12/20 法國 0760122
 (71) 申請人：湯姆生特許公司 (法國) THOMSON LICENSING (FR)
 法國
 (72) 發明人：拉梅爾 奧利維 LE MEUR, OLIVIER (FR)；尼納西 亞歷山卓 NINASSI,
 ALEXANDRE (FR)；雪維 尚克勞迪 CHEVET, JEAN-CLAUDE (FR)
 (74) 代理人：陳詩經
 (56) 參考文獻：
 TW 200719319A TW 200738041A
 US 2005/0078172A1 US 2005/0141728A1
 US 2006/0233442A1
 審查人員：林文琦
 申請專利範圍項數：8 項 圖式數：12 共 0 頁

(54) 名稱

聲影文件突起映圖之決定方法和裝置

METHOD AND DEVICE FOR CALCULATING THE SALIENCE OF AN AUDIO VIDEO DOCUMENT

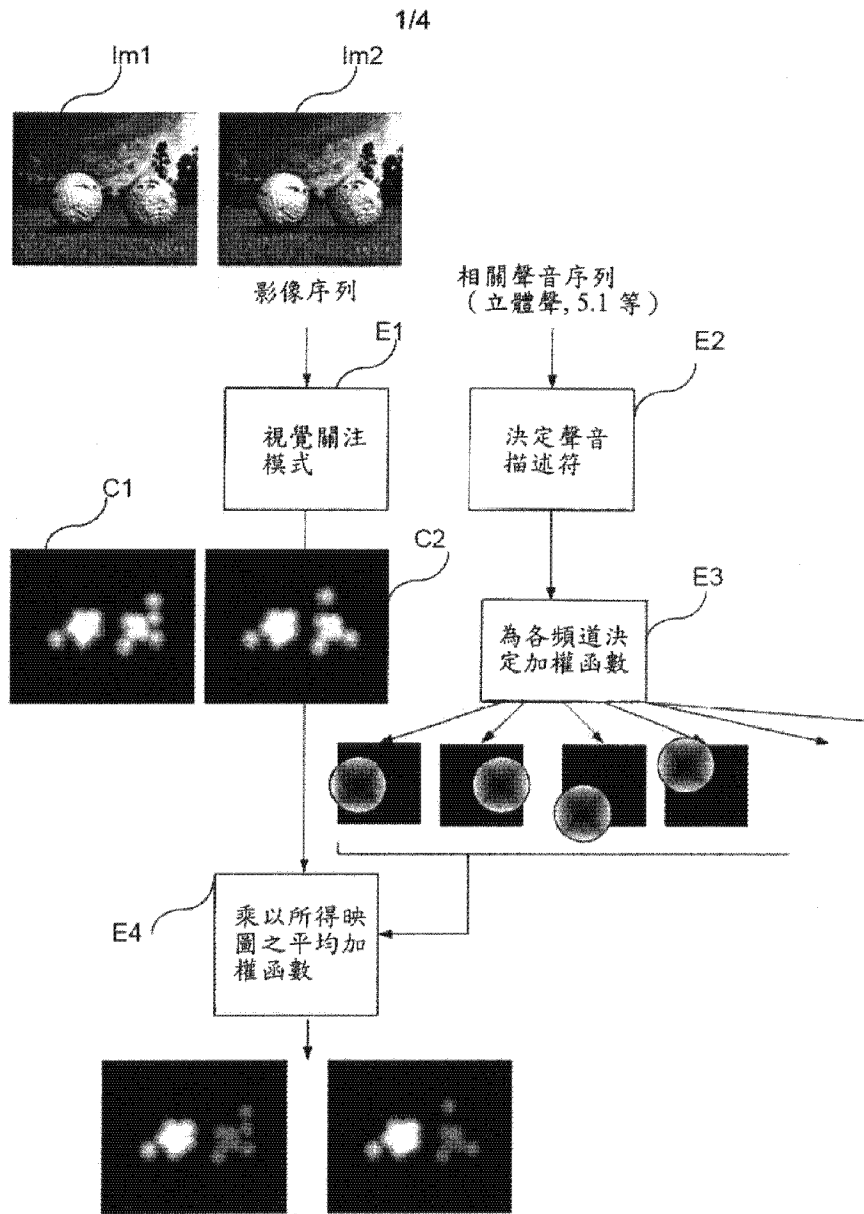
(57) 摘要

本發明係關於聲影文件突起映圖之決定方法，包括如下步驟：

- 計算(E1)突起映圖與該文件影像部份之關係；
- 決定(E2)展示該文件之至少一聲音描述符；
- 以該聲音描述符變化為函數，加權(E4)影像突起映圖。

The invention relates to a method for determining the salience map of an audio video document. The method comprises the steps for:

- calculating (E1) a salience map in relation to the video part of said document,
- determining (E2) at least one audio descriptor representing said document,
- weighting (E4) the video salience map as a function of the variation of said audio descriptor.



C1,C2 . . . 突起映圖
lm1,lm2 . . . 突起映圖

E1 . . . 利用影像文件從視覺關注模式決定突起映圖之步驟

E2 . . . 決定代表聲影文件的聲音內容的聲音描述符之步驟

E3 . . . 由聲音描述符決定加權函數之步驟

E4 . . . 乘所得映圖平均加權函數之步驟

圖 1

發明專利說明書

(本說明書格式、順序，請勿任意更動，※記號部分請勿填寫)

※申請案號：97147218

※申請日：97.12.5

※IPC 分類：G06T5/00 (2006.01)

G10L19/00 (2006.01)

一、發明名稱：(中文/英文)

聲影文件突起映圖之決定方法和裝置

METHOD AND DEVICE FOR CALCULATING THE SALIENCE OF
AN AUDIO VIDEO DOCUMENT

二、中文發明摘要：

本發明係關於聲影文件突起映圖之決定方法，包括如下步驟：

- 計算 (E1) 突起映圖與該文件影像部份之關係；
- 決定 (E2) 展示該文件之至少一聲音描述符；
- 以該聲音描述符變化為函數，加權 (E4) 影像突起映圖。

三、英文發明摘要：

The invention relates to a method for determining the salience map of an audio video document. The method comprises the steps for:

- calculating (E1) a salience map in relation to the video part of said document,
- determining (E2) at least one audio descriptor representing said document,
- weighting (E4) the video salience map as a function of the variation of said audio descriptor.

四、指定代表圖：

(一)本案指定代表圖為：第 (1) 圖。

(二)本代表圖之元件符號簡單說明：

C1,C2 突起映圖

lm1,lm2 突起映圖

E1 利用影像文件從視覺關注模式決定突起映圖之步驟

E2 決定代表聲影文件的聲音內容的聲音描述符之步驟

E3 由聲音描述符決定加權函數之步驟

E4 乘所得映圖平均加權函數之步驟

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

六、發明說明：

【發明所屬之技術領域】

本發明係關於聲影文件突起映圖之計算裝置和方法。本發明尤指辨識聲音特徵，以決定聲影文件之突起。

【先前技術】

決定聲影文件的突起區，可用於不同用途，諸如壓縮（突起區是例如以較多數位元寫碼，或該區之量化步驟經修飾）、定索引（可用此等突起區內的資訊，定文件指數）等。

有若干方法供決定影像文件之突起。Thomson Licensing 在 2004 年 12 月 14 日申請的歐洲專利 EP1695288，記載建立影像文件突起映圖特別有效方法。

惟在某些情況下，影像不能以充分正確方式獲得聲影文件之突起。

【發明內容】

為改進現時方法之適切性，本發明擬議令聲音特徵與視覺特性關聯。

為此目的，本發明擬議聲影文件突起映圖之決定方法，其特徵為，包括下列步驟：

- 計算突起映圖與文件影像部份之關係；
- 決定至少一聲音描述符，代表文件；
- 以聲音描述符變化為函數，加權影像突起映圖。

聲音描述符宜從下列選擇：

- 文件之總體音響水準；
- 感音環境；
- 各聲音頻道之音響水準；
- 音響水準之時間性變化；
- 音響水準之空間 / 時間性變化；

以及此等成份之任何組合。

更好是，當各聲音頻道之音響水準和總體音響水準，用做聲音描述符，

—各聲音頻道之增益，是以音響水準和全部頻道的音響水準為函數，加以計算 (E3)。

有益的是，

—加權視窗是對各圖元決定，具有大於預定水準之視覺突起，稱為視覺突起圖元；

—視覺突起圖元位置之空間性位置，在圖像內決定；

—各頻道的加權函數，係以各聲音頻道的增益和圖像內圖元的空間性位置增益之函數，為各視覺突起圖元計算。

按照較佳具體例，

—加權函數是以圖像內空間性位置和聲音頻道增益值之函數，為各頻道和圖像之各圖元計算。

各頻道之加權函數，以線性函數為佳，使位於頻道側之圖元，加權比位於空間性遠離頻道之圖元為大。

若聲音為立體型，位於圖像左方的圖元對左方頻道的加權函數有較高加權，而對位於圖像右方的圖元加權較低。

按照較佳具體例，

—突起映圖是相對於影像部份，乘以複數加權函數，而得複數突起映圖；

—聲影突起映圖是由該突起映圖平均而得。

突起映圖宜以視音響水準的時間性變化而定之係數加權。

本發明亦關於聲影文件突起映圖之決定裝置，包括之機構供：

—計算突起映圖與該文件影像部份之關係；

—決定至少一聲音描述符代表文件；

—以聲音描述符變化為函數，加權影像突起映圖。

【實施方式】

本發明參照附圖所示非限制性具體例和實施之說明，即可更為明白。

圖示模組是功能性單位，相當於實體有別之單位與否均可。例如，此等模組或其部份，可在單一組件內集合在一起，或構成同樣軟體之功能。反之，有些模組可由分開之實體組成。

第 1 圖表示本發明較佳模態之具體例，應用於影像序列，表示一高爾夫球和另一高爾夫球在講話。

在步驟 E1，突起映圖是利用影像文件由視覺關注模式決定。如此所得突起映圖，是在圖示 C1 和 C2 中分別為各圖像 $lm1$ 和 $lm2$ 決定。在所得二突起映圖 C1 和 C2 觀察到，與各球關聯之突起很相似。誠然，二球有很像的特性，惟與此等球關聯之音響很不同。誠然，右方球在聽，而左方球在講。觀看聲影文件的人員關注，自然聚焦在左方球上。所得突起映圖 C1 和 C2，即不足以代表聲影文件。

因此，在步驟 E2 中，決定聲音描述符，代表聲影文件之聲音內容。

在描述符當中，來自聲音磁軌之描述符係如下決定：

- 各頻道 (CSL) (右(r)、左(l)、底(b)、頂(t)、後右(rl)、後左(bl)等) 之音響水準；
- 總體音響位準 (OSL) (沉默期、活性期)，即 CSL 頻道音響水準之總體合計；
- 感音環境 (AE)；
- 音響水準的時間性變化 (SLTV) (從沉默通到強烈活性)；
- 音響水準的空間 / 時間性變化 (SLSTV) (從左通到右等)。

由此等聲音描述符，在步驟 E3 中決定加權函數。此加權函數使各圖元與較高或較低權關聯，視是否需要強調或減輕突起而定。所以，此函數有支援，以圖像的解像度為維度。利用從缺值，即無聲音資料可用時，加權函數均勻，即全圖像之值均為"1"。

考量各頻道之聲音水準

若立體音響可行，則有二頻道，圖像各側有一。在編校影像上之音響時，若需強調右側，例如人在圖像右側說話，而左側人員不說話，則可對右側給予比左側更多的音響。

因此，音響不一定賦予圖像全體，而是可以更為局部性。例如就立體音響而言，左右可以截然有別。以 Dolby 5.1 音響而言，左、右、前、後等可以分辨。

各頻道界定加權函數，其增益因聲音遮蔽容量而定。低波幅音響別人聽不見，故載有此音響之頻道增益低。加權函數以單位波幅之高斯為優，稱為 Gauss_σ (σ 代表高斯波封之空間性程度)，具有不同增益。

加權函數之應用中心，可對各頻道利用預定光柵設定。其例如第 2 圖所示（例如有 6 頻道成點狀均勻分佈於圖像上）。

此圖表示 9 點，加全函數即對其中心有此 9 點。

加權函數之應用中心宜位在各象限之突起最大值，如第 3 圖所示。決定只保存預定數突起點時，可用突起超出預定臨限值之突起點，甚至在突起點位置之後，若圖像之縮小區含大部份突起點，且圖像含有若干分佈之突起區，則只保存一突起點。

在其他具體例中，可用 Laplace 型函數，以定中於尺寸固定或可變動的點上之四方形視窗展示。

各聲音頻道的增益如下式所示：

$$\text{增益權值}_i = \frac{CSL_i^p}{\varepsilon + \sum_j CSL_j^p}$$

「增益權值_i」表示頻道 i 之增益，而 CSL_i 表示頻道 i 的音響位準。

P 是冪數，以數值 2 為佳；

ε 是低值常數，防止除以 0。若頻道有同樣音響位準或

很接近的音響位準，則與各頻道關聯之增益相同。然而，頻道的音響與其他頻道的音響相較愈高者，增益愈大。

各頻道增益乘以先前界定的加權視窗增益，稱為權值。此加權視窗的波幅視頻道的增益而定。

$$\text{「增益權值 } 2_i\text{」} = \text{「增益權值 } i\text{」} * \text{「權值」}$$

因此，當加權函數之應用中心位於突起最大值，則此最大值周圍的加權視窗如第 3 圖所界定。在立體模態中，突起最大值的位置對應用此加權函數具重要性。誠然，以位於頻道右方的突起最大值而言，重點在對其應用右方頻道之「增益權值」，反之，對位於圖像左方的最大值，重點在對右方頻道應用「增益權值」。因此，突起最大值之距離，係就圖像的邊緣計算。以右方頻道之加權函數言，對位於接近圖像右方邊緣的最大值，應用高「增益權值」係數，而對位於接近圖像左方邊緣之圖元，給予低係數。

以左方頻道的加權函數言，對位於接近圖像左方邊緣的最大值，應用高「增益權值」係數，而對位於接近圖像的右方邊緣之圖元，給予低係數。

當聲音並非立體聲而是 Dolby 5.1，則距離並非就圖像的邊緣，而是圖像內之位置計算。按照此位置，決定頻道對此點之影響。

關於距離，於前述方程式：「增益權值 2_i 」 = 「增益權值 i 」 * 「權值」，可增加另一加權因數。

因距離概念而立的方程式，則變成：

$$\text{「增益權值 } 2_i\text{」} = \alpha * \text{「增益權值 } i\text{」} * \text{「權值」}$$

α 視相對於中心 (x_c, y_c) 的偏心率而定，即現時點 (x, y) ， $\alpha(x, y) = (x - x_c)^2 - (y - y_c)^2$ 。故，若 $(x, y) = (x_c, y_c)$ ，則 α 趨向 0，否則 α 隨與中心的距離而遞增。

第 4a 至 7b 圖表示在立體音響情況下，增益權值 2 之不同具體例，包括右方頻道和左方頻道。此等具體例只有在上述視覺映圖的最大值突起不變，但在圖像之所有點之突起均

改變。

第 4a 圖表示右方頻道之第一具體例。「增益權值 $2d$ 」是位於圖像右方諸點之最大值，對中心點和位於中心點左方的圖像點變成等於 1。

第 4b 圖表示左方頻道之第一具體例。「增益權值 $2g$ 」是位於圖像左方諸點之最大值，對中心點和位於中心點右方的圖像點變成等於 1。「權值」即代表線性函數。

此線性函數預先界定，其變化可視使用者需要之調變強度而定。

在第 4a 和 4b 圖中，亦知左方頻道的增益比右方頻道增益大。因此，當乘以全部加權函數時，左方諸點比右方諸點可得更強調之突起。

第 5a 和 5b 圖表示在左、右有對數型下降。

第 5a 圖表示右方頻道之第二具體例。「增益權值 $2d$ 」是位於圖像右方諸點的最大值，對中心點和位於中心點左方的圖像點變成等於 1。「權值」即代表遞增冪數函數 $\exp(x)$ 。

第 5b 圖表示左方頻道之第二具體例。「增益權值 $2g$ 」是位於圖像左方諸點的最大值，對中心點和位於中心點右方的圖像點變成等於 1。「權值」即代表遞降冪數函數 $\exp(x)$ 。

此遞增或遞降冪數函數經預先界定，變化係視使用者需要的調變強度而定。

第 6a 圖表示右方頻道之第三具體例。「增益權值 $2d$ 」是位於圖像右方諸點之最大值，對中心右方的圖像區，以及例如位於圖像右方四分之一內，保持常值。否則其值為 "1"。

第 6b 圖表示左方頻道之第三具體例。「增益權值 $2g$ 」是位於圖像左方諸點之最大值，對中心左方的圖像區，以及例如位於圖像左方四分之一內，保持常值。否則其值為 "1"。

此項函數經預界定，變化可視使用者需要的調變強度而定。

第 7a 圖表示右方頻道之第四具體例。「增益權值 $2d$ 」是

位於圖像右方諸點之最大值，線性遞減至圖像左方。

第 7b 圖表示左方頻道之第四具體例。「增益權值 $2g$ 」是位於圖像左方諸點之最大值，線性遞減至圖像右方。

此二最後曲線的坡度一致。右方頻道增益比左方頻道增益為大，位於圖像右方諸點增加其突起，比位於圖像左方諸點為大。此函數經預先界定，變化可視使用者所需調變強度而定。

於步驟 E4 當中（第 1 圖），視覺突起映圖即乘以全部 N 加權函數。所得 N 映圖經平均而得中間映圖 $SMod$ 。

設 SM 為視覺突起映圖，

設 $SMod$ 為聲影突起映圖，

設 N 為音響頻道，

設 $Norm$ 為比例因數，

則由下式可得聲影突起映圖：

$$SMod = SM(1 + \sum_{i=0}^{N-1} \text{增益權值} 2_i^{NORM})$$

，其中增益權值 2_i^{NORM}

是「增益權值 2_i 」值利用加權視窗值合計加以常態化者。

考量音響水準之時間性變化

於步驟 E5 當中，考量音響水準之時間性變化。

當音響水準突然變化，尤其是水準從低水準前往高水準，視覺關注提升，在高水準。而且，當通過突起時，視覺關注當然較不強烈。

所以，對時間上衰減係數，擬在時間上應用突起之調變係數，稱 TAC。第 3 圖表示 TAC 係數以總體音響水準之函數而變化。

時間上衰減係數 TAC 視總體音響水準 OSL 值之時間變化性而定，經數秒整合，其波幅亦經時間性平均化。此係數均勻應用於突起映圖。以從缺值計，為 1。

易言之，在突起期間，突起以趨向 0 之係數加權，故對一區並未賦予勝過另一區之優點。

$$SMod2 = CTA \times SMod$$

又知上述不同具體例所述發明，傾向減少圖像中心之突起。惟有時重點不在減少此突起，而是加以強調。

以從缺值而言，若不同頻道之增益相似（音響平均分配於頻道之間），不是無作為，便是中心以從缺值強調。

上述應用特別應用於二人間之對話脈絡，更一般情況是，圖像中心保持吸引人的優先區。

【圖式簡單說明】

第 1 圖表示本發明具體例和本發明應用於聲影文件之效果；

第 2 圖表示加權函數應用中心之例；

第 3 圖代表較佳具體例，表示應用中心、加權函數、影像映圖之突起點，及其關聯之加權視窗；

第 4a 和 4b 圖分別表示立體聲情況下，加權函數第一變化例之右方和左方頻道；

第 5a 和 5b 圖分別表示立體聲情況下，加權函數第二變化例之右方和左方頻道；

第 6a 和 6b 圖分別表示立體聲情況下，加權函數第三變化例之右方和左方頻道；

第 7a 和 7b 圖分別表示立體聲情況下，加權函數第四變化例之右方和左方頻道；

第 8 圖表示總體音響水準和時間上衰減係數間之關係。

【主要元件符號說明】

C1,C2 突起映圖

lm1,lm2 突起映圖

E1 利用影像文件從視覺關注模式決定突起映圖之步驟

E2 決定代表聲影文件的聲音內容的聲音描述符之步驟

E3 由聲音描述符決定加權函數之步驟

E4 乘所得映圖平均加權函數之步驟

七、申請專利範圍：

1. 一種聲影文件突起映圖之決定方法，其特徵為，包括：

- 計算 (E1) 突起映圖與該文件影像部份之關係；
- 決定 (E2) 至少一聲音描述符，代表該文件；
- 以該聲音描述符之變化為函數，加權 (E4) 影像突起映圖；
- 其中各聲音頻道之音響水準和總體音響水準，用做聲音描述符，各聲音頻道之增益，係以音響水準和全部頻道之音響水準為函數計算；又其中加權視窗係對視覺突起大於預定水準之各圖元決定，稱為視覺上突起圖元，該視覺上突起圖元位置之空間性位置，係在圖像上決定，各頻道之加權函數，係以各聲音頻道的增益和圖像內圖元之空間性位置為函數，為各視覺上突起圖元計算者。

2. 如申請專利範圍第 1 項之方法，其中聲音描述符係從：

- 該文件之總體音響水準；
 - 感音環境；
 - 各聲音頻道之音響水準；
 - 音響水準之時間性變化；
 - 音響水準之空間 / 時間性變化；
- 及上述任何組合選出者。

3. 如申請專利範圍第 1 或 2 項之方法，其中：

- 加權函數係以其在圖像內之空間性位置和該聲音頻道之增益值為函數，對各頻道和圖像之各圖元計算者。

4. 如申請專利範圍第 3 項之方法，其中各頻道之該

加權函數係線性函數，使位於頻道側之圖元，加權較位於空間上遠離該頻道之圖元為大者。

5.如申請專利範圍第 4 項之方法，其中聲音係立體聲，位於圖像左方之圖元對左方頻道的加權函數有較大加權，而位於圖像右方的圖像有較低加權者。

6.如申請專利範圍第 1 或 2 項之方法，其中：

—突起映圖係相對於影像部份，乘以多數加權函數，而得複數突起映圖；

—該聲影突起映圖係由該突起映圖平均而得者。

7.如申請專利範圍第 3 項之方法，其中突起映圖以因音響水準的時間性變化而定之係數加權者。

8.一種聲影文件突起映圖之決定裝置，其特徵為，包括：

計算機構，計算突起映圖與該文件影像部份之關係；

決定機構，決定至少一聲音描述符，代表該文件；

加權機構，以該聲音描述符為函數，加權該影像突起映圖，

計算機構，當使用各聲音頻道的音響水準和全體音響水準，做為聲音描述符，以音響水準和全部頻道的音響水準為函數，計算各聲音頻道之增益；

決定機構，決定具有視覺突起大於頻定水準的各圖元，稱為視覺突起圖元之加權視窗，

決定機構，決定該視覺突起圖元在圖像位置之空間位置，

計算機構，以各聲音通道之增益和圖元加權圖像之空間位置為函數，為各視覺突起圖元計算，各通道之加權函數者。

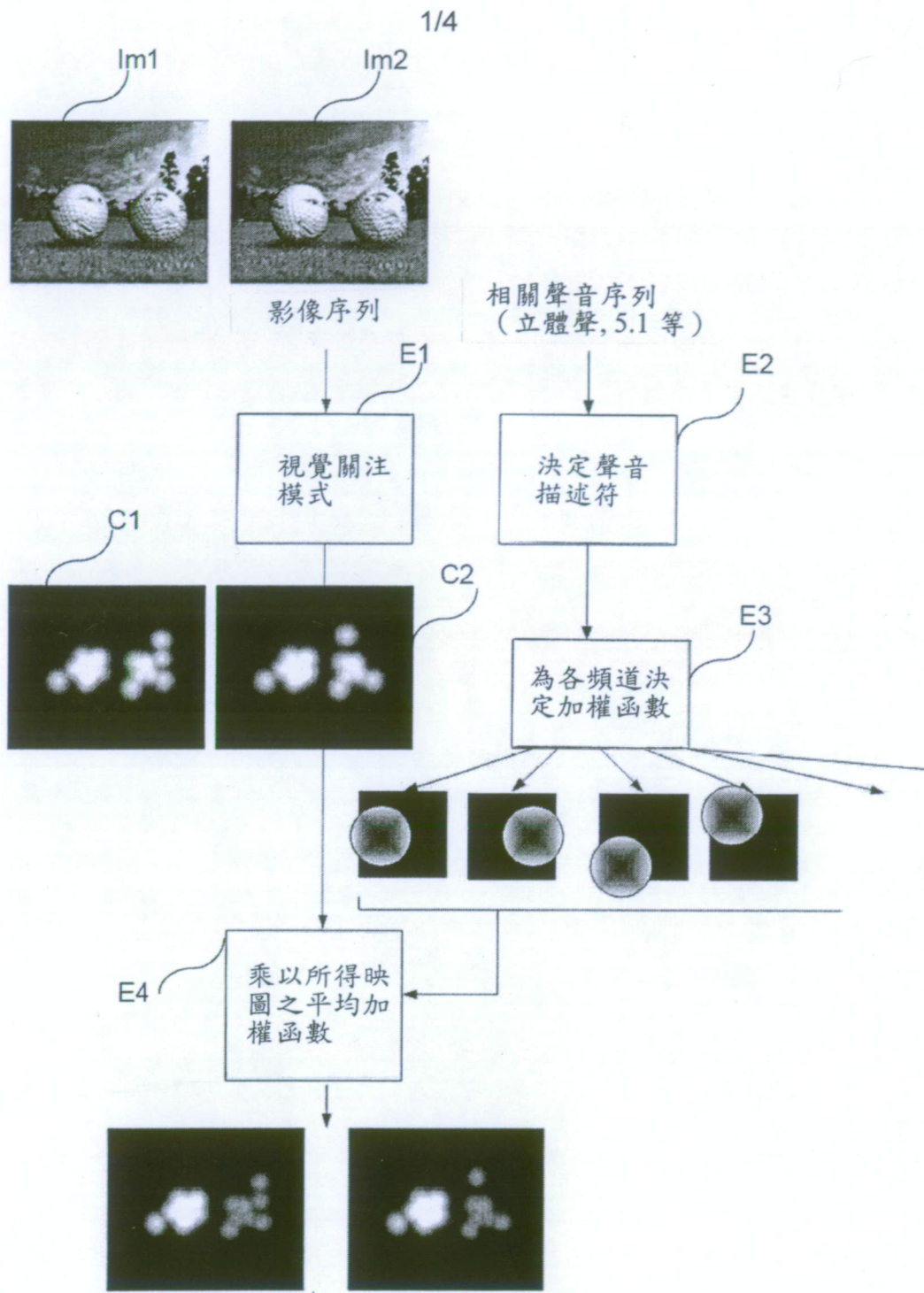


圖 1

視頻明顯映圖

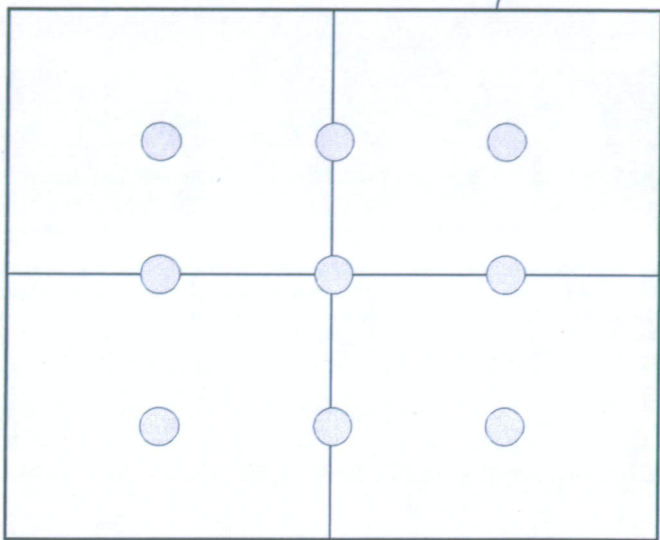
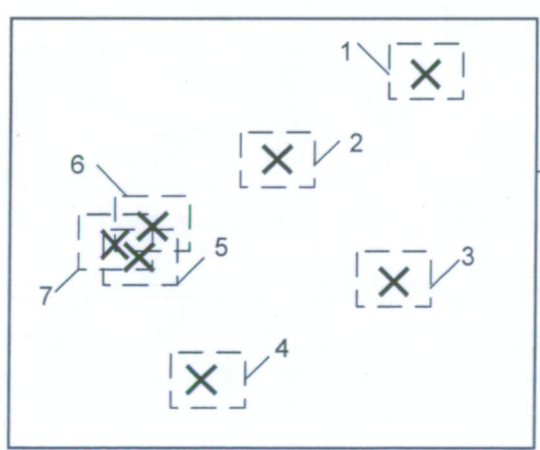


圖 2

視頻明顯映圖



X: 明顯圖元



加權視窗

圖 3

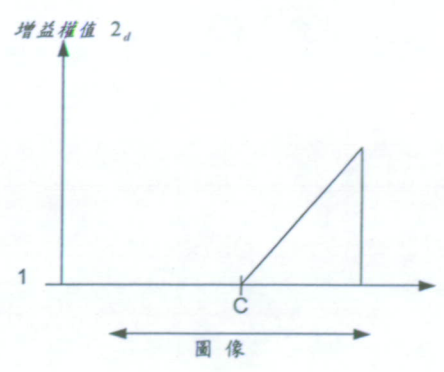


圖 4a

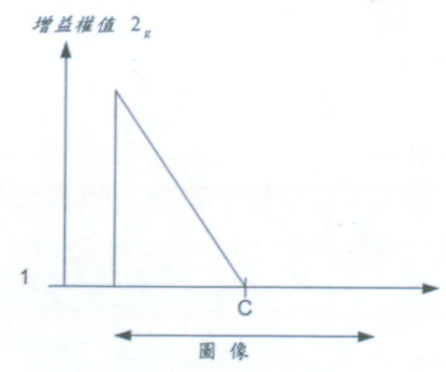


圖 4b

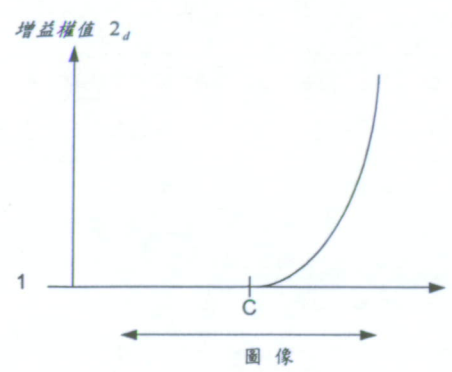


圖 5a

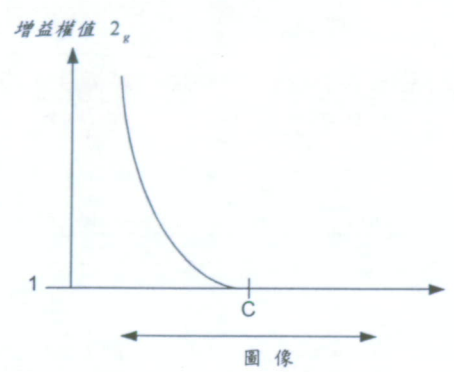


圖 5b

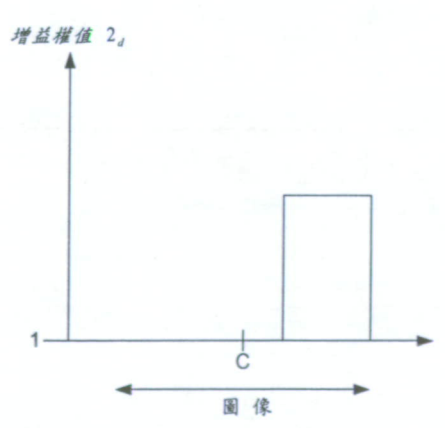


圖 6a

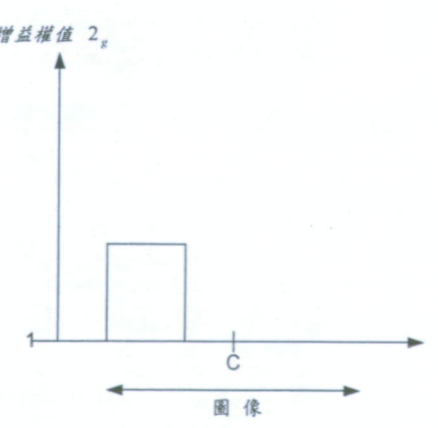


圖 6b

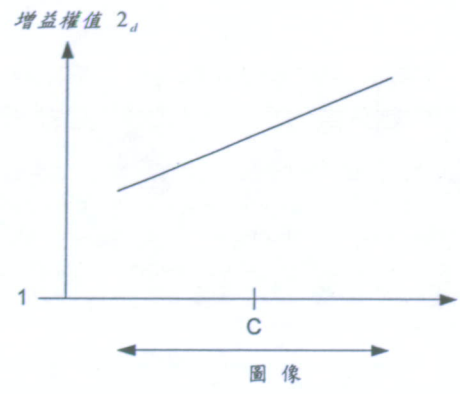


圖 7a

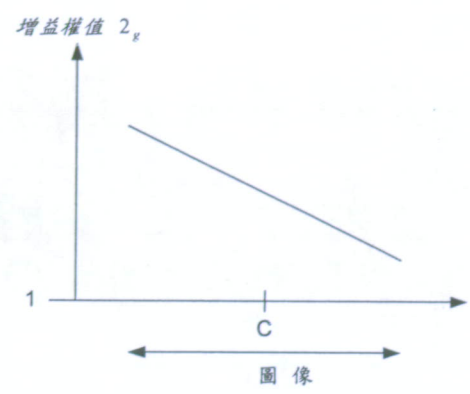


圖 7b

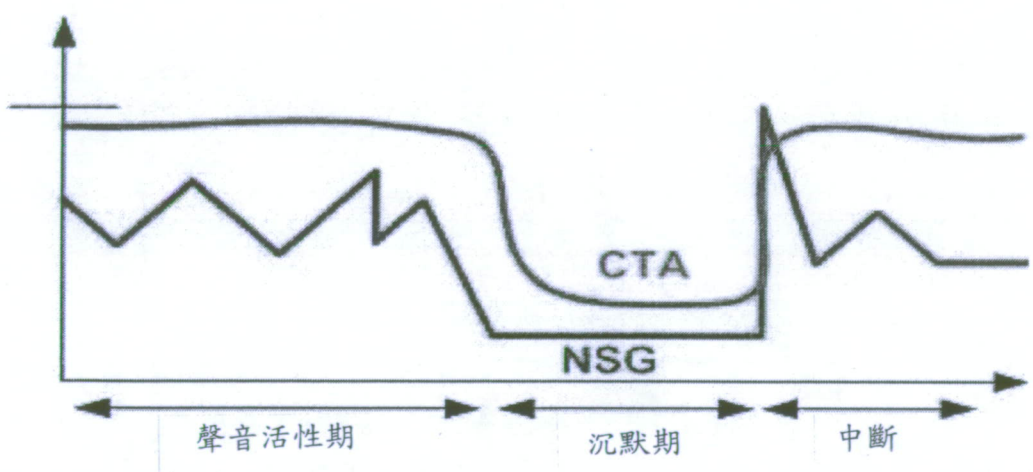


圖 8