

第三章 教學影片主要畫面判斷技術

3.1 研究目標

本研究所提出的研究方法均以下列四項目標為發展主軸：

1. 直接使用壓縮格式下的教學影片進行處理：

現今數位攝影機已非常普及，所錄製的影片均已直接存成壓縮的影片格式，為求節省處理時間，本研究定位在直接處理壓縮格式的教學影片，故以 MPEG I/II 格式教學影片的壓縮資訊作為判斷依據。

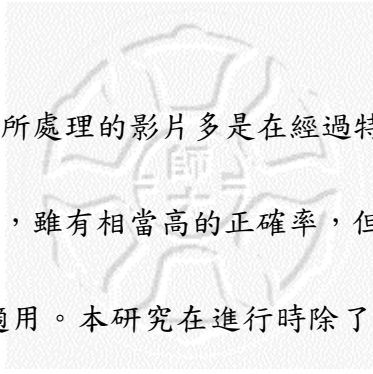
2. 找出研究定義的教學影片特性加入判斷以提高正確率：

由第二章的文獻探討可以得知，一般單純判斷連續畫面間的片段變化偵測方法並不適用於教學影片，而教學影片也因各個研究環境不同而有不同的定義，故必須詳細觀察本研究所定義的教學影片，利用其本身所具的特性加入判斷的技術中，才能提高正確率。

3. 減少處理循環，以達成即時性：

為了實用性考量，本研究的技術希望在可接受範圍內的正確率下，能在一個循環內即完成所有判斷，即在存取完所有連續畫面資訊後即能得到結果。因為影片所耗費的計算量龐大，且連續畫面的存取是依畫面順序進行，若是像一般研究還要再分成數個階段進行，雖然能提高正確率，但相對的必須耗費額外空間存放候選畫面資訊，就即時性的觀點來看並不實際。

4. 以實際教學環境錄製的教學影片作為實驗影片：



一般教學影片的研究，所處理的影片多是在經過特殊安排的教學環境下進行，或是在固定教室內進行錄製，雖有相當高的正確率，但一到不同教學環境，或是真正的教學現場，往往便不適用。本研究在進行時除了為研究所錄製的實驗影片之外，也以研究者本身所在學校實際教學現場所錄製的教學影片進行探討，期能處理現實環境的教學活動。

3.2 教學影片的觀察

如同研究[29]中所提及，教學影片的投影片變換偵測是較屬於具語意(semantic)上的判斷，有一些不必要的移動會造成畫面產生劇烈的變化，然而這些變化卻是在語意上無意義的，因此必須除去；而相反的，有一些重要的改變雖然並未在畫面組成上造成巨量改變，卻是必須定位出來的。就觀察本研究所定義的教學影片（如圖 1.2 所示），存在兩項困難點如下：

1. 教師或學生突然移動至畫面中，造成誤判：

在教學情境中，常會發生教師講解到一個部份，為了指出投影片中的某個重點，而走入投影片的畫面中，造成影片畫面形成劇烈改變；又或者學生因遲到或是座位關係，出現在影片的鏡頭中，造成連續幾張投影片間產生相當的變化，如圖 3.1 所示。以上所示兩種情況，不管是在像素、histogram 或是 motion vector 部份都會造成巨量改變，用一般影片偵測方法都會被視為發生片段變化，但在教學影片中，這樣的改變並不是投影片內容發生改變，因此就意義上來說，不應被視為發生片段變化。

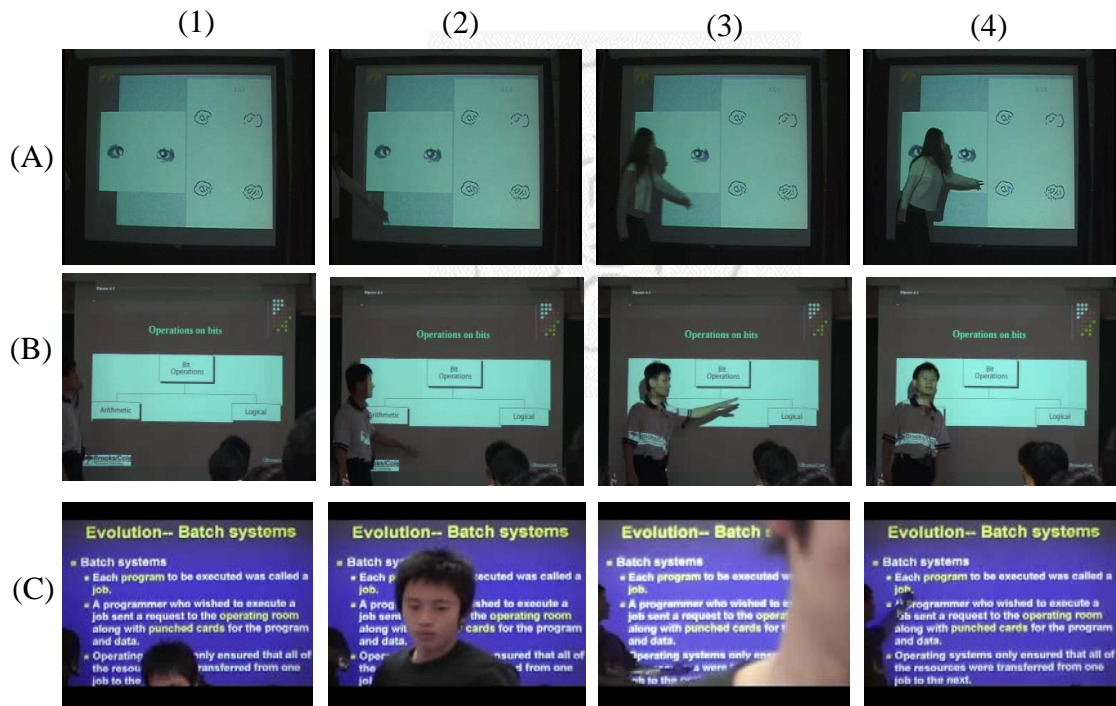


圖 3.1 無關教學內容的畫面改變。(A)(B)兩部份各連續四張圖(由 1 到 4)表示講師為講解某個概念走進影片畫面中；(C)部份連續四張圖(由 1 到 4)表示學生於上課中間起身移動而影響影片畫面。

2. 投影片內容改變量相對較少，造成遺漏：

教學影片的第二個困難點在於大部份的講師在設計教學投影片時，通常都以同一系列主題或樣式來呈現每一張教學投影片，這樣的方式造成每張教學投影片的樣式均相當類似。而有時連續兩張不同內容的投影片，在畫面的分布上極為類似，僅有少部份區域不同，如圖 3.2 所示，雖然屬於不同內容投影片，但因其所造成的投

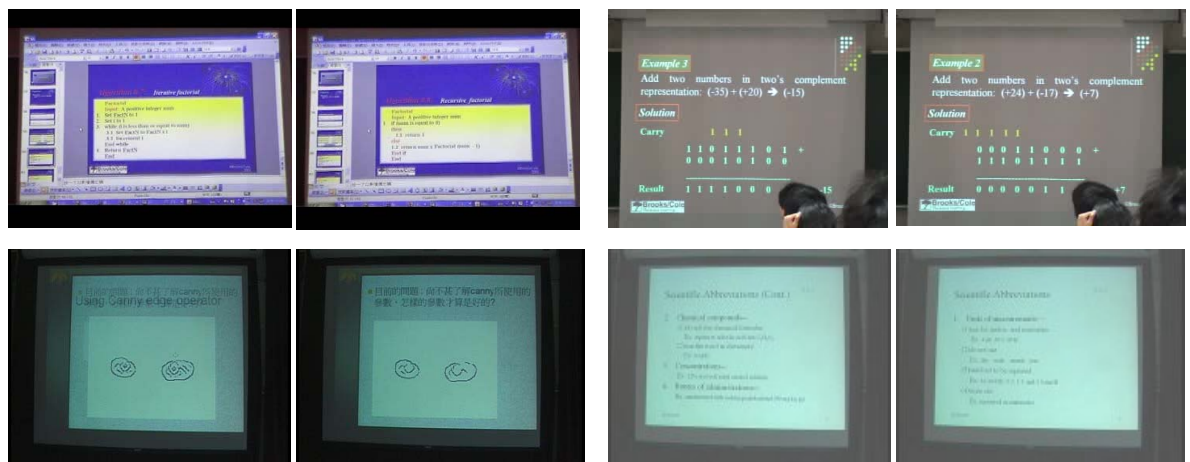
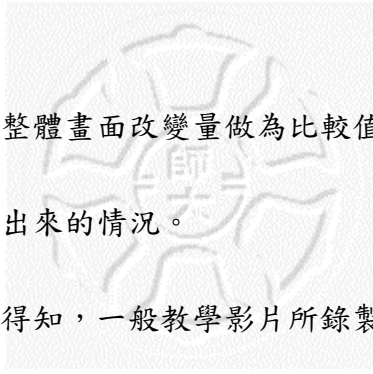


圖 3.2 細微的投影片變化。左上、左下、右上及右下共四對圖片，每對均屬同一個教學影片中連續的兩個畫面，但卻分屬不同的投影片。



影片畫面改變極少，若單以整體畫面改變量做為比較值，很可能因為所設的判斷門檻值過高，而造成無法偵測出來的情況。

詳細檢視教學影片可以得知，一般教學影片所錄製的畫面，畫面中絕大部份均為投影布幕，而因為取景的角度及攝影機鏡頭形狀的關係，會再多取到一部份非投影影片布幕的區域，如圖 3.3 所示，圖中四張不同教學影片的畫面，白色框所框住的部份內部即為投影片的畫面區域，白色框以外的部份即為非投影片布幕的區域。

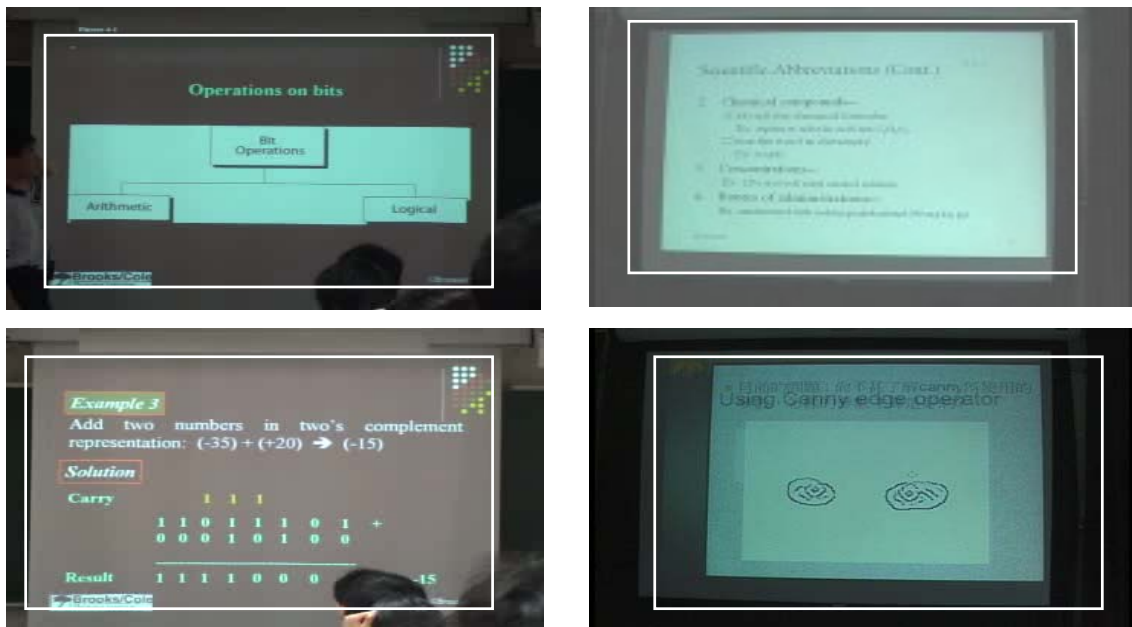


圖 3.3 教學影片取景範例，圖中四張畫面的白色框部份為投影片布幕部份，白色框以外的部份為非投影片布幕部分。

由於教學投影片乃是透過投影機投射在投影布幕上，當投影片內容改變時，投影布幕的內容當然也隨之改變；而大部份時候，非投影片布幕的部份並不會有太多變化，只有在教師或學生出現在畫面中時，這些區域才会有變化，在此我們先將包含投影片布幕的區域，如圖 3.3 的白色框部份區域設定為「內部區域」，而將畫面中白色框以外的部份設為「外部區域」。如圖 3.4 所示，並重新定義如下：

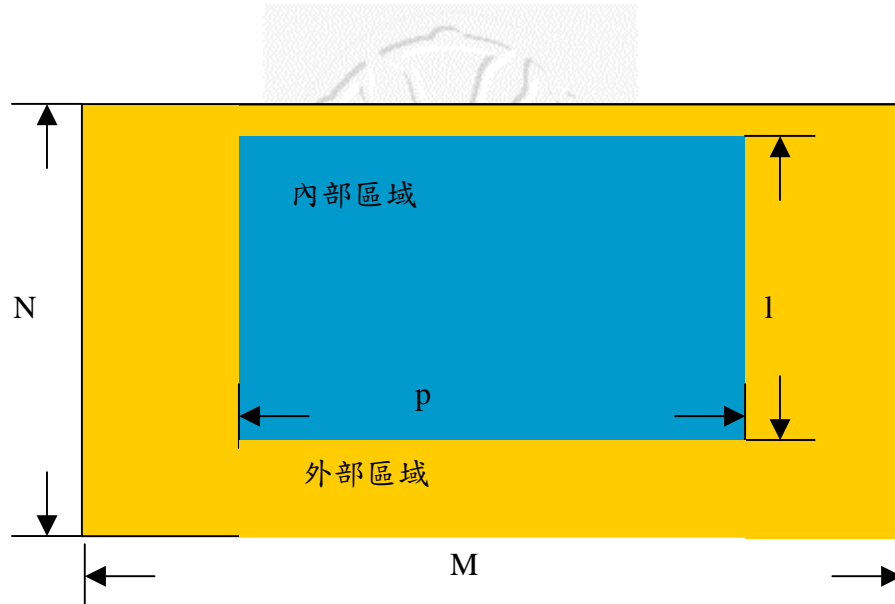


圖 3.4 內部區域及外部區域示意圖。

1. 「內部區域」(inner-region)：

如圖 3.4 所示，假設影片畫面大小為 $M \times N$ 個像素，則令「內部區域」為大小 $l \times p$ 的矩形區域，其中 $l < N$ 且 $p < M$ ，此塊矩形區域包含大部份投影片內容。

2. 「外部區域」(outer region)：

相對於「內部區域」，「外部區域」乃是指在大小為 $M \times N$ 個像素的教學影片畫面中，扣除「內部區域」所剩下的部份，其大小為 $M \times N - l \times p$ ，如圖 3.4 所示。

檢視圖 3.2 及圖 3.3 可以發現，所有的投影片內容發生的改變，幾乎都是發生在「內部區域」。且由圖 3.1 及圖 3.3 可知，講師或學生出現在畫面中時，必定先經過「外部區域」而後再進到「內部區域」。由以上觀察可以歸納出以下兩個現象：

現象 1. 當發生投影片切換時，「內部區域」應具有較大的改變量。

現象 2. 當教師或學生開始出現在投影片內容畫面時，教學影片畫面的「外部區域」應具有較大的改變量。

第一個現象可以用來協助判斷是否發生投影片變化，第二個現象則可以用來避



免因教師或學生出現所產生的誤判，本研究將依循以上兩個現象，來發展投影片切換的偵測技術。

3.3 特徵值的選用

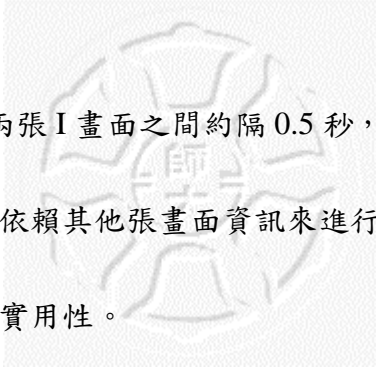
由於本研究希望能直接在壓縮格式下直接進行偵測，因此將以 MPEG 格式中可用的資訊作為特徵值。而本研究所定義的教學影片除了 3.2 節所敘述的兩項觀察之外，另外存在幾項特性可以作為進行特徵值選用的參考。分述如下：

(1) 投影片畫面變化實為亮度改變

因為教學影片中的投影片內容，乃是以 LCD 光學投影機投射至教室中的布幕上，故投影片發生內容改變時，即是投影機的光學投影過程發生變化，因此在本研究中將直接採用亮度的資訊來做為特徵值。而在 MPEG 中的色彩空間是採用 YCrCb，其中 Y 值代表亮度，CrCb 則是代表色差，故本研究直接取 Y 值的資訊做為主要判斷的特徵值。

(2) 移動為慢速變化

在教學影片中，畫面資訊會產生位移的原因主要來自兩方面：投影片內容改變、教師及學生移動。這兩種變化都不是快速的變化，經過大量觀察所拍攝的教學影片，大部份的投影片切換均是在 0.5~2 秒內完成，教師及學生的移動更是多達數十秒。而影片乃是以每秒數十張連續畫面所組成，故要進行判斷投影片內容改變或是教師、學生移動，並不需使用所有畫面資訊。本研究將以 MPEG 格式中的 I 畫面做為判斷的畫面，因為 MPEG 格式影片每秒約有 30 張畫面，其中每隔約 15 張即有



一張 I 畫面，換句話說，每兩張 I 畫面之間約隔 0.5 秒，這提供了一個非常合理的取樣值。同時 I 畫面本身不需依賴其他張畫面資訊來進行解碼，使用 I 畫面將可加速本研究的處理時程，以提高實用性。

(3)攝影機鏡頭固定

教學影片的拍攝現場，經常是工作人員事先架設好固定式攝影機，然後從開始到結束完整錄製，中途除非特殊狀況，否則不會有任何人去移動攝影機或調整鏡頭，故 3.2 節所設定之「內部區域」及「外部區域」為固定的，並不會隨著時間而改變，因此我們可以利用畫面中的位置資訊來進行判斷。在 MPEG 格式中，是以區域壓縮的方式來儲存資訊，也就是畫面在編碼時即先被分割成固定大小的區塊，在 I 畫面中，這些區塊先經過 DCT 轉換及量化（quantization）的程序將空間資訊轉換為頻率資訊儲存起來，因此可以利用這些區塊資訊來直接定義「內部區域」及「外部區域」，以縮短處理時間。

另外，為了加速處理的速度，本研究採用研究[2]中所使用的 DC-image 進行判斷。在 MPEG 影片格式的 I 畫面中，所儲存的資料皆已經過 DCT 轉換，因此並不須重複計算即可取出 DC 值，在 I 畫面中，所有的區塊大小均為 8x8，令每一個原始區塊中的像素值為 $f(i,j)$ ， i,j 表示像素的座標，由式 (1) 可知，經過 DCT 轉換後，令 DC 值為 $c(0,0)$ ，則 DC 值和原始區塊的像素值存在下列關係：

$$c(0,0) = \frac{1}{8} \sum_{x=0}^7 \sum_{y=0}^7 f(x,y) \quad \text{式(2)}$$

由式 (2) 可以看出，DC 值約為原始區塊像素平均值的 8 倍，而由 DC 值所組



成的 DC-image(如圖 3.5 (a) 所示)，大小為原圖的 1/64，則可以視為原圖的縮小版，如圖 3.5 (b) 所示，不但保留了原始影像大部份的特質，也減少因物體移動或鏡頭晃動所造成的差異，同時因為 DC-image 為原影像的 1/64，處理的計算量減少了非常多，對整體速度提升有相當可觀的幫助。

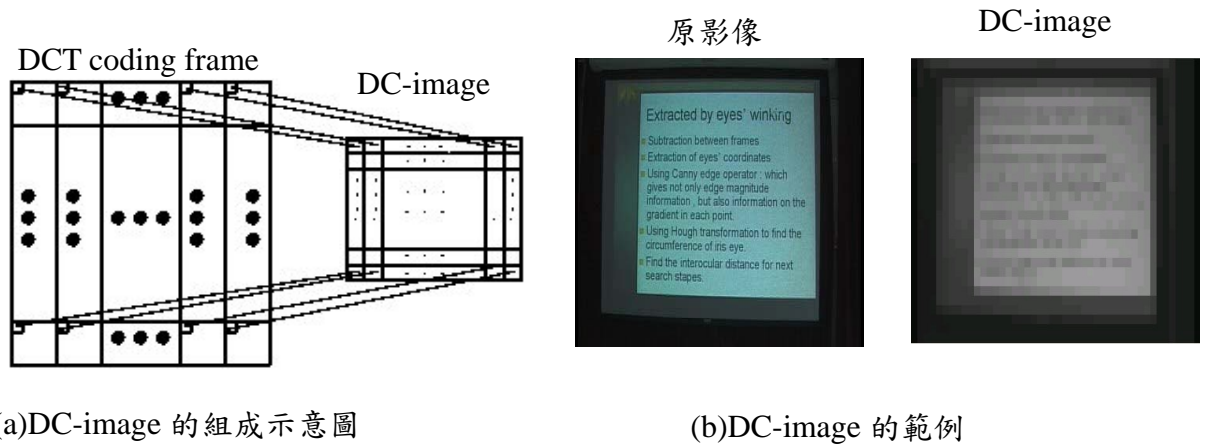


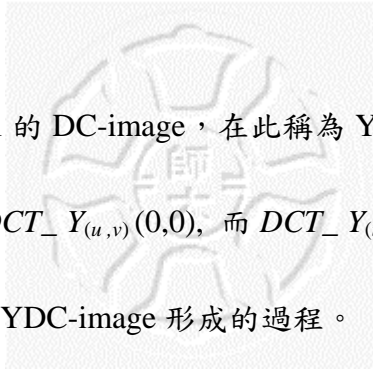
圖 3.5 DC-image 的組成及範例。
(a)部份為 DC-image 組成示意圖，DC-image 為原圖大小的 1/64。(b)部份為 DC-image 的範例，右方的 DC-image 為放大 64 倍後的影像，由圖 3.5(b)可看出，DC-image 保留了原圖大部份的特徵。

綜合以上所述，本研究將以 MPEG 格式中的 I 畫面作為主要取樣的來源，並取出每張 I 畫面的 Y-channel 資訊，再以其中的 DC 值所形成的 DC-image 作為比較依據。

3.4 判斷方法詳述

3.4.1 YDC-image

在此我們針對本研究所使用的 I 畫面中 Y-channel 的 DC-image 進行定義，令 $Y(x,y)$ 代表原始 I 畫面中的 Y-channel 影像，而 $Y_{(u,v)}(i,j)$ 代表 $Y(x,y)$ 中第 (u, v) 個



macroblock，於是 Y-channel 的 DC-image，在此稱為 YDC-image，即可以被定義成 $YDC(u,v)$ 其中 $YDC(u,v) = DCT_Y(u,v)(0,0)$ ，而 $DCT_Y(u,v)$ 即是 $Y(u,v)$ 經過 DCT 轉換之後的結果。圖 3.6 表示了 YDC-image 形成的過程。

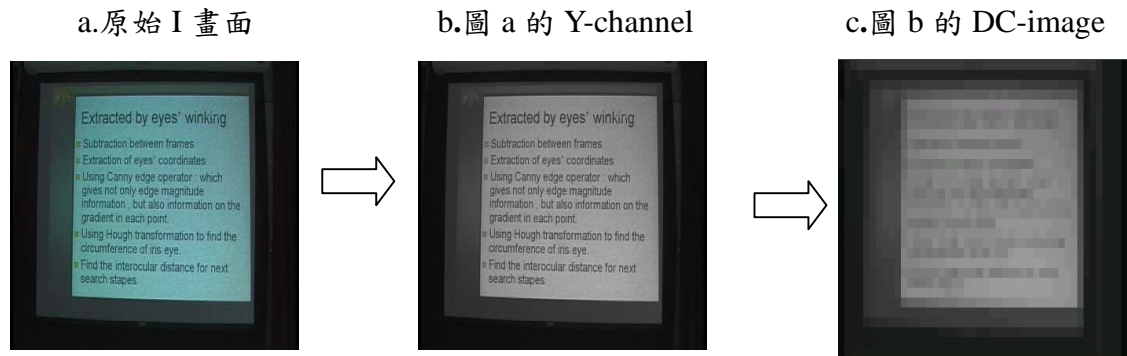


圖 3.6 YDC-image 的形成範例。

圖 a 部份為原始 I 畫面，圖 b 為圖 a 的 Y-channel 部份，可以看出圖 b 因為只紀錄亮度，只具備灰階值，彩色的部份已不存在，而圖 c 則為圖 b 每 8×8 個像素形成區塊，每一個區塊經過 DCT 轉換後取出 DC 值所形成的 DC-image（已放大 64 倍），也就是圖 a 的 YDC-image。

以 YDC-images 來取代 I 畫面之後，YDC-image 的畫面大小為原本的 $1/64$ ，雖然 YDC-image 的畫面品質並不是非常好，但因為 YDC-image 每一點像素都可視為原本 I 畫面的中每一個 8×8 大小的 macroblock 的亮度平均值，因此雖然 YDC-image 無法提供細微的畫面內容分析，但卻已提供了足夠的資訊來進行投影片畫面的變化偵測。

3.4.2 判斷運算式說明

令 $YDC^i(u,v)$ 代表原始影片中第 i 張 I 畫面所取出的 YDC-image，而由 3.2 節中兩個觀察所得的現象，我們將每個 YDC-image 也區分成內部區域(inner-region)及外部區域(outer-region)，則連續兩張 YDC-image($YDC^i(u,v)$ 及 $YDC^{i+1}(u,v)$) 的內部區域

及外部區域平均變化量表示如下兩個式子：

$$\Delta YDC_{inner}^i = \frac{1}{\#of(u,v) \in inner \text{ reg.}} \sum_{(u,v) \in inner \text{ reg.}} |YDC^i(u,v) - YDC^{i+1}(u,v)| \quad \text{式(3)}$$

$$\Delta YDC_{outer}^i = \frac{1}{\#of(u,v) \in outer \text{ reg.}} \sum_{(u,v) \in outer \text{ reg.}} |YDC^i(u,v) - YDC^{i+1}(u,v)| \quad \text{式(4)}$$

其中式 (3) 代表連續兩張 YDC-image 內部區域平均像素差異值，式 (4) 則為外部區域平均像素差異值。

接下來則根據 3.2 節中觀察所得到的兩個現象，當發生投影片內容變化時，必須同時滿足以下兩個運算式：

$$1. \Delta YDC_{inner}^i > \alpha \Delta YDC_{outer}^i \quad \text{式(5)}$$

其中 α 為一個可調整的常數。式 (5) 即是根據 3.2 節中兩個觀察所得的現象：當發生投影片切換時，內部區域的變化必定大於外部區域的變化，以及當教師走入投影片畫面時，外部區域的變化會大於內部區域的變化。因此當發生投影片變化時，連續兩張 YDC-image 的內部區域變化量，應該要大於連續兩張 YDC-image 外部區域的變化量，因為外部區域經常是比較單調且固定的部份，因此若內部區域發生一點細微的變化，很可能就會大於外部區域所發生的變化，這樣的情況將會造成過多不必要的誤判，因此以 α 值作為調整判斷的敏感度， α 值愈大，代表設定投影片變化的門檻值愈高。

式 (5) 的另一層意義代表若發生 $\Delta YDC_{inner}^i < \Delta YDC_{outer}^i$ 時，表示外部區域的變化大於內部區域變化，這很可能是發生在教師的移動造成了畫面的改變，若不將畫面區分成內部、外部區域，則教師移動所造成的畫面整體改變量則可能大於投影片

內容改變造成的變化。因此透過式(5)可有效降低因教師移動所造成的誤判。

$$2. \Delta YDC_{inner}^i > \beta(\Delta YDC_{inner}^{(i-k)\sim(i-1)} + \Delta YDC_{inner}^{(i+1)\sim(i+k)}) \quad \text{式(6)}$$

其中 β 為一個可調整的常數， $\Delta YDC_{inner}^{p\sim q} = \sum_{j=p}^q \Delta YDC_{inner}^j$ ，表示從第 p 張到第 q 張

間連續 $(q-p+1)$ 張 YDC-image 的變化平均值。 $\Delta YDC_{inner}^{(i-k)\sim(i-1)}$ 表示前面連續 k 張

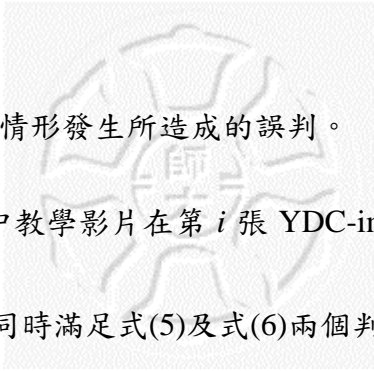
YDC-image 的平均變化值， $\Delta YDC_{inner}^{(i+1)\sim(i+k)}$ 表示後面連續 k 張 YDC-image 的平均變化

值，式 (6) 乃是限定發生投影片切換時，內部區域改變量必須是相對於前後相鄰

畫面群的突然且劇烈的變化，其意義有兩點：

(1) 教師在講解每一張投影片時應該都持續一段時間，因此影片在投影片切換前及切換後，應該都是持續一段時間畫面沒有明顯改變，在發生投影片切換時，內部區域的改變量相對於前後沒有投影片切換的改變量應該是較為劇烈的，因此式(6)的作用在確保投影片切換必須是在時間軸上，連續 YDC-image 間的變化量為突然改變，且變化量較前後為高的狀況。

(2) 若是在一段短時間內，連續畫面間均發生變化，則可能是教師並未按照原本投影片編排順序授課，為了到達某一張投影片，因此以快速切換方式瀏覽投影片，如此而造成了連續畫面間發生了一連串的劇烈變化；另一種可能是影片中的教師進入投影片的畫面區域中，因為當發生教師走入投影片畫面時，由於教師移動為慢速的移動，而連續的兩張 YDC-image 的間隔約為 0.5 秒，因此教師的移動從開始到結束，所造成的畫面的改變絕對不會只發生在兩張連續的 YDC-image 之間，而是發生在連續數張的 YDC-images 中，而這兩種情況都很可能滿足式(5)，因此再加上式(6)



的條件，即能避免上述兩種情形發生所造成的誤判。

綜上所述，在本研究中教學影片在第 i 張 YDC-image 發生投影片的切換時， ΔYDC_{inner}^i 及 ΔYDC_{outer}^i 必須同時滿足式(5)及式(6)兩個判斷式。

3.5 投影片切換偵測範例

3.4 節敘述了本研究最重要的兩個判斷式，用來偵測教學影片中投影片切換的時間點，在本節中將以一段 MPEG 格式教學影片範例來說明式(5)及式(6)運用的過程。令範例影片名為「影片一」，影片一總長度為 1 分 53 秒，共有 3411 張畫面，其中有 214 張 I 畫面，也就是有 214 張 YDC-image，在這段教學影片中，共有 11 張不同內容的投影片，如圖 3.7 所示。每張投影片第一次出現在影片中的 YDC-image



圖 3.7 範例影片一的 11 張不同投影片內容畫面。圖下方兩組數字，括號外的數字為該投影片首次出現所代表的 YDC-image 的畫面編號，括號內為其 MPEG 影片中的畫面編號

畫面編號及實際 MPEG 中的畫面編號分別標注在每張圖的下方。

偵測的過程中，先將影片一中每張 YDC-image 取出，每張 YDC-image 畫面先依長寬各取成四等分，分成 16 個區塊，令中間六個區塊組成的區域為內部區域，外圍十個區塊組成的區域為外部區域，如圖 3.8 所示。

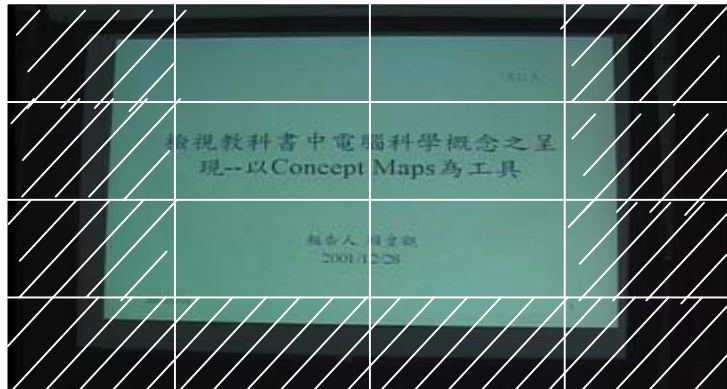


圖 3.8 斜線部份為外部區域，其餘中間部份為內部區域

在此直接以 YDC-image 的像素值作為特徵值， ΔYDC_{inner}^i 即為圖 3.8 中的內部區域平均像素差異值， ΔYDC_{outer}^i 即為圖 3.8 中的外部區域平均像素差異值。 k 值設為 5， α 值設為 2，圖 3.9 顯示了這段影片的 ΔYDC_{inner}^i 及 ΔYDC_{outer}^i 的數據分布圖。

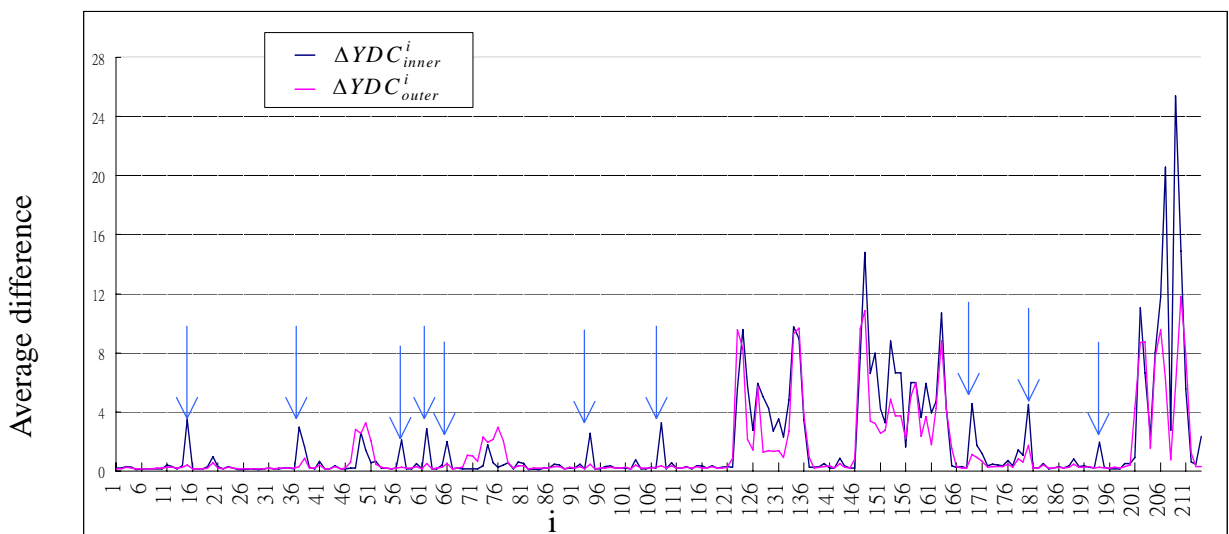


圖 3.9 影片一 ΔYDC_{inner}^i 及 ΔYDC_{outer}^i 數據分布圖，橫座標即 YDC-image 畫面編號
黑色的折線代表內部區域發生的變化，粉紅色折線代表外部區域的變化。

由 3.7 圖中可以看出，發生投影片切換的時間點分別在 YDC-image 的畫面編號為 15、37、57、62、66、94、108、169、180 及 194 的地方，如圖 3.9 中箭頭所指的數據圖上的尖端點，而這些點確實同時滿足本研究的兩個判斷式式(5)及式(6)。為了再進一步分析，我們將圖 3.9 分割成前半及後半兩張圖，分別如圖 3.10 及 3.11 所示。

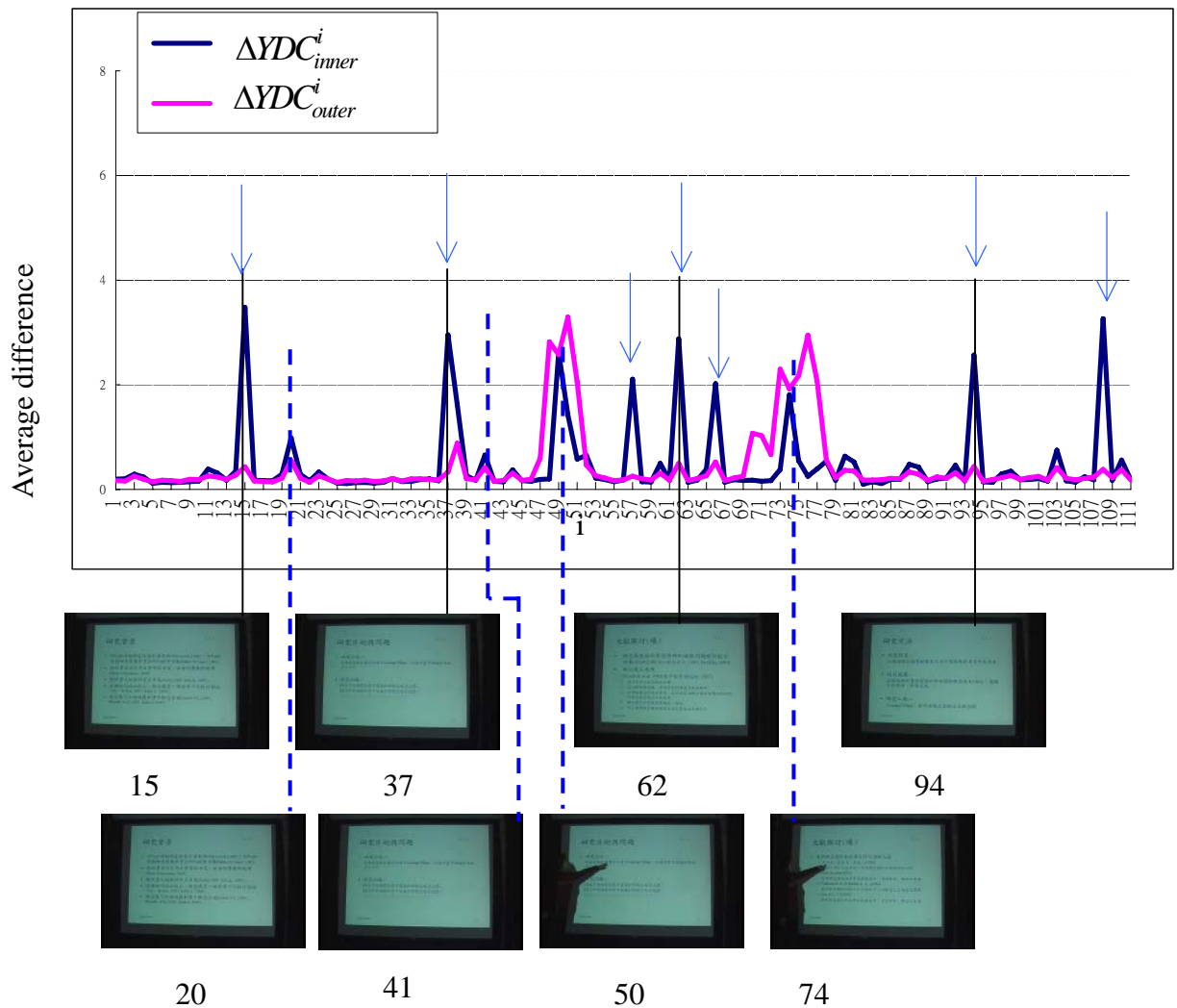


圖 3.10 前半部 ΔYDC_{inner}^i 及 ΔYDC_{outer}^i 數據分布圖 (1~111) 及部份相對應的圖片

圖 3.10 及圖 3.11 中有藍色箭頭所指的地方為發生投影片切換的地方，圖 3.10 上下方兩排對應的圖片中，上排圖片為部份發生投影片切換的說明畫面 ($i=15、37、$

62、94)，這些點的數據呈現可以發現，當發生投影片切換時，內部區域的 ΔYDC_{inner}^i 顯然是一個較為巨量的峰值；而圖 3.10 下排 3 張有藍色虛線的圖片則是因其他原因所造成的畫面改變。在 $i=20、41$ 時，雖然 $\Delta YDC_{inner}^i > \Delta YDC_{outer}^i$ 但並未到達所設定的門檻值 (α)，故並不會被偵測為投影片切換。另外在 $i=47\sim 55$ 以及 $i=71\sim 78$ 時， ΔYDC_{inner}^i 及 ΔYDC_{outer}^i 均有很大的變化，但這兩段時間內正好是講師進出投影片畫面區域，從圖 3.10 中看出，兩段時間所呈現出來的數據為 $\Delta YDC_{inner}^i < \Delta YDC_{outer}^i$ ，表示其外部區域的變化大於內部區域的變化。以上圖 3.10 中的一些非投影片切換所造成的改變以式 (5) 就可避免誤判。

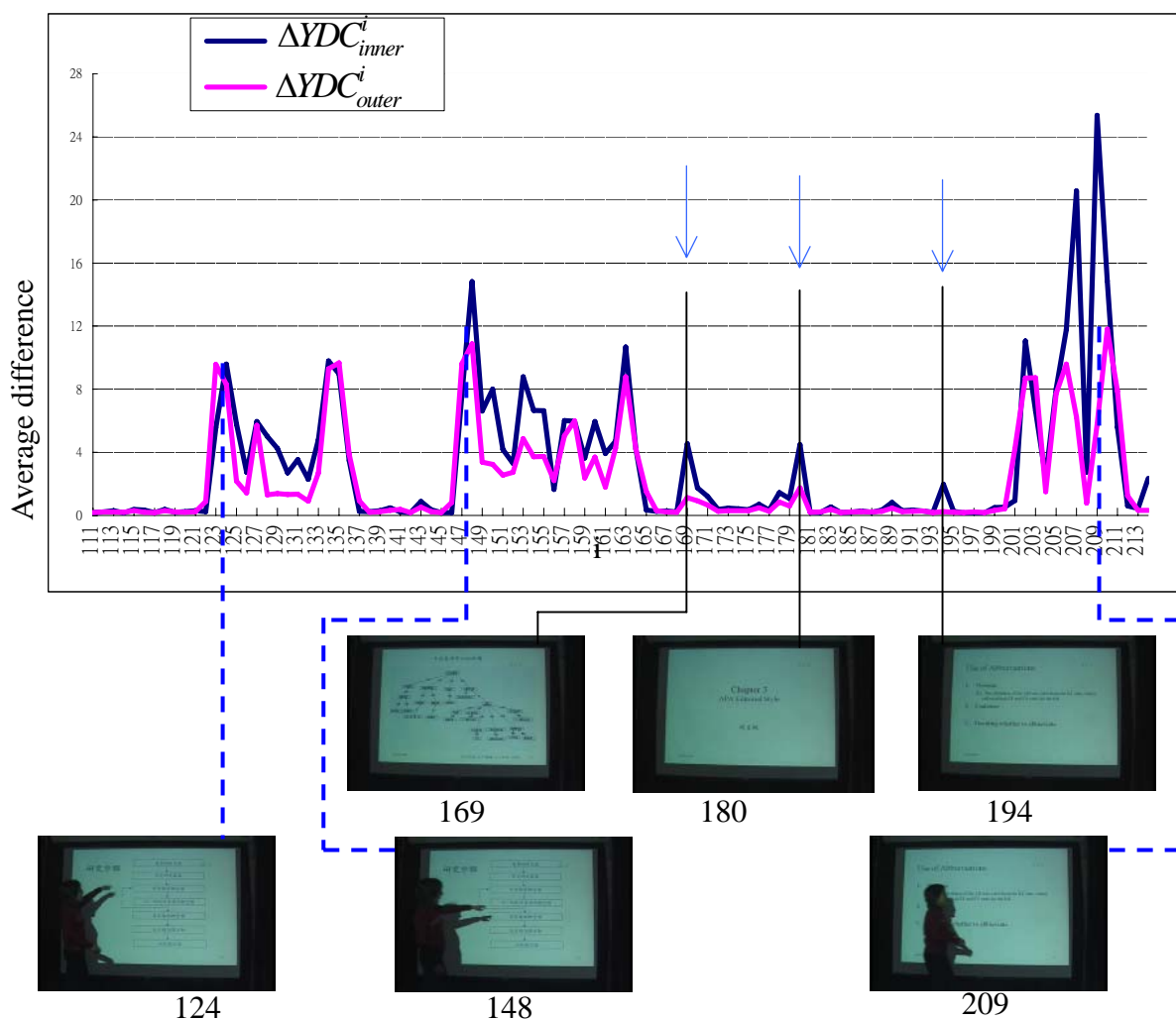


圖 3.11 後半部 ΔYDC_{inner}^i 及 ΔYDC_{outer}^i 數據分布圖 (111~213) 及部份相對應圖片

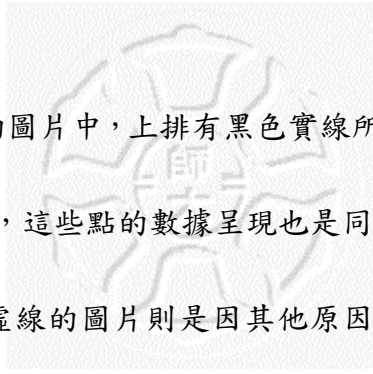


圖 3.11 下方兩排對應的圖片中，上排有黑色實線所指的圖片為發生投影片切換的畫面 ($i=169、180、194$)，這些點的數據呈現也是同時滿足式 (5) 及式 (6)，而圖 3.11 下排 3 張有藍色虛線的圖片則是因其他原因所造成的畫面改變。另外在 $i=123\sim 137$ 、 $i=148\sim 157$ 以及 $i=203\sim 213$ 時，這三段時間也是講師進出投影片畫面區域進行講解的時間，所呈現出來的數據可看出 ΔYDC_{inner}^i 及 ΔYDC_{outer}^i 均有劇烈的變化，且並不同時滿足式 (5) 及式 (6)，因此不會被偵測為投影片變化。其中在 $i=209$ 時，雖然滿足了式 (5)，但因其前後 ΔYDC_{inner}^i 均有連續劇烈變化，故不滿足式 (6) 的條件，不會被視為發生投影片切換。

3.6 教學影片主要畫面判斷技術總結

由 3.5 節的範例影片所呈現的數據可以看出，本研究所推導出的兩個判斷式能有效地來偵測何時發生投影片切換，而當找出投影片切換的時間點所在的第 i 張 YDC-image 畫面時，因為第 $i+1$ 張 YDC-image 畫面為切換後的不同內容投影片畫面第 1 張，故第 $i+1$ 張 YDC-image 畫面即為研究所需要的主要畫面。圖 3.12 為本研究的完整主要畫面偵測流程。

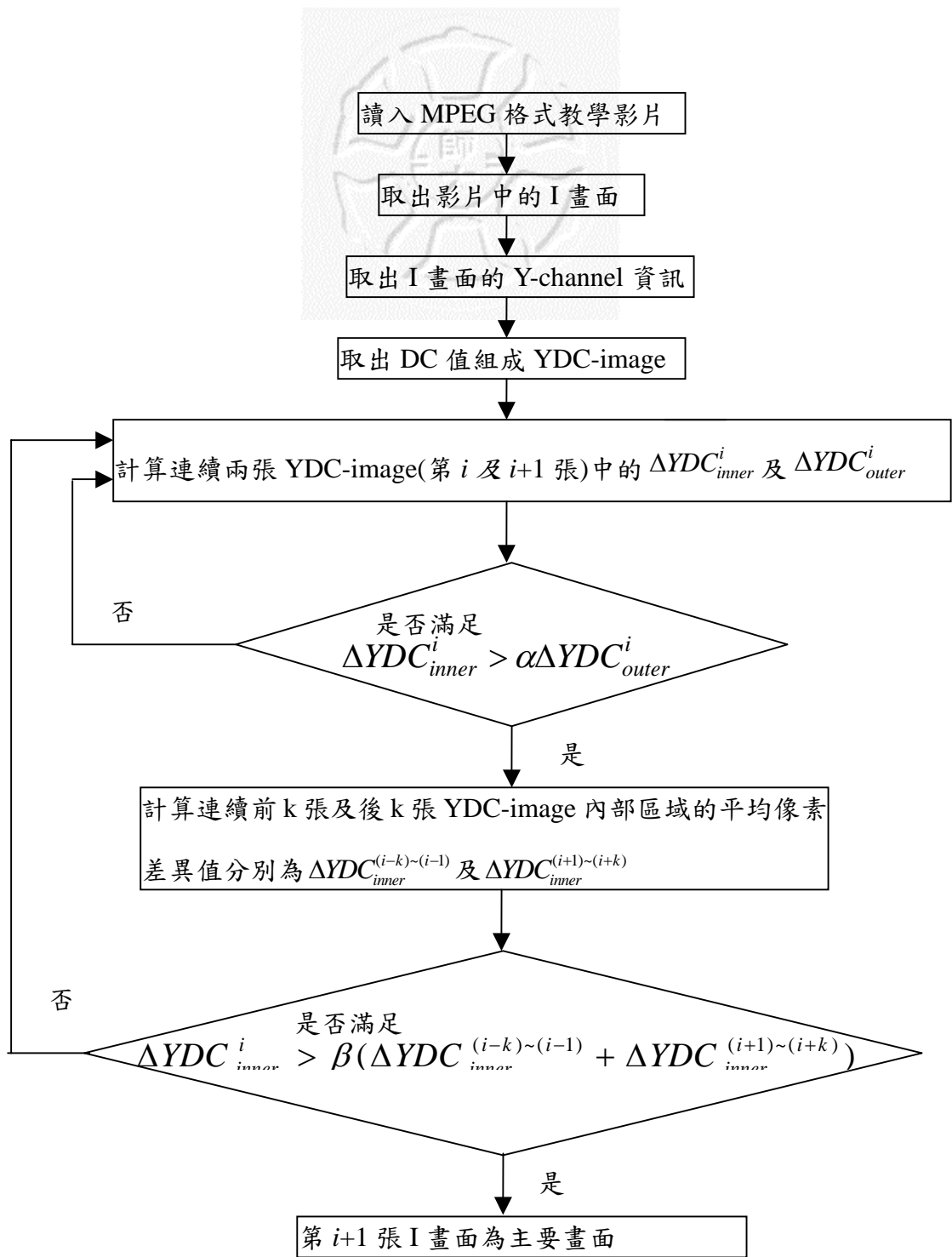


圖 3.12 完整主要畫面偵測流程圖