

國立交通大學

資訊學院 資訊學程

碩士論文

可調性視訊之人臉偵測

Face Detection in Scalable Video



研究生：林美亨

指導教授：傅心家 教授

中華民國九十七年十一月

可調性視訊之人臉偵測  
Face Detection in Scalable Video

研究生：林美亨

Student : Mei-Heng Lin

指導教授：傅心家

Advisor : Dr. Hsin-Chia Fu

國立交通大學  
資訊學院 資訊學程  
碩士論文



A Thesis  
Submitted to College of Computer Science  
National Chiao Tung University  
in partial Fulfillment of the Requirements  
for the Degree of  
Master of Science  
in  
Computer Science  
Nov 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年十一月

# 可調性視訊之人臉偵測

學生：林美亨

指導教授：傅心家

國立交通大學

資訊學院

資訊學程碩士班

## 摘要

本論文研製之 3D 小波可調性編碼(3D Wavelet Scalable Video Coding)擁有頻寬可調性、時間軸可調性與空間軸可調性等優點，可分為三個主要的模組:時間小波 T\_Module、空間小波 S\_Module 與 Entropy\_Coding 模組以產生最後壓縮資料串流。本論文我們希望結合小波可調性編碼與人臉偵測系統，一次性嵌入人臉偵測於不同等級資料串中，針對不同的網路需求，讓接收端用最小資料量分析最快偵測出人臉影像。對可調性視訊接收端而言：資訊接收端不再受限於單一撥放環境，接收端可根據自身的網路環境，選擇觀看人臉偵測影像。對可調性視訊服務端而言：網路伺服器只需提供一個檔案供所有使用端下載，無論接收端的網路環境為何，對伺服器都只要求同一個檔案，不但可以提供不同使用端的需求，更可節省檔案伺服器對硬體需求上的負擔。

由實驗結果驗證，在不同時間壓縮、空間壓縮下之不同等級資料串流偵測人臉影像，均可達到可調性視訊人臉偵測目標，並達到最小資料量分析最快偵測結果。

# Face Detection in Scalable Video

Student : Mei-Heng Lin

Advisors : Dr. Hsin-Chia Fu

Degree Program of Computer Science  
National Chiao Tung University

## ABSTRACT

3D wavelet scalable video coding has the advantages of bandwidth, temporal, and spatial scalability, which includes three main modules : T\_Module 、 S\_Module 、 Entropy\_Coding, those produce final compressed output bit stream.

In this thesis we would like to combine face detection in 3D wavelet scalable video. The advantages are: 1. User only download one bit stream and extract it for what they need. 2. Server only provides one bit stream for all users to download and it also saves lots of disk space in network.

We propose a strategy that makes the face can be detected under different environment and detect the face correctly and quickly.

## 誌 謝

交大畢業多年，能在忙碌的工作之餘，繼續完成研究所學業是自己多年的夢想。一路走來，覺得自己非常幸運，能夠有一位好老師傅教授與徐永煜學長願意不斷地指導我們，還有一群可以互相討論課業的好伙伴們。從小填鴨式的教育讓自己在創造力與獨立思考問題能力較為微弱，這四年研究生活正是訓練自己這些能力的好時機。我想這四年來努力的過程遠大於最後學位的取得，最後真心感謝有你們一路相伴。



# 目 錄

中文摘要	.....	i
英文摘要	.....	ii
誌謝	.....	iii
目錄	.....	iv
圖目錄	.....	vi
表目錄	.....	vii
Chapter1	簡介.....	1
1.1	研究動機與目標.....	1
1.2	相關研究與主要貢獻.....	1
Chapter2	先前研究.....	2
2.1	SVC 可調性編碼研究.....	2
2.1.1	時間小波 T_Module 研究.....	4
2.1.2	空間小波 S_Module 研究.....	8
2.1.3	壓縮編碼研究.....	11
2.1.4	圖例說明.....	11
2.2	人臉偵測研究.....	12
2.2.1	顏色判斷.....	12
2.2.2	外型特徵.....	13
Chapter3	可調性視訊之人臉偵測.....	14
3.1	系統與架構.....	14
3.2	人臉偵測嵌入位置設計.....	15
3.2.1	空間小波解碼嵌入分析.....	15
3.2.2	時間小波解碼嵌入分析.....	17
3.2.3	人臉偵測嵌入位置.....	19
3.3	可調性視訊人臉偵測設計.....	20
3.3.1	膚色 YUV 值取得.....	21
3.3.2	膚色區塊篩選.....	24
3.3.3	外型特徵判斷.....	26
3.3.4	可調性人臉偵測結果.....	27

## 目 錄

Chapter4	實驗結果與討論.....	28
4.1	實驗影片.....	28
4.2	實驗程式與環境.....	29
4.3	可調變實驗 I.....	30
4.3.1	時間調變實驗 I.....	30
4.3.2	空間調變實驗 I.....	33
4.4	可調變實驗 II.....	36
4.4.1	時間調變實驗 II.....	37
4.4.2	空間調變實驗 II.....	38
4.4.3	實驗修改差別分析.....	39
Chapter5	結論與未來發展.....	40
參考文獻	.....	41



## 圖目錄

2.1	可調性視訊編碼	3
2.2	Lifting-based 小波轉換	5
2.3	5-3 時間濾波器	6
2.4	多層次視訊編碼	7
2.5	Frame after S_Module	10
2.6	兩次小波轉換的 Lena	10
2.7	可調性視訊編碼圖例說明	11
3.1	可調性視訊人臉偵測架構	14
3.2	兩次 S_Module 轉換後示意圖	15
3.3	Spatial scalable 1	16
3.4	Spatial scalable 2	16
3.5	空間小波人臉偵測嵌入設計	16
3.6	兩次 T_Module 轉換後示意圖	17
3.7	Temporal scalable 1	17
3.8	Temporal scalable 2	18
3.9	時間小波人臉偵測嵌入設計	18
3.10	人臉偵測位置嵌入	19
3.11	人臉偵測流程	20
3.12	Get LL2 ( Y ) Before Transform QCIF	21
3.13	Get LL2 ( U ) Before Transform QCIF	22
3.14	Get LL2 ( Y ) After Transform QCIF	23
3.15	Get LL2 ( U ) After Transform QCIF	23
3.16	可調適編碼中之人臉偵測	24
3.17	膚色點偵測	24
3.18	光罩去除小雜點後	25
3.19	眼睛判斷方式	26
3.20	可調性人臉偵測結果 1	27
3.21	可調性人臉偵測結果 2	27
4.1(a)	時間調變實驗 1	30
4.1(b)	實驗結果 1	31
4.1(c)	時間調變實驗 2	31
4.1(d)	實驗結果 2	32
4.1(e)	人臉偵測失敗圖	32



## 圖目錄

4.2(a)	空間調變實驗 3	33
4.2(b)	實驗結果 3	34
4.2(c)	空間調變實驗 4	34
4.2(d)	實驗結果 4	35
4.2(e)	人臉偵測失敗圖	35
4.3	十字架干擾圖	36
4.4	架構更改	36
4.5(a)	時間調變實驗 1	37
4.2(b)	時間調變實驗 2	37
4.6(a)	空間調變實驗 1	38
4.6(b)	空間調變實驗 2	38
4.7	實驗修改差別分析	39



## 表 目 錄

4.1	測試影片.....	28
-----	-----------	----



# chapter1 簡介

## 1.1 研究動機與目標

寬頻網路的普及，影音資訊可廣泛透過網路上傳送著，目前線上影音娛樂已成為家庭娛樂重要的一環。考慮網路環境迥異，資訊接收端會有不同的影音視訊需求，如網路頻寬 (bits/sec)、播放解析度 (resolution) 或播放速度 (frame/sec)，為了在不同的網路環境上達到接收端所希望的視訊品質，在此應用下因而產生了可調性的視訊編碼 (Scalable Video Coding)。

近年來人們對於影像處理，有關人臉偵測、追蹤等有相當多的研究及文獻，其中又以保安系統上應用最具代表性。針對兩種趨勢，我們希望結合可調性視訊優點與人臉偵測系統，設計一套可調性視訊人臉偵測系統：

對可調性視訊接收端而言：資訊接收端不再受限於單一撥放環境，可依其網路環境，選擇觀看人臉偵測影像。對可調性視訊服務端而言：網路伺服器只需提供一個檔案供所有使用端下載，無論接收端的網路環境為何，對伺服器都只要求同一個檔案，不但可以提供不同使用端的需求，更可節省檔案伺服器對硬體需求上的負擔。

## 1.2 相關研究與主要貢獻

目前對於影像偵測的研究大多未提及網路環境不同調變 (Scalable) 時的影音播放品質。本論文以微軟提出的 MSSVC 可調性視訊編碼為架構，加入人臉影像偵測功能，一次性將人臉偵測程式嵌入可調性視訊編碼中。接收端可根據自身的網路環境，選擇觀看人臉偵測影像，是本論文的主要貢獻。


我們將分為五個章節說明，第一章為研究動機與相關研究、第二章為 3D Wavelet Scalable Video Coding 可調性編碼研究、第三章為可調性視訊之人臉偵測，第四章實驗結果與討論，第五章為結論與未來展望。

## chapter2 先前研究

本章中我們先對視訊編碼 (Video coding) 作分類，在對可調性視訊編碼各個模組作較詳細的研究討論。

### 2.1 SVC 可調性編碼研究

可調性視訊編碼 (Scalable Video Coding, 簡稱 SVC) 是以小波轉換 (Wavelet Transform) 達到可調性編碼的演算法。基於小波轉換的視訊編碼，可大略的分為兩類：一類是二維轉換後移動補償 (2D+Motion Compensation) 的視訊編碼，另外一類是三維 (3D) 視訊編碼。



二維轉換後移動補償的視訊編碼，主要先針對連續視訊中的每一個畫面 (Frame) 作空間小波轉換，以期讓每一個畫面轉換到頻率域，在對在頻率域的畫面作時間前後上的移動補償，進行編碼。此種編碼方式有個較不易解決的問題，即是在空間域上移動一個點 (Pixel)，在小波轉換過到頻率域上卻差異甚大，如此造成移動補償不易取得其先後畫面的相關性，造成編碼的效率較差。

三維視訊編碼，主要先針對時間上的前後畫面做小波轉換，轉換後的資料再對空間做小波轉換。在作時間小波轉換的同時，並將移動向量 (Motion Vector) 求出，分別編碼，如此即可較有效率的對時間上點的移動描述，並有效率的編碼。

圖 2.1 為三維視訊編碼的例子，可分為時間軸小波分解與空間軸小波分解；此方法是求出前後畫面的移動向量，進行移動向量模式選擇及編碼。另一方面參考移動向量 (MV & Mode Coding)，先將每一個畫面作時間小波 (T\_Module) 分解，先產生出高低頻相間的畫面，在經由二維空間小波 (S\_Module) 分解，以期每一個畫面能再分

解出高低頻部份。時間與空間小波壓縮處理後，最後我們在對相似碼進行壓縮編碼（Entropy Coding），產生出最後的資料串流。

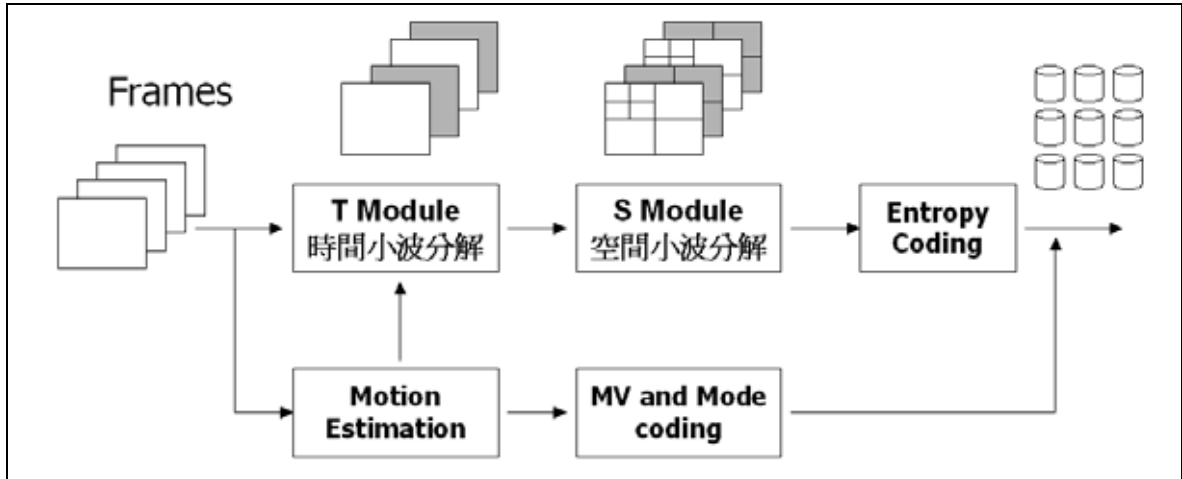


圖 2.1 可調性視訊編碼



## 2.1.1 時間小波 T\_Module 研究

影像是由許多張連續的 frames 在短時間內播放出來，在影像壓縮編碼中，爲了達到良好的壓縮效果，我們可以壓縮相鄰 frames 的資料。考慮 frames 彼此間會有許多相似的地方，爲了節省資料量傳遞，我們在傳 frame 時，並不需要傳送整張 frame，而只要傳送跟 previous frame 差異的地方。在三維視訊編碼是先由 T\_Module 作時間軸 decomposition 處理，其會參照 motion-vector 來表示相鄰間相似 frames 間的關係，兩個 frames 間可用 frame 與移動向量(motion-vector information)來表示，其可以節省大量的資料儲存。

時間小波採用 Lifting based 方式，這種方式移動向量不再是由第一個畫面到最後一個畫面連接產生，它是固定將數個畫面爲一組，前後產生進階移動緒，即是說移動向量不止只單方向的與下一個畫面有關，移動向量也與上一個畫面有關。

以下爲 Lifting based 方式說明，假設有一組輸入訊號  $\{x_0, x_1 \dots x_6\}$ ，使用高通濾波器與低通濾波器，以求出兩組輸出的小波係數 (Wavelet coefficients) 訊號，一組爲低通係數  $\{L_0, \dots, L_3\}$  與另一組高通係數  $\{H_0, \dots, H_2\}$ 。

$$\begin{cases} H_i = x_{2i+1} + a(x_{2i} + x_{2i+2}) \\ L_i = x_{2i} + b(H_{i-1} + H_i) \end{cases} \quad (1)$$

with  $a = -1/2, b = 1/4$

圖 2.2 為此 Lifting based 小波轉換的示意圖，左圖 (a) 為正轉換，右圖 (b) 為逆轉換。

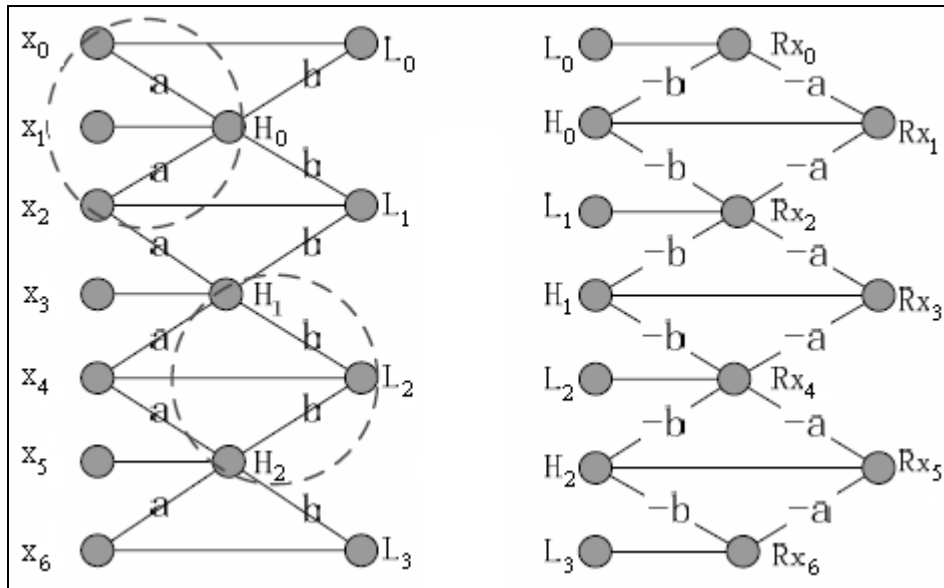


圖 2.2 Lifting-based 小波轉換 (a) 正轉換 (b) 逆轉換

將此 Lifting-based 小波轉換應用到視訊上，進階移動緒的高低頻濾波器可改寫成

$$\begin{cases} P_{H_i} = P_{F_{2i+1}} + a \times (MT(F_{2i}) + MT(F_{2i+2})) \\ P_{L_i} = P_{F_{2i}} + b \times (MT(H_{i-1}) + MT(H_i)) \end{cases} \quad (2)$$

$P_{F_{2i}}$  &  $P_{F_{2i+1}}$  : current pixel to be updated.

$MT(.)$  : the mapping pixels in the adjacent frames.

$P_{H_i}$  &  $P_{L_i}$  : wavelet coefficients of the current pixel

即是將目前畫面的每一點求出上一個畫面與下一個畫面的移動向量，找出前後畫面最接近的點，再求出經過此高低頻濾波器的結果。

圖 2.3 是 5-3 時間濾波器的示意圖，以Frame<sub>1</sub>中x<sub>4</sub>為例，經由移動向量找出前後畫面最接近的點，前為Frame<sub>0</sub>點x<sub>3</sub>，後為Frame<sub>2</sub>點x<sub>4</sub>，經由高頻濾波器可產生出高頻訊號H<sub>0</sub>點x<sub>4</sub>。以此例而言，我們可看出進階移動緒與 5-3 時間濾波器解決了移動緒中多對一點的限制，如Frame<sub>2</sub>點x<sub>4</sub>，其它如移動緒的結束點限制，如Frame<sub>1</sub>點x<sub>4</sub>，移動緒的無參考點限制，如Frame<sub>4</sub>點x<sub>5</sub>，均可由進階移動緒配合 5-3 時間濾波器的方法解決。

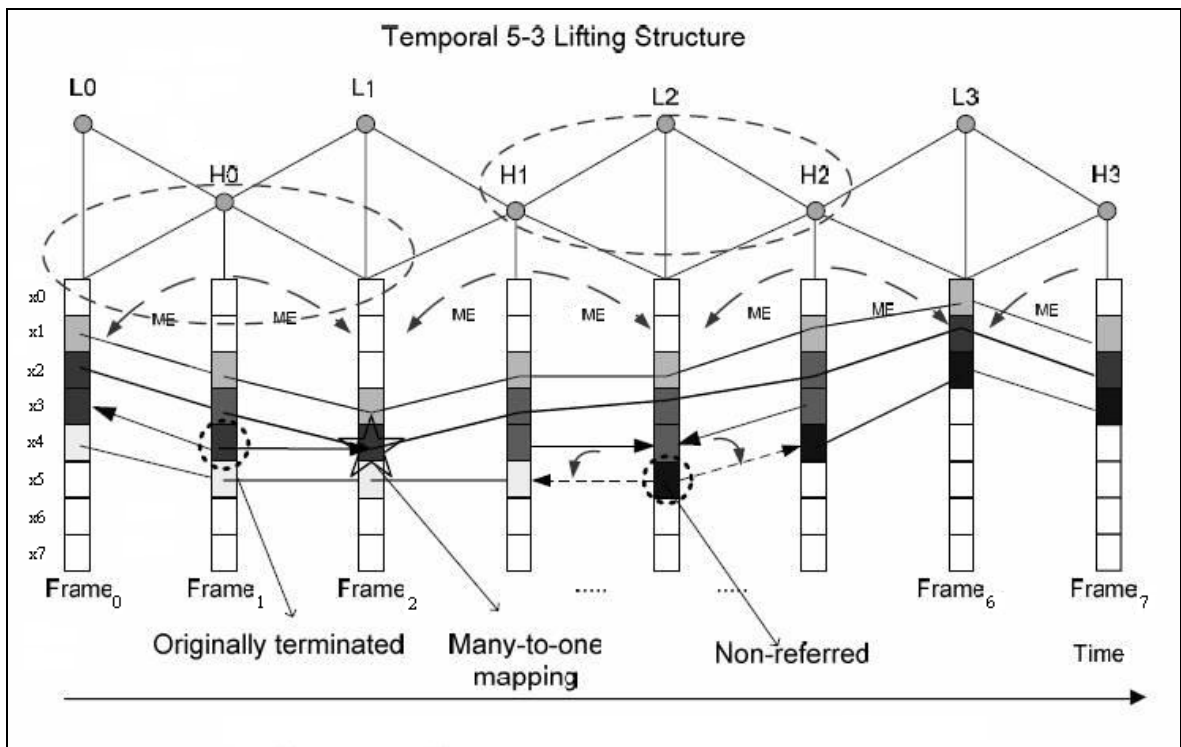


圖 2.3 5-3 時間濾波器

爲了滿足視訊在時間上的壓縮，達到可調變（Scalability）的目標，通常會將原先只有一層的編碼方式，再分解成多個層次，以因應不同的需求，越上層則品質越精細（Detail），越下層則品質越粗糙（Coarse），上層與下層同樣存在著相關性，視訊編碼利用這些相關性以進行有效率的編碼。



圖 2.4 為視訊編碼下多層次架構圖：最上層為無任何時間壓縮下，每單位時間需播出固定的畫面，每個畫面的前後存在著進階移動緒，利用移動向量找尋最接近的點；第二層為每單位時間需播出第一層一半數量的畫面，此時畫面減少為原來的一半，進階移動緒也因此往前或往後兩個畫面去尋找最接近的點，第三層更壓縮成原來的四分之一，最低層為原來的八分之一。另外，在經過時間小波分解後，原先為偶數的畫面（第一個畫面為偶數Frame<sub>0</sub>）產生出低頻畫面L Frame（Temporal Low-pass frame），原先為奇數的畫面產生出高頻畫面H Frame（Temporal High-pass frame）。在產生出第一層偶數的低頻畫面（L<sup>(1)</sup>）為第二層的低頻畫面（L<sup>(2)</sup>），第一層奇數的低頻畫面為第二層的高頻畫面（H<sup>(2)</sup>），以下類推。

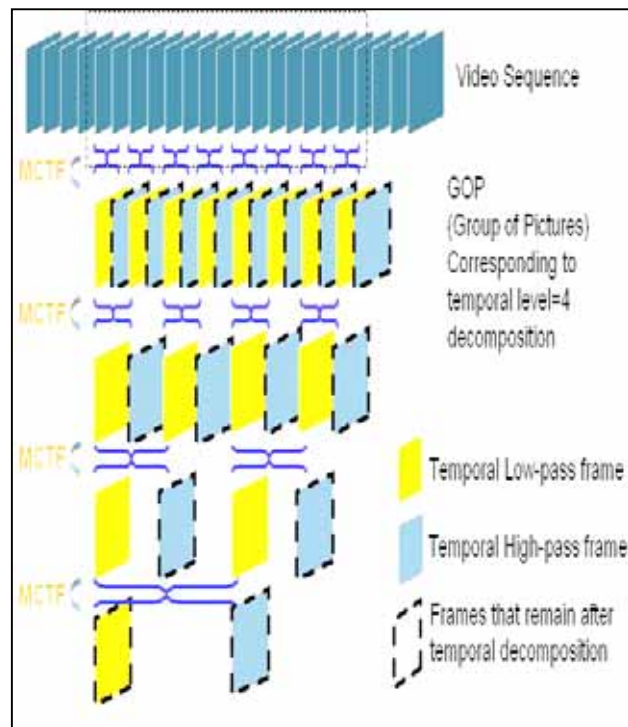


圖 2.4 多層次視訊編碼

進階移動緒在多層的架構下，為了編碼的效率，通常我們會將較粗糙（下層）的移動向量，來預測較精細（上層）的移動向量，並定義出不同的模式，考慮最佳編碼的情況下，選取適當的模式。

## 2.1.2 空間小波 S\_Module 研究

經由三維視訊 T\_Module 時間小波分解產生出高低頻相間的畫面後，接著由 S\_Module 空間小波再對高低頻相間畫面進行二維空間小波分解，以期讓每一個畫面能再分解出高低頻資料。如果說我們將原始的影像，看成是由影像中具有一致性的資料以及具有高度變化性的資料，S\_Module 主要就是將原始資料中分解成兩筆高低頻獨立的資料。我們將原始數列經過低頻分析濾波的步驟之後，所產生的低頻數列，將保留原始數列的一致性的資料：比如說人臉部特寫的影像中，臉部會有一大片區域的顏色值相近，或者是風景照中的背景天空的一大片藍色，就是原始資料中具有一致性的部分。我們將原始數列經過高頻分析濾波的步驟之後，所產生的高頻數列，則將保留原始數列的高度變化性的資料，比如說在影像中物體的輪廓部分，邊界內外的顏色值相差甚大，或者是有許多顏色參雜在一起的區域，都是原始資料中具有高度變化性的部分。



接著我們以常用的 Haar 小波轉換來舉例資料處理：假設我們有下列數列

$$\{ X_{n,j} \} = \{ 10 \quad 13 \quad 25 \quad 26 \quad 29 \quad 21 \quad 7 \quad 15 \}$$

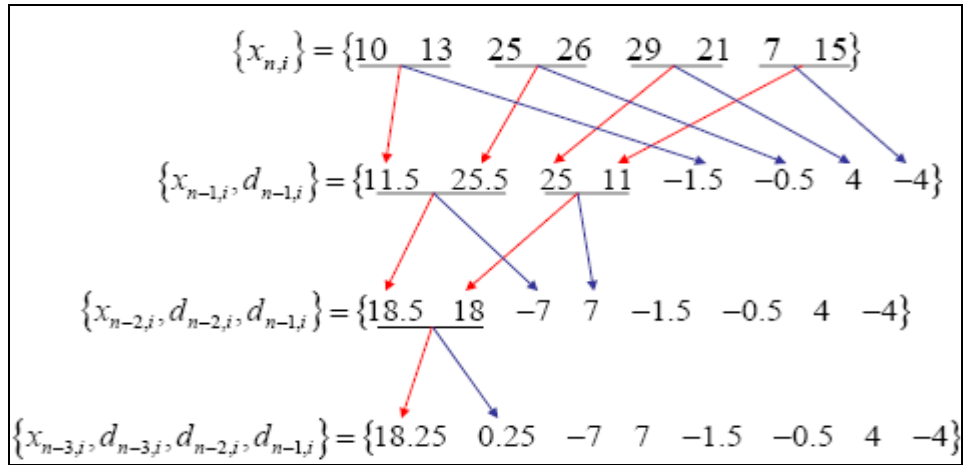
考慮把原數列以成對的  $X_{n-1,i}$  的平均值跟  $D_{n-1,i}$  的差值來取代原數列

$$X_{n-1,i} = (X_{n,2i} + X_{n,2i+1}) / 2 \quad ; \quad D_{n-1,i} = (X_{n,2i} - X_{n,2i+1}) / 2$$

假設把平均值放在前半部，差值放在後半部原來數列就變成

$$\{ X_{n-i}, d_{n-1,j} \} = \{ 11.5 \quad 25.5 \quad 25 \quad 11 \quad -1.5 \quad -0.5 \quad 4 \quad -4 \}$$

我們可以發現平均值的數量剛好是原數列量的一半，差值的數量也是原來數列的一半，所以經過轉換後的數量並沒改變，所以對影像來說經轉換後不會增加也不會減少要處理的資料數目。因為經過轉換後前半部是由平均值組成，後半部為差值組成，所以我們可以說上半部是原始資料的大概近似值，而後半部便是原始資料的誤差值。下列是原數列做三次小波轉換後的結果。



由上我們可以非常容易的驗證,經由以下的關係把轉換過後的數列還原成原來的數列

$$\begin{aligned}
 x_{n,2i} &= x_{n-1,i} + d_{n-1,i} \\
 x_{n,2i+1} &= x_{n-1,i} - d_{n-1,i}
 \end{aligned}
 \tag{3}$$

要將低頻數列與高頻數列還原成原始數列，我們並不需要保留低頻數列與高頻數列上所有的數值，由小波轉換的可還原性，我們只要保留低頻數列與高頻數列裡面一半的資料就可以達到還原的目的，我們可以將低頻數列與高頻數列裡面所有的數值選擇保留數列中的奇數項或者是偶數項。

壓縮過的資料進行解壓縮之後，其取樣數只有原來的一半，不能進行小波合成濾波的動作，因此我們要將被降低取樣的數列提高取樣，被保留下來的資料就擺在奇數項或者是偶數項，沒有被保留的偶數項或者是奇數項由於資料遺失，所以沒有被保留的項都設定為 0。

以量化的觀點來看，我們將原始的影像資料，經過小波分解得到低頻數列與高頻數列，高頻數列裡大部分的值都將非常接近 0，因此高頻數列具有相當大的壓縮空間。

圖 2.5 為 Frame 經過 S\_Module 轉換後示意圖，我們分別對行與列進行第一次小波轉換後，會產生 LL, LH, HL 和 HH 四組高低頻資料區。接著可以在對 LL 進行第二次小波轉換，如圖中標號 1 即為 LL Frame 經過兩次 S\_Module 小波轉換所得到 LL2 subband。

LL2 1	LH2 3	LH 6
HL2 2	HH2 4	
HL 5		HH 7

圖 2.5 Frame after S\_Module

圖 2.6 為 Lena 經過兩次小波轉換的結果，我們可以發現低頻數列，則將保留原始數列的一致性的資料，高頻數列，則將保留原始數列的高度變化性的資料，比如影像中物體的輪廓部分

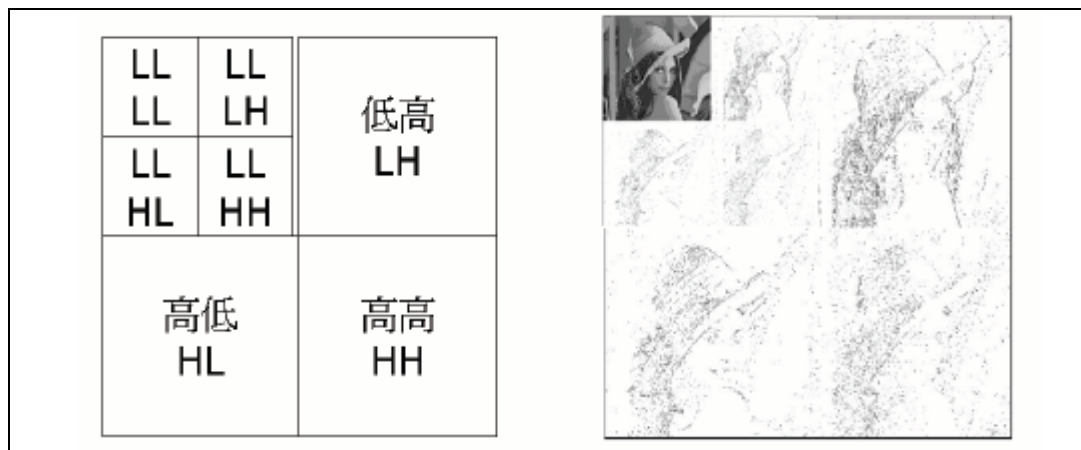


圖 2.6 兩次小波轉換的 Lena

### 2.1.3 壓縮編碼研究

經過時間與空間小波轉換後，接著對相似碼進行壓縮編碼，其採用 J. Xu, S. Li, Y.-Q. Zhang, and Z. Xiong 提出的 3D ESCOT (EmbeddedSub-band Coding with Optimized Truncation) 方法來做壓縮編碼。基本上 3D ESCOT 將經過時間和空間維度轉換過的影格，分割成一個個的 GOP (Group Of Pictures)，在編碼和解碼的過程中針對每一個 GOP 做處理。

### 2.1.4 圖例說明

經過時間小波第一次分解後，會產生高低頻畫面，高頻 H frames 主要由奇數畫面產生，並參考前後偶數畫面，低頻 L frames 主要由偶數畫面產生，並參考前後奇數畫面。接著我們將第一層產生的低頻畫面可在進行時間小波第二次分解，原先為第一層偶數的低頻畫面（為第二層的低頻畫面（LL），第一層奇數的低頻畫面為第二層的高頻畫面（LH），以下類推。時間小波完成接著進行二維空間小波分解，最後經過 entropy coding 產生出可調性視訊串流。

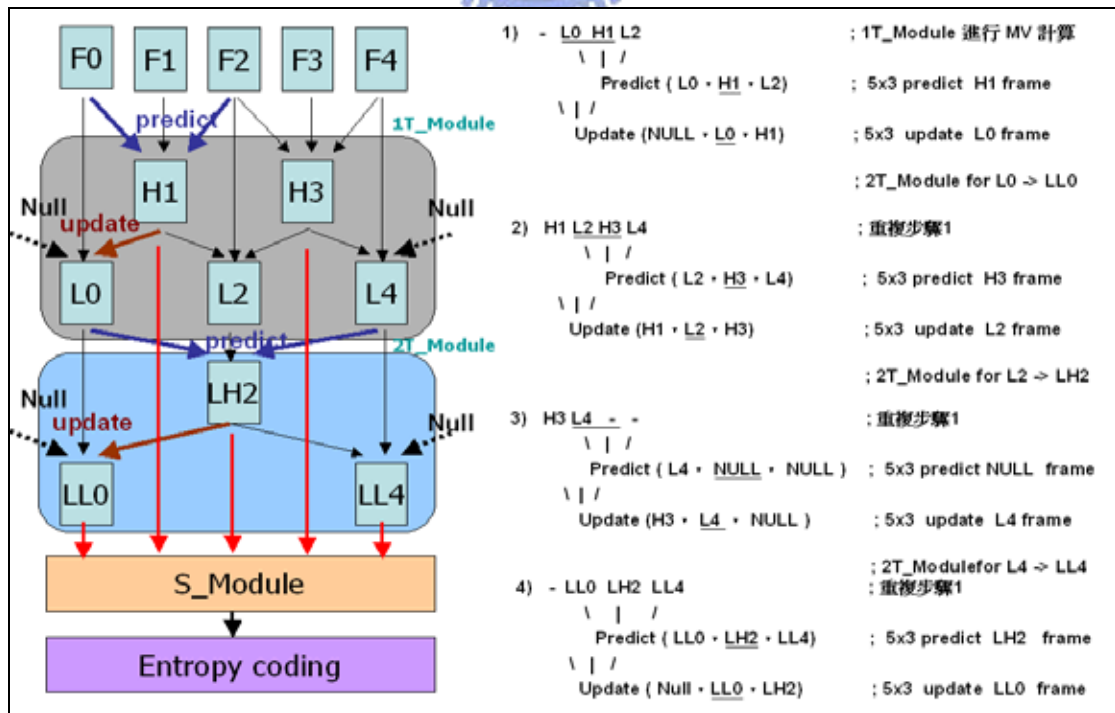


圖 2.7 可調性視訊編碼圖例說明

## 2.2 人臉偵測研究

人臉有一定的特徵，我們可以以人臉中的膚色、形狀、臉部特徵(眼睛和鼻子或其他)及臉部動態(如表情)等資訊，作為判斷人臉的依據。本章將對這些研究作簡單的敘述，介紹目前許多系統用到的臉部辨識的方法。

### 2.2.1 顏色判斷

大多關於人臉辨識的研究中，因為人臉膚色跟一般周圍環境的色彩會有顯著的差別，所以通常會以膚色作為人體在影像中的依據。為了將膚色表達出來，我們可採用顏色分割分析膚色分部區域。利用顏色分割把影像之中的各種顏色予以分類，我們可以將影像的顏色條方圖(color histogram)裡做多重閾值(multi threshold)，對不同的顏色分佈做選擇性的過濾，或者就針對不同的膚色做膚色(Skin Color)分析，建立起相關的膚色資料，將來只要有類似的顏色就判斷為膚色。這樣就可把目標以顏色作判別，將目標從影像中分離出來，如要把臉部跟背景分辨出來，就是要分析臉部的顏色分布。

在使用膚色作為人部辨識時，背景往往是影響判斷人臉正確性的要素之一，複雜的背景資訊往往會有和目標物相似的特徵，容易因顏色的相同而造成誤判，所以我們可以先作影像的前置處理，如直接把兩張連續影像相減，將靜止的背景去除。

在以顏色作為判斷人臉時，需考慮到由於大部份的電腦擷取系統都是以 RGB 作為基本顏色的表示方法。除了 RGB 彩色系統外，彩色系統另有好幾個系統，例如：YIQ、HSV、YUV 和 YCbCr 等知名的彩色系統。任兩種系統間都存在聯繫彼此的數學關係。如以 YUV 和 RGB 互相轉換系統中，存在著如下的數學關係：

$$R = Y + 1.371*V \quad ; \quad G = Y - 0.336*U - 0.698*V \quad ; \quad B = Y + 1.732*U$$

其中 Y 代表的是亮度，U 代表的是藍色和亮度的差，V 代表的是紅色和亮度的差。U 和 V 也可視為彩度。

## 2.2.2 外型特徵

顏色分割後找出可能的膚色區塊後，通常並不只有人臉膚色區塊，還會受到週遭環境同是膚色背景影響人臉偵測判斷。接著，我們可再依人的臉部特徵來和週遭環境之影像作進一步區隔，如人臉中的五官有眼睛、鼻子、嘴巴，這些都是明顯易辨的特徵。下面是一些系統辨識臉部特徵時用到的方法：

### 1. 眼睛的邊界及形狀

Xie 用 kalman filter 以及一組由不同曲線組成的眼睛模型，以影像的邊界去追蹤眼睛及眼珠的位置。Xin 則是利用 Hough Transform 來找尋灰階影像中臉部的圓形達到偵測眼珠的目的。

### 2. 五官的顏色、邊界及形狀

Wu 用顏色找出臉部大概位置後，再用一組曲線配合動態輪廓(Active Contour)的技術追蹤人臉、眉毛、眼睛、鼻子以及嘴唇。

### 3. 五官的顏色及灰階值

Sobotta 以臉部的顏色來找出人臉後，以灰階影像中眼珠為灰階值最小的特徵來判斷眼珠位置。用顏色找出臉部大概位置後，再用一組曲線配合動態輪廓(Active Contour)的技術追蹤人臉、眉毛、眼睛、鼻子以及嘴唇。



## chapter3 可調性視訊之人臉偵測

本章節提出可調性視訊編碼中，人臉偵測嵌入位置設計與可調性視訊人臉偵測方法。基於小波函數，我們在解碼端一次性嵌入人臉偵測系統，並達到最小資料量分析最快偵測結果。

### 3.1 系統與架構

解碼端收到資料串流後，按照編碼端之反方向，經由 Entropy coding 相似性解碼在經由二維小波空間(In\_S\_Module)執行合成處理，會產生兩種不同畫面：低頻畫面(L Frame)與高頻畫面(H Frame)。

我們選擇在空間小波合成 LL2 低低頻畫面過程嵌入人臉偵測系統，高頻畫面則不嵌入人臉偵測系統。低低頻畫面經由人臉偵測系統分析所產生的新低低頻畫面(LL2' Frame)將取代原低低頻畫，接著資料繼續流經可調性視訊編碼二維時間小波(In\_T\_Module)執行合成處理。

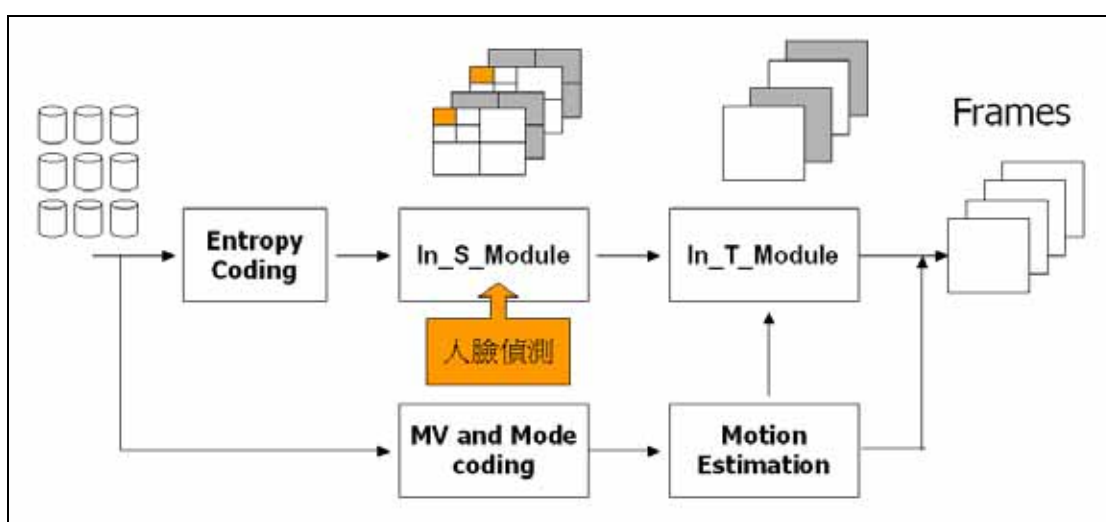


圖 3.1 可調性視訊人臉偵測架構



## 3.2 人臉偵測嵌入位置設計

爲了達到快速有效的可調性視訊人臉偵測系統，我們依可調性解碼時間小波與空間小波解碼特性，設計人臉偵測嵌入位置。首先，爲了不破壞原始資料，我們選擇在解碼 Decoder 端嵌入人臉偵測系統。接著，我們希望用最小資料量分析以達到快速偵測目標，我們考量以時間小波與空間小波解碼所產生的低頻資料區作爲人臉偵測嵌入切入點考量。

我們先以可調性編碼在 Encoder 端經過兩次空間/時間小波轉換  $T\_Module=2$ ， $S\_Module=2$  所產生的 bitstream 爲例，在 Decoder 端嵌入人臉偵測系統時，在空間小波反轉換  $In\_S\_module$  或時間小波反轉換  $In\_T\_module$  分析偵測人臉所需資料量。

### 3.2.1 空間小波解碼嵌入分析

假設一個 CIF 格式的視訊檔案，畫面解析度爲  $352 \times 288$ ，下圖 3.2 爲經過兩次  $S\_Module$  轉換後示意圖，標號 1.subband 代表最低解析度 LL2 Frame。

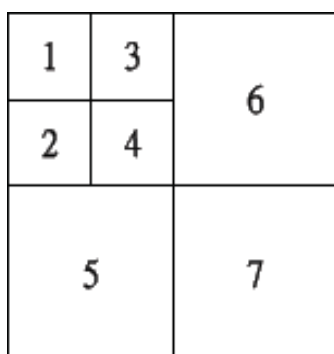


圖 3.2 兩次  $S\_Module$  轉換後示意圖

對可調性編碼 Decoder 進行空間小波反轉換  $In\_S\_module$  解碼時，其可還原出兩種不同可調性畫面解析度如圖 3.3 與圖 3.4。

Spatial scalable 1 :

還原 QCIF (176 X 144)格式視訊，其為編號 1+ 編號 2-4 組合

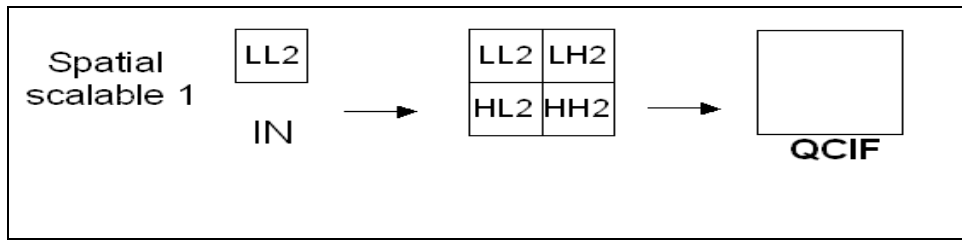


圖 3.3 Spatial scalable 1

Spatial scalable 2 :

還原原 CIF (352 X 288)格式視訊為例，其為編號 1+ 編號 2-4 組合 + 編號 5-7 組合

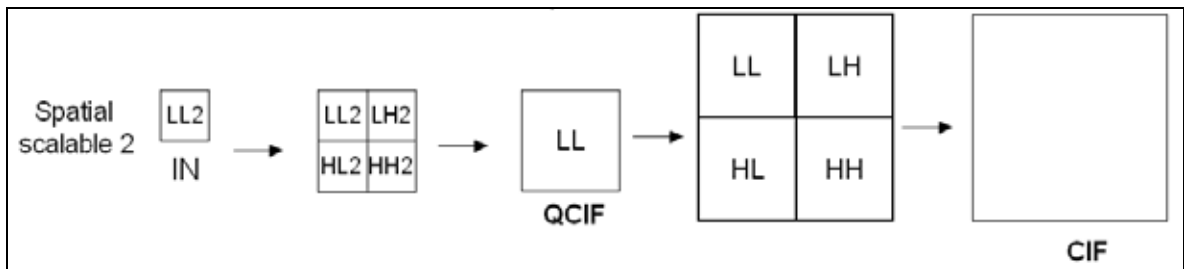


圖 3.4 Spatial scalable 2

針對可調性編碼空間小波反轉換 In\_S\_module 特性，經過兩次 S\_Module 轉換後，LL2 subband 代表最低解析度約佔單一 Frame  $1/4 \times 4 = 1/16$  大小，若選擇在 IN\_S\_LL2 嵌入人臉偵測僅需 1 / 16 資料量分析。

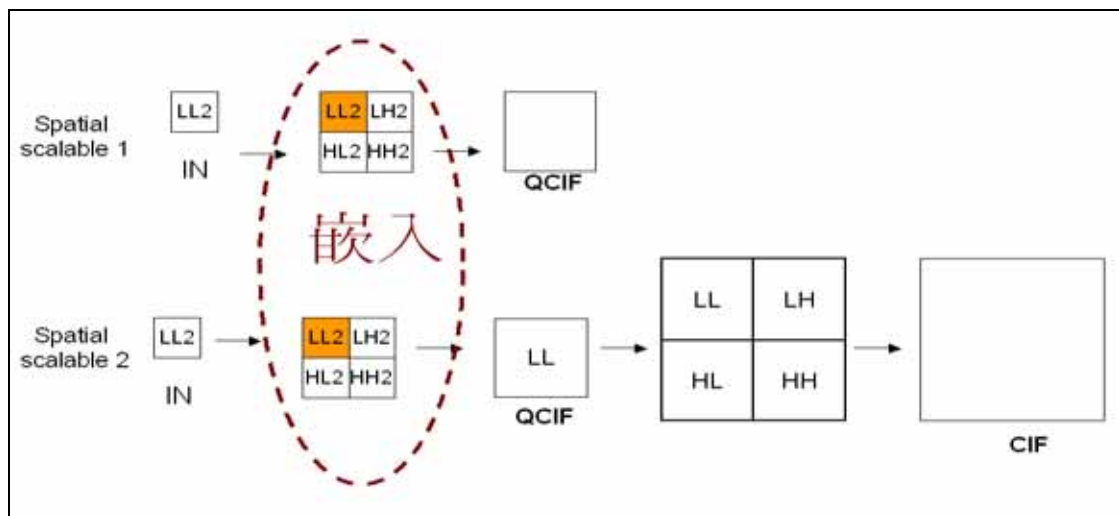


圖 3.5 空間小波人臉偵測嵌入設計

### 3.2.2 時間小波解碼嵌入分析

假設一個 video 有 5 Frame，下圖為經過兩次 T\_Module 轉換後示意圖。經過第一次時間小波分解後，原先為偶數的畫面會產生出低頻畫面 L Frame，原先為奇數的畫面則產生出高頻畫面 H Frame。經過第二次時間小波分解後，原第一層偶數的低頻畫面 (L0) 又可為第二層的低頻畫面 (LL)，第一層奇數的低頻畫面則為第二層的高頻畫面 (LH)。

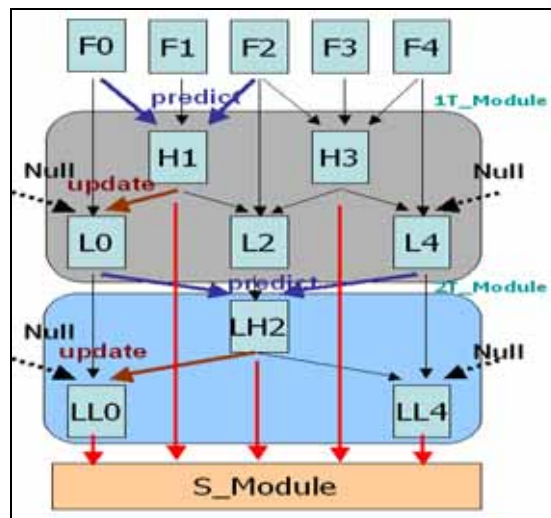


圖 3.6 兩次 T\_Module 轉換後示意圖

對可調性編碼 Decoder 進行時間小波反轉換 In\_T\_module 解碼時，其可還原出兩種不同 Frame 畫面個數

Temporal scalable 1 :

不解出原 H Frame 改以 Null 取代 -> 原 Frame 數還原減半

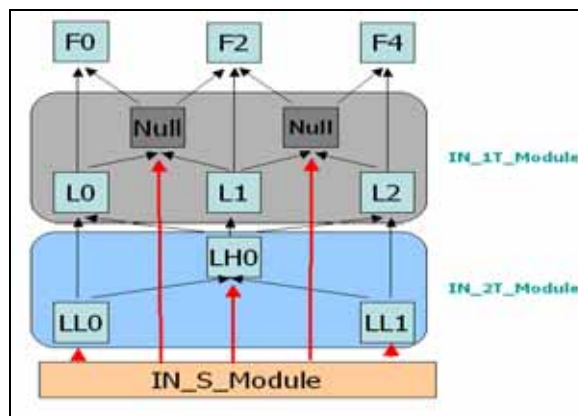


圖 3.7 Temporal scalable 1

Temporal scalable 2 :

解出 H Frame -> 還原回原 Frame 數

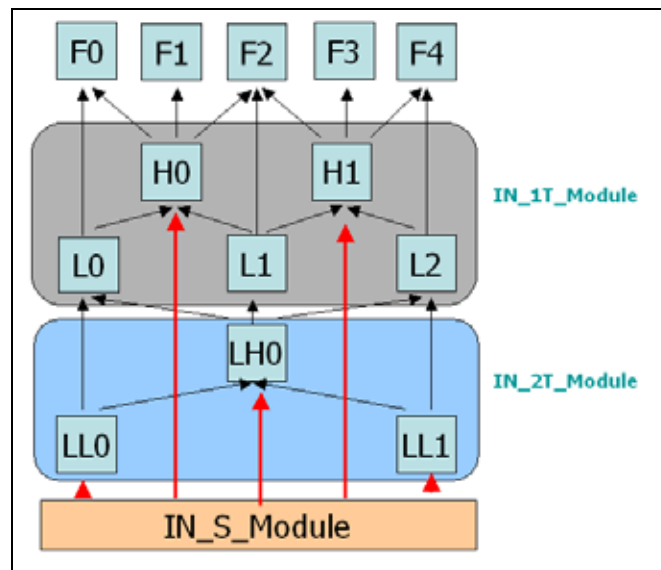


圖 3.8 Temporal scalable 2

針對可調性編碼空間小波反轉換 In\_T\_module 特性，經過兩次 T\_Module 轉換後，假設總 Frame 數為  $n$  個，選擇在 IN\_1T\_Module  $L_i$  嵌入人臉偵測，共有  $1/2n$  個  $L_i$ ，約需  $1/2$  資料量；選擇在 IN\_2T\_Module  $LL_i$  嵌入人臉偵測，共有  $1/2 \cdot 2n$   $LL_i$ ，約需  $1/4$  資料量

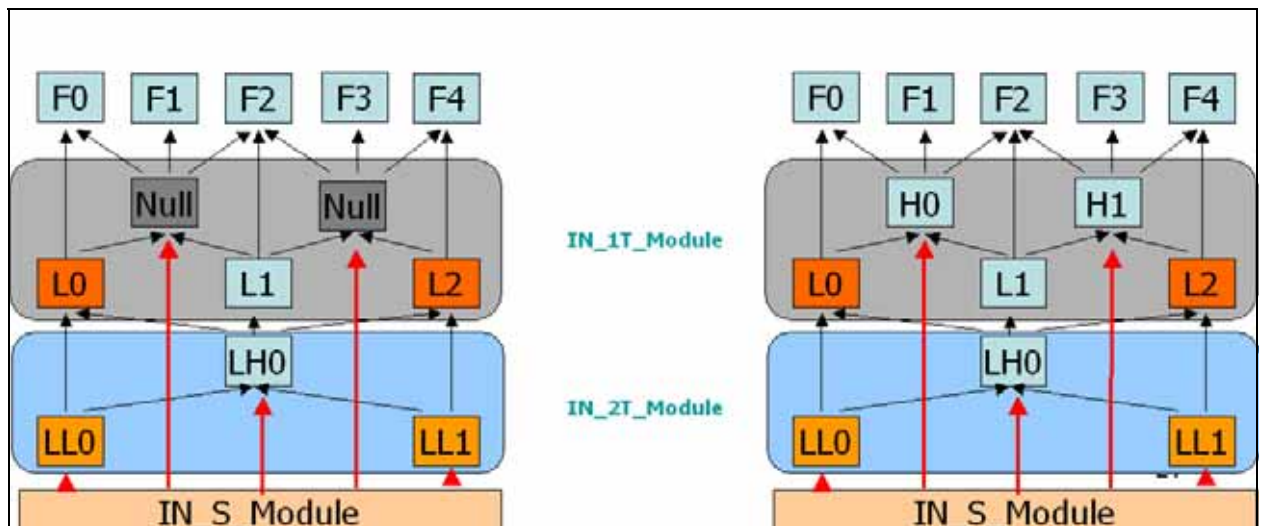


圖 3.9 時間小波人臉偵測嵌入設計

### 3.2.3 人臉偵測位置嵌入

綜合空間小波與時間小波可調性編碼特性，針對空間軸壓縮，選擇在空間小波 IN\_S\_Module LL2 分析，假設偵測假設總資料量為  $N$ ，僅需  $N/16$  資料量；針對時間軸壓縮，選擇在時間小波 IN\_2T\_Module LLi 分析，假設偵測假設總資料量為  $N$ ，僅需  $N/4$  資料量。

結合空間小波與時間小波特性的，我們可選擇時間小波與空間小波的低頻區 IN\_2T\_Module LLi 之 IN\_S\_LL2 嵌入人臉偵測。如圖 3.9，假設總資料量為  $N$ ，僅需  $N/4/16 = N/64$  資料量分析，達到最小資料量分析最快速偵測人臉。

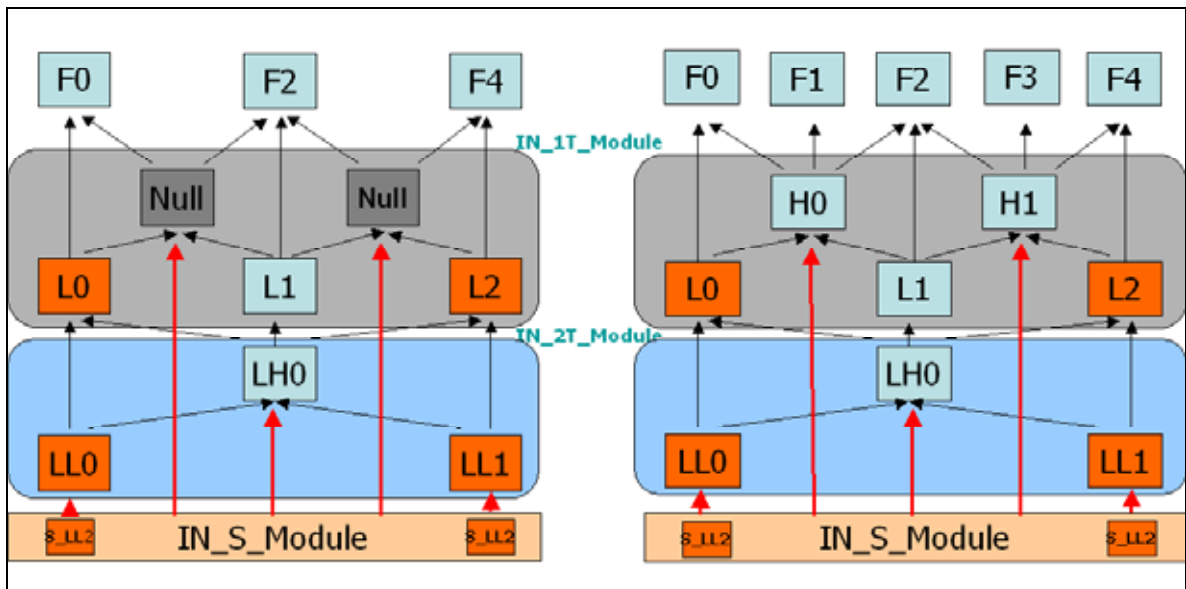


圖 3.10 人臉偵測位置嵌入

### 3.3 可調性視訊人臉偵測設計

人臉偵測嵌入位置確定後，接著我們將設計可調性視訊人臉偵測方式。我們以膚色過濾與外型特徵作為人臉偵測方式，可調性視訊人臉偵測流程如圖 3.11：

1. In\_S\_Module 端取得之 In\_S\_LL2 YUV 值  
主要為 In\_S\_LL2 端取得 YUV 值以進行人臉偵測第一步
2. In\_S\_LL2 進行人臉偵測  
於 LL2 嵌入人臉偵測技術，進行人臉偵測
3. 偵測人臉結果紀錄十字框位置  
將偵測結果以十字框紀錄位置並更新 LL2 Frame
4. Spatial scalable 完成  
繼續進行可調性解碼直到 Spatial scalable 完成
5. In\_T\_Module 執行  
接著繼續時間小波 In\_T\_Module 解碼

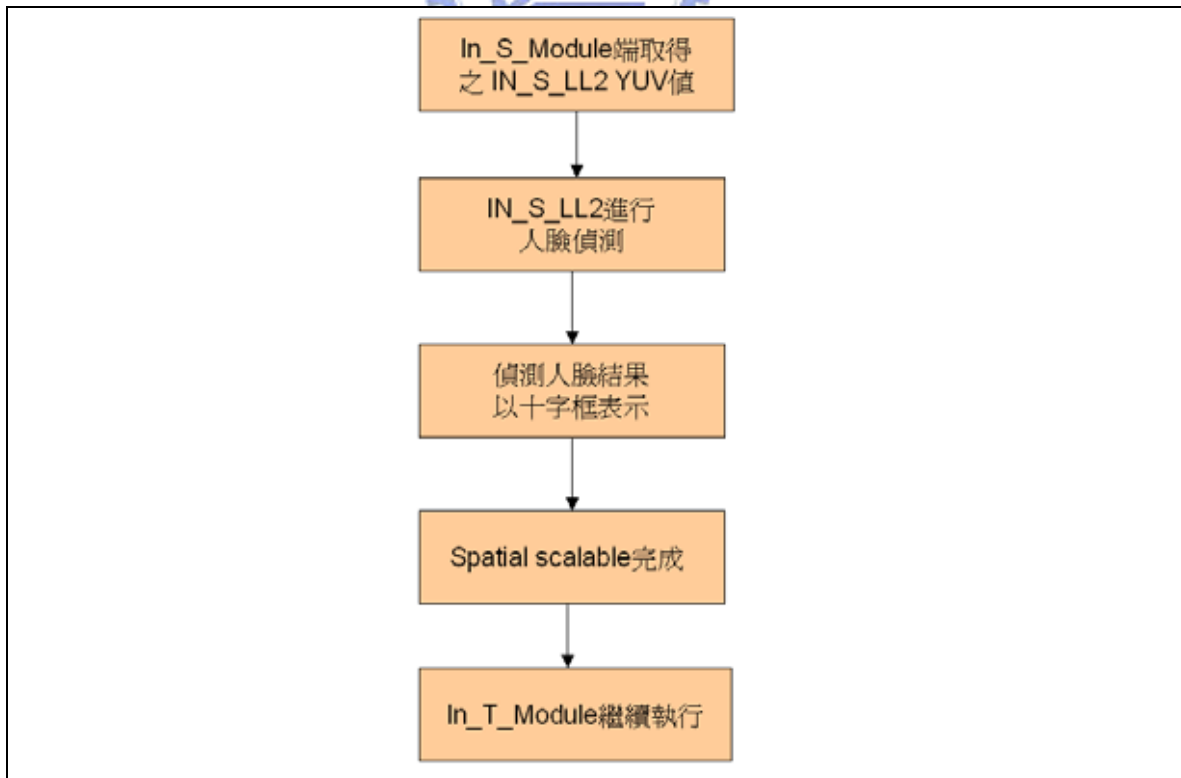


圖 3.11 人臉偵測流程

### 3.3.1 膚色 YUV 值取得

人臉偵測有很多方式，我們是以膚色過濾作為人臉偵測第一步。我們先從可調性編碼組合 QCIF 檔案大小過程，在 In\_2T\_Module LLi(QCIF)組合前取得 LL2 膚色 YUV 值來分析。可調適編碼中格式為 YUV，Y 代表的是亮度，U 代表的是藍色和亮度的差，V 代表的是紅色和亮度的差，U 和 V 可視為彩度，取的 YUV 值後，我們可再透過 YUV 與 RGB 轉換後過濾膚色。

圖 3.12 為轉換 QCIF 過程中取得 Y 值的縮小圖，橘色框為 LL2 低頻資料中，Y 值轉換前資料區與放大資料值顯示。我們可取此區 LL2 資料作為人臉偵測膚色值判斷。在圖 3.12 Y 值縮小圖中，橘色框資料值之外為 LH2、HL2、HH2 高頻資料區，高頻數列裡大部分的值都將非常接近 0。若我們將整個 Y 值區縮小，LH2、HL2、HH2 可明顯的表現出影像中物體的輪廓部分。

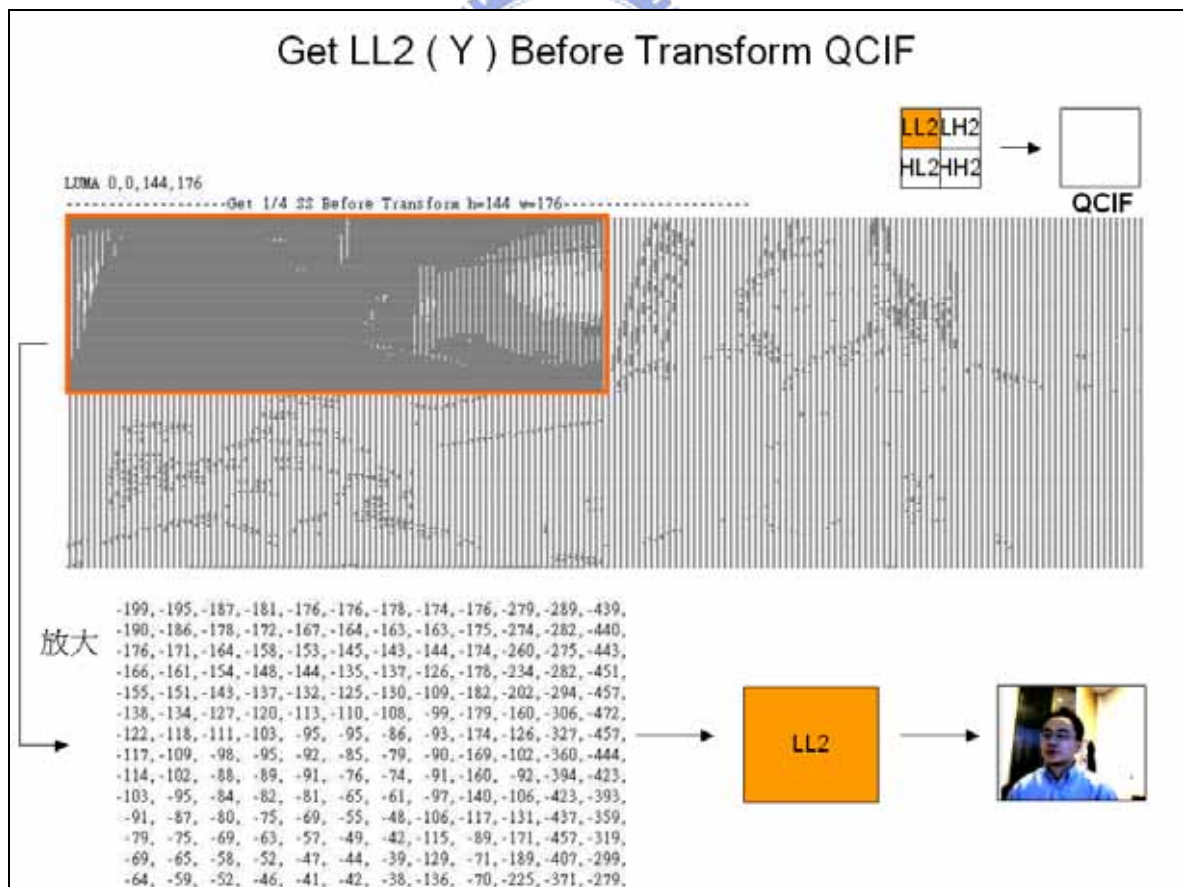


圖 3.12 Get LL2 ( Y ) Before Transform QCIF

圖 3.13 為轉換 QCIF 過程中取得 U 值的縮小圖，橘色框為 LL2 低頻資料中，U 值轉換前資料區與放大資料值顯示。我們可取此區 LL2 資料作為人臉偵測膚色值判斷。在圖 3.12 Y 值縮小圖中，橘色框資料值之外為 LH2、HL2、HH2 高頻資料區，高頻數列裡大部分的值都將非常接近 0。

我們可將 LL2 YUV 值收集分析，其資料展現會如右下圖中的人物，大小為 LL2 大小。針對 LL2 YUV 值分佈，我們可以對此 LL2 區塊進行膚色收集，分析膚色 model 以進行下一步過濾膚色區塊

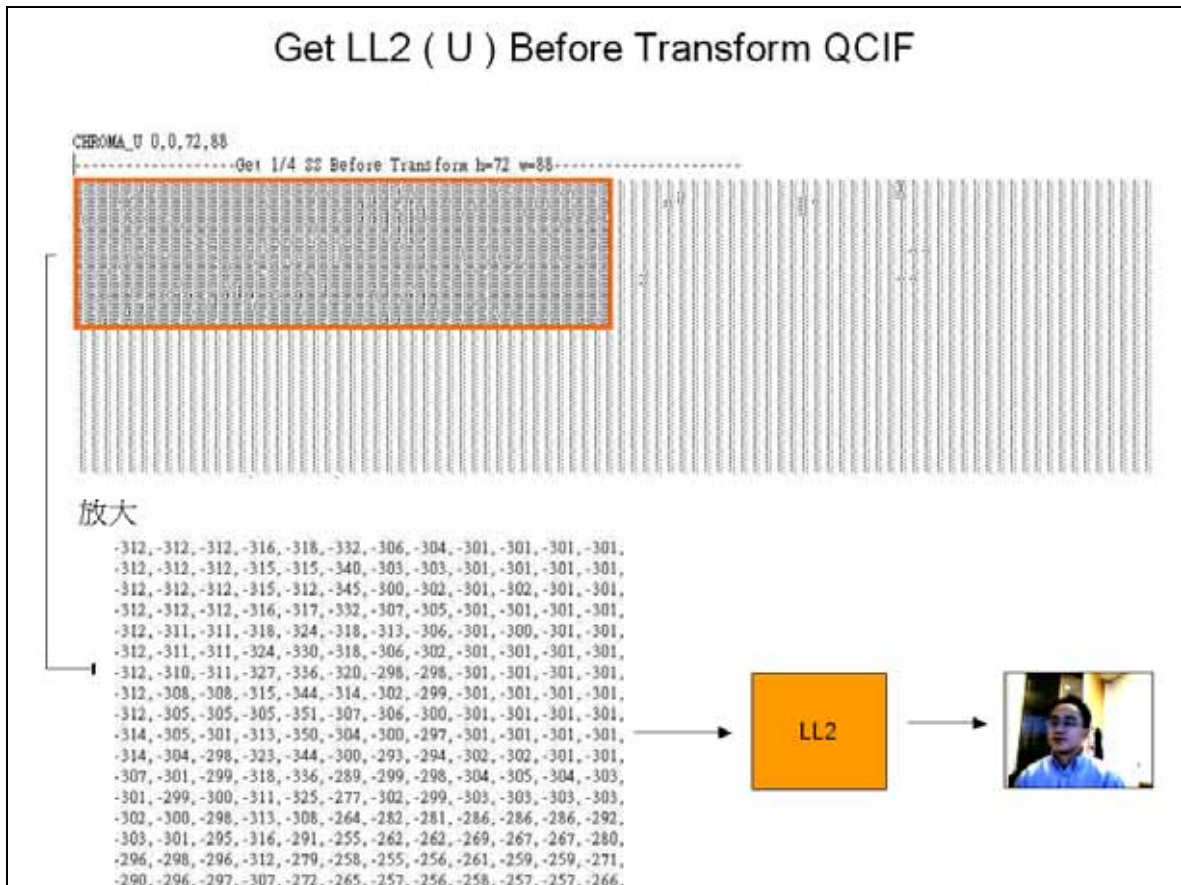


圖 3.13 Get LL2 ( U ) Before Transform QCIF

圖 3.14 與圖 3.15 為轉換完成 QCIF 之 Y 值 或 U 值縮小圖與放大圖，若我們將 QCIF YUV 值收集分析，其資料展現會如右下圖中的人物，其大小為 QCIF 大小。其中圖 3.14 為 Y 值轉換後縮小圖，可明顯的看出整個影像中物體的輪廓部分。



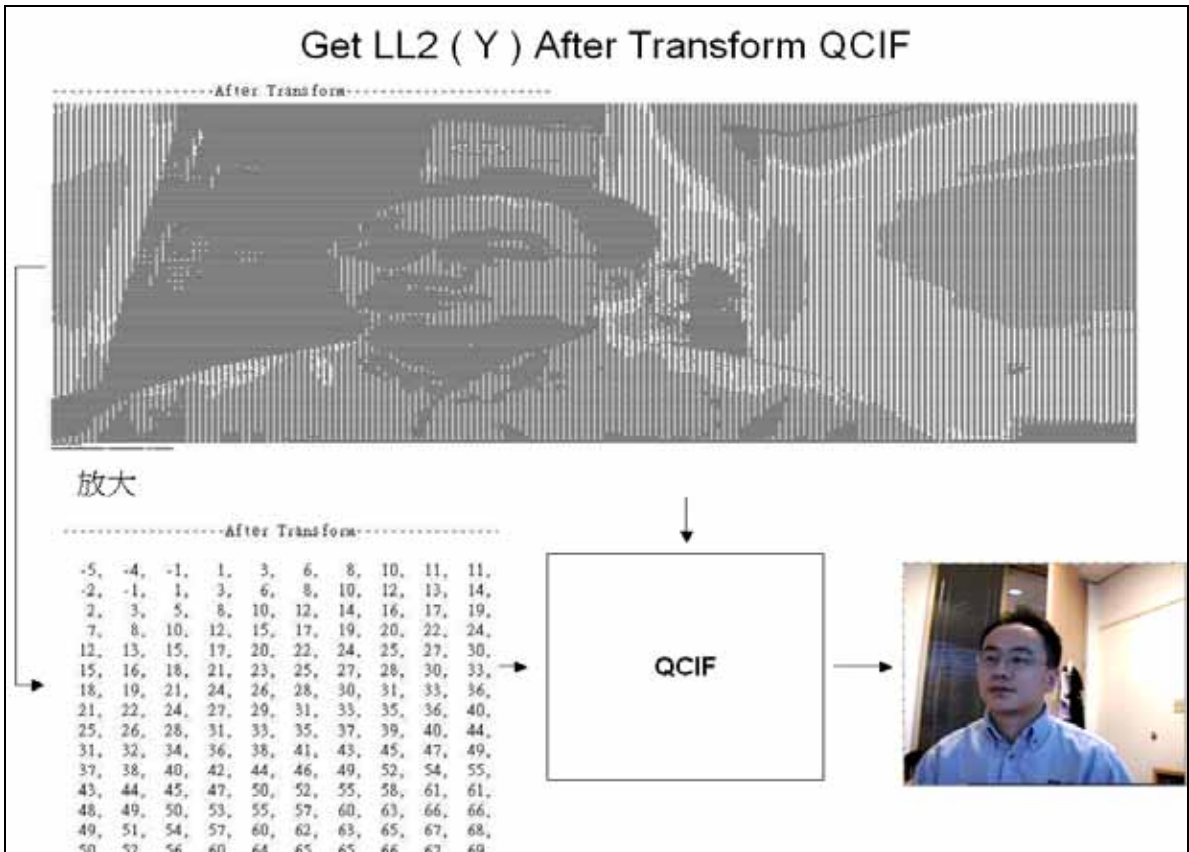


圖 3.14 Get LL2 ( Y ) After Transform QCIF

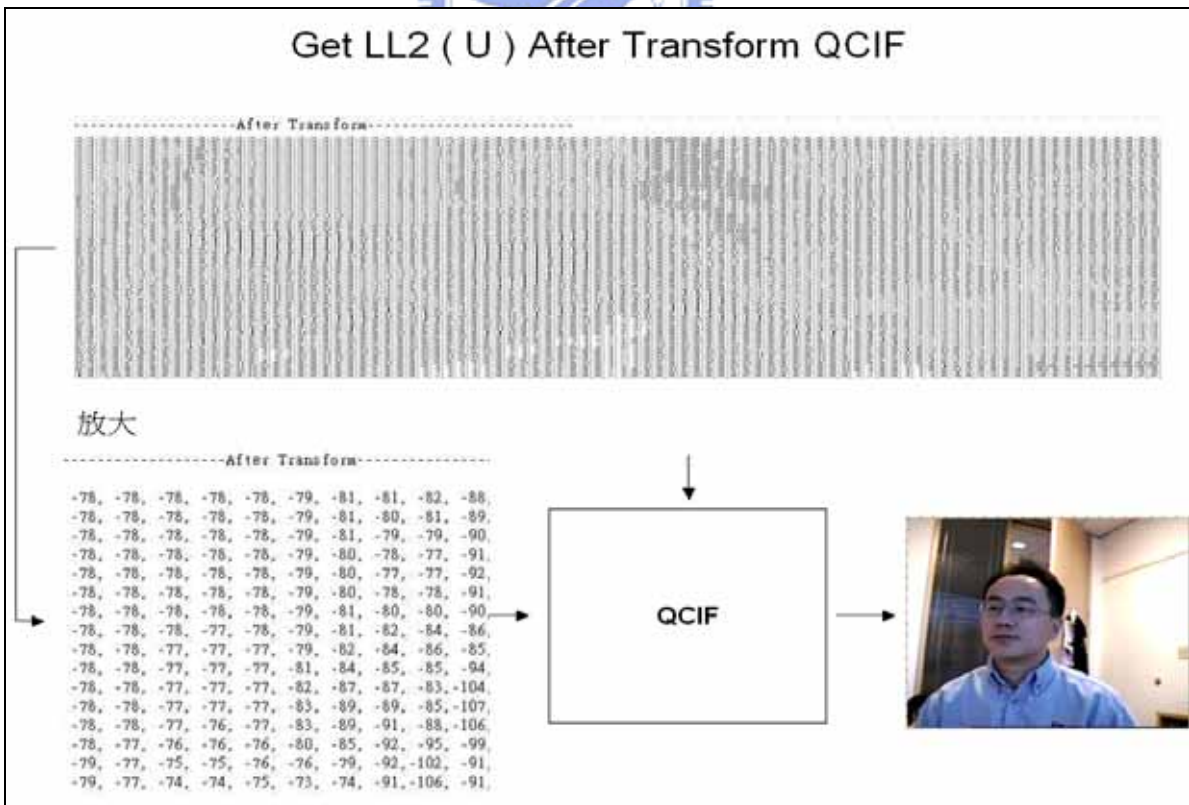


圖 3.15 Get LL2 ( U ) After Transform QCIF

### 3.3.2 膚色區塊篩選

取得 LL2 YUV 值後，接著我們可透過公式將 YUV 轉換 RGB 在轉成 YCbCr color space，透過膚色 model 過濾膚色。經過膚色分析處理後所找出屬於膚色亮點，通常除了人臉以外，還會有零星膚色雜點區塊，如下兩圖



圖 3.16 可調適編碼中之人臉偵測

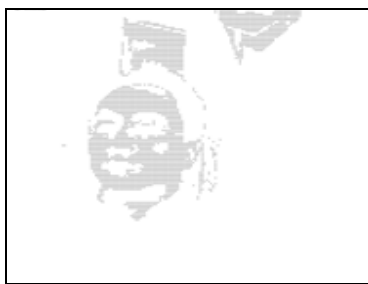


圖 3.17 膚色點偵測

接著我們將對非人臉區塊零星雜點的作過濾去除動作，我們設計是以光罩方式去除雜點，其演算法分別簡述如下。

光罩去除雜點：

1. 由圖中由上而下由左而右地搜尋膚色點數進行”光罩”過濾雜點
2. 我們以  $12*12$  pixel 為單位面積，每單位面積若有大於 50 個膚色點，保留此單位面積的所有膚色點。
3. 我們以  $12*12$  pixel 為單位面積，每單位面積若小於 50 個膚色點，則去除此單位面積的所有膚色點。

下圖 3.18 為進行光罩過濾膚色雜點的實驗結果，與上圖 3.17 比較，光罩可有效去除週邊零星的膚色雜點。



圖 3.18 光罩去除小雜點後



### 3.3.3 外型特徵判斷

通常得到膚色區域後，若背景含有膚色牆，將會影響到人臉判斷。於是我們在加入人臉眼睛的特徵作更精確人臉膚色區塊判斷。我們知道眼睛部份必定在膚色區域中，加上人的眼睛最大特徵為對稱，是人臉很好的辨別特徵。我們可以從已知的膚色區域找非膚色部分來判斷眼睛存在。

我們可根據膚色樣本分析，在 LL2 大小影像中，找出眼睛最小寬度與額頭與眼睛的長度比例，利用此特性在膚色區塊中尋找符合此特性的‘對稱’且‘非膚色區塊’。我們設計眼睛演算法簡述如下：

1. 找出一條連續的膚色 pixel 其長度為最小眼睛長度的 2 倍。
2. 以此膚色線開始往下最多六次開始掃描尋找非膚色區塊寬度。
3. 在同一水平線中，若能同時找到兩條非膚色區塊寬度為最小眼睛長度，則此水平線有眼睛存在。

下圖 3.5 為進行眼睛篩檢說明，在原圖中 line 50 行可找到有一條連續的膚色長度  $W$ ，其長度  $W \geq 2 * \text{最小眼睛寬度}$ ，接著我們開始掃描，我們可以在往下第六次掃描找到 line 55 有兩塊非膚色區寬度至少為眼睛寬度 9 pixel，我們便可判斷 line 55 為眼睛範圍區。

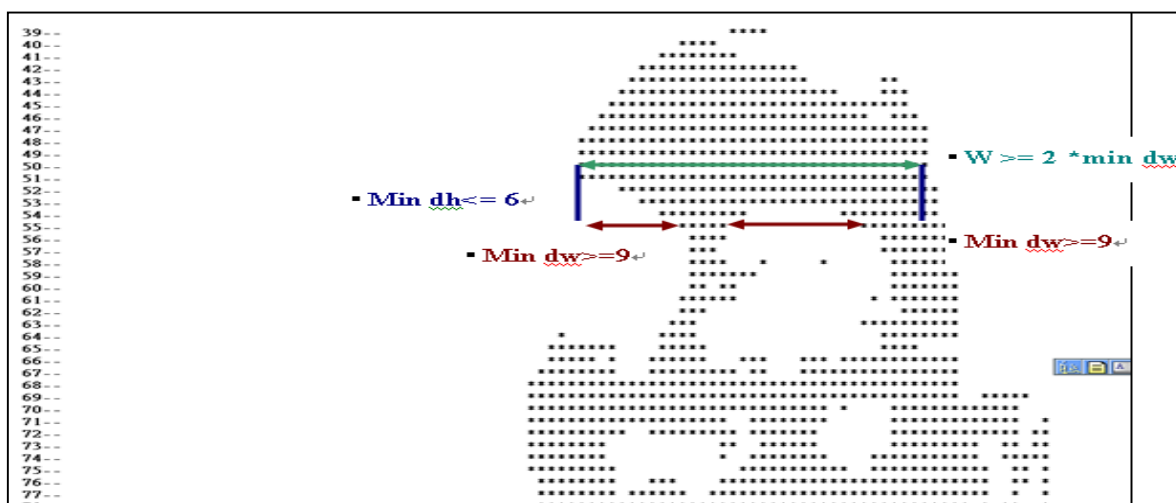


圖 3.19 眼睛判斷方式

### 3.3.4 可調性人臉偵測結果

LL2 subband 人臉偵測完成後，我們會對偵測人臉結果以十字框表示並更新 LL2' subband，接著繼續執行 In\_S\_Module Spatial scalable 與 In\_T\_Modul。

我們以還原原 QCIF(176 X 144 )格式為例，結果顯示如圖 3.21

1. 編號 1 LL2 subband 嵌入人臉偵測並紀錄人臉偵測位置
2. Update 編號 1 LL2 subband
3. 編號 1+編號 2-4 組合完成 QCIF 格式 (176 X 144 )



圖 3.20 可調性人臉偵測結果 1

我們以還原原原 CIF( 352 X 288 )格式為例，結果顯示如圖 3.22

1. 編號 1 LL2 subband 嵌入人臉偵測並紀錄人臉偵測位置
2. Update 編號 1 LL2 subband
3. 編號 1+編號 2-4 組合還原 QCIF 格式 (176 X 144 )
4. 編號 1 + 編號 2-7 組合還原原 CIF 格式( 352 X 288 )



圖 3.21 可調性人臉偵測結果 2

## chapter4 實驗結果與討論

本章將驗證可調性視訊人臉偵測系統，接收端可依其網路環境選擇人臉偵測系統，觀看人臉偵測影像。包含兩大實驗：

1. 時間可調變實驗
2. 空間可調變實驗

### 4.1 實驗影片

我們以 student video 為實驗對象，實驗影片格式如表 4.1：

1. student 解析度：352 x 288 (寬 x 高).
2. 播放畫面速率：10 (frames/sec)
3. 畫面數：30 (frames)
4. 檔案格式：YUV format



表 4.1 測試影片

## 4.2 實驗程式與環境

程式分為編碼端與解碼端，解碼端又可分為兩大部份。

編碼端主要部份是可調性視訊編碼，負責將原本的視訊影片經過三大模組與移動向量的編碼，產生出單一個資料串流送至解碼端。解碼端根據所收到的資料串流，反向經過三大模組與移動向量的解碼，產生出解碼影片。另一部份為人臉偵測系統部份，解碼端收到的資料串流後，負責在資訊串流透過可調性視訊解碼第一個模組（二維空間小波）反向產生出 L 低頻畫面過程進行人臉偵測。我們選在組合 L 低頻畫面過程之 LL2 低低頻畫面嵌入人臉偵測系統。透過膚色與外型偵測偵測出人臉位置後，更新 LL2 低低頻畫面，完成空間可調變後繼續進行視訊解碼第二個模組（時間小波）。

基於程式的兩大部份，輸入參數根據不同條件之要求，詳述如下：

- 可調性視訊編碼參數：

- 時間模組 (T Module)：在可調性視訊編碼中，經過幾次時間小波轉換，每增加一次，每秒播放畫面數減半。T = 0 代表時間無調變，T=1 代表時間調變一次。
- 空間模組 (S Module)：在可調性視訊編碼中，經過幾次空間小波轉換，每增加一次，每張畫面的寬度減半，高度也減半。S=0 代表空間無調變，S=1 代表空間調變一次。
- 速率 (Rate)：在可調性視訊編碼中，經過相似性編碼後，所產生的資料串流以不超過此傳輸頻寬 (bits/sec) 為主。

### 4.3 可調變實驗 I

本節實驗主要有兩小節，目的在驗證本論文中有關於可調變（空間、時間）的狀況下，於解碼端嵌入人臉偵測系統，可正確偵測出人臉實驗。

#### 4.3.1 時間調變實驗 I

本實驗目的是在不同時間壓縮下，人臉影像偵測結果的變化。

在可調性視訊編碼參數中，假設時間模組設為 2 ( $T = 0$  或  $T = 1$ )，空間模組設為 1 ( $S = 0$ )，即在本實驗下，原視訊在編碼端只經過兩次時間小波分解，當  $T=0$  時每秒播出的畫面數不變 (10 frames/sec)， $T=1$  時每秒播出的畫面數為原來的一半 (5 frames/sec)，此時畫面減少為原來的一半。另外原視訊在空間小波為無壓縮分解，解析度不變。

圖 4.1 為 student video 實驗結果，其中 (a)時間調變實驗 1： $S = 0$ ， $T = 1$  (b) 實驗結果 1 (c) 時間調變實驗 2： $S = 0$ ， $T = 0$  (d)實驗結果 2



圖 4.1(a)：時間調變實驗 1： $S = 0$ ， $T = 1$



# 實驗結果 1

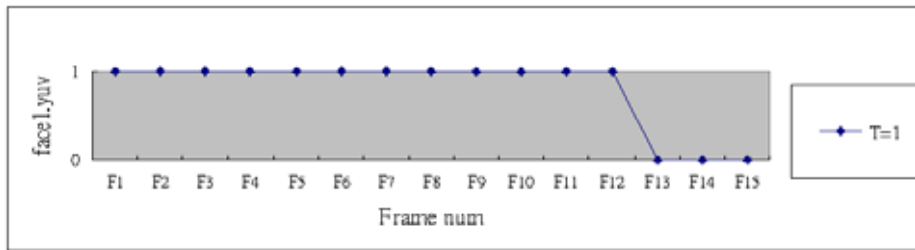


圖 4.1(b)：實驗結果 1



圖 4.1(c)：時間調變實驗 2：  $S = 0$ ， $T = 0$

## 實驗結果 2

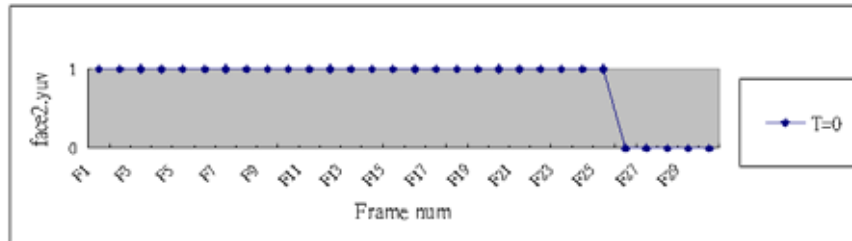


圖 4.1(d): 實驗結果 2

針對無法偵測出人臉影片，我們分析為圖片人臉過小原因所致。若原圖為  $352 \times 288$ ，我們選擇在  $S_{LL2}$  為  $88 \times 44$  分析人臉，雖可減少資料量分析達到快速偵測目標，但若是原圖人臉過小， $88 \times 44$  中再縮小的人臉雖可偵測出膚色區塊，但是區塊過小眼睛特徵不明顯，會造成人臉偵測結果變差，如下圖 4.1(e)

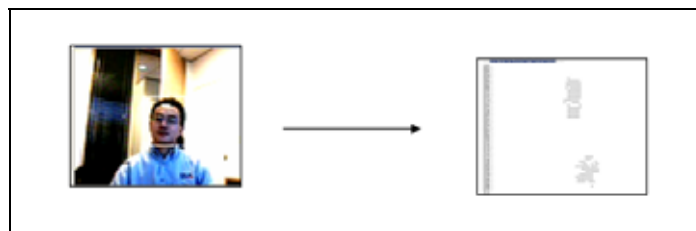


圖 4.1(e): 人臉偵測失敗圖

### 4.3.2 空間調變實驗 I

本實驗目的是在不同空間壓縮下，人臉影像偵測結果的變化。

在可調性視訊編碼參數中，假設空間模組設為 2 ( $S = 0$  或  $S = 1$ )，時間模組設為 0 ( $T = 0$ )，即在本實驗下，原視訊在編碼端經過兩次的空間的小波分解， $S = 0$  時解析度為  $352 \times 288$ ， $S = 1$  時解析度為  $176 \times 144$ 。另外原視訊在時間小波為無壓縮分解，每秒播出的畫面數不變。

圖 4.2 為 student video 實驗結果，其中 (a) 空間調變實驗 3： $S = 1$ ， $T = 0$  (b) 實驗結果 3 (c) 空間調變實驗 4： $S = 0$ ， $T = 0$  (d) 實驗結果 4

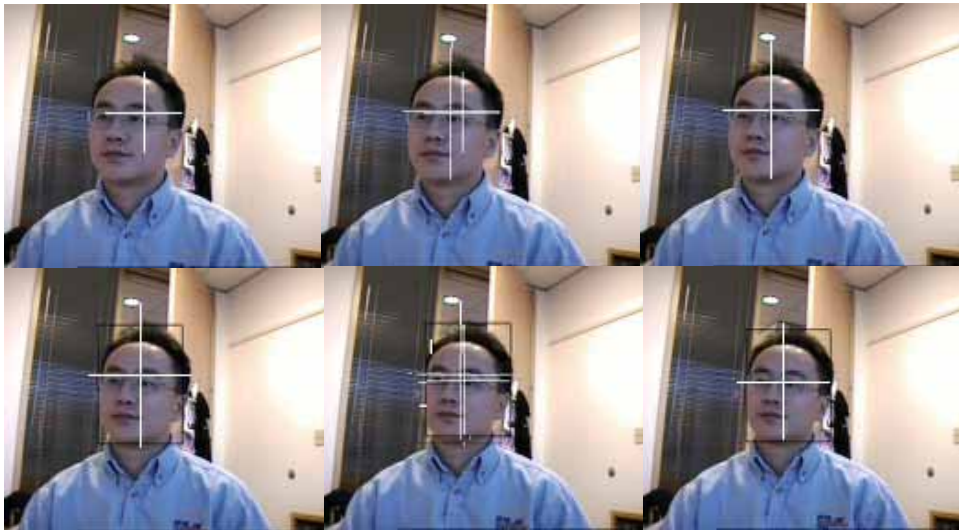


圖 4.2(a)：空間調變實驗 3： $S = 1$ ， $T = 0$

# 實驗結果 3

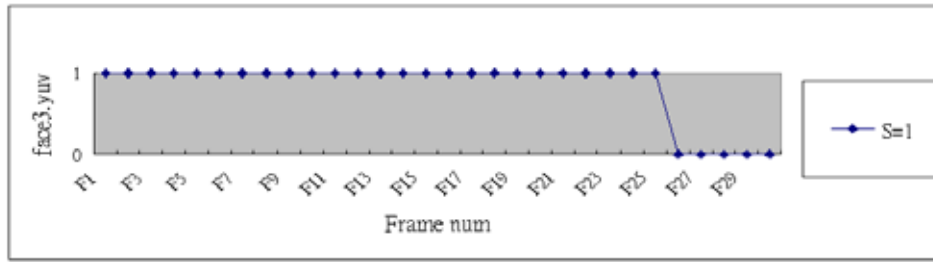


圖 4.2(b)：實驗結果 3



圖 4.2(c)：空間調變實驗 4：  $S = 0$ ， $T = 0$

## 實驗結果 4

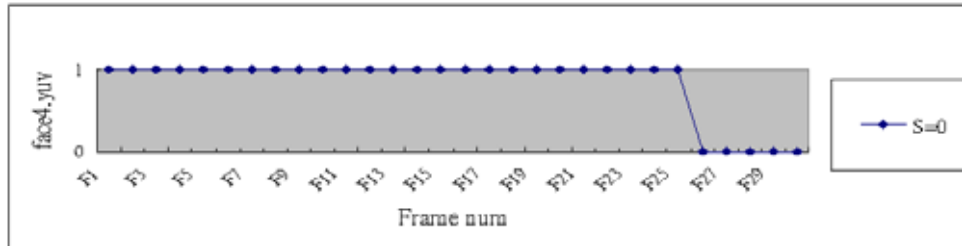


圖 4.2(d): 實驗結果 4

針對無法偵測出人臉影片，我們分析為圖片人臉過小原因所致。若原圖為  $352 \times 288$ ，我們選擇在  $S_{LL2}$  為  $88 \times 44$  分析人臉，雖可減少資料量分析達到快速偵測目標，但若是原圖人臉過小， $88 \times 44$  中再縮小的人臉雖可偵測出膚色區塊，但是區塊過小眼睛特徵不明顯，會造成人臉偵測結果變差，如下圖 4.2(e)

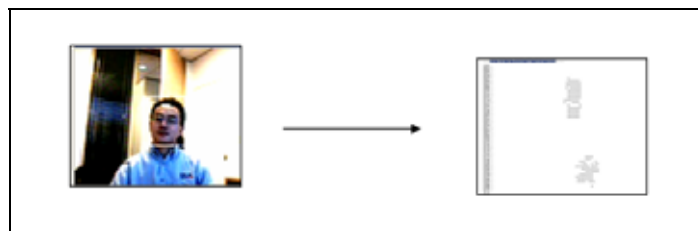


圖 4.2(e): 人臉偵測失敗圖

## 4.4 可調變實驗 II

我們選擇在空間小波與時間小波的低頻區：IN\_2T\_Module LLi 之 IN\_S\_LL2 嵌入人臉偵測，我們發現雖可用較少的資料量快速偵測出人臉，但在顯示人臉偵測結果容易有兩組十字架干擾問題，分析原因為若人臉移動顯著，IN\_T\_module 前後 LL2 Frame 人臉分析會有位差產生，由 LL2 還原 LH2 Frame 與 H Frame 時，便會造成兩組十字架，如下圖 4.3 左圖中 F1-F3 還原資料時會參考 LL0 與 LL2，若 LL0 與 LL2 人臉分析會有位差產生，F1-F3 便會有兩組十字架產生如右圖顯示：

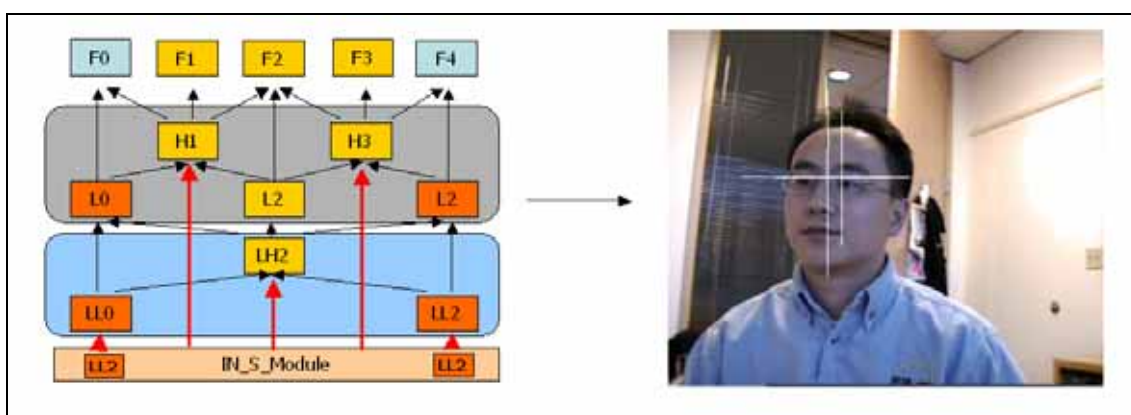


圖 4.3：十字架干擾圖

針對十字架干擾問題，我們想到若改在 IN\_T\_module 偶數 LL2 中還原 IN\_S\_module LL2 畫面進行人臉偵測，奇數 LL2 則不進行人臉偵測，將可改善此干擾問題。整個架構更改如下圖 4.4：如此 F1-F3 將只受 LL0 影響，同理，F5-F7 將只受 LL8 影響，可有效改善之前十字架干擾問題。接著我們將進行更改架構後的時間/空間調變實驗。

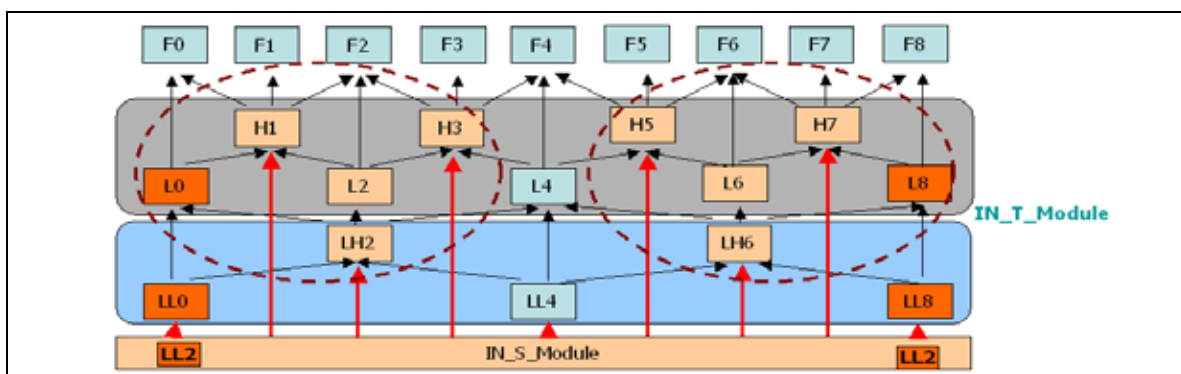


圖 4.4：架構更改

#### 4.4.1 時間調變實驗 II

圖 4.5 為 student video 實驗結果，其中 (a)時間調變實驗 1： $S = 0, T = 1$  (b) 時間調變實驗 2： $S = 0, T = 0$



圖 4.5(a)：時間調變實驗 1： $S = 0, T = 1$



圖 4.5(b)：時間調變實驗 2： $S = 0, T = 0$

#### 4.4.2 空間調變實驗 II

圖 4.6 為 student video 實驗結果，其中 (a) 空間調變實驗 1： $S = 1, T = 0$  (b) 空間調變實驗 2： $S = 0, T = 0$



圖 4.6(a)：空間調變實驗 1： $S = 1, T = 0$



圖 4.6(b)：空間調變實驗 2： $S = 0, T = 0$



### 4.4.3 實驗修改差別分析

我們發現修改架構後的實驗除了可降低十字架干擾問題，更可進階在縮小 1/2 資料分析量：若選擇在偶數 IN\_2T\_Module LL (LL8i)分析， 假設總資料量為 N， 僅需  $N/4/16/2 = N/128$  資料量分析，可再減少資料量 1/2。但是缺點為若在某 LL 偵測不到人臉，將影響前後 Frame 偵測結果，如下圖 LL8 偵測不出人臉，將會影響到 F5-F11 偵測結果。

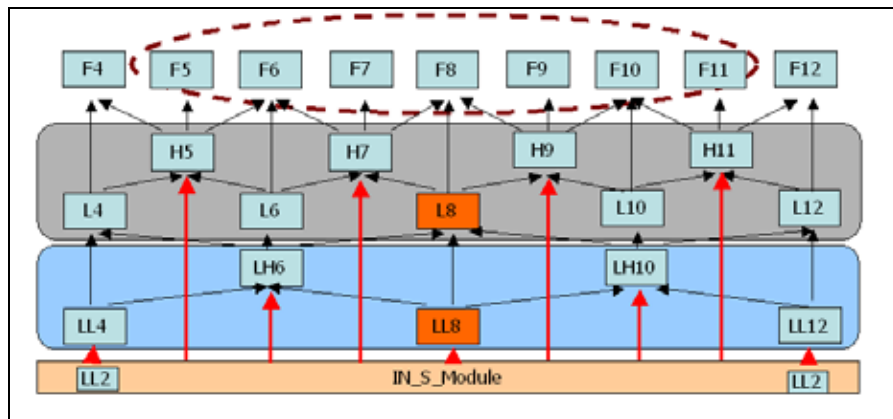


圖 4.7:實驗修改差別分析



## chapter5 結論與未來發展

本論文利用離散小波轉換多層次解析空間/時間特性，我們選擇於 Decoder 端 IN\_T\_module 偶數 LL 中還原 IN\_S\_module LL2 畫面 Frame 嵌入人臉偵測。達到在解碼端一次嵌入，資訊接收端可於不同網路環境下，依收到的不同等級資料串流偵測人臉影像，達到可調性視訊人臉偵測目標並能用最小資料量分析最快偵測結果。

由實驗結果顯示，產生出一個含有不同時間壓縮、或不同空間壓縮的資料串流在網路上傳輸。在接收端根據不同條件，儘量地還原到最佳品質，不同時間壓縮、不同空間壓縮的資料串流，內嵌的人臉偵測均可偵測出人臉影像。未來 LL2 偵測可搭配移動向量 MV 進行人臉偵測，我們僅偵測第一組 LL2 Frame，不但可再減少人臉資料量分析並進階達到人臉追蹤目標。

爲了達到最小資料量分析最快偵測結果，我們選擇在 IN\_S\_module LL2 畫面 Frame 嵌入人臉偵測，若是原圖人臉過小，LL2 中 88\*44 中再縮小的人臉雖可偵測出膚色區塊，但是區塊過小眼睛特徵不明顯，會造成人臉偵測結果變差，在未來展望方面，也必需考慮人臉過小時的影像偵測。



## 參考文獻：

1. Wenxian Yang, Feng Wu, Yan Lu, Jianfei Cai, King Ngi Ngan, Shipeng Li, Scalable Multiview Video Coding Using Wavelet, 2005
2. Xiangyang Ji, Jizheng Xu, Debin Zhao, Feng Wu, "Architectures of incorporating MPEG-4 AVC into three-dimensional wavelet video coding," Picture Coding Symposium, San Francisco, CA, USA, Dec 2004.
3. J.-Z. Xu, S. Li and Y.-Q. Zhang, "Three-dimensional shape-adaptive discrete wavelet transforms for efficient object-based video coding," to be appeared in Proceedings of SPIE VCIP'2000, Perth, Australia, July, 2000.
4. Lin Luo, Feng Wu, Shipeng Li, Zhenquan Zhuang, and Ya-Qin Zhang, A two-pass optimal motion-threading technique for 3D wavelet video coding, IEEE Intl. Symposium on Circuits and Systems (ISCAS 2002), pp.819-822, Scottsdale, AZ
5. Lin Luo, Feng Wu, Shipeng Li, and Zhenquan Zhuang, Advanced lifting-based Motion-Threading (MTh) techniques for 3D wavelet video coding, invited paper, to appear in SPIE/IEEE Visual Communications and Image Processing (VCIP2003), Lugano, Switzerland
6. Lin Luo, Feng Wu, Shipeng Li, and Zhenquan Zhuang, Layer-correlated motion estimation and motion vector coding for the 3D-wavelet video coding, IEEE International Conference on Image Processing (ICIP2003), Barcelona, Spain
7. Lin Luo, Feng Wu, Shipeng Li, Zixiang Xiong, Zhenquan Zhuang, "Advanced motion threading for 3D wavelet video coding", Signal Processing: Image Communication, special issue on subband/wavelet video coding, 2004
8. Wanzeng Kong, Shan'an Zhu, "A New Method of Single Color Face Detection Based on Skin Model and Gaussian Distribution 2006"

9. Kumar, C. N. Ravi Bindu, A., An Efficient Skin Illumination Compensation Model for Efficient Face Detection, 2006
10. Tsung-Ying Sun\*, Shang-Jeng Tsai and Vincent Chan, "HSI Color Model Based Lane Marking Detection," ITSC2006 – The 9th International IEEE Conference on Intelligent Transportation Systems, Toronto, Canada, Sept. 17-20, 2006.
11. Z. F. Liu, Z. S. You, A. K. Jain, and Y. Q. Wang, "Face detection and facial feature extraction in color image," in Proc. of the 5th International Conference on Computational Intelligence and Multimedia Applications, Sept. 2003, pp. 126-130
12. G. L. Foresti, C. Micheloni, L. Snidaro, and C. Marchiol, "Face detection for visual surveillance," in Proc. of the 12th IEEE International Conference on Image Analysis and Processing, Sept. 2003, pp. 115-120.
13. R. L. Hsu, M. A. Mottaleb, and A. K. Jain, "Face detection in color images," IEEE Trans. Pattern Anal. Machine Intell., vol. 24, pp. 696–706, 2002.
14. Xingming Zhang, Huangyuan Zhan: An Illumination Independent Eye Detection Algorithm. ICPR (1) 2006: 392-395
15. Fok Hing Chi Tivive and Abdesselam Bouzerdoum, "A Fast Neural-Based Eye Detection System," in Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems ISPACS2005, 2005, pp. 641-644.
16. Zhang L.M., Lenders P., Knowledge-based eye detection for human face recognition, In: Proc. Knowledge-Based Intelligent Engineering Systems and Allied Technologies[C], 2000(1): 117~120.