

行政院國家科學委員會專題研究計畫 成果報告

三維立體深度估測技術開發及其晶片設計(第2年) 研究成果報告(完整版)

計畫類別：個別型
計畫編號：NSC 96-2221-E-216-039-MY2
執行期間：97年08月01日至98年07月31日
執行單位：中華大學資訊工程學系

計畫主持人：鄭芳炫

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫涉及專利或其他智慧財產權，2年後可公開查詢

中華民國 98年09月24日

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

三維立體深度估測技術開發及其晶片設計

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC96-2221-E-216-039-MY2

執行期間：2007年8月1日至2009年7月31日

計畫主持人：鄭芳炫

共同主持人：

計畫參與人員：張佑維、張正園、林宣德、邱文冠、呂建德、鄭明輝

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：中華大學資訊工程學系

中 華 民 國 98 年 8 月 25

摘要

在進入數位資訊的時代中，科技日新月異，數位電視，手機，Webcam 等多種 3C 產品中皆有影像播放功能，平面影像已經漸漸不符合大眾所需求。目前市場上需要多元化的應用以及能為移動用戶帶來更多附加價值的服務，立體顯示便是其中之一。在一般人的觀念中，3D 與立體影像是相同的東西，其實不然，3D 是擁有正確的深度值，立體成像則是人們觀看具有立體感稱之為立體影像。立體影像並不需要絕對的深度值，只需擁有相對的深度值即可。為了使立體顯示可大眾化，本計畫是從單張影像中尋找消失線與消失點去估測深度，再利用原圖及深度圖計算左右雙眼影像給予立體顯示器使用。在實驗結果部分雖說有些缺點，但是在估測深度方法上已提供了一個基礎的方法。從實驗結果得知，求出的深度圖已經可以大部分估測出相對的深度，而錯誤的部分是由於光線、或者影像清晰度等問題所造成的影響。我們利用所做出的深度圖以及原圖去計算出雙眼影像配合 3D 立體顯示器去顯示立體影像，所呈現的效果非常不錯，也間接證明了我們所估測的深度圖有一定的準確性。計畫中我們亦設計了一個利用單眼影像所相對應的影像深度地圖產生雙眼立體影像的硬體架構。我們使用 DIBR 的演算法來產生出雙眼影像，再使用簡化的影像填補演算法填補產生的空洞。實驗中我們以 150MHz 的時脈效能就可達到即時的運算(320 x 240 @30 fps)。整個設計所使用的邏輯暫存器(logic registers)個數一共有 264,924 個邏輯暫存器。

關鍵字：消失點，消失線，立體影像，深度圖，影像填補，深度為基礎之影像重繪。

Abstract

In entering the age of digit information, most of 3C products such as digit TV, the cell-phone, and cameras have built-in display function. Since the function of display is mainly applied on 2D image, it is unable to fulfill the user's requirement at present. It needs pluralistic application and more additional value of service for the mobile subscriber on the market, e.g. the stereo image. In people's knowledge, 3D and stereo image are mis-recognized as the same things. In fact, 3D image has correct depth information and stereo image only has relative depth information. In order to make stereo image more popular, we focus on how to estimate depth information from single image in this project. We utilize original image and depth map to calculate the binocular image and use 3D LCD to display the stereo image. The concept of vanishing point is used to estimate the depth information. From the experimental results, the estimation of depth map is not perfect but it indeed offers the idea of depth estimation. Some defects of depth map estimation are due to the illumination and sharpness problems. By combining the depth map and 2D image to generate the binocular image and then display on a 3D LCD monitor, the stereo image appearance is satisfactory. This proves the feasibility of the proposed method. In order to achieve real time application, an image inpainting technique with its hardware design is also developed in this project.

Keyword: Vanishing point, Vanishing line, Stereo image, Depth map, Image inpainting, DIBR (depth-image-based rendering).

目錄

摘要.....	i
1. 前言.....	1
2. 研究目的.....	2
3. 文獻探討.....	3
3.1. 以多張的圖片去估測深度.....	3
3.2 以動態影像估測.....	4
3.3 取得相機參數.....	4
3.4 虛擬雙眼立體影像產生.....	4
3.4.1 立體攝影機與DIBR(depth-image-based rendering).....	4
3.4.2 影像填補.....	7
3.4.3 相關硬體設計.....	9
4. 研究方法.....	10
4.1 消失線的介紹.....	10
4.2 消失點的介紹.....	10
4.3 尋找消失線與消失點.....	12
4.4 建立深度漸層圖.....	12
4.5 估測深度圖.....	14
4.6 硬體設計及架構.....	15
4.6.1 整體架構.....	15
4.6.2 雙眼立體影像產生.....	16
5. 結果與討論.....	18
5.1. 拍攝條件限制.....	18
5.2. 實驗結果.....	18
5.2.1 深度估測.....	18
5.2.2 影像填補.....	19
5.3. 結論與建議.....	31
參考文獻.....	32
可供推廣之研發成果資料表.....	35

1. 前言

近年來，影像的品質、效果越來越讓人們注重，在科幻電影中通話以及許多有關影像的播放皆以立體影像方式呈現，在現實的生活中我們還是生活在平面影像，人們便漸漸的開始追求更立體的影像視覺。

立體影像的應用十分廣泛，特別是應用在許多產業中例如航太工業、生物醫學、軍事演習、地質探勘、環境探勘、建築行銷、汽車展示等。配合立體影像的環境能創造出更加具有效果的服務，如醫學上醫生可藉由多方面的角度觀察病因以提供更好的醫療，軍事演習上擁有立體的圖像可達到多方面考量，因此立體影像一定是未來的趨勢之一。目前市面上的立體顯示必須有特殊的器材與特殊的資料才可以呈現立體影像的效果，因此如何大眾化、便利化是本計畫的重點之一。

立體成像的形成是利用人的視覺原理，人類的雙眼就像兩台照相機，各自拍攝形成影像，成像落於視網膜中。視網膜就像照相機的底片，人的雙眼是高度相同且平行的。雙眼距離相差約 6-10 公分，所以左右眼中的景色不太相同。從一張圖片來說，以左眼而言，圖片最左邊某部分是右眼所沒有的，而右眼所看到的最右邊某部分是左眼所看不到，這個部分稱之為視差[1]。左右雙眼中相同景色的部分也會因為雙眼的距離差異而在左右眼中的成像位置也不同，人們的大腦中便是利用這些差異讓我們所看到的景色成為立體影像。目前在市場上許多立體顯示器是利用以上所敘述的原理所製造出來的，因此資料的輸入部分至少需要 2 張不同的圖片再藉由硬體的特殊設計所去呈現出立體影像。

擁有立體顯示功能的產品目前有很多種，最早的技術為紅藍眼鏡。現階段在市面上最常見的立體顯示器是以兩台的單槍配合特殊的立體轉換器，並且需要兩台攝影器材在固定的角度及距離下拍攝才可以顯示，而觀看者必須配戴特殊的立體眼鏡才可觀看到立體影像。在幾年前如 DTI、SHARP 等大公司已經成功研發出裸眼觀看立體影像的液晶螢幕，其原理如圖一所示，在液晶面版前面加上光柵（遮光棒），讓左眼只能看到左眼影像，右眼只能看到右眼影像，配合眼睛的視差讓大腦產生立體影像。

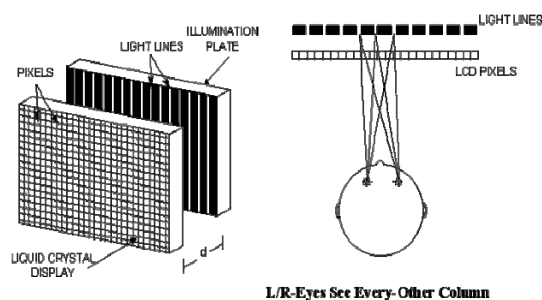


圖 1-1 為裸眼立體顯示器設計原理

雖然現在3D立體視覺的產品越來越多，但是還是沒有辦法普及到一般的使用者；因為立體影像取得不易，拍攝的儀器也非常昂貴，且一般的電視業者並沒有辦法提供這種立體影片的服務，因此現在大多都只有應用在遊戲以及電影上。所以為了讓一般的使用者更方便的就能享受立體視覺，將原有的2D影像與影片直接轉換成3D立體影像與影片是一種很好的方式。這種的轉換方式分為兩個部分：

一、 找出原始圖的深度資訊：可利用深度攝影機取得深度資訊，或是利用立體攝影機來計算相對的深度資訊，或是以影像分析的方式利用顏色、清晰度、幾何位置、物體大小...等方法，來估算影像中的深度資訊。

二、 利用原圖的深度資訊與原圖產生雙眼立體影像：利用深度資訊與虛擬立體攝影機的設定以及影像填補來產生雙眼立體影像。

當我們產生立體影像的過程中，由於移動了影像內物體的位置，造成影像中有空洞的問題，必須再利用影像填補(image inpainting)的方法將空洞填補起來，由於影像填補必須花費長時間的計算，並無法即時的產生，因此我們希望能夠將產生立體影像的處理加以硬體化，使得立體影像能夠即時的產生，讓使用者能夠更容易的就可以使用立體視覺的產品。

2. 研究目的

為了使立體顯示可大眾化，本計畫是直接以單張影像去估測深度，再利用原圖及深度圖計算左右雙眼影像給予立體顯示器使用。首先在單張圖片中先找尋消失線與消失點，在本報告的文獻探討中有介紹何謂消失線與消失點，且與深度又有何關係。在尋找消失點方面因為會受到某些因素的影響，所以本計畫在尋找方法上做了一些改良。而消失點是由圖片中找出的最遠距離的點，且位於水平線上，因此我們利用消失點的特性做出漸層的深度圖。再利用連通原理以及左右顏色比對辨別區塊，在相同區塊中以消失點為起始點給予相同的深度值。在計畫中我們對實際拍攝的影像做深度估測，所估測的深度雖說無法達到完全正確，但是相同物體或平面中已經可以給予正確的深度值。我們也利用立體顯示器去觀看我們所估測的深度圖與原圖所計算出的雙眼影像，在效果中主體部分已經有很明顯的可以浮現在螢幕前面了。因此在圖片背景比較簡單的情形下，依照本計畫的方法所估測出深度值已具有相對的準確度。當然本計畫中所估測的深度圖只是初步的成果而已，仍然還有許多地方待改進，如陰影、光線的影響、背景複雜度也會影響正確性。在實驗結果圖中可以看見錯誤的部分大多是光線或陰影的影響。而在單張估測深度方面還是有許多困難的地方得解決，如資訊過少。因此這部分在未來的工作中提出了一些簡單解決的看法。當我們產生立體影像的過程中，由於移動了影像內物體的位置，造成影像中有空洞的問題，必須再利用影像填補(image inpainting)的方法將空洞填補起來，由於影像填補必須花費長時間的計算，並無法即時的產生，因此我們希望能夠將產生立體影像的處理加以硬體化，使得立體影像能夠即時的產生，讓使用者能夠更容易的就可以使用立體視覺的產品。

3. 文獻探討

立體成像的原理主要分成主動式與被動式兩種：

主動式：透光切換眼鏡 (shutter glasses) 是以兩片 LCD 膜片來製作，利用 LCD 通電與否能產生膜片透光與否的特性，並且讓眼鏡與影片播映端都以高速交替切換左右畫面的方式，同步放映 (同步訊號校準利用有線或紅外線感應無線方式)，即當螢幕播映左眼畫面時剛好讓眼鏡的左眼透光，反之播映右眼畫面時剛好眼鏡的右眼透光；如此利用視覺殘留讓左右眼看到不同的畫面而產生立體感。

被動式：主要運用偏極光濾片將投影之光束分成水平或垂直兩種方向，將兩種角度不同的濾片分別裝在兩台投影機的鏡頭前，配合以兩片偏極光膜片製作的偏光眼鏡；而只讓水平偏極光的那一眼看到水平偏極光那一台投影出的畫面，讓垂直偏極光那一眼看到垂直偏極光那一台投影出的畫面，再運用視差方式而產生立體影像。

估測深度的方法可以分很多種，目前研究主要可分為三種：

1. 以多張的圖片去估測深度
2. 以動態影像估測
3. 取得相機參數

當我們得到這些深度的資訊後，就可利用深度資訊來做各種的應用，例如虛擬雙眼立體影像、建立 3D 立體模組....等。而一個深度地圖的表示方式如圖 3-1 所示。

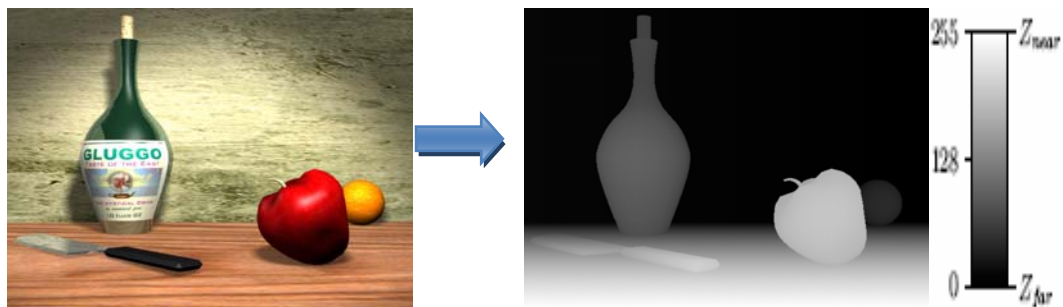


圖 3-1 影像深度地圖表示圖，一張影像中物體的遠近關係，我們可以用灰階值的大小來表示，當物體離觀看人越近，則灰階值越大，反之則越小

3.1 以多張的圖片去估測深度

在 2D 影像中關於深度資訊是非常少的，若可獲得稍有差異的圖片便可從中估測深度，因此有人利用視角相同和焦距不同拍攝影像從中估測深度值。在[2]、[3]、[4]、[6]這些論文中即是用相同角度及不同焦距所拍攝的圖片去估測深度。[2]是利用清晰與模糊兩張影像中邊緣的粗細與輪廓鮮明度並且取得焦距來估測深度值。我們可以知道當焦距不同時在每點的 Pixel 的清晰度就不一樣，尤其時物體邊緣的部分，若可以得知焦距那麼每點的清晰度的變化便可配合焦距去計算出來，再配合物體邊緣的清晰度便可以判斷每區塊的深度相對值。[3]是使用模糊理論 (Fuzzy Logic)，而[4]論文中是使用拓撲，[3][4]

兩篇論文中計算方法不一樣其原理是相同的，皆是利用 DFF(Depth-from-focus)和 DFD((Depth-from-defocus)兩種方法去改進。DFF 是針對清晰影像去作分析的，而 DFD 是針對模糊的影像去作分析。[3]是利用相機拍攝的模糊影像中每點 Pixel 的 PSF(point spread function)的情況跟 DFF 做轉換並比較出之間的差異，在得知焦距的情況下利用模糊理論去估測其深度。[4]與[3]的方法非常相似，一開始的步驟是相同的，都是利用 DFD 與 DFF 之間的特性先做轉換，再利用拓樸原理計算出深度圖。在這些方法中是以兩張圖片或以兩張以上同視角但拍攝焦距卻不同，利用圖片中清晰模糊以及邊緣等資訊去估測深度。[6]是[4]的改進，主要是使得估測準確性提高。

3.2 以動態影像估測

從視訊檔案中估測深度與多視角拍攝影像估測深度是一樣的，這與火車原理相同。當人坐在車廂中向外看，遠處的山移動的非常慢，而近處的景物很快速的移動。從每張 frame 中可以估測物體在畫面中移動多少，再利用三角原理計算出其深度。在攝影機不動時人物動，以及攝影機動人物不動兩種皆可以用火車原理來估測深度。首先得得知移動的速度，若是景色不動拍攝者動得知道移動者的速度以及移動的方向。若是物體移動拍攝者就不行動，同時也得取得物體的移動速度。以上兩種拍攝法皆得比較每張 Frame 之間差異的點在得知速度的情況下利用物體較近在 frame 間移動量就比較大，物體較遠則移動量就較小，可從這之間去估測其深度值。在[5]中便是利用物體在眼睛視角中移動多少來估測深度的。其優點是資料比較容易拍攝，缺點則是每 frame 中必須有固定的差異，如移動物體移動的方向需固定速度也得固定。

3.3 取得相機參數

因為多視角拍攝不方便且又得拍攝視角不能誤差太大，因此有人以一張圖片加上可獲得當時拍攝時相機的參數來估測深度。在這部分中得獲得相機的許多參數，如焦距，鏡頭的曲率，ISO 值等許多參數。這部分雖然是以單張圖片來估測深度，但是在他假設相機參數方面實際上是不能輕易取得的，因此在這部分的研究是非常少的。大多都是假設相機的參數以及用手動方式調整相機參數等方法來取得。雖說是以單張影像作估測深度，可是在取得相機參數方面便困難許多，每家廠商所做的相機中參數即使相同但拍攝的影像還是有差異的，而且在某些參數方面根本是不存在的，如鏡頭的曲率。

3.4 虛擬雙眼立體影像產生

當我們取得影像中的深度資訊後，就可以利用立體攝影機的設定與 DIBR 產生虛擬的雙眼立體影像，而當我們產生虛擬雙眼立體影像後就會出現空洞的問題，這些空洞會造成我們觀看時的障礙，因此我們必須使用影像填補的方式將立體影像中的空洞填補起來。

3.4.1 立體攝影機與 DIBR(depth-image-based rendering)

立體攝影機的設置方式[7~11]有很多種，一般使用的方式通常為 Toe-in 以及 Off-axis 這兩種方式。

Toe-in：兩個攝影機有相同固定與對稱的孔徑，鏡頭向內指向相同的焦點。缺點是

當物體在兩個攝影機的焦點上時，物體會失去立體的效果；當兩個攝影機焦距的距離與播放距離的比例如果沒有設定好，會造成影像物體比實際物體大或小的結果。而 Toe-in 攝影機設定的示意圖如圖 3-2 所示。

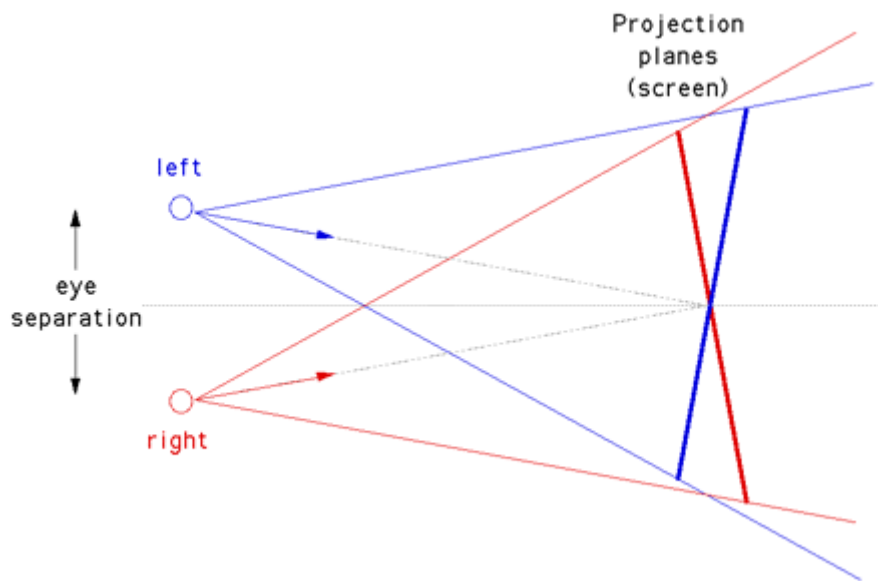


圖 3-2 Toe-in 法之虛擬攝影機設定，鏡頭向內拍攝物體

Off-axis :

兩個攝影機有相同固定與對稱的孔徑，鏡頭以平行的方式設置。而 Off-axis 攝影機設定的示意圖如圖 3-3 所示。

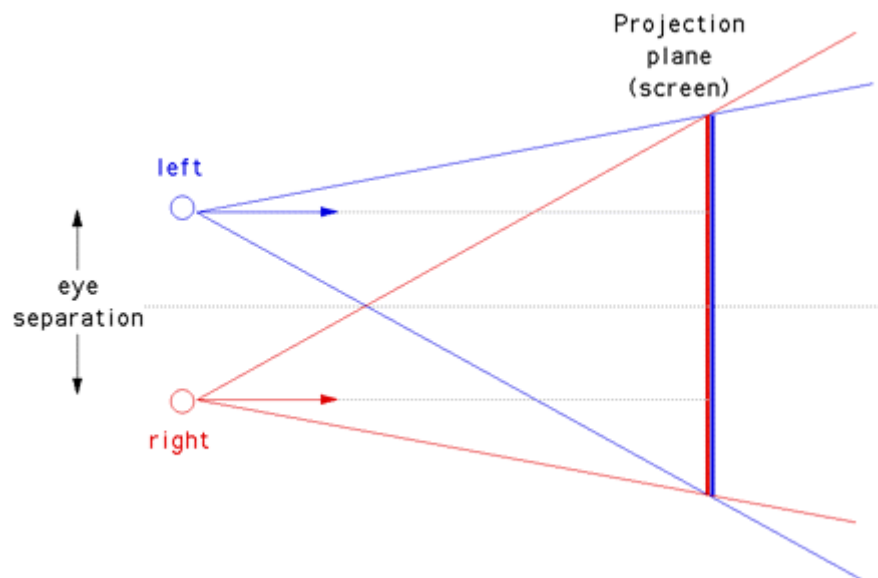


圖 3-3 Off-axis 法之虛擬攝影機設定，鏡頭平行拍攝物體

當我們選定好立體攝影機的模型後，就可以用 DIBR 來產生出虛擬的立體影像了。DIBR (depth-image-based rendering)[12~13]是以原始影像所相對應的影像深度地圖為基

礎，使用一個 Shift-sensor 立體攝影機模型(Off-axis)[12~13]，利用深度的資訊與立體攝影機的幾何關係，來改變影像內物體的位置，產生出一組有視差的虛擬雙眼影像。

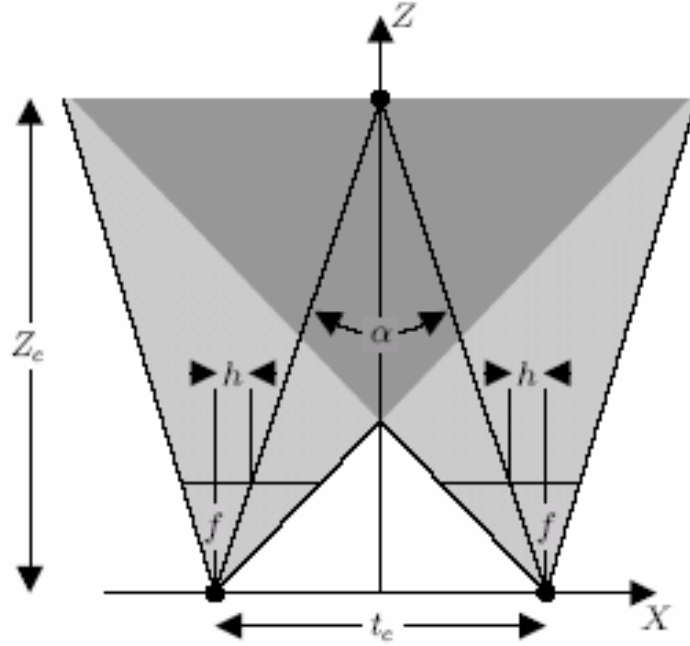


圖 3-4 Shift-sensor 立體攝影機設定圖[12~13]

$$u^* = u + \frac{ft_x}{Z} + h \quad v^* = v \quad (3-1)$$

u^* 代表雙眼視角像素的 X 軸座標， u 為原始影像像素的 X 軸座標， v^* 代表雙眼視角像素的 Y 軸座標， v 為原始影像像素的 Y 軸座標， f 為攝影機焦距， t_c 為攝影機間的間距 Z_c 為兩攝影機光軸與拍攝物的焦點距離， h 為攝影機光軸與拍攝物的偏移量。

$$t_x = \begin{cases} -\frac{t_c}{2} & , \text{ left - eye view} \\ \frac{t_c}{2} & , \text{ right - eye view} \end{cases} \quad (3-2)$$

將 $Z = Z_c$ 帶入方程式(3-1)中，則 $u^* = u$ ，可以得到方程式(3-3)

$$h = -t_x \frac{f}{Z_c} \quad (3-3)$$

將原始影像像素每一點的座標，依照虛擬攝影機模型的座標公式以及深度資訊，來決定雙眼視角影像像素所對應的座標位置。Shift-sensor 的優點是可調整 f 攝影機焦距與 t_c 攝影機間的間距這兩個參數值，來改變視差的大小。

3.4.2 影像填補

當我們產生出虛擬的立體影像後，所衍生的空洞問題必須用影像填補將空洞都填補起來，成為一個完整的虛擬立體影像，而影像填補演算法[14~20]基本的概念為從空洞區域周圍尋找填補方塊並將它們的內容複製至空洞區域。在圖 3-5(a)中， Ω 為目標區域，也就是空洞區域； $\delta\Omega$ 為目標區的輪廓； Φ 為來源區域。在圖 3-5(b)中，我們要合成由中心點為 p 的填補方塊 ψ_p 所包含的區域，在圖 3-5 (c)中，最相像的的填補方塊很可能出現在兩塊不同顏色或紋路的邊界上，如 ψ'_p 與 ψ''_p 。圖 3-5 (d) 為找到最相似填補方塊後，將其相對位置的像素複製至 ψ_p 中。

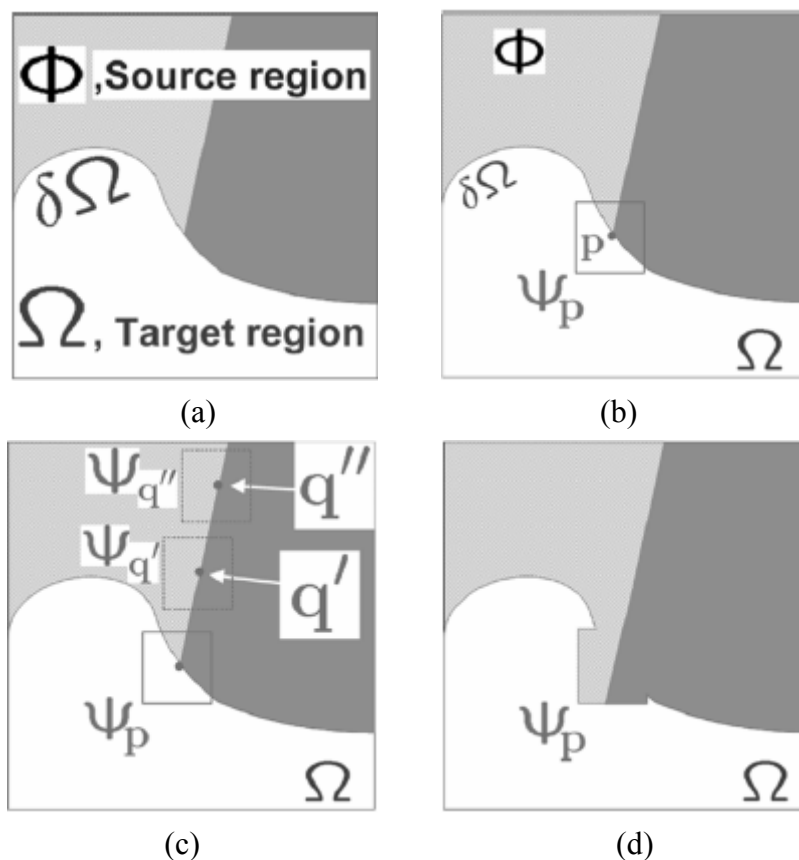


圖3-5 填補演算法示意圖 (a) 原始影像 (b) 目的填補區塊(c) 搜尋來源填補區塊 (d) 填補後的結果

步驟 1: 計算優先權 $P(p) \quad \forall p \in \delta\Omega$ ，以找出目標區域的輪廓上各像素優先填補順序。

步驟 2: 找出來源填補方塊 $\Psi_{\bar{q}} \in \Phi$ 且擁有 $d(\Psi_{\bar{p}}, \Psi_{\bar{q}})$ 最小值，將其影像資料從 $\Psi_{\bar{q}}$ 複製至 $\Psi_{\bar{p}}$ 。

$$\Psi_{\hat{q}} = \arg \min_{\Psi_q \in \Phi} d(\Psi_{\hat{p}}, \Psi_q) \quad (3-4)$$

步驟 3: 更新信心值 $C(p) = C(\hat{p}) \quad \forall p \in \Psi_{\hat{p}} \cap \Omega$ 。

在計算優先權時，我們必須給定一個填補方塊 Ψ_p ，其中心點位於 p ， $p \in \delta\Omega$ ，如圖 2.6 所示。我們將優先權計算方程式定義如下

$$P(p) = C(p)D(p) \quad (3-5)$$

其中信心值 $C(p)$ 與資料值 $D(p)$ 定義如下

$$C(p) = \frac{\sum_{q \in \Psi_p \cap (I-\Omega)} C(q)}{|\Psi_p|} \quad D(p) = \frac{|\nabla I_p^\perp \cdot n_p|}{\alpha} \quad (3-6)$$

其中 $|\Psi_p|$ 代表 Ψ_p 的面積， α 為正規化係數(對一般 256 階灰階影像，其值為 255)， n_p 為 p 點的法向量， $C(q)$ 為 Ψ_p 內非空洞的像素值。

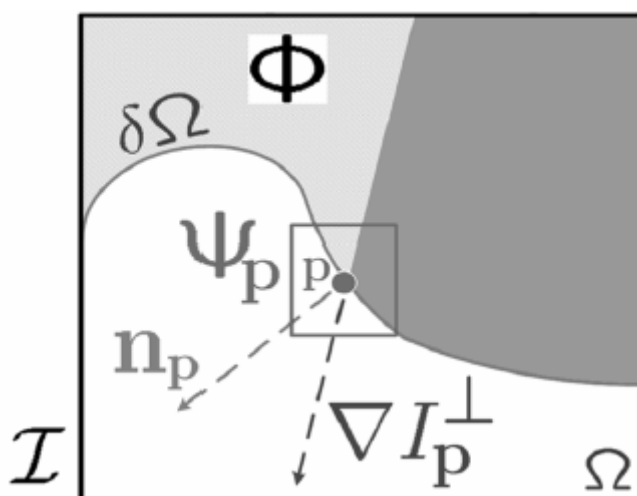


圖 3-6 填補優先權示意圖，任意一目的填補方塊 ψ_p ， n_p 為目的區塊 Ω 輪廓之法向

量，而 ∇I_p^\perp 為 p 點的 isophote 方向，整個影像以 I 來表示

影像填補演算法由於必須不斷的更新填補優先權，所以計算量非常之龐大，且計算方式是屬於變動式的，計算量是跟所輸入的影像有關係，這其實是不適合硬體設計的方法，一個適合硬體設計的方法必須是比較固定式的計算方式。

3.4.3 相關硬體設計

Hong-Ming Wang 提出了一個對於不同紋理背景以子區塊紋理合成與加權內差法來填補移除的區域[21]。子區塊紋理合成可一次填補一排的目標填補區域，而加權內差法可將區塊平滑化。並且分析整個演算法的效能，來設計填補搜尋的積體電路，將其以 pipeline 的方法讀入欲計算的像素，來提升搜尋時的計算速度。整體搜尋的時間如下

$$\text{Cycle number} = [5 + (N - 1)] * [h * k - l * c] * l \quad (3-7)$$

其中的 5 是因為一個計算必須經過五個時脈週期， N 是代表所要比對的點數， $h * k$ 為影像的大小， $l * c$ 為空洞的大小， l 為空洞的行數。其 pipeline 架構的設計方式如下圖 3-7 所示。

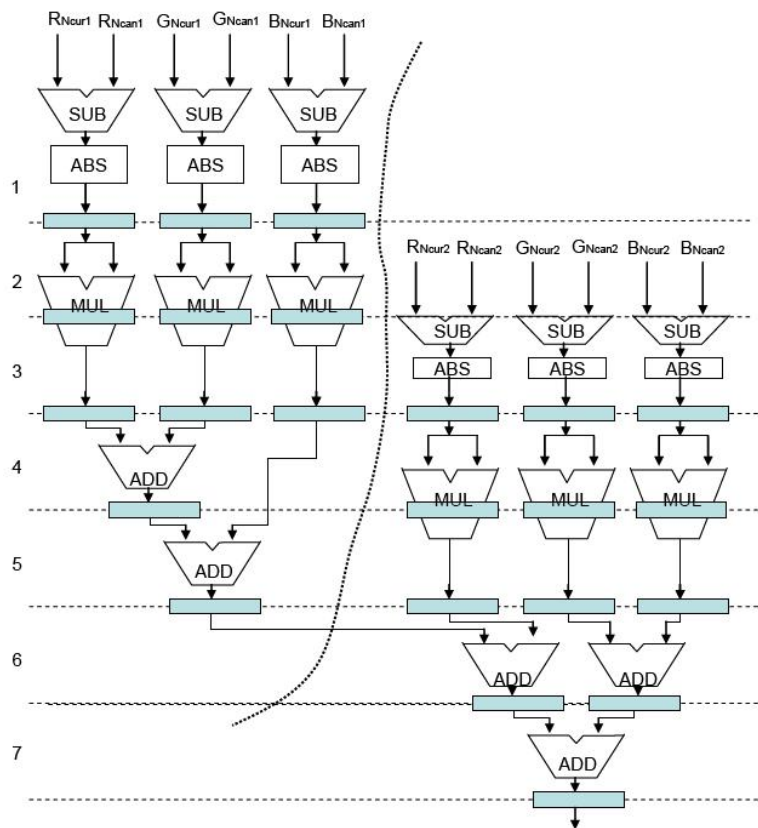


圖 3-7 搜尋區塊的 pipeline 結構圖[21]

Wan-Yu Chen 提出了一個即時 DIBR 的硬體架構[22~23]，首先會對影像深度地圖做 Edge-dependent gaussian filter，讓深度的變化比較連續，來減少空洞的數量。由於使用 Edge-dependent gaussian filter 而讓計算量上升，為了提升運算速度，所以以 pipeline 的方式來設計 Edge-dependent gaussian filter 的計算，以提升計算速度，使整個設計符合 ATTEST(720 x 576 @25fps)的標準，並且以一個 80MHz 的時脈效能就可以達到即時的運算。而其中所使用的影像填補方式，是以空洞旁的像素來向內填補的一個簡單的方式來做填補。其硬體架構圖如下圖 3-8 所示。

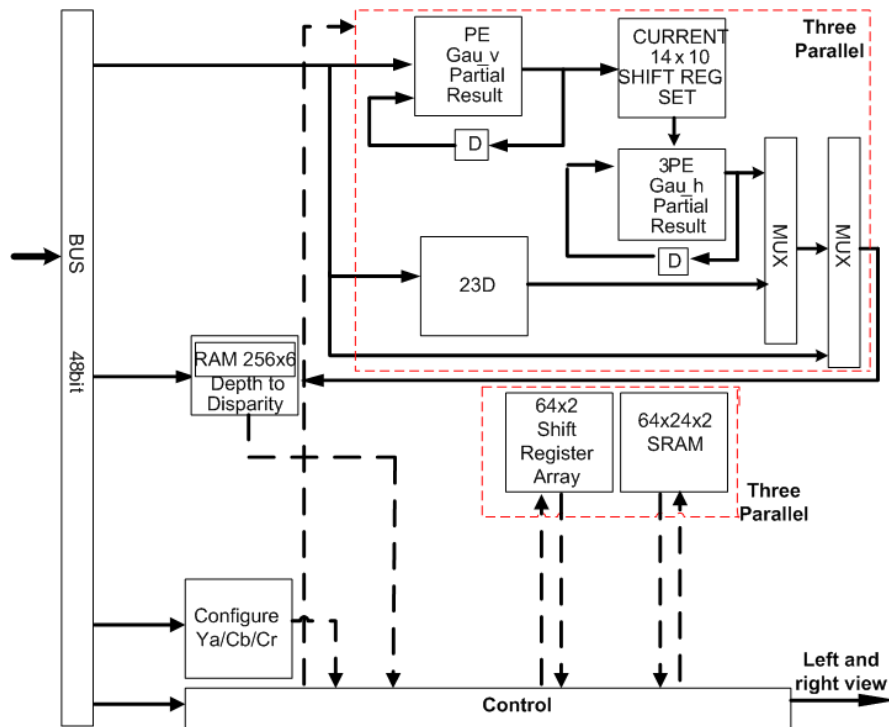


圖 3-8 DIBR 硬體架構圖[23]，圖中 D 為 D 型正反器

4. 研究方法

在 2D 影像中關於深度資訊是非常稀少的，只能從色彩，影像清晰度，邊緣銳利度，物件完整性，物件之間的重疊等少數資訊中獲得大概的深度資訊。

4.1 消失線的介紹

在畫圖時首先應該是先觀察一下實景；打算畫多大的圖像、主要的主體是什麼(圖中所要表現的物件、表現的物體)，接著用最簡單的直線分割成幾個大區塊或是幾何圖形，這幾條簡單的直線就是主軸線，如圖 4-1 中 1.2.3 黃色的線即為主軸線。只要主軸線之間的角度與關係抓住了，這張圖的比例就不會「失真」，如圖 4-1 中的 2 是指圖片中的樹幹。接著再找出次要的軸線，如圖 4-1 中 4.5.6 粉紅色的線為次要的軸線，其目的是輔助主軸線來將圖面內的景物粗略定型，如圖 4-1 中 5.6 兩條線構成山與天，這是畫圖的原理。那麼在照片中也有所謂的主軸線與輔助主軸線，而這些軸線就是我們所謂的消失線。消失線以上面敘述來定義是物體的中心線以及物體邊緣所形成的平行線。

4.2 消失點的介紹

在多數的圖片中一定有貫穿圖面的水平直線，稱之為地平線。當然在某些圖中是看不到地平線的，例如室內(如圖 4-2)。這條虛擬的地平線只是代表繪圖者或者照相者的視高(換句話說即是眼睛觀看的高度)。因為地平線是我們人眼的高度，若先忽略物體大小不一的問題，則物體距離我們越遠基本上是會變的越小最後都會消失在地平線上。

在圖三中除了地平線之外還有另外六條線，這六條線互相交錯且與地平線交錯於一點，而且是在地平線上，這些線就是延伸所有平行於地平面的線，這裡面包括了任何有直線邊緣，稜角，形狀的東西，像是樓房、牆、畫框、鐵道、家具、地面、人、車等等許多東西。由這些所構成的線都會與平行線交錯且會交於同一點上，圖 4-2 中的這六條線是由下列三對線條;1.延伸虛擬的天花板／牆面交界之線，2.畫框頂緣 3.牆面／地面交界之線所構成，且又與平行線交錯於同一點上，因此交錯的點稱為消失點，此點為圖片中最遠的一點。

圖 4-2 是虛擬的室內示意圖[7]，在敘述若不考慮物體大小時，假設物體大小一樣時，越近觀看者，觀看者所看到的物體則越大，反之則越小。圖中所畫的線是消失線，雖說實際上我們可尋找的到的，可是無法像圖中這麼準確。

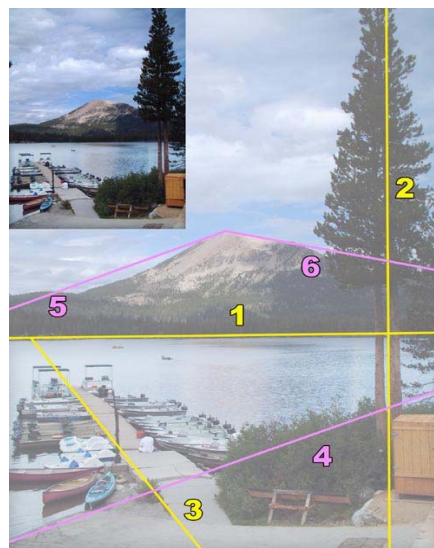


圖 4-1，繪圖中主軸線與消失線示意圖

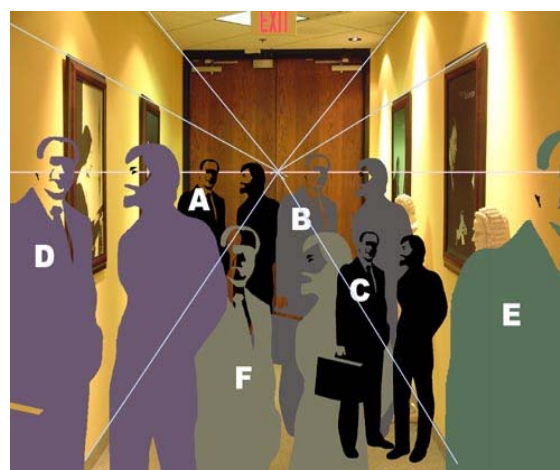


圖 4-2，室內消失線與消失點示意圖

4.3 尋找消失線與消失點

在上節中得知欲找消失點得先尋找消失線，而消失線與圖中物體的邊緣有關係，主要分為四個步驟。第一步驟：首先我們將原圖先使用 3×3 Sobel masks 計算出水平與垂直的邊緣影像，再將原圖中的邊緣尋找出來。第二步驟：有了邊緣資訊圖之後，我們接著要選擇主要的消失線，我們利用霍夫轉換（公式(4-1)）先計算 ρ 值。公式(4-1)是極座標系之直線方程式，其中 ρ 表與原點之距離， θ 表與 X 軸之交角。第三步驟：我們擁有 ρ 和 θ 值之後，累加相同的 ρ 和 θ （公式(4-2)，其中 Acc 表累加值）。第四步驟：在眾多的消失線中，有些是重要資訊有些是不需要的，因此我們選擇累加值前十名者作為重要數據，利用公式(4-1)反算 X、Y 座標所形成的線即為消失線。

從上一節之說明我們得知消失線是由物體的邊緣，形狀等所計算出來的。因此邊緣圖中表現邊緣的粗細會影響到累加值，且 ρ 經過公式(4-1)的計算原本應為實數值，但為了計算量以及 X、Y 座標為整數等問題，我們取 ρ 為整數，因此會使得原本累加值推算回去的消失線應為 10 條，卻因為這些因素使得原本一條的消失線變為好多條。可能互相平行或交錯接近平行，使得原本交錯點最多的消失點卻因為這些因素而使得可能交錯點最多的不是消失點。為了解決這些問題，因此我們使用了 11×11 大小的範圍來做累加。如此即使消失線互相平行或交錯接近平行也會朝向消失點的方向聚集。因此我們用 11×11 的範圍做累加，兩條消失線交於一點記為 1，三條消失線交於一點記為 2，依此推算累加，在 11×11 的範圍裡總值累加為最多的中心點即為消失點。此方法雖然無法很正確的尋找到消失點但不會差異太大，在圖片上可能的差異為 4、5 個像素。

$$\rho = x \cos \theta + y \sin \theta \quad -90^\circ \leq \theta \leq 90^\circ \quad (4-1)$$

$$Acc[\rho, \theta] = Acc[\rho, \theta] + 1 \quad (4-2)$$



圖 4-3 (a) 消失線及消失點實驗範



圖 4-3 (b) 消失線及消失點實驗範例

圖 4-3 是實際所拍攝的相片，(a)為白天所拍攝而(b)為夜晚所拍攝，其中黑色線為本計畫方法所找到的消失線，紅點則是所搜尋到的消失點。

4.4 建立深度漸層圖

再尋找到消失點後，我們可以確定此點為圖中最遠的一點。我們利用消失線的交錯來尋找消失點，而消失線為直線交錯點可能位於圖片的範圍之外，因此消失點可能在圖

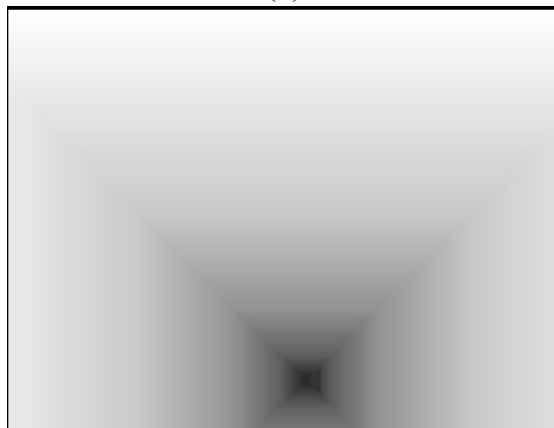
中也可能在圖之外。一般來說知道消失點後可以大概確定整張圖的深度走向，若消失點在圖片的右方那麼可以大概得知圖的左邊比較近圖的右邊比較遠，深度的走向由左向右漸遠，因此我們可以依照消失點的方向做出深度漸層圖。我們將深度漸層圖以灰階圖來表示，越白表示越近越黑表示越遠。若灰階值為 256 階，且消失點在圖右邊，則圖中最右邊的灰階值為 0 向左遞增，每向左移動一個 Pixel 點則增加 $255/Width$ ，其中 255 為灰階最大值，Width 為圖片的寬度，所以漸層圖會成為長條狀的漸層圖。圖 4-4 是由本方法所估測出的深度漸層圖及與參考文獻[7]之比較。



(a)



(b)



(c)

圖 4-4 (a)參考文獻(7)測試圖片 (b)採用本計畫方法所做出的深度漸層圖 (c) 採用參考文獻(7)所做出的深度漸層圖

4.5 估測深度圖

深度的漸層圖只是圖片中深度大概的走向，並非是完全正確的深度值。物體有分前後，而背景在圖中是從左到右貫穿圖中的。我們提出使用顏色作區塊的判別，利用連通的原理配合區塊顏色，然後依照消失點的方向做判斷。若消失點在圖片右方，我們就先從右方開始做判斷。通常物體是位於圖片的中央，而背景貫穿整張圖片且是最遠的部分，因此要從消失點所在的方向開始做判斷。首先利用连通原理作以顏色為指標做區塊判斷，我們將 R、G、B 三顏色先平方相加之後再開根號，若小於某臨界值則判斷為相同的 Pixel。整體之判斷可分為六個步驟，說明如下：

第一步先做邊緣的判斷：

原因在於先確定圖片邊緣正確的深度值，若超過消失點 Y 座標則深度圖遞增。在上節中提到是因為消失點基本上是位於水平線上，而水平線以上的通常為背景部分屬於遠區塊的，而水平線以下會漸近，如地板、桌子等。

第二步以向上、向右上、向右顏色的區塊做判斷：

低於臨界值且又是最小值那麼認定為相同區塊內的 Pixel，所以給予相同的深度值，若超過消失點的 Y 座標一樣做遞增的動作。

第三步是做左右區塊的比對。

背景會被物體切為左右兩邊，因此要將圖中左方部分與右方部分做區塊的比對。消失點在右邊則要左邊區域與右邊比對，而考慮到可能背景為斜的，因此本計畫尋找 Y 座標範圍為 $Y-20 \sim Y+20$ ，X 座標則為從最右邊開始照片的寬度 $\sim X+1$ 。此步驟可以將左右區塊做部分的判斷。

第四步是從左邊開始做连通原理

因為有些部分區塊的深度值錯誤的，在做過左右的比對後可將左邊與右邊相同的區塊給予相同的深度值。而某些點部分會錯誤因此從左邊開始做连通原理可將這些錯誤的部分給予修正。當第三步驟做完左右比對後，因為某些點受亮度或邊緣的影響使得原為相同區塊內的 Pixel 判斷為不同點時，這時與第二步驟的比對方法相同只是以左邊開始做连通原理的判斷讓這些錯誤的點補正回來。

第五步是從下而上做连通原理

一般照片中下方的區塊為地板桌子之類的物體，也就是說水平線下方的區塊是屬於漸近的區塊，我們所做的漸層的深度圖中給予的深度可能為相同。因此這部分要由下往上比對，而我們比對的方法是用不同的色彩空間，這部分是使用 cr、cb 的值來比對。比對方式也是 cr、cb 的平方相加開根號小於臨界值則視為相同區塊內的 Pixel。這部分之所以要用不同的色彩空間是因為光線會影響到色彩，而我們用 YCrCb 的色彩空間，去除 Y 的值（即是亮度）後即不會受到亮度的影響。之前比對的部分之所以不使用 YCrCb 空間是因為在某些顏色下會判斷為相同的，例如透明的瓶子與白色的桌子會判斷為相同。而我們所使用的连通原理只跟周圍相鄰的 Pixel 相比，所以亮度的影響便減少許多。因此之前使用 R、G、B 比對效果比較佳，而這部分要連續性的與大區域比對因此得考慮亮度的因素。比對的方法是由最下的點開始往上比對，比對只到水平線，在比對為相同區塊內給予深度的漸層。

第六步是做 7x7 大小的平滑化(Smoothing)

在某些點因為亮度的因素使得之前所進行的區塊判斷錯誤，給予錯誤的深度值。為了減少這部分的錯誤因此做平滑化。平滑化可以使區塊內的深度值均化，使得給予錯誤深度值的點可以減少誤差值，而有正確深度的點不會因為平滑化後而錯誤。

4.6 硬體設計及架構

4.6.1 整體架構

整個系統簡單的架構是利用深度地圖由原始影像產生雙眼立體影像，接著再各自做影像的填補，填補完畢後將雙眼立體影輸出。系統架構圖如圖 4-5 所示。

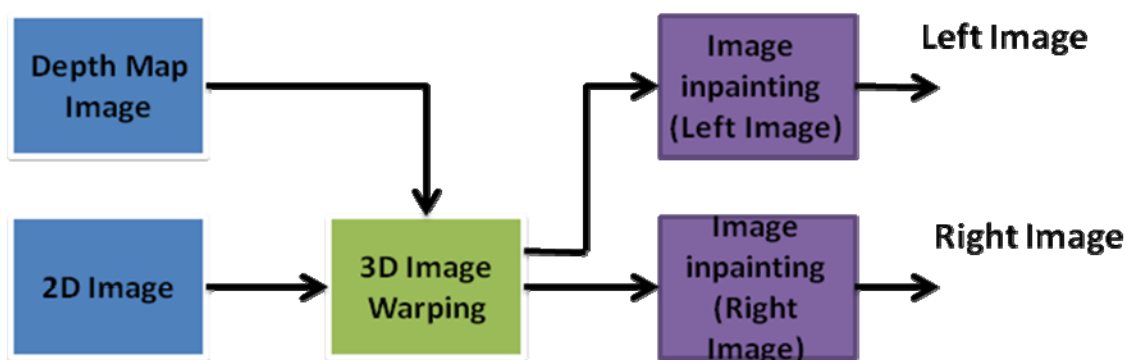


圖 4-5 系統架構圖

我們整個系統的流程是，首先利用 DIBR 的方法產生出雙眼立體影像，接著利用 3 x 3 的 mask 來填補因為深度連續變化時所產生的細小空洞區域，填補完細小的空洞區域後，再計算影像的梯度方向，以便接下來的影像填補法來使用，接著再搜尋空洞邊緣的參考點，找到邊緣參考點後再以邊緣參考點為顏色基礎，以方向性搜尋顏色最接近的填補參考點，再以填補參考點的方向性的點做為填補的點填補回邊緣點旁的空洞內，重覆搜尋與填補的動作直到沒有空洞為止，最後將填補完畢的影像輸出，流程圖如圖 4-6 所示。我們會在接下來的小節內分別介紹每個步驟的電路設計方式與電路動作流程。

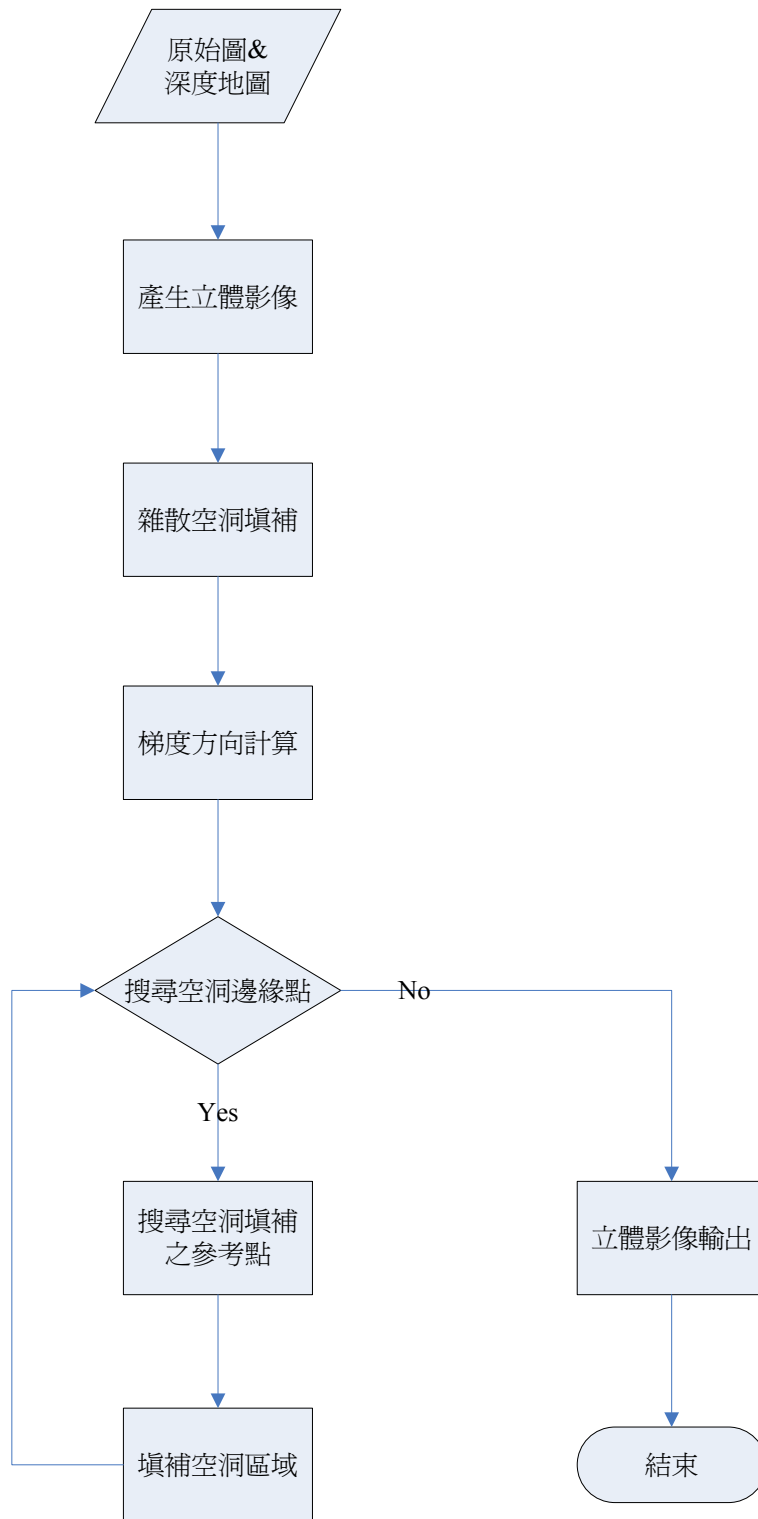


圖 4-6 系統流程圖

4.6.2 雙眼立體影像產生

我們使用 Shift-sensor 立體攝影機設定方式，將方程式(3-1)做整理化簡，可以得到下面的方程式(4-3)：

$$\text{Left} : u + \left(\frac{ft_c}{2Z_c} - \frac{ft_c}{2Z} \right) \quad \text{Right} : u - \left(\frac{ft_c}{2Z_c} - \frac{ft_c}{2Z} \right) \quad (4-3)$$

我們以方程式(4-3)來設計 3D 影像位移產生器(圖 4-7 中的 3D image warping)的功能。此區塊電路動作的流程是將位址產生器(圖 4-7 中的 Address Generator)所產生的位址輸入給 3D 影像位移產生器，3D 影像位移產生器將所輸入的位址加上計算的位移量，接著將輸入的像素儲存到對應的記憶體位址內。此區塊電路如圖 4-7 所示。

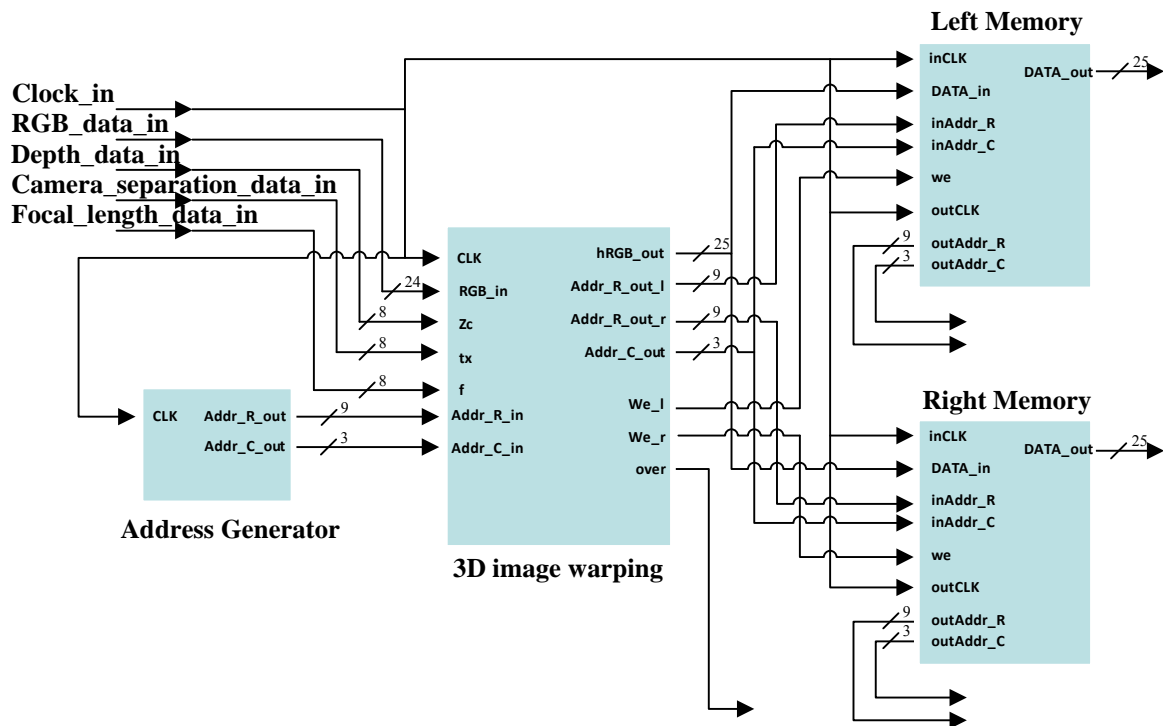


圖 4-7 立體影像產生區塊電路圖

5. 結果與討論

5.1 拍攝條件限制

本計畫中所拍攝的影像資料必須注意以下幾點;第一圖片必須要有個方向性。因為我們估測深度圖前必須要尋找到消失線與消失點，消失點是由消失線所尋找的，消失線則是由圖片中物體的邊緣、主軸等所計算出來的。因此物體的擺設以及背景拍攝的角度都有關係。整體上來說拍攝方面盡量必須稍微有個角度拍攝。第二是圖片不能太複雜，尤其是背景的部分。因為本計畫是使用顏色來判斷區塊，因此區塊內顏色不能差異太大或者顏色混雜。第三是物體或背景必須有一定的排列，物體不能重疊，否則有些深度值會計算錯誤。

5.2 實驗結果

5.2.1 深度估測

我們以實際拍攝在不同場景的照片去進行實驗，並求出其深度圖，結果如圖5-1所示。我們以實際照片作實驗，我們去觀察物體，背景，以及環境等三因素的影響，發現物體影響所產生的估測錯誤遠小於背景以及環境，而背景中雖說有影響但遠小於環境因素。主要是因為環境因素所造成圖片中的差異性較大。光以亮度來說，不同的亮度使得判斷錯誤率大增許多。即使亮度相同，但如背景貫穿全圖，將使得左右邊背景在判斷上略有錯誤，因此未來還是有很大的改善空間。



圖5-1 (a) 原圖以及所估測的深度圖



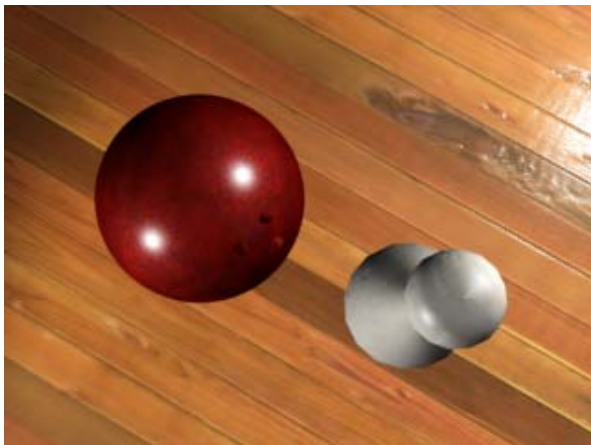
圖5-1 (b) 原圖以及所估測的深度圖

5.2.2 影像填補

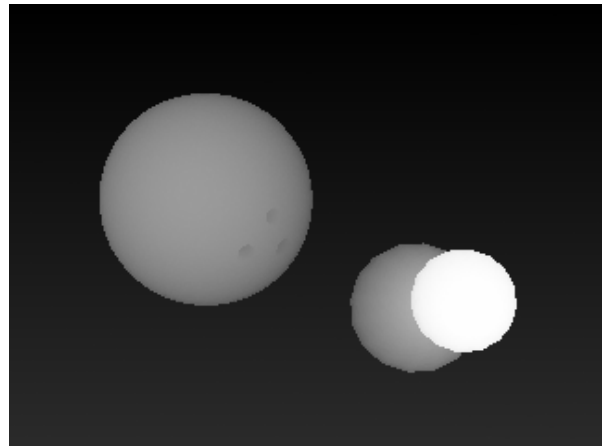
我們以六種不同背景類型的影像做結果的分析，分為人工背景影像、天空背景影像、棋盤背景影像、前景物件互相遮蓋影像、室內照片影像、以及戶外自然風景。

人工背景的背景影像

來源影像為保齡球及保齡球瓶放置於木製球道上，球道上各木片的顏色及紋路皆不同。我們可以看到填補結果的顏色與紋路大致都是正確的，但是邊圍的木板紋理會有一點偏移的狀況，這是由於我們將填補演算法化簡的關係。



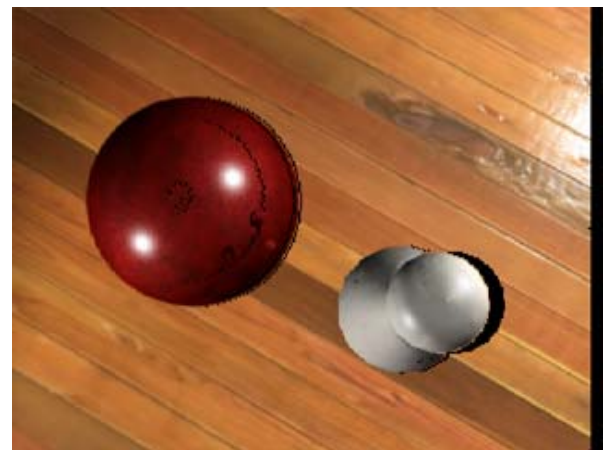
(a)



(b)



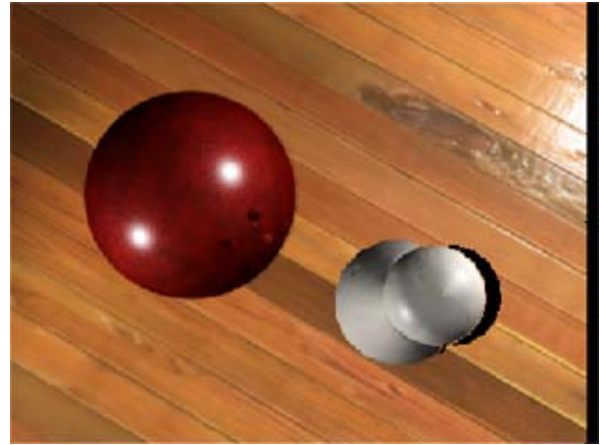
(c)



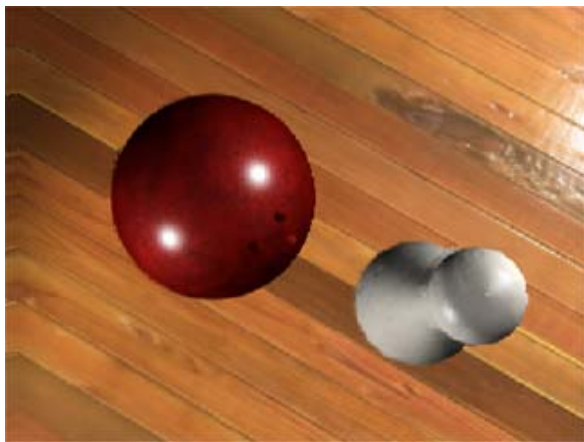
(d)



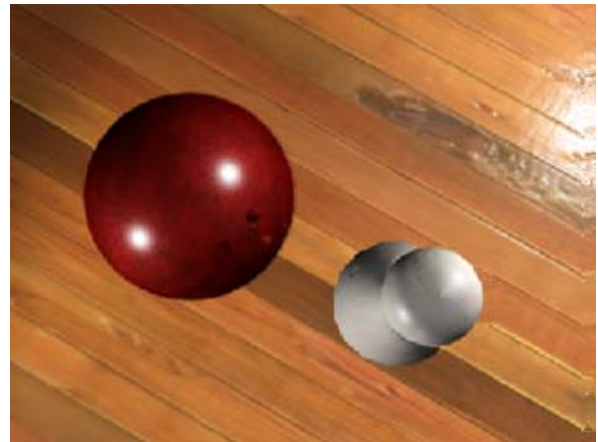
(e)



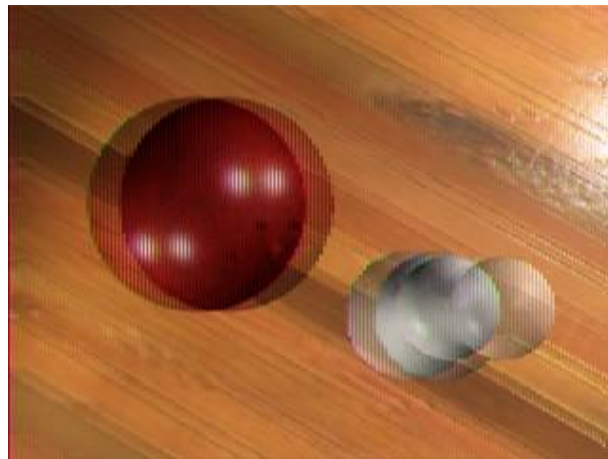
(f)



(g)



(h)



(i)

圖 5-2 人工背景の影像結果圖(a)原始影像(b)深度圖(c)未做任何填補の左眼圖(d) 未做任何填補の右眼圖(e)含有空洞區域の左眼圖(f)含有空洞區域の右眼圖(g)完成填補之左眼圖(h)完成填補之右眼圖(i)左右眼插排圖

天空背景的影響

我們選擇了一個生物飛行於天空中的影像，背景是屬於自然不規則的影像。我們可以明顯的看出背景的填補較人工背景好，這是因為背景是屬於不規則的背景，所以填補優先順序的影響較小。我們可以看到翅膀附近以及兩腳中間處有些地方填補不完美，這是因為翅膀附近的前景區域非常的窄造成這種填補錯誤的現象，以及兩腳中間區域的背景參考點太少，以至於填補錯誤。



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)



(i)

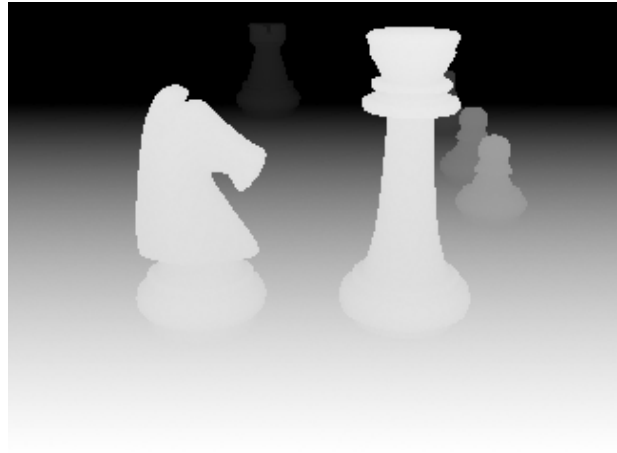
圖 5-3 自然背景の影像結果圖(a)原始影像(b)深度圖(c)未做任何填補の左眼圖(d) 未做任何填補の右眼圖(e)含有空洞區域の左眼圖(f)含有空洞區域の右眼圖(g)完成填補之左眼圖(h)完成填補之右眼圖(i)左右眼插排圖

棋盤式背景的影響

我們選擇了棋盤背景的影響，棋盤影像是由深色及淺色的方塊所組成。我們可以看出填補還是有偏移的情形。



(a)



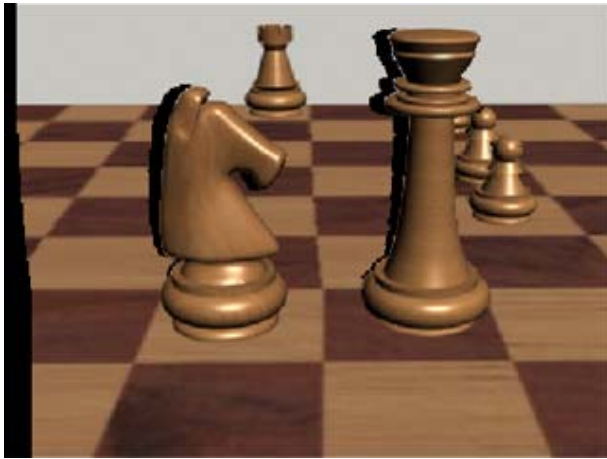
(b)



(c)



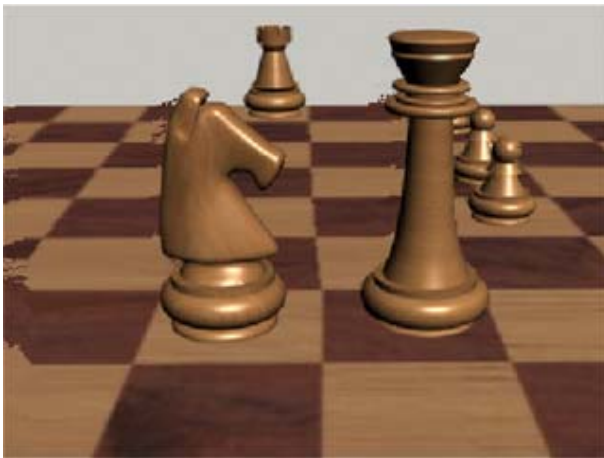
(d)



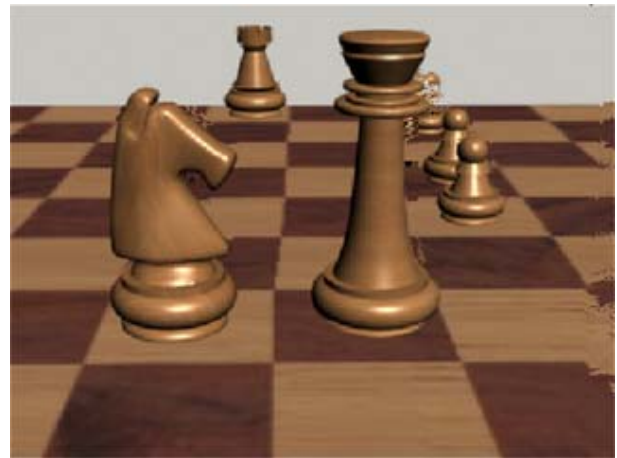
(e)



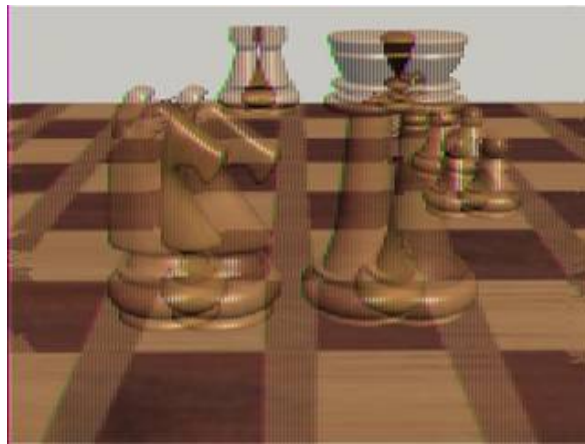
(f)



(g)



(h)



(i)

圖 5-4 棋盤式背景的影像結果圖(a)原始影像(b)深度圖(c)未做任何填補的左眼圖(d)未做任何填補的右眼圖(e)含有空洞區域的左眼圖(f)含有空洞區域的右眼圖(g)完成填補之左眼圖(h)完成填補之右眼圖(i)左右眼插排圖

前景物件相互遮蓋的影像

我們選擇了一個有蘋果和柳橙重疊的影像，蘋果旁的空洞區域包含了背景及柳橙。我們可以看出紋裡還是有一點點的偏移，以及柳橙的結果沒有那麼圓，但基本上顏色都是正確的。



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)



(i)

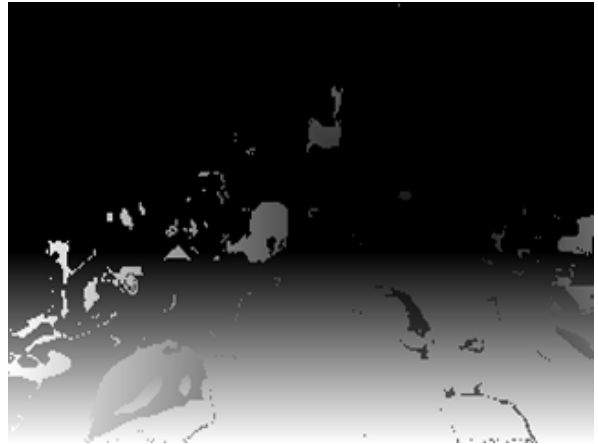
圖 5-5 前景物件相互遮蓋的影像結果圖(a)原始影像(b)深度圖(c)未做任何填補的左眼圖(d)未做任何填補的右眼圖(e)含有空洞區域的左眼圖(f)含有空洞區域的右眼圖(g)完成填補之左眼圖(h)完成填補之右眼圖(i)左右眼插排圖

室內照片的影像

我們以一個實際拍攝的室內照片為實驗影像。我們可以看出只有邊緣的填補結果有一點偏移，但整體的效果是很好的。



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)



(i)

圖 5-6 室內照片的影像結果圖(a)原始影像(b)深度圖(c)未做任何填補的左眼圖(d) 未做任何填補的右眼圖(e)含有空洞區域的左眼圖(f)含有空洞區域的右眼圖(g)完成填補之左眼圖(h)完成填補之右眼圖(i)左右眼插排圖

戶外自然風景照片的影像

我們以一個戶外自然風景照片為實驗影像。我們可以看出只有邊緣的填補結果有一點偏移，尤其是欄杆的部分特別明顯，但整體的效果是很好的。



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)



(i)

圖 5-7 戶外自然風景照片的影像結果圖(a)原始影像(b)深度圖(c)未做任何填補的左眼圖(d)未做任何填補的右眼圖(e)含有空洞區域的左眼圖(f)含有空洞區域的右眼圖(g)完成填補之左眼圖(h)完成填補之右眼圖(i)左右眼插排圖

5.3 結論與建議

在本計畫所做的主體是以單張照片為主。以單張的2D影像來說可獲得相關的深度資訊是非常稀少的，而只能從單張圖片中搜尋有關深度的資訊來估測相對的深度值。在本計畫中所做的只是初步去估測單張圖片的深度值，尚有許多地方待改進。

例如本計畫所估測的深度圖是在相同物件中給予相同深度值。而基本上物體有輪廓以及背景拍攝有角度時應是一邊較近而另一邊則較遠，而本計畫中都是給予相同的深度值。另環境因素會影響到判斷，以及物體的陰影所造成的判斷錯誤等地方仍需加以改善。大多估測深度的研究都是以獲得不同照片或獲取相機等設定的參數去進行研究，因此本計畫的重點是在於如何從單張影像中估測其深度值。

所以所估測的深度圖只是初步的結果，而在實驗結果部分配合本實驗室另一位學長的研究中利用原圖以及深度圖做出的雙眼影像配合立體顯示器(Sharp 的)所呈現的立體影像效果非常好。當然還是有許多缺點得改進，而主要改善的地方我們大概可以分為幾點。第一判斷區塊的正確性應予提高，我們所做的實驗室皆以單純的背景做實驗，因此在背景複雜時估測錯誤便會增加許多，這是因為在區塊判斷上面判斷錯誤，因此區塊判斷的正確性可以提高深度的準確性。第二必須去除陰影，物體在光線照射下一定會有陰影的部分，而在陰影的部分中會誤認為是不同的區塊，因此去除陰影是其中一項可以改進的工作。第三本計畫中給予同一物件或區塊是相同的深度值，因此在判斷上面得能分辨物體上是同一平面或是有深淺應給予不同的深度值，以及背景是平面或是斜面。這部分可用清晰度、銳利度等方法來加以改進。第四點是在於環境的許多因素，如最常見的光線照射以及亮度等問題，這部分影響最大，因此未來最主要的工作也在於此。第五照片中清晰與模糊的部分判斷。在照片中清晰的地方大多為主體部分，而背景大多為模糊部分，因此在模糊部分的判斷上面就更加困難。而清晰的部分又得區分其深淺。

以上五點是大略總結歸於本計畫未來必須改進的地方度的領域。立體影像是未來的趨勢，因此如何利用大眾中多數擁有的單眼拍攝器材去估測深度必定是未來主要研究的課題之一。由於立體影像的產生使人們之間不但可從聲音獲得對方的訊息更可藉由立體影像增進與對方的互動，而科幻電影中不再是科幻而是實際的生活了。

參考文獻

- [1] 李獻仁，”立體照片原理與製作(Lenticular 立體照片)”。
- [2] Christophe Simon, Frederique Bicking and Thierry Simon, 2002, ” Estimation of depth on thick edges from sharp and blurred images”, IEEE Instrumentation and Measurement Technology Conference, Volume 1 pp.323- 328.
- [3] Cassandra Swain, Alan Peters and Kazuhiko Kawamura, 1994, ” Depth Estimation from Image Defocus using Fuzzy Logic”, Proceedings of the 3rd IEEE International Conference on Fuzzy System, Volume 1, P9,94-99, Orlando, USA.
- [4] F. Deschen, D. Ziou, P. Fuchs, 2002, ”Homotopy-Based Estimation of Depth Cues in Spatial Domain”, IEEE International Conference on Pattern Recognition, Vol.3 pp.627-630.
- [5] Satoko OHTSUKA, Shinya SAIDA, 1994, ” Depth Perception from Motion Parallax in the Peripheral Vision”, IEEE international Workshop on Robot and Human Communication, pp.72-77.
- [6] F. Deschênes, D. Ziou, 2003, ”Homotopy-Based Computation of Defocus Blur and Affine Transform”, in Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03), Vol.1, pp.I-398-404.
- [7] S. Battiatoo, A. Caprab, S. Curtib, M. La Casciac, 2004, ”3D Stereoscopic Image Pairs by Depth-Map Generation”, in Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04), pp.124-131.
- [8] Yamanoue, H., “The Differences Between Toed-in Camera Configurations and Parallel Camera Configurations in Shooting Stereoscopic Images”, Multimedia and Expo, 2006 IEEE International Conference, page(s): 1701-1704, 9-12 July 2006.
- [9] Yamanoue Hirokazu, Nagayama Masaru , Bitou Mineo , Tanada Jun, “**Subjective study on the orthostereoscopic conditions for 3D-HDTV**”, 映像情報メディア学会技術報告, 21(63), pp.7-12, 1997.
- [10] Hoonjong Kang, Namho Hur, Seunghyun Lee, Hiroshi Yoshikawa, “**Horizontal parallax distortion in toed-in camera with wide-angle lens for mobile device**”, [Optics Communications](#), [Volume 281](#), [Issue 6](#), Pages 1430-1437, 15 March 2008.
- [11] William and Craig pp. 115 – 164 “**Seeing 3D from 2D Images**”.
- [12] Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut (HHI), “**Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV**”, Stereoscopic Displays and Virtual Reality Systems XI, Proceedings of the SPIE, Volume 5291, pp. 93-104 (2004).

- [13] C. Fehn. “**A 3D-TV Approach Using Depth-Image-Based Rendering (DIBR)**” In Proceedings of 3rd IASTED Conference on Visualization, Imaging, and Image Processing , pp. 482-487, Benalmádena, Spain, Sep. 2003.
- [14] Marcelo Bertalmio, [Guillermo Sapiro](#), Vincent Caselles, Coloma Ballester, “**Image inpainting**”, International Conference on Computer Graphics and Interactive Techniques, Proceedings of the 27th annual conference on Computer graphics and interactive techniques, Pages: 417- 424, 2000.
- [15] Bertalmio, M., Bertozzi, A.L., Sapiro, G., “**Navier-stokes, fluid dynamics, and image and video inpainting**”, Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference, Volume 1, page(s): I-355-I-362 vol.1, 2001.
- [16] Bertalmio, M., Vese, L., Sapiro, G., Osher, S., “Simultaneous structure and texture image inpainting”, Image Processing, IEEE Transactions, Volume 12, Issue 8, page(s): 882- 889, Aug. 2003.
- [17] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama, “**Region filling and object removal by exemplar-based image inpainting**”, Image Processing, IEEE Transactions on, Volume 13, Issue 9, page(s): 1200-1212, Sept. 2004.
- [18] BianRu Li, Yue Qi, XuKun Shen, “**An image inpainting method**”, Computer Aided Design and Computer Graphics, 2005. Ninth International Conference, page(s): 6 pp.-, 7-10 Dec. 2005.
- [19] Shantanu D. Rane, Guillermo Sapiro, and Marcelo Bertalmio , “**Structure and Texture Filling-In of Missing Structure and Texture Filling-In of Missing Compression Applications**”, Image Processing, IEEE Transactions on Volume 12, Issue 3, page(s): 296- 303, March 2003.
- [20] Tauber, Z., Ze-Nian Li, Drew, M.S., “**Review and Preview: Disocclusion by Inpainting for Image-Based Rendering**”, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, Volume 37, Issue 4, page(s): 527-540, July 2007.
- [21] Hong-Ming Wang, Jhing-Fa Wang, “**Object Removal Algorithm Hardware in Image Processing**”, 國立成功大學 電機工程學系碩士論文 2004.
- [22] Wan-Yu Chen, Yu-Lin Chang, Shyh-Feng Lin, Li-Fu Ding, and Liang-Gee Chen, “**Efficient Depth Image Based Rendering with Edge Dependent Depth Filter and Interpolation**”, Multimedia and Expo ,2005.ICME 2005, IEEE International Conference on July 2005
- [23] Wan-Yu Chen, Yu-Lin Chang, Hsu-Kuang Chiu, Shao-Yi Chien, and Liang-Gee Chen,

“Real-Time Depth Image Based Rendering Hardware Accelerator for Advanced Three Dimensional Television System”, Multimedia and Expo,2006 IEEE International Conference on July 2006.

可供推廣之研發成果資料表

■可申請專利

■可技術移轉

日期：98年8月25日

國科會補助計畫	計畫名稱：三維立體深度估測技術開發及其晶片設計 計畫主持人：鄭芳炫 計畫編號：NSC 96-2221-E-216-039-MY2 學門領域：資訊工程
技術/創作名稱	1. 二維影像深度估測技術 2. 影像填補技術及其硬體設計
發明人/創作人	鄭芳炫
技術說明	<p>中文：</p> <ol style="list-style-type: none"> 本研究是從單張影像中尋找消失線與消失點去估測深度，再利用原圖及深度圖計算左右雙眼影像給予立體顯示器使用。 左右雙眼影像所產生之影像空洞填補技術及其硬體架構之晶片設計 <p>英文：</p> <ol style="list-style-type: none"> In order to make stereo image more popular, we focus on how to estimate depth information from single image in this project. We utilize original image and depth map to calculate the binocular image and use 3D LCD to display the stereo image. The concept of vanishing point is used to estimate the depth information. An image inpainting technique with its hardware architecture and chip design is developed to achieve real time application..
可利用之產業 及 可開發之產品	光電產業，顯示器產業，電視產業 立體數位相框，三維立體顯示器，三維立體電視
技術特點	<ol style="list-style-type: none"> 利用所做出的深度圖以及原圖去計算出雙眼影像配合 3D 立體顯示器去顯示立體影像 利用硬體設計達到即時左右雙眼影像空洞填補以完成真實 3D 立體影像顯示

推廣及運用的價值	
----------	--

- ※ 1. 每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送 貴單位研發成果推廣單位（如技術移轉中心）。
- ※ 2. 本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。
- ※ 3. 本表若不敷使用，請自行影印使用。

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

98 年 5 月 27 日

附件三

報告人姓名	鄭芳炫	服務機構及職稱	中華大學資工系教授
時間 會議地點	2009/5/19~2009/5/23 日本橫濱 慶應大學	本會核定 補助文號	NSC96-2221-E-216-039-MY2
會議 名稱	(中文) 2009 年國際圖形識別聯盟機器視覺應用研討會 (英文) 2009 IAPR Conference on Machine Vision Applications		
發表 論文 題目	(中文) 針對新一代視訊編碼的畫框間模式決定演算法 (英文) Inter Mode Decision Algorithm For Advanced Video Coding		

一、參加會議經過

會議的開幕典禮由主辦單位與會議的委員會主席簡單的致歡迎詞後，隨即展開。由於本會議為一個專業之研討會，為了讓與會之學者專家不會錯過任何一個場次之研討，會議只安排一個場地進行。本屆會議由於受到新流感的影響參加人數比上屆少，共有 144 篇論文投稿，經過嚴格的審查後，最後通過 122 篇論文，其中 39 篇安排口頭報告(oral)，而 83 篇為海報展示(poster)。大會安排了三天的會議議程，共分為十五個 section，其中安排了三個場次之專題報告：(1) Large Scale Image Search (2) Focal Stack Photography: High-Performance Photography with a Conventional Camera (3) Integration of Earth Observation Data: Challenge of GEOSS (Global Earth Observation System of Systems)，分別由 Dr. Cordelia Schmid, INRIA, France，Prof. Kyros Kutulakos, University of Toronto, Canada 以及 Prof. Ryosuke Shibasaki, The University of Tokyo, Japan 做精彩的專題報告。本人之論文『Inter Mode Decision Algorithm For Advanced Video Coding』被安排在第一天的下午 section 3 之場次以海報的方式發表，如下圖所示。本次會議尚有台灣之其他論文發表，也大都是以海報的方式發表。經過三天完整會議研討，與會者均有豐富的收穫。



二、與會心得

此次會議之地點在慶應大學的橫濱校區，慶應大學目前仍是日本國內最好的學府之一，共有六個校區，因此參加此次會議充份感受到一流學府之氣息，的確有許多值得學習的地方。本會議之定位是以專業精緻之研討會自許，與一般大雜燴式之大型研討會不同。主要目的是讓與會之學者能真正達到充份的學術交流，而不是走馬看花。三天的會議安排得十分緊湊，每天都是從上午 9:00 至下午 6:00 止。本會議在第二天晚上安排了晚宴，晚宴中除了報告此次會議的相關數據外如註冊人數 198 人，也頒發了過去十屆以來之最佳論文五篇，其中有四篇都是日本的論文，此點不禁讓我好奇此次的最佳論文是如何選出的。畢竟台灣的研究水準也不錯，同時參予的人數一直是僅次於日本，但卻沒有一篇最佳論文被選出，個人感到有些遺珠之憾。

三、考察參觀活動(無是項活動者省略)

本會議的定位是專業精緻之研討會，三天的會議安排得十分緊湊，每天都是從上午 9:00 至下午 6:00 止，因此並無時間做考察參觀活動。

四、建議

每次參加研討會常常會在會場碰到許多台灣去的學者教授，若在出國前就可以互相聯繫一起出席，不僅在費用上可以比較節省，在會議上也可以整合力量為台灣之學術界出聲讓國際能充份了解台灣在學術領域之實力。本次會議共有 29 個國家的研究學者參加，台灣大概有十幾位教授及學生參加，除本校中華大學外，尚有清華大學、中正大學、北科大、虎尾科大、中原大學、大葉大學、亞洲大學等。也許國科會可以在現有之網站上另闢一個出席國際會議之交流園地，讓國內之研究學者可以互通訊息，不僅可以整合大家的力量，也可知道國內在國際學術界之活動能量。

五、攜回資料名稱及內容

本次會議攜回一本紙本的會議論文集，資料名稱為 Proceedings of the IAPR Conference on Machine Vision Applications。另有一片資料光碟為本次會議論文集之光碟版。

六、其他

Inter Mode Decision Algorithm For Advanced Video Coding

Fang-Hsuan Cheng

Department of Computer Science & Information
Engineering, Chunag Hua University
Hsinchu, Taiwan 300
fhcheng@chu.edu.tw

Yea-Shuan Huang

Department of Computer Science & Information
Engineering, Chunag Hua University
Hsinchu, Taiwan 300

Abstract

Variable block size used for inter coding is one of the key technologies in H.264/AVC. When different objects contained in the same macroblock have different motions, smaller block sizes probably achieve better predictions. However, this feature results in extremely high computational complexity when all the block sizes are considered to decide a best one. This paper proposes a new inter mode decision algorithm to reduce the number of inter modes that has to be checked, and then encoding time is reduced. We use the co-located macroblock in previous frame and its neighbors as candidates, and check whether an edge of moving object is crossing the middle of these candidates by using the score given to the modes. The experimental results show that the proposed algorithm is able to reduce 31%-41% total encoding time and about 41%-54% motion estimation time with a negligible PSNR loss of 0.05 dB and bit-rate increment of 2% on the average.

Keyword: Variable Block Size; Motion Estimation; Mode Decision

1. Introduction

Video Compression plays an important role in digital video communication, transmission and storage. H.264/AVC [1-4] is the latest video coding standard developed by the JVT (Joint Video Team) of ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Expert Group (VCEG). While the H.264 belongs to H.26L family of VCEG and the AVC (Advanced Video Coding) belongs to MPEG-4 part 10. This standard has been designed in order to provide higher coding efficiency and network adaptation, which includes a Video Coding Layer (VLC) and a Network Abstraction Layer (NAL). While the VCL represents the video content, and the NAL provides a network-friendly interface.

Comparing to the previous video coding standards, H.264/AVC achieves significant improvement in coding efficiency. This is due to the fact that a number of new techniques are adopted in this standard such as variable block size (VBS) motion estimation, multiple reference frames, quarter-pixel motion estimation, directional prediction of intra coded blocks, in-loop deblocking filter, integer DCT transform and context-based adaptive binary arithmetic coding

(CABAC) etc.. As a result, H.264 can save over half bitrate compared with that of MPEG-2 under the same quality.

Motion estimation (ME) is used as a main method for removing the redundantly information between frames in many video coding standards. H.264, like other video encoders, adopts block-based motion estimation to find a best block matching from a pre-defined search area, and performs variable block size motion estimation to indicate individual motion object in a macroblock. Figure 1 shows the seven block sizes and corresponding mode number/symbol in H.264. We can divide these seven block sizes into two levels which are macroblock level and sub-macroblock level. In macroblock level, there are four inter modes and an additional skip mode (mode 0) which uses the same size with mode 1. If the macroblock is processed in sub-macroblock level, it can be further partitioned into 8x8, 8x4, 4x8 and 4x4 block sizes. The same works will be done in the four sub-macroblocks, and the order of process is from left to right and top to bottom even in mode 2 or 3.

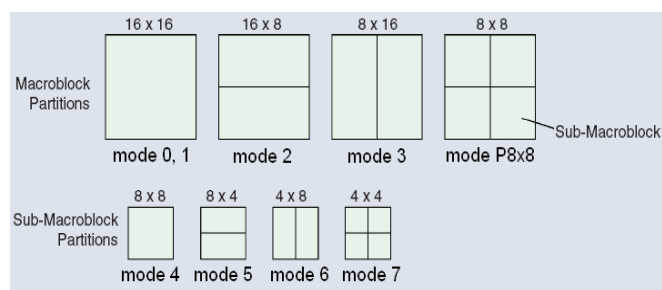


Figure 1. Variable block sizes and corresponding mode number.

According to the new technique described above, H.264/AVC achieves higher coding efficiency than prior video coding standards. However, the large amount of computation makes the encode time extremely increase, thus, it is difficult to be used in practical applications especially in real-time environment. It can be seen that inter modes still take the biggest part of computation. For the reason, we propose a new inter mode decision algorithm to reduce the encoding time with negligible loss of coding efficiency.

The rest of this paper is organized as follows. Section 2 introduces some related works of inter mode decision in H.264. The proposed new inter mode decision algorithm is described in Section 3. The experimental results are shown in Section 4. And a conclusion will be given in Section 5.

2. Mode Decision in Video Coding

There are a number of fast mode decision algorithms proposed for H.264/AVC video coding [5-15] recently. In [7], they select five neighboring macroblocks as candidates by using the concept of 3DRS motion estimation algorithm which is proposed in prior standard. The modes and motion vectors belong to these candidates are used for checking the best prediction in current macroblock. By calculating the rate/distortion cost, the candidate with minimum cost will be chosen as the prediction result.

Reference [6] identifies a macroblock into two types. One is called stationary block and the other is homogeneous block. Stationary block refers to the “stillness” between consecutive frames in the temporal dimension. A region is homogeneous if the textures in the region have similar spatial property. This kind of blocks in the picture would have similar motion and are very seldom split into smaller blocks. The way of determining homogeneous block is to use edge information. An edge map is created for each frame using Sobel operator. Then, they calculate the amplitude of edge by summing up the edges of x and y direction. Finally, they check whether it is a homogeneous block according to the amplitude.

3. Proposed Method

Taking the properties of the seven block sizes into consideration, we can observe that small block sizes usually appear at the edge of motion object. This is due to the size of macroblock is fixed to 16x16 and the context may not just belong to one object, for instance, an object and the background are included in one macroblock. Furthermore, if the motions of objects are more complex and half partitioning (mode 2 or mode 3) a macroblock is not enough, small block sizes will achieve better prediction by partitioning the macroblock into four 8x8 sub-macroblock.

Similar to the idea of the method [7] described in Section 2, a region with complex textures and the motion is not stillness, should be partitioned into small block size. That is, a macroblock, which belongs to the edge of a moving object, is more likely to be partitioned using small block size. Although the mode may be chosen to mode 2 or mode 3 when the edges of object just fall into one side of the macroblock (i.e. one of the two blocks in mode 2 or mode 3), it is not often occurred because the shape of objects belonged to the nature is usually irregular. It means that if a macroblock can be identified that it contains the edge of a moving object, we can determine that this macroblock should be partitioned into small block size (i.e. P8x8 mode).

According to the property of a common video sequence, there is less change between two frames. The co-located macroblock in previous frame has the highest correlation, so it can be used to predict the mode for current macroblock. For the enhancement of mode prediction, not only the co-located macroblock but also the neighbors of that are taken as the candidates. Referring to figure 2, the macroblock named ‘C’ is the co-located one in previous frame, and the eight neighbors surrounded C are also chosen as candidates. For convenience, we assign number 1 to 9 for each of the nine candidates from top left to bottom right. In order to find the edge, we use the four patterns to check whether the ‘C’ is at the edge of moving

object. The four patterns are shown in figure 2. We collect the three macroblocks, which belong to each pattern, as a MB set, and the middle of them definitely are ‘C’. That is, we can obtain four MB sets, and each one indicates its own direction.

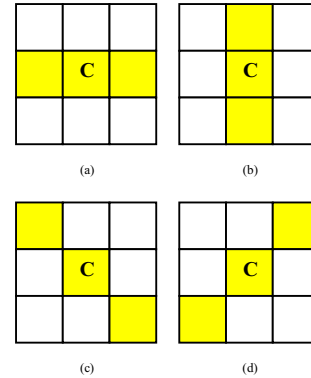


Figure 2. The four directional MB sets.

Considering (a) in figure 2, macroblock ‘4’, ‘5’ and ‘6’ are collected to a MB set in horizontal direction. It means that we can check whether ‘C’ is crossing a horizontal edge of moving object by using this MB set. If so, current macroblock is probably at the edge. (b), (c) and (d) have the same meanings in vertical, diagonal from top left to bottom right and diagonal from top right to bottom left respectively.

To find out which MB set may be the edge, we assign a score for each mode first. Then, sum up the total value by using the score of modes corresponding to the three neighbors in each MB set and put it into a variable called “center_value”. Each MB set has its own center_value. Finally, choose the maximum one from the four center_value and check whether the center_value is greater than a threshold. If so, it can be thought as an edge of object, and smaller block sizes such as P8x8 mode is used for the current macroblock. At this time, large block sizes such as mode 1, 2 and 3 are disabled. Table I shows the scores corresponding to each mode. And we can observe that more P8x8 macroblock appear in the neighbors, the probability of choosing the current macroblock to P8x8 is higher. Furthermore, if P8x8 neighbors concentrate in one of the four MB sets, there must be one moving edge crossing this direction.

On the other hand, if the maximum center_value is small enough, that is, all the neighbors contain large block sizes especially skip mode, it can be thought as the background and more modes can be disabled. Another threshold is also used for this checking.

Table 1. Modes and Corresponding Scores

mode	score
0	0
1	1
2	2
3	2
P8x8	4
9 (intra)	4
10 (intra)	4

The two thresholds are called *upper thd* and *lower thd*. The former is used to divide all the inter mode into large block sizes and small block size, which the small one is P8x8

mode and the large ones are 16x16, 16x8 and 8x16. The latter is used to divide the 16x16 mode from the three ones. The algorithm is shown below:

```

if center_value > upper_thd
    disable mode 1, 2 and 3
else if center_value < lower_thd
    disable mode 2, 3 and P8x8
    else
        disable mode P8x8

```

According to experimental results, the upper value is set to 8 and the lower value is set to 2 will obtain the best performance. It means that there must be more than two P8x8 macroblock in the MB set which has the maximum center value, and there must be at most one 16x16 macroblock in the MB set respectively.

Here we discuss all the possible combinations of the three neighbors that belong to the same MB set. If the center_value is greater than upper_thd, there are three situations which are 9, 10 and 12. Note that the value of center_value is impossible to be 11 under this score of modes. When 9, there must be two P8x8 (or Intra) macroblocks and one 16x16 macroblock. When 10, there must be two P8x8 (or Intra) and one 16x8 or 8x16, because the scores of mode 2 or mode 3 are the same. When 12, there must be three P8x8 (or Intra). On the other hand, if the center_value is smaller than lower_thd, there are two situations which are 0 and 1. When 0, it means that all the three neighbors must be skip mode. When 1, only one 16x16 macroblock in the MB set is allowed.

4. Experimental Results

The proposed mode decision method is implemented in H.264 reference software JM12.2 [16]. We select seven QCIF sequences which are commonly used for video compression testing in our experiments. The form of QCIF is 176 pixels in width and 144 pixels in height, that is, there are eleven macroblocks in width and nine macroblocks in height, and totally ninety nine macroblocks contained in one frame. The first 100 frames of every sequence are tested by our method. The GOP (Group of Picture) structure is IPPP, i.e. the first picture is coded as I-picture and remaining pictures are coded as P-pictures. The Full Search algorithm is used in all experiments with a ± 16 pixels search window.

Firstly, total encoding time and total motion estimation time derived from the JM reference software for each sequence are listed in Table II. The total encoding time is counted all the needed time from the first frame to the last one for encoding, and the total motion estimation time is just only counted the time performing motion search in P-frame. The main difference of them is the time performing the reconstruction of macroblock for mode decision which includes inter modes and intra modes. From the table, we can observe that the time of motion estimation is about 50%-60% of total encoding time and not the 80% as we known in the prior standard. This may be caused by the amount of intra modes in H.264 and the improvement of Full Search.

In the second experiment, we inspect the PSNR change, bit rate change in percentage and time change in percentage. The experimental results are shown in Table III. The positive values mean increments whereas negative values mean decrements. In the time change, we list both total encoding time and motion estimation time. For the comparison, we also implement the Wu's method [6] and list the results in Table III. These changes are calculated by the following equations.

$$\Delta PSNR(dB) = PSNR_{proposed} - PSNR_{reference}$$

$$\Delta Bitrate(\%) = \frac{Bitrate_{proposed} - Bitrate_{reference}}{Bitrate_{reference}} \times 100\% \quad (6)$$

$$\Delta Time(\%) = \frac{Time_{proposed} - Time_{reference}}{Time_{reference}} \times 100\%$$

Table 2. Total Encoding Time and ME Time

Time Sequences	Total Encoding Time (sec)	Total ME Time (sec)
Akiyo	127.686	52.030
Coastguard	207.802	119.152
Container	153.245	76.112
Foreman	169.844	89.693
M & D	138.190	64.234
Silent	153.073	73.527
Suzie	156.942	86.407

We can observe that the time saving of total encoding time in our proposed method exceeds 30% in all sequences with negligible quality loss and bit rate increase. The increase of bit rate in Akiyo is a little high because the motion of this sequence is too small. In fact, at the beginning of this

Table 3. Changes of PSNR, Bitrate and Time

sequences	Wu's				Proposed			
	$\Delta PSNR$ (dB)	$\Delta Bitrate$ (%)	$\Delta Total$ Encoding Time (%)	ΔME Time(%)	$\Delta PSNR$ (dB)	$\Delta Bitrate$ (%)	$\Delta Total$ Encoding Time (%)	ΔME Time(%)
Akiyo	-0.02	0.71	-24.73	-32.80	-0.07	2.17	-36.22	-49.81
Coastguard	-0.02	1.28	-26.82	-36.01	-0.03	1.13	-33.59	-44.57
Container	-0.01	1.44	-20.69	-21.09	-0.05	1.46	-41.36	-54.14
Foreman	-0.09	2.14	-27.13	-33.31	-0.08	2.84	-31.39	-41.32
M & D	-0.08	0.75	-37.90	-48.35	-0.08	0.46	-36.41	-47.30
Silent	-0.06	2.46	-26.36	-37.23	-0.04	4.14	-36.53	-50.82
Suzie	-0.04	2.18	-35.89	-47.37	-0.01	2.75	-33.71	-43.87

sequence, Akiyo did not move last a while time. As a result, the motion is difficult to detect. Otherwise in Silent, small-sized blocks are concentrated on her arms because they move in a very high speed. But it is hard to distinguish inside or outside from her arms because there is about only one macroblock in width of that.

In the third experiment, we calculate the average processed modes per macroblock and show it in Table IV. If one of macroblocks is decided as P8x8 mode by the upper threshold, it performs four modes, else three ones will be performed. On the other hand, if the lower threshold works, only one mode will be performed. Figure 3 shows RD curves of four video sequences.

Table 4 Average Processed Modes per Macroblock

Sequence	modes
Akiyo	2.01
Coastguard	3.07
Container	1.89
Foreman	3.13
M & D	2.38
Silent	2.54
Suzie	3.03

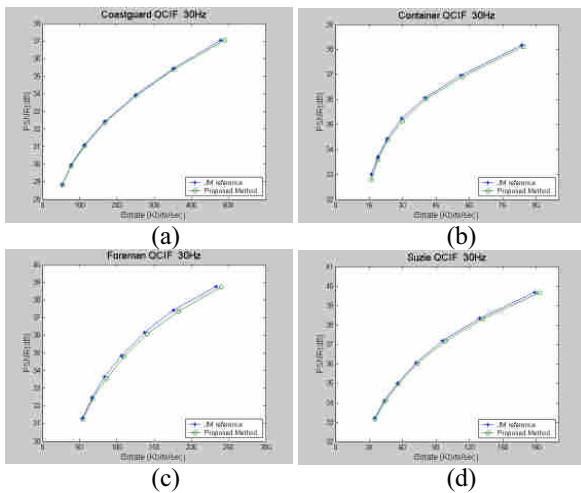


Figure 3. RD curves. (a)Akiyo (b) Container (c)Foreman (d)Suzie

5. Conclusion

This paper presented a new inter mode decision algorithm for video coding. Because the less change between frames, the co-located macroblocks with its neighbors in previous frame are used for predicting the mode of the current macroblock. This fast inter mode decision algorithm uses at most 3 modes as the candidates for checking the best one, and reduces 31%-41% of total encoding time with negligible quality loss about 0.05 dB and bit-rate increase about 2% on average. And the additional memory space used for storing the mode number could be less as possible that it just only enough to denote the mode of each macroblock.

Acknowledgement

We would like to thank National Science Council of Taiwan for her partial support of this research with grand no. NSC96-2221-E-216-039-MY2.

References

- [1] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, T. Wedi, "Video coding with H.264/AVC: Tools, Performance, and Complexity," IEEE Circuits and Systems Magazine, vol. 4, no. 1, pp. 7-28, First Quarter 2004.
- [2] Gary J. Sullivan, and Thomas Wiegand, "Video Compression – From Concepts to the H.264/AVC Standard," Proceedings of the IEEE, vol. 93, no. 1, pp. 18-31, Jan. 2005.
- [3] D. Marpe, T. Wiegand, Gary J. Sullivan, "The H.264/MPEG-4 Advanced Video Coding Standard and its Applications," IEEE Communications Magazine, vol. 44, no. 8, pp. 134-143, Aug. 2006.
- [4] Iain E. G. Richardson, "H.264/MPEG-4 Part 10," in H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia, New York: John Wiley & Sons, 2003, ch. 6, pp.159-223.
- [5] Libo Yang, Keman Yu, Jiang Li, and Shipeng Li, "An Effective Variable Block-Size Early Termination Algorithm for H.264 Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 6, pp. 784-788, Jun. 2005.
- [6] D. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, and C. C. Ko, "Fast Inter-mode Decision in H.264/AVC Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 6, pp. 953-958, Jul. 2005.
- [7] N.A. Khan, S. Masud, A. Ahmad, "A Variable Block Size Motion Estimation Algorithm for Real-time H.264 Video Encoding," Signal Processing: Image Communication, vol. 21, no. 4, pp. 306-315, Apr. 2006.
- [8] Christos Grecos, Ming Yuan Yang, "Fast Inter Mode Prediction for P Slices in the H.264 Video Coding Standard," IEEE Transactions on Broadcasting, vol. 51, no. 2, pp. 256-263, Jun. 2005.
- [9] Jongmin You, Wonkyun Kim, and Jechang Jeong, "16x16 Macroblock Partition Size Prediction for H.264 P Slices," IEEE Transactions on Consumer Electronics, vol. 52, no. 4, pp. 1377-1383, Nov. 2006.
- [10] Bin Feng, Guang-Xi Zhu, Wen-Yu Liu, "Fast Adaptive Inter-Prediction Mode Decision Method for H.264 Based on Spatial Correlation," IEEE International Symposium on Circuits and Systems 2006, pp. 21-24, May 2006.
- [11] Donghyung Kim, and Jechang Jeong, "A Fast Mode Selection Algorithm in H.264 Video Coding," IEEE International Conference on Multimedia and Expo 2006, pp. 1709-1712, Jul. 2006.
- [12] Yun Cheng, Silian Xie, Jianjun Guo, Zhiying Wang, Minlian Xiao, "A Fast Inter Mode Selection Algorithm for H.264," 2006 1st International Symposium on Pervasive Computing and Applications, pp. 821-824, Aug. 2006.
- [13] Zhenyu Wei, and King Ngi Ngan, "A Fast Macroblock Mode Decision Algorithm for H.264," IEEE Asia Pacific Conference on Circuits and Systems 2006, pp. 772-775, Dec. 2006.
- [14] Chih-Hung Kuo, Meiyin Shen, and C.-C. Jay Kuo, "Fast Motion Search with Efficient Inter-prediction Mode Decision for H.264," Journal of Visual Communication & Image Representation, vol. 17, pp. 217-242, Jul. 2005.
- [15] Zhi Zhou, Jun Xin, Ming-Ting Sun, "Fast Motion Estimation and Inter-mode Decision for H.264/MPEG-4 AVC Encoding," Journal of Visual Communication & Image Representation, vol. 17, pp. 243-263, Jul. 2005.
- [16] H.264/AVC Reference Software JM version 12.2, http://iphome.hhi.de/suehring/tml/download/old_jm/