行政院國家科學委員會補助專題研究計畫成果報告

多點視訊會議技術之研究(I)

Research in Multipoint Videoconferencing Technologies

計畫編號：NSC 92-2219-E-009-009
執行期限：92 年 8 月 1 日至 93 年 7 月 31 日
主持人：林大衛 交通大學電子工程學系 教授
計畫參與人員：詹益鎬、林岳賢、蔡鎮宇、簡志凱、劉明瑋
交通大學電子工程學系 研究生

摘要

近年來，桌上型視訊會議技術已愈趨實用與可即。然而目前一般的系統仍不具有近似當面開會的視聽感覺。本計畫主旨在研究分散式桌上型多點視訊會議技術，其中特別著重視訊的處理。我們擬在每個會議端點的電腦螢幕上顯示一個虛擬的會議室場景，其中呈現所有其他端點的與會人員。為此，每一端點需先將本地輸入視訊加以分割，取出與會者影像予以編碼，然後傳到其他端點。每一端點也需將所有接收到的視訊予以解碼及合成。本計畫之研究係建構在 MPEG-4 規範的基礎上，採用個人電腦為實現平台。計畫之研究子題可分四大組：會議系統、網路傳輸、傳送視訊處理、與接收視訊處理，預定以三年時間進行研究。本報告係針對第一年之研究，其中著重視訊分割技術、有效率之 MPEG-4 視訊編碼、及 MPEG-4 規範之了解。在視訊分割方面，我們提出了一些做法，並使用個人電腦實作了一個即時視訊輸入與分割系統。在 MPEG-4 視訊編碼方面，我們使用一個公眾領域的軟體加以改進，在個人電腦上實現了一個較快速的 MPEG-4 視訊編碼器。在其他 MPEG-4 相關規範之了解方面，我們研讀了有關場景合成與訊號傳輸介面的文獻，也試取得相關軟體並檢驗其功能。

關鍵詞：視訊分割、視訊合成、MPEG-4 視訊、軟體即時視訊編碼、多點視訊會議

Abstract

Desktop videoconferencing is becoming more practical and accessible in recent years. But typical systems today still lack a nearly face-to-face conference look and feel. The purpose of this project is to research into technologies for distributed desktop multipoint videoconferencing, wherein we especially emphasize the processing of video signals. On the computer screen of each conference terminal, we intend to show a virtual conference room where the images of conferees at all other terminals are shown. For this, each terminal will first need to segment the local input video to extract the image of the local conferee, encode it, and transmit it to the other terminals. Each terminal will also need to decode all the received videos and form a composition. The research is conducted based on the MPEG-4 specifications, utilizing personal computers as the realization platform. Subjects in this research can be divided into four major groups: conferencing system, network transport, transmitter video processing, and receiver video processing. The intended period of research is three years. This report is concerned with research done in year one, where we have emphasized video segmentation techniques, efficient MPEG-4 video coding, and the understanding of MPEG-4 specifications. In video segmentation, we proposed several methods and implemented a real-time video input and segmentation system on a personal computer. In MPEG-4 video coding, we implemented a faster MPEG-4 video encoder on a personal computer by improving a public-domain software. In understanding of other MPEG-4-related specifications, we studied literature about scene composition and the signal transmission interface, and we tried to acquire some relevant software packages and examine their functionalities.

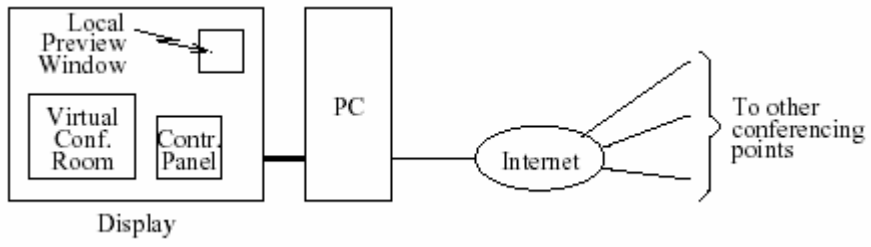
Keywords: Video Segmentation, Video Composition, MPEG-4 Video, Real-time Software Video Coding, Multipoint Videoconferencing

目錄 Table of Contents

一、計畫緣由與目的

過去所謂視訊會議，通常是在特定的視訊會議室、使用特用的視訊會議設備來進行。近年來，隨著視訊壓縮技術與數位積體電路的發展，透過個人電腦與網際網路來在辦公室或家中打視訊電話或進行多點視訊會議已經不再是未來的夢想。如 NetMeeting (Microsoft), IVS (INRIA), NV (Xerox Parc), vic (U.C. Berkeley and Lawrence Berkeley National Laboratory), CU-SeeMe (Cornell University), 及 iVisit 等系統，都提供即時互動視訊通訊的功能。這種型態的互動式視訊通訊，有時被稱為桌上型(desktop)視訊通訊。目前一般的桌上型多點視訊會議(multipoint videoconferencing)系統，仍未提供近似於當面開會的視聽感覺。其實，即使是使用特定房間與設備的傳統視訊會議，也未能達到近似於當面開會的特質。最近有一些研究，即在探討視訊會議所呈現的畫面等課題，以期能使視訊會議更具有趨近於當面開會的視聽特質。這些研究中考慮將由數個會議端點所收到的視訊合成一個類似會議室的場景。當然，在合成之前可能需要先將這些會議端點的與會人員的影像分割出來。幾個此類研究的例子見[1]-[4]。國內人士過去亦有相關之研究[5]-[7]，但相關之技術議題仍有不少待進一步探討之處。

本計畫主旨在研究桌上型多點視訊會議之相關技術，並建構一個實驗性之系統。本計畫之研究標的，可透過圖一說明之。圖中左方呈示之個人電腦(PC)及螢幕(display)，為每一會議點所使用之設備。螢幕顯示一個虛擬之會議室(virtual conference room)，其中之視訊為所有其他會議點所傳來之與會人員視訊的一個合成(composition)。會議室場景及各人的位置是由主席安排。螢幕上另有二個視窗，即控制台(control panel)與本地視訊之預覽(local preview)。我們發現：最近擬訂的 MPEG-4 標準，其中的若干規範相當適合本研究之所需。例如：其視訊編碼部分容許將視訊分割後再編碼，其資料結構與合成部分定義了一個相當有效率的 BIFS (Binary Format for Scenes)，其網路傳輸部分定義了 DMIF (Delivery Multimedia Integration Framework)等[8]。故本計畫之研究係基於 MPEG-4 之規範。



圖一：多點視訊會議系統架構示意圖

本計畫之研究子題可分為四個群組，即會議系統、網路傳輸、傳送視訊處理、及接收視訊處理，預定以三年時間進行相關研究。本報告係針對第一年之研究，其中重點為第三群組子題之深入研究，以及第二與第四兩群組子題之初步探討。申言之，在第三群組子題，即傳送視訊處理方面，我們研究了視訊分割技術與有效率之 MPEG-4 視訊編碼。關於視訊分割，我們提出了一些適用於會議型態影像的視訊分割方法，並使

用個人電腦實作了一個簡單的即時視訊輸入與分割系統，目前在繼續改進其運算速度中。關於 MPEG-4 視訊編解碼，我們使用一個公眾領域的軟體加以改進，在個人電腦上實現了一個較快速的 MPEG-4 視訊編碼器。而在第二及第四群組子題，即網路傳輸與接收視訊處理兩方面，我們研讀了有關 MPEG-4 場景合成與訊號傳輸介面之規範的文獻，也試取得相關軟體並檢驗其功能。在第二年度之研究中，我們將較著重第二及第四群組子題的探討，也擬開始第一群組子題，即會議系統之研究。上述視訊編碼器改進過程中所累積的經驗，可為第四群組子題中，視訊解碼器實現之參考。

二、結果與討論

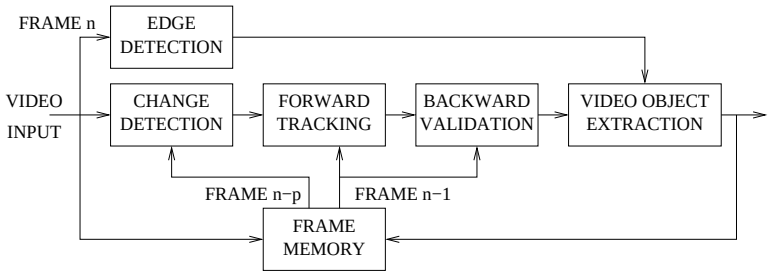
本節中，我們分項討論視訊分割之研究、有效率之 MPEG-4 視訊編碼之研究、及場景合成與訊號傳輸介面之初步研究等三個主題。在視訊分割方面，我們的研究係分兩途進行：一是考慮如何更精確的界定移動物件的邊界，並就使用分割後的物件做場景合成的方法及效果做初步探討；二是使用背景建構的觀念，發展出一個視訊分割方法，並使用個人電腦建構一個包括視訊輸入、即時分割、與分割後視訊顯示的即時視訊分割系統。前者所提出的界定移動物件邊界的方法，可為後者未來改進之參考。前者對場景合成之研究成果，則亦可為前節所述第四群組子題後續研究之參考。

因此，以下就分四小節分別簡述上開三主題之研究，其中視訊分割依上述兩途分為二小節討論之。

A. 視訊分割：更精確界定物件邊界及將分割結果用於場景合成之研究

視訊分割有兩大議題尚需繼續研究，一是物件邊界之精確界定，二是運算複雜度。關於物件邊界之界定，幾個常見的途徑為分水嶺分析(watershed analysis) [9], [10]、輪廓演化(contour evolution) [11], [12]、及邊緣連結(edge linking) [13], [14]。本研究考慮第三者，並提出一個有效率的運算方法。

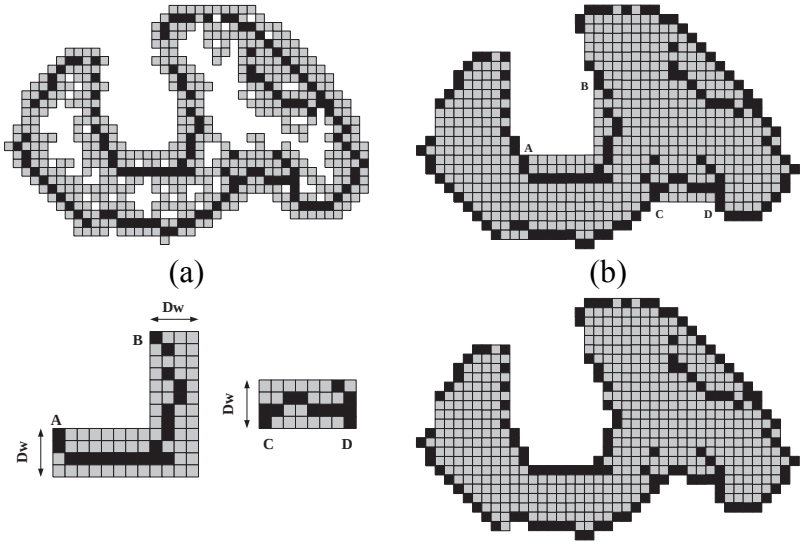
本研究所提出的視訊分割演算法架構如圖二所示。其中 “Edge Detection” 使用 Canny edge detector [15]； “Change Detection” 係將兩畫框(frame)間相異程度較高的像素取出稱為 changed pixels； “Forward Tracking” 及 “Backward Validation” 係使用階層式運動估計，等效之搜尋範圍為正負 14 像素，複雜度約與全尋法相似。(但我們亦另外發展了更快速的運動估計法，在投稿中。) “Video Object Extraction” 為最創新之部分，使用形態學式(morphological)之處理，以獲得一相當逼近實際物件邊界之物件草型(object mask)，然後用 Dijkstra 最短路徑演算法[16]連結草型外緣邊界之「斷裂」之處，以得到最後萃取出之物件。細節可參[17]。由於 Dijkstra 演算法之複雜度與其所需搜尋的像素數目成平方關係，故當物件草型相當逼近實際物件邊界時，Dijkstra 演算法之複雜度可以降低。



圖二：所提出之視訊物件萃取與追蹤方法之一

圖三為一任意之例，用以說明 “Video Object Extraction” 方塊運作之效果。其中圖三(a)所示為 Video Object Extraction 方塊之輸入。圖中之小格表示像素，灰格及黑格共同表示經 backward validation 後之物件草型，黑格則表示草型中屬邊緣的部份(為

Edge Detection 方塊之輸出)。圖三(b)所示為上述所謂相當逼近實際物件邊界之物件草型，可見其外緣相當貼近最外方的邊緣(黑格)。Dijkstra 演算法在 A-B 及 C-D 間的搜尋範圍見圖三(c)，其中 Dw 為搜尋深度，為一可調之參數。經使用 Dijkstra 演算法連結所有外緣邊界斷裂之處後，結果如圖三(d)所示。

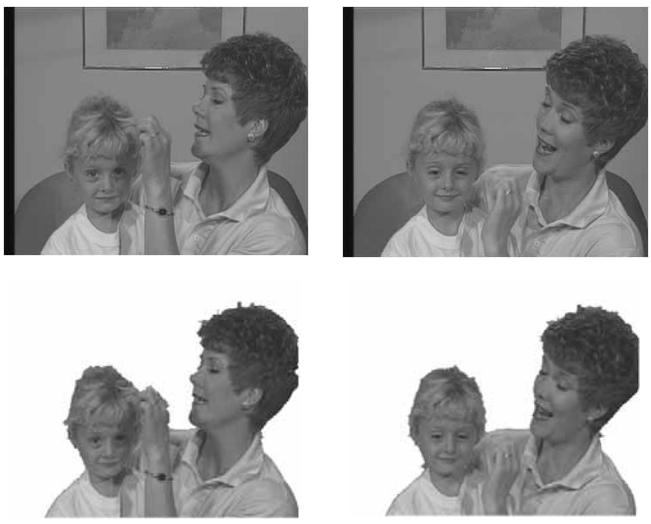


(a) (b)

(c) (d)

圖三：任意之圖例，用以說明 “Video Object Extraction” 方塊運作之效果

以上 video object extraction 方法，在物件形狀高度非凸狀(highly nonconvex)時，特別能顯出其效用。圖四顯示一些視訊分割結果，其中最後之 Dijkstra 演算法所用之搜尋深度為 Dw = 5。配以較高速之運動估計法，整個分割法可對 CIF (352x288)視訊在現有一般個人電腦中達每秒數十張畫框之即時執行速度。



80 150

圖四：一些 Mother-and-Daughter 視訊之分割結果。上列：原始圖框；下列：分割出之移動物件。底部數字為圖框序號

其次我們考慮使用分割出之物件做場景合成。此種處理應該常會需要做物件之放大或縮小。為此我們使用 MPEG-2 中規範的二倍縮放濾波器，設計了一個簡單的多倍數連續放大或縮小的作法。這種連續性的縮小或放大在作物件大小可調式編碼(spatial scalable coding)時，可能相當有用。MPEG-2 之二倍放大及縮小濾波器係數各為 [-12, 140, 140, -12]/256 及 [-29, 0, 88, 138, 88, 0, -29]/256。若放大倍數 m 為 2 的整數次方，則只要使用 MPEG-2 放大濾波器整數倍即可。若 m 為整數但不為 2 的整數次方，則我們先將物件放大至比 m 小的、2 的整數次方倍，然後使用相鄰的點作線性內插，以求得在 m 被放大時的各像素數值。縮小倍數為整數倍時，處理方式相似。若縮放倍數為有理數但非整數，亦可做類似之處理，見[17]。圖五顯示幾個使用分割後之物件與其他影像合成後的場景。圖面可稱相當自然。

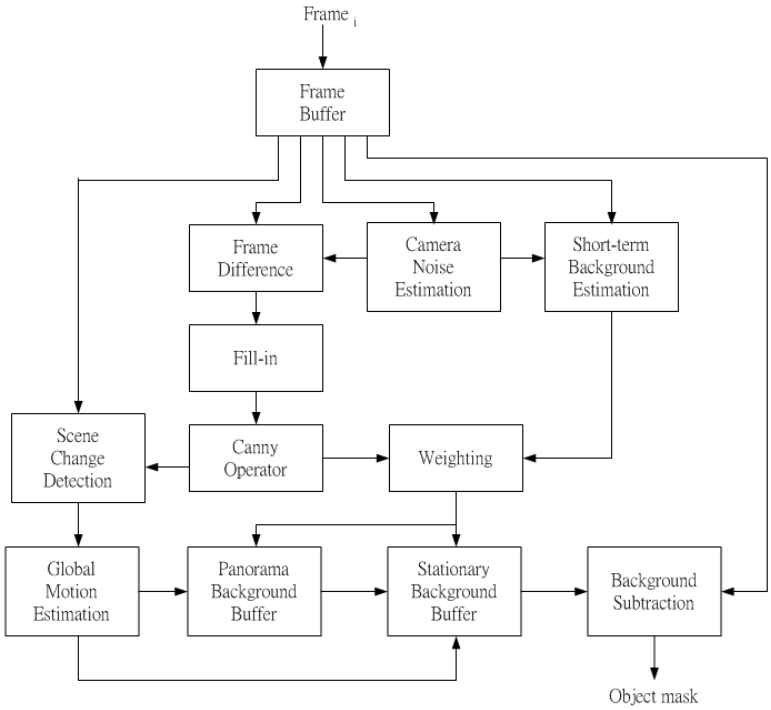
圖五：使用分割而得之物件(Akiyo, Claire, Salesman)，在縮小二倍後，與其他影像合成之場景

B. 視訊分割：使用背景建構法分割視訊以及即時視訊分割系統之實現

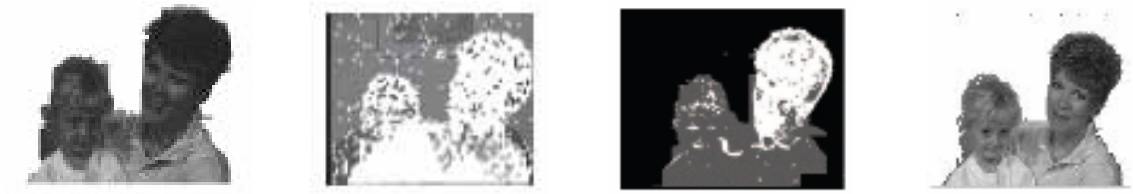
在此，我們設計一種方法來收集各個畫框中的背景部份，建構出一個盡量完整的背景圖。然後將現在收到的畫框與背景圖比較，把差異很大的部份取出並做一些修飾，就可以將移動的前景物件分割出來了。

演算法的架構如圖六所示。其中我們首先分析視訊中的攝影機雜訊量(Camera Noise Estimation)。我們設計了一個二級的方法來估計雜訊的變異數，以減低移動物件對估計精確度的負面影響。細節見[18]。其次，我們構建一個暫時性的前景物件草型(temporary foreground mask)。這是透過圖六中的 Frame Difference、Fill-In、及 Canny Operator 三個功能方塊達成。其中 Frame Difference 取得畫框中變異較大的像素，Fill-in 將像素間的空白處填滿，使其涵蓋移動物件的區域，Canny Operator 及相隨的運算則使區域內縮，使之更接近實際的物件形狀(但可能仍有相當差異)。圖七(a)所示為一個結果的例子。第三，我們構建一個短期的背景(Short-term Background Estimation)。這是透過分析連續六張畫框來達成。如果某一像素值在這些畫框中變化不大，則暫將之算為背景像素。圖七(b)所示為一個結果的例子。第四，我們使用以上結果來構建一個靜態背景畫面(Stationary Background Buffer)。由於移動物件若是內部的亮度與色彩相當平滑，則在簡單的分析中，有可能被誤判為背景，如圖七(b)所暗示。

所以我們使用之前獲得的暫時性前景物件草型來將短期背景像素予以加權，如果其累積之加權值超過某一門檻，再將之放入最終的靜態背景畫面。繼續上例，圖七(c)所示為根據圖七(a)所得之權值，其中黑色表示最高的權值(比較可靠的背景部份)，白色表示最低的權值(零)，灰色表示中等的權值。



圖六：使用背景建構法來分割視訊的演算法架構



(a) (b) (c) (c)

圖七：演算過程中之部分結果

以上結果，在靜態背景的情況下，已可用。但有時攝影機會被移動，此時就需調整背景。若要將既有背景畫面刪除，重新計算出一個，自然可以。但這會需要較多的畫框來重新建構一個背景畫面。但在視訊電話或視訊會議的應用中，攝影機移動之後的背景畫面，可能有很大一部分是和之前的畫面重疊的。所以我們考慮估計攝影機的運動，以使我們可以由之前已建構好的背景畫面中獲得此重疊的部份，以加速新背景畫面的建構。這就是圖六中 Scene Change Detection、Global Motion Estimation、及 Panorama Background Buffer 的作用所在了。其中 Scene Change Detection 是用以偵測攝影機是否有移動，Global Motion Estimation 係使用 affine motion model，而 Panorama Background Buffer 則收存迄今獲得之完整背景畫面(在 MPEG-4 中可稱 sprite)。由此 Panorama Background Buffer 可協助建構現在的 Stationary Background

Buffer。最後，將各個要分割的畫框與 Stationary Background Buffer 相減，並稍加修飾，即可得所分割出之移動物件。圖七(d)為一例。

根據以上演算法，我們在個人電腦上實現了一個即時視訊輸入、分割、及結果顯示的系統。圖八為視窗介面的一個 snapshot。由於視窗程式佔去不少時間，以 2.4 GHz P-4 CPU 而言，目前尚未最佳化的程式所能達到的速度，在攝影機靜止的情況下，約每秒 5 張 CIF 畫框，在攝影機有移動的情況下，則約每秒 2 張畫框。

圖八：即時視訊分割系統視窗介面之 snapshot

C. 有效率之 MPEG-4 視訊編碼

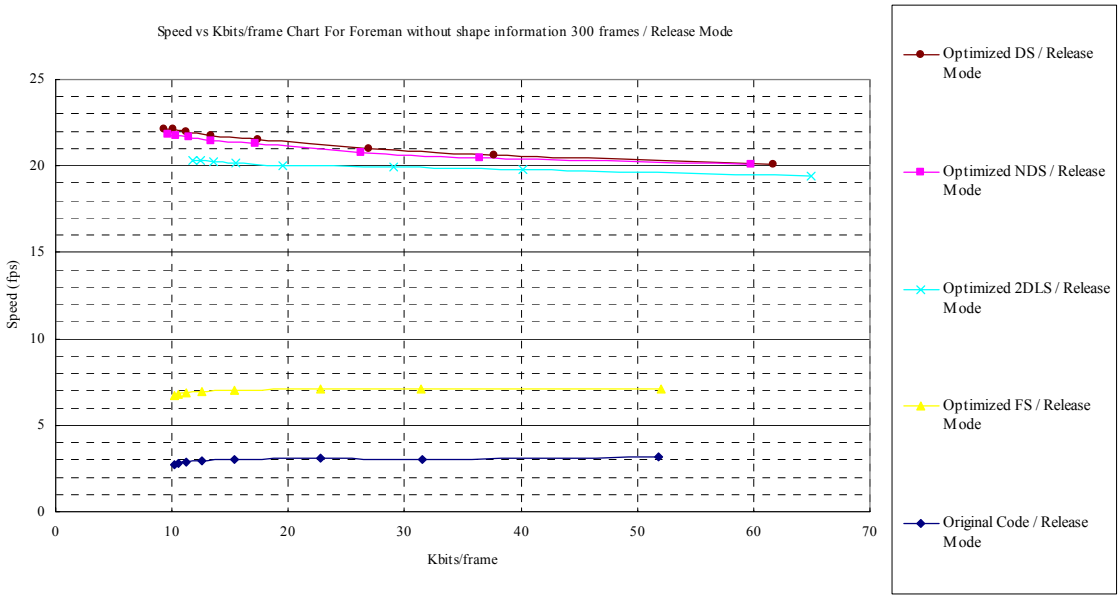
由於我們考慮使用個人電腦做會議系統的平台，所以需要在其上建構一個有效率的 MPEG-4 視訊編碼器。過去有些公眾領域的 MPEG-4 視訊編碼軟體，並非特別為個人電腦上的執行效率而設計，所以其執行效率有大幅提升的空間。此處我們使用 Microsoft 公司開發的一個公眾領域的 MPEG-4 視訊編碼軟體，在個人電腦上，試改良程式寫法，並使用 Intel CPU 的 MMX (multimedia extension) 處理器單元以加速其運算。上述編碼軟體係處於 MPEG-4 視訊定義之 Main Profile 和 Simple Scalable Profile 之層次。

Intel 之 MMX 處理器單元，是在 Pentium CPU 以外的處理單元，第一代稱為 MMX，有八個 64 bits 的 registers，可儲存 8 bytes 或 4 words 或 2 doublewords 或 1 quadword。第二代稱 SSE (streaming SIMD extensions)，加了八個 128 bits 的 registers，可進行一些浮點運算。不過視訊編解碼中，不太需要用到浮點運算，所以此一功能效用不大。第三代稱 SSE2 (streaming SIMD extension 2)，其中八個 128 bits 的 registers 可做定點運算，這對視訊編解碼就很有用了。

總而言之，MMX 處理器中共有 16 個大小不一的 registers，而 Intel 也提供了一些可在此處理器中執行的高效率平行計算組合語言指令。由於此處理器係與 Pentium CPU 分開運作，加上 Intel 的 C++ compiler 可以處理這些指令與 C++程式併用的情形，所以即使使用組合語言來寫部分的程式，也不太困難。我們遂分析 MPEG-4 視訊編碼軟體，找到其中較耗時的部份。然後使用適用的 MMX 處理器指令來加速。上述分析，主要是藉助於 Intel 的 VTune performance analyzer。此軟體工具提供 tree-structured call

graph、各函式使用時間百分比及使用之 clockticks 等資料，相當有助於了解程式的瓶頸所在，並可用以比較修改前及修改後之程式運作差異。

程式中第一耗時的，如一般可預期，是運動估計。對有的視訊而言，可佔 90%以上的運算時間。我們也使用較快速(但效果稍遜)的運動估計法作實驗。圖九顯示 Foreman 視訊的編碼速率，其中原始程式約為每秒 3 張畫框，修改後但使用相同運動估計法(FS, 即 full search)的程式約每秒 7 張，改用較快速的運動估計法(DS = diamond search, NDS = new diamond search, 2DLS = two-dimensional logarithmic search)則可達每秒 20 張。

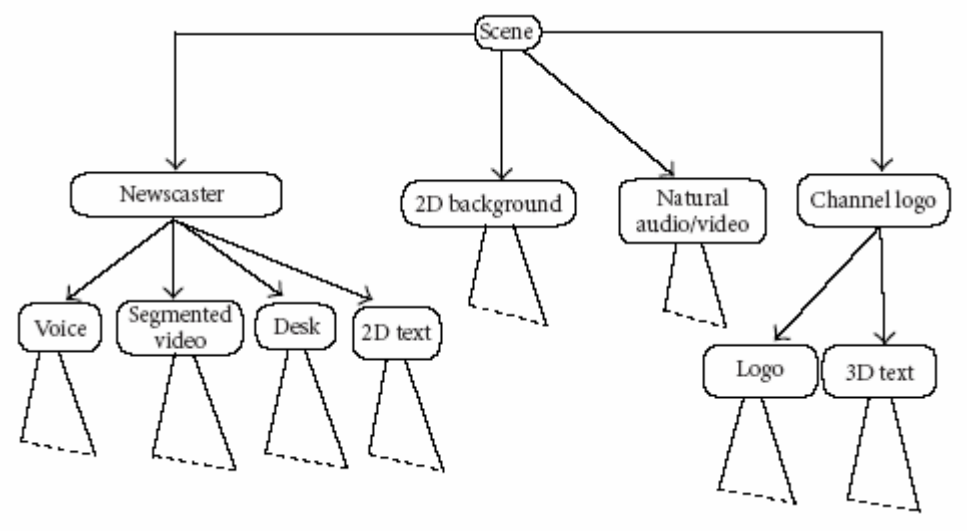


圖九：修改前後之程式效率比較

D. MPEG-4 場景合成及訊號傳輸介面之初步研究

MPEG-4 中的場景合成，可使用其二進位格式場景(Binary Format for Scenes，簡稱 BIFS)規範達成。BIFS 是根據 VRML 發展而來，用來表示預先定義的視聽物件(Media Object)的行為及時空關係。BIFS 攜帶了場景描述的資訊，規定了如何重現 MPEG-4 的場景圖，實現物件的動畫和互動行為，以及對這些元素的生成加以時序化和同步化；此外還定義了事件的處理、物件組合及運行規則等。

場景描述的結構採用如圖十所示的樹狀結構，透過各個節點(Node)來描述一個完整的場景，節點又由表示節點特性的一組「域(Field)」所組成。域可以為某一特定值，如球節點的半徑大小等，有時可以表示許多值，如定義一個多邊型頂點的表列。許多節點都包含在其他節點的域中，這就是場景描述構成樹狀結構的原因。



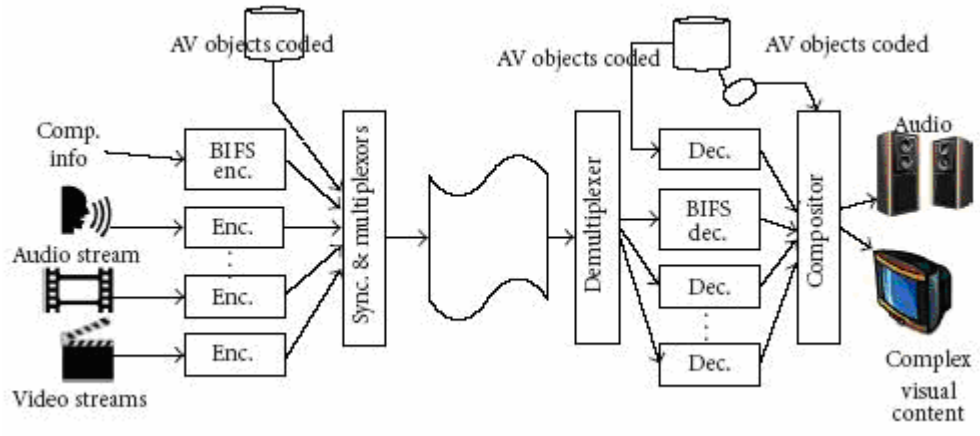
圖十：場景描述結構(取自[19])

BIFS–Command 的設定使得場景圖中的屬性可以改變，用來修改場景在某個給定時間的一系列屬性。為了能夠在一個單一訪問單元中發送幾個命令，命令被組合到 CommandFrames 中。BIFS-Command 一共定義了 4 個基本命令，分別為：Replacement of an entire scene；Insertion；Deletion；Replacement。

BIFS–Anim 則提供了場景中節點的某些域的連續更新，它被用來組合各種動畫。儘管 BIFS-Anim 和 BIFS-Command 都有相同的基本串流(Bitstream)類型，但它們不會在同一個基本串流中傳輸。BIFS-Anim 資訊在一個獨立的基本串流中傳輸，和傳輸 BIFS-Command 的串流分開。

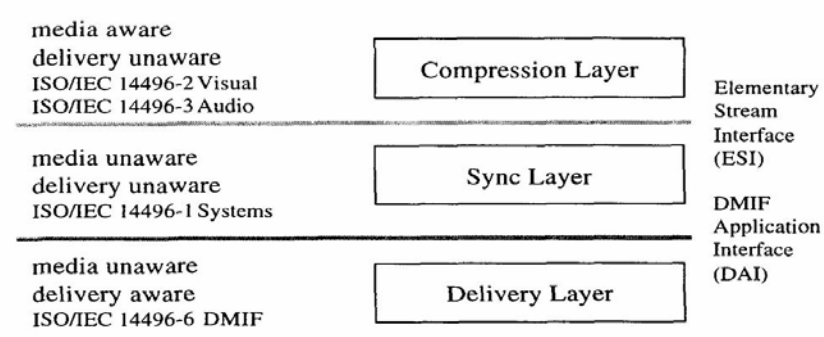
此外，BIFS 導入了量化(Quantization)的概念，可以對各種不同類型的域量化，降低了資料量，使網路傳輸能更有效的進行。

圖十一顯示了 MPEG-4 標準中描述的一個完整的多媒體系統，從圖中可以看出，MPEG-4 將視聽物件、場景描述資訊作為基本串流進行傳輸，依靠 BIFS 的場景描述資訊將視聽物件組合，進而生成多媒體互動式場景。



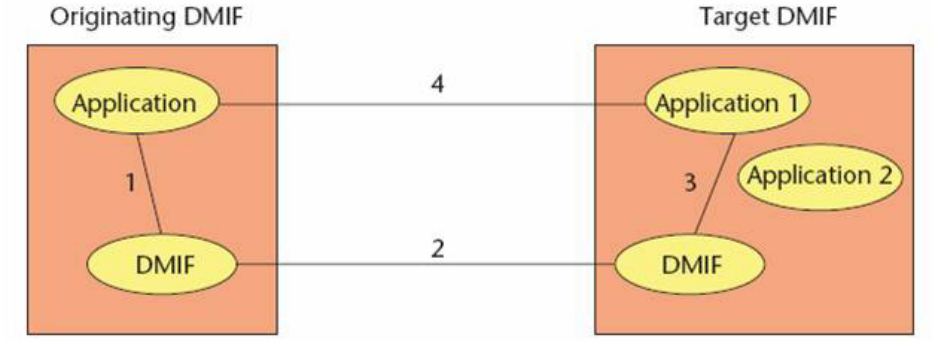
圖十一：MPEG-4 系統架構圖(取自[19])

關於訊號傳輸，MPEG-4 為了能夠更廣泛地運用在不同的網路上，因此制定了一個處理網路傳送的部份，稱為多媒體傳輸整合骨架(DMIF, Delivery Multimedia Integration Framework)。圖十二便是闡述 MPEG-4 將資料和網路傳送分層的精神。將資料壓縮層放在最上層，處理後再經過系統層將資料做同步的處理再傳給下層的 DMIF 負責網路傳送。MPEG-4 主要是由系統層去掌控大局。



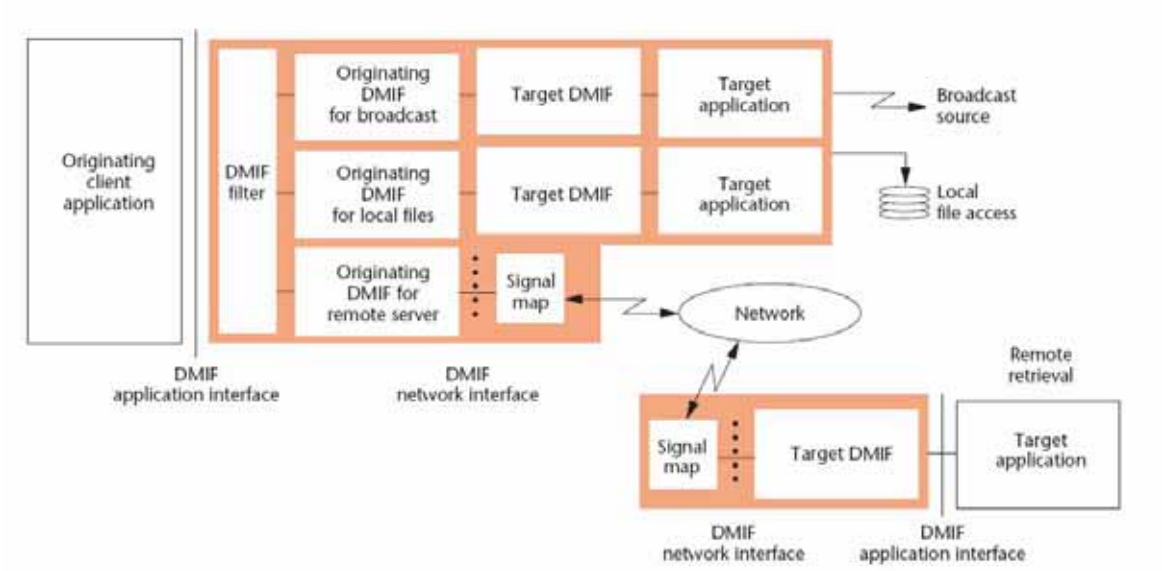
圖十二：闡述 MPEG-4 將資料壓縮及訊號傳輸分層處理的基本精神與架構(取自[20])

接下來，我們來看 DMIF 如何動作以及溝通。最宏觀地來看(圖十三)，當有一個應用程式要和遠端工作站應用程式互相傳輸時，是由 DMIF 去做事前溝通的動作，等溝通完畢時，會再建立一條應用程式之間的通道，資料由此真正傳送。



圖十三：DMIF 運作之宏觀示意圖(取自[21])

那真正運用 DMIF 又是怎麼樣的情況呢？我們拿微軟出的測試檔案 IM1 (implementation 1)來做分析(圖十四)，可以發現 DMIF 把在硬碟上儲存以及在網路內廣播都視作是一種網路傳輸方式，使 DMIF 能夠更加地全面。而 DMIF 之所以能夠適用於各種網路上，主要都是靠 DMIF Filter 的部份來根據系統層傳出來的資訊做選擇。而再接下來的 DMIF Instance 的部份則約和 OSI 的 session layer 相當，做一些檢查，授權等的動作，而將資訊真正地傳輸出去。



圖十四：IM1 軟體中的實現的 DMIF 功能(取自[21])

為了分層，讓資料能夠獨立處理而不管網路傳輸的部份，在 DAI (DMIF-Application Interface)的部份便規定幾個參數，只要資料在系統層加上這些參數為 header 的格式之一，便能夠傳送資料了，換句話說，DAI 只認定幾個參數；這些參數各是服務(service)，通道(channel)以及資料(data)參數。首先，先依照服務參數決定和何種遠端工作站連線以及進行何種服務；接者便靠通道參數建立通道；最後再靠資料參數建立傳送資料的通道。

那在 DMIF Filter 的部份，又是如何選擇那一部份的網路做傳送呢？在 IM1 內做的並不明智，它是將 DMIF Instance 寫成 dll 檔，再一個個掛上 DMIF Filter，讓 DMIF 一個個嘗試。最後，當資料通道建好後，要傳送資料時，TransMux 會負責將所需網路相仿的資料都集合在一起，再傳送出去，而接收端收到這些資料時則再由 TransMux 負責將資料還原為原本的獨立資料。如此一來，IM1 便能夠有效實現 MPEG-4 的 DMIF 規格了。

三、參考文獻

[1] M. E. Lukacs and D. G. Boyer, "A universal broadband multipoint teleconferencing service for the 21st century," *IEEE Commun. Mag.*, vol. 33, no. 11, pp. 36-43, Nov. 1995.

[2] D. G. Boyer, M. E. Lukacs, and M. Mills, "The personal presence system experimental research prototype," in *IEEE Int. Conf. Commun. Conf. Rec.*, pp. 1112-1116, 1996.

[3] O. Schreer, M. Karl, and P. Kauff, "PCI-based multi-processor system for immersive videoconference terminals," in *IEEE Int. Conf. Multimedia Expo*, pp. 181-184, 2002.

[4] O. Schreer, M. Karl, and P. Kauff, "A Trimedia based multi-processor system using PCI technology for immersive videoconference terminals," in *Int. Conf. Digital Signal Processing*, pp. 289-293, 2002.

[5] Y.-J. Chang, C.-C. Chen, J.-C. Chou, and Y.-C. Chen, "Implementation of a virtual chat room for multimedia communications," in *IEEE Workshop Multimedia Signal Processing*, pp. 599-604, 1999.

[6] Y.-J. Chang, C.-C. Chen, J.-C. Chou, and Y.-C. Chen, "Virual Talk: a model-based virtual phone using layered audio-visual integration," in *IEEE Int. Conf. Multimedia Expo*, pp. 415-418, 2000.

[7] C.-W. Lin, W.-H. Wang, M.-T. Sun, and J.-N. Hwang, "Implementation of H.323 multipoint video conference systems with personal presence control," in *IEEE Int. Conf. Consumer Electron. Digest of Tech. Papers*, pp. 108-109, 2000.

[8] S. Battista, F. Casalino, and C. Lande, "MPEG-4: a multimedia standard for the third millenniem," in two parts, *IEEE Multimedia*, vol. 6, no. 4, pp. 74-83, Oct.-Dec. 1999, and vol. 7, no. 1, pp. 76-84, Jan.-Mar. 2000.

[9] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 539-546, Sep. 1998.

[10] S.-Y. Chien, Y.-W. Huang, and L.-G. Chen, "Predictive watershed: a fast watershed algorithm for video segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 5, pp. 453-461, May 2003.

[11] S. Sun, D. R. Haynor, and Y. Kin, "Semiautomatic video object segmentation using VSnakes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 1, pp. 75-82, Jan. 2003.

[12] A.-R. Mansouri and J. Konrad, "Multiple motion segmentation with level sets," *IEEE Trans. Image Processing*, vol. 12, no. 2, pp. 201-220, Feb. 2003.

[13] T. Meier and K. N. Ngan, "Video segmentation for content-based coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1190-1203, Dec. 1999.

[14] C. Kim and J. N. Hwang, "Fast and automatic video object segmentation and tracking for content-based applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no.

2, pp. 122-129, Feb. 2002.

[15] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679-698, Nov. 1986.

[16] D. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269-271, 1959.

[17] Y.-H. Jan and D. W. Lin, "Edge-and-motion-based semantic video object extraction and application to scene composition," in *Proc. IEEE Int. Symp. Consumer Electronics*, pp. 340-344, Sep. 2004.

[18] Y.-H. Lin and D. W. Lin, "Real-time video segmentation based on background modeling applicable to videoconferencing," to appear in *Proc. Workshop Consumer Electronics*, Hsinchu, Taiwan, ROC, Nov. 2004.

[19] P. Daras *et al.*, "MPEG-4 authoring tool using moving object segmentation and tracking in video shots," *EURASIP J. Applied Signal Processing*, vol. 2003, no. 9. pp. 1-18.

[20] http://ailab.chonbuk.ac.kr/~sjmun/mpeg4/tutorial/6-DMIF_paper.html.

[21] G. S. Tselikis, "An overview of the Delivery Multimedia Integration Framework for broadband networks," *IEEE Commun.*, vol. 2, no. 4, 1999.

四、計畫成果自評

研究內容與原計畫相符程度：符合計畫主題，達成之主要成果包括：視訊分割之方法及軟體實現、有效率之 MPEG-4 視訊編碼、及 MPEG-4 場景合成與訊號傳輸介面之初步研究等。

達成預期目標情況：本子計畫達成之貢獻形式，含創新之發現、技術水準之提升、實驗系統之建立、人才培育。

成果之學術與應用價值等：視訊分割方面之若干成果已發表或投稿為會議論文，並在進行期刊論文之撰稿與投稿。MPEG-4 視訊編碼實作方面之經驗與成果將成為我們後續相關研究的參考。以上成果亦皆可供相關業界參考，惟暫不擬以收費技轉方式處理。

綜合評估：本計畫獲得一些具有學術與應用價值的成果，並達人才培育之效。成效良好。

五、發表論文

以下附二篇論文，如下列：

[1] Y.-H. Jan and D. W. Lin, “Edge-and-motion-based semantic video object extraction and application to scene composition,” in *Proc. IEEE Int. Symp. Consumer Electronics*, pp. 340-344, Sep. 2004.

[2] Y.-H. Lin and D. W. Lin, “Real-time video segmentation based on background modeling applicable to videoconferencing,” to appear in *Proc. Workshop Consumer Electronics*, Hsinchu, Taiwan, ROC, Nov. 2004.

Edge-and-Motion-Based Semantic Video Object Extraction and Application to Scene Composition

Yih-Haw Jan and David W. Lin, Senior Member, IEEE

Abstract — *We consider automatic segmentation of natural video for content-based video applications. In particular, we present a segmentation algorithm employing both motion and edge information. It employs the edge-linking approach for accurate determination of object boundaries and bidirectional motion estimation for robust tracking of object motion and shape change. The algorithm is designed with computational complexity also in mind. We further consider scene composition using the segmented objects. We describe and evaluate a method to enlarge or shrink objects for such purpose. The method is amenable to spatial-scalable coding.*[1]

Index Terms — Object resizing, object tracking, scene composition, video segmentation.

I. Introduction

VIDEO segmentation for object-based coding and video content manipulation has received much recent attention. In this work, we consider employing the segmented video objects in scene composition. It is noted that current segmentation techniques are still in need of improvement in two areas to make them more practical for various applications. The two areas are accuracy in identification of object boundaries (especially when the objects are grossly nonconvex in shape) and computational complexity.

In terms of object boundary identification, several major approaches are watershed analysis [1], [2], contour evolution [3], [4], and edge linking [5], [6]. We take the edge linking approach, which appears to be able to result in reasonably accurate identification of object boundaries. We present a method that performs edge linking efficiently. To track robustly the motion and potential shape changes of the segmented objects, we conduct both forward and backward motion estimation. The motion estimation method is designed with computational complexity in mind.

In using the segmentation results in scene composition, we frequently need to enlarge or shrink the objects. We consider how these can be done efficiently with good performance, in a manner that is also suitable for spatial-scalable coding.

In what follows, Section II describes the video segmentation algorithm and the associated experimental results. Section III discusses the use of the segmentation results in scene composition. And Section IV is the conclusion.

[1] This work was supported by the National Science Council of Republic of China under grant no. NSC 92-2219-E-009-009 and by the Lee and MTI Center for Networking Research at National Chiao Tung University.

The authors are with the Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan 30010, ROC (e-mails: yhjan.ee86g@nctu.edu.tw, dwlin@mail.nctu.edu.tw).

II. Proposed Video Segmentation Algorithm

The proposed video segmentation algorithm is shown in Fig. 1. The algorithm employs both motion analysis and edge analysis, where the motion analysis consists in the blocks marked "change detection," "forward tracking," and "backward validation," and the edge analysis consists in the blocks marked "edge detection" and "video object extraction." In edge detection, we employ the Canny edge detector [7]. In change detection, we employ a statistical significance test on the interframe pixel value differences [8], [9]. Details are omitted. The primary novelty of the algorithm consists in the forward tracking, backward validation, and video object extraction blocks.

A. Video Object Tracking

"Forward tracking" and "backward validation" apply to the second and subsequent frames in a video sequence. They are omitted for the first frame, for there is no prior presence of objects to track in the first frame.

"Forward tracking" tracks the motion and shape change of each object. It employs hierarchical, block-based motion estimation [10] to find the motion vectors, where the finest block size is 2×2. It is carried out for the segmented objects only and the effective search range is ± 14 pixels. Hierarchical motion estimation can capture true object motion better at a reduced complexity compared to straightforward block-matching motion estimation [10]. Nevertheless, because we have used a small final block size for added accuracy in the estimated motion, the complexity of motion estimation becomes relatively high and it amounts to roughly 97% of full-search motion estimation with an equal search range (but only over the areas where motion estimation is conducted). Since subsequent backward validation and video object extraction will adjust the object shape, we can tolerate some inaccuracy in forward tracking. This can be used to reduce the complexity of motion estimation, but it is not pursued in the

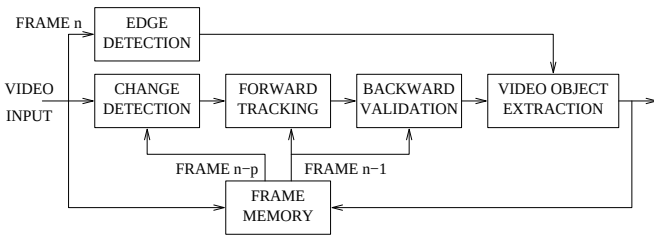


Fig. 1. Proposed structure of video extraction and tracking algorithm.

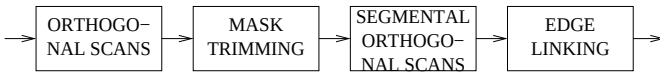


Fig. 2. Procedure of video object extraction.

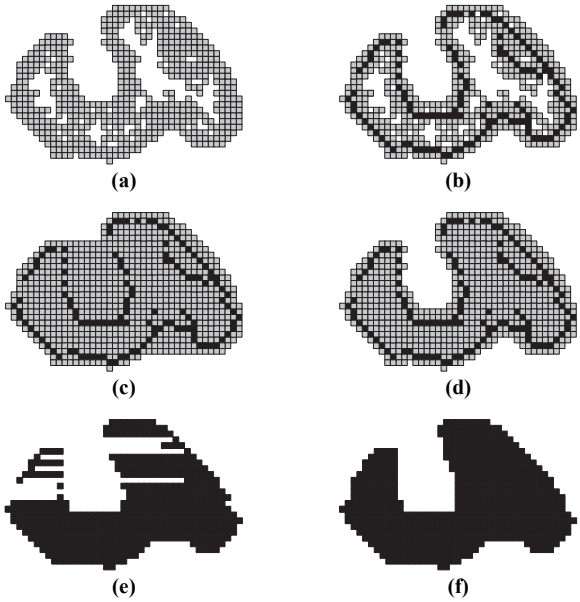


Fig. 3. Arbitrary example for illustration of the first three steps in video object extraction. (a) Backward-validated object mask *M*. (b) Same as (a) with edge pixels marked in black. (c) M_o, the mask after orthogonal scans. (d) M_t, the mask after mask trimming. (e) After segmental horizontal scan. (f) M_s, the mask after segmental vertical scan.

present work.

"Backward validation" projects (after motion compensation) those pixels in the change detection mask but not in the forward-tracked object footprint into the previous frame and verifies if they are part of the segmented object. If not, they are deleted. This requires backward motion estimation, which is also done using the above-described hierarchical motion estimation method. But the complexity is low because it is needed only for a small portion of the pixels in a frame.

B. Video Object Extraction

After motion tracking, the resulting pixel mask for an object may not match the object accurately. For example, it may contain some holes in the interior and its boundary may contain cracks or bulges. "Video object extraction" attempts to rectify these problems by edge-based morphological processing. The process involves four steps as illustrated in Fig. 2, where the purpose of the first three steps are to attain a suitably tight sketch of the object mask so that the last step (edge linking) can obtain the final, precise object mask efficiently.

We illustrate the steps using an arbitrary example of the backward-validated object mask shown in Fig. 3(a) (the gray pixels). For convenience, denote the mask *M*. Let the edge pixels (obtained by the Canny edge detector) in *M* be as shown in black in Fig. 3(b). We assume that the edge pixels in *M* and near its outer perimeter are the ones that define the

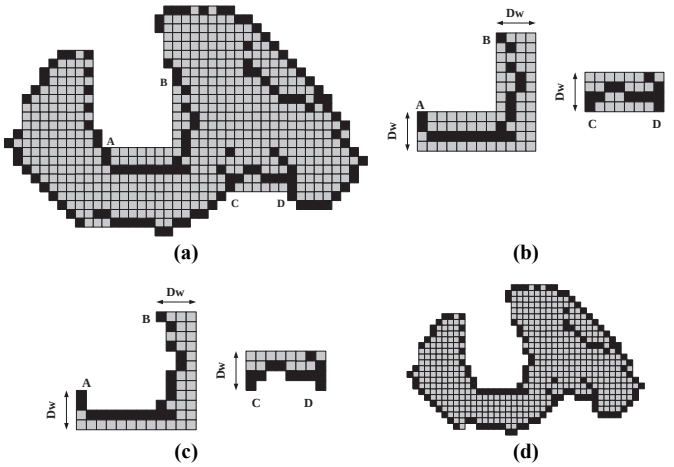


Fig. 4. Illustration of edge linking method. (a) M_s, the mask after segmental vertical scan. (b) Search windows for gap regions AB and CD, respectively, at search depth *Dw* = 4. (c) Edge search results for gaps AB and CD, respectively. (d) Final object mask after edge linking.

object boundary. The problem is to identify these pixels and bridge all the gaps.

For this, we first stop the "holes" inside *M* through the "orthogonal scans" step. In this step, similar to [11], we first conduct a "horizontal scan" over each row of *M* to fill in the space between the leftmost and the rightmost pixels. Then the result is subject to a "vertical scan" that fills in the space between the topmost and the bottommost pixels of each column. For the arbitrary example, we obtain Fig. 3(c) as the result. Denote it M_o for convenience.

As can be seen, holes inside *M* are stopped in M_o, but the mask may also be significantly expanded. Thus in the "mask trimming" step, we trim the overgrowth by eroding M_o from the outer side inwards, deleting every pixel that is not an eight-connected neighbor of the original backward-validated mask *M*. This can patch up any one- and two-pixel-wide "cracks" that may remain in M_o, but will also leave a one-pixel-wide "coating" around M_o that is outside the original perimeter of *M*. Therefore, we conduct the erosion for one more time, trimming away any pixel on the outer boundary of the remaining M_o that does not belong to *M*. The result for the arbitrary example is shown in Fig. 3(d). Denote it M_t for convenience.

The next step, "segmental orthogonal scans," tries to tighten up the mask further for the benefit of the last step, "edge linking." It first examines each connected horizontal line segment in M_t and fills in the space between the two farthest edge pixels (see Fig. 3(e) for the arbitrary example). Then a similar operation is conducted in the vertical direction, but this time regarding pixels in the horizontal result as equivalent to edge pixels. Denote the final result M_s for convenience. Fig. 3(f) illustrates this result for the arbitrary example.

After the foregoing steps, we have now obtained an object mask that is solid inside and relatively tight around the (assumed) object boundary edges on the outer side. This can be seen from Fig. 4(a), which redraws M_s in gray and black

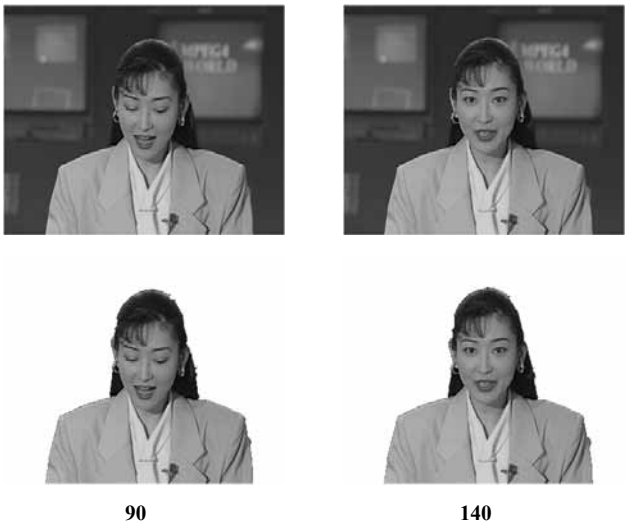


Fig. 5. Segmentation result of Akiyo. Top row: original frames; bottom row: extracted moving object. Numbers below are frame numbers.

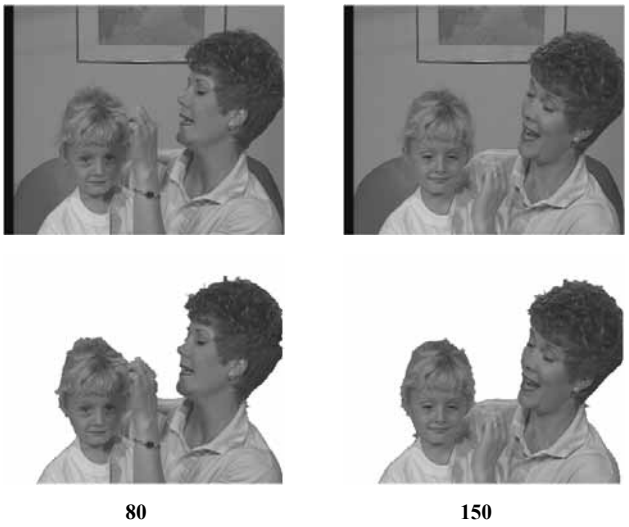


Fig. 7. Segmentation result of Mother and Daughter. Top row: original frames; bottom row: extracted moving object. Numbers below are frame numbers.

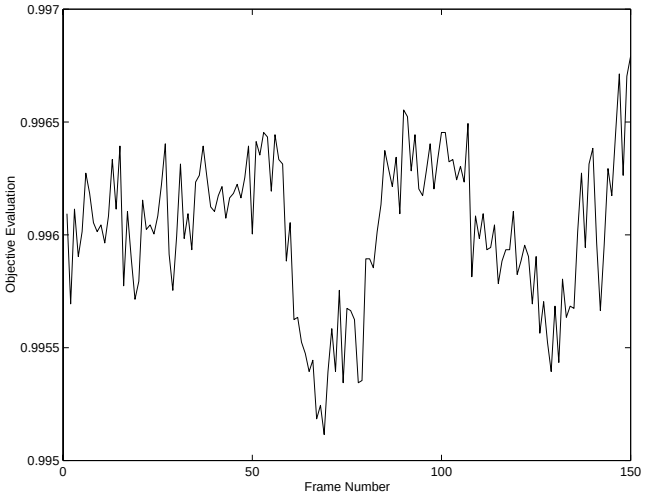


Fig. 6. Fractional agreement between segmentation and reference mask for Akiyo sequence.

(the latter denoting edge pixels). We now perform the last step of work, "edge linking," to finalize the boundary of the extracted object. For this we resort to the well-recognized Dijkstra shortest-path algorithm [12] to bridge all the gaps, including the "apparent gaps" AB and CD.

In application of the Dijkstra algorithm, each gap is treated separately. An edge pixel in M_s is assigned a distance d_0 and each nonedge pixel a distance d_1. (We set $d_0 = 1$ and $d_1 = 10$ in the experiments reported below.) A suitable search depth into M_s (denoted *Dw*) is chosen. Fig. 4(b) illustrates the edge search windows for gaps AB and CD under *Dw* = 4. Since the complexity of the Dijkstra algorithm is typically $O(n^2)$ where *n* is the number of pixels in the search window, the smaller the window, the higher the algorithm efficiency. Because we have suitably tightened the outer perimeter of M_s around the (assumed) object boundary edges, the search depth *Dw* can be quite small (experiments show that $Dw \le 5$ is enough) and yet we can find most of the desired edges. Fig. 4(c) shows the edge search result for gaps AB and CD, and Fig. 4(d) the final object mask obtained by video object extraction.

C. Experimental Results

We show some results from segmenting the Akiyo and the Mother-and-Daughter sequences in CIF format (352×288). Fig. 5 shows the results for Akiyo at *Dw* = 5. Fig. 6 shows the fractional agreement of the segmented object mask compared to a reference mask, calculated as proposed in [13]. The agreement is always above 0.995. Fig. 7 shows the results for Mother and Daughter, also at *Dw* = 5. The identified object boundaries are rather accurate.

III. APPLICATION IN SCENE COMPOSITION

A. Method for Object Enlargement and Shrinkage

Consider scene composition using the extracted video objects. This is expected to require often enlargement or shrinkage of the video objects. For convenience, we consider employing the interpolation and decimation filters specified in MPEG-2, which are [-12, 140, 140, -12]/256 for interpolation and [-29, 0, 88, 138, 88, 0, -29]/256 for decimation. The frequency responses of these filters are shown in Fig. 8.

Several modifications are necessary, however. First, MPEG-2 only considers enlargement and shrinkage by a factor of two (both horizontally and vertically), but we consider arbitrary-factor interpolation and decimation. And second, MPEG-2 only considers interpolation and decimation for rectangular video frames, whereas we have arbitrarily shaped video objects.

Consider enlargement by arbitrary integer factors first. We consider a simple interpolation method illustrated by example in Fig. 9. In essence, enlargement by an integer-power-of-2

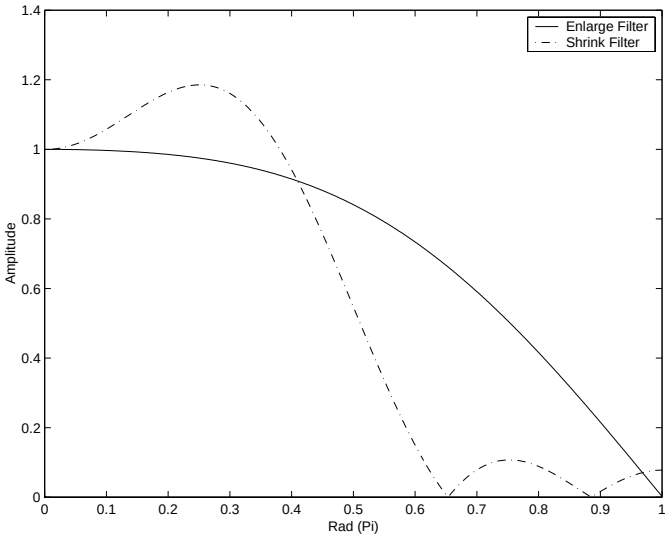


Fig. 8. Amplitude responses of the MPEG-2 enlargement and shrinkage filters.

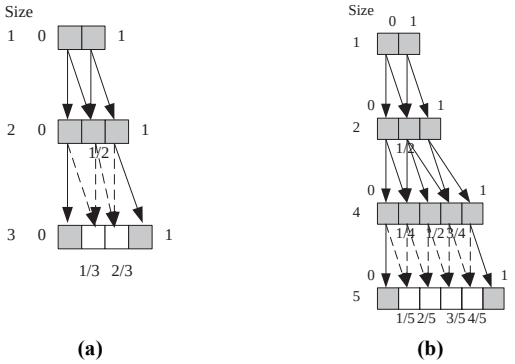


Fig. 9. Proposed enlargement algorithm, illustrated by example. (a) Enlargement by two or three times. (b) Enlargement by two, four, or five times. (Gray squares denote pixels in the original object mask or as obtained by interpolation using the MPEG-2 filter; white squares denote linearly interpolated pixels. Dashed lines indicate linear interpolation between nearest pixels.)

(say 2^n) times is accomplished by repeated application of the MPEG-2 interpolation filter n times. Enlargement by a factor that is not an integer power of 2 is accomplished by enlarging to the nearest lower integer power of 2, followed by linear interpolation between two nearest pixels according to the relative pixel distances. Shrinkage by arbitrary integer factors operates on a similar principle and is illustrated also by an example in Fig. 10. Again, shrinkage by an integer-power-of-2 times is accomplished by repeated application of the MPEG-2 decimation filter. Shrinkage by a factor that is not an integer power of 2 is accomplished by shrinking to the nearest lower integer power of 2, followed by linear interpolation between two nearest pixels according to the relative pixel distances.

A rational-factor (say, q/p where p and q are coprime integers) enlargement or shrinkage can be obtained as follows. Find the nearest lower powers of 2 for p and q, say, 2^m and 2^n, respectively. Enlarge or shrink the object by 2^{n-m} times.

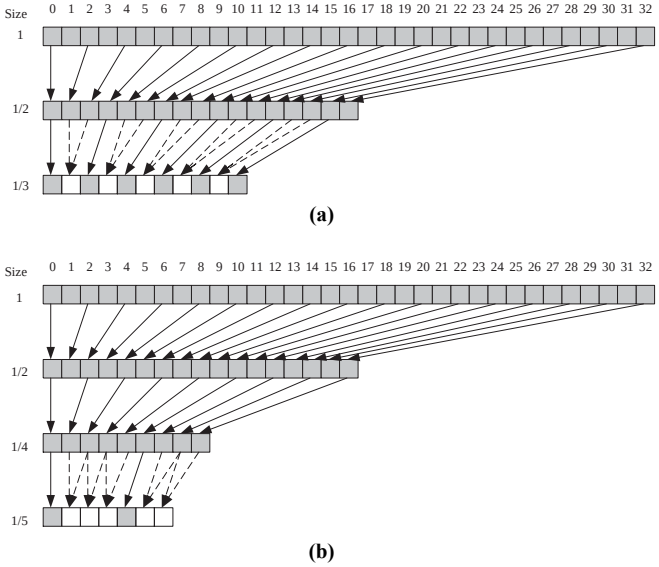


Fig. 10. Proposed shrinkage algorithm, illustrated by example. (a) Shrinkage by two or three times. (b) Shrinkage by two, four, or five times. (Gray squares denote pixels in the original object mask or as obtained by decimation using the MPEG-2 filter; white squares denote linearly interpolated pixels. Dashed lines indicate linear interpolation between nearest pixels.)

Then perform linear interpolation between nearest pixels according to the factor q/p in a similar way as that for integer-factor enlargement or shrinkage.

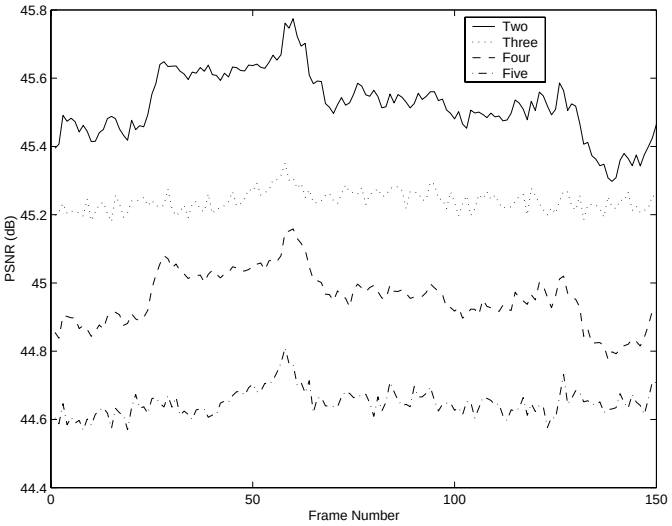
In either enlargement or shrinkage, when the filter memory extends beyond the object mask (this happens when operating on pixels near the object boundary), we repeat the boundary pixel values for filtering purpose, similar to the principle used in H.263 and MPEG-4 for motion estimation.

Note also that the way object enlargement and shrinkage are conducted fits well into a context of spatial-scalable coding.

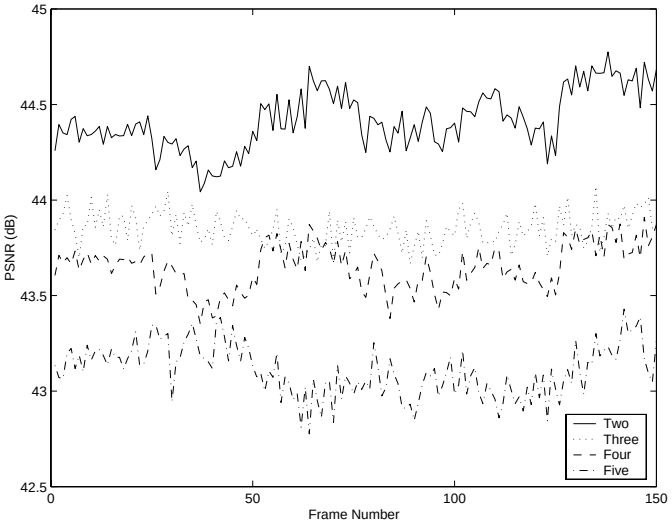
B. Experimental Results

Since we deal with arbitrarily shaped objects, the interpolation and decimation performance at object boundaries is of particular interest. In Fig. 11 we show some results on the similarity (in PSNR) between the original segmented object and the resulting one that is enlarged X times and shrunk to the original size, where X is between 2 and 5. As expected, the interior pixels (inside a three-pixel-wide band at object boundary) are subject to less distortion from enlargement and shrinkage compared to pixels near the object boundary. But overall, the PSNR is reasonably high.

Fig. 12 shows two synthesized scenes (two frames each) obtained by superimposing the moving objects segmented (and shrunk by two times) from the Claire, the Akiyo, and the Salesman sequences onto some background and middle ground, and with an additional foreground overlaid on them. The synthesized scenes appear natural.



(a)



(b)

Fig. 11. Similarity, in PSNR, between the original segmented video object in CIF Mother-and-Daughter sequence and that enlarged *X* times and shrunk to original size, where *X* is between 2 and 5. (a) For interior of object (inside a three-pixel-wide band at object boundary). (b) For the three-pixel-wide band at object boundary.

IV. CONCLUSION

We proposed an automatic video segmentation (or video object extraction) algorithm for content-based video applications. The algorithm employed edge analysis for accurate determination of object boundaries and it employed bidirectional motion estimation for robust tracking of object motion and shape change. The algorithm had been designed with computational complexity in mind, but further reduction in computational complexity is still highly desirable, especially in motion estimation.

We considered using the segmented video objects in scene composition. For this we discussed a way to enlarge or shrink arbitrarily shaped video objects. The method worked in a spatial-scalable manner. The synthesized scenes appeared rather natural.

Fig. 12. Synthesized scenes using segmented (and shrunk by two times) video objects and other video contents.

REFERENCES

[1] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 539-546, Sep. 1998.

[2] S.-Y. Chien, Y.-W. Huang, and L.-G. Chen, "Predictive watershed: a fast watershed algorithm for video segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 5, pp. 453-461, May 2003.

[3] S. Sun, D. R. Haynor, and Y. Kin, "Semiautomatic video object segmentation using VSnakes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 1, pp. 75-82, Jan. 2003.

[4] A.-R. Mansouri and J. Konrad, "Multiple motion segmentation with level sets," *IEEE Trans. Image Processing*, vol. 12, no. 2, pp. 201-220, Feb. 2003.

[5] T. Meier and K. N. Ngan, "Video segmentation for content-based coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1190-1203, Dec. 1999.

[6] C. Kim and J. N. Hwang, "Fast and automatic video object segmentation and tracking for content-based applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 2, pp. 122-129, Feb. 2002.

[7] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679-698, Nov. 1986.

[8] M. Kim and J. Kim, "Moving video object segmentation using statistical hypothesis testing," *Electron. Lett*., vol. 36, no. 2, pp. 128-129, Jan. 2000.

[9] M. Kim, J. G. Choi, D. Kim, H. Lee, M. H. Lee, C. Ahn, and Y. S. Ho, "A VOP generation tool: automatic segmentation of moving objects in image sequences based on spatio-temporal information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1216-1226, Dec. 1999.

[10] M. Bierling, "Displacement estimation by hierarchical blocking," *SPIE* vol. 1001, *Visual Commun. Image Processing*, pp. 387-403, 1986.

[11] T. Meier and K. N. Ngan, "Automatic segmentation of moving objects for video object plane generation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 525-538, Sep. 1998.

[12] D. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269-271, 1959..

[13] M. Wollborn and R. Mech, "Refined procedure for objective evaluation of VOP generation algorithms," Doc. ISO/IEC JTC1/SC29/WG11 MPEG98/3448, Mar. 1998.

REAL-TIME VIDEO SEGMENTATION BASED ON BACKGROUND MODELING APPLICABLE TO VIDEOCONFERENCING

Yueh-Hsien Lin and David W. Lin

Dept. of Electronics Engineering and Center for Telecommunications Research
National Chiao Tung University
Hsinchu, Taiwan 30010, R.O.C.
E-mails: yhlin.ee91g@nctu.edu.tw, dwlin@mail.nctu.edu.tw

ABSTRACT

We design a video segmentation algorithm and conduct a real-time implementation on a personal computer (PC). The algorithm is based on background subtraction. To facilitate the setting of some thresholds used in the algorithm, we develop a novel method to estimate the camera noise. The algorithm builds a stationary background buffer by considering short-term background and temporary foreground masks. When camera motion occurs, we recover the stationary background from panorama background buffer through global motion estimation. Simulation results show that the proposed techniques perform reasonably well. The real-time PC-software implementation employs a graphical user interface. With a P-4 2.4-GHz CPU and 512-MB RAM, the current un-optimized implementation yields a speed of about 5 frames per second (fps) for CIF-size video when the camera is still. In presence of camera motion, the speed is about 1.7 fps.

1. INTRODUCTION

We consider the design and implementation of a video segmentation algorithm on a personal computer (PC). The intended application is to support PC-based multipoint videoconferencing.

Our segmentation algorithm is based on the "background subtraction" approach. One way to obtain the background image is to find and remove all moving objects, but in many situations this may not be easy. In our method, the background is obtained by gathering the stationary regions. Because flat inner regions in moving objects may be mistaken to be background, we use a temporary object mask to alleviate this problem. In the event of camera motion, we need to rebuild a new background. In order to reduce the rebuilding delay, we use the idea of mosaic ("sprite") to salvage most of the existing background.

This work was supported by the National Science Council of R.O.C. under grant no. NSC 92-2219-E-009-009.

In deciding whether a pixel should be considered background, some thresholds are needed which should be set taking camera noise into account. We thus develop a novel two-stage method for camera noise estimation.

In what follows, Section 2 describes the proposed segmentation method. Section 3 describes the PC-based implementation. Section 4 presents some simulation results. And finally, Section 5 contains the conclusion.

2. PROPOSED SEGMENTATION METHOD

Figure 1 shows a block diagram of the proposed segmentation method. We explain its functioning in the following subsections.

2.1. Two-Stage Noise Estimation

To facilitate the choice of the many noise-dependent thresholds needed in the segmentation algorithm, we develop a two-stage method for accurate camera noise estimation.

Assume that the camera noise obeys zero-mean Gaussian distribution. Since the camera noise is uncorrelated between frames, the interframe difference of a stationary pixel obeys zero-mean Gaussian distribution with variance σ^2 that is twice the variance of the camera noise. This is similar to [1] and other works.

To estimate σ^2, we should use the pixels belonging to stationary background and exclude those belonging to moving objects. In our experience (and as is intuitively reasonable), the pixels with larger interframe differences usually form a cluster when they belong to moving objects. The larger frame differences caused by camera noise are usually randomly distributed. Therefore, similar to [2], we check for existence of directional structure in the interframe difference at each pixel to detect pixels belonging to moving objects. Specifically, for each pixel, we calculate the four directional sums in frame difference map using the masks shown in Fig. 2. If one of them is larger than a certain

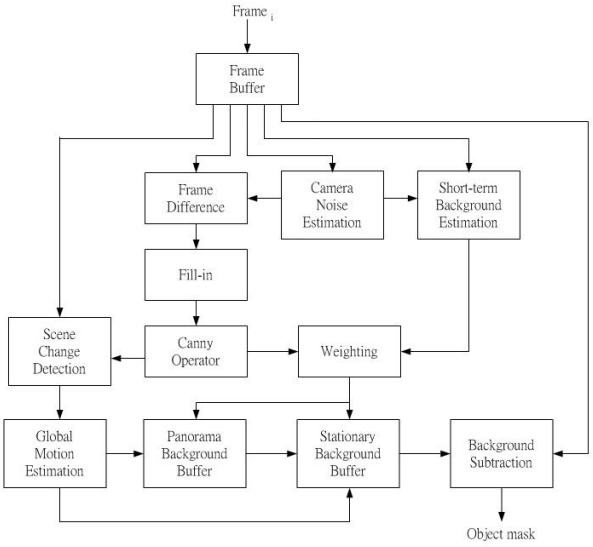


Fig. 1. Block diagram of the proposed segmentation method.

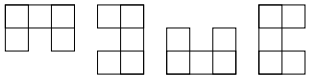


Fig. 2. Four masks for directional sums.

threshold, we assume that the pixel belongs to a moving object.

Now we face a problem: how can we set the threshold without knowing the amount of camera noise in the first place? To solve this problem, we calculate the interframe pixel variance σ_G^2 over the entire frame and use it to set an initial threshold $\alpha\sigma_G^2$ to test the directional sums. After we remove the pixels with large directional interframe differences, the remaining pixels are used to compute a second interframe variance σ_1^2. Then we use $\beta\sigma_1^2$ as the threshold to classify the pixels and obtain the final estimate for σ^2.

To verify the validity of our method, we first estimate the true variance σ^2 from manually chosen pixels in video

Fig. 3. Image areas used to estimate the true camera noise variance for (a) Mother-and-Daughter and (b) Claire sequences.

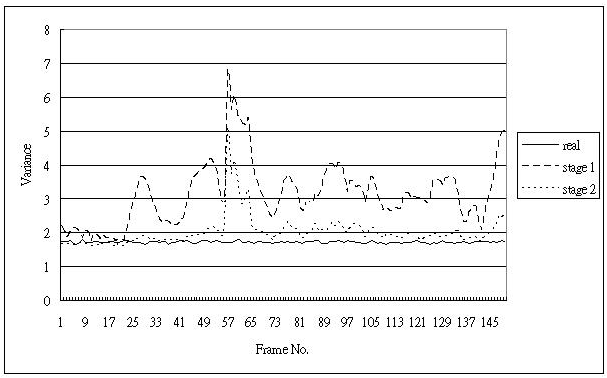


Fig. 4. Noise estimation for Mother-and-Daughter sequence.

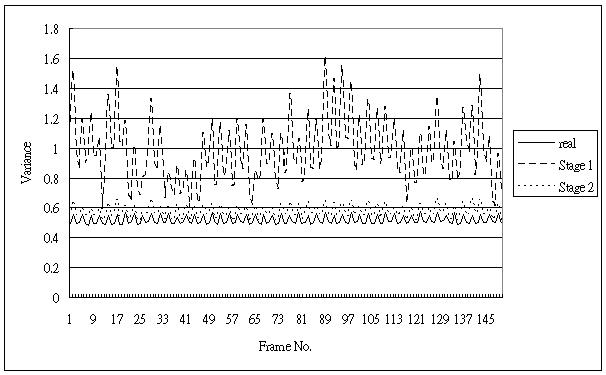


Fig. 5. Noise estimation for Claire sequence.

frames that belong to the stationary background. For example, the white areas in Fig. 3 mark the pixels used to estimate σ^2 for the two images. Figures 4 and 5 (curves labeled "Stage 1") show that using $\alpha\sigma_G^2$ we can remove most pixels belonging to moving objects, but not completely satisfactorily. The result of stage 2 is closer to the exact value.

2.2. Temporary Foreground Mask

We now generate a temporary foreground mask to be used in obtaining the stationary background buffer, detecting scene change, and global motion estimation.

2.2.1. Get Initial Object Mask

At first, we use the 3×3 window to calculate the mean-square frame difference at each pixel. If the result is larger than a threshold $\gamma\sigma^2$, then the pixel is classifed as belonging to a moving object. An example of the thresholded frame difference map is shown in Fig. 6(a). Next, we use the fill-in technique proposed in [3] to get a rough mask. For this the pixels between the first and last white pixels (indicating pixels in moving objects) in each row is made white. Then

Fig. 6. (a) Thresholded frame difference map. (b) Fill-in for each row. (c) Fill-in for each column. (d) Second fill-in for each row.

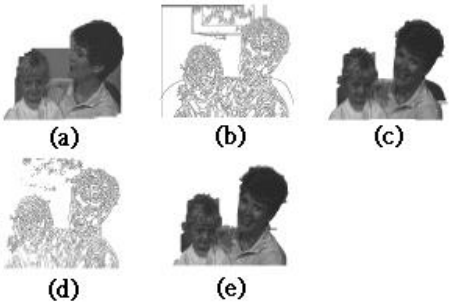


Fig. 7. (a) Initial object mask. (b) Edge map. (c) Refined mask. (d) Edge map after removing background edges. (e) Final object mask.

this is doen for each column and once again for each row. The step-by-step results are shown in Fig. 6.

2.2.2. Refine Initial Object Mask

A rough mask obtained above is enough in some cases but not in others. In Fig. 7, for instance, there are two persons sitting side by side. The background between the two persons is included in the mask by the fill-in process. Hence we use edge information to refine the initial mask, where the edge information is obtained using the Canny operator [4].

The Canny operator performs a gradient operation on the image that has been convolved with a Gaussian filter. Then nonmaximum suppression is applied to thin the edges. Lastly, thresholding with hysteresis is used to find and link edges. The edge map after applying Canny operator is shown in Fig. 7(b). The code for it is obtained from [5].

We refine the initial object mask by shrinking the initial mask to fit the edge map. Figure 7(c) shows the shrunk mask for Fig. 7(a). The example shows that the edge map may include many background edges and these edges may impact the result adversely. To reduce their influence, we use a buffer to store the background edges. When an edge always appears at a certain position, we assume that it is a background edge. The edge map after removing background edges is shown in Fig. 7(d) and the resulting object mask is shown in Fig. 7(e). Comparing Figs. 7(c) and (e), we see that the overgrowth due to background edges can be effectively removed.

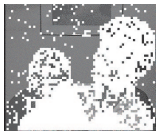


Fig. 8. Result of short-term background estimation.

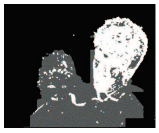


Fig. 9. Weighting mask for the Mother-and-Daughter sequence.

2.3. Short-Term Background Estimation

We consider six consecutive frames $f_k(i)$ $(1 \leq k \leq 6)$ and calculate the frame differences $d_m(i) = f_6(i) - f_m(i)$ $(1 \leq m \leq 5)$ at each pixel i. For every pixel i, we calculate the variance of $d_m(i)$ $(1 \leq m \leq 5)$ in a 3×3 window. If the variance is smaller than a threshold given by $\lambda\sigma^2$, then we consider the changes in the six frames small and regard pixel i in the sixth frame as background. The result is shown in Fig. 8 for the earlier example.

2.4. Construct Stationary Background Buffer

Most wrong decisions in short term are due to flat inner object regions because they do not show significant variations across frames. Such effects can be seen from Fig. 8. To reduce this problem, we use the temporary foreground mask to weigh every pixel before putting the short-term background into the final background buffer.

A weighting mask for the earlier example is shown in Fig. 9, where black pixels indicate reliable background pixels and are given higher weight while white pixels indicate moving objects and are given zero weight. A pixel marked gray is one in the short-term background and also in the temporary foreground mask. It may suffer from the flat inner region problem and we give it a lower weight. We accumulate the weights at every pixel and the short-term background is put into the stationary background buffer when the accumulated weight meets a threshold.

2.5. Deal with Camera Motion

In background subtraction, the background should ideally be stationary. If the camera moves, the background buffer should be rebuilt. Typically, there may be a large overlap between the old and the new backgrounds. Thus we employ the image mosaic technique to make use of the overlap and speed up background reconstruction.

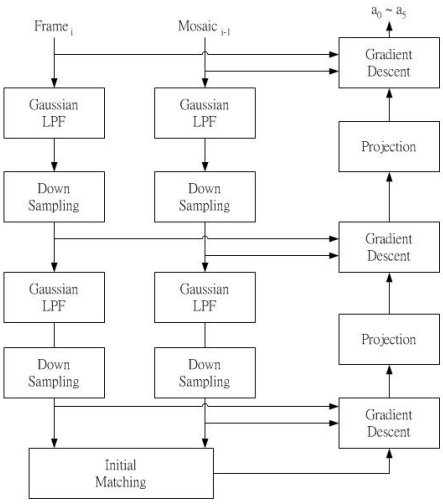


Fig. 10. Global motion estimation.

2.5.1. Scene Change Detection

We use scene change detection to detect camera motion. When the difference between background frames at different times is large, we assume that camera motion has occured. The background here is obtained by excluding the temporary foreground mask. Since a flat region may yield no frame difference in small camera motion, we only consider the regions near edges.

2.5.2. Global Motion Estimation

Figure 10 illustrates the method we use to estimate the global motion due to camera motion, which is based on the hierarchical architecture of [6] and [7]. The advantage of a hierarchical architecture is that it can handle large displacements and reduce computational complexity.

We minimize the sum-squared difference E between the current frame I and the displaced reference frame I':

$$E = \sum_{i=1}^{N} e_i^2$$

where $e_i = I'(x_i', y_i') - I(x_i, y_i)$. Considering both performance and simplicity, we adopt the affine motion model:

$$x_i' = a_0 + a_2 x_i + a_3 y_i, \;\; y_i' = a_1 + a_4 x_i + a_5 y_i,$$

where $a_0, \ldots, a_5$ are the motion parameters. The gradient descent method [6], [8] needs a set of initial values for a_k. We use stepped search to obtain the initial a_0 and a_1. The search range is ± 15 in both coordinates and therefore the range in full size is ± 60. The others a_k's are set to $a_2 = a_5 = 1$ and $a_3 = a_4 = 0$.

The gradient descent iterations are carried out according to

$$\mathbf{a}^{t+1} = \mathbf{a}^t + \mathbf{H}^{-1}\mathbf{b}$$

where $\mathbf{a}^t$ denotes $\mathbf{a}$ at iteration t, $\mathbf{H}$ is an 6×6 matrix equal to one-half times the Hessian of E:

$$H_{kl} = \sum_{i=1}^{N} \frac{\partial e_i}{\partial a_k} \frac{\partial e_i}{\partial a_l},$$

and $\mathbf{b}$ is an 6-vector equal to minus one-half times the gradient of E:

$$b_k = -\sum_{i=1}^{N} e_i \frac{\partial e_i}{\partial a_k}.$$

Besides, in the first iteration of each level, the histogram of $|e_i|$ is computed to find a threshold T such that the number of $|e_i|$ bigger than T is about 15% of considered pixels. In the following iterations, the pixels whose $|e_i|$ are larger than T are excluded in gradient descent operation. By observing the speed of convergence for the Stefan sequence and the related results in [6], we set the maximum number of iterations at each level to 34. The transform between (x_i, y_i) and (x_i', y_i') is usually non-integer and therefore bilinear interpolation is used.

The projection of motion parameters from one level to the next is effected by multiplying a_0 and a_1 by two and keeping the others the same.

2.5.3. Panorama Background and Background Recovery

After we have obtained the camera motion, the background can be stored in the panorama background buffer according to the motion parameters. When camera motion occurs, the stationary background buffer can be rebuilt from the panorama background quickly. The bilinear interpolation is adopted to deal with non-integer motion.

2.6. Background Subtraction

The final object mask is obtained by differencing the current frame and the stationary background buffer. In general, the background in the current frame may be subject to lighting change and contain shadows, and the stationary background may contain some wrongly identified background pixels. Therefore, the differencing result may still contain some error regions. We therefore remove the small isolated pixel groups outside and inside the resulting object mask. Figure 11 shows an example.

3. PC-BASED IMPLEMENTATION

The PC-based implementation employs a capturing camera and a PC, where the PC is used for system control, video

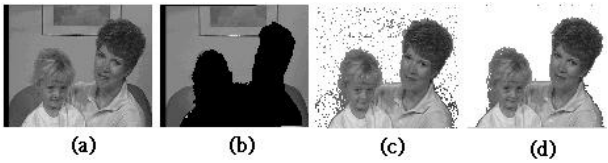


Fig. 11. (a) 255th frame. (b) Stationary background buffer. (c) Mask after subtraction and thresholding. (d) Final object mask.

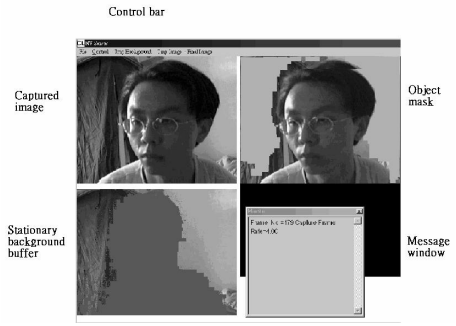


Fig. 12. The application program interface.

segmentation, and result display. We develop a GUI (graphical user interface) employing the Windows SDK (software development kit) from Microsoft.

We use the VfW (Video for Windows) 1.0 library, originally released by Microsoft in November 1992 for the Windows 3.1 operating system, to interface with the camera. The captured frame is embedded in an AVI (Audio-Video Interleaved) file with its beginning marked by "##db".

After the video segmentation, we display the result in a window created through the Windows SDK.

We implement two major control units: capture control and threshold control. The former controls the digital camera, such as start, stop, image size, and luminance. The latter is used to adjust the thresholds in temporary foreground mask, short-term background, and background subtraction.

The application program interface is illustrated in Fig. 12.

With a P-4 2.4-GHz CPU and 512-MB RAM, the current un-optimized implementation yields a speed of about 5 frames per second (fps) for CIF-size (352×288) video when the camera is still. In the presence of camera motion, the speed is about 1.7 fps.

4. SIMULATION RESULTS

4.1. Segmented Image Masks

Figure 13 shows some results of the Mother-and-Daughter, the Claire, and the Akiyo sequences. The required number of frames to obtain enough background for them is observed to be about 260, 150, and 10, respectively. We can see that

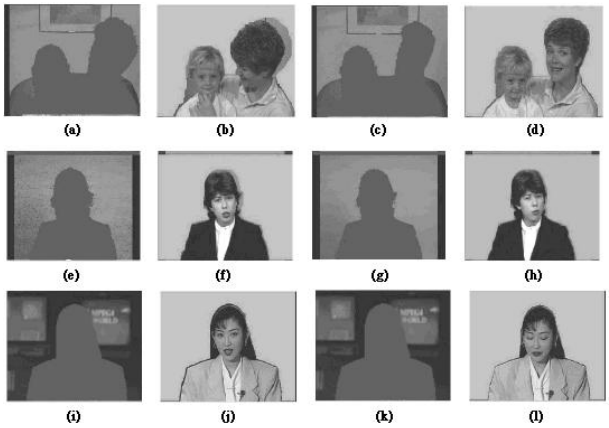


Fig. 13. Some segmentation results of the Mother-and-Daughter ((a)–(d)), the Claire ((e)–(h)), and the Akiyo ((i)–(l)) sequences. Panels alternately show stationary background buffer and segmented object mask. The frame numbers are as follows. (a), (b) 140. (c), (d) 260. (e), (f) 60. (g), (h) 150. (i), (j) 10. (k), (l) 165.

the object masks are more accurate as we obtain more background.

Two major factors influence the required number of frames to gather enough background. First, if the background is covered by moving objects for a long time, then of course we would need to wait for a long time until the background becomes uncovered to gather it. Second, it also depends on the amount of camera noise. For a low camera noise sequence, we can set more critical thresholds for the short-term background and the temporary foreground. This can lead to a shorter time in gathering the stationary background buffer. The amount of camera noise in the three test sequences, in descending order, is Mother-and-Daughter, Claire, and Akiyo. We see that the required number of frames to obtain enough background, in descending order, is also the same for the three sequences.

4.2. Global Motion Estimation and Mosaic

We now examine the effect of global motion estimation and background mosaic. First, Figure 14 shows two panorama background buffers obtained using the Stefan test sequence, for which a reference segmentation exists

Next, we show the benefit of background salvaging using a sequence captured in our lab. Figure 15 shows the segmented image masks for several frames after a camera motion when we just reset the background upon camera motion. If we use background mosaic to salvage existing background, the resulting image masks are as shown in Figure 16. It is obvious that the result with background mosaic is more accurate during the rebuilding of new background.

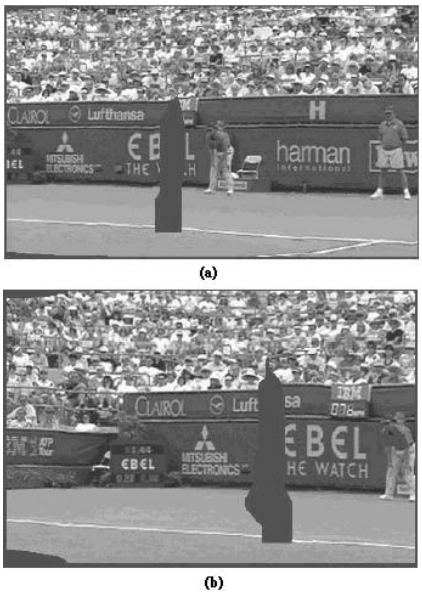


Fig. 14. Mosais results for the Stefan sequence. (a) From 1st to 13th frames. (b) From 40th to 73th frames.

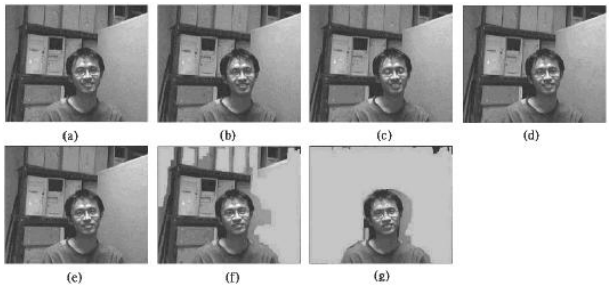


Fig. 15. Segmented image masks without background mosaic. Frame numbers are: (a) 145 (where camera motion is detected), (b) 146, (c) 147, (d) 148, (e) 149, (f) 150, and (g) 151.

5. CONCLUSION

We developed a video segmentation algorithm based on the background subtraction approach and implemented a real-time video segmentation system on PC based on the algorithm. The intended application was PC-based multipoint videoconferencing.

The algorithm used a temporary foreground mask to reduce the influence in background construction of the inner flat regions in the moving objects. It also used a panorama background buffer (mosaic or sprite) to improve the segmentation accuracy immediately after camera motion. A two-stage method for camera noise estimation was developed to facilitate the setting of various algorithm thresholds. Simulation results show that the algorithm performs relatively well, although further improvements are still desirable.

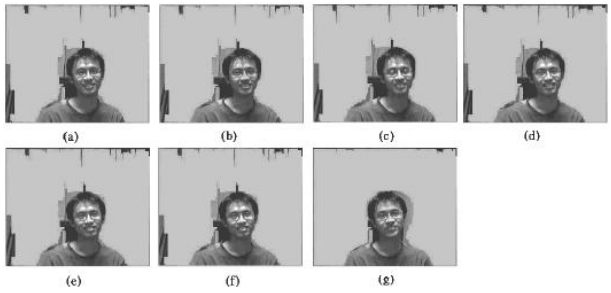


Fig. 16. Segmented image masks with background mosaic. Frame numbers are as in previous figure.

With a P-4 2.4-GHz CPU and 512-MB RAM, the current un-optimized implementation yields a speed of about 5 fps for CIF-size video when the camera is still. In the presence of camera motion, the speed is about 1.7 fps.

6. REFERENCES

[1] T. Aach, A. Kaup ,and R. Mester, "Statistical model-based change detection in moving video," *Signal Processing*, vol. 31, pp. 165–180, Mar. 1993.

[2] Y. H. Jan and D. W. Lin, "Video segmentation with extraction of overlaid objects via multi-tier spatio-temporal analysis," *Int. J. Elec. Eng.*, vol. 11, no. 3, Aug. 2004.

[3] T. Meier and K. N. Ngan, "Video segmentation for content-based coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 525–538, Dec. 1999.

[4] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 679–698, Nov. 1986.

[5] "Canny operator code," http://ouray.cudenver.edu/na0alber/DataCompressionPaper.htm.

[6] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Trans. Image Processing*, vol. 9, pp. 497–501, Mar. 2000.

[7] Y. Lu, W. Gao, and F. Wu, "Fast and robust sprite generation for MPEG-4 video coding," in *Proc. IEEE Pacific-Rim Conf. Multimedia*, Oct. 2001, pp. 118–125.

[8] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, 1992.

附件：出席國際學術會議報告(含論文)

報告人：林大衛 國立交通大學電子工程學系 教授

國科會計畫編號：NSC 92-2219-E-009-009

壹、前言

在本計畫及交大電子工程系補助之下，本人今年共參加二項國際學術會議如下。其中第一項會議之支出由本計畫補助，而第二項會議支出則由本計畫補助小部分，交大電子工程系補助大部分。因此，本報告主體亦分為二部分，分別報告二項會議的出席經過及心得等。兩項會議共發表三篇論文，其中二篇係國科會計畫成果，不過不是本計畫，而是另一個計畫。由於出國經費係核定於本計畫之下，故由其負擔出國經費。

1. **會議名稱：**

(中文) 2004 年 IEEE 國際聲學、語音、與訊號處理會議

(英文) 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)

時間及地點：2004 年 5 月 17-21 日 加拿大魁北克蒙特婁

發表論文題目：

(1) (中文) 使用邊界均方差估計以變化步階之高階 QAM 之多模盲目判定回授等化器 (論文全文見 Appendix A)

(英文) Variable-Step-Size Multimodulus Blind Decision-Feedback Equalization for High-Order QAM Based on Boundary MSE Estimation

(2) (中文) 使用平行干擾消除接收器之片碼間插寬頻分碼多重進接在多路徑瑞利衰落通道中之效能 (國科會計畫 NSC 92-2219-E-009-018 成果) (論文全文見 Appendix B)

(英文) Chip-Interleaved WCDMA with Parallel-Interference-Cancellation Receiver in Multipath Rayleigh Fading Channels

2. **會議名稱：**

(中文) 2004 年 IEEE 無線通訊領域之訊號處理進展國際研討會

(英文) 2004 IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2004)

時間及地點：2004 年 7 月 11-14 日 葡萄牙里斯本

發表論文題目：

(中文) 一個可消除多用戶干擾並降低符際干擾而可用低複雜度之接收處理的片碼間插同步直接序列分碼多重進接技術 (國科會計畫 NSC 92-2219-E-009-018 成果) (論文全文見 Appendix C)

(英文) A chip-interleaved synchronous DS-CDMA technique enabling MAI-free and reduced-ISI transmission with low complexity receiving

貳、第一部分：ICASSP 2004

一、參加會議經過

IEEE Signal Processing Society 每年一度的 International Conference on Acoustics, Speech, and Signal Processing (ICASSP)自 1976 年起舉辦，迄今年已是第 29 屆。是世界上最主要的大型國際訊號處理學術研討會之一，素受訊號處理學術界之重視。今年投稿之論文達 2434 篇，獲得接受者 1261 篇。由於去年原定在香港舉行的 ICASSP 2003 因 SARS 之故臨時取消，所以今年的會議中也收錄了一些原定去年發表的論文。我與學生合撰之論文二篇獲得接受於會中發表，分別是 "Variable-Step-Size Multimodulus Blind Decision-Feedback Equalization for High-Order QAM Based on Boundary MSE Estimation" (與洪崑健同學及工研院電通所柯俊男先生合撰)及 "Chip-Interleaved WCDMA with Parallel-Interference-Cancellation Receiver in Multipath Rayleigh Fading Channels" (與林郁男同學合撰)。

本次會議係於 2004 年 5 月 17-21 日在加拿大蒙特婁舉辦，其中第一天(5 月 17 日)主要是一些短期課程(tutorials)，報名費另收，我未參加。主要會議(technical sessions)在後四天(5 月 18-21 日)，除四場 plenary talks(含一場 keynote speech)之外，共分 13 個 parallel sessions 同時進行，其中包括 5 個 oral sessions 及 8 個 poster sessions，而 poster sessions 的論文數目通常又較 oral sessions 為多，故大部分的論文係以 poster 方式發表。在 technical sessions 中，與我的研究領域比較直接相關的是視訊處理與傳輸訊號處理等方面的主題。

我於 5 月 17 日自中正機場出發，經溫哥華轉蒙特婁，在 18 日上午 7 點多抵達，單程約 20 小時。回程係於 5 月 21 日晚上自蒙特婁機場出發，經溫哥華轉機返台，在 23 日早上 6 時左右抵達，單程約 22 小時。除聆聽 plenary talks 及參加 technical sessions 之外，由於我前幾年擔任 IEEE Taipei Signal Processing Chapter 主席，故也代表該 chapter 參加了 chapter chairs 的午餐會，了解有關 chapters 的事宜，並與其他 chapters 的代表有點交流。

參加本會議之差旅支出全由本計畫補助。

二、與會心得

1. Plenary Talks

Prof. Nikil Jayant (GeorgiaTech) 主講 "Pervasive Broadband: Opportunities for Signal Processing"。通訊網路正在由過去的電話時代進入多媒體寬頻通訊的時代(這就是標題中的 broadband 一字之所指)。本演講主要目的是論及此發展中有關訊號處理的議題與應用。Prof. Jayant 在訊號處理領域世界知名。他早期(1970 年代)任職 Bell Laboratories，並在語音編碼方面著有貢獻，是在 1998 年起轉任職於 GeorgiaTech。

Dr. Gene Frantz (Texas Instruments) 主講 "Human Speech: The Alpha and Omega of Signal Processing"。二十餘年前，TI 推出了一個稱為 Speak-and-Spell 的玩具，寓教於樂，可幫助兒童們學習英文拼字，很受歡迎。該產品是大量使用語音訊號處理的消費電子產品的早期而成功的例子。Dr. Frantz 在其中扮演重要的角色。本演講回顧過去 25 年來語音訊號處理技術的發展，含語音編碼、語音朗誦(text-to-speech conversion)、語音辨認(speech recognition)等。也談及相關的數位訊號處理硬體(DSP)與積體電路之發展。

Prof. Bill Dally (Stanford University) 主講 “Streaming Signal Processors”。其中比較了一般 DSP、FPGA、與 streaming DSP 的架構，也討論了如何用 streaming DSP 的架構以使用今後數年間縮小尺寸的 IC 元件來達成高速低耗能的訊號處理。

Prof. Metin Akay (Dartmouth College) 主講 “Biocomplexity from System to Neuron and Beyond”。他主要是討論 SIDS (sudden infant deaths syndrome)的研究中所進行的相關分析與訊號處理。SIDS 主要是因為呼吸系統運作失調而致窒息。此講題領域為生物訊號處理，非我所熟悉，但聽了此演講，有開廣見聞之效。

2. Technical Sessions

與我目前研究領域比較直接相關的是視訊處理與傳輸訊號處理方面的主題。

在視訊處理方面，與我研究領域較有關有以下的 technical sessions：

- IMDSP-P1: Video Coding
- IMDSP-P3: Robust and Scalable Coding
- IMDSP-P5: Motion Estimation
- SPCOM-P3: Joint Source/Channel Coding and Quantization
- IMDSP-L1: Image and Video Analysis
- MSP-L1: Multimedia Communication, Networking and Security I
- ITT-L1: Image Processing Applications
- IMDSP-P10: Image Analysis
- IMDSP-L6: Interframe Subband/Wavelet Video Coding
- IMDSP-P11: Video Analysis
- MSP-P2: Multimedia Systems and Applications
- MSP-P3: Multimedia Communication, Networking and Security II

有關的論文主要是與視訊分割(video segmentation)、位元率控制(rate control)、軟體實現等議題相關者。

在傳輸訊號處理方面，與我研究領域較有關有以下的 technical sessions：

- SPCOM-LI: Frequency/Channel Estimation in OFDM
- SS-3: Wireless Communication over Rapidly Time-Varying Channels
- SAM-L2: Blind Source Estimation and Channel Identification
- SPTM-P4: Adaptive Filters I
- SPCOM-P6: OFDM and Multi-Carrier Systems
- SPCOM-L4: (Semi)blind Techniques I
- SPTM-P8: Adaptive Filters II
- SPCOM-P8: Synchronization and Timing Recovery
- SPCOM-P9: (Semi)blind Techniques II
- SPCOM-P11: CDMA and Spread Spectrum Systems
- SPTM-P12: Estimation
- SPCOM-P12: Detection, Estimation, and Demodulation
- SPCOM-P13: DMT and xDSL

有關的論文主要是與 CDMA、OFDM(或 DMT)、equalization、synchronization 等議題相關者。其他還有許多有關 turbo coding、space-time (MIMO) signal processing、ultra wideband (UWB)等議題的論文，以後的研究可能會朝這些項目發展，但目前還不是主要研究方向，故沒有特別注意之。

綜合上述各議題的論文觀察，發現視訊處理方面的論文似較傾向於實用面的探討，而傳輸訊號處理方面的論文則較具理論性。後者中，有些使用 Cramer-Rao bound 等觀念來討論訊號處理方法的極限效能。這促使我複習該觀念，也促使我要提醒博士生們能對 detection and estimation theory 有基本的了解。其他關於細項技術的心得，不在此一一列述。

三、考察參觀活動：無。

四、建議事項：無。

五、攜回資料名稱及內容：會議手冊(主要為議程細節)一份，論文集光碟一份。

叁、第二部分：SPAWC 2004

一、參加會議經過

IEEE International Workshop on Signal Processing Advances in Wireless Communications(簡稱 SPAWC)，原係隔年舉辦，自今年起改為每年舉辦，可旁證本領域受重視之程度與研究成果的豐富。本會議本次為第五屆，共收錄約 130 篇論文，並有 9 場 plenary talks。本會議一個獨特之處，是除了 plenary talks 之外，所有的論文皆以 poster 方式發表。林郁男同學與我合著的論文 “A chip-interleaved synchronous DS-CDMA technique enabling MAI-free and reduced-ISI transmission with low complexity receiving” 獲接受於其中發表。

本會議本次係於 2004 年 7 月 11-14 日在葡萄牙里斯本 Hotel Tivoli Tejo 舉辦。我搭機從國內出發，經曼谷及阿姆斯特丹轉抵里斯本。總計自中正機場起飛至里斯本降落，連轉機時間在內，但不連前後往返機場與候機的時間，共花去 25 小時。回程則耗時較短，但也要約 21 小時。

在本次會議中，除了技術領域上的切磋學習外，見到了一位多年不見的老友，也認識了一些新朋友，日後可有研究交流。

參加本會議之差旅支出由本計畫補助小部分，交大電子工程系補助大部分。

二、與會心得

1. Plenary Talks

Prof. Jose Moura (Carnegie Mellon University) 講 “Sensor networks: a graph based approach”。其中討論 parallel architecture 與 web architecture 兩種架構下的 sensor network 設計與效能，包括 sensor 數目與 quantization accuracy of sensor outputs 之間的取捨、及 suboptimal sensor quantization threshold 之使用與 sensor network performance 之間的關係等。

Prof. Ali H. Sayed (UCLA) 講 “Signal processing and communications challenges in wireless location”。由於手機使用的普遍，而由手機撥出的緊急求助電話(在美國為 911)也大幅增加。但許多時候，手機用者對本身位置沒有確切掌握。因此美國 FCC 在 1996 年對美國行動電話公司提出要求，要在若干年內分階段達成某些精確度的定位功能。除了緊急求助之用外，無線定位還有許多其他的商業應用。Prof. Sayed 介紹了其研究團隊近幾年間在此領域的研究與成果，其精確度相當高，也顯示出其研究團隊能力極可觀。

Prof. Babak Hassibi (CalTech) 講 “Distributed space-time processing in wireless networks”。其中討論三個主題，即 broadcast channel capacity、power efficiency of sensor networks and ad hoc networks、與 diversity gains through distributed space-time coding。所提出的一個最主要的基本理念，是將任意分佈的 wireless transceivers 當作一種天線陣列結構，使其發揮多入多出(MIMO)的陣列天線傳收效能。此基本理念相當獨特而具創意。

Prof. Amos Lapidoth (Swiss Federal Institute of Technology) 講 “The truth about capacity calculations for fading channels”。其中對近年來若干有關 fading channels 之傳輸能量之探討中所用的之假設(主要是 block fading 及 availability of perfect channel estimation)提出質疑，並根據較真實的假設，來計算在高 SNR 時的傳輸能量之 asymptotic values。他並未提出推導的過程，只是提出其數學結果，而這些結果相當顛覆一般的認知。

Prof. Luc Vandendorpe (Universite Catholique de Louvain, Belgium) 講 “Soft information aided parameter estimation”。Turbo coding 是近十年來在訊號傳輸方面的一個重要發展。它可以在 SNR 很低的狀況下運作。但其所需的通道參數，如 carrier phase、timing、及 channel coefficients 能否在這樣的 SNR 之下獲得精確的估計，則成為一個問題。近年來，有人也考慮用類似 turbo coding 的觀念來做這些參數的 iterative estimation。本演講討論其使用 EM (expectation-maximization) algorithm 來做通道參數之 iterative estimation 的方法與效果。正如 turbo coding 的情形，雖然 iterations 之初始條件(initial conditions)會影響最後的結果，且並不能夠保證最後必會收斂到 global optimum，但常常是可以接近理論上最佳的結果，即 Cramer-Rao lower bound。

另外四場 plenary talks 是一個 special session，介紹微分幾何的一些觀念(主要是 manifolds 的觀念)及其近來在訊號處理領域的應用。以下簡單描述之。

Prof. Jonathan H. Manton (University of Melbourne, Australia) 講 “Theory of optimization on manifolds”。Manifold 是任意維度空間中的曲面，具有連續性並具有一些可以用參數方式描述的特性。之所以考慮 manifolds，是因為在訊號處理領域的 optimization，常要處理 subspace 的集合。此一集合在幾何結構上等同於一個 Grassmann manifold，其中每一 subspace 為 manifold 上的一個點。此外，有些 constrained optimization 問題可化成 manifolds 上的 unconstrained optimization 問題來處理。

Prof. Anuj Srivastava (Florida State University) 講 “Bayesian estimation and tracking of dynamic signal subspaces”。Optimization 問題通常是求取一個函數的最小值或最大值，其中常用 iterative 逼近的方式來求取答案。在問題解答中，常用到均值(mean)與 covariance 等統計量。Prof. Srivastava 討論這些統計量在 manifold 上的定義，並介紹在

non-Gaussian 環境中作 Bayesian estimation 時，如何追蹤機率分佈的演變。

Prof. Joao, Xavier (Instituto Superior Tecnico, Portugal) 講 "Statistical bounds: a Riemannian-geometric viewpoint"。在 parameter estimation 問題中，parameter space 的結構常非呈現 Euclidean space 的結構，而是 manifold 曲面的結構。在 Euclidean space 中的 parameter estimation 之精確度之極限，有一些重要的理論結果，如：估計誤差必不小於 Cramer-Rao lower bound、估計誤差之 covariance 必不小於 Fisher information matrix。本演講討論了在 manifold 與 quotient space 中的 estimation 精確度之極限，提出一個稱為 IVLB 的 lower bound，並以 blind channel estimation 為例來做討論。

Prof. Jean Claude Belfiore (Ecole Nationale Superieure des Telecommunications, France) 講 "Constructive coding on the Grassmann manifold: application to noncoherent space-time communication"。此一問題為 discrete optimization 的問題，故較 continuous optimization 為困難。本演講考慮到在 MIMO 傳輸系統中幾種不同傳輸方法的自由度問題，結論是 noncoherent transmission 比 differential detection 的自由度要高，所以決定探討 noncoherent transmission 之下的 coding，最後獲得一些甚佳的結果。

2. Poster Papers

本次會議中，我比較注意 channel estimation for OFDM and other modulation、blind equalization and frequency-domain equalization、OFDM synchronization、synchronization in general、及 CDMA single- and multi-user detection 等方面的論文，因為這些題目是我最近較專注研究的主題。此外，有一篇關於視訊傳輸的論文，但作者並未出席發表，不知何故。

在 channel estimation for OFDM and other modulation 方面，與我們的研究相關性較高的論文有：

- R. Ramakrishna, S. Bhashyam, and G. Krishnamurthi, "Improving channel estimation in OFDM systems for sparse multipath channels."
- F. D. Nunes and J. M. Leitao, "Innovation-based channel estimation and symbol detection of multi-level OFDM signals."
- A. Rontogiannis and K. Berberidis, "Bandwidth efficient transmission through sparse channels using a parametric channel estimation-based DFE."
- S. Kalyani and K. Giridhar, "Spectrally weighted frequency domain equalizer for mobile OFDM."

在 blind equalization and frequency-domain equalization 方面，與我們的研究相關性較高的論文有：

- S. P. Neugebauer and Z. Ding, "A turbo-driven known-modulus blind equalization method."
- I. Ghaleb, O. A. Alim, and K. Seddik, "A new finite alphabet based blind channel estimation for OFDM systems."
- S. Colonnese, G. Panci, and G. Scarano, "Using SOS in blind Gussgang equalization for QAM communication systems."

- C. M. Panazio and M. Bellanger, “On the frequency domain approach for spread spectrum: towards a convergence of DS-CDMA, MC-CDMA and OFDM.”
- D. Darsena, G. Gelli, L. Paura, and F. Verde, “Subspace-based blind channel identification for noncircular multicarrier transmissions.”

在 OFDM synchronization 方面，與我們的研究相關性較高的論文有：

- S. Attalah, “Blind estimation of residual carrier offset in OFDM systems.”
- Y. Yao and G. B. Giannakis, “Blind CFO estimation in OFDM systems.”

在 synchronization in general 方面，與我們的研究相關性較高的論文有：

- N. Noels, H. Steendam, and M. Moeneclaey, “On the Cramer-Rao lower bound and the performance of synchronizers for (turbo) encoded systems.”

在 CDMA single- and multi-user detection 方面，與我們的研究相關性較高的論文有：

- M. A. Ouameur and D. Massicotte, “A low complexity turbo detection for coded DS-CDMA systems in multipath channels at rake computational load.”
- T. Huovinen and T. Ristaniemi, “DS-CDMA capacity enhancement using blind source separation based group-wise successive interference cancellation.”
- H. Ge, L. L. Scharf, and M. Lundberg, “Reducced-rank multiuser detectors based on vector and matrix conjugate gradient Wiener filters.”
- G. P. White, Y. Zakharov, and A. G. Burr, “Chip-equalised UMTS downlink for fast fading channels.”
- B. Mouhouche, P. Loubaton, and W. Hachem, “Asymptotic analysis of reduced rank chip-level MMSE equalizers in the downlink of CDMA systems.”
- B. Mouhouche, K. Abed-Meraim, N. Ibrahim, and P. Loubaton, “Combined MMSE equalization and parallel interference cancellation for downlink multirate CDMA aystems.”
- L. Fety, D. Z. Filho, and M. Terre, “A comparison of chip rate-MMSE and symbol rate-MMSE in multiuser uplink long code DS-CDMA.”

將請我指導的學生注意這些論文。

三、考察參觀活動：無。

四、建議事項：無。

五、攜回資料名稱及內容：會議論文光碟一份及會議手冊(含議程與論文摘要)一冊。

Appendix A

VARIABLE-STEP-SIZE MULTIMODULUS BLIND DECISION-FEEDBACK EQUALIZATION FOR HIGH-ORDER QAM BASED ON BOUNDARY MSE ESTIMATION

Kun-Chien Hung[*†], *David W. Lin*[*†]

Dept. of Electronics Engineering and
Center for Telecommunications Research
National Chiao Tung University
Hsinchu, Taiwan 30010, ROC

Chun-Nan Ke[†]

Computer and Communications Research
Laboratories
Industrial Technology Research Institute
Chutung, Hsinchu, Taiwan 310, ROC

ABSTRACT

We consider blind decision-feedback equalization (DFE) under high-order modulation, which presents more difficult operating requirements than lower-order modulation. We base our design on the multimodulus algorithm (MMA). To attain fast convergence speed and low steady-state mean-square error (MSE), we consider varying the adaptation step size according to the presently achieved MSE. For this we investigate the properties of the MSE under blind MMA-based DFE and, based on the results, propose a method to estimate its value. The estimate is obtained by analyzing those equalizer filter outputs whose values fall outside the boundary of the modulation's constellation. Simulation results demonstrate the effectiveness of the proposed scheme.

1. INTRODUCTION

Blind equalization is of use in transmission systems where there exist no (or insufficient) known signal patterns that can be used for equalizer training. An example is downlink cable modem transmission where the known signal patterns are quite sparse that, if they are used to adapt equalizers in conventional training-sequence-based ways, the convergence may be very slow. A number of blind equalization algorithms have been proposed in the last few decades, of which many are of the stochastic gradient type. We consider the recently proposed multimodulus algorithm (MMA) [1], [2], which has relatively good performance.

A well-known design issue of stochastic-gradient type of adaptive algorithms is the choice of the adaptation step size, which has to strike a balance between convergence speed and steady-state SNR. In QAM-based transmission, this issue is more acute for higher-order modulations than for lower-order ones, because the former require higher SNR values to attain a given error performance than the latter and thus the convergence speed has to be sacrificed more. A way to alleviate this problem is to employ a variable step size (VSS) [3], [4]. For automatic adjustment of the step size, however, a mechanism to determine the current state of convergence is required, where the state of convergence may be characterized, for example, by the estimated mean-square error (MSE) at equalizer output [4]. Herein lies another problem that is more serious for blind equalization under higher-order modulations than under lower-order ones. That is, the equalizer output errors are relatively large before final convergence. Thus, for higher-order modulations, tentative decisions are liable to greater error probabilities and simplistic MSE estimates may suffer greater inaccuracy. In this work, we propose a method for reliable MSE estimation and an associated VSS multimodulus algorithm for blind decision-feedback equalization (DFE) under high-order QAM-based transmission, such as 1024-QAM over downlink digital cable channels.

The remainder of the paper is organized as follows. Section 2 presents the MMA algorithm for blind DFE and motivates our VSS design based on MSE estimation. Section 3 discusses the proposed MSE estimator. Section 4 further discusses the design of the VSS MMA algorithm and elucidates it with an example that also illustrates the superiority of the VSS MMA over a single-stage MMA. And Section 5 is the conclusion.

[*] This work was supported by Computer and Communications Research Laboratories, ITRI, R.O.C., under grant no. T1-92019-6.

[†] E-mails: hkc.ee90g@nctu.edu.tw, dwlin@mail.nctu.edu.tw, jnko@itri.org.tw

2. BLIND DFE EMPLOYING MMA

2.1. System Structure

The MMA [1], [2] seeks to minimize a cost function given by

$$\Psi(y) = E[(y_r^L - R_m^L)^2 + (y_i^L - R_m^L)^2] \tag{1}$$

with

$$R_m^L = \frac{E\left(a_r^{2L}\right)}{E\left(|a_r|^L\right)} = \frac{E\left(a_i^{2L}\right)}{E\left(|a_i|^L\right)}, \tag{2}$$

where y is the filter output in the equalizer, with y_r being its real part and y_i its imaginary part, a_r and a_i are, respectively, the real part and the imaginary part of the QAM symbol a, L is a postive integer, and R_m is called the constraint value of the algorithm. In practice, $L = 2$ is a good choice to compromise between implementation complexity and performance [2]. Letting $L = 2$ and taking the gradient of $\Psi(y)$ with respect to y yield

$$\psi(y) = y_r\left(y_r^2 - R_m^2\right) + j\, y_i\left(y_i^2 - R_m^2\right). \tag{3}$$

With fractionally-spaced DFE, we have

$$y(n) = \sum_{i=0}^{L-1}\sum_{k=0}^{N_f-1} f_{k,i}\, x_i(n-k) - \sum_{k=1}^{N_b} b_k\, \widehat{a}(n-k) \tag{4}$$

where we now associate a time index n (in number of QAM symbols) with the equalizer filter output y, $f_{k,i}$ is the kth coefficient

of the ith phase of the feedforward filter (FFF), b_k is the kth coefficient of the feedback filter (FBF), $x_i(n)$ is the equalizer input in phase i of symbol n, L is the oversamping factor, N_f is the length of the FFF in number of symbols, N_b is the number of taps in the FBF, and $\widehat{a}(n)$ is the decision output of the equalizer. From (3), we obtain the MMA blind DFE adaptation equations as

$$f_{k,i}(n) = f_{k,i}(n-1) - \mu\,\psi(y)\,x_i^*(n-k), \quad (5)$$

$$b_k(n) = b_k(n-1) + \mu\,\psi(y)\,\widehat{a}^*(n-k), \quad (6)$$

where μ is the step size.

2.2. Steady-State MSE of the MMA

We start by considering the simpler case of linear equalization. For this, let $\underline{w}$ be the vector of equalizer coefficients and $\underline{x}$ be the vector of input signal samples stored in the equalizer tapped delay line. The adaptation equation is given by

$$\underline{w}(n) = \underline{w}(n-1) - \mu\,\psi(y(n))\,\underline{x}^*(n). \quad (7)$$

Let $\underline{w}_{opt}$ be the optimal filter coefficient vector and define $\underline{\tilde{w}}(n) = \underline{w}_{opt} - \underline{w}(n)$. Then the a priori and the a posteriori estimation errors are given by, respectively,

$$e_a(n) = \underline{x}(n)\,\underline{\tilde{w}}(n-1), \quad (8)$$

$$e_p(n) = \underline{x}(n)\,\underline{\tilde{w}}(n) = e_a(n) - \mu\,\|\underline{x}(n)\|^2\,\psi(y(n)). \quad (9)$$

In the steady state where $E\{\underline{\tilde{w}}(n)\} = E\{\underline{\tilde{w}}(n-1)\}$, the mean-squares of the estimation errors are related by [5]

$$E\left\{\frac{|e_a(n)|^2}{\|\underline{x}(n)\|^2}\right\} = E\left\{\frac{\left|e_a(n) - \mu\|\underline{x}(n)\|^2\psi(y(n))\right|^2}{\|\underline{x}(n)\|^2}\right\}, \quad (10)$$

which can be simplified to

$$E\{\Re[e_a(n)\psi(y(n))]\} = \frac{\mu}{2}E\{\|\underline{x}(n)\|^2\;|\psi(y(n))|^2\}. \quad (11)$$

Now, assume that the residual error is small when the equalizer is converged. Then first-order approximation as in [6] gives

$$\psi(y(n)) \approx \psi(a(n)) + \psi'(a(n))\,e_a(n), \quad (12)$$

where $a(n)$ is the transmitted symbol at time n. (Without loss of generality, the transmission and filtering delays are disregarded.) Substituting (12) into (11) and assuming independence among $\psi(a(n))$, $e_a(n)$, and $\|\underline{x}(n)\|^2$ as in [5], we obtain the MSE as

$$E\{|e_a|^2\} = \frac{\mu}{2}\frac{E(\|\underline{x}\|^2)\cdot E(|\psi(a)|^2)}{E\{\Re[\psi'(a)]\}}. \quad (13)$$

Since $E(\|\underline{x}\|^2) = P_x\,N$ where P_x is the mean-square value of the equalizer input and N is the length of the equalizer, we have

$$E\{|e_a|^2\} = \frac{\mu}{2}\frac{P_x\cdot N\cdot E(|\psi(a)|^2)}{E\{\Re[\psi'(a)]\}}. \quad (14)$$

Now we turn to the case of DFE. The update equations (5) and (6) can be combined into

$$\begin{bmatrix}\underline{f}(n)\\ \underline{b}(n)\end{bmatrix} = \begin{bmatrix}\underline{f}(n-1)\\ \underline{b}(n-1)\end{bmatrix} - \mu\,\psi(y(n))\begin{bmatrix}\underline{x}^*(n)\\ \widehat{\underline{a}}^*(n-1)\end{bmatrix}, \quad (15)$$

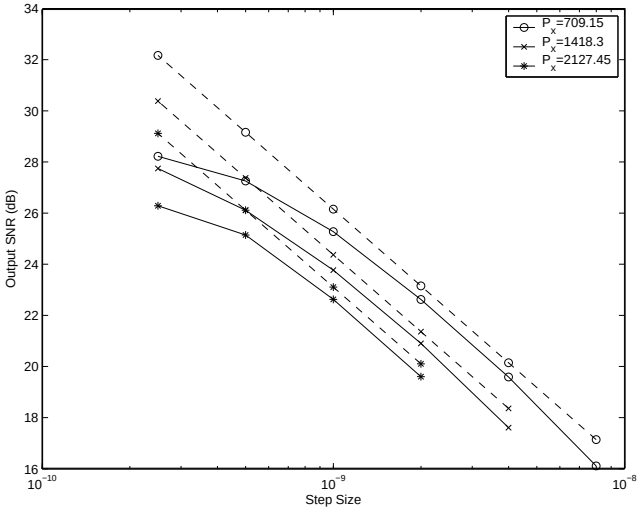


Fig. 1. Steady-state SNR of MMA blind DFE for 1024-QAM transmission over the channel in [7]. Solid lines: simulation; dashed lines: theory.

with obvious definitions for $\underline{f}(n)$, $\underline{b}(n)$, $\underline{x}(n)$, and $\widehat{\underline{a}}$. Extending the result (14) for linear equalization to the case of DFE yields

$$E\{|e_a|^2\} = \frac{\mu}{2}\frac{E\{|\psi(a)|^2\}}{E\{\Re[\psi'(a)]\}}\cdot(N_f L P_x + N_b E_s) \quad (16)$$

where $E_s = E\{|a|^2\}$ is the QAM symbol energy. Now since

$$\psi'(a) = \frac{3}{2}|a|^2 - R_m^2, \quad (17)$$

we get

$$E\{|e_a|^2\} = \frac{\mu\cdot E(|\psi(a)|^2)}{3E_s - 2R_m^2}\cdot(N_f L P_x + N_b E_s). \quad (18)$$

To verify the above results, we simulate 1024-QAM transmission over the (rather bad) cable channel in [7]. Theoretical analysis of the achievable SNR under MMSE (minimum MSE) DFE shows that $N_f = 10$ and $N_b = 20$ should be suitable choices for $L = 2$ (i.e., $T/2$-spaced FFF). The equalizer input SNR is 36 dB. The results for several adaptation step sizes and several equalizer input signal power levels are depicted in Fig. 1. The figure shows that the theory agrees reasonably well with the simluation results at larger step sizes that yield smaller steady-state SNR. The discrepancy at smaller step sizes (larger steady-state SNR) should be due to that the assumptions made earlier do not fully capture the dynamic behavior of the algorithm.

2.3. Approach to Variable-Step-Size MMA

The above results may be used in the following way: At any time, we calculate the actual SNR and compare it with the steady-state SNR for the presently used adaptation step size. If the latter has been reached, then we may switch to a smaller step size. Finally, when the SNR is high enough, we may switch out of blind mode of operation and enter decision-directed (DD) mode of equalizer operation.

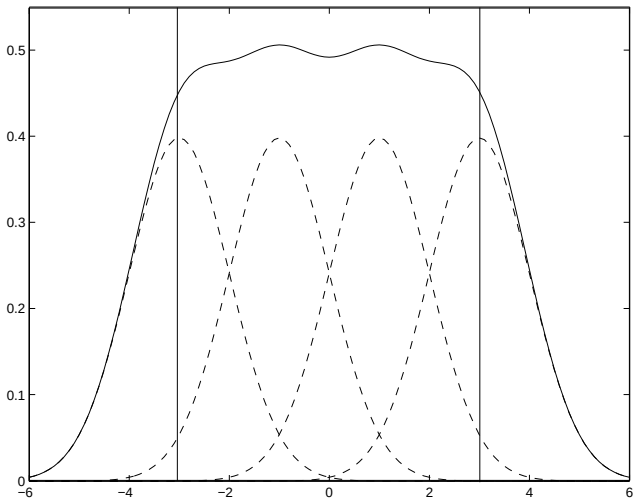


Fig. 2. Motivation and principle of boundary MSE estimation, illustrated for the case of 4-PAM (applicable also to 16-QAM). The constellation points are at ± 1 and ± 3. Dashed lines illustrate PDFs of equalizer filter output $y(n)$ corresponding to different values of $a(n)$; solid line their sum.

Key in this procedure is the estimation of actual SNR, or equivalently, the estimation of the actual MSE given by $E\{|y(n) - a(n)|^2\}$, the mean-square difference between equalizer filter output and the transmitted QAM symbol. To see how this can be accomplished, consider Figure 2 which illustrates the situation of 4-PAM (applicable also to 16-QAM), where the signal points are at ± 1 and ± 3. In the figure, the dashed lines illustrate the PDFs of the equalizer filter output $y(n)$ corresponding to different values of $a(n)$ and the solid line is their sum. An estimator of the MSE can be obtained from analyzing the PDF of $y(n)$. However, when the SNR is not high, the center part of the PDF is relatively flat. The variation in this part with changes in SNR (when the SNR is not high) is relatively small. Thus this part does not contribute significantly to the ability of MSE estimation. Numerical results also verify this observation. PDF variation outside the boundary symbol values is greater. Thus we base our MSE estimation on analysis of values of $y(n)$ that fall outside the boundary symbol values. Accordingly, we call this approach boundary MSE estimation. It is further described and analyzed in the next section.

From Fig. 2, we also see that only about $1/M$ of the equalizer input samples will be used in performing the estimate, but not all samples. This is a price we pay to have good sensitivity in MSE estimation.

3. BOUNDARY MSE ESTIMATION

Treat a QAM symbol as the direct sum of two PAM symbols. Let $\bar{y}$ denote the value of either dimension of the equalizer filter output y. The boundary MSE, for PAM, is defined as

$$\text{BMSE} = E\left\{ (\bar{y} - \bar{a}_{\max})^2 \Big| \, |\bar{y}| > \bar{a}_{\max} \right\}, \qquad (19)$$

where $\bar{a}_{\max}$ is the largest symbol value in the PAM constellation.

For convenience, let the constellation points of M-PAM have values $\pm 1, \pm 3, \cdots, \pm(M-1)$. Due to symmetry, in theoretical analysis we only need to consider the positive side. Let P_k be the probability that the transmitted symbol value is $(M-1-2k)$ but $\bar{y} > \bar{a}_{\max}$. Then

$$P_k = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x+2k)^2}{2\sigma^2} \right) dx = \text{Q}\left(\frac{2k}{\sigma} \right), \quad (20)$$

where we have assumed that the sum of the residual intersymbol interference (ISI) and additive noise is Gaussian and let σ^2 denote its variance. Note that σ^2 is the target of estimation. The corresponding mean-square boundary error is given by

$$\begin{aligned} V_k &= \int_0^\infty \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x+2k)^2}{2\sigma^2} \right) dx \\ &= (4k^2 + \sigma^2) P_k - \sqrt{\frac{2}{\pi}}\, k\sigma \exp\left(-\frac{2k^2}{\sigma^2} \right). \end{aligned} \quad (21)$$

Assume all constellation points are transmitted with equal probability. Then the total boundary error probability and total mean-square boundary error, on the positive side, are given by

$$P = \frac{1}{M} \sum_{k=0}^{M-1} P_k \quad \text{and} \quad V = \frac{1}{M} \sum_{k=0}^{M-1} V_k, \qquad (22)$$

respectively. The BMSE is thus given by

$$\begin{aligned} \frac{V}{P} &= \sigma^2 + \frac{4 \sum_k k^2 Q(2k/\sigma)}{\sum_k Q(2k/\sigma)} \\ &\quad - \frac{\sqrt{2/\pi}\, \sigma \sum_k k \, \exp(-2k^2/\sigma^2)}{\sum_k Q(2k/\sigma)}. \end{aligned} \qquad (23)$$

Figure 3 plots the ratio of σ^2 to BMSE in log scale for 32-PAM (applicable to 1024-QAM) over a range of SNR values (where SNR $= E\{\bar{a}^2\}/\sigma^2$ with $\bar{a}$ being PAM symbol value). Note that the ratio is nearly unity in large SNR. This is because when the true MSE σ^2 is small, the last two terms in the RHS of (23) are close to zero. Even for an SNR as low as 0 dB, the difference is only about 1.65 dB. In any case, this difference is compensated for in our variable-step-size MMA.

Practical estimation of the BMSE, for QAM, may be effected by time averaging, such as

$$\text{BMSE}(n) = \beta \cdot \text{BMSE}(n-1) + (1-\beta) \cdot 2 (|\bar{y}(n)| - \bar{a}_{\max})^2, \quad (24)$$

where β is the forgetting factor, the factor 2 is to account for the difference between QAM and PAM, and the recursion is executed only when $|\bar{y}(n)| > \bar{a}_{\max}$.

4. VARIABLE-STEP-SIZE MMA BLIND DFE AND SIMULATION RESULTS

Our proposed variable-step-size method works in a multistage, gear-shifting fashion rather than employing a continuously varying step size as some other researchers have considered. First, we decide an SNR that is safe to switch to DD mode with little concern of divergence afterwards. In the case of 1024-QAM, for example, simulation results indicate that 27 dB appears to be proper. Hence, the objective of blind equalization is set to be 27 dB. Figure 1 can then be used to find a suitable step size. For example, when $P_x = 709$, a suitable step size is around 5×10^{-10}. This constitutes the last stage of blind equalization.

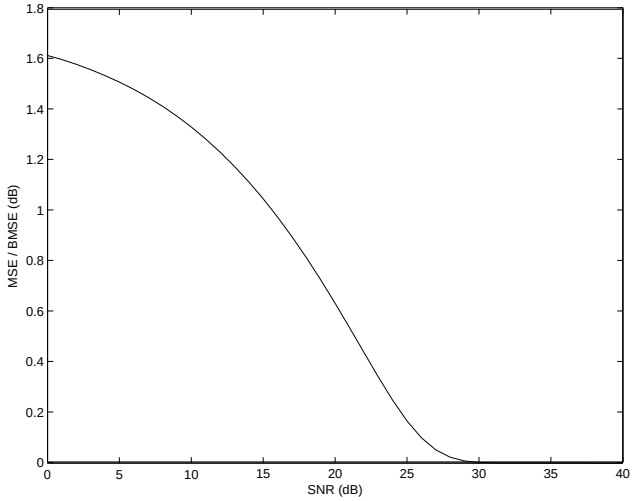


Fig. 3. Ratio of MSE to BMSE for 32-PAM (and 1024-QAM).

Table 1. Algorithm Parameters for the Example Design

Stage Index	1	2	3
Objective SNR (dB)	16	23	27
Step Size	8e–9	2e–9	5e–10
Stage Transition Threshold	13.6	2.9	1.3

Prior to the last blind stage, we can have one or more stages with larger (but diminishing with stage index) adaptation step sizes to effect fast initial convergence. We here present an example using two additional stages. For this, note again from Fig. 1 that a step size of 8×10^{-9} can yield an SNR of 16 dB after convergence and a step size of 2×10^{-9}, 23 dB. The BMSE value corresponding to these SNR values can be obtained as

$$TH = \frac{E_s}{SNR \cdot R(SNR)}, \tag{25}$$

where $R(SNR)$ is the ratio of MSE to BMSE at the given SNR value as can be obtained from (23) and shown in Fig. 3. For example, at SNR = 16 dB the ratio is about 1.1 dB. In the case of 1024-QAM for which $E_s = 682$, we get $TH = 13.6$. Once the estimated BMSE of the first stage reaches this level (call it threshold), we can switch to the second stage. And this continues to the last stage of blind equalization. The resulting algorithm parameters for this example are summarized in Table 1.

Figure 4 shows some simulation results for blind DFE under 1024-QAM transmission over the cable channel in [7] where the equalizer input SNR is 36 dB. The parameters of the multistage, VSS MMA blind DFE are as given above. The forgetting factor used in BMSE estimation is 0.99. The single-stage MMA uses a step size of 5×10^{-10}. The multistage algorithm clearly outperforms the single-stage algorithm. It only requires about 50,000 samples to converge and switch to DD mode whereas the single-stage MMA requires about 125,000 samples. Hence the former can provide a much faster startup speed than the latter.

An alternative algorithm design that sidesteps Fig. 1 is under investigation. Nevertheless, we note that the parameters in Table 1 are found applicable to all the cable channels that we simulated.

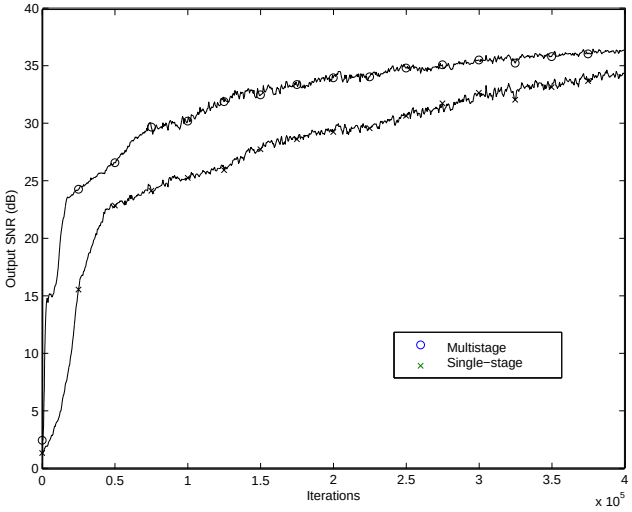


Fig. 4. Convergence of multistage VSS MMA and single-stage MMA.

5. CONCLUSION

Higher-order modulations lead to more stringent design requirements for blind equalization than lower-order ones. We considered the problem of blind DFE employing the multimodulus algorithm under high-order QAM, and developed a novel variable-step-size adaptation scheme based on a new way of MSE estimation. Simulation results show that the proposed method is effective in achieving fast startup.

6. REFERENCES

[1] J. Yang, J.-J. Werner and G. A. Dumont, "The multimodulus blind equalization algorithm," in *Proc. Int. Conf. Digital Signal Processing*, vol. 1, pp. 127–130, July 1997.

[2] J. Yang, J.-J. Werner and G. A. Dumont, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. Select. Areas Commun.*, vol. 20, no. 5, pp. 997–1015, June 2002.

[3] R. H. Kwong and E. W. Johnson, "A variable step size LMS algorithm," *IEEE Trans. Signal Processing*, vol. 40, no. 7, pp. 1633–1642, July 1992.

[4] F.-B. Ueng and Y. T. Su, "Adaptive VSS blind equalizers," *IEEE Signal Processing Lett.*, vol. 4, no. 4, pp. 100–102, Apr. 1997.

[5] J. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally spaced blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 48, no. 1, pp. 80–91, Jan. 2000.

[6] Y. Li and K. J. R. Liu, "Static and dynamic convergence behavior of adaptive blind equalizers," *IEEE Trans. Signal Processing*, vol. 44, no. 11, pp. 2736–2745, Nov. 1996.

[7] C.-I Hwang and D. W. Lin, "Joint low-complexity blind equalization, carrier recovery, and timing recovery with application to cable modem transmission," *IEICE Trans. Commun.*, vol. E82-B, no. 1, pp. 120–128, Jan. 1999.

Appendix B

CHIP-INTERLEAVED WCDMA WITH PARALLEL-INTERFERENCE-CANCELLATION RECEIVER IN MULTIPATH RAYLEIGH FADING CHANNELS

Yu-Nan Lin and David W. Lin

Department of Electronics Engineering and Center for Telecommunications Research
National Chiao Tung University, Hsinchu, Taiwan 30010, ROC
E-mails: ynlin.ee87g@nctu.edu.tw, dwlin@mail.nctu.edu.tw

ABSTRACT

Multiple access interference (MAI) is a major limiting factor of DS-CDMA system capacity. Parallel interference cancellation (PIC) is an effective means to mitigate MAI. However, if the initial decisions of which PIC is based on have high error probabilities, then PIC cannot perform well. To improve the performance of PIC, we propose chip-interleaving to provide intra-bit diversity. Then the initial decisions can be more accurate so as to enhance the PIC's performance. We design a chip-interleaving scheme based on some 3GPP WCDMA system features, such as the use of the Q-branch to transmit control bits and the I-branch to transmit data bits in a QPSK-like modulation scheme. Simulation results show that the proposed chip-interleaved WCDMA (CI-WCDMA) transmission can have significant performance advantage compared to simple WCDMA in transmission over fading multipath channels with MAI.

1. INTRODUCTION

Direct-sequence code division multiple access (DS-CDMA) has been employed in the second- and the third-generation mobile communication systems, and it has been envisioned for use in other contexts as well. Since it is hard to maintain orthogonality among simultaneously transmitting user signals, multiple access interference (MAI) has been a limiting factor to the capacity of DS-CDMA systems. Various multiuser detection (MUD) technologies have been proposed. Among them parallel interference cacellation (PIC) [1], [2] is an efficient yet practical choice due to its lower complexity. However, the performance of PIC is highly dependent on the correctness of tentative decisions in previous stages. In the transmission over fading channels, the decision signals may fluctuate enormously in magnitude, which hampers the interference cancellation capability of PIC.

Bit-interleaving is a common way to remedy channel fading in a channel coded system. However, most MUDs including PIC perform interference cancellation before channel decoding, for otherwise the complexity becomes overwhelming. In this case, the inter-bit diversity provided by bit-interleaving provides no help to the MUDs. Chip-interleaved DS-CDMA (CI-CDMA), which spreads out the chips of each single bit, has been proposed by several research groups recently. And some discussions of its performance in fading channels can be found in [3], [4]. Due to the resulting intra-bit diversity, chip-interleaving should provide great help to the performance of PIC.

This work was supported by the National Science Council of R.O.C. under grant no. NSC 92-2219-E-009-018.

The wideband code division multiple access (WCDMA) wireless communication standard completed recently by the Third Generation Partnership Project (3GPP) is now entering the stage of commercial operation. Based on some features of 3GPP WCDMA, we propose a chip-interleaved WCDMA (CI-WCDMA) technique in this paper and examine its transmission performance in multipath Rayleigh fading channels both with and without use of PIC.

This paper is organized as follows. Section 2 describes the CI-WCDMA signals. Two receiving strategies, including a rake-like receiver and a PIC receiver, are discussed in Sec. 3. Some simulation results are presented in Sec. 4. They demonstrate the effectiveness of the proposed scheme in addressing the MAI. Finally, Sec. 5 is the conclusion.

2. CI-WCDMA SIGNALING

In a chip-interleaved DS-CDMA system, bits are first spread as in conventional DS-CDMA and are then transmitted with interleaved chips. Consider the spreading and modulation method defined in the FDD mode of 3GPP WCDMA [5]. Figure 1 shows the proposed CI-WCDMA signaling scheme. As in WCDMA system, the data and control bits are first spread by channelization codes and carried in I- and Q-branches, respectively. A scrambling code is applied after block interleaving of the complex output chips. Let M be the interleaving depth in number of data bits and N the spreading factor of each data bit. Then we have NM chips in one interleaving block. The interleaving output forms N chip-blocks, with M chips per chip-block, for each interleaving block. With this chip-interleaver, we send the chips of each data bit in different chip-blocks and hence increase the intra-bit diveristy. The control channel always uses 256 as the spreading factor and therefore $R = 256/N$ is the spreading factor ratio of data channel to control channel. And each interleaving block contains M/R control bits in addition to M data bits.

Consider one interleaving block. The transmitted signal by the kth user can be expressed as

$$\begin{aligned} s^{(k)}(t) &= \sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\Big[\Big(b_d^{(k)}[m]C_I[n] \\ &\quad + j\beta^{(k)}\cdot b_c^{(k)}[\lfloor\frac{m}{R}\rfloor]C_Q[(mN+n)\%256]\Big) \\ &\quad \cdot c_s^{(k)}[nM+m]p(t-(nM+m)T_c)\Big], \end{aligned} \tag{1}$$

where $b_d^{(k)}[m]$ and $b_c^{(k)}[m]$ denote the mth data and control bits for the user, respectively, $C_I[n]$ and $C_Q[n]$ are the orthogonal

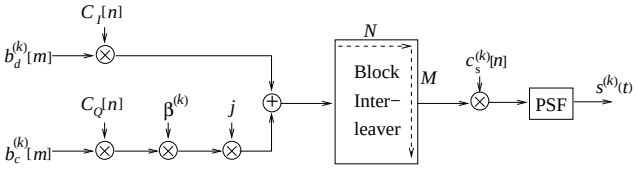


Fig. 1. The proposed CI-WCDMA signaling scheme.

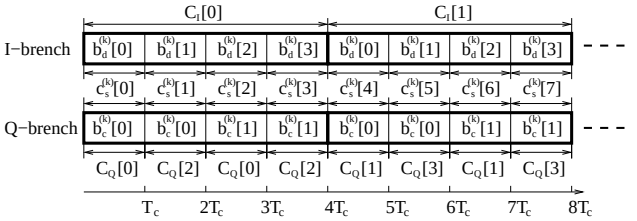


Fig. 2. An example of CI-WCDMA signal.

channelization codes for the data and the control channel, respectively, $c_s^{(k)}[n]$ is the scrambling code, T_c is the chip period, $p(t)$ is a square pulse of chip duration that is normalized so that $\int_0^{T_c} p^2(t)dt = 1$, and % denotes the modulo operation. The weighting factor $\beta^{(k)}$ usually should be less than 1 for power efficiency. The channelization codes are Walsh-Hadamard sequences and the scrambling code is a pseudo-random sequence with much longer period than the spreading factor. Figure 2 gives a illustrative example of the CI-WCDMA signal with $M = 4$ and $N = 2$, where each thick rectangle marks a chip-block.

In an asynchronous system, transmission over a multipath fading channel results in a received signal as

$$r(t) = \sqrt{2P}\sum_{k=0}^{K-1}\left\{\sum_{l=0}^{L-1}\alpha_l^{(k)}(t)s_k(t-\tau_k T_c - lT_c)\right\} + \eta(t), \tag{2}$$

where $\sqrt{2P}$ is the normalized signal amplitude of each user, K is the number of users in the system, L is the number of multipaths of each user, $\tau_k T_c$ is the relative signaling delay of the kth user, $\eta(t)$ is the additive noise (assumed white Gaussian), and $\alpha_l^{(k)}(t)$ are time-varying channel coefficients which, under a Rayleigh fading assumption, are zero-mean, complex-valued Gaussian random variables. In additon, let $\sum_{l=0}^{L-1} E\{|\alpha_l^{(k)}(t)|^2\} = 1$ for all k so that no user is at a disadvantaged position. Thus, we are considering a perfect slow power-controlled systems.

3. RECEVING OF CI-WCDMA SIGNALS

3.1. Rake-like Receiver

A PIC receiver is usually built on a rake receiver. Like in conventional DS-CDMA, a rake-type receiver which can provide multipath diversity is also an efficient receiver structure for CI-WCDMA signals. Figure 3 shows such a receiver, where z^{-M} denotes delay by M chips. Channel estimation is performed before chip deinterleaving using the control bits transmitted in the Q-branch. If the channel remains constant during each chip-block but varies independently randomly between chip-blocks, then with known control bits (pilot bits in 3GPP WCDMA language) and known channelization code, the maximum likelihood estimate of the channel coefficient in Gaussian noise and interference for each chip-block is the normalized correlation between the input signal samples and the product of the control bits and the channelization code in the chip-block. This is what shown in Fig. 3. Practical fading channels usually behave differently from the above and usually are not independent from chip-block to chip-block, and accordingly samples from several chip-blocks can be combined in a more complicated way to achieve better channel estimation. But such is not considered in the present work.

Without loss of generality, consider the detection of user 0's signal. After matched filtering and descrambling, the input signal to the lth finger of the rake-like receiver is given by

$$s_l^{(0)}[n] = c_s^{*(0)}[n]\cdot\int_{(n+\tau_0+l)T_c}^{(n+\tau_0+l+1)T_c} r(t)\cdot p(t-(n+\tau_0+l)T_c)dt. \tag{3}$$

The control bits $b_c^{(0)}[m]$ in 3GPP WCDMA may include pilot bits, power control bits and other control bits. In the following discussion, we treat them as only consisting of pilot bits (that is, all bits are known). The results can be easily extended to other situations. To estimate the channel coefficient of the lth path during the n_dth chip-block, we therefore calculate

$$\begin{aligned}\hat{\alpha}_l^{(0)}[n_d] &= \frac{1}{j2M\beta^{(0)}}\sum_{m=0}^{M-1}\Big(s_l^{(0)}[n_dM+m] \\ &\quad\cdot b_c^{(0)}[\lfloor\frac{m}{R}\rfloor]C_Q[(mN+n_d)\%256]\Big) \\ &= \sqrt{2P}\cdot\alpha^{(0)}[n_d] + \sum_{l'\neq l}^{L-1}\zeta_{l,l'}^{(0)}[n_d] \\ &\quad + \sum_{k=1}^{K-1}\sum_{l'=0}^{L-1}\zeta_{l,l'}^{(k)}[n_d] + \hat{\eta}_l[n_d],\end{aligned} \tag{4}$$

where $\hat{\eta}_l[n_d]$ summarizes the effect of the AWGN (additive white Gaussian noise), and $\zeta_{l,l'}^{(k)}$ is the interference from user k's l'th path to the estimation of user 0's lth path coefficient. Thus, the second and third terms in (4) are the interpath interference (IPI) and MAI, respectively. It can also be shown that for channel estimation, the "processing gain" provided by chip-interleaving in reducing the effect of IPI and MAI is M. Therefore, a larger interleaving depth increases the robustness of the channel estimate. But this is on the condition that the channel stays relatively constant over the chip-block. For fast fading channels a large M may not be beneficial.

Unlike channel estimation, data detection is performed after chip deinterleaving. Let $\hat{b}_d^{(0)}[m] = \text{sign}\{\Re\{y^{(0)}[m]\}\}$ be the decision for $b_d^{(0)}[m]$, we have the decision signal $y^{(0)}[m]$ as

$$\begin{aligned}y^{(0)}[m] &= \sum_{l=0}^{L-1}\sum_{n_d=0}^{N-1}\left\{s_l^{(0)}[n_dM+m]\left(\hat{\alpha}_l^{*(0)}[n_d]C_I[n_d]\right)\right\} \\ &= 2P\cdot b_d^{(0)}[m]\cdot\left(\sum_{l=0}^{L-1}\sum_{n_d=0}^{N-1}2\cdot\hat{\alpha}_l^{*(0)}[n_d]\alpha_l^{(0)}[n_d]\right) \\ &\quad + \sum_{l=0}^{L-1}\sum_{l'\neq l}^{L-1}\xi_{l,l'}^{(0)}[m] + \sum_{l=0}^{L-1}\sum_{k=1}^{K-1}\sum_{l'=0}^{L-1}\xi_{l,l'}^{(k)}[m] \\ &\quad + \hat{\eta}^{(0)}[m],\end{aligned} \tag{5}$$

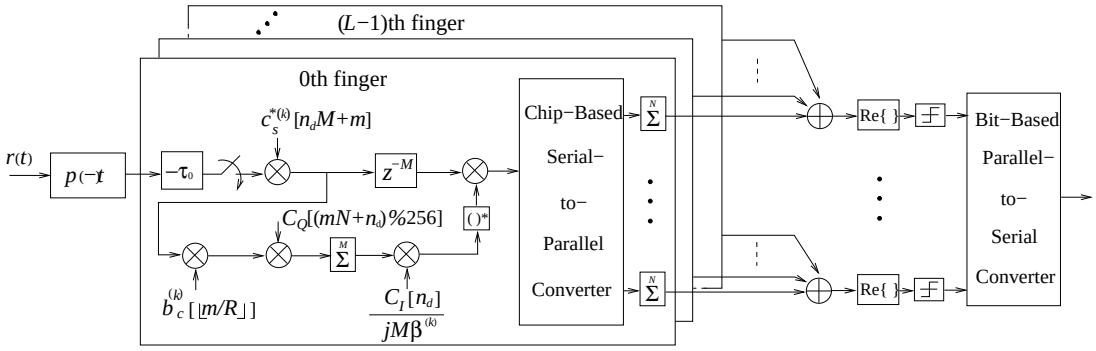


Fig. 3. A rake-like receiver for CI-WCDMA signals.

where $\xi_{l,l'}^{(k)}[m]$ has a similar interpretation as $\zeta_{l,l'}^{(k)}[m]$ but the processing gain here is $2N$. By (5), it is clear that the decision signal is the combination of NL independently faded coefficients $\alpha_l^{(0)}[n_d]$. Thus, there is an intra-bit diversity of order NL in CI-WCDMA systems.

3.2. PIC Detector

The basic idea behind PIC is quite simple: regenerate the interference to all received user signals from all others and substract such interference from the received signals simultaneously. For simplicity, only wideband hard-decision PIC is described here. In the first stage of PIC, the regenerated received signal of user k can be written as

$$\hat{r}^{(k)}(t) = \sum_{l=0}^{L-1} \hat{\alpha}_l^{(k)} \hat{s}^{(k)}(t - \tau_k T_c - lT_c), \tag{6}$$

where $\hat{s}^{(k)}(t)$ has the same expression as (1) except that $b_d^{(k)}[m]$ is replaced by $\hat{b}_d^{(k)}[m]$. Then for user 0, we pass the following signal again to the rake-like receiver: $r(t) - \sum_{k=1}^{K-1} \hat{r}^{(k)}(t)$. Subsequent PIC stages regenerate $\hat{s}^{(k)}(t)$ and $\hat{r}^{(k)}(t)$ using the respectively previous stages' output in a similar fashion, in hope that better detection results can be obtained with more repetitions.

Evidently, the correctness of the tentative decisions $\hat{b}_d^{(k)}[m]$ will affect the PIC performance greatly. With chip-interleaving, the output of rake receiver is more resistant to channel fading effects and hence the regenerated signals are more correct. An alternative to wideband PIC is narrowband PIC, which does not regenerate the received signals but the resultant interference in the despread signals directly [2]. That is, $\xi_{l,l'}[m]$ is regenerated and substracted from $y^{(0)}[m]$. Narrowband PIC has exactly the same performance as wideband PIC but with lower complexity especially when the spreading factor is high. Variants of these basic PIC schemes also exist. No matter what the kind of PIC is performed, CI-WCDMA can always help the interference cancellation ability.

4. SIMULATION RESULTS

We compare the performance of CI-WCDMA and nonchip-interleaved WCDMA by computer simulation. Similar to the 3GPP WCDMA, we let the chip rate be 3.84 Mcps. The interleaving depth M is set arbitrarily to 2560. Therefore, there are 10 control bits in each chip-block. (As indicated previously and will be touched on again in a later paragraph, a good choice of M has something to do with the channel's coherence time.) Let the data spreading factor be $N = 32$ and SNR be 13 dB. For the channel, let the number of multipaths be $L = 4$. And we perform only one stage of PIC.

Consider first quasi-static (or slow fading) channels, where the path coefficients remain constant during a chip-block and vary independently randomly from block to block. Figures 4 and 5 show some performance results for rake and PIC receivers, respectively. Clearly, CI-WCDMA has much better performance than simple WCDMA in both cases and the difference is more pronounced with PIC-based reception. The influence of β is also included in these simulations. It is interesting to note that, in the case of rake reception, the performance with $\beta = 4/15$ is worse than that with $\beta = 1$. This is due to the fact that the control channels also interfere with the data channel and a larger β induces a larger interference although it can facilitate a more accurate channel estimation. With PIC, however, the converse is true because the increased interference from having a larger β can be dealt with by its interference cancellation capability.

Next, we consider a more practical channel condition. We generate correlated Rayleigh fading channels by the baseband Doppler filtering method [6]. The coherence time of a fading channel can be defined as $B_c = 0.423/f_d$ where f_d is the maximum Doppler shift. In 3GPP WCDMA parameters, if the carrier frequency is equal to 2 GHz and if the coherence time is equal to the chip-block length $M = 2560$, then the corresponding mobile velocity is $v = 340$ km/h. Let $\beta = 1$. The performance of PIC is shown in Fig. 6. Again, CI-WCDMA has much better detection results than WCDMA. Another observation is that although the fading channel with velocity $v = 240$ km/h should yield a larger diversity order, its performance is inferior to that with $v = 120$ km/h. This is reasonable because our way of channel coefficient estimation by simple correlation does not yield very accurate results in relatively fast fading.

Additionally, we consider a channel with unequal path energies, where the path energies show exponential decay and are equal to 0, -3, -6, and -9 dB, respectively. The results are shown in Fig. 7. The performance of CI-WCDMA in this channel is quite similar to that in the equal-energy channel, but the performance of WCDMA is worse. This is because the main diversity exploited in WCDMA is path diveristy and hence the diversity gain is reduced when the paths have larger energy disparity. For CI-WCDMA, there is an additional time diversity and it is more robust.

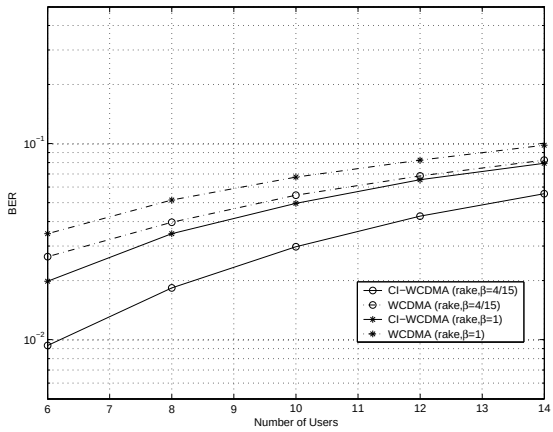


Fig. 4. Performance of rake receiver in multiple access communication over quasi-static channels for different systems and different choices of β with equal-energy multipaths.

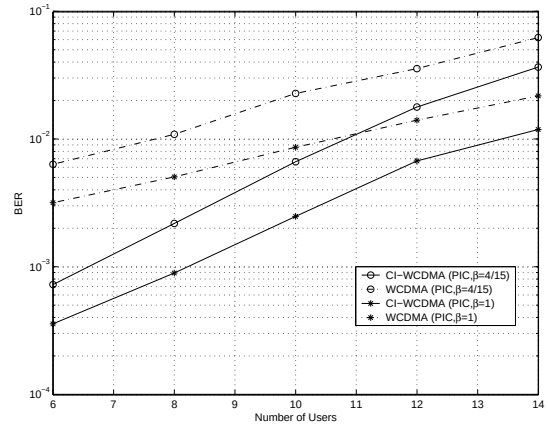


Fig. 5. Performance of PIC in multiple access communication over quasi-static channels for different systems and different choices of β with equal-energy multipaths.

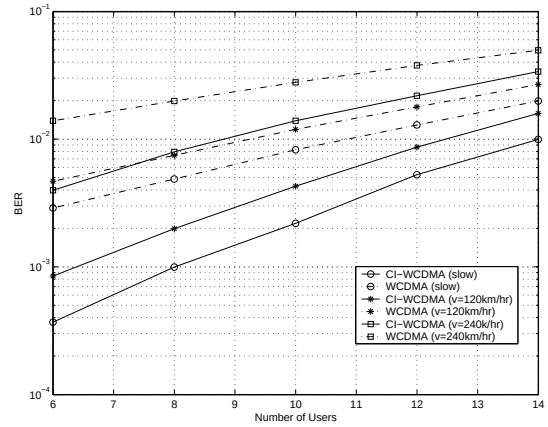


Fig. 6. Performance of PIC in multiple access communication over various fading channels for different systems with equal-energy multipaths.

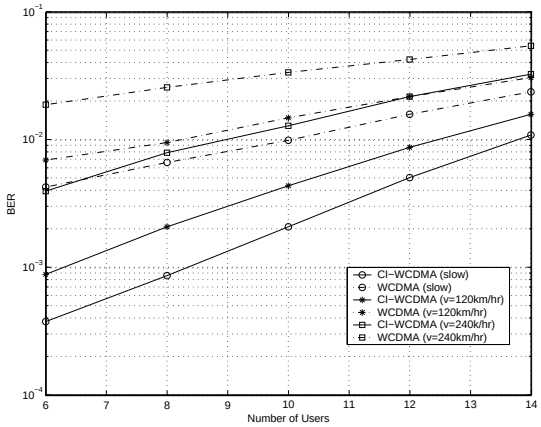


Fig. 7. Performance of PIC in multiple access communication over various fading channels for different systems with unequal-energy multipaths.

5. CONCLUSION

We proposed a chip-interleaved WCDMA (CI-WCDMA) transmission scheme and considered two ways of signal reception, namely, rake-like receiver and PIC. Due to the intra-bit diversity provided by chip-interleaving, CI-WCDMA is shown to have better performance than simple WCDMA in various conditions. The performance gain is especially significant with PIC reception. Essentially, the performance gain of CI-WCDMA comes from the increased time diversity due to chip-interleaving and, intuitively, more diversity is always beneficial to performance in multipath fading channels.

We have used a simple channel estimation scheme in the simulations. More sophisticated channel estimation methods should lead to even better performance of CI-WCDMA.

6. REFERENCES

[1] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.

[2] A. Nahler, R. Irmer, and G. Fettweis, "Reduced and differential parallel interference cancellation for CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 20, no. 2, pp. 237–247, Feb. 2002.

[3] P. Frenger, P. Orten, and T. Ottosson, "Coded-spread CDMA using maximum free distance low-rate convolutional codes," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 135–144, Jan. 2000.

[4] Y.-N. Lin and D. W. Lin, "Chip interleaving for performance improvement of coded DS-CDMA systems in Rayleigh fading channels," in *IEEE 58th Veh. Technol. Conf.*, Oct. 2003.

[5] 3GPP, *Technical Specification Group Radio Access Network; Spreading and Modulation (FDD)*. Doc. 3G TS 25.213 version 4.1.0, June 2001.

[6] T. S. Rappaport, *Wireless Communication: Principles and Practice, 2nd ed.* Upper Saddle River, NJ: Prentice-Hall, 2002.

Appendix C

Fifth IEEE Workshop on Signal Processing Advances in Wireless Communications, Lisboa, Portugal, July 11-14, 2004

A Chip-Interleaved Synchronous DS-CDMA Technique Enabling MAI-Free and Reduced-ISI Transmission with Low Complexity Receiving

Yu-Nan Lin[1] and David W. Lin
Dept. of Electronics Engineering
and Center for Telecommunications
Research
National Chiao Tung University
Hsinchu, Taiwan 30010, ROC
E-mails: ynlin.ee87g@nctu.edu.tw,
dwlin@mail.nctu.edu.tw

Abstract — **The capacity of conventional direct-sequence code-division multiaccess (DS-CDMA) systems is limited by multiaccess interference (MAI) that arises from correlation among user signals, especially in multipath channels. While inducing an intersymbol-interference (ISI) problem, chip-interleaved DS-CDMA (CIDS-CDMA) can preserve the orthogonality among user signals and effect MAI-free transmission. We propose a multicode CIDS-CDMA (MCIDS-CDMA) technique which can also remove ISI in addition to MAI when a sufficient number of multicodes are assigned. This design facilitates a low complexity receiver structure. In addition, a low complexity ISI canceller can be used to deal with the residual ISI, should the number of codes be insufficient.**

I. Introduction

Due to its high spectrum efficiency, direct-sequence code-division multiaccess (DS-CDMA) is a favored choice for high-speed wireless communication systems [1]. The capacity of a DS-CDMA system is highly dependent on the design of the spreading codes. In multipath channels, the codes' odd correlation functions [2] affect the system performance as well as their periodic correlation functions. While the commonly considered spreading sequences, such as the Walsh-Hadmard, the Gold, the Kasami, and the m-sequences have good periodic correlation functions, their odd correlation functions are not as desirable. As a result, multipath channels can exert serious impact on the performance of DS-CDMA. This problem has inspired the design of some new spreading sequences, such as the LS-codes [3], [4].

Unlike conventional DS-CDMA, chip-interleaved DS-CDMA (CIDS-CDMA) transmits a signal with the chips in interleaved order. By zero-padding between successive data blocks, CIDS-CDMA can achieve perfect multiaccess-interference-free (MAI-free) transmission [5]. However, due to the small chip duration needed for high-speed CDMA systems, CIDS-CDMA may be subject to large intersymbol interference (ISI). The ISI can be remedied by single-user equalization, which can be much simpler than multiuser detection [5]. Nevertheless, equalization is still a high-complexity operation.

In this paper, we propose a multicode CIDS-CDMA (MCIDS-CDMA) technique which gluttonously assigns multiple spreading codes to each user. With enough codes for each user, MCIDS-CDMA can translate a multiuser, multipath channel into an additive white Gaussian noise (AWGN) channel. By using a higher-order modulation with a larger spreading factor, MCIDS-CDMA can serve the same number of users as the single-code CIDS-CDMA (SCIDS-CDMA) technique proposed in [5]. The multiple spreading codes assigned to each user signal are applied periodically. One benefit of this arrangement (to be explained below) is that, if the number of codes is insufficient, then the residual ISI that remains will be contributed by much fewer symbols and can thus be highly reduced. Hence we may employ a simple receiver structure that combines a rake receiver and a simple ISI canceller (ISIC). In fact, a wireless system is free to employ both MCIDS-CDMA and SCIDS-CDMA without mutual interference. This gives much flexibility in system design.

This paper is organized as follows. Section II describes the basic CIDS-CDMA signals. In Section III, we present the proposed design. Section IV compares the proposed design with the recently proposed DS-CDMA system employing LS-codes. Some numerical results are presented in Section V. And finally, Section VI is the conclusion.

II. Basic CIDS-CDMA Signals

The modulated signal under CIDS-CDMA can be described as follows. First, information symbols are partitioned into blocks. After spreading, the symbols in each data block are transmitted with their chips interleaved. The lowpass-equivalent transmitted signal for the kth user is given by

$$x_k(t) = \sum_{g=-\infty}^{\infty} \sum_{m=0}^{M-1} s_k[gM+m] \cdot \sum_{n=0}^{N-1} c_k[n] p(t-(gNM+nM+m)T_c) \quad (1)$$

where M is the chip-interleaving depth or block length (in number of symbols), g is the block index, $s_k[h]$ are the user symbols, N is the spreading factor, $c_k[n]$ is the spreading code (or spreading sequence) for the kth user, T_c is the chip period, and $p(t)$ is a square pulse of chip duration that is normalized so that $\int_0^{T_c} p^2(t)dt = 1$. We consider short-code spreading, that is, the period of $c_k[n]$ is equal to N. Also, orthogonal spreading codes are assumed in the following discussion.

Synchronous transmission over a static multipath channel results in a received signal as

$$r(t) = \sqrt{2P} \sum_{k=0}^{K-1} \left\{ \sum_{l=0}^{L} \alpha_k[l] x_k(t-lT_c) \right\} + \zeta(t) \quad (2)$$

[1]This work was supported by National Science Council of R.O.C. under grant no. NSC 92-2219-E-009-018.

where $\sqrt{2P}$ is the normalized signal amplitude of each user, K is the number of users in the system, L is the number of multipaths, $\alpha_k[l]$ is the channel coefficient of the kth user's lth path, and $\zeta(t)$ is white Gaussian noise.

III. MAI-Free and Low-ISI Transmission

A. Despreading Result of CIDS-CDMA Signals

Like in conventional DS-CDMA, rake receiver is also an efficient receiver structure for CIDS-CDMA. Without loss of generality, consider the detection of user 0's signal. Then a rake combiner would yield the following decision signal for $s_0[h]$, where $h = gM + m$:

$$\begin{aligned} y_0[h] &= \left(\sqrt{2P}N\sum_{l=0}^{L-1}|\alpha_0[l]|^2\right)s_0[h] \\ &\quad + \xi_{\rm ISI}[h] + \xi_{\rm MAI}[h] + \xi_{\rm AWGN}[h], \end{aligned} \tag{3}$$

where

$$\xi_{\rm ISI}[h] = \sqrt{2P}\sum_{l=0}^{L-1}\left\{\sum_{l'\neq l}^{L-1}(\alpha_0^*[l]\alpha_0[l'])\cdot\phi_{l,l'}^{(0)}[h]\right\}, \tag{4}$$

$$\xi_{\rm MAI}[h] = \sqrt{2P}\sum_{k=1}^{K-1}\left\{\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}(\alpha_0^*[l]\alpha_k[l'])\phi_{l,l'}^{(k)}[h]\right\}, \tag{5}$$

and

$$\xi_{\rm AWGN}[h] = \sum_{l=0}^{L-1}\alpha_0^*[l]\sum_{n=0}^{N-1}c_0[n]\int_{iT_c}^{iT_c}\zeta(t)p(t-iT_c)dt, \tag{6}$$

with

$$\phi_{l,l'}^{(k)}[h] = \begin{cases} s_k[h-d]c_0[0]c_k[N-1] \\ +s_k[h-d+M]\sum_{n=1}^{N-1}c_0[n]c_k[n-1], \\ \qquad d>0 \text{ and } m<d-1; \\ s_k[h-M-d]\sum_{n=0}^{N-2}c_0[n]c_k[n+1] \\ +s_k[h-d]c_0[N-1]c_k[0], \\ \qquad d<0 \text{ and } m>M-1+d; \\ s_k[h-d]\sum_{n=0}^{N-1}c_0[n]c_k[n], \\ \qquad \text{otherwise}; \end{cases} \tag{7}$$

$i = gMN + m + nM + l$, and $d = l' - l$. Note that $\xi_{\rm ISI}[h]$, $\xi_{\rm MAI}[h]$, and $\xi_{\rm AWGN}[h]$ are contributions of ISI, MAI, and AWGN, respectively.

B. Design for Interference-Free Transmission

The first and the second righ-hand-side expressions in (7) are the interference contributed from the previous and the following blocks. They give the inter-block interference (IBI). By padding M_0 zero symbols in each data block, the IBI can be fully removed as long as $M_0 \geq L-1$. Let $M_d = M - M_0$ be the number of symbols in each data block carrying actual user data, the transmission efficiency can be computed as

$$\eta_{\rm CI} = M_d/M. \tag{8}$$

High efficiency can be obtained as the block length, M, is large.

Since IBI is removed by zero-padding, only the last condition in (7) needs to be considered. Thus we obtain

$$\xi_{\rm ISI}[h] = \sqrt{2P}\sum_{l=0}^{L-1}\sum_{l'\neq l}^{L-1}(\alpha_0^*[l]\alpha_0[l'])\cdot N\cdot s_0[h-(l'-l)] \tag{9}$$

and

$$\begin{aligned} \xi_{\rm MAI}[h] &= \sqrt{2P}\sum_{k=1}^{K-1}\left\{\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}(\alpha_0^*[l]\alpha_k[l'])\right. \\ &\quad \left.\cdot\left(\sum_{n=0}^{N-1}c_0[n]c_k[n]\right)s_k[h-(l'-l)]\right\}. \end{aligned} \tag{10}$$

Equation (10) shows that the zero-shift crosscorrelation of the spreading sequences are preserved in the transmission over multipath channels. Since we assume use of orthogonal sequences, the MAI is perfectly annulled.

While CIDS-CDMA has advantage in controlling the MAI, (9) brings to light an important problem, that is, CIDS-CDMA also maintains the zero-shift autocorrelation of the spreading codes. This is indicated by the factor N in (9), which arises from correlating $c_0[n]$ with itself, i.e., $\sum_{n=0}^{N-1}c_0[n]c_0[n]$. Therefore, CIDS-CDMA under short-code spreading may suffer a serious ISI problem. This phenomenon has been reported in [5] where the authors propose to remedy the ISI through equalization. While equalization can be simpler than multiuser detection, it may still be a rather complex operation. An even simpler solution is desirable and, in fact, possible.

Consider use of different short spreading sequences for successive symbols, which can be viewed as treating different symbols belonging to the same user as from different users. Let $c_k^{(h)}[n]$ be the spreading code for the kth user's hth symbol. Then (9) is changed to

$$\begin{aligned} \xi_{\rm ISI}[h] &= \sqrt{2P}\left\{\sum_{l=0}^{L-1}\sum_{l'\neq l}^{L-1}\alpha_0^*[l]\alpha_0[l']\right. \\ &\quad \left.\cdot\left(\sum_{n=0}^{N-1}c_0^{(h)}[n]c_0^{(h-(l'-l))}[n]\right)s_0[h-(l'-l)]\right\}. \end{aligned} \tag{11}$$

Comparing (11) with (9), we see that ISI is totally eliminated with the use of orthogonal spreading codes. The MAI term (10) is changed to

$$\begin{aligned} \xi_{\rm MAI}[h] &= \sqrt{2P}\sum_{k=1}^{K-1}\left\{\sum_{l=0}^{K-1}\sum_{l'\neq l}^{L-1}(\alpha_0^*[l]\alpha_k[l'])\right. \\ &\quad \left.\cdot\left(\sum_{n=0}^{N-1}c_0^{(h)}[n]c_k^{(h-(l'-l))}[n]\right)s_k[h-(l'-l)]\right\}, \end{aligned} \tag{12}$$

where, again, the MAI can be eliminated as long as no users use the same spreading codes.

Note that we do not need to assign a distinct spreading sequence for every symbol. To deal with multipath delay spread up to $L-1$ chips, it suffices to have L spreading sequences for each user, to be used periodically. This is evident from (11), where note that the index difference between two spreading codes that are correlated together is $l'-l$, and $|l'-l| \leq L-1$. For ease of reference, we abbreviate the multicode and the single-code CIDS-CDMA as MCIDS-CDMA and SCIDS-CDMA, respectively. A simple MCIDS-CDMA transmitter is shown in Fig. 1 where chip-interleaving is accomplished by simple block interleaving.

Though MCIDS-CDMA can remove MAI and ISI without equalization, one drawback is that less users can be served due

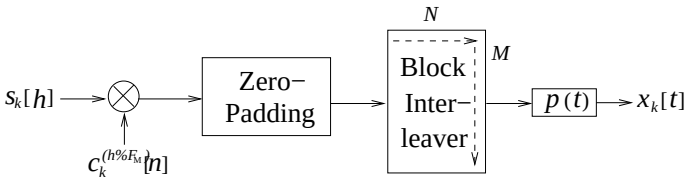


Fig. 1: A simple MCIDS-CDMA transmitter.

to the gluttonous code usage. To increase the system capacity, a larger spreading factor may be used. But this would reduce the user data rate. One solution to this problem is to use high-order modulations. Use of high-order modulations in conventional DS-CDMA is not an efficient design due to the resulting MAI from correlation (i.e., nonorthogonality) among user signals [4]. In the above MCIDS-CDMA, however, since orthogonality among user signals can be preserved, high-order modulations do not incur an interference problem.

When the user number and the multipath number are so many that there are not enough spreading codes to assign, ISI would occur whereas the MAI is still zero. As a high-order modulation is used, the resulting ISI may lead to more serious performance degradation than with a low-order modulation. An equalizer can be used in MCIDS-CDMA as in SCIDS-CDMA to mitigate the ISI. But due to the structure of MCIDS-CDMA signals, the equalizer can be much simpler. We elaborate on this point presently.

C. A Low Complexity Receiver Structure

Assume that each user uses F_M different spreading codes and therefore the rake receiver has F_M finger inputs. The residual ISI in (11) becomes

$$\xi_{\rm ISI}[h] = \sqrt{2P}\left\{\sum_{l=0}^{F_M-1}\sum_{l'\neq l}^{L-1} s_0[h-(l'-l)]\cdot(\alpha_0^*[l]\alpha_0[l']) \cdot\left(\sum_{n=0}^{N-1} c_0^{(h\%F_M)}[n]c_0^{((h-(l'-l))\%F_M)}[n]\right)\right\}. \quad (13)$$

Since $\sum_{n=0}^{N-1} c_0^{(h\%F_M)}[n]c_0^{((h-(l'-l))\%F_M)}[n] = 0$ for all $(l'-l)\%F_M \neq 0$, only those paths whose relative delays with a desired finger are multiples of F_M would induce ISI at this finger. These interfering signals all come from the same symbol. Thus the resulting residual ISI is given by

$$\xi_{\rm ISI}[h] = \sum_{f=1}^{\lceil\frac{L-F_M}{F_M}\rceil} \sqrt{2P}N\beta_f\cdot s_0[h-fF_M] \quad (14)$$

where

$$\beta_f \triangleq \sum_{l=0}^{F_M-1} \Re\{\alpha_0^*[l]\alpha_0[l+fF_M]\} \quad (15)$$

is the interference from symbol $h-fF_M$. It is seen that the number of multipaths contributing to ISI is reduced to one-F_Mth of the channel delay spread. This can simplify the receiver dramatically.

With F_M spreading codes, assume the energies of the first F_M paths can be perfectly collected and we only need to cancel the ISI from later paths. One way to accomplish this is by decision-feedback equalization (DFE) with a suitable number of filter taps. However, rather than referring to it as DFE, we shall term it an ISI cancellation (ISIC) technique as we shall also consider another ISIC scheme inspired by multiuser interference cancellation.

The main idea of multiuser interference cancellation is to regenerate the MAI and cancel it from the received signal. In ISIC, we regenerate the ISI $\xi_{\rm ISI}[h]$ and substract it from the rake output. We consider two ISI regeneration methods: linear and hard-decision-based; the latter corresponds to DFE but the former is not linear equalization. Assume perfect knowledge of the channel coefficients (i.e., the channel state information). Then the decision signal in linear ISIC is given by

$$\hat{y}_0[h] = y_0[h] - \frac{1}{\sum_{l=0}^{F_M-1}|\alpha_0[l]|^2}\sum_{f=1}^{\lceil\frac{L-F_M}{F_M}\rceil}\beta_f\cdot\hat{y}_0[h-fF_M]. \quad (16)$$

The hard ISIC has a similar form for the decision signal except that the decision-circuit outputs are used in the cancellation process:

$$\hat{y}_0[h] = y_0[h] - \sqrt{2P}N\sum_{f=1}^{\lceil\frac{L-F_M}{F_M}\rceil}\beta_f\cdot\hat{s}_0[h-fF_M], \quad (17)$$

where $\hat{s}_0[h-fF_M]$ are the decision-circuit outputs. Figure 2 shows the receiver structure employing hard ISIC, which can handle multipath delay spread upto $3F_M-1$ chips due to the two block delays of F_M symbols in the decision-feedback path. With B block delays, the receiver can handle $(B+1)F_M-1$ chips of delay spread. Note that we cannot set F_M to be a too small value because only the energies of paths whose delays are within F_M chips can be collected. If F_M is too small, unreliable interference estimates would be regenerated and it may worsen the decision performance than improves it.

IV. Comparison with LAS-CDMA

It is of interest to note that the proposed MCIDS-CDMA technique has some similar properties to LAS-CDMA [3], [4]. LAS-CDMA is similar to conventional DS-CDMA except that the so-called LS-codes are used as the spreading codes. Since the aperiodic autocorrelation sidelobes and crosscorrelation of LS-codes are all zero within the designed interference free window (IFW), it is also an MAI- and ISI-free system over multipath channels. Stańczack *et al.* [3] provide systematic methods for the construction of LS-codes with IFW width equal to $2F_M-1$, which can effect interference-free receiving as the multipath delay sprad is less than F_M-1 chips. The number of LS-codes is bounded above by N/F_M, which is the same as the maximum user number in MCIDS-CDMA. To use LS-codes, $2(F_M-1)$ out of $N+F_M-1$ chips must be zero and the transmission efficiency becomes

$$\eta_{\rm LS} = \frac{N-F_M+1}{N+F_M-1}, \quad (18)$$

which is dependent on both the spreading factor and the IFW width, whereas that of MCIDS-CDMA, for the same F_M, is dependent on the block length only.

Another difference is the influence of under-designed parameters. When the actual channel delay spread is larger than the number of codes, MCIDS-CDMA still provides MAI-free transmission with reduced number of ISI terms. However, in DS-CDMA with LS-codes spreading, MAI appears and may come from more symbols than for the ISI in MCIDS-CDMA. MAI

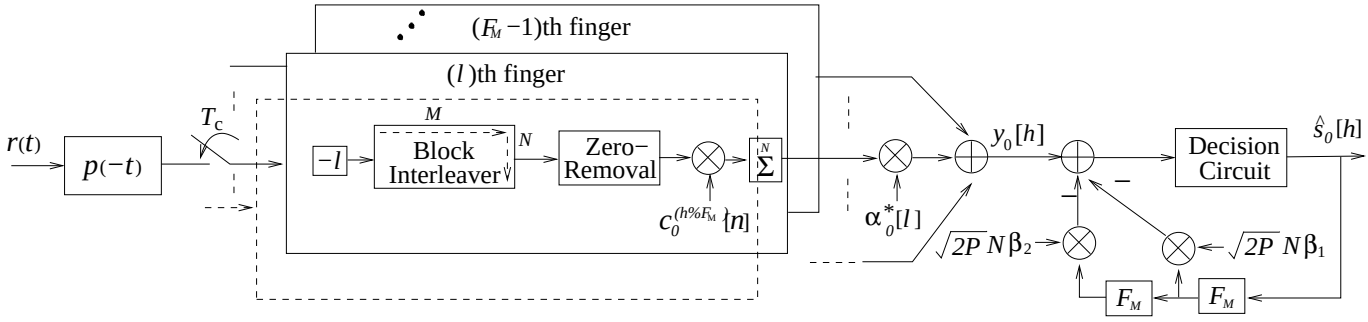


Fig. 2: Block diagram of hard ISIC receiver.

cancellation is usually a high-complexity operation and is normally only considered for uplink transmission. Further, MCIDS-CDMA has more flexibility on spreading codes assignment and can assign more or less codes to each individual user according to the channel condition for the user or the processing capability of the transceivers involved. The construction of LS-codes does not allow the codes to have different IFW widths. Nevertheless, LAS-CDMA can preserve the code property better than MCIDS-CDMA in fast fading. In conclusion, MCIDS-CDMA appears to be more robust in dealing with variations in delay spread while the DS-CDMA with LS-codes spreading may be better able to handle fast fading.

V. Numerical Results

This section presents some numerical results for the proposed MCIDS-CDMA system. The chip rate is set to 3.84 Mcps, similar to 3GPP WCDMA [1]. For MCIDS-CDMA, the spreading factor is set to $N = 64$ and 16-QAM modulation is applied to increase the data rate. Let the user data block length $M_d = 128$. And let the number of zeros padded be 8, which can eliminate IBI for multipath delay spreads up to 8 chips. In the MCIDS-CDMA system, each user is assigned 4 different Walsh-Hardmard sequences and therefore the maximum allowed user number is 16. For comparison, both conventional DS-CDMA and SCIDS-CDMA using a spreading factor of 16 are also simulated. We assume that the path coefficients do not change during a block and that there is perfect channel estimation. Each path coefficient follows Rayleigh distribution, but the system exercises perfect power control so that the total path energy is normalized to 1 in each block.

Figure 3 compares the BER performance between DS-CDMA and MCIDS-CDMA at SNR = 13 dB for a 4-path channel with delays 0, 1, 2, and 3 chips and relative powers 0, −3, 0, −3 dB, respectively. As mentioned, DS-CDMA cannot preserve the orthogonality among user signals in multipath channels, even when orthogonal spreading sequences are used. Hence it is not surprising to see in Fig. 3 that DS-CDMA with Walsh-Hadmard spreading sequences has poor performance even when the user number is small. The result for DS-CDMA with random-code spreading is also shown and it outperforms the system using Walsh-Hadmard sequences as less users are served. With MCIDS-CDMA, the performance is irrespective of user number and is much better than the other systems. Another observation is that the performance of MCIDS-CDMA in such a channel condition is exactly the same as 16-QAM modulation over the AWGN channel though the system is one with multiuser transmission and multipath propagation.

The influence of SNR is shown in Fig. 4. The results of DS-CDMA is for a 4-user condition while the results can be applied to up to 16 users for the other systems. Apparently, MCIDS-CDMA performs much better than DS-CDMA even when the user number is not large. For comparison, performance of SCIDS-CDMA using MMSE linear equalization (LE) is also shown, where "LE(n_s)" denotes LE with n_s taps. We simply set the decision delay to $n_s/2$. As shown in Fig. 4, even with much more equalizer taps, SICDS-CDMA has worse performance than MCIDS-CDMA at high SNR while performs slightly better at low SNR.

To examine the effect of insufficient code number for the MCIDS-CDMA system, the 3GPP typical urban channel model (TUx) [6] is considered and adapted. To simplify the model, the time resolution is set to 1 chip duration by the way mentioned in Annex B of [6] and the resultant path delays are 0, 1, 2, 3, 5, 6, 7, and 8, with the relative path energies being 0, −1.9, −0.3045, −3.6367, −5.183, −8.7855, −10.2, and −15.16 dB, respectively. From Fig. 5, it is clear that the basic MCIDS-CDMA has much worse performance in this channel because the code number is insufficient and hence cannot perfectly eliminate the ISI. Since the maximum delay spread is 8 chips, only an ISIC of $\lceil \frac{9-4}{4} \rceil = 2$ taps is required. The decision results for different ISIC methods are shown in Fig. 5. As can be expected, nonlinear (hard) ISIC has better performance than linear ISIC. Different from multiuser interference cancellation, the nonlinear ISIC results are quite close to perfect ISIC which uses the transmitted symbols to regenerate interference. This means that nonlinear ISIC is near optimal in such a receiving structure, which is because the interference in MCIDS-CDMA is from few sources and therefore the ISIC converges relatively easily.

Figure 6 compares the performance of MCIDS-CDMA and SCIDS-CDMA with equalization under the TUx channel. Both LE and DFE are considered for SCIDS-CDMA, where "DFE(n_f,n_d)" denotes DFE with n_f feedforward taps and n_d feedback taps. The decision delay is set to $n_f/2$. In terms of complexity, our ISIC receiver is comparable to DFE(4,2). Clearly, at a lower complexity, MCIDS-CDMA with rake-ISIC receiver can deliver a higher performance than SCIDS-CDMA with DFE. Although the performance of the latter can be improved by lengthening the equalizer, to better the performance of the former at high SNR would require very significant increase in equalizer complexity.

VI. Conclusion

We proposed an MCIDS-CDMA technique that employed multicode spreading with chip interleaving. The ability of this

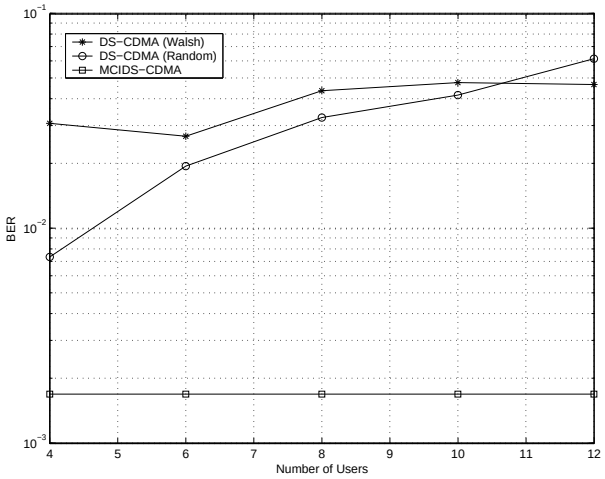


Fig. 3: BER performance versus user number for different systems at SNR = 13 dB with a 4-path channel.

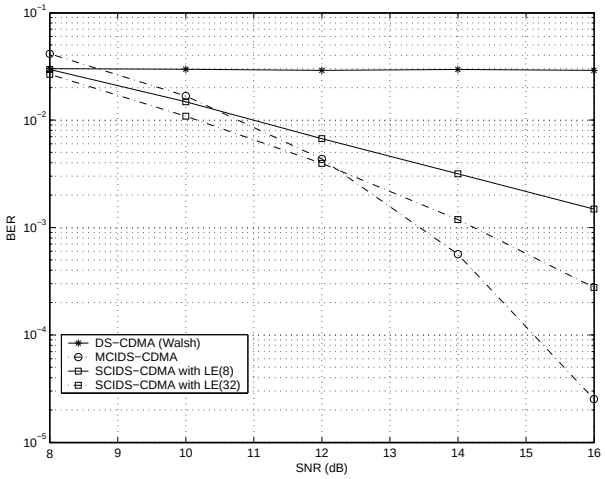


Fig. 4: BER performance versus SNR for different systems with a 4-path channel.

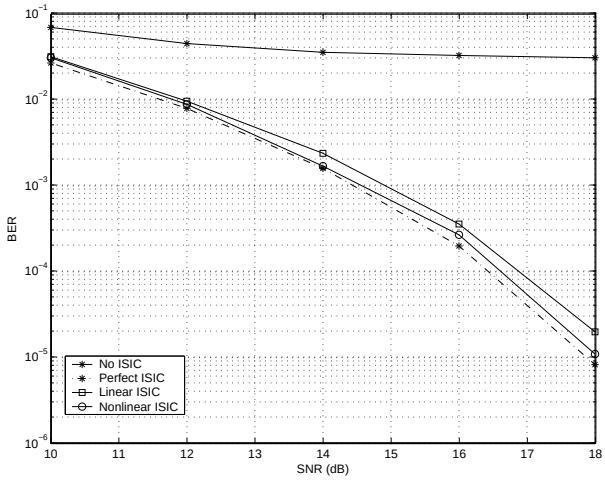


Fig. 5: BER performance versus SNR of MCIDS-CDMA systems for the TUx channel.

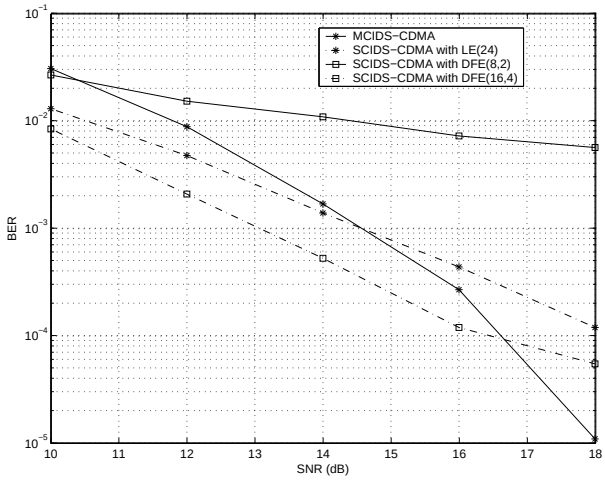


Fig. 6: BER performance versus SNR of CIDS-CDMA systems for the TUx channel.

technique to remedy MAI and ISI was shown. With high-order modulation, such a system can support the same number of users as SCIDS-CDMA. Only an ISIC with a few taps is required to deal with the ISI under the condition of insufficient number of codes. It was shown that the proposed technique could outperform SCIDS-CDMA with equalization at comparable complexity by a large margin. Moreover, the proposed MCIDS-CDMA can be employed jointly with SCIDS-CDMA in a wireless system without mutual interference. This gives much flexibility in system design.

Synchronous tranmission has been assumed in this paper. However, MCIDS-CDMA actually permits asynchronous transmission to a certain degree if enough zeros are padded.

References

[1] M. Haardt, A. Klein, R. Koehn, S. Oestreich, M. Purat, V. Sommer, and T. Ulrich, "The TD-DMA based UTRA TDD mode," *IEEE J. Select. Areas Commun.*, vol. 18, no. 8, pp. 1375–1385, Aug. 2000.

[2] M. B. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication-part I: system analysis," *IEEE Trans. Commun.*, vol. 25, no. 8, pp. 795–799, Aug. 1977.

[3] S. Stańczack, H. Boche, and M. Haardt, "Are LAS-codes a miracle?" in *IEEE Global Telecommun. Conf.*, Nov. 2001, pp. 589–593.

[4] D. Li, "The perspectives of large area synchronous CDMA technology for the fourth-generation mobile radio," *IEEE Commun. Mag.*, pp. 114–118, Mar. 2003.

[5] S. Zhou, G. B. Giannakis, and C. Le Martret, "Chip-interleaved block-spread code division multiple access," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 235–248, Feb. 2002.

[6] TSG RAN WG4, 3GPP, "Deployment aspects," Tech. Rep. 3G TR 25.943 v5.1.0, June 2002.