行政院國家科學委員會補助專題研究計畫 ■ 成果報告
□期中進度報告

# 基於 MPEG 標準之多媒體通訊整合平台及其應用(I) —子計畫五：MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬(I)

# MPEG IPMP System and Robust Video Decoder Design and Simulation (I)

計畫類別：□ 個別型計畫 ■ 整合型計畫
計畫編號： NSC 92-2219-E-009-008

執行期間： 92 年 8 月 1 日至 93 年 7 月 31 日

計畫主持人：杭學鳴
計畫參與人員：張峰誠 唐之瑋 蔡家揚 范振韋 洪朝雄

成果報告類型(依經費核定清單規定繳交)：□精簡報告 ■完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：國立交通大學電子工程學系

中 華 民 國 93 年 9 月 30 日

# 行政院國家科學委員會專題研究計畫成果報告

# MPEG 智財管理與保護系統及強韌視訊解碼器之設計與模擬(I)

# MPEG IPMP System and Robust Video Decoder Design and Simulation (I)

計畫編號：NSC 92-2219-E-009 -008

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：杭學鳴　　國立交通大學電子工程學系教授

計畫參與人員：張峰誠 唐之瑋 蔡家揚 范振韋 洪朝雄

國立交通大學電子研究所

## 中文摘要

過去數年間，由於多媒體通訊突飛猛進，網際網路迅速普及，以多媒體物件製作、處理、編輯及傳輸為目標的 MPEG-4/MPEG-21 標準活動受到大家的關注。我們在此研究計畫的目的為(1)深入研究並模擬 MPEG-4 IPMP 延伸系統，並將其概念實作於 MPEG-21 Test Bed 上，(2)數位浮水印與密碼方法結合應用於多層次編碼上，以及(3)畫面間小波視訊編碼（Interframe Wavelet）之開發與研究，利用 AVC 中之移動估測方法改善移動補償時間濾波，並提出一個適用於 AVC 移動估測之可調式移動向量技術。其中 IPMP System 在 2004 年 7 月加入 MPEG-21 Multimedia Test Bed 中，已成為 Test Bed 軟體的一部份，全案期望在 2005 年左右完成標準化。Interframe Wavelet 成果在 2004 年 3 月與 7 月提案 MPEG 標準組織，參加 scalable video coding Call-for-Proposal 競賽，與後續 Reference Model 之改進。

**關鍵詞**：多媒體通訊, MPEG-4, IPMP, MPEG-21, Digital Watermarking, Crypto, Multi-layer, 畫面間小波視訊編碼

## 英文摘要

Due to tremendous advances in multimedia communication and the aggressive expansion of Internet in the past a few years, the MPEG standards activity, which aims at establishing a comprehensive specifications for multimedia object construction, manipulation, editing and delivery, receives a lot of attentions. Our goals in this project are to (1) investigate the MPEG-4 Systems and its IPMP extension, and implement the extension on the MPEG-21 Test Bed, (2) investigate digital watermarking and crypto combined systems, and apply them to the multi-layer coding structure, and (3) study and design interframe wavelet coding algorithms, improve its motion estimation part, and propose the scalable motion techniques. The MPEG-4 IPMP system has been included as a part of the MPEG-21 Multimedia Test Bed. Our inter-frame wavelet scheme has been submitted to the MPEG committee in response to the scalable video coding Call-for-Proposal. Also, we continue participating in the core experiments and refinement of the MPEG interframe wavelet reference model.

**Keywords**: Multimedia communication, MPEG-4, IPMP, MPEG-21, Digital Watermarking, Crypto, Multi-layer Coding, Interframe Wavelet

# 目錄 Table of Contents

# 第一部分 MPEG-4 IPMP 系統的研究

## A. 背景與目的

每件多媒體，例如書本、音樂、電影、電視節目等等，被數位化前或者後，都是智慧財產的一種。現今網際網路的蓬勃發展，涉及數位化多媒體財產的電子商務日漸普遍。因此，能保護智慧財產和使得智慧財產可以透過更方便的形式作商務交易的技術 ─ 智慧財產的數位化管理和保護(Intellectual Property Management & Protection，簡稱 IPMP)，將會成為電子商務系統之中不可或缺的一環。

由 1997 年開始，MPEG 組織開始嘗試為 IPMP 訂立標準的公用介面和相關協定，以解決數位化媒體在網路上交易所衍生的問題，目標是把 IPMP 整合到現有的 MPEG 標準之中。其中利用密碼方法或數位浮水印的技術以有效的提高非法剽竊的困難度。目前，類似但簡化之技術已被應用在數位電視系統中。截至本報告撰寫時，MPEG-4 和 MPEG-2 標準中的 IPMP 訂定已經完成，而 MPEG 組織也開始投入更多的資源訂定 MPEG-21 IPMP 標準。

MPEG-4 IPMP 架構已經歷重大改進，由最初只能作有限範圍保護的 IPMP hook 介面，進展為更具彈性與共通操作性的 IPMP Extension (IPMPX)介面。MPEG-21 IPMP 尚在起草階段，許多介面方式仍有待討論，不過可以預期將來與 MPEG-21 主架構整合，可以提供比 MPEG-4 IPMPX 更為完善的數位資訊保護。

## B. 研究步驟 – IPMP 標準與架構

我們的研究承續之前對於 MPEG-4 系統[1]的了解，本期專注於 MPEG-4 標準 (ISO/IEC 14496-1) 中所提供的 IPMP 架構，並輔以部分 MPEG-21 IPMP 及 MPEG-21 Test Bed[2]，以期對 MPEG IPMP 之運作有更前瞻的認識。

MPEG-4 Systems ver.1 中所定義的 IPMP 架構提供了一個標準化的介面讓 MPEG-4 players 可以使用不同的 IPMP System[3]，MPEG-4/AMD3[5]提出後，為了區別兩個版本中差異頗大的 IPMP 子系統，前者稱 IPMP Hook，而後者稱為 IPMP Extension (IPMPX)。IPMP Hook 運作主由 IPMP ES (Elementary Stream) 以及 IPMP 描述子 (descriptor) 所構成。IPMP ES 傳遞時間上變化較快的週期性的資訊給一個或是多個 IPMP 系統。IPMP 描述子由 object descriptor stream 所攜帶並傳送出去。圖 1 是 MPEG-4 中的 IPMP Hook 架構。

IPMP Hook 架構的最大不足之處在於工具間的溝通方式並沒有正式定義，這使得各工具的實作者無法利用他人的實作成果。有鑑於此，IPMPX 架構觀念改為虛擬終端機(Virtual Terminal)，與既有的 MPEG-4 系統以 Message 互相溝通，如圖 2 所示。IPMP 虛擬終端主要由兩大概念合成，一為 Message Router (MR)，另一為 Tool Manager

(TM)。Message Router 負責將所有的IPMP Message傳送至對應的IPMP Tool或終端機本身，而接收message的一方則根據Message內容負責串流的資料處理或控制，例如解碼或是權限控管。Tool Manager 負責 Tool 的建立、消滅、與關聯等功能，當需要時，可由MPEG-4系統或Message Router發出管理需求。

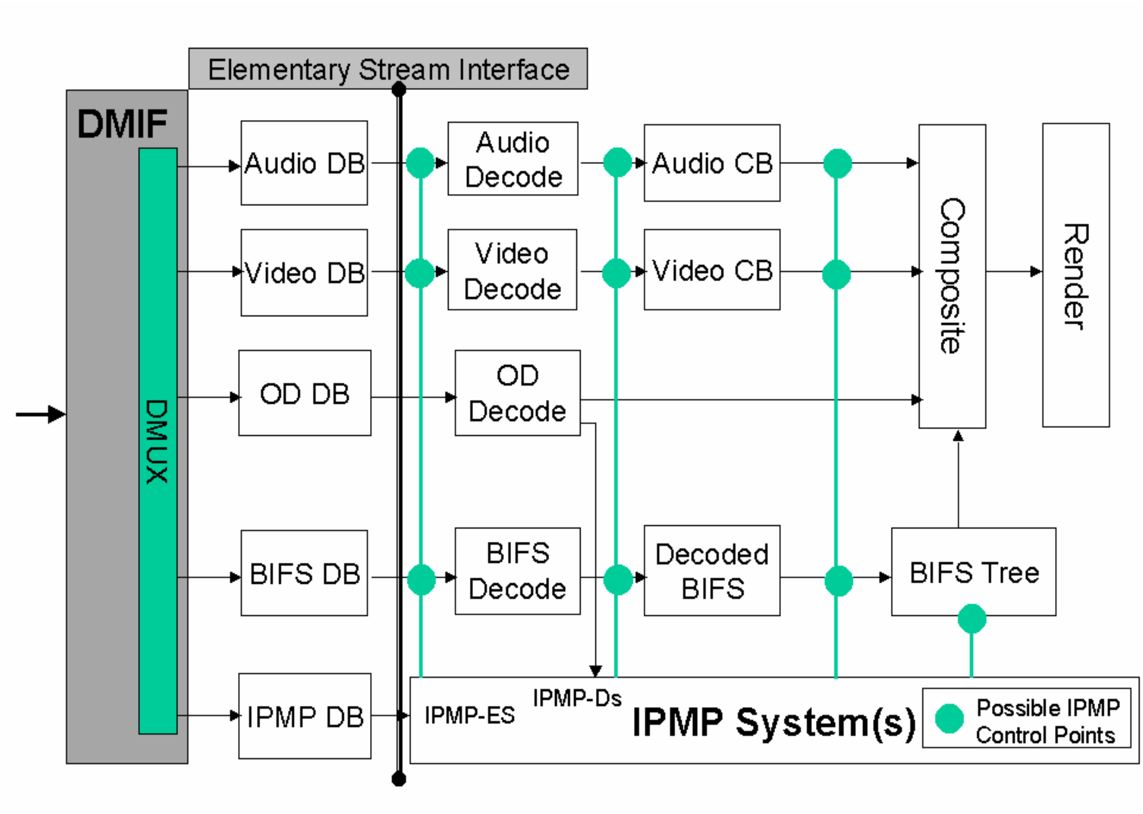


圖 1 MPEG-4 的 IPMP Hook 架構[4]

在IPMPX下，依據IPMP Tool Identifier，系統可以使用的Tool可分為下列幾類：

- Unique Implementation -- 代表唯一的獨立實作
- Parametric Description
  - 根據給定參數描述，符合特定條件的實作
  - 實作所需的參數可透過Parametric Configuration Information傳入
- List of Alternative Tools
  - List中指明的所有Tool皆視為等效實作
  - 可符合Parametric Description描述特定條件的所有等效實作

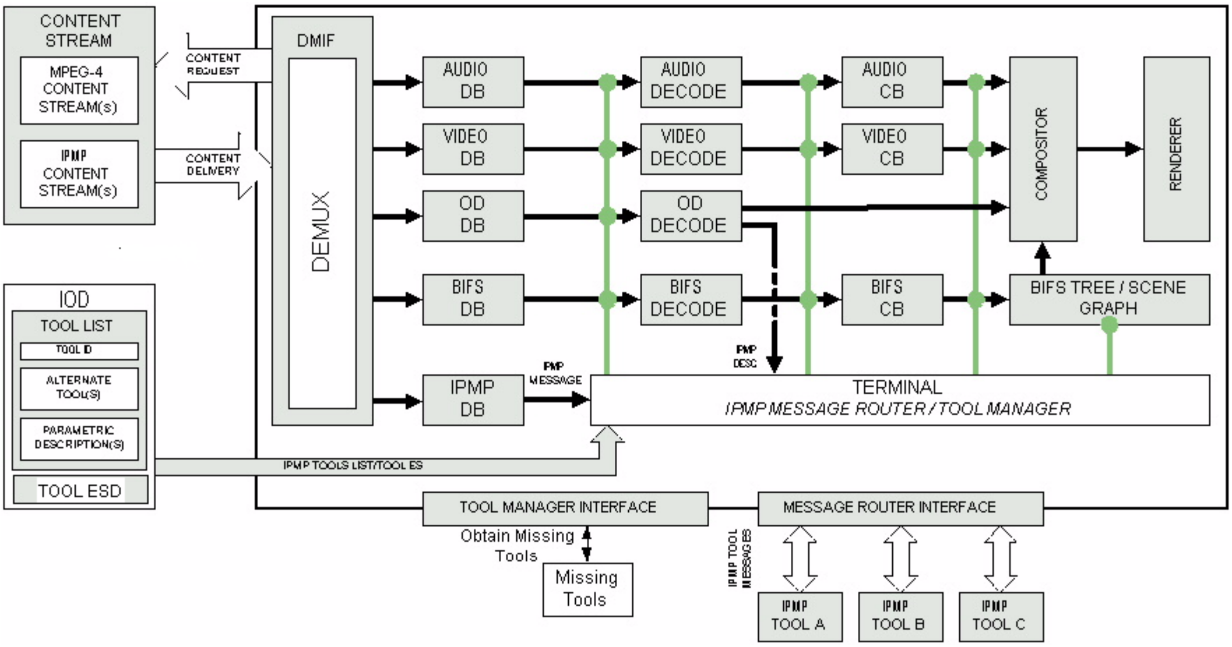


圖 2 IPMPX 與 MPEG-4 系統[5]

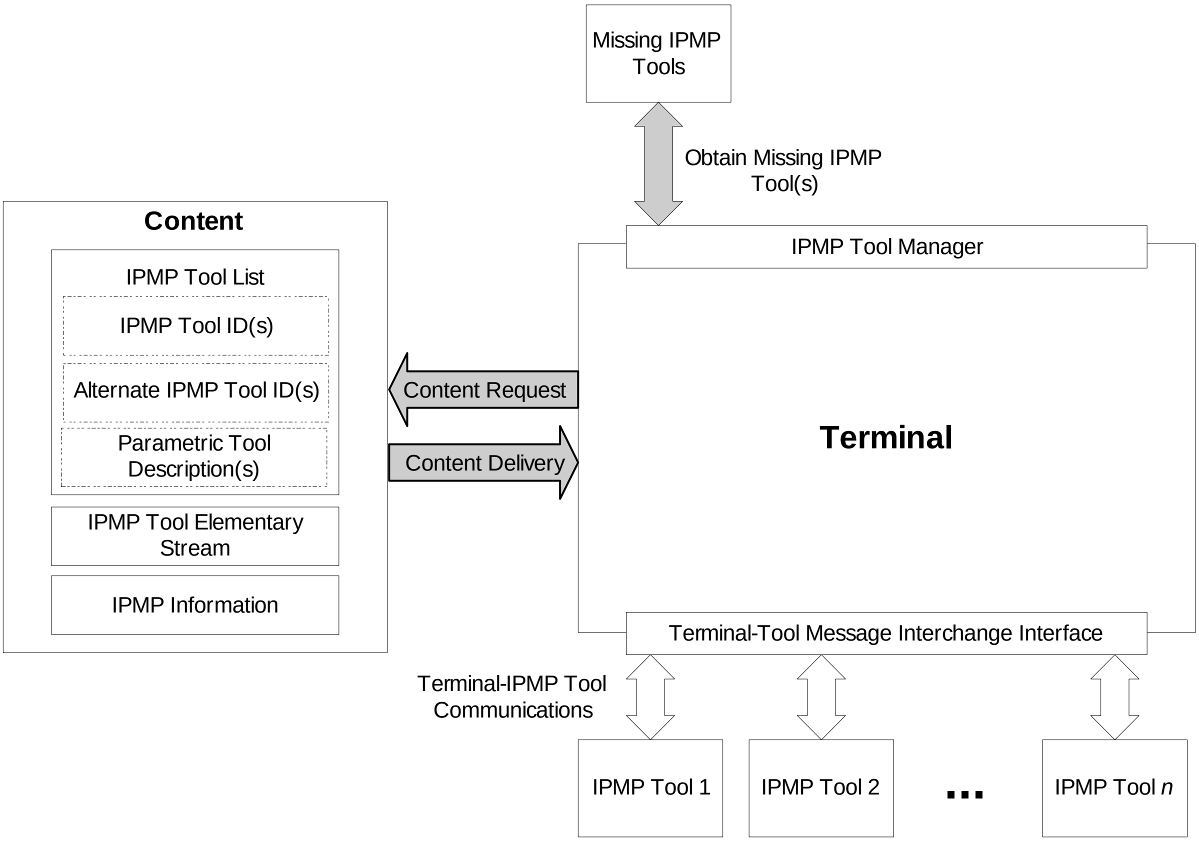

Missing IPMP Tools
Obtain Missing IPMP Tool(s)
Content
IPMP Tool List
IPMP Tool ID(s)
Alternate IPMP Tool ID(s)
Parametric Tool Description(s)
IPMP Tool Elementary Stream
IPMP Information
Content Request
Content Delivery
IPMP Tool Manager
Terminal
Terminal-Tool Message Interchange Interface
Terminal-IPMP Tool Communications
IPMP Tool 1
IPMP Tool 2
...
IPMP Tool n

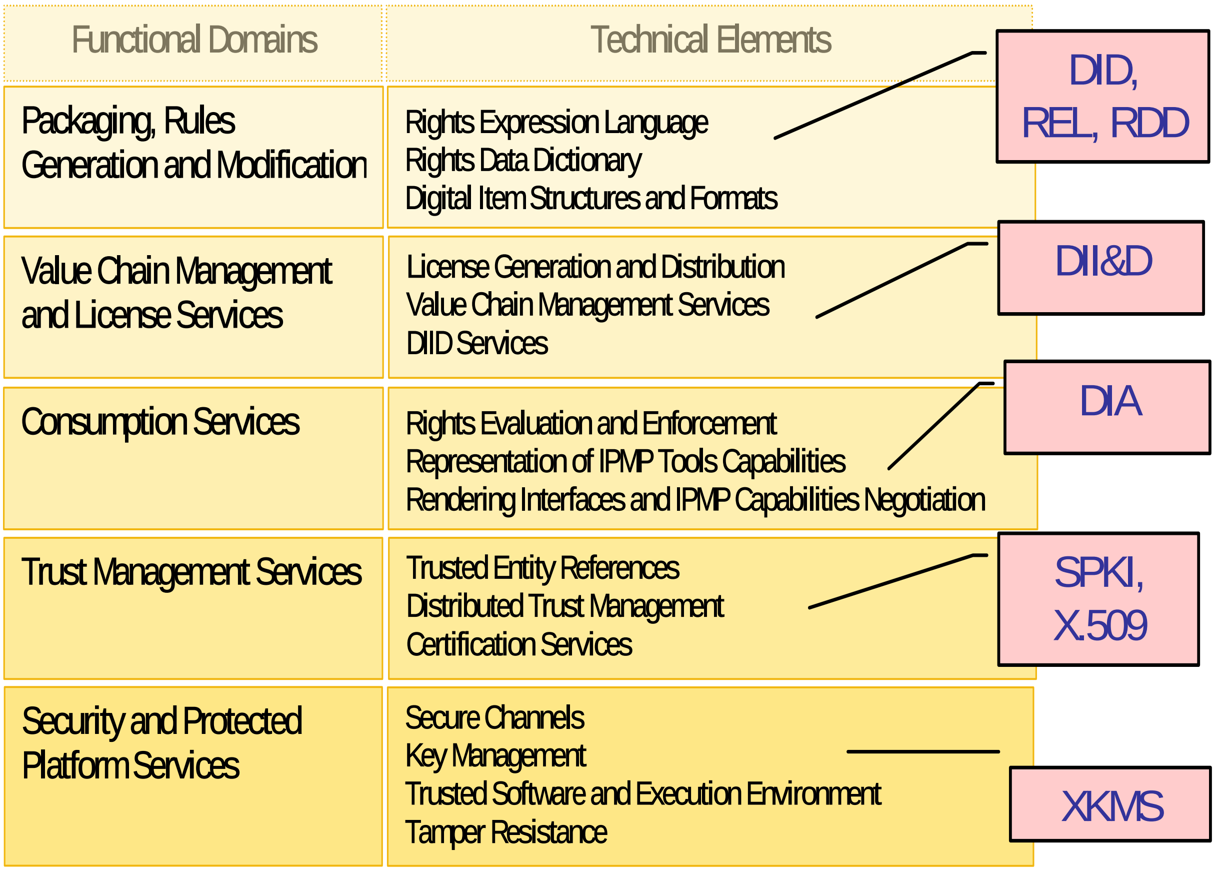
圖 3 IPMP 使用流程示意圖[5]

IPMPX 的起始資訊也是透過 MPEG-4 系統串流傳遞至終端。接收端對於這部分的處理流程如圖 3 所示，以下將逐步解說：

1. 使用者要求特定的數位內容
   - IPMP 需求先於內容需求
   - 在傳送內容前先完成 access control
2. 取得 IPMP Tools description
   - 從內容中取得所需的 Tool List
   - 根據 Tool List 解析可建立的 Tool
3. 取得 Tool
   - 找尋並下載無法在終端機找到的 Tool，這項功能不在標準定義內
4. 建立 Tool instance
   - 依據參數於對應的 control point 實際建立 Tool
   - 從內容中取得其他 IPMP information
5. 初始化 Tool 與更新 Tool
   - 給予各 Tool 起始 IPMP information
   - 終端開始消化允許處理的內容
   - 開始處理內容後，終端機可根據收到的 IPMP 命令，重新初始化或更新 IPMP Tool

MPEG-21[6] IPMP[7]需求文件尚在草稿階段，不過可從其中窺見 MPEG 委員會對於未來IPMP 系統的構想。在 MPEG-21 的數位內容描述與散佈架構下，MPEG-21 IPMP 涵蓋的範圍非常廣，從某個特定數位內容的生命期來看，IPMP 作用的範圍包含所有觸及權限管理的軟硬體及技術。從技術範圍來看，IPMP 同時具有系統性與切割性。就系統性而言，IPMP 必須整合散佈在整個商業價值鏈的各個部分，包含各點之間信賴關係的建立與管理，商業模式的操作方式與架構等等。另一方面，切割性來自於 IPMP 與許多技術領域相關，例如應用整合、元件化軟體、行動程式碼、作業系統、分散式服務、數位內容管理、安全服務等等。

圖 4 為 IPMP 的抽象系統模型，這個模型從兩個維度來說明。垂直方向代表 IPMP 技術的層次，越上面的越高層。水平方向代表各功能與相對應的技術，可以看出 IPMP 的切割性。最右側的說明方塊標示在 MPEG-21 中對應的標準工具。以下將對圖中五大功能簡略描述：



| Functional Domains | Technical Elements |
|---|---|
| Packaging, Rules Generation and Modification | Rights Expression Language<br>Rights Data Dictionary<br>Digital Item Structures and Formats |
| Value Chain Management and License Services | License Generation and Distribution<br>Value Chain Management Services<br>DIID Services |
| Consumption Services | Rights Evaluation and Enforcement<br>Representation of IPMP Tools Capabilities<br>Rendering Interfaces and IPMP Capabilities Negotiation |
| Trust Management Services | Trusted Entity References<br>Distributed Trust Management<br>Certification Services |
| Security and Protected Platform Services | Secure Channels<br>Key Management<br>Trusted Software and Execution Environment<br>Tamper Resistance |

圖 4 Abstract IPMP System Model[7]

- Packaging, Rules Generation and Modification 的相關技術：
  - 數位內容包裝技術，通常包含內容與 meta-data、智慧財產權資訊的整合包裝，最後產出為受到保護的數位內容。
  - 著作權與相關資料表示法的標準規格。
  - 數位內容的產生與修改。
- Value Chain Management and License Services 的相關技術：
  - 在商業模式下數位內容的散佈與使用。
  - 包裝好的內容除了可以直接實體散佈出去外，也可以採用類似 URL 的方式只將參考連結散佈到使用者手上。
- Consumption Services
  - 內容的使用者，包含一般使用者及價值鏈中的內容處理者，皆是透過 Consumption Service 使用數位內容及服務。

- Trust Management Services
    - 可用於管理 IPMP 各元件間的信賴關係，這些元件包含執行程序與結構。
    - 這部分的功能可能包含下列幾種形式：
        - 信賴模式，例如點對點模式、web 模式、階層模式
        - 價值鏈中的軟硬體認證
        - 價值鏈中的軟硬體註冊
        - 安全平台或服務的個人化設定
        - 安全機制的生命期，包含更新與廢止等管理
        - 使用者憑證與密碼管理
        - 可信賴的對時服務
- Security and Protected Platform Services
    - 要與受到保護的系統或服務溝通，或是使用受保護的內容，本身必須是安全機制中的一環，也必須能夠向對方證明自己的安全能力。
    - 這些安全機制可能為：
        - 對抗攻擊的能力
        - 作業系統執行環境的安全管理
        - 軟體個人化
        - 對使用者存取遠端或近端資源的認證

由以上說明，我們可以了解 MPEG IPMP 的目標是建立一個從數位內容產生到消失，整個過程都加以保護的散佈與使用機制。現階段相對完備的為 MPEG-4 IPMPX[10]與 MPEG-2 IPMP[10]，這二者採用 Virtual Terminal 概念，在現有系統之外，加上 IPMP 介面。為了使 IPMP 工具之間溝通順暢，採用 Message 方式交換 IPMP 資訊及命令。MPEG-4 IPMPX 與 MPEG-21 IPMP 最大不同點在於，前者作用範圍在 Terminal 端，後者則散佈在整個傳遞架構各環節中。

MPEG-21 目前有一個整合型的參考軟體，我們的目標之一，是把 IPMPX 實現於這個軟體架構之中。圖 5 為 MPEG-21 Resource Delivery Test Bed[2]的架構，主要分為三部份：Server 端、Client 端、以及網路模擬器。網路模擬是用 Linux 機器加上 NistNet 核心模組作成的 router，可以透過程式控制兩張網路卡之間的流量。使用者介面以 Java 完成，可以即時顯示 Network profile 指定的頻寬與實際使用的頻寬。Server 端從 Media database

讀取 media data 後，經過 DIA 模組針對給定條件作 adaptation，再經由 streamer 切割成網路封包往外傳送。Client 端將接收到的封包重整，供 decoder 取用，decoder 將解碼的結果放到 output buffer 給 player 播放。Server 端與 Client 端的 controller 負責訊息的溝通，例如起始參數或網路頻寬狀況，必要時可透過 controller 要求重新傳送遺失的封包。至於圖中的 IPMPX 相關部份，我們將於下節詳述。

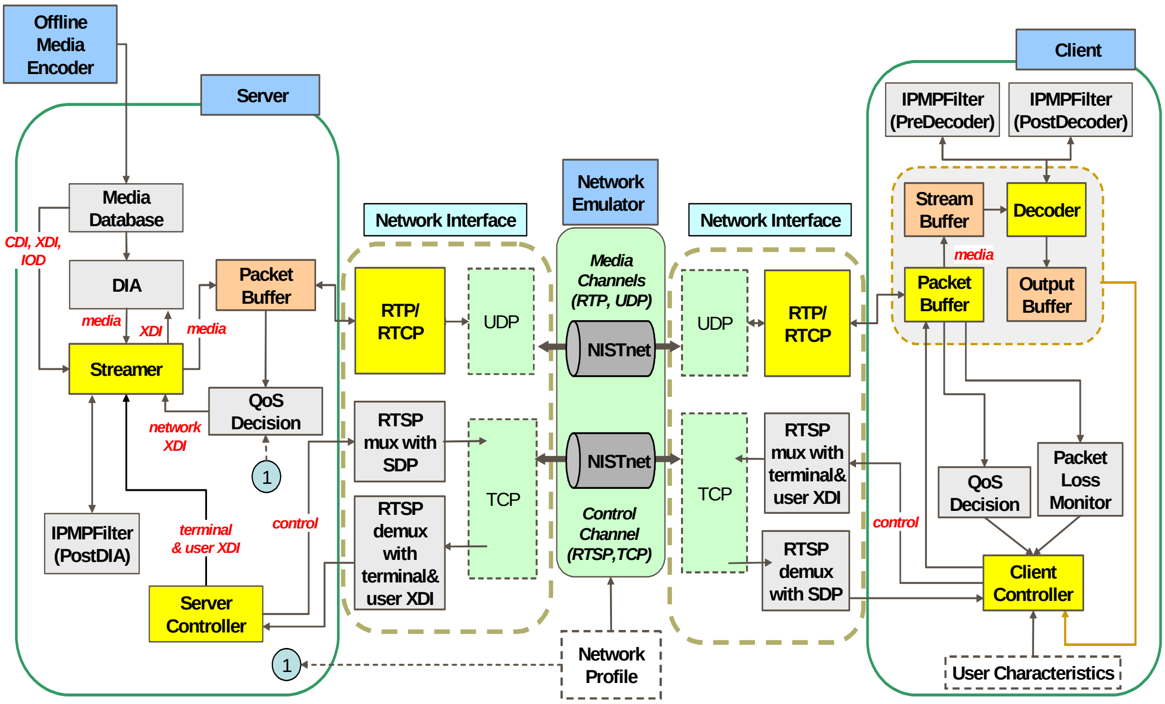


圖 5 MPEG-21 Resource Delivery Test Bed Architecture[2]

## C. 模擬與實驗

本模擬重點為 IPMPX[9][10]，相關的參考軟體不多，一為 Craig A. Schultz 所作，另一個為 MOSES。兩者皆利用 IM1 為 MPEG-4 player。IM1[8]，或是稱做 AHG (Ad-hoc group) on Systems Reference Software Implementation，是 MPEG 委員會中負責從事開發以及整合 MPEG-4 System 軟體的團體。IM1 軟體包括了所有 MPEG-4 System 中標準化的部分。MOSES 至今未有正式公開版本，但根據架構圖(圖 6)顯示，設計上像是 IM1 的加強版，因此與 Craig 的 IPMPX 實作架構(圖 7)有些許不同。

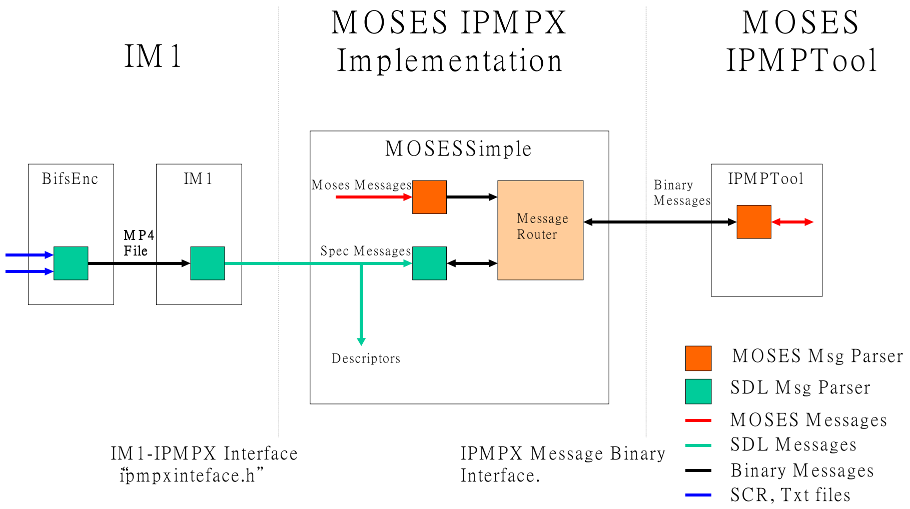
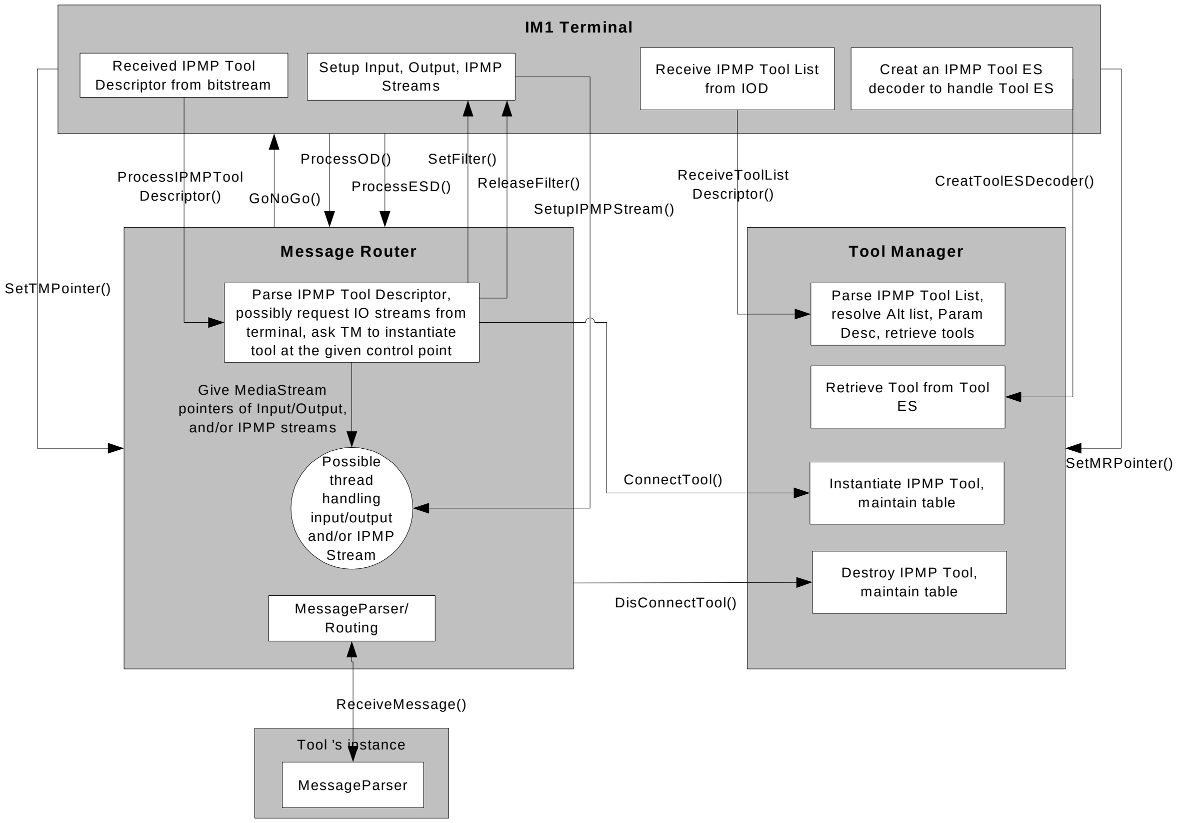


圖 6 MOSES IPMPX



IM1 Terminal
Received IPMP Tool Descriptor from bitstream
Setup Input, Output, IPMP Streams
Receive IPMP Tool List from IOD
Creat an IPMP Tool ES decoder to handle Tool ES
ProcessIPMPTool Descriptor()
GoNoGo()
ProcessOD()
ProcessESD()
SetFilter()
ReleaseFilter()
SetupIPMPStream()
ReceiveToolList Descriptor()
CreatToolESDecoder()
SetTMPointer()
SetMRPointer()
Message Router
Parse IPMP Tool Descriptor, possibly request IO streams from terminal, ask TM to instantiate tool at the given control point
Give MediaStream pointers of Input/Output, and/or IPMP streams
Possible thread handling input/output and/or IPMP Stream
MessageParser/ Routing
Tool Manager
Parse IPMP Tool List, resolve Alt list, Param Desc, retrieve tools
Retrieve Tool from Tool ES
Instantiate IPMP Tool, maintain table
Destroy IPMP Tool, maintain table
ConnectTool()
DisConnectTool()
ReceiveMessage()
Tool 's instance
MessageParser

圖 7 IM1 IPMPX (Craig's version) [10]

在之前的研究中，我們分析了 IM1 Core 中有關 IPMPX 的部分，了解 IPMPX 系統

的操作模式，並據此以循序漸進的方式，展示 IPMPX 的運作。IM1 IPMPX 的實作方式，是將 IPMPX 散佈於 IM1 各功能模組中。儘管這種架構可以達到前述 Virtual Terminal 概念，卻使得 IPMPX 缺乏模組化，不利後續研究。且分析程式碼後發現，IM1 IPMPX 無法獨立於 IM1 core 之外，這對未來與 MPEG-21 架構結合產生不少困擾。因此，我們根據 IPMPX 架構，重新設計軟體模組並實作，以達到下列目標：

- IPMPX 模組化，以程式庫型態與主系統連結。
- 不引用現有 reference software 的核心函式，盡量降低與特定系統的相依性。
- 預留與 terminal 銜接的機制，減少 porting 的困難。

軟體模組的設計上，我們盡可能保留文件上所描述的概念性物件，例如 MessageRouter 與 ToolManager 等等(圖 8)。而在 API 的設計上，我們則參考 IM1 IPMPX 的設計，並稍加延伸，例如加上 timestamp 參數以利 terminal 傳遞系統時間給 tool。在實作上，為求與標準格式相容，我們採用 PSL MPEG-2 IPMPX[11]中的 MessageInterface 函式庫產生與解析 ToolMessage。較為特別的是 context 的設計，ToolMessage 利用 context ID 決定 routing 方式，但 IPMPX 標準中並未定義 context 的實際結構，因此我們根據可能的使用情況，將 context 設計成樹狀結構，依照 top、object、elementary stream、IPMPTool 四個階層加以關聯，並自動給予各 context 識別編號。另一個標準中沒有定義的物件為 IPMPFilter，根據相關描述，我們歸納出 IPMPFilter 應該是一個 container，允許至多 256 個 IPMPTool 串接，而資料串流則依序通過各 Tool。

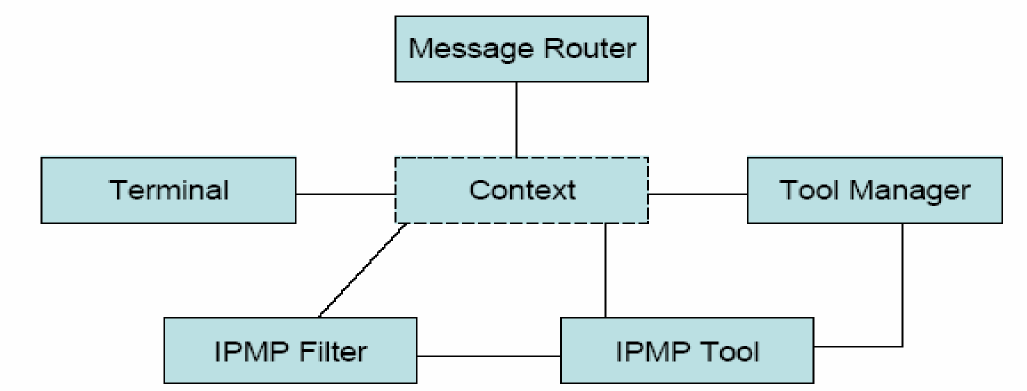


圖 8 重新設計的 IPMPX 模組關係圖

圖 9 所示為 IPMPX 模組與 MPEG-21 Testbed 各模組間的關係。為了整合 IPMPX，MPEG-21 Testbed 作了如下修改：

- ServerController 必須能送出 initial object descriptor (IOD)。
- ClientController 必須能接收 IOD。
- Server/Client Controller 必須能收發 IPMP Device Message。
- Streamer 取得資料後，必須先經 PostDIAFilter 處理才送出。

- Decoder 取得資料後，必須先經 PreDecoderFilter 處理才能 decode。
- Decode 完成後，必須經 PostDecoderFilter 處理才能送至 output buffer。
- Server/Client 必須在開始傳送 media 前 initialize IPMP 子系統。

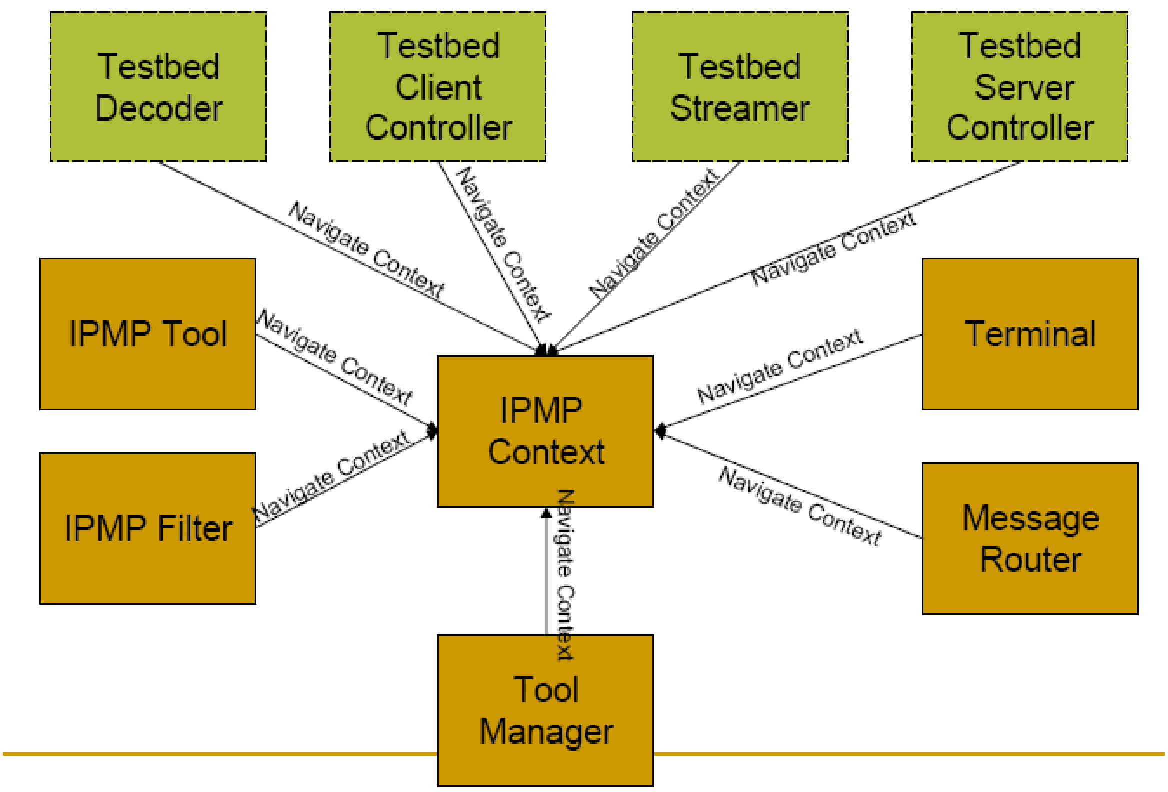
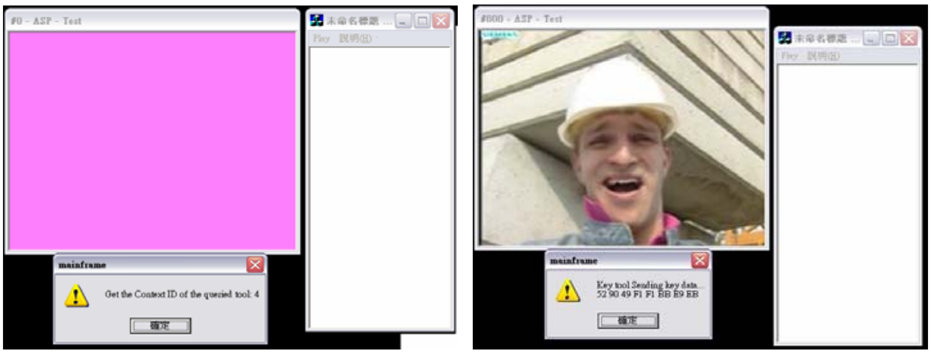


圖 9 IPMPX 模組與 MPEG-21 Testbed 其他模組的關係

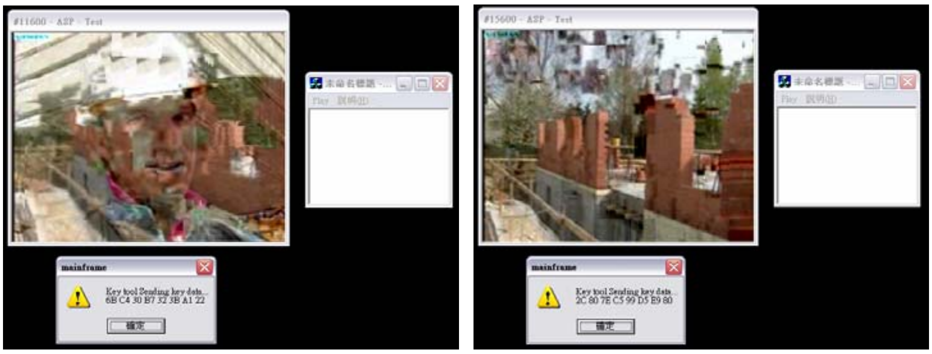
最後，我們用一個加密傳送影像的系統展示我們的研究與實作成果。圖 10 所示 client 端程式啟動 IPMPX 子系統，並週期性換 key 進行 DES 解密的過程，就使用者而言，這些動作都是隱藏在 virtual terminal 完成，只會看到播放程式播放短片。接著我們刻意在某個時間區間給定錯誤的 key，對 decoder 而言，因為無法經由解密過程取得正確的影片資料，所以解碼過程發生錯誤，如圖 11 左側照片所示。當恢復取得正確的 key 之後，decoder 根據正確的資料逐漸修復播放的畫面(圖 11 右側照片)。



Get the Context ID of the queried tool: 4

Key tool Sending key data...
52 90 49 F1 F1 BB B9 EB

圖 10 程式啟動 IPMPX 與正常解密



Key tool Sending key data...
6B C4 30 B7 32 3B A1 22

Key tool Sending key data...
2C 80 7E C5 99 D5 E9 80

圖 11 解密用的 key 錯誤造成解碼錯誤

## D. 結論

本研究的主要目的是探討MPEG IPMP，除了以MPEG-4 IPMPX為主要研究對象外，我們也對MPEG-21 IPMP的設計進行了解。IPMPX以Message為交換IPMP資訊的方法，使得各Tool間的互通性大增，也讓IPMPX比IPMP Hook介面更為可行、更具彈性。然而以現今的多媒體系統而言，要達到 MPEG-21 那種理想並不容易，因此我們以MPEG-4 IPMPX為參考，在追蹤過IM1程式架構與了解IPMPX運作原理後，我們發現IPMPX 與 IM1 間的相關性仍然太高，如果想把 IPMPX 抽離 IM1，成為像概念圖上的MPEG-4 Terminal與IPMP Virtual Terminal，有很多困難。因此我們重新設計具模組化的 IPMPX 系統，作為未來研究 IPMP 的基礎。除了把標準文件中抽象的元件對應到實際的架構之外，也考量未來整合的需求，盡可能降低與現有 reference software 的相依性。

接著，我們將IPMPX整合到MPEG-21 Testbed中，並修改相對應的部分，而修改過程也間接印證我們當初的設計，的確讓 IPMPX 子系統保留相當的獨立性，並未因MPEG-21 Testbed的架構而造成IPMPX子系統架構上的變動。研究的最後，我們以一個加密傳送影片的系統展示實作成果，這個系統用來追蹤 IPMPX 的動作是否如標準所描述，從啟動系統、解析IPMPTool、掛載tool、傳遞message、以及加密解密tool的配對運作等等。總結說來，本次研究將我們過去對IPMP的了解進行實作，不僅設計概念上力求與MPEG IPMPX相容，並結合MPEG-21 Testbed驗證與MPEG標準格式的相容性。

這部分成果在2004年7月加入MPEG-21 Multimedia Test Bed中，已成為Test Bed軟體的一部份。Multimedia Test Bed提案至MPEG標準組織，2003年12月成為委員會草案，全案期望在一年左右完成標準化。

# 第二部分 數位浮水印與密碼方法結合應用於多層次編碼

## A. 背景與目的

數位媒體的壓縮與傳輸技術持續演進，使得今日處於數位媒體爆炸的時代。對使用者而言，不僅藉此得到更高品質的影音，也可經由更多樣化的傳遞途徑獲得資訊。而對於提供內容的一方，經由數位傳播，無形中也獲得更廣的觀眾群。然而，數位傳播也衍生出新的技術課題。傳統的媒體傳播方式為類比媒介，例如錄影帶或膠捲片，這類型媒介在轉錄的過程中會一再失真，且媒介本身的價格無法消除，因此盜版的成本與品質受到一定的限制。反觀數位媒體，媒體資料並不依附在特定媒介上，因此可透過任意傳播管道，完全不失真地轉存至任意媒介，也就是說，盜版者可以用更低廉的價格取得與原始媒體一模一樣的品質。

儘管數位媒體帶來這類問題，但是技術的進步與使用者的需求，已經使得數位媒體成為傳播的趨勢。至於保護版權的方法，目前較流行的主題有兩類，一是限制使用者存取媒體的權限，二是在媒體中嵌入數位浮水印。前者主要作用於傳遞過程，以特定方式，排除不具權限的使用者。在網路傳播上，最常使用的就是密碼方法，簡單說來，傳送一端先將資料加密，送到用戶端，經解密後才能得到正確的媒體資料。只要使用的方法適當，即使傳遞過程中資料被攔截或偷聽，竊聽者是無法還原資料的，此外，經由特定密碼方法保護的媒體，也可對抗惡意第三者送出假造資料欺騙用戶端。至於數位浮水印的應用，則是著重在媒體使用過程，讓具有存取權限的使用者，無法任意轉換或散佈影音資訊。從嵌入強度來分，數位浮水印可分為可見與不可見兩類，可見的數為浮水印類似傳統類比浮水印，肉眼即可識別，而在數位應用上，多半討論不可見的浮水印，利用特定演算法，在不影響肉眼識別範圍內，嵌入浮水印資料。從特性來分，數位浮水印可分為兩類，一類稱為 fragile watermark[12]，當媒體資料受到變動而連帶稍許影響嵌入資訊時，浮水印將被大幅破壞到無法辨識的地步，此類浮水印可用於驗證媒體的完整性。另一類稱為 robust watermark[13]，其特點是當媒體在一定破壞程度內，浮水印維持在相當的可辨識度，此類浮水印主要用於版權識別。從商業角度來看，robust watermark 可以用來嚇阻惡意轉手散佈的使用者，因此應用空間較 fragile watermark 大。

數位影音內容應用在廣播環境中，面臨另一個問題，就是用戶端的接收解碼設備並非單一規格，例如電視或手機，就有不同的解碼速度與播放速度。為了應付各式各樣的接收設備，一個方法是應用多層次編碼將原始影音資料預先解析成各種解析度，並加以編碼，接收端接收與處理越多資料，就能合成越高品質的影音資訊[14][15]。多層次編碼除了用於解決上述問題外，也可用於提供不同品質的多媒體給不同使用者，例如依據付費的多寡，決定能收看的畫面解析度。這裡牽涉到的問題是，如何在多層次編碼的廣播架構下，進行存取權限控管[16][17]。其衍生的問題是，廣播環境中，當雜訊干擾以致資料流失時，是沒有辦法重送資料的，只能設法補救，然而密碼方法往往不允許資料錯誤。因此，如何在資料錯誤時盡量不影響解碼，讓使用者得到最接近期望的品質，將

是本研究的重點。

## B. 研究步驟 – 利用數位浮水印達到多層次存取限制

如前節所述，多層次編碼適用於異質接收端的廣播環境，且適於提供多重品質的應用。照多層次編碼的特性來看，原始資料被解析分割為多個層次，下層資料代表低解析度，上層資料代表對底層的加強，經過合成運算，可以得到較高解析度的資料。在廣播環境中，我們無法針對特定使用者群送出適合其播放能力的資料，而是將所有層次廣播出去，由使用者端自行決定接收哪些層次。在此種模式下，使用者分群是以接收的層數為依據。若想保護資料排除沒有權限讀取的人，一個直接的想法就是：將不同層次資料分別加密，權限越高的使用者能解密的層數越多。就密碼方法的角度來看，就是不同層用不同密碼方法，或使用同方法但不同 key，不論是哪一種方式，都是為了區別不同層次的存取權。

如果只是單純為了區別資料存取權，那麼分層加密已經足夠。但是對多媒體資料而言，有些需求與一般資料不太一樣，最明顯的一點就是，在即時性多媒體串流應用中，少部分資訊漏失是不可避免的，而依照傳輸架構的不同，有些環境可以要求重新傳輸漏失的資訊，不過對廣播而言，因為必須符合即時性與缺乏回授管道，系統無法知道哪些資料已經漏失，所以遺失的資訊是無法重傳的。而對於數位廣播系統，除了資料串流外，加解密用的 key 如何傳輸與保護也是一個重要課題。例如數位電視廣播標準(DVB[18])所採用的是將 key 的有效時間分為三類，有效時間長的用來保護有效時間短的，而且其加密方法也較複雜(強度較高)。用來保護資料串流的密碼方法最簡單，且換 key 頻率最高。另一個則是 key 與 content 的同步問題，在很多狀況下，在網路上傳播的封包並不保證接收時間，也就是說，受到保護的串流資料可能比 key 早到，最糟的狀況是，key 的封包遺失，又無法重傳，就會出現整段資訊無法解密的情形。

我們的研究，是利用 watermark 技術，把上層解密所需的 key 嵌入下層中，如此一來，key 與 content 同步的問題得以解決。再藉由 robust watermark 的特性，讓整個解密與解碼過程受到相當程度保護，換言之，如果發生傳輸錯誤，我們仍能由下層取出內嵌的 key，因此不會影響對上層的解密。當然，下層資訊遺失必然對合成最終資訊有所影響，但我們認為只要能妥善利用 error concealment 技術，應可修復至可接受的範圍，這種做法會比出錯時直接捨棄所有 enhancement 資訊更能滿足使用者。

圖 12 為接收端架構，$X_i$ 為收到的加密資料，經解密後得到 $E_i$，再與之前解碼所得的 base layer ($B_{i-1}$)合成得到新的 base layer ($B_i$)，其中解密用的 key ($K_i$)的來源，是由 $B_{i-1}$ 取出 watermark ($W_i$)，經過 user key ($G_i$)解密後，得到真正隱藏的資訊($F_i$)，包括解密用的 key ($K_i$)與下次抓取 watermark 的參數($P_i$)。以上敘述可以用下列各式表示：

- $B_i$ = compose($B_{i-1}$, $E_i$)

- $E_i$ = decrypt$_e$($X_i$, $K_i$)
- $W_i$ = extract($B_{i-1}$, $P_{i-1}$)
- $F_i$ = decrypt$_f$($W_i$, $G_i$)
- $K_i$ = key($F_i$)
- $P_i$ = param($F_i$)

整個解密與解碼過程是反覆進行，後一次動作需要的參數皆由前一層 base layer 而來。至於起始參數則有如下考量：

- 當整個媒體串流受到密碼保護，則解密所需的 $K_0$ 必須由其他管道是先獲得。如果 $B_0$ 當成 preview layer 不加密，則沒有這個問題，只需跳過解密步驟即可。
- 如果抓取 watermark 需要起始參數，則 $P_0$ 也必須由其他管道是先獲得。
- 用來保護 $K_i$ 與 $P_i$ 的 user key $G_i$，是本系統中必須事先透過其他管道取得的參數，這取得管道可伴隨訂購行為發生。

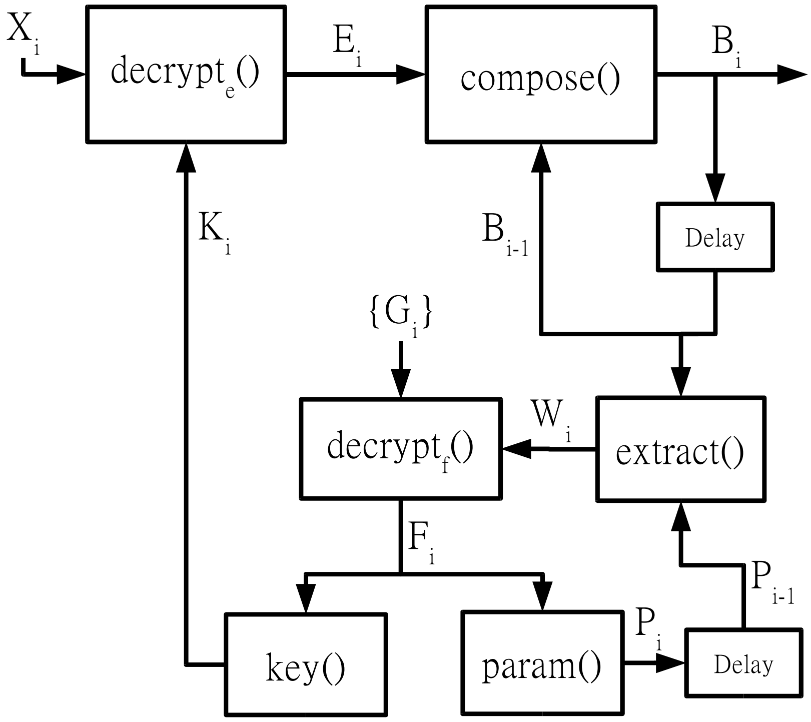


圖 12 多層次解密與解碼

隨著不同的多層次編碼方法，傳送端的架構也會有所不同，圖 13 所示為其中一種可能的架構。先把 $K_i$ 與 $P_i$ 合併成 $F_i$，再用 $G_i$ 加密產生 $W_i$，利用 $P_i$ 參數將 $W_i$ 嵌入 $B'_{i-1}$ 後得到 $B_{i-1}$，然後分析 $B_i$ 與 $B_{i-1}$ 得到 enhancement $E_i$，最後將 $E_i$ 以 $K_i$ 加密送出 $X_i$。

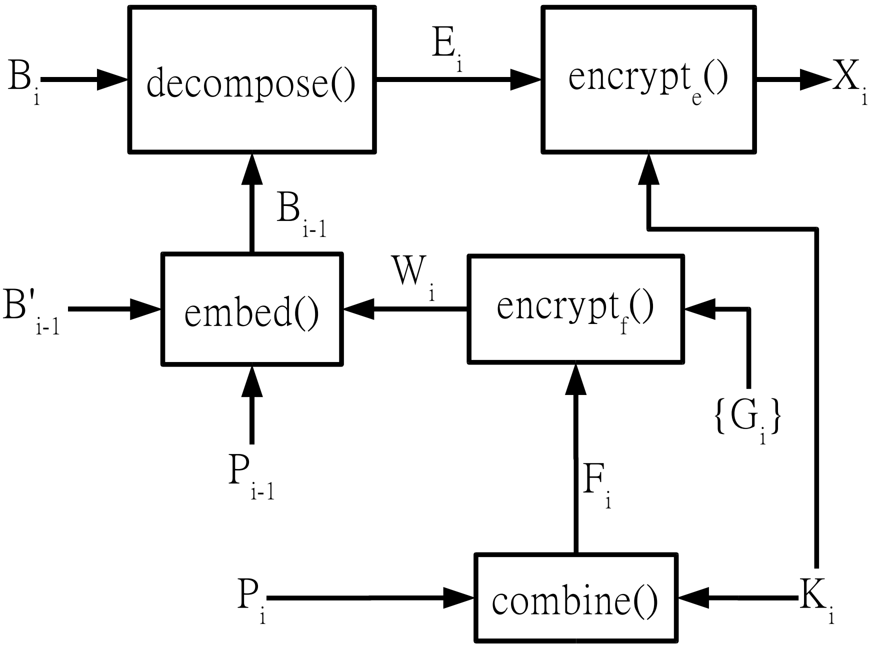


圖 13 多層次加密與編碼

## C. 模擬與實驗

本次研究我們用單張照片進行概念性的模擬。首先，我們從 1024x1024 的照片($B_1$)解析出 512x512 的 base layer ($B'_0$)，另外將一個 64-bit 的字串(**NCTU-DEE**)當成加密用的 key，產生 binary watermark (圖 14)後嵌入 base layer，接著將含有 watermark 的 512x512 照片($B_0$)當成最終的 base layer，利用 $B_0$ 與 $B_1$ 就可得到 enhancement layer ($E_1$)，最後我們用 DES[19]方法對 $E_1$ 加密。

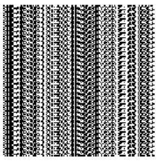
**NCTU-DEE ➔** 

圖 14 Binary watermark for embedding

圖 15 所示為傳送前對應的 base layer 與 enhancement layer，其中 base layer 因為嵌入了 watermark，使得 PSNR 變為 39.24 dB，但在視覺上與原始 base layer 沒有區別。

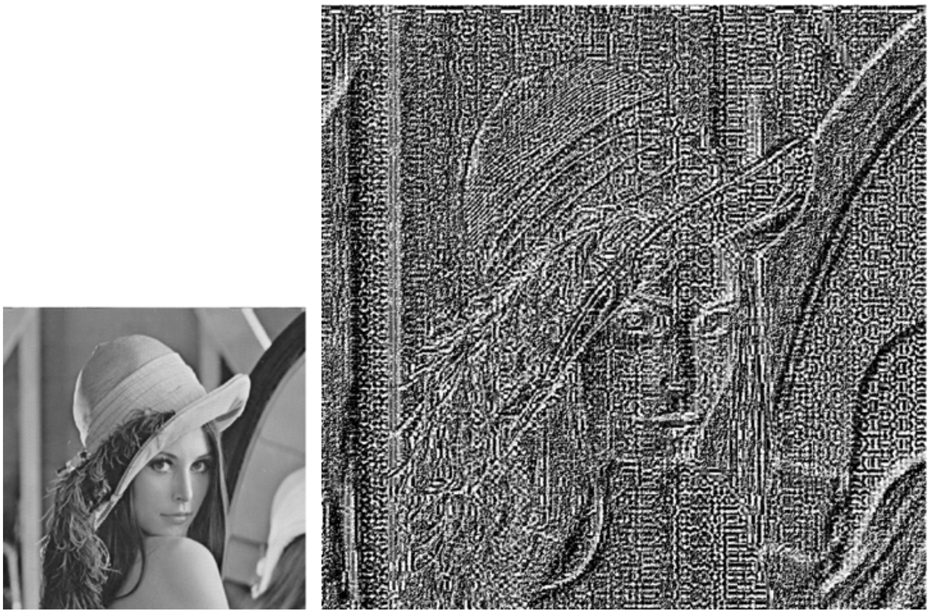
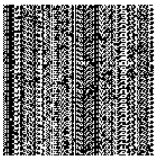


圖 15 傳送前的 512x512 base layer 與 1024x1024 enhancement layer

因為此次為概念驗證，所以我們並未將處理過的照片以特定標準進行壓縮。為了模擬傳輸錯誤，我們將 base layer 切割成 8x8 的區塊，並以亂數選取 10%的區塊丟棄[20]。從破壞後的 base layer 抓取出來的 binary watermark 如圖 16 所示，其 bit-correct rate 為 92.74%，而從這個 watermark 中，我們依然可以推出原本的 64-bit key，因此接收到的 enhancement layer 資料可以正常解密。圖 17 為不做任何 error concealment 的合成結果，可以看到除了被破壞的 10%區域之外，其他部分皆可解碼回原本的解析度。



➔ **NCTU-DEE**

圖 16 從傳輸錯誤的 base layer 抓取出的 watermark，其 bit-correct rate 為 92.74%

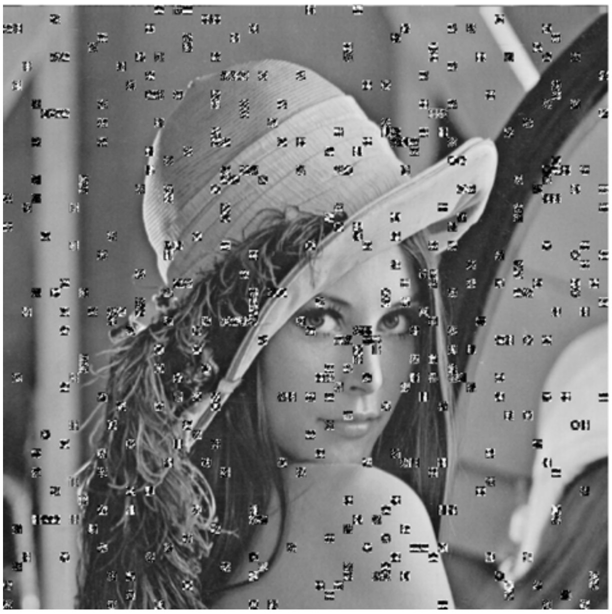


圖 17 傳輸錯誤情形下解碼的結果

## D. 結論

在本研究中，我們提出一個方法，將多層次編碼、密碼方法、與數位浮水印技術結合，試圖解決數位廣播環境下面臨的幾個問題：

- 使用者存取權限：經由多層次加密概念加以解決，將不同層次的資料以不同 key 保護起來，權限低的人無法解開較高層的加密資料。
- key-content 同步：將 key 以 watermark 形式嵌入 content 中，取得 content 同時也取得 key，不會有同步問題。
- 傳輸錯誤影響解碼品質：使用 robust watermark 對抗錯誤，在一定程度破壞下，key 仍能正確取出，不致影響 enhancement layer 的解密與解碼。傳統加密方式可能因為小部分錯誤而導致無法取得 enhancement layer，相較之下，我們的方法可以獲得較多正確資訊，對於後續 error concealment 或 post-processing 較有幫助。

透過初步的模擬實驗，我們驗證了想法的可行性，接下來的目標將是結合標準的壓縮格式，並研究傳輸錯誤時可能發生的狀況。

# 第三部分 畫面間小波視訊編碼之開發與研究

## A. 背景與目的

過去幾年來，隨著網際網路以及廣播事業的發展與普及，多媒體的應用的標準日趨重要。多媒體視訊的應用包括了視訊會議、網路電視、數位電視等。對於各種不同的應用，也有不同的標準。目前制定標準的主要兩個機構是 ITU 以及 MPEG，分別是針對於視訊會議以及數位電影不同的應用來規劃。但近幾年，因為應用的變化幅度大，要求的規格也有許多變化，因此新訂定的標準也就沒有特別為某種應用做設計，而反以可調性為主。可調性則又可分類為三種，一是所謂的空間可調層次性(Spatial Scalability)，二是時間可調層次性(Temporal Scalability)，三是位元率可調層次性(Bitrate Scalability)。MPEG-4 的 FGS 則是頻寬可調性的一個例子。這種可調層次性的壓縮方法，就可以應用到行動視訊(Mobile Video)這種低頻寬的應用，或是頻寬變化頻繁的網路電視等等。視訊畫面間小波編碼技術(Interframe Wavelet)是近幾年被提出的壓縮演算法，其優點就在於全可調層次性(Full Scalability) — 同時兼具三種可調層次性，所以利用一次壓縮好的視訊，就可以提供給各種應用。而其壓縮效能在高位元率時跟目前標準制定的先進視訊編碼技術(ISO/IEC 14496-10 Advanced Video Coding, AVC)十分接近，如圖 18。

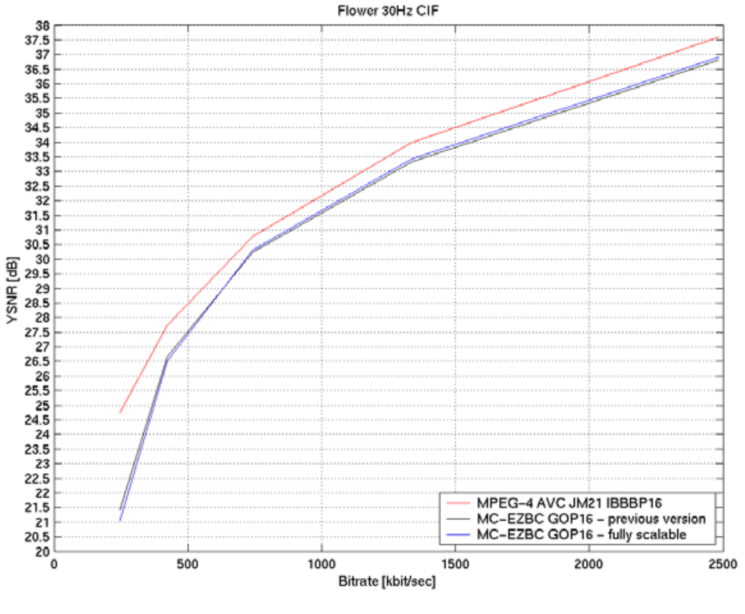


圖 18:視訊畫面間小波編碼技術與先進視訊編碼技術之壓縮效能比較

在視訊壓縮的研究上，演算法的發展已經過幾十年的發展。許多的演算法經過標準會議嚴密的測試訂定而成為標準，例如 MPEG-1、MPEG-2 等多媒體標準。現行的標準當中，演算法主要採用的架構是混合式時間與空間域編碼法(hybrid coding)。在時間軸上利用動態補償的方式，將時間軸上多餘的訊號移除，空間則是利用轉換編碼的方式(Transform Coding)將能量集中，最後經過熵編碼(entropy coder)將資料壓縮。在空間軸

上，最常採用的轉換編碼的方式是以離散餘弦轉換(discrete cosine transform)為主。而近幾年，採用次頻寬轉換(subband coding)或小波轉換(wavelet transform)的越來越多。以JPEG2000為例，則完全採用小波轉換。利用次頻寬轉換或是小波轉換可以省去離散餘旋轉換的方塊狀瑕疵(blocking artifact)。也有人研究過在時間軸上直接做轉換編碼，不過其編碼效果在動作大的影像並不好。視訊畫面間小波編碼技術(Interframe Wavelet)在時間域沿著物體移動軌跡，用Haar Filter在時間軸上做濾波(Temporal Filtering)，把每兩幅影像就可分為和(Sum)與差(Difference)，之後再做單一影像的壓縮，其主要架構如圖19所示。

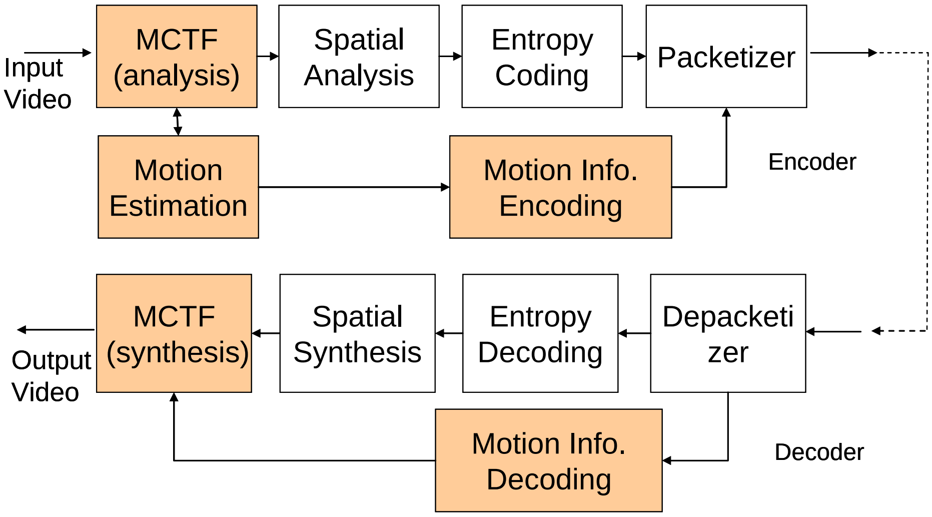
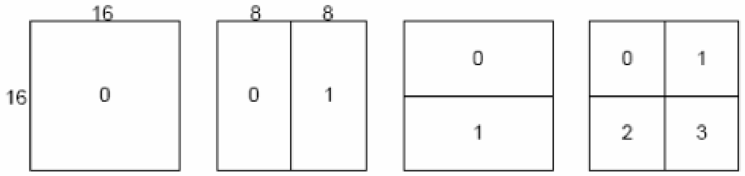


圖 19: 視訊畫面間小波編碼技術之架構圖

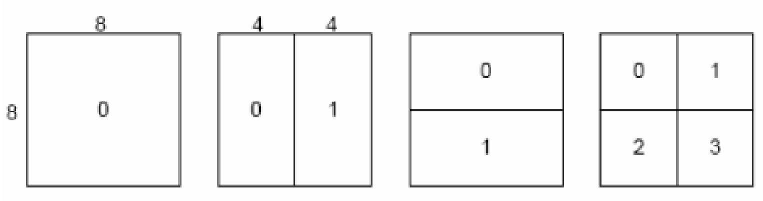
## B. 研究步驟 – 移動補償時間濾波之改良與可調式移動資訊

### B.1 移動估測

本計畫利用AVC中之移動估測方法對移動補償時間濾波改善，在AVC中的移動估測所使用的基本單位為16x16大小方塊，稱macroblock。此基本單位可再細分為16x16, 16x8, 8x16及8x4大小長方塊，如圖 20(a)所示。而8x8之sub-macroblock可細分為8x8, 8x4, 4x8及4x4大小，如圖 20(b)。



(a)



(b)

圖 20: (a) sub-macroblock 之四種型態. (b) 8×8 sub-macroblock 之切割.

在此移動搜尋的過程中，為了增加搜尋的精確度，我們將參考畫面經填補濾波器(interpolation filter)，進行 1/2 像素與 1/4 像素精確度的移動搜尋。由於搜尋方塊切割較為細緻，也造成了移動向量之位元率過高，因此我們利用相鄰方塊移動向量的相關性，進行預測。我們用被預測方塊的左方、左上方、上方及右上方進行被預測方塊的移動向量預測。藉由細緻的搜尋方塊切割與移動向量預測，我們可以精確有效率的將畫面切割成適當的搜尋方塊，並找到對應的移動向量。圖 21 為經過移動搜尋之後的畫面切割圖例。

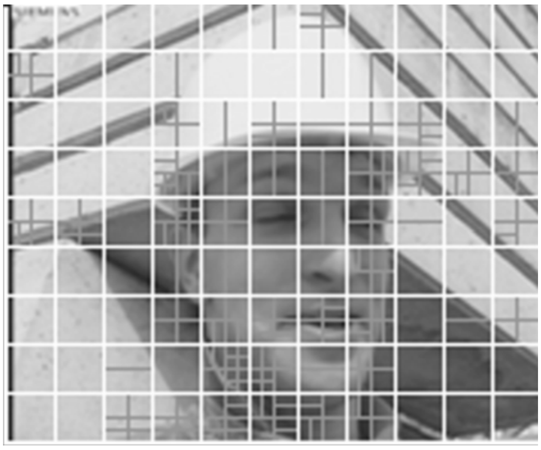


圖 21: 經過移動搜尋之後的畫面切割

## B.2 I-Block 與雙向移動估測

I-Block 與雙向移動估測的概念曾在[22]描述過。我們利用此概念並經適當的修改將其用於改善移動補償時間濾波。

時間軸之低通畫面是基於其中每個點的連接狀態，利用方程式(2)及(3)所產生的。一般來說，對於判斷為「連接狀態點」(此點在前後兩張畫面有很好對應)，移動補償效果不錯。然而，缺乏對應的比較差的移動估測結果，會對這些低通畫面造成畫質的缺陷。而低通畫面對於在時間軸可調應用(temporal scalability)極為重要，因此，我們必須將這些較差的移動估測之點強制宣告為「非連接狀態」以改善低通影像效率。

$$H[m,n]=\left(A[m,n]-\widetilde{B}[m-d_m,n-d_n]\right)/\sqrt{2} \qquad (1)$$

$$L[m,n] = \widetilde{H}[m+d_m, n+d_n] + \sqrt{2}B[m,n] \qquad (2)$$

$$L[m,n] = \sqrt{2}B[m,n] \qquad (3)$$

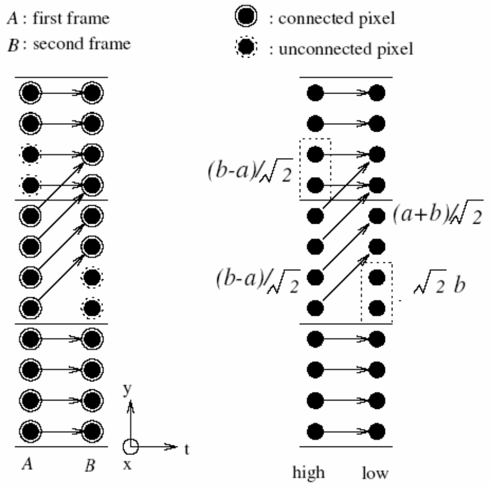


圖 22: 前後兩張畫面連接狀態示意圖

我們將 I-block 的大小訂為 16x16。如同圖 22 所示，$A[m,n]$ 為在 A 畫面中座標為(m,n)之方塊，$B[m-d_m,n-d_n]$ 為 B 畫面沿著($d_m$,$d_n$)移動向量所對應之參考方塊。我們先比較兩者之變異數，然後選擇較小的值稱為 $V_{min}$；若兩者的均方差直大於 $F*V_{min}$，此方塊將被宣告為「非連接狀態」，其中 $F$ 為一個可調之參數。根據我們的實驗結果，$F$ 值為 0.7 為一個理想值。

此外，除了單方向的搜尋外，我們亦考慮了另一方向搜尋到較好的預測值的可能性。因此畫面 A 將具有兩個方向的移動向量。雙向移動估測較為準確，可有效降低高通畫面的數值，使編碼效率提高。

## B.3 移動成本函數調整

在移動向量的模式決定(mode decision)方面，我們利用此 R-D 成本估計函數 $J = D + \lambda * R$ ，其中 $D$ 是差值，$R$ 為移動向量之估測的位元率。然而，隨著移動補償時間濾波的時間軸階層(temporal level)的增加，低通畫面之能量隨之增加。因此 λ 值必須隨著階層做調整，使得在較高的時間軸階層還能適當的 RD 關係。我們將 λ 值隨著每次時間軸階層的增加而增加為原來的 $\sqrt{2}$ 倍。

## B.4 可調式移動資訊

在文獻[23]中，提出一個適用於 MC-EZBC 之可調式移動向量方法。在此計畫中，我們將此概念做沿伸與適當的修改，提出一個可適用於 AVC 移動估測之可調式移動向量技術。

在傳統的小波轉換中，擁有空間軸、時間軸及畫質之可調式設計，移動資訊在空間軸與畫質可調式設計中是無法切割的。當可允許傳送之位元率太低時，抽取器(Extractor or Puller)很可能會因移動資訊過大，而沒有足夠可分配的位元率進行小波係數切割而失敗。此外，在非常低位元率的情況下，我們會希望在省去部份的移動資訊將較多的位元率分配給小波係數，以換取較好的畫質。因此，在移動估測之後，我們對移動資訊進行切割。

在AVC的畫面間預測中，基本的處理大小為16x16之macroblock。每一個macroblock可再細分為16x16, 16x8, 8x16, 8x8 ,8x4, 4x8 及 4x4，並且所對應到的移動向量具有 1/4 像素之精確度。我們依下列步驟將移動向量做切割。

第一步：進行 16x16 大小之整數像素精確度之移動搜尋，所產生之移動向量為“基本層”之移動向量。

第二步：進行 16x16 與 8x8 大小之 1/2 像素精確度之移動搜尋。與基本層之差值將被保留編碼，稱“第一加強層”。

第三步：進行所有搜尋方塊大小之 1/4 像素精確度之移動搜尋。將其與基本層與第一加強層之和的差值將被保留編碼,稱“第二加強層”。

第四步：將所有的移動向量階層分別利用 CABAC 編碼

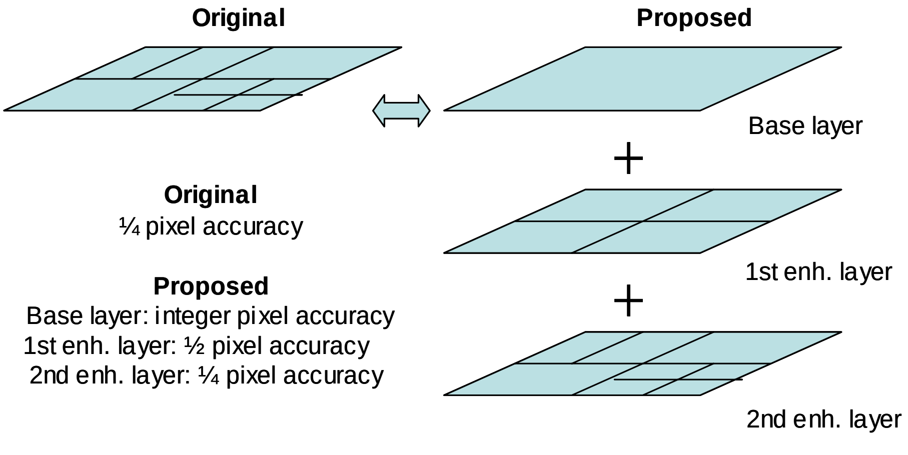

Original
Proposed
Base layer
+
Original
¼ pixel accuracy
Proposed
Base layer: integer pixel accuracy
1st enh. layer: ½ pixel accuracy
2nd enh. layer: ¼ pixel accuracy
1st enh. layer
+
2nd enh. layer

圖 23: The base and enhancement layer motion vectors.

如圖 23 所示，原始單層的移動向量被分為三層。每張畫面的移動資訊都會在移動補償時間濾波的過程中被分成基本層與適當的加強層，因此，所有的移動向量資訊都會被集中，如圖 24 所示。因為所有時間軸階層之移動資訊基本層必需被傳送，以產生所有時間軸解析度畫面，所以，基本層資料非常重要不可遺失。

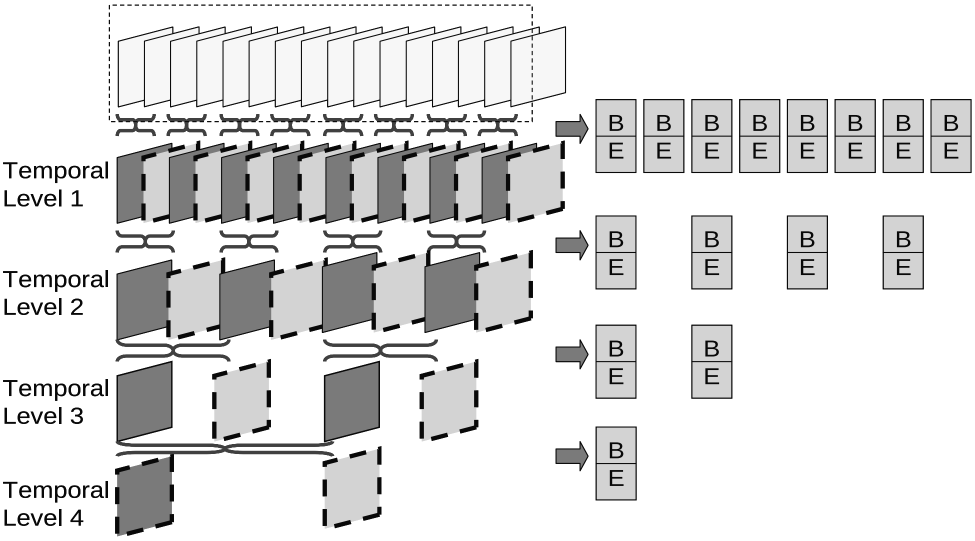
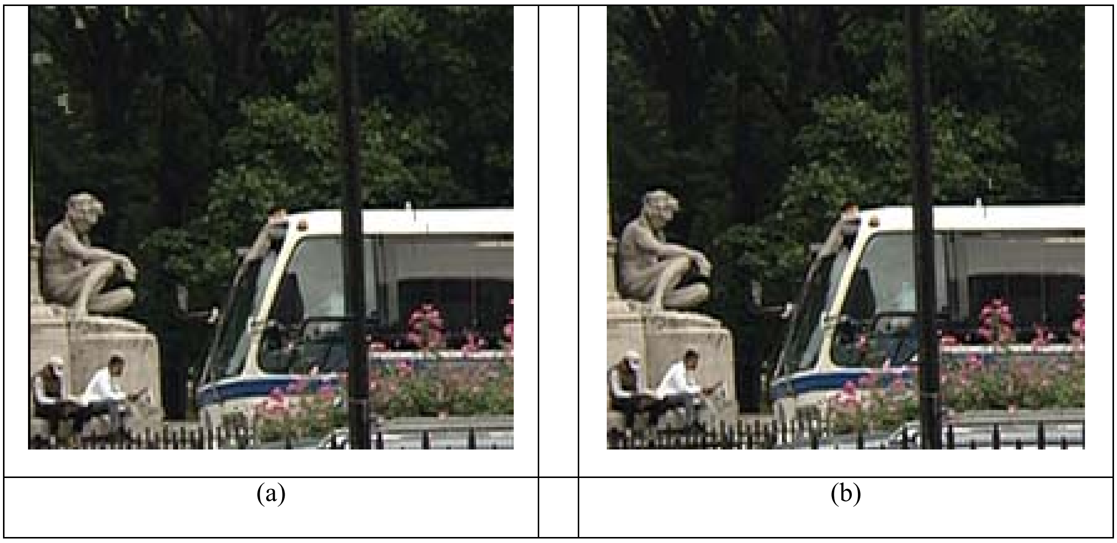


圖 24: GOP 內之移動資訊示意圖

若所允許傳送之位元率過低時，抽取器將會根據情況切除一或兩層的加強層移動資訊。此外，若使用者希望進行空間軸可調式步驟時，抽取器亦可切除適當的加強層。若所需要之可調式幅度小時，可將所有加強層移動資訊編碼合為一層以節省位元率消耗。

## C. 模擬與實驗

利用前述之移動補償濾波技術的改善，我們可以得到以下低通影像之差異：


(a)

(b)

圖 25: 第二時間軸階層之低通影像比較：(a)無 I-block 技術 (b) 利用 I-block.

在圖 25(a)中為原始未經 I-block 技術應用之低通影像。我們可以發現,在圖中左上角可見到一些因移動估測表現不良而造成的缺陷，經過 I-block 技術之後，可在圖 25(b)中見到明顯改善。

## D. 結論

目前的可調式技術可分為時間軸、空間軸及畫質可調式編碼。畫面間小波轉換視訊編碼為一個利用小波轉換達成完全可調式(full scalable)編碼之嶄新技術。DCT 混合式(hybrid)架構為近年來最廣為被使用的視訊編碼方式。然而，由於過去在小波視訊編碼上的幾項困難問題近五、六年獲得解決，包括分數精確度之移動補償技術等。如今小波視訊編碼已成為可在高位元率與 H.264 具有相似表現。

在低位元率時，畫面間小波轉換視訊編碼尚不如 H.264。這是由於 H.264 使用混合式架構並且可針對特定位元率做最佳化，而畫面間小波轉換視訊編碼可適用於各位元率之位元資料流(bitstream)切割，因此，在最低位元率上之表現並未如同 H.264 可最佳化。在許多的情況下，我們也注意到了由於大量的移動資訊對在低位元率下之壓縮資料所造成的不利影響。

在低位元率時，我們亦發現畫面間小波視訊轉換編碼在一些特定的測試影像無法表現得很好，而這些影像內容都具大幅度的移動。雖然分數精確度之移動估測可改善編碼效率，但在移動補償上並未做最佳化。時間軸之低通影像有時會有移動估測造成的缺陷。

本計畫中所提出之技術可有效改善編碼效率。我們利用了在 AVC 中高效率之移動估測來進行移動補償時間濾波，以及利用 I-block 偵測以改善在時間軸低通影像之畫質表現。此外，所提出的多階層式移動估側可達成可調式移動資訊，並且在低位元率之下有相當明顯之編碼改善。畫面間小波視訊編碼仍有許多結構與參數有待進一步最佳化。

這部分成果分成兩部分，分別在 2004 年 3 月與 7 月提案 MPEG 標準組織。3 月之提案為參加 scalable video coding Call-for-Proposal 競賽，在 14 個提案中經視覺主觀評審，成績中等。在國際大企業環伺下，此結果似乎尚可。7 月之提案為參加 Core Experiments，改善目前的 Reference Model。MPEG 標準組織在未來兩年將持續改良 Reference Model，最後成為標準。

## 參考文獻

[1] ISO/IEC JTC1/SC29/WG11 N3747. *MPEG-4 Overview* - (V.16 – La BauleVersion), Contribution for La Baule, October 2000.

[2] C.J. Tsai, M. van der Shaar and Y.K. Lim, "*Working Draft 3.0 of ISO/IEC TR2100-12 Multimedia Test Bed for Resource Delivery,*" ISO/IEC JTC1/SC29/WG11 MPEG2003/M10299, Hawaii, December 2003.

[3] ISO/IEC JTC1/SC29/WG11 N3850. *ISO/IEC 14996-1 ,COR1, AMD1.*

[4] ISO/IEC JTC1/SC29/WG11 N2614 *MPEG-4 Intellectual Property Management & Protection (IPMP) Overview & Applications.*

[5] ISO/IEC JTC1/SC29/WG11 N5068, *Study of FPDAM ISO/IEC 14496-1:2001/AMD3, Jul. 2002.*

[6] ISO/IEC JTC1/SC29/WG11 N5333, *MPEG-21 Requirements v.14, Dec. 2002.*

[7] ISO/IEC JTC1/SC29/WG11, N5535, *Requirement for MPEG-21 Intellectual Property Management and Protection, Pattaya, Mar. 2003.*

[8] ISO/IEC JTC1/SC29/WG11, Part 5 – *Reference Software – Systems (ISO/IEC 14496-5 Systems)*

[9] ISO/IEC JTC1/SC29/WG11 N4702, *MPEG-4 IPMP Extension Reference Software Architecture based on IM1, Jeju, Mar. 2002.*

[10] ISO/IEC JTC1/SC29/WG11 N4850, *MPEG-2 and MPEG-4 IPMP Extension Reference Software Architecture based on IM1, May. 2002.*

[11] J. Liu, et al., "*WD1.0 of ISO/IEC 13818-5:1997/AMD2:2003 MPEG-2 IPMP Reference Software*," ISO/IEC JTC1/SC29/WG11 M9840, Trondheim, July 2003.

[12] Y. Lim, C. Xu, and D.D. Feng, "*Web-based image authentication using invisible fragile watermark*," in Conference in Research and Practice in Information Technology, 2002, pp.31-34.

[13] C.S. Shieh, H.C. Huang, F.H. Wang, and J.S. Pan, "*Genetic watermarking based on transform domain techniques*," Pattern Recognition, 2004, pp.555-565.

[14] X. Sun, F. Wu, S. Li, W. Gao, and Y.-Q. Zhang, "*Seamless switching of scalable video bitstreams for efficient streaming*," IEEE Trans. On Multimedia, 2004, pp.291-303.

[15] J.M. Almeida, D.L. Eager, M.K. Vernon, and S.J. Wright, "*Minimizing delivery cost in scalable content distribution systems*," IEEE Trans. On Multimedia, 2004, pp.356-365.

[16] R. Parviainen and P. Parnes, "*Large scale distributed watermarking of muticast media through encryption*," in Proceedings of the International Federation for Information Processing, Communications and Multimedia Security Joint Working Conference IFIP TC6 and TC11, 2001, pp.149-158.

[17] X. Xu, S. Dexter, and A.M. Eskicioglu, "*A hybrid scheme for encryption and watermarking*," IS&T/SPIE Symposium on Electronic Imaging 2004, Security, Steganography, and Watermarking of Multimedia Contents IV Conference, 2004, pp.723-734.

[18] Digital video broadcasting project (DVB): http://www.dvb.org (2004).

[19] Data Encryption Standard (DES): http://www.itl.nist.gov/fipspubs/fip46-2.htm (1993)

[20] V. Chande and N. Farvardin, "*Progressive transmission of images over memoryless noisy channels*," IEEE Journal on Selected Areas in Communications, 2000, pp.850-860.

[21] P. Chen and J. W. Woods, "Comparison of MC-EZBC and H.26L TML 8 on Digital Cinema Test Sequences," ISO/IEC JTC1/SC29/WG11, MPEG2002/8130, Cheju Island, March 2002

[22] P. Chen, *Fully scalable subband/wavelet coding*, Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, New York, May 2003.

[23] H.-M. Hang, S. S. Tsai, and Tihao Chiang, "Motion information scalability for MC-EZBC", ISO/IEC/JTC1 SC29/WG11 doc. No. M9756, July 2003

## 計畫成果自評

本計畫有以下幾類成果。第一類為 MPEG-4 IPMP System 與 Interframe Wavelet 所發展出的技術、經驗及成品與國際 MPEG 標準直接相關，極具實用價值，可促進國內工業研發技術開發。第二類為將上述技術提案至 MPEG 標準組織，有助我國技術之進入國際舞台，共有六篇 MPEG 標準提案。其中 IPMP System 在 2004 年 7 月加入 MPEG-21 Multimedia Test Bed 中，已成為 Test Bed 軟體的一部份，全案期望在 2005 年左右完成標準化。Interframe Wavelet 在 2004 年 3 月與 7 月提案 MPEG 標準組織，參加 scalable video coding Call-for-Proposal 競賽，與後續 Reference Model 之改進。第三類為計畫執行過程所獲得之研究成果論文四篇，已發表於國內外學術會議。其四，參與計畫之同學可獲得國際多媒體最先進的 MPEG-4 與 MPEG-21 相關技術及多媒體系統設計經驗，畢業後進入產業，直接有助於產業界開發新產品，提昇我國工業技術能力。達到人才培育之目的。

綜合評估：本計畫產出相當多具有學術與應用價值的成果，特別是直接參與國際標準會議，在國際上展示成果。並培育高科技人才培育，整體成效良好。已發表學術論文四篇，碩士學位論文一冊，以及六篇 MPEG 標準提案如下表。

## Publications

(1) H.-M. Hang, “Next generation MPEG video and system,” (invited talk), *2003 Workshop on Consumer Electronics*, Nov. 27 – 28, Tainan, Taiwan 2003.

(2) H.-K. Hsu, H.-C. Huang, and H.-M. Hang, ``An enhanced entropy coding scheme for interframe wavelet," in *2004 Conf. on Computer Vision, Graphics, and Image Processing*, Hualin, Taiwan, Aug. 2004.

(3) C.-Y. Tsai, H.-K. Hsu, H.-C. Huang, H.-M. Hang and G.-Z. Wu, “Enhanced motion estimation for interframe wavelet video coding,” *IEEE International Conf. on Image Processing '04*, Singapore, Oct. 2004

(4) F.-C. Chang, H.-C. Huang and H.-M. Hang, “Combined Encryption and Watermarking Approaches for Scalable Multimedia Coding,” *Pacific Rim Conference on Multimedia 2004*, Tokyo Japan, Dec. 2004

(5) Chen-Wei Fan 范振韋, *MPEG-4 IPMPX Design and Implementation on MPEG-21 Testbed,* MS Thesis, NCTU, June 2004.

## MPEG Standard Contributions

1. C.-Y. Tsai, H.-C. Chuang, J.-H. Chen, J.-C. Ma, C.-Y. Liu, C.-W. Fan, F.-C. Chang, C.-N. Wang, C.-J. Tsai, Tihao Chiang, S.-Y. Lee, and H.-M. Hang, “ISO/IEC JTC1/SC29/WG11 M10160: Scalable Multimedia Streaming Test Bed for Media Coding and Testing in Streaming Environments,” October 2003 (66th, Brisbane, Australia)

2. C.-N. Wang, C.-Y. Tsai, H.-C. Chuang, J.-H. Chen, J.-C. Ma, C.-N. Chiu, C.-Y. Liu, C.-W. Fan, F.-C. Chang, C.-J. Tsai, Tihao Chiang, S.-Y. Lee, and H.-M. Hang, “ISO/IEC

JTC1/SC29/WG11 M10298: Scalable Multimedia Streaming Test Bed for Media Coding and Testing in Streaming Environments," December 2003 (67th, Kona, Hawaii, USA)

3. C.-Y. Tsai, H.-K. Hsu, H.-M. Hang, and T. Chiang, "ISO/IEC JTC1/SC29/WG11 M10569/S08, Response to Cfp on Scalable Video Coding Technology: Proposal S08 -- A Scalable Video Coding Scheme Based on Interframe Wavelet Technique," March 2004 (68th, Munich, Germany)

4. H.-C. Huang, W.-H. Peng, Y.-C. Lin, C.-N. Wang, T. Chiang and H.-M. Hang, "ISO/IEC JTC1/SC29/WG11 M10569/S07, Response to Cfp on Scalable Video Coding Technology: Proposal S07 -- A Robust Scalable Video Coding Technique," March 2004 (68th, Munich, Germany)

5. C.-N. Wang, C.-H. Li, C.-W. Fan, Y.-T. Shih, J.-P. Ho, C.-L. Lin, F.-C. Chang, G.-C. Li, C.-N. Chiu, C.-Y. Tsai, H.-C. Chuang, J.-H. Chen, J.-C. Ma, Chi-Yu Liu, C.-C. Chen, C.-J. Tsai, Tihao Chiang, S.-Y. Lee, and H.-M. Hang, "ISO/IEC JTC1/SC29/WG11 M11117: Scalable Multimedia Streaming Test Bed for Media Coding and Testing in Streaming Environments," July 2004 (69th, Redmond, Washington, USA)

6. H.-K. Hsu, C.-Y. Tsai, H.-C. Huang, H.-M. Hang, and T. Chiang, "ISO/IEC JTC1/SC29/WG11 M10934: Response to Core Experiments in SVC 1b Spatial Transform & Entropy Coding," July 2004 (69th, Redmond, Washington, USA)

## 附錄

(1) H.-K. Hsu, H.-C. Huang, and H.-M. Hang, "An enhanced entropy coding scheme for interframe wavelet," in *2004 Conf. on Computer Vision, Graphics, and Image Processing*, Hualien, Taiwan, Aug. 2004.

(2) C.-Y. Tsai, H.-K. Hsu, H.-C. Huang, H.-M. Hang and G.-Z. Wu, "Enhanced motion estimation for interframe wavelet video coding," *IEEE International Conf. on Image Processing '04*, Singapore, Oct. 2004

(3) F.-C. Chang, H.-C. Huang and H.-M. Hang, "Combined Encryption and Watermarking Approaches for Scalable Multimedia Coding," *Pacific Rim Conference on Multimedia 2004*, Tokyo Japan, Dec. 2004

# AN ENHANCED ENTROPY CODING SCHEME FOR INTERFRAME WAVELET

*Han-Kuang Hsu, Hsiang-Cheh Huang, and Hsueh-Ming Hang*

Department of Electronics Engineering,
National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
hmhang@mail.nctu.edu.tw

## ABSTRACT[1]

An enhanced entropy coding scheme is incorporated into the interframe wavelet coding architecture in this paper. Interframe wavelet coding has the advantage of SNR, temporal, and spatial scalability, and is a potential candidate for the on-going MPEG-21 scalable video coding (SVC) standard. Motion-Compensated Temporal Filtering (MCTF) and Wavelet Transform Coding are two most essential components in the interframe wavelet coding architecture. The arithmetic entropy subsystem is an indispensable element in Wavelet Transform Coding. It produces the final output bitrate. In this paper, we modify the entropy coding syntax/scheme originally specified in the MPEG SVC Core Experiment (CE) reference software. We observe some bit savings of this technique in our simulations based on the conditions specified by the MPEG core experiments; however, the full potential of this technique is yet to be further explored.

## 1. INTRODUCTION

Video compression is an essential element in multimedia applications. Conventional video coding systems, including MPEG-1, MPEG-2, H.261 and H.263 international standards, employ the so-called *hybrid coding* structure. In these schemes, the reconstructed previous frame is used to predict the current frame after motion compensation.

The on-going MPEG-21 Scalable Video Coding (SVC) standard employs a new approach different from the hybrid coding structure, Motion-Compensated Temporal Filtering (MCTF) with Wavelet Transform Coding, to achieve SNR, spatial, and temporal scalability. Ohm first proposed a motion-compensated t+2D coding structure [1], as shown in Figure 1 [2]. The major difference between the hybrid coding and the t+2D coding is that the latter does not contain the closed-loop (interframe) DPCM. In addition, the t+2D coding scheme fulfills for the scalable video coding requirements. One of the improved and highly efficient realizations of this concept is the interframe wavelet video coder proposed by Woods and his co-workers [2]. This scheme is called Motion Compensated Temporal Filtering – Embedded Zero Block Coding (MCTF-EZBC or MC-EZBC). Its architecture is shown in Figure 2 [4][6]. Essentially, the same basic structure was adopted by the MPEG committee in March 2004 as the first reference model of SVC.

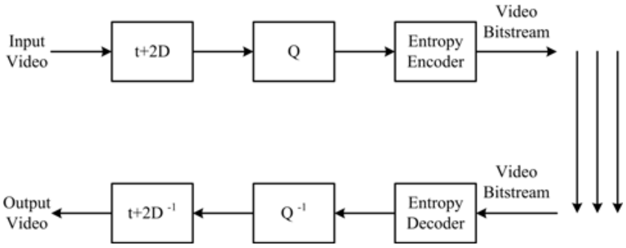


**Figure 1.** Block diagram of t+2D transform coding system.

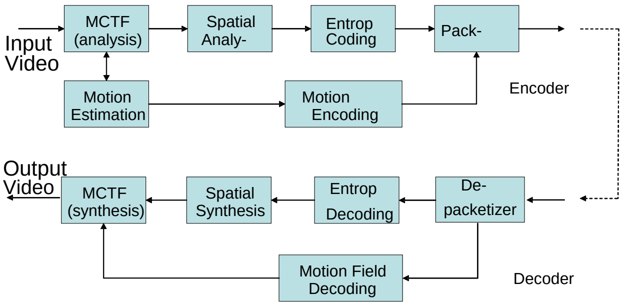


**Figure 2.** The interframe wavelet video coder.

In this paper, we focus on improving the entropy coding scheme in the aforementioned interframe wavelet coding structure. As illustrated in Figure 1 and Figure 2, no matter how the motion compensation is performed in SVC, entropy coding is a must to further reduce the bits in the output bitstream.

The motivation behind our scheme is the observation of clusters of “1”-bits on the wavelet coefficient bitplanes. Thus, we develop our entropy coding based on the quadtree concept. At the end, we compare the performance between our proposed scheme and that in [7] and [8], and show a somewhat better performance of the proposed scheme.

This paper is organized as follows. In Sec. 2, we outline the motivation behind our proposed scheme. The conventional 3D EBCOT technique is summarized in

[1] This work is supported in parts by National Science Council (Taiwan, ROC) under Grant NSC 92-2219-E-009-003 and OES, Industrial Technology Research Institute (Taiwan, ROC) under Grant 93B-05.

Sec. 3. In Sec. 4, we describe the coding process of new entropy coding scheme by modifying the existing 3D EBCOT. The changes on CE software for integrating the proposed algorithm are described in Sec. 5. Simulation results are shown in Sec. 6, in which we compare the results with the existing scheme. We conclude this paper in Sec. 7.

## 2. MOTIVATION

In the SVC core experiment one software, the 3D EBCOT entropy coding procedure is applied after MCTF and spatial transform [7][8]. We observe that high energy wavelet coefficients often cluster together [10][9]. In order to save coding bits, we propose a modified coding procedure as described in Section 3. Essentially we construct another layer that records the bitplane locations of the Significant Bits (SB) of all coefficients. We observe bit savings of this technique in our simulation; however, the full potential of this technique is yet to be further explored.

## 3. 3D EMBEDDED BLOCK CODING WITH OPTIMAL TRUNCATION SCHEME

In the MPEG SVC core experiment reference software [8], the coefficients are coded by the 3D Embedded Block Coding with Optimal Truncation (3D EBCOT) process after the temporal and spatial subband transform. Each subband, generated by temporal and spatial transforms, is divided into 3D codeblocks, which is coded independently. Next, the entropy coding module is applied to these codeblocks. It encodes each bitplane sequentially using context-based arithmetic coding.

Three coding operations are employed to encode the samples in a bitplane: [8] [8]

- **Zero Coding (ZC)**: When a sample is not yet significant in the previous bitplane, this primitive operation is utilized to code the new information of the sample. The definition of "significance" is described in Sec. 4.
- **Sign Coding (SC)**: Once a sample becomes significant in the current bitplane, Sign Coding operation is performed to encode the sign of the sample. Sign Coding also utilizes an adaptive context-based arithmetic coder to compress the sign symbols.
- **Magnitude Refinement (MR)**: Magnitude Refinement is employed to encode the new information of a sample, which has already become significant in the previous bitplane.

For each bitplane, the EBCOT coding procedure consists of three distinct passes, applied in turn. The three passes are:

- **Significant Propagation (SP) pass:** In this pass, samples which are not yet significant but have significant neighbor sample(s) are processed.
- **Magnitude Refinement (MR) pass:** Significant samples are coded in this pass.
- **Normalization pass:** During this pass, those samples which are not yet coded in SP and MR passes (insignificant samples) are coded. So zero coding and sign coding primitives are applied in this pass.

## 4. PROPOSED ENTROPY CODING SCHEME

We propose the so-called SB-reach method in this Section. In the 3D EBCOT described in Sec. 3 (also in the core experiment software [8]), all wavelet coefficients are initially "insignificant". A coefficient becomes "significant" when its non-zero bit is first found. The first non-zero bit will thus be called Significant Bit (SB) (of a coefficient). For each bitplane, we construct another binary bitplane – so-called SB-reach plane. As shown in Figure 3, a single sample in the SB-reach plane represents a square *mapping block* of *n* by *n* coefficients. The size of the SB-reach plane thus decreases as its representing *mapping block* becomes larger. The binary sample on an SB-reach plane is set to 1, if its square mapping block contains one or more significant coefficients. On the other hand, if the binary sample on the SB-reach plane is 0, it means that all its associated bits in the coefficient bit plane are zero.

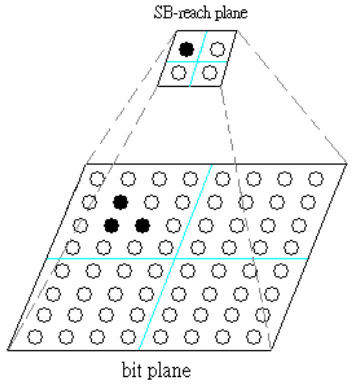


**Figure 3.** One binary sample on the SB-reach plane is associated with 4x4 mapping block bits.

In this modified coding process, we first construct all the SB-reach bit planes up to the selected "SB-reach depth", as illustrated in Figure 4. Each SB-reach bitplane is associated with one bit plane of the original coefficients. We first encode an SB-reach bit plane before encoding its associated coefficient bit plane using the core experiment software (CES) procedure. In encoding an SB-reach plane, we perform the Significant Propagation pass and the Normalization pass following the scanning order in CES. If a sample is classified significant in a previous SB-reach plane, it must be a "1" bit in the current SB-reach plane and thus is not coded.

After coding one SB-reach plane, we code its associated coefficient bitplane. The coefficients on the bit plane are not coded, if its corresponding SB-reach plane bit is zero (insignificant). If a bit on the SB-reach plane is 1, and then its associated coefficient mapping block bits are coded in the order shown in Figure 5. We perform three coding passes as the original CES does on these coefficient bits. One example is given in Figure 6 to illustrate the procedure in our proposed algorithm. The “coding bits” in this figure are the bits (samples) to be coded by the context-based arithmetic coder.

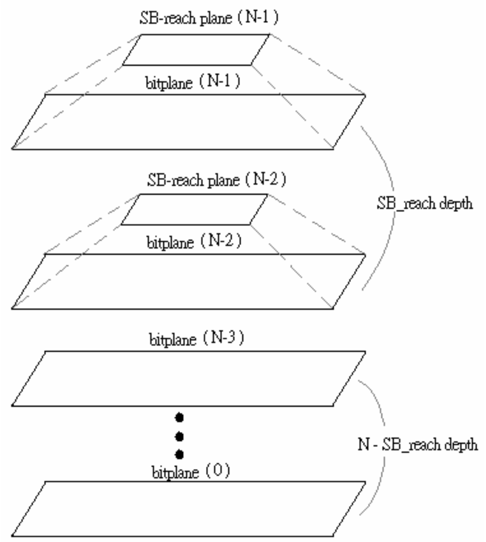


**Figure 4.** Illustration of “SB-reach depth”.

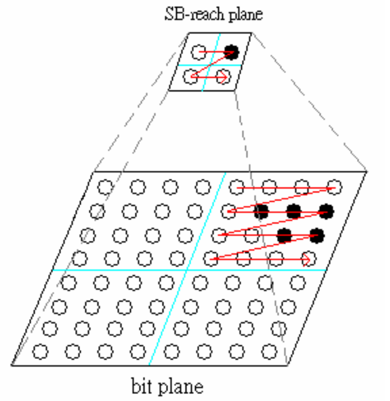
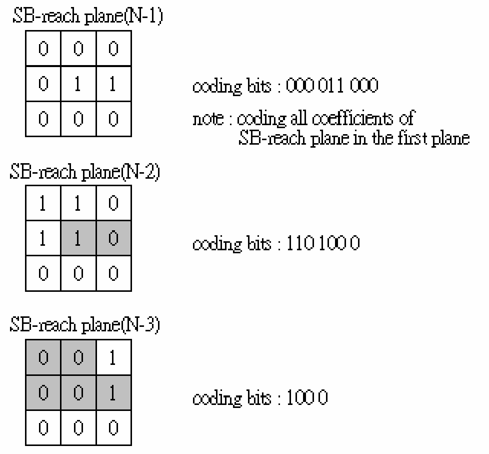


**Figure 5.** The encoding process of an SB-reach plane and its associated bitplane.



SB-reach plane(N-1)

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 0 | 0 | 0 |

coding bits : 000 011 000

note : coding all coefficients of SB-reach plane in the first plane

SB-reach plane(N-2)

| | | |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 1 | 0 |
| 0 | 0 | 0 |

coding bits : 110 100 0

SB-reach plane(N-3)

| | | |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 0 | 1 |
| 0 | 0 | 0 |

coding bits : 100 0

**Figure 6.** An example of coding steps with our algorithm.

With the method described above, we try all combinations of mapping block size and SB-reach depth, and we then compare the resulting coded bits of all combinations. The best combination of mapping block size and SB-reach depth is retained and coded.

## 5. CHANGES ON CE SOFTWARE FOR INTEGRATING THE PROPOSED ALGORITHM

On the top of the core experiment software, we changed some syntax and decoding procedure as follows. We add the SB-reach plane architecture to the original 3D EBCOT. The information for the mapping block size, the SB-reach plane depth, and the SB-reach planes is added to the original syntax as shown in Figure 7.

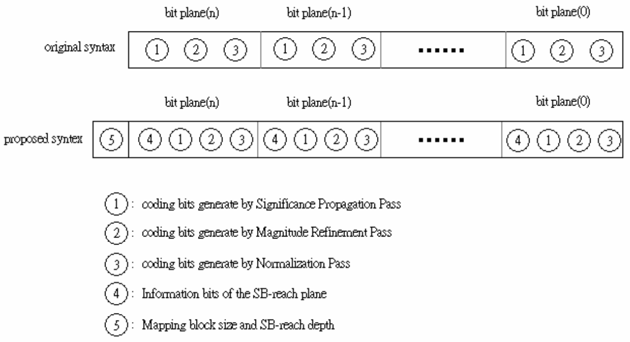

bit plane(n)
bit plane(n-1)
bit plane(0)
original syntax
1 2 3 1 2 3 ...... 1 2 3
proposed syntex
5 4 1 2 3 4 1 2 3 ...... 4 1 2 3
1 : coding bits generate by Significance Propagation Pass
2 : coding bits generate by Magnitude Refinement Pass
3 : coding bits generate by Normalization Pass
4 : Information bits of the SB-reach plane
5 : Mapping block size and SB-reach depth

**Figure 7.** Changes between the original and the proposed syntax.

Here are some definitions in the newly added terms.

- Mapping_blk_size: The mapping block size information is defined in Figure 8. Bit pattern “00” = size 4x4; “01” = 8x8; “10” = 16x16; and “11” = 32x32.
- SB-reach depth: The depth of SB-reach planes. Bit pattern “00” = depth 2; “01” = depth 3; “10” = depth 4; and “11” = depth 5.
- SB_plane: Record SB-reach bits of the corresponding bitplane.

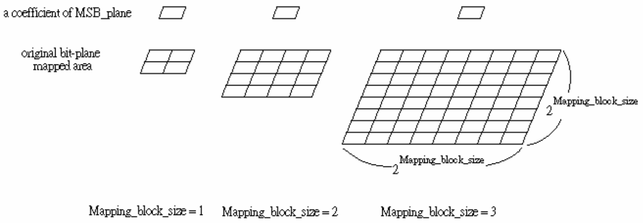

a coefficient of MSB_plane
original bit-plane mapped area
$2^{Mapping\_block\_size}$
$2^{Mapping\_block\_size}$
Mapping_block_size = 1
Mapping_block_size = 2
Mapping_block_size = 3

**Figure 8.** Square mapping block size.

## 6. SIMULATION RESULTS

We evaluate the performance of our algorithm by measuring the bitrate savings between the proposed algorithm and the core experiment software. We follow the core experiment (CE) specifications to conduct a series of experiments and to test the effectiveness of the proposed algorithm. Eight sequences are tested, namely, CREW, HARBOR, SOCCER, CITY, BUS, FOOTBALL, FOREMAN, and MOBILE, under different spatial, temporal and bitrate test points. Spatial resolutions are QCIF, CIF, and 4CIF, temporal resolutions are 15, 30 and 60 frame/sec, and bitrates vary from 96 kbit/sec to 3 Mbit/sec. The objective image quality, or the PSNR values, are almost the same between our results and the results from the core experiment software. Besides, the subjective qualities are almost identical. Therefore, we compare the resulting bits generated by our algorithm and those by the CE reference software. Some savings in bits with our algorithm are observed.

In Table I to Table III, the bitrate savings are expressed in percentage. In these tables, each entry is the total bitrate saving accumulated from the 1st bitplane to the current one. For example, the cumulative biplane 2 means the total bits saved for the 1st, 2nd and 3rd bitplanes together. The positive numbers denote bitrate savings, while the negative numbers mean bitrate loss. The LL, LH, HL, and HH bands in these tables are the spatial subbands of all spatial resolutions accumulated.

**Table I.** Bitrate savings (in percentage) for the FOREMAN and BUS sequences of the H frames at temporal levels 1 and 2.

| Cumulative bitplane | FOREMAN | | | | BUS | | | |
|---|---|---|---|---|---|---|---|---|
| | LL | LH | HL | HH | LL | LH | HL | HH |
| 2 | 0.22% | 0.17% | 0.27% | 0.18% | -1.86% | -0.63% | -0.59% | -0.17% |
| 3 | 0.67% | 0.45% | 0.51% | 0.37% | 0.36% | 0.51% | 0.30% | 0.45% |
| 4 | 0.46% | 0.25% | 0.28% | 0.23% | 0.18% | 0.23% | 0.17% | 0.22% |

**Table II.** Bitrate savings (in percentage) of the H frames at temporal levels 3 and 4.

| Cumulative bitplane | FOREMAN | | | | BUS | | | |
|---|---|---|---|---|---|---|---|---|
| | LL | LH | HL | HH | LL | LH | HL | HH |
| 2 | 1.04% | 1.04% | 1.11% | 0.91% | -0.71% | -0.32% | -0.49% | 0.10% |
| 3 | 1.47% | 1.26% | 1.21% | 1.20% | 0.53% | 0.61% | 0.29% | 0.68% |
| 4 | 0.81% | 0.66% | 0.59% | 0.83% | 0.27% | 0.30% | 0.14% | 0.37% |

**Table III.** Bitrate savings (in percentage) at the bottommost temporal level.

| Cumulative bitplane | FOREMAN | | | | BUS | | | |
|---|---|---|---|---|---|---|---|---|
| | LL | LH | HL | HH | LL | LH | HL | HH |
| 2 | -0.05% | 0.91% | 0.39% | 0.97% | 0.31% | 0.28% | -0.06% | 0.05% |
| 3 | 0.13% | 0.13% | 0.99% | 1.03% | 0.24% | 0.47% | 0.24% | 0.61% |
| 4 | 0.14% | 0.80% | 0.67% | 0.68% | 0.15% | 0.25% | 0.10% | 0.30% |

As shown in the simulation results regarding to the output bitrates, our algorithm performs somewhat better than the CE software. In general, we gain more at the cumulative biplane 3. Particularly, the HH bands at higher temporal levels perform better. Even better results may be obtained by selecting good context and probability models for arithmetic coding. Also, we should tune carefully the parameter values in our algorithm.

## 7. CONCLUSIONS AND FUTURE WORK

In this paper, we propose an enhanced entropy coding scheme to further increase the compression efficiency of the interframe wavelet coding algorithm. We modify the entropy coding unit by adding an extra SB-reach layer. Several test conditions specified by the core experiment are tested. So far, our proposed algorithm has somewhat better performance at low- to mid-bitrates comparing to the MPEG Core Experiment (CE) reference software. Further parameter tuning should provide better results, and the full potential of this technique is yet to be further explored.

## 8. REFERENCES

[1] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[2] S.-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Processing: Image Communications*, vol. 16, pp. 705–724, May 2001.

[3] T. Rusert, et al., *Recent improvements to MC-EZBC*, ISO/IEC/JTC1 SC29/WG11 doc. M9232, Dec. 2002.

[4] S. S. Tsai, *Motion information scalability for interframe wavelet video coding*, MS thesis, National Chiao Tung University, Hsinchu, Taiwan, R.O.C., Jun. 2003.

[5] C. Y. Tsai, H. K. Hsu, H.-C. Huang, H.-M. Hang, and G. Z. Wu, "Enhanced motion estimation for interframe wavelet video coding," *IEEE Int'l Conf. Image Processing*, 2004.

[6] C. Y. Tsai, H. K. Hsu, H.-M. Hang, and T. Chiang, "A scalable video coding scheme based on interframe wavelet technique," ISO/IEC/JTC1 SC29/WG11 MPEG2004/M10569/S08, Mar. 2004.

[7] J. Xu, Z. Xiong, S. Li, and Y. Q. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *Applied and Computational Harmonic Analysis*, vol. 10, pp.290–315, May 2001.

[8] J. Xu, R. Xiong, B. Feng, G. Sulliman, M. C. Lee, F. Wu, and S. Li, "3D sub-band video coding using barbell lifting," ISO/IEC/JTC1 SC29/WG11 MPEG2004/ M10569/S05, Mar. 2004.

[9] W. H. Peng, T. Chiang and H.-M. Hang, "Context-based binary arithmetic coding for fine granuality scalability," *Proc. Seventh Int'l Symp. Signal Processing and Its Applications*, vol. 3, pp.105–108, 2003

[10] D. Taubman, "EBCOT (Embedded Block Coding with Optimized Truncation) A complete reference", ISO/IEC JTC1/SC29/WG1 N988, Sep.1998.

# ENHANCED MOTION ESTIMATION FOR INTERFRAME WAVELET VIDEO CODING

*Chia-Yang Tsai, Han-Kuang Hsu, Hsiang-Cheh Huang, Hsueh-Ming Hang and Guo-Zua Wu**

National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
*OES, Industrial Technology Research Institute, Hsinchu, Taiwan, R.O.C.
hchuang@mail.nctu.edu.tw, hmhang@mail.nctu.edu.tw

## ABSTRACT [†]

An enhanced motion estimation scheme is incorporated into the interframe wavelet coding architecture in this paper. Interframe wavelet coding has the advantage of SNR, temporal, and spatial scalability, and is a potential candidate for the on-going MPEG-21 scalable video coding standard. Motion-compensated temporal filtering (MCTF) is one of its essential components. Therefore, motion estimation plays an important role in deciding the coding performance. In this paper, we modified the motion estimation syntax/scheme originally specified in the MPEG Advanced Video Coding (AVC) and use it in the interframe wavelet structure. Besides, the techniques of I-block, bi-directional motion estimation, λ-value adjustment and motion information partitioning are employed. Simulation results show very promising performance particularly on subjective quality.

## 1. INTRODUCTION

Video compression is an essential element in multimedia applications. Conventional video coding systems, including MPEG-1, MPEG-2, H.261 and H.263 international standards, employ the so-called *hybrid coding* structure. In these schemes, the reconstructed previous frame is used to predict the current frame after motion compensation.

Different from the aforementioned schemes, Ohm proposed a motion-compensated t+2D frequency coding structure [1]. The major difference between the hybrid coding and the t+2-D coding is that in the latter case, it does not contain the closed DPCM loop. In addition, the t+2-D coding is suitable for scalable video coding. One of the successful example of this concept is the interframe wavelet video coder proposed by Woods and his co-workers [2][3][4]. This scheme is called Motion Compensated Temporal Filtering – Embedded Zero Block Coding (MCTF-EZBC or MC-EZBC). The architecture of the interframe wavelet video coder is shown in Figure 1.

In this paper, we focus on improving the motion estimation scheme in the described interframe wavelet coding structure. At the end, we compare the performance between our proposed scheme and that in [3], and show the effectiveness of the proposed scheme.[‡]

This paper is organized as follows. In Section 2, we outline the motion estimation syntax/scheme in AVC. In Section 3, we incorporate the motion estimation scheme described in Section 2 into the interframe wavelet coding structure. Various techniques have been used to improve the subjective image quality such as I-block, bi-directional motion estimation and λ-value adjustment. Motion information partitioning is described in Section 4. Simulation results are shown in Section 5, in which we compare the results with the existing scheme.

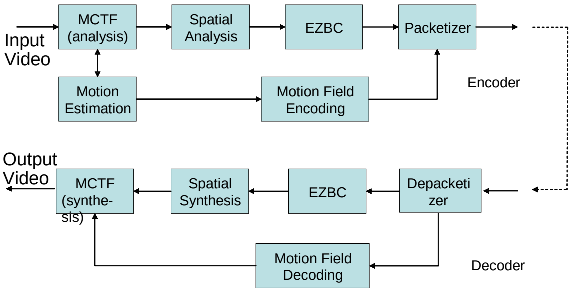


**Figure 1.** The interframe wavelet video coder.

## 2. MOTION ESTIMATION IN AVC

Motion estimation plays an important role in interframe wavelet coding. In this paper, we replace the hierarchical motion estimation algorithm in [3] by a modified version of the motion estimation scheme in the advanced video coding (AVC) standard [5].

[†] This work is partially supported by National Science Council (Taiwan, ROC) under Grant NSC 92-2219-E-009-008 and OES, Industrial Technology Research Institute (Taiwan, ROC) under Grant C92144.

[‡] We would like to thank Woods and Chen for providing us the source codes of their MC-EZBC algorithm. The experiments in this paper are based on our modified version of these codes.

The motivation is to improve the motion compensated filter in [3]. This is because in the interframe wavelet coding structure, the error between the original and reconstruction frames accumulate due to inaccurate motion estimation. In addition, the motion-compensated temporal filtered frames are reference pictures in temporal scalability. They are decoded pictures shown in the temporally down-sampled playback.

The motion estimation scheme in AVC has three main parts: (i) tree structured motion compensation, (ii) sub-pel motion vectors, and (iii) motion vector prediction. They are outlined below.

### 2.1. Tree structured motion compensation

The basic unit in AVC motion estimation is the $16\times16$ macroblock structure. The luminance part of each macroblock can be divided into four types of sub-macroblocks, namely, $16\times16$, $16\times8$, $8\times16$, and $8\times8$.Besides, the $8\times8$ sub-macroblocks can further be partitioned into $8\times8$, $8\times4$, $4\times8$, and $4\times4$ blocks.

### 2.2. Sub-pel motion vectors

After completing motion search, the border of the reference picture is used for padding, and the full-pel motion estimation finds the best-matched motion vector and the mode with the least cost. After the full-pel search, the interpolated picture is used for $\frac{1}{2}$-pel and $\frac{1}{4}$-pel motion search.

### 2.3. Motion vector prediction

The motion vector of one block is highly correlated with those of its neighboring blocks. This phenomenon becomes more apparent when the block sizes get smaller. Thus, we can make use of the left, upper-left, upper, and upper-right blocks to reduce the correlation among near-by motion vectors.

## 3. ENHANCED MOTION ESTIMATION FOR MCTF

We adopt the afore-mentioned motion estimation scheme for the MCTF component [6][7] in MC-EZBC [3]. Temporal subband decomposition is achieved by applying high-pass and low-pass filtering along the temporal axis. Motion compensated techniques are necessary to produce better compression performance by effectively removing the temporal redundancy.

### 3.1 MCTF structure

The MC-EZBC coder processes one group of picture (GOP) at a time. Each GOP contains $2^n$ frames, where $n$ equals to the levels of temporal subband decompositions in one GOP. The temporal subband decomposition process is performed by first constructing the motion vector map between two consecutive frames, and then the motion compensated temporal filtering (MCTF) is applied to these two frames to generate the temporal high- and low-pass frames. The temporal low-pass frames are grouped as another sub-set of GOP, and these frames are further temporally decomposed again. Decomposition process, illustrated in Figure 2 [8], is iterated until there is only one temporal low-pass frame, and a temporal filtering pyramid is thus constructed.

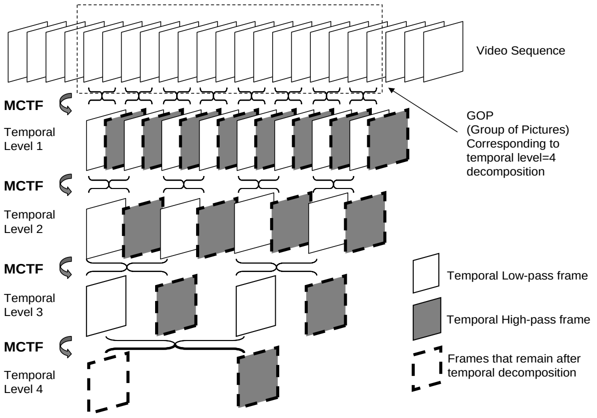


**Figure 2.** Temporal filtering pyramid

### 3.2 Lifting scheme temporal filtering

The temporal filtering operation in interframe wavelet coding is the so-called lifting scheme [9], which can achieve perfect reconstruction even when sub-pel motion estimation is used. Before temporal filtering, we find the connection relationship of pixels between the reference and the predicted frames. Then we follow the motion trajectory to generate temporal low-pass ($L[m,n]$) and high-pass ($H[m,n]$) frames. Figure 3 [3] shows the state of each pixel defined in MCTF as specified by Eqs.(1)-(5).

After the detection of the connection state of each pixel, Eq. (1) is used to generate the high-pass frame and Eq. (2) to generate the low-pass frame for connected pixels, and Eq. (3) is used for unconnected pixels [4].

$$H[m,n]=\left(A[m,n]-\tilde{B}[m-d_m,n-d_n]\right)/\sqrt{2} \tag{1}$$

$$L[m,n]=\tilde{H}[m+d_m,n+d_n]+\sqrt{2}B[m,n] \tag{2}$$

$$L[m,n]=\sqrt{2}B[m,n] \tag{3}$$

At the decoder, we can do the same interpolation on $H$ and reconstruct $A$ exactly by using Eq. (4) for connected pixels and the inverse process in Eq. (3) for unconnected ones if there is no quantization error.

$$B[m_m,n]=\left(L[m,n_n]-\tilde{H}[m+d_m,n+d_n]\right)/\sqrt{2} \tag{4}$$

Then, $A$ can be reconstructed by Eq. (5).

$$A[m,n]=\sqrt{2}H[m,n]+\tilde{B}[m-d_m,n-d_n] \tag{5}$$

The interpolator we use is the one for generating $\frac{1}{4}$-pel accuracy reference frame in AVC motion estimation.

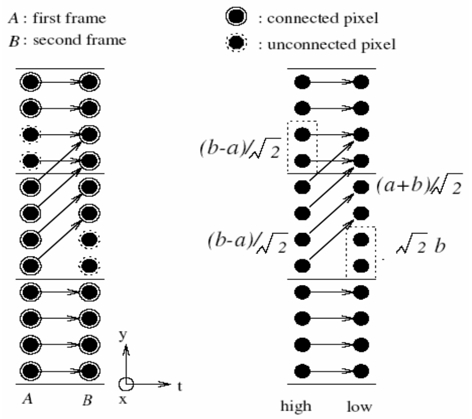


**Figure 3.** State of connection of each pixel

### 3.3 I-block and bi-directional motion estimation

The concepts of I-block and bi-directional motion estimation for MCTF were described in [3]. We adopted these concepts with some modifications in our MCTF scheme.

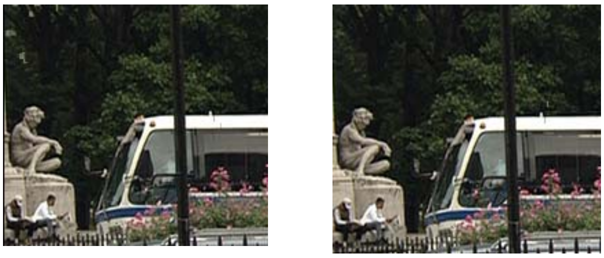
The temporal low-pass frame is generated by Eqs. (2) and (3) based on the state of connection of each pixel. Typically, motion compensation works well on the connected pixels. However, it is possible to have connected blocks with a poor match after motion estimation. These blocks tend to produce artifacts in the temporal low-pass frame, which lead to poor visual quality for temporal scalability. These blocks are forced to be unconnected as proposed in [3][4].

Our I-block size is 16x16. As shown in Figure 3, let $A[m,n]$ be the block with connected state at the location $(m,n)$ of the A frame and $B[m-d_m,n-d_n]$ be the motion-compensated block with motion vector $(d_m,d_n)$ in the B frame. We compute the variance of these two blocks, and choose the minimum as $V_{min}$. If the mean squared prediction error between these two blocks is larger the threshold $F*V_{min}$, this block is declared as an unconnected block, where F is an adjusting parameter. Based on our experiments, F is taken around 0.7. Figure 4 shows the subjective improvement (left upper corner, for sample) using the I-blocks.

Furthermore, an A frame block may find a better match (motion compensation) from the previous B frame. Thus, frame A has both forward and backward motion vectors. The use of bi-directional motion estimation reduces high-pass frame magnitude and thus increases coding efficiency.

### 3.4 Motion cost function adjustment

The rate-distortion cost function, $J=D+\lambda R$, is used to decide the best motion vectors in the AVC motion estimation, in which $D$ is the frame difference, and $R$ is the estimated motion vector coding bits. However, as the temporal level increases in MCTF, the energy of temporal low-pass frame is also increased. Therefore, the $\lambda$ value should be increased to maintain a constant rate-distortion relation at the higher temporal levels. Therefore, the $\lambda$ value is increased by a factor of $\sqrt{2}$ for each additional temporal level.



(a) (b)

**Figure 4.** The 2nd temporal low-pass frame: (a) without I-block (b) with I-block.

## 4. MOTION INFORMATION PARTITIONING

In [8], Hang and Tsai proposed the concept of motion information scalability for MC- EZBC. In this paper, we adopted this concept with some modifications to partition the motion information generated by the AVC inter-frame-prediction in MCTF. If the required bitrate is very slow, the extractor may fail to extract the bitstream because the motion information bits are larger than the specified bits. Also, at low rates, we may want to save some bits from motion information and use these bits for wavelet coefficients to achieve acceptable quality. Therefore, we partition motion information after motion estimation according to the steps below.

*Step 1*: Do 16x16 block size motion search with integer-pixel accuracy. The generated motion vectors are the *base layer* motion vectors.

*Step 2*: Do 16x16 and 8x8 block size motion search with 1/2-pixel accuracy. The difference between these motion vectors and the base-layer is the *first enhancement layer* motion vectors.

*Step 3*: Do all sub-block size motion search with 1/4-pixel accuracy. The difference between these motion vectors and the base-layer plus the first-enhancement-layer is the *second enhancement layer* motion vectors.

*Step 4*: Encode the above three layers motion information using CABAC separately.

If the required bitrate is too small, the extractor will drop one or two enhancement layers according to the given conditions. Furthermore, if one likes to extract the spatially down-sampled bitstream, the extractor can also drop proper the enhancement layers. When the codec scalability range is small, we can reduce the enhancement layers to one to save bits in encoding motion vectors.

The proposed algorithm can provide an acceptable video quality at very low bit rates, especially for high-motion cases. However, if not all the motion vectors are used in reconstruction, the "mismatch errors" would occur. That is, the residual image data calculated at the encoder are based on the complete set of motion vectors but only "partial" motion vectors are available at the decoder if they are truncated. This problem will be further studied.

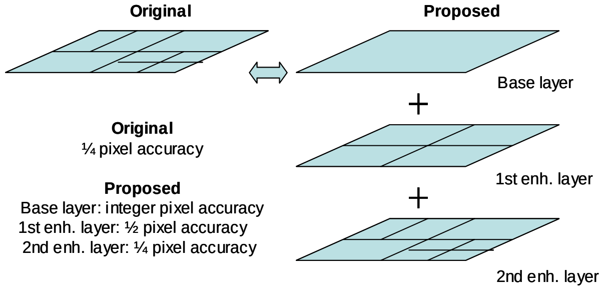

Original
Proposed
Base layer
Original
¼ pixel accuracy
1st enh. layer
Proposed
Base layer: integer pixel accuracy
1st enh. layer: ½ pixel accuracy
2nd enh. layer: ¼ pixel accuracy
2nd enh. layer

**Figure 5.** The base and enhancement layer motion vectors.

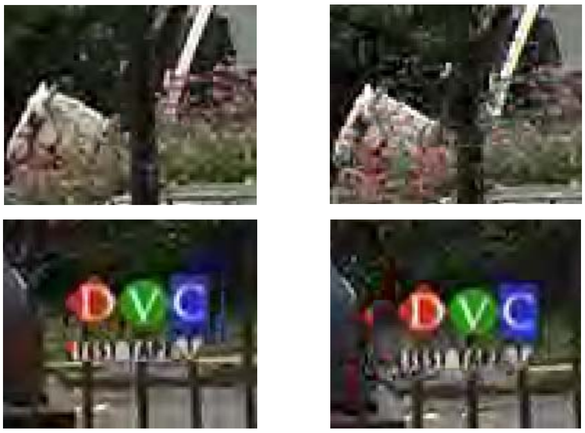
## 4. SIMULATION RESULTS

We perform two sets of simulations to show the effectiveness of the algorithm proposed in this paper. In our MCTF, we adopt both the motion vector (MV) syntax and the arithmetic coding for MV in AVC. We compare the results with those in [3]. We found improvements both objectively and subjectively but the subjective performance is more important because the final judgment of an image processing algorithm is the subjective picture quality.

In the MPEG scalable video coding call-for-proposal [10], the MPEG committee specifies three main test conditions. For the test sequence Bus_CIF, one test point in Test 2b is 30 frames per second (fps) at 512kbps [10]. The subjective quality of these two coding schemes at this test point is compared. AVC has a very complicated interframe prediction scheme, and the motion block size could be one of seven block types. Therefore, the motion estimation is very accurate. As we can see from Figure 6, the proposed coding scheme has a better subjective quality. But the PSNR value is not much changed.

## 6. CONCLUSION AND FUTURE WORK

In this paper, we propose an enhanced motion estimation scheme to improve the existing interframe wavelet coding algorithm (MC-EZBC). We modify the motion estimation syntax/scheme specified in AVC to fit into the motion compensated temporal filtering (MCTF) structure. Various additional techniques such as I-block, bi-directional motion estimation and λ-value adjustment are incorporated. Also, we propose the motion information partitioning technique for AVC interframe-prediction to improve coding performance at low rates. Preliminary simulation results indicate that this new motion estimation algorithm has improved the subjective image quality. Further parameter tuning should provide even better results.



(a) (b)

**Figure 6.** Subjective quality of the 2nd frame of Bus_CIF.yuv sequence at 512kbps with GOP=4: (a) proposed scheme, (b) MC_EZBC [3].

## 6. REFERENCES

[1] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[2] S.-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Processing: Image Communications*, vol. 16, pp. 705–724, May 2001.

[3] P. Chen, *Fully scalable subband/wavelet coding*, Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, New York, May 2003.

[4] T. Rusert, et al., *Recent Improvements to MC-EZBC*, ISO/IEC/JTC1 SC29/WG11 doc. M9232, Dec. 2002.

[5] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[6] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[7] T. Kronander, "Motion compensated 3-dimensional wave-form image coding," *Int'l Conf. Acoustic, Speech, and Signal Processing*, vol. 3, pp1921–1924, 1989

[8] S. S. Tsai, *Motion information scalability for interframe wavelet video coding*, MS thesis, National Chiao Tung University, Hsinchu, Taiwan, R.O.C., Jun. 2003.

[9] B. Pesquet-Popescu, V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," *Int'l Conf. Acoustic, Speech, and Signal Processing*, vol. 3, pp. 1793–1796, 2001.

[10] Call for proposals on scalable video coding technology, ISO/IEC JTC1/SC29/WG11 MPEG2003/N6193, Dec. 2003.

# Combined Encryption and Watermarking Approaches for Scalable Multimedia Coding

Feng-Cheng Chang, Hsiang-Cheh Huang, and Hsueh-Ming Hang

Department of Electronics Engineering, National Chiao Tung University,
Hsinchu 300, Taiwan, R.O.C.,
hmhang@mail.nctu.edu.tw

**Abstract.** Intellectual Property (IP) protection is a critical element in a multimedia transmission system. Conventional IP protection schemes can be categorized into two major branches: *encryption* and *watermarking*. In this paper, a structure to perform layered access protection by combining encryption and robust watermarking is proposed and implemented. By taking advantage of the nature of cryptographic schemes and digital watermarking, the copyright of multimedia contents can be well protected. We employ scalable transmission over the broadcasting environment, and the embedded watermark can be extracted with certain confidence measure, while the next-layer secrets can be perfectly decrypted and derived. This proves the effectiveness of the proposed structure.

## 1 Introduction

With the widespread use of multimedia broadcasting, the digital media, including images, audio and video clips, are easily acquired in our daily life. The current network environments make scalable coding of multimedia a necessary requirement when multiple users try to access the same information through different communication links [1, 2]. Scalability means that a multimedia data bitstream is partitioned into layers in such a way that the base layer is independently decodable into a content with reduced quality. The reduction may be in spatial resolution, temporal resolution, or signal-to-noise ratio (SNR). To reproduce the original content, enhancement layers provide additional data to restore the original quality from the base layer. Enhancement layers represent the scalability of the content coding, namely, spatial, temporal, or SNR scalability. Therefore, scalable coding of multimedia is suitable to deliver digital contents to different uses and devices with various capabilities [3].

In many cases, it requires to deliver multimedia content securely. However, the channel for multimedia broadcasting is an open environment, thus, if the user data and information are not protected, it might be illegally used and altered by hackers even if partial information is received by them. To protect privacy and intellectual property (IP) right, people often use cryptographic techniques to encrypt data, and thus the contents protected by encryption are expected to be securely transmitted over the Internet [4, 5].

In cryptography, the contents to be encrypted are called *plaintext*, while the encrypted contents are called *ciphertext*. Although cryptographic schemes provide secure data exchange among peers, it implies that the ciphertext cannot be altered during transmission [6]. If any one bit is received erroneously, the plaintext cannot be decrypted correctly. This is not a good property when we deliver protected contents in a broadcasting environment. Moreover, if we do not partition the protected content well, a one-bit error may cause a totally useless content. To meet this deficiency for multimedia broadcasting, we apply watermarking to aid encryption, because it allows the watermarked contents to experience some kinds of *attacks*, including signal processing, geometric distortion, and transmission errors. In this paper, we combine both the cryptographic and watermarking schemes for layered content protection. On the one hand, the message for protection of multimedia contents can be perfectly decrypted by cryptography, while on the other hand, the encrypted message can be further protected by robust watermarking algorithms to resist transmission errors.

This paper is organized as follows. Sec. 2 describes the concepts and issues of layered content protection. In Sec. 3, we propose a layered protection structure with combined cryptographic and watermarking schemes. We give an example application and some simulations in Sec. 4. And Sec. 5 concludes this paper.

## 2 Layered Protection Concepts

As mentioned in Sec. 1, scalable coding is a solution to broadcast contents to devices with various playback capabilities. With the nature of layered coding, the whole media can be treated as partitions of data. Thus, it is straightforward to group receivers of different playback capabilities by sending different combinations of data partitions. However, the conditional access problem is dealt with a different way in a broadcast environment. To distinguish different groups of users, a common solution is to encrypt data by a certain group-shared key. Thus, this issue can be solved by encrypting data partitions, and a granted user has the corresponding decryption keys to the assigned data partitions.

The next issue is how to distribute the keys. Depending on the delivery infrastructure, two problems may arise. One is how to protect keys from malicious listeners. There are methods to protect keys from malicious listeners, such as the one proposed in the DVB standard [7]. The other problem is how to synchronize a key with the content having a proper key delivery method. For example, to broadcast a protected content on Internet, we may send the key to users via a reliable channel (such as RTSP connection [8]), while the content goes through an unreliable channel (such as RTP sessions [9]). A reliable channel guarantees information correctness by sacrificing delivery speed, and it is very likely that the key information is out-of-sync to the corresponding content.

A possible solution to eliminate synchronization problem is to transmit the key information with the content, such as inserting it into the optional header fields of the coded stream. However, it is sensible to transmission errors or transcoding. Our proposed method has better resistance to this kind of problem:

we embed the key information into the content with robust watermarking techniques. Since the key information is available at the same time as we reconstruct the content, the synchronization problem is implicitly resolved. The drawback is that if packet loss or transcoding occurs, the reconstructed content is different from the original one, and the key information may not be extracted accurately. To reduce the impact of unreliable or distorted delivery, we incorporate robust digital watermarking methods [10] to make the embedded key information more robust.

An overall description of the layered protection is organizing secrets (keys and necessary parameters) into a watermark, robustly watermarking the base layer, and encrypting the enhancement layer. A granted user receives the base layer, extracts and derives the decryption key, decrypts the enhancement layer, and compose the layers to produce the contents with better quality. In the following sections, we will describe our proposed method in detail.

## 3 Proposed Method

In this section, we describe the layered decryption and decoding operations on the receiver side. Because the corresponding encryption and encoding operations vary depending on the scalable coding, we provide a possible scheme at the end of this section. We first demonstrate the receiver architecture in our proposed method, then we describe the corresponding transmitter architecture in the following paragraphs.

### 3.1 Receiver Architecture

Scalable coding is composed of one *base layer* and several *enhancement layers* to adapt with the network environment for transmission. The enhancement operation is illustrated in Fig. 1. Assuming that the initial base layer $B_0$ has been received, the subsequent composing operations can be expressed by

$$B_i = \text{compose}\ (B_{i-1}, E_i), \tag{1}$$

where

$$E_i = \text{decrypt}_e\ (X_i, K_i). \tag{2}$$

In Eq. (1), $B_{i-1}$ is the available base layer, and $E_i$ is the enhancement layer to improve quality from $B_{i-1}$ to $B_i$. During transmission, $E_i$ is protected by a cryptic algorithm with $K_i$ as the key, and the transmitted data is $X_i$ in Eq. (2).

There are some secret information to be obtained prior to decrypting $E_i$, and the operations can be expressed as follows:

$$W_i = \text{extract}\ (B_{i-1}, P_{i-1}) \tag{3}$$
$$F_i = \text{decrypt}_f\ (W_i, G_i) \tag{4}$$
$$K_i = \text{key}\ (F_i) \tag{5}$$
$$P_i = \text{param}\ (F_i) \tag{6}$$

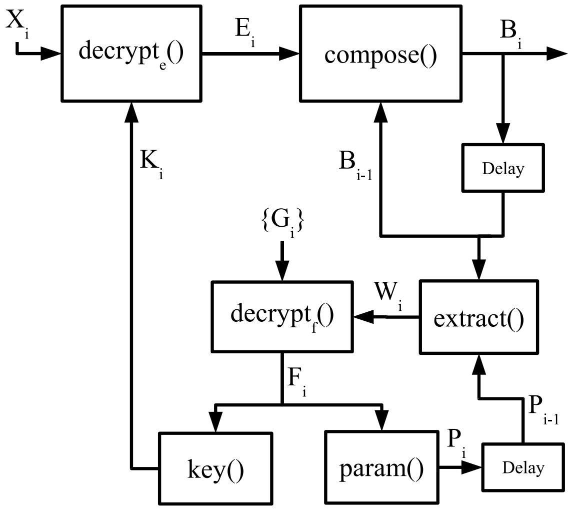


**Fig. 1.** Decryption and decoding of layer-protected content

$W_i$ is the digital watermark extracted from the constructed base layer $B_{i-1}$ with extraction parameter $P_{i-1}$. As described in Sec. 2, $W_i$ represents the protected secret information. Thus, we have the secret information $F_i$ by decrypting the watermark using user-specific key $G_i$. After parsing $F_i$, we obtain the decryption key $K_i$ and the next watermark extraction parameter $P_i$.

As Fig. 1 illustrates, the decryption and composition blocks are iterative processes. There are several initial parameters required to activate these processes. We will discuss how to obtain the initial parameters in the following paragraphs.

- When the whole content is protected, namely, $B_0$ is encrypted, we need $K_0$ to decrypt $X_0$. In this case, $K_0$ should be obtained in a separate channel. A possible way is to obtain it through the channel, and we get $\{G_i\}$.
- Some scenario provides $B_0$ as the "preview" layer, i.e., $B_0$ is not encrypted, we simply bypass the block of decrypt$_e$ ().
- Depending on the watermarking algorithm, the extraction process may requires specific parameters. If it does, the first watermark extraction parameter $P_0$ should be obtained in a separate channel to activate subsequent extraction process.
- All the key-protection keys $\{G_i\}$ should be obtained before receiving the media, for instance, by manually or automatically update after subscription.

### 3.2 Transmitter Architecture

Depending on the scalable coding algorithm, the design of transmitter side varies with different situations. Fig. 2 shows one of the possible designs. The architecture is almost the inverse of the receiver architecture in Fig. 1. The watermark $W_i$ is the encrypted version of the key $K_i$ and the embedding parameter $P_i$. The

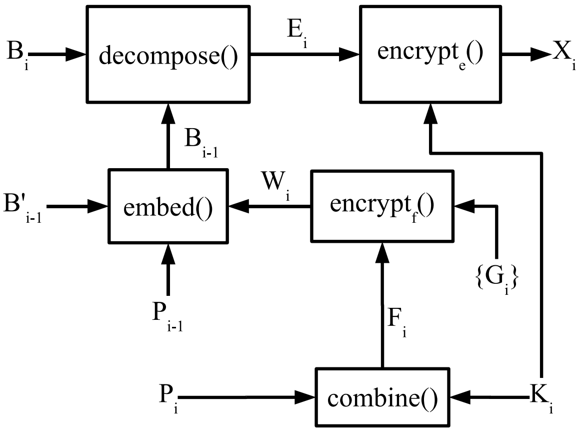


**Fig. 2.** Encryption and encoding of layer-protected content

$B'_{i-1}$ is the un-watermarked base layer with lower quality. After embedding $W_i$ into $B'_{i-1}$, we have the transmitting base layer $B_{i-1}$. The enhancement layers are generated as the differences between $B_i$ and $B_{i-1}$. All the $\{K_i\}$, $\{P_i\}$, and $\{G_i\}$ are known in advance.

## 4 Simulation Results

In this paper, we use the test image Lena with size 1024 × 1024 to conduct the simulations in this section. The original Lena is first converted into 512 × 512 base layer. The secret for representing the original image is the 8-byte (or 64-bit) message **NCTU–DEE**, shown in Fig. 3(a), to represent the authors' affiliation. The secret can be any message with 8-byte length. The 8-byte message is repeated for 32 times, and we employ DES [11] to encrypt the repeated message into ciphertext. Next, the ciphertext is converted into a binary watermark, shown in Fig. 3(b).

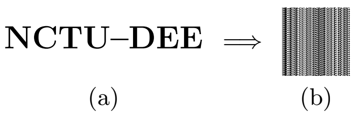
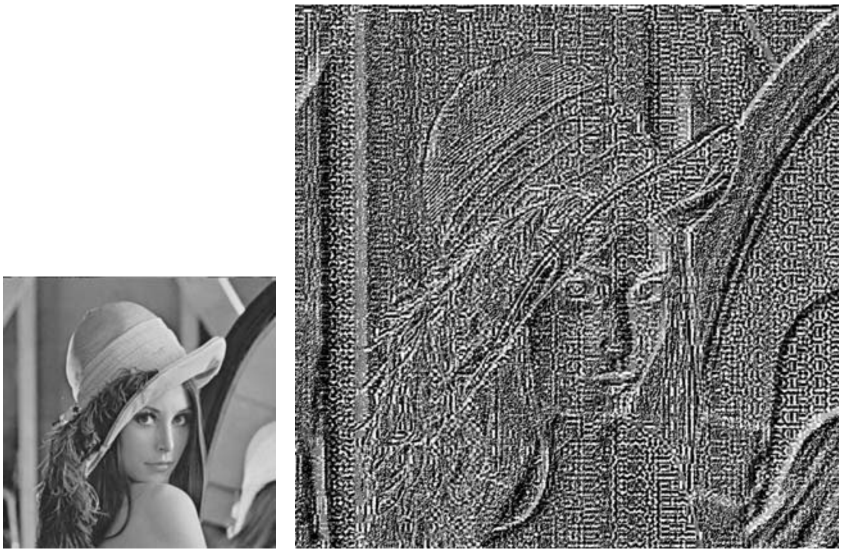


**Fig. 3.** Plaintext encryption and watermark generation. (a) The 8-byte plain-text. (b) The converted binary watermark with size 128 × 128.

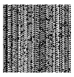
Fig. 4 presents the data in transmitted base layer and the enhancement layers. Before transmission, the watermarked base layer has acceptable visual quality, with the PSNR of 39.24 dB in Fig. 4(a). We can extract the watermark from the base layer picture, derive the decryption key, decrypt the transmitted enhancement data in the next layer, and finally reconstruct the original 1024 × 1024 picture.



(a) (b)

**Fig. 4.** (a) 512 × 512 base layer. (b) 1024 × 1024 enhancement layer.

We then perform packet loss on the base layer over the random packet loss channel [12]. The packet loss rate in our simulations is set to 10%. The extracted watermark is shown in Fig. 5(a). The distortion is within the tolerance range of the extracted watermark, with the bit-correct ratio of 92.74%. We then use majority vote technique to produce the 8-byte, extracted ciphertext, and decrypt the extracted ciphertext. Finally, we can recover the original key information correctly in Fig. 5(b). In addition, the 1024 × 1024 picture thus can be reconstructed to a certain extent as shown in Fig. 6.



$\Longrightarrow$ **NCTU–DEE**

(a) (b)

**Fig. 5.** Watermark extraction and cipher-text decryption. (a) The extracted watermark, with the bit-correct ratio of 92.74% to compare with Fig. 3(b). (b) The decrypted cipher-text, which is identical to that in Fig. 3(a).

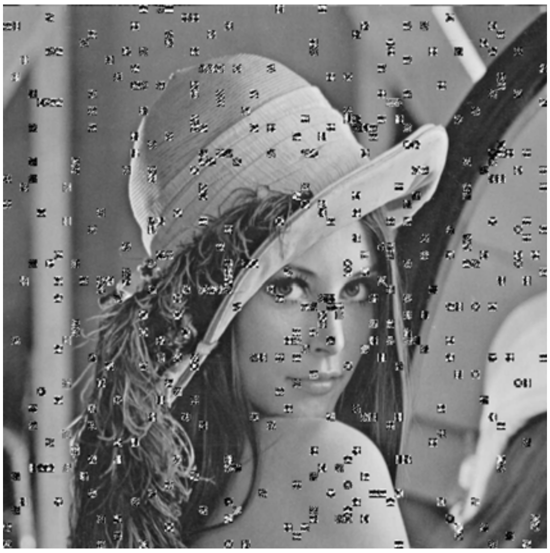


**Fig. 6.** The encrypted and watermarked image after transmission, with best-effort reconstruction.

## 5 Conclusion

In this paper, we proposed a structure to protect the layered (scalable) content in a broadcast environment. By combining cryptographic schemes and robust watermarking techniques, the secret for decrypting enhancement data streams can be safely embedded in each base layer. Watermarking allows to embed some bits of information directly in some multimedia content, and the embedded bits can be extracted even after the watermarked media experiencing attacks during transmission. In contrast, cryptography provides confidentiality, but even when one bit is altered in the encrypted media, the secret message therein cannot be correctly decrypted. The contribution in this paper is to build a link in the two main categories, and employ the advantages of both for intellectual property protection.

In the proposed scheme, the encryption concept guarantees the access control, keeping away malicious eavesdroppers. Also, the embedding concept solves the key-content synchronization problem, and the robust watermarking concept raise the resistant ability to transmission errors and distortions. Comparing with conventional cipher-block chaining encryption, our method not only provides a way to guarantee access controls, but also synchronously transmits decryption information. Moreover, robust watermarking implicitly gives higher data integrity protection for keys than that of contents. Although the protection is not as good as a reliable channel, it is a trade-off solution in a broadcast environment. The scalable image application and the simulation results prove the effectiveness of the proposed structure.

In our future work, we will modify our structure with scalable video coding. We will also integrate our proposed structure with MPEG IPMP (Intellectual Property Management and Protection) [13, 14].

## References

1. Sun, X., Wu, F., Li, S., Gao, W., Zhang, Y.-Q.: Seamless switching of scalable video bitstreams for efficient streaming. IEEE Transactions on Multimedia **6** (2004) 291–303
2. Almeida, J.M., Eager, D.L., Vernon, M.K., Wright, S.J.: Minimizing delivery cost in scalable streaming content distribution systems. IEEE Transactions on Multimedia **6** (2004) 356–365
3. Wiegand, T., Sullivan, G.J., Bjntegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Transactions on Circuits and Systems for Video Technology **13** (2003) 560–576
4. Parviainen, R., Parnes, P.: Large scale distributed watermarking of multicast media through encryption. Proceedings of the International Federation for Information Processing, Communications and Multimedia Security Joint Working Conference IFIP TC6 and TC11 (2001) 149–158
5. Lim, Y., Xu, C., Feng, D.D.: Web-based image authentication using invisible fragile watermark. Conferences in Research and Practice in Information Technology (2002) 31–34
6. Xu, X., Dexter, S., Eskicioglu, A.M.: A hybrid scheme for encryption and watermarking. IS&T/SPIE Symposium on Electronic Imaging 2004, Security, Steganography, and Watermarking of Multimedia Contents VI Conference (2004) 723–734
7. Digital video broadcasting project (DVB): http://www.dvb.org/ (2004)
8. Real time streaming protocol: http://www.rtsp.org/ (2004)
9. Schulzrinne, H.: http://www.cs.columbia.edu/~hgs/rtp/ (2004)
10. Shieh, C.S., Huang, H.C., Wang, F.H., Pan, J.S.: Genetic watermarking based on transform domain techniques. Pattern Recognition **37** (2004) 555–565
11. Data Encryption Standard (DES): http://www.itl.nist.gov/fipspubs/fip46-2.htm (1993)
12. Chande, V., and Farvardin, N.: Progressive transmission of images over memoryless noisy channels. IEEE Journal on Selected Areas in Communications **18** (2000) 850–860
13. Avaro, O., Eleftheriadis, A., Herpel, C., Rump, N., Swaminathan, V., Zamora, J., Kim, M.: MPEG systems (1-2-4-7) FAQ, version 17.0. ISO/IEC JTC1/SC29/WG11 N4291 (2001)
14. Huang, C.C., Hang, H.M., Huang, H.C.: MPEG IPMP standards and implementation. IEEE PCM'02 (2002) 344–352