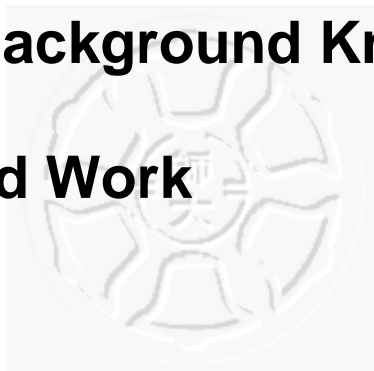


# Chapter 2 Background Knowledge and Related Work



## 2.1 Virtual Museum over Internet

在古蹟文物的展示上，由於博物館空間的限制往往難以展示全部的典藏品，同時基於保護與維護典藏品之需要，典藏品也多半置於櫥窗內以供民眾欣賞，基於安全性的考量民眾很難進一步欣賞或是觸摸這些典藏品，自然也無法與這些文物有不同的互動。因此許多博物館致力於建構 3D 數位化虛擬博物館 [3][4][5][6][13][14][15][16]，透過虛擬博物館展示(Virtual Museum Exhibitions)提供觀賞者與文物有別於傳統展示所無法達到的互動效果。同時隨著網路的四通八達，3D 數位典藏品除了透過固定的場館或地點來觀賞，更可以透過網路作遠距離的展示，而讓不同地方的觀賞者觀賞這些收藏於博物館中的珍貴古物或收藏品，使虛擬博物館成爲一個超越時間與空間的理想展示平台。

博物館採用 3D 虛擬展示一般都會遭遇到兩個問題，第一如何有效建立典藏品的三維模型；第二如何根據這些模型建構出虛擬展示環境。過去建立 3D 虛擬文物或展示環境都是一件費時且不經濟的工作，形成博物館在建構 3D 虛擬展示上的阻礙。隨著近幾年在三維建模技術的進步，使得數位化的工作變得比較容易且符合經濟效益。有了典藏品的 3D 數位化模型之後，如何根據這些物品簡單且有效建構出虛擬實境(Virtual Reality) 或由電腦視覺角度出發所研究的增添式虛擬實境(Augmented Reality)的展示環境就成爲重要的研究課題，這樣的系統不但

要簡便使得博物館人員可以輕易操作，更要能提供直覺化的使用者介面，讓觀賞者能以自然的方式與典藏品互動。

目前有許多博物館 3D 虛擬展示的例子，像 Wojciechowski et. al.[4] 提出的 Augmented Representation of Cultural Objects (ARCO)是一套以 X-VRML 為基礎輔助博物館建立、操作與管理數位化文化資產之系統。此系統採用的建模技術會根據物體的複雜度而不同，簡單物體的建模是利用 3ds max 之類的軟體搭配 ARCO 發展的 plug-ins；複雜的物體則利用 stereo digital camera system 來建立模型。數位化後的資料會以 XML 格式存放在資料庫，並以 ARCO Content Management Application(ACMA)加以管理。在視覺化呈現方面，ARCO 使用 X-VRML 技術在使用者介面上動態產生虛擬展覽，X-VRML 為高階 XML-based 語法，類似 VRML 或 X3D 語言，並能夠動態的控制虛擬場景及模型。

在目前的研究中，相對於傳統的三維圖學技術，影像式描繪技術(Image-Based Rendering, IBR)[8]也被應用於建構虛擬博物館。環場影像與環物影片都是屬於影像式描繪技術(Image-Based Rendering, IBR)的一種，透過物體或場景的影像達到虛擬 3D 模型的效果。對於複雜的物體而言，相較於以往三維建模的方式，以影像式描繪技術來達成文物 3D 數位化是比較快速且經濟的方法，同時 3D 數位化文物又可有如照片般真實的視覺效果。在國內數位典藏技術研發分項計畫中，Huang et. al.[7]發展了一個以影像為基礎(image-based)的高擬真虛擬環境展示技術稱為增添式環場(Augmented Panorama)來建構 3D 虛擬博物館。增添式環場技術將環物影像 (Object Movie) 以幾何一致性的方式逼真地合成於環場影像 (panorama) 中使得觀察者在觀看虛擬場景時有逼真的視覺效果。除了一般透過滑鼠操控介面之外，有別於傳統博物館與使用者之間的靜態互動關係，為了提昇

使用者操作時的融入感與真實感，並同時提供更多數位影音的文物導覽與介紹，他們更進一步建構出一套以三維視覺追蹤技術建構互動式擬真博物館系統（Tangible Photo-Realistic Virtual Museum）[3]。

影像式描繪技術雖能提供文物逼真之視覺效果，但是爲了提高觀賞者與虛擬物體之間的互動性，勢必須要大量的影像才能有良好的操作自由度。以上述的系統爲例，每個環物影片需要 360 張影像，每張影像的大小爲 512x512 pixel。沒有任何處理的情況下，一個環物影片的資料量就多達 276 MB，即使用 JPEG 技術壓縮一個環物影片也要 30MB 左右，這樣的資料量大小對於網路傳輸來說都是相當沈重的負擔。因此如何藉由壓縮處理大量的影像資料，一向都是 IBR 技術中重要的研究課題[8]。本文將基於這個系統，採用 H.264/AVC 視訊壓縮技術降低環物影片資料量大小，並進一步利用環物影片的特性提升 H.264 壓縮與解壓縮的效能，最後結合 cache 架構達成即時描繪(just-in-time rendering)的系統。

## 2.2 Object Movie

環物影片(Object Movie) 是由許多物體影像所組成的集合，每張影像對應到不同相機拍攝的方位。每個方位是相機朝向物體中心所拍的角度，通常以相機水平角度(pan)與垂直角度(tilt)來描述。因此環物影片除了為物體的影像集合之外，每張影像還存在著空間上的二維關係，也就是說，根據觀賞的角度可以選出一張相對應的影像。

Apple 的 QuickTime VR(QTVR)為環物影片在第一個商業化軟體，藉由 QTVR 提供的編輯軟體以及播放軟體(QuickTime 2.0+)，可以在 Windows 或 Macintosh™ 觀看環物影片。一般而言，環物影片的拍攝需藉由所謂的 motion-controllable 相機，或者利用特別的裝置，例如 Kaidan[13]特別為拍攝環物影片而設計的拍攝架(如圖 2.1)，才能獲得播放平順的環物影片。

在 QTVR 中環物影片是以類似視訊的方式被儲存，並在檔頭的部分加上每一列、每一行的個數以及每張的角度資訊。QTVR 對於影像的安排是以每列優先方式，使得影片水平旋轉時有最短搜尋時間。QTVR 提供了三種壓縮環物影片的方式，以降低環物影片的檔案大小。這三種方式分別是 Photo JPEG、Cinepak、Sorenson：Photo JPEG 將每張影像以 JPEG 格式獨立壓縮存放，通常有最好的影像品質但檔案大小也比較大。Cinepak 則利用了影像間的相似性，產生較小的檔案，但是在低頻寬時只能提供普通的品質。Sorenson 則更進一步利用相似性，在低頻寬時提供較佳的影像品質。Cinepak 和 Sorenson 均使用了類似視訊的方式性來壓縮環物影片，視訊的壓縮只考慮了影像間一維序列的關係，但環物影片的是存在著二維的相似性，但這兩個壓縮法都忽略了這樣的關係。

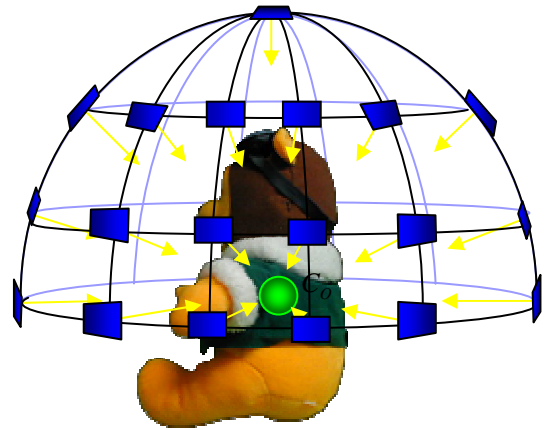


圖 2.1 環物影片拍攝架

## 2.3 H.264/AVC

在之前的章節中，我們提到環物影像壓縮之必要性與目前常見之壓縮方法，接下來我們將探討為何以 H.264 壓縮環物影片以及介紹 H.264 演算法之特性。本文採用目前最新的 H.264/AVC (MPEG4-part 10) 視訊壓縮技術，原因有兩個：第一，H.264 比之前的視訊壓縮標準在壓縮的效率上要增進許多，例如在相同的壓縮品質下比 H.263v2 節省 50% 以上的 bitrate，符合博物館對於虛擬展示時高品質影像的要求；第二，H.264/AVC 壓縮標準除了提出視訊編碼層外，還提出了網路提取層，使得 H.264/AVC 的串流適合透過網路傳輸。基於以上的兩點，使得 H.264/AVC 適合用在博物館的虛擬展示上，透過高品質的壓縮效果，呈現精美的典藏物品，並且能夠將虛擬展示透過網路傳送到世界各地。以下簡單介紹 H.264/AVC 這個新的視訊壓縮標準。

### 2.3.1 前言

H.264/AVC 或者 MPEG-4 Part10 是 2001 年 12 月由 ITU-T VCEG 與 ISO MPEG 共同組成聯合視訊小組(Joint Video Team, JVT)所研訂新的視訊壓縮格式。此新格式在 ITU-T 組織中稱為 H.264，在 ISO 組織中則納入 MPEG-4 Part-10 (ISO/IEC 14496-10) 並命名為 Advanced Video Coding (AVC)，通常合併稱為 H.264/AVC [17]，其國際標準的第一版於 2003 年公佈，而增修的第二版也於 2005 年 3 月定案，稱之為 Fidelity Range Extensions (FRExt)。

FRExt 支援更高的像素精度(包括 10 bits 和 12 bits 像素精度)和支持更高的色度精度(包括 YUV 4:2:2 和 YUV 4:4:4)來支持更高精度的視頻編碼。同時還加入

了一些新的特性，比如 8x8 區塊大小的 prediction 和 integer transform、使用者自定加權量化矩陣，無失真編碼，支持新增的色度空間和色度參差變換(residual color transform)。

相關研究顯示 H.264/AVC 與 MPEG-2 及 MPEG-4 相較之下，無論是壓縮率或視訊品質皆有大幅的提升[19]。而且 H.264/AVC 也首次將視訊編碼層(Video Coding Layer, VCL)與網路提取層(Network Abstraction Layer, NAL)的概念涵蓋進來(如圖 2.2)，以往視訊標準著重的是壓縮效能部分，而 H.264/AVC 包含一個內建的 NAL 網路協定適應層，藉由 NAL 來提供網路的狀態，可以讓 VCL 有更好的編解碼彈性與糾錯能力，使得 H.264/AVC 適用於多媒體串流(multimedia streaming)、行動電視(mobile TV)或是高品質 DVD 的資料壓縮的相關應用。

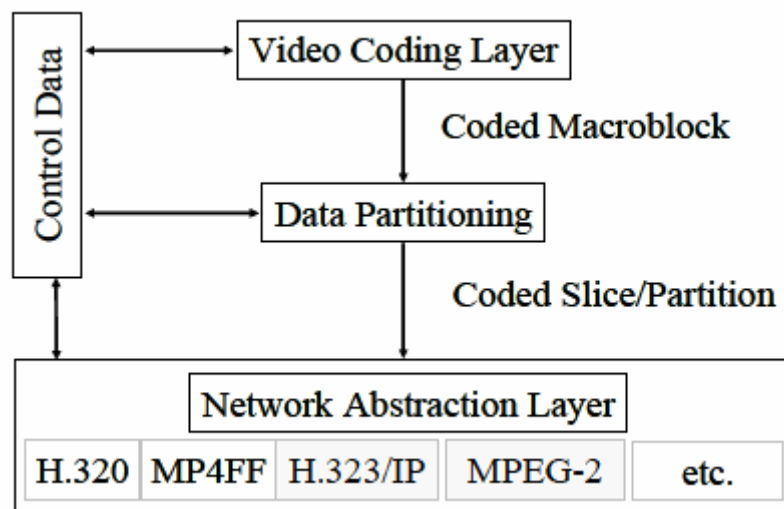


圖 2.2 H264/AVC 視訊壓縮架構圖

## 2.3.2 視訊編碼層

視訊壓縮的原理是利用影像在時間與空間上存有相似性，這些相似的資料經過壓縮演算法處理之後，可以將人眼無法感知的部分抽離出來，這些稱為視覺冗餘 (visual redundancy) 的部分在去除之後，就可以達到視訊壓縮的目的。如圖 2.3 所示，H.264/AVC 的視訊編碼機制是以區塊(block-based) 為基礎單元，也就是說先將整張影像分割成許多矩形的小區域，稱之為 macroblock (MB)，再將這些 MB 進行編碼。首先，利用運動估測 (motion estimation) 與運動補償 (motion compensation) 的方式，去除影像內 (intra-prediction) 與影像間 (inter-prediction) 的相似性，得到所謂的差餘影像 (residual)，再將差餘影像施以空間轉換 (transform) 與量化 (quantize) 來去除視覺冗餘，最後視訊編碼層會輸出編碼過的位元流 (bitstream)，之後再包裝成網路提取層的單元封包 (NAL-unit)，經由網路傳送到遠端或儲存在儲存媒體中。H.264/AVC 允許視訊影片以 frame 或是以 filed 的方式來進行編碼，兩者可以共存，而 frame 可以是 progress 或是 interlace 形式，對同一段影片來說也可使用兩者來混合編碼。

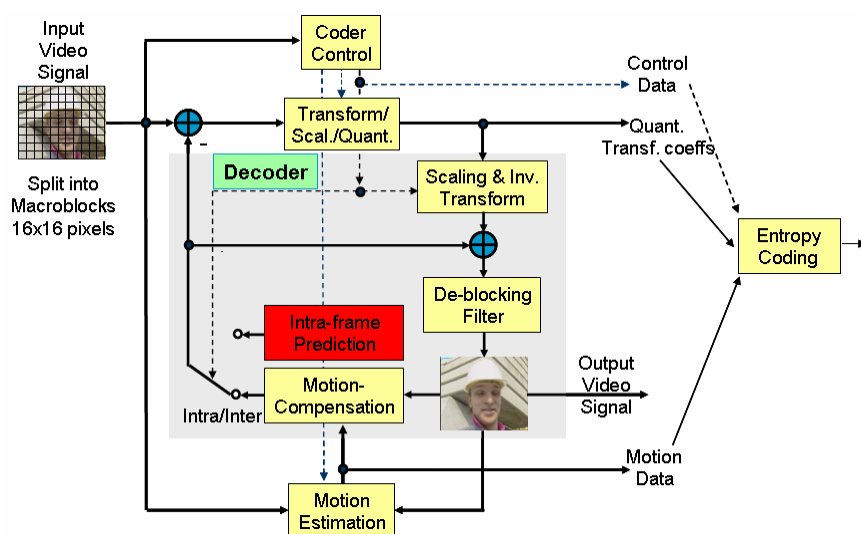


圖 2.3 H.264/AVC 區塊壓縮結構示意圖



H.264/AVC 包含了一系列新的特徵，使得它比起以前的編解碼器不但能夠更有效的進行編碼，還能在各種網絡環境下的應用中使用。主要的特性包括：

1. 多重參考 frame 的運動補償，在先前的視訊壓縮標準(如 MPEG-2)，每張 P frame 都只能有一張的參考 frame，從 H.263++開始允許選擇多張以壓縮的影像當作參考影像。
2. 可變區塊大小的運動補償，根據影像內容可以選擇不同大小的區塊進行運動估測，最小的區塊為 4x4 pixel。
3. 1/4 像素精度的運動補償，能夠提供更高精度的運動區塊預測，由於色度(chroma)通常是亮度(luma)抽樣的 1/2，這時運動補償的精度就達到了 1/8 像素精度。
4. 加權的運動估測(weighted prediction)，指在運動補償時可以在 encoder 使用增加權重(weighted)和偏移(offset)的辦法。它能在一些特殊的場合，如淡入、淡出、淡出而後淡入等情況提供相當大的編碼效益。
5. 增加 skipped 和 direct 的運動估測方式，在之前的標準，一個被 skipped 的區域是不能有任何的移動，這樣對於當影像有穩定的全域的移動時，是沒有任何幫助的。因此 H.264 提供在 skipped 的區域中可以藉由之前的預測結果推論出區塊的移動，而 direct 模式是用在 bi-predictively 的區域(B slice)，對於往前或往後的預測直接推論出區塊的運動。
6. In-Loop Deblocking Filter，能夠減少因為離散餘弦變換(DCT)所造成的視頻編解碼器的區塊效應。

7. 在經過 DCT 轉換之後，對 DC 係數再進行一次 Hadamard Transform，使得在平滑區域得到更好的壓縮效果。
8. 支援 4x4 區塊大小的 transform，過去的標準 block transform 的大小都為 8x8，H.264 則基於 4x4 的區塊進行 transform。如此可以減少鋸齒現象。
9. 算術編碼技術(Arithmetic Entropy Coding)，在 H.263 標準中算術編碼技術列為選擇性項目，在 H.264 此一技術已經包含在標準之中，並進一步改進算術編碼提供更好的壓縮效能，稱之為 CABAC(context-adaptive binary arithmetic coding)。
10. Context-Adaptive Entropy Coding，在 H.264 標準中兩種的 entropy 的壓縮法都採用基於上下文的方式調整壓縮，分別為 CAVLC (context-adaptive variable-length coding) and CABAC，相較於之前壓縮標準都提供更好的效能。

上述這些特色的結合，使得 H.264 比起以前的視頻編解碼標準能夠在性能上具有顯著提升，並更廣泛地應用在各種不同的環境。H.264 在壓縮性能上比起 MPEG-2 有很大的提高，在相同的 PSNR 下可以將碼率減少到一半或者更少。和 MPEG 的其它視頻標準一樣，H.264/AVC 也提供了一個參考軟體 JVT reference model software,[20]。

### 2.3.3 網路提取層

H.264/AVC 標準的另一個重要的特色是在 VCL 之上定義了網路提取層 (Network Abstraction Layer, NAL)，亦即以 NAL 封包做為 VCL 編解碼的處理基本單位。傳輸層拿到 NAL 封包之後不需要再進行切割，只需附加該傳輸協定的檔頭資訊(adding header only)就可以交由底層傳送出去，可以將 NAL 當成是一個專作封裝(packaging)的模組，用來將 VCL 壓縮過的 bitstream 封裝成適當大小的封包單位，並在 NAL-unit Header 中的 NAL-unit Type 欄位記載此封包的型式。每種型式分別對應到 VCL 中不同的編解碼工具。

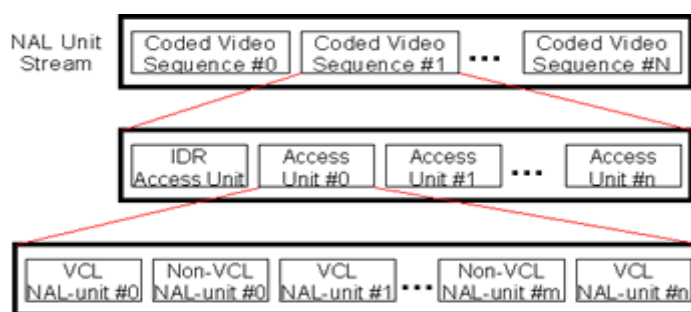


圖2. Hierarchical structure of NAL unit stream.

圖 2.4 NAL unit stream 階層架構圖

NAL 另外一個重要的功能為當網路發生壅塞而導致封包錯誤或接收次序錯亂 (out-of-order)的狀況時，傳輸層協定會在 Reference Flag 作設定的動作，接收端的 VCL 在收到這種 NAL 封包時，就知道要進行所謂的糾錯運算(error concealment)，在解壓縮的同時也會嘗試將錯誤修正回來。如圖 2.4 所示，一個完整的 H.264/AVC bitstream 是由多個 NAL-units 所組成的，所以此 bitstream 也稱之為 NAL unit stream。一個 NAL unit stream 內可以包含多個壓縮視訊序列 (coded video sequence)，一個單獨的壓縮視訊序列代表一部視訊影片，而壓縮視訊序列又是由多個 access units 所組成，當接收端收到一個 access unit 後，可以完

整地解碼成單張的畫面，而每個壓縮視訊序列的第一個 `access unit` 必須為 `Instantaneous Decoding Refresh (IDR) access unit`，`IDR access unit` 的內容全是採用 `intra-prediction` 編碼，所以自己本身即可完全解碼，不用參考其他 `access unit` 的資料。

`Access unit` 是由多個 `NAL-units` 所組成，標準中總共規範 12 種的 `NAL-unit` 型式，這些型式可以進一步分類成 `VCL NAL-unit` 及 `non-VCL NAL-unit`：所謂的 `VCL NAL-unit` 純粹是壓縮影像的內容，而所謂的 `non-VCL NAL-unit` 則有兩種：`Parameter Sets` 與 `Supplemental Enhancement Information (SEI)`，`SEI` 可以存放影片簡介、版權宣告、使用者自行定義的資料...等；`Parameter Sets` 主要是描述整個壓縮視訊序列的參數，例如：長寬比例、影像顯現的時間點(`timestamp`)、相關解碼所需的參數...等，這些資訊非常重要。萬一在傳送的過程中發生錯誤，會導致整段影片無法解碼，以往像 `MPEG-2/-4` 都把這些資訊放在一般的 `packet header`，所以很容易隨著 `packet loss` 而消失，現在 `H.264/AVC` 將這些資訊獨立出來成為特殊的 `parameter set`，以使用最高層級的通道編碼(`channel coding`)保護機制，來保證傳輸的正確性。