

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

Mai Thị Lan Oanh

**CÁC CHUẨN NÉN VÀ ỨNG DỤNG TRUYỀN VIDEO
TRÊN MẠNG INTERNET**

KHOÁ LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành : Điện Tử - Viễn Thông.

HÀ NỘI - 2005

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

Mai Thị Lan Oanh

**CÁC CHUẨN NÉN VÀ ỨNG DỤNG TRUYỀN VIDEO
TRÊN MẠNG INTERNET**

KHOÁ LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành : Điện Tử - Viễn Thông.

**Cán bộ hướng dẫn: TS. Ngô Thái Trí
Cán bộ đồng hướng dẫn: TS. Đinh Quốc Tuấn**

HÀ NỘI - 2005

LỜI CẢM ƠN

Em xin bày tỏ lòng kính trọng và biết ơn sâu sắc tới thầy giáo, TS.Ngô Thái Trị, người thầy đã trực tiếp dạy và hướng dẫn em rất tận tình, cho em những định hướng và ý kiến quý báu về công nghệ truyền hình. Em cũng xin bày tỏ lòng biết ơn sâu sắc tới TS. Đinh Quốc Tuấn, người thầy đã giúp đỡ em rất nhiều, cho em rất nhiều kiến thức bổ ích trong thời gian em thực tập tại Phòng Công Nghệ Thông Tin – ĐTHVN cũng như trong thời gian em làm luận văn. Đồng thời, em xin gửi lời cảm ơn chân thành tới tất cả các Anh, các Chị tại Trung Tâm Tin Học và Đo lường – ĐTHVN đã tạo điều kiện và giúp đỡ em rất tận tình trong thời gian em thực tập tại trung tâm.

Em xin chân thành cảm ơn tất cả các thầy cô trong Trường Đại Học Công Nghệ- Đại Học Quốc Gia Hà Nội đã nhiệt tình giảng dạy và giúp đỡ em trong thời gian em học tập tại trường.

Con xin gửi đến Bố Mẹ và gia đình tình thương yêu và lòng biết ơn. Bố Mẹ và gia đình luôn là nguồn động viên của con và là chỗ dựa vững chắc cho cuộc đời con. Tôi xin cảm ơn các bạn cùng lớp đã động viên và giúp đỡ tôi trong thời gian thực hiện luận văn này.

Vì thời gian có hạn , bài luận văn của em không tránh khỏi khiếm khuyết. Rất mong nhận được sự góp ý của thầy cô và các bạn.

Hà Nội 30/5/ 2005

Sinh viên

Mai Thị Lan Oanh.

TÓM TẮT NỘI DUNG

Trong kỹ thuật truyền hình, việc nén Video là một vấn đề qua trọng cho việc truyền tải các chương trình truyền hình. Và nó đã trở thành vấn đề nóng hổi khi truyền hình số ra đời, với tín hiệu Video sau khi số hoá (8 bit) có tốc độ bit bằng 216 Mbít/s, không thể truyền trên một kênh truyền hình thông thường nếu không được nén. Khi đó, chuẩn nén MPEG-2 với tốc độ mã hoá từ 4Mbít/s đến 30Mbít/s đã được đưa ra để thực hiện nhiệm vụ trên. Ngày nay, khi truyền hình trên mạng Internet được phát triển, chuẩn nén Video H.264/MPEG -4 Part 10 được đưa ra, với tốc độ mã hoá 1.5Mbít/s và có khả năng tương tác tới từng đối tượng , phù hợp với môi trường truyền tải trên mạng Internet hiện nay. Nghiên cứu các chuẩn nén MPEG (MPEG -1, MPEG -2, MPEG -4, MPEG -7) - được giới thiệu trong chương 1 và chương 2 và đặc biệt là chuẩn nén H.264/MPEG -4 Part 10 cho ứng dụng nén Video truyền trên mạng Internet- được giới thiệu trong chương 3. Chương 3 sẽ nói chi tiết cách mã hoá, giải mã video và các ưu việt của MPEG -4 Part 10 so với các chuẩn nén trước đó.

DANH SÁCH CÁC CHỮ VIẾT TẮT

ADSL	: Asymmetric Digital Subscriber Line.
CABAC	: Context-based Adaptive Binary Arithmetic Coding.
CAVLC	: Context-based Adaptive Variable Length Coding.
DCT	: Discrete Cosine Transform .
DVB	: DVB – Terrestrial
DVB-C	: DVB – Cable
DVB-S	: DVB – Satellite
DVB-T	: Digital Video Broadcasting
ES	: Elementary Stream.
FMO	: Flexible Macro-Block Order.
GOP	: Group of Picture.
IEC	: International Electrotechnical Commission (Part of the ISO)
ISO	: International Standard Organization
ITU	: International Telecommunication Union
MB	: Macro-Block .
MPEG	: Moving Picture Expert Group
NTSC	: National Television System Committee.
PAL	: Phase Alternative Line
RLC	: Run Length Coding
RVLC	: Reversible Variable Length Codes.
SAD	: Summation of Absolute Difference
SI	: Switching Intra Picture
SIF	: Source Intermediate Format
SP	: Switching Prediction Picture
VLC	: Variable Length Coding
VO	: Video Object

1. DANH SÁCH CÁC HÌNH

Hình 1: Cấu trúc dòng Bít MPEG Video	5
Hình 2: Cấu trúc ảnh MPEG	7
Hình 3: Nén MPEG	8
Hình 4: Giải nén MPEG	9
Hình 5: Quá trình biến đổi sang định dạng SIF và kích thước mảng các điểm ảnh.....	12
Hình 6: Tính toán giá trị cho các điểm ảnh trong bộ lọc thập phân	12
Hình 7: Chuẩn nén MPEG -2	16
Hình 8: Giải mã phân cấp theo SNR	17
Hình 9: Giải mã phân cấp theo không gian	18
Hình 10: Sự tổ hợp khung hình trong MPEG -4	22
Hình 11: Cấu trúc của bộ mã hoá và giải mã Video MPEG -4.....	23
Hình 12: Profile và Level trong MPEG -4	25
Hình 13: Phạm vi của MPEG -7.....	28
Hình 14: Truyền hình trực tuyến trên mạng.....	32
Hình 15: Cấu trúc của bộ mã hoá Video H264.....	36
Hình 16: Các phần mã hoá riêng của Profile trong H264.....	37
Hình 17: Sơ đồ mã hoá Video của H264/MPEG Part 10.....	39
Hình 18: Các mode trong MPEG-4.....	42
Hình 19: Phân chia Macro-Block cho bù chuyển động	43
Hình 20: Bù chuyển động nhiều Frame – ngoài Vector chuyển động, các tham số tham chiếu ảnh (Δ) cũng được truyền đi.....	45
Hình 21: Ảnh nội suy B (dự đoán hai chiều)	46
Hình 22: chuyển đổi sử dụng ảnh SP.....	49
Hình 13: Sơ đồ khối của CABAC	55
Hình 24: Tác dụng của bộ lọc tách khối đối với ảnh được nén nhiều	56
Hình 25: Sơ đồ giải mã Video H264/MPEG -4 Part 10	57
Hình 26: Ví dụ về Mã hoá chiều dài biến đổi ngược	59

2. DANH SÁCH CÁC BẢNG BIỂU

Bảng 1: <i>Tính chất các định dạng ảnh SIF cơ bản:</i>	13
Bảng 2: <i>Tham số theo tiêu chuẩn MPEG -1:</i>	15
Bảng 3 : <i>Bảng thông số chính Profiles và Levels của tín hiệu chuẩn MPEG -2</i>	20

MỤC LỤC

LỜI CẢM ƠN.....	iii
TÓM TẮT NỘI DUNG.....	iv
DANH SÁCH CÁC CHỮ VIẾT TẮT	v
1. DANH SÁCH CÁC HÌNH.....	vi
2. DANH SÁCH CÁC BẢNG BIỂU	vi
LỜI MỞ ĐẦU	1
Chương I: KHÁI QUÁT VỀ NỀN TÍN HIỆU VIDEO	3
VÀ CHUẨN NÉN MPEG	3
1.1. Mục đích nén Video	3
1.2. Chuẩn nén MPEG	4
1.2.1. Khái quát về nén MPEG	4
1.2.2. Cấu trúc dòng bit MPEG video	5
1.2.3. Các loại ảnh trong chuẩn MPEG :	7
1.2.4. Nguyên lý nén MPEG	8
1.2.5. Nguyên lý giải nén MPEG	9
Chương II: CÁC CHUẨN NÉN MPEG.....	11
2.1 Chuẩn nén MPEG-1	11
2.1.1 Giới thiệu khái quát	11
2.1.2 Định dạng trung gian SIF (Source Intermediate Format).	11
2.1.3 Cấu trúc dòng bit và các tham số của MPEG-1.	14
2.2 Chuẩn nén MPEG-2	16
2.2.1 Giới thiệu về MPEG-2	16
2.2.2 Mã hoá và giải mã video	16
2.2.3 Profiles và Levels	18
2.2.4 MPEG -2 với phát sóng và sản xuất chương trình	21
2.3 Chuẩn nén MPEG-4	21
2.3.1 Khái quát về MPEG-4	21
2.3.2 Công nghệ mã hoá và giải mã video trong MPEG-4	22
2.3.3 Các Profiles và Levels trong chuẩn MPEG-4	24
2.4 Tiêu chuẩn MPEG-7.....	26
2.4.1 Giới thiệu về chuẩn MPEG-7	26
2.4.2 Đối tượng (Objectives) và cách miêu tả dữ liệu của MPEG -7.....	27
2.4.3 Phạm vi ứng dụng của tiêu chuẩn MPEG-7.....	28
Chương III: CHUẨN NÉN VIDEO MPEG-4 VÀ ỨNG DỤNG TRONG TRUYỀN HÌNH TRÊN MẠNG INTERNET.	29
3.1 Giới thiệu tổng quan về truyền hình trên Internet	29
3.2 Lựa chọn H.264/ MPEG -4 part 10 cho truyền hình trên mạng Internet.....	33
3.2.1 Giới thiệu chung về H.264 /MPEG-4 part 10	33
3.2.2 Tính kế thừa của chuẩn nén H.264/MPEG- 4 part 10.....	35
3.3. Tiêu chuẩn H.264/MPEG - 4 Part 10	35
3.3.1. Lớp trừu tượng mạng NAL (Network Abstraction Layer).....	35
3.3.2. Các Profile và các Level.....	36
3.3.3. Kỹ thuật mã hoá video	39

3.3.3.2. Các ảnh và bù chuyển động dùng trong H264/MPEG Part 10.....	40
3.3.3.3. Xác định Vector chuyển động (Motion Estimation)	49
3.3.3.4. Nén video.....	51
3.3.3.5. Bộ lọc tách khối.....	55
3.3.4. Kỹ thuật giải mã video.....	57
3.3.4.1. Bù chuyển động.....	57
3.3.4.2. Khôi phục lỗi (Error Resiliency)	57
3.3.5 So sánh hiệu quả mã hoá của H264/MPEG Part 10 với các tiêu chuẩn trước đó.....	59
KẾT LUẬN	61
TÀI LIỆU THAM KHẢO	63

LỜI MỞ ĐẦU

Truyền hình Việt Nam đã trải qua nhiều giai đoạn phát triển, từ truyền hình đen trắng, truyền hình màu và hiện nay truyền hình số đang phát triển mạnh mẽ trên tất cả mọi lĩnh vực: truyền hình cáp, truyền hình số mặt đất DVB-T, truyền hình số qua vệ tinh. Khi Internet phát triển mạnh và trở thành hệ thống có quy mô toàn cầu, trở nên phổ cập rất nhanh trong mọi lĩnh vực, bằng việc kết nối các chương trình hình với hệ thống viễn thông – Internet, một công nghệ truyền hình mới ra đời đó là truyền hình Internet.

Thực tế hiện nay, có rất nhiều hãng ở Việt Nam và trên thế giới đang cung cấp dịch vụ truyền hình trực tuyến (Online Television), khán giả chỉ cần truy cập vào địa chỉ Web- Site của nhà cung cấp dịch vụ là có thể xem trực tiếp các chương trình truyền hình theo thời thực real-time(còn gọi là phương thức Dowload and Play)hay có thể tải File các chương trình truyền hình về máy tính cá nhân (gọi là phương thức Dowload Stream-File).

Nói về kỹ thuật truyền hình thì có rất nhiều kỹ thuật như: kỹ thuật ghi hình, kỹ thuật dựng hình, các kỹ thuật nén Video, kỹ thuật truyền tải... trong khuôn khổ bài luận văn này, em xin được tìm hiểu kỹ thuật nén Video. Vấn đề nén Video trong truyền hình không phải là một vấn đề mới mẻ. Với mỗi công nghệ truyền hình mới ra đời, sẽ có một công nghệ nén Video phù hợp. Nén Video từ những năm 1950 được thực hiện bằng công nghệ tương tự với tỷ số nén thấp. Ngày nay công nghệ nén đạt được hiệu quả cao hơn nhờ chuyển đổi tín hiệu Video từ tương tự sang số. Với đề tài “Các chuẩn nén và ứng dụng truyền Video trên mạng Internet”, mục đích của bài khoá luận của em là tìm hiểu một số các chuẩn nén MPEG ứng dụng nén video đã được sử dụng, đặc biệt là chuẩn nén H264/MPEG Part 10 . Nội dung của bài gồm 3 chương:

Chương 1: Lý do phải nén tín hiệu Video và nén MPEG (Moving Picture Expert Group) là nhóm chuyên gia về hình ảnh, với nhiệm vụ xây dựng tiêu chuẩn cho tín hiệu Audio và Video số. Trong việc nén video, MPEG đã đạt được một tỷ số nén tốt hơn so với các chuẩn nén trước đó như JPEG, M-JPEG, DV...

Chương 2: Giới thiệu khái quát về các chuẩn nén mà nhóm MPEG đã xây dựng để nén video và lưu trữ. Trong đó, MPEG -1 với mục đích là mã hoá Video và âm thanh kèm theo trong các môi trường lưu trữ như đĩa CD-ROM, đĩa quang... với tốc độ bit là 1.5 Mbit/s; MPEG -2 có kế thừa các tiêu chuẩn của MPEG -1 và mục đích nhằm hỗ trợ việc truyền Video số tốc độ bit trong khoảng 4 – 30 Mbít/s; sau đó khi truyền hình Internet ra đời thì chuẩn nén tương ứng là MPEG -4 với nhiệm vụ nhằm

phát triển các chuẩn xử lý, mã hoá và hiển thị ảnh động, audio và các tổ hợp của chúng. Còn MPEG -7 là một chuẩn dùng để mô tả nội dung Multimedia, chứ không phải là một chuẩn dùng để nén và mã hoá audio hay ảnh động như các chuẩn trước đó.

Chương 3: là nội dung chính của bài khoá luận. Em đã nghiên cứu và tìm hiểu kỹ thuật mã hoá Video/ ảnh động sử dụng chuẩn nén H.264/MPEG -4 Part 10 ứng dụng cho truyền hình trên mạng Internet . H264/MPEG Part 10 có nhiều ưu việt trong việc nén Video so với chuẩn MPEG-2 – đã rất thành công trong việc nén video trong truyền hình kỹ thuật số đã ra đời trước đó.

Phần cuối là phần kết luận, là phần tổng kết lại những gì mà em đã làm được trong bài khoá luận này. Đồng thời, Em cũng nêu lên một vài nhận định của mình về hướng phát triển tiếp theo của đề tài.

Chương I: KHÁI QUÁT VỀ NÉN TÍN HIỆU VIDEO VÀ CHUẨN NÉN MPEG

1.1. Mục đích nén Video

Tín hiệu video sau khi được số hoá 8 bit có tốc độ 216 Mb/s. Để có thể truyền trong một kênh truyền hình thông thường, tín hiệu video số cần phải được nén trong khi vẫn phải đảm bảo chất lượng hình ảnh.

Nén video trong những năm 1950 được thực hiện bằng công nghệ tương tự với tỷ số nén thấp. Ngày nay công nghệ nén đã đạt được những thành tựu cao hơn bằng việc chuyển đổi tín hiệu video từ tương tự sang số. Công nghệ nén số (Digital Compressed) đòi hỏi năng lực tính toán nhanh. Song ngày nay với sự phát triển của công nghệ thông tin, điều này không còn trở ngại.

Như chúng ta biết tín hiệu video có dải phổ từ 0 – 6 MHz, tuy nhiên trong nhiều trường hợp năng lượng phổ chủ yếu tập trung ở miền tần số thấp và chỉ có rất ít thông tin chứa đựng ở miền tần số cao.

Đối với tín hiệu video số, số lượng bit được sử dụng để truyền tải thông tin đối với mỗi miền tần số khác nhau, có nghĩa là: miền tần số thấp, nơi chứa đựng nhiều thông tin, được sử dụng số lượng bit lớn hơn và miền tần số cao, nơi chứa đựng ít thông tin, được sử dụng số lượng bit ít hơn. Tổng số bit cần thiết để truyền tải thông tin về hình ảnh sẽ giảm một cách đáng kể và dòng dữ liệu được “nén” mà chất lượng hình ảnh vẫn đảm bảo. Thực chất của kỹ thuật “nén video số” là loại bỏ đi các thông tin dư thừa. Các thông tin dư thừa trong nén video số thường là:

- + Độ dư thừa không gian giữa các pixel;
- + Độ dư thừa thời gian do các ảnh liên tiếp nhau;
- + Độ dư thừa do các thành phần màu biểu diễn từng pixel có độ tương quan cao;
- + Độ dư thừa thống kê do các kí hiệu xuất hiện trong dòng bit với xác suất xuất hiện không đều nhau;
- + Độ dư thừa tâm lý thị giác (các thông tin nằm ngoài khả năng cảm nhận của mắt).vv...

Như vậy, mục đích của nén tín hiệu video là :

- Giảm tốc độ dòng bit của tín hiệu gốc xuống một giá trị nhất định đủ để có thể tái tạo ảnh khi giải nén;
- Giảm dung lượng dữ liệu trong lưu trữ cũng như giảm băng thông cần thiết;

- Tiết kiệm chi phí trong lưu trữ và truyền dẫn dữ liệu trong khi vẫn duy trì chất lượng ảnh ở mức chấp nhận được.

Với nguyên nhân và mục đích của việc nén tín hiệu được trình bày như ở trên, ngày nay có nhiều các chuẩn nén đã ra đời như: JPEG, M-JPEG, MPEG, DV... Trong đó chuẩn nén MPEG được sử dụng nhiều trong nén video trong truyền hình với thành công của chuẩn nén video MPEG-2 trong truyền hình số và chuẩn nén MPEG-4 trong truyền hình trên mạng Internet.

1.2. Chuẩn nén MPEG

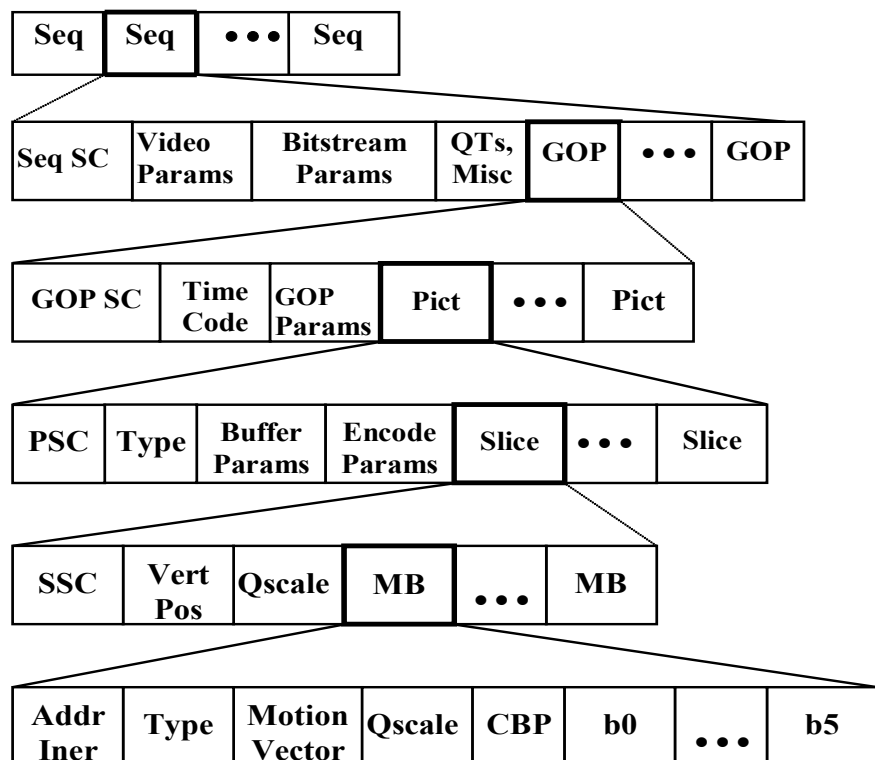
1.2.1. Khái quát về nén MPEG

- MPEG (Moving Picture Expert Group) là nhóm chuyên gia về hình ảnh, được thành lập từ tháng 2 năm 1988 với nhiệm vụ xây dựng tiêu chuẩn cho tín hiệu Audio và Video số. Ngày nay, MPEG đã trở thành một kỹ thuật nén Audio và Video phổ biến nhất vì nó không chỉ là một tiêu chuẩn riêng biệt mà tùy thuộc vào yêu cầu cụ thể của từng thiết bị sẽ có một tiêu chuẩn thích hợp nhưng vẫn trên cùng một nguyên lý thống nhất.
- Tiêu chuẩn đầu tiên được nhóm MPEG đưa ra là MPEG-1, mục tiêu của MPEG-1 là mã hoá tín hiệu Audio-Video với tốc độ khoảng 1.5Mb/s và lưu trữ trong đĩa CD với chất lượng tương đương VHS.
- Tiêu chuẩn thứ 2 : MPEG-2 được ra đời vào năm 1990, không như MPEG-1 chỉ nhằm lưu trữ hình ảnh động vào đĩa với dung lượng bit thấp. MPEG-2 với “công cụ” mã hoá khác nhau đã được phát triển. Các công cụ đó gọi là “Profiles” được tiêu chuẩn hoá và có thể sử dụng để phục vụ nhiều mục đích khác nhau.
- Tiêu chuẩn tiếp theo mà MPEG đưa ra là MPEG-4, được đưa ra vào tháng 10 năm 1998, đã tạo ra một phương thức thiết lập và tương tác mới với truyền thông nghe nhìn trên mạng Internet, tạo ra một phương thức sản xuất, cung cấp và tiêu thụ mới các nội dung video trên cơ sở nội dung và hướng đối tượng (content/object-based).
- MPEG-7: là một chuẩn dùng để mô tả các nội dung Multimedia, chứ không phải là một chuẩn cho nén và mã hoá audio/ảnh động như MPEG-1, MPEG-2 hay MPEG-4. MPEG-7 sử dụng ngôn ngữ đánh dấu mở rộng XML(Extensible Markup Language) để lưu trữ các siêu dữ liệu Metadata, đính kèm timecode để gắn thẻ cho các sự kiện, hay đồng bộ các dữ liệu. MPEG-7 bao gồm 3 bộ chuẩn sau:

- + Bộ các sơ đồ đặc tả (Description Schemes) và các đặc tả (Description).
- + Ngôn ngữ xác định DDL (Description Definition Language) để định nghĩa các sơ đồ đặc tả.
- + Sơ đồ mã hoá quá trình đặc tả.

Việc kết hợp MPEG-4 và MPEG-7 sẽ tạo ra các giải pháp lý tưởng cho các dịch vụ Streaming Media, các hệ thống lưu trữ và sản xuất Streaming Media trong thời gian tới.

1.2.2. Cấu trúc dòng bit MPEG video



Hình 1: Cấu trúc dòng bit MPEG Video

Trong đó :

❖ **Sequence:** Thông tin về chuỗi bit

- Video Params: chứa thông tin về chiều cao, bề rộng, tỷ lệ khuôn hình các phần tử ảnh.
- Bitstream Params: Thông tin về tốc độ bit và các thông số khác.
- QTs: có 2 loại QTs :
 - Nén trong ảnh (ảnh I – I Frame)
 - Nén liên ảnh (ảnh P – P Frame)

Đoạn video và thông tin đầu đoạn tạo thành một dòng bit được mã hoá gọi là dòng cơ sở (Elementary Stream).

❖ **GOP (Group Of Picture):** Thông tin về nhóm ảnh

Là tổ hợp của nhiều các khung I, P, B. Cấu trúc nhóm ảnh gồm 2 tham số là: m và n (tham số m xác định số khung hình B và P xuất hiện giữa 2 khung hình I gần nhau nhất, tham số n xác định số khung B xuất hiện giữa 2 khung P). Mỗi một nhóm ảnh bắt đầu bằng một khung I và xác định điểm bắt đầu để tìm kiếm và biên tập.

Các tham số của đoạn mào đầu của GOP:

- Time code: mã định thời, xác định giờ, phút, giây, ảnh.
- GOP Params: miêu tả cấu trúc GOP.

❖ **Pict :** thông tin về ảnh, các tham số trong phần mào đầu của Pict:

- Type: Cho phép bộ giải mã xác định ảnh được mã hoá là ảnh I, P hay B.
- Buffer Params: thông tin về Buffer(chỉ thứ tự truyền khung để bộ giải mã có thể sắp xếp các loại ảnh theo một thứ tự đúng).
- Encode Params: chứa thông tin về đồng bộ, độ phân giải và phạm vi của vector chuyển động.

❖ **Slice:** Mảng bao gồm một vài cấu trúc khối kề nhau.

Kích thước lớn nhất của mảng có thể bao gồm toàn bộ bức ảnh và kích thước nhỏ nhất của mảng là một cấu trúc khối. Các thông số của đoạn mào đầu của Slice gồm:

- Vert PoS: Slice bắt đầu từ dòng nào.
- Qscale: Thông tin về bảng lượng tử.

Kích thước thông tin đầu của mảng được xác định bằng số lỗi cho phép xuất hiện trong mảng đối với một ứng dụng nhất định, do đó bộ giải mã có thể bỏ qua các mảng có nhiều lỗi và xác định bằng tính hiệu quả của phương pháp nén ảnh. Do đó hệ số cân bằng lượng tử có thể được điều chỉnh thường xuyên với việc sử dụng các mảng có kích thước nhỏ hơn. Hệ số DCT tham chiếu dùng trong mã hóa DPCM sẽ được so chuẩn tại mỗi mảng.

❖ **MB (Macroblock)**

Một cấu trúc khối là một nhóm các khối tương ứng với lượng thông tin chứa đựng trong kích thước 16x16 điểm trên bức ảnh.

Các tham số của đoạn mào đầu của nhóm MB:

- Addr Iner: Số lượng MB được bỏ qua.
- Type : Loại vector chuyển động dùng cho Macroblock.

- Qscale : Bảng lượng tử dùng cho Macroblock.
- Coded Block Pattern (CBP): chỉ rõ Block nào được mã hoá.

1.2.3. Các loại ảnh trong chuẩn MPEG :

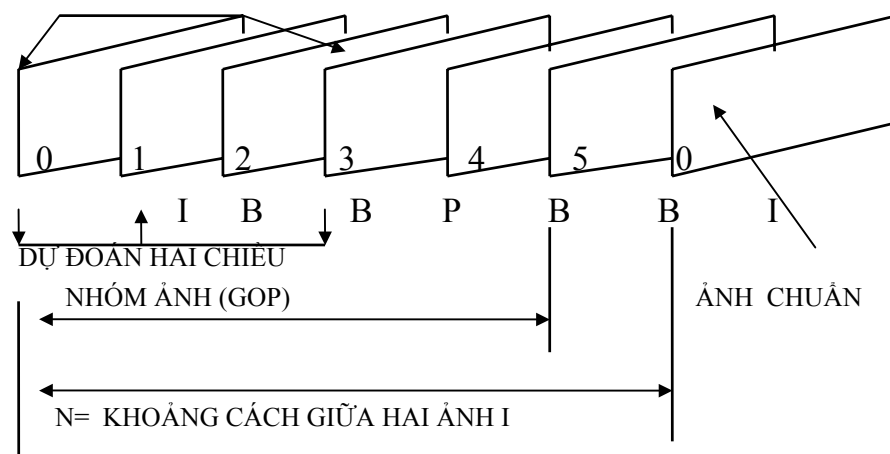
Trong nén MPEG người ta sử dụng 3 loại ảnh sau:

- **Ảnh I (Intra Pictures):** được mã hóa mà không có sự so sánh tham khảo các ảnh khác, dùng trong nén trong ảnh. Chúng chứa tất cả các thông tin cần thiết để tái tạo lại ảnh sau giải mã, nên tỷ lệ nén các ảnh I tương đối thấp. Vì vậy, ảnh I là điểm nút quan trọng phục vụ việc truy cập vào một đoạn Video.

- **Ảnh P (Predicted Pictures):** được mã hoá từ ảnh I, ảnh P trước đó, nhờ sử dụng các thuật toán dự đoán bù chuyển động. Các ảnh P có thể được sử dụng như là cơ sở dữ liệu cho việc dự đoán ảnh tiếp theo. Tuy nhiên do hạn chế của kỹ thuật bù chuyển động, số ảnh P giữa hai ảnh I không thể quá lớn. Tỷ lệ nén của các ảnh P tương đối lớn so với tỷ lệ nén các ảnh I.

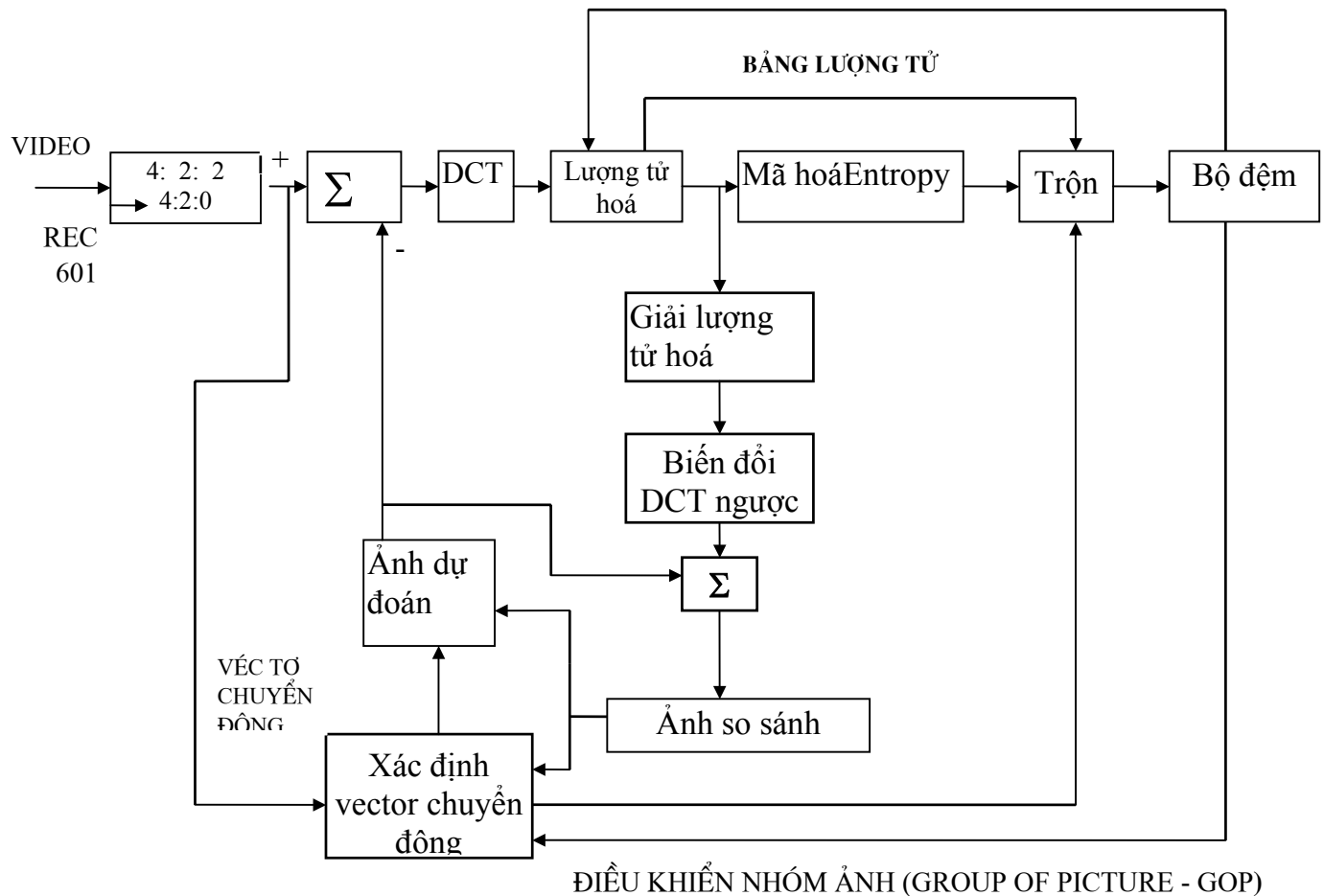
- **Ảnh B (Bidirectionally Predicted Pictures):** được mã hoá bởi phép nội suy giữa các ảnh I và P ở trước và sau đó. Vì không được sử dụng để mã hoá các ảnh tiếp theo, ảnh B không phải là nguồn gốc sinh ra các lỗi ảnh trong quá trình mã hoá. Các ảnh B cho tỷ lệ nén cao nhất.

M = KHOẢNG CÁCH GIỮA
HAI ẢNH SO SÁNH



Hình 2: Cấu trúc ảnh MPEG

1.2.4. Nguyên lý nén MPEG



Hình 3: Nén MPEG

* Cơ sở của công nghệ nén video MPEG là sự kết hợp giữa nén trong ảnh (Intra-Frame Compression) và công nghệ nén liên ảnh (Inter-Frame Compression). Trong đó:

- **Nén trong ảnh (Intra -Frame Compression):** là loại nén nhằm giảm bớt thông tin dư thừa trong miền không gian. Nén trong ảnh sử dụng cả hai quá trình có tổn hao và không có tổn hao để giảm bớt dữ liệu trong ảnh. Quá trình này không sử dụng thông tin của các ảnh trước và sau ảnh đang xét.
- **Nén liên ảnh (Inter-Frame Compression):** Trong tín hiệu video có chứa thông tin dư thừa trong miền thời gian. Nghĩa là với một chuỗi liên tục các ảnh, lượng thông tin chứa đựng trong mỗi ảnh thay đổi rất ít từ ảnh này sang ảnh khác. Tính toán sự dịch chuyển vị trí của nội dung ảnh là một phần rất quan trọng trong kỹ thuật nén liên ảnh. Trong thuật nén MPEG, quá trình xác định Vector chuyển động được thực hiện bằng cách chia hình ảnh thành

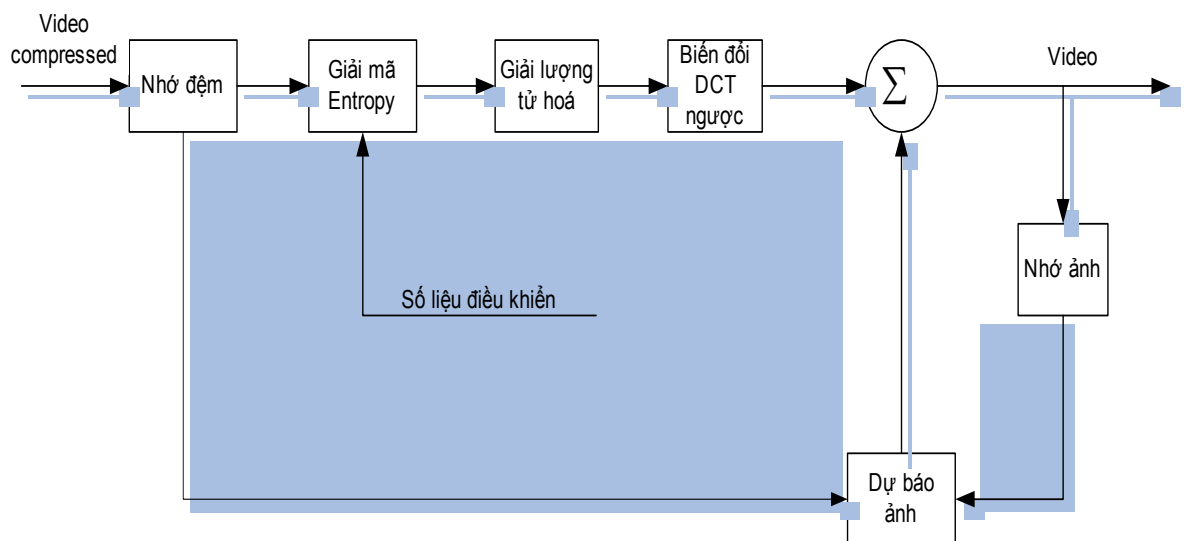
các Macro-Block, mỗi Macro-Block có 16 x 16 phần tử ảnh (tương đương với 4 Block, mỗi Block có 8 x 8 phần tử ảnh). Để xác định chiều chuyển động, người ta tìm kiếm vị trí của Macro-Block trong ảnh tiếp theo, kết quả của sự tìm kiếm sẽ cho ta Vector chuyển động của Macro-Block.

*** Nguyên lý nén MPEG :**

Dạng thức đầu vào là Rec- 601 4:2:2 hoặc 4:2:0. Ảnh hiện tại được so sánh với ảnh trước tạo ra ảnh khác biệt. Ảnh này sau đó lại được nén trong ảnh qua các bước : biến đổi DCT, lượng tử hóa, mã hoá. Dữ liệu của ảnh khác biệt và vector chuyển động (được xác định như trên) mang thông tin về ảnh sau nén liên ảnh được đưa đến bộ đệm ở đầu ra.

Tốc độ bit của tín hiệu video được nén không cố định, phụ thuộc vào nội dung ảnh đang xét (ví dụ một phần nén ít hơn hoặc nhiều hơn), nhưng tại đầu ra bộ mã hoá dòng bit phải cố định để xác định tốc độ cho dung lượng kênh truyền.

1.2.5. Nguyên lý giải nén MPEG



Hình 4: Giải nén MPEG

*** Nguyên lý giải nén MPEG :**

- Đầu tiên là giải mã Entropy, sau đó tách dữ liệu ảnh (hệ số biến đổi DCT) ra khỏi các vector chuyển động. Dữ liệu ảnh sẽ được giải lượng tử hoá và biến đổi DCT ngược.

- Nếu ảnh là ảnh loại I bắt đầu ở mỗi nhóm ảnh trong chuỗi, ở đầu ra sẽ nhận được ảnh hoàn chỉnh bằng cách trên (vì ảnh loại I chỉ là nén trong ảnh, không có bù chuyển động, không dùng dữ liệu của ảnh khác). Nó được lưu trữ trong bộ nhớ ảnh và được dùng để giải mã các ảnh tiếp theo.
- Nếu ảnh là ảnh loại P thì cũng thực hiện giải lượng tử hóa và biến đổi DCT ngược kết hợp với việc sử dụng vector chuyển động và lưu vào bộ nhớ ảnh sớm hơn. Trên cơ sở đó xác định được dự đoán ảnh đang xét. Ta nhận được ảnh ra sau khi cộng dự đoán ảnh (ảnh dự đoán) và kết quả biến đổi DCT ngược. Ảnh này cũng được lưu vào bộ nhớ để có thể sử dụng như là chuẩn khi giải mã các ảnh tiếp theo.

Chương II:

CÁC CHUẨN NÉN MPEG.

2.1 Chuẩn nén MPEG-1

2.1.1 Giới thiệu khái quát

MPEG -1 được hình thành vào năm 1988, là tiêu chuẩn của nhóm chuyên gia về hình ảnh MPEG ở trong giai đoạn đầu tiên (tương ứng với tiêu chuẩn ISO/IEC 11172 của ITU). Mục đích của MPEG -1 là nghiên cứu một tiêu chuẩn mã hoá video và âm thanh kèm theo trong các môi trường lưu trữ như: CD-ROM, đĩa quang ... Tốc độ mã hoá trong khoảng 1.5 Mb/s.

Chuẩn nén MPEG -1 bao gồm 4 phần :

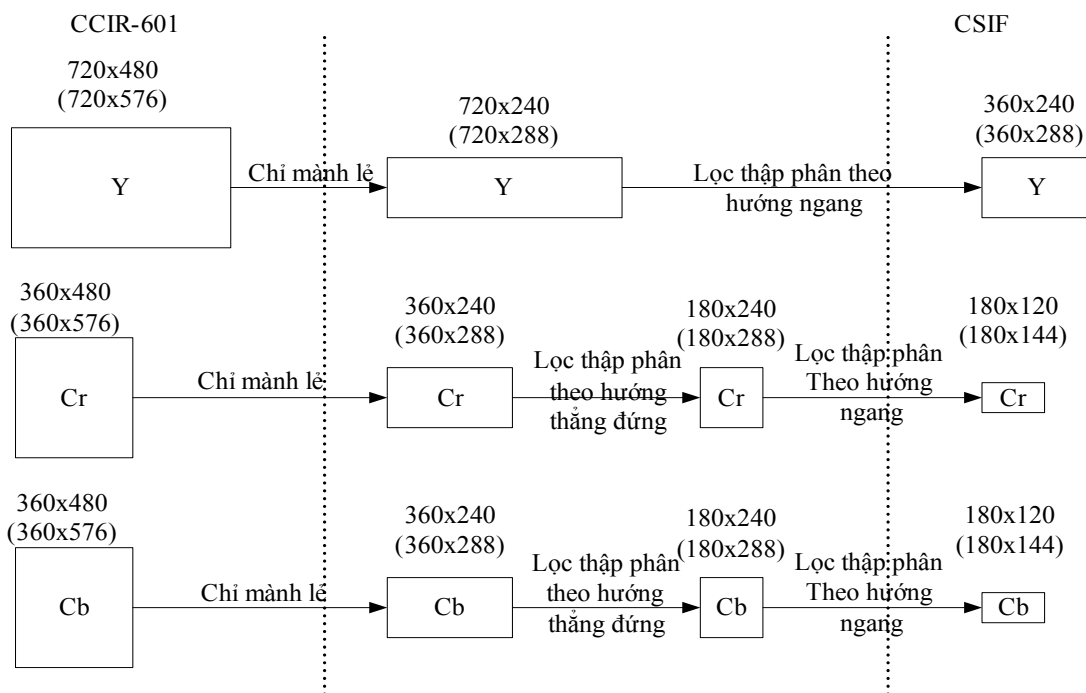
- Các hệ thống : ISO/IEC 11172 -1
- Video : ISO/IEC 11172 -2
- Audio : ISO/IEC 11172 -3
- Hệ thống kiểm tra: ISO/IEC 11172 -4

Trong các phần trên ta nghiên cứu một vài thông số trong phần Video (ISO/IEC 11172 -2).

2.1.2 Định dạng trung gian SIF (Source Intermediate Format).

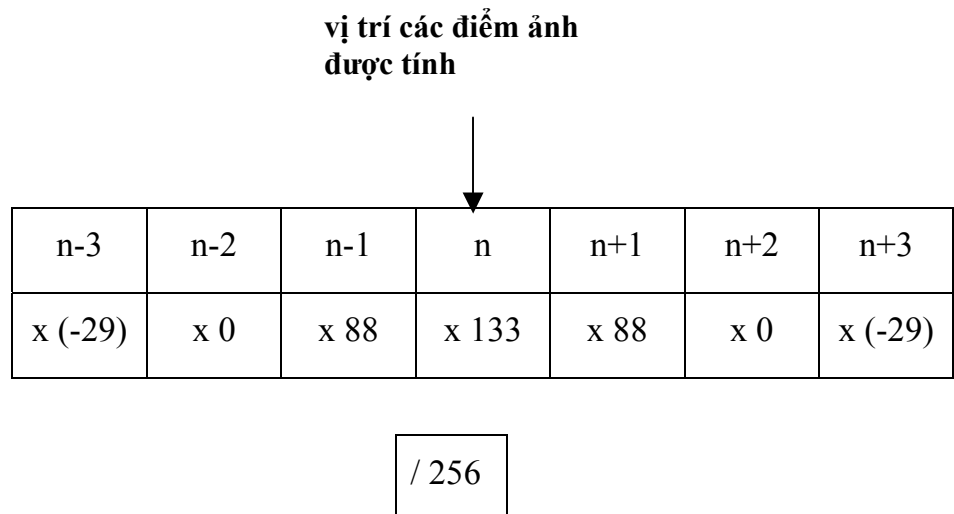
Khi truyền hình màu phát triển, xuất hiện nhiều hệ truyền hình khác nhau như: NTSC (ở Mỹ), PAL (châu Âu)... với các hệ thống quét truyền hình khác nhau như hệ 525/60 và 625/50. Do đó cần có một định dạng chung cho nguồn tín hiệu dùng cho bộ mã hoá nén số liệu và các xác định riêng khác nhau phù hợp mỗi hệ thống. Định dạng trung gian cho nguồn tín hiệu được gọi là SIF (Source Intermediate Format).

Trong định dạng chung này, tần số lấy mẫu được lấy theo xác định của chuẩn CCIR-601. Do đó số mẫu trên một dòng tích cực của cả hai tiêu chuẩn 525/60 và 625/50 là bằng nhau. Quá trình chuyển đổi từ định dạng theo tiêu chuẩn CCIR-601 sang định dạng SIF được thực hiện bằng cách sử dụng một bộ lọc thập phân theo chiều ngang cho các mảnh lẻ của tín hiệu Y, một bộ lọc theo chiều ngang và một bộ lọc theo chiều thẳng đứng cho các mảnh số lẻ cho các tín hiệu Cr và Cb như sau:



Hình 5: Quá trình biến đổi sang định dạng SIF và kích thước mảng các điểm ảnh

- Quá trình tính toán giá trị cho các điểm ảnh trong lọc thập phân như sau:



Hình 6: Tính toán giá trị cho các điểm ảnh trong bộ lọc thập phân

Giá trị điểm ảnh tại vị trí n được tính bằng: tích số của các giá trị điểm ảnh từ (n-3) đến (n+3) với các hệ số của bộ lọc tương ứng tại vị trí này trên hình vẽ trên.

Tổng các kết quả này được chia cho 256 và thu được giá trị điểm ảnh ở vị trí n . Phép tính tiếp theo được thực hiện cho điểm ảnh ở vị trí $n+2$.

Một quá trình lọc tương tự áp dụng theo chiều thẳng đứng tạo ra giá trị thập phân tín hiệu Cr và Cb theo chiều này.

+ Số các điểm ảnh trên một dòng tích cực được giảm từ 360 xuống 352 để thu được một bội số của 16 nhằm tổ chức thuận lợi các cấu trúc khối điểm ảnh 16×16 với cấu trúc 4:2:0. Ảnh tích cực đã được làm giảm xuống (352×240) được gọi là vùng điểm ảnh xác định (có ý nghĩa) cho SIF. Định dạng SIF phối hợp với cấu trúc lấy mẫu 4:2:0 làm giảm thêm số liệu tín hiệu màu. Các thông số cho định dạng SIF đối với các tiêu chuẩn truyền hình được cho trong bảng sau:

Bảng 1: Tính chất các định dạng ảnh SIF cơ bản:

	CCIR-601	SIF-525	CCIR-601	SIF-625
	525	4:2:0	625	4:2:0
Số điểm ảnh trên dòng tích cực				
Điểm chói Y	720	352	720	352
Điểm màu Cr, Cb	360	176	360	176
Tần số lấy mẫu (MHz)				
Điểm chói Y	13.5	6.75	13.5	6.75
Điểm màu Cr, Cb	6.75	3.38	6.75	3.38
Số dòng tích cực				
Điểm chói Y	480	240	576	288
Điểm màu Cr, Cb	480	120	576	144
Tần số trung bình				
	30	30	25	25
Cỡ ảnh				
	4:3	4:3	4:3	4:3

Trước khi mã hoá MPEG -1, tốc độ số liệu ban đầu cần được giảm nhờ bộ biến đổi 4:2:2 sang định dạng SIF, từ dòng số có tốc độ 166 Mb/s (98 bit biểu diễn mẫu) xuống dòng số có tốc độ 31.5 Mb/s. Do đó quá trình giải mã sẽ cần một bộ chuyển đổi ngược lại quá trình này.

2.1.3 Cấu trúc dòng bit và các tham số của MPEG-1.

Cấu trúc dòng bit của MPEG -1 cũng tương tự như cấu trúc dòng bit của MPEG, nó được phân thành các lớp như:

- **Sequence (chuỗi ảnh) :** gồm nhiều nhóm ảnh GOP, có chức năng là dòng bit video.
- **GOP (Group of Picture) :** gồm từ 1- n ảnh bắt đầu bằng ảnh I, có chức năng là đơn vị truy xuất.
- **Picture I, P, B:** gồm nhiều Slice, chức năng là đơn vị mã hoá cơ bản.
- **Slice :** gồm nhiều các Macro Block, là đơn vị để tái đồng bộ phục hồi lỗi.
- **Macro-Block :** gồm 16 x 16 pixel, là đơn vị bù chuyển động.
- **Block :** gồm 8 x 8 pixel, là đơn vị tính DCT.

Và một vài thông số của chuẩn nén MPEG -1 là:

Bảng 2: Tham số theo tiêu chuẩn MPEG -1:

Tham số	Đặc điểm
Tín hiệu mã hoá	Y và Cr,Cb
Cấu trúc lấy mẫu	4:2:0
Kích thước ảnh tối đa(điểm ảnh x điểm ảnh)	4095 x 4095
Biểu diễn mẫu	8 bit
Độ chính xác của quá trình lượng tử hoá và biến đổi DCT	9 bit
Phương pháp lượng tử hoá hệ số DCT	DPCM tuyến tính
Cấu trúc khối trong quá trình lượng tử hoá thích nghi	16 x 16 bit
Độ chính xác cực đại của hệ số DC	8 bit
Biến đổi RLC	Mã Huffman
Bảng VLC	Không thể truyền tải
Hệ số cân bằng các khối	Có thể biến đổi
Bù chuyển động	Trong khung hình và giữa các khung hình
Quét	Tuần tự
Độ chính xác dự đoán chuyển động	$\frac{1}{2}$ điểm ảnh
Tốc độ khi nén	1.85 Mb/s cho nén tham số 100 Mb/s cho dòng đầy đủ tham số

2.2 Chuẩn nén MPEG-2

2.2.1 Giới thiệu về MPEG-2

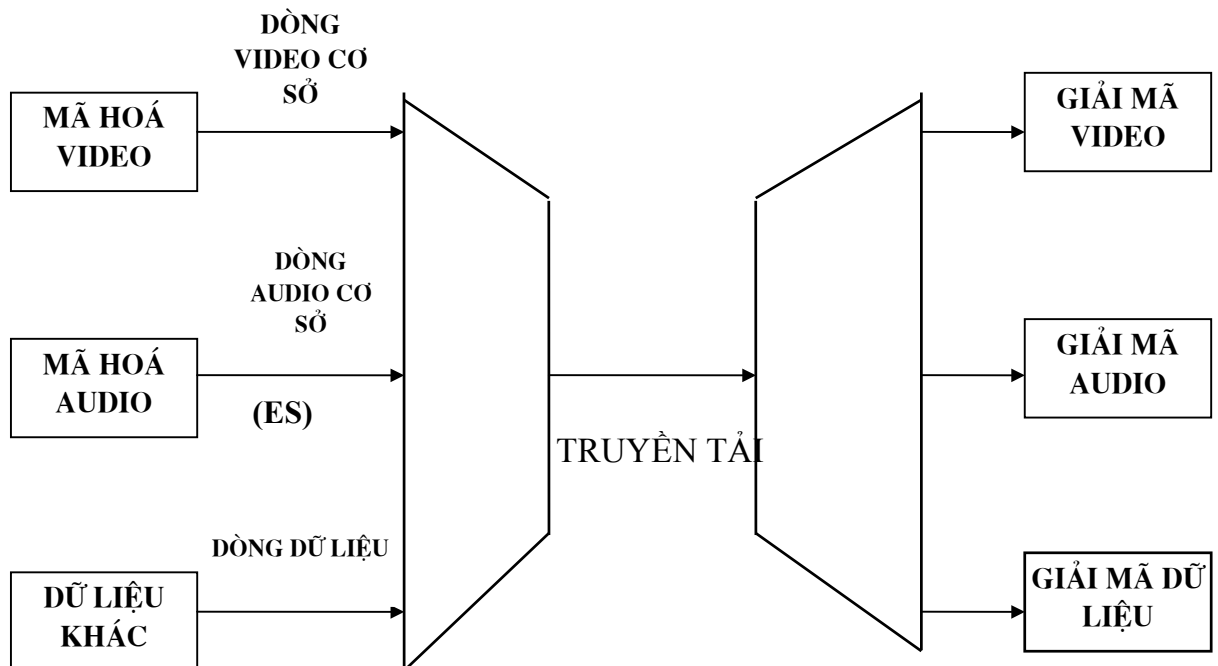
Chuẩn nén MPEG -2 là chuẩn nén phát triển tiếp sau MPEG -1, có kế thừa tất cả các tiêu chuẩn của MPEG -1 và mục đích là nhằm hỗ trợ việc truyền video số, tốc độ bit lớn hơn 4 Mb/s, bao gồm các ứng dụng DSM (phương tiện lưu trữ số), Các hệ thống truyền hình hiện tại (NTSC, PAL, SECAM), cáp, thu lượm tin tức điện tử, truyền hình trực tiếp từ vệ tinh, truyền hình mở rộng (EDTV), truyền hình độ phân giải cao (HDTV)...

Chuẩn MPEG -2 bao gồm 4 phần chính:

- Các hệ thống : ISO/IEC 13818 -1.
- Video : ISO/IEC 13818 -2
- Audio : ISO/IEC 13818 -3
- Các hệ thống kiểm tra: ISO/IEC 13818 -4.

2.2.2 Mã hoá và giải mã video

- Mã hoá MPEG -2:

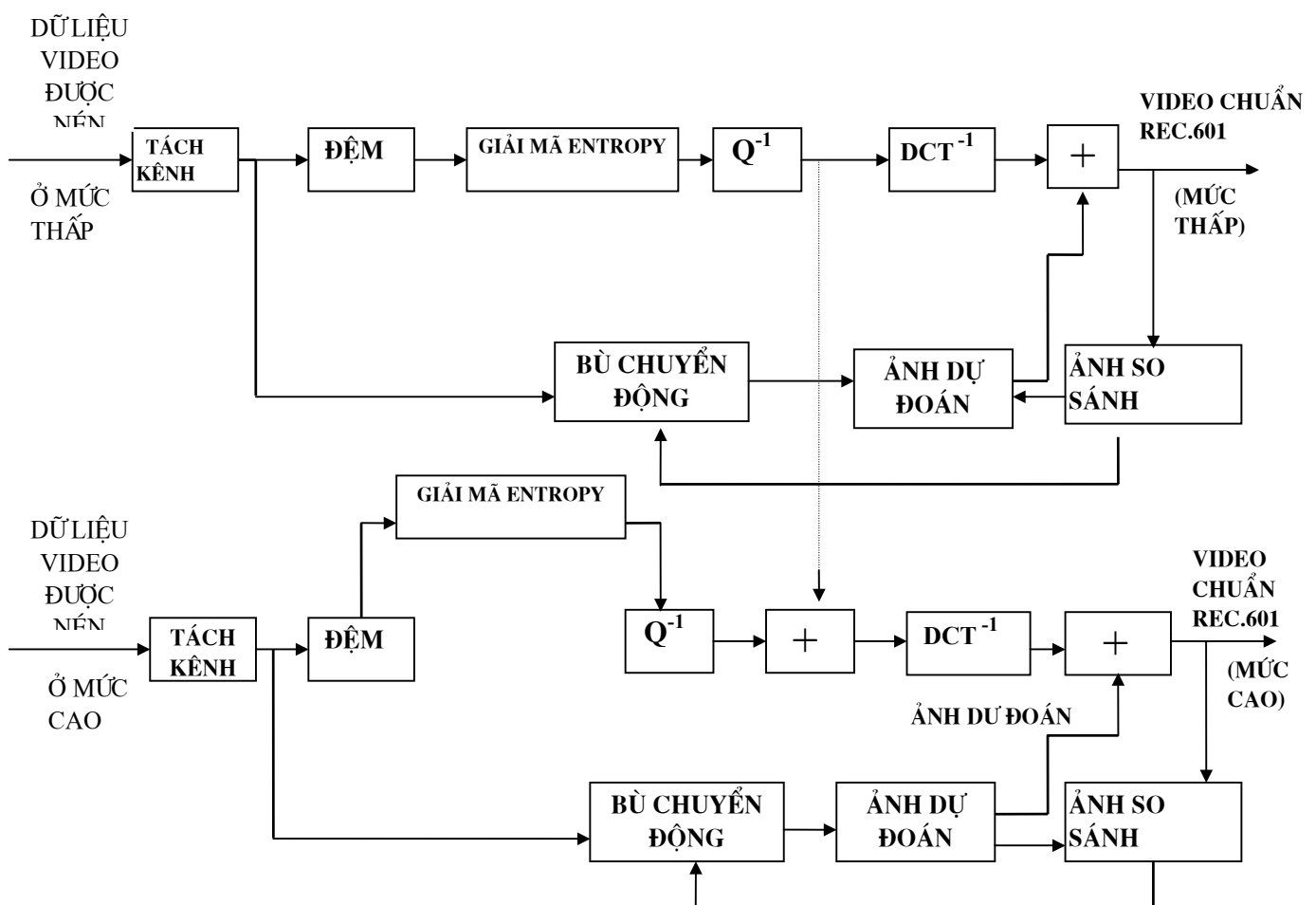


Hình 7: Chuẩn nén MPEG-2

Tín hiệu Video và Audio được nén (theo như nguyên lý nén MPEG) và tạo thành các dòng dữ liệu cơ sở ES (Elementary Stream). Dòng ES được sử dụng để tạo nên dòng dữ liệu cơ sở được đóng gói PES (Packetized Elementary Stream). Dòng PES lại được tiếp tục đóng gói tạo thành dòng truyền tải TS (Transport Stream).

- **Giải mã MPEG -2:**

MPEG -2 Coder và Decoder không nhất thiết phải có cùng cấp chất lượng. Tính phân cấp cho phép các bộ giải mã MPEG đơn giản, rẻ tiền, có khả năng giải mã một phần của toàn bộ dòng bit và như vậy có khả năng tạo được hình ảnh tuy chất lượng có thấp hơn các bộ giải mã toàn bộ dòng bit.

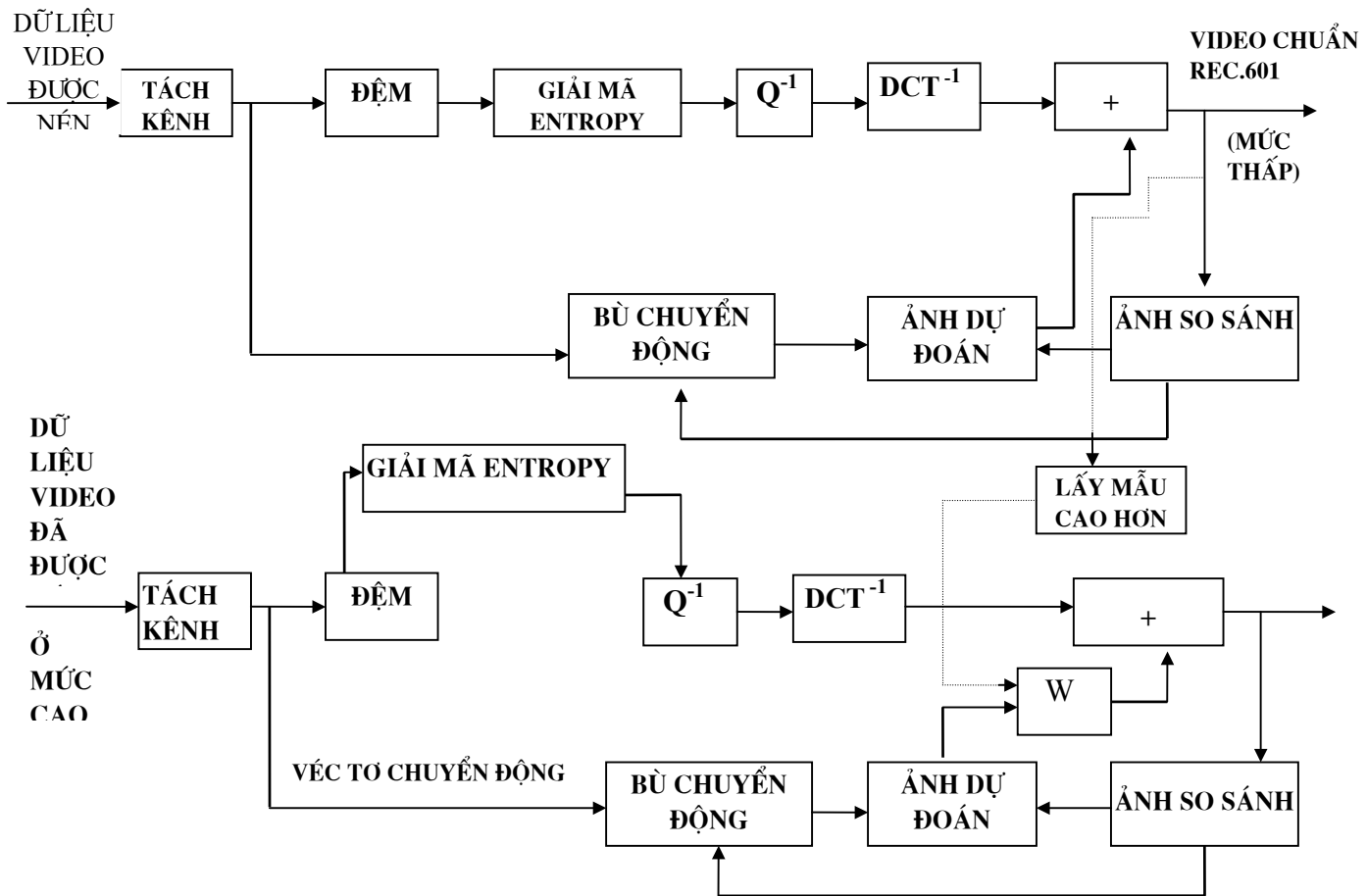


Hình 8: Giải mã phân cấp theo SNR

Tiêu chuẩn MPEG cho phép phân cấp theo tỷ số tín hiệu trên tạp âm (SNR) và theo độ phân giải. Trong đó :

- Tính phân cấp theo tỷ số tín hiệu trên tạp âm (SNR Scalability) có nghĩa là chất lượng hình ảnh và tỷ số tín hiệu trên tạp âm có tính thỏa hiệp. Một bộ giải mã có tốc độ bit thấp, có thể có đầy đủ độ phân giải nhưng tỷ số tín hiệu trên tạp âm (SNR) thấp hơn so với bộ giải mã có tốc độ bit cao.

- Tính phân cấp theo không gian (Spatial Scalability) có nghĩa là có sự thỏa hiệp đối độ phân giải. Một máy thu có tốc độ bit thấp cho hình ảnh có độ phân giải thấp hơn so với máy thu có khả năng giải mã toàn bộ dòng bit.



Hình 9:Giải mã phân cấp theo không gian

2.2.3 Profiles và Levels

Chuẩn MPEG -2 có 4 Levels (mức) và 5 Profiles (bộ công cụ). Trong đó:

- Profiles: Là khái niệm cho ta biết cấp chất lượng bộ công cụ nén được sử dụng chuẩn nén này. Ở đây có sự thỏa hiệp giữa tỷ số nén và giá thành bộ giải nén. Có 5 định nghĩa Profiles :

- **Simple Profiles (Profiles đơn giản):** Số bước nén thấp nhất, chỉ cho phép mã hoá các ảnh loại I và P. Do có tổn thất cao về tốc độ bit, nó không được sử dụng trong nén với kỹ thuật chuẩn đoán ảnh hai chiều (các ảnh B).
- **Main Profiles (Profiles chính):** Là sự thỏa hiệp tốt nhất giữa tổn hao tốc độ bit và chi phí, do nó sử dụng tất cả các ảnh I, P và B trong nén. Chất lượng tốt hơn Profiles đơn giản nhưng nó đòi hỏi phải sử dụng các thiết bị mã hoá và giải mã phức tạp hơn.
- **SNR Profiles Scalable (Profiles phân cấp theo SNR):** Tiêu chuẩn MPEG-2 cho phép phân cấp tỷ số tín hiệu trên tạp âm, có nghĩa là chất lượng hình ảnh và tỷ số tín hiệu trên tạp âm có tính thỏa hiệp. Chuỗi ảnh chia thành hai lớp phân biệt nhau về chất lượng. Các lớp thấp bao gồm ảnh có chất lượng cơ sở, ví dụ như chứa tín hiệu theo chuẩn 4:2:0. Các lớp cao bao gồm lớp hoàn thiện hơn so với lớp thấp hơn, như với tín hiệu video trong chuẩn 4:2:2. Có thể mã hoá kênh khác nhau cho các lớp riêng.
- **Spatially Scalable Profiles (phân cấp theo không gian):** Tính phân cấp theo không gian có nghĩa là có sự thỏa hiệp với độ phân giải. Chuỗi ảnh được chia ra thành hai lớp tương ứng với các độ phân giải khác nhau của ảnh. Lớp thấp hơn bao gồm ảnh có độ phân giải thấp như truyền hình tiêu chuẩn. Còn lớp cao hơn bao gồm ảnh có độ phân giải cao hơn như truyền hình độ phân giải cao (HDTV).
- **High Profiles (Profiles cao):** Cho phép cả hai loại thang mức được ứng dụng trong truyền hình HDTV với các định dạng 4:2:0 hay 4:2:2. Nó bao gồm toàn bộ các công cụ của Profiles trước cộng thêm khả năng mã hoá các tín hiệu màu khác nhau cùng một lúc.

Như vậy, giữa các Profiles nói trên có sự tương thích cao dần, nghĩa là các bộ giải mã của một Profiles cao cấp hơn có khả năng giải mã tất cả các Profiles ở cấp đó và cấp thấp hơn.

- **Levels :** Khái niệm Levels trong chuẩn MPEG-2 cho ta biết mức độ phân giải của ảnh, bao gồm từ định dạng trung gian cho nguồn tín hiệu SIF (Source Intermediate Format), định dạng cơ sở MPEG -1 (360 x 288 @ 25Hz hay 360 x 240 @ 30Hz), đến truyền hình số phân giải cao HDTV (hệ thống truyền hình với trên 1000 dòng quét). Theo quan điểm ứng dụng có 4 mức Levels trong MPEG -2 được mô tả như sau:

- **Low Levels (mức thấp):** phù hợp với độ phân giải SIF được sử dụng trong MPEG -1 (cho đến 360 x 288 pixel).
- **Main Levels (mức chính):** phù hợp với độ phân giải chuẩn 4:2:2 (tới 720 x 576 pixel).
- **High Levels 1440 (mức cao 1440):** nhằm vào truyền hình phân giải cao HDTV (độ phân giải tới 1440 x 1152 pixel).
- **High Levels (mức cao):** được tối ưu hoá đối với HDTV màn ảnh rộng (độ phân giải tới 1920 x 1152 pixel).

Trong thực tế ứng dụng, các nhà chế tạo đã chọn sẵn một số thoả hiệp giữa các mức Levels và các Profiles cho người sử dụng. Kết hợp 4 Levels và 5 Profiles ta được 20 tổ hợp khả năng và hiện nay đã có 11 khả năng được ứng dụng như trong Bảng 3. Trong các ô trong bảng 3 lần lượt từ trên xuống là: tỷ lệ lấy mẫu, điểm ảnh theo chiều ngang x điểm ảnh theo chiều dọc, vận tốc cao nhất của dòng dữ liệu sau khi nén, dòng cuối là các loại ảnh sử dụng để nén.

Bảng 3 :Bảng thông số chính Profiles và Levels của tín hiệu chuẩn MPEG -2

Profiles Levels	Simple (Đơn giản)	Main (Chính)	SNR (Phân cấp theo SNR)	Spatially Theo không gian	High (Cao)
Low (thấp)		4:2:0 352 x 288 4 Mb/s	4:2:0 352 x 288 4 Mb/s I, P, B		
Main (chính)	4:2:0 720 x 576 15 Mb/s I, P	4:2:0 720 x 576 15 Mb/s I, P, Block	4:2:0 720 x 576 15 Mb/s I, P, B		4:2:0 720 x 576 20 Mb/s I, P, B
High 1440 (Cao 1440)		4:2:0 1440x1152 60 Mb/s I, P, B		4:2:0 1440x1152 60 Mb/s I, P, B	4:2:2 ;4:2:0 1440x1152 80 Mb/s I, P, B
High (cao)		4:2:0 1920x1152 80 Mb/s I, P, B			4:2:0;4:2:2 1920x1152 100Mb/s I, P, B

2.2.4 MPEG -2 với phát sóng và sản xuất chương trình

Trong lĩnh vực phát sóng và sản xuất hậu kỳ, chuẩn nén MPEG -2 cuối cùng đã giải quyết được bị kịch đa dạng thức xưa nay.

Để đạt được khả năng này, cần nắm vững yêu cầu và đặc tính của từng ứng dụng khác nhau.

- Sản xuất tin thời sự:

Chất lượng ảnh đòi hỏi cao khi đi quay ngoại cảnh, nhưng tín hiệu không phải in đi in lại nhiều lần. Thiết bị nhỏ, gọn, xách tay và dùng ắc quy. Do vậy các thiết bị này có thể sử dụng tốc độ bit thấp 18Mb/s (đỡ tốn năng lượng, giảm giá thành). Cấu trúc GOP đơn giản, thuận tiện cho công đoạn dựng hình.

- Lưu trữ:

Cần chất lượng cao hoặc ít nhất bằng ảnh gốc. Công nghệ lưu trữ còn cần sử dụng tốc độ bit lớn và cấu trúc IB-GOP. Tốc độ lưu trữ là 30 Mb/s.

- Sản xuất hậu kỳ:

Sản xuất hậu kỳ đòi hỏi cả chất lượng và mức độ cao về tính năng dựng hình. Tốc độ 50 Mb/s sản xuất và cấu trúc I-GOP là sự lựa chọn thích hợp.

- Phân phối chương trình:

Phân phối chương trình đòi hỏi khả năng lưu trữ về chất lượng phát sóng quảng bá. Tốc độ 20 Mb/s (hoặc 15 Mb/s để cung cấp cho máy phát NTSC/PAL thậm chí thấp hơn cho dịch vụ phát trực tiếp từ vệ tinh DBS).

2.3 Chuẩn nén MPEG-4

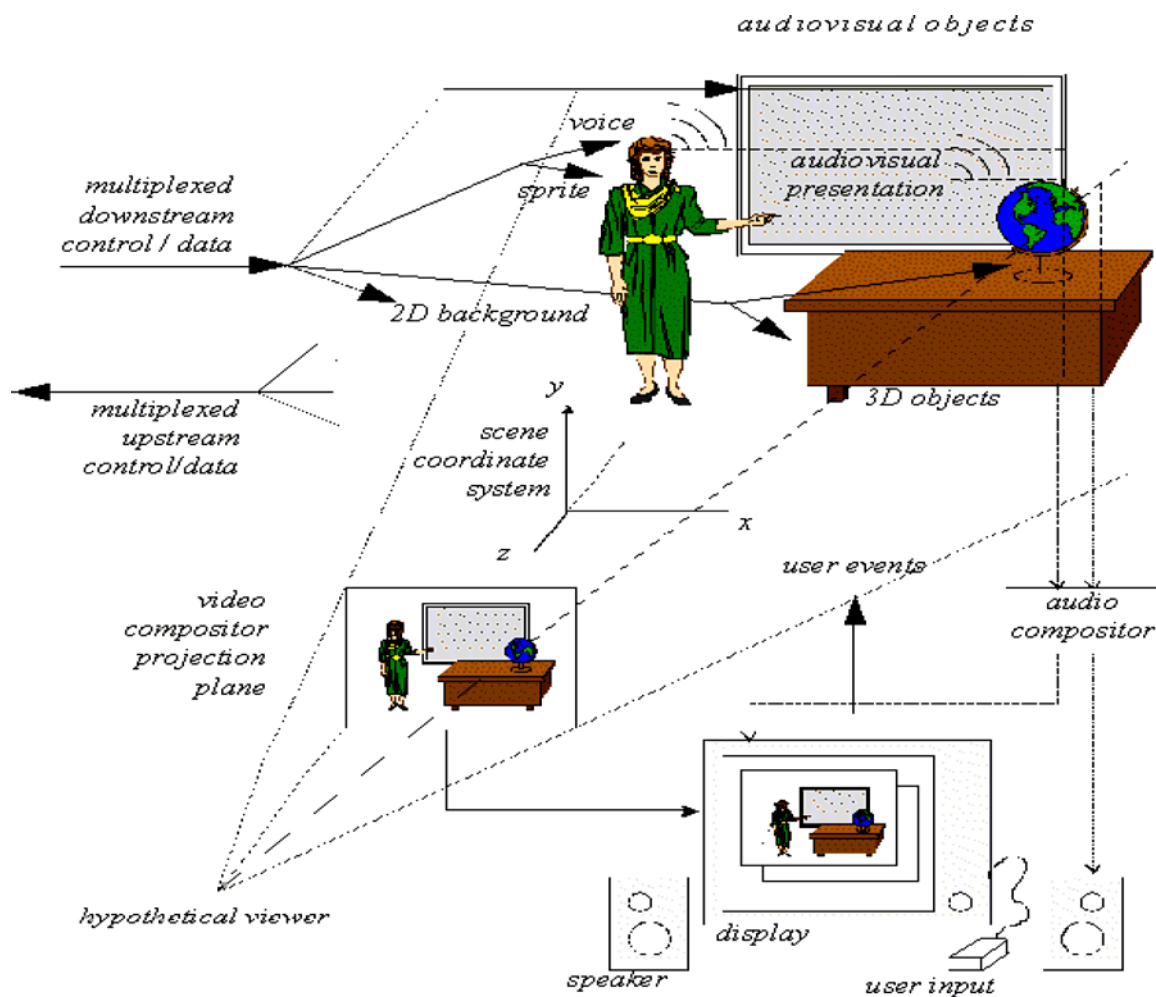
2.3.1 Khái quát về MPEG-4

Ngày nay, khi nhu cầu truyền phát các ứng dụng video và đa phương tiện mới trên hạ tầng kỹ thuật Internet đã làm nảy sinh các yêu cầu chức năng mới không có trong các chuẩn nén MPEG -1 và MPEG -2 hay các chuẩn nén trước đó. Tháng 10 năm 1998 với sự xuất hiện của chuẩn nén MPEG -4 đã tạo ra một phương thức thiết lập và tương tác mới với truyền thông nghe nhìn trên mạng Internet, tạo ra một phương thức sản xuất, cung cấp và tiêu thụ mới các nội dung video trên cơ sở nội dung và hướng đối tượng (content/object-based). Đây chính là một công nghệ trình diễn truyền

thông đa phương tiện phức hợp, có khả năng truyền thông tại các môi trường truyền thông đa phương tiện phức hợp, có khả năng truyền thông tại các môi trường : truyền hình số, đồ hoạ tương tác, World Wide Web. Vì vậy, nhiệm vụ của MPEG -4 là nhằm phát triển các chuẩn xử lý, mã hoá và hiển thị ảnh động, audio và các tổ hợp của chúng. MPEG -4 đang được triển khai bởi nhiều nhà vận hành mạng và dịch vụ trên thế giới với các dịch vụ mới đang được bổ xung để chiếm các lợi thế cấu trúc hạ tầng băng rộng đang phát triển.

2.3.2 Công nghệ mã hoá và giải mã video trong MPEG-4

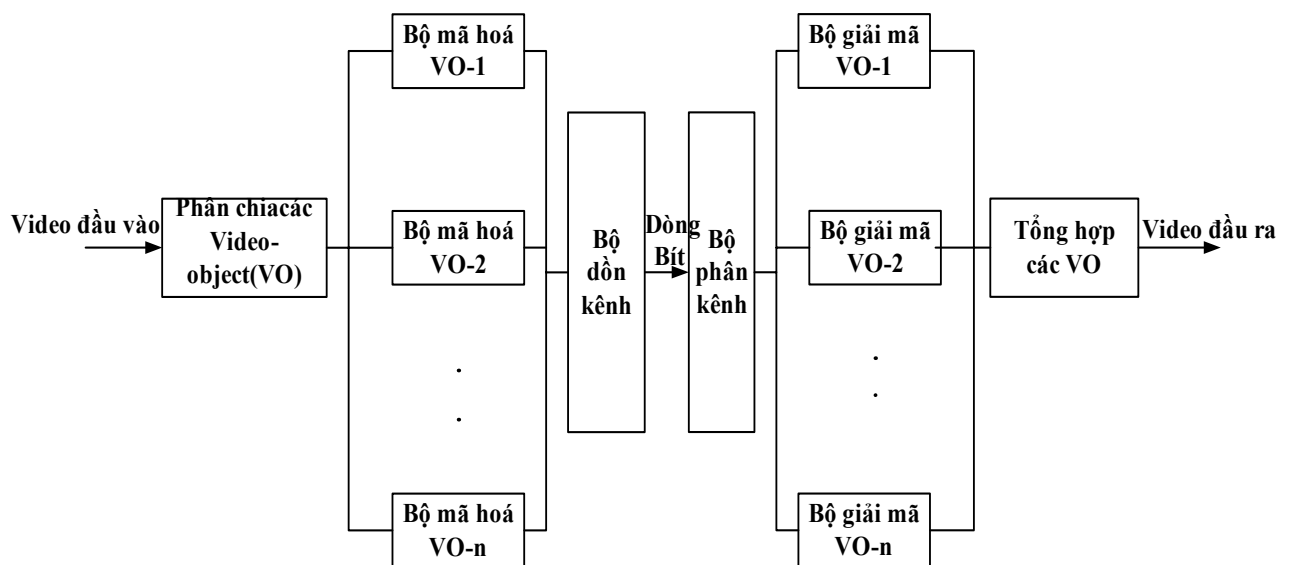
Chuẩn MPEG -4 là một chuẩn động, dễ thay đổi: với MPEG -4 các đối tượng khác nhau trong một khung hình có thể được mô tả, mã hoá và truyền đi một cách riêng biệt đến bộ giải mã trong các dòng cơ bản ES (Elementary Stream) khác nhau.



Hình 10: Sự tổ hợp khung hình trong MPEG -4

Cũng nhờ xác định, tách và xử lý riêng các đối tượng (như nhạc nền, âm thanh xa gần, đồ vật, đối tượng ảnh video như con người hay động vật, nền khung hình...) nên người sử dụng có thể loại bỏ riêng từng đối tượng khỏi khuôn hình. Sự tổ hợp lại thành khung hình chỉ được thực hiện sau khi giải mã các đối tượng đó.

Trên hình 10 là ví dụ về sự tổ hợp khuôn hình MPEG -4. Trong hình có nhiều đối tượng như: bàn, quả cầu, bảng đen, người hướng dẫn và audio được đặt vào một hệ thống toạ độ không gian 3 chiều (3-D) đối với vị trí người xem giả định.



Hình 11: Cấu trúc của bộ mã hoá và giải mã Video MPEG - 4

Trên hình 11 là cấu trúc của bộ mã hoá và giải mã Video MPEG - 4, các thiết bị mã hoá và giải mã video đều áp dụng sơ đồ mã hoá như nhau cho mỗi đối tượng video (video-object) riêng biệt. Ví dụ khi ta mã hoá và tổng hợp một khung hình, nhiều đối tượng đầu vào như: ô tô, nhà, người... được tách ra khỏi video đầu vào. Mỗi đối tượng video sau đó được mã hoá riêng rẽ bởi bộ mã hoá đối tượng video VO (video object) và được truyền đi trên mạng. Tại vị trí thu, những đối tượng này được giải mã riêng rẽ nhờ bộ giải mã VO decoder và gửi đến bộ tổng hợp Compositor. Vì vậy người sử dụng có thể thực hiện các hoạt động tương tác riêng với từng đối tượng (thay đổi tỷ lệ, di chuyển, kết nối, loại bỏ, bổ xung các đối tượng...) ngay tại vị trí giải mã hay mã hoá. Ngoài ra, người dùng có thể download các đối tượng khác từ thư viện cơ sở dữ liệu (có

sẵn trên thiết bị hay từ xa thông qua mạng LAN, WAN hay Internet) để chèn thêm vào hay thay thế các đối tượng có trong khuôn hình gốc.

Các bộ phận chức năng chính trong các thiết bị MPEG -4 bao gồm:

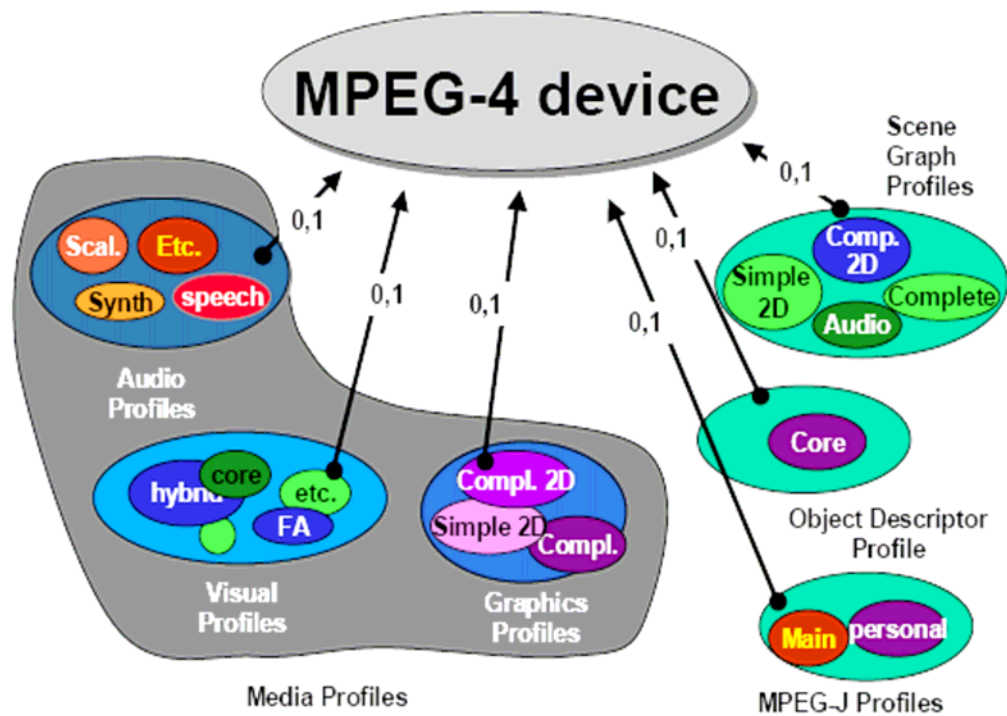
- Bộ mã hoá hình dạng ngoài Shape coder dùng để nén đoạn thông tin, giúp xác định khu vực và đường viền bao quanh đối tượng trong khung hình scene.
- Bộ dự đoán và tổng hợp động để giảm thông tin dư thừa theo thời gian.
- Bộ kết cấu mặt ngoài Texture coder dùng để xử lý dữ liệu bên trong và các dữ liệu còn lại sau khi đã bù chuyển động.

Trong MPEG -4, tất cả các đối tượng có thể có thể được mã hoá với sơ đồ mã hoá tối ưu riêng của nó: video được mã hoá theo kiểu video, text được mã hoá theo kiểu text, các đồ hoạ được mã hoá theo kiểu đồ hoạ... thay vì việc xử lý tất cả các phần tử ảnh pixels như là mã hoá ảnh động. Do các quá trình mã hoá đã được tối ưu hoá cho từng loại dữ liệu thích hợp, nên chuẩn MPEG -4 sẽ cho phép mã hoá với hiệu quả cao tín hiệu ảnh video, audio và cả các nội dung tổng hợp như các bộ mặt và cơ thể hoạt hình.

2.3.3 Các Profiles và Levels trong chuẩn MPEG-4

Chuẩn nén MPEG -4 bao gồm nhiều tính năng khác nhau và không phải bất kỳ ứng dụng nào cũng đòi hỏi tất cả các tính năng của MPEG -4. Để sử dụng công cụ MPEG -4 một cách hiệu quả nhất, mỗi thiết bị chuẩn MPEG -4 chỉ được trang bị một số tính năng phù hợp với một phạm vi ứng dụng nhất định và để tạo điều kiện cho người sử dụng lựa chọn công cụ MPEG -4, các thiết bị MPEG -4 chia thành các nhóm công cụ gọi là các Profiles, mỗi nhóm Profiles chỉ chứa một vài tính năng cần thiết của chuẩn mã hoá thích hợp cho một phạm vi ứng dụng nào đó. Mỗi Profiles lại chỉ có một số các mức Levels khác nhau, thể hiện mức độ phức tạp xử lý tính toán dữ liệu của công cụ đó (thông qua việc xác định rõ tốc độ bit, con số tối đa của các đối tượng trong khung hình, độ phức tạp của quá trình giải mã audio...)

Hình sau nói về một số bộ công cụ của thiết bị MPEG -4. Có nhiều Profiles như: Media Profiles , Scene Graph Profiles, MPEG-J Profiles ...



Hình 12: Profiles và Levels trong MPEG -4

Ta xét Media Profiles có: Audio Profiles, Visual Profiles, Graphics Profiles. Trong đó Visual Profiles gồm các Profiles như sau:

- **Simple Profiles (Profiles đơn giản) :** chỉ cho phép các loại vật đơn giản (tỷ lệ chiều cao/chiều rộng là tùy ý và tỷ lệ بیت thấp) và được tạo ra với các ứng dụng có độ phức tạp không cao trong việc nhận ra vật. Ứng dụng Simple Profiles cho di động, truyền các hình ảnh Video có độ phức tạp không cao trên Internet hay các thiết bị camera ghi các hình ảnh động như: đĩa hay chip nhớ. Có 3 mức (Levels) được sử dụng cho Simple Profiles với tốc độ بیت vào khoảng 64 - 384 kb/s.
- **Simple Scalable Profile (Profiles phân mức đơn giản) :** có thể phân chia việc mã hoá trong cùng một môi trường hoạt động như việc dự đoán trước cho Simple và có 2 mức được định nghĩa ở Profiles này.
- **Core Profiles (Profiles lõi) :** dùng cho các dịch vụ tương tác chất lượng cao, phối hợp chất lượng tốt với việc hạn chế độ phức tạp và hỗ trợ tùy ý hình dạng của vật. Vì vậy các dịch vụ Broadcast di động được hỗ trợ cho Profiles này. Tốc độ بیت tối đa là 384kb/s cho mức 1 (Levels 1) và 2Mb/s cho mức 2 (Level 2).

- **Main Profile (Profiles chính)** : được tạo ra với các dịch vụ quảng bá broadcast, địa chỉ hoá tiên tiến. Nó phối hợp chất lượng cao nhất với tính linh hoạt của các vật có hình dạng tự do mà sử dụng đến thang màu nâu để mã hoá. Mức cao nhất cho phép 32 vật (có nhiều mức) và tốc độ bit lớn nhất là 38 Mb/s.
- **N-bit Profiles** : hữu ích cho những vùng mà sử dụng những người mô tả nhiệt độ (thermal imagers), như các ứng dụng kiểm soát. Vì vậy các ứng dụng trong y học muốn sử dụng để làm tăng độ sâu của điểm ảnh đưa ra một miền ảnh động lớn trong tín hiệu màu và tín hiệu chói.
- **Scalable Texture Profiles (Profiles phân mức kết cấu)** : là giá trị trung bình cho các ứng dụng âm thanh đồ hoạ. Nó được yêu cầu bởi các công ty muốn xây dựng các thiết bị di động, mà có kết nối đồng thời âm thanh và hình ảnh hiển thị, đồ hoạ có dạng thức BIFS cơ sở trong các thiết bị đầu cuối rất đơn giản.
- **Simple Face Profiles (Profiles bề mặt đơn giản)** : chỉ cho các vật có bề mặt đơn giản (các hoạt ảnh). Phụ thuộc vào từng mức, song trong một khuôn hình tối đa có 4 bề mặt có thể xuất hiện. Tốc độ bit còn thấp, 32 kb/s là đủ cho việc truyền tối đa 4 bề mặt.
- **Hybird Profiles (Profiles lai ghép)** : cho phép liên kết cả các vật tự nhiên và nhân tạo trong cùng một khuôn hình mà vẫn giữ độ phức tạp hợp lý. Với các vật tự nhiên, nó so sánh với Core Profiles, còn với các vật nhân tạo, nó cộng gộp các hoạt ảnh lưới, kết cấu phân lớp và các bề mặt hoạt ảnh - thiết lập nên bộ công cụ mà tạo ra sự lai ghép thú vị nội dung của các vật tự nhiên và nhân tạo. Profiles này được sử dụng cho những nơi có vật thật trong thế giới nhân tạo hay ngược lại kết hợp các vật thật vào trong môi trường tự nhiên.
- **Basic Animated Texture Profiles (Profiles kết cấu hoạt ảnh cơ bản)** : cho phép hoạt ảnh của các hình ảnh tĩnh và các mặt hoạt ảnh. Dùng cho các nội dung mà được tạo ra với tốc độ bit rất thấp.

2.4 Tiêu chuẩn MPEG-7

2.4.1 Giới thiệu về chuẩn MPEG-7

MPEG (ISO/IEC SC29/WG11) cho ra đời một tiêu chuẩn mới là MPEG-7 với mục đích để mô tả các nội dung Multimedia, chứ nó không phải là chuẩn nén và mã hoá audio, video, ảnh động như các chuẩn nén đã ra đời trước đó.

2.4.2 Đối tượng (Objectives) và cách miêu tả dữ liệu của MPEG -7

Hiện nay trên thế giới thông tin nghe nhìn được số hoá trở nên phổ biến và được nhiều người ưa chuộng. Trước khi sử dụng một thông tin nào đó thì cần phải thiết lập chúng. Đồng thời việc tìm dữ liệu cần thiết ngày càng khó khăn hơn. Ngày nay xuất hiện một giải pháp cho việc tìm kiếm các thông tin nguyên bản trên trang Web mở rộng toàn cầu (World Wide Web), nhiều các nghiên cứu đã ra đời và có nhiều người quan tâm. Tuy nhiên, trong lĩnh vực nghe nhìn, việc nhận dạng các thông tin là cực khó. Các kênh truyền số phát triển mạnh mẽ gây khó khăn cho việc lựa chọn ra một kênh (TV, radio) hoàn hảo. Khi MPEG -7 ra đời đã đưa ra một giải pháp cho vấn đề trên, với tên gọi “Giao diện mô tả truyền thông đa phương tiện – Multimedia Content Description Interface”- và được gọi tắt là MPEG -7.

Một mục đích khác của MPEG -7 là đưa ra một tiêu chuẩn cho bộ mô tả có thể được dùng để mô tả nhiều loại thông tin truyền thông khác nhau. Sự mô tả này được kết hợp với chính nội của nó, cho phép tìm kiếm dữ liệu cho quyền lợi của người dùng một cách nhanh chóng và hiệu quả.

MPEG -7 không dựa vào cách mã hoá của dữ liệu được xây dựng trong MPEG-4, tiêu chuẩn cung cấp cách mã hoá dữ liệu nghe nhìn của vật thể có mối quan hệ nào đó trong thời gian và không gian (trên màn hình tín hiệu video và trong phòng với tín hiệu audio). Khi dùng kiểu mã hoá MPEG -4, nó có thể gán sự mô tả các yếu tố bên trong cảnh vật, cho phép truy cập riêng rẽ các yếu tố này. Nhưng với MPEG -7, nó sẽ thừa nhận sự khác nhau ở chính trong cách miêu tả chính nó và đưa ra các mức khác nhau đối với mỗi sự phân biệt đó.

Vì các đặc trưng miêu tả phải mang đầy đủ ý nghĩa trong phạm vi của các ứng dụng, chúng sẽ khác với mỗi lĩnh vực người dùng khác nhau và ứng dụng khác nhau. Trong cách miêu tả của dữ liệu thì những dữ liệu giống nhau được dùng những kiểu khác nhau của các đặc trưng, phù hợp với mỗi phạm vi ứng dụng. Ví dụ về dữ liệu thị giác: một mức dưới của khái niệm trừu tượng sẽ được mô tả bởi hình dáng, kích thước, màu sắc, sự cử động và vị trí (nơi ở trên màn hình có thể tìm thấy vật). Còn đối với dữ liệu audio: giọng nói, tâm trạng, nhịp độ, thay đổi nhịp độ, vị trí ở trong không gian âm thanh. Ở mức cao sẽ gửi nghĩa của thông tin như : “đây là cảnh của một con chó nâu đang sủa ở bên trái và một quả bóng xanh đang rơi xuống ở bên phải, cùng với tiếng động đang phát ra từ những chiếc ô tô bên dưới”. Tất cả những sự mô tả này là tiến trình mã hoá theo cách có hiệu quả (hiệu quả cho việc tìm kiếm đó). Mức trung gian của khái niệm trừu tượng cũng có thể tồn tại.

Các mức trên được liên hệ với các đặc trưng: một vài đặc trưng mức thấp được sao chép tự động đầy đủ, đặc trưng mức cao lại cần thiết hơn nhiều cho hoạt động của con người.

2.4.3 Phạm vi ứng dụng của tiêu chuẩn MPEG-7

MPEG-7 sẽ gửi ứng dụng có thể được lưu trữ (trực tuyến – on line hay ngoại tuyến – off line) và khai thác ở trong hai môi trường thời gian thực (real time) và không thực (non real time). Một môi trường thời gian thực có nghĩa là thông tin được kết hợp với nội dung trừ khi nó bị lưu giữ lại.



Feature Extraction : trích một đặc trưng
MPEG -7 Description: miêu tả MPEG -7
Search Engine: Dụng cụ tìm kiếm

Hình 13: Phạm vi của MPEG-7

Trong đó, phân trích dẫn một đặc trưng (Feature Extraction) thì bao gồm các công việc như: phân tích nội dung, trích dẫn đặc trưng, công cụ chú giải, công cụ tác chủ; phần miêu tả MPEG -7 bao gồm: lược đồ miêu tả, bộ miêu tả, ngôn ngữ, khái niệm MPEG-7 còn phần công cụ tìm kiếm gồm: việc tìm kiếm và lọc lựa, việc phân loại, các thao tác, việc tóm tắt, việc chỉ thị. Việc khai thác đầy đủ khả năng sự mô tả dữ liệu của MPEG -7 và việc tự động khai thác như vậy thì rất hữu ích, tuy nhiên không phải lúc nào cũng làm được. Do đó chuẩn hoá này được quy định để cho phép sử dụng chúng một cách hiệu quả trong phạm vi cho phép.

ChươngIII: CHUẨN NÉN VIDEO MPEG-4 VÀ ỨNG DỤNG TRONG TRUYỀN HÌNH TRÊN MẠNG INTERNET.

3.1 Giới thiệu tổng quan về truyền hình trên Internet

Truyền hình Việt Nam đã trải qua nhiều giai đoạn phát triển, bắt đầu từ truyền hình đen trắng, truyền hình màu, sau đó đến công nghệ truyền hình số việc chuyển đổi từ công nghệ truyền hình tương tự sang công nghệ truyền hình số là một bước tiến bộ vượt bậc của truyền hình Việt Nam. Hiện nay truyền hình số đang được phát triển mạnh mẽ trên tất cả các lĩnh vực: truyền hình cáp, truyền hình số mặt đất DVB-T, truyền hình số qua vệ tinh DTH ...

Khi Internet được phát triển thành hệ thống có qui mô toàn cầu, nó trở nên phổ cập rất nhanh trong mọi lĩnh vực. Ngoài vấn đề cung cấp dữ liệu dưới dạng văn bản, ảnh đồ họa ... Internet cũng hỗ trợ truyền dẫn audio và video, cơ sở của các dịch vụ Multimedia trên qui mô toàn cầu. Với sự ra đời của truyền hình kỹ thuật số, các thành tựu mới trong lĩnh vực truyền dẫn Internet tốc độ cao - đặc biệt là công nghệ đường thuê bao số không đối xứng ADSL (Asymmetric Digital Subscriber Line) sử dụng kỹ thuật truyền dẫn băng rộng Broadband đã cho phép nâng tốc độ tải của các kết nối Internet lên 500 Kbit/s, 2Mbit/s, 5Mbit/s và cao hơn nữa – và chuẩn nén MPEG -4 cho phép giảm bớt tốc độ truyền tải cần thiết của truyền hình số từ 4 – 6 Mbit/s hiện nay xuống 1.5 – 2 Mbit/s là những tiền đề kỹ thuật tạo điều kiện cho sự hội tụ giữa truyền hình và Internet. Truyền hình Internet – Internet TV là cụm từ dùng để chỉ dịch vụ cung cấp các chương trình truyền hình như thời sự, văn nghệ, thể thao, phim truyện ... qua mạng Internet, vì vậy còn được là “truyền hình trực tuyến (Online Television)”. Hiện có hai phương thức xem các chương trình truyền hình qua mạng Internet là:

- Xem trực tiếp theo thời gian thực Real-time (còn gọi là phương thức Download and Play). Việc xem trực tiếp cho phép khách hàng không cần tải file chương trình về máy tính của mình, nhưng cho chất lượng hình ảnh thấp hơn, vì vậy chỉ phù hợp với những kết nối Internet tốc độ cao hoặc yêu cầu không cao về chất lượng hình ảnh.
- Tải file chương trình về máy tính cá nhân (Download stream-file). Phương thức này sẽ đảm bảo được chất lượng hình ảnh chương trình,

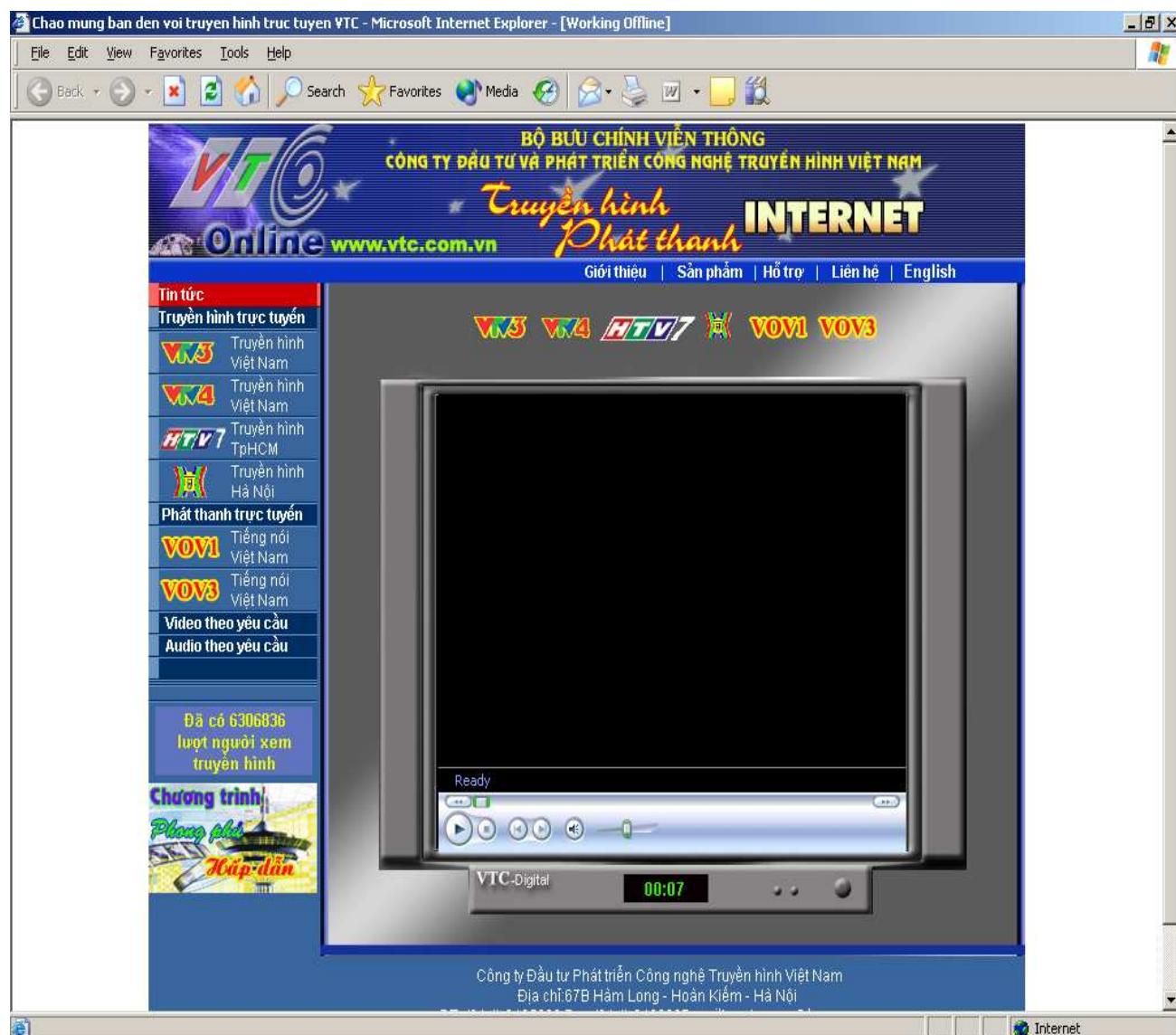
nhưng lại mất thời gian chờ đợi và không áp dụng được cho các chương trình truyền hình trực tiếp.

Để sử dụng dịch vụ truyền hình trực tuyến, người dùng sẽ phải truy cập vào Website của nhà cung cấp dịch vụ trên Internet. Tại đây sẽ có các đường dẫn cho khách hàng lựa chọn chương trình cần xem. Hệ thống truyền hình trực tuyến sử dụng công nghệ Video Streaming, Webstreaming ... hay nói cách khác chính là sự kết giữa công nghệ Web thông thường với công nghệ Streaming.

Hiện nay, hầu hết các hãng truyền hình, thông tấn lớn trên thế giới đều đã sử dụng công nghệ truyền hình trực tuyến như một phương thức đăng tải thông tin hữu hiệu nhất đến với mọi người. Chương trình truyền hình trực tuyến đầu tiên chính thức xuất hiện ngày 13/5/2000 tại Stockholm (Thụy Điển), truyền đi chương trình ca nhạc “EUROVISION SONG CONTEST” do EBU tổ chức hàng năm.

Ở Việt Nam, Đài truyền hình Hà Nội (HTV) và công ty VDC đã phối hợp và đưa vào sử dụng dịch vụ truyền hình trực tuyến lần đầu tiên ngày 15/10/2002, chương trình truyền hình trực tiếp đêm chung kết cuộc thi Tiếng hát trên sóng truyền hình Hà nội lần thứ 5, đã phát thử nghiệm trên Internet tại địa chỉ Website <http://www.hanoitv.org.vn/> hay tại địa chỉ <http://www.vdconline.com/>. Ngoài ra còn nhiều nhà cung cấp dịch vụ khác như Công ty đầu tư và phát triển công nghệ truyền hình Việt Nam - thuộc Bộ Bưu Chính Viễn Thông có địa chỉ Website là: <http://www.vtc.com.vn> có giao diện như sau:

Chương 3: Chuẩn nén Video MPEG -4 và ứng dụng trong truyền hình trên mạng Internet





Hình 14: Truyền hình trực tuyến trên mạng.

Đài truyền hình Việt Nam cũng đang tiến hành đưa vào triển khai công nghệ truyền hình qua mạng Internet cho chương trình VTV4 (chương trình truyền hình cho kiều bào Việt Nam đang định cư ở nước ngoài).

Để đưa được các chương trình truyền hình lên mạng Internet phải có rất nhiều công nghệ và kỹ thuật mới, phức tạp hỗ trợ. Trong khuôn khổ cho phép của bài khoá luận, em xin được đề cập đến công nghệ được dùng để nén video cho việc truyền trên mạng Internet.

3.2. Lựa chọn H.264/ MPEG -4 part 10 cho truyền hình trên mạng Internet

3.2.1 Giới thiệu chung về H.264 /MPEG-4 part 10

Tổ chức tiêu chuẩn quốc tế/ Ủy ban kỹ thuật điện tử quốc tế (ISO/IEC) đã phát triển các tiêu chuẩn mã hoá Video được gọi là MPEG-x (như là MPEG – 1, MPEG – 2 và MPEG - 4). Trong đó, MPEG -1 tốc độ mã hóa 1.5 Mbit/s với mục đích mã hoá video và âm thanh kèm theo, được ứng dụng để lưu trữ trong CD-ROM, đĩa quang... Còn MPEG-2 đã rất thành công trong việc mã hoá video và audio trong truyền hình kỹ thuật số, với tốc độ mã hoá từ 3 – 40 Mbit/s. Chúng được ứng dụng để quảng bá video số trên mạng cáp, đường truyền số xDSL, truyền hình qua vệ tinh ...

Tuy nhiên, các ứng dụng trên mạng Internet hiện nay vẫn gặp vấn đề khó khăn đó là sự hạn chế về băng thông, nghẽn mạng và sự đa dạng các thiết bị hiển thị của người dùng. Nhu cầu truyền phát các ứng dụng video và đa phương tiện (multimedia) mới trên hạ tầng kỹ thuật Internet là nảy sinh các yêu cầu chức năng mới mà các chuẩn nén video trước đó không thể đáp ứng nổi. Vào tháng 10 /1998, MPEG -4 đã ra đời và đã giải quyết được các yêu cầu mới đặt ra. Với tốc độ mã hoá dữ liệu khoảng 1.5 Mbit/s chuẩn nén MPEG -4 đã giải quyết phần nào về việc tắc nghẽn mạng và sự hạn chế của băng thông. Đồng thời sự đa dạng của các thiết bị của người dùng cũng được giải quyết vì trên thực tế đa số các hãng cung cấp dịch vụ truyền thông theo dòng Multimedia đều hỗ trợ chuẩn MPEG -4 trong các cấu trúc hạ tầng và sản phẩm đang triển khai của mình như: Quick Time 6 (của Apple), Real Networks, DivX5 (của DivXNetwork) hay sản phẩm nổi tiếng của Microsoft là Windows Media Player đều cho phép người dùng xem nội dung MPEG-4 trong các thiết bị hiển thị này.

MPEG -4 bao gồm các bộ phận riêng rẽ, có quan hệ chặt chẽ với nhau và có thể được triển khai ứng dụng riêng hay tổ hợp với các phần khác. Các phần cơ bản ở đây là:

- Phần 1 : System.

- Phần 2: MPEG -4 ASP (Advanced Simple Profile).
- Phần 3: Audio.
- Phần 4: Conformance – xác định việc thử nghiệm triển khai một MPEG -4 sẽ như thế nào.
- Phần 5: Các phần mềm tham chiếu, đưa ra một nhóm các phần mềm tham chiếu quan trọng, được sử dụng để triển khai MPEG -4 và phục vụ như một ví dụ demo về các bước phải thực hiện khi triển khai.
- Phần 6: Khung chuẩn cung cấp truyền thông đa phương tiện tích hợp DMIF (Delivery Multimedia Integration Framework), xác định một giao diện giữa các ứng dụng và mạng/ lưu trữ.
- Phần 7: Các đặc tính của một bộ mã hóa video tối ưu (bổ sung phần mềm tham chiếu, nhưng không phải là các triển khai tối thiểu cần thiết).
- Phần 8: Giao vận, cần để ánh xạ như thế nào các dòng MPEG -4 vào giao vận IP.
- Phần 9: mô tả phần cứng tham chiếu (Reference Hardware Description).
- Phần 10: MPEG -4 Advanced Video Coding /H.264 (mã hoá Video tiên tiến tương ứng với tiêu chuẩn H.264 của ITU).
- Phần 11: mô tả khung hình
- Phần 12: Định dạng file truyền thông ISO (ISO Media File Format).
- Phần 13: Quản lý bản quyền nội dung IPMP (Intellectual Property Managenment and Protection Extensions).
- Phần 14: Định dạng File MP4(trên cơ sở phần 12).
- Phần 15: Định dạng file AVC(trên cơ sở phần 12).
- Phần 16: AFX (Animation Framwork Extensions)và MuW(Multi-uer Worlds).

Đồng thời, Hiệp hội viễn thông quốc tế (ITU) đã cho ra đời các khuyến nghị gọi tắt là chuẩn H.26x (H.261, H.262, H.263 và H.264). Những khuyến nghị này được thiết kế cho các ứng dụng truyền thông Video thời gian thực như Video Conferencing(truyền hình tương tác) hay điện thoại truyền hình.

Với đối tượng để truyền dẫn video là mạng Internet thì ứng cử viên hàng đầu là chuẩn nén MPEG-4AVC hay còn gọi là H.264/MPEG-4 Part 10.

3.2.2 Tính kế thừa của chuẩn nén H.264/MPEG- 4 part 10

Mục tiêu chính của chuẩn nén H.264 đang phát triển nhằm cung cấp Video có chất lượng tốt hơn nhiều so với những chuẩn nén Video trước đây. Điều này có thể đạt được nhờ sự kế thừa các ưu điểm của các chuẩn nén video trước đó và một số ưu điểm như sau:

- Phân chia mỗi hình ảnh thành các Block bao gồm nhiều điểm ảnh, do vậy quá trình xử lý từng ảnh có thể được tiếp cận tới mức Block.

- Khai thác triệt để sự dư thừa về mặt không tồn tại giữa các hình ảnh liên tiếp bởi một vài mã của những Block gốc thông qua dự đoán về không gian, phép biến đổi, quá trình lượng tử và mã hoá Entropy.

- Khai thác sự phụ thuộc tạm thời của các Block của hình ảnh liên tiếp, do đó chỉ cần mã hoá những chi tiết thay đổi giữa các ảnh liên tiếp. Việc này được thực hiện thông qua dự đoán và bù chuyển động. Với bất kỳ Block nào cũng có thể được thực hiện từ một hoặc vài ảnh mã hoá trước đó hay ảnh được mã hoá sau đó để quyết định Vector chuyển động, các Vector được sử dụng trong bộ mã hoá và giải mã để dự đoán các loại Block.

- Khai thác tất cả sự dư thừa về không gian còn lại trong ảnh bằng việc giải mã các Block dư thừa. Ví dụ như sự khác biệt giữa các Block gốc và Block dự đoán sẽ được mã hoá thông qua quá trình biến đổi, lượng tử hoá và mã hoá Entropy.

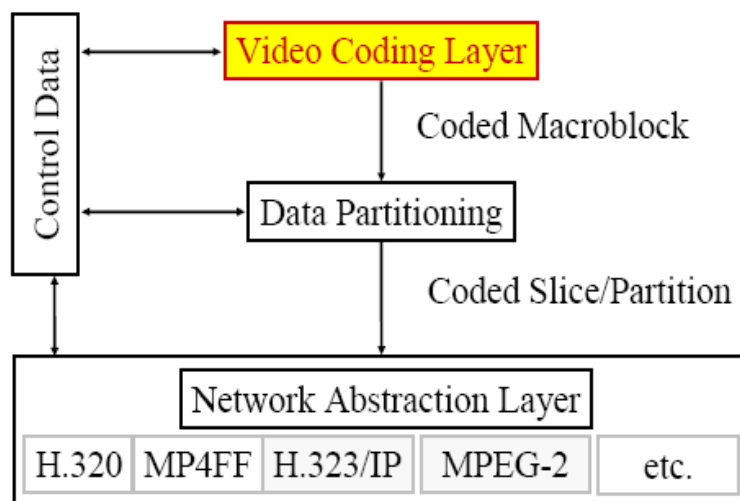
3.3. Tiêu chuẩn H.264/MPEG - 4 Part 10

Như ta đã biết, đặc điểm nổi bật của MPEG-4 trong việc mã hoá hình ảnh là mô tả các đối tượng trong ảnh thành các đối tượng riêng biệt để mã hoá. Sau đó các đối tượng đó được mã hoá (tương tự như trong kỹ thuật mã hoá MPEG-2), tạo dòng và truyền đi trên mạng.

3.3.1. Lớp trừu tượng mạng NAL (Network Abstraction Layer)

Lớp trừu tượng mạng NAL (Network Abstraction Layer) được xác định để định dạng dữ liệu và cung cấp thông tin Header trong cách thích hợp cho việc chuyển trở bởi các lớp môi trường lưu trữ. Tất cả dữ liệu được chứa trong các khối NAL, mỗi khối chứa một số nguyên byte. Một khối NAL xác định định dạng chung cho việc sử dụng trong cả hệ thống định hướng gói (Packet-oriented) và hệ thống định hướng dòng bit. Định dạng

của các khối NAL là đồng nhất cho cả việc phân phối dòng truyền tải định hướng gói và định hướng dòng bit, ngoại trừ rằng mỗi khối NAL trong lớp truyền tải định hướng dòng bit có thể có một tiền tố mã hoá bắt đầu (start code prefix).



Hình 15: Cấu trúc của bộ mã hoá Video H264

3.3.2. Các Profile và các Level

Các Profile và các Level xác định các điểm bắt buộc. Các điểm bắt buộc này được thiết kế để tạo thuận tiện cho sự linh hoạt giữa các ứng dụng khác nhau của tiêu chuẩn H264/MPEG Part 10.

Profile xác định tập các công cụ mã hoá hoặc các thuật toán mã hoá có thể được sử dụng để tạo ra các dòng bit tương thích.

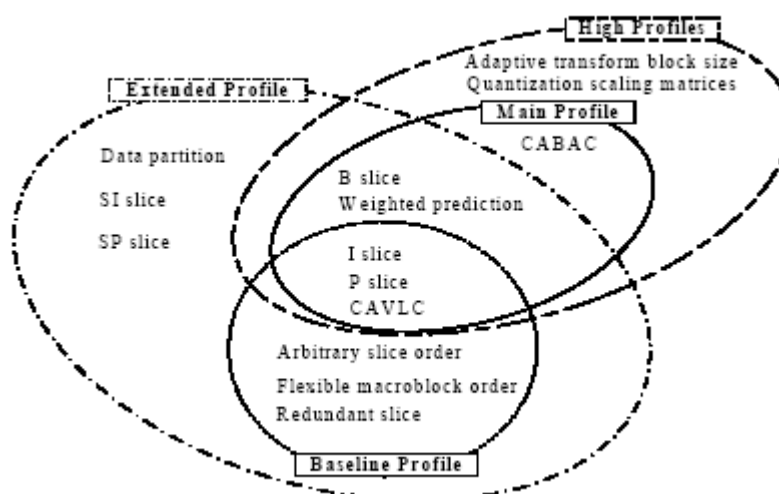
Level đặt ra những giới hạn cho những thông số chủ yếu nhất định của dòng bit.

Có 3 Profile trong phiên bản thứ nhất: Baseline, Main, Extended. Baseline Profile thì được ứng dụng cho thời gian thực như các dịch vụ đàm thoại như truyền hình tương tác và điện thoại hình. Main Profile được thiết kế cho các thiết bị lưu trữ số và truyền hình quảng bá. Extended Profile với mục đích là ứng dụng cho các dịch vụ đa phương tiện trên mạng Internet. Ngoài ra còn có 4 High Profile đã được định nghĩa trong độ trung thực của vùng mở rộng cho các ứng dụng như: sự đóng góp nội dung, sự phân phối nội dung, việc dàn dựng studio và xử lý quá trình gửi thông tin. Các High Profile đó là : High, High 10, High 4:2:2, High 4:4:4. Trong đó:

- High Profile: hỗ trợ cho việc lấy mẫu 8- bit video với chuẩn 4:2:0.
- High 10 Profile: hỗ trợ cho việc lấy mẫu 10-bit video với chuẩn 4:2:0.

- High 4:2:2 Profile: hỗ trợ cho việc lấy mẫu 10 –bít cho thành phần màu với chuẩn 4:2:2.
- High 4:4:4 Profile được hỗ trợ cho việc lấy mẫu 12-bít cho thành phần màu với chuẩn 4:4:4.

Các Profile đều có các phần mã hoá chung và các phần mã hoá riêng được chỉ trong hình sau.



Hình 16: Các phần mã hoá riêng của Profile trong H264.

• **Các phần chung cho tất cả các Profile :**

- Slice I (Slice mã hoá trong ảnh): Slice đã mã hoá bằng việc sử dụng dự đoán chỉ từ các mẫu đã mã hoá bên trong cùng một Slice .
- Slice P (Predictive-coded Slice): các Slice được mã hoá bằng việc sử dụng dự đoán liên ảnh từ các ảnh tham chiếu đã được mã hoá trước đó sử dụng hầu hết cho Vector chuyển động và hệ số tham chiếu để dự đoán các giá trị mẫu cho mỗi Block .
- CAVLC (Context-based Adaptive Variable Length Coding): mã hoá độ dài biến đổi tương thích với tình huống, cho mã hoá Entropy.

• **Baseline Profile (Profile mức cơ bản):**

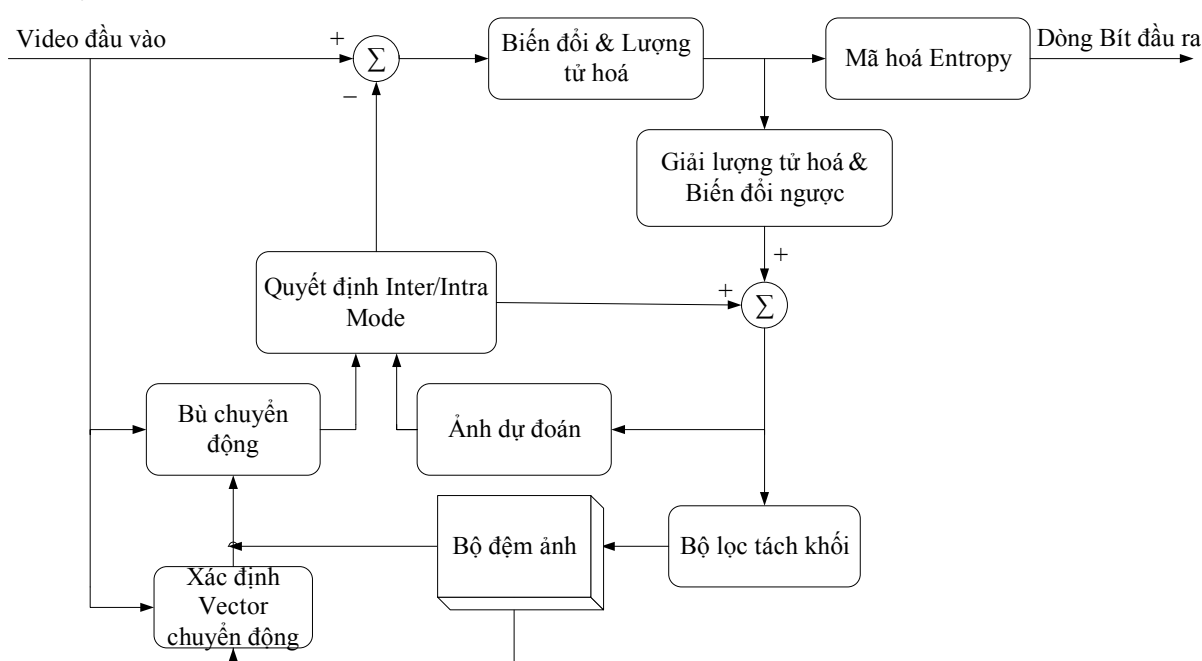
- Thứ tự Macro-Block linh hoạt (Flexible Macro-Block order): các Macro-Block có thể không cần thiết trong thứ tự quét mảnh. Bản đồ định vị các Macro-Block cho một nhóm Slice .

- Thứ tự Slice bất kỳ (Arbitrary Slice order): địa chỉ Macro-Block của Macro-Block đầu tiên của một Slice của một ảnh có thể nhỏ hơn địa chỉ Macro-Block của đầu tiên của một vài Slice khác có trước của cùng ảnh đã mã hoá.
- Slice dư thừa (Redundant Slice): Slice này thuộc về dữ liệu đã mã hoá dư thừa thu được từ cùng hay khác một tỷ lệ mã hoá, so với dữ liệu đã mã hoá trước đó của cùng một Slice .
- **Main Profile(Profile chính):**
 - B Slice (Bi-directionally predictive-coded Slice): Slice đã mã hoá bằng việc sử dụng dự đoán liên ảnh từ các ảnh tham chiếu đã mã hoá trước đó, sử dụng tại hầu hết hai Vector chuyển và các hệ số tham chiếu để dự đoán các giá trị mẫu của từng Block .
 - Dự đoán có trọng số (weighted prediction): phân chia các hoạt động bằng việc áp dụng một thừa số trọng số để lấy mẫu dữ liệu dự đoán bù chuyển động của mỗi Block .
 - CABAC (Context-based Adaptive Binary Arithmetic Coding): mã hóa thuật toán nhị phân tương thích với tình huống, cho mã hoá Entropy.
- **Extended Profile (Profile mở rộng):**
 - Bao gồm tất cả các phần của Baseline Profile: thứ tự Macro-Block linh hoạt, thứ tự Slice bất kỳ, Slice dư thừa.
 - Slice SP: Slice mã hoá đặc biệt để có khả năng chuyển đổi giữa các dòng video, tương tự như việc mã hoá một Slice P.
 - Slice SI: Slice đã chuyển đổi tương tự như việc mã hoá một Slice I.
 - Sự phân chia dữ liệu: dữ liệu đã mã hoá được đặt trong sự phân chia dữ liệu tách rời, mỗi sự phân chia có thể được đặt trong đơn vị lớp khác nhau.
 - Slice B.
 - Dự đoán có trọng số.

- **High Profile (Profile cao):**

- Bao gồm tất cả các phần của Main Profile: Slice B, dự đoán có trọng số, CABAC.
- Kích cỡ khối biến đổi tương thích: biến đổi nguyên khối 4 x 4, 8 x 8 cho việc lấy mẫu thành phần chói.
- Các ma trận phân chia lượng tử hoá: sự phân chia khác nhau tùy theo các tần số riêng được kết hợp với các hệ số biến đổi trong quá trình lượng tử hoá để làm tối ưu chất lượng chủ thể.

3.3.3. Kỹ thuật mã hoá video



Hình 17: Sơ đồ mã hoá Video của H264/MPEG Part 10

Lớp mã hoá video của H264/MPEG Part 10 là sự kết hợp của mã hoá không gian, mã hoá thời gian và mã chuyển vị. Ảnh được tách thành các khối, ảnh đầu tiên của dãy hoặc điểm truy cập ngẫu nhiên thì được mã hoá “Intra”- mã hoá trong ảnh, có nghĩa là không dùng thông tin của các ảnh khác mà chỉ dùng thông tin chứa trong ảnh đó. Mỗi mẫu của một khối trong một Frame Intra được dự đoán nhờ dùng các mẫu không gian bên cạnh của các khối đã mã hoá trước đó. Đối với tất cả các ảnh còn lại của dãy hoặc giữa các điểm truy cập ngẫu nhiên, mã hoá “Inter” được sử dụng, dùng dự đoán bù

chuyển động từ các ảnh được mã hoá trước đó. Quá trình mã hoá cho dự đoán liên ảnh (bù chuyển động) gồm việc lựa chọn dữ liệu chuyển động, các ảnh tham chiếu và sự dịch chuyển không gian được ứng dụng cho tất cả việc lấy mẫu của khối.

Bộ mã hoá có thể lựa chọn giữa mã hoá Intra và Inter cho miền hình dạng khối của mỗi ảnh. Mã hoá Intra có thể chỉ ra điểm truy cập của chuỗi được mã hoá, tại đó việc giải mã có thể bắt đầu và tiếp tục một cách chính xác. Mã hoá Intra sử dụng các mode dự đoán không gian riêng rẽ để làm giảm độ dư thừa không gian trong tín hiệu gốc của mỗi ảnh đơn. Mã hoá Inter (dự đoán một chiều hay nhiều chiều) thì việc sử dụng dự đoán liên ảnh hiệu quả hơn cho mỗi block của giá trị lấy mẫu từ một vài ảnh được giải mã trước đó.

Mã hoá Inter sử dụng các Vector chuyển động cho các block cơ sở dự đoán liên ảnh (Inter prediction) để làm giảm sự dư thừa thời gian giữa các ảnh (picture) khác nhau. Việc dự đoán được thu được từ tín hiệu đã lọc tách khối của các ảnh được thiết lập lại trước đó.

Bộ lọc tách khối làm giảm sự nhiễu khối tại các đường biên của block. Các vector chuyển động và các mode dự đoán trong ảnh (intra prediction) có thể (theo lý thuyết) làm biến đổi kích thước block trong ảnh. Sự dự đoán thẳng dư được nén tốt hơn bằng việc sử dụng một phép biến đổi để loại bỏ sự tương quan theo không gian trong một block trước khi được lượng tử hoá.

Cuối cùng, Vector chuyển động hay các mode dự đoán liên ảnh được liên kết với thông tin của hệ số biến đổi lượng tử hóa và được mã hoá sử dụng mã Entropy như mã hoá chiều dài biến đổi thích ứng theo tình huống CAVLC (context-adaptive variable length code) hay mã hoá theo số học nhị phân thích ứng theo tình huống CABAC (context-adaptive binary arithmetic coding).

3.3.3.2. Các ảnh và bù chuyển động dùng trong H264/MPEG Part 10

1. Chia ảnh thành các Macro-Block .

Mỗi ảnh video, Frame hoặc Field được chia thành các Macro-Block có kích thước cố định bao trùm một diện tích ảnh hình chữ nhật gồm 16 x 16 mẫu cho thành phần chói(luma) và 8 x 8 mẫu cho một trong hai thành phần màu(chroma). Tất cả các mẫu Macro-Block của thành phần chói hoặc của thành phần màu được dự đoán theo không gian hoặc thời gian và kết quả sai số dự đoán được truyền đi bằng việc sử dụng mã hoá biến đổi. Do đó, mỗi thành phần màu của sai số dự đoán(*prediction residual*) được chia

nhỏ thành các khối. Mỗi khối được biến đổi nhờ dùng một phép biến đổi nguyên, và các hệ số biến đổi được lượng tử hoá và được biến đổi sử dụng phương pháp mã hoá Entropy.

Các Macro-Block được tổ chức thành các Slice, biểu diễn các tập con của ảnh đã cho và có thể được giải mã độc lập. Thứ tự truyền của các Macro-Block trong dòng bit phụ thuộc vào “Bản đồ định vị Macro-Block” - Macro-Block Allocation Map- và không nhất thiết phải theo thứ tự quét.

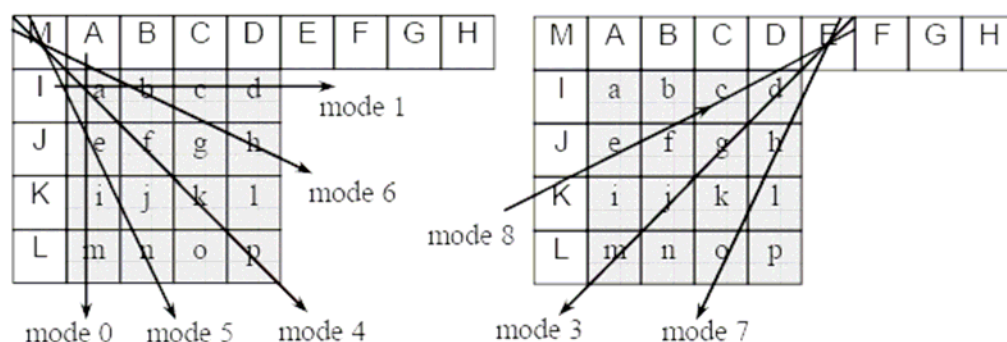
H264/MPEG Part 10 hỗ trợ 5 dạng mã hóa Slice khác nhau. Đơn giản nhất là Slice I, trong đó tất cả Macro-Block được mã hoá không có sự tham chiếu tới các ảnh khác trong dãy video. Tiếp theo là hai dạng Slice P và Slice B có sự tham chiếu tới các ảnh khác; với Slice P thì chỉ tham chiếu tới các ảnh trước đó; còn Slice B thì tham chiếu tới cả ảnh trước và ảnh sau nó. Hai dạng khác mà mới xuất hiện ở H264/MPEG Part 10 đó là SI (Switching I) và SP (Switching P) được dùng để cho chuyển mạch hiệu quả giữa các dòng bit được mã hoá ở các tốc độ bit khác nhau.

Để cung cấp các phương pháp che giấu hiệu quả trong các kênh có xu hướng bị lỗi với các ứng dụng độ trễ thấp, H264/MPEG Part 10 hỗ trợ một đặc điểm gọi là thứ tự Macro-Block linh hoạt FMO (Flexible Macro-Block Ordering). FMO định rõ một giản đồ ấn định các Macro-Block trong ảnh vào một hoặc vài nhóm Slice. Mỗi nhóm Slice được truyền riêng biệt. Nếu một nhóm Slice nào bị mất, các mẫu trong các Macro-Block bên cạnh về mặt không gian, thuộc về các nhóm Slice được thu đúng, có thể được sử dụng cho che giấu hiệu quả lỗi. Các giản đồ được phép trải rộng từ các giản đồ hình chữ nhật tới các giản đồ phân tán theo các quy tắc khác nhau.

2. Dự đoán trong ảnh Intra- Frame.

Các tiêu chuẩn trước đó đã chấp nhận Macro-Block mã hoá trong ảnh, mã hoá bằng chính nó mà không có dự đoán thời gian. Macro-Block mã hoá trong ảnh xảy ra trong các mảng (slice) hay các Macro-Block không chấp nhận sự hiệu chỉnh thời gian của việc dự đoán bù chuyển động. Về bản chất, Macro-Block mã hoá trong ảnh đưa ra một số lượng lớn các bit được mã hoá. H.264 sử dụng phương pháp dự đoán các Macro-Block mã hoá trong ảnh để giảm một lượng lớn các bit được mã hoá bằng chính bản thân tín hiệu gốc đưa vào. Để mã hoá một block hay một Macro-Block trong mode mã hoá trong ảnh, một block được định dạng cơ sở trên các block được khôi phục trước đó (nhưng không qua bộ lọc). Tín hiệu dư thừa giữa các block hiện tại và dự đoán cuối cùng được mã hoá. Để lấy mẫu tín hiệu chói, Block dự đoán có thể được định dạng là: cho mỗi

Block nhỏ (Subblock) là 4 x 4, mỗi Block là 8 x 8, hay mỗi Macro-Block là 16 x 16. Trong trường hợp được lựa chọn từ 9 mode, cho các Block luma (khối tín hiệu chói) là 4 x 4 và 8 x 8; 4 mode cho một Block luma 16 x 16; và 4 mode cho mỗi Block chroma (khối tín hiệu màu).



Hình 18: Các mode trong MPEG-4

Hình 18 chỉ ra một Block luma 4 x 4 được dự đoán. Để dự đoán các mẫu [a,b, ..., p] cho Block hiện tại, các mẫu khôi phục lại ở phía trên và bên trái trước đó [A,B, ...,M] được dùng theo định hướng các mode. Các mũi tên trong hình 18 chỉ định hướng dự đoán của mỗi mode. Với mode 0 (vertical-theo chiều dọc) và mode 1 (horizontal-theo chiều dọc), các mẫu dự đoán được định dạng bằng phép ngoại suy từ các mẫu cao [A,B,C,D] và từ các mẫu bên trái [I, J, K, L], tách biệt nhau. Với mode 2 (DC- một chiều), tất cả các mẫu dự đoán được định dạng bằng cách lấy trung bình của các mẫu phía trên và bên trái [A, B, C, D, I, J, K, L]. Với mode 3 (đường chéo xuống phía trái), mode 4 (đường chéo xuống phía phải), mode 5 (theo chiều dọc phía phải), mode 6 (theo chiều ngang xuống), mode 7 (theo chiều dọc phía trái) và mode 8 (theo chiều ngang lên), các mẫu dự đoán được định dạng từ trọng số trung bình của các mẫu dự đoán từ A-M. Ví dụ, các mẫu a và d được dự đoán lần lượt bằng cách tính làm tròn $(I/4 + M/2 + A/4)$ và $(B/4 + C/2 + D/4)$ trong mode 4, và bằng $(I/2 + J/2)$ và $(J/4 + K/2 + L/4)$ trong mode 8. Bộ mã hoá có thể lựa chọn mode dự đoán cho mỗi Block để số dư giữa các Block được mã hoá và dự đoán là nhỏ nhất.

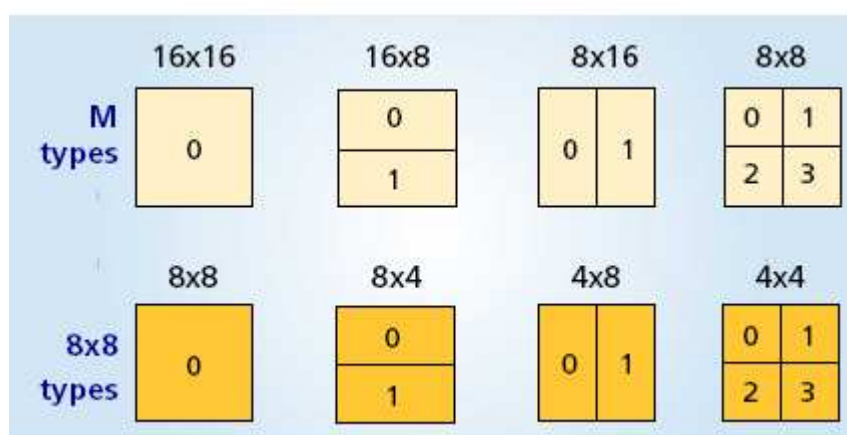
Để dự đoán cho mỗi Block luma 8 x 8, một mode được lựa chọn từ 9 mode, tương tự như việc dự đoán khối trong ảnh 4 x 4. Để dự đoán cho tất cả thành phần độ chói luma 16 x 16 của một Macro-Block, thì ta phải dùng 4 mode. Với mode 0 (theo chiều dọc), mode 1 (theo chiều ngang), mode 2 (DC), việc dự đoán thì tương tự như trong các trường hợp của Block luma 4 x 4. Với mode 4 (mặt phẳng-Plane), một hàm mặt phẳng tuyến tính được làm khớp với các mẫu phía trên và bên trái. Mỗi thành phần màu của một

Macro-Block được dự đoán từ các mẫu thành phần màu ở phía trên và/hay đến bên trái mà vừa được mã hoá hay khôi phục lại trước đó. Việc dự đoán thành phần màu được định nghĩa cho 3 kích cỡ Block có thể là: thành phần màu 8×8 định dạng 4:2:0, thành phần màu 8×16 trong định dạng 4:2:2 và thành phần màu 16×16 trong định dạng 4:4:4. Với 4 mode dự đoán cho tất cả các trường hợp thì rất giống với mode dự đoán thành phần chói 16×16 , ngoại trừ các mode có thứ tự sau là khác: mode 0, mode 1, mode 2 và mode 3.

3. Bù chuyển động trong các Slice P (Prediction Inter Frame)

Ngoài các dạng mã hoá Macro-Block Intra, các dạng mã hoá bù chuyển động hoặc dự đoán khác được xác định cho các Macro-Block Slice P. Dự đoán liên ảnh được làm giảm với sự tương quan theo thời gian với sự trợ giúp của việc xác định Vector chuyển động (Motion Estimation) và bù chuyển động (compensation).

* **Chia Macro-Block thành các Block** : Trong H.264, ảnh hiện tại có thể được phân chia thành các Macro-Block hay các Block nhỏ hơn. Một Macro-Block của các mẫu thành phần màu 16×16 có thể chia nhỏ hơn thành các Block kích cỡ từ 4×4 . Với Macro-Block mode 16×16 , có 4 trường hợp là: 16×16 , 16×8 , 8×16 , hay 8×8 , ngoài ra cũng có 4 trường hợp cho mode 8×8 là: 8×8 , 8×4 , 4×8 , hay 4×4 . Một Block kích cỡ nhỏ hơn yêu cầu một số lượng bit lớn để truyền Vector chuyển động và dữ liệu thêm vào của việc phân chia, tuy nhiên dữ liệu dư bù chuyển động có thể được giảm. Do đó, việc lựa chọn kích cỡ phân chia phụ thuộc vào các đặc điểm video đầu vào. Sự phân chia Macro-Block thành các Block được minh hoạ trong hình sau:



Hình 19: Phân chia Macro-Block cho bù chuyển động

Trên: Phân chia các Macro-Block .

Dưới: Phân chia các phần 8×8 .

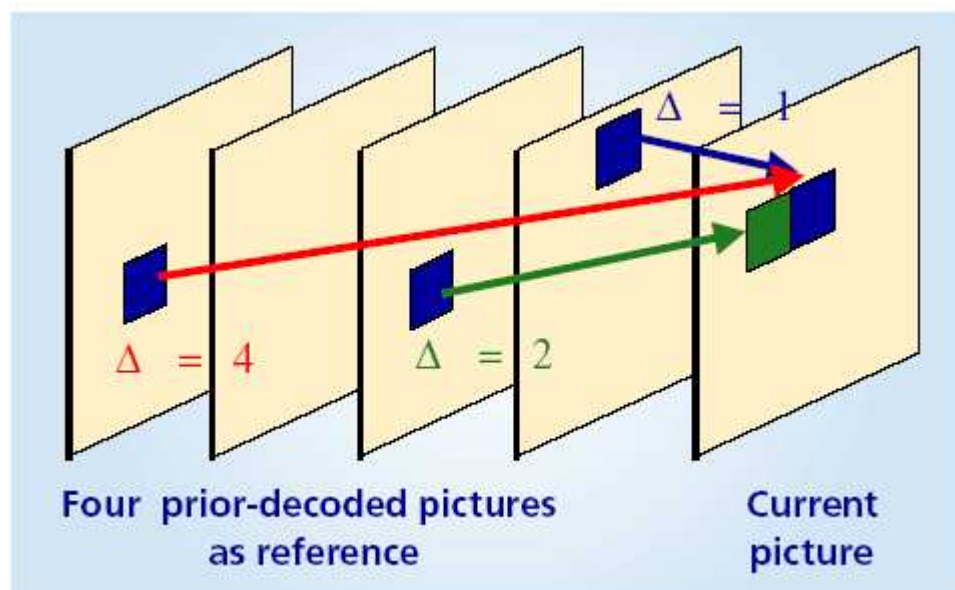
*** Các giá trị dự đoán cho thành phần chói (Luma) và thành phần màu (Chroma):**

Tín hiệu dự đoán cho mỗi khối luma $m \times n$ đã mã hoá dự đoán thu được bằng việc đổi chỗ diện tích của ảnh tham chiếu tương ứng, nó được xác định bởi Vector chuyển động biến đổi và chỉ số ảnh tham chiếu. Như vậy, nếu Macro-Block được mã hoá khi dùng dạng Sub-Macro-Block (Macro-Block phụ) Inter 4×4 , nhiều nhất 16 Vector chuyển động có thể được truyền đi cho một Macro-Block P Slice. Độ chính xác bù chuyển động bằng $\frac{1}{4}$ khoảng cách mẫu. Trong trường hợp Vector chuyển động nhắm trúng vị trí của một mẫu nguyên vẹn (Integer- Sample), các tín hiệu dự đoán là các mẫu tương ứng của ảnh tham chiếu; nếu khác đi, chúng được nội suy ở các vị trí của mẫu con (Sub-Sample). Trong thành phần tín hiệu chói, các mẫu phần tử ảnh phụ tại vị trí một phần hai phần tử ảnh được phát đầu tiên và được nội suy từ các mẫu phần tử ảnh nguyên bên cạnh sử dụng một bộ lọc FIR 6 đầu ra với các trọng số (1, -5, 20, 20, -5, 1)/32. Đôi khi tất cả các mẫu một phần hai phần tử ảnh là sẵn có, với mỗi mẫu một phần tư phần tử ảnh được sinh ra khi sử dụng phép nội suy tuyến tính hai chiều giữa các mẫu một phần hai hay cả phần tử ảnh bên cạnh (lấy trung bình ở các vị trí nguyên mẫu và nửa mẫu).

Với thành phần tín hiệu màu (Chroma): Việc lấy mẫu nguồn Video định dạng 4:2:0, $\frac{1}{8}$ các mẫu phần tử ảnh được dùng trong các thành phần màu (tương ứng với $\frac{1}{4}$ các mẫu phần tử ảnh trong thành phần màu). Các mẫu này được nội suy hai chiều (nội suy tuyến tính) giữa các mẫu nguyên phần tử ảnh màu.

*** Bù chuyển động:** H264/MPEG Part 10 cho phép các Vector chuyển động không hạn chế, tức là chúng có thể nhắm ra ngoài miền ảnh. Trong trường hợp này các Frame tham chiếu được mở rộng ra ngoài biên ảnh bằng việc lặp lại các Pixel biên trước khi nội suy. Các thành phần Vector chuyển động được mã hoá vì sai khi dùng hoặc là giá trị trung bình (median) hoặc là dự đoán định hướng từ các khối xung quanh. Không có dự đoán thành phần Vector chuyển động ở các biên của Slice.

H264/MPEG Part 10 hỗ trợ dự đoán bù chuyển động đa ảnh (Multi-Picture). Điều này có nghĩa là có nhiều hơn một ảnh được mã hoá trước đó có thể được sử dụng để tham chiếu cho dự đoán bù chuyển động. Hình sau minh hoạ khái niệm này



Hình 20: Bù chuyển động nhiều Frame – ngoài Vector chuyển động, các tham số tham chiếu ảnh (Δ) cũng được truyền đi.

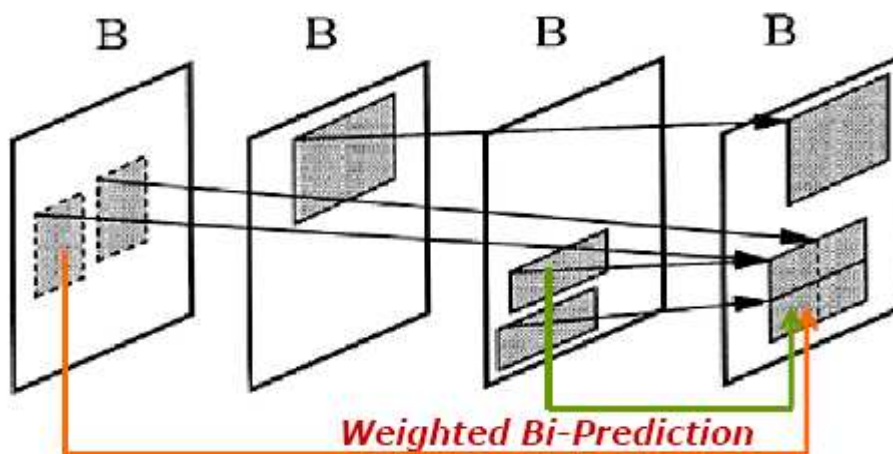
Cả bộ mã hoá và bộ giải mã phải lưu trữ các ảnh tham chiếu được sử dụng cho dự đoán ảnh Inter trong bộ nhớ đệm đa ảnh (Multi-Picture). Bộ giải mã sao lưu lại bộ nhớ đệm đa ảnh của bộ mã hoá, theo dạng nhớ đệm ảnh tham chiếu và các hoạt động điều khiển quản lý bộ nhớ bất kỳ được xác định trong dòng bit. Trừ khi kích thước bộ nhớ đệm đa ảnh được đặt cho mỗi ảnh, chỉ số mà ảnh tham chiếu được định vị ở đó bên trong bộ nhớ đệm đa ảnh thì phải được ký hiệu. Tham số chỉ số tham chiếu cho mỗi khối luma bù chuyển động 16 x 16, 16 x 8, 8 x 16 hoặc 8 x 8.

Ngoài các mode bù chuyển động được mô tả ở trên, Macro-Block P- Slice cũng có thể được mã hoá trong mode gọi là SKIP. Đối với mode này, cả tín hiệu lỗi dự đoán lượng tử hóa lẫn Vector chuyển động hoặc tham số chỉ số tham chiếu đều không được truyền đi. Các tín hiệu thiết lập lại thì thu được tương tự như cách với tín hiệu dự đoán của Macro-Block Inter 16 x 16, tức là tham chiếu tới ảnh đặt ở vị trí có chỉ số 0 trong bộ nhớ đệm đa ảnh. Nhìn chung Vector chuyển động được sử dụng cho việc thiết lập lại Macro-Block SKIP là đồng nhất với bộ dự đoán Vector chuyển động cho khối 16 x 16. Nếu các điều kiện đặc biệt được duy trì, Vector chuyển động Zero được sử dụng thay cho nó.

4. Bù chuyển động trong các Slice B (Bi-Direction Prediction Inter Frame)

Dự đoán hai chiều rất có hiệu quả để giảm sự tương quan theo thời gian bằng việc sử dụng các ảnh tham chiếu. Các chuẩn nén hiện nay với các ảnh B sử dụng mode dự

đoán hai chiều, tức là chỉ cho phép kết hợp các tín hiệu dự đoán trước đó với các tín hiệu dự đoán sau. Một tín hiệu dự đoán được nhận từ một ảnh trong ảnh (Inter Picture) tiếp sau, hay từ một ảnh dự đoán trước, hoặc từ tín hiệu trung bình tuyến tính của hai tín hiệu dự đoán bù chuyển động.



Hình 21: Ảnh nội suy B (dự đoán hai chiều)

So với các tiêu chuẩn trước đó, H264/MPEG Part 10 đã tổng quát khái niệm Slice B và không những chỉ hỗ trợ một cặp dự đoán theo hướng forward/backward (tiền/lùi) mà còn cả hai cặp theo hướng forward/forward (tiền/tiến) và backward/backward (lùi/lùi). Tham chiếu tiến hai bước (two forward) có thể có lợi cho dự đoán bù chuyển động của một vùng vừa thay đổi cảnh trước đó, và tham chiếu lùi hai bước mà cảnh vừa thay đổi sau đó. Các Slice đã mã hoá dự đoán hai chiều có thể cũng được sử dụng để tham chiếu cho việc mã hoá trong ảnh của các ảnh khác. Như vậy, các ảnh khác có thể tham chiếu các ảnh B cho dự đoán chuyển động, phụ thuộc vào hoạt động điều khiển quản lý bộ nhớ của việc nhớ đệm đa ảnh. Do đó, sự khác nhau cơ bản giữa Slice B và Slice P là: các Slice B được mã hoá theo cách trong đó một số Macro-Block hoặc Block có thể dùng trung bình trọng số của hai giá trị dự đoán bù chuyển động riêng biệt cho việc hình thành các tín hiệu dự đoán. Nhìn chung các Slice B dùng hai bộ nhớ đệm ảnh tham chiếu khác nhau gọi là bộ nhớ đệm ảnh tham chiếu thứ nhất và thứ hai tương ứng. Các ảnh nào được xấp đặt đúng vị trí thực tế trong mỗi bộ nhớ đệm ảnh tham chiếu thì được đưa ra điều khiển bộ nhớ đệm ảnh.

* **Các mode dự đoán trong Slice Block:** Trong các Slice B, có 4 dạng dự đoán ảnh Inter khác nhau được hỗ trợ: dự đoán List 0, List 1, hai hướng (bi-predictive) và trực tiếp (direct). Trong đó, dự đoán List 0 hiển thị rằng tín hiệu dự đoán được tạo thành nhờ dùng bù chuyển động từ ảnh của bộ nhớ đệm tham chiếu đầu tiên, thì ảnh của bộ nhớ đệm ảnh tham chiếu thứ hai được sử dụng cho việc xây dựng tín hiệu dự đoán nếu dự đoán List 1 được sử dụng.

H264/MPEG Part 10 giới thiệu mode trực tiếp (direct-mode), nó không yêu cầu thông tin kèm theo (bên cạnh), nhưng nhận ảnh tham chiếu, kích cỡ khối, dữ liệu Vector chuyển động từ ảnh Inter tiếp theo. Dự đoán trọng số được làm tăng thêm do sự chuyển tiếp dần dần từ cảnh (scene) này sang cảnh kia.

Trong mode dự đoán hai hướng, tín hiệu dự đoán được tạo thành bởi trung bình trọng số của tín hiệu dự đoán bù chuyển động List 0 và List 1. Dự đoán có trọng số sẽ được ta đề cập đến sau đây.

* **Dự đoán có trọng số (weighted Prediction)**

Tất cả các chuẩn nén đang có hiện nay đều coi các ảnh tham chiếu có trọng số bằng nhau, ví dụ như tín hiệu dự đoán thu được bằng cách lấy trung bình các tín hiệu tham chiếu có trọng số bằng nhau. Nhưng sự biến đổi dần dần từ cảnh này sang cảnh khác thì cần các trọng số khác nhau.

Sự biến đổi dần dần là rất phổ biến trong các chuyển động, chuyển cảnh từ mờ dần sang đen (fade to black) (fade to black : các mẫu thành phần chói của cảnh dần dần xấp xỉ zero, các mẫu thành phần màu dần dần xấp xỉ 128), một cảnh chuyển đổi sang mờ dần từ đen (fade from black).

H264/MPEG Part 10 sử dụng phương pháp dự đoán có trọng số cho một Macro-Block của Slice P hay Slice B. Một tín hiệu dự đoán là p cho Slice B được thu được bằng các trọng số khác nhau từ hai tín hiệu tham chiếu là r1 và r2, ta có:

$$P = w1 \times r1 + w2 \times r2$$

Trong đó w1 và w2 là các trọng số. Nó được xác định khác nhau theo hai loại ẩn (Implicit) và hiện (explicit) trong bộ mã hoá. Trường hợp hiện, các thừa số được truyền trong phần đầu của Slice (header Slice). Trong trường hợp ẩn, các thừa số được tính toán dựa trên cơ sở khoảng cách thời gian giữa các ảnh. Trọng số nhỏ hơn được áp dụng nếu khoảng cách thời gian giữa ảnh tham chiếu và ảnh hiện tại là gần; còn trọng số lớn hơn cho khoảng cách thời gian dài.

*** Việc phân chia các Macro-Block** : Các Slice B dùng sự chia nhỏ Macro-Block tương tự như với các Slice P. Tức là với Macro-Block 16×16 có thể phân chia thành các phần với kích thước như: 16×16 , 16×8 , 8×16 và 8×8 ; còn với khối có kích thước 8×8 thì có thể chia thành các phần với kích thước như: 8×8 , 8×4 , 4×8 và 4×4 . Ngoài ra với các mode dự đoán khác nhau (như dự đoán trực tiếp hay dự đoán hai chiều) có thể dùng các kích thước ảnh khác nhau.. Nếu không có tín hiệu dự đoán được truyền đi cho mode Macro-Block trực tiếp, nó sẽ được dùng đến mode SKIP Slice Block và có thể được mã hoá rất hiệu quả, tương tự như với mode SKIP trong các Slice P.

Các Vector bù chuyển động cũng tương tự như đối với Slice P với những biến đổi thích hợp vì các khối bên cạnh có thể được mã hoá khi dùng mode dự đoán khác nhau.

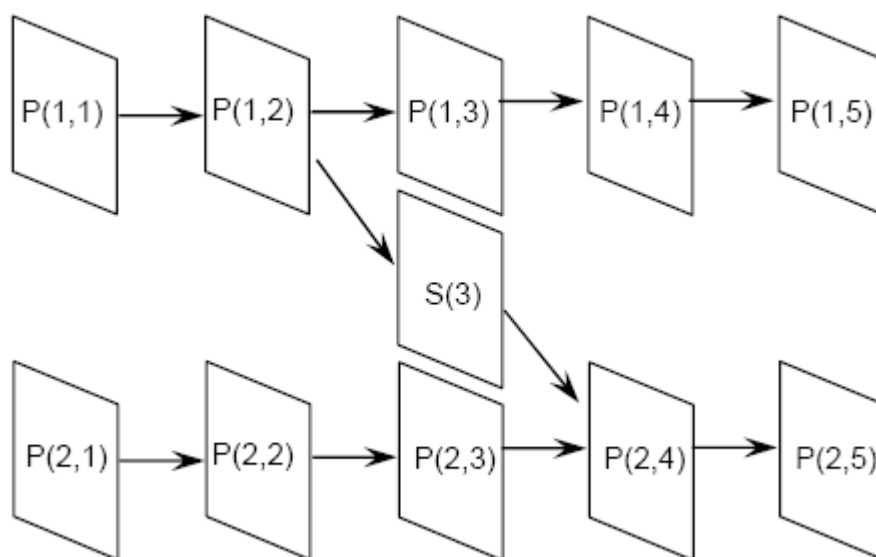
5. Các mảng SP và SI.

Trong các tiêu chuẩn trước đó, việc chuyển đổi hoàn toàn giữa các dòng bit (bitstream) thì có thể chỉ trong một ảnh I. Việc khôi phục lại các ảnh I tại các khoảng thời gian cố định cho phép truy cập ngẫu nhiên hay hiển thị lại nhanh. Tuy nhiên, mặt hạn chế của ảnh I là nó yêu cầu số lượng bit lớn, vì vậy các ảnh I không khai thác được sự dư thừa về mặt thời gian. H.264 giới thiệu các mảng chuyển đổi là SP (Switching P) và SI (Switching I) để chuyển đổi giữa các dòng bit được mã hoá ở các tốc độ bit khác nhau. Các tín hiệu dự đoán Inter của các dòng bit cho một Frame SP được chọn được lượng tử hoá trong miền biến đổi, sau đó được đưa vào dải biên độ thô hơn để cho phép mã hoá tốc độ bit thấp tín hiệu khác nhau giữa các dòng bit. Các Frame SI được xác định để thực hiện sự thích nghi hoàn thiện cho các Frame SP trong trường hợp mà dự đoán Inter không thể được sử dụng do các lỗi truyền dẫn.

Trong hình 22 chỉ ra một ví dụ về cách sử dụng các ảnh SP chuyển đổi giữa các dòng bit khác nhau như thế nào. Chúng ta thừa nhận rằng có 2 dòng bit là $P(1,k)$ và $P(2,3)$ tương ứng với cùng một chuỗi bit được mã hoá tại các tốc độ bit khác nhau. Trong khoảng thời gian mỗi dòng bit được mã hoá, các ảnh SP được đặt tại các vị trí chuyển đổi từ một dòng bit đến dòng bit khác sẽ được cho phép.

Trong trường hợp chuyển đổi từ dòng bit $P(1,3)$ ở trên đến $P(2,3)$, một ảnh SP là $S(3)$ cho phép đưa ra ảnh $P(2,3)$ đã giải mã bằng việc sử dụng $P(1,2)$ trong dòng bit khác, thậm chí là bao gồm các bù chuyển động.

Mảng SI được sử dụng tương tự như mảng SP, nhưng dự đoán được định dạng bằng việc sử dụng các mode dự đoán trong ảnh 4×4 từ các mẫu đã giải mã trước đó của ảnh đã được khôi phục lại.



Hình 22: chuyển đổi sử dụng ảnh SP

3.3.3.3. Xác định Vector chuyển động (Motion Estimation)

Trong khi thực hiện DCT và lượng tử hóa phục vụ cho nén không gian bên ngoài của một Frame, xác định vector chuyển động được sử dụng để nén sự dư thừa về thời gian, ví dụ như trong miền thời gian qua hai Frame liên tiếp.

Để làm một cảnh tiêu biểu, từ một phim truyện ví dụ từ 25 – 30 frame liên tiếp được hiển thị trong mỗi giây bằng tivi hay máy chiếu phim. Tuy nhiên, không may là trong một vài giây bất kỳ, một chuỗi các hình ảnh lưu lại giống nhau. Phong nền có thể không thay đổi tất cả, các đặc tính còn lại cũng tương tự và vì vậy các Frame liên tiếp rất giống nhau.

Việc xác định vector chuyển động thực hiện phép phân tích giữa hai frame liên tiếp và xác định những miền của hình ảnh có thay đổi hay chuyển động giữa các ảnh. Trong nhiều trường hợp một miền lưu lại chính xác vì nó đã ở trong các ảnh dự đoán vì thế nó đủ để cho từ mã hoá đến giải mã để hiển thị miền này khi nó đã ở trong ảnh dự đoán. Nếu miền di chuyển theo một hướng xác định, thuật toán xác định vector chuyển động điều khiển quá trình giải mã để sử dụng các mảnh (piece) của hình ảnh như trong ảnh dự đoán, nhưng để di chuyển nó một lượng nhất định trong một hướng đã được xác định. Trên thực tế điều này được hoàn thành bằng việc gửi vector chuyển động trong dòng bit MPEG -4. Vector này sẽ chỉ dẫn để giải mã trong việc chọn các phần thích hợp của các ảnh giải mã trước đó được sử dụng trong sự thiết lập lại của một dòng khung hình. Nó được làm sạch vì thế tỷ lệ nén rất cao. Trên thực tế, một vài loại nội dung có thể được nén với phạm vi lớn dẫn tới thiếu các hành động trong hình ảnh. Một ví dụ là “lời mở

đầu” của nội dung , như là một người phát thanh viên, cái đó tạo ra một dòng MPEG -4 nén rất chặt.

Đó có thể là một điều mong đợi, việc xác định vector chuyển động là một hàm ước lượng rất cao. Việc tìm kiếm qua một hình ảnh cho tất cả các vật có thể (các vùng) thì có thể thay đổi bất kỳ vị trí sẽ cần đến nhiều sự tính toán. Tuy nhiên chỉ có 1/15 đến 1/30 giây để làm điều này trước khi khung tiếp theo đến để xử lý.

Để hiểu tốt hơn các cách khác nhau của việc thực hiện xác định vector chuyển động, hãy quan sát kỹ thao tác bàn tay. Về cơ bản, việc xác định vector chuyển động cần chia hình ảnh thành các khối nhỏ hơn và lấy mỗi khối tại các thời gian và vị trí khác nhau trên phía trên cùng của ảnh trước đó để xác định nếu đó là cái khớp nhau. Việc so khớp có thể được làm bằng việc tính toán sự khác nhau giữa mỗi điểm ảnh (pixel) trong khối ảnh và việc so khớp vị trí trong các ảnh trước đó.

Như vậy, một con số (giá trị SAD “Summation of Absolute Difference” = tổng sự khác nhau tuyệt đối) thu được cái đó biểu thị “how well- tốt như thế nào” đó là khối riêng biệt thích hợp với vị trí nào đó trong ảnh trước đó. Nếu giá trị SAD là không (zero) nghĩa là mỗi điểm ảnh (pixel) là chính xác trong cùng một vị trí như trong ảnh trước đó, cho nên vị trí mới cho khối đó tìm được. Nếu không một vị trí nào là khớp hoàn toàn, thì thuật toán có hai lựa chọn :

- Đầu tiên : nó kết luận rằng tất cả sự khác nhau là quá lớn nghĩa là một câu hỏi đặt ra là một thực thể mới không tồn tại trong ảnh trước đó hay nó di chuyển quá xa so với vị trí trước hay vị trí sau đó.
- Trong trường hợp thứ hai, giá trị SAD nhỏ khác không (nonezero) nó chấp nhận sự thoả mãn nhất tuy nhiên sự thoả mãn đó không hoàn toàn.

Trong tiêu chuẩn MPEG -4, với mỗi 16 x 16 điểm ảnh trong một khối thì được thoả mãn cho tất cả các vị trí trong một vùng tìm kiếm, phạm vi đó xa bao nhiêu từ vị trí gốc một khối (block) có thể di chuyển giữa hai khung (frame). Diện tích vùng tìm kiếm là +/- 16 điểm ảnh. Khi đó, với mỗi vị trí tìm kiếm là 256(=16 x 16) điểm ảnh của mọi block thì được so sánh với ảnh trước đó. Với định dạng CIF độ phân giải là (352 x 288) tại 30 ảnh (frame) trên một giây (frame/s) cho dòng video, con số so sánh như sau:

- Mỗi Macro-Block có: $16 \times 16 = 256$ pixel;
- Mỗi Macro-Block được thoả mãn trong mỗi vị trí tìm kiếm là: $16 \times 16 = 256$;
- Mỗi khung hình (frame) bao gồm 396 block;

- 30 khung hình được xử lý trong mỗi giây.

Như vậy, tổng sự khác nhau (cả cộng và trừ) trong một giây kết quả là:

$$256 \times 256 \times 396 \times 30 = 778.567.680.$$

Ví dụ bộ xử lý đa năng có khả năng thực hiện cả việc cộng và trừ trong một chu kỳ đồng hồ, bộ xử lý có thể cần chạy với tần số là 779 MHz để thực hiện xác định vector chuyển động. Có hai giải pháp tiêu biểu như sau:

1. Nhẹ nhàng hơn là dùng thuật toán tối ưu nhất.
2. Có sự trợ giúp của phần cứng là xác định vector chuyển động.

3.3.3.4. Nén video

1. Nén theo miền thời gian.

Những dư thừa về mặt thời gian là những hình ảnh giống nhau lặp đi lặp lại từ khung này sang khung khác, ví dụ như khung nền không chuyển động của một chương trình đối thoại trên truyền hình. Vì vậy, để giảm bớt độ dư thừa này ta phải tiến hành nén theo miền thời gian.

Khi bộ mã hoá đang hoạt động ở chế độ “giữa khối”, khối này sẽ phải qua công đoạn hiệu chỉnh chuyển động. Quá trình này sẽ phát hiện ra bất kỳ chuyển động nào diễn ra giữa khối đó và một khối tương ứng ở một hoặc hơn một ảnh tham chiếu đã được lưu trữ từ trước, sau đó tạo ra khối “chênh lệch” hoặc “lỗi”. thao tác này làm giảm bớt dữ liệu trong mỗi Block một cách hiệu quả do chỉ phải biểu diễn chuyển động của nó mà thôi. Tiếp đến là công đoạn biến đổi Cosine rồi rạc DCT để bắt đầu nén theo miền không gian. Khi bộ mã hoá hoạt động ở chế độ “trong khối”, khối này sẽ bỏ qua công đoạn hiệu chỉnh chuyển động và tới thẳng công đoạn DCT.

2. Nén theo miền không gian.

Dư thừa về mặt không gian là các khối có chứa các điểm ảnh tương tự nhau hoặc giống hệt nhau. Trong nhiều trường hợp các điểm ảnh thường không thay đổi nhiều. Như vậy có nghĩa là tần số thay đổi giá trị điểm ảnh trong khối này là rất thấp. Những khối như thế được gọi là khối có tần số không gian thấp. Bộ lập mã lợi dụng đặc điểm này bằng cách chuyển đổi các giá trị điểm ảnh của khối thành các thông tin tần số trong công đoạn biến đổi Cosine rồi rạc.

*** Biến đổi Cosine rời rạc (DCT)**

Biến đổi Cosine là một hàm mà làm biến đổi dữ liệu hình ảnh được thể hiện trong hệ tọa độ X-Y sang miền tần số. Công đoạn DCT biến đổi các giá trị điểm ảnh của khối thành một mạng lưới gồm các hệ số ngang dọc đặt trong không gian tần số. Khi khối ban đầu có tần số không gian thấp, DCT sẽ tập hợp năng lượng tần số vào góc tần số thấp của mạng lưới. Nhờ vậy, những hệ số tần số thấp ở góc đó sẽ có giá trị cao hơn. Một số lượng lớn các hệ số khác còn lại trên mạng lưới đều là các hệ số có tần số cao, năng lượng thấp và có giá trị thấp. Tại đây, hệ số DC và một vài hệ số tần số thấp sẽ hàm chứa phần lớn thông tin được mô tả trong khối ban đầu. Có nghĩa là bộ lập giải mã có thể loại bỏ phần lớn hệ số tần số cao còn lại mà không làm giảm chất lượng hình ảnh của khối. Bộ lập mã chuẩn bị các hệ số cho công đoạn này bằng cách quét chéo mạng lưới theo đường zig-zag, bắt đầu từ hệ số DC và qua vị trí của hệ số ngang dọc tăng dần. Do vậy nó tạo ra được một chuỗi hệ số được sắp xếp theo tần số.

Ưu điểm độc đáo hơn của H264/MPEG Part 10 là ở chỗ nó sử dụng biến đổi không gian nguyên(gần giống với DCT) đối với các khối 4 x 4 điểm ảnh. Ta xét biến đổi nguyên cho cấu trúc 4 x 4 như sau:

+ Cho hệ số DCT của khối dữ liệu đầu vào 4 x 4 là F, công thức chính xác được cho là :

$$X=HFH^T \quad (1)$$

Trong đó ma trận H là

$$H = \begin{bmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{bmatrix}$$

H^T là ma trận chuyển vị của ma trận H.

Các giá trị a,b, c là :

$$a = \frac{1}{2}, b = \sqrt{\frac{1}{2}} \cos\left(\frac{\pi}{8}\right), c = \sqrt{\frac{1}{2}} \cos\left(\frac{3\pi}{8}\right)$$

Tuy nhiên để cho đơn giản ta có thể cho giá trị $=0.5$ và để đảm bảo tính trực giao thì các giá trị của a, b được lựa chọn như sau:

$$a = \frac{1}{2}, b = \sqrt{\frac{2}{5}}, c = \frac{1}{2}$$

Do phép nhân trong quá trình biến đổi được tránh do phải kết hợp với quá trình lượng tử hoá, vì vậy phương trình (1) được viết thành :

$$X = \overline{H} F \overline{H}^T \otimes SF$$

Trong đó :

+ Ma trận H được cho là :

$$\overline{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

+ Ma trận SF là :

$$SF = \begin{bmatrix} a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \\ a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \end{bmatrix}$$

Ký hiệu \otimes biểu diễn phép nhân tích trực tiếp phần tử với phần tử (*the element by element multiplication*) của ma trận tương ứng.

So sánh với MPEG-2, thì ta thấy trong từng Slice sẽ có 1 khối đầu tiên chứa giá trị hệ số DC và các khối tiếp theo sẽ có những giá trị sẽ có những giá trị khác nhau, những khối này được gán thêm giá trị dự đoán khối, tạo thành chuỗi dòng khối. Còn với MPEG-4 có thể dự đoán xa hơn. Chỉ cần một khối cũng có thể dự đoán toàn bộ hàng trên cùng của hệ số hoặc có thể dự đoán toàn bộ cột bên trái của hệ số từ 1 khối đầu tiên.

* Lượng tử hoá

Quá trình lượng tử hoá là quá trình biến đổi có mất thông tin, làm giảm bớt số lượng bit cần thiết để biểu diễn các hệ số. Dựa trên một hệ số tỷ lệ xích (có thể điều chỉnh bởi bộ mã hoá), bộ lượng tử hoá sẽ cân đối tất cả các giá trị hệ số. Do phần lớn các hệ số đi ra từ DCT đều mang năng lượng cao nhưng giá trị thấp nên bộ lượng tử hoá bắt đầu bằng một số giá trị cao ở đầu chuỗi, theo sau là một hàng dài các hệ số đã được

lượng tử hoá về 0. Bộ lập mã Entropy có thể theo dõi số lượng các giá trị 0 liên tiếp trong một chuỗi mà không cần mã hoá chúng, nhờ vậy giảm bớt được khối lượng dữ liệu trong mỗi chuỗi. Để lượng tử hóa các hệ số biến đổi, H264/MPEG Part 10 dùng phương pháp lượng tử hóa vô hướng. Các bộ lượng tử hoá được lựa chọn cho mỗi Macro-Block là dựa vào các tham số lượng tử hoá QP (Quantization Parameter). Các bộ lượng tử hoá được sắp xếp sao cho có sự tăng khoảng 12.5% trong kích thước bước lượng tử hoá khi QP tăng một đơn vị. Nhìn chung các hệ số biến đổi được lượng tử hoá của khối được quét zig-zag và được truyền đi nhờ dùng phương pháp mã hoá Entropy.

*** Mã hoá Entropy**

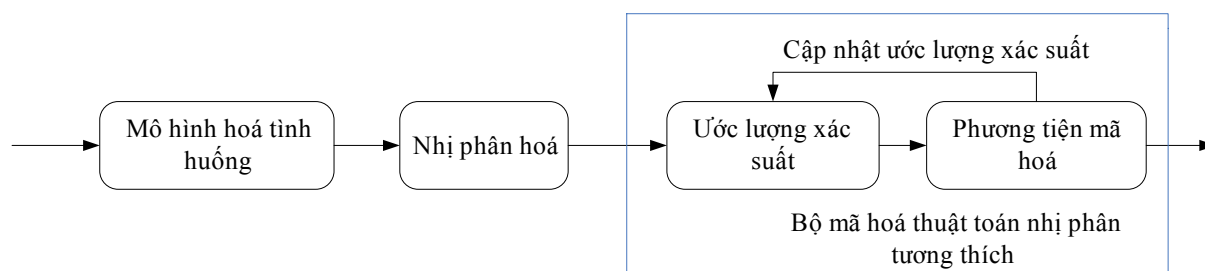
Mã hoá Entropy trong các tiêu chuẩn trước đó như MPEG -1,2,4, H.261, và H.263 thì cơ bản là trên các bảng cố định mã hoá biến đổi theo chiều dài (VLC). Các tiêu chuẩn đó xác định các bộ mã hoá từ là cơ bản trên sự phân bố xác suất của các video chung thay cho mã Huffman chính xác đến các chuỗi video. Tuy nhiên H.264 sử dụng các VLC để mã khớp với một biểu tượng được mã hoá cơ bản trên các đặc trưng của ngữ cảnh. Tất cả các phần tử cú pháp, ngoại trừ các dữ liệu dư thừa, được mã hoá bằng mã Exp-Golomb. Để mã hoá được các dữ liệu dư thừa (các hệ số biến đổi đã lượng tử hoá) thì ta sử dụng phương pháp quét Zig-Zag (xen kẽ nhau) hay quét lần lượt (không xen kẽ hay phân trường). Để mã hoá dữ liệu dư thừa, một phương pháp phức tạp hơn gọi là CAVLC (mã hoá chiều dài biến đổi tương thích theo tình huống) được phát triển. Ngoài ra, CABAC (mã hoá thuật toán nhị phân tương thích theo tình huống) được phát triển trong Main Profile và High Profile, CABAC có khả năng mã hoá tốt hơn nhưng độ phức tạp cao hơn so với CAVLC.

+ Mã hoá chiều dài biến đổi tương thích theo tình huống - Context-based Adaptive Variable Length Coding (CAVLC)

Sau khi biến đổi và lượng tử hoá, xác suất các hệ số là zero hay ± 1 là rất lớn. CAVLC xử lý các hệ số zero và ± 1 theo cách khác nhau với các mức của các hệ số. Tổng số các số zero và ± 1 được mã hoá. Các hệ số khác các mức của chúng được mã hoá.

+ Mã hoá thuật toán nhị phân tương thích theo tình huống – Context –based Adaptive Binary Arithmetic Coding (CABAC)

CABAC dùng thuật toán để mã hoá, để đạt được hiệu quả nén tốt hơn, mô hình có khả năng cho mỗi phần tử biểu tượng được cập nhật như trong hình 28. Quá trình xử lý mã hoá CABAC bao gồm 3 bước cơ bản sau đây:



Hình 23: Sơ đồ khối của CABAC

- Bước 1: Nhị phân hoá; một biểu tượng giá trị của nó không phải là nhị phân (như một hệ số biến đổi hay Vector chuyển động) là bản đồ duy nhất để chuỗi nhị phân trước đó được mã hoá thuật toán. Quá trình xử lý này thì giống như việc xử lý của việc biến đổi một biểu tượng dữ liệu thành mã hoá chiều dài biến đổi, nhưng mã nhị phân được mã hoá tốt hơn bằng bộ mã hoá thuật toán trước đó để truyền.
- Bước 2: phạm vi làm mô hình: một phạm vi mô hình thì có khả năng cho một hay nhiều phần tử của biểu tượng đã nhị phân. Một mô hình có khả năng được lựa chọn như để việc lựa chọn sự tương ứng có thể phụ thuộc vào các phần tử cú pháp đã mã hoá trước đó.
- Bước 3: thuật toán mã hoá nhị phân: Một bộ mã hoá thuật toán mã hoá mỗi phần tử theo mô hình có khả năng lựa chọn cùng với một việc cập nhật mô hình tiếp sau.

3.3.3.5. Bộ lọc tách khối

Một đặc trưng riêng của mã hoá dựa trên cơ sở khối là có thể nhìn thấy các cấu trúc khối. Các mép của khối được cấu trúc lại với độ chính xác kém hơn các phần tử ảnh (pixel) bên trong và nhìn chung dạng khối (blocking) được xem là một trong những nhiều “artifact” dễ nhìn thấy nhất với các phương pháp nén hiện tại. Do nguyên nhân này mà H.264/MPEG-4 Part 10 sử dụng bộ lọc tách khối (Deblocking Filter) để làm giảm hiện tượng tách khối, ngăn chặn việc truyền của tập âm mã hoá được tích lũy. Tại bộ lọc này, cường độ lọc được điều khiển bởi giá trị của nhiều phần tử cấu trúc.

Các chuẩn nén trước đó đã không sử dụng bộ lọc tách khối bởi vì việc bổ sung rất phức tạp, mặt khác việc chia các nhiễu khối có thể được làm giảm bằng việc sử dụng MC chính xác một nửa phần tử ảnh. Một nửa phần tử ảnh thu được bằng cách lọc tuyến tính (bilinear filtering) của các phần tử ảnh nguyên vẹn bên cạnh đã phát huy vai trò làm “nhẵn” của mã hoá tập âm trong miền phần tử ảnh nguyên vẹn.

H.264 sử dụng bộ lọc tách khối để việc thực hiện việc mã hoá cao hơn mặc dù việc thực hiện rất phức tạp. Việc lọc được áp dụng cho các mép của các Block 4 x 4 trong một Macro-Block. Quá trình điều khiển bộ lọc tách khối thành phần chói được thực hiện trên 4 cạnh của mẫu 16 x 16 (16-sample) và quá trình xử lý bộ lọc tách khối cho mỗi thành phần màu được thực hiện trên 2 cạnh của mẫu 8 x 8.

Bộ lọc tách khối được áp dụng tương thích với một vài mức(level) sau:

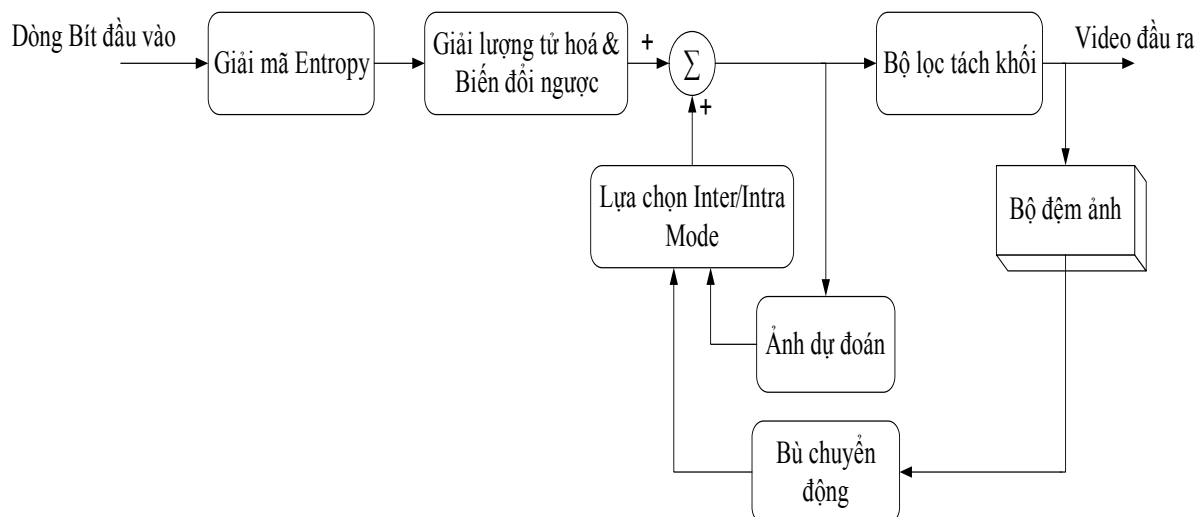
- Mức mảnh (Slice Level): Mặt mạnh của việc lọc toàn bộ có thể được điều chỉnh tới các đặc điểm riêng của chuỗi video.
- Mức cạnh của khối (Block – edge Level): mặt mạnh của việc lọc là phụ thuộc vào việc quyết định dự đoán Inter/Intra, các chuyển động khác nhau, và việc hiển thị của các dư thừa mã hoá trong 2 Block tham gia. Đặc biệt việc lọc mạnh được áp dụng cho các Macro-Block với các đặc điểm rất phẳng để loại bỏ “sự nghiêng nhân tạo (tilting artifacts)”.
- Mức lấy mẫu (Sample Level): các giá trị mẫu và các điểm bắt đầu phụ thuộc vào bộ lượng tử hóa có thể dừng việc lọc cho mỗi mẫu riêng biệt.

Tính khối bị giảm mà không ảnh hưởng nhiều đến độ sắc nét của nội dung, kết quả là đã cải thiện được đáng kể chất lượng của hình ảnh.



Hình 24: Tác dụng của bộ lọc tách khối đối với ảnh được nén nhiều
Trái: không có bộ lọc tách khối; Phải: có bộ lọc tách khối

3.3.4. Kỹ thuật giải mã video



Hình 25: Sơ đồ giải mã Video H264/MPEG -4 Part 10

3.3.4.1. Bù chuyển động

Bù chuyển động thực hiện việc thiết lập lại các khung hình (frame) trên cơ sở các Vector đã nhận được, dữ liệu khung hình delta đã nhận (dữ liệu khác nhau giữa hai khung hình liên tiếp) và hình ảnh đã mã hoá trước đó. Vì vậy, nếu dữ liệu delta được nhận, thì dòng khung hình được thiết lập lại bằng việc cộng dữ liệu khung hình delta với dữ liệu từ khung hình đã được giải mã trước đó trong vị trí riêng đã được chỉ dẫn bằng Vector chuyển động đã nhận được.

3.3.4.2. Khôi phục lỗi (Error Resiliency)

Một nhiệm vụ quan trọng nhất của bộ giải mã là làm thích hợp và khôi phục lại từ các lỗi chắc chắn xảy ra, đặc biệt là khi kết nối qua các liên kết dễ xảy ra lỗi như mạng không dây. Lỗi đàn hồi bao gồm ba loại kỹ thuật khác nhau được sử dụng khi một lỗi được phát hiện. Việc phát hiện lỗi có thể xảy ra trong 3 trường hợp sau:

- Trong trường hợp có một lỗi, lỗi đó thực sự được phát hiện;
- Trong trường hợp có một lỗi, lỗi đó không được phát hiện;
- Trong trường hợp không có lỗi, bộ giải mã biết là không có lỗi;

Nếu lỗi được phát hiện, hệ thống bộ giải mã lỗi đàn hồi cố gắng che giấu lỗi. Các đặc trưng của lỗi đàn hồi có thể được bổ sung trong hệ thống MPEG -4 bằng việc sử dụng

dụng tiêu chuẩn thiết lập các công cụ được cung cấp bằng tiêu chuẩn MPEG -4 để phát hiện lỗi. Cho một hệ thống MPEG -4 tương thích, có khả năng cùng làm việc với hệ thống MPEG -4 khác, nó phải tuân theo bộ công cụ chuẩn hoá phát hiện lỗi. Tuy nhiên, phương pháp mà các bộ công cụ được bổ sung làm tăng nơi cung cấp bộ Codec, theo cách đó, nó có thể làm cho tốt hơn hoặc xấu hơn khi thiết lập các thuật toán trong hệ thống tương thích hoàn toàn với tiêu chuẩn MPEG -4.

1. Đồng bộ lại (Resynchronisation)

Bộ công cụ Đồng bộ lại cố gắng để có thể đồng bộ lại giữa bộ giải mã và dòng bit sau khi một lỗi hay một loạt lỗi được phát hiện. Nhìn chung, dữ liệu giữa điểm đồng bộ trước bị lỗi và điểm đầu tiên nơi mà sự đồng bộ đã được thiết lập lại, thì được loại bỏ. Nếu như phương pháp đồng bộ lại có ảnh hưởng đến việc xác định số lượng dữ liệu đã bị loại bỏ bởi bộ giải mã, khi đó khả năng của các loại công cụ khác nhau để khôi phục dữ liệu hay để che giấu các ảnh hưởng của lỗi thì phải được nâng cấp rất nhiều.

2. Khôi phục dữ liệu (Data Recovery)

Sau khi đồng bộ vừa được thiết lập lại, các công cụ khôi phục dữ liệu cố gắng khôi phục lại dữ liệu thường bị mất. Các công cụ đó không chỉ đơn giản là mã hoá đúng các lỗi, mà kỹ thuật đòi hỏi cao hơn là mã hoá dữ liệu cho một loại lỗi đàn hồi. Ví dụ, một bộ công cụ riêng vừa được xác nhận bằng nhóm Video (Video Group) là mã hoá chiều dài biến đổi ngược RVLC (Reversible Variable Length Codes). Trong phương pháp này, các từ mã hoá chiều dài biến đổi được thiết kế để đọc cả hướng tiến cũng như hướng lùi.

Hình 26 minh hoạ cho ví dụ trên sử dụng một RVLC được chỉ ra trong hình dưới đây. Nhìn chung, trong một tình huống như vậy, nơi mà một sự tăng vọt của các lỗi làm sai lạc một phần của dữ liệu, tất cả dữ liệu giữa hai điểm đồng bộ có thể bị mất. Tuy nhiên, được chỉ ra trong hình 26, một RVLC làm cho một vài dữ liệu được khôi phục. Chú ý đến các thông số QP và HEC trong hình sau thể hiện các trường dành riêng (Field Reserved) trong phần đầu header gói video cho thông số lượng tử hoá và mã hoá mở rộng header, theo thứ tự định sẵn.

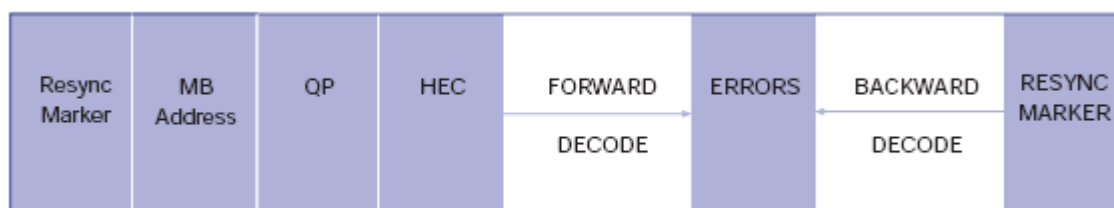


Figure 4: Example of Reversible Variable Length Code

Hình 26: Ví dụ về Mã hoá chiều dài biến đổi ngược

3.3.5 So sánh hiệu quả mã hoá của H264/MPEG Part 10 với các tiêu chuẩn trước đó

Qua quá trình tìm hiểu kỹ thuật mã hoá của H264/MPEG Part 10 và của MPEG-2, ta thấy một số ưu điểm của H264/MPEG Part 10 so với MPEG-2 là:

- Dự đoán hệ số AC/DC mà ở đó các hệ số Macro-Block được dự đoán từ các hệ số trong các Macro-Block bên cạnh và từ một số dữ liệu trong bản thân Macro-Block. Việc giải mã Macro-Block phải thay đổi khác với thuật toán dòng bit đơn giản trước đó để đưa vào một số tính toán cho mỗi hệ số.
- Các công cụ toán học mới như CAVLC và CABAC cho sự cải thiện đáng kể trong mã hoá Entropy.
- Trong khi bù chuyển động trong MPEG-2 Part 2 được hạn chế đến nội suy hai chiều $\frac{1}{2}$ pixel thì H264/MPEG Part 10 cho phép các Vector chuyển động chính xác đến $\frac{1}{4}$ pixel và sau đó dùng nội suy nhiều chiều (Bi-cubic). Nội suy nhiều chiều tạo ra sự thích hợp hơn cho Macro-Block, do vậy giảm năng lượng được lưu trong ảnh lỗi, làm giảm số bit cần phải mã hoá. Tuy nhiên nội suy nhiều chiều đòi hỏi nhiều thuật toán hơn, sự phức tạp thực hiện bù chuyển động cao hơn.
- Việc dùng bộ lọc tách khối ở cả hai phía bộ mã hoá và bộ giải mã làm giảm sự không liên tục ở các biên của khối do các hệ số chất lượng khác nhau được sử dụng cho các khối cạnh nhau gây ra. Điều này làm giảm nhiễu khối thường thấy trong mã hoá MPEG-2 tốc độ bit thấp vì nó nằm trong vùng bù chuyển động, bộ mã hoá và bộ giải mã vẫn còn trong sự đồng bộ. Công cụ này có ảnh hưởng đến tính phức tạp của bộ mã hoá và bộ giải mã vì số biên khối và cũng vì thực tế là bộ lọc tách khối cũng không thể được thực hiện như một module riêng biệt.

- Kích thước khối thay đổi từ 16 x 16 xuống còn 4 x 4. Khi kích thước khối giảm cung cấp độ lợi mã hoá mà không có sự tăng đáng kể tính phức tạp.
- Các Header tương đối lớn trong MPEG-2(trong Sequence, Picture, Slice) để biểu diễn một OverHead cố định trong dòng dữ liệu. OverHead này là không đáng kể trong tốc độ 6 Mb/s nhưng ở tốc độ vài trăm Kb/s thì chúng dễ nhận thấy. Vì vậy H264/MPEG Part 10 thực hiện các Header hiệu quả hơn.

Qua quá trình thực nghiệm đã cho rằng sự tiết kiệm tốc độ bit trung bình của H264/MPEG Part 10 so với MPEG-2 là khoảng 65%. Như vậy hiệu quả của nén H264/MPEG Part 10 tăng lên đáng kể.

Hiệu quả nén tăng của H264/MPEG Part 10 tạo ra các phạm vi ứng dụng và các cơ hội kinh doanh mới

KẾT LUẬN

Khoá luận “Các chuẩn nén và ứng dụng truyền Video trên mạng Internet” đã đạt được một số kết quả sau:

Nêu lên được thế nào là truyền hình trên mạng Internet và cách thức truy cập để xem được các chương trình truyền hình.

Tìm hiểu được nguyên tắc chung trong việc nén tín hiệu video của nhóm MPEG là nén trong ảnh và nén liên ảnh kết hợp với dự đoán bù chuyển động- đây là sự khác biệt so với các chuẩn nén JPEG, M-JPEG... trước đó. Và trong tập hợp các chuẩn nén MPEG, mỗi chuẩn nén khác nhau có một nhiệm vụ khác nhau được đặt ra.

Với MPEG-1 nhiệm vụ là mã hoá âm thanh và hình ảnh để lưu trữ trong các thiết bị lưu trữ như: CD-ROM, đĩa quang... sử dụng định dạng trung gian cho nguồn tín hiệu là SIF để phù hợp với mỗi hệ thống quét truyền hình khác nhau hiện nay.

MPEG -2 kế thừa tất cả các tiêu chuẩn của MPEG -1 và mục đích nhằm hỗ trợ cho việc truyền Video số, MPEG -2 chia ảnh thành các Macro-Block và sử dụng phương pháp nén trong ảnh với ảnh I và nén liên ảnh kết hợp dự đoán bù chuyển động với các ảnh P và ảnh B. Tỷ số nén của MPEG -2 tốt hơn nhiều so với MPEG -1.

MPEG -4 với nhiệm vụ là nhằm phát triển các chuẩn xử lý, mã hoá và hiển thị ảnh động, audio và các tổ hợp của chúng. MPEG -4 khác so với MPEG -2 là trong một khung hình thì nó không mã hoá toàn bộ khung hình mà nó miêu tả từng đối tượng riêng rẽ và sau đó mới mã hoá từng đối tượng đó. MPEG -4 chia các Macro-Block thành các Block nhỏ hơn. Trong khi bù chuyển động trong MPEG-2 Part 2 được hạn chế đến nội suy hai chiều $\frac{1}{2}$ pixel thì H264/MPEG Part 10 cho phép các Vector chuyển động chính xác đến $\frac{1}{4}$ pixel và sau đó dùng nội suy nhiều chiều(Bi-cubic). Và còn nhiều điểm ưu việt khác nữa. Do đó hiệu quả nén của chuẩn nén MPEG -4 tốt hơn so với MPEG -2.

MPEG -7 là một chuẩn dùng để mô tả các nội dung Multimedia, chứ không phải là một chuẩn cho nén và mã hoá audio/ảnh động như MPEG-1, MPEG-2 hay MPEG-4. MPEG-7 sử dụng ngôn ngữ đánh dấu mở rộng XML(Extensible Markup Language) để lưu trữ các siêu dữ liệu Metadata, đính kèm timecode để gắn thẻ cho các sự kiện, hay đồng bộ các dữ liệu.

Với nhiều tính ưu việt như trên, một hướng đặt ra là: liệu MPEG -4 có thể sử dụng cho việc lưu trữ như MPEG -1 hay dùng trong công nghệ truyền hình số như MPEG -2. Thực tế cho thấy, ngoài các ứng dụng trên môi trường mạng nói chung và truyền hình tương tác nói riêng – theo APB(04/2004)- các Đài Truyền hình Nhật Bản như NHK, TBS, NTV TV Asahi và TV Tokyo đã chính thức chấp nhận đưa chuẩn

H264/MPEG Part 10 cho phát sóng số mặt đất, cho thu di động, đây cũng là một khả năng trong những khả năng còn tiềm ẩn trong chuẩn mã hoá và trình diễn video của H264/MPEG Part 10. Do vậy, đề tài này có thể nghiên cứu bước tiếp theo là nghiên cứu ứng dụng H264/MPEG Part 10 cho truyền hình số và cho thu di động hay có thể cải tạo chất lượng truyền hình Internet để đạt được chất lượng hình ảnh tốt như truyền hình số hiện nay.

TÀI LIỆU THAM KHẢO

- [1] Đinh Quốc Tuấn - Đề tài nghiên cứu khoa học cấp ngành “*Truyền hình trên mạng Internet và giải pháp cung cấp chương trình VTV qua mạng Internet*”, Trung tâm tin học và Đo lường – ĐTHVN, Hà Nội – 7/2004.
- [2] Đỗ Hoàng Tiến, Vũ Đức Lý - *Truyền hình số*, Nhà xuất bản Khoa Học và Kỹ Thuật, Hà Nội – 2001.
- [3] Ngô Thái Trị - *Truyền hình số*, Hà Nội – 2002.
- [4] Nguyễn Minh Hồng – Chuyên đề “*Chuẩn nén MPEG -4 và khả năng ứng dụng trong truyền hình*”, Trung tâm Tin học và Đo lường – ĐTHVN, Hà Nội-1/2001.
- [5] Nội san Khoa học Kỹ thuật Truyền hình – các số trong các năm từ 2001-2005.
- [6] *H.264 & IPTV Over DSL – White Paper On H264/MPEG-4* -2004, <http://www.envivio.com>.
- [7] Hantro Products Oy, Oulu, Finland “ *MPEG4 Codec Overview* ” 1 ISO/IEC JTC1/SC29/WG11 N4030 March 2001- <http://www.hantro.com>
- [8] José M. Martínez (UAM-GTI, ES) ,“*MPEG-7 Overview*“- ISO/IEC JTC1/SC29/WG11N5525 Pattaya, March 2003 - <http://www.mp7c.org/>
- [9] Ralf Schäfer, Thomas Wiegand and Heiko Schwarz - “*The emerging H.264/AVC Standard* “- EBU TECHNICAL REVIEW – January 2003
- [10] Soon-kak Kwon, A. Tamhankar, K.R. Rao - “*Overview of H.264 / MPEG-4 Part 10*” <http://www.mpeg.org>.
- [11] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra “*Overview of the H.264 / AVC Video Coding Standard*” - IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, JULY 2003.
- [12] <http://www.vtc.com.vn>., <http://www.vdconline.com>