HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

XỬ LÝ ÂM THANH, HÌNH ẢNH

(Dùng cho sinh viên hệ đào tạo đại học từ xa)

Lưu hành nội bộ

HÀ NỘI - 2007

XỬ LÝ ÂM THANH, HÌNH ẢNH

Biên soạn : **TS. NGUYỄN THANH BÌNH**

THS. VÕ NGUYỄN QUỐC BẢO

LỜI NÓI ĐẦU

Tài liệu hướng dẫn học tập môn "Xử lý âm thanh và hình ảnh" dành cho khối đào tạo từ xa chuyên ngành điện tử viễn thông. Tài liệu này sẽ giới thiệu những kiến thức cơ bản về xử lý âm thanh và hình ảnh. Đặc biệt, tác giả chú trọng tới vấn đề xử lý tín hiệu ứng dụng trong mạng viễn thông: đó là các phương pháp nén tín hiệu, lưu trữ, các tiêu chuẩn nén tín hiệu âm thanh và hình ảnh. Những kiến thức được trình bày trong tài liệu sẽ giúp học viên tiếp cận nhanh với các vấn đề thực tiễn thường gặp trong mạng viễn thông.

Vì khối lượng kiến thức trong lĩnh vực xử lý âm thanh cũng như hình ảnh rất lớn, và với quỹ thời gian quá eo hẹp dành cho biên soạn, tài liệu hướng dẫn này chưa thâu tóm được toàn bộ kiến thức cần có về lĩnh vực xử lý âm thanh và hình ảnh. Để tìm hiểu về một số vấn đề có trong đề cương môn học đòi hỏi học viên phải nghiên cứu thêm trong số sách tham khảo được tác giả đề cập tới trong phần cuối của tài liệu này.

Nội dung cuốn sách được chia làm hai chương:

- Chương 1: Kỹ thuật xử lý âm thanh
- Chương 2: Kỹ thuật xử lý hình ảnh.

Để có thể học tốt môn này, sinh viên cần phải có kiến thức cơ bản về xử lý tín hiệu số. Các kiến thức này các bạn có thể tìm hiểu trong cuốn “Xử lý tín hiệu số” dành cho sinh viên Đại học từ xa của Học viện.

Đây là lần biên soạn đầu tiên, chắc chắn tài liệu còn nhiều sơ sót, rất mong các bạn đọc trong quá trình học tập và các thày cô giảng dạy môn học này đóng góp các ý kiến xây dựng. Trong thời gian gần nhất, tác giả sẽ cố gắng cập nhập, bổ xung thêm để tài liệu hướng dẫn được hoàn chỉnh hơn.

Mọi ý kiến đóng góp đề nghị gửi về theo địa chỉ email**: binhntptit@yahoo.com**

Tp. Hồ Chí Minh 19/05/2007

Nhóm biên soạn

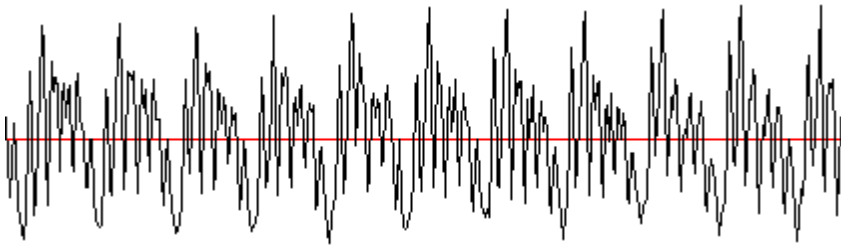
CHƯƠNG 1 KỸ THUẬT XỬ LÝ ÂM THANH

1.1 TỔNG QUAN VỀ XỬ LÝ ÂM THANH

1.1.1 Giới thiệu sơ lược về âm thanh & hệ thống xử lý âm thanh

1.1.1.1 Đặc tính của âm thanh tương tự [1]

Mục đích của lời nói là dùng để truyền đạt thông tin. Có rất nhiều cách mô tả đặc điểm của việc truyền đạt thông tin. Dựa vào lý thuyết thông tin, lời nói có thể được đại diện bởi thuật ngữ là *nội dung thông điệp,* hoặc là *thông tin.* Một cách khác để biểu thị lời nói là *tín hiệu mang nội dung thông điệp,* như là *dạng sóng âm thanh.*



Hình 1.1 Dạng sóng của tín hiệu ghi nhận được từ âm thanh của người

Kỹ thuật đầu tiên dùng trong việc ghi âm sử dụng các thông số về cơ, điện cũng như trường có thể làm nên nhiều cách thức ghi âm ứng với các loại áp suất không khí khác nhau. Điện áp đến từ một microphone là tín hiệu tương tự của áp suất không khí (hoặc đôi khi là vận tốc). Dù được phân tích bằng cách thức nào, thì các phương pháp khi so sánh với nhau phải dùng một tỉ lệ thời gian.

Trong khi các thiết bị tương tự hiện đại trông có vẻ xử lý âm thanh tốt hơn những thiết bị cổ điển, các tiêu chuẩn xử lý thì hầu như không có gì thay đổi, mặc dù công nghệ có vẻ xử lý tốt hơn. Trong hệ thống xử lý âm thanh tương tự, thông tin được truyền đạt bằng thông số liên tục biến thiên vô hạn.

Hệ thống xử lý âm thanh số lý tưởng có những tính năng tương tự như hệ thống xử lý âm thanh tương tự lý tưởng: cả hai hoạt động một cách "trong suốt" và tạo lại dạng sóng ban đầu không lỗi. Tuy nhiên, trong thế giới thực, các điều kiện lý tưởng rất hiếm tồn tại, cho nên hai loại hệ thống xử lý âm thanh hoạt động sẽ khác nhau trong thực tế. Tín hiệu số sẽ truyền trong khoảng cách ngắn hơn tín hiệu tương tự và với chi phí thấp hơn. Trong giáo trình này, tập trung đề cập đến hệ thống số xử lý âm thanh.

Thông tin dùng để truyền đạt của âm thoại về bản chất có tính rời rạc [2], và nó có thể được biểu diễn bởi một chuỗi ghép gồm nhiều phần tử từ một tập hữu hạn các ký hiệu (symbol). Các ký hiệu từ mỗi âm thanh có thể được phân loại thành các âm vị (phoneme). Mỗi ngôn ngữ có các tập âm vị khác nhau, được đặc trưng bởi các con số có giá trị từ 30 đến 50. Ví dụ như tiếng Anh được biểu diễn bởi một tập khoảng 42 âm vị.

Tín hiệu thoại được truyền với tốc độ như thế nào? Đối với tín hiệu âm thoại nguyên thủy chưa qua hiệu chỉnh thì tốc độ truyền ước lượng có thể tính được bằng cách lưu ý giới hạn vật lý của việc nói lưu loát của người nói tạo ra âm thanh thoại là khoảng 10 âm vị trong một giây. Mỗi

một âm vị được biểu diễn bởi một số nhị phân, như vậy một mã gồm 6 bit có thể biểu diễn được tất cả các âm vị của tiếng Anh. Với tốc độ truyền trung bình 10 âm vị/giây, và không quan tâm đến vấn đề luyến âm giữa các âm vị kề nhau, ta có thể ước lượng được tốc độ truyền trunh bình của âm thoại khoảng 60bit/giây.

Trong hệ thống truyền âm thoại, tín hiệu thoại được truyền lưu trữ và xử lý theo nhiều cách thức khác nhau. Tuy nhiên đối với mọi loại hệ thống xử lý âm thanh thì có hai điều cần quan tâm chung là:

1. Việc duy trì nội dung của thông điệp trong tín hiệu thoại
2. Việc biểu diễn tín hiệu thoại phải đạt được mục tiêu tiện lợi cho việc truyền tin hoặc lưu trữ, hoặc ở dạng linh động cho việc hiệu chỉnh tín hiệu thoại sao cho không làm giảm nghiêm trọng nội dung của thông điệp thoại.

Việc biểu diễn tín hiệu thoại phải đảm bảo việc các nội dung thông tin có thể được dễ dàng trích ra bởi người nghe, hoặc bởi các thiết bị phân tích một cách tự động.

1.1.1.2 Khái niệm tín hiệu

Là đại lượng vật lý biến thiên theo thời gian, theo không gian, theo một hoặc nhiều biến độc lập khác, ví dụ như:

- Âm thanh, tiếng nói: dao động sóng theo thời gian (t)
- Hình ảnh: cường độ sáng theo không gian (x, y, z)
- Địa chấn: chấn động địa lý theo thời gian

Biểu diễn toán học của tín hiệu: hàm theo biến độc lập

Ví dụ:

- $u(t) = 2t^2 - 5$
- $f(x, y) = x^2 - 2xy - 6y^2$

Thông thường các tín hiệu tự nhiên không biểu diễn được bởi một hàm sơ cấp, cho nên trong tính toán, người ta thường dùng hàm xấp xỉ cho các tín hiệu tự nhiên.

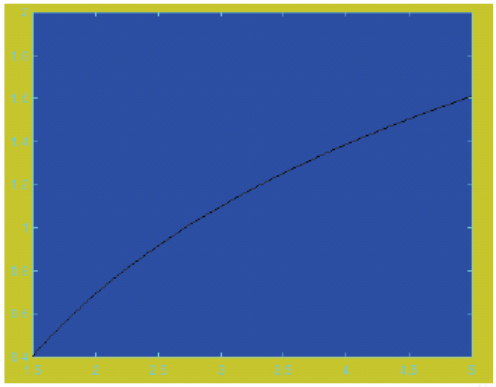
Hệ thống: là thiết bị vật lý, thiết bị sinh học, hoặc chương trình thực hiện các phép toán trên tín hiệu nhằm biến đổi tín hiệu, rút trích thông tin, … Việc thực hiện phép toán còn được gọi là xử lý tín hiệu.

1.1.1.3 Phân loại tín hiệu:

Tín hiệu đa kênh: gồm nhiều tín hiệu thành phần ,cùng chung mô tả một đối tượng nào đó (thường được biểu diễn dưới dạng vector, ví dụ như tín hiệu điện tim (ECG-ElectroCardioGram) , tín hiệu điện não (EEG – ElectroEncephaloGram), tín hiệu ảnh màu RGB.

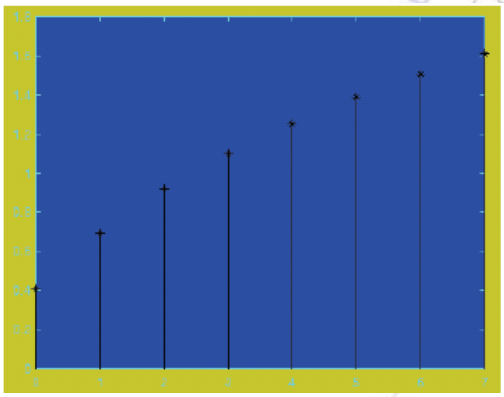
Tín hiệu đa chiều: biến thiên theo nhiều hơn một biến độc lập, ví dụ như tín hiệu hình ảnh, tín hiệu tivi trắng đen.

Tín hiệu liên tục theo thời gian: là tín hiệu được định nghĩa tại mọi điểm trong đoạn thời gian [a,b], ký hiệu $x(t)$.



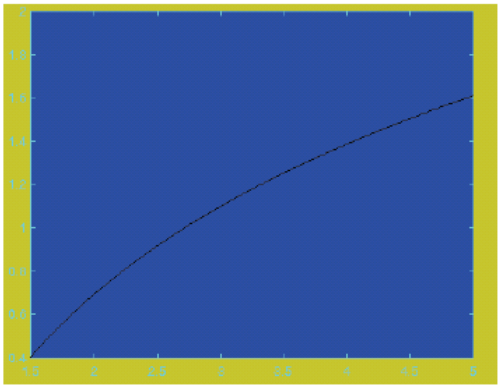
Hình 1.2 Tín hiệu liên tục theo thời gian

Tín hiệu rời rạc thời gian: là tín hiệu chỉ được định nghĩa tại những thời điểm rời rạc khác nhau, ký hiệu $x(n)$.



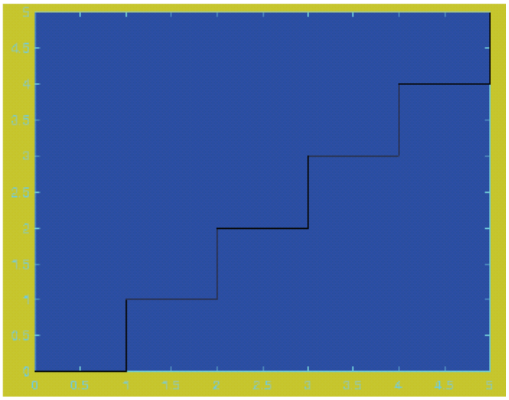
Hình 1.3 Tín hiệu rời rạc theo thời gian

Tín hiệu liên tục giá trị: là tín hiệu có thể nhận trị bất kỳ trong đoạn $[Y_{\min}, Y_{\max}]$, ví dụ tín hiệu tương tự (analog).



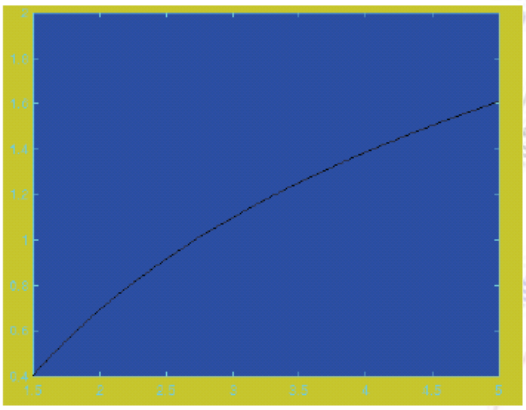
Hình 1.4 Tín hiệu liên tục giá trị

Tín hiệu rời rạc giá trị: tín hiệu chỉ nhận trị trong một tập trị rời rạc định trước (tín hiệu số).



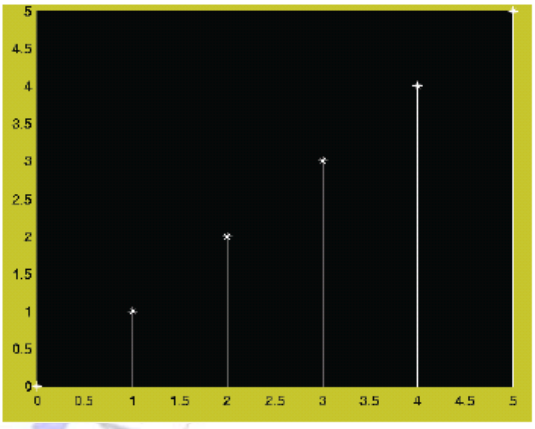
Hình 1.5 Tín hiệu rời rạc giá trị

Tín hiệu analog: là tín hiệu liên tục về thời gian, liên tục về giá trị.



Hình 1.6 Tín hiệu analog

Tín hiệu số: là tín hiệu rời rạc về thời gian, rời rạc về giá trị.



Hình 1.7 Tín hiệu số

Tín hiệu ngẫu nhiên: giá trị của tín hiệu trong tương lai không thể biết trước được. Các tín hiệu trong tự nhiên thường thuộc nhóm này

Tín hiệu tất định: giá trị tín hiệu ở quá khứ, hiện tại và tương lai đều được xác định rõ, thông thường có công thức xác định rõ ràng

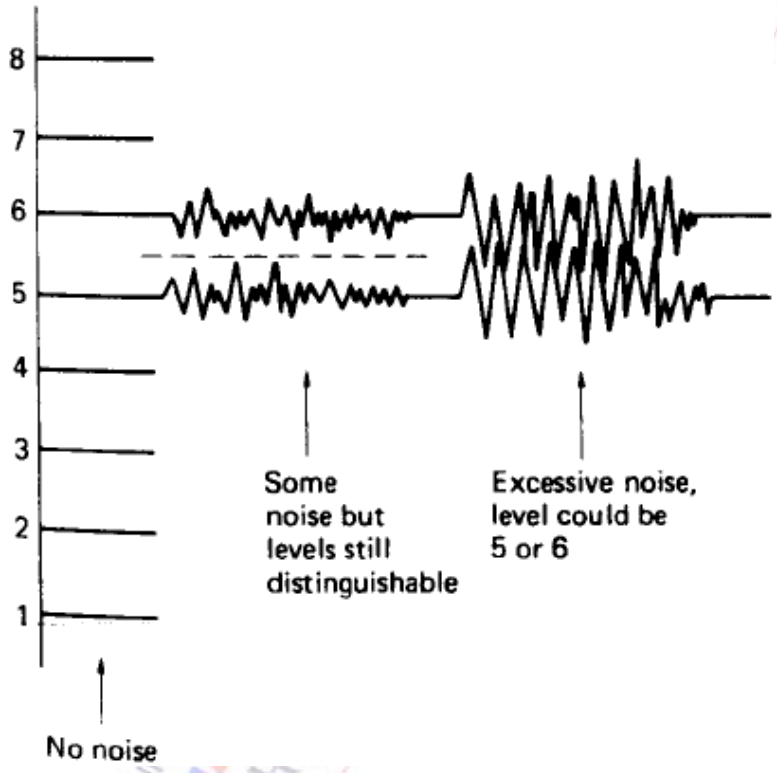
1.1.1.4 Phân loại hệ thống xử lý

Gồm hai loại hệ thống là hệ thống tương tự và hệ thống số. Trong đó hệ thống xử lý số: là hệ thống có thể lập trình được, dễ mô phỏng, cấu hình, sản xuất hàng loạt với độ chính xác cao, giá thành hạ, tín hiệu số dễ lưu trữ, vận chuyển và sao lưu, nhược điểm là khó thực hiện với các tín hiệu có tần số cao

1.1.1.5 Hệ thống số xử lý âm thanh [3]

Độ nhạy của tai người rất cao, nó có thể phân biệt được số lượng nhiễu rất nhỏ cũng như chấp nhận tầm biên độ âm thanh rất lớn. Các đặc tính của một tín hiệu tai người nghe được có thể được đo đạc bằng các công cụ phù hợp. Thông thường, tai người nhạy nhất ở tầm tần số 2kHz và 5kHz, mặc dù cũng có người có thể nhận dạng được tín hiệu trên 20kHz. Tầm động nghe được của tai người được phân tích và người ta nhận được kết quả là có dạng đáp ứng logarith.

Tín hiệu âm thanh được truyền qua hệ thống số là chuỗi các bit. Bởi vì bit có tính chấtt rời rạc, dễ dàng xác định số lượng bằng cách đếm số lượng trong một giây, dễ dàng quyết định tốc độ truyền bit cần thiết để truyền tín hiệu mà không làm mất thông tin.



Hình 1.8 Để nhận được tám mức tín hiệu khác nhau một cách phân biệt, tín hiệu đỉnh-đỉnh của tín hiệu nhiễu phải nhỏ hơn hoặc độ sai biệt giữa các mức độ. Tỉ số tín hiệu trên nhiễu phải tối thiểu là 8:1 hoặc là 18dB, truyền bởi 3 bit.Ở 16 mức thì tỉ số tín hiệu trên nhiễu phải là 24dB, truyền bởi 4 bit.

1.1.1.6 Mô hình hóa tín hiệu âm thanh [4]

Có rất nhiều kỹ thuật xử lý tín hiệu được mô hình hóa và áp dụng các giải thuật trong việc khôi phục âm thanh. Chất lượng của âm thoại phụ thuộc rất lớn vào mô hình giả định phù hợp với dữ liệu. Đối với tín hiệu âm thanh, bao gồm âm thoại, nhạc và nhiễu không mong muốn, mô hình phải tổng quát và không sai lệnh so với giả định. Một điều cần lưu ý là hầu hết các tín hiệu âm thoại là các tín hiệu động trong thực tế, mặc dù mô hình thực tiễn thì thường giả định khi phân tích tín hiệu là tín hiệu có tính chất tĩnh trong một khoảng thời gian đang xét.

Mô hình phù hợp với hầu hết rất nhiều lãnh vực trong việc xử lý chuỗi thời gian, bao gồm việc phục hồi âm thanh là mô hình Autoregressive (viết tắt AR), được dùng làm mô hình chuẩn cho việc phân tích dự đoán tuyến tính.

Tín hiệu hiện tại được biểu diễn bởi tổng giá trị của P tín hiệu trước đó và tín hiệu nhiễu trắng, P là bậc của mô hình AR:

$$s[u] = \sum_{i=1}^{P} s[n-i]a_i + e[n] \tag{1.1}$$

Mô hình AR đại diện cho các quá trình tuyến tính tĩnh, chấp nhận tín hiệu tương tự nhiễu và tín hiệu tương tự điều hòa. Một mô hình khác phù hợp hơn đối với nhiều tình huống phân tích là mô hình auto regressive moving-average (ARMA) cho phép các điểm cực cũng như điểm 0. Tuy nhiên mô hình AR có tính linh động hơn trong phân tích hơn mô hình ARMA, ví dụ một tín hiệu nhạc phức tạp cần mô hình có bậc $P > 100$ để biểu diễn dạng sóng của tín hiệu, trong khi các tín hiệu đơn giản hơn chỉ cần biểu diễn bằng bậc 30. Trong nhiều ứng dụng, việc lựa chọn bậc của mô hình phù hợp cho bài toán sao cho đảm bảo việc biểu diễn tín hiệu là thỏa việc không làm mất đi thông tin của tín hiệu là việc hơi phức tạp. Có rất nhiều phương pháp dùng để ước lượng bậc của mô hình AR như phương pháp *maximum likelihood/least-squares [Makhoul, 1975],* và phương pháp *robust to noise [Huber, 1981, Spath, 1991]*, v.v… Tuy nhiên, đối với việc xử lý các tín hiệu âm nhạc phức tạp thì thông thường sử dụng mô hình Sin (Sinusoidal) rất có hiệu quả trong các ứng dụng âm thoại. Mô hình Sin rất phù hợp trong các phương pháp dùng để giảm nhiễu. Tín hiệu được cho bởi công thức sau

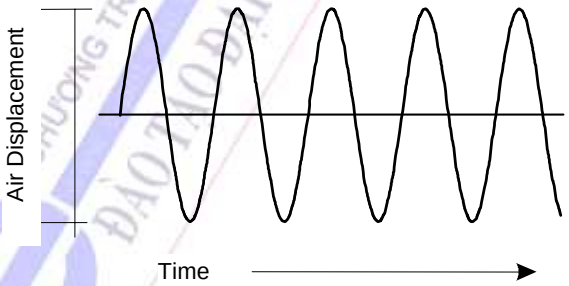
$$s[n] \sum_{i=1}^{P_n} a_i[n] \sin\left(\int_0^{nT} \omega_i(t)dt + \phi_i \right) \tag{1.2}$$

Đây là mô hình tổng quát đối với các điều chế biên độ và điều chế tần số, tuy nhiên lại không phù hợp đối với các tín hiệu tương tự nhiễu, mặc dù việc biểu diễn tín hiệu nhiễu có thể được biểu diễn bởi số lượng hàm sin rất lớn.

1.1.1.7 Kiến trúc hệ thống số xử lý âm thanh

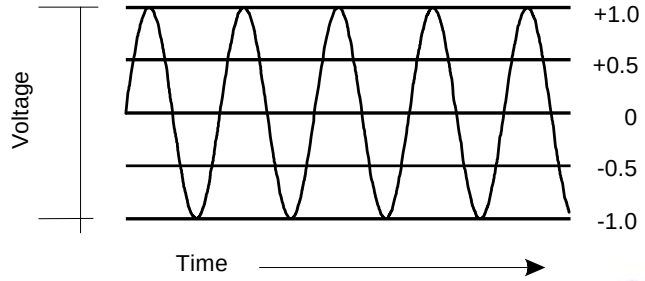
Đối với máy tính số xử lý âm thanh, người ta thường dùng phương pháp Điều chế xung (Pulse Code Modulation , viết tắt PCM). Dạng sóng âm thanh được chuyển sang dãy số PCM như sau, xét tín hiệu hình sin làm ví dụ:

- Tín hiệu gốc là tín hiệu như Hình 1.9



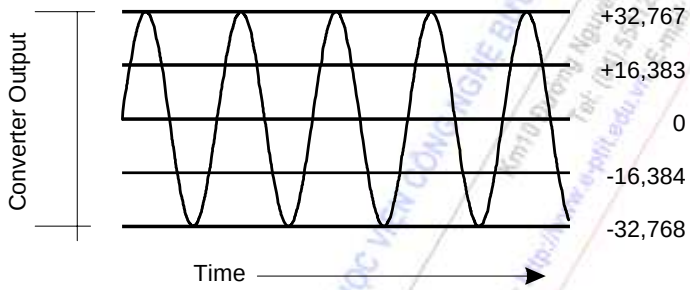
Hình 1.9 Dạng sóng âm thanh nguyên thủy

- Kế đến, sử dụng một microphone để thu tín hiệu âm thanh (trong không khí) và chuyển đổi thành tín hiệu điện, tầm điện áp ngõ ra của microphone ±1 volt như Hình 1.10.



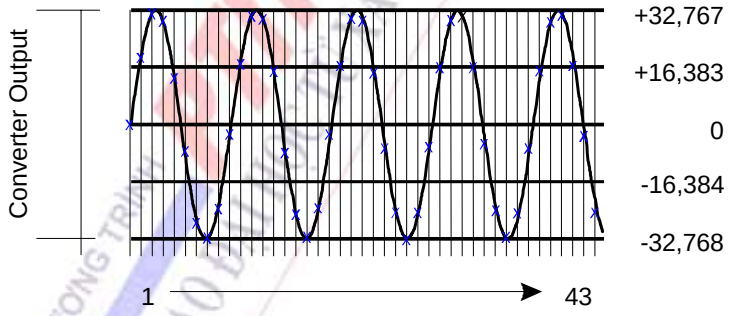
Hình 1.10 Dạng sóng của tín hiệu điện

- Tín hiệu điện áp dạng tương tự sau đó được chuyển thành dạng số hóa bằng thiết bị chuyển đổi tương tự-số (*analog-to-digital converter)*. Khi sử dụng bộ chuyển đổi 16bit tương tự-số, tầm số nguyên ngõ ra có giá trị –32,768 đến +32,767, được mô tả như hình 1.11.



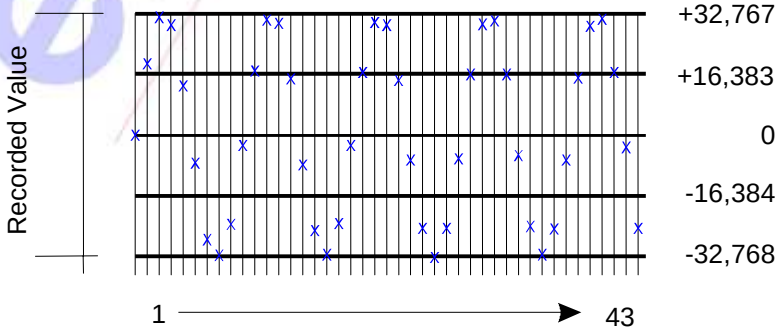
Hình 1.11 Ngõ ra của bộ chuyển đổi tín hiệu tương tự sang tín hiệu số

- Vì số lượng điểm dữ liệu là vô hạn nên không thể lấy tất cả các điểm thuộc trục thời gian, việc lấy mẫu sẽ được thực hiện trong một khoảng thời gian đều đặn. Số lượng mẫu trong một giây được gọi là tần số lấy mẫu (*sampling rate*). Hình 1.12 mô tả 43 mẫu được lấy



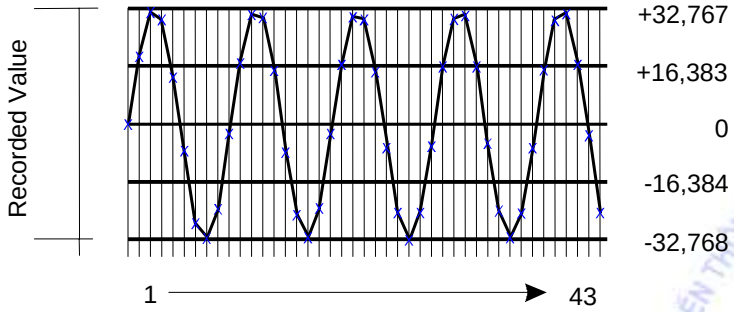
Hình 1.12 Thực hiện việc lấy mẫu

- Kết quả của việc lấy mẫu là một chuỗi gồm 43 chữ số biểu diễn cho các vị trí của dạng sóng ứng thời gian gian là một chu kỳ (hình 1.13).



Hình 1.13 Kết quả của việc lấy mẫu các giá trị

Máy tính sau đó sẽ xây dựng lại dạng sóng của tín hiệu bằng việc kết nối các điểm dữ liệu lại với nhau. Dạng sóng kết quả được mô tả ở Hình 1.14.



Hình 1.14 Dạng sóng được tái tạo lại

Lưu ý rằng có một vài điểm khác biệt giữa dạng sóng nguyên thủy và dạng sóng tái tạo (Hình 1.9 và Hình 1.14), lý do:

A. Các giá trị được tạo ra tại bộ chuyển đổi tín hiệu tương tự sang tín hiệu số là các số nguyên và được làm tròn giá trị.

B. Hình dáng của tín hiệu tái tạo phụ thuộc vào số lượng mẫu được ghi nhận.

Tổng quát, một dãy số hữu hạn (đại diện cho tín hiệu số) chỉ có thể biểu diễn cho một dạng sóng tín hiệu tương tự với độ chính xác hữu hạn.

1.1.1.8 Tần số lấy mẫu

Khi chuyển đổi một âm thanh sang dạng số, điều cần lưu ý là tần số lấy mẫu của hệ thống xử lý phải đảm bảo tính trung thực và chính xác khi cần phục hồi lại dạng sóng tín hiệu ban đầu.

Theo định lấy mẫu Nyquist và Shannon, tần số lấy mẫu quyết định tần số cao nhất của tín hiệu phục hồi. Để tái tạo lại dạng sóng có tần số là F, cần phải lấy $2F$ mẫu trong một giây. Tần số này còn được gọi là tần số Nyquist. Tuy nhiên, định lý Nyquist không phải là tối ưu cho mọi trường hợp. Nếu một dạng sóng hình Sin có tần số là 500Hz, thì tần số lấy mẫu 1000Hz. Nếu như tần số lấy mẫu cao hơn tần số Nyquist sẽ gây ra tình trạng "hiệu ứng là" ảnh hưởng đến biên độ của tín hiệu và tín hiệu bị cộng nhiễu, tuy nhiên lúc đó thì các thành phần hài tần số thấp lại có tín hiệu chính xác hơn khi được phục hồi.

1.1.2 Nhắc lại một số khái niệm toán học trong xử lý âm thanh

1.1.2.1 Phép biến đổi z [5]

Phép biến đổi *z* của một chuỗi được định nghĩa bởi cặp biểu thức

$$X(z)=\sum_{n=-\infty}^{\infty}x(n)z^{-n} \tag{1.3a}$$

$$x(n)=\frac{1}{2\pi j}\oint_C X(z)z^{n-1}dz \tag{1.3b}$$

Biến đổi z của $x(n)$ được định nghĩa bởi biểu thức (1.6a). $X(z)$ còn được gọi là dãy công suất vô hạn theo biến z^{-1} với các giá trị của $x(n)$ chính là các hệ số của dãy công suất. Miền hội tụ ROC là $\{z\mid |X(z)|<\infty\}$, là những giá trị của z sao cho chuỗi hội tụ, hay nói cách khác

$$\sum_{n=-\infty}^{\infty}\left|x(n)\right|\left|z^{-n}\right| \quad < \quad \infty \tag{1.4}$$

Thông thường, miền hội tụ của z có dạng:

$$R_1 < |z| < R_2 \tag{1.5}$$

Ví dụ: Cho $x(n)=\delta(n-n_0)$. Theo công thức (1.3a), ta có $X(z)=z^{-n_0}$

Ví dụ: Cho $x(n)=u(n)-u(n-N)$. Theo công thức (1.3a), ta có

$$X(z)=\sum_{n=0}^{N-1}(1).z^{-n}=\frac{1-z^{-N}}{1-z^{-1}}$$

Ví dụ: Cho $x(n)=a^n.u(n)$. Suy ra $X(z)=\sum_{n=0}^{\infty}a^n z^{-n}=\frac{1}{1-az^{-1}}, |a|<|z|$

Ví dụ: Cho $x(n)=-b^n u(-n-1)$. Then $X(z)=\sum_{n=-\infty}^{-1}b^n z^{-n}=\frac{1}{1-bz^{-1}}$, $|z|<|b|$

Bảng 2.1 Chuỗi tín hiệu và biến đổi *z* tương ứng

	Chuỗi tín hiệu	Biến đổi *z*
1. Tuyến tính	$ax_1(n)+bx_2(n)$	$aX_1(z)+bX_2(z)$
2. Dịch	$x(n+n_0)$	$z^{n_0}X(z)$
3. Hàm mũ	$a^n x(n)$	$X(a^{-1}z)$
4. Hàm tuyến tính	*nx(n)*	$-z\frac{dX(z)}{dz}$
5. Đảo thời gian	*x(-n)*	$X(z^{-1})$
6. Tương quan	*x(n)*h(n)*	*X(z)H(z)*
7. Nhân chuỗi	*x(n)w(n)*	$\frac{1}{2\pi j}\oint_C X(v)W(z/v)v^{-1}dv$

1.1.2.2 Phép biến đổi Fourier

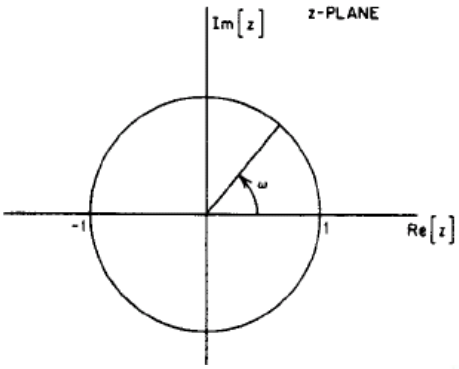
Biến đổi Fourier của tín hiệu rời rạc thời gian được cho bởi biểu thức

$$X(e^{jw})=\sum_{n=-\infty}^{\infty}x(n)e^{-jwn} \tag{1.6a}$$

$$x(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}X(e^{jw})e^{jwn}dw \tag{1.6b}$$

Biến đổi Fourier là trường hợp đặc biệt của phép biến đổi z bằng cách thay thế $z=e^{jw}$. Như mô tả trong Hình 1.4, trong mặt phẳng *z*, tần số w là góc quay. Điều kiện đủ để tồn tại biến đổi Fourier là $|z|=1$, như vậy

$$\sum_{n=-\infty}^{\infty}|x(n)| < \infty \tag{1.7}$$



Hình 1.15 Vòng tròn đơn vị thuộc mặt phẳng z

Một đặc tính quan trọng của biến đổi Fourier của một chuỗi là $X(e^{iw})$ là hàm điều hòa *w*, với chu kỳ là 2π .

Bằng cách thay $z = e^{iw}$ ở bảng 2.1, có có được bảng biến đổi Fourier tương ứng.

1.1.2.3 Phép biến đổi Fourier rời rạc

Trong trường hợp tín hiệu tương tự, tuần hoàn với chu kỳ *N*

$$\tilde{x}(n) = \tilde{x}(n+N) \qquad -\infty < n < \infty \tag{1.8}$$

Với $\tilde{x}(n)$ có thể có dạng là tổng rời rạc các tín hiệu sin thay vì tích phân như ở công thức (1.9b). Phép biến đổi Fourier cho chuỗi tuần hoàn như sau

$$\tilde{X}(k) = \sum_{n=0}^{N-1}\tilde{x}(n)e^{-j\frac{2\pi}{N}kn} \tag{1.9a}$$

$$\tilde{x}(k) = \frac{1}{N}\sum_{k=0}^{N-1}\tilde{X}(k)e^{j\frac{2\pi}{N}kn} \tag{1.9b}$$

Chuỗi *x(n)* hữu hạn, có giá trị bằng 0 với $0 \le n \le N-1$, có phép biến đổi *z* là.

$$X(z) = \sum_{n=0}^{N-1}x(n)z^{-n} \tag{1.10}$$

Nếu chia $X(z)$ thành *N* điểm trên vòng tròn đơn vị, $z_k = e^{j2\pi k/N}$, $k = 0,1,...,N-1$, ta có:

$$X(e^{j\frac{2\pi}{N}k}) = \sum_{n=0}^{N-1}x(n)e^{-j\frac{2\pi}{N}kn}, \; k = 0,1,...,N-1 \tag{1.11}$$

Chuỗi tuần hoàn vô hạn $\tilde{x}(n)$ có công thức từ *x(n)* như sau

$$\tilde{x}(n) = \sum_{r=-\infty}^{\infty}x(n+rN) \tag{1.12}$$

Ta nhận thấy rằng các mẫu $X(e^{j\frac{2\pi}{N}k})$ từ phương trình (1.9a) và (1.11) chính là các hệ số Fourier của chuỗi tuần hoàn $\tilde{x}(n)$ trong phương trình (1.12). Như vậy, một chuỗi có chiều dài N có thể được biểu diwnx bởi phép biến đổi Fourier rời rạc (DFT) như sau:

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{-j\frac{2\pi}{N}kn},\ k=0,1,...,N-1 \tag{1.13a}$$

$$x(n)=\frac{1}{N}\sum_{k=0}^{N-1}X(k)e^{j\frac{2\pi}{N}kn},\ n=0,1,...,N-1 \tag{1.13b}$$

Điều khác biệt duy nhất giữa biểu thức (1.12) và (1.9) là ký hiệu (loại bỏ ký hiệu ~ khi nói đến tín hiệu tuần hoàn) và giới hạn hữu hạn $0\le k\le N-1$ và $0\le n\le N-1$. Lưu ý một điều là chỉ dùng phép biến đổi DFT cho tín hiệu tuần hoàn có tính chất là module của N.

$$\begin{aligned}x(n)&=\sum_{k=-\infty}^{\infty}x(n+rN)=x(n \quad module\ N)\\&=x((n))_N\end{aligned} \tag{1.14}$$

Bảng 2.2 Chuỗi và biến đổi DFT

	Chuỗi tín hiệu	Biến đổi N điểm DFT
1. Tuyến tính	$ax_1(n)+bx_2(n)$	$aX_1(k)+bX_2(k)$
2. Dịch	$x((n+n_0))_N$	$e^{j\frac{2\pi}{N}kn_0}X(k)$
3. Đảo thời gian	$x((-n))_N$	$X^*(k)$
4. Kết hợp	$\sum_{m=0}^{N-1}x(m)h((n-m))_N$	*X(k)H(k)*
5. Nhân chuỗi	*x(n)w(n)*	$\frac{1}{N}\sum_{r=0}^{N-1}X(r)W((k-r))_N$

1.2 MÔ HÌNH XỬ LÝ ÂM THANH

1.2.1 Các mô hình lấy mẫu và mã hoá thoại

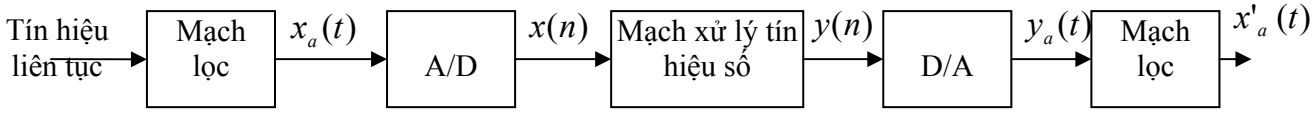
1.2.1.1 Lấy mẫu tín hiệu ở miền thời gian và tái tạo tín hiệu liên tục [6]

Để xử lý một tín hiệu liên tục bằng các phương tiện xử lý tín hiệu số, ta phải đổi tín hiệu liên tục đó ra dạng một chuỗi số bằng các lấy mẫu tín hiệu liên tục một cách tuần hoàn có chu kỳ là T giây. Gọi $x(n)$ là tín hiệu rời rạc hình thành do quá trình lấy mẫu, tín hiệu liên tục $x_a(t)$, ta có

$$x(n)=x_a(nT)\ -\infty<n<\infty \tag{1.15}$$

Các mẫu $x(n)$ phải được lượng hóa thành một tập các mức biên độ rời rạc rồi mới được đưa vào bộ xử lý số. Hình 1.16 minh họa một cấu hình tiêu biểu cho hệ thống xử lý tín hiệu tương

tự bằng phương pháp số. Trong các phần sau, ta bỏ qua sai số lượng hóa phát sinh trong quá trình biến đổi A/D



Hình 1.16 Cấu hình hệ thống xử lý tín hiệu tương tự bằng phương pháp số

Để xác định quan hệ giữa phổ của tín hiệu liên tục và phổ của tín hiệu rời rạc tạo ra từ quá trình lấy mẫu tín hiệu, liên tục đó, ta chú ý đến quan hệ giữa biến độc lập t và n của tín hiệu $x_a(t)$ và $x(n)$

$$t = nT = \frac{n}{F_s} \tag{1.16}$$

Định lý lấy mẫu: một tín hiệu liên tục có băng tần hữu hạn, có tần số cao nhất là B Hertz có thể khôi phục từ các mẫu của nó với điều kiện tần số lấy mẫu $F_s \geq 2B$ mẫu / giây

1.2.1.2 Lấy mẫu tín hiệu ở miền tần số và tái tạo tín hiệu liên tục

Ta đã biết tín hiệu liên tục có năng lượng hữu hạn thì có phổ liên tục. Trong phần này, ta sẽ xét quá trình lấy mẫu của các tín hiệu loại đó một cách tuần hoàn và sự tái tạo ín hiệu từ các mẫu của phổ của chúng

Xét một tín hiệu liên tục $x_a(t)$ với một phổ liên tục $X_a(F)$. Giả sử ta lấy mẫu $X_a(F)$ tại các thời điểm cách nhau ∂F Hertz. Ta muốn tái tạo $X_a(F)$ hoặc $x_a(t)$ từ các mẫu $X_a(F)$

Nếu tín hiệu tương tự $x_a(t)$ có giới hạn thời gian là $\Im$ giây và T_s được chọn để $T_s > 2\Im$ thì aliasing không xảy ra và phổ $X_a(F)$ có thể được khôi phục hoàn toàn từ các mẫu.

1.2.1.3 Lấy mẫu tín hiệu ở miền tần số và tái tạo tín hiệu rời rạc

Xét một tín hiệu rời rạc không tuần hoàn $x(n)$ có phép biến đổi Fourier:

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \tag{1.17}$$

Giả sử ta lấy mẫu $X(\omega)$ tuần hoàn tại các điểm cách nhau $\partial\omega$ rad. Vì $X(\omega)$ tuần hoàn với chu kỳ 2π, chỉ có các mẫu trong phạm vi tần số cơ bản là cần thiết. Để thuận tiện, ta lấy N mẫu cách đều nhau trong khoảng $0 \leq \omega \leq 2\pi$ theo khoảng cách $\partial\omega = 2\pi / N$

Xét $\omega = 2\pi k / N$, ta được $X\left(\frac{2\pi}{N}k\right) = \sum_{n=-\infty}^{\infty} x(n)e^{-j2\pi kn/N}$ $k = 0,1,...,N-1$ (1.18)

Xét tín hiệu $x_p(n) = \sum_{l=-\infty}^{\infty} x(n-lN)$ nhận được bằng cách lặp lại tuần hoàn $x(n)$ tại mỗi N mẫu, tín hiệu này tuần hoàn với chu kỳ N, do đó có thể được triển khai theo khai triển Fourier

$$x_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(\frac{2\pi}{N}k)e^{j2\pi kn/N},\ n = 0,1,...,N-1 \tag{1.19}$$

Từ công thức $x_p(n)$ trên, ta nhận thấy có thể khôi phục tín hiệu $x_p(n)$ từ các mẫu của phổ $X(\omega)$. Như vậy, ta phải tìm ra mối tương quan giữa $x_p(n)$ và $x(n)$ để có thể thực hiện khôi phục $x(n)$ từ $X(\omega)$

Vì $x_p(n)$ là sự mở rộng tuần hoàn của $x(n)$, nên $x(n)$ có thể được khôi phục từ $x_p(n)$ nếu không có aliasing ở cõi thời gian, nghĩa là nếu $x(n)$ có thời gian giới hạn nhỏ hơn hoặc bằng chu kỳ N của $x_p(n)$.

1.2.1.4 Các chuẩn mã hóa âm thoại trong các hệ thống xử lý thoại [7]

Chuẩn mã hóa âm thoại thông thường được nghiên cứu và phát triển bởi một nhóm các chuyên gia đã giành hết thời gian và tâm huyết thực hiện các công việc kiểm nghiệm, mô phỏng sao cho đảm bảo một tập các yêu cầu đưa ra đáp ứng được. Chỉ có các tổ chức với nguồn tài nguyên khổng lồ mới có thể thực hiện được các công việc khó khăn này, thông thường, thời gian tối thiểu cần thiết để hoàn thành một chuẩn trong trường hợp gặp nhiều thuận lợi trong quá trình là khoảng bốn năm rưỡi.

Điều này không có nghĩa là một chuẩn được đưa ra thì "không có lỗi" hoặc không cần phải cải tiến. Do đó, các chuẩn mới luôn luôn xuất hiện sao cho tốt hơn chuẩn cũ cũng như phù hợp với các ứng dụng trong tương lai.

Hội đồng chuẩn là các tổ chức có trách nhiệm trong việc giám sát việc phát triển các chuẩn cho một ứng dụng cụ thể nào đó. Sau đây là một số hội đồng chuẩn nổi tiếng được nhiều nhà cung cấp sản phẩm tuân theo

- Liên minh viễn thông quốc tế - International Telecommunications Union (ITU): Các chuẩn viễn thông của ITU (chuẩn ITU-T) có uy tín trong việc định ra các chuẩn mã hóa âm thoại cho hệ thống mạng điện thoại, bao gồm các mạng vô tuyến lẫn hữu tuyến.
- Hiệp hội công nghiệp viễn thông - Telecommunications Industry Association (TIA): có trách nhiệm ban hành các chuẩn mã hóa thoại cho các ứng dụng cụ thể, là một thành viên của Viện tiêu chuẩn quốc gia Hoa Kỳ - National Standards Institute (ANSI). TIA đã thành công trong việc phát triển các chuẩn sử dụng trong các hệ thống tổng đài tế bào số Bắc Mỹ, bao gồm các hệ thống sử dụng chuẩn đa kết phân thời gian - Time division multiple access (TDMA) và Đa truy nhập phân chia theo mã - Code division multiple access (CDMA).
- Viện tiêu chuẩn viễn thông châu Âu - European Telecommunications Standards Institute (ETSI): ETSI có các hội viên từ các nước cũng như các công ty Châu Âu, là tổ chức đưa ra các chuẩn sản xuất thiết bị tại Châu Âu. ETSI được thành lập bởi nhóm có ảnh hưởng nhất trong lãnh vực mã hóa âm thoại là nhóm di động đặc biệt - Groupe Speciale Mobile (GSM), đã đưa ra rất nhiều chuẩn hữu dụng và được triển khai rất nhiều trên thế giới
- Bộ quốc phòng Hoa Kỳ - United States Department of Defense (DoD). DoD có liên quan đến việc sáng lập các chuẩn mã hóa thoại, được biết đến với các chuẩn liên bang Hoa Kỳ (U.S. Federal) dùng nhiều cho các ứng dụng quân sự
- Trung tâm phát triển và nghiên cứu các hệ thống vô tuyến của Nhật Bản - Research and Development Center for Radio Systems of Japan (RCR). Các chuẩn tế bào số được phát hành bởi RCR.

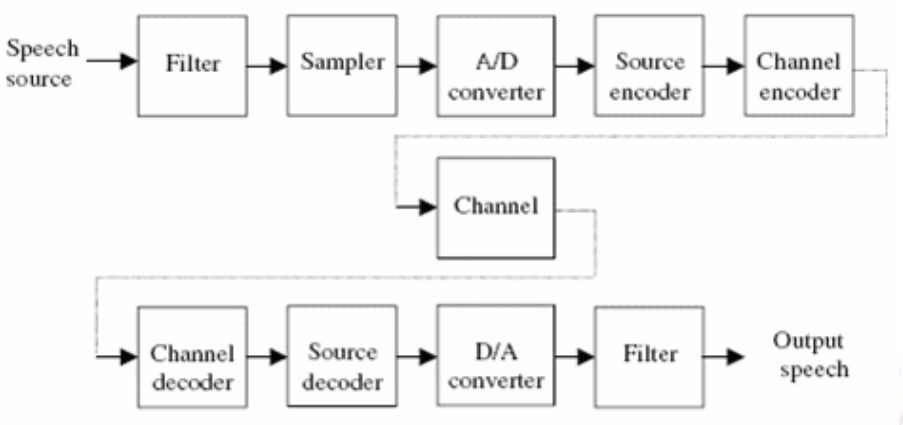
Bảng 2.3 Các chuẩn mã hóa âm thoại chính

Năm hoàn thành	Tên chuẩn	Tốc độ bit truyền (kbps)	Các ứng dụng
1972[a]	ITU-T G.711 PCM	64	Sử dụng công cộng
1984[b]	FS 1015 LPC	2.4	Liên lạc bảo mật
1987[b]	ETSI GSM 6.10 RPE-LTP	13	Vô tuyến di động số
1990[c]	ITU-T G.726 ADPCM	16, 24, 32, 40	Sử dụng công cộng
1990[b]	TIA IS54 VSELP	7.95	Hệ thống thoại tế bào số TDMA Bắc Mỹ
1990[c]	ETSI GSM 6.20 VSELP	5.6	Hệ thống tế bào GSM
1990[c]	RCR STD-27B VSELP	6.7	Hệ thống tế bào Nhật
1991[b]	FS1016 CELP	4.8	Liên lạc bảo mật
1992[b]	ITU-T G.728 LD-CELP	16	Sử dụng công cộng
1993[b]	TIA IS96 VBR-CELP	8.5, 4, 2, 0.8	Hệ thống thoại tế bào số CDMA Bắc Mỹ
1995[a]	ITU-T G.723.1 MP-MLQ/ACELP	5.3, 6.3	Liên lạc đa phương tiện, điện thoại truyền hình
1995[b]	ITU-T G.729 CS-ACELP	8	Sử dụng công cộng
1996[a]	ETSI GSM EFR ACELP	12.2	Sử dụng công cộng
1996[a]	TIA IS641 ACELP	7.4	Hệ thống thoại tế bào số TDMA Bắc Mỹ
1997[b]	FS MELP	2.4	Liên lạc bảo mật
1999[a]	ETSI AMR-ACELP	12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15, 4.75	Sử dụng công cộng viễn thông

[a] là được mô tả một phần
[b] là được giải thích đầy đủ
[c] là được mô tả ngắn gọn mà không có mô tả kỹ thuật chi tiết

1.2.1.5 Kiến trúc của hệ thống mã hóa âm thoại [8]

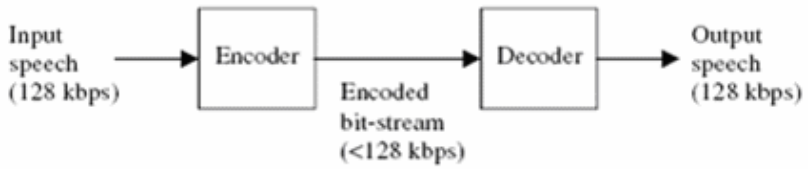
Hình 1.17 mô tả sơ đồ khối của hệ thống mã hóa âm thoại. Tín hiệu âm thoại tương tự liên tục có từ nguồn cho trước sẽ được số hóa bởi bộ một bộ lọc chuẩn, bộ lấy mẫu (bộ chuyển đổi thời gian rời rạc), và bộ chuyển tín hiệu tương tự sang tín hiệu số. Tín hiệu ngõ ra là tín hiệu âm thoại thời gian rời rạc với các giá trị lấy mẫu cũng rời rạc hóa. Tín hiệu này được xem là tín hiệu âm thoại số.



Hình 1.17 Sơ đồ khối của hệ thống xử lý tín hiệu thoại

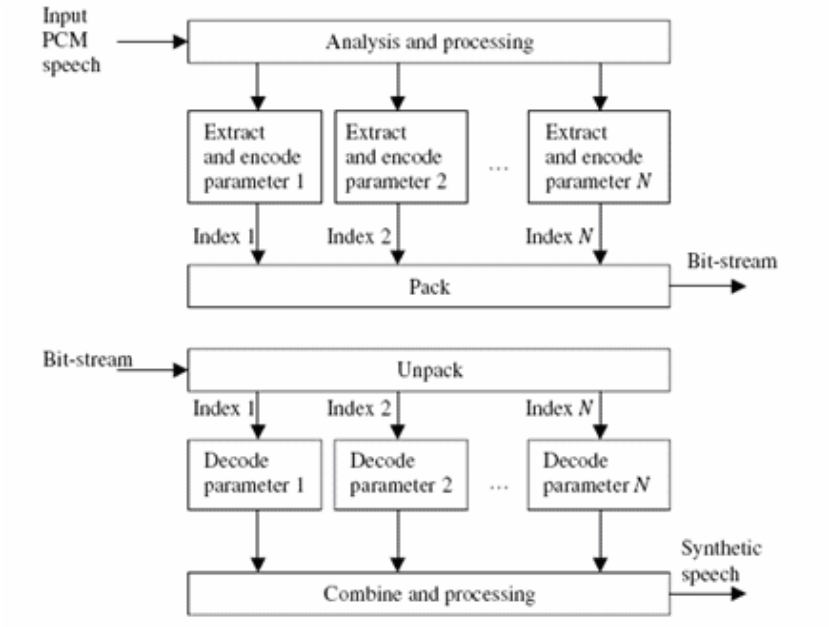
Thông thường, hầu hết các hệ thống mã hóa âm thoại được thiết kế để hỗ trợ các ứng dụng viễn thông, với tần số giới hạn giữa 300 và 3400Hz. Theo lý thuyết Nyquist, tần số lấy mẫu tối thiểu phải lớn hơn hai lần băng thông của tín hiệu liên tục thời gian. Giá trị 8kHz thường được lựa chọn là tần số lấy mẫu chuẩn cho tín hiệu thoại. Bộ mã hóa kênh thực hiện việc mã hóa hiệu chỉnh lỗi của chuỗi bit truyền trước khi tín hiệu được truyền trên kênh truyền, nơi mà tín hiệu sẽ bị thay đổi do nhiễu cũng như giao thoa tín hiệu…. Bộ giải mã thực hiện việc hiệu chỉnh lỗi để có được tín hiệu đã mã hóa, sau đó tín hiệu được đưa vào bộ giải mã để có được tín hiệu âm thoại số có cùng tốc độ với tín hiệu ban đầu. Lúc này, tín hiệu số sẽ được chuyển sang dạng tương tự thời gian liên tục. Bộ phận thực hiện việc xử lý tín hiệu thoại chủ yếu của mô hình hệ thống xử lý thoại là bộ mã hóa và giải mã. Thông thường, khi xử lý các bài toán về truyền thoại, mô hình được đơn giản hóa như Hình 1.18

Ví dụ tín hiệu thoại ngõ vào là tín hiệu rời rạc thời gian có tốc độ bit là 128kbps được đưa vào bộ mã hóa để thực hiện mã hóa chuỗi bit hoặc thực hiện nén dữ liệu thoại. Tốc độ của chuỗi bit thông thường sẽ có tốc độ thấp hơn tốc độ của tín hiệu ngõ vào bộ mã hóa. Bộ giải mã nhận chuỗi bit mã hóa này và tạo ra tín hiệu thoại có dạng là rời rạc thời gian và có tốc độ bằng với tốc độ của tín hiệu ban đầu truyền vào hệ thống.



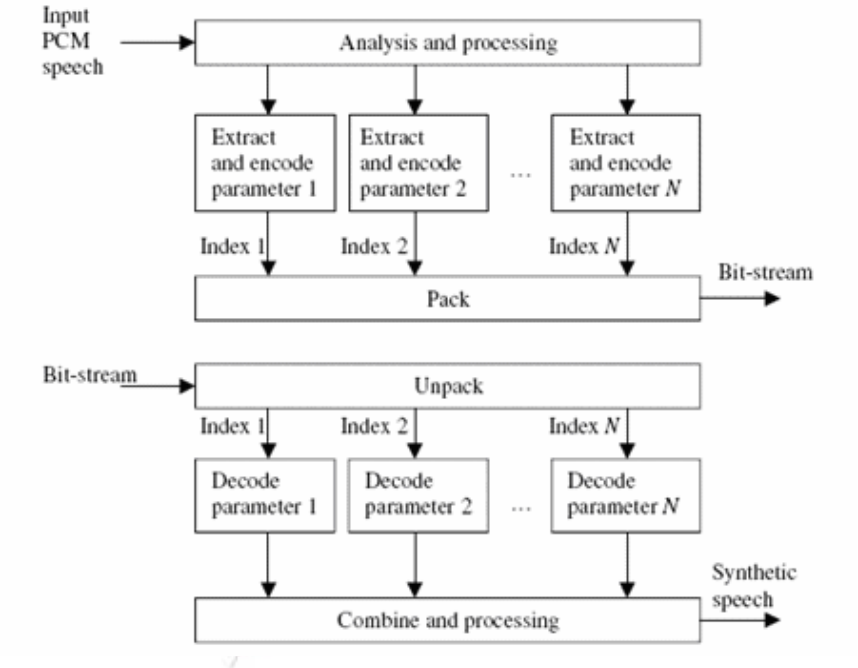
Hình 1.18 Sơ đồ khối đơn giản hóa của bộ mã hóa âm thoại

1.2.1.6 Kiến trúc tổng quát của bộ mã hóa – giải mã âm thoại [9]



Hình 1.19 Mô tả sơ đồ khối tổng quát của bộ mã hóa và giải mã âm thoại.

Đối với bộ mã hóa, tín hiệu âm thoại đầu vào được xử lý và phân tích nhằm thu được các thông số đại diện cho một khung truyền. Các thông số ngày được mã hóa và lượng tử với mã chỉ số nhị phân và được gửi đi như là một chuỗi bit đã được nén. Các chỉ số này được đóng gói và biểu diễn thành chuỗi bit, chúng được sắp xếp thứ tự truyền dựa vào các thông số đã quyết định trước và được truyền đến bộ giải mã.



Hình 1.20 Mô hình chung của bộ mã hóa âm thoại. Hình trên: bộ mã hóa, hình dưới: bộ giải mã.

Bộ giải mã thực hiện việc phân tích chuỗi bit nhận được, các chỉ số nhị phân được phục hồi sau quá trình phân tích và dùng để kết hợp với các thông số tương ứng của bộ giải mã để có

được các thông số đã được lượng tử. Các thông số giải mã này sẽ kết hợp với nhau và được xử lý để tạo lại tín hiệu âm thoại tổng hợp.

1.2.1.7 Các yêu cầu cần có của một bộ mã hóa âm thoại [10]

Mục tiêu chính của của mã hóa thoiạ là tối đa hóa chất lượng nghe tại một tốc độ bit nào đó, hoặc tối thiểu hóa tốc độ bit ứng với một chất lượng đặc thù. Tốc độ bit tương ứng với âm thoại nào sẽ được truyền hoặc lưu trữ phụ thuộc vào chi phí của việc truyền hay lưu trữ, chi phí của mã hóa tín hiệu thoại số, và các yêu cầu về chất lượng của âm thoại đó. Trong hầu hết các bộ mã hóa âm thoại, tín hiệu được xây dựng lại sẽ khác với tín hiệu nguyên thủy. Tốc độ bit truyền bị giảm bởi việc biểu diễn tín hiệu âm thoại (hoặc các thông số trong mô hình tạo âm thoại) với độ chính xác bị giảm, và bởi quá trình loại bỏ các thông tin dư thừa của tín hiệu. Các yêu cầu lý tưởng của một bộ mã hóa thoại bao gồm:

- Tốc độ bit thấp: đối với chuỗi bit mã hóa có tốc bit tỉ lệ thuận với băng thông cần cho truyền dữ liệu. Điều này dẫn đến nếu tốc độ bit thấp sẽ làm tăng hiệu suất của hệ thống. Yêu cầu này lại xung đột với các các đặc tính tốt khác của hệ thống, như là chất lượng của âm thoại. Trong thực tế, việc đánh đổi giữa các lựa chọn phụ thuộc vào áp dụng vào ứng dụng gì.
- Chất lượng thoại cao: tín hiệu âm thoại đã giải mã phải có chất lượng có thể chấp nhận được đối với ứng dụng cần đạt. Có rất nhiều khía cạnh về mặt chất lượng bao gồm tính dễ hiểu, tự nhiên, dễ nghe và cũng như có thể nhận dạng người nói.
- Nhận dạng tiếng nói / ngôn ngữ khác nhau: kỹ thuật nhận dạng tiếng nói có thể phân biệt được giọng nói của người lớn nam giới, người lớn nữ giới và trẻ con cũng như nhận dạng được ngôn ngữ nói của người nói.
- Cường độ mạnh ở trong kênh truyền nhiễu: đây là yếu tố quan trọng đối với các hệ thống truyền thông số với các nhiễu ảnh hưởng mạnh đến chất lượng của tính hiệu thoại.
- Hiệu suất cao đối với các tín hiệu phi thoại (ví dụ như tín hiệu tone điện thoại): trong hệ thống truyền dẫn kinh điển, các tín hiệu khác có thể tồn tại song song với tín hiệu âm thoại. Các tín hiệu tone như là đa tần tone đôi – Dual tone multifrequency(DTMF) của tín hiệu âm bàn phím và nhạc thông thường bị chèn vào trong đường truyền tín hiệu. Ngay cả những bộ mã hóa thoại tốc độ thấp cũng có thể không thể tạo lại tín hiệu một cách hoàn chỉnh.
- Kích thước bộ nhớ thấp và độ phức tạp tính toán thấp: nhằm mục đích sử dụng được bộ mã hóa âm thoại trong thực tế, chi phí thực hiện liên quan đến việc triển khai hệ thống phải thấp, bao gồm cả việc bộ nhớ càn thiết để hỗ trợ khi hệ thống hoạt động cũng như các yêu cầu tính toán. Các nhà nghiên cứu mã hóa âm thoại đã nổ lực trong việc tìm kiếm hiện thực bài toán triển khai trong thực tiễn sao cho có hiệu quả nhất.
- Độ trễ mã hóa thấp: trong quá trình xử lý mã hóa và giải mã thoại, độ trễ tín hiệu luôn luôn tồn tại, chính là thời gian trượt giữa âm thoại ngõ vào của bộ mã hóa với tín hiệu ngõ ra của bộ giải mã. Việc trễ quá mức sẽ sinh ra nhiều vấn đề trong việc thực hiện trao đổi tiếng nói hai chiều trong thời gian thực.

1.2.2 Các mô hình dùng trong xử lý âm thanh [11]

1.2.2.1 Mô hình quang phổ

1.2.2.1.1 Mô hình sin

Tín hiệu âm thanh có thể được triển khai từ tập hợp các mô hình sin nếu như có có dạng

$$y(t) = \sum_{i=1}^{I} A_i(t) e^{j\phi_i(t)} \tag{1.20}$$

Với $\phi_i(t) = \int_{-\infty}^{t} \omega_i(\tau) d\tau$, $A_i(t)$ và $\omega_i(t)$ là thành phần biên độ và tần số tương ứng của thành phần sin thứ i. Trong thực tế, tín hiệu được xem xét là tín hiệu rời rạc thời gian thực, như vậy ta có thể viết lại

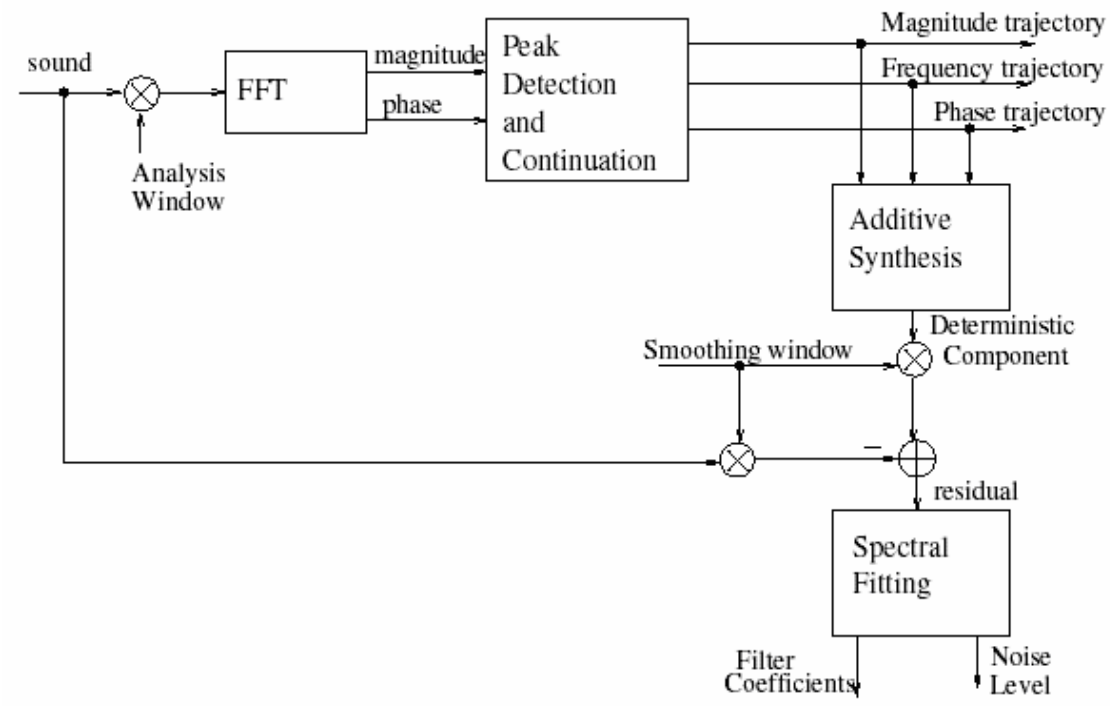
$$y(n) = \sum_{i=1}^{I} A_i(n) \cos(\phi_i(n)) \tag{1.21}$$

$$\text{Với } \phi_i(n) = \int_{0}^{nT} \omega_i(\tau) d\tau + \phi_{0,i}. \tag{1.22}$$

Về cơ bản, nếu như I có giá trị vô cùng lớn, thì bất cứ tín hiệu âm thanh nào cũng có thể được triển khai từ mô hình sin, phép tính gần đúng được áp dụng tính toán trong mô hình này. Thực tế, tính hiệu nhiễu cũng được triển khai thành vô số các tín hiệu sin, và ta tách việc xử lý riêng tín hiệu này thành phần xử lý Stochastic (Λ) được ký hiệu là $e(n)$.

$$y(n) = \underbrace{\sum_{i=0}^{I} A_i(n) \cos(\phi_i(n))}_{\Gamma} + \underbrace{e(n)}_{\Lambda} \tag{1.23}$$

Thành phần Λ có thể được tính bằng phép biến đổi Short-Time Fourier sử dụng lưu đồ ở hình 1. Phương pháp này được ứng dụng trong các phần mềm sms, viết tắt của tổng hợp mô hình phổ - spectral modeling synthesis.



Hình 1.21 Phân tích các thành phần hình sin của phần stochastic

Phát hiện đỉnh và ghép (Peak detection and continuation): để thực hiện việc phân tích các thành phần hình sin từ tín hiệu thặng dư, ta phải tìm được và ghi chú lại các đỉnh tần số nổi trội, tức là các thành phần hình sin nắm vai trò chính trong công thức phân tích được. Một chiến thuật được sử dụng để thực hiện điều này là vẽ "bảng chỉ dẫn" trong các khung STFT.

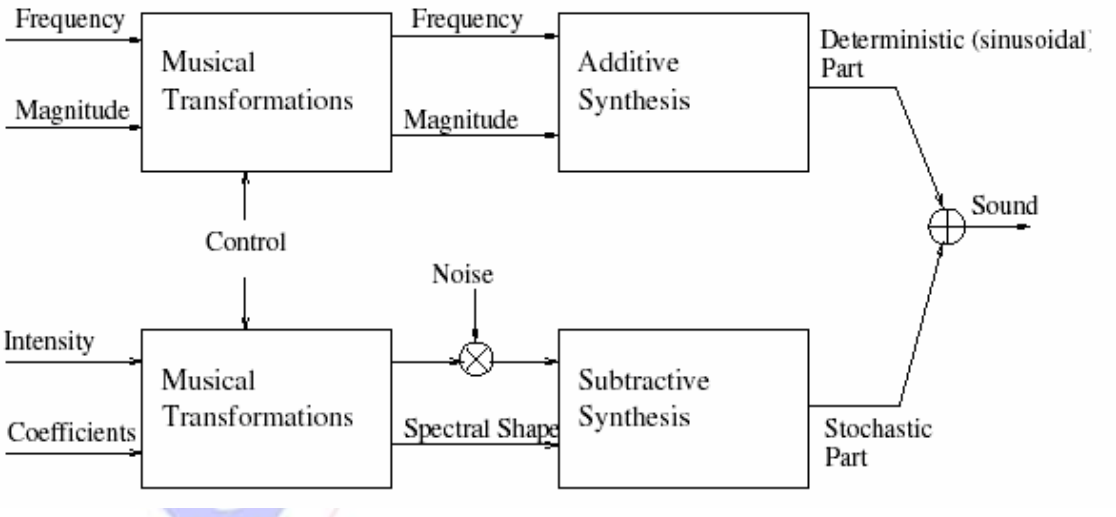
Để thực hiện việc phân chia phần nào là tín hiệu, phần nào là nhiễu, các tần số và pha phải được xác định một cách chính xác. Ngoài ra, để quá trình tổng hợp lại hai tín hiệu đó được đơn giản, biên độ của các thành phần nên được nội suy giữa các khung tín hiệu, và phép nội suy tuyến tính thường được sử dụng. Các tần số cũng như pha của tín hiệu cũng có thể được nội suy, tuy nhiên cần phải lưu ý là phép nội suy tần số có ảnh hưởng chặt chẽ đến phép nội suy pha.

Tổng hợp lại các thành phần sin: Trong giai đoạn tổng hợp lại, các thành phần sin có thể được tạo bởi bất kỳ phương pháp nào như máy tạo dao động số, máy tạo dao động bảng sóng hoặc tổng hợp lấy mẫu bảng sóng, hoặc kỹ thuật dựa trên cơ sở FFT. Kỹ thuật FFT được sử dụng nhiều do tính tiện lợi khi tín hiệu có nhiều thành phần hình sin.

Trích tín hiệu thặng dư (Extraction of the residual): Việc trích phổ của tín hiệu nhiễu thặng dư có thể được thực hiện ở miền tần (được mô tả trong hình 1) hoặc trực tiếp từ miền thời gian.

Sự hiệu chỉnh phổ thặng dư (Residual spectral fitting): thành phần stochastic được mô hình hóa là tín hiệu nhiễu băng rộng, được lọc bởi khối đặc trưng tuyến tính. Phổ cường độ của tín hiệu thặng dư có thể được xấp xỉ bằng giá trị trung bình của hàm piecewise-linear. Việc tổng hợp trong miền thời gian có thể được thực hiện bằng phép đảo FFT, sau khi đã ấn định được một tập cường độ mong muốn và một tập pha ngẫu nhiên.

Hiệu chỉnh âm thanh: mô hình sin là một mô hình hữu dụng vì nó cho phép áp dụng việc truyền các âm thanh nhạc lấy từ việc ghi băng thực tế. Hình 1.22 mô tả một các bước thực hiện cho việc hiệu chỉnh tín hiệu âm nhạc



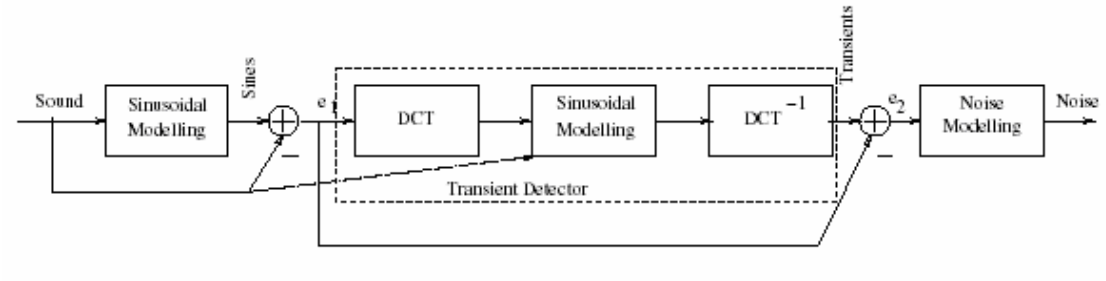
Hình 1.22 Cơ cấu tổ chức cho việc biểu diễn việc truyền tín hiệu âm nhạc

1.2.2.1.2 Tín hiệu sin + nhiễu + nốt đệm

Trong mô hình sin + nhiễu, điều cơ bản là các tín hiệu âm thanh là tổng hợp của nhiều tín hiệu sin tần số thấp và các loại nhiễu băng rộng hầu như ở dạng tĩnh. Khi đó, một thành phần của âm thanh không được xem xét đến, đó là nốt đệm. Việc hiệu chỉnh âm thanh có thể được thực

hiện dễ dàng bằng cách tách riêng thành phần nốt đệm để xét riêng. Thực tế, hầu hết các dụng cụ âm nhạc mở rộng trường độ của một nốt nhạc không làm ảnh hưởng đến chất lượng xử lý.

Với lý do này, một mô hình mới là sin + nhiễu + nốt đệm được phát họa dùng trong việc phân tích âm thanh. Ý tưởng chính của việc trích âm đệm trong thực tế từ việc quan sát rằng, các tín hiệu hình sin trong miền thời gian được ánh xạ qua miền tần thành các đỉnh có vị trí xác định, trong khi đó các xung ngắn đối ngẫu trong miền thời gian khi được ánh xạ qua miền tần lại có dạng hình sin. Như vậy, mô hình sin có thể được ứng dụng trong miền tần số biểu diễn các tín hiệu hình sin. Sơ đồ của việc phân tích SNT được mô tả trong Hình 1.23.



Hình 1.23 Phân tích tín hiệu âm thanh theo mô hình sin + nhiễu + nốt đệm

Khối DCT trong Hình 1.23 mô tả hoạt động của phép rời rạc cosin.

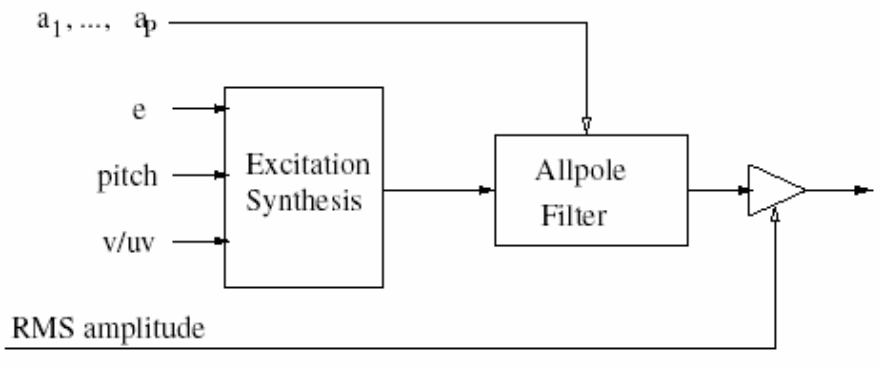
Phép biến đổi, được định nghĩa như sau:

$$C(k) = \alpha \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)k\pi}{2N} \right) \qquad (1.24)$$

Phép biến đổi DCT thực hiện việc một xung được biến đổi thành dạng cosin và ngược lại.

1.2.2.1.3 Mô hình LPC

Mã hóa dự đoán tuyến tính có thể được sử dụng để mô hình phổ tĩnh. Tổng hợp LPC được mô tả trong lưu đồ trong Hình 1.24. Về bản chất, mô hình chính là giải thuật trừ tổng hợp thực hiện một tính hiệu có phổ "đặc" được lọc bởi một bộ lọc cực. Tín hiệu kích thích có thể sử dụng chính tín hiệu thặng dư e có được qua quá trình phân tích, hoặc có thể dử dụng các thông tin của tín hiệu thoại/phi thoại.



Hình 1.24 Tổng hợp LPC

1.2.2.2 Mô hình miền thời gian

Việc mô tả âm thanh trong miền tần rất có hiệu quả, tuy nhiên trong một vài ứng dụng, để tiện việc nghiên cứu việc tổng hợp âm thanh, việc phân tích trong miền thời gian lại có ưu thế hơn.

1.2.2.2.1 Máy tạo dao động số

Ta nhận thấy một âm thanh phức tạp được tổng hợp từ nhiều thành phần hình sin bằng phép tổng hợp FTT^{-1}. Nếu như các thành phần hình sin không quá nhiều, việc tổng hợp từng thành phần được thực hiện bằng cách lấy giá trị trung bình của máy tạo dao động số.

$$e^{j\omega_0(n+1)} = e^{j\omega_0}e^{j\omega_0 n} \tag{1.25}$$

Với $e^{j\omega_0 n} = x_R(n) + jx_I(n)$ ở dạng số phức, mỗi bước nhảy thời gian được định nghĩa như sau:

$$x_R(n+1) = \cos\omega_0 x_R(n) - \sin\omega_0 x_I(n) \tag{1.26}$$

$$x_I(n+1) = \sin\omega_0 x_R(n) + \cos\omega_0 x_I(n) \tag{1.27}$$

Thông số biên độ và pha ban đầu có thể tính dựa theo pha ban đầu $e^{j\omega_0 0}$ và thực hiện việc lệch pha vào số mũ. Tín hiệu $x_R(n+1)$ có thể được tính theo công thức sau

$$x_R(n+1) = 2\cos\omega_0 x_R(n) - x_R(n-1) \tag{1.28}$$

Đáp ứng xung của bộ lọc như sau

$$H_R(z) = \frac{1}{1 - 2\cos\omega_0 z^{-1} + z^{-2}} = \frac{1}{\left(1 - e^{-j\omega_0 z^{-1}}\right)\left(1 - e^{j\omega_0 z^{-1}}\right)} \tag{1.29}$$

Giá trị cực của bộ lọc biểu thức 10 nằm trên chu vi đường tròn đơn vị.

Gọi x_{R1}, x_{R2} là hai biến trạng thái của hai mẫu trước đó của tín hiệu ngõ ra x_R, pha ban đầu ϕ_0 có thể được tính theo hệ phương trình sau

$$x_{R1} = \sin(\phi_0 - \omega_0) \tag{1.30}$$

$$x_{R2} = \sin(\phi_0 - 2\omega_0) \tag{1.31}$$

Máy tạo dao động số đặc biệt hữu ích trong việc biểu diễn tổng hợp tín hiệu đối với các bộ vi xử lý đa mục đích, khi các phép toán trên dấu chấm động được triển khai. Tuy nhiên, phương pháp này dùng cho việc tạo tín hiệu sin có hai bất lợi:

- Việc cập nhật thông số yêu cầu tính toán trên hàm cosin. Đây là một điều khó đối với điều chế tốc độ âm thanh, do phải thực hiện phép tính cosin ứng với từng mẫu trong miền thời gian
- Thay đổi tần số của máy dao động số sẽ làm thay đổi biên độ tín hiệu sin. Khi đó bộ phận logic điều khiển biên độ cần được sử dụng để điều chỉnh hạn chế này.

1.2.2.2.2 Máy tạo dao động bảng sóng

Trong phương pháp kinh điển và linh động nhất về tổng hợp các dạng sóng có chu kỳ (bao gồm tín hiệu dạng sin) là việc đọc lặp đi lặp lại một bảng chứa nội dung của một dạng sóng đã được lưu trữ trước. Nếu dạng sóng được tổng hợp ở dạng sin, đối xứng thì việc lưu trữ cho phép chỉ cần lưu trữ ¼ chu kỳ, và việc tính toán số học sẽ được nội suy cho cả chu kỳ.

Đặt $buf[\ ]$ là bộ đệm có nội dung chứa là chu kỳ của dạng sóng, hoặc bảng dạng sóng. Máy tạo dao động dạng sóng hoạt động lặp lại theo chu kỳ quét bảng dạng sóng là bội số của gia số I và đọc nội dung của bảng dạng sóng tại vị trí đó.

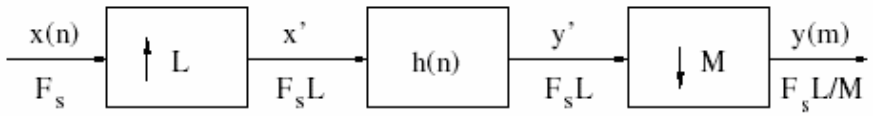
Gọi B là chiều dài của bộ đệm, f_0 là tần số mà ta muốn tạo tần số lấy mẫu F_s, khi đó giá trị của gia số I là:

$$I = \frac{Bf_0}{F_s} \tag{1.32}$$

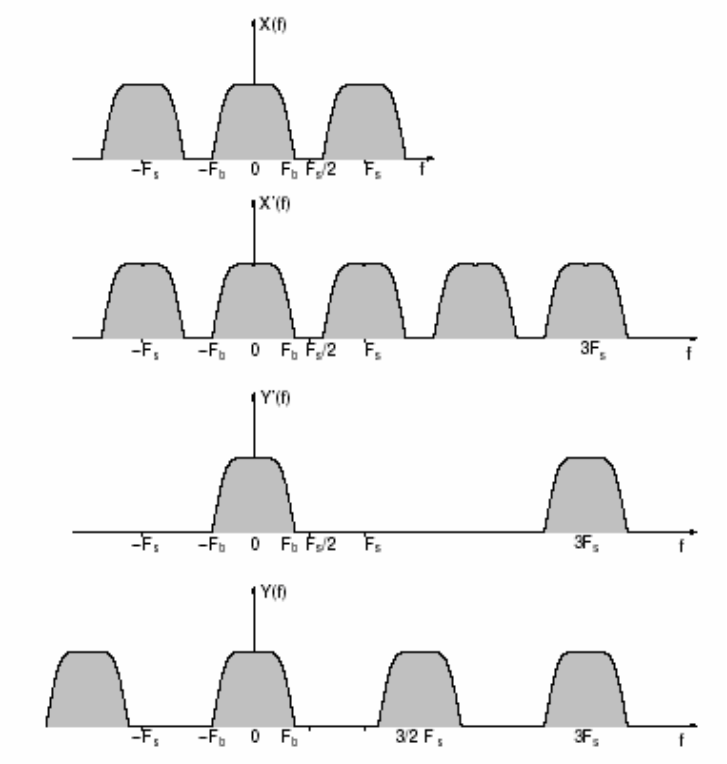
Sự thay đổi tần số lấy mẫu

Bài toán thiết kế máy tạo dao động bảng sóng có thể chuyển thành bài toán thay đổi tần số lấy mẫu, ví dụ như biến đổi tín hiệu được lấy mẫu tại tần số lấy mẫu $F_{s,1}$ thành tín hiệu được lấy mẫu tại tần số $F_{s,2}$. Nếu $\frac{F_{s,2}}{F_{s,1}} = \frac{L}{M}$, với L và M là các số nguyên tối giản, việc thực hiện thay đổi tần số lấy mẫu có thể được như hiện bằng các bước:

1. Tăng tần số lấy mẫu bằng hệ số L
2. Sử dụng bộ lọc thông thấp
3. Giảm tần số lấy mẫu bằng hệ số M



Hình 1.25 Sơ đồ khối phân rã của quá trình thay đổi tần số lấy mẫu



Hình 1.26 Ví dụ về thay đổi tần số lấy mẫu với $L/M = 3/2$

1.2.2.2.3 Tổng hợp lấy mẫu bảng sóng

Tổng hợp lấy mẫu bảng sóng là phần mở rộng của máy dao động bảng sóng đối với

- Dạng sóng phân tích không phải dạng sin
- Bảng dạng sóng được lưu trữ với nhiều chu kỳ

Các tín hiệu điều khiển rất quan trọng trong việc nhận được âm thanh tự nhiên

1.2.2.2.4 Tổng hợp hạt (với Giovanni De Poli)

Các bảng sóng ngắn có thể được đọc với nhiều tốc độ khác nhau, và kết quả là âm điệu có thể chồng chéo vào nhau trong miền thời gian. Trong phương pháp miền thời gian, việc tổng hợp âm thanh này được gọi là tổng hợp hạt. Tổng hợp hạt bắt đầu từ ý tưởng việc phân tích âm thanh trong miền thời gian được thay thế bằng biểu diễn chúng thành một chuỗi các thành phần ngắn được gọi là "hạt". Các thông số của kỹ thuật này là các dạng sóng của hạt thứ $g_k(\cdot)$, vị trí trong miền thời gian l_k và biên độ a_k

$$s_g(n) = \sum_k a_k g_k(n - l_k) \tag{1.33}$$

Khi số lượng "hạt" lớn, thì việc tính toán sẽ trở nên phức tạp. Tính chất của các hạt và các vị trí trong miền thời gian quyết định âm sắc của âm thanh. Việc lựa chọn các thông số tùy thuộc vào các tiêu chuẩn đưa ra bởi các mô hình thể hiện. Việc lựa chọn các mô hình biểu diễn liên quan đến các quá trình hoạt động mà các quá trình này có thể ảnh hưởng đến âm thanh nào đó theo nhiều cách khác nhau.

Loại cơ bản và quan trọng nhất của tổng hợp hạt (tổng hợp hạt bất đồng bộ) là phân phối các hạt không theo quy luật trong miền tần số -thời gian. Dạng sóng hạt có dạng

$$g_k(i) = \omega_d(i)\cos(2\pi f_k T_s i) \tag{1.34}$$

Với $\omega_d(i)$ là cửa số có chiều dài là d mẫu, dùng để điều khiển nhịp thời gian và băng tần phổ f_k.

1.2.2.3 Các mô hình phi tuyến

1.2.2.3.1 Điều pha và điều tần

Kỹ thuật tổng hợp phi tuyến thông dụng nhất là điều tần (FM). Trong liên lạc thông tin, FM được dùng trong các thập kỹ gần đây, nhưng ứng dụng của nó trong giải thuật tổng hợp âm thanh trong miền thời gian rời rạc được biết đến với cái tên John Chowning. Về bản chất, Chowning đã thực hiện các nghiên cứu trên các phạm vi khác nhau của việc tạo tiếng rung bằng các bộ tạo dao động đơn giản, và thu được kết quả là các tần số rung nhanh sẽ tạo ra các thay đổi đầy kịch tính. Như vậy, điều chế tần số của một máy tạo dao động cũng đủ tạo ra tín hiệu âm thanh có phổ phức tạp. Mô hình FM của Chowning như sau:

$$x(n) = A\sin(\omega_c n + I\sin(\omega_m n)) = A\sin(\omega_c n + \phi(n)) \tag{1.35}$$

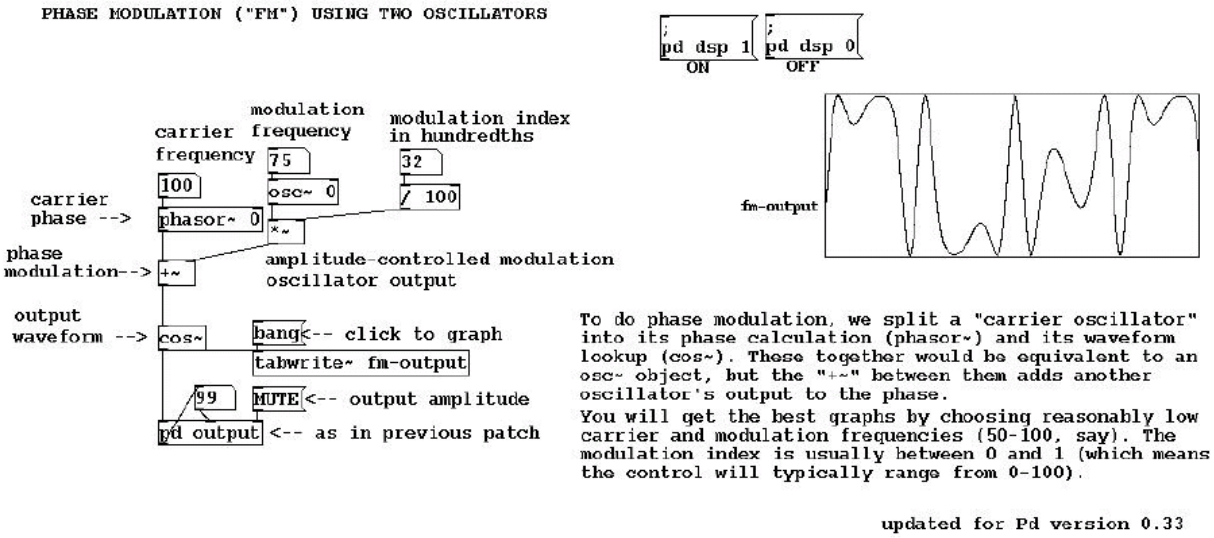
Với ω_c là tần số sóng mang và ω_m là tần số điều chế, I là chỉ số điều chế. Phương trình (16) thực tế cũng là phương trình điều pha. Tần số tức thời của phương trình (16)

$$\omega(n) = \omega_c - I\omega_m\cos(\omega_m n) \tag{1.36}$$

Hoặc: $$f(n) = f_c - If_m\cos(2\pi f_m n) \tag{1.37}$$

Hình 1.27 mô tả việc triển khai pd của giải thuật FM đơn giản. Tần số điều chế được dùng để điều khiển trực tiếp bộ tạo dao động, trong khi tần số sóng mang dùng để điều khiển bộ

tạo pha đơn vị, tạo pha theo chu kỳ. Với tần số sóng mang, tần số điều chế và chỉ số điều chế cho trước, ta có thể dễ dàng dự đoán các thành phần ở phổ tần số của âm thanh kết quả.



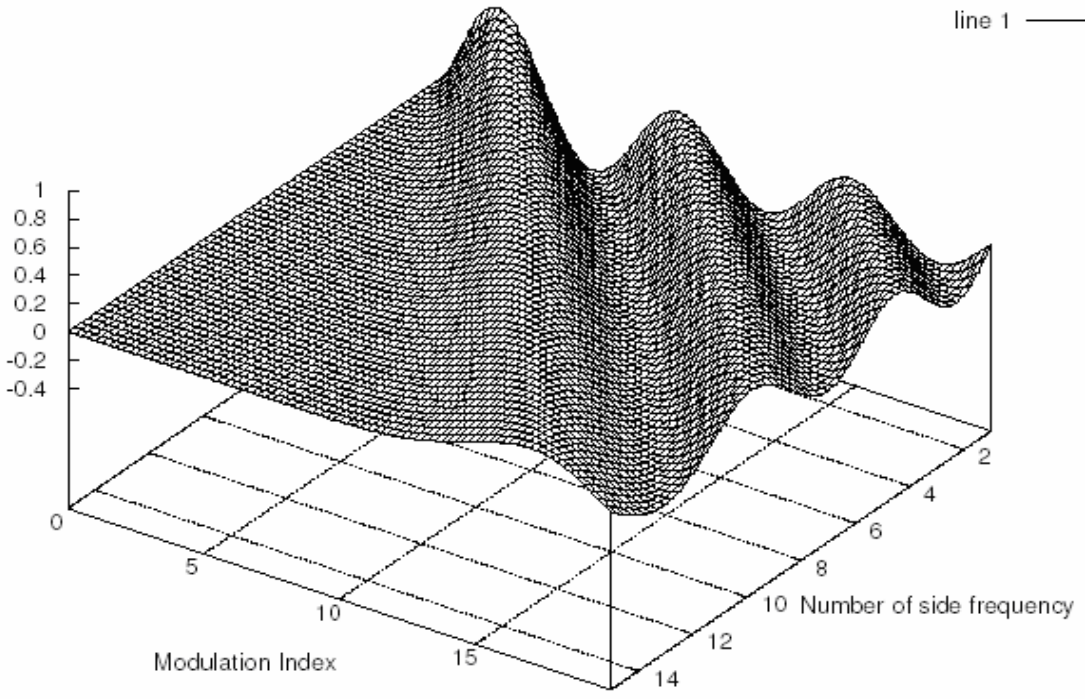
Hình 1.27 Phần triển khai phân phối *pd* của điều pha.

Việc phân tích dựa trên đặc tính lượng giác

$$x(n) = A\sin\left(\omega_c n + I\sin\left(\omega_m n\right)\right)$$

$$= A\left\{\underbrace{J_0(I)\sin(\omega_c n)}_{carrier} + \underbrace{\sum_{k=1}^{\infty} J_k(I)\left[\sin\left(\left(\omega_c + k\omega_m\right)n\right) + \left(-1\right)^k \sin\left(\left(\omega_c - k\omega_m\right)n\right)\right]}_{side-frequencies}\right\} \quad (1.38)$$

Với $J_k(I)$ là bậc thứ k của hàm Bessel. Các hàm Bessel được vẽ trên hình 9 ứng với nhiều giá trị k trên trục số lượng side-frequencies và giá trị I trên trục chỉ số điều chế.



Hình 1.28 Các giá trị của hàm Bessel.

Băng thông có giá trị xấp xỉ bằng

$$BW = 2\left(I + 0.24I^{0.27}\right)\omega_m \approx 2I\omega_m \tag{1.39}$$

1.2.2.3.2 Méo phi tuyến

Khái niệm tổng hợp âm thanh bằng méo phi tuyến – Nonlinear distortion (NLD) rất đơn giản: ngõ ra của mạch tạo dao động được dùng như là thông số của một hàm phi tuyến. Trong miền thời gian rời rạc số, hàm phi tuyến được lưu trữ trong một bảng, và ngõ ra của bộ dao động được dùng như là chỉ số để truy nhập vào bảng. Điều thú vị của NLD là lý thuyết này cho phép thiết kế một bảng méo cho bởi các đặc điểm kỹ thuật của một phổ mong muốn.

Nếu bộ tạo dao động có dạng tín hiệu sin, ta có thể tính toán NLD như sau

$$x(n) = A\cos(\omega_0 n) \tag{1.40}$$

$$y(n) = F\left(x(n)\right) \tag{1.41}$$

Với hàm số phi tuyến, dùng đa thức Chebyshev. Đa thức Chebyshev cấp độ n được định nghĩa đệ quy như sau:

$$T_0(x) = 1 \tag{1.42}$$

$$T_1(x) = x \tag{1.43}$$

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \tag{1.44}$$

và có tính chất:

$$T_n(\cos\theta) = \cos n\theta \tag{1.45}$$

Như vậy, với tính chất (31), nếu hàm méo phi tuyến là đa thức Chebyshev cấp độ m, giá trị ngõ ra y có được bằng cách sử dụng bộ dao động sin $x(n) = \cos\omega_0 n$, như vậy $y(n) = \cos\left(m\omega_0 n\right)$ là hài bậc m của x.

Phổ của y(n) với:

$$y(n) = \sum_k h_k \cos(k\omega_0 n) \tag{1.46}$$

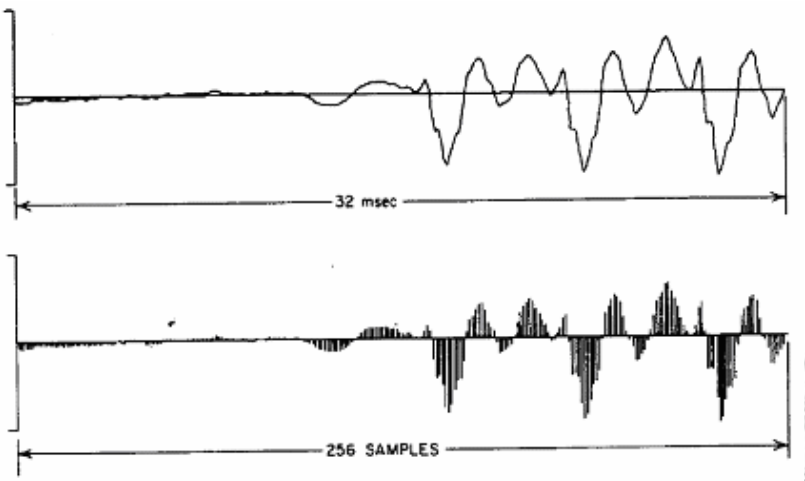
là:

$$F(x) = \sum_k h_k T_k(x) \tag{1.47}$$

Ngoài các mô hình trên, các mô hình vật lý cũng được áp dụng trong việc tổng hợp, xử lý âm thanh như mạch dao động vật lý, mạch dao động đôi và mạch phân phối cộng hưởng một chiều

1.2.3 Mô hình thời gian rời rạc [12]

Trong hầu hết các trường hợp liên quan đến xử lý thông tin, việc biểu diễn tín hiệu sao cho đảm bảo tính tiện lợi trong phân tích mà vẫn không làm mất đi tính chất của tín hiệu là điều mà các nhà khoa học quan tâm. Sóng âm thanh xuất phát từ lời nói của người có tính chất tự nhiên và ngẫu nhiên nhất. Phân tích toán học thuận tiện nhất là xem sóng âm thanh là một hàm số theo biến thời gian t. Ta ký hiệu $x_a(t)$ là dạng sóng tương tự theo thời gian t.



Hình 1.29 Biểu diễn tín hiệu âm thoại

Trong giáo trình này, ta dùng ký hiệu $x(n)$ mô tả cho chuỗi số. Trong trường hợp lấy mẫu tín hiệu âm thoại, một chuỗi có thể được xem như là một dãy các mẫu của tín hiệu tương tự được lấy mẫu một cách đều đặn với thời gian lấy mẫu là T, khi đó tín hiệu sau khi lấy mẫu được ký hiệu bởi $x_a(nT)$. Hình 1.1 mô tả một ví dụ của việc tín hiệu âm thoại được biểu diễn ở cả hai dạng là tín hiệu tương tự và dạng chuỗi các mẫu được lấy mẫu ở tần số là 8kHz.

Xung đơn vị được định nghĩa như sau:

$$\delta(n) = 1 \; n = 0$$
$$= 0 \text{ ngược lại} \tag{1.48}$$

Chuỗi bước đơn vị được ký hiệu

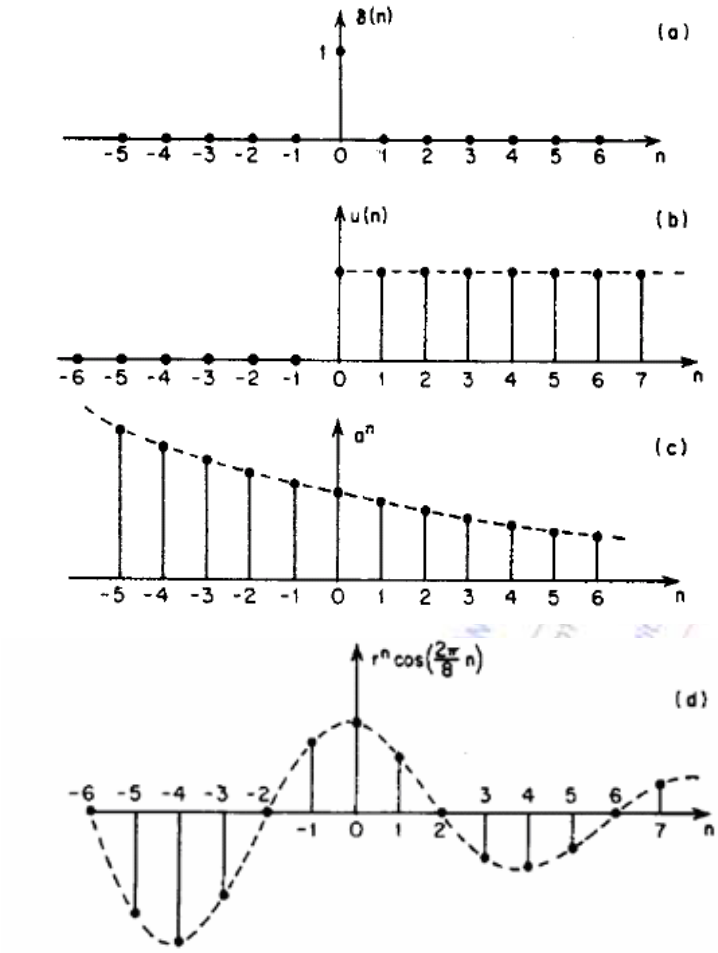
$$u(n) = 1 \; n \geq 0$$
$$= 0 \; n < 0 \tag{1.49}$$

Hàm mũ

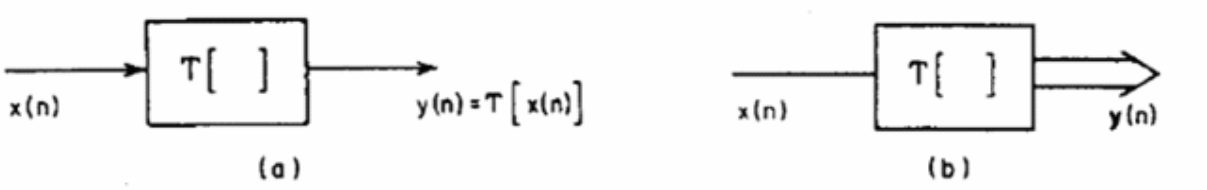
$$x(n) = a^n \tag{1.50}$$

Nếu a ở dạng số phức, $a = re^{jw_0}$, thì

$$x(n) = r^n e^{jw_0 n} = r^n(\cos \omega_0 n + j \sin \omega_0 n) \tag{1.51}$$



Hình 1.30 (a) Lấy mẫu đơn vị, (b) đơn vị bước, (c) hàm mũ thực và (d) hàm sin suy giảm



Hình 1.31 Sơ đồ khối (a) hệ thống đơn ngõ vào/đơn ngõ ra; (b) hệ thống đơn ngõ vào/đa ngõ ra

Khi hệ thống gồm nhiều ngõ ra, tín hiệu chuỗi ngõ ra sẽ được biểu diễn bằng một vector được mô tả như ở Hình 1.31.

Hệ thống tuyến tính dịch bất biến là hệ thống đặc biệt hữu dụng cho việc xử lý tín hiệu âm thoại. Hệ thống được đặc trưng bởi đáp ứng xung, $h(n)$, khi đó tín hiệu ngõ ra được tính bởi công thức

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) = x(n) * h(n) \qquad (1.52a)$$

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) = h(n) * x(n) \qquad (1.52b)$$

với * là phép chập hai tín hiệu

1.3 LÝ THUYẾT VÀ CÁC BÀI TOÁN CƠ BẢN

1.3.1 Phân tích dự đoán tuyến tính [12]

Dự đoán tuyến tính (Linear prediction, viết tắt là LP) là một phần không thể thiếu của hầu hết tất cả giải thuật mã hóa thoại hiện đại ngày nay. Ý tưởng cơ bản là một mẫu thoại có thể được xấp xỉ bằng một kết hợp tuyến tính của các mẫu trong quá khứ. Trong một khung tín hiệu, các trọng số dùng để tính toán kết hợp tuyến tính được tìm bằng cách tối thiểu hóa bình phương trung bình lỗi dự đoán; các trọng số tổng hợp, hoặc các hệ số dự đoán tuyến tính (LPC) được dùng đại diện cho một khung cụ thể.

Trong phần chương 3, sự sắp xếp LP theo hệ thống dựa trên mô hình ngược tự động

Trong thực tế, phân tích dự là một tiến trình ước lượng để tìm các thông số của AR, mà các thông số này được cho bởi các mẫu của tín hiệu. Như vậy, LP là một kỹ thuật nhận dạng với các thông số của một hệ thống đực tìm từ việc quan sát. Với giả định là tín hiệu thoại được mô hình như là tín hiệu AR, điều này đã được chứng minh tính đúng đắn của nó trong thực tiễn.

Một cách biểu diễn LP khác là phương pháp ước lượng phổ. Như đã trình bày ở trên, phân tích LP cho phép việc tính toán các thông số của AR, đã được định nghĩa trong mật độ phổ công suất (PSD) của chính bản thân tín hiệu. Bằng cách tính toán LPC của một khung tín hiệu, ta có thể tạo ra một tín hiệu khác theo cách thức có nội dung phổ gần như tương đồng với tín hiệu gốc.

LP cũng có thể được xem như là một quá trình loại bỏ các dư thừa khi thông tin bị lặp lại trong một sự trường hợp cần khử. Sau cùng, việc truyền dữ liệu có thể không cần thiết nếu như dữ liệu cần truyền có thể được dự đoán trước. Bằng cách thức chuyển chỗ các dư thừa trong một tín hiệu, số lượng bit cần thiết để mang thông tin sẽ ít hơn và như thế sẽ đạt được mục tiêu nén dữ liệu.

Trong phần này sẽ đề cập đến bài toán cơ bản của phân tích LP đã được định rõ, kết hợp với việc hiệu chỉnh lại cho phù hợp theo hướng các tín hiệu động, cũng như ví dụ và các giải thuật cần thiết cho quá trình dự đoán tuyến tính.

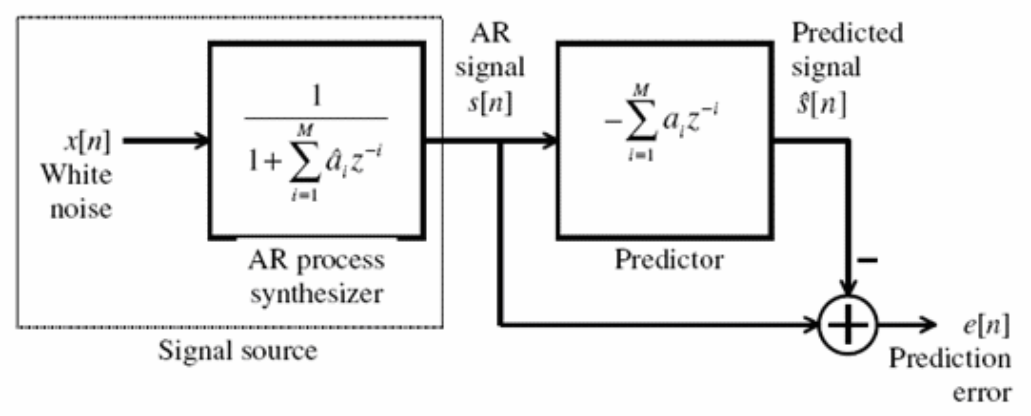
1.3.1.1 Bài toán dự đoán tuyến tính

Dự đoán tuyến tính được mô tả như là một bài toán nhận dạng hệ thống, với các thông số của một mô hình AR được ước lượng từ bản thân tín hiệu. Mô hình được trình bày ở Hình 1.32. Tín hiệu nhiễu trắng $x[n]$ được lọc bởi quá trình tổng hợp AR để có được tín hiệu AR $s[n]$, với các thông số AR được ký hiệu là $\hat{a}_i$. Dự đoán tuyến tính thực hiện ước đoán $s[n]$ dựa vào M mẫu trong quá khứ:

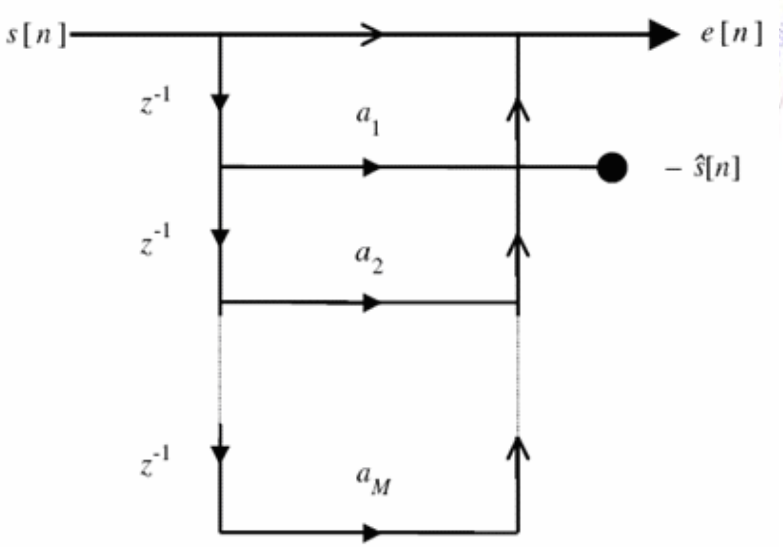
$$\hat{s}[n] = -\sum_{i=1}^{M} a_i s[n-i] \qquad (1.53)$$

Với a_i là các ước lượng của các thông số AR được xem là các hệ số dự đoán tuyến tính (LPC). Hằng số M trong công thức là bậc dự đoán. Như vậy, việc dự đoán dự trên tổ hợp tuyến tính của M mẫu trong quá khứ của tín hiệu, chính vì thế việc dự đoán mang tính tuyến tính. Lỗi dự đoán được tính bằng công thức:

$$e[n] = s[n] - \hat{s}[n] \tag{1.54}$$



Hình 1.32 Hệ thống nhận dạng dưới dạng dự đoán tuyến tính



Hình 1.33 Bộ lọc lỗi dự đoán

Lỗi dự đoán chính là độ sai biệt giữa mẫu thật sự và mẫu ước lượng. Hình 1.33 mô tả lưu đồ tín hiệu thực hiện bộ lọc lỗi dự đoán. Bộ lọc có ngõ vào là tín hiệu AR và ngõ ra chính là tín hiệu lỗi dự đoán.

Tối thiểu hoá lỗi

Bài toán nhận dạng hệ thống bao gồm việc ước lượng các thông số AR $\hat{a}_i$ từ $s[n]$. Để thực hiện việc ước lượng, tiêu chuẩn phải được thiết lập. Trong đó, bình phương trung bình lỗi dự đoán được tính bởi công thức:

$$J = E\{e^2[n]\} = E\left\{\left(s[n] + \sum_{i=1}^{M} a_i s[n-i]\right)^2\right\} \tag{1.55}$$

Được tối thiểu hóa bằng cách lựa chọn LPC thích hợp. Thông số LPC tối ưu có thể được tìm bằng cách thiết lập các đạo hàm riêng phần của J khi a_i tiến tới zero:

$$\frac{\partial J}{\partial a_k} = 2E\left\{\left(s[n] + \sum_{i=1}^{M} a_i s[n-i]\right)s[n-k]\right\} = 0 \qquad (1.56)$$

Với $k = 1,2,...,M$, khi (4.4) xảy ra thì $a_i = \hat{a}_i$, lúc này LPC chính bằng các thông số AR.

Độ lợi dự đoán

Độ lợi dự đoán của bộ dự đoán được cho bởi công thức

$$PG = 10\log_{10}\left(\frac{\sigma_s^2}{\sigma_e^2}\right) = 10\log_{10}\left(\frac{E\{s^2[n]\}}{E\{e^2[n]\}}\right) \qquad (1.57)$$

Là tỉ số giữa biến tín hiệu ngõ vào và biến của lỗi dự đoán theo đơn vị decibels (dB). Độ lợi dự đoán là thông số đo lường chất lượng của bộ dự đoán. Một bộ dự đoán tốt hơn có khả năng tạo ra lỗi dự đoán nhỏ hơn với độ lợi cao hơn.

Tối thiểu hóa bình phương trung bình lỗi dự đoán

Từ Hình 1.33, ta có thể nhận xét khi $a_i = \hat{a}_i$, thì $e[n] = x[n]$; như vậy lỗi dự đoán tương tự như dùng tín hiệu nhiễu trắng để tạo ra tín hiệu AR $s[n]$. Đây là trường hợp tối ưu khi lỗi bình phương trung bình được tối thiểu hóa, với

$$J_{\min} = E\{e^2[n]\} = E\{x^2[n]\} = \sigma_x^2 \qquad (1.58)$$

Khi đó, độ lợi dự đoán đạt giá trị lớn nhất.

Điều kiện tối ưu có thể đạt được khi bậc của bộ dự đoán lớn hơn hoặc bằng bậc của quá trình tổng hợp AR. Trong thực tế, M thường là số chưa biết trước. Một phương pháp đơn giản để có thể ước lượng được giá trị M từ tín hiệu nguồn là vẽ biểu đồ độ lợi dự đoán như là một hàm của bậc dự đoán. Với phương pháp này, ta có thể quyết định được bậc của dự đoán ứng với độ lợi bão hòa, khi đó khi tăng bậc dự đoán thì độ lợi không tăng. Giá trị của bậc dự đoán tại điểm thỏa điều kiện bão hòa này được xem là giá trị ước lượng tốt nhất cho bậc của tín hiệu AR.

Sau khi đã xác định được giá trị M, hàm chi phí J đạt giá trị tối thiểu khi $a_i = \hat{a}_i$, dẫn đến $e[n] = x[n]$. Và khi đó, lỗi dự đoán sẽ bằng với giá trị tín hiệu đầu vào của bộ tổng hợp quá trình AR.

1.3.1.2 Phân tích dự đoán tuyến tính cho tín hiệu động

Tín hiệu thoại trong thực tế là tín hiệu động, nên LPC phải được tính ứng với từng khung tín hiệu. Trong một khung tín hiệu, một tập LPC được tính toán và dùng để đại diện cho các thuộc tính của tín hiệu trong một chu kỳ cụ thể, với giả định rằng số liệu thống kê của tín hiệu vẫn không thay đổi trong một khung. Quá trình tính toán LPC từ dữ liệu tín hiệu được gọi là phân tích dự đoán tuyến tính.

Bài toán dự đoán tuyến tính cho tín hiệu động được phát biểu lại như sau: đây là bài toán thực hiện việc tính các giá trị LPC ứng với N điểm dữ liệu với thời gian kết thúc là m: $s[m-N+1]$, $s[m-N+2]$, …, $s[m]$. Vector LPC được viết như sau:

$$a[m] = [a_1[m] \quad a_2[m] \quad ... \quad a_M[m]]^T \qquad (1.59)$$

Với M là bậc dự đoán

Độ lợi dự đoán

Độ lợi dự đoán của bộ dự đoán được cho bởi công thức

$$PG[m] = 10\log_{10}\left(\frac{\sum_{n=m-N+1}^{m} s^2[n]}{\sum_{n=m-N+1}^{m} e^2[n]}\right) \tag{1.60}$$

Với

$$e[n] = s[n] - \hat{s}[n] = s[n] + \sum_{i=1}^{M} a_i[m]s[n-i];\ n = m-N+1, \ldots, m \tag{1.61}$$

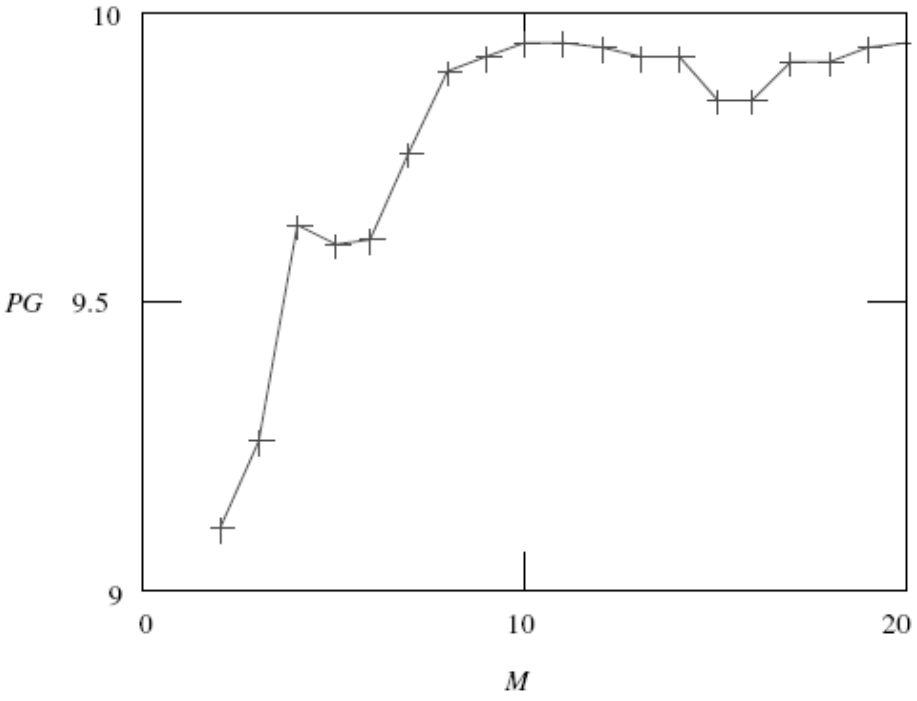
Các LPC $a_i[m]$ được tính toán từ các mẫu trong chu kỳ. Độ lợi dự đoán định nghĩa ở công thức (4.23) là một hàm theo biến thời gian m.

Ví dụ: Nhiễu trắng được tạo ra bởi bộ tạo số ngẫu nhiên phân phối đều, sau đó được lọc bởi bộ tổng hợp AR với

$a_1 = 1.534$ $a_2 = 1$ $a_3 = 0.587$ $a_4 = 0.347$ $a_{56} = 0.08$

$a_6 = -0.061$ $a_7 = -0.172$ $a_8 = -0.156$ $a_9 = -0.157$ $a_{10} = -0.141$

Khung tổng hợp của tín hiệu AR được dùng cho phân tích LP, với tổng cộng là 240 mẫu. Ước lượng tự động tương quan không hồi qui sử dụng cửa sổ Hamming. Phân tích LP được thực hiện với bậc từ 2 đến 20. Hình 1.34 tóm tắt kết quả, với độ lợi dự đoán được tính toán tại $M = 2$ và đạt giá trị cao nhất tại $M = 10$. Các bậc lớn hơn 10 không cho được độ lợi cao hơn nữa, cho nên ta có thể chỉ cần xét đến $M = 10$.



Hình 1.34 Độ lợi dự đoán (PG) là một hàm theo biến bậc dự đoán M

1.3.1.3 Giải thuật Levison-Durbin

Thông thường, việc tính toán ma trận nghịch đảo không đơn giản, tuy nhiên đối với bài toán này, các hệ số giải thuật có thể được tính thông qua tính toán ma trận tương quan. Hai giải thuật Levison-Durbin và Leroux-Gueguen là hai giải thuật rất phù hợp cho việc tính toán LP của các hệ thống triển khai trong thực tế.

Giải thuật Levison-Durbin thực hiện việc tìm bộ dự đoán bậc thứ M từ bộ dự đoán bậc thứ $(M-1)$. Đây là quá trình lặp đệ quy cho đến khi tìm được lần đầu tiên bộ dự đoán bậc zero, sau đó sẽ dùng bộ bậc zero sẽ được dùng để tính bộ dự đoán bậc 1 và quá trình tiếp tục cho đến khi tính toán được bộ dự đoán có bậc cần tìm.

Giải thuật: biến đầu vào là các hệ số tự tương quan $R[l]$, giá trị tính được là các LPC và RC

- Định trị ban đầu: $l=0$, tập $J_0 = R[0]$
- Thực hiện đệ quy, $for \quad l = 1,2,...,M$
  - Bước 1: Tính toán RC thứ l, $k_l = \frac{1}{J_{l-1}}\left(R[l]\right) + \sum_{i=1}^{l-1} a_i^{(l-1)} R[l-i]$,
  - Bước 2: Tính toán các LPC ứng với bộ dự đoán bậc l

    $a_i^{(l)} = -k_l;$

    $a_i^{(l)} = a_i^{(l-1)} - k_l a_{l-i}^{(l-1)}; \quad i-1,2,...,l-1$

    Dừng nếu $l = M$
  - Bước 3: Tính giá trị bình phương trung bình lỗi dự đoán tương ứng với lời giải tại bậc l

    $J_l = J_{j-1}\left(1-k_l^2\right)$

    Gán $l = l+1$, quay lại bước 1

1.3.1.4 Giải thuật Leroux-Gueguen

Bài toán sử dụng giải thuật Levinson-Durbin dựa trên các giá trị của các LPC, bởi vì chúng có thuộc một tầm vực rộng và giá trị biên của biên độ của các LPC không thể tính được ứng với cơ sở lý thuyết. Vấn đề xảy ra khi giải thuật được áp dụng cho tính toán trên dấu chấm tĩnh. Giải thuật Leroux-Gueguen khắc phục điểm yếu này của giải thuật Levison-Durbin.

Leroux và Gueguen [1979] đã đề xuất một phương pháp tính toán các RC từ các giá trị tự tương quan mà không cần phải tính thông qua các LPC. Do đó, bài toán liên quan đến tầm động với điều kiện dấu chấm tĩnh đã được giải quyết. Xét thông số sau

$$\varepsilon^{(l)}[k] = E\left\{e^{(l)}[n]s[n-k]\right\} = \sum_{i=0}^{l} a_i^{(l)} R[i-k], \tag{1.62}$$

Với

- $e^{(l)}[n]$ = lỗi dự đoán sử dụng bộ lọc dự đoán lỗi bậc thứ l
- $a_i^{(l)}$ = LPC của bộ dự đoán bậc thứ l

- $R[k]$ = giá trị tự tương quan của tín hiệu $s[n]$

Định lý:

$$\varepsilon^{(l)}[k] \le R[0] \tag{1.63}$$

Sinh viên có thể tự chứng minh

Bảng 1.4 mô tả các thông số ε cần thiết ứng với mỗi bậc l trong giải thuật Leroux-Gueguen

l	Các thông số cần thiết
M	
$M-1$	$\varepsilon^{(M-1)}[0], \varepsilon^{(M-1)}[M]$
$M-2$	$\varepsilon^{(M-2)}[-1], \varepsilon^{(M-2)}[0], \varepsilon^{(M-2)}[M-1], \varepsilon^{(M-2)}[M]$
$M-3$	$\varepsilon^{(M-3)}[-2], \ldots, \varepsilon^{(M-3)}[0], \varepsilon^{(M-3)}[M-2], \ldots, \varepsilon^{(M-3)}[M]$
$M-4$	$\varepsilon^{(M-4)}[-3], \ldots, \varepsilon^{(M-4)}[0], \varepsilon^{(M-4)}[M-3], \ldots, \varepsilon^{(M-4)}[M]$
$\vdots$	
1	$\varepsilon^{(1)}[-M+2], \ldots, \varepsilon^{(1)}[0], \varepsilon^{(1)}[2], \ldots, \varepsilon^{(1)}[M]$
0	$\varepsilon^{(0)}[-M+1], \ldots, \varepsilon^{(0)}[0], \varepsilon^{(0)}[1], \ldots, \varepsilon^{(0)}[M]$

Giải thuật:

- Định trị ban đầu: $l = 0$, tập $\varepsilon^{(0)}[k] = R[k], k = -M+1,...,M$
- Thực hiện đệ quy, *for* $l = 1,2,...,M$
  - Bước 1: Tính toán RC thứ l, $k_l = \dfrac{\varepsilon^{(l-1)}[l]}{e^{(l-1)}[0]}$, dừng khi $l = M$
  - Bước 2: Tính toán các thông số

$$\varepsilon^{(l)}[k] = \varepsilon^{(l-1)}[k] - k_l \varepsilon^{(l-1)}[l-k]; \quad k = -M+l+1,...,0,l+1,...M.$$

  - Gán $l = l+1$, quay lại bước 1

1.3.1.5 So sánh giải thuật Levison-Durbin và Leroux-Gueguen

Giải thuật Leroux-Gueguen phù hợp hơn cho các bài toán dấu chấm tĩnh đối với các biến trung gian có biên đã được biết trước. Nhược điểm của giải thuật này là chỉ có các giá trị RC là kết quả trả về, là kết quả không cần thiết đối với bộ lọc lưới. Đối vóic các bộ lọc có dạng trực tiếp, các giá trị LPC có thể có được nếu dùng một trong hai giải thuật.

Việc sử dụng bộ lọc mắt cao thường trong việc tính toán LP thường không đơn giản do số lượng tính toán. Ngoài ra, đối với trường hợp thời gian biến đổi, các hệ số được cập nhật từ khung thời gian này đến khung thời gian khác sẽ làm cho việc tính toán càng phức tạp hơn đối với cấu trúc lưới. Ngoài ra, phương pháp Leroux-Gueguen sử dụng biến đổi RC-sang-LPC không cung cấp việc lưu trữ lại các bước tính toán quan trọng so với giải thuật Levinson-Durbin. Tất cả các điều trên làm cho giải thuật Levinson-Durbin thông dụng hơn trong thực tiễn, đặc biệt là đối với các bài toán số.

Trong các bài toán ứng dụng thực tế, giải thuật Levison-Durbin dùng trong điều kiện dấu chấm tĩnh phải được cân nhắc kỹ sao cho đảm bảo các biến phải nằm trong tầm vực cho phép.

1.3.2 Dự đoán tuyến tính trong xử lý thoại [13]

Đối với việc đơn giản hóa mô hình xử lý thoại, giải thuật dự đoán tuyến tính (LPC) là một trong những giải thuật áp dụng tạo các bộ mã hóa chuẩn cho việc xử lý âm hoạt động ở tần số thấp. Ở tốc độ 2.4kbps, bộ mã hóa FS1015 LPC [Hãng Tremain, 1982] là một bước tiến vượt bật trong ngành xử lý âm thanh; mặc dù chất lượng của âm thanh được giải mã không cao, nhưng hệ thống giải mã đơn giản và dễ hiểu. Thuật ngữ "mã hóa dự đoán tuyến tính" xuất hiện từ khi việc tạo ra âm thanh thoại sử dụng bất kỳ giải thuật ứng dụng mô hình LPC, trong đó chuẩn FS1015 là chuẩn điển hình.

Ban đầu, trong việc phát triển cho việc truyền thông bảo mật thuộc các ứng dụng quân sự, bộ mã hóa FS1015 được đặc trưng bởi tín hiệu thoại mã tổng hợp ngõ ra thường cần đến các nhân viên vận hành tổng đài đã được huấn luyện sử dụng. Mặc dù hầu hết các bộ mã hóa thoại dựa vào công nghệ LP đạt được hiệu suất cao hơn ngày nay, nhưng về cơ bản, hoạt động của chúng là có nguồn gốc từ LPC, việc cải tiến nhằm mục đích đạt được chất lượng tốt hơn và hiệu suất mã hóa tối ưu hơn.

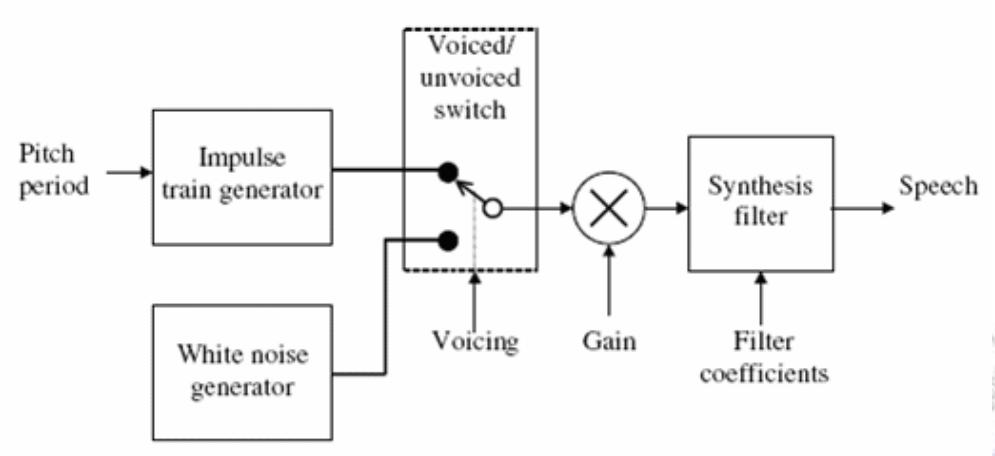
1.3.2.1 Mô hình xử lý tín hiệu thoại

Mô hình xử lý thoại dựa vào mô hình mã hóa dự đoán tuyến tính được mô tả trong Hình 1.35. Mô hình được dựng dựa vào việc quan sát các đặc tính cơ bản của tín hiệu thoại và bắt chước kỹ thuật tạo âm thanh thoại của người. Bộ lọc tổng hợp được mô phỏng theo sự phát âm, khẩu âm của miệng người. Tín hiệu lái ngõ vào của bộ lọc hoặc tín hiệu kích thích mạch được mô phỏng theo dạng xung truyền động (âm thanh thoại) hoặc là nhiễu ngẫu nhiên (âm thanh phi thoại). Như vậy, phụ thuộc vào trạng thái âm thanh thoại hay phi thoại của tín hiệu, mạch chuyển được thiết lập ở vị trí thích hợp sao cho ngõ vào tương ứng sẽ được chọn đưa vào mạch. Mức năng lượng của tín hiệu ngõ ra đwojc điều khiển bởi thông số độ lợi.

Làm cách nào mô hình phù hợp với ngữ cảnh của mã hóa âm thoại? Xét các mẫu thoại một cách riêng lẻ ứng với từng khung tín hiệu không chồng lên nhau. Ứng với từng đoạn khung đủ ngắn, thuộc tính của tín hiệu về cơ bản là hằng số. Trong mỗi khung, các thông số của mô hình được ước lượng từ các mẫu thoại, các thông số bao gồm:

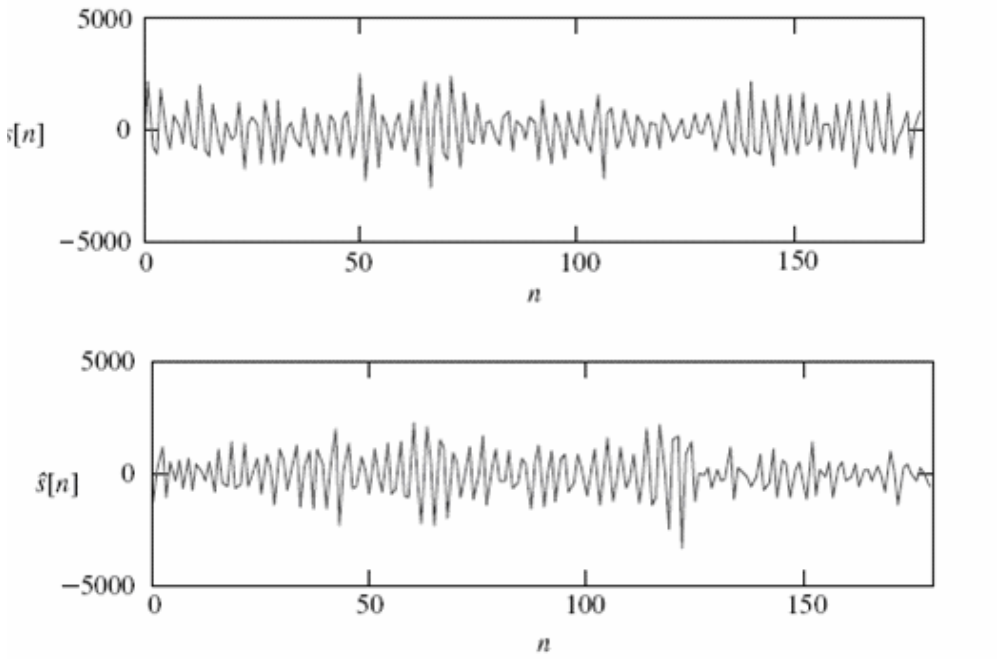
- Dạng: tín hiệu thuộc khung là thoại hay phi thoại
- Độ lợi: liên quan chủ yếu đến mức năng lượng của khung
- Hệ số lọc: định rõ đáp ứng của bộ lọc tổng hợp
- Chu kỳ âm thanh: trong trường hợp đối với khung thoại, là chiều dài thời gian giữa các xung kích thích liên tiếp nhau.

Quá trình ước lượng thông số được thực hiện ứng với từng mỗi khung, các kết quả chính là các thông tin của khung. Như vậy, thay vì truyền các xung PCM, các thông số của mô hình sẽ được gửi đi. Giảm giảm thiểu nhiễu và sự méo tín hiệu, các bít truyền được cấp phát theo chỉ định ứng với từng thông số, và tỉ số nén tối ưu có thể đạt được.



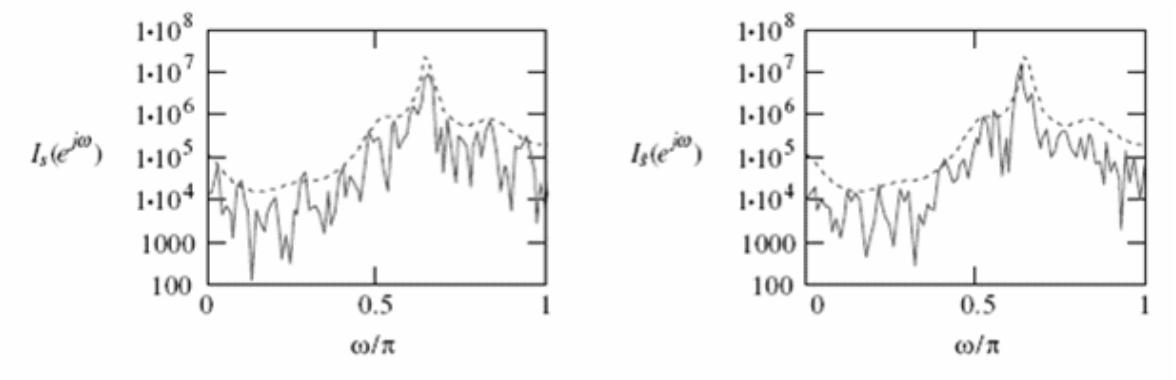
Hình 1.35 Mô hình LPC tổng hợp tiếng nói

Việc ước lượng các thông số là nhiệm vụ của bộ mã hóa. Bộ giải mã sẽ sử dụng các thông số ước lượng này và dùng mô hình tạo thoại để tổng hợp âm thoại.

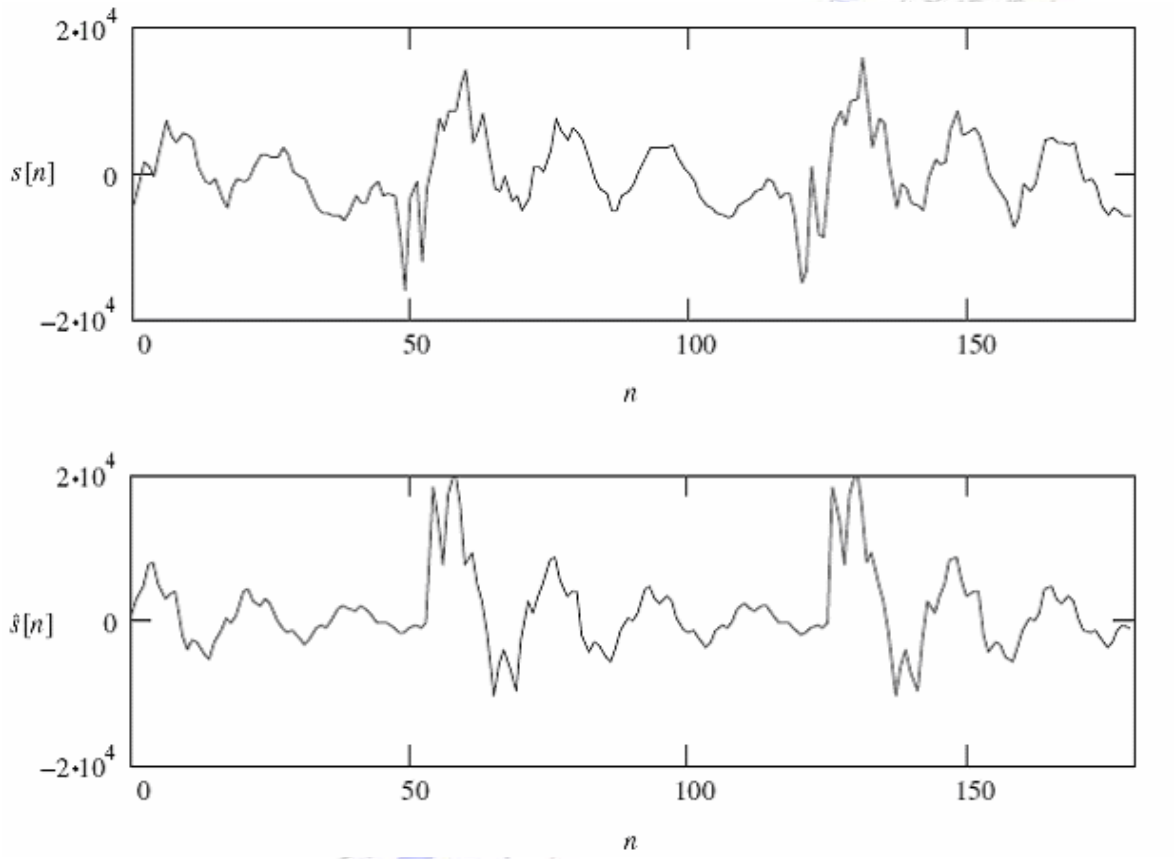


Hình 1.36 Hình vẽ các các khung phi thoại

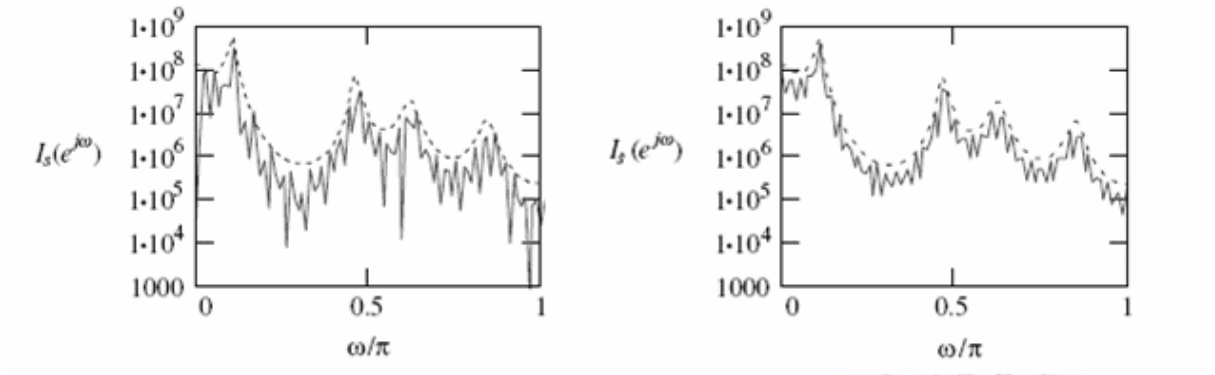
Hình 1.36 mô tả một khung tín hiệu phi thoại có 180 mẫu (sử dụng bộ mã hóa FS1015). Các mẫu nguyên thủy được xử phân tích LP qua quá trình tổng hợp LPC dùng cho việc tổng hợp âm thoại dựa trên mô hình Hình 1.35. Tín hiệu của tín hiệu nguyên thủy và tín hiệu sau khi tổng hợp có vẻ giống nhau do mật độ phổ cố suất có dạng tương đương, được mô tả trong Hình 1.37.



Hình 1.37 Sơ đồ của một khung âm thanh phi thoại, Hình bên trái: tín hiệu nguyên thuỷ; Hình bên phải: tín hiệu tổng hợp. Đường nét đứt là giá trị mật độ phổ công suất dùng phương pháp dự đoán LPC.



Hình 1.38 Sơ đồ khung tín hiệu âm thanh thoại. Hình trên: tín hiệu nguyên thủy; Hình dưới: tín hiệu tổng hợp.



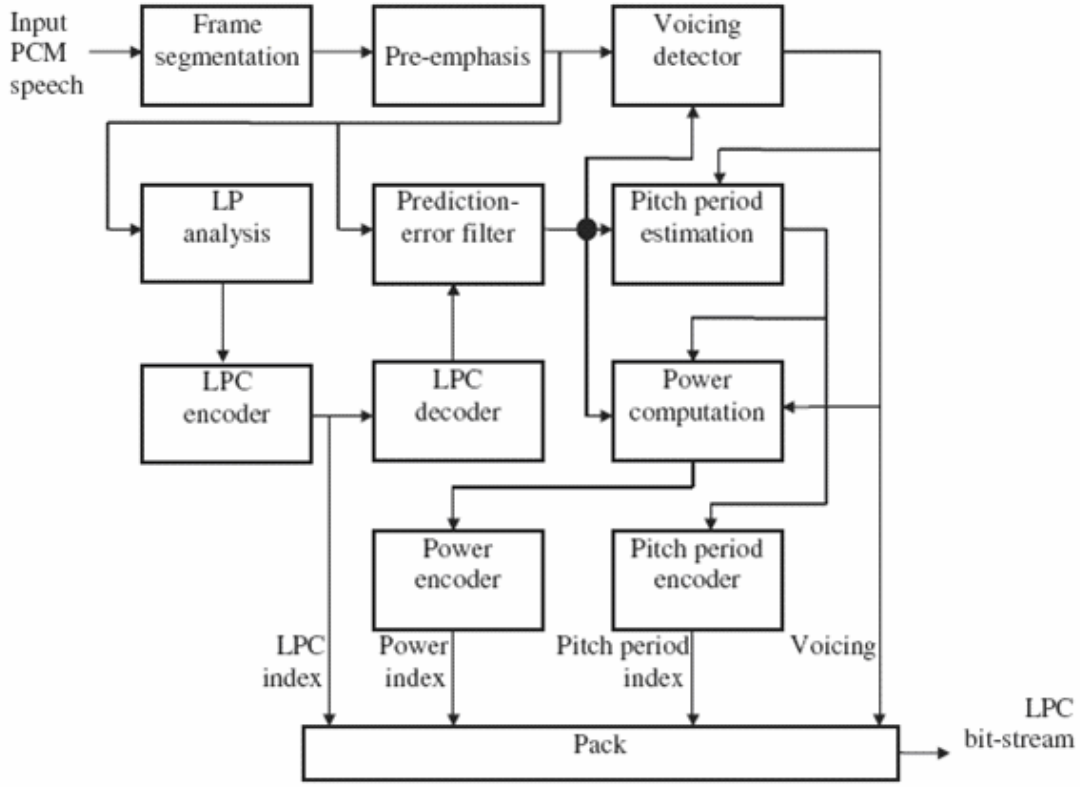
Hình 1.39 Sơ đồ của một khung âm thanh thoại, Hình bên trái: tín hiệu nguyên thuỷ; Hình bên phải: tín hiệu tổng hợp. Đường nét đứt là giá trị mật độ phổ công suất dùng phương pháp dự đoán LPC.

1.3.2.2 Cấu trúc của giải thuật dùng mô hình LPC

1.3.2.2.1 Bộ mã hóa (Encoder)

Hình 1.40 mô tả sơ đồ khối của bộ mã hóa. Tín hiệu thoại ngõ vào đầu tiên sẽ được phân cắt thành các khung tín hiệu không chồng lên nhau. Bộ lọc đầu dùng để hiệu chỉnh phổ của tín hiệu ngõ vào, bộ nhận dạng tiếng nói phân loại khung hiện tại đang xử lý là tín hiệu thoại hay phi thoại và các ngõ ra một bit biểu thị trạng thái của âm thoại.

Tín hiệu ra của bộ lọc đầu được dùng để phân tích LP, mạch bao gồm mười LPC . Các hệ số này sẽ được lượng tử hóa với các chỉ số được truyền như là thông tin của khung. Các LPC được lượng tử hoá dùng để xây dựng bộ lọc dự đoán lỗi, dùng để lọc các tín hiệu âm thanh từ bộ lọc đầu để có được tín hiệu dự đoán lỗi ở ngõ ra.



Hình 1.40 Sơ đồ khối của bộ mã hóa LPC

1.3.2.2.2 Tính toán công suất

Công suất của chuỗi lỗi dự đoán ứng với hai trường hợp khung thoại và khung phi thoại là khác nhau. Ký hiệu chuỗi lỗi dự đoán là $e[n]$, $n \in [0, N-1]$ với N là chiều dài của khung.

Trường hợp tín hiệu là phi thoại:

$$p = \frac{1}{N}\sum_{n=0}^{N-1} e^2[n] \tag{1.64}$$

Trường hợp tín hiệu là âm thoại, T là chu kỳ lớn nhất của tín hiệu thành phần

$$p = \frac{1}{[N/T]T}\sum_{n=0}^{[N/T]T-1} e^2[n] \tag{1.65}$$

Với $[\cdot]$ là hàm tính giá trị nhỏ hơn hoặc bằng với toán hạng. Giả sử rằng $N > T$ thì việc dùng $[\cdot]$ luôn đảm bảo rằng việc tính toán luôn nằm trong vùng biên của khung.

1.3.2.2.3 Bộ giải mã

Hình 1.41 mô tả sơ đồ khối của bộ giải mã theo mô hình tạo âm LPC với các thông số được điều khiển bởi luồng bit. Giả sử rằng của bộ tạo chuỗi xung tạo ra các xung có biên độ đơn vị, trong khi bộ tạo nhiễu trắng có tín hiệu ngõ ra có biên độ khác đại lượng đơn vị.

Việc tính toán độ lợi được thực hiện như sau: Đối với tín hiệu là phi thoại, công suất của tín hiệu của bộ lọc tổng hợp phải bằng với lỗi dự đoán của bộ mã hóa. Ký hiệu độ lợi là g, ta có

$$g = \sqrt{p} \tag{1.66}$$

1.3.2.2.4 Giới hạn của mô hình LPC

Giới hạn 1: Trong một số trường hợp, một khung âm thanh có được phân loại là tín hiệu dạng thoại hay phi thoại.

Giới hạn 2: Việc sử dụng hoàn toàn nhiễu ngẫu nhiên hoặc hoàn toàn chuỗi xung có chu kỳ tạo kích thích không phù hợp với thực tế là sử dụng tín hiệu âm thoại thực.

Giới hạn 3: Thông tin về pha của tín hiệu nguyên thủy không được xem xét.

Giới hạn 4: Phương pháp thực hiện việc tổng hợp các khung thoại, trong khi một chuỗi xung dùng để kích thích bộ lọc tổng hợp với các hệ số có được từ việc phân tích LP vi phạm nền tảng của mô hình AR.

1.4 PHÂN TÍCH CHẤT LƯỢNG XỬ LÝ THOẠI

1.4.1 Các phương pháp mã hoá

Dịch vụ thoại là dịch vụ cơ bản và quan trọng nhất trong các dịch vụ cung cấp cho khách hàng của các nhà khai thác di động ở Việt Nam cũng như trên thế giới. Để đảm bảo hỗ trợ tốt khách hàng nhằm đạt được lợi thế cạnh tranh, các nhà khai thác di động cần hỗ trợ tốt dịch vụ cơ bản này. Do đó, việc đánh giá các chỉ tiêu chất lượng chất lượng thoại có vai trò rất quan trọng. Các phương thức đánh giá chất lượng thoại đã được nhiều tổ chức viễn thông (như ITU,ETSI...) nghiên cứu, xây dựng. Trong phần này trình bày một số phương pháp đánh giá chất lượng thoại, đặc biệt là cho mạng viễn thông (cố định, di động).

Việc đánh giá chất lượng thoại có vai trò rất quan trọng đối với các nhà khai thác mạng thông tin di động và cố định. Vì thoại là dịch vụ thông tin cơ bản cho nên việc đảm bảo cung cấp dịch vụ này với chất lượng ổn định là một yếu tố cạnh tranh của các nhà khai thác mạng.

Phương pháp đánh giá chất lượng thoại đã được nhiều tổ chức tiêu chuẩn như ITU-T, ETSI, 3GPP thực hiện chuẩn hóa. Bài báo phân tích bản chất của một số phương pháp đánh giá chất lượng thoại cơ bản: phương pháp đánh giá theo thang điểm MOS (Mean Opinion Score) dựa trên khuyến nghị ITU-T P.800 [1], các phương pháp đánh giá dựa trên mô hình giác quan PSQM (Perceptual Speech Quality Measurement) theo khuyến nghị ITU-T P.861 [2], PESQ (Perceptual Evaluation of Speech Quality) theo khuyến nghị ITU-T P.862 [3] và phương pháp dựa trên mô hình đánh giá truyền dẫn E-model theo tiêu chuẩn ETR 250 [4] của ETSI. Các phương pháp này được so sánh về ưu nhược điểm và phạm vi ứng dụng.

1.4.2 Các tham số liên quan đến chất lượng thoại

Các tham số truyền dẫn cơ bản liên quan đến chất lượng thoại là:

- Tham số đánh giá cường độ âm lượng/tổn hao tổng thể (OLR-Overall Loudness Rating): OLR của hệ thống phải không được vượt quá giới hạn được định nghĩa trong khuyến nghị G.111 của ITU-T. Các giá trị đánh giá tổn hao phía phát và thu (SLR và RLR) đối với hệ thống GSM được đánh giá cho đến giao diện POI. Tuy nhiên, tham số ảnh hưởng chính là đặc tính của MS gồm cả bộ chuyển đổi tương tự - số (ADC) và số tương tự (DAC). Do vậy, thông thường, người ta đánh giá OLR của giao diện vô tuyến.
- Trễ: thời gian truyền dẫn tín hiệu giữa hai đầu cuối gây ra những khó khăn trong việc hội thoại. Trễ bao gồm: trễ chuyển mã thoại, trễ mã hóa kênh, trễ mạng và trễ xử lý tín hiệu thoại để loại bỏ tiếng vọng và giảm nhiễu ở chế độ Handsfree.
- Tiếng vọng (echo).
- Cắt ngưỡng (clipping): là hiện tượng mất phần đầu hoặc phần cuối của cụm tín hiệu thoại.
- Các tính chất liên quan đến độ nhạy tần số.
- Xuyên âm (sidetone loss).
- Nhiễu nền...

1.4.3 Các phương pháp đánh giá chất lượng thoại cơ bản

Việc đánh giá chất lượng thoại trong mạng GSM cũng như các hệ thống thông tin khác (cố định và vô tuyến) có thể được thực hiện bằng cách đánh giá các tham số truyền dẫn có ảnh hưởng đến chất lượng thoại và xác định tác động của các tham số này đối với chất lượng tổng thể. Tuy nhiên, việc đánh giá từng tham số rất phức tạp và tốn kém. Hiện nay, việc đánh giá chất lượng thoại được dựa trên một tham số chất lượng tổng thể là MOS (Mean Opinion Score). Những phương pháp sử dụng MOS đều mang tính chất chủ quan do chúng phụ thuộc vào quan điểm của người sử dụng dịch vụ. Tuy vậy, chúng ta có thể phân chia các phương pháp đánh giá chất lượng thoại ra làm hai loại cơ bản:

- Các phương pháp đánh giá chủ quan: việc đánh giá theo quan điểm của người sử dụng về mức chất lượng được thực hiện trong thời gian thực. Phương pháp này được quy định trong khuyến nghị ITU-T P.800.
- Các phương pháp đánh giá khách quan: sử dụng một số mô hình để ước lượng mức chất lượng theo thang điểm MOS.

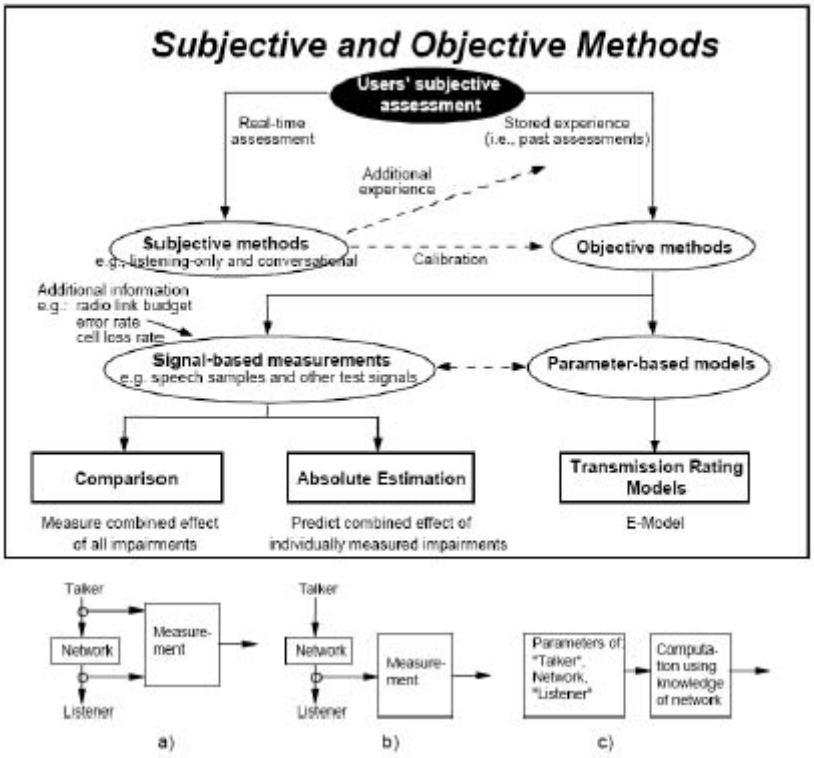
  Phương pháp đánh giá khách quan có thể được phân thành:

  a) *Các phương pháp so sánh:* dựa trên việc so sánh tín hiệu thoại truyền dẫn với một tín hiệu chuẩn đã biết.

b) *Các phương pháp ước lượng tuyệt đối:* dựa trên việc ước lượng tuyệt đối chất lượng tín hiệu thoại (phương pháp này không sử dụng các tín hiệu chuẩn đã biết); vd: INMD (sử dụng trong khuyến nghị P.561 của ITU-T).

c) *Các mô hình đánh giá truyền dẫn:* phương pháp này xác định giá trị chất lượng thoại mong muốn dựa trên những hiểu biết về mạng; vd: mô hình ETSI Model.

Việc phân chia các phương pháp đánh giá được cho trên Trên thực tế, các thiết bị đo có thể sử dụng kết hợp nhiều phương pháp đánh giá chất lượng thoại. Tuy vậy, trong các phương pháp này thì phương pháp so sánh (hay còn gọi là intrusive methods) có kết quả đánh giá chính xác nhất. Các phương pháp đánh giá các cũng có thể được sử dụng cho một số ứng dụng đặc thù.



Hình 1.41 Phân loại các phương pháp đánh giá chất lượng thoại: a) Các phương pháp so sánh, b) Các phương pháp ước lượng tuyệt đối, c) Các mô hình đánh giá truyền dẫn

1.4.3.1 Phương pháp đánh giá chủ quan (MOS)

Kỹ thuật này đánh giá chất lượng thoại sử dụng đối tượng là một số lượng lớn người nghe, sử dụng phương pháp thống kê để tính điểm chất lượng. Điểm đánh giá bình quân của nhiều người được tính là điểm Mean Opinion Scoring (MOS). Kỹ thuật thực hiện tính điểm MOS được mô tả trong khuyến nghị P.800 của ITU. Khuyến nghị P.830 thể hiện các phương pháp cụ thể để đánh giá chất lượng thoại cho các bộ mã hóa. Cả hai khuyến nghị ITU này mô tả: phương thức đánh giá, cách tính điểm theo phương thức đánh giá chủ quan, giá trị của điểm, tính chất của các mẫu thoại được sử dụng để đánh giá và các điều kiện khác mà việc kiểm tra chất lượng được thực hiện.

Phương thức đánh giá theo MOS có thể được thực hiện theo các bài kiểm tra hội thoại hai chiều hoặc bài nghe một chiều. Các bài kiểm tra nghe một chiều sử dụng các mẫu thoại chuẩn. Người nghe nghe mẫu truyền qua một hệ thống và đánh giá chất lượng tổng thể của mẫu dựa trên

thang điểm cho trước. P.800 định nghĩa một số hình thức đánh giá chất lượng thoại theo phương pháp chủ quan:

- Bài kiểm tra hội thoại (Conversation Opinion Test).
- Đánh giá phân loại tuyệt đối (Absolute Category Rating (ACR) Test).
- Phương thức phân loại theo suy hao (Degradation Category Rating (DCR)).
- Phương thức phân loại so sánh (Comparison Category Rating (CCR)).

Mỗi phương thức trên có một thang điểm đánh giá. Ví dụ: phương thức đánh giá hội thoại và ACR đều có thang điểm tương tự gọi là điểm hội thoại và điểm chất lượng nghe. Trong phương thức hội thoại, người nghe được hỏi về quan điểm của họ đối với kết nối đang sử dụng. ACR hỏi chủ thể về chất lượng thoại. Thang điểm cho cả hai phương thức trên như sau:

Điểm đánh giá	Chất lượng thoại
5	Rất tốt
4	Tốt
3	Chấp nhận được
2	Tồi
1	Rất tồi

Đây là thang điểm từ 1-5 thông thường được sử dụng để tính MOS.

Ví dụ thứ hai là điểm nỗ lực nghe trong phương thức ACR (ACR Listening Effort Score). Trong phương thức này, chủ thể được yêu cầu đánh giá nỗ lực của họ thực hiện để hiểu ngữ nghĩa của các câu chuẩn sử dụng làm mẫu. Thang điểm được cho như sau:

Điểm đánh giá	Mức độ cố gắng cần thực hiện để hiểu câu
5	Không cần cố gắng
4	Cần chú ý nhưng không cần cố gắng nhiều
3	Cần tương đối tập trung
2	Cần tập trung
1	Không hiểu câu mẫu

Hiển nhiên, các thương thức cho điểm theo MOS có một số nhược điểm như sau:

- Phương thức này mang tính chất chủ quan vì kết quả phụ thuộc vào nhiều yếu tố không thể kiểm soát của chủ thể như: trạng thái tâm lý, thái độ đối với bài kiểm tra và trình độ văn hóa. Trên thực tế, phương thức đánh giá chất lượng thoại theo thang điểm MOS không phải là phương thức nhất quán.

- Phương thức này rất tốn kém, đòi hỏi nhiều người tham gia và thiết lập phức tạp.
- Khi cần thực hiện đo thường xuyên các tham số chất lượng thì việc sử dụng phương pháp đánh giá chất lượng này là không thực tế.

Những hạn chế của phương pháp đánh giá chất lượng thoại dựa trên MOS cho thấy cần có một phương thức đánh giá khách quan, phương pháp này có thể thực hiện một cách tự động để đánh giá chất lượng thoại.

1.4.3.2 Các phương pháp so sánh dựa trên mô hình giác quan

1.4.3.2.1 Phương pháp PSQM

PSQM là kỹ thuật đánh giá chất lượng thoại được phát triển bởi John G. Beerends và J. A. Stemerdink thuộc Trung tâm nghiên cứu KPN ở Hà Lan. Trong khoảng từ 1993-1996, nhiều kỹ thuật đánh giá chất lượng thoại đã được ITU so sánh để xác định kỹ thuật có độ chính cao nhất (ước lượng gần nhất với phương pháp đánh giá chủ quan). Theo ITU, PSQM là kỹ thuật đánh giá chất lượng thoại có tương quan lớn nhất với các kết quả theo phương pháp đánh giá chủ quan. PSQM sau đó đã được ITU-T Study Group 12 thông qua và đã được công bố trong khuyến nghị P.861 năm 1996. Kỹ thuật này đã được sử dụng rộng rãi và thể hiện độ chính xác tương đối cao.

PSQM là một phương pháp tính toán nhằm ước lượng chất lượng thoại theo kết quả của phương pháp đánh giá chủ quan theo khuyến nghị P.830 (MOS). Tuy nhiên, PSQM tính theo thang điểm khác so với MOS. Điểm PSQM thể hiện độ lệch giữa tín hiệu chuẩn và tín hiệu truyền dẫn.

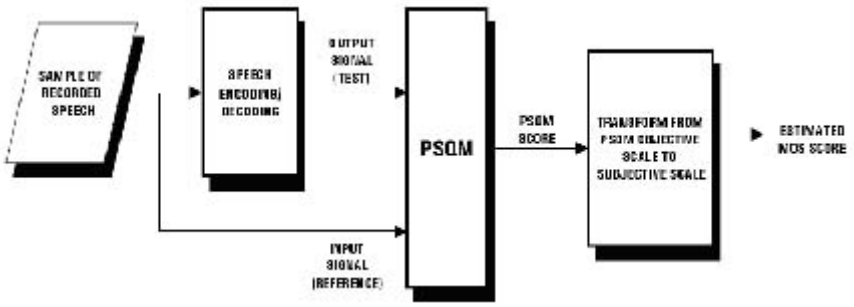
PSQM được thiết kế để sử dụng cho tín hiệu thoại (300-3400 Hz) qua các bộ mã hóa thoại. Phương thức này được sử dụng để đo tổn hao của các bộ mã hóa thoại này dựa trên các thông số nhận thức của con người. Phương thức này sử dụng hiệu quả đối với các bộ mã hóa thoại tốc độ thấp. Việc xử lý trong phương thức PSQM được thể hiện trên *Error! Reference source not found.*.

Để thực hiện phép đo PSQM, một mẫu tiếng nói được đưa vào hệ thống và được xử lý bởi một bộ mã hóa thoại bất kỳ. Những tính chất của tín hiệu vào giống như của các tín hiệu sử dụng cho phép đánh giá MOS được định nghĩa trong chuẩn ITU P.830.

Các tín hiệu vào có thể là mẫu tiếng nói thật hoặc tiếng nói nhân tạo theo khuyến nghị ITU P.50. ITU-T khuyến nghị tín hiệu vào được lọc theo modified IRS (Intermediate Reference System trong khuyến nghị ITU P.48) có những tính chất được định nghĩa trong phụ lục của D/P.830. Nó mô phỏng đặc tính tần số của máy điện thoại đầu cuối.

Khi nhận được, tín hiệu ra được ghi lại. Sau đó, nó được đồng bộ về mặt thời gian với tín hiệu vào. Hai tín hiệu này được thực hiện so sánh bởi các thuật toán PSQM. So sánh được thực hiện theo từng phân đoạn thời gian (khung thời gian) trong miền tần số (được biết đến là các phần tử thời gian - tần số) hoạt động dựa trên các tham số lấy từ mật độ phổ công suất của tín hiệu vào và ra của các phần tử thời gian - tần số. Việc so sánh dựa trên các tham số nhận thức của con người như: tần số và độ nhạy âm lượng (không chỉ phụ thuộc vào Mật độ phổ công suất - Spectral Power Densities (SPD)).

Điểm PSQM nằm trong dải từ 0 đến vô cùng. Điểm số này thể hiện độ lệch về mặt cảm nhận giữa tín hiệu ra và tín hiệu vào. VD: điểm 0 thể hiện tín hiệu ra và tín hiệu vào hoàn toàn trùng khớp, đánh giá là mức chất lượng hoàn hảo. Điểm PSQM càng cao thì thể hiện mức tổn hao càng lớn và đánh giá là mức chất lượng thấp. Trên thực tế, giới hạn trên đối với thang điểm PSQM trong khoảng từ 15-20.



Hình 1.42 Phương thức đánh giá chất lượng thoại PSQM

1.4.3.2.2 Phương pháp PESQ

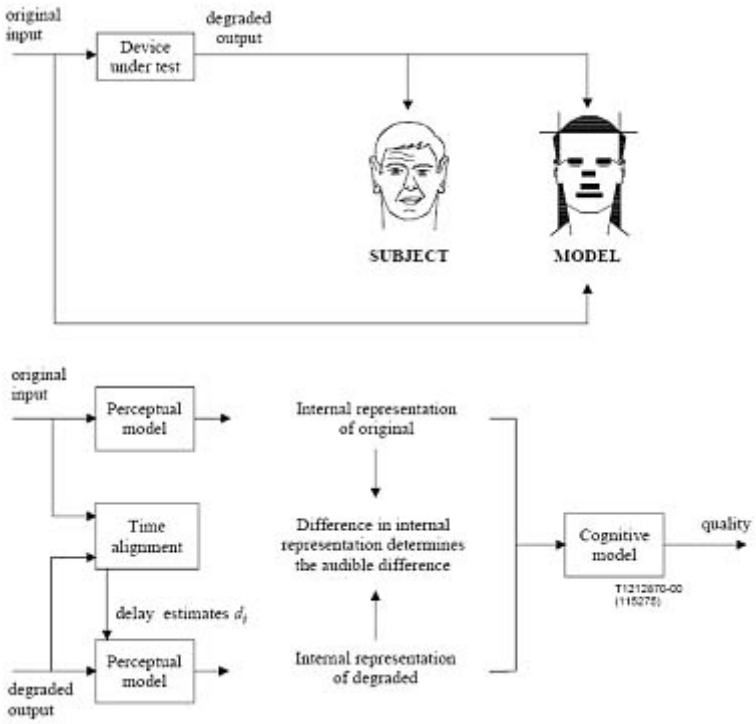
PESQ là phương pháp đánh giá chất lượng thoại so sánh; phương pháp này được mô tả trong khuyến nghị ITU-T P.862 [] được sử dụng thay thế cho khuyến nghị ITU-T P.861.

PESQ so sánh tín hiệu gốc X(t) với tín hiệu suy giảm Y(t) là kết quả của việc truyền tín hiệu X(t) qua hệ thống thông tin. Đầu ra của PESQ là một ước lượng về chất lượng thoại nhận được của tín hiệu Y(t).

Trong bước đầu tiên của PESQ, một loạt các trễ tín hiệu giữa tín hiệu vào ban đầu và tín hiệu ra được xác định; mỗi giá trị trễ được tính cho một khoảng thời gian mà có sự khác biệt về độ trễ so với phân đoạn thời gian trước đó. Ứng với mỗi phân đoạn thời gian, điểm bắt đầu và kết thúc được xác định. Một thuật toán sắp xếp dựa trên nguyên tắc so sánh giữa khả năng có hai trễ trong một đoạn thời gian với khả năng có một trễ trong đoạn thời gian đó. Thuật toán này có thể xử lý thay đổi về trễ trong cả khoảng lặng và trong thời gian tích cực thoại. Dựa trên tập các trễ đã xác định được, PESQ so sánh tín hiệu vào ban đầu với tín hiệu ra đã được sắp xếp bằng cách sử dụng một mô hình giác quan. Điểm mấu chốt của quá trình này là chuyển đổi cả tín hiệu gốc và tín hiệu đã bị suy giảm thành dạng biểu diễn của tín hiệu âm thanh trong hệ thống thính giác của con người có tính đến tần số thính giác và cường độ âm. Quá trình này được thực hiện theo nhiều giai đoạn: sắp xếp về mặt thời gian, sắp xếp mức tín hiệu về mức tín hiệu nghe đã được căn chỉnh, ánh xạ thời gian - tần số, frequency warping và căn chỉnh cường độ âm.

Trong PESQ, hai tham số lỗi được tính toán trong mô hình kinh nghiệm; chúng được kết hợp lại để ước lượng điểm MOS.

Một mô hình máy tính của chủ thể bao gồm mô hình giác quan và mô hình kinh nghiệm được sử dụng để so sánh tín hiệu đầu ra với tín hiệu gốc sử dụng các thông tin sắp xếp lấy được từ các tín hiệu định thời trong môđun sắp xếp định thời.



Hình 1.43 Mô tả phương pháp đánh giá chất lượng thoại PESQ

Phương pháp PESQ là có thể sử dụng không chỉ để đánh giá các bộ mã hóa thoại mà còn để đánh giá chất lượng thoại đầu cuối đến đầu cuối. Các hệ thống thông tin trên thực tế có thể bị ảnh hưởng của nhiều khâu lọc, trễ khả biến và các tổn hao do lỗi kênh truyền dẫn và việc sử dụng các bộ codec tốc độ thấp. Phương pháp PSQM được mô tả trong khuyến nghị ITU-T P.861 chỉ được khuyến nghị sử dụng để đánh giá các bộ codec thoại mà không tính đến các yếu tố như lọc, trễ khả biến ... PESQ tính đến các yếu tố này nhờ sử dụng cân bằng hàm truyền dẫn, sắp xếp định thời, và một thuật toán mới để thực hiện xác định tổn hao trung bình. PESQ đã được kiểm tra trong điều kiện kết hợp nhiều yếu tố như: lọc, trễ khả biến, tổn hao mã hóa và lỗi kênh truyền dẫn. Phương pháp này được khuyến nghị sử dụng thay thế cho PSQM để đánh giá chất lượng thoại từ đầu cuối đến đầu cuối.

1.4.3.2.3 Mô hình đánh giá truyền dẫn E-Model

E-model (tham khảo ETR 250 [4], EG 201 050 [2] và khuyến nghị ITU-T G.107 [7]) được sử dụng làm một công cụ để quy hoạch truyền dẫn trong các mạng điện thoại. Nó hỗ trợ việc ước lượng chất lượng tín hiệu thoại từ một kết hợp của nhiều yếu tố can nhiễu. E-model khác so với các phương pháp đánh giá chất lượng đã phân tích ở trên:

- Đây không phải là một công cụ đo mà là một công cụ quy hoạch mặc dù nó có thể sử dụng kết hợp với các phép đo.
- Nó ước lượng chất lượng thoại hai chiều và tính đến các yếu tố như: tiếng vọng, trễ ...

Đầu vào của E-model bao gồm các tham số được sử dụng tại thời điểm quy hoạch. Lưu ý rằng việc quy hoạch có thể được thực hiện trước và sau khi triển khai mạng. E-model có tính đến các tham số như: nhiễu, trễ, tiếng vọng và tính chất của thiết bị đầu cuối mà đã được chuẩn hóa hoặc đã được xác định, có thể đo được. Ngoài ra, E-model xác định trọng số đối với ảnh hưởng của thiết bị số hiện đại (các bộ codec tốc độ thấp, các bộ ghép kênh ...) đến chất lượng truyền dẫn.

Trong nhiều trường hợp, số lượng và chủng loại các thiết bị này được xác định tại thời điểm quy hoạch.

E-model dựa trên giả thiết là các tổn hao truyền dẫn có thể được chuyển đổi thành "psychological factors" và các hệ số này có tính cộng dồn trên một "psychological scale". Nói cách khác, nhận thức chủ quan về chất lượng thoại được coi như là tổng hợp của các tổn hao truyền dẫn.

E-model đầu tiên thực hiện tính toán một "giá trị gốc" về chất lượng (giá trị này được xác định từ nhiễu trên mạng). Mỗi tổn hao thêm vào được biểu diễn dưới dạng một giá trị tổn hao. Kết quả của phép trừ giá trị gốc với các giá trị tổn hao thể hiện ước lượng chất lượng thoại cho một mạng cụ thể. Cuối cùng, kết quả chất lượng thoại thu được được sử dụng để ước tính tỷ lệ thuê bao đánh giá chất lượng là tốt hay tồi. Cụ thể, E-model tính một hệ số đánh giá truyền dẫn R như sau:

$$R = Ro - Is - Id - Ie + A \quad (1.67)$$

Hệ số này bao gồm: giá trị gốc *Ro*, các tổn hao *Is*, *Id* và *Ie* và một hệ số thuận lợi (Advantage factor) như sau:

- *Ro* mô tả tỷ số tín hiệu trên nhiễu (SNR) của kết nối. Nó bao gồm tạp âm trong mạng, trong môi trường phía người nói và người nghe và ảnh hưởng của tạp âm tại phía người nghe, SNR được coi là một tham số biểu diễn chất lượng cơ bản.
- *Is* thể hiện các tổn hao nhất thời bao gồm: mức cường độ âm, mức xuyên âm vượt quá phạm vi cho phép và tổn hao lượng tử (mã hóa PCM).
- *Id* chứa các tổn hao do trễ và tiếng vọng.
- *Ie* bao gồm các tổn hao gây ra bởi các kỹ thuật nén thoại (codec tốc độ thấp).
- *A* cho phép điều chỉnh chất lượng trong những trường hợp đặc biệt nhờ thêm vào các yếu tố phi kỹ thuật để đánh giá chất lượng.

Cuối cùng, E-model sử dụng một ánh xạ phi tuyến tính để chuyển giá trị *R* thành giá trị MOS tương đương.

Như vậy, E-model cho phép xác định chất lượng thoại nhờ phân tích tác động của nhiều tham số truyền dẫn. Nhờ đó có thể đánh giá ảnh hưởng của các tham số này đối với mức chất lượng tổng thể.

1.4.3.2.4 Kết luận

Phần 2.4.3.2 đã phân tích các phương pháp đánh giá chất lượng có thể sử dụng để đánh giá chất lượng thoại trong mạng GSM. Như đã phân tích ở trên, điểm MOS là chỉ tiêu chất lượng tổng thể được sử dụng để đánh giá chất lượng thoại. Phương pháp đánh giá chủ quan sử dụng số liệu vào là nhận xét của khách hàng về mức chất lượng từ đó tính toán ra điểm đánh giá bình quân MOS. Các phương pháp đánh giá khách quan sử dụng các mô hình tính toán để ước lượng ra mức chất lượng quy đổi về MOS.

Dựa trên những ưu nhược điểm và phạm vi ứng dụng của các phương pháp này, để sử dụng đánh giá chất lượng thoại cho mạng GSM của VNPT có thể sử dụng các phương pháp đánh giá như sau:

- Sử dụng PESQ để đánh giá chất lượng thoại một chiều từ đầu cuối đến đầu cuối.
- Mô hình đánh giá E-Model có thể được sử dụng để phân tích hệ thống nhằm xác định các yếu tố ảnh hưởng đến chất lượng thoại.

- Ngoài ra, nếu có điều kiện có thể sử dụng kết hợp phương pháp đánh giá chủ quan để kiểm chứng lại việc đánh giá theo PESQ.

1.5 MÔ HÌNH ỨNG DUNG XỬ LÝ THOẠI

1.5.1 Mô hình thời gian động [14]

1.5.1.1 Tổng quan

Nhận dạng tiếng nói tự động (Automatic speech recognition-ASR) là một lãnh vực nghiên cứu quan trọng và có nhiều ứng dụng trên thực tế, dựa trên việc lưu trữ một hay nhiều mẫu âm thanh (*template*) ứng với từng từ trong bảng từ vựng nhận dạng. Quá trình nhận dạng thực hiện việc so trùng tiếng nói nhận được với các mẫu lưu trữ. Các mẫu có khoảng cách đo lường thấp nhất so với mẫu tiếng nói nhận được chính là từ được nhận dạng. Giải thuật dùng để tìm được sự tương thích tốt nhất là dựa trên lập trình động (Dynamic Programming - DP), và một trong các giải thuật là giải thuật mô hình thời gian động (Dynamic Time Warping-DTW).

Để có thể nắm bắt được kiến thức về DTW một cách nhanh chóng, có hai khái niệm cần làm rõ

→ **Điểm đặc trưng**: là thông tin của từng tín hiệu được biểu diễn dưới dạng nào đó.

→ **Sai biệt**: dạng đo lường nào đó được dùng để tính toán được sự tương thích, có hai dạng:

1. Cục bộ: độ tính toán sai biệt giữa điểm đặc trưng của một tín hiệu một tín hiệu khác.
2. Toàn cục: độ tính toán sai biệt tổng giữa một tín hiệu tổng với một tín hiệu khác có thể có sai biệt.

Việc phân tích điểm đặc trưng bao gồm việc tính toán vector đặc trưng với khoảng thời gian thông thường. Đối với việc phân tích dự đoán tuyến tính, vector đặc trưng bao gồm việc tính toán các hệ số dự đoán (hoặc các phép biến đổi giữa chúng). Một loại vector đặc trưng thôngdụng dùng trong nhận dạng tiếng nói là Mel Frequency Cepstral Coefficients (MFCCs).

Vì các vector đặc trưng có thể có nhiều phần tử phức tạp, nên giá trị trung bình của việc tính toán cần được thiết lập. Phép đo sai biệt giữa 2 vector đặc trưng được tính toán bằng đơn vị theo hệ Euclidean. Như vậy độ sai biệt cục bộ giữa vector đặc trưng x của tín hiệu 1 và vector đặc trưng y của tín hiệu 2 được cho bởi

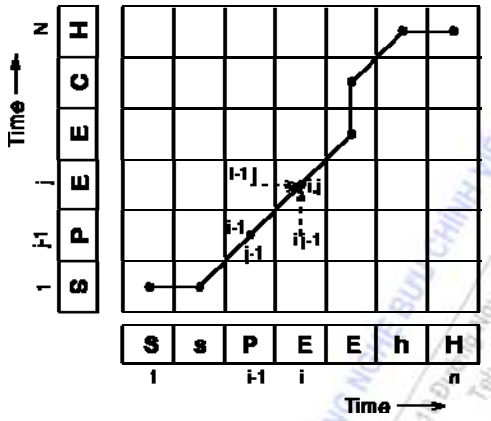
$$d(x,y)=\sqrt{\sum_i (x_i-y_i)^2} \tag{1.68}$$

Mặc dù sử dụng hệ đơn vị Euclidean cho việc tính toán sẽ tăng độ phức tạp hơn so với các hệ đo lường khác, nhưng nó lại cho tác dụng nhiều hơn với độ sai biệt lớn đối với một đặc trưng đơn. Nếu như việc quay lui trong quá trình tìm kiếm sự tương thích nhất cần thiết thì một dãy có nhiệm vụ lưu trữ các entry trước đó trong quá trình xử lý tìm kiếm, được gọi là *backtrace array*.

1.5.1.2 Giải thuật DTW đối xứng

Tiếng nói là một quá trình phụ thuộc vào thời gian. Tiếng nói có nhiều âm tiết tương tự nhau nhưng với khoảng thời gian phát âm khác nhau, hoặc có âm tiết đồng âm nhưng khác ở chỗ nhấn âm, v.v… Để phân tích được sự sai biệt toàn cụ giữa hai mẫu tiếng nói (đại diện bởi một chuỗi các vector), vấn đề về thời gian phải được xem xét.

Bài toán được mô tả ở hình 0, mô tả một ma trận hai chiều theo thời gian sử dụng cho việc canh chỉnh theo thời gian. Cột là mô tả cho tiếng nói mẫu (*template*) và dòng là tiếng nói thu được cần nhận dạng. Trong hình minh họa, tín hiệu vào "SsPEEhH" được xem là một dạng "nhiễu" của tiếng nói mẫu. Tín hiệu vào này sẽ được so sánh trùng với tất cả các mẫu tiếng nói được lưu trữ trong hệ thống. Mẫu có độ tương thích tốt nhất sẽ có độ sai biệt nhỏ nhất so với tín hiệu vào cần so sánh. Giá trị độ sai biệt toàn cục là tổng các sai biệt cục bộ của việc so sánh.



Hình 1.44 Mô tả canh chỉnh thời gian giữa mẫu tiếng nói "SPEECH" và tín hiệu tiếng nói đầu vào "SsPEEhH"

Làm cách nào để có thể tính được độ tương thích tốt nhất (có giá trị độ sai biệt toàn cục nhỏ nhất) giữa tín hiệu cần so sánh và tiếng nói mẫu? Việc này được thực hiện bằng cách ước lượng tất cả khoảng cách có thể có, nhưng cách này không hiệu quả khi số lượng khoảng cách có dạng hàm mũ theo chiều dài của tín hiệu ngõ vào. Thay vào đó, ta xem xét những ràng buộc tồn tại trong quá trình so trùng (hoặc có thể áp đặt các ràng buộc này) và dùng những ràng buộc này để có được giải thuật hiệu quả hơn. Các ràng buộc được thiết lập phải không phức tạp và cũng không hạn chế nhiều, như:

- → Các khoảng cách so trùng không thể thực hiện việc đi lui;
- → Mọi khung của tín hiệu cần so trùng phải được dùng trong quá trình so trùng;
- → Các giá trị sai biệt cục bộ được kết hợp bằng phương pháp cộng dồn vào giá trị sai biệt toàn cục.

Mọi khung trong tín hiệu cần so trùng với mẫu tiếng nói được xem xét ứng với từng tính toán độ sai biệt. Nếu thời điểm đang xét là (i, j) , với *i* là chỉ số của khung tín hiệu ngõ vào, *j* là của khung tiếng nói mẫu, thì các vị trí trước đó là $(i-1, j-1)$, $(i-1, j)$, $(i, j-1)$. Ý tưởng chính của lập trình động là tại vị trí (i, j), việc tính toán dự trên độ sai biệt nhỏ nhất của các vị trí $(i-1, j-1)$, $(i-1, j)$, $(i, j-1)$.

Giải thuật lập trình động thực hiện cần phải đồng bộ thời gian: mỗi cột của ma trận thời gian-thời gian được xem như là một sự kế vị các tính toán trước đó, do đó, ứng với một mẫu tiếng nói có chiều dài N, số lượng bước so trùng tối đa là N.

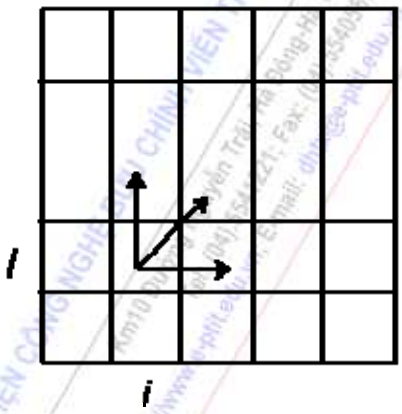
Gọi $D(i, j)$, $d(i, j)$ tương ứng là độ sai biệt toàn cục và độ sai biệt cục bộ tại vị trí (i, j).

$$D(i, j) = \min\left[D(i-1, j-1), D(i-1, j), D(i, j-1)\right] + d(i, j) \qquad (1)$$

Với $D(1,1) = d(1,1)$ là giá trị khởi tạo ban đầu, giải thuật ứng dụng đệ qui vào việc tính toán các độ sai biệt tại $D(i, j)$. Giá trị cuối $D(n, N)$ chính là giá trị chênh lệch giữa template và tín hiệu cần so sánh, lưu ý rằng N sẽ khác nhau ứng với mỗi template.

Đối với việc nhận dạng tiếng nói, giải thuật DP không cần phải chạy trên các máy tính có bộ nhớ lớn, việc lưu trữ được thực hiện bởi một array, lưu giữa từng cột đơn trong ma trận thời gian-thời gian. Ma trận có vị trí đầu tiên có giá trị là 0, như vậy chỉ những hướng di chuyển trong ma trận được mô tả ở hình 1 mới có thể xuất phát từ vị trí (i, j).

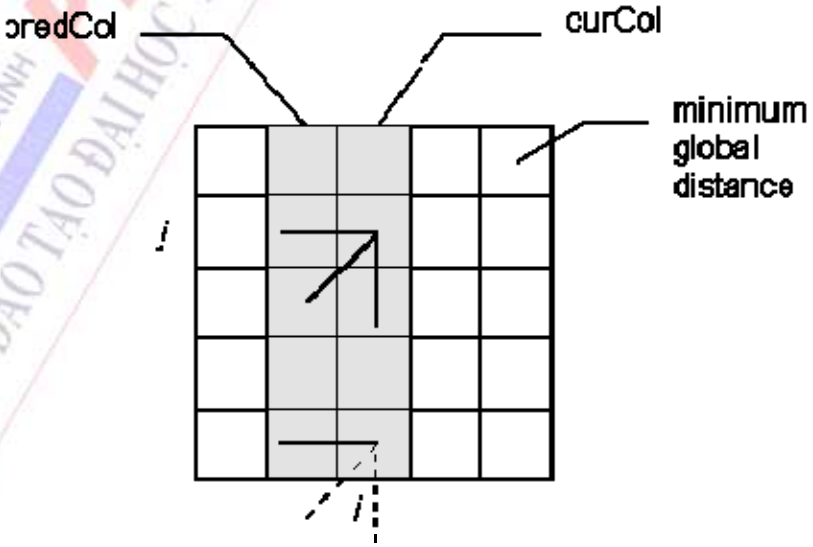
Figure 1. The three possible directions in which the best match path may move from cell (i,j) in symmetric DTW.



Hình 1.45 Ba hướng có độ tương thích tốt nhất có thể đi đến từ ô (i, j) trong giải thuật DTW đối xứng.

Phương trình (1) được thực hiện bằng phương pháp đệ quy. Tuy nhiên, trừ khi ngôn ngữ đã được tối ưu cho việc đệ quy, phương pháp này có thể tốn nhiều thời gian ngay cả đối với những tín hiệu cần so sánh có kích thước nhỏ. Một phương pháp khác có thể cải tiến được tốc độ xử lý cũng như cần bộ nhớ thực thi nhỏ hơn là dùng hai vòng lặp, sử dụng hai array để lưu trữ các cột kề với ma trận thời gian-thời gian.

Figure 2. The cells at (i,j) and (i,0) have different possible originator cells. The path to (i,0) can only originate from (i-1,0). However, the path to (i,j) can originate from the three standard locations as shown.



Hình 1.46 Vị trí ô (i, j) và $(i,0)$ có các ô định hướng khác nhau. Từ $(i,0)$ chỉ có thể đi từ ô $(i-1,0)$. Tuy nhiên tại ô (i, j) thì có thể đi đến 3 ô như mặc định.

Giải thuật tìm chi phí toàn cục nhỏ nhất:

1. Tính tại cột 0, bắt đầu từ đáy của ô. Giá trị chi phí toàn cụ của ô bằng giá trị chi phí cục bộ. Sau đó, tính toán giá trị chi phí toàn cục của các ô có khả năng cho được giá trị thấp bằng cách lấy giá trị chi phí cục bộ của ô cộng thêm cho giá trị toàn cục của ô ngay dưới đó, ô này được gọi là predCol (*predecessor column*).
2. Tính giá trị chi phí toàn cục của ô đầu tiên của cột kế tiếp là curCol. Giá trị cục bộ của ô cộng thêm cho giá trị toàn cục của ô phía dưới cùng của cột trước nó.
3. Tính giá trị toàn cục của các ô còn lại của curCol.
4. curCol được định là predCol và lặp lại bước 2 cho đến khi tất cả các cột được tính toán xong.
5. Giá trị chi phí toàn cục tại vị trí cột cuối cùng, dòng trên cục là giá trị cần tìm.

Mã giải của quá trình như sau:

```
calculate first column (predCol);
for i=1 to number of input feature vectors
    curCol[0] = local cost at (i,0) + global cost at (-1,0)
    for j=1 to number of template feature vectors
        curCol[j]=local cost at (i,j) + minimum of global costs at (i-1,j), (i-1,j-1) or (i,j-1).
    end
    predCol=curCol
end
minimum global cost is value in curCol[number of template feature vectors]
```

1.5.1.3 Giải thuật DTW bất đối xứng

Mặc dù giải thuật cơ bản DP có ưu điểm là đối xứng (tất cả các khung của tín hiệu cần so trùng và mẫu tiếng nói cần tham khảo được xem xét), tuy nhiên giải thuật vẫn còn yếu điểm là tại các vị trí cột và hàng lệ thuộc vào các vị trí đường chéo có thể sinh lỗi.

Một cách để tránh việc này là thực hiện việc dùng $d(i,j)$ hai lần trong mỗi bước tại vị trí đường chép, điều này dẫn đến loại bỏ lỗi tại các vị trí cột và hàng, gọi giá trị lỗi cô lập d_h, d_v tương ứng cho các bước di chuyển theo hàng và cột. Phương trình (1) sẽ trở thành

$$D(i,j) = \min\left[D(i-1,j-1) + 2d(i,j), D(i-1,j) + d(i+j) + d_h, D(i,j-1) + d(i,j) + d_v\right] \quad (2)$$

Với giá trị d_h, d_v có được qua thực nghiệm

Các ràng buộc trong việc nhảy đến các ô kế tiếp:

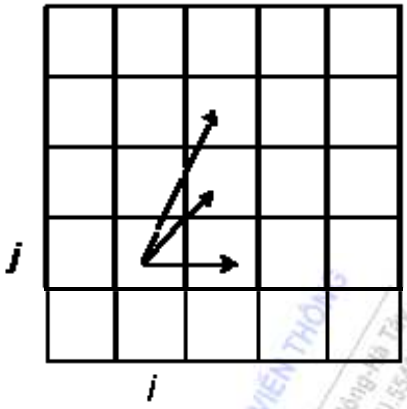
→ $(i-1, j-2)$ đến vị trí (i,j) - gọi là đường chéo mở rộng (độ đốc là 2)

→ $(i-1, j-1)$ đến vị trí (i,j) - gọi là đừong chéo chuẩn (độ dốc là 1)

→ $(i-1, j)$ đến vị trí (i,j) - gọi là đường ngang (độ dốc là 0)

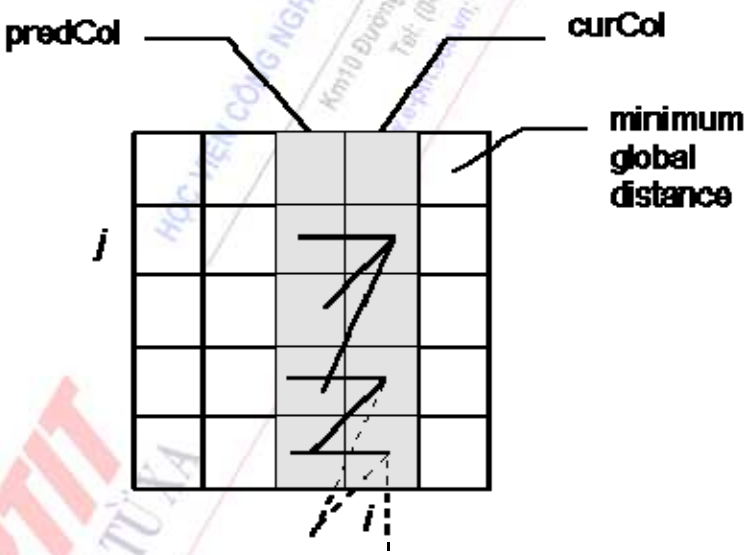
Giả sử rằng mỗi khung của tín hiệu ngõ vào chỉ được xét đến một lần duy nhất, như vậy ta có thể không xét đến việc chuẩn hóa độ dài của mẫu tiếng nói. Do việc tính toán tại từng ô khác nhau nên giải thuật được gọi là giải thuật lập trình động bất đối xứng.

Figure 3. The three possible directions in which the best match path may move from cell (i,j) in asymmetric DTW.



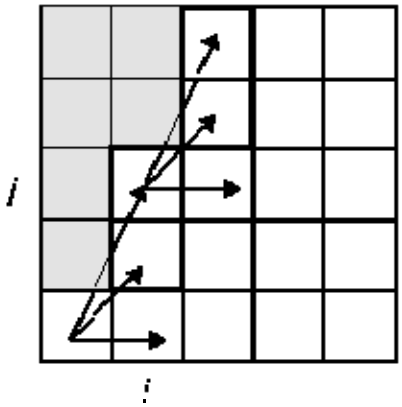
Hình 1.47 Ba hướng có giá trị so trùng tốt nhất có thể đi đến từ ô (i, j) trong giải thuật DTW bất đối xứng.

Figure 4. The cells at (i,j), (i,0) and (i,1) have different possible originator cells. The path to (i,0) can only originate from (i-1,0). The path to (i,1) can originate at either (i-1,1) or (i-1,0). However, the path to (i,j) can originate from the three standard locations as shown.



Hình 1.48 Các vị trí ô tại (i, j), $(i,0)$ và $(i,1)$ có các ô định hướng khác nhau. Tại ô $(i,0)$ chỉ có thể đi đến ô $(i-1,0)$; tại ô $(i,1)$ có thể đi đến $(i-1,1)$ hoặc $(i-1,0)$; tại ô (i, j) có thể đi đến các vị trí chuẩn.

Figure 5 The shaded area shows the region into which the path can never move. The rectangular boxes show the 'special cases' for the path. As in symmetrical DTW, row 0 is a special case as normal. Additionally, in asymmetric DTW, row 1 is also treated specially.



Hình 1.49 Vùng tô màu mô tả vùng mà các hướng di chuyển không bao giờ đi đến. Các hình chữ nhật mô tả là các trường hợp đặc biệt. Trong giải thuật DTW đối xứng, dòng 0 là trường

hợp đặc biệt, còn trong giải thuật DTW bất đối xứng, dòng 1 được xử lý khác với các trường hợp còn lại.

Các trường hợp đặc biệt thường xảy ra tại vị trí $j = 2i - 1$, và $j = 2i$, giá trị chi phí toàn cục cho từng trường hợp được tính như sau:

→ $2i - 1$: chi phí cục bộ + giá trị nhỏ nhất tại vị trí predCol$[j-1]$ và predCol$[j-2]$

→ $2i$: chi phí cục bộ + giá trị nhỏ nhất tại vị trí predCol$[j-2]$

Mã giả của quá trình như sau:

```
predCol[0]=local cost at (0,0)
for i=1 to number of input feature vectors
    curCol[0] = local cost at (i,1) + global cost at (i-1 0)
    for j=1 to (minimum of number of template feature vectors and 2i+1)
        store j in highestJ
        If the cell (i,j) is a special case then
            If it is row 1 then
                curCol[j]=local cost at (i,j) + global cost at (i-1,j-1).
            else
                If the cell (i,j) is the lower of the special case pair
                    curCol[j]=local cost at (i,j) + minimum of global costs at (i-1,j-1) or (i-1,j-2).
                else
                    curCol[j]=local cost at (i,j) + global cost at (i-1,j-2).
                end
            end
        else
            If it is row 1 then
                curCol[j]=local cost at (i,j) + minimum of global costs at (i-1,j) or (i-1,j-1).
            else
                curCol[j]=local cost at (i,j) + minimum of global costs at (i-1,j). (i-1,j-1) or (i-1,j-2).
            end
        end
    end
    predCol=curCol
end
minimum global cost is value in curCol[highestJ]
```

Giá trị chi phí nhỏ nhất là cột cuối cùng lưu trong *highestJ*

1.5.2 Mô hình chuỗi markov ẩn [15]

1.5.2.1 Tổng quan

Mô hình Markov (Hidden Markov Model - HMM) ẩn được sử dụng trong việc thống kê mô hình tạo âm thoại. Tính hiệu quả của mô hình được thể hiện trong việc có thể mô tả đặc điểm của tín hiệu âm thoại theo dạng toán học dễ dàng cho việc xử lý tín hiệu.

Các trạng thái của HMM có được trước khi thực hiện việc xử lý các trạng thái (trích các thông số). Như thế, ngõ vào của HMM chính là chuỗi các thông số vector rời rạc theo thời gian.

1.5.2.2 Định nghĩa mô hình Markov ẩn

Mô hình Markov ẩn là một tập các trạng thái hữu hạn, mà mỗi trạng thái có liên quan đến hàm phân phối xác suất. Việc chuyển tiếp giữa các trạng thái được định nghĩa bởi một tập xác suất được gọi là xác suất chuyển tiếp *(transition probability)*. Trong một trạng thái cụ thể, kết quả có thể được tạo ra dựa trên hàm phân phối xác suất tương ứng. Kết quả này không phải là một

trạng thái có thể nhìn thấy được thông qua việc quan sát các trạng thái, cho nên được gọi là mô hình Markov ẩn.

Trong mô hình Markov ẩn, các ký hiệu sau đây được sử dụng

- Số lượng trạng thái của mô hình, N.
- Số lượng ký hiệu quan sát theo thứ tự, M. Nếu việc quan sát là liên tục thì có giá trị M là vô hạn
- Tập các trạng thái xác suất chuyển tiếp $\Lambda = \{a_{ij}\}$

$a_{ij} = p\{q_{t+1} = j | q_t = i\}, 1 \le i, j \le N$ với q_t là trạng thái hiện tại.

- Xác suất trạng thái chuyển tiếp phải thỏa mãn ràng buộc trực giao sau

$$a_{ij} \ge 0, \quad 1 \le i, j \le N \text{ và } \sum_{j=1}^{N} a_{ij} = 1, \quad 1 \le i \le N$$

- Hàm phân phối xác suất của mỗi trạng thái $B = \{b_j(k)\}$

$b_j(k) = p\{a_t = v_k | q_t = j\}, \quad 1 \le j \le N, \quad 1 \le k \le M$ với v_k định nghĩa cho ký hiệu quan sát thứ *k* theo thứ tự alphabet, và a_t là vector thông số hiện tại. Điều kiện

$$b_j(k) \ge 0, \quad 1 \le j \le N, \quad 1 \le k \le M \text{ và } \sum_{k=1}^{M} b_j(k) = 1, \quad 1 \le j \le N$$

- Nếu việc quan sát là liên tục thì phải dùng hàm mật độ xác suất liên tục thay cho xác suất rời rạc. Trong trường hợp này, các thông số của hàm mật độ xác suất liên tục phải được định rã. Thông thường mật độ xác suất xấp xỉ với trọng số tổng M của phân bố Gaussian N
- $b_j(a_t) = \sum_{m=1}^{M} c_{jm} \mathrm{N}(\mu_{jm}, \Sigma_{jm}, a_t)$ với
- c_{jm} = hệ số trọng số
- μ_{jm} = vector trung bình
- Σ_{jm} = ma trận đồng biến

c_{jm} thỏa các điều kiện $c_{jm} \ge 0, \quad 1 \le j \le N, \quad 1 \le m \le M$ và $\sum_{m=1}^{M} c_{jm} = 1, \quad 1 \le j \le N$

- Trạng thái ban đầu của hàm phân phối $\pi = \{\pi_i\}$ với $\pi_i = p\{q_i = i\}, \quad 1 \le i \le N$

Ký hiệu $\lambda = (\Lambda, B, \pi)$ dùng cho HMM với hàm phân phối xác suất rời rạc, và

$\lambda = (\Lambda, c_{jm}, \mu_{jm}, \Sigma_{jm}, \pi)$ dùng cho HMM với hàm mật độ xác suất liên tục

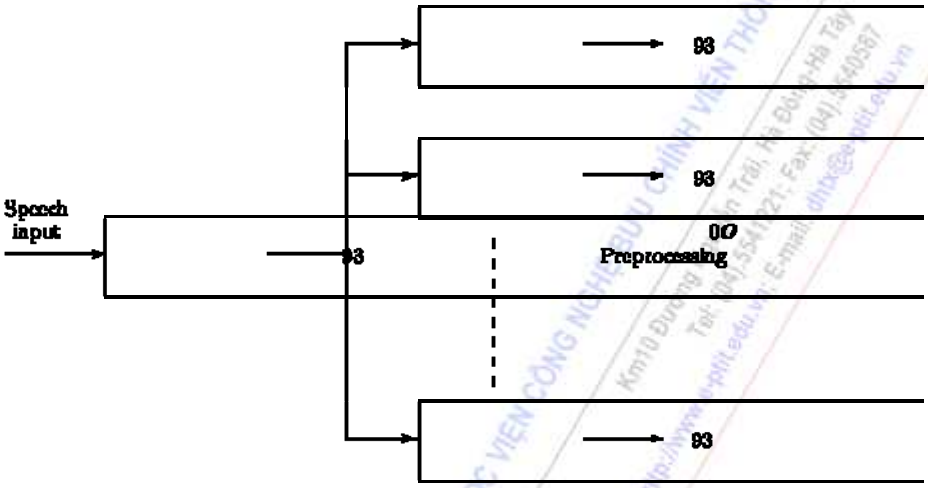
Một số ứng dụng mô hình Markov ẩn trong thực tiễn là:

- Ứng dụng HMM trong việc nhận dạng âm thoại cô lập
- Ứng dụng HMM trong việc nhận dạng âm thoại liên tục
- Ứng dụng HMM trong việc xử lý cấu trúc đa cấp độ cho văn phạm tiếng Anh
- Ứng dụng HMM trong học bản thảo viết tay

1.5.2.3 Ứng dụng HMM trong việc nhận dạng âm thoại cô lập

Nhận dạng cô lập với nghĩa tổng quát là nhận dạng âm thoại dựa trên bất kỳ loại đơn vị âm thoại cô lập, có thể là một từ hoặc là một phần của từ hoặc là một số từ liên tục nhau. Đặc biệt việc nhận dạng một phần của từ trong chế độ cô lập có thể có được kết quả tốt trong việc nhận dạng các từ liên tục nhau ứng với cùng kỹ thuật nhận dạng.

Trong vấn đề nhận dạng âm thoại đơn vị cô lập, giả sử từ vựng chứa N âm thoại đơn vị, ta dùng hệ thống mô tả ở hình 1.1



Hình 1.50 Bộ nhận dạng âm thoại đơn vị cô lập

Có rất nhiều giải pháp cho việc nhận dạng này, vì có rất nhiều các tiêu chuẩn lựa chọn tối ưu, trong đó tiêu chuẩn MMI là phương pháp dựa vào Gradient. Phương pháp này gồm hai phần nhỏ: huấn luyện và nhận dạng, sử dụng mỗi một HMM cho việc nhận dạng từng đơn vị âm thoại.

1.5.2.4 Ứng dụng HMM trong việc nhận dạng âm thoại liên tục

Trong chế độ cô lập, ta sử dụng một HMM cho từng đơn vị âm thoại, còn trong trường hợp nhận dạng liên tục, hệ thống cần nhận dạng một chuỗi các âm đơn vị kết nối lại với nhau, đôi khi cần nhận diện cả một câu, hoặc nhiều câu. Khi đó số lượng câu có thể rất lớn. Phương pháp thực hiện cũng tương tự như trong nhận dạng âm thoại cô lập, bao gồm hai bước huấn luyện và nhận dạng. Bước huấn luyện có thể dùng hoặc là tiêu chuẩn MMI hoặc là ML, và bước nhận dạng có thể sử dụng các phương pháp như nhận dạng trên cơ sở Veterbi, xây dựng cấp độ, tìm kiếm N-tốt nhất và tính toán hiệu suất bộ nhận dạng.

1.5.3 Mạng nơron

1.5.3.1 Tổng quan

Mạng neural nhân tạo (Artificial Neural Network - ANN) là một mô hình xử lý thông tin dựa trên cơ chế hoạt động của hệ thống thần kinh sinh học, như não bộ. Thành phần chính yếu của mô hình này là cấu trúc đặc biệt của hệ thống này. Nó tập hợp một số lượng lớn các phần tử xử lý kết hợp nội tại (được gọi là các neuron) hoạt động hợp nhất để giải quyết các bài toán cụ thể. Một ANN sẽ được cấu hình cho một ứng dụng cụ thể nào đó, ví dụ như nhận dạng mô hình hoặc phân loại dữ liệu thông qua quá trình học. Việc học trong hệ thống nhằm mục đích điều chỉnh các kết nối thuộc kỳ tiếp hợp được phân chia trong tế bào mà đã có sẵn giữa các neuron.

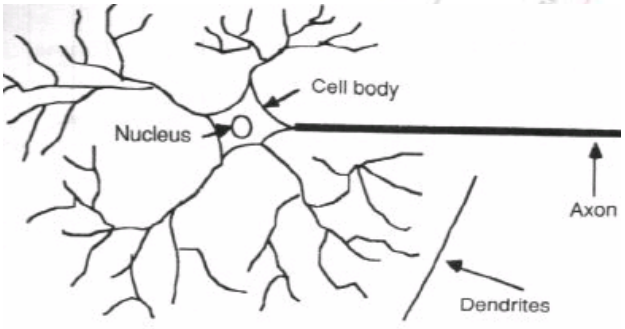
Neuron nhân tạo đầu tiên được tạo ra vào năm 1943 bởi nhà nghiên cứu neuron học Warren McCulloch và nhà logic học Walter Pits. Nhưng kỹ thuật thời đó không cho phép neuron phát triển được các thế mạnh của nó. Mạng neuron này nay có nhiều cải tiến cũng như đáp ứng

được các yêu cầu đặt ra của các bài toán, một số ưu điểm của mạng neuron ngày nay so với thời trước là:

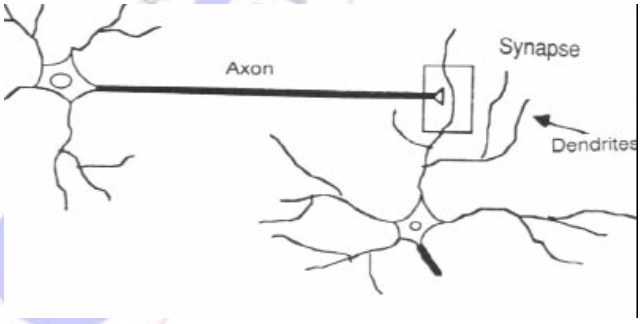
1. Học thích ứng: có khả năng học cách thức thực hiện công việc dựa trên các dữ liệu cho sẵn trong quá trình huấn luyện hoặc định các thông số ban đầu
2. Khả năng tự tổ chức: một ANN có thể tự thân tổ chức hoặc miêu tả các thông tin mà nó nhận được trong suốt thời gian học
3. Hiệu chỉnh lỗi thông qua mã hóa thông tin dư thừa: có thể hủy một phần mạng làm cho hiệu suất hệ thống giảm. Tuy nhiên, một số mạng có khả năng nhớ được phần mạng đã hủy.

1.5.3.2 Phương pháp học của não người

Trong não người, một neuron sẽ thực hiện nhiệm vụ thu thập các tín hiệu từ các neuron khác thông qua các cấu trúc thần kinh phức tạp được gọi là *dạng cây*. Neuron gửi cac hoạt động điện thông qua sợ mỏng, dài, gọi là sợi trục thần kinh *axon*, được phân chia thành hàng ngàn nhánh nhỏ. Tại cuối mỗi nhánh, một cấu trúc được gọi là khớp thần kinh *synapse* sẽ chuyển đổi các hoạt động từ *axon* thành các hiệu ứng điện thực hiện việc ức chế hoặc kích thích hoặc động này đối với các neuron kết nối tới nhánh. Khi một neuron nhận được tín hiệu kích thích đầu vào có mức độ so sánh tương đối lớn so với tín hiệu cấm ngõ vào, neuron sẽ gửi một gai điện đến axon của nó. Việc học xảy ra theo cách thức thay đổi hiệu lực của khớp thần kinh dẫn đến việc truyền thông tin từ một neuron đến một neuron khác về sự thay đổi.



Hình 1.51 Các thành phần của một neuron

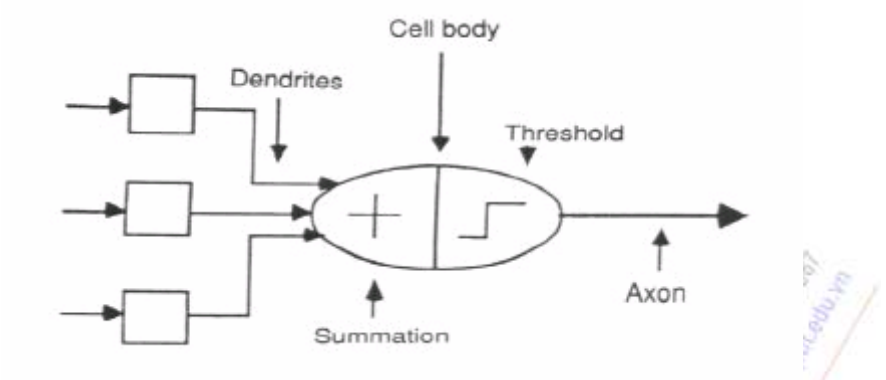


Hình 1.52 Khớp thần kinh

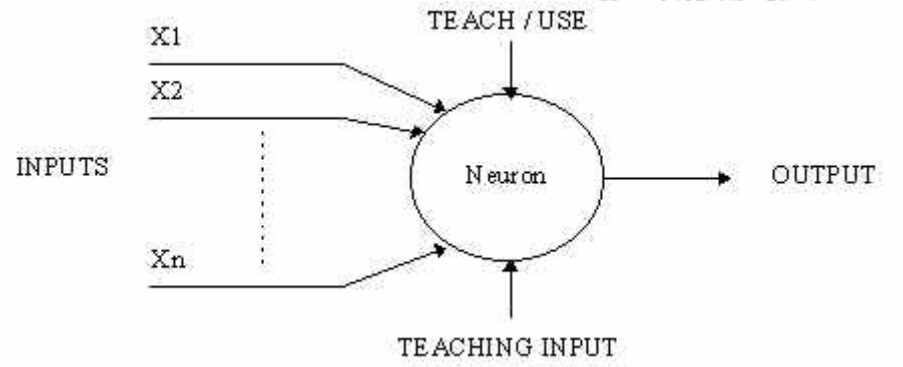
1.5.3.3 Từ neuron người đến neuron nhân tạo

Thực hiện quản lý các mạng neuron bằng cách đầu tiên tìm cách suy luận bản chất của các neuron và các kết nối nội tại bên trong. Sau đó thực hiện việc lập trình để giả lập các đặc tính này. Tuy nhiên, do nhận thức về các neuron không đầy đủ cũng như năng lực của việc tính toán là có

giới hạn, cho nên mô hình mạng neuron nhân tạo so với mạng neuron người thuộc dạng "lý tưởng" và đơn giản hơn.

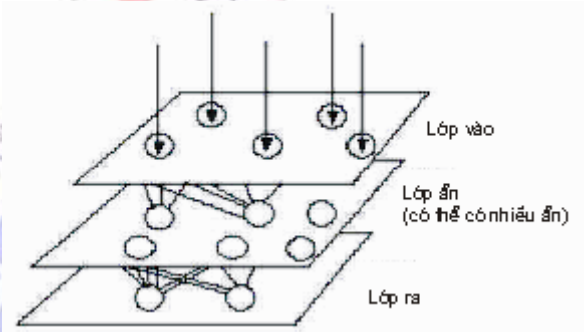


Hình 1.53 Mô hình neuron



Hình 1.54 Mô hình hóa mô hình neuron

Mô phỏng hoạt động của các neuron thần kinh, mạng neuron nhân tạo là hệ thống bao gồm nhiều phần tử xử lý đơn giản (neuron) hoạt động song song. Tính năng của hệ thống này tùy thuộc vào cấu trúc của hệ, các trọng số liên kết neuron và quá trình tính toán tại các neuron đơn lẻ. Mạng neuron có thể từ dữ liệu mẫu và tổng quát hoá dựa trên các dữ liệu mẫu học.



Hình 1.55 Mô hình mạng Neuron theo các lớp

Một nhóm các neuron được tổ chức theo một cách sao cho tất cả chúng đều nhận cùng một vector vào X để xử lý tại cùng một thời điểm. Việc sản sinh ra tín hiệu ra của mạng xuất hiện cùng một lúc. Vì mỗi neuron có một tập trọng số khác nhau nên có bao nhiêu neuron sẽ sản sinh ra bấy nhiêu tín hiệu ra khác nhau. Một nhóm các neuron như vậy được gọi là một lớp mạng. Chúng ta có thể kết hợp nhiều lớp mạng tạo ra một mạng có nhiều lớp, lớp nhận tín hiệu đầu vào (vector tín hiệu vào x) được gọi là lớp vào (input layer). Trên thực tế chúng thực hiện như một bộ

đệm chứa tín hiệu đầu vào. Các tín hiệu đầu ra của mạng được sản sinh ra từ lớp ra của mạng (output layer). Bất kỳ lớp nào nằm giữa 2 lớp mạng trên được gọi là lớp ẩn (hidden layer) và nó là thành phần nội tại của mạng và không có tiếp xúc nào với môi trường bên ngoài. Số lượng lớp ẩn có thể từ 0 đến vài lớp. Mô hình neuron nhân tạo đòi hỏi 3 thành phần cơ bản sau:

- Tập trọng số liên kết đặc trưng cho các khớp thần kinh.

- Bộ cộng (Sum) để thực hiện phép tính tổng các tích tín hiệu vào với trọng số liên kết tương ứng

- Hàm kích hoạt (squashing function) hay hàm chuyển (transfer function) thực hiện giới hạn đầu vào của neuron.

Trong mô hình neuron nhân tạo mỗi neuron được nối với các neuron khác và nhận được tín hiệu x_i từ chúng với các trọng số w_i. Tổng thông tin vào có trọng số là: $Net = \sum w_j x_j$.

1.5.3.4 Ứng dụng mạng neuron trong nhận dạng tiếng nói

Mạng neuron (Neuron Network) là một công cụ có khả năng giải quyết được nhiều bài toán khó, thực tế những nghiên cứu về mạng neuron đưa ra một cách tiếp cận khác với những cách tiếp cận truyền thống trong lý thuyết nhận dạng. Trong thực tế, mạng neuron được triển khai có hiệu quả trong nhận dạng tiếng nói thường dùng mạng neuron lan truyền ngược hướng (Back-propagation Neural Network) hoặc kết hợp với phương pháp mã dự đoán tuyến tính LPC (Linear Predictive Coding).

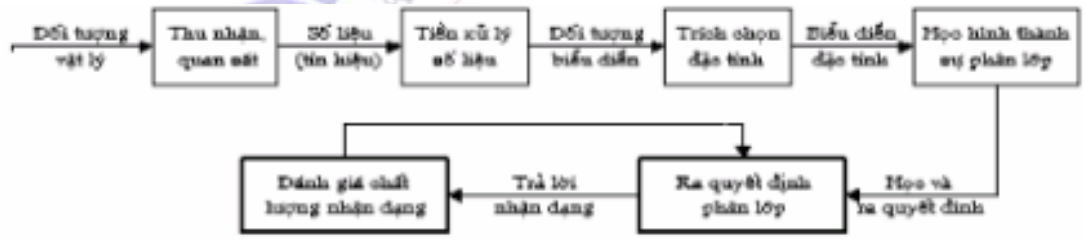
1.5.3.4.1 Sơ lược về lý thuyết nhận dạng

Lý thuyết nhận dạng là phương pháp để xây dựng một hệ thống tin học có khả năng: cảm nhận-nhận thức-nhận biết các đối tượng vật lý gần giống khả năng của con người. Nhận dạng có gắn chặt với 3 khả năng trên là một lĩnh vực hết sức rộng có liên quan đến việc xử lý tín hiệu trong không gian nhiều chiều, mô hình, đồ thị, ngôn ngữ, cơ sở dữ liệu, phương pháp ra quyết định... Hệ thống nhận dạng phải có khả năng thể hiện được quá trình nhận thức của con người qua các mức:

- Mức 1- mức cảm nhận: cảm nhận được sự tồn tại các đối tượng quan sát, hay đối tượng mà hệ thống cần nhận dạng. Mức này cũng đưa ra quá trình thu nhận số liệu qua các bộ cảm biến trong hệ thống nhận dạng, ví dụ trong hệ thống nhận dạng tiếng nói: đối tượng ở đây là tiếng nói (speech) và thu nhận đầu vào qua Micro hoặc các file âm thanh .wav.

- Mức 2- mức nhận thức: ở đây biểu diễn quá trình học, mô hình hoá đối tượng để tiến tới hình thành sự phân lớp (classification) các đối tượng cần nhận dạng.

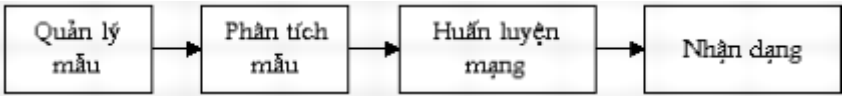
- Mức 3- mức nhận biết: từ đối tượng quan sát có thể trả lời nhận biết đối tượng là gì ? Hay đây là quá trình ra quyết định.



Hình 1.56 Sơ đồ tổng quan của hệ thống nhận dạng

1.5.3.4.2 Nhận dạng tiếng nói dùng mạng Neuron

Hình 1.57 mô tả chức năng cơ bản của hệ thống nhận dạng tiếng nói



Hình 1.57 Sơ đồ khối mô tả hệ thống nhận dạng tiếng nói

Phương án lựa chọn số nút của từng lớp trong mạng: theo kinh nghiệm của các chuyên gia về mạng neuron trong các bài toán phân lớp có sử dụng mạng lan truyền ngược hướng, sử dụng 1 lớp tính toán là lớp mạng Kohonen làm lớp ẩn. Việc xác định số neuron cho từng lớp.

+ Số neuron lớp vào = số chiều của vector vào

+ Số neuron lớp Kohonen = số giá trị các tập trả lời

+ Số neuron lớp ra = số lượng kết quả đầu ra, sử dụng phương pháp mã hoá bằng số bit biểu diễn số lượng kết quả

Phương pháp học cạnh tranh của lớp ẩn và quá trình học có chỉ đạo tại lớp ra của mạng theo các bước sau:

+ Khởi tạo trọng số: các thành phần ma trận trọng số được khởi tạo bởi giá trị ngẫu nhiên

+ Đọc tín hiệu vào cho mạng: dữ liệu trong file mẫu chứa thông tin mẫu học và cho kết quả gồm 2 thành phần: mảng 1 chiều chứa vector tín hiệu vào và mảng 2 chiều chứa ma trận trọng số liên kết ban đầu của lớp Kohonen

+ Hiệu chỉnh ma trận trọng số lớp Kohonen: hiệu chỉnh trọng số liên kết neuron lớp ẩn Kohonen sao cho mạng có thể học mẫu tốt nhất.

Phương pháp nhận dạng

- Đầu vào: file chứa dữ liệu tín hiệu tiếng nói cần nhận dạng và file chứa thông tin trọng số liên kết neuron lớp ẩn và lớp ra. Ngoài ra đầu vào nguồn âm cũng có thể là từ micro thông qua sound card, lúc này dữ liệu tiếng nói được đọc trong buffer dữ liệu của Windows.

- Đầu ra: kết quả cần nhận dạng

Quá trình nhận dạng tiếng nói được thực hiện qua các bước:

+ Đọc tín hiệu vào: đọc dữ liệu từ file wav hoặc từ buffer dữ liệu âm thanh

+ Xử lý tín hiệu giống như chức năng phân tích LPC ở trên

+ Đọc ma trận trọng số liên kết lớp ẩn và lớp ra của mạng

+ Xác định neuron trung tâm

+ Tra cứu kết quả: tra cứu trên bản đồ topo mạng neuron để đưa ra giá trị cần nhận dạng.

CHƯƠNG 2: KỸ THUẬT XỬ LÝ ẢNH

2.1 TỔNG QUAN VỀ XỬ LÝ ẢNH VÀ VIDEO SỐ

Xử lý ảnh số là lĩnh vực khoa học tương đối mới mẻ và được quan tâm nhiều hiện nay. Hai ứng dụng cơ bản của xử lý ảnh là nâng cao chất lượng hình ảnh và xử lý ảnh cũng như video số với mục đích lưu trữ hoặc truyền qua các hệ thống truyền dẫn hình ảnh.

Trong phần này, chúng ta sẽ đề cập tới nhưng vấn đề sau:

1- Giới thiệu khái niệm cơ bản về ảnh số và xử lý video số, xác định ranh giới của lĩnh vực xử lý ảnh.

2- Giới thiệu các ứng dụng quan trọng của xử lý ảnh trong một số lĩnh vực khoa học

3- Xác định các giai đoạn cơ bản trong quá trình xử lý ảnh;

4- Giới thiệu các thành phần của hệ thống xử lý ảnh tổng quát.

2.1.1 Khái niệm cơ bản về xử lý ảnh

Hình ảnh tĩnh có thể được biểu diễn bởi hàm hai chiều f(x,y), trong đó, x và y là tọa độ không gian phẳng (2 chiều). Khi xét ảnh "đen-trắng", giá trị hàm f tại một điểm được xác định bởi tọa độ (x,y) được gọi là độ chói (mức xám) của ảnh tại điểm này. Nếu x,y,và f là một số hiện hữu các giá trị rời rạc, chúng ta có ảnh số. Xử lý ảnh số là quá trình biến đổi ảnh số trên máy tính (PC). Như vậy, ảnh số được tạo ra bởi một số hữu hạn các điểm ảnh, mỗi điểm ảnh nằm tại một vị trí nhất định và có 1 giá trị nhất định. Một điểm ảnh trong một ảnh còn được gọi là một pixel.

Hệ thống thị giác là cơ quan cảm nhận hình ảnh quang học tương đối hoàn hảo, cho phép con người cảm nhận được hình ảnh quang học trong thiên nhiên. Ứng dụng quan trọng nhất của xử lý ảnh là biến đổi tính chất của ảnh số nhằm tạo ra cảm nhận về sự gia tăng chất lượng hình ảnh quang học trong hệ thống thị giác.

Tuy nhiên, mắt người chỉ cảm nhận được sóng điện từ có bước sóng hạn chế trong vùng nhìn thấy được, do đó ảnh theo quan niệm thông thường gắn liền với hình ảnh quang học mà mắt người có thể cảm nhận. Trong khi đó "ảnh" đưa vào xử lý có thể được tạo ra bởi các nguồn bức xạ có phổ rộng hơn, từ sóng vô tuyến tới tia gamma, ví dụ: ảnh do sóng siêu âm hoặc tia X tạo ra. Nhiều hệ thống xử lý ảnh có thể tương tác với những "ảnh" nêu trên, vì vậy trên thực tế, lĩnh vực xử lý ảnh có phạm vi tướng đối rộng, và liên quan tới nhiều lĩnh vực khoa học khác.

Có thể tạm phân biệt các hệ thống xử lý ảnh theo mức độ phức tạp của thuật toán xử lý như sau:

1- Xử lý ảnh mức thấp: đó là các quá trình biến đổi đơn giản như thực hiện các bộ lọc nhằm khử nhiễu trong ảnh, tăng cường độ tương phản hay độ nét của ảnh. Trong trường hợp này, tín hiệu đưa vào hệ thống xử lý và tín hiệu ở đầu ra là ảnh quang học.

2- Xử lý ảnh mức trung: quá trình xử lý phức tạp hơn, thường được sử dụng để phân lớp, phân đoạn ảnh, xác định và dự đóan biên ảnh, nén anh để lưu trữ hoặc truyền phát. Đặc điểm của các hệ thống xử lý ảnh mức trung là tín hiệu đầu vào là hình ảnh, còn tín hiệu đầu ra là các thành phần được tách ra từ hình ảnh gốc, hoặc luồng dữ liệu nhận được sau khi nén ảnh.

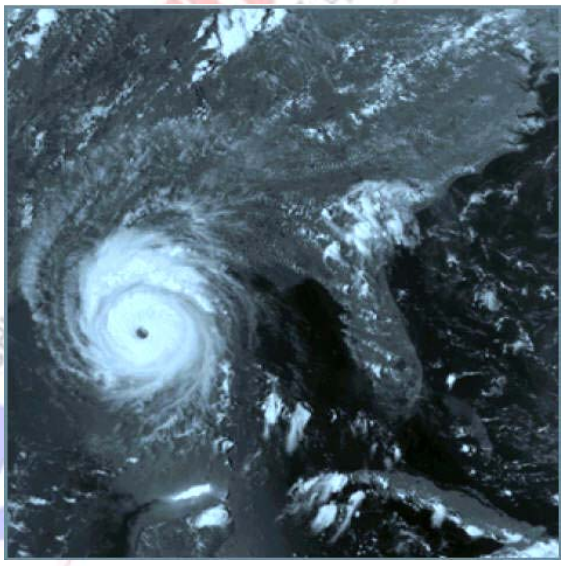
3- Xử lý ảnh mức cao: là quá trình phân tích và nhận dạng hình ảnh. Đây cũng là quá trình xử lý được thực hiện trong hệ thống thì giác của con người.

2.1.2 Lĩnh vực ứng dụng kỹ thuật xử lý ảnh

Như đã nói ở trên, các kỹ thuật xử lý ảnh trước đây chủ yếu được sử dụng để nâng cao chất lượng hình ảnh, chính xác hơn là tạo cảm giác về sự gia tăng chất lượng ảnh quang học trong mắt người quan sát. Thời gian gần đây, phạm vi ứng dụng xử lý ảnh mở rộng không ngừng, có thể nói hiện không có lĩnh vực khoa học nào không sử dụng các thành tựu của công nghệ xử lý ảnh số .

Trong y học các thuật tóan xử lý ảnh cho phép biến đổi hình ảnh được tạo ra từ nguồn bức xạ X -ray hay nguồn bức xạ siêu âm thành hình ảnh quang học trên bề mặt film x-quang hoặc trực tiếp trên bề mặt màn hình hiển thị. Hình ảnh các cơ quan chức năng của con người sau đó có thể được xử lý tiếp để nâng cao độ tương phản, lọc, tách các thành phần cần thiết (chụp cắt lớp) hoặc tạo ra hình ảnh trong không gian ba chiều (siêu âm 3 chiều).

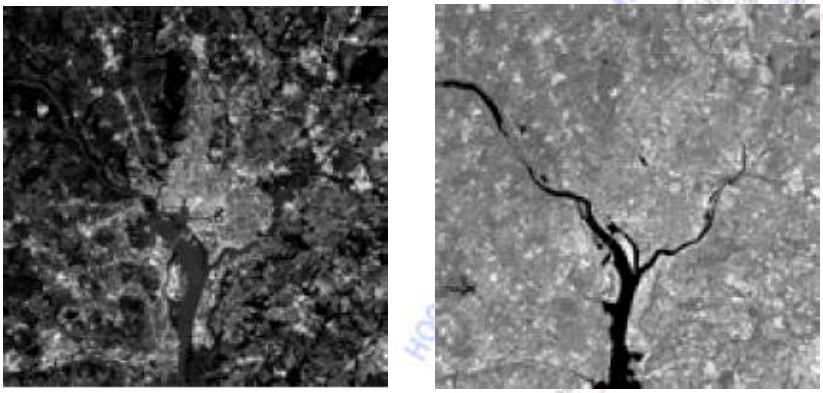
Trong lĩnh vực địa chất, hình ảnh nhận được từ vệ tinh có thể được phân tích để xác định cấu trúc bề mặt trái đất. Kỹ thuật làm nổi đường biên (image enhancement) và khôi phục hình ảnh (image restoration) cho phép nâng cao chất lượng ảnh vệ tinh và tạo ra các bản đồ địa hình 3-D với độ chính xác cao.



Hình 2.1.1 Ảnh nhận được từ vệ tinh dùng trong khí tượng học

Trong ngành khí tượng học, ảnh nhận được từ hệ thống vệ tinh theo dõi thời tiết cũng được xử lý, nâng cao chất lượng và ghép hình để tạo ra ảnh bề mặt trái đất trên một vùng rộng lớn, qua đó có thể thực hiện việc dự báo thời tiết một cách chính xác hơn. Dựa trên các kết quả phân tích ảnh vệ tinh tại các khu vực đông dân cư còn có thể dự đóan quá trình tăng trưởng dân số, tốc độ ô nhiễm môi trường cũng như các yếu tố ảnh hưởng tới môi trường sinh

thái. Ảnh chụp từ vệ tinh có thể thu được thông qua các thiết bị ghi hình cảm nhận được tia sáng quang học ($\lambda = 450 - 520\ nm$) (hình 2a), hoặc tia hồng ngoại ($\lambda = 760 - 900\ nm$) (hình 2b). Trên hình 2a và 2b lần lượt là ảnh bề mặt trái đất nhận được từ 2 ống ghi hình nói trên, dễ dàng nhận thấy sự khác biệt rõ ràng giữa hai ảnh. Đặc biệt trên ảnh 2b, hình con sông được tách biệt rất rõ ràng so với vùng ảnh hai bên bờ. Thiết bị thu hình nhạy cảm với vật thể bức xạ các tia trong miền hồng ngoại sẽ cho ra những bức ảnh trong đó vật thể có nhiệt độ thấp sẽ được phân biệt rõ ràng so với vật thể có nhiệt độ cao hơn. Như vậy việc lựa chọn các thiết bị ghi hình khác nhau sẽ tạo ra ảnh có đặc tính khác nhau, tùy thuộc vào mục đích sử dụng trong các lĩnh vực khoa học cụ thể .



2.2.1a 2.2.1b

Hình 2.1.2 - Ảnh bề mặt trái đất thu được từ hai camera khác nhau

Xử lý ảnh còn được sử dụng nhiều trong các hệ thống quản lý chất lượng và số lượng hàng hóa trong các dây truyền tự động, ví dụ như hệ thống phân tích ảnh để phát hiện bọt khí bên vật thể đúc bằng nhựa, phát hiện các linh kiện không đạt tiêu chuẩn (bị biến dạng) trong quá trình sản xuất hoặc hệ thống đếm sản phẩm thông qua hình ảnh nhận được từ camera quan sát.

Xử lý ảnh còn được sử dụng rộng rãi trong lĩnh vực hình sự và các hệ thống bảo mật hoặc kiểm soát truy cập: quá trình xử lý ảnh với mục đích nhận dạng vân tay hay khuôn mặt cho phép phát hiện nhanh các đối tương nghi vấn cũng như nâng cao hiệu quả hệ thống bảo mật cá nhân cũng như kiểm soát ra vào. Ngoài ra, có thể kể đến các ứng dụng quan trọng khác của kỹ thuật xử lý ảnh tĩnh cũng như ảnh động trong đời sống như tự động nhận dạng, nhận dạng mục tiêu quân sự, máy nhìn công nghiệp trong các hệ thống điều khiển tự động, nén ảnh tĩnh, ảnh động để lưu và truyền trong mạng viễn thông v.v.

2.1.3 Các giai đoạn chính trong xử lý ảnh

1- Thu nhận hình ảnh: đây là giai đoạn đầu tiên và quan trọng nhất trong tòan bộ quá trình xử lý ảnh. Ảnh nhận được tại đây chính là ảnh gốc để đưa vào xử lý tại các giai đoạn sau, trường hợp ảnh gốc có chất lượng kém hiệu quả của các bước xử lý tiếp theo sẽ bị giảm. Thiết bị thu nhận có thể là các ông ghi hình chân không (vidicon, plumbicon v.v.) hoặc thiết bị cảm biến quang điện bán dẫn CCD (Charge-Coupled Device).

2- Tiền xử lý ảnh: giai đoạn xử lý tương đối đơn giản nhằm nâng cao chất lượng ảnh để trợ giúp cho các quá trình xử lý nâng cao tiếp theo, ví dụ: tăng độ tương phản, làm nổi đường biên, khử nhiễu v.v.

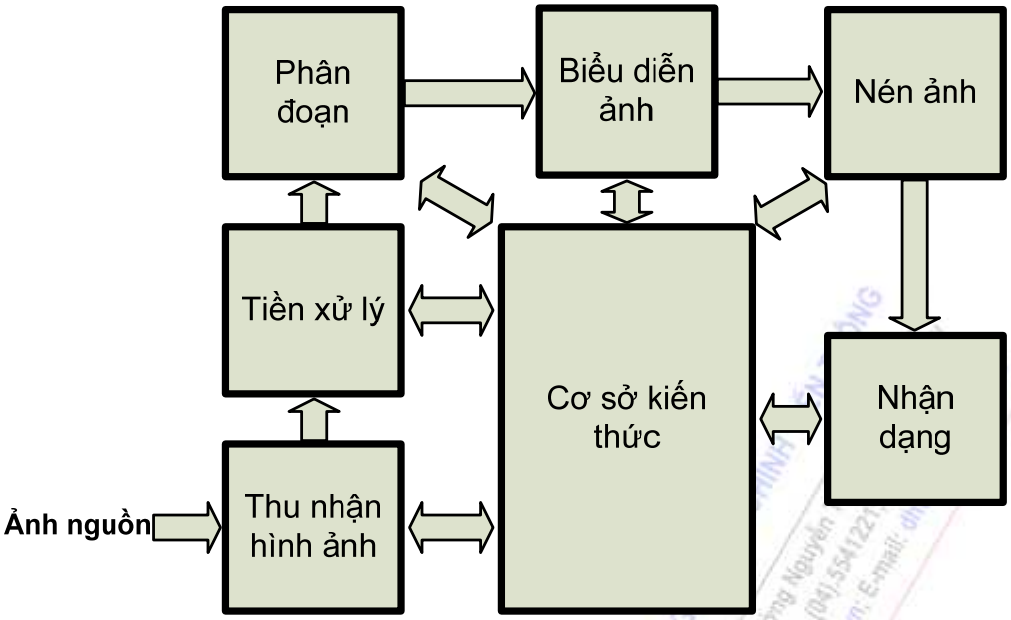
3- Phân đoạn: là quá trình tách hình ảnh thành các phần hoặc vật thể riêng biệt. Đây là một trong nhưng vấn đề khó giải quyết nhất trong lĩnh vực xử lý ảnh. Nếu thực hiện tách quá chi tiết thì bài toán nhận dạng các thành phần được tách ra trở nên phức tạp, còn ngược lại nếu quá trình phân đoạn được thực hiện quá thô hoặc phân đoạn sai thì kết quả nhận được cuối cùng sẽ không chính xác.

4- Biểu diễn và mô tả: là quá trình xử lý tiếp sau khâu phân đoạn hình ảnh. Các vật thể sau khi phân đoạn có thể được mô tả dưới dạng chuỗi các điểm ảnh tạo nên ranh giới một vùng, hoặc tập hợp tất cả các điểm ảnh nằm trong vùng đó. Phương pháp mô tả thông qua ranh giới vùng thường được sử dụng khi cần tập trung sự chú ý vào hình dạng bên ngòai của chi tiết ảnh như độ cong, các góc cạnh v.v. Biểu diễn vùng thường được sử dụng khi chúng ta quan tâm tới đặc tính bên trong của vùng ảnh như đường vân (texture) hay hình dạng skeletal.

5- Nén ảnh - bao gồm các biện pháp giảm thiểu dung lượng bộ nhớ cần thiết để lưu trữ hình ảnh, hay giảm băng thông kênh truyền, cần thiết để truyền tín hiệu hình ảnh số.

6- Nhận dạng: là quá trình phân loại vật thể dựa trên cơ sở các chi tiết mô tả vật thể đó (ví dụ các phương tiện giao thông có trong ảnh).

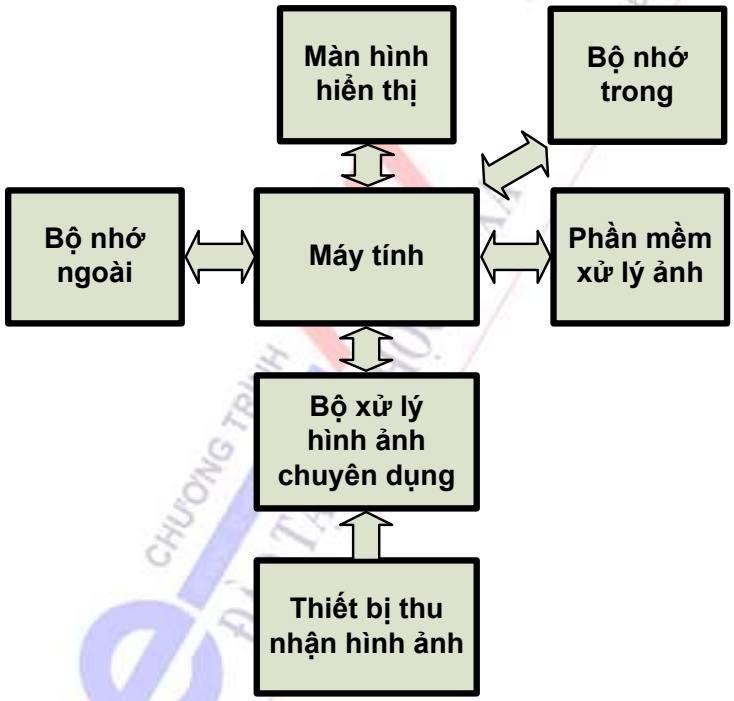
Các quá trình xử lý liệt kê ở trên đều được thực hiện dưới sự giám sát và điều khiển dựa trên cơ sở các kiến thức về lĩnh vực xử lý ảnh. Các kiến thức cơ bản có thể đơn giản như vị trí vùng ảnh nơi có những thông tin cần quan tâm, như vậy có thể thu nhỏ vùng tìm kiếm. Trường hợp phức tạp hơn, cơ sở kiến thức có thể chứa danh sách tất cả những hư hỏng có thể gặp trong quá trình kiểm sóat chất lượng thành phẩm hoặc các ảnh vệ tinh có độ chi tiết cao trong các hệ thống theo dõi sự thay đổi môi trường trong một vùng. Ngoài việc điều khiển hoạt động của từng modul xử lý ảnh (hình 2.1.3), cơ sở kiến thức còn sử dụng để thực hiện việc điều khiển tương tác giữa các modules. Trong hình 2.1.3, quá trình điều khiển nói trên được biểu diễn bằng mũi tên hai chiều.



Hình 2.1.3 Các giai đoạn xử lý ảnh số

2.1.4 Các phần tử của hệ thống xử lý ảnh số

Cấu trúc một hệ thống xử lý ảnh đa dụng dùng để thực hiện các giai đoạn xử lý ảnh đề cập ở trên được mô tả trên hình 2.1.4.

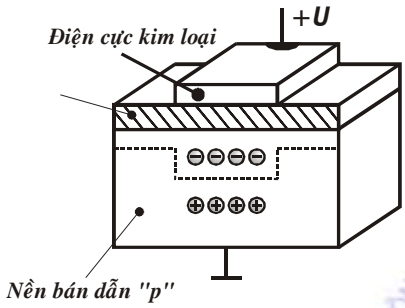


Hình 2.1.4 Các thành phần chính của hệ thống xử lý ảnh

Thiết bị thu nhận hình ảnh: là thiết bị biến đổi quang-điện, cho phép biến đổi hình ảnh quang học thành tín hiệu điện dưới dạng analog hay trực tiếp dưới dạng số. Có nhiều dạng cảm biến cho phép làm việc với ánh sáng nhìn thấy hoặc hồng ngoại. Hai loại thiết bị biến đổi quang – điện chủ yếu thường được sử dụng là đèn ghi hình điện tử và chip CCD (Charge Couple Device – linh kiện ghép điện tích).

Ống vidicon là đại diện tiêu biểu cho họ đèn ghi hình điện tử được sử dụng tương đối rộng rãi trong camera màu cũng như đen trắng. Ống Vidicon có kích thước nhỏ gọn (đường kính 18-25 mm, chiều dài 10-12 cm), nhẹ, cấu tạo đơn giản, dễ sử dụng. Đèn hình này sử dụng nguyên lý hiệu ứng quang điện trong và nguyên lý tích lũy điện tích.

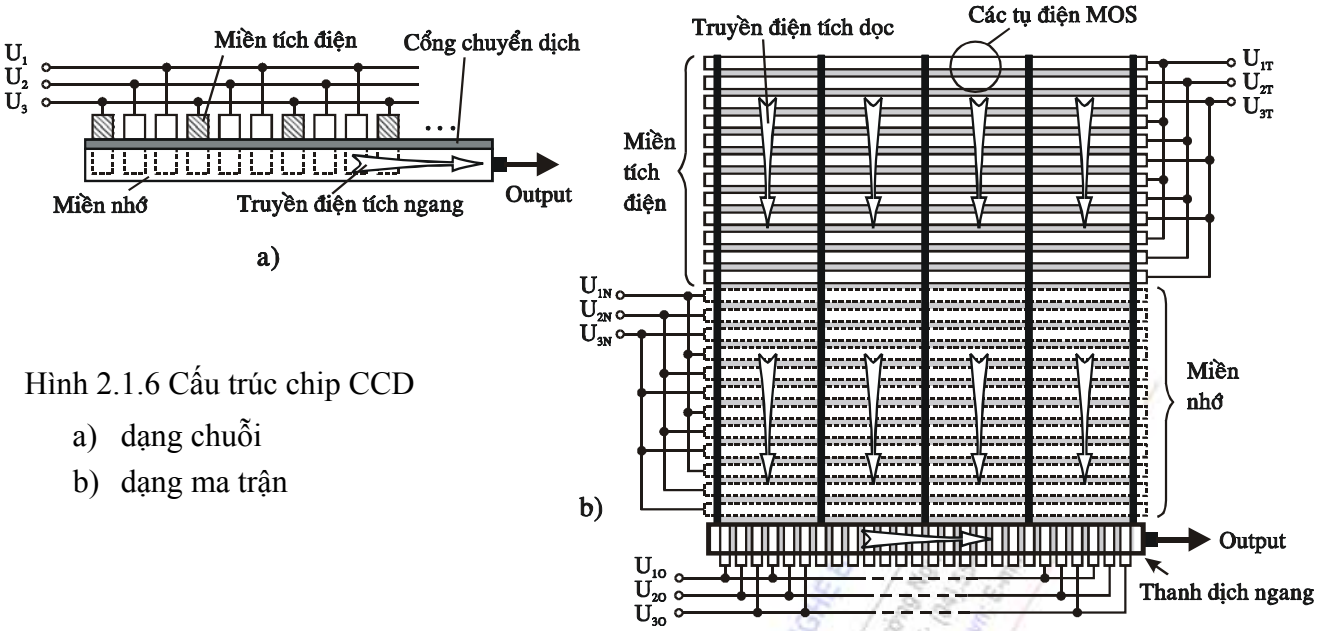
Chip CCD là linh kiện bán dẫn có khả năng biến đổi năng lượng quang phổ thành tín hiệu điện. Thành phần chính của chip CCD là các tụ điện MOS (Metal-Oxide-Semiconductor). Tụ điện MOS được hình thành bởi ba lớp: một má tụ bằng kim loại, chất điện môi nằm giữa là lớp SiO_2 và một má tụ bằng lớp bán dẫn loại p hoặc n (hình 2.1.5).



Hình 2.1.5 Cấu trúc tụ điện MOS

Một chuỗi tụ điện MOS phân bố đều trên bề mặt chip CCD được biểu diễn trên hình 2.1.6a, mỗi tụ điện với bề mặt cảm quang là má bán dẫn sẽ tạo ra một điểm trên hình ảnh thu được.Theo phương pháp dịch chuyển điện tích, các chip CCD có thể chia ra làm hai loại: CCD dạng chuỗi (một chiều) và dạng ma trận (hai chiều).

Trên Hình 2.1.6a là cấu trúc chip CCD dạng chuỗi, quá trình ghi (tích điện) và đọc được thực hiện tại hai khu vực khác nhau, gọi là miền tích điện và miền nhớ. Hai khu vực trên được ngăn cách bởi cổng chuyển dịch. Sau khi kết thúc quá trình tích điện tại các phần tử cảm quang, điện tích sẽ được truyền song song qua cổng chuyển dịch vào thanh dịch ngang (không nhạy cảm với ánh sáng) tức miền nhớ. Sau khi cổng chuyển dịch đóng lại, quá trình ghi và đọc tại hai miền nói trên sẽ được tiến hành song song.



Hình 2.1.6 Cấu trúc chip CCD

a) dạng chuỗi

b) dạng ma trận

Chip CCD sử dụng trong máy quay video thường có cấu trúc ma trận (hình 2.16b). Các phần tử cảm quang trong CCD tập hợp thành ma trận hai chiều, quá trình "đọc" tín hiệu được thực hiện theo chiều ngang và chiều dọc. Có nhiều cách tổ chức quá trình ghi và đọc tín hiệu trong CCD, nhưng phổ biến nhất là phương pháp dịch chuyển từng ảnh. Khi sử dụng phương pháp này, trong chip CCD được thiết kế một miền nhớ, không tiếp xúc với ánh sáng và có diện tích bằng miền tích lũy – là ma trận các phần tử cảm quang.

Điện tích thu được tại miền tích lũy được chuyển về miền nhớ. Sau đó, quá trình ghi ảnh tại miền tích lũy và đọc ảnh từ miền nhớ vào thanh dịch ngang sẽ được tiến hành song song.Từng dòng ảnh được dịch chuyển xuống thanh dịch ngang, sau đó các gói điện tích ứng với các điểm trong dòng ảnh sẽ được đẩy ra lần lượt khỏi thanh dịch. Sau khi toàn bộ ảnh trong miền nhớ được đọc ra hết, một ảnh mới từ miền tích lũy sẽ lại được chuyển về đây.Với những tính năng vượt trội trước ống ghi hình điện tử cổ điển, linh kiện biến đổi - quang điện CCD được sử dụng rất rộng rãi trong công nghệ truyền hình và ảnh số. Hầu hết các camera quay video dân dụng và bán chuyên nghiệp (semi-professional) được thiết kế trên cơ sở chip CCD.

Bộ nhớ trong và ngoài trong các hệ thống xử lý ảnh số thường có dung lượng rất lớn dùng để lưu trữ ảnh tĩnh và động dưới dạng số. Ví dụ, để lưu một ảnh số đen trắng kích thước 1024x1024 điểm, mỗi điểm được mã hóa bằng 8 bits cần bộ nhớ ~1MB. Để lưu một ảnh màu không nén, dung lượng bộ nhớ phải tăng lên gấp 3. Bộ nhớ số trong hệ thống xử lý ảnh có thể chia làm 3 loại: 1- bộ nhớ đệm trong máy tính để lưu ảnh trong quá trình xử lý. Bộ nhớ này phải có khả năng ghi/đọc rất nhanh (ví dụ 25 hình/s); 2- bộ nhớ ngoài có tốc độ truy cập tương đối nhanh, dùng để lưu thông tin thường dùng. Các bộ nhớ ngoài có thể là ổ cứng, thẻ nhớ flash v.v.. 3- Bộ nhớ dùng để lưu trữ dữ liệu. Loại bộ nhớ này thường có dung lượng lớn, tốc độ truy cập không cao. Thông dụng nhất là đĩa quang ghi 1 lần (ROM) hoặc nhiều lần (ROM) như đĩa DVD có dung lượng 4.7GB (một mặt). Ngoài ra trong hệ thống xử lý ảnh còn sử dụng các thiết bị cho phép lưu ảnh trên vật liệu khác như giấy in, giấy in nhiệt, giấy trong, đó có thể là máy in phun, in laser, in trên giấy ảnh đặc biệt bằng công nghệ nung nóng v.v.

Bộ xử lý ảnh chuyên dụng:

Xử dụng chip xử lý ảnh chuyên dụng, có khả năng thực hiện nhanh các lệnh chuyên dùng trong xử lý ảnh. Cho phép thực hiện các quá trình xử lý ảnh như lọc, làm nổi đường bao, nén và giải nén video số v.v.. Trong bộ xử lý ảnh thường tích hợp bộ nhớ đệm có tốc độ cao.

Màn hình hiển thị: Hệ thống biến đổi điện - quang hay đèn hình (đen trắng cũng như màu) có nhiệm vụ biến đổi tín hiệu điện có chứa thông tin của ảnh (tín hiệu video) thành hình ảnh trên màn hình. Có hai dạng display được sử dụng rộng rãi là đèn hình CRT (Cathode-Ray Tube) và màn hình tinh thể lỏng LCD (Liquid Crystal Display). Đèn hình CRT thường có khả năng hiển thị màu sắc tốt hơn màn hình LCD nên được dùng phổ biến trong các hệ thống xử lý ảnh chuyên nghiệp.

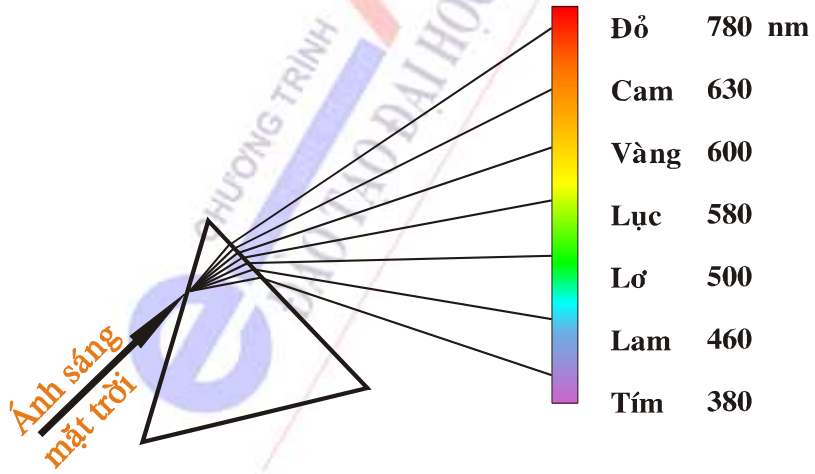
Máy tính: có thể là máy tính để bàn cũng như siêu máy tính có chức năng điều khiển tất cả các bộ phận chức năng trong hệ thống xử lý ảnh số.

2.1.5 Biểu diễn ảnh số

Trong phần này, chúng ta sẽ đề cập tới một số những kiến thức cơ bản và ký hiệu được sử dụng trong lĩnh vực xử lý ảnh. Đó là các vấn đề về ánh sáng, màu sắc, khả năng tiếp thu hình ảnh quanh học của hệ thống thị giác. Tiếp theo là quá trình biến đổi ảnh analog thành tín hiệu ảnh số, cách biểu diễn hình ảnh số, ảnh hưởng của quá trình lấy mẫu và lượng tử hóa tới chất lượng ảnh số. Ngoài ra, trong phần này sẽ xét tới quan hệ tương quan giữa các điểm ảnh, những kiến thức cơ bản này sẽ được sử dụng rộng rãi trong các phần sau của bài giảng này.

2.1.5.1 Ánh sáng, màu sắc và hình ảnh

Phổ của các sóng điện từ trong thiên nhiên trải dài từ tia gamma (10^{-12} m) đến sóng radio (10^{-4}-10^{4} m). Mắt người chỉ cảm nhận được những sóng điện từ có bước sóng từ 380 nm (tia màu tím) đến 780 nm (tia màu đỏ) (hình 2.1.7).



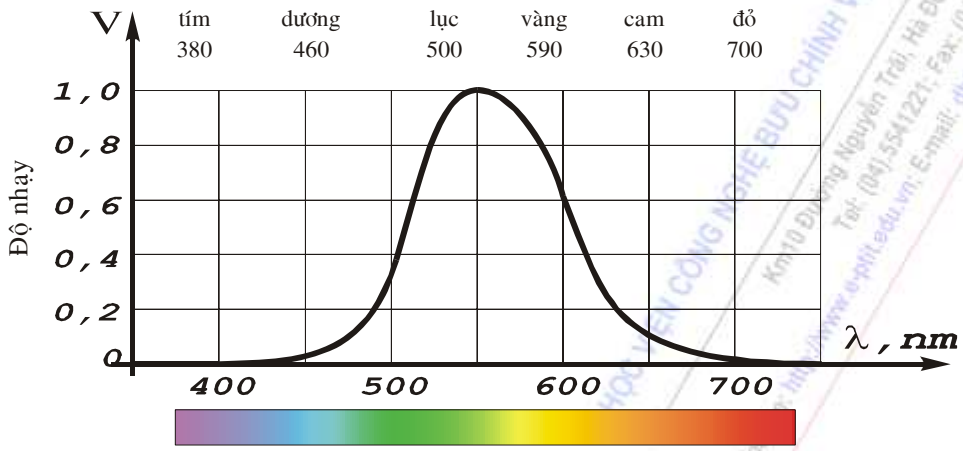
Hình 2.1.7 Các màu quang phổ trong ánh sáng mặt trời

Các bức xạ điện từ đặc biệt nói trên được gọi là ánh sáng. Trong lĩnh vực xử lý ảnh, người ta chỉ quan tâm đến phần năng lượng bức xạ mà mắt người cảm nhận được. Các đại lượng trắc quang được sử dụng để đánh giá tính chất của nguồn sáng: quang thông, độ sáng,

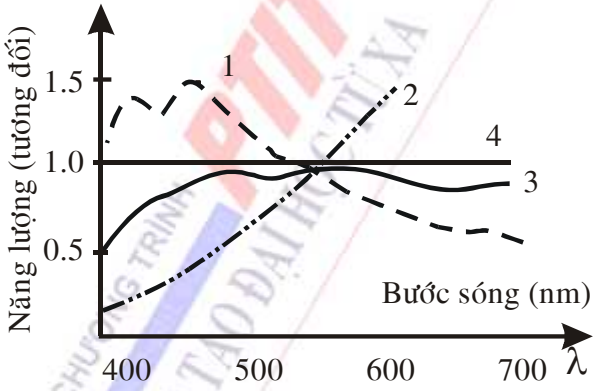
độ rọi và độ chói. Để đánh giá được tác động của ánh sáng lên mắt người, chúng ta phải lưu ý đến hai yếu tố quan trọng:

1 – Mắt có độ nhậy cảm không đồng đều đối với các tia bức xạ có bước sóng khác nhau. Trên đồ thị độ nhạy của mắt người (Hình 2.1.8) ta thấy vùng lục – vàng (λ ~ 555 nm) là nơi nhậy cảm nhất của mắt.

2 – Mật độ phân bố công suất của các nguồn sáng trong thiên nhiên không đồng đều trên trục tần số (hình 2.1.9).



Hình 2.1.8 Đáp ứng phổ (độ nhạy) của mắt người



Hình 2.1.9 Đồ thị phân bố công suất của các nguồn bức xạ:

1- Bầu trời phía bắc 2- Đèn điện 3- Mặt trời

4- Nguồn sáng đẳng năng E

Trường hợp tổng quát, một nguồn bức xạ có thể đặc trưng bởi hàm mật độ phân bố công suất trên trục tần số $\rho(\lambda)$:

$$\rho(\lambda) = \frac{dP(\lambda)}{d\lambda} \qquad \text{(Watt/µm)} \qquad (2.1.1)$$

λ - bước sóng (µm);

$P(\lambda)$ – công suất nguồn bức xạ có bước sóng λ (Watt);

Công suất toàn phần của nguồn ánh sáng có phổ liên tục (ánh sáng mặt trời, áng sáng đèn đốt nóng v.v.) sẽ bằng:

$$P_{\sum} = \int_{370}^{780} \rho(\lambda) \cdot d\lambda \qquad (2.1.2)$$

Để đặc trưng cho phần năng lượng bức xạ có ích (cảm nhận được bằng mắt) ta đưa ra khái niệm quang thông F

$$F = K \int_{370}^{780} V(\lambda)\rho(\lambda) \cdot d\lambda \qquad \text{(lumen}^{1}\text{)} \qquad (2.1.3)$$

$V(\lambda)$ - hàm độ nhạy phổ tương đối của mắt người (không có đơn vị).

Trên đồ thị $V(\lambda)$ (hình 2.1.8) ta thấy mắt người cảm nhận tốt nhất tia bức xạ có bước sóng 555 nm, do đó $V(555nm) = 1$.

K là hệ số tỷ lệ giữa quang thông và công suất bức xạ.

Một số ví dụ về đơn vị quang thông:

1- Bóng đèn sợi tóc có thường có hệ số phát sáng là 8 – 15 lumen/watt, khi công suất bóng là P=100 watt, quang thông của đèn sẽ bằng $F \cong 800 \div 1500$ *lumen* .

2- Để có hình ảnh đủ độ chói trên màn hình 6x8m, quang thông của đèn chiếu phải đạt là 8000 lumen.

Nói chung, quang thông của một nguồn sáng có thể phân bố không đồng đều trong mọi phương hướng. Do đó ta định nghĩa đại lượng độ sáng I đặc trưng cho khả năng phát sáng của nguồn sáng theo một hướng nào đó (hình 2.1.10):

$$I = \frac{dF}{d\Omega}, \text{(candela}^{2}\text{)} \qquad (2.1.4)$$

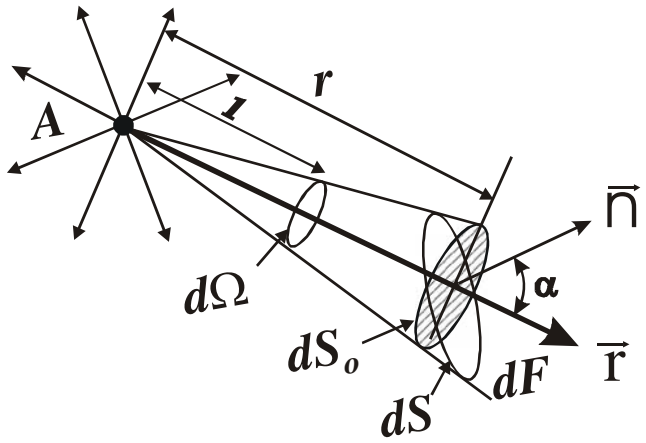
Ω - góc khối; $d\Omega = \frac{dS}{r^2}$ (sr –steradian)

dF – quang thông truyền qua góc khối dΩ.

---

[1] Lumen (lat.) – nguồn sáng

[2] Đơn vị độ sáng là candela (cd) – là 1 trong sáu đơn vị cơ bản hệ SI. 1 candela là độ sáng đo theo phương vuông góc của bề mặt platin, ở nhiệt độ đông đặc, (2046,5 K), có diện tích 1/60π=0.0053 cm^2.



Hình 2.1.10 Minh họa độ sáng của nguồn điểm

Góc khối lớn nhất là $\Omega = \frac{4\pi r^2}{r^2} = 4\pi$, do đó độ sáng trung bình của nguồn điểm[3] sẽ bằng toàn bộ quang thông chia cho góc 4π:

$$I_{tb} = \frac{F_A}{\Omega} = \frac{F_A}{4\pi}$$

Thí dụ: Bóng đèn sợi tóc công suất P=100 watt, $F \cong 800 \div 1500 \ \ lumen$, sẽ cho độ sáng trung bình là:

$$I_{tb} = \frac{F_a}{4\pi} = \frac{800 \div 1500}{4\pi} = 60 \div 120 \ \ candela$$

Quang thông và độ sáng là hai đại lượng đặc trưng cho nguồn sáng.

Độ rọi E là đại lượng đặc trưng cho bề mặt được chiếu sáng.

Độ rọi là mật độ phân bố quang thông trên bề mặt được chiếu sáng:

$$E = \frac{dF}{dS} \ (\text{lux}^4) \qquad (2.1.5)$$

1 lux là độ rọi lên một bề mặt khi 1 m^2 bề mặt đó nhận được quang thông bằng 1 lumen.

Xét bề mặt được chiếu sáng dS bởi nguồn điểm A (Hình 2.1.10). Diện tích bề mặt hình cầu giới hạn trong góc khối dΩ là dS_0., α là góc giữa pháp tuyến của dS và pháp tuyến dS_0.

[3] Nguồn điểm là nguồn sáng có kích thước (d) nhỏ hơn nhiều lần so với khoảng cách (l) đến bề mặt được chiếu sáng (l ≥ 10d).

[4] lux (lat.) – chiếu sáng

$$d\Omega \cong \frac{dS_0}{r^2} = \frac{dS \cdot \cos\alpha}{r^2};$$

$$E_S = \frac{dF}{dS} = \frac{Id\Omega}{dS} = \frac{I}{dS}\frac{dS \cdot \cos\alpha}{r^2} \qquad (2.1.6)$$

$$E_S = \frac{I \cdot \cos\alpha}{r^2}$$

Như vậy độ rọi của bề mặt được chiếu sáng bởi nguồn điểm tỷ lệ nghịch với bình phương khoảng cách giữa nguồn sáng và bề mặt đó.

Bảng dưới đây cho ta độ rọi trong một số trường hợp:

2.1.6 Bảng 2.1.1

Vật được rọi sáng	Độ rọi (lx)
Màn hình chiếu bóng (kino)	40-200
Trường quay (studio)	2000
Trang sách lúc đọc	30
Vật thể trong bóng râm (ban ngày)	1000
Vật thể ngoài nắng	100000

Độ chói L là đại lượng đặc trưng cho bề mặt phát sáng (trong khi độ rọi đặc trưng cho bề mặt được chiếu sáng).

Độ chói là mật độ độ sáng trên bề mặt phát sáng. Độ chói đặc trưng cho mức độ sáng của nguồn sáng. Cho bề mặt phát sáng S_0. Theo hướng trực giao với S_0, độ chói sẽ bằng:

$$L_0 = \frac{I_0}{S} \qquad (candel/m^2) \qquad (2.1.7)$$

Đơn vị độ chói còn gọi là Nít (Nít là độ chói của nguồn sáng có diện tích 1 m^2 và cường độ sáng là 1 candela theo hướng vuông góc với bề mặt nguồn sáng)

Dưới đây là độ chói của một số nguồn sáng:

Bảng 2.1.2

Vật phát sáng	Độ chói (cd/m^2)
Màn hình chiếu phim	10-30
Bóng hình TV	40-80
Sợi tóc đèn chiếu sáng	$5\ 10^6$- 10^7
Mặt trời	$1.5\ 10^9$

2.1.6.1 Màu sắc và các thông số đặc trưng

Cảm nhận về màu sắc là kết quả của sự nhận biết chủ quan của mắt người dưới tác động của ánh sáng. Để giải thích cơ chế cảm nhận màu của mắt người, các nhà khoa học đã đưa ra nhiều giả thuyết khác nhau, trong đó, thuyết ba thành phần cảm thụ màu, do nhà bác học Nga Lômônôxốp đưa ra năm1756 được công nhận rộng rãi hơn cả (T. Young năm 1801 – cũng đưa ra giả thiết về "ba nhóm tế bào cảm nhận" trong mắt và mô hình mắt người). Theo thuyết này, các tế bào hình nón trên võng mạc có thể chia ra ba loại. Mỗi loại tế bào đặc biệt nhạy cảm với những vùng phổ nhất định trong dải quang phổ – vùng sóng ngắn (màu xanh lam - B), vùng sóng trung (màu lục - G) và vùng sóng dài (màu đỏ - R). Nếu kích thích riêng rẽ từng loại tế bào, mắt sẽ cảm nhận được các màu sắc bão hòa tương ứng. Khi quán sát hình ảnh không có những màu bão hòa, nguồn sáng của ảnh có chứa tất cả các thành phần quang phổ tác động cùng một lúc lên cả ba loại tế bào. Sóng điện từ có bước sóng khác nhau tác động lên các tế bào không đồng đều, sự khác biệt về tỷ lệ kích thích sẽ tạo nên cảm nhận về các sắc màu khác nhau. Thuyết ba thành phần cảm thụ màu được chứng minh qua nhiều thực nghiệm và phù hợp với luật pha trộn màu mà chúng ta sẽ đề cập tới dưới đây.

Trên phương diện sinh học (cảm giác chủ quan), ánh sáng được cảm nhận thông qua ba đại lượng chính là: độ sáng, sắc màu và độ bão hòa màu.

Độ sáng – phụ thuộc vào công suất của nguồn sáng, nguồn sáng trắng 500 W sẽ có độ sáng lớn hơn nguồn sáng trắng 15 W.

Sắc màu (sắc điệu) là tính chất đặc trưng của màu mà qua đó ta nhận biết được màu đỏ, xanh, vàng v.v.

Độ bão hòa màu là cường độ về sắc màu qua đó ta có thể phân biệt được màu đỏ thẫm hay màu đỏ nhạt v.v.

Khi đánh giá về số lượng của các đại lượng trên (bằng cách đo lường khách quan) chúng ta sẽ sử dụng ba khái niệm tương đương là: độ chói, bước sóng trội và độ sạch của màu. Sắc điệu của nguồn sáng tương đương với bước sóng λ có năng lượng lớn nhất trong phổ của nguồn sáng đó, đại lượng này được gọi là bước sóng trội λ. Độ bão hòa của một màu có bước sóng trội λ được tính bằng:

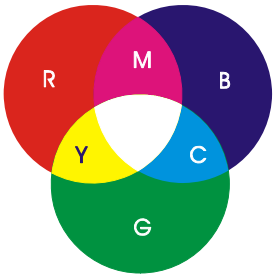
$$p = F_\lambda / \left(F_\lambda + F_E \right) \qquad (2.1.8)$$

p – là độ sạch màu, đại lượng xác định số lượng ánh sáng trắng trong nguồn sáng hay độ pha loãng của nguồn sáng đó. Như vậy, màu quang phổ sẽ có độ sạch là $p = 100\%$, đối với màu trắng $p = 0$.

Dựa trên những tính chất của hệ thị giác như độ phân giải, tính lưu ảnh và thuyết ba thành phần cảm nhận màu, để tạo ra cảm giác về một màu nào đó, người ta có thể thực hiện pha trộn các màu cơ bản theo những phương pháp sau:

1. Phương pháp cộng màu quang học

Phương pháp này dựa trên khả năng tổng hợp màu khi các nguồn bức xạ được chiếu lên một mặt phẳng. Các nguồn bức xạ có thể được rọi cùng một lúc hay nối tiếp nhau với một tốc độ tương đối lớn (còn gọi là phép trộn màu theo thời gian), khi đó, ánh sáng thứ cấp phản xạ từ bề mặt của mặt phẳng trên sẽ mang lại cho người quan sát cảm nhận về một màu sắc tổng hợp. Sắc độ màu mới phụ thuộc vào tỷ lệ công suất của các bức xạ thành phần (hình 2.1.1).



Hình 2.1.11 Trộn các màu cơ bản

2. Phương pháp trộn màu không gian

Khi trộn màu không gian, các phần tử ảnh mang sắc màu cơ bản nằm độc lập với nhau trong không gian, nếu những phần tử này nằm gần nhau và có kính thước nhỏ thì mắt cảm nhận chúng như một điểm ảnh, màu sắc của điểm ảnh này phụ thuộc vào tỷ lệ công suất của các màu cơ bản. Hình dạng của các phần tử có thể là các điểm tròn hay vạch màu nhỏ. Tỷ lệ công suất của các màu cơ bản có thể thay đổi bằng cách thay đổi độ chói của các phần tử ảnh tham gia trộn màu hay thay đổi kích thước của chúng. Việc tái tạo hình ảnh màu trên màn hình vô tuyến thường được thực hiện bằng phương pháp trộn màu trong không gian.

3. Phương pháp trừ

Để tạo ra một màu mới, ngoài phương pháp cộng các màu đơn sắc, người ta còn có thể sử dụng phương pháp loại bỏ bớt một số màu từ ánh sáng trắng. Thí dụ, nếu cho ánh sáng trắng qua môi trường hấp thụ (kính lọc) các tia màu đỏ, ta sẽ nhận được màu lơ. Phương pháp này thường dùng trong kỹ thuật ảnh màu, in ấn và trong hội họa. Đặc điểm của phương pháp trừ là độ chói của màu được tạo ra bao giờ cũng nhỏ hơn độ chói của màu trắng ban đầu.

Ngoài ra, trong kỹ thuật truyền hình stereo, người ta còn sử dụng phương pháp trộn màu mang tên "binocular". Người ta pha trộn hai hay nhiều màu bằng cách tác động các màu đó riêng rẽ lên mắt phải và mắt trái của người quan sát, kết quả là hệ thị giác sẽ cảm nhận được một màu mới.

2.1.6.2 Các định luật trộn màu cơ bản

Trên cơ sở các kết quả nhận được qua nhiều công trình thực nghiệm về cảm nhận màu sắc của hệ thống thị giác, nhà bác học người Đức H.Grassmann đã đưa ra bốn định luật về trộn màu:

Định luật thứ nhất: Bất kỳ một màu sắc nào cũng có thể tạo được bằng cách trộn 3 màu cơ bản độc lập tuyến tính với nhau.

Ba màu được gọi là độc lập tuyến tính khi một trong những màu đó không thể tạo ra bằng cách pha trộn hai màu còn lại được. Như vậy, ta có thể viết ra được phương trình so màu như sau:

$$f'F = r'R + g'G + b'B \qquad (2.1.9)$$

$f'F$ - nguồn ánh sáng bất kỳ, có đơn vị là F và số lượng ánh sáng là f'; R,G,B – đơn vị màu cơ bản; r', g', b' - số lượng các màu cơ bản R, G, B, còn gọi là modul của các màu đó.

Năm 1931, theo quy định của tổ chức quốc tế CIE (Commission Internationale d'Eclairage - International Commission on Ilumination – ủy ban quốc tế về ánh sáng) ba nguồn bức xạ đơn sắc màu đỏ, lục và lam tương ứng có bước sóng:

$$\lambda_R = 700\ nm$$
$$\lambda_G = 546{,}1\ nm$$
$$\lambda_B = 435{,}8\ nm$$

Ba màu trên được gọi là ba màu cơ bản. Mỗi màu cơ bản sẽ có một màu bổ xung tương ứng, khi pha trộn màu cơ bản và màu bổ xung của nó chúng ta sẽ nhận được màu trắng. Các cặp màu cơ bản và màu bổ xung là: Đỏ – Lơ (Cyan), Lục – Mận chín (Magenta), Lam – Vàng (Yellow).

Định luật thứ hai: Sự biến đổi liên tục của các hệ số công suất của các màu cơ bản sẽ dẫn đến sự biến đổi liên tục của màu sắc tổng hợp, nó chuyển từ màu này sang màu khác.

Khi thay đổi công suất của các nguồn sáng cơ bản nhưng giữ nguyên tỷ lệ công suất thì màu tổng hợp sẽ không thay đổi sắc độ, chỉ có sự thay đổi về độ chói mà thôi. Vì vậy, tỷ lệ tương đối giữa ba màu cơ bản R:B:G sẽ quyết định màu sắc của màu tổng hợp.

Định luật thứ ba: Màu sắc tổng hợp của nhiều nguồn sáng chỉ xác định bởi các thành phần màu sắc của từng nguồn sáng chứ không phụ thuộc vào thành phần phổ của chúng.

Theo định luật này ta có thể dễ dàng định lượng màu sắc của ánh sáng tổng hợp được tạo ra khi pha trộn nhiều nguồn sáng với nhau.

Cho phương trình so màu của 2 nguồn sáng là:

$$F_1 = r_1'R + g_1'G + b_1'B$$
$$F_2 = r_2'R + g_2'G + b_2'B$$

Phương trình so màu của nguồn sáng tổng sẽ là:

$$F = F_1 + F_2 = \left(r_1' + r_2'\right)R + \left(g_1' + g_2'\right)G + \left(b_1' + b_2'\right)B \qquad (2.1.10)$$

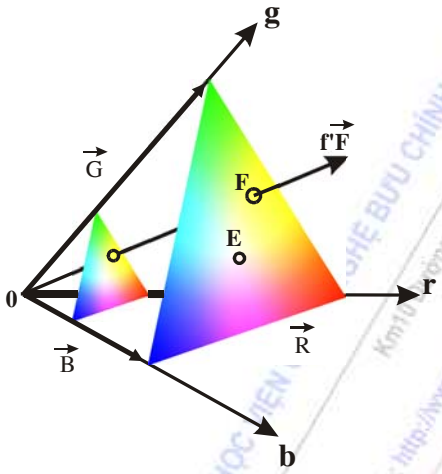
Từ phương trình (2.1.10) ta thấy rằng, tọa độ màu của màu được tạo ra bằng tổng tọa độ màu tương ứng của các màu dùng để trộn.

Định luật thứ tư: độ chói của màu tổng hợp bằng tổng số độ chói của các thành phần màu.

$$L_{\sum} = L_R + L_G + L_B$$

2.1.6.3 Không gian màu RGB

Để xác định một màu F, theo định luật trộn màu thứ nhất, ta chỉ cần biết số lượng của ba màu cơ bản trong phương trình (2.1.9). Vì thế màu F có thể được biểu diễn bằng một điểm trong không gian màu ba chiều rgb hay như một vector nối từ gốc tọa độ tới điểm đó, các vector màu đơn vị sẽ là $\vec{R},\vec{G},\vec{B}$ (Hình 2.1.12). Độ chói của màu F sẽ bằng chiều dài (modul) của vector $\vec{F}$, sắc – tương ứng với phương hướng của $\vec{F}$ trong không gian rgb. Tổng ba vector đơn vị sẽ cho ta màu trắng chuẩn.



Hình 2.1.12 Không gian màu **rgb**

Xét một màu bất kỳ trong không gian màu rgb, xác định bởi phương trình:

$$f'F = r'R + g'G + b'B \qquad (2.1.11)$$

Ta thấy ba hệ số r',g',b' cho ta biết cả về số lượng lẫn chất lượng của nguồn sáng. Nếu chỉ cần xét đến chất lượng hay thành phần "sắc" của màu, chúng ta không cần biết đến giá trị tuyệt đối r',g',b', mà chỉ cần biết đến số lượng tương đối giữa các thành phần màu cơ bản R, G, B, tìm được qua các phương trình sau:

$$\left.\begin{aligned} r &= \frac{r'}{r'+g'+b'} = \frac{r'}{m};\\ g &= \frac{g'}{r'+g'+b'} = \frac{g'}{m};\\ b &= \frac{b'}{r'+g'+b'} = \frac{b'}{m} \end{aligned}\right\} \qquad (2.1.12)$$

$m = r' + g' + b' = f'$ - độ chói của màu.

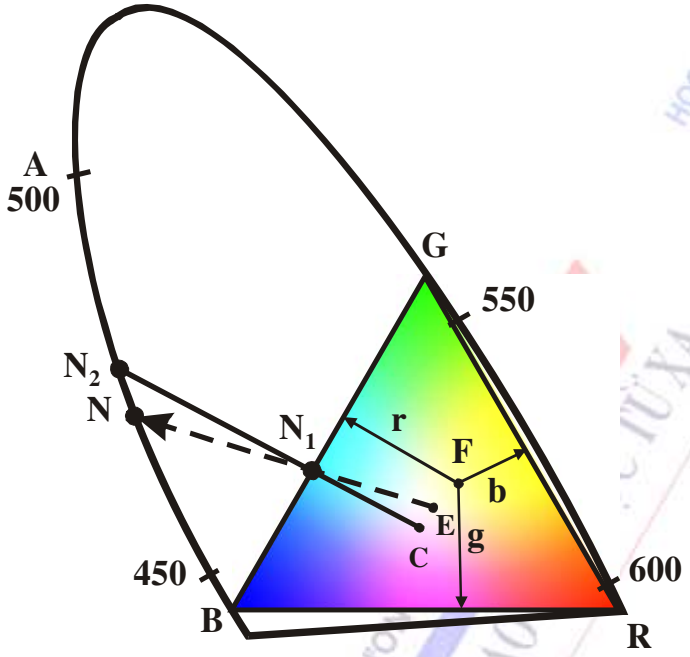
Chia (2.1.11) cho m ta nhận được màu F:

$$F = rR + gG + bB \qquad (2.1.13)$$

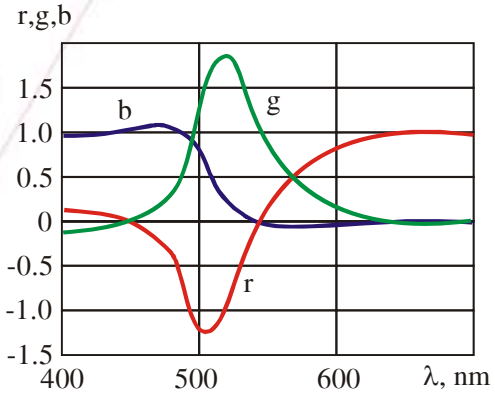
r,g,b - được gọi là tọa độ màu. Các tọa độ màu không cho ta biết về chiều dài của vector màu, nhưng hoàn toàn xác định phương hướng của nó trong không gian màu ba chiều.

Đẳng thức (2.1.13) cho phép chúng ta biểu diễn sắc của một nguồn sáng bất kỳ như một điểm trên hình tam giác đều RGB có chiều cao bằng 1 (Hình 2.1.13). Mặt phẳng RGB còn gọi là mặt phẳng đơn vị. Nếu vị trí của điểm màu F trong tam giác màu được biết trước, ta có thể tìm tọa độ màu bằng cách kẻ các đoạn thẳng vuông góc từ F đến hai cạnh và đo chiều dài của chúng. Nếu cho trước tọa độ màu, người ta tìm vị trí của điểm màu theo luật tìm trọng tâm của tam giác màu. Sắc màu của những điểm nằm ngoài tam giác RGB (như điểm N trên Hình 2.1.13) không thể nhận được khi ta cộng ba màu cơ sở, để nhận được sắc màu điểm N, một trong các tọa độ màu (tọa độ màu đỏ - r) sẽ phải là âm.

Kẻ đường thẳng nối điểm N là màu quang phổ có bước sóng λ_N và điểm trắng đẳng năng E, điểm N_1 là điểm cắt của đường EN và BG sẽ có sắc điệu tương đương điểm N, nhưng độ sạch màu thấp hơn ($p_{N_1} < p_N$). Nói cách khác, bước sóng trội của tất cả các điểm màu nằm trên đường thẳng NE sẽ bằng λ_N - tức bước sóng của màu quang phổ N. Như vậy khi cộng ba màu R, G, B ta có thể tạo ra tất cả các sắc điệu, nhưng không thể tạo ra mọi độ bão hòa.



Hình 2.1.13 Biểu đồ màu RGB



Hình 2.1.14 Đặc tuyến tọa độ màu trong hệ RGB

Màu quang phổ là các màu có độ sạch màu tuyệt đối : $p_\lambda = 100\%$, màu trắng có độ bão hòa $p_E = 0$, độ sạch màu tại điểm N_1 có giá trị:

$$p_{N_1} = \frac{N_1E}{NE} \cdot 100\% \qquad (2.1.14)$$

Sử dụng phương trình so màu (2.1.11) trong các thí nghiệm quang học người ta xác định được quan hệ của r, g, b theo bước sóng λ (Hình 2.1.14). Qua đó, ví dụ, để nhận được nguồn sáng có sắc màu tương đương màu quang phổ với bước sóng $\lambda = 500\ nm$, cần trộn

ba màu R, G, B theo tỷ lệ $r = -1{,}17,\ g = 1{,}39,\ b = 0.78$. Sử dụng tọa độ màu trên hình (2.1.14) ta có thể tìm vị trí của tất cả màu quang phổ trên mặt phẳng màu đơn vị. Các màu này nằm trên đường cong hình móng ngựa RGAB. Hai đầu cuối của đường cong là điểm R và B. Sắc màu nằm trên đường thẳng RB (đỏ mận chín) không phải là màu quang phổ, những màu này thường gặp trong thiên nhiên. Các điểm nằm ngoài đường màu quang phổ là những màu không có thực, vì độ sạch màu của chúng lớn hơn 100%.

Sử dụng tam giác màu, ta có thể xác định được lượng sắc màu (tương đương tọa độ màu) của các màu khác nhau. Đây cũng là cơ sở để đáng giá một cách khách quan độ trung thực của tín hiệu màu, cũng như xác định sai số cho phép đối với các thông số của hệ truyền hình màu.

Mặc dù hệ màu RGB tương đối tiện lợi cho việc thí nghiệm vì các màu R, G, B là các màu thực, nhưng ta có thể nêu ra một số nhược điểm của hệ màu này:

1. Để tạo ra các màu quang phổ cũng như một số màu khác (nằm ngoài tam giác RGB) một trong những toạ độ màu trong phương trình trộn màu (2.1.13) sẽ có giá trị âm. Điều này dễ gây nhầm lẫn khi tính toán.

2. Trong ba tọa độ r', g', b' không có một tọa độ nào cho ta biết trực tiếp độ chói hay quang thông của nguồn sáng.

3. Để tìm được độ chói của màu tổng hợp khi pha trộn nhiều màu, cần phải biết cả ba tọa độ màu r', g', b' của tất các màu đó.

Để khắc phục những nhược điểm của hệ màu RGB, năm 1931, CIE đã đề xuất một không gian màu mới, ký hiệu là không gian XYZ. Khi thiết kế biểu đồ màu mới người ta đặt ra những yêu cầu như sau:

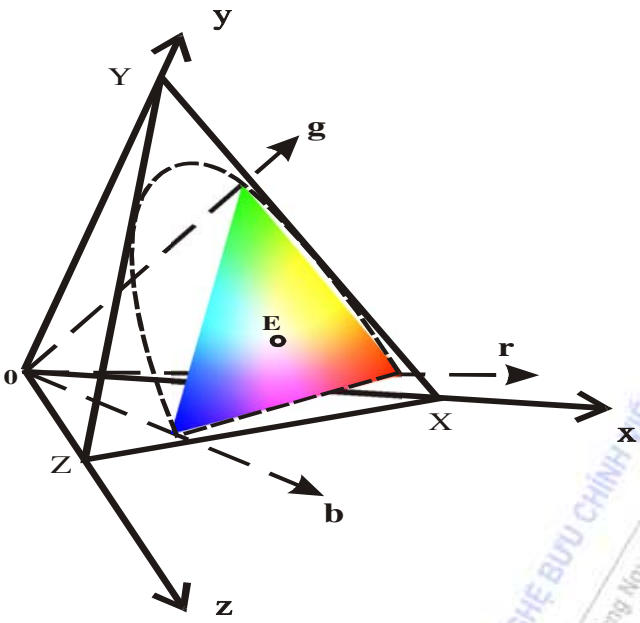
1. Tất cả các màu thực phải được xác định bởi các modul màu có giá trị dương. Như vậy tọa độ màu của tất cả các màu thực (nằm phía trong đường màu quang phổ) phải nằm trong tam giác màu XYZ.

2. Độ chói phải được biểu diễn bằng một trục (Y)

3. Điểm trắng đẳng năng phải nằm ở trọng tâm tam giác màu XYZ.

Không gian XYZ do CIE công bố với ba màu cơ bản X,Y,Z đáp ứng được các yêu cầu trên. Hệ toạ độ không gian XYZ được chọn làm sao cho các vector màu thực (nằm bên trong đường màu quang phổ) đều đi qua tam giác màu đơn vị XYZ (Hình 2.1.15). Như vậy, trong phương trình màu $f'F = x'X + y'Y + z'Z$ các thành phần x', y', z' sẽ có giá trị dương cho tất cả các màu thực.

Các màu đơn vị X,Y,Z đều không có thực vì độ sạch màu của chúng lớn hơn 100%. Để đánh giá màu sắc của một nguồn sáng (không tính đến độ chói của nguồn sáng đó), người ta sử dụng tam giác màu đơn vị với các tọa độ màu $x, y, z:\ x + y + z = 1$ (Hình 2.1.16).



Hình 2.1.15 Không gian màu RGB và XYZ

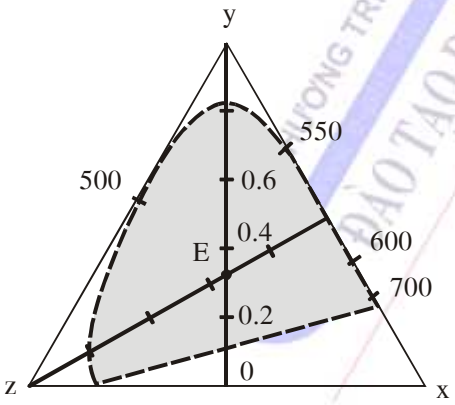
Quan sát tam giác màu đơn vị XZY ta thấy rằng:

1- Tất cả các màu thực đều nằm bên trong tam giác XYZ

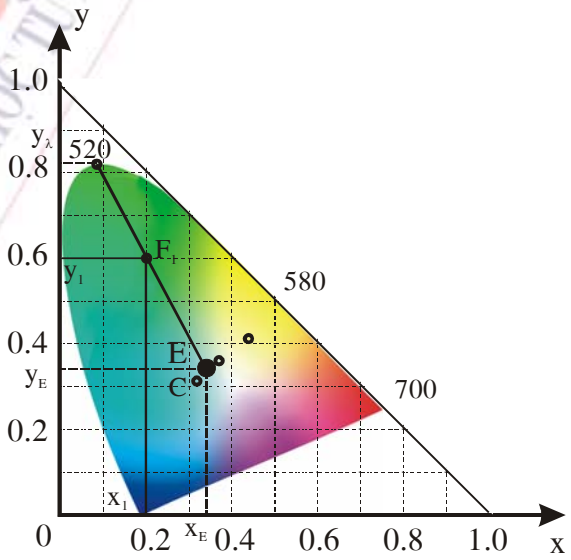
2- Toạ độ của màu trắng đẳng năng là: $x_E = y_E = z_E = 1/3$

3- Các màu đơn vị X.Y.Z nằm ngoài đường cong các màu quang phổ nên có độ bão hòa lớn hơn 100%.

Để thuận tiện cho việc sử dụng, tam giác màu đơn vị XYZ được biến đổi thành biểu đồ màu trong hệ tọa độ vuông góc xy (Hình 2.1.17).



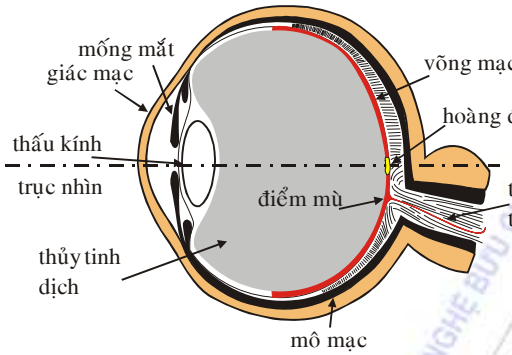
Hình 2.1.16 Biểu đồ màu trên mặt phẳng đơn vị XYZ



Hình 2.1.17 Biểu đồ màu XYZ theo CIE

2.1.6.4 Hệ thống thị giác

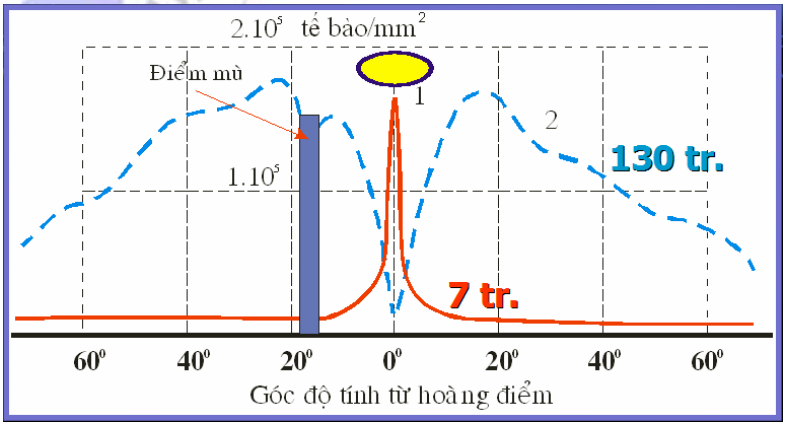
Tuy lý thuyết xử lý ảnh chủ yếu dựa trên nền tảng lý thuyết toán và xác suất thống kê, nhưng việc lựa chọn các phương pháp xử lý khác nhau cũng như việc đánh giá chất lượng hình ảnh ở đầu ra của hệ thống chủ yếu dựa trên cảm nhận chủ quan của cơ quan thị giác. Vì vậy sau đây chúng ta sẽ làm quen với một số vấn đề cơ bản về cấu tạo và khả năng phân tích hình ảnh của hệ thống thị giác.



Hình 2.1.18 Cấu tạo của mắt người

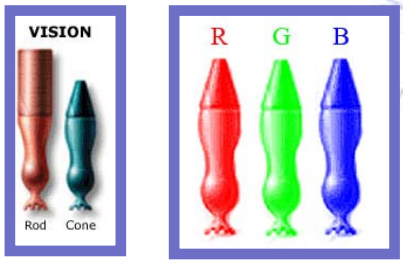
Cấu tạo của mắt người (cắt dọc) được trình bày trên Hình 2.1.18. Tròng mắt có dạng hình cầu, đường kính khoảng 24 mm. Phía trước mắt là giác mạc (cornea) – là màng trong suốt , có cấu trúc tương đối cứng, có đường kính khoảng 12 mm. Hai lớp tiếp theo của mắt là mô mạc (choroid) và võng mạc (retina). Mô mạc bao gồm một mạng mạch máu nhỏ để nuôi dưỡng các cơ quan của mắt. Nối tiếp với mô mạc là hệ thống cơ mắt và mống mắt. Mống mắt có thể co giãn để điều tiết lượng ánh sáng đi vào mắt, độ mở mống mắt thay đổi với đường kính từ 2 đến 8 mm. Hệ thống cơ mắt cho phép thay đổi độ cong của hai bề mặt thấu kính, do đó thay đổi tiêu cự thấu kính nhằm hội tụ hình ảnh lên võng mạc. Thấu kính chứa 60-70% nước, có màu vàng nhạt, nó hấp thụ khoảng 8% ánh sáng nhìn được (độ hấp thụ tỷ lệ nghịch với bước sóng), các tia có bước sóng ngắn (hồng ngoại, cực tím …) được hấp thụ bởi protein trong cấu trúc của thấu kính.

Phần lớn ánh sáng được hội tụ ở vùng hoàng điểm (fovea) trên võng mạc, nơi mật độ các tế bào thần kinh thị giác lớn nhất. Có hai loại tế bào cảm nhận ánh sáng (receptors) là tế bào hình nón (cones) và tế bào hình que (rods). Trong thành phần võng mạc có khoảng 7 triệu tế bào hình nón và 130 triệu tế bào hình que.



Hình 2.1.19 Mật độ phân bố tế bào thần kinh thị giác trên võng mạc

Mật độ phân bố các tế bào thần kinh thị giác trên võng mạc không đồng đều: Đồ thị trên Hình 2.4.19 cho ta thấy mật độ phân bố các tế bào hình nón (đồ thị 1) và hình que (đồ thị 2) trên 1 mm^2 võng mạc (tính từ tâm hoàng điểm). Có thể thấy các tế bào hình nón tập trung tại vùng hoàng điểm, còn gọi là vùng "nhìn rõ nhất". Vùng này có hình bầu dục rộng 0.8 mm, dài 2 mm). Tế bào hình que phân bố xung quanh hoàng điểm. Các tế bào hình que nhạy cảm với ánh sáng hơn tế bào hình nón, nhưng chúng không có cảm thụ về màu sắc. Tế bào hình nón ngược lại kém nhậy cảm với sự kích thích của ánh sáng, nhưng có khả năng phân biệt màu sắc. Theo thuyết ba thành phần cảm thụ màu của mắt người, trong võng mạc tồn tại 3 loại tế bào hình nón, các tế bào này có phản ứng khác nhau đối với các màu khác nhau. Cụ thể, qua thí nghiệm ta thấy rằng ba loại tế bào hình nón nhạy cảm với ba màu khác nhau là Đỏ, Lục, và Lam. Sự cảm thụ màu của mắt sẽ phụ thuộc vào tỷ lệ mức độ kích thích của mỗi loại tế bào nói trên.



Hình 2.1.20 Tế bào hình que và hình nón trong võng mạc

Các tế bào cảm quang biến đổi năng lượng ánh sáng thành các xung điện để truyền đến bộ phận xử lý hình ảnh trong não qua hệ thống dây thần kinh thị giác (khoảng 800000 dây). Một vùng nhỏ trên võng mạc không nhạy cảm với ánh sáng là nơi tập hợp các dây thần kinh thị giác. Vùng này gọi là điểm mù của mắt. Sở dĩ số dây thần kinh ít hơn số tế bào cảm quang vì mỗi dây được nối với hàng trăm tế bào hình que và hàng chục tế bào hình nón. Riêng các tế bào hình nón trong vùng hoàng điểm được nối trực tiếp với tế bào hạch, do đó, độ phân giải của mắt tại vùng này là cao nhất.

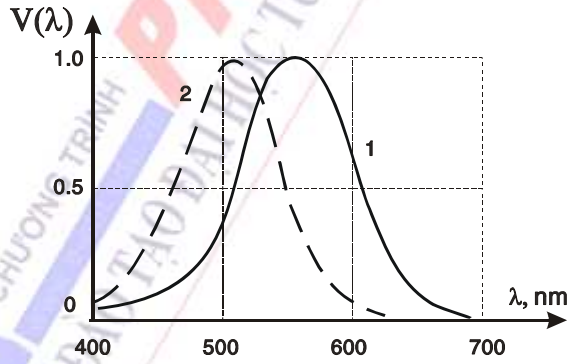
Theo một số công trình nghiên cứu về mắt, người ta thấy rằng các tế bào cảm nhận ánh sáng trong mắt biến đổi năng lượng ánh sáng thành các xung điện, năng lượng nhiệt và năng lượng hoá học. Được biết, năng lượng các xung điện đi vào dây thần kinh thị giác lớn hơn nhiều so với năng lượng áng sáng rọi vào mắt, do đó có thể kết luận là ở võng mạc tín hiệu hình ảnh không chỉ được tiếp nhận mà còn được biến đổi và ánh sáng là tín hiệu điều khiển quá trình biến đổi đó. Quá trình biến đổi quan trọng nhất trong mắt là quá trình biến đổi quang-hóa học. Dưới tác động của ánh sáng, chất cảm quang rodopxin có trong những tế bào que tách ra thành retinen và opxin tự do. Retinen là một dạng vitamin A. Ở nơi ánh sáng càng mạnh thì rodopxin bị phân huỷ càng nhiều. Phản ứng phân huỷ này kích thích các tế bào thị giác làm cho các tế bào thị giác phát ra luồng xung động truyền vào trung ương thần kinh. Khi ánh sáng tắt, retinen lại kết hợp với opxin để tái tạo rodopxin. Phản ứng phục hồi đòi hỏi một thời gian ngắn; sau thời gian đó, tế bào thị giác sẽ hưng phấn trở lại. Trong khoảng thời gian này hình ảnh được lưu lại trên võng mạc (thời gian lưu ảnh kéo dài khoảng 0.1-0.3s). Khả

năng nhìn phụ thuộc vào lượng rodopxin, nên khi mới ở nơi sáng vào chỗ tối, chúng ta hầu như không nhìn thấy gì. Sau vài phút, lượng rodopxin tăng lên, khả năng nhìn có thể tăng lên hàng trăm lần.

Phạm vi các mức sáng mà mắt có thể cảm nhận được rất rộng. Các tế bào que bắt đầu cảm nhận được hình ảnh có độ chói từ 10^{-4}-10^{-5} cd/m^2, các tế bào nón từ 1 cd/m^2. Khi độ chói xấp xỉ 10 cd/m^2 các tế bào que bị "loá" dần vì lúc này, tốc độ phân hủy rodopxin lớn hơn tốc độ tái tạo, do đó lượng rodopxin trong các tế bào hình que giảm đi nhanh chóng. Ở độ chói từ 10-10^4 cd/m^2, chỉ còn các tế bào nón làm việc. Phản ứng quang - hoá học nói trên và cơ chế tự điều chỉnh lượng ánh sáng đi vào võng mạc là nguyên nhân để mắt có phạm vi cảm nhận ánh sáng rộng như vậy (~$10^{9)}$.

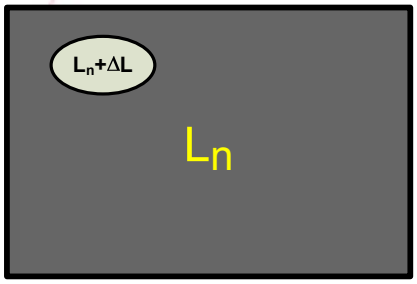
Tuy nhiên mắt không thể cảm nhận được cùng một lúc tất cả mức sáng trong phạm vi rộng như đã nói trên. Trên thực tế, mắt người chỉ có thể cảm nhận một khoảng nhỏ giới hạn từ $L_{\min} \div L_{\max}$ xung quanh mức chói trung bình của ảnh, khoảng này ta gọi là phạm vi động của mắt. Đối với hình ảnh có mức chói trung bình nào đó, tất cả mức chói lớn hơn $L_{\max}$ sẽ cảm nhận như mức trắng, tất cả mức chói nhỏ hơn $L_{\min}$ sẽ được cảm nhận như mức đen. Khi mức chói trung bình thay đổi, mắt người sẽ tự động điều tiết để di chuyển phạm vi động theo mức chói trung bình. Đây chính là tính chất thích nghi với độ sáng của mắt người. Thí nghiệm cho thấy, khi mức sáng tăng lên, thời gian mắt điều tiết để thích nghi với mức mới rất nhanh (khoảng vài giây). Ngược lại, khi mức chiếu sáng giảm thì mắt điều tiết để thích nghi tương đối chậm (vài phút).

Như đã nói ở trên, mắt có độ nhậy khác nhau với các tia bức xạ có bước sóng khác nhau (đồ thị 1, Hình 2.1.21). Nhưng khi cường độ ánh sáng nhỏ (vùng scotopic) đồ thị độ nhậy của mắt di chuyển về phía ánh sáng có bước sóng ngắn hơn (đồ thị 2, Hình 2.1.21).



Hình 2.1.21 Đáp ứng phổ (độ nhạy) của mắt người.

Khả năng mắt người cảm nhận sự thay đổi độ chói là không liên tục.



Hình 2.1.22 Khảo sát khả năng cảm nhận độ chói của mắt người

Nếu tăng dần độ chói của chi tiết trong một ảnh từ mức chói nền (Hình 2.1.22), lúc đầu mặc dù đã có sự khác biệt về độc chói giữa chi tiết và nền, nhưng người quan sát không phát hiện ra chi tiết này. Khi sự chênh lệch đạt tới mức ngưỡng, người quan sát bắt đầu nhận dạng được chi tiết ảnh.

Người ta định nghĩa ngưỡng cảm nhận ánh sáng tuyệt đối của mắt ε là đại lượng ngược với giá trị độ chói nhỏ nhất của điểm sáng trên nền đen mà mắt phát hiện được trong bóng tối: $\varepsilon = 1/L_{\min}$. Trên thực tế ta thường gặp hình ảnh có khoảng chói động là $L_{\min} \div L_{\max}$ và có độ chói của nền là L_n. Độ tương phản của ảnh là tỷ lệ $k = L_{\max}/L_{\min}$.

Các chi tiết ảnh có độ chói khác với độ chói nền $\Delta L = (L - L_n)$, nếu $\Delta L_{\min}$ là mức khác biệt nhỏ nhất mà mắt còn nhận biết được, thì tỷ lệ $\Delta L_{\min}/L_n = (\Delta L/L_n)_{min} = \sigma$ gọi là ngưỡng tương phản. Giá trị σ phụ thuộc vào kích thước của chi tiết hình ảnh và độ chói của nền

Kết luận quan trọng rút ra được ở dây là giá trị ngưỡng tương phản của mắt người $\sigma > 0$, hay nói cách khác, khả năng cảm nhận độ tương phản của mắt mang tính rời rạc (tương tự như độ phân giải của mắt). Chính vì vậy, số lượng các mức xám cần có là hữu hạn trong dải động các mức chói $L_{\min} \div L_{\max}$ của ảnh số.

Số lượng mức xám mà mắt người cảm nhận được cùng một lúc phụ thuộc vào giá trị ngưỡng tương phản và độ tương phản của ảnh:

$$m = \frac{\ln k}{\ln(1+\sigma)} + 1 \qquad (2.1.15)$$

Thay vào công thức (2.1.15) giá trị độ tương phản trung bình của hình ảnh trên display $k = 100$, giá trị ngưỡng tương phản $\sigma = 0.03...0.04$, ta nhận được số sọc xám cực đại để mắt cảm nhận được sẽ là $m = 100 \div 150$.

Trên thực tế, độ tương phản k và số lượng mức xám m bị hạn chế bởi:

- thông số kỹ thuật của màn hình hiển thị: kích thước, độ chói cực đại, đặc tuyến gamma v.v.
- chế độ làm việc của màn hình: độ chói, độ tương phản;
- điều kiện quan sát: khoảng cách từ nơi quan sát đến màn hình, ánh sáng bên ngoài.

Nguồn ánh sáng bên ngoài L_{ng} chiếu vào màn hình sẽ làm giảm độ tương phản của ảnh gốc, vì độ tương phản trong trường hợp này là:

$$k' = \frac{L_{\max} + L_{ng}}{L_{\min} + L_{ng}} < k = \frac{L_{\max}}{L_{\min}},$$

do đó số mức xám tính theo (2.1.15) cũng sẽ giảm đi.

2.1.6.5 Biểu diễn tín hiệu hình ảnh trong không gian và thời gian

2.1.5.6.1 Hình ảnh tương tự

Như đã đề cập tới ở phần trên, hình ảnh có thể biểu diễn bằng hàm 2 chiều $f(x,y)$. Giá trị hàm f tại điểm có tọa độ không gian (x,y) là độ chói của điểm ảnh (x,y). Đa số ảnh sử dụng trong sách này là ảnh đen – trắng, độ chói của các điểm ảnh nằm trong phạm vi nhất định từ $L_{\min}$ tới L_{max}. Nếu ảnh được tạo ra bởi quá trình vật lý thì giá trị các điểm ảnh sẽ tỷ lệ thuận với năng lượng của nguồn bức xạ, ví dụ năng lượng sóng điện từ, khi đó hàm $f(x,y)$ khác không và hữu hạn: $0 < f(x,y) < \infty$.

Hàm $f(x,y)$ có thể được đặc trưng bởi hai thành phần đó là lượng ánh sáng rọi lên cảnh vật và số lượng ánh sáng phản xạ lại từ cảnh vật đó:

$$f(x,y) = i(x,y)r(x,y)$$

với $0 < i(x,y) < \infty$, $0 < r(x,y) < 1$

$i(x,y)$ - Hàm biểu diễn độ rọi sáng của nguồn lên bề mặt cảnh vật.

$r(x,y)$ - Hàm mô tả tính phản xạ (hay hấp thụ) ánh sáng của các vật thể trong cảnh vật.

Giá trị độ lớn của điểm ảnh đen-trắng có tọa độ (x_0, y_0) được gọi là mức xám hay độ chói của ảnh tại điểm này: $l = (x_0, y_0)$; độ chói nằm trong khoảng $L_{\min} < l < L_{max}$ - được gọi là thang xám. Thường mức xám nhỏ nhất được quy về mức 0 (mức đen) , còn mức trắng sẽ tương ứng với giá trị độ chói lớn nhất $l = L - 1$.

2.1.5.6.2 Quá trình lấy mẫu và lượng tử hóa tín hiệu hình ảnh

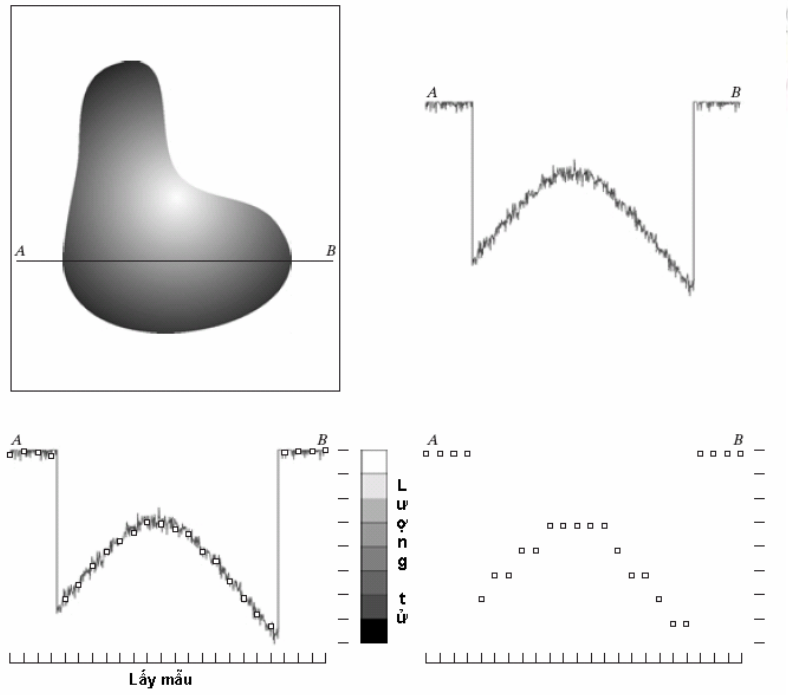
Tín hiệu hình ảnh nhận được từ cảm biến quang điện thường có dạng tương tự, ví dụ tín hiệu điện áp có biên độ thay đổi liên tục theo độ chói của ảnh nguồn. Để có thể đưa tín hiệu hình ảnh vào xử lý bằng máy tính cần thực hiện quá trình số hóa thông qua hai giai đoạn: lấy mẫu và lượng tử hóa.

Lấy mẫu tín hiệu: Quá trình lấy mẫu tín hiệu được mô tả trên Hình 2.1.23. Tín hiệu video ứng với một dòng ảnh AB là tín hiệu một chiều liên tục theo thời gian và có biên độ biến đổi liên tục (Hình 2.1.23). Khi lấy mẫu, thời gian truyền dòng AB được chia ra thành nhiều đoạn bằng nhau. Giá trị tín hiệu tại các điểm lấy mẫu được đánh dấu ô vuông trên đồ thị. Theo định lý lấy mẫu Nyquist, nếu tần số lấy mẫu lớn hơn (hoặc bằng) hai lần tần số lớn nhất trong phổ tín hiệu tương tự, thì tập hợp các mẫu rời rạc nhận được hoàn toàn xác định tín hiệu đó.

Để biến đổi tiếp tín hiệu thành dạng số, chúng ta phải thực hiện giai đoạn lượng tử hóa các mẫu vừa nhận được. Đây là quá trình rời rạc tín hiệu theo biên độ. Trên Hình 2.1.23 thang xám được chia thành 8 mức rời rạc từ mức trắng tới mức đen. Lượng tử hóa được thực

hiện đơn giản bằng cách tìm giá trị mức lượng tử gần giống nhất với giá trị thực của mẫu và gán giá trị này cho mẫu ảnh.

Kết quả nhận được sau khi lấy mẫu và lượng tử hóa là chuỗi số rời rạc mô tả biến đổi độ chói trong một dòng ảnh. Nếu thực hiện quá trình số hóa cho tất cả các dòng ảnh từ trên xuống dưới, chúng ta sẽ nhận được ảnh số trong không gian hai chiều.



Hình 2.1.23 Quá trình số hóa tìn hiệu video

Khi sử dụng chip cảm biến CCD, tín hiệu hình ảnh đã được rời rạc trong không gian hai chiều. Vùng ảnh được lấy mẫu phụ thuộc vào số lượng các điểm cảm quang phân bố theo chiều ngang và chiều dọc trên bề mặt CCD (Hình 2.1.24). Chất lượng hình ảnh số nhận được phụ thuộc vào số lượng điểm ảnh cũng như số mức lượng tử được sử dụng trong quá trình mã hóa.

a) b)

Hình 2.1.24 Quá trình hình thành ảnh rời rạc trong chip CCD

a - Ảnh tương tự

b - Ảnh rời rạc trên bề mặt CCD

2.1.6.6 Tín hiệu video

Thông tin thị giác về một vật thể được truyền đi bao gồm tin tức về độ chói, màu sắc và vị trí của vật đó trong không gian. Khi vật thể đó chuyển động hay khi nguồn ánh sáng chiếu lên vật thể thay đổi, các tin tức trên đều thay đổi. Như vậy, mô hình toán học của tín hiệu hình ảnh là các hàm phân bố độ chói L, sắc màu λ và độ bão hoà màu p trong không gian và thời gian:

$$\left.\begin{aligned} L &= f_L(x,y,z,t);\\ \lambda &= f_\lambda(x,y,z,t);\\ p &= f_p(x,y,z,t). \end{aligned}\right\} \qquad (2.1.16)$$

x, y, z - tọa độ trong không gian 3 chiều,

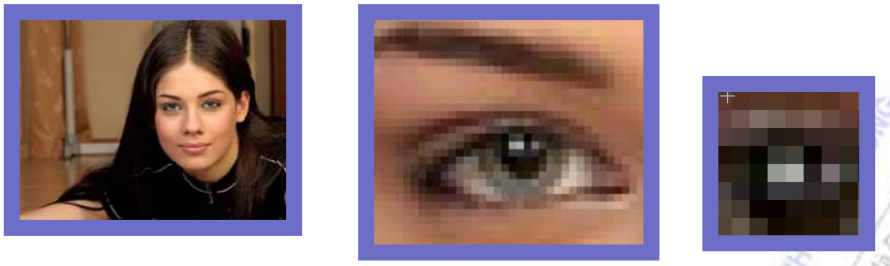
t - thời gian.

Những phương trình trong (2.1.16) xác định độ chói (L) và màu sắc (λ, p) cho từng điểm di chuyển trong không gian và thời gian. Hệ thống truyền hình hiện nay là hệ thống truyền hình phẳng, do đó khi truyền đi các ảnh đen trắng, phân bố độ chói sẽ là hàm ba chiều: $L = f_L(x,y,t)$.

Điều này cho ta thấy, ngoài giá trị độ chói tức thời L cần phải xác định chính xác vị trí của điểm sáng trong không gian (hai chiều) màn hình.

Khi biến đổi tín hiệu hình ảnh 3 chiều thành tín hiệu điện 1 chiều người ta dựa trên 2 nguyên tắc chính là rời rạc hình ảnh (trong không gian và thời gian) và quét hình. Rời rạc hình ảnh trong không gian là phương pháp chia nhỏ hình ảnh ra thành một số hữu hạn các thành phần rời rạc. Một *phần tử* của hình ảnh là chi tiết nhỏ nhất của ảnh có độ chói và sắc màu không thay đổi trên diện tích chi tiết đó. Về mặt lý thuyết, số lượng phần tử ảnh càng nhiều thì độ nét của ảnh càng cao. Nhưng trên thực tế, do sự hạn chế về độ phân giải của mắt người,

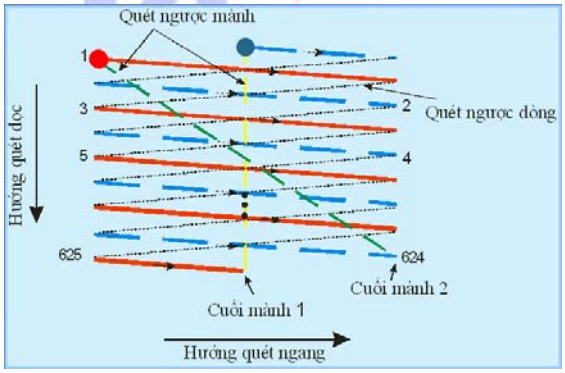
các hình ảnh chỉ cần phân chia ra thành các phần tử có kích thước nhất định đủ để mắt người không nhận ra cấu trúc rời rạc của hình ảnh (Hình 2.1.25). Chia nhỏ thêm những phần tử này không làm cho hình ảnh rõ nét thêm (theo cảm nhận của mắt), trong khi đó, lượng thông tin sẽ tăng lên nhiều lần.



Hình 2.1.25 Ma trận các điểm ảnh rời rạc ảnh và "phần tử" ảnh

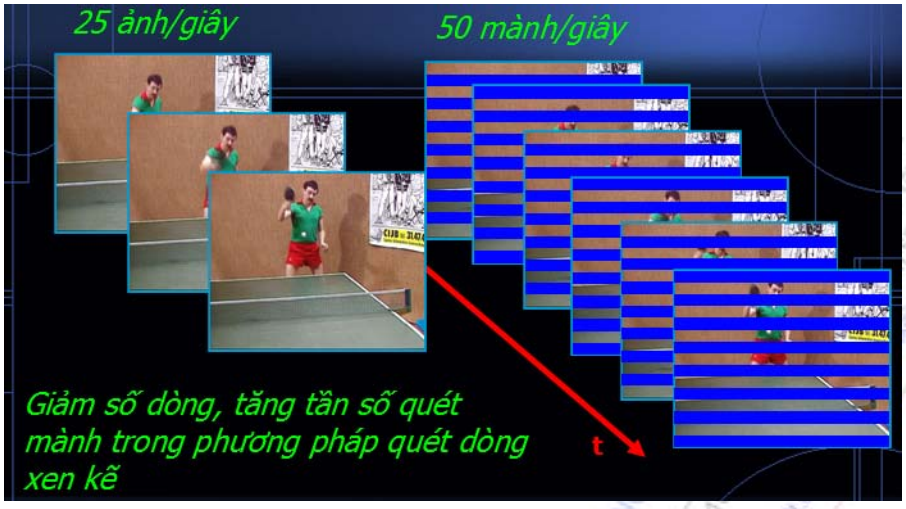
Sau khi hình ảnh được rời rạc, các phần tử có thể được mã hoá và truyền đi riêng rẽ sang bên thu. Nhưng chúng ta không thể truyền song song tất cả các phần tử vì khi đó cần đến rất nhiều kênh truyền. Để giải quyết vấn đề này, trong hệ thống truyền hình người ta sử dụng nguyên tắc quét hình: nguyên tắc truyền lần lượt theo thời gian từng phần tử hình ảnh. Nguyên tắc này dựa trên đặc điểm lưu ảnh của mắt người. Sự lưu ảnh là khả năng mà người xem nhớ lại ấn tượng về ảnh trong một thời gian nào đó (~0.1- 0.3 giây) sau khi tác động của ảnh đó đã chấm dứt. Chính vì vậy, để truyền đi một hình ảnh tĩnh, ta "chiếu" lần lượt tất cả các phần tử của một ảnh tĩnh lên màn hình, vào đúng vị trí tương đương của các phần tử đó như trong hình ảnh đã được truyền đi. Nếu tốc độ "chiếu" một hình nhanh hơn thời gian lưu ảnh thì mắt người xem sẽ thu nhận và lưu lại tất cả các phần tử đã truyền đi để tái tạo ra một ảnh tĩnh hai chiều. Quá trình truyền lần lượt các phần tử của ảnh gọi là quá trình quét (scanning) ảnh.

Tiếp theo, khi "chiếu" nhiều ảnh tĩnh nhận được bằng phương pháp trên với tần số tương đối lớn (trên 10 hình/giây), trong đó mỗi ảnh là một pha của hình ảnh chuyển động, thì người xem sẽ có cảm giác như đang quan sát chuyển động liên tục. Tần số ảnh được lựa chọn để đáp ứng hai yếu cầu: 1- Tạo cảm giác về quá trình chuyển động liên tục của ảnh; 2- Ảnh động tái tạo trên màn hình không bị chớp. Trong các hệ truyền hình đại chúng, tần số được chọn là 25 (hoặc 30) ảnh/giây. Khi quét theo phương pháp xen kẽ, người ta chia ảnh thành 2 mành, trong mành đầu tiên sẽ được truyền đi các dòng lẻ 1, 3, 5 …, trong mành tiếp theo truyền đi các dòng chẵn 2, 4, 6 …(hình 3.1.26). Như vậy toàn bộ ảnh sẽ được chia ra làm 2 mành. Tần số ảnh sẽ là 25 (30) Hz, tần số mành là 50 (60) Hz.



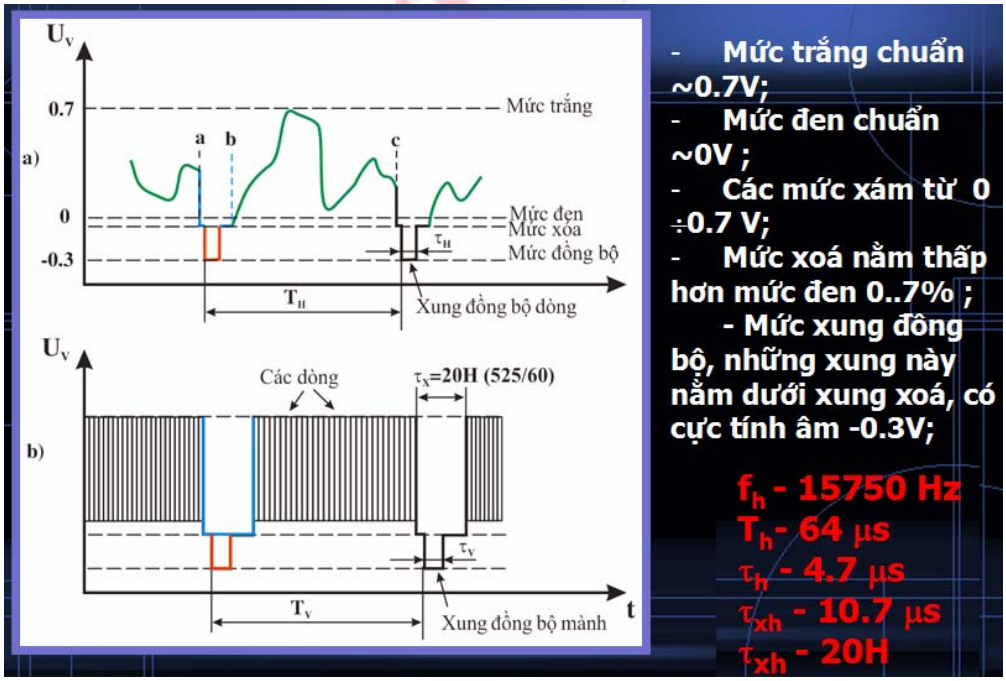
Hình 2.1.26 Quá trình quét hình xen kẽ

Tín hiệu video là tín hiệu được phân tích (rời rạc) cả trong miền tần số và miền thời gian:



Hình 2.1.27 Tín hiệu hình ảnh rời rạc trong không gian (theo dòng) và thời gian (theo mành)

Tín hiệu video được tạo ra tại ống ghi hình bằng phương pháp quét xen kẽ, tuyến tính từ trái sang phải, trên xuống dưới là hàm của thời gian, giá trị hàm tỷ lệ thuận với độ chói của các phần tử ảnh truyền hình. Tín hiệu video đầy đủ (Hình 2.1.28) sẽ bao gồm các thành phần sau: tín hiệu video, tín hiệu đồng bộ dòng và mành, tín hiệu xoá. Trong tín hiệu video màu còn có thêm thành phần mang tin tức về màu sắc của các dòng ảnh.



Hình 2.1.28 Hình dạng tín hiệu video

Tín hiệu video có các đặc điểm sau:

- Tín hiệu video là tín hiệu mang tính chất xung: ngoài các xung đồng bộ và xung xóa, trong tín hiệu video thường có sự thay đổi biên độ đột ngột, tạo ra biên trước và biên sau của các "xung hình";

- Tín hiệu video là tín hiệu đơn cực, có thành phần một chiều;

- Tín hiệu video có thể được coi là tín hiệu tuần hoàn với tần số lặp lại là $f_H = 1/T_H ; f_V = 1/T_V$;

Tín hiệu video tương tự cũng như tín hiệu ảnh tĩnh phải được số hóa trước khi đưa vào hệ thống xử lý số.Cũng như trong các hệ thống xử lý tín hiệu một chiều, quá trình số hóa tín hiệu hình ảnh cũng chia thành 3 giai đoạn:

1- Rời rạc tín hiệu trong miền không gian 2 chiều, đây là quá trình lấy mẫu

2- Số lượng vô hạn các mức xám trong tín hiệu hình ảnh tương tự được thay bằng số lượng hữu hạn các mức lượng tử đây là quá trình lượng tử hóa tín hiệu

3- Mỗi mức lượng tử được biểu diễn bằng một số nhị phân - mã hóa tín hiệu

So với tín hiệu một chiều, quá trình số hóa tín hiệu hình ảnh trong không gian hai chiều có thể được thực hiện với nhiều cấu trúc lấy mẫu khác nhau và các bước lượng tử khác nhau nhằm giảm dung lượng tín hiệu số nhận được. Tuy nhiên, trên thực tế cấu trúc lấy mẫu trong đa số trường hợp có dạng trựuc giao (hình chữ nhật) với giá trị bước lượng tử không thay đổi, vì khi đó quá trình số hóa sẽ đơn giản nhất. Khi sử dụng cấu trúc lấy mẫu trực giao, ảnh số nhận được dưới dạng ma trận các điểm ảnh phân bố theo dòng và cột.

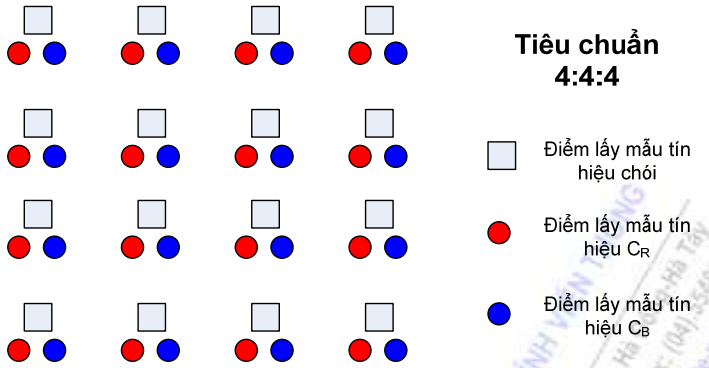
Quá trình lấy mẫu tín hiệu video phải thỏa mãn định lý lấy mẫu Nyquist. Ví dụ: Tín hiệu video hệ PAL có bề rộng phổ $\mathbf{B_{PAL}} = 5.0\,\mathbf{MHz}$ do đó tần số lấy mẫu theo Nyquits phải $\geq 10\text{MHz}$.

Trên thực tế, tần số lấy mẫu thường được lựa chọn cao hơn để tăng khoảng cách giữa dải phổ chính và phổ phụ của tín hiệu video rời rạc, khi đó thành phần phổ chính có thể được tách ra (trong quá trình khôi phục ảnh gốc) bằng các mạch lọc thông thấp đơn giản. Ngoài ra, tín hiệu video tổng hợp (bao gồm thành phần màu) được lấy mẫu với tần số là bội số của tần số sóng mang phụ f_s (sóng mang màu). Với hệ PAL, tần số lấy mẫu sẽ là $3f_s$ (13,3 MHz) hoặc $4f_s$ (17,7 MHz).

Trong hệ thống số hóa tín hiệu video theo thành phần, ba tín hiệu R, G, B hoặc thành phần chói Y và hai tín hiệu hiệu màu R-Y, B-Y sẽ được lấy mẫu với tần số đáp ứng định lý Nyquist và là bội số của tần số dòng theo cả 2 tiêu chuẩn 525 và 625 dòng/ ảnh. Tiêu chuẩn CCIR-601 cho phép sử dụng tần số lấy mẫu là 13,5 MHz. Số bít để mã hóa tín hiệu video là 8 hoặc 10 bít.

Các tiêu chuẩn lấy mẫu video thành phần: có nhiều tiêu chuẩn lấy mẫu theo thành phần, điểmkhác nhau chủ yếu ở tỷ lệ giữa tần số lấy mẫu và phương pháp lấy mẫu tín hiệu chói và tín hiệu màu (hoặc hiệu màu): đó là các tiêu chuẩn 4:4:4, 4:2:2, 4:2:0, 4:1:1.

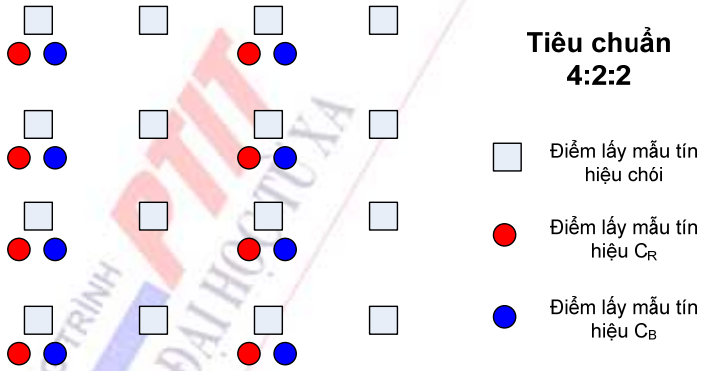
- Tiêu chuẩn 4:4:4: Tín hiệu chói và màu được lấy mẫu tại tất cả các điểm lấy mẫu trên dòng tích cực của tín hiệu video. Cấu trúc lấy mẫu trực giao (hình 3.1.29)



Hình 2.1.29 Cấu trúc lấy mẫu theo chuẩn 4:4:4

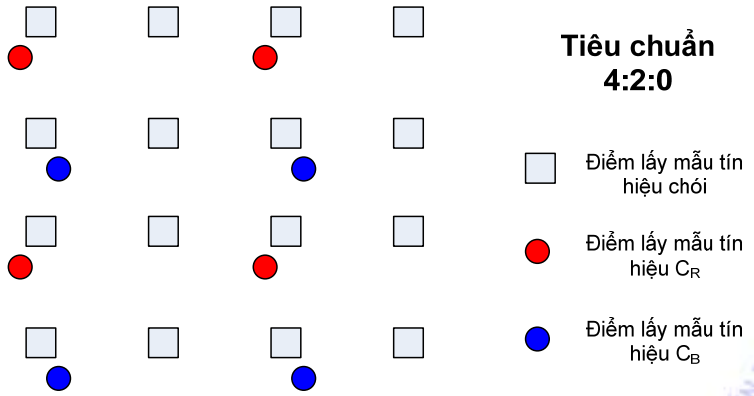
Tiêu chuẩn lấy mẫu 4:4:4 cho chất lượng hhình ảnh tốt nhất, rthuận tiện cho việc xử lý tín hiệu video số. Tuy nhiên, với phương pháp lấy mẫu này, tốc độ dòng dữ liệu video số sẽ tương đối cao, ví dụ khi số hóa tín hiệu video có độ phân giải 720x576 (hệ PAL), 8 bít lượng tử /điểm ảnh, 25 ảnh/s luồng dữ liệu số nhận được sẽ có tốc độ : 3x720x576x8x25= 249Mbits/s

-Tiêu chuẩn 4:2:2: Tín hiệu chói được lấy mẫu tại tất cả các điểm lấy mẫu trên dòng tích cực của tín hiệu video. Tín hiệu màu trên mỗi dòng được lấy mẫu với tần số bằng nửa tần số lấy mẫu tín hiệu chói (Hình 2.1.30)



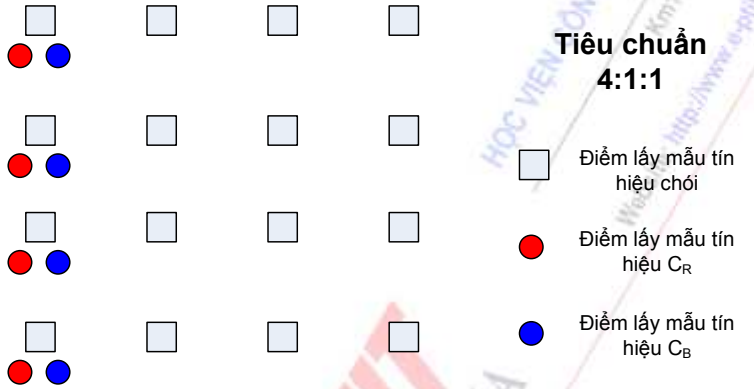
Hình 2.1.30 Cấu trúc lấy mẫu theo chuẩn 4:2:2

-Tiêu chuẩn 4:2:0: Tín hiệu chói được lấy mẫu tại tất cả các điểm lấy mẫu trên dòng tích cực của tín hiệu video. Cách một điểm lấy mẫu một tín hiệu màu. Tại dòng chẵn chỉ lấy mẫu tín hiệu màu C_R, tại dòng chẵn lấy mẫu tín hiệu C_B. Như vậy, nếu tần số lấy mẫu tín hiệu chói là f_D, Thì tần số lấy mẫu tín hiệu màu sẽ là $f_D/2$.



Hình 2.1.31 Cấu trúc lấy mẫu theo chuẩn 4:2:0

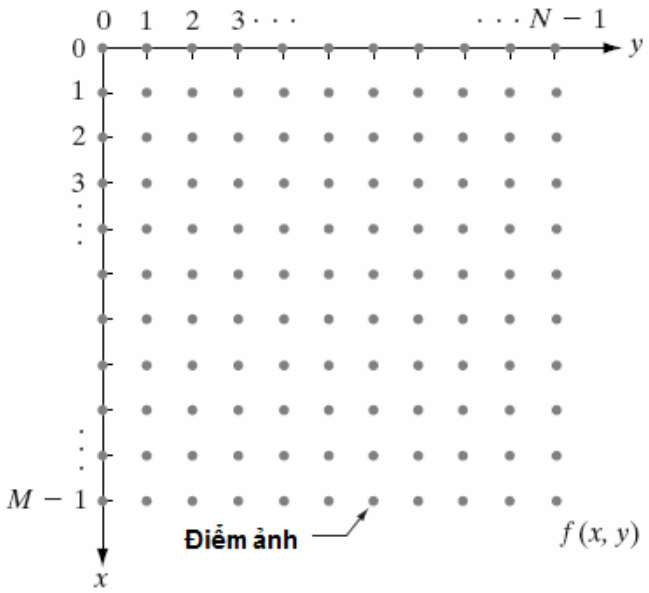
-Tiêu chuẩn 4:1:1: Tín hiệu chói được lấy mẫu tại tất cả các điểm lấy mẫu trên dòng tích cực của tín hiệu video. Tín hiệu màu trên mỗi dòng được lấy mẫu với tần số bằng một phần tư tần số lấy mẫu tín hiệu chói (Hình 2.1.32) Như vậy, nếu tần số lấy mẫu tín hiệu chói là f_D, thì tần số lấy mẫu tín hiệu màu C_R và C_B sẽ là $f_D/4$.



Hình 2.1.32 Cấu trúc lấy mẫu theo chuẩn 4:1:1

2.1.6.7 Biểu diễn tín hiệu ảnh số

Sau khi số hóa tín hiệu hình ảnh theo các phương pháp đã nêu ở trên, chúng ta nhận được ma trận giá trị mức xám của các điểm ảnh. Chúng ta sẽ sử dụng 2 cách biểu diễn tín hiệu ảnh số. Cách thứ nhất, các điểm ảnh rời rạc được sắp xếp theo cột và hàng như trên hình 2.1.33. Tọa độ của các điểm ảnh (x,y) là rời rạc. Gốc tọa độ nằm tại góc trên bên trái của ảnh $(\mathbf{x},\mathbf{y}) = (1,1)$.



Hình 2.1.33 Hệ tọa độ để biểu diễn ảnh số

Như vậy, chúng ta có thể biểu diễn ảnh số nói trên như ma trận kích thước MxN:

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix} \quad (2.1.17)$$

Mỗi phần tử của ma trận được gọi là 1 điểm ảnh (image element hay pixel).

Trong một số trường hợp, chúng ta có thể sử dụng phương pháp mô tả ảnh số như một ma trận thông thường:

$$A=\begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \dots & a_{1,N-1} \\ \vdots & \vdots & \vdots & \\ a_{M-1,0} & a_{M-1,1} & \dots & a_{M-1,N-1} \end{bmatrix} \quad (2.1.18)$$

Với cách biểu diễn trên, $a_{i,j}=f(x=i,y=j)=f(i,j)$, do đó hai ma trận trên hoàn toàn giống nhau.

Đối với ảnh số, giá trị M và N phải là số nguyên dương. Số lượng mức xám có thể gán cho 1 điểm ảnh L thường được lựa chọn sao cho $L=2^k$, k là số nguyên dương.

Như vậy, số lượng bits được sử dụng để biểu diễn 1 ảnh số sẽ được xác định theo công thức: $b=M\times N\times k$.

Ví dụ: ảnh số hiển thị trên màn hình VGA có kích thước 640x480 điểm, số lượng các mức xám là 256 (8 bits/mẫu) có thể được lưu lại trong bộ nhớ có kích thước bằng: $b = 640 \times 480 \times 8 = 2457600$ bits.

2.1.7 Lý thuyết toán ứng dụng trong xử lý ảnh và video số

Tín hiệu hình ảnh tĩnh sau khi được số hóa có thể được lưu trữ dưới dạng ma trận 2 chiều các bít. Các dòng và cột của ma trận sẽ tương ứng với dòng và cột các phần tử ảnh (pixel). Đối với ảnh động (video), kết quả quá trình số hóa sẽ là ma trận 3 chiều cho thấy phân bố các điểm ảnh trong không gian theo hàng và cột cũng như quá trình biến đổi hình ảnh trong miền thời gian.

Quá trình biến đổi tín hiệu trong hệ thống xử lý ảnh số có thể được mô tả bằng các thuật toán trong miền không gian và thời gian hoặc các thuật toán trong không gian tín hiệu khác dựa trên phép biến đổi ánh xạ không gian, ví dụ biến đổi Fourier, biến đổi Karhumen Loeve v.v. Trong phần này chúng ta sẽ làm quen với công cụ toán học thường dùng để mô tả quá trình xử lý ảnh trong không gian và các phép biến đổi không gian một và hai chiều (được sử dụng rộng rãi trong các hệ thống lọc và nén ảnh).

Song song với việc trình bày lý thuyết toán, trong phần này sẽ đưa ra các ví dụ minh họa một số phép biến đổi hình ảnh cụ thể. Nhiều ví dụ sẽ được thực hiện dựa trên phần mềm Matlab. Đây là một công cụ tính toán được xây dựng trên cơ sở các phép xử lý ma trận rất thích hợp cho việc mô tả các giải thuật xử lý ảnh số. Trong tài liệu này, tác giả sử dụng Matlab 7.04 SP2. Dấu ">>" là ký hiệu khởi đầu 1 hàm trong môi trường Matlab.

2.1.7.1 Các toán tử không gian

a) Hệ thống tuyến tính

Hệ thống xử lý tín hiệu số nói chung và xử lý ảnh nói riêng đều có thể được mô tả thông qua phương trình sau:

$$y(m,n) = T\left[x(m,n)\right] \qquad (2.1.19)$$

$x(m,n)$ - ảnh số đưa vào hệ thống (là tín hiệu 2 chiều);

$y(m,n)$ - ảnh số tại đầu ra hệ thống;

T – toán tử đặc trưng của hệ thống.

Trong giáo trình này, chúng ta sẽ quan tâm chủ yếu đến các hệ thống tuyến tính. Hệ thống biểu diễn bởi (2.1) được gọi là tuyến tính khi và chỉ khi:

$$T\left[ax_1(m,n) + bx_2(m,n)\right] = aT\left[x_1(m,n)\right] + bT\left[x_2(m,n)\right] = \\ = ay_1(m,n) + by_2(m,n) \qquad (2.1.20)$$

a và b là các hằng số bất kỳ.

Các toán tử thực hiện với ảnh 2 chiều thường có tính chất tuyến tính, ví dụ các phép dịch chuyển trong không gian, phép chập, các phép biến đổi cũng như nhiều quá trình lọc tuyến tính mà chúng ta sẽ xét ở các chương sau.

b) Xung đơn vị trong không gian 2 chiều

Xung đơn vị được sử dụng rộng rãi để mô tả các tác động trực tiếp lên điểm ảnh trong không gian.

$$\delta(m,n)=\begin{cases}1 & khi\ m=n\\ 0 & khi\ m\neq n\end{cases} \qquad (2.1.21)$$

$\delta(m-A,n-B)$ là điểm ảnh có mức chói tối đa (1) tại vị trí (A,B) trong không gian.

Đáp ứng xung của hệ thống là tín hiệu nhận được khi xung đơn vị được đưa vào hệ thống:

$$h[m,n]=T[\delta(m,n)] \qquad (2.1.22)$$

c) Mô tả quá trình biến đổi tín hiệu trong không gian 2 chiều

Cho ảnh số gốc là ma trận các điểm ảnh có kích thước NxN. Trong trường hợp tổng quát, đáp ứng của hệ thống tuyến tính đối với tín hiệu vào có thể tìm được thông qua đáp ứng xung như sau:

$$y(m,n)=\sum_{l=0}^{N-1}\sum_{k=0}^{N-1}x(l,k)h(m,l;n,k) \qquad (2.1.23)$$

Khi hệ thống xử lý số là tuyến tính và bất biến, ta có thể tìm được ảnh ra thông qua ảnh gốc nói trên và đáp ứng xung của hệ thống sử dụng tích chập:

$$y(m,n)=\sum_{l=0}^{N-1}\sum_{k=0}^{N-1}x(l,k)h(m-l;n-k) \qquad (2.1.24a)$$

hay $$y(m,n)=x(m,n)\otimes h(m,n) \qquad (2.1.24b)$$

2.1.7.2 Các phép tính với vector và ma trận

Đối với tín hiệu hình ảnh, các thuật toán nói trên thường được thực hiện trên ma trận các điểm ảnh hai chiều, do đó phần này sẽ giới thiệu sơ lược về ma trận và các phép toán thực hiện trên ma trận.

a) Vector

Vector cột (ma trận cột) f , kích thước Nx1 là tập hợp các phần tử $f(n)$ với n=1, 2, ..., N sắp xếp theo cột dọc:

$$f = \begin{bmatrix} f(1) \\ f(2) \\ . \\ f(j) \\ . \\ f(N) \end{bmatrix} \qquad (2.1.25)$$

Vector dòng (ma trận dòng) h, kích thước 1xN là tập hợp các phần tử $f(n)$ với n=1, 2, ...,N sắp xếp theo dòng ngang: $h = \left[h(1), h(2)..h(j)..h(N)\right]$ (2.1.26)

b) Ma trận

Ma trận F, kích thước MxN là tập hợp các phần tử F(m,n) với m=1,2,..,M, n=1,2,...,N được sắp xếp thành M hàng và N cột như sau:

$$F = \begin{bmatrix} F(1,1) & F(1,2) & ..... & F(1,N) \\ F(2,1) & F(2,2) & ..... & F(2,N) \\ ..... & & & \\ F(M,1) & F(M,2) & ..... & F(M,N) \end{bmatrix} \qquad (2.1.27)$$

Lưu ý rằng, trong Matlab, địa chỉ của mỗi điểm ảnh được xác định theo vị trí hàng và cột trong ma trận của điểm ảnh đó, ví dụ F(2,1) là điểm ảnh nằm ở hàng thứ 2, cột thứ 1 trong ma trận F. Các biểu diễn này khác với phương pháp biểu diễn ảnh số được xét ở phần .

Ma trận NxN được gọi là ma trận vuông cấp N.

Trong ma trận vuông, tập hợp các phần tử F(1,1), F(2,2),..F(NN) được gọi là đường chéo chính, đường chéo còn lại gọi là đường chéo phụ.

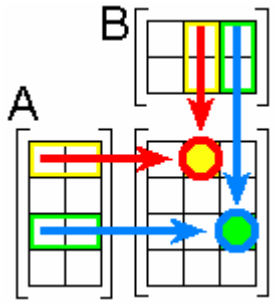
Ma trận vuông có các phần tử ngoài đường chéo chính bằng 0 gọi là ma trận chéo. Ma trận chéo với các phần tử trên đường chéo bằng 1 gọi là ma trận đơn vị, ký hiệu là I_n.

c) Cộng ma trận

Tổng ma trận C=A+B chỉ xác được định khi A và B có cùng kích thước MxN. C cũng có kích thước MxN, các phần tử của C là: C(m,n)= A(m,n)+B(m,n).

d) Nhân ma trận

Tích hai ma trận C=AB chỉ xác định khi số lượng cột của A bằng số dòng của B. Khi nhân ma trận A có kích thước MxP với B - PxN ta nhận được C có kích thước MxN:



$$C(m,n)=\sum_{p=1}^{P}A(m,p)B(p,n) \qquad (2.1.28)$$

Tính của hai ma trận không có tính giao hoán.

Ví dụ 1.

Sử dụng Matlab để tạo ma trận và nhân ma trận

>> A=ones(2,3) A = 1 1 1 1 1 1	>> B= magic(3) B = 8 1 6 3 5 7 4 9 2	>> A*B ans = 15 15 15 15 15 15
>> A= magic(3) A = 8 1 6 3 5 7 4 9 2	>> B=eye(3,3) B = 1 0 0 0 1 0 0 0 1	>> A*B ans = 8 1 6 3 5 7 4 9 2

Tính của ma trận vuông A và ma trận đơn vị cùng cấp B chính là ma trận A.

e) Ma trận nghịch đảo

Ma trận nghịch đảo của ma trận vuông A là ma trận A^{-1} nếu: $AA^{-1}=I$ và $A^{-1}A=I$.

Nếu tồn tại ma trận nghịch đảo của ma trận A cấp n thì A được gọi là khả nghịch.

>> A=[1 2;3 4] A = 1 2 3 4	>> inv(A) ans = -2 1 1.5 -0.5	>> A*inv(A) ans = 1 0 0 1

Ma trận đơn vị I có nghịch đảo là chính nó.

f) Ma trận chuyển vị

Ma trận chuyển vị của A thu được bằng cách đổi chỗ hàng thành cột và cột thành hàng và giữ nguyên thứ tự các phần tử trên hàng. Ma trận chuyển vị của A ký hiệu là A^T.

Nếu $A = A^T$, ma trận A được gọi là ma trận đối xứng. Ma trận nhận được khi cộng $A + A^T$ và nhân AA^T là ma trận đối xứng.

Cho ma trận bất kỳ:	Ma trận chuyển vị A^T	Tích AA^T là ma trận đối xứng:
>> A=[1 2 3;4 5 6]	>> A'	>> A*A'
A =	ans =	ans =
1 2 3	1 4	14 32
4 5 6	2 5	32 77
	3 6	

g) Tích vô hướng (scalar product) hai vector f và g kích thước Nx1:

$$k = g^T f = f^T g \quad ,$$
$$k = \sum_{n=1}^{N} g(n) f(n) \qquad (2.1.29)$$

Ví dụ: $\begin{bmatrix}1\\2\\3\end{bmatrix}.\begin{bmatrix}1\\3\\5\end{bmatrix} = [1\ 2\ 3]\begin{bmatrix}1\\3\\5\end{bmatrix} = [22]$

h) Tích ma trận của hai vector f kích thước Mx1 và g kích thước Nx1 là ma trận:

$$A = gf^T \quad ,$$
$$A(m,n) = g(m) f(n) \qquad (2.1.30)$$

Ví dụ: $\begin{bmatrix}1\\2\\3\end{bmatrix} X \begin{bmatrix}1\\3\\5\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix}[1\ 3\ 4] = \begin{bmatrix}1 & 3 & 4\\2 & 6 & 8\\3 & 9 & 12\end{bmatrix}$

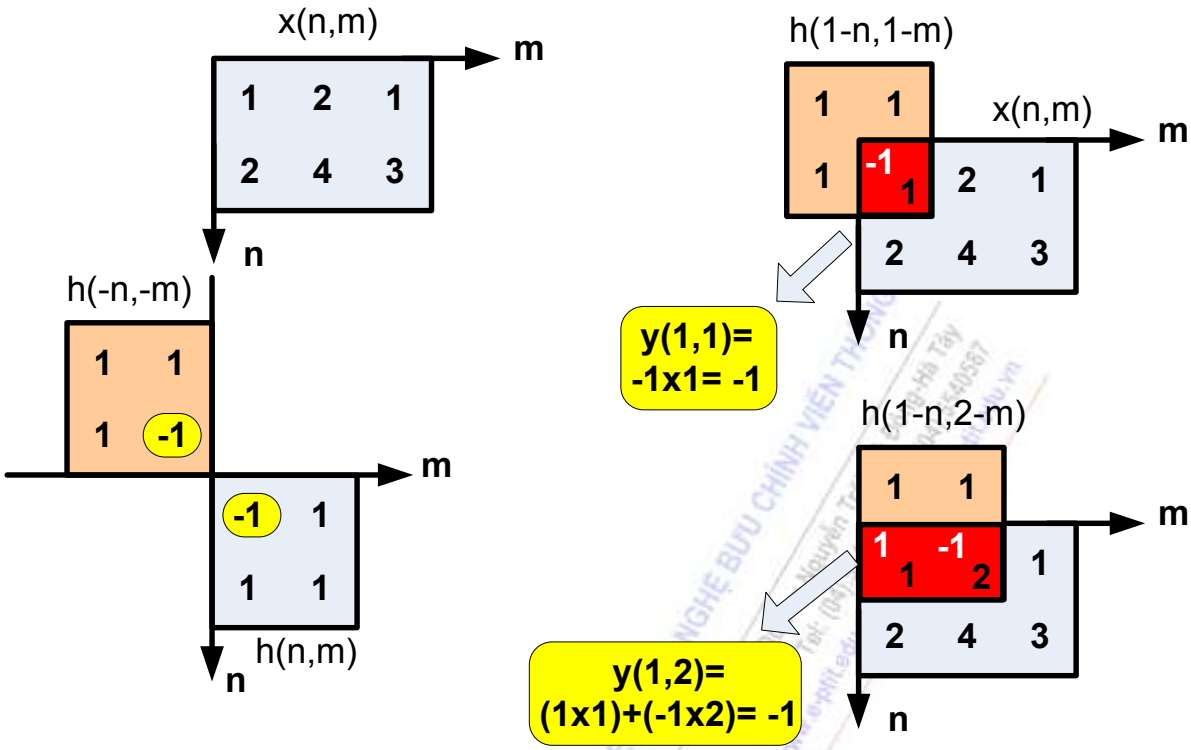
i) Tích chập hai ma trận

Như chúng ta đã biết, đối với các hệ thống xử lý tín hiệu rời rạc tuyến tính và bất biến quan hệ giữa đáp ứng (dãy ra) và kích thích (dãy vào) của hệ thống được mô tả theo (2.1.23):

$$y(m,n) = \sum_{l=0}^{N-1}\sum_{k=0}^{N-1} x(l,k) h(m-l;n-k) \qquad (2.1.31)$$

Đối với hệ thống xử lý ảnh, tín hiệu vào và đáp ứng xung thường được biểu diễn dưới dạng ma trận hai chiều, do đó để mô tả tác động của hệ thống lên tín hiệu ta cần tìm tích chập hai ma trận. Tích chập hai ma trận kích thước $M_1 \times N_1$ và $M_2 \times N_2$ sẽ là ma trận có kích thước $(M_1 + M_2 - 1) \times (N_1 + N_2 - 1)$.

Ví dụ:



Hình 2.1.23 Tích chập hai ma trận

x =	h =	>> y = convn(x,h)
1 2 1	-1 1	y=
2 4 3	1 1	-1 -1 1 1
		-1 1 4 4
		2 6 7 3

Khi tìm giá trị tích chập cho các điểm nằm tại biên của ảnh ví dụ điểm y(1,1) (Hình 2.1.23), các điểm ảnh không tồn tại trong x(m,n) phải được gán các giá trị nhất định. Có nhiều quy tắc chèn giá trị mức xám như: mặc định bằng 0, lặp lại các giá trị mức xám trên đường biên của ảnh v.v. Ta sẽ xét các trường hợp này khi nói về các phương pháp lọc ảnh.

k) Biến đổi ma trận thành một vector (stacking operator)

Trong một số trường hợp, việc phân tích hình ảnh 2 chiều sẽ đơn giản hơn khi ma trận F các điểm ảnh 2 chiều ($N_1 \times N_2$) được biến đổi thành vector cột có kích thước ($N_1 N_2$,1), để làm được như vậy, chúng ta sắp xếp lần lượt các cột (hay hàng) của F thành 1 vector dài.

Thao tác trên có thể được mô tả thông qua vector v_n ($N_2 \times 1$) và ma trận N_n ($N_1 N_2 \times N_1$):

$$v_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} 1 \\ \vdots \\ n-1 \\ n \\ n+1 \\ \vdots \\ N_2 \end{matrix} \tag{2.1.32}$$

$$N_n = \begin{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} & \left\{ (n-1)\ zeros\ \ matrix\ (\mathrm{N}_1 \times N_1) \right. \\ \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} & \left\{ n^{th}\ \ unit\ matrix \right. \\ \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} & \left\{ (N_2 - n)\ zeros\ \ matrix\ (\mathrm{N}_1 \times N_1) \right. \end{bmatrix} \tag{2.1.33}$$

Ma trận F sẽ được biến đổi thành vector f như sau:

$$\mathrm{f} = \sum_{\mathrm{n}=1}^{\mathrm{N}_2} \mathrm{N}_\mathrm{n} \mathrm{F} \mathrm{v}_\mathrm{n} \tag{2.1.34}$$

Biến đổi nghịch từ f thành F là:

$$\mathrm{F} = \sum_{\mathrm{n}=1}^{\mathrm{N}_2} \mathrm{N}_n^T \mathrm{f}\ \mathrm{v}_\mathrm{n}^\mathrm{T} \tag{2.1.35}$$

Sở dụng công thức (2.1.34) và (2.1.35) có thể dễ dàng xác định quan hệ giữa hai phương pháp biểu diễn hình ảnh 2 chiều thông qua ma trận và vector. Phương pháp biểu diễn dưới dạng vector giúp thu gọn đáng kể các công thức mô tả quá trình xử lý ảnh và cho phép chúng ta áp dụng những phương pháp xử lý tín hiệu 1 chiều trong xử lý ảnh.

Ví dụ: Biến đổi ma trận F (3x3) thành vector f

$$F=\begin{bmatrix}1&4&7\\2&5&8\\3&6&9\end{bmatrix}, \mathrm{v}_1=\begin{bmatrix}1\\0\\0\end{bmatrix}; \mathrm{v}_2=\begin{bmatrix}0\\1\\0\end{bmatrix}; \mathrm{v}_3=\begin{bmatrix}0\\0\\1\end{bmatrix},$$

$$N_1=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\end{bmatrix}; N_2=\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\\1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\end{bmatrix}; N_3=\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\1&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

$$\mathrm{f}=\sum_{n=1}^{N_2}\mathrm{N_nFv_n}=\mathrm{N_1Fv_1}+\mathrm{N_2Fv_2}+\mathrm{N_3Fv_3}$$

$$N_1Fv_1=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}1&4&7\\2&5&8\\3&6&9\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix}=\begin{bmatrix}1\\2\\3\\0\\0\\0\\0\\0\\0\end{bmatrix}$$

Tương tự ta có: $\mathrm{N_2Fv_2}=\begin{bmatrix}0\\0\\0\\4\\5\\6\\0\\0\\0\end{bmatrix}$; $\mathrm{N_3Fv_3}=\begin{bmatrix}0\\0\\0\\0\\0\\0\\7\\8\\9\end{bmatrix}$ $\Rightarrow$ $f=\begin{bmatrix}1\\2\\3\\4\\5\\6\\7\\8\\9\end{bmatrix}$

2.1.7.3 Biểu diễn hệ thống tuyến tính bằng ma trận

Việc phân tích các toán tử tuyến tính trong xử lý ảnh sẽ thuận tiện hơn khi chúng ta sử dụng phương pháp biểu diễn bằng ma trận đã đề cập tới ở trên.

Khi tìm đáp ứng hệ thống, biết hàm đáp ứng xung, chúng ta sử dụng (2.1.23):

$$y(m,n)=\sum_{l=0}^{N-1}\sum_{k=0}^{N-1}x(l,k)h(m-l;n-k)$$

Cho tín hiệu vào $x(m,n)$ và đáp ứng xung $h(m,n)$ là các ma trận vuông 3x3, ta có:

$$\begin{aligned}y(m,n)=&x(0,0)h(m-0,n-0)+x(1,0)h(m-1,n-0)+x(2,0)h(m-2,n-0)+\\&x(0,1)h(m-0,n-1)+x(1,1)h(m-1,n-1)+x(2,1)h(m-2,n-1)+\\&x(0,2)h(m-0,n-2)+x(1,2)h(m-1,n-2)+x(2,2)h(m-2,n-2)\end{aligned}$$

(2.1.36)

Có thể thấy rằng, vế phải của (2.1.36) là tích vô hướng của hai vector cột x và $h_{m,n}$:

$$x=\begin{bmatrix}x(0,0)\\x(1,0)\\x(2,0)\\x(0,1)\\x(1,1)\\x(2,1)\\x(0,2)\\x(1,2)\\x(2,2)\end{bmatrix}\quad h_{m,n}=\begin{bmatrix}h(m-0,n-0)\\h(m-1,n-0)\\h(m-2,n-0)\\h(m-0,n-1)\\h(m-1,n-1)\\h(m-2,n-1)\\h(m-0,n-2)\\h(m-1,n-2)\\h(m-2,n-2)\end{bmatrix}$$

Có thể thấy rằng, ma trận các điểm ảnh X ứng với hàm 2 chiều $x(m,n)$ - được biến đổi theo (2.1.34) để nhận được vector x. Nếu ma trận ảnh Y ứng với $y(m,n)$ cũng được biểu diễn tương tự thì vector $h_{m,n}$ sẽ biến đổi thành ma trận H:

$$H=\left[\begin{array}{ccc|ccc|ccc}h(0,0)&h(-1,0)&h(-2,0)&h(0,-1)&h(-1,-1)&h(-2,-1)&h(0,-2)&h(-1,-2)&h(-2,-2)\\h(1,0)&h(0,0)&h(-1,0)&h(1,-1)&h(0,-1)&h(-1,-1)&h(1,-2)&h(0,-2)&h(-1,-2)\\h(2,0)&h(1,0)&h(0,0)&h(2,-1)&h(1,-1)&h(0,-1)&h(2,-2)&h(1,-2)&h(0,-2)\\\hline h(0,1)&h(-1,1)&h(-2,1)&h(0,0)&h(-1,0)&h(-2,0)&h(0,-1)&h(-1,-1)&h(-2,-1)\\h(1,1)&h(0,1)&h(-1,1)&h(1,0)&h(0,0)&h(-1,0)&h(1,-1)&h(0,-1)&h(-1,-1)\\h(2,1)&h(1,1)&h(0,1)&h(2,0)&h(1,0)&h(0,0)&h(2,-1)&h(1,-1)&h(0,-1)\\\hline h(0,2)&h(-1,2)&h(-2,2)&h(0,1)&h(-1,1)&h(-2,1)&h(0,0)&h(-1,0)&h(-2,0)\\h(1,2)&h(0,2)&h(-1,2)&h(1,1)&h(0,1)&h(-1,1)&h(1,0)&h(0,0)&h(-1,0)\\h(2,2)&h(1,2)&h(0,2)&h(2,1)&h(1,1)&h(0,1)&h(2,0)&h(1,0)&h(0,0)\end{array}\right]$$

(2.1.37)

Khi đó, phương trình (2.1.23) được rút ngắn như sau:

$$y = Hx \qquad (2.1.38)$$

Đây là phương trình cơ bản trong lĩnh vực xử lý ảnh tuyến tính. Ma trận H có thể được chia thành 9 ma trận kích thước 3×3 như sau:

$$\mathrm{H}=\begin{bmatrix}\mathrm{H}_{00} & \mathrm{H}_{01} & \mathrm{H}_{02}\\ \mathrm{H}_{10} & \mathrm{H}_{11} & \mathrm{H}_{12}\\ \mathrm{H}_{20} & \mathrm{H}_{21} & \mathrm{H}_{22}\end{bmatrix} \qquad (2.1.39)$$

Khi $x(m,n)$ và $y(m,n)$ có kích thước $N\times N$, Ma trận H có kích thước $\mathrm{N}^2\times\mathrm{N}^2$, các ma trận nhỏ $H_{\mathrm{m,n}}$ sẽ có kích thước $N\times N$. Trong trường hợp tổng quát, H là ma trận circulant khối, được tạo ra bởi $N\times N$ ma trận circulant theo cách sau:

$$H=\begin{bmatrix}\begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=0\\ n=0\end{pmatrix}\end{bmatrix} & \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=1\\ n=0\end{pmatrix}\end{bmatrix} & \cdots & \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=N-1\\ n=0\end{pmatrix}\end{bmatrix}\\ \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=0\\ n=1\end{pmatrix}\end{bmatrix} & \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=1\\ n=1\end{pmatrix}\end{bmatrix} & \cdots & \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=N-1\\ n=1\end{pmatrix}\end{bmatrix}\\ \vdots & \vdots & & \vdots\\ \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=0\\ n=N-1\end{pmatrix}\end{bmatrix} & \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=1\\ n=N-1\end{pmatrix}\end{bmatrix} & \cdots & \begin{bmatrix} l\rightarrow \\ m\downarrow\begin{pmatrix}k=N-1\\ n=N-1\end{pmatrix}\end{bmatrix}\end{bmatrix} \qquad (2.1.40)$$

Theo (2.1.34) vector y tương đương ma trận ảnh Y có thể tìm được thông qua ma trận X như sau:

$$\mathrm{y}=\sum_{\mathrm{n}=1}^{\mathrm{N}}\mathrm{HN}_{\mathrm{n}}\mathrm{Xv}_{\mathrm{n}} \qquad (2.1.41)$$

Ngoài ra, sử dụng (2.1.35) chúng ta có thể biểu diễn ma trận Y thông qua vector y của ảnh:

$$\mathrm{Y}=\sum_{\mathrm{m}=1}^{\mathrm{N}}\mathrm{M}_{\mathrm{m}}^{\mathrm{T}}\mathrm{y}\,\mathrm{u}_{\mathrm{m}}^{\mathrm{T}} \qquad (2.1.42)$$

Từ (2.1.41) và (2.1.42) chúng ta có thể tìm ra quan hệ giữa ma trận ảnh vào và ra của hệ thống tuyến tính:

$$\mathrm{Y}=\sum_{\mathrm{m}=1}^{\mathrm{N}}\sum_{\mathrm{n}=1}^{\mathrm{N}}\left(\mathrm{M}_{\mathrm{m}}^{\mathrm{T}}\mathrm{HN}_{\mathrm{n}}\right)\mathrm{X}\left(\mathrm{v}_{\mathrm{n}}\mathrm{u}_{\mathrm{m}}^{\mathrm{T}}\right) \qquad (2.1.43)$$

Có thể chứng minh được rằng việc nhân H với ma trận $\mathrm{M}_{\mathrm{m}}^{\mathrm{T}}$ và N_{n} sẽ tách ra các ma trận circulant $\mathrm{H}_{\mathrm{m,n}}$, như vậy ta có:

$$Y=\sum_{m=1}^{N}\sum_{n=1}^{N}H_{m,n}X\left(v_n u_m^T\right) \qquad (2.1.44)$$

Đối với hệ thống tuyến tính tách được (separable), quá trình tìm tổng chập (2.1.23) có thể được thực hiện lần lượt bằng cách tính tổng theo m, sau đó theo n. Ta có thể viết:

$$h\left(m,l;n,k\right)=h_c\left(m,l\right)h_r\left(n,k\right) \qquad (2.1.45)$$

$$y\left(m,n\right)=\sum_{l=0}^{N-1}h_c\left(m,l\right)\sum_{k=0}^{N-1}x\left(l,k\right)h_r\left(n,k\right) \qquad (2.1.46)$$

Theo (2.1.45) ta thấy trong các ma trận circulant (2.1.40) thành phần $h_r\left(n,k\right)$ là constant và có thể đưa ra ngoài ma trận. Do đó ma trận circulant khối H có thể biến đổi như sau (để rút gọn, chúng ta viết tắt $h_{n,k}=h\left(n,k\right)$):

$$H=\begin{bmatrix} h_{r0,0}\begin{pmatrix} h_{c0,0}\dots h_{c\,N-1,0}\\ \vdots \quad\quad \vdots\\ h_{c0,N-1}\dots h_{c\,N-1,N-1}\end{pmatrix} & \cdots & h_{rN-1,0}\begin{pmatrix} h_{c0,0}\dots h_{c\,N-1,0}\\ \vdots \quad\quad \vdots\\ h_{c0,N-1}\dots h_{c\,N-1,N-1}\end{pmatrix}\\ h_{r0,1}\begin{pmatrix} h_{c0,0}\dots h_{c\,N-1,0}\\ \vdots \quad\quad \vdots\\ h_{c0,N-1}\dots h_{c\,N-1,N-1}\end{pmatrix} & \cdots & h_{rN-1,1}\begin{pmatrix} h_{c0,0}\dots h_{c\,N-1,0}\\ \vdots \quad\quad \vdots\\ h_{c0,N-1}\dots h_{c\,N-1,N-1}\end{pmatrix}\\ \vdots & & \vdots\\ h_{r0,N-1}\begin{pmatrix} h_{c0,0}\dots h_{c\,N-1,0}\\ \vdots \quad\quad \vdots\\ h_{c0,N-1}\dots h_{c\,N-1,N-1}\end{pmatrix} & \cdots & h_{rN-1,N-1}\begin{pmatrix} h_{c0,0}\dots h_{c\,N-1,0}\\ \vdots \quad\quad \vdots\\ h_{c0,N-1}\dots h_{c\,N-1,N-1}\end{pmatrix}\end{bmatrix}=h_r^T\otimes h_c^T \qquad (2.1.47)$$

Ma trận H được gọi là tích Kronecker của hai ma trận h_r^T và h_c^T.

2.1.7.4 Biến đổi không gian tín hiệu

Ngoài phương pháp phân tích tín hiệu trong không gian thực, trong nhiều trường hợp

2.1.7.4.1 *Biến đổi Fourier liên tục*

Cặp biến đổi Fourie liên tục một chiều được định nghĩa như sau:

$$F\left(u\right)=\int_{-\infty}^{\infty}f\left(x\right)e^{-j2\pi ux}dx \qquad (2.1.48)$$

$$f\left(x\right)=\int_{-\infty}^{\infty}F\left(u\right)e^{j2\pi ux}du \qquad (2.1.49)$$

$f\left(x\right)$ là hàm liên tục, có biến thực x, $j=\sqrt{-1}$.

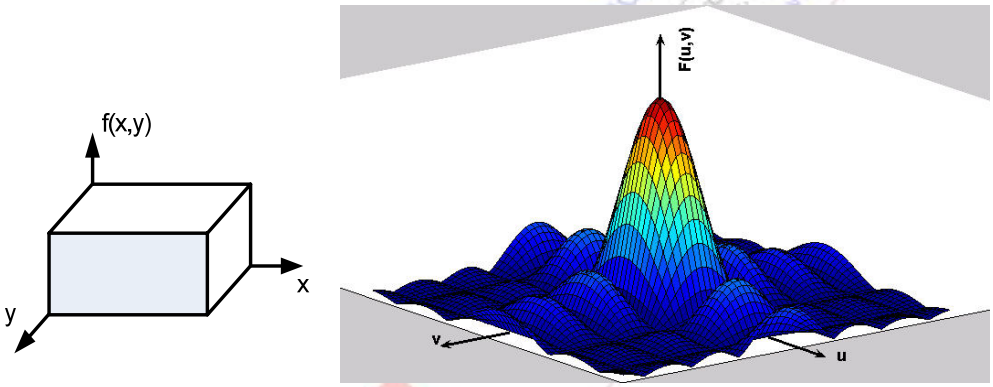
Công thức (2.1.48) được gọi là khai triển thuận Fourier còn (2.1.49) – khai triển nghịch Fourier.

Khai triển Fuorier 2 chiều (2D) được sử dụng tương đối rộng rãi trong các quá trình xử lý ảnh như lọc, xác định biên ảnh, nén ảnh v.v. Cặp biến đổi Fourier liên tục 2 chiều thuận và nghịch sẽ là:

$$F(u,v) = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy \qquad (2.1.50)$$

$$f(x,y) = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} F(u,v)e^{j2\pi(ux+vy)}dudv \qquad (2.1.51)$$

Ví dụ: Biến đổi Fourier của xung hình chữ nhật trong không gian 3 chiều (hình 3.1.24a) có dạng như trên (Hình 2.1.24b).



a) b)

Hình 2.1.24 Phổ Fourier của xung hình chữ nhật trong không gian 3 chiều.

Đối với tín hiệu rời rạc, một biến, cặp khai triển Fourier thuận và nghịch có dạng như sau:

$$F(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(x)e^{-j2\pi\frac{ux}{M}} \quad \text{với } u = 0,1,2...M-1, \qquad (2.1.52)$$

$$f(x) = \sum_{u=0}^{M-1} F(u)e^{j2\pi\frac{ux}{M}} \quad \text{với } x = 0,1,2...M-1 \qquad (2.1.53)$$

Để thực hiện khai triển Fourier, chúng ta phải sử dụng M^2 phép nhân và phép cộng. Cũng như tín hiệu rời rạc $f(x)$, khai triển Fourier của nó cho kết quả là M các thành phần rời rạc. Dễ dàng nhận thấy rằng, mỗi thành phần rời rạc trong F(u) là tổng của tích tất cả các giá trị của hàm $f(x)$ nhân với các hàm cosin và sin có M tần số khác nhau. Như vậy có thể nói F(u) là biểu diễn tín hiệu f(x) trong miền tần số vì biến u xác định các tần số tạo nên tín hiệu rời rạc f(x). Có thể nói, khai triển Fourier cho phép chúng ta mô tả một hàm thông qua các thành phần tần số chứa trong hàm đó. Chính vì vậy khai triển Fourie có thể được sử dụng như 1 công cụ quan trọng để mô tả và phân tích các phương pháp lọc tuyến tính.

Trong trường hợp tổng quát, hàm F(u) là hàm phức, do đó nó có thể được biểu diễn dưới dạng:

$$F(u) = |F(u)| e^{j\Phi(u)} \quad (2.1.54)$$

Modul $|F(u)| = \left[R^2(u) + I^2(u)\right]^{1/2}$ được gọi là phổ biên độ, còn hàm

$\Phi(u) = \text{arctg}\left[\frac{I(u)}{R(u)}\right]$ gọi là phổ pha của hàm f(u). I(u) và R(u) là thành phần thực và ảo của F(u).

Một đại lượng khác có thể suy ra từ phổ Fourier là phổ công suất của tín hiệu P(u):

$$P(u) = |F(u)|^2 = R^2(u) + I^2(u) \quad (2.1.55)$$

Phổ năng lượng cho chúng ta thấy phân bố công suất của tín hiệu trong miền tần số.

Biến đổi Fourier có thể được mở rộng cho hàm $f(x,y)$ có 2 biến. Khi $f(x,y)$ liên tục và lấy tích phân được thì cặp biến đổi Fourier 2 chiều thuận và nghịch sẽ là:

$$F(u,v) = \int_{-\infty}^{\infty}\int f(x,y) e^{-j2\pi(ux+vy)} dxdy \quad (2.1.56)$$

$$f(x,y) = \int_{-\infty}^{\infty}\int F(u,v) e^{j2\pi(ux+vy)} dudv \quad (2.1.57)$$

u, v là biến tần số.

Cũng như trường hợp biến đổi Fourier 1 chiều, ta có phổ biên độ, phổ pha và phổ công suất cho trường hợp 2 chiều:

$$|F(u,v)| = \left[R^2(u,v) + I^2(u,v)\right]^{1/2} \quad (2.1.58)$$

$$\Phi(u,v) = \text{arctg}\left[\frac{I(u,v)}{R(u,v)}\right] \quad (2.1.59)$$

$$P(u,v) = \left|F(u,v)^2\right| = R^2(u,v) + I^2(u,v) \quad (2.1.69)$$

2.1.7.4.2 *Biến đổi Fourier rời rạc 2 chiều*

Biến đổi Fourier thuận 2 chiều của hàm rời rạc f(x,y) (mô tả ảnh số kích thước MxN) được biểu diễn như sau:

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \quad (2.1.70)$$

Nếu có F(u,v) chúng ta có thể tìm ra f(x,y) bằng khai triển Fourier thuận:

$$f(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)e^{j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \quad (2.1.71)$$

Phổ biên độ, phổ pha và phổ công suất cũng được xác định như sau:

$$\left|F(u,v)\right|=\left[R^2(u,v)+I^2(u,v)\right]^{1/2} \quad (2.1.72)$$

$$\Phi(u,v)=\text{arctg}\left[\frac{I(u,v)}{R(u,v)}\right] \quad (2.1.73)$$

$$P(u,v)=\left|F(u,v)^2\right|=R^2(u,v)+I^2(u,v) \quad (2.1.74)$$

Giá trị của phổ Fourier tại điểm u=v=0 bằng:

$$F(0,0)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y) \quad (2.1.75)$$

Nếu f(x,y) hàm biểu diễn độ chói của ảnh số thì F(0,0) chính là giá trị trung bình của độ chói ảnh. Vì (0,0) là điểm gốc tọa độ không gian tần số, nơi tần số bằng 0, nên thành phần F0,0) còn được gọi là thành phần một chiều (DC) của phổ tín hiệu.

Ví dụ: Trên Hình 2.1.25a là ảnh chi tiết hình chữ nhật màu trắng, kích thước 20x40 nằm trên ảnh phông màu đen. Phổ 2 chiều của ảnh trên nhận được bằng khai triển Fourie (3.1.70) biểu diễn trên Hình 2.1.25b. Trong trường hợp này, các thành phần phổ của tín hiệu sẽ được đánh số theo thứ tự từ u=1 tới u=M, v=1 tới v=N.

Để thành phần một chiều của phổ nằm tại trung tâm của ảnh, chúng ta phải thực hiện dịch phổ trong không gian hai chiều, bằng cách nhân hàm f(x,y) với $(-1)^{x+y}$, khi đó theo tính chất của khai triển Fourier, phổ Fuorier phức sẽ dịch chuyển đến vị trí $u=(M/2)$ và $v=(N/2)$:

$$f(x,y)(-1)^{x+y}\leftrightarrow F\left(u-\frac{M}{2};v-\frac{N}{2}\right) \quad (2.1.76)$$

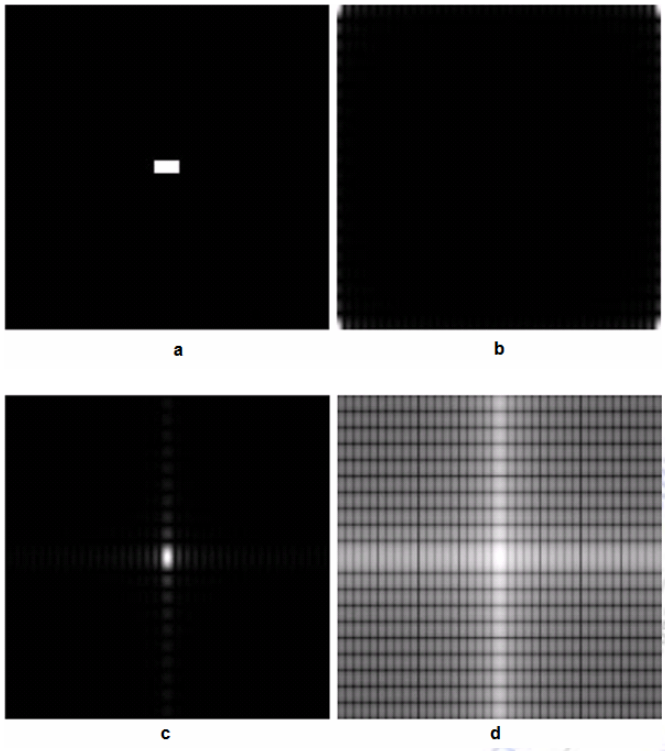
Phổ 2 chiều của ảnh sau khi dịch chuyển được biểu diễn trên Hình 2.1.25c

Phạm vi dải động (khoảng biến thiên) của giá trị các hệ số khai triển Fourier có thể cao hơn nhiều so với giá trị của thành phần chói. Vì vậy, để ảnh phổ hiển thị rõ hơn trên màn hình, đôi khi người ta biến đổi các giá trị phổ theo hàm logarit như sau:

$$D(u,v)=c\log\left[1+\left|F(u,v)\right|\right] \quad (2.1.77)$$

c- hằng số.

Ảnh phổ sau khi biến đổi bằng (3.1.77) biểu diễn trên Hình 2.1.25d.



Hình 2.1.25 Phổ Fourier của hình ảnh 2D

2.2 PHÂN TÍCH CÁC KỸ THUẬT XỬ LÝ ẢNH VÀ VIDEO

2.2.1 Khái niệm về quan hệ giữa các điểm ảnh

2.2.1.1 Các điểm ảnh lân cận

Mỗi điểm ảnh p tại tọa độ (x,y) sẽ có 4 điểm ảnh được gọi là lân cận theo chiều ngang và dọc, đó là các điểm $(x+1,y),(x-1,y),(x,y+1)(x,y-1)$. Tập hợp 4 điểm lân cận trên được ký hiệu là $N_4(p)$. Mỗi điểm lân cận nằm cách điểm (x,y) 1 đơn vị. 4 điểm ảnh lân cận với điểm (x,y) theo đường chéo ký hiệu là $N_D(p)$, đó là các điểm: $(x+1,y+1),(x+1,y-1),(x-1,y+1)(x-1,y-1)$. Tập hợp 8 điểm $N_4(p)$ và $N_D(p)$ được gọi là 8 điểm lân cận: $N_8(p)$. Trong trường hợp khi điểm p nằm ở biên của ảnh, các điểm lân cận có thể nằm bên ngoài ảnh.

2.2.1.2 Mối liên kết (connectivity)

Mối liên kết giữa các điểm ảnh là khái niệm quan trọng, cho phép xác định các giới hạn của chi tiết hay các vùng trong một ảnh. Hai điểm ảnh có sự liên kết với nhau nếu chúng là các điểm lân cận và giá trị mức xám của chúng đáp ứng 1 tiêu chuẩn nào đó (thí dụ nếu chúng giống nhau). Ví dụ, đối với ảnh nhị phân, 2 điểm ảnh có liên kết, khi chúng năm trong bộ 4 lân cận và có giá trị giống nhau.

Cho V là tập các mức xám dùng để định nghĩa mối liên kết, ví dụ trong ảnh đen-trắng, giá trị các mức xám thay đổi từ 0-255, thì V có thể là 1 tập bất kỳ trong số 255 giá trị này. Chhúng ta có 3 loại liên kết:

1) Liên kết 4: hai điểm ảnh p và q có các giá trị từ V có liên kết 4 nếu q nằm trong tập $N_4(p)$.

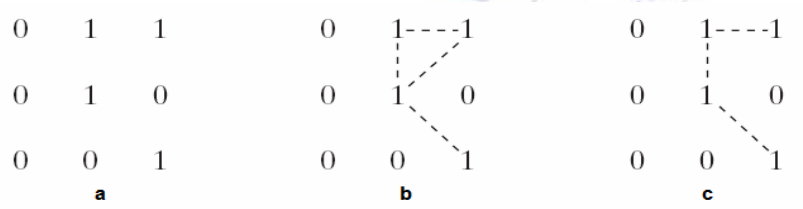
2) Liên kết 8: hai điểm ảnh p và q có các giá trị từ V có liên kết 8 nếu q nằm trong tập $N_8(p)$.

3) Liên kết m (hỗn hợp): hai điểm ảnh p và q có các giá trị từ V có liên kết m nếu:

- q nằm trong tập $N_4(p)$, hoặc

- q nằm trong tập $N_D(p)$ và tập $N_4(p) \cap N_4(q)$ không chứa các giá trị trong V.

Liên kết m là biến thể của liên kết 8 dùng để loại trừ các mối liên kết đa hướng (không rõ ràng) có thể gặp khi dùng liên kết 8. Điều này minh họa trên Hình 2.2.1b. Liên kết 8 trên Hình 2.2.1 được biểu diễn bằng đường đứt nét là liên kết đa hướng, trong khi đó nếu sử dụng khái niệm liên kết m, ta sẽ xác định 1 đường liên kết duy nhất giữa các điểm ảnh có giá trị bằng 1 (Hình 2.2.1c).



Hình 2.2.1 Minh họa liên kết 8 và liên kết m.

Đường kết nối (rời rạc) giữa hai điểm ảnh p có tọa độ (x,y) và q có tọa độ là (s,t) là chuỗi các pixel khác nhau với các tọa độ: $(x_0,y_0),(x_1,y_1),...,(x_n,y_n)$, trong đó:

$$(x_0,y_0)=(x,y)$$
$$(x_n,y_n)=(s,t)$$

(x_i,y_i) và (x_{i-1},y_{i-1}) là các điểm liên kết với $1 \le i \le n$. n được gọi là chiều dài của đường kết nối. Khi $(x_0,y_0)=(x_n,y_n)$, đường nối được gọi là khép kín.

Tùy theo phương pháp xác định liên kết,chúng ta có đường kết nối 4, 8 hoặc đường kết nối m. Trên Hình 2.2.1 mô tả 2 loại đường kết nối m và 8.

Cho S là một tập các điểm ảnh, hai pixel p và q được gọi là liên kết trong S, nếu tồn tại đường kết nối giữa hai điểm p và q được tạo ra chỉ bởi các điểm trong tập S.

Với bất cứ điểm p nào từ S, tất cả các pixel liên kết với p trong S sẽ được gọi là thành phần liên kết của S. Nếu S trong chỉ tồn tại một thành phần liên kết, thì tập S gọi là tập liên kết.

Cho R là tập con các điểm ảnh, R được gọi là 1 vùng ảnh nếu R là tập liên kết. Đường biên của vùng R được tạo ra từ tập nhỏ các điểm ảnh. Các điểm này có 1 hoặc nhiều hơn các điểm lân cận không nằm trong tập R.

2.2.1.3 Toán tử xử lý điểm ảnh

Trong các phần sau, chúng ta sẽ nói đến các phép tính thực hiện với hình ảnh, thí dụ chia một ảnh cho ảnh khác. Như đã giới thiệu ở trên, ta có thể biểu diễn ảnh số như ma trận các điểm ảnh,tuy nhiên trong trường hợp tổng quát, hai ma trận không chia được cho nhau. Vì vậy, trên thực tế, khi thực hiện toán tử chia 2 ảnh cho nhau, người ta chia các pixel tướng ứng của 1 ảnh cho ảnh khác (với điều kiện các điểm ảnh của anh chia khác 0). Ví dụ, khi chia ảnh f cho ảnh g, điểm ảnh thứ nhất nhận được sẽ là giá trị điểm ảnh thứ nhất của ảnh f chia cho giá trị điểm ảnh thứ nhất của ảnh g. Tương tự như vậy, các toán tử số học và logic sẽ được thực hiện cho các pixel tương ứng giữa 2 ảnh.

2.2.1.3.1 Khoảng cách giữa các điểm ảnh

Đối với các pixels p,q,z với các tọa độ (x,y), (s,t), (u,v), D là hàm khoảng cách hoặc metric, nếu:

$$D(p,q) \geq 0, \quad D(p,q) = 0 \text{ nếu p=q}$$

$$D(p,q) = D(q,p)$$

$$D(p,z) \leq D(p,q) + D(q,z)$$

2.2.1.3.2 Khoảng cách Euclide giữa p va q được định nghĩa:

$$D_e(p,q) = \left[(x-s)^2 + (y-t)^2 \right]^{1/2} \qquad (2.2.1)$$

2.2.1.3.3 Khoảng cách D_4 giữa p va q được định nghĩa:

$$D_4(p,q) = |x-s| + |y-t| \qquad (2.2.2)$$

Các pixels nằm cách điểm (x,y) một khoảng D_4 nhỏ hơn hoặc bằng giá trị r sẽ tạo ra hình thoi có tâm điểm tại (x,y).Ví dụ: pixels nằm cách (x,y) một khoảng $D_4 \leq 2$ tạo ra hình thoi sau:

```
      2
    2 1 2
  2 1 0 1 2
    2 1 2
      2
```

2.2.1.3.4 Khoảng cách D_8 giữa p va q được định nghĩa như sau:

$$D_8(p,q) = \max\left(|x-s|, |y-t|\right) \qquad (2.2.3)$$

Ví dụ: pixels nằm cách (x,y) một khoảng $D_8 \leq 2$ tạo ra hình vuông có tâm điểm tại điểm (x,y):

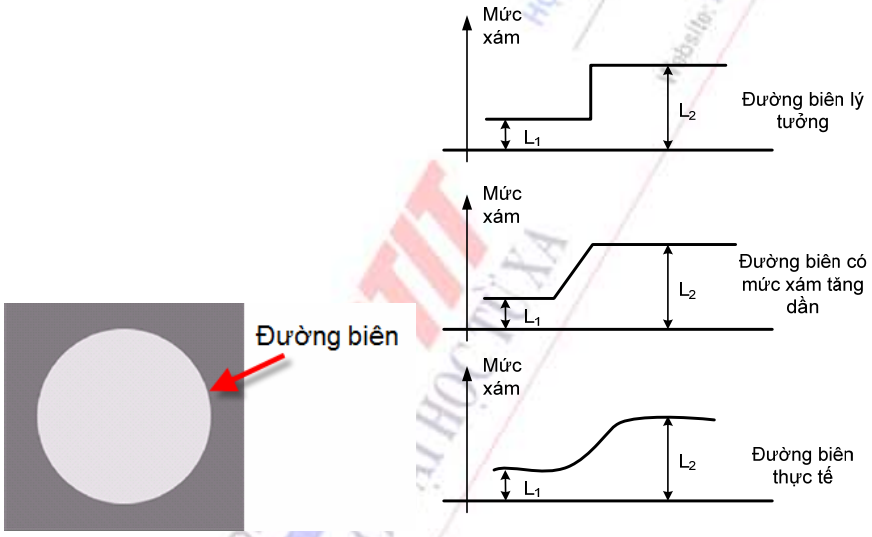
2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

Pixels với khoảng cách $D_8 = 1$ là 8 điểm lân cận của pixel (x,y).

2.2.2 Các phương pháp xác định và dự đoán biên ảnh

2.2.2.1 Cơ sở phát hiện đường biên ảnh

Làm nổi, phát hiện và dự đoán biên ảnh là vấn đề quan trọng trong phân tích ảnh. Như đã được định nghĩa ở phần trước, đường biên của một vùng ảnh R được tạo ra bởi các điểm ảnh có một hoặc nhiều điểm lân cận không nằm trong tập liên kết R. Nói cách khác, một điểm ảnh được coi là nằm trên đường biên nếu tại vị trí điểm ảnh đó có sự thay đổi đột ngột của mức xám. Như vậy, đường biên là đường nối các điểm ảnh nằm trong khu vực ảnh có thay đổi đột ngột về độ chói, đường biên thường ngăn cách hai vùng ảnh có các mức xám gần như không đổi.



Hình 2.2.2 Minh họa khái niệm đường biên của ảnh

Đường biên giữa hai vùng ảnh (có độ chói khác nhau) trong không gian 2 chiều và sự thay đoori độ chói trên đường biên.

Trong trường hợp lý tưởng, độ chói giữa 2 vùng ảnh thay đổi đột ngột hoặc tăng dần đều. Tuy nhiên, trên thực tế, mức xám giữa các vùng ảnh thay đổi tương đối ngẫu nhiên. Chính vì vậy quá trình phát hiện đường biên thường không đơn giản và kết quả thường không hòan toàn chính xác.

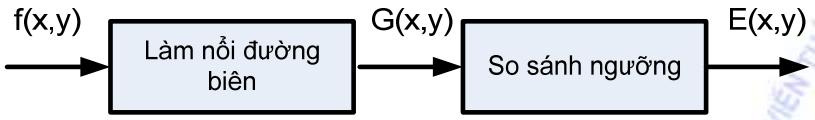
Có nhiều phương pháp phát hiện đường biên khác nhau:

1- Phương pháp phát hiện đường biên trực tiếp dựa trên các phân tích về sự thay đổi độ chói của ảnh. Kỹ thuật chủ yếu dùng để phát hiện biên là dùng đạo hàm. Khi lấy đạo hàm bậc nhất của ảnh ta có phương pháp gradient, khi lấy đạo hàm bậc hai ta có kỹ thuật Laplace.

2- Phương pháp phát hiện đường biên trong ảnh màu: phân tích ảnh màu thành các ảnh đơn sắc (R,G,B) và xác định đường biên trên cơ sở sự thay đổi màu sắc trong các ảnh đơn sắc nói trên.

3- Phân tích ảnh thành vùng theo các đặc điểm đặc trưng (thí dụ kết cấu bề mặt (texture)), ranh giới giữa các vùng chính là đường biên của ảnh.

2.2.2.2 Phương pháp Gradient



Hình 2.2.3 Sơ đồ khối tổng quát của hệ thống phát hiện đường biên

Sơ đồ khối tổng quát của hệ thống phát hiện đường biên biểu diễn trên Hình 2.2.3.

Ảnh gốc f(x,y) được đưa vào khối làm nổi đường biên. Ở đây, bằng phương pháp xử lý tuyến tính hoặc phi tuyến ảnh F(x,y) được làm tăng mức chênh lệch độ chói giữa các vùng ảnh. Ảnh G(x,y) là ảnh gốc đã được tăng cường biên độ đường biên giữa các vùng ảnh. Sau đó, tại khối so sánh, người ta so sánh giá trị các điểm ảnh G(x,y) với mức ngưỡng T để xác định vị trí các điểm có mứuc thay đổi độ chói lớn. Nếu:

$G(x,y) < T_L$ - có sự thay đổi mức chói từ cao xuống thấp

$G(x,y) > T_H$ - có sự thay đổi mức chói từ thấp lên cao.

T_L và T_H là giá trị mức ngưỡng thấp và cao.

Việc lựa chọn giá trị ngưỡng rất quan trọng trong quá trình xác định đường biên. Khi giá trị T quá cao, các đường biên có độ tương phản thấp sẽ bị mất đi, ngược lại, khi T quá thấp, dễ xảy ra hiện tượng xác định biên sai dưới tác động của nhiễu.

Phương pháp gradient là phương pháp dò biên cục bộ dựa vào giá trị cực đại của đạo hàm. Gradient là vector cho thấy tốc độ thay đổi giá trị độ chói của các điểm ảnh theo hướng nhất định. Các thành phần của gradient được tính bởi:

$$\frac{\partial f(x,y)}{\partial x} = \frac{f(x+dx,y) - f(x,y)}{dx} \qquad (2.2.4)$$

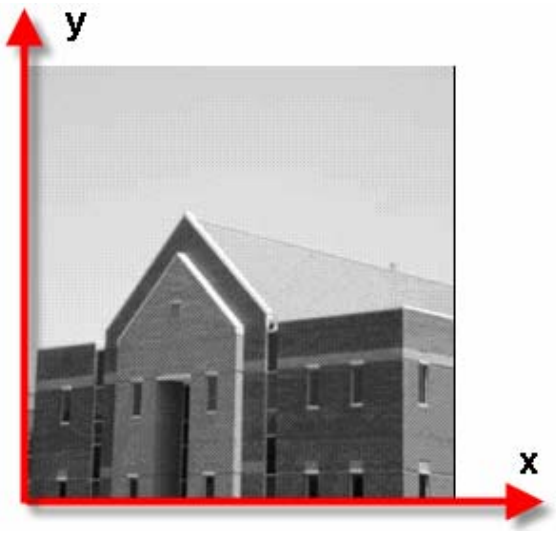
$$\frac{\partial f(x,y)}{\partial y} = \frac{f(x,y+dy) - f(x,y)}{dy} \qquad (2.2.5)$$

dx và dy là khoảng cách giữa các điểm theo hướng lấy x và y. Trên thực tế thường dùng dx=1, dy=1. Khi ảnh số được biểu diễn như ma trận các điểm ảnh phân bố theo dòng và cột, gradient rời rạc theo hướng x sẽ bằng:

$$G_x(x,y) = f(x+1,y) - f(x,y) \qquad (2.2.6)$$

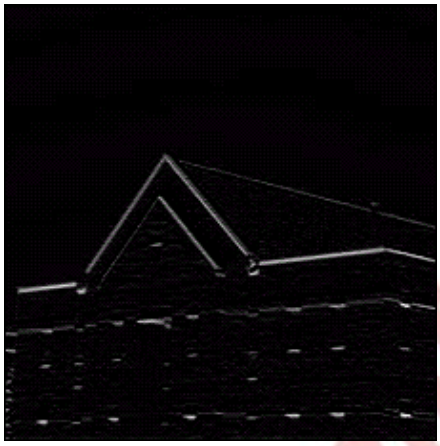
Gradient theo hướng y sẽ là:

$$G_y(x,y) = f(x,y+1) - f(x,y) \qquad (2.2.7)$$



a)

b)



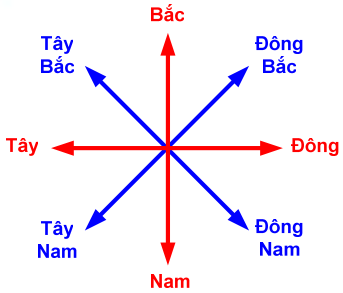
c)

Hình 2.2.4 Làm nổi đường biên sử dụng phương pháp gradient, a- ảnh gốc; b- lấy đạo hàm riêng theo chiều y; c- lấy đạo hàm riêng theo chiều x;

Để làm nổi những đường biên dọc, ta phải lấy đạo hàm rời rạc theo chiều ngang và ngược lại. Kết quả quá trình làm nổi đường biên theo phương pháp gradient thể hiện trên hình 2.2.4.

2.2.2.3 Làm nổi biên bằng toán tử la bàn

Đạo hàm rời rạc hai chiều có thể được thực hiện bằng cách lấy tổng chập giữa ảnh gốc và mặt nạ gradient. Một số mặt nạ được sử dụng để làm nổi các đường biên theo hướng nhất định [ ]. Tên mặt nạ được đặt theo hướng thay đổi độ chói mà mặt nạ có đáp ứng lớn nhất.



Bắc

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

Đông

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

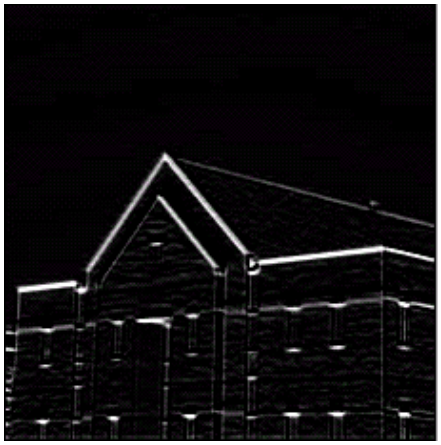
Đông-bắc

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

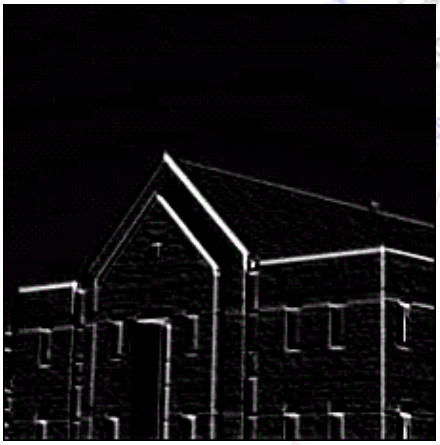
Đông-nam

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

Ảnh đã được làm nổi biên bằng các mặt nạ trên biểu diễn ở các hình sau:



Bắc



Đông Bắc

Đông

Đông Nam

Có thể thấy rằng, các mặt nạ trên có tổng giá trị các hệ số bằng 0, do đó đáp ứng của mặt nạ tại vùng ảnh có độ chói không đổi sẽ bằng 0.

2.2.2.4 Kỹ thuật Laplace

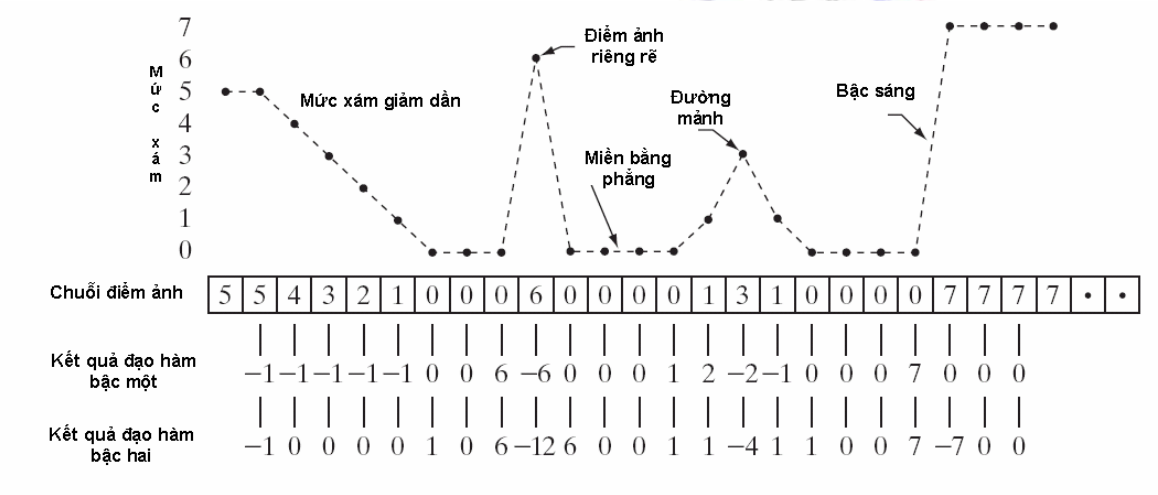
Phương pháp gradient làm việc tốt khi độ sáng thay đổi tương đối rõ nét. Khi mức xám thay đổi chậm, các đường biên không rõ nét, miền chuyển tiếp tương đối rộng, phương pháp hiệu quả hơn là dùng đạo hàm bậc hai mà chúng ta gọi là phương pháp Laplace. Toán tử Laplace được định nghĩa như sau:

$$\nabla^2 f = \frac{\partial^2 f}{dx^2} + \frac{\partial^2 f}{dy^2} \qquad (2.2.8)$$

Việc xấp xỉ đạo hàm bậc hai cho tín hiệu rời rạc (tạm thời xét tín hiệu một chiều) được thực hiện như sau:

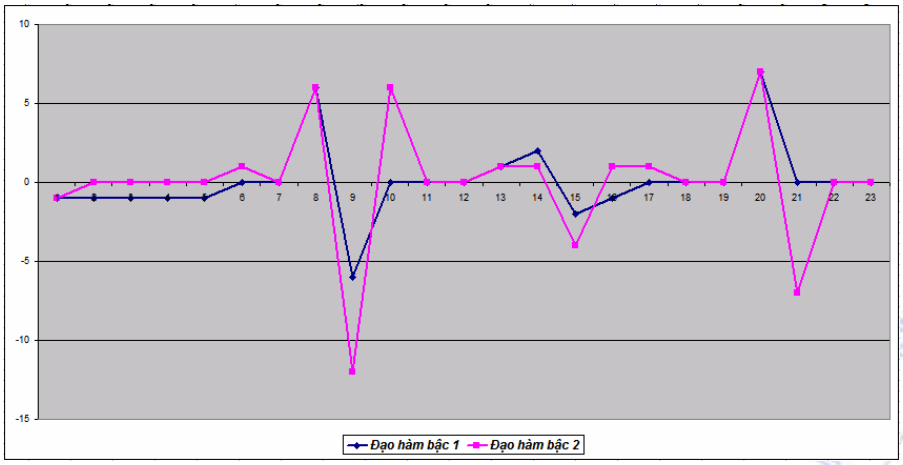
$$\frac{\partial^2 f}{\partial x^2} = f(x+1) + f(x-1) - 2f(x) \qquad (2.2.9)$$

Để thấy sự khác biệt, chúng ta quan sát tác động của đạo hàm bậc một và bậc hai tới tín hiệu:



Hình 2.2.5 Minh họa ứng dụng đạo hàm bậc một và bậc hai trong quá trình tách biên

Nhìn trên đồ thị kết quả đạo hàm bậc một và hai (hình 2.2.5 và 2.2.6) ta thấy đạo hàm bậc một lấy trong miền mức xám giảm dần khác không, trong khi đạo hàm bậc hai chỉ khác không ở đoạn đầu và đoạn cuối. Đường biên giữa các vùng trên ảnh thực thường có sự thay đổi độ chói như trên, do đó có thể thấy đạo hàm bậc hai cho phép xác định vị trí đường biên tốt hơn đạo hàm bậc một. Tại vị trí điểm ảnh riêng rẽ, đạo hàm bậc 2 cho đáp ứng mạnh hơn, do đó các chi tiết nhỏ (và cả nhiễu) sẽ được làm nổi rõ nét hơn. Đường nét tương đối mảnh trên hình ảnh cũng sẽ được làm nổi nhiều hơn khi sử dụng đạo hàm bậc hai. Cuối cùng, đối với tính hiệu dạng bậc sáng độ lớn đáp ứng của cả đạo hàm bậc một và hai gần giống nhau. Tuy nhiên đạo hàm bậc 2 cho kết quả âm và dương. Do đó trên hình ảnh sẽ có hiệu ứng đường biên đôi.



Hình 2.2.6 Dạng tín hiệu nhận được sau khi lấy đạo hàm bậc 1 và bậc 2

Đối với tín hiệu 2 chiều, đạo hàm riêng theo trục x và y bằng:

$$\frac{\partial^2 f}{\partial x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y) \qquad (2.2.10)$$

$$\frac{\partial^2 f}{\partial y^2} = f(x,y+1) + f(x,y-1) - 2f(x,y) \qquad (2.2.11)$$

Tóan tử Laplace hai chiều rời rạc tính theo (2.2.8) có dạng:

$$\nabla^2 f = \left[f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) \right] - 4f(x,y) \qquad (2.2.12)$$

Tóan tử trên có thể được thức hiện bởi các mặt nạ sau:

0	1	0
1	−4	1
0	1	0

a

1	1	1
1	−8	1
1	1	1

b

Hình 2.2.7 Mặt nạ thực hiện toán tử Laplace và ảnh nhận được ở đầu ra

2.3 CÁC KỸ THUẬT NÉN ẢNH

2.3.1 Giới thiệu chung về kỹ thuật nén ảnh

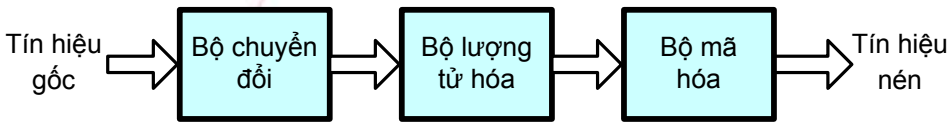
Các hệ truyền hình (tương tự) NTSC, PAL, SECAM sử dụng nén thông tin video bằng cách giảm độ rộng băng tần thành phần màu xuống ≤ 1,2 MHz do độ nhạy của mắt người có giới hạn đối với tín hiệu màu ở tần số cao. Tiêu chuẩn định dạng video thành phần 4:2:2 (CCIR-601) dùng độ rộng băng tần tín hiệu chói và màu là 5,75 MHz và 2,75 MHz (± 0,1 dB). Sau khi số hóa, tốc độ bit tổng cộng của tín hiệu chói và màu là 270 Mbps. Giá trị này rất cao, không thể thực hiện được việc truyền tín hiệu video số qua vệ tinh với độ rộng dải tần một kênh 27 MHz hoặc qua hệ thống truyền hình quảng bá trên mặt đất với tiêu chuẩn 7÷8 MHz cho một kênh truyền hình tiêu chuẩn. Do vậy, nén tín hiệu video là công đoạn không thể thiếu để khắc phục được những khó khăn trên.

Nén video trong những năm 1950 được thực hiện bằng công nghệ tương tự với tỉ số nén thấp. Sự phát triển của kỹ thuật số và việc sử dụng công nghệ số vào kỹ thuật truyền hình làm cho khái niệm " nén video" trở thành đề tài nóng hổi trong những năm gần đây. Từ những năm 1980, các nhà khoa học đã đạt được những thành tựu quan trọng trong việc nén tín hiệu video và audio. Có rất nhiều hãng sản xuất thiết bị nén, nhưng đều dựa trên hai định dạng nén rất phổ biến là JPEG và MPEG.

Trong lĩnh vực truyền thông video, kỹ thuật xử lý tín hiệu chủ yếu tập trung vào mục đích nén. Người ta thường sử dụng 3 phương pháp nén đối với hình ảnh dựa vào các loại độ dư: dư thừa không gian, dư thừa phổ và dư thừa tâm sinh lý nhìn.

Nén về cơ bản là một quá trình trong đó số lượng số liệu (data) biểu diễn lượng thông tin của một ảnh hoặc nhiều ảnh được giảm bớt bằng cách loại bỏ những số liệu dư thừa trong tín hiệu video. Các chuỗi ảnh truyền hình có nhiều phần ảnh giống nhau. Vậy tín hiệu truyền hình có chứa nhiều dữ liệu dư thừa, ta có thể bỏ qua mà không làm mất thông tin hình ảnh. Đó là các phần xóa dòng, xóa mành, vùng ảnh tỉnh hoặc chuyển động rất chậm,vùng ảnh nền giống nhau, mà ở đó các phần tử liên tiếp hoặc khác nhau rất ít. Ngoài ra, để tăng hệ số nén ảnh động, chuyển động trong ảnh truyền hình phải được dự báo, khi đó, ta chỉ cần truyền các thông tin về hướng và mức độ (vector) chuyển động của các vùng ảnh khác nhau. Các phần tử lân cận trong ảnh thường giống nhau, do đó chỉ cần truyền các thông tin biến đổi. Các hệ thống nén sử dụng đặc tính này của tín hiệu video và các đặc trưng của mắt người (là kém nhạy với sai số trong hình ảnh có nhiều chi tiết và các phần tử chuyển động). Quá trình giải nén ảnh là quá trình xấp xỉ để khôi phục ảnh gốc (thường thực hiện ở phía thu).

Một hệ thống nén video tiêu biểu (hay bộ mã hoá nguồn) bao gồm: bộ chuyển đổi, bộ lượng tử hoá, bộ mã hoá (Hình 2.4.1)



Hình 2.3.1: Sơ đồ khối hệ thống nén ảnh tiêu biểu

- Bộ chuyển đổi: thường dùng phép biến đổi Cosin rời rạc để tập trung năng lượng tín hiệu vào một số lượng nhỏ các hệ số khai triển để thực hiện phép nén hiệu quả hơn là dùng tín hiệu nguyên thủy.

- Bộ lượng tử hoá: tạo ra một lượng ký hiệu giới hạn cho ảnh nén với hai kỹ thuật: lượng tử vô hướng (thực hiện lượng tử hoá cho từng phần dữ liệu) và lượng tử vectơ (thực hiện lượng tử hoá một lần một khối dữ liệu). Quá trình này không thuận nghịch.

- Bộ mã hoá: gán một từ mã, một dòng bit nhị phân cho mỗi ký hiệu.

Các hệ thống nén được phân biệt dựa trên sự kết hợp khác nhau giữa 3 bộ xử lý trên và được phân loại như sau:

- Hệ thống nén không mất thông tin (lossless data reduction): thực hiện tối thiểu tốc độ bit mà không làm méo ảnh, hệ thống còn gọi là nén toàn bit hay có tính chất thuận nghịch.

- Hệ thống nén có mất thông tin (loss data reduction): đạt được độ trung thực tốt nhất đối với tốc độ bit cho trước, hệ thống phù hợp áp dụng cho tín hiệu âm thanh và hình ảnh vì có hệ số nén cao.

Trong sơ đồ Hình 2.3.1, tầng chuyển đổi và tầng mã hoá là nơi tín hiệu xử lý không bị tổn thất, tầng lượng tử là có tổn thất. Ngoài ra, dưa trên quan điểm về tổn thất chúng ta có thể phân biệt hai loại mã hoá như sau: mã hoá entropy (mã hoá không tổn thất) và mã hoá nguồn (mã hoá có tổn thất).

2.3.1.1 Các kỹ thuật mã hoá entropy

Kỹ thuật này chỉ quan tâm đến độ đo tin trong dữ liệu mà không quan tâm đến ngữ nghĩa của tin. Sau đây là một số kỹ thuật mã hoá entropy hay dùng trong hệ thống xử lý video:

- Mã hoá chiều dài dải liên tục (RLC - Run Length Coding): các chuỗi điểm ảnh có cùng độ chói (mức màu) sẽ được mã hoá bằng cặp thông tin (độ chói, chiều dài chuỗi).

- Mã hoá bằng các loại bỏ trùng lặp: các chuỗi đặc biệt được thay thế bằng cờ và số đếm lặp.

- Mã hoá dùng mẫu thay thế: đây là dạng mã hoá thống kê mà nó thay thế các mẫu hay lặp lại bằng một mã.

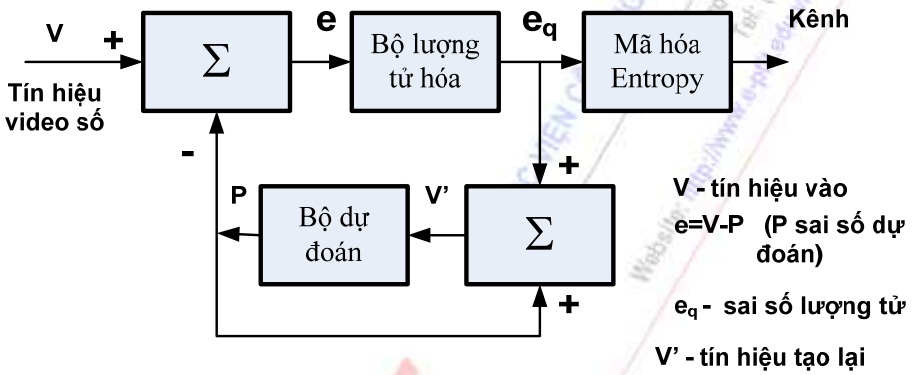
- Mã hóa với độ dài (của từ mã) thay đổi (VLC- Variable-Length Coding)

Phương pháp này còn được gọi là mã hóa Huffman. Nguyên lý của nó dựa trên xác suất xuất hiện của các giá trị biên độ trùng hợp trong một bức ảnh và việc gán một từ mã ngắn cho các giá trị có tần suất xuất hiện cao nhất và từ mã dài hơn cho các giá trị còn lại. Khi thực hiện giải nén, các thiết lập mã trùng hợp sẽ được sử dụng để tạo lại giá trị tín hiệu ban đầu. Mã hóa và giải mã Huffman có thể thực hiện một cách dễ dàng bằng cách sử dụng các bảng tìm kiếm. Như vậy, mã Huffman dựa trên nguyên tắc "ký tự có tần số suất hiện càng cao thì số bit mã hoá càng ngắn".

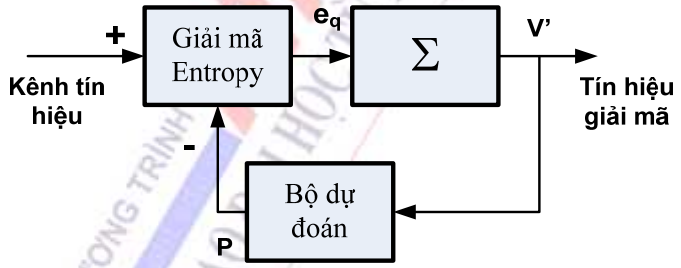
2.3.1.2 Phương pháp mã dự đoán

Phương pháp mã dự đoán hay còn gọi là điều xung mã vi sai (DPCM – Differential Pulse Code Modulation) không mã hóa thông tin có biên độ ở mỗi mẫu, mà mã hóa thông tin có biên độ ở vi sai (biên độ chênh lệch) giữa mẫu đã cho và trị dự báo (được tạo từ các mẫu trước đó). Phân tích thống kê về phân bố biên độ tín hiệu video, ta thấy phân bố biên độ các mẫu tương ứng với các điểm ảnh (pixel) về nguyên tắc là phân bố đều, ngược lại phân bố về độ chênh lệch biên độ các điểm ảnh có đồ thị hình chuông xung quanh điểm 0. Nếu dựa trên các đặc trưng thống kê ảnh, thì sự khác nhau này là không lớn lắm và để mã hóa nó chỉ cần giảm số bit là đủ (so với việc mã hóa toàn bộ biên độ các mẫu).

Phương pháp DPCM còn sử dụng đặc điểm của mắt người (kém nhạy với mức lượng tử có chênh lệch về độ chói giữa điểm ảnh gần nhau, so với mức lượng tử hóa chênh lệch nhỏ) và cho phép dùng đặc trưng phi tuyến về lượng tử hóa. Hình 2.3.2 minh họa sơ đồ khối bộ mã hóa và giải mã DPCM.



a) mã hóa DPCM



b) giải mã DPCM

Hình 2.3.2 Sơ đồ khối bộ mã hóa và giải mã DPCM

Nhằm tránh các lỗi có thể xuất hiện trong khi truyền, một mẫu đầy đủ được gửi đi theo chu kỳ nhất định, cho phép cập nhật được các giá trị chính xác. Mã hóa DPCM cũng sử dụng thêm các kỹ thuật dự đoán và lượng tử hóa thích nghi để hoàn thiện thêm kỹ thuật này.

2.3.1.3 Các kỹ thuật mã hoá nguồn

Các kỹ thuật này dựa trên ngữ nghĩa của ảnh nên còn gọi là kỹ thuật mã hoá hướng ngữ nghĩa. Có 3 loại cơ bản là mã hoá chuyển đổi, mã hoá sai phân và lượng tử hoá vectơ.

- Mã hoá chuyển đổi: dùng phép biến đổi Fourier hay Cosin để chuyển từ miền thời gian hay miền không gian sang miền tần số. Ở đây, chúng ta quan tâm đến phép chuyển

đổi Cosin rời rạc (DCT - Discrete Cosine Transform). Đặc điểm của phép biến đổi này là tín hiệu ảnh trong miền không gian chuyển sang miền tần số thì các thành phần DC và các thành phần AC mang hầu hết các thông tin chứa trong ảnh gốc. Trong đó, DC là thành phần quan trọng nhất mang độ chói trung bình của ảnh, các thành phần AC chứa các thông tin về chi tiết của ảnh. Sau đó, khi qua tầng lượng tử hoá, các hệ số ít quan trọng sẽ bị loại bỏ bớt và chỉ giữ lại một số hệ số hệ số đầu tiên gọi là hệ số DCT.

- Mã hoá sai phân: cũng được gọi là mã hoá ước đoán do chỉ mã hoá sự khác biệt giữa giá trị mẫu thực và giá trị ước đoán, mã hoá sai phân thường dùng cho video hình ảnh động. Lớp kỹ thuật này bao gồm: điều mã xung sai phân, điều chế delta, điều mã xung thích nghi.

- Lượng tử hoá vectơ: mã hoá từng khối hai chiều kích thước cố định (gọi là vectơ) và tra bảng tìm mã phù hợp nhất. Kỹ thuật chỉ thích hợp cho dữ liệu có cấu trúc biết trước.

2.3.1.4 Các phương pháp nén ảnh trong hệ thống video thời gian thực

Việc lựa chọn kỹ thuật nén phụ thuộc vào chất lượng ảnh và giới hạn thời gian trễ. Các tiêu chuẩn về các hệ thống nén ảnh dựa trên tuỳ chọn này để đưa ra các chuẩn phù hợp. Trong hệ thống truyền ảnh động (video), người ta thường sử dụng phương pháp nén theo tiêu chuẩn MPEG (như MPEG-1, MPEG-2, MPEG-4). Trong đó, điểm ảnh là thành phần cơ bản nhất và được nhóm thành từng khối 8×8 điểm ảnh (block), một nhóm 4×4 block này hình thành một khối 16×16 điểm ảnh gọi là Macroblock(MB). Một slice là một dãy các MB liên tiếp giữa hai ký hiệu đánh dấu tái đồng bộ (sync.marker). Các thành phần cấu trúc cao hơn của chuỗi video là khung ảnh (frame), đây chính là các ảnh thực sự của chuỗi video. Có 3 khung ảnh tiêu biểu: khung I, khung P, và khung B. cuối cùng là nhóm các khung (GOP) bắt đầu với khung I và kết thúc với khung P hoặc B.

2.3.1.5 Các tiêu chuẩn nén ảnh

Các tiêu chuẩn quan trọng bao gồm:

- JPEG: dùng cho nén ảnh tĩnh, phát triển bởi sự kết hợp giữa ITU-TS và ISO.

- MPEG-1, MPEG-2, MPEG-4, MPEG-7: do Ủy ban ISO IEC/JTC1/SC29-/WG11 phát triển cho mã hoá kết hợp giữa video và audio.

- H.261: do Nhóm nghiên cứu XI phát triển và được biết rộng rãi như tiêu chuẩn mã hoá video cho các dịch vụ nghe nhìn tốc độ n×64Kbps.

- ITU-TS H.263 cho các ứng dụng điện thoại thấy hình dưới tốc độ dưới 64Kbps.

Mặc dù các tiêu chuẩn được giới thiệu ở trên phù hợp cho từng loại ứng dụng riêng biệt. Tuy nhiên, chúng cũng có chung các nguyên tắc cơ bản. Sự khác biệt giữa các chuẩn phụ thuộc chủ yếu vào yêu cầu đặc biệt của từng ứng dụng. Trong đó, tiêu chuẩn MPEG-4 được xem như một chuẩn tổng quát hoá của chuẩn H.263, vì vậy, việc khảo sát dựa trên chuẩn này sẽ có tính chất áp dụng chung cho cả hai chuẩn.

2.3.1.6 Xác định hiệu quả của quá trình nén tín hiệu số

Hiệu quả nén được xác định bằng tỉ lệ nén, nghĩa là tỉ số giữa số lượng dữ liệu của ảnh gốc trên số liệu của ảnh nén.

Độ phức tạp của thuật toán nén, được xác định bằng số bước tính toán trong cả hai quá trình mã hóa và giải mã. Thông thường thì thuật toán nén càng phức tạp bao nhiêu thì hiệu quả nén càng cao nhưng ngược lại giá thành và thời gian để thực hiện lại tăng. Đối với thuật toán nén có tổn thất thì độ sai lệch được xác định bằng số thông tin bị mất đi khi ta tái tạo lại hình ảnh từ dữ liệu nén. Với nén không tổn thất thì chúng ta có thể có những thuật toán mã hóa càng gần với entropy của thông tin nguồn bởi vì lượng entropy của nguồn chính là tốc độ nhỏ nhất mà bất cứ một thuật toán nén không tổn thất nào cũng có thể đạt được.

Ngược lại, trong nén có tổn thất thì mối quan hệ giữa tỉ lệ nén và độ sai lệch thông tin được Shannon nghiên cứu và biểu dinễ dưới dạng hàm $R_{(D)}$ (hàm về độ sai lệch thông tin). Lý thuyết của ông cũng chỉ ra rằng với thuật toán nén có tổn thất thì chúng ta sẽ có hiệu quả nén cao nhất nhưng ngược lại ta lại bị mất thông tin trong quá trình tái tạo lại nó từ dữ liệu nén. Trong khi đó nén không tổn thất, mặc dù đạt được hiệu quả thấp, nhưng ta lại không bị mất thông tin trong quá trình tái tạo lại nó. Vì vậy ta phải tìm ra một giải pháp nhằm trung hòa giữa hai thuật toán nén này để tìm ra một thuật toán nén tối ưu sao cho hiệu quả cao mà lại không bị mất mát thông tin.

2.3.1.7 Độ dư thừa số liệu

Nén số liệu là quá trình giảm lượng số liệu cần thiết để biểu diễn cùng một lượng thông tin cho trước. Cần phải phân biệt giữa số liệu và thông tin. Thực tế số liệu và thông tin không đồng nghĩa với nhau. Số liệu (và do đó tín hiệu) chỉ là phương tiện dùng để truyền tải thông tin. Cùng một lượng thông tin cho trước có thể biểu diễn bằng các lượng số liệu khác nhau. Ví dụ, nếu hai người khác nhau dùng số từ khác nhau để kể cùng một câu truyện, sẽ có hai version khác nhau của câu truyện và một có chứa số liệu không chủ yếu; nó bao gồm số liệu hoặc từ không cho thông tin thích hợp lẫn xác định đã biết. Đó là do nó đã chứa độ dư thừa số liệu.

Độ dư thừa số liệu là vấn đề trung tâm trong nén ảnh số. Đánh giá cho quá trình thực hiện giải thuật nén là tỉ lệ nén (C_N) được xác định như sau: Nếu N_1 và N_2 là lượng số liệu trong hai tập hợp số liệu cùng được dùng để biểu diễn lượng thông tin cho trước thì độ dư thừa số liệu tương đối R_D của tập số liệu thứ nhất so với tập số liệu thứ hai có thể được định nghĩa như sau:

$R_D = 1- 1/C_N$

trong đó:

$C_N = N_1/N_2$

Trong trường hợp N1 = N2 thì $C_N = 1$ và có nghĩa là so với tập số liệu thứ hai thì tập số liệu thứ nhất không chứa số liệu dư thừa. Khi N2 << N1 thì CN tiến tới vô cùng và R_D tiến tới một, có nghĩa là độ dư thừa số liệu tương đối của tập số liệu thứ nhất là khá lớn hay tập số liệu thứ hai đã được nén khá nhỏ.

Ở đây có sự kết hợp giữa tỉ lệ nén và chất lượng hình ảnh. Tỉ lệ nén càng cao sẽ làm giảm chất lượng hình ảnh và ngược lại. Chất lượng và quá trình nén có thể thay đổi tùy theo đặc điểm của hình ảnh nguồn và nội dung ảnh. Đánh giá chất lượng ảnh được đề nghị tính số

bit cho một điểm trong ảnh nén (Nb). Nó được xác định là tổng số bit ở ảnh nén chia cho tổng số điểm:

Nb = Số bit nén/Số điểm

Trong nén ảnh số, ba loại dư thừa số liệu có thể được nhận dạng và phân biệt.

- Dư thừa mã (Coding Redundancy)

Nếu các mức của tín hiệu video được mã hóa bằng càc symbol nhiều hơn cần thiết (tuyệt đối) thì kết quả là có độ dư thừa mã. Để giảm độ dư thừa mã, trong nén ảnh thường sử dụng các mã VLC như mã Huffman, mã RLC v.v...Lượng thông tin về hình ảnh có xác suất thấp hơn.

- Dư thừa trong pixel (Interpixel Redundancy)

Vì giá trị của bất kỳ một pixel nào đó, cũng có thể được dự báo từ giá trị của các lân cận của nó, nên thông tin từ các pixels riêng là tương đối nhỏ. Sự tham gia của một pixel riêng vào một ảnh là dư thừa. Nhiều tên (bao gồm: dư thừa không gian, dư thừa hình học, dư thừa trong ảnh) được đặt ra để phân biệt sự phụ thuộc này của các pixels. Ta dùng độ dư thừa trong pixel để chỉ tất cả các tên trên. Để giảm độ dư thừa trong pixel của một ảnh, dãy pixel hai chiều dùng cho việc nhìn và nội suy, phải được biến đổi thành một dạng có hiệu quả hơn. Trong các phương pháp nén ảnh sẽ trình bày trong chương này, ta dùng phép biến đổi cosin rời rạc (DCT) biến đổi pixel từ miền không gian sang miền tần số, bằng cách này sẽ giảm được độ dư thừa số liệu trong pixel ở miền tần số cao.

- Dư thừa tâm sinh lý

Bằng trực quan ta thấy, sự thu nhận cường độ sáng thay đổi chỉ giới trong một phạm vi nhất định. Hiện tượng này xuất phát từ sự thật là mắt không đáp ứng với cùng độ nhậy của tất cả các thông tin nhìn thấy. Thông tin đơn giản có tầm quan trọng ít hơn thông tin khác trong vùng nhìn thấy. Thông tin này được gọi là độ dư thừa tâm lý nhìn. Nó có thể được loại bỏ mà không ảnh hưởng đáng kể đến chất lượng thu nhận ảnh. Khác với độ dư thừa mã và dư thừa trong pixel, độ dư thừa tâm sinh lý có liên quan đến thông tin theo định lượng, nó có quan hệ tới việc lượng tử hóa. Điều đó có nghĩa là ánh xạ một khoảng rộng các giá trị đầu vào lên một số hữu hạn các giá trị đầu ra. Đó là toán tử không đảo ngược (mất thông tin) cho kết quả nén số liệu có tổn hao.

2.3.1.8 Sai lệch bình phương trung bình

Phương pháp đánh giá chất lượng ảnh nén thông dụng nhất là dựa trên mức sai lệch bình phương trung bình so với ảnh gốc - RMS (Root Mean Square) được tính bởi biểu thức:

$$RMS = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(X_i - X_i^{'}\right)^2}$$

trong đó: RMS- sai lệch bình phương trung bình

X_i - giá trị điểm ảnh gốc

$X_i^{'}$ - giá trị điểm ảnh sau khi nén

n - tổng số điểm ảnh trong một ảnh

Thông thường, khi giá trị RMS thấp, chất lượng ảnh nén sẽ tốt. Tuy nhiên, trong một số trường hợp chất lượng hình ảnh nén không nhất thiết phải tỷ lệ thuận với giá trị RMS.

2.3.2 Phương pháp nén ảnh JPEG

JPEG (Joint Photographic Expert Group) là tên của một tổ chức nghiên cứu về các chuẩn nén ảnh (trước đây là ISO) được thành lập vào năm 1982. Năm 1986, JPEG chính thức được thiết lập nhờ sự kết hợp giữa nhóm ISO/IEC và ITV. Tiêu chuẩn này có thể được ứng dụng trong nhiều lĩnh vực : lưu trữ ảnh, Fax màu, truyền ảnh báo chí, ảnh cho y học, camera số v.v...

Tiêu chuẩn JPEG được định ra cho nén ảnh tĩnh đơn sắc và màu. Tuy nhiên cũng được sử dụng cho nhiều ứng dụng với ảnh động bởi vì nó cho chất lượng ảnh khôi phục khá tốt và ít tính toán hơn so với nén MPEG. Nén JPEG có thể thực hiện bởi bốn mode mã hóa đó là:

a) Mã tuần tự (sequential DCT-based) : ảnh được mã hóa theo kiểu quét từ trái qua phải, từ trên xuống dưới dựa trên khối DCT.

b) Mã hóa lũy tiến (progressive DCT-based) : ảnh được mã hóa bằng kiểu quét phức hợp theo chế độ phân giải không gian cho các ứng dụng trên kiểu băng hẹp và do đó thời gian truyền dẫn có dài.

c) Mã hóa không tổn thất (lossless) : ảnh được đảm bảo khôi phục chính xác cho mỗi giá trị mẫu của nguồn. Thông tin không cần thiết sẽ mới cắt bỏ cho nên hiệu quả nén thấp hơn so với phương pháp có tổn thất.

d) Mã hóa phân cấp (hierarchical) : ảnh được mã hóa ở chế độ phân giải không gian phức hợp, để cho những ảnh có độ phân giải thấp có thể được truy xuất và hiển thị mà không cần giải nén như những ảnh có độ phân giải trong không gian cao hơn.

Mã hóa không tổn thất không sử dụng cho video động bởi vì nó cung cấp một tỉ lệ nén không đủ cao. Tỉ lệ nén ảnh tĩnh có thể đạt từ 1/10 đến 1/50 mà không làm ảnh hưởng đến chất lượng hiển thị của ảnh. Khai triển DCT được chọn là kỹ thuật then chốt trong JPEG vì nó cho ảnh nén chất lượng tốt nhất tại tốc độ bit thấp và giải thuật chuyển đổi nhanh và dễ dàng thực hiện bằng phần cứng. Trên Hình 2.4.3 là sơ đồ mã hoá tiêu biểu của JPEG.

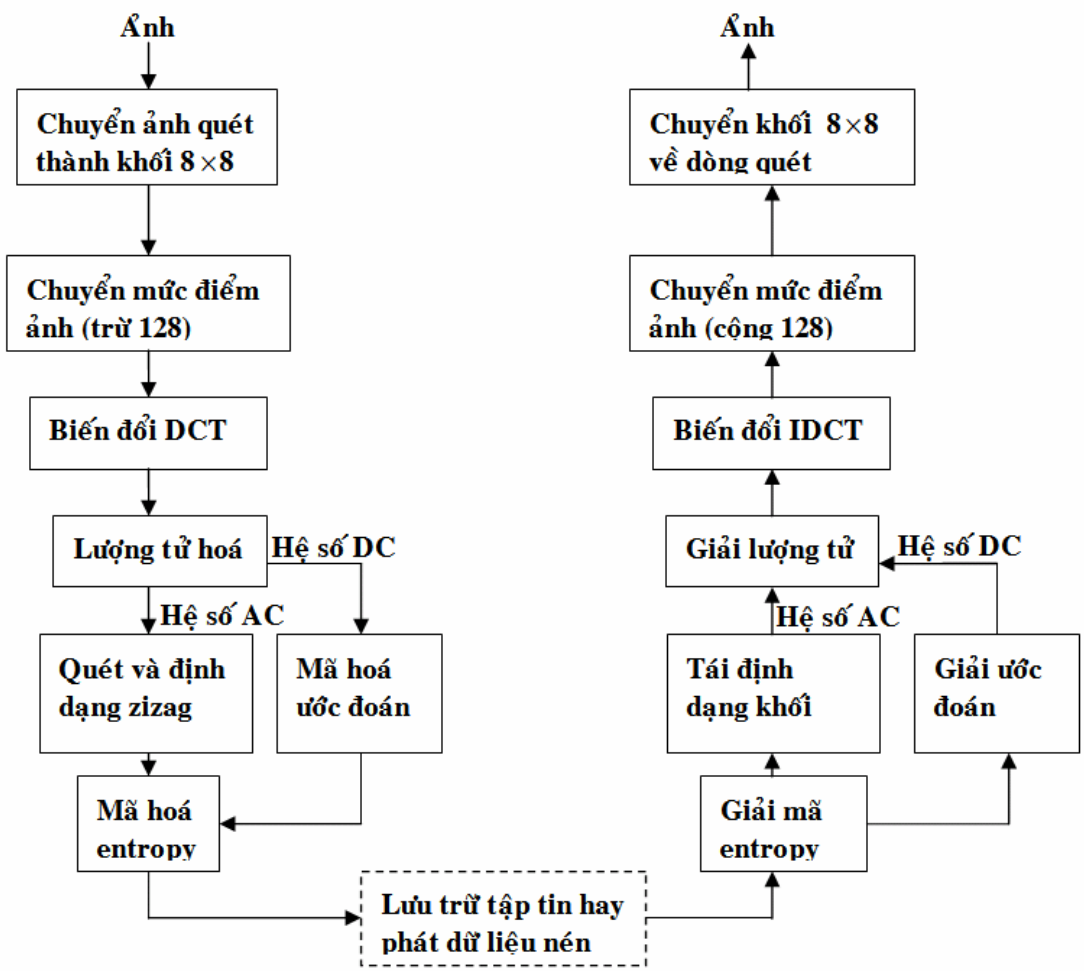
Kỹ thuật nén ảnh JPEG cho phép sử dụng hoặc mành (field) hoặc ảnh (frame) như một ảnh gốc. Nếu kỹ thuật nén dùng mành thì nén trong ảnh sẽ tạo ra hai ảnh trong mỗi ảnh truyền hình. Vì vậy, khi bàn về nén, thuật ngữ "ảnh" không luôn luôn đồng nghĩa với thuật ngữ ảnh trong lĩnh vực truyền hình.

Trước khi đưa vào chuyển đổi DCT, ảnh gốc phải được xử lý để nén dải tần tín hiệu hiệu màu và chia ảnh thành các block. Việc nén phổ tín hiệu hiệu màu làm giảm độ dư thừa tâm sinh lý. Kỹ thuật này dựa vào đặc trưng hệ thống thị giác của con người (HVS: human visual system). Mắt người kém nhậy với sự thay đổi tín hiệu màu hơn sự thay đổi tín hiệu

chói. Vì vậy, ta không cần thiết truyền đi thông tin của tín hiệu màu với tần số như truyền thông tin tín hiệu chói.

Như đã giới thiệu ở trên, theo khuyến cáo CCIR 601-2, có rất nhiều phương pháp lấy mẫu thông tin tín hiệu hiệu màu. Tỷ lệ lấy mẫu thông dụng là 4:2:2 và 4:1:1. Định dạng 4:2:2 nghĩa là cứ 4 mẫu tín hiệu chói thì có 2 mẫu cho mỗi loại tín hiệu hiệu màu. Nói cách khác, cứ 2 mẫu tín hiệu chói có 1 mẫu tín hiệu hiệu màu. Định dạng 4:1:1 nghĩa là cứ 4 mẫu tín hiệu chói thì có 1 mẫu cho mỗi loại tín hiệu hiệu màu. Giả sử tín hiệu hiệu màu chỉ được lấy mẫu theo chiều dọc và mỗi mẫu có 8 bit, số bit trung bình trên một pixel theo tỷ lệ lấy mẫu 4:2:2 là 8×4/2, hay 16 bit/pixel. Theo tỷ lệ 4:1:1 là 8×6/4, 12bit/pixel. Kỹ thuật lấy mẫu tín hiệu hiệu màu được áp dụng cả hai chiều ngang và dọc. Dĩ nhiên, điều này làm giảm hơn nữa lượng thông tin về tín hiệu hiệu màu.

Trước khi thực hiện biến đổi DCT, cả ảnh được chia thành các khối lớn riêng biệt không chồng nhau (MB-Marco Block). Mỗi MB bao gồm 4 block các tín hiệu chói (Y) và 2,4 hoặc 8 block các mẫu tín hiệu hiệu màu (Cr,Cb). Số các block của tín hiệu hiệu màu phụ thuộc vào tiêu chuẩn lấy mẫu của tín hiệu video: 4:2:2, 4:1:1 hay 4:2:0 v.v.



Hình 2.3.3 Sơ đồ mã hóa và giải mã theo JPEG

Tất cả các block có cùng kích thước và mỗi block là một ma trận điểm ảnh 8×8 pixel được lấy từ một ảnh màn hình theo chiều từ trái sang phải, từ trên xuống dưới. Kích thước block là 8×8 được chọn bởi hai lý do sau:

a) Thứ nhất, qua việc nghiên cứu cho thấy hàm tương quan suy giảm rất nhanh khi khoảng cách giữa các pixel vượt quá 8.

b) Thứ hai, là sự tiện lợi cho việc tính toán và thiết kế phần cứng. Nói chung, độ phức tạp về tính toán sẽ tăng nếu kích thước block tăng.

Ví dụ về việc chia thành các block của hình ảnh đối với hệ PAL. Phân tích cực của tín hiệu video với độ phân giải 576×720 sẽ được chia làm 72×90 block tín hiệu chói. Và như vậy sẽ có 36×45 MB.

Cấu trúc của MB cũng phụ thuộc vào loại ảnh quét. Nếu quét liên tục thì các block bao gồm các mẫu từ các dòng liên tục (lúc này nén ảnh theo-frame). Ngược lại, trong trường hợp quét xen kẽ, trong một block chỉ có các mẫu của một nửa ảnh (nén ảnh theo-mành). Tóm lại, việc chia hình ảnh thành các ảnh con (block, MB) sẽ thực sự có ý nghĩa cho bước chuyển vị tiếp theo.

2.3.2.1 Biến đổi cosin rời rạc DCT

Công đoạn đầu tiên của hầu hết các quá trình nén là xác định thông tin dư thừa trong miền không gian của một mành hoặc một ảnh của tín hiệu video. Nén không gian được thực hiện bởi phép biến đổi cosin rời rạc DCT (Discrete Cosine Transform). DCT biến đổi dữ liệu dưới dạng biên độ thành dữ liệu dưới dạng tần số. Mục đích của quá trình biến đổi là thay đổi dữ liệu biểu diễn thông tin: dữ liệu của ảnh con tập trung vào một phần nhỏ các hệ số hàm truyền. Việc mã hóa và truyền chỉ thực hiện đối với các hệ số năng lượng này, và có thể cho kết quả tốt khi tạo lại tín hiệu video có chất lượng cao. DCT đã trở thành tiêu chuẩn quốc tế cho các hệ thống mã chuyển vị bởi nó có đặc tính gói năng lượng tốt, cho kết quả là số thực và có các thuật toán nhanh để thể hiện chúng.

Các phép tính DCT được thực hiện trong phạm vi các khối 8×8 mẫu tín hiệu chói Y và các khối tương ứng của tín hiệu hiệu màu. Việc chia hình ảnh thành các block đã được thực hiện ở khối tiền xử lý. Hiệu quả của việc chia này rất dễ thấy. Nếu ta tính toán DCT trên toàn bộ frame thì ta xem như toàn bộ frame có độ dư thừa như nhau. Đối với một hình ảnh thông thường, một vài vùng có một số lượng lớn các chi tiết và các vùng khác có ít chi tiết . Nhờ đặc tính thay đổi của các ảnh khác nhau và các phần khác nhau của cùng một ảnh, ta có thể cải thiện một cách đáng kể việc mã hóa nếu biết tận dụng nó.

a) DCT một chiều

DCT một chiều biến đổi biên độ tín hiệu tại các điểm rời rạc theo thời gian hoặc không gian thành chuỗi các hệ số rời rạc, mỗi hệ số biểu diễn biên độ của một thành phần tần số nhất định có trong tín hiệu gốc. Hệ số đầu tiên biểu diễn mức DC trung bình của tín hiệu. Từ trái sang phải, các hệ số thể hiện các thành phần tần số không gian cao hơn của tín hiệu và được gọi là các hệ số AC. Thông thường, nhiều hệ số AC có giá trị sẽ gần hoặc bằng 0.

Quá trình biến đổi DCT thuận (FDCT) dùng trong tiêu chuẩn JPEG được định nghĩa như sau:

$$X(k)=\sqrt{\frac{2}{N}}C(k)\sum_{m=0}^{N-1}x(m)\cos\frac{(2m+1)k\pi}{2N} \qquad (2.3.1)$$

Hàm biến đổi DCT ngược (một chiều):

$$x(m)=\sqrt{\frac{2}{N}}\sum_{k=0}^{N-1}X(k)C(k)\cos\frac{(2m+1)k\pi}{2N} \qquad (2.3.2)$$

trong đó: X(k) là chuỗi kết quả.

x(m) là giá trị của mẫu m.

k-chỉ số của hệ số khai triển.

m-chỉ số của mẫu.

N- số mẫu có trong tín hiệu

$$C(k)=\begin{cases}1/\sqrt{2} & k=0\\ 1 & k\neq 0\end{cases}$$

b) DCT hai chiều

Để tách tương quan nội dung ảnh cao hơn, mã hóa DCT hai chiều (2-D) được dùng cho các khối 8×8 giá trị các điểm chói. Quá trình biến đổi DCT tiến FDCT (forward DCT) dùng trong tiêu chuẩn JPEG được định nghĩa như sau:

$$F(u,v)=\frac{C(u)C(v)}{4}\sum_{j=0}^{7}\sum_{k=0}^{7}f(j,k)\cos\frac{(2j+1)u\pi}{16}\cos\frac{(2k+1)v\pi}{16} \qquad (2.3.3)$$

trong đó:

f(j,k)- các mẫu gốc trong khối 8×8 pixel.

F(u,v)-các hệ số của khối DCT 8×8.

$$C(u),C(v)=\begin{cases}1/\sqrt{2} & u,v=0\\ 1 & u,v\neq 0\end{cases}$$

Phương trình trên là một liên kết của hai phương trình DCT một chiều, một cho tần số ngang và một cho tần số đứng. Giá trị trung bình của block 8x8 chính là hệ số thứ nhất (khi u,v= 0)

$$F(0,0)=\frac{1}{8}\sum_{j=0}^{7}\sum_{k=0}^{7}f(j,k) \qquad (2.3.4)$$

Phương trình này cộng tất cả các giá trị pixel trong khối 8×8 và chia kết quả cho 8. Kết quả phép tính bằng 8 lần giá trị pixel trung bình trong khối. Do đó hệ số thứ nhất được gọi là

hệ số DC. Các hệ số khác, dưới giá trị thành phần một chiều, biểu diễn các tần số cao hơn theo chiều dọc. Các hệ số ở về phía bên phải của thành phần một chiều biểu thị các tần số cao hơn theo chiều ngang. Hệ số trên cùng ở cận phải (0,7) sẽ đặc trưng cho tín hiệu có tần số cao nhất theo phương nằm ngang của ma trận 8×8, và hệ số hàng cuối bên trái (7,0) sẽ đặc trưng cho tín hiệu có tần số cao nhất theo phương thẳng đứng. Còn các hệ số khác ứng với những phối hợp khác nhau của các tần số theo chiều dọc và chiều ngang.

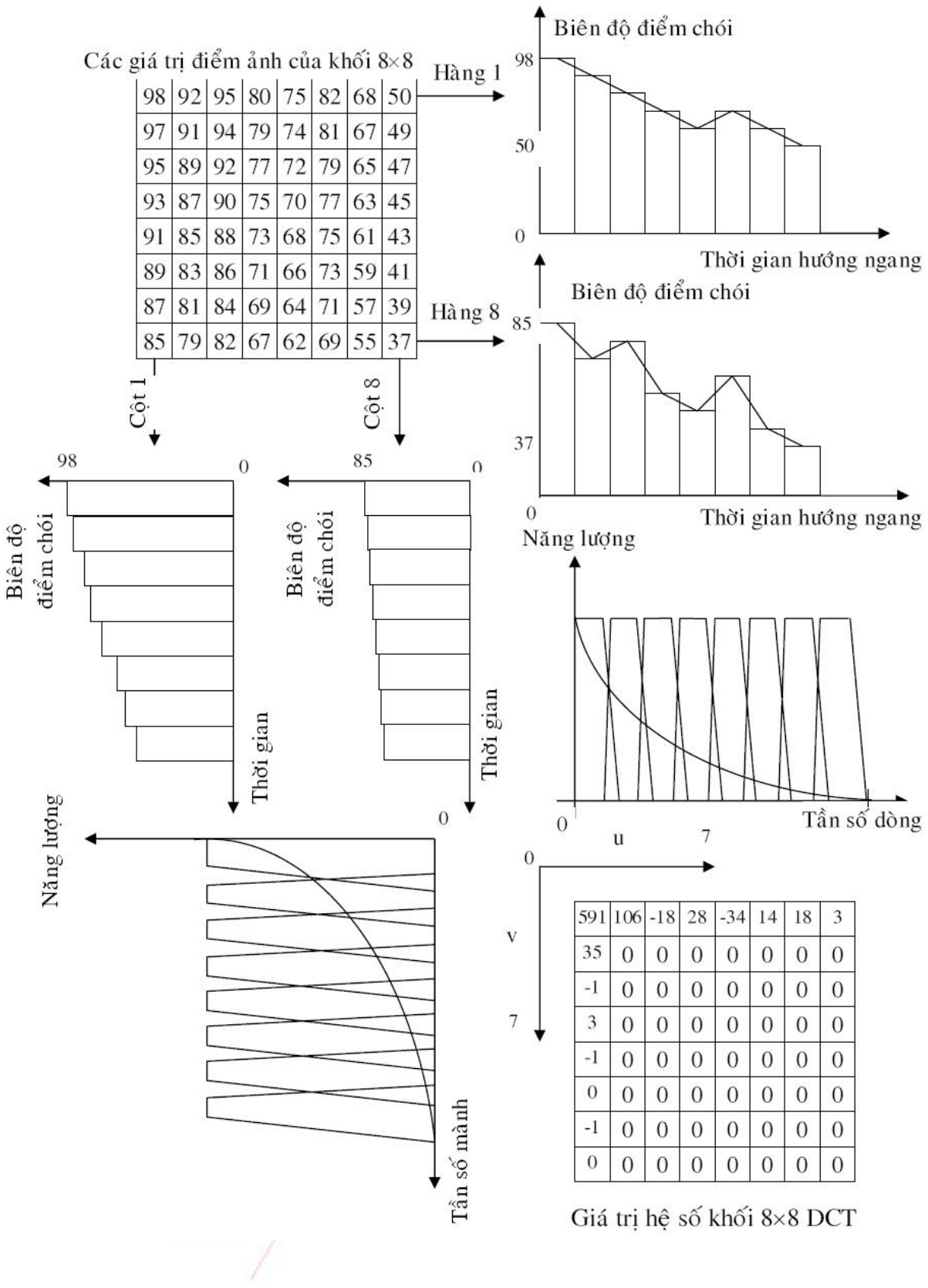
Phép biến đổi DCT hai chiều là biến đổi đối xứng và biến đổi nghịch có thể tạo lại các giá trị mẫu f(j,k) trên cơ sở các hệ số F(u,v) theo công thức sau:

$$f(j,k)=\sum_{u=0}^{7}\sum_{v=0}^{7}\frac{C(u)C(v)}{4}F(u,v)\cos\frac{(2j+1)u\pi}{16}\cos\frac{(2k+1)v\pi}{16} \qquad (2.3.5)$$

Như vậy, biến đổi DCT giống như biến đổi Fourier và các hệ số F(u,v) cũng giống nhau về ý nghĩa. Nó biểu diễn phổ tần tín hiệu được biểu diễn bằng các mẫu f(j,k). Bản thân phép biến đổi DCT không nén được số liệu, từ 64 mẫu ta nhận được 64 hệ số. Tuy nhiên, phép biến đổi DCT thay đổi phân bố giá trị các hệ số so với phân bố các giá trị mẫu.

Do bản chất của tín hiệu video, phép biến đổi DCT cho ta giá trị DC tức F(0,0) thường lớn nhất và các hệ số trực tiếp kề nó ứng với tần số thấp có giá trị nhỏ hơn, các hệ số còn lại ứng với tần số cao có giá trị rất nhỏ.

Hình vẽ 2.3.4 là một ví dụ minh họa quá trình DCT hai chiều của một khối 8×8 điểm ảnh (chói) được trích ra từ một ảnh thực. Nếu dùng quá trình DCT cho các tín hiệu số thành phần Y, C_R,C_B thì các tín hiệu C_B,C_R có biên độ cực đại ±128 (giá trị nhị phân trong hệ thống lấy mẫu 8 bit), còn tín hiệu Y có một khoảng cực đại từ 0 đến 255 giá trị nhị phân. Để đơn giản việc thiết kế bộ mã hóa DCT, tín hiệu Y được dịch mức xuống dưới bằng cách trừ 128 từ từng giá trị pixel trong khối để có khoảng cực đại của tín hiệu giống như đối với các tín hiệu C_R và C_B. Ở phần giải mã DCT, giá trị này (128) được cộng vào các giá trị pixel chói. Giá trị hệ số DC của khối DCT có một khoảng từ –1024 đến 1016.

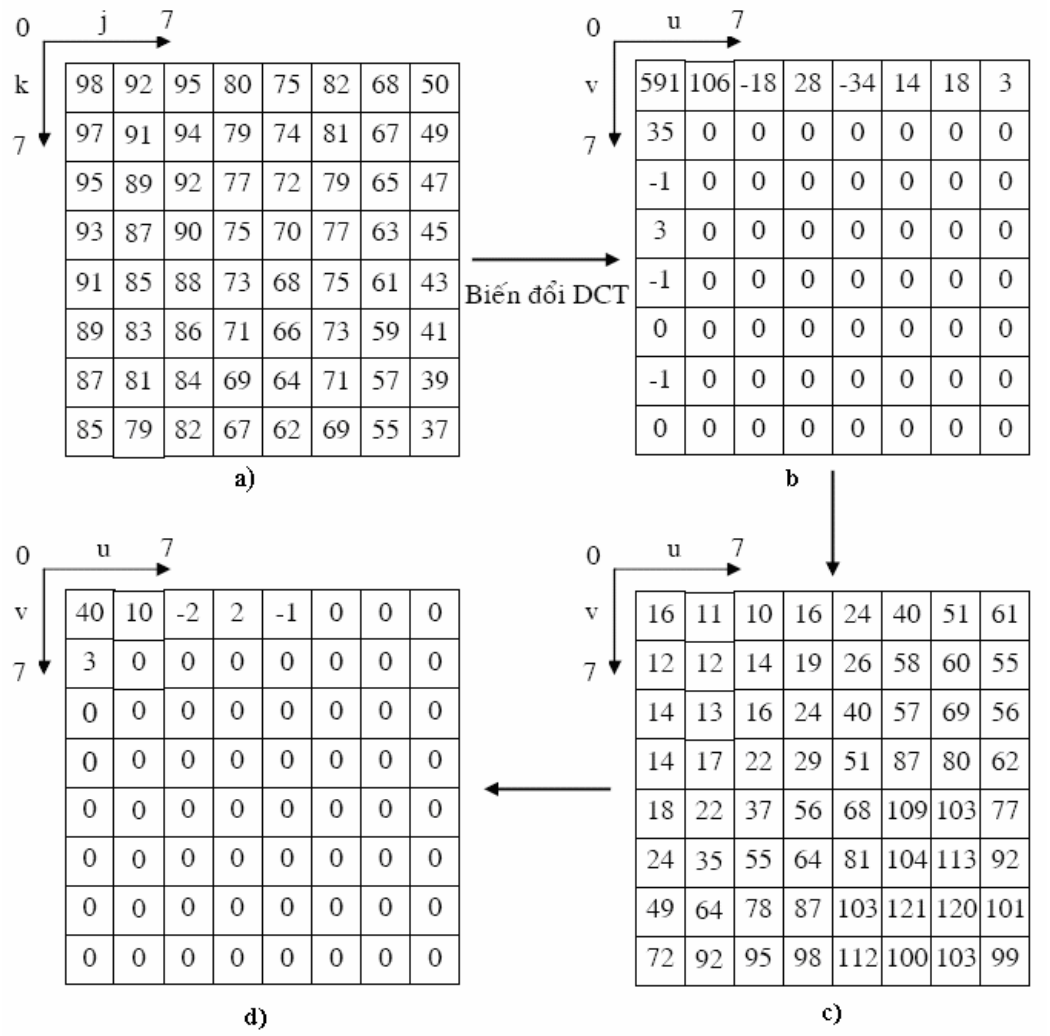


Hình 2.3.4 Mã hóa khối 8x8 bằng DCT 2 chiều

Đối với hệ số AC (với u,v=1,2,...,7), C(u) và C(v)=1 và các giá trị cực đại của nó nằm trong khoảng ±1020 theo phương trình FDCT. Khối 8×8 các giá trị của hệ số DCT đưa ra 1 giá trị DC lớn (ví dụ =591), biểu diễn độ sáng trung bình của khối 8×8 và các giá trị rất nhỏ của các thành phần tần số cao theo chiều ngang và chiều đứng.

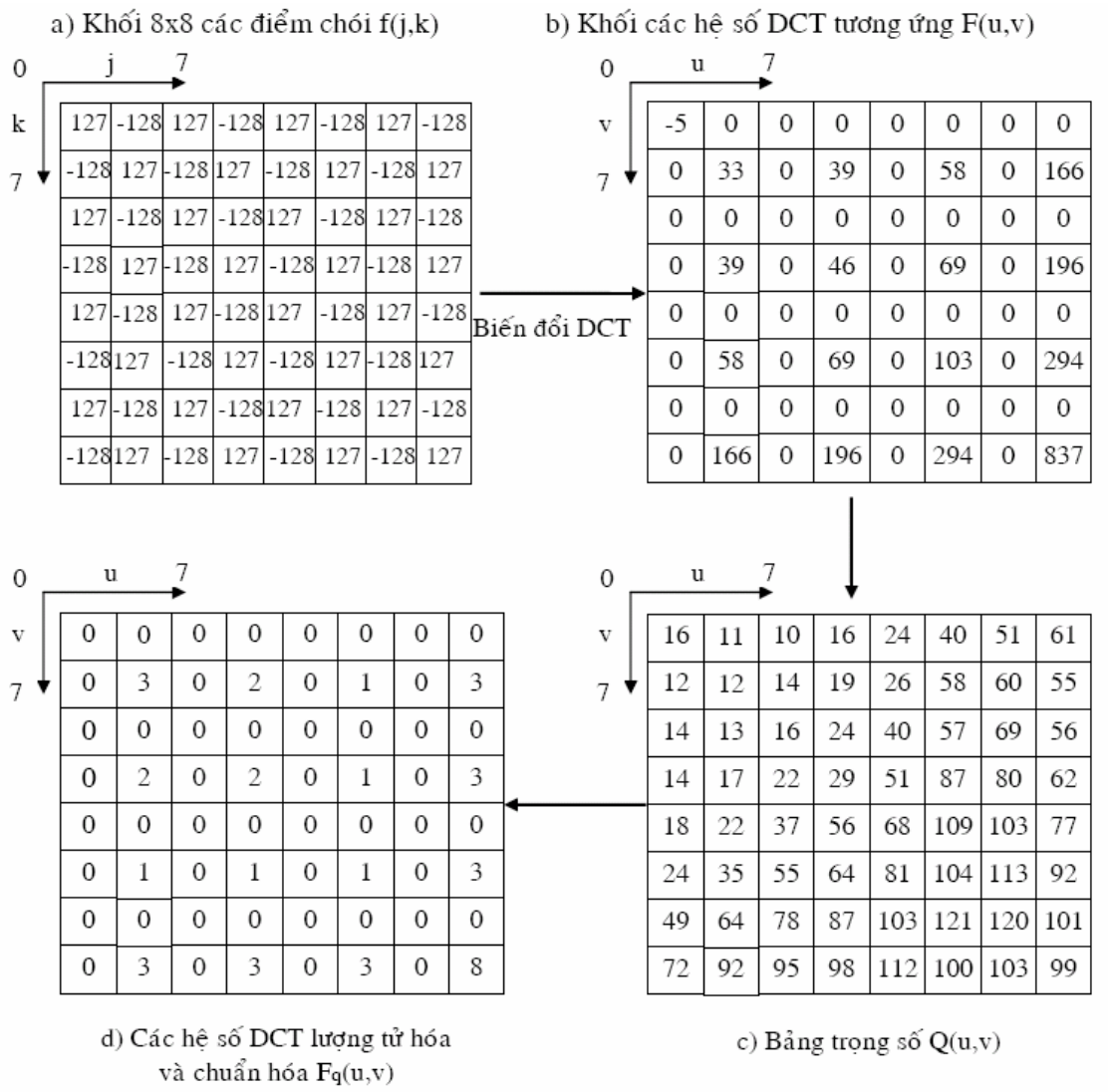
Nguyên tắc chung là nếu có sự thay đổi nhiều giá trị pixel-đến-pixel theo 1 chiều của khối pixel (ngang, đứng, chéo) sẽ tạo ra các giá trị hệ số cao theo các chiều tương ứng của khối hệ số DCT.

Khi nén ảnh theo JPEG, ma trận các hệ số khai triển sau DCT phải được nhân với bảng trọng số Q(u,v) để loại bỏ một phần các hệ số có biên độ nhỏ (thường là các tàhnh phần cao tần). Trên Hình 2.3.5 mô tả quá trình biến đổi DCT từ ảnh gốc (a) và ma trận hệ số DCT trước (b) và sau (d) khi nhân với Q(u,v) (c).



Hình 2.3.5 Khai triển DCT và bảng trọng số Q(u,v).

Ví dụ được trình bày trong hình vẽ 2.3.6, quá trình biến đổi DCT một khối pixel có các giá trị pixels đen (giá trị=0) và trắng (giá trị=255) xen kẽ. Sau khi giảm với –128 thu được các giá trị pixels trong dải động +127 và –128. Đây là khối ảnh có sự biến đổi lớn nhất về giá trị biên độ các điểm ảnh và các hệ số DCT kết quả xác định nhận xét này. Mặc dù có nhiều hệ số tần số bằng 0, giá trị của các hệ số có tần số cao đóng vai trò quan trọng trong việc tái tạo lại hình ảnh khi biến đổi DCT ngược.



Hình 2.3.6 Khai triển DCT cho khối ảnh có độ chói dạng bàn cờ.

Tóm lại, DCT làm giảm độ tương quan không gian của thông tin trong block. Điều đó cho phép biểu diễn thích hợp ở miền DCT do các hệ số DCT có xu hướng có phần dư thừa ít hơn. Điều này có nghĩa là DCT gói một phần lớn năng lượng tín hiệu vào các thành phần biến đổi có tần số tương đối thấp để lưu trữ hoặc truyền dẫn, tạo 0 và các giá trị rất thấp đối với thành phần tần số cao. Nhờ đặc tính của hệ thống nhìn của mắt người, các hệ số DCT có thể được mã hóa phù hợp, chỉ các hệ số DCT quan trọng nhất mới được mã hóa và truyền đi. DCT thuận kết hợp với DCT nghịch sẽ không cho tổn thất nếu độ dài từ mã của hệ số là 13 đến 14 bits cho tín hiệu video đầu vào được số hóa bằng các mẫu dài 8 bit. Nếu hệ số được lượng tử hóa bằng 11 bit (hoặc ngắn hơn), thì nén bằng DCT sẽ có tổn hao.

2.3.2.2 Lượng tử hóa

Bước tiếp theo của quá trình nén trong ảnh là lượng tử hóa các hệ số F(u,v) sao cho làm giảm được số lượng bit cần thiết. Các hệ số tương ứng với tần số thấp có các giá trị lớn hơn, và như vậy nó chứa phần năng lượng chính của tín hiệu, do đó phải lượng tử hóa với độ chính xác cao. Riêng hệ số một chiều đòi hỏi độ chính xác cao nhất, bởi lẽ nó biểu thị giá trị độ chói trung bình của từng khối phần tử ảnh.

Bất kỳ một sai sót nào trong quá trình lượng tử hệ số một chiều đều có khả năng nhận biết dễ dàng bởi nó làm thay đổi mức độ chói trung bình của khối. Ngược lại, với các hệ số tương ứng với tần số cao và có các giá trị nhỏ, thì có thể biểu diễn lại bằng tập giá trị nhỏ hơn hẳn các giá trị cho phép.

Chức năng cơ bản của bộ lượng tử hóa là chia các hệ số F(u,v) cho các hệ số ở vị trí tương ứng trong bảng lượng tử Q(u,v) để biểu diễn số lần nhỏ hơn các giá trị cho phép của hệ số DCT. Các hệ số có tần số thấp được chia cho các giá trị nhỏ, các hệ số ứng với tần số cao được chia cho các giá trị lớn hơn. Sau đó, các hệ số được làm tròn (bỏ đi các phần thập phân).

Kết quả ta nhận được bảng Fq(u,v) mới, trong đó phần lớn các hệ số có tần số cao sẽ bằng 0. Hệ số lượng tử hóa thuận được xác định theo biểu thức:

$$Fq(u,v)=\left[\frac{F(u,v)}{Q(u,v)}\right]=\text{số nguyên tố gần nhất}\left[\frac{F(u,v)+\frac{Q(u,v)}{2}}{Q(u,v)}\right] \qquad (2.3.6)$$

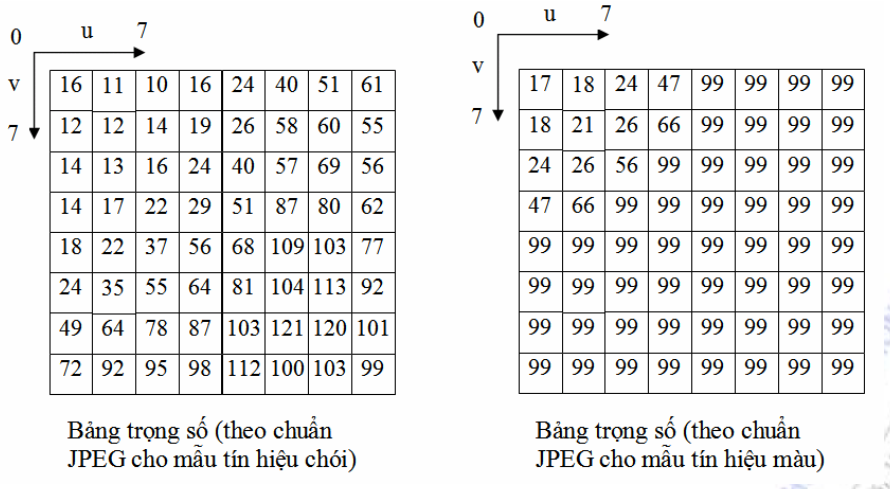
Các giá trị Fq(u,v) sẽ được mã hóa trong các công đoạn tiếp theo.

Cần phải xác định là trong quá trình lượng tử hóa có trọng số có xảy ra mất thông tin, gây tổn hao. Đây là bước tổn hao duy nhất trong thuật toán nén. Mức độ tổn hao phụ thuộc vào giá trị các hệ số trên bảng lượng tử. Sau khi nhân các hệ số lượng tử hóa Fq(u,v) với Q(u,v) và biến đổi ngược DCT sẽ không nhận được block sơ cấp các mẫu f(j,k). Tuy nhiên, trong trường hợp ảnh tự nhiên và lựa chọn các giá trị Q(u,v) thích hợp, sự khác nhau sẽ nhỏ đến mức mà mắt người không phân biệt được giữa ảnh gốc và ảnh biểu diễn.

Các thành phần DC và tần số thấp là các thông số nhạy cảm nhất của khối pixel gốc. Hệ số DC sẽ được lượng tử với độ chính xác 12 bit nhằm tránh các nhiễu xuất hiện giữa các khối điểm ảnh. Ngược lại, các hệ số tần số cao có thể lượng tử hóa thô với độ chính xác 2 bit-do khả năng cảm nhận của mắt người giảm ở tần số cao. Theo đó, hệ số chia trong bảng lượng tử hóa là nhỏ đối với các hệ số có tần số thấp và tăng từ từ đối với các hệ số có tần số cao hơn.

Trong hình vẽ dưới đây, giá trị khối xác định cho phép các giá trị tín hiệu chói và tín hiệu màu được lượng tử khác nhau. Nhiễu lượng tử đối với tín hiệu màu khó nhìn thấy hơn đối với tín hiệu chói, cho nên có thể thực hiện lượng tử hóa thô tín hiệu màu.

Như vậy, khối DCT đóng vai trò quan trọng trong quá trình lượng tử hóa khi thiết kế hệ thống nén video vì nó ảnh hưởng trực tiếp đến việc cho lại chất lượng ảnh khôi phục tốt hay xấu.



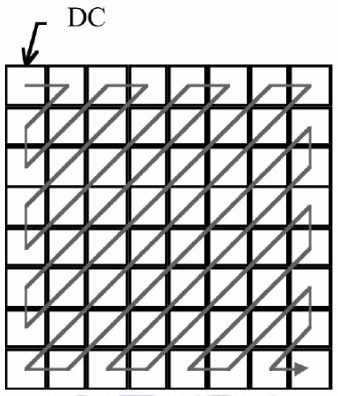
Hình 2.3.7 Các bảng lượng tử cho tín hiệu chói và màu theo chuẩn JPEG

Việc biến đổi sao cho chất lượng hình ảnh do mắt người cảm nhận tốt, phụ thuộc vào các thành phần tần số và sự biến đổi chi tiết ảnh từng vùng trong miền không gian. Các ảnh càng chi tiết thì hệ số thành phần tần số cao càng lớn.

2.3.2.3 Quét zig-zag

Để mã hóa entropy các hệ số được lượng tử hóa Fq(u,v), trước hết, cần biến đổi mảng hai chiều của các hệ số Fq(u,v) thành chuỗi số một chiều bằng cách quét zig-zag.

Việc xử lý 64 hệ số của khối 8x8 pixel bằng cách quét zig-zag làm tăng tối đa chuỗi các giá trị 0 và do vậy làm tăng hiệu quả nén khi dùng RLC.



Hình 2.3.8 Quét zig-zag các hệ số lượng tử hóa DCT

2.3.2.4 Mã hóa độ dài chạy (RLC)

Các giá trị lượng tử hóa có thể chỉ biểu diễn nhờ các từ mã có độ dài cố định hay đồng đều, tức là các giá trị lượng tử hóa biểu diễn bằng cùng một số bit. Tuy nhiên hiệu quả của việc mã hóa không cao. Để cải tiến hiệu quả người ta dùng mã hóa entropy. Mã hóa entropy dùng những đặc tính thống kê của tín hiệu được mã hóa. Một tín hiệu, ở đây là giá trị pixel hoặc các hệ số chuyển vị, có chứa một lượng thông tin (entropy) tùy theo những xác suất của những giá trị hay sự kiện khác nhau xuất hiện. Ví dụ những từ mã nào ít xảy ra hơn sẽ có nhiều thông tin hơn từ mã hay xảy ra.

Khi dùng mã hóa entropy có hai vấn đề đặt ra: thứ nhất, mã hóa entropy làm tăng độ phức tạp và yêu cầu bộ nhớ lớn hơn so với mã độ dài cố định. Thứ hai, mã hóa entropy gắn

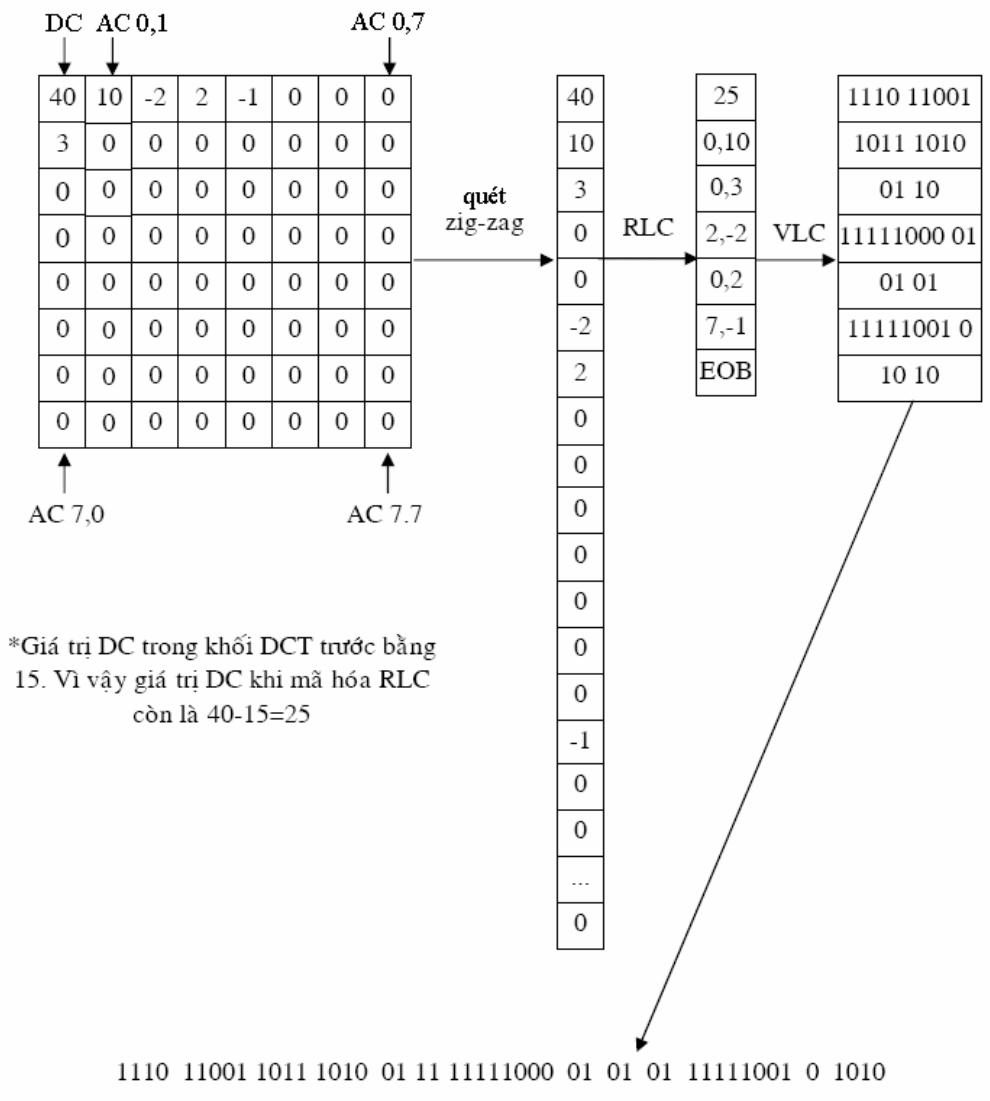
liền với tính không ổn định của tín hiệu video sẽ làm tốc độ bit thay đổi theo thời gian. Do đó, cần một cơ cấu điều khiển bộ đệm khi mã hóa nguồn tốc độ bit biến đổi được ghép với kênh tốc độ bit hằng.

Bộ mã hóa entropy làm giảm độ dư thừa thống kê trong các phần tử được mã hóa để truyền. Sự dư thừa chính là phân bố xác suất không đồng đều trên giá trị của mỗi phần tử. Phân bố xác suất càng lệch khỏi phân bố đều thì hiệu suất mã hóa entropy càng tăng. Mã Huffman là một trong những sơ đồ mã được sử dụng phổ biến. Ngoài ra, trong mã hóa entropy còn sử dụng mã RLC sẽ cho hiệu suất nén rất cao.

Kỹ thuật RLC được dùng để mã hóa có hiệu quả các hệ số DCT đã lượng tử hóa hơn là dùng trực tiếp cho số liệu ảnh. Sau quá trình quét zig-zag ở trên, RLC sẽ được thực thi. Một hệ số khác 0 sau giá trị DC được mã hóa bằng 1 từ mã bao gồm 2 thông số: số lượng 0 chạy trước 1 hệ số riêng khác 0 và mức của nó sau khi lượng tử hóa. RLC thực chất là việc thay thế các hệ số có giá trị 0 bằng số lượng các chữ số 0 xuất hiện.

Hình vẽ 2.3.9 là một ví dụ về mã hóa entropy. Trong ví dụ này, chuỗi một chiều các hệ số DCT sau khi quét zig-zag với các giá trị giống nhau được gom lại với nhau bằng mã RLC. Lúc này, chuỗi một chiều có các đoạn chuỗi dài có cùng giá trị là các symbol có dạng:< chiều dài chuỗi 0, giá trị>.

Ở đây, giá trị 10 không có giá trị 0 nào trước đó được biểu diễn bằng <0,10>; giá trị –2 có hai giá trị 0 đứng trước được biểu diễn bằng <2,-2>v.v... Riêng một dấu đặc biệt là End of Block (EOB) được dùng để cho biết tất cả các hệ số tiếp theo trong khối bằng 0. Trong ví dụ này, ta có một chuỗi 49 từ mã với giá trị 0. Như vậy chỉ xét riêng 49 từ mã giá trị 0 được nén xuống chỉ còn 3 từ mã. Điều này chứng tỏ hiệu suất nén rất cao của mã hóa RLC. Nén bằng mã RLC là quá trình nén không tổn hao.



Hình 2.3.9 Quá trình mã hóa RLC

2.3.2.5 Mã hóa độ dài thay đổi VLC

Các từ mã RLC tiếp tục được mã hóa bằng cách đặt các từ mã ngắn cho các mức có xác suất xuất hiện cao và các từ mã dài cho các mức có xác suất xuất hiện thấp. Bảng 2.3.1 minh họa các phân nhóm các hệ số AC.

Bảng 2.3.2 là một ví dụ về bảng mã Huffman tương ứng cho các nhóm. Từ mã ngắn báo hiệu kết thúc khối (EOB) cho biết tất cả các hệ số còn lại trong khối mang giá trị 0. Trong ví dụ khối hệ số DCT, hệ số DCT (40) được mã hóa DPCM bằng cách dùng giá trị DC (25) của khối DCT trước. Mã hóa DPCM mở rộng thang biểu diễn tín hiệu Y từ (-1024 đến 1016) đến (-2048 đến 2032).

Bảng 2.3.1 Phạm vi giá trị các hệ số trong các nhóm (category).

Loại	Phạm vi hệ số	
NA	0	
1	-1	1
2	-3,-2	2, 3
3	-7,-6,-5,-4	4, 5, 6, 7
4	-15,.....,-8	8,.....,15
5	-31,.....,-16	16,.....,31
6	-63,.....,-32	32,.....,63
7	-127,.....,-64	64,.....,127
8	-255,.....,-128	128,.....,255
9	-511,.....,-256	256,.....,511
10	-1023,.....,-512	512,.....,1023
11	-2047,.....,-1024	1024,.....,2047

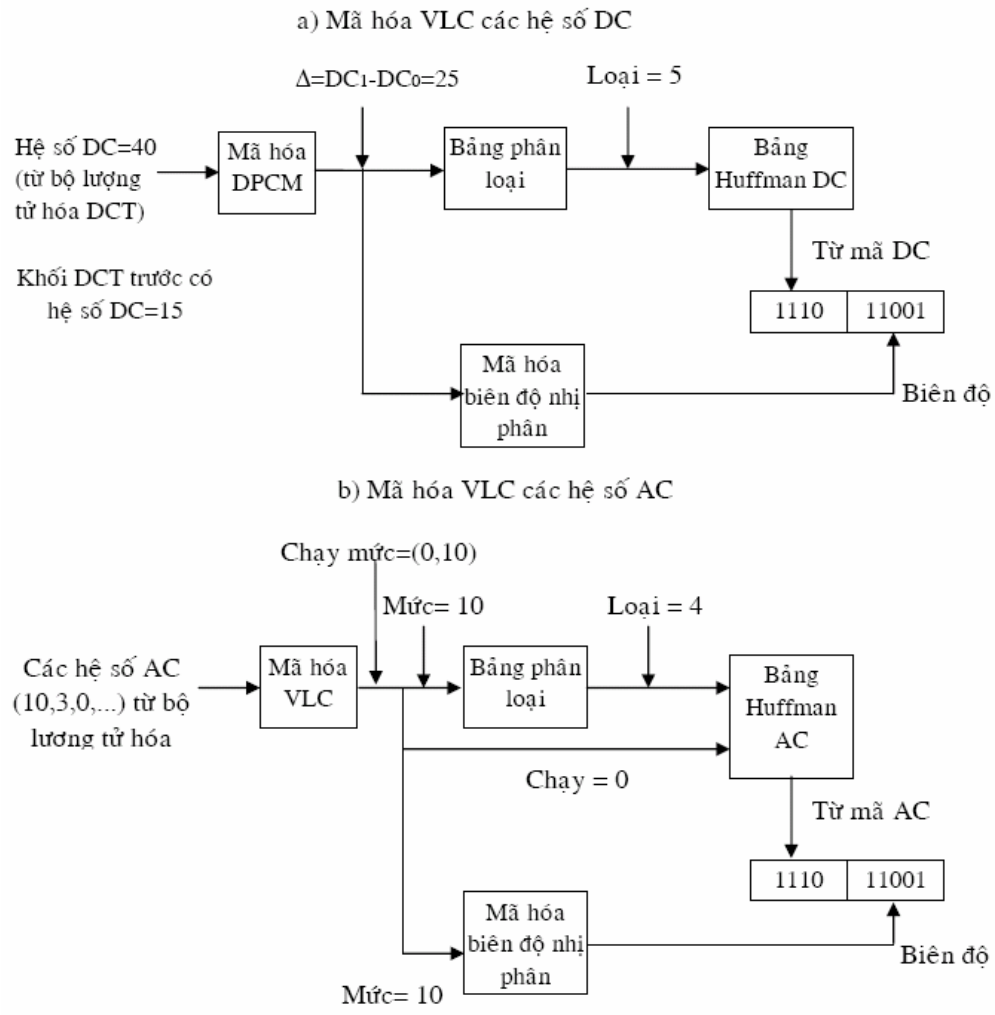
Bảng 2.3.2 Ví dụ bảng Huffman cho hệ số AC

0 chạy	loại	Độ dài mã	Từ mã
0	1	2	00
0	2	2	01
0	3	3	100
0	4	4	1011
0	5	5	11010
0	6	6	111000
0	7	7	1111000
.	.	.	.
1	1	4	1100
1	2	6	111001
1	3	7	1111001
1	4	9	111110110
.	.	.	.
2	1	5	11011
2	2	8	11111000
.	.	.	.
3	1	6	111010
3	2	9	111110111
.	.	.	.
4	1	6	111011
5	1	7	1111010
6	1	7	1111011
7	1	8	11111001
8	1	8	11111010
9	1	9	111111000
10	1	9	111111001
11	1	9	111111010
.	.	.	.
EOB		4	1010

Giá trị chênh lệch hệ số DC được mã hóa VLC nhờ một bảng tìm kiếm (lookup table). Đầu ra của nó là một số nhị phân Huffman dựa trên giá trị chênh lệch các hệ số DC này. Các hệ số AC biểu diễn bởi các từ mã RLC được mã hóa Huffman bằng các bảng tìm kiếm. Đầu ra kết hợp với giá trị chạy (số lượng số 0 trước hệ số AC) để tạo một số nhị phân Huffman

biểu diễn giá trị hệ số AC tương ứng. Trong cả hai trường hợp mã hóa giá trị sai lệch hệ số DC và độ lớn các hệ số AC đều sử dụng từ mã nhị phân ngắn nhất để biểu diễn chúng.

Tại đầu ra VLC, tất cả các từ mã của cùng một khối DCT được kết hợp tạo thành một dòng tín hiệu ra. Trong ví dụ trên, số liệu tương ứng với khối DCT ban đầu (8x8x8 bit =512 bit) được giảm thành 48 bits sau khi mã hóa VLC. Hệ số nén trong trường hợp này bằng 512/48=10,6. Hệ số nén cũng thường được tính bằng số bit biểu diễn điểm ảnh. Trong ví dụ trên, 48 bit biểu diễn cho 64 điểm ảnh, theo đó thu được hệ số nén tương ứng là 48/64=0,75 (bit/điểm ảnh). Mã hóa VLC tự nó là một kỹ thuật mã hóa không tổn thất, nó cho phép giảm thêm tốc độ dòng bit (đã được giải tương quan, làm tròn, và giảm qua quá trình lượng tử hóa DCT). Quá trình mã hóa VLC cho hệ số DC và các hệ số AC được mổ tả trong sơ đồ khối Hình 2.3.10.



Hình 2.3.10 Sơ đồ khối hệ thống mã VLC cho hệ số DC (a) và AC (b)

2.3.2.6 Bộ nhớ đệm

Sau khi mã hóa entropy, ta nhận được chuỗi bit có tốc độ thay đổi, phụ thuộc vào độ phức tạp của ảnh. Nếu phải truyền qua kênh có tốc độ cố định thì ta cần phải có bộ nhớ đệm. Cơ chế điều khiển bộ nhớ là đảm bảo bộ nhớ không trống (underflow) hoặc không tràn (overflow) bằng cơ cấu hồi tiếp. Bộ mã hóa kiểm tra trạng thái đầy của bộ nhớ đệm. Khi số liệu trong bộ nhớ đệm gần bằng dung lượng cực đại, thì các hệ số biến đổi DCT được lượng tử hóa ít chính xác hơn (tăng tỉ số nén). Trong trường hợp ngược lại, có nghĩa là bộ nhớ đệm

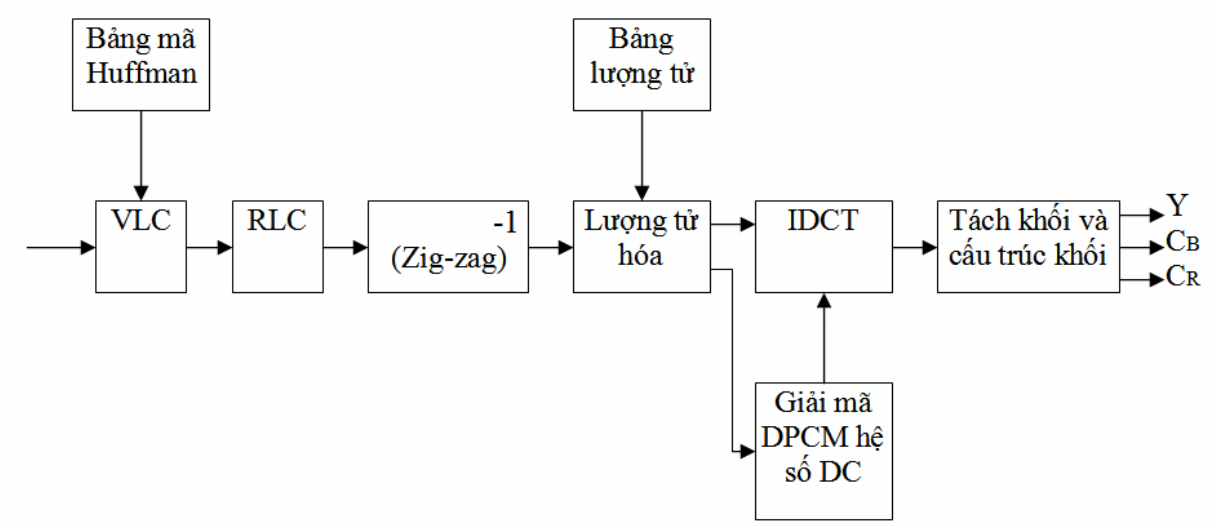
chứa số liệu quá ít, thì độ chính xác của việc lượng tử hóa các hệ số sẽ tăng lên. Quá trình này được thực hiện một cách tự động.

2.3.2.7 Quá trình giải nén

Quá trình giải nén trong ảnh dựa trên cơ sở thực hiện thuật toán ngược với quá trình nén.

Hình vẽ 2.3.11 minh họa sơ đồ khối của quá trình giải nén DCT. Các bảng Huffman và lượng tử hóa giống như các bảng của bộ mã hóa DCT được dùng để tạo lại các giá trị hệ số DCT của một khối 8x8 pixel. Quá trình lượng tử hóa ngược R(u,v) được tiến hành theo biểu thức :

$$R(u,v)=Fq(u,v)Q(u,v) \qquad (2.3.7)$$



Hình 2.3.11 Sơ đồ khối hệ thống giải mã JPEG

Quá trình biến đổi DCT ngược (IDCT) tạo lại khối giá trị các điểm ban đầu theo biểu thức:

$$f^*(j,k)=\sum_{u=0}^{7}\sum_{v=0}^{7}\frac{C(u)C(v)}{4}F(u,v)\cos\frac{(2j+1)u\pi}{16}\cos\frac{(2k+1)v\pi}{16} \qquad (2.3.8)$$

Quá trình tính toán IDCT cũng tương tự như quá trình tính toán FDCT. Ảnh được khôi phục $f^*(j,k)$ (ma trận trên Hình 2.3.12d) khác với ảnh gốc (trên Hình 2.3.5a). Sai số giữa các giá trị khôi phục và giá trị gốc được tính như sau:

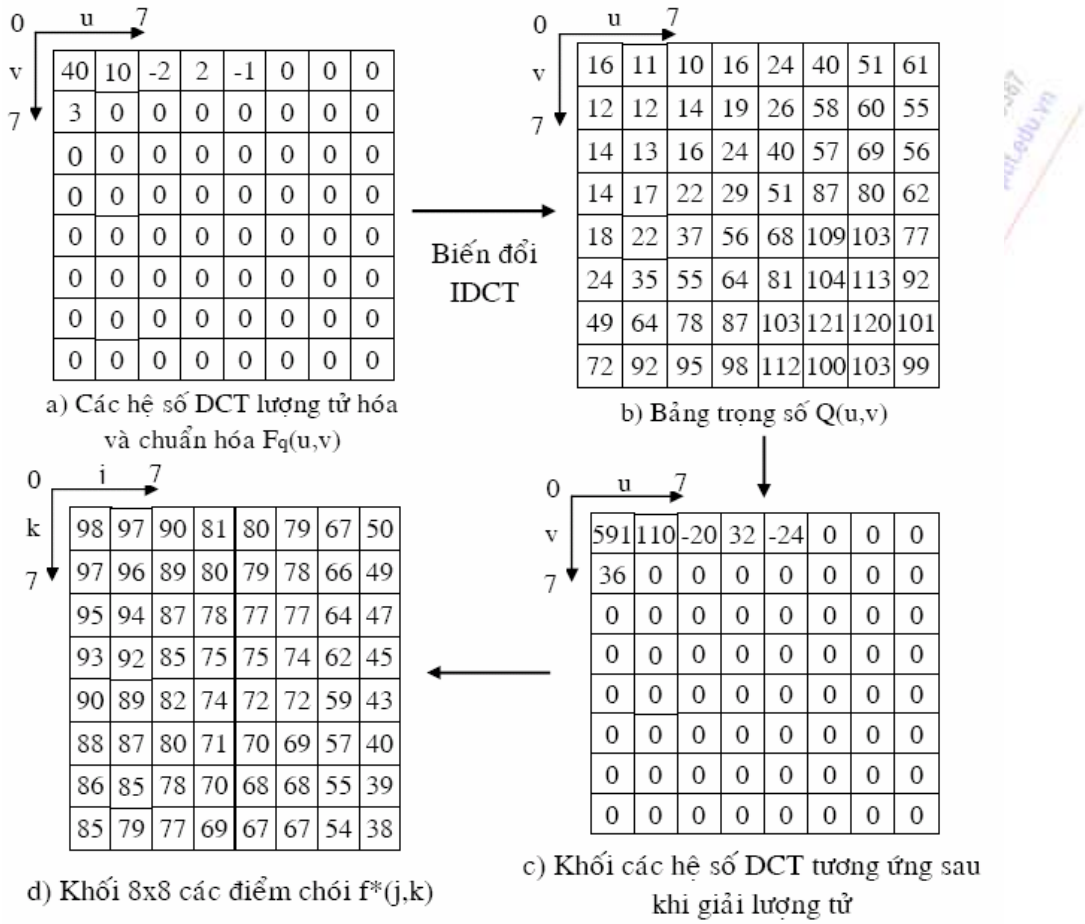
$$e(j,k)=f(j,k)-f^*(j,k) \qquad (2.3.9)$$

Để đánh giá chất lượng ảnh khôi phục, ta sử dụng các đại lượng đo là giá trị sai số trung bình bình phương (RMSE) và hệ số biên độ đỉnh tín hiệu trên nhiễu (PSNR: peak signal-to-noise ration):

$$RMSE=\sqrt{\frac{1}{64}\sum_{k=0}^{7}\sum_{j=0}^{7}e(j,k)e(j,k)} \qquad (2.3.10)$$

$$PSNR=20\lg\left(\frac{255}{RMSE}\right) \qquad (2.3.11)$$

Giá trị sai số tuyệt đối của ảnh giải nén và ảnh gốc cho hai hình ảnh có độ chi tiết khác nhau được biểu diễn trên Hình 2.3.13a và 3.3.13b. Như vậy có thể thấy rằng ảnh có phân bố độ chói dạng bàn cờ "khó nén" hơn do các thành phần DCT tần số cao có biên độ lớn.



Hình 2.3.12 Khôi phục các điểm ảnh trong khối 8x8

0	-5	5	-1	-5	3	1	0
0	-5	5	-1	-5	3	1	0
0	-5	5	-1	-5	2	1	0
0	-5	5	-1	-5	3	1	0
1	-4	6	-1	-4	3	-28	0
1	-4	6	0	-4	4	2	1
1	-4	6	-1	-4	3	2	0
0	0	5	-2	-5	2	1	-1

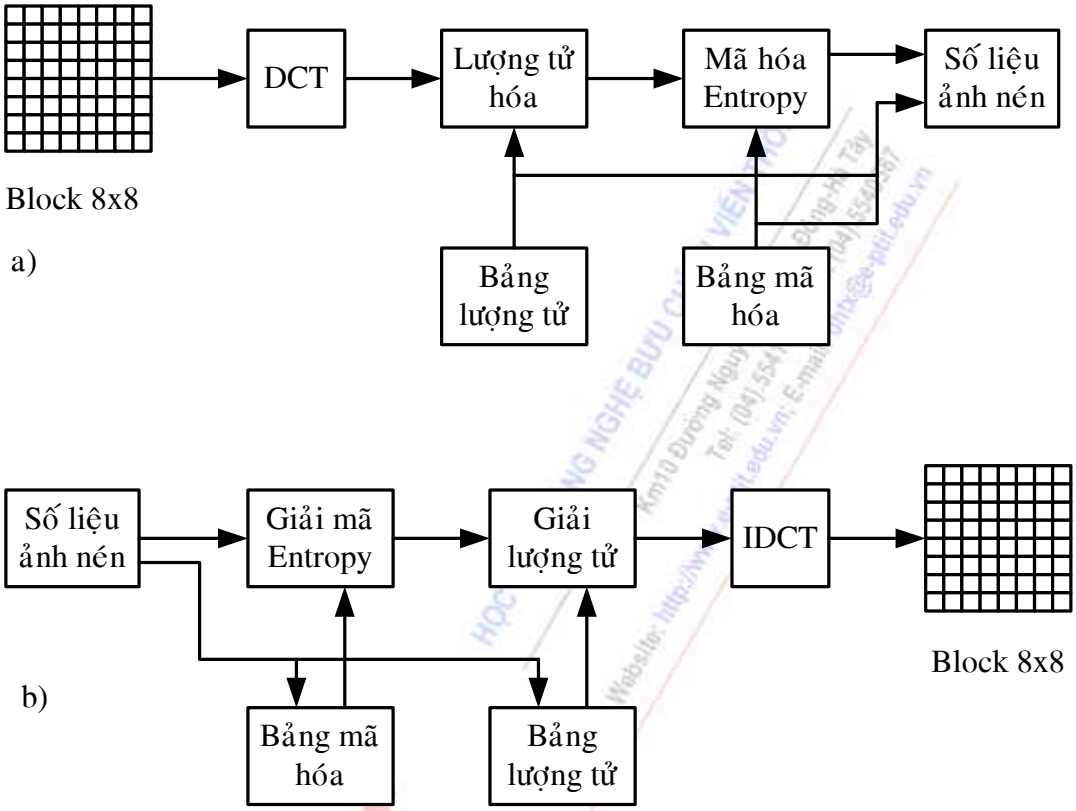
-5	-2	5	-11	10	-6	1	4
2	7	-6	13	-14	5	-8	-3
0	-14	0	-9	8	-1	13	-1
3	1	-9	10	-11	8	-2	-4
-4	-2	8	-11	10	-9	1	3
-1	13	-1	8	-9	0	-14	0
-3	-8	5	-14	13	-6	7	2
4	1	-6	10	-11	5	-2	-5

Hình 2.3.13 Sai số của các điểm ảnh được khôi phục so với giá trị ban đầu

a) ảnh (2.3.5a) có mức độ chi tiết thấp b) ảnh hình bàn cờ

2.3.2.8 Mã hóa và giải mã JPEG tuần tự

Thuật toán nén và giải nén theo từng block 8x8 theo JPEG được phân tích ở trên được gọi là phương pháp nén tuần tự. Sơ đồ khối thuật toán mã hóa và giải mã tiêu chuẩn JPEG tuần tự mô tả trên Hình 2.3.14.



Hình 2.3.14 Sơ đồ khối mã hóa (a) và giải mã (b) JPEG

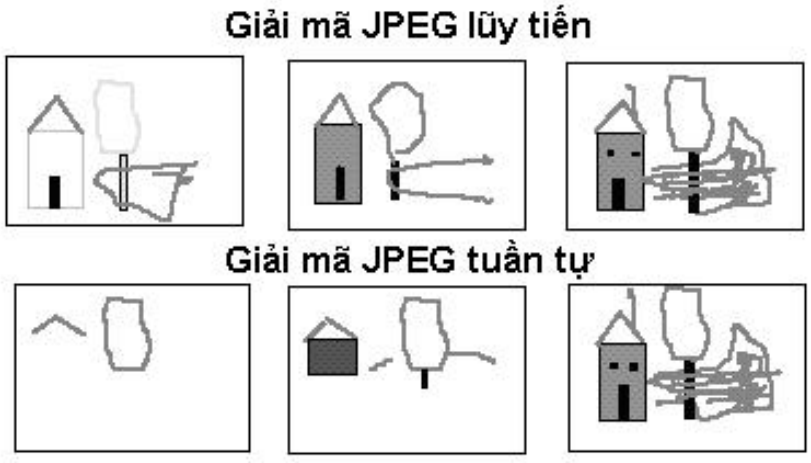
Để mã hóa JPEG, các mẫu pixels đầu vào được nhóm thành 8×8 blocks. Mỗi block được biến đổi FDCT thành một tập 64 giá trị, gọi là các hệ số DCT. Sau đó, 64 hệ số này được lượng tử hóa. Điều này có nghĩa là các hệ số không còn chính xác và một số hệ số sẽ bằng 0. Các hệ số được quét zig-zag trước khi mã hóa entropy. Hiệu quả của quét zig-zag là biến đổi ma trận 2 chiều thành ma trận 1 chiều. Hệ số đầu tiên trong chuỗi zig-zag là hệ số DC. Tất cả các hệ số còn lại là hệ số AC. Đối với hệ số DC, sai lệch giữa giá trị hiện tại và giá trị DC trước đó được mã hóa. Mã hóa entropy sử dụng mã hóa Huffman.

Quá trình giải mã ngược lại với quá trình mã hóa. Các tiến trình giống nhau nhưng ngược thứ tự. Bảng tham số sử dụng cho quá trình mã hóa được mang đi cùng với luồng dữ liệu sau khi nén và được sử dụng cho giải mã. Các hệ số DCT được nhân với các hệ số giải lượng tử và chuyển đến khối biến đổi DCT ngược. Đầu ra của quá trình là 8×8 block của pixels. Dĩ nhiên, khối 8×8 block pixels này không giống hoàn toàn khối pixels ban đầu bởi vì đã có mất mát thông tin trong quá trình mã hóa. Do đó quá trình này được coi là nén có tổn thất.

Trên quan điểm coi hình ảnh động là một chuỗi liên tiếp các hình ảnh tĩnh. Khi đó, tiêu chuẩn JPEG được áp dụng cho việc nén ảnh động và nó có tên gọi M – JPEG.

2.3.2.9 Nén JPEG lũy tiến

Trong một số ứng dụng, một ảnh có thể có số lượng điểm ảnh lớn và việc giải mã, bao gồm truyền ảnh nén qua mạng, có thể mất nhiều phút. Trong những ứng dụng như vậy, cần thiết có tiến trình tạo ảnh thô nhanh chóng và sau đó cải thiện chất lượng của nó bằng nhiều lần quét. Kiểu mã hóa JPEG lũy tiến thực hiện nhiều lần quét, mỗi lần quét mã hóa một tập con các hệ số DCT. Vì vậy, mã hóa JPEG lũy tiến phải có bộ đệm phụ tại đầu ra của bộ lượng tử và trước bộ mã hóa entropy. Kích thước bộ đệm phải đủ lớn để chứa tất cả các hệ số DCT của ảnh.



Hình 2.3.15 Quá trình giải mã JPEG lũy tiến và tuần tự

Hình vẽ 2.3.15 mô tả sự khác nhau trong quá trình giải nén JPEG lũy tiến và JPEG tuần tự.

Nén JPEG lũy tiến có thể thực hiện 3 giải thuật sau:

1) Giải thuật lựa chọn phổ (progressive spectral selection algorithm).

2) Giải thuật xấp xỉ thành công (progressive successive approximation algorithm).

3) Giải thuật kết hợp (combined progresssive algorithm).

Trong giải thuật lựa chọn phổ, các hệ số DCT được nhóm lại theo nhiều nhóm băng tần. Nhìn chung, hệ số DCT tần số thấp được gửi trước, và sau đó là các hệ số DCT tần số cao. Ví dụ, một chuỗi 4 nhóm như sau:

Band 1: chỉ có hệ số DC.

Band 2: hệ số AC_1 và AC_2.

Band 3: hệ số AC_3,AC_4,AC_5,AC_6.

Band 4: hệ số AC_7,…, AC_{63}.

Trong giải thuật xấp xỉ, tất cả các hệ số DCT được gởi đi trước với độ chính xác thấp hơn, và sau đó được cải tiến lại trong các lần quét sau. Ví dụ, một chuỗi ba band như sau:

Band 1: tất cả các hệ số DCT (chia 4).

Band 2: tất cả các hệ số DCT (chia 2).

Band 3: tất cả các hệ số DCT .

Giải thuật kết hợp kết hợp cả hai giải thuật chia phổ và xấp xỉ. Hệ thống JPEG lũy tiến hiệu quả cho việc truyền các ảnh phức tạp. Hệ thống này hướng đến những ứng dụng yêu cầu truyền nhanh các ảnh có độ phân giải cao, chất lượng cao và phức tạp qua mạng có băng thông giới hạn. Những ứng dụng đó bao gồm truyền ảnh y học, ứng dụng khám phá trái đất và vũ trụ cũng như các ứng dụng trên Internet.

2.3.2.10 Các tham số tiêu chuẩn của JPEG

Tiêu chuẩn JPEG xác định các tham số trong bảng 2.3.3.

Bảng 2.3.3. Tham số theo tiêu chuẩn JPEG.

Tham số	Đặc điểm
Tín hiệu mã hóa	RGB hoặc Y và C_R, C_B
Cấu trúc lấy mẫu	4:4:4, 4:2:2 và 4:2:0
Kích thước ảnh tối đa (điểm ảnh x điểm ảnh)	65 536 x 65 536
Biểu diễn mẫu	8 bit cho hệ thống cơ bản 8÷12 bit cho hệ thống mở rộng DCT
Độ chính xác của quá trình lượng tử hóa và biến đổi DCT	9 bit
Phương pháp lượng tử hóa hệ số DC	DPCM
Cấu trúc khối trong quá trình lượng tử hóa thích nghi	16x16 bit
Độ chính xác cực đại của hệ số DC	11 bit
Bảng lượng tử	Sai lệch giữa các giá trị Y và C_R ,C_B
Biến đổi RLC	Mã Huffman
Hệ số cân bằng các khối	Có thể biến đổi
Bù chuyển động	Không
Quét	Tuần tự hay xen kẽ
Kênh truyền	Được quản lý lỗi

2.3.2.11 Phương pháp nén ảnh động M – JPEG

M – JPEG là sự mở rộng của JPEG. Vì nén M – JPEG chỉ thực hiện trong mỗi ảnh, điều đó dẫn đến hiệu quả (tỉ số nén) thấp hơn so với các phương pháp nén ảnh động MPEG sẽ được xét sau đây.

Nén ảnh động theo phương pháp M – JPEG có dặc điểm như sau:

- Tín hiệu 48 Mbit/s (hệ số nén 3,5) cho kết quả ảnh rất tốt.
- Tín hiệu 36 Mbit/s (hệ số nén 4,7) cho kết quả ảnh có nhiễu với mức độ chất lượng nhất định.
- Tín hiệu 24 Mbit/s cho kết quả ảnh có nhiễu nhìn thấy, chất lượng ảnh khôi phục không đủ dùng cho mục đích chuyên dùng.

Trong trường hợp nén với tỉ số cao sẽ xuất hiện các ô vuông (Artifacts) trên ảnh khôi phục, đó là các đặc trưng của các hệ số DC. Nếu mã hóa nhiều lần thì hiệu ứng trên sẽ tăng lên.

Với những đặc điểm trên, chuẩn M – JPEG có ưu điểm khi sử dụng trong công nghệ sản xuất chương trình truyền hình. Vì các ảnh được mã hóa độc lập với nhau nên việc thực hiện dựng chính xác tới từng ảnh là hoàn toàn có thể thực hiện được. Đây chính là điểm mạnh của M – JPEG sử dụng trong các thiết bị sản xuất chương trình tiện dụng cho studio và dựng hậu kỳ, làm kỹ xảo với giá thành hệ thống phù hợp, không gây tổn hao trong quá trình dựng.

Tuy nhiên, đối với các thiết bị sử dụng định dạng nén M – JPEG có các nhược điểm :

- Mặc dù sử dụng cùng một phương pháp nén M – JPEG trong các thiết bị của mình, các sản phẩm của các nhà máy khác nhau cũng không hoàn toàn giống nhau về mặt biểu diễn cũng như phương pháp xử lý đối với tín hiệu video được nén. Chính vì vậy các thiết bị này rất khó có thể trao đổi trực tiếp số liệu cho nhau.
- Các thiết bị sử dụng phương pháp nén theo định dạng M – JPEG không thể sử dụng cho truyền dẫn, phát sóng vì tốc độ dòng bit sau khi được nén còn cao.

2.3.3 Chuẩn nén MPEG

2.3.3.1 Giới thiệu chung về MPEG

Nén tín hiệu video theo chuẩn MPEG (Moving Picture Experts Group) là phương pháp nén ảnh động không những làm giảm dư thừa không gian (như JPEG) mà còn làm giảm dư thừa thời gian giữa các khung ảnh, đây là khác biệt so với JPEG là chuẩn nén ảnh tĩnh chỉ làm giảm dư thừa thông gian trong một khung ảnh.

Chuẩn MPEG định nghĩa một khái niệm mới là “nhóm các khung ảnh” (GOP) để giải quyết dư thừa thời gian và cho phép truy xuất ngẫu nhiên khi mã hoá MPEG dùng để lưu trữ. Trong chuẩn MPEG, người ta quy định 3 loại khung ảnh phụ thuộc vào phương pháp nén: nén trong khung ảnh (khung I), nén ước đoán (khung P) và nén nội suy hai chiều theo thời gian (khung B). Khung I luôn luôn là khung ảnh đầu tiên trong nhóm GOP, tạo điểm truy xuất ngẫu nhiên chuẩn.

Chuẩn nén MPEG bao gồm các tiêu chuẩn nén video có tốc độ luống bít (tương đương chất lượng ảnh video) khác nhau:

a) MPEG – 1: (chuẩn ISO/IEC 11172) có từ tháng 11-1992. MPEG – 1 dùng cho nén động có kích thước 320x240 và tốc độ bit còn từ 1 Mbit/s đến 1,5 Mbit/s dùng cho ghi hình trên băng từ và đĩa quang (CD), đồng thời truyền dẫn trong các mạng (ví dụ như mạng máy tính). Đối với mã hóa audio stereo, tốc độ khoảng 250 Kbps. MPEG – 1 có các thành phần chính sau đây :

- Phần 1 : ISO/IEC 11172 – 1 : Hệ thống ghép kênh Video, Audio MPEG – 1 .
- Phần 2 : ISO/IEC 11172 – 2 : Nén video MPEG – 1.
- Phần 3 : ISO/IEC 11172 – 3 : Nén audio MPEG – 1.
- Phần 4 : ISO/IEC 11172 – 4 : Conformance.
- Phần 5 : ISO/IEC 11172 – 5 : Software.

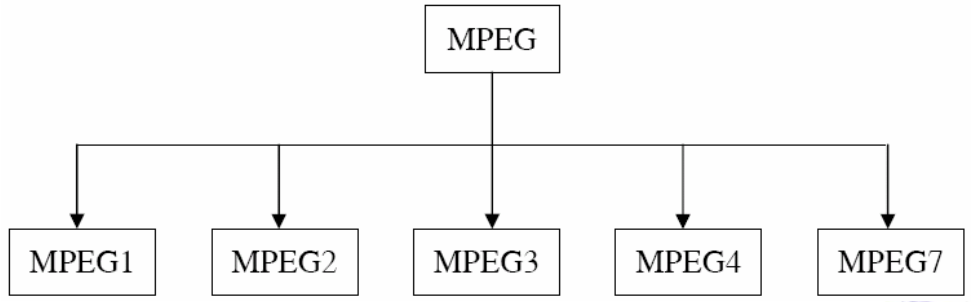
b) MPEG – 2 : là dự án giai đoạn hai của ISO/IEC, được sử dụng cho các ứng dụng cao hơn với tốc độ còn ≤ 10 Mbit/s cho viễn thông, truyền hình thông thường.

c) MPEG – 3 : tiêu chuẩn nén tín hiệu số xuống còn ≤ 50 Mbps để truyền tín hiệu truyền hình có độ phân giải cao (HDTV). Sau đó, nhập chung vào MPEG – 2 và thành tiêu chuẩn quốc tế MPEG – 2 vào tháng 11 năm 1994 (ISO/IEC 13818). Tiêu chuẩn MPEG – 2 này dùng cho truyền hình thông thường và truyền hình có độ phân giải cao.

d) MPEG – 4 : được thiết kế để mã hóa Video/Audio với tốc độ thấp (khoảng 9÷14 Kbps) chủ yếu ứng dụng trong điện thoại video, multimedia. MPEG – 4 hoàn thiện vào tháng 10 – 1998.

e) MPEG – 7 : chuẩn này được đề nghị vào tháng 10 – 1998 và thành chuẩn quốc tế vào tháng 9 – 2001. MPEG – 7 sẽ là chuẩn mô tả thông tin của rất nhiều loại đa phương tiện. Mô tả này sẽ kết hợp với chính nội dung của nó cho phép khả năng tìm kiếm nhanh và hiệu quả theo yêu cầu người dùng. MPEG-7 đặc trưng cho một tập tiêu chuẩn biểu diễn nhiều loại thông tin multimedia khác nhau. Chính vì vậy, MPEG-7 còn được gọi là " Giao thức mô tả nội dung đa phương tiện".

Tiêu chuẩn MPEG là sự kết hợp giữa nén trong ảnh và nén liên ảnh. Tức là phương pháp nén có tổn hao dựa trên sự biến đổi DCT và bù chuyển động. MPEG dùng biểu diễn màu bằng YC_rC_b.

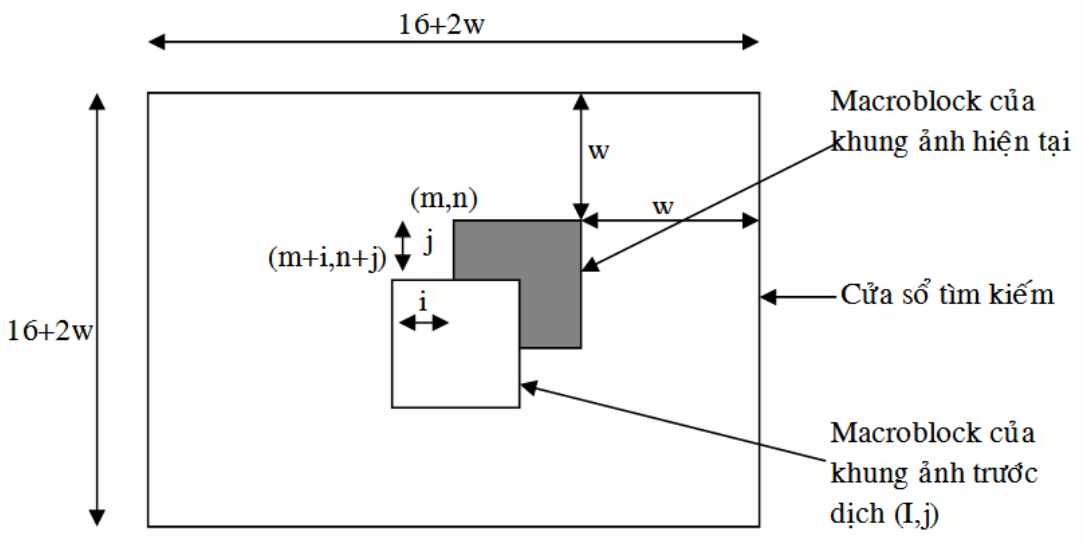


Hình 2.3.16 Hệ thống các chuẩn MPEG

2.3.3.2 Bù chuyển động trong chuẩn nén MPEG

Trong tất cả các trường hợp, khi một ảnh mã hoá dùng khung tham khảo thì luôn dùng kỹ thuật bù chuyển động để nâng cao hiệu suất nén. Sau đây, chúng ta sẽ đề cập đến phương pháp bù chuyển động.

Các phương pháp bù chuyển động: có hai cách là bù chuyển động ước đoán và nội suy. Phương pháp ước đoán bù chuyển động giả thiết ảnh hiện tại là một phép biến đổi từ ảnh trước đó, nghĩa là biên độ và hướng dịch chuyển không cần thiết phải giống ảnh trước đó. Phương pháp nội suy bù chuyển động là kỹ thuật nhiều độ phân giải: chỉ mã hoá một tín hiệu phụ với độ phân giải thấp (khoảng 1/2 đến 1/3 tốc độ khung). Ảnh có độ phân giải đầy đủ sẽ được xây dựng lại qua nội suy ảnh có độ phân giải thấp cộng thêm thành phần sửa sai. Đơn vị xử lý ảnh mà MPEG sử dụng là macroblock (MB) 16×16 điểm ảnh. Trong ảnh mã hoá nội suy, các MB có thể là loại nén trong khung hay nén liên khung. Trong kỹ thuật ước đoán chuyển động, nếu sử dụng kỹ thuật so sánh khối (BMA - Block Matching Algorithm) thì sẽ thu được các vectơ chuyển động theo tiêu chí tối thiểu hoá sai số giữa khối cần tìm vectơ chuyển động và mỗi khối ứng cử.



Hình 2.3.17 Minh họa quá trinh bù chuyển động theo giải thuật BMA

2.3.3.3 Các cấu trúc ảnh

MPEG định nghĩa các loại ảnh khác nhau cho phép sự linh hoạt để cân nhắc giữa hiệu quả mã hóa và truy cập ngẫu nhiên. Các loại ảnh đó như sau:

2.3.3.3.1 Ảnh loại I (Intra-picture)

Là ảnh được mã hóa riêng, tương tự như việc mã hóa ảnh tĩnh trong JPEG. Ảnh I chứa đựng dữ liệu để tái tạo lại toàn bộ hình ảnh vì chúng được tạo thành bằng thông tin của chỉ một ảnh và để dự báo cho ảnh B,P. Ảnh I cho phép truy cập ngẫu nhiên, tuy nhiên cho tỷ lệ nén thấp nhất.

2.3.3.3.2 Ảnh loại P (Predicted-picture)

Là ảnh được mã hóa có bù chuyển động từ ảnh I hoặc ảnh P phía trước. Ảnh P cung cấp cho hệ số nén cao hơn ảnh I và có thể sử dụng làm một ảnh so sánh cho việc bù chuyển động cho các ảnh P và B khác.

2.3.3.3.3 Ảnh loại B (Bi-directional predicted picture)

Là ảnh được mã hóa sử dụng bù chuyển động từ các ảnh I hoặc P ở phía trước và ở phía sau. Ảnh B cho tỷ lệ nén cao nhất.

2.3.3.3.4 Ảnh loại D (Dc-coded picture)

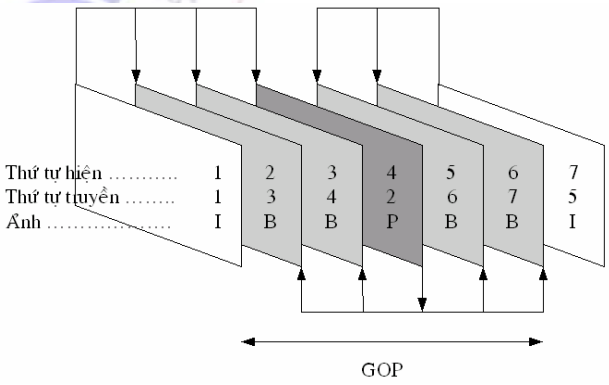
Là ảnh được sử dụng trong MPEG-1 và MPEG-4 nhưng không được sử dụng trong MPEG-2. Nó giống như ảnh I, tuy nhiên chỉ có thành phần một chiều ở đầu ra DCT được thể hiện. Nó cho phép dò tìm nhanh nhưng chất lượng ảnh thấp.

2.3.3.4 Nhóm ảnh (GOP)

Đối với chuẩn MPEG, chất lượng ảnh không những phụ thuộc vào tỷ lệ nén trong từng khuôn hình mà còn phụ thuộc vào độ dài của nhóm ảnh. Nhóm ảnh (GOP-Group of picture) là khái niệm cơ bản của MPEG. Nhóm ảnh là đơn vị mang thông tin độc lập của MPEG.

MPEG sử dụng ba loại ảnh I, B, P. Trong đó, ảnh P, B không phải là một ảnh hoàn chỉnh mà chỉ chứa sự khác biệt giữa ảnh đó và ảnh xuất hiện trước nó (đối với ảnh P) hay sự khác biệt đối với cả khuôn hình xuất hiện trước và sau nó (đối với ảnh B). Để có một khuôn hình hoàn chỉnh ảnh P và B cần có dữ liệu từ các ảnh lân cận, chính vì vậy đối với MPEG một khái niệm mới là GOP (nhóm ảnh) được sử dụng. Mỗi GOP bắt buộc phải bắt đầu bằng một ảnh hoàn chỉnh I và tiếp sau nó là một loại các ảnh P và B. Nhóm ảnh có thể mở (Open) hoặc đóng (Closed).

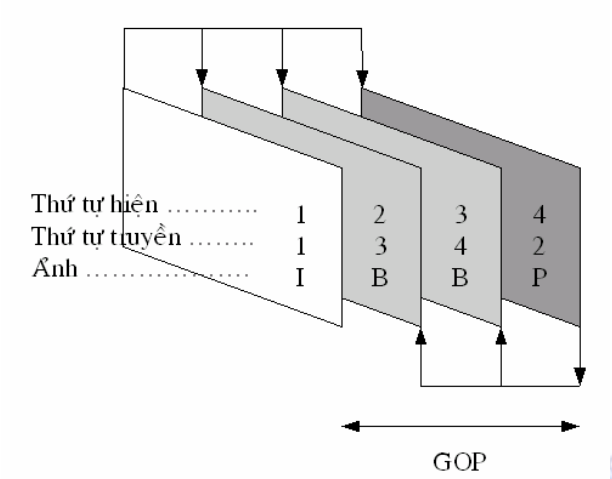
Nhóm ảnh mở luôn bắt đầu từ một ảnh I và kết thúc ở một ảnh trước ảnh trước ảnh I tiếp theo, tức là ảnh cuối cùng của GOP dùng ảnh đầu tiên của GOP tiếp theo làm ảnh chuẩn, Hình 2.3.18.



Hình 2.3.18 Cấu trúc GOB mở

Trong Hình 2.4.18, ảnh P (ảnh 4) được dự báo trước trên cơ sở ảnh I (ảnh 1). Ảnh B được dự đoán từ hai hướng, ảnh B (ảnh 2) và ảnh B (ảnh 3) được dự đoán từ hai ảnh I (ảnh 1) và ảnh P (ảnh 4). Ảnh B (ảnh 5,6) được dự đoán từ ảnh P (ảnh 4) và ảnh I tiếp theo (ảnh 6). Một điều chú ý là thứ tự truyền ảnh và hiện ảnh trên màn hình là không giống nhau.

Đối với cấu trúc khép kín (đóng), việc dự đoán ảnh không sử dụng thông tin của GOP khác. Trong trường hợp này, theo quy định, ảnh cuối cùng của một GOP bao giờ cũng là ảnh P (Hình 2.3.19).



Hình 3.3.19 Cấu trúc GOB đóng

Nhóm ảnh được xác định bởi hai thông số m và n. Thông số m xác định số khung hình P và khung hình B xuất hiện giữa hai khung hình I gần nhau nhất. Số n xác định số khung hình B giữa hai khung hình P.

Tỷ lệ nén video của MPEG phụ thuộc rất nhiều vào độ dài của GOP. Tuy nhiên, GOP dài thường gây khó khăn cho quá trình tua, định vị, sửa lỗi... Do đó tùy thuộc vào từng khâu (sản xuất, dựng hình, truyền dẫn, phát sóng v..v) mà ta chọn độ dài GOP thích hợp. Trong sản xuất hậu kỳ, nếu có yêu cầu truy cập ngẫu nhiên vào bất cứ ảnh nào, điều đó cũng có nghĩa là yêu cầu dựng chính xác đến từng ảnh, GOP đương nhiên sẽ phải chỉ có duy nhất ảnh I. Trong trường hợp này, tỷ lệ nén sẽ đạt rất thấp. Để tăng tỷ lệ nén cho truyền dẫn và phát sóng, trong GOP số lượng ảnh P, B sẽ phải tăng lên. Lúc này không cho phép việc dựng hình cũng như làm các kỹ xảo trên chuỗi hình ảnh đó. Trong trường hợp này ta có thể có GOP gồm 12 ảnh.

2.3.3.5 Cấu trúc dòng bit MPEG

Để tạo khả năng chống lỗi khi truyền tín hiệu qua kênh có nhiễu, bộ ước đoán phải được xác lập lại (reset) thường xuyên và mỗi ảnh nén trong khung hay nén ước đoán được phân đoạn thành nhiều lát nhỏ (slice) cho việc tái đồng bộ tại bộ giải mã phía thu. Cấu trúc dòng MPEG gồm 6 lớp: lớp dãy ảnh (sequence), lớp nhóm ảnh (GOP), lớp ảnh (pictrue), lớp cắt lát dòng bit (slice), lớp macroblock, lớp khối (Block). Mỗi lớp này hỗ trợ một chức năng nhất định: một là chức năng xử lý tín hiệu (DCT, bù chuyển động) hai là chức năng logic (tái đồng bộ, điểm truy xuất ngẫu nhiên). Quá trình tạo ra dòng bit MPEG là ghép kênh: kết hợp các dòng dữ liệu vào, dòng dữ liệu ra, điều chỉnh đồng bộ và quản lý bộ đệm. Cú pháp dòng MPEG bao gồm: lớp dòng bit (stream), lớp gói (pack) và lớp gói tin (packet) như trong Hình 2.3.20:

1. Khối. Khối 8x8 các điểm ảnh tín hiệu chói và tín hiệu màu dùng cho phương pháp nén DCT.

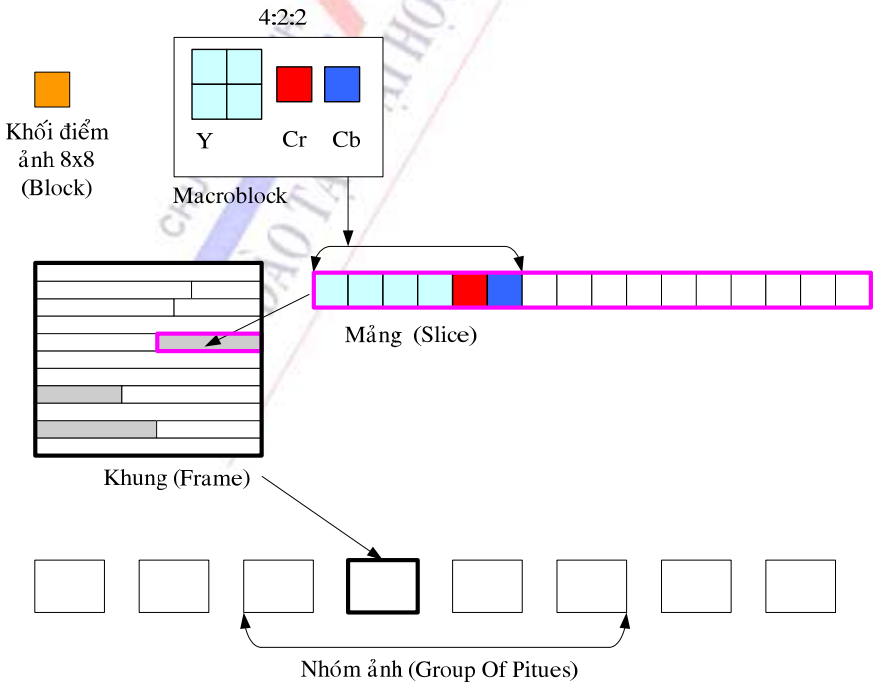
2. Tổ hợp cấu trúc khối (macroblock): một cấu trúc khối là một nhóm các khối tương ứng với lượng thông tin chứa đựng trong kích thước 16x16 điểm trên bức ảnh. Cấu trúc khối này cũng xác định lượng thông tin chứa trong đó sẽ thay đổi tùy theo cấu trúc mẫu được sử dụng. Thông tin đầu tiên trong cấu trúc khối mang dạng của nó (là cấu trúc khối Y hay Cr, Cb) và các vector bù chuyển động tương ứng.

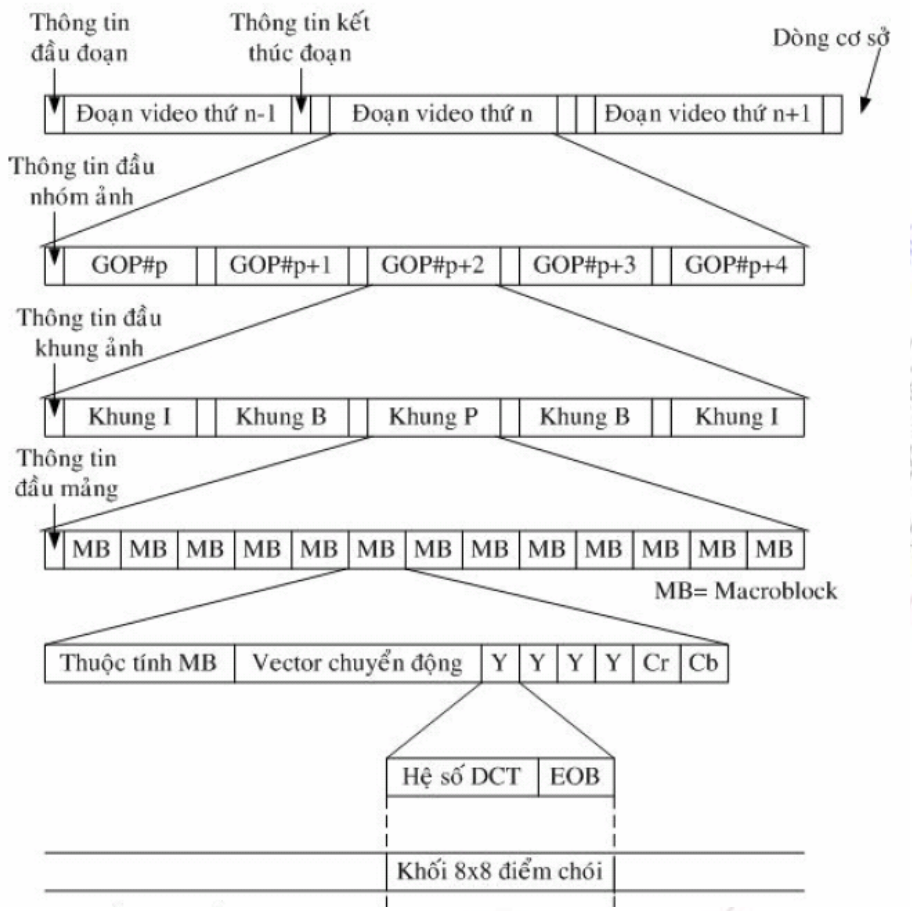
3. Mảng (Slice): mảng bao gồm một vài cấu trúc khối kề nhau. Kích thước lớn nhất của mảng có thể bao gồm toàn bộ bức ảnh và kích thước nhỏ nhất của mảng là một cấu trúc khối. Thông tin đầu của mảng chứa đựng vị trí của mảng trong toàn bộ ảnh, và hệ số cân bằng lượng tử.

4. Ảnh (Picture): lớp ảnh cho phép bộ giải mã xác định loại của ảnh được mã hóa là ảnh P, I hay ảnh B. Thông tin đầu dùng để chỉ thứ tự truyền khung để bộ giải mã có thể sắp xếp các ảnh lại theo một thứ tự đúng. Trong thông tin đầu của ảnh còn chứa các thông tin về đồng bộ, độ phân giải và phạm vi của vector chuyển động.

5. Nhóm ảnh (GOP): nhóm ảnh là tổ hợp của nhiều các khung I, P và B. Cấu trúc nhóm ảnh được xác định bằng hai tham số m và n. Mỗi một nhóm ảnh bắt đầu bằng một khung I cho phép xác định điểm bắt đầu để tìm kiếm và biên tập. Thông tin đầu gồm 25 bit chứa mã định thời và điều khiển.

6. Đoạn (chương trình) video: đoạn video bao gồm thông tin đầu, một số nhóm ảnh và thông tin kết thúc đoạn. Thông tin đầu của đoạn video chứa đựng kích thước mỗi chiều của ảnh, kích thước của điểm ảnh, tốc độ bit của dòng video số, tần số ảnh và bộ đệm tối thiểu cần có. Đoạn video và thông tin đầu tạo thành một dòng bit được mã hóa gọi là dòng cơ bản (Elementary Stream).





Hình 2.3.20 Kiến trúc dòng dữ liệu MPEG

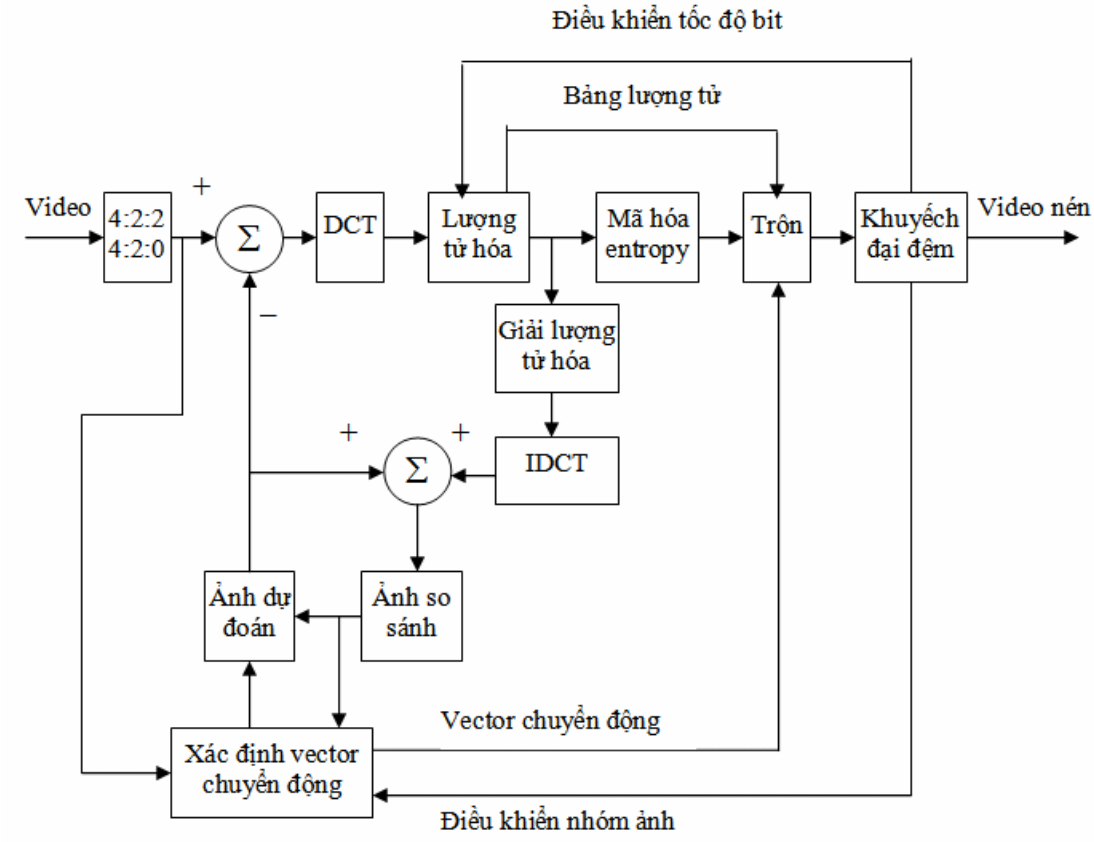
2.3.3.6 Sơ đồ khối bộ mã hóa MPEG

Quá trình nén theo chuẩn MPEG là sự kết hợp giữa nén trong ảnh và nén liên ảnh. Tín hiệu đầu vào có dạng 4:2:2 hoặc 4:2:0 được nén liên ảnh nhằm tạo ra ảnh khác biệt ở đầu ra bộ cộng. Ảnh khác biệt này sau đó được nén trong ảnh qua các bước: biến đổi DCT, lượng tử hóa, mã hóa. Cuối cùng ảnh này được trộn cùng với vector chuyển động đưa đến bộ khuyếch đại đệm sẽ thu được ảnh đã nén. Ta xét ví dụ bộ nén theo phương pháp trên, dùng ảnh I và P trong cấu trúc GOP (Hình 2.3.21).

Ảnh thứ nhất trong nhóm phải được mã hóa như ảnh loại I. Trong trường hợp này, sau khi lấy mẫu lần đầu, tín hiệu video được truyền đến khối biến đổi DCT cho các MB riêng, sau đó bộ lượng tử và mã hóa entropy. Tín hiệu ra từ bộ lượng tử hóa được đưa đến bộ lượng tử hóa ngược và biến đổi DCT ngược, sau đó được lưu vào bộ nhớ ảnh.

Trong trường hợp mã hóa ảnh loại P, mạch nén chuyển động làm việc. Trên cơ sở so sánh ảnh đang xét và ảnh trong bộ nhớ, sẽ xác định được các vector chuyển động, sau đó dự báo ảnh. Sự chênh lệch giữa ảnh đang xét và dự báo ảnh của nó được biến đổi DCT, lượng tử hóa và mã hóa entropy. Cũng như trong trường hợp các ảnh loại I, tín hiệu ra từ bộ lượng tử hóa được giải lượng tử hóa và biến đổi DCT ngược rồi cộng với ảnh dự báo đang xét và lưu vào bộ nhớ.

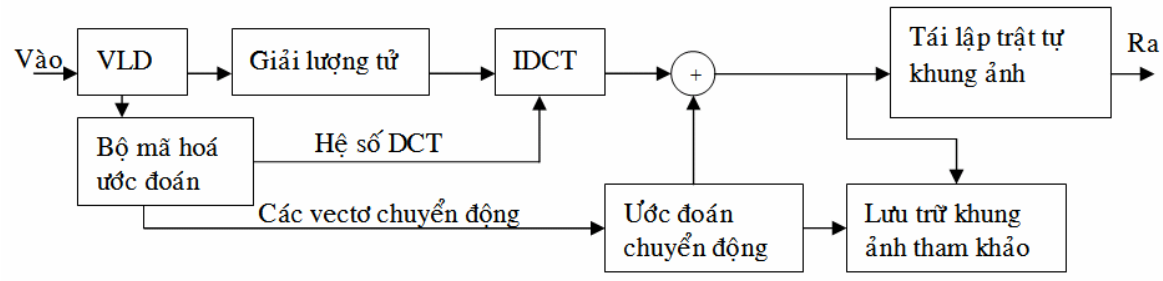
Tốc độ bit của tín hiệu video được nén không cố định, phụ thuộc vào nội dung ảnh đang xét. Ngược lại, tại đầu ra bộ mã hóa, dòng bit phải cố định để xác định tốc độ cho dung lượng kênh truyền. Do đó, tại đầu ra bộ mã hóa phải có bộ nhớ đệm đủ lớn. Bộ mã hóa phải kiểm tra trạng thái đầy của bộ nhớ đệm. Khi số liệu trong bộ nhớ đệm gần bằng dung lượng cực đại, thì các hệ số biến đổi DCT ngược được lượng tử hóa ít chính xác hơn. Trong trường hợp ngược lại, có nghĩa là bộ nhớ đệm chứa số liệu quá ít, thì việc lượng tử hóa các hệ số sẽ tăng lên.



Hình 2.3.21 Bộ mã hóa MPEG tiêu biểu

2.3.3.7 Quá trình giải mã

Quá trình giải mã, theo lý thuyết, là ngược lại với quá trình mã hóa và được minh họa trên Hình 2.3.22.



Hình 2.3.22 Bộ giải mã MPEG tiêu biểu

Chuỗi tín hiệu vào được giải mã entropy tại VLD (Variable-Length Decoder). Sau đó tách số liệu ảnh (hệ số biến đổi DCT) ra khỏi các vector chuyển động. Số liệu sẽ được giải lượng tử hóa và biến đổi DCT ngược. Trong trường hợp ảnh loại I bắt đầu ở mỗi nhóm ảnh trong chuỗi, sẽ nhận được ảnh đầu ra hoàn chỉnh bằng cách trên. Nó được lưu trong bộ nhớ ảnh và được sử dụng để giải mã các ảnh tiếp theo.

Trong trường hợp ảnh loại P sẽ thực hiện giải lượng tử và biến đổi DCT ngược với việc sử dụng các vector chuyển động và ảnh lưu vào bộ nhớ ảnh. Trên cơ sở đó xác định được dự báo ản đang xét. Ta nhận được ảnh ra sau khi cộng dự báo ảnh 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.

2.3.3.8 Tiêu chuẩn MPEG-1

Chuẩn MPEG-1 cho phép vận dụng ảnh động linh hoạt như một dạng dữ liệu của máy tính. Do đó, ta có thể truyền và nhận ảnh động thông qua máy tính và mạng viễn thông. MPEG-1 chọn các khối (MB-Macro Block) 16x16 để thực hiện quá trình bù chuyển động. Kích thước này là sự hài hòa giữa hiệu quả nén bằng cách bù chuyển động và việc lưu trữ các khối MB. Các khối MB này lại có thể chia ra làm nhiều loại khác nhau như : Intra coded, Forward prediction coded, Backward prediction coded và Bi-directional prediction coded. Dựa trên các khối MB, thông tin các vector chuyển động được lưu trữ cùng với tín hiệu khác biệt (giữa ảnh nén và ảnh dự báo). Sự khác nhau giữa vector động hiện tại và vector động truyền đi trước được mã hóa bằng mã entropy.

Tín hiệu video số MPEG-1 vào bao gồm 1 tín hiệu chói Y’ và 2 tín hiệu hiệu màu Cb và Cr. Tỷ số tần số lấy mẫu tín hiệu chói so với tần số lấy mẫu hai tín hiệu hiệu màu Cb và Cr là 2:1 theo cả hai chiều dòng và mành như một tín hiệu không chèn. Trước khi mã hóa các ảnh có thể được sắp xếp lại theo trật tự giải mã bởi vì bộ giải mã chỉ có thể giải mã được ảnh B sau khi đã giải mã ảnh I và P. Sau quá trình giải mã thì trật tự của các ảnh sẽ được sắp xếp lại như cũ.

Sau khi chọn kiểu ảnh cho một ảnh vào, bộ mã hóa sẽ đánh giá chuyển động cho mỗi khối MB của ảnh. Với mỗi khối MB này bộ mã hóa sẽ tạo ra một vector chuyển động cho 1 ảnh P và 2 vector chuyển động cho 1 ảnh B.

Tùy thuộc vào từng kiểu ảnh mà tín hiệu sai lệch (giữa ảnh nén và ảnh dự báo) được nhận dạng bằng cách tìm ra sự khác nhau giữa dự đoán bù chuyển động và dữ liệu thực sự của MB hiện tại. Tín hiệu sai lệch này được chuyển đến khối DCT 8x8 và lượng tử hóa khi đi qua bộ lượng tử hóa. Các hệ số lượng tử hóa DCT được quét theo trật tự zig-zag và mã hóa bằng mã entropy.

Một bộ điều khiển cùng với bộ đệm có nhiệm vụ điều chỉnh tốc độ dữ liệu đưa ra thông qua điều chỉnh bước lượng tử. Để có thể tạo ra được ảnh I và ảnh P trong bộ đệm trong quá trình mã hóa, thì bộ giải mã lượng tử và bộ chuyển đổi ngược DCT 8x8 được đưa vào nhằm tạo ra tín hiệu sai lệch.

2.3.3.9 Cấu trúc dòng bit và các tham số chính của chuẩn nén MPEG-1

MPEG-1 là thuật toán chỉ định nghĩa cú pháp (syntax) biểu diễn dòng bit mã hóa và giải mã. Cú pháp dòng bit được cấu tạo bằng 6 lớp : Sequence (chuỗi ảnh), GOP = Group of

Picture (nhóm ảnh), Picture (ảnh), Slice, macro block (MB), block (khối). Cấu tạo và chức năng của mỗi lớp được chỉ ra trong bảng 2.3.4. Các tham số chính của tiêu chuẩn MPEG-1 được minh họa trong bảng 2.3.5.

Bảng 2.3.4 Các thông số MPEG-1.

Lớp	Cấu tạo	Chức năng
Sequence	Gồm nhiều GOP	Dòng bit video
GOP	Gồm từ (1÷n) ảnh bắt đầu bằng ảnh I	Đơn vị truy xuất
Picture I, B, P Slice	Gồm nhiều Slice Gồm nhiều MB	Đơn vị mã hóa cơ bản Đơn vị tái đồng bộ để phục hồi lỗi
Macro Block (MB)	Với 4:2:2 gồm : 4 block Y, 1 block Cr và 1 block Cb	Đơn vị bù chuyển động
Block	Gồm 8x8 pixel	Đơn vị tính DCT

Bảng 2.3.5. Tham số theo tiêu chuẩn nén MPEG-1.

Tham số	Đặc điểm
Tín hiệu mã hóa	Y và Cr, Cb
Cấu trúc lấy mẫu	4:2:0
Kích thước ảnh tối đa(điểm ảnh x điểm ảnh)	4095x4095
Biểu diễn mẫu	8 bit
Độ chính xác của quá trình lượng tử hóa và biến đổi DCT	9 bit
Phương pháp lượng tử hóa hệ số DC	DPCM tuyến tính
Cấu trúc khối trong quá trình lượng tử hóa thích nghi	16x16 bit
Độ chính xác cực đại của hệ số DC	8 bit
Biến đổi VLC	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	½ điểm ảnh
Tốc độ khi nén	1,85 Mbps cho nén tham số 100 Mbps cho dòng đầy đủ tham số

Phương pháp nén MPEG-1 cho phép truy cập ngẫu nhiên các khung hình video, tìm kiếm thuận và nghịch trên dòng tín hiệu nén, biên tập và phát lại trên dòng tín hiệu nén. MPEG-1 là tập con của MPEG-2, nên tất cả các bộ giải mã MPEG-2 đều có thể giải mã được dòng tín hiệu MPEG-1.

2.3.3.10 Tiêu chuẩn MPEG-2

2.3.3.10.1 Giới thiệu về chuẩn MPEG-2

MPEG-2 là dự án giai đoạn 2 của ủy ban ISO/PEC MPEG. MPEG-2 hướng tới các ứng dụng rộng rãi hơn và có tốc độ bit cao hơn MPEG-1, bao gồm điện tử viễn thông và truyền hình số thế hệ kế tiếp. Nội dung kỹ thuật đã được đúc kết vào 11/1993 thành dự thảo ISO/IEC 13818 tên gọi " Mã hóa chung ảnh động và audio đi kèm" gồm ba phần chính: Hệ thống; Video; Thử nghiệm. MPEG-2 được tiến hành ngay sau MPEG-1, nhằm hỗ trợ việc truyền video số tốc độ bit lớn hơn 4 Mbps, bao gồm các ứng dụng DSM (phương tiện lưu trữ số), các hệ thống TV hiện đại (PAL, NTSC, SECAM), cáp, thu lượm tin tức điện tử, truyền hình trực tiếp từ vệ tinh, EDTV (truyền hình mở rộng), HDTV (truyền hình có độ phân giải cao) v...v.

MPEG-2 là chuẩn nén video có tổn thất. Công ty Nethold's Multichoice đã truyền 20 kênh truyền hình số cho Bỉ, Hà Lan, Luxembourge, Scandinavia, Trung Đông, Châu Phi, ... qua vệ tinh Pan Amsat vào tháng 10/1995. Hệ thống sử dụng trên một triệu bộ giải mã MPEG-2 set-top của Phillips, Panasonic, Pace. Mạng truyền hình Dish của Echostar có kế hoạch truyền 150 kênh truyền hình số. Cả châu Âu (DVB), Mỹ (ATV), và nhiều hãng khác trên thế giới (Galaxy, Shinawatra Satellite, ...) dùng MPEG-2 trong các hệ truyền hình có độ phân giải cao để có thể phát sóng truyền hình số trên mặt đất.

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).

Phần 1 đưa ra cấu trúc kết nối phức tạp giữa dữ liệu audio và video và đồng bộ thời gian thực. Phần 2 đưa ra cách mã hóa tín hiệu video và cũng chỉ ra quá trình giải mã để tái tạo lại

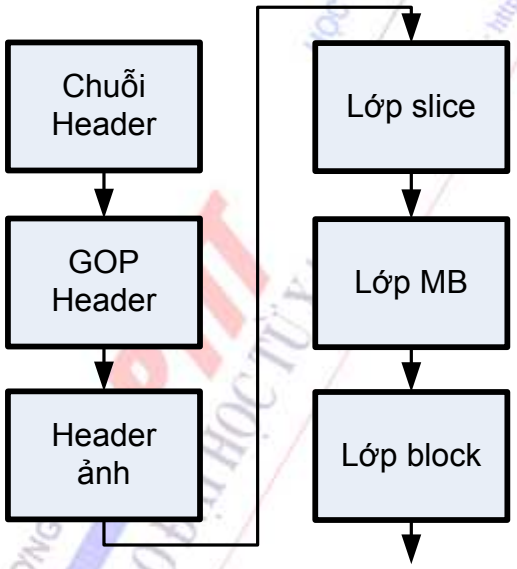
các khung ảnh. Phần 3 là cấu trúc mã hóa của tín hiệu audio và phần 4 là các hệ thống kiểm tra các yêu cầu đặt ra của MPEG-2.

MPEG-2 hoạt động hoàn toàn dựa trên các cơ sở MPEG-1, tuy nhiên có thêm một vài yêu cầu tạo thành một hệ thống đầy đủ cho các dịch vụ nghe nhìn:

- Hỗ trợ xử lý file hoặc frame.
- Áp dụng vào nhiều lĩnh vực từ tốc độ bit rất thấp đến rất cao (từ chất lượng ảnh rất thấp đến rất cao). MPEG-2 đưa ra các dạng thức cơ bản ứng với 6 profiles và một số level.
- MPEG-2 có tính tương hợp (compatibility) và tính co dãn (scalability) cao theo không gian và thời gian.

2.3.3.10.2 Cấu trúc dòng bit video MPEG-2

Một trong những khác biệt chính giữa hai tiêu chuẩn MPEG-2 và MPEG-1 là ở chỗ MPEG-2 có khả năng xử lý chuỗi video xen kẽ, ví dụ như dạng thức ITU-R601. Sơ đồ mã hóa có thể thích nghi với sự lựa chọn field hoặc frame, trong đó MPEG-1 chỉ có một mode cố định. Hình 2.3.23 chỉ ra cấu trúc dòng bit video MPEG-2.



Hình 2.3.23 Cấu trúc dòng bit vdeo MPEG-2

- Chuỗi video được mã hóa bắt đầu bằng Sequence Header, sau đó là chuỗi mở rộng (nếu có) và các nhóm ảnh. Nếu phần chuỗi mở rộng (Sequence extension) không được xác định (không có mã báo có thành phần mở rộng), các lớp tiếp theo khi đó sẽ thực hiện một quy trình giống như MPEG-1 và đó là tương hợp thuận. Khi có thêm phần mở rộng thì phải có thêm các đặc tính mở rộng để mã hóa hữu hiệu hơn.
- Header của nhóm ảnh (GOP) có chức năng tương tự như header của MPEG-1. Các thông số quan trọng dùng để mã hóa ảnh mở rộng được định nghĩa trong extension header của ảnh. Vì có hai loại ảnh, liên tục và xen kẻ nên cấu trúc ảnh cần phải được xác định rõ field trên hay field dưới hoặc frame.

2.3.3.10.3 Đặc tính và các mức trong MPEG-2

Nén MPEG-2 có một chuỗi các mức (Level) và đặc tính (Profile), được dùng cho nhiều ứng dụng khác nhau.

Cấu trúc tín hiệu số trong tiêu chuẩn MPEG-2 rất phức tạp. Việc sử dụng tiêu chuẩn MPEG-2 không phải lúc nào cũng cần thiết hoặc có ý nghĩa. Vì thế dẫn đến việc phân chia cấu trúc thành các tập con gọi là profiles. Có 6 định nghĩa về profile:

- Simple profile (profile đơn giản): có số công cụ thấp nhất và sử dụng tốc độ bit thấp và không dùng B frame. Nó tương đương với đặc điểm kỹ thuật MPEG-1, phù hợp với các ứng dụng low-delay bởi không cần thiết sắp xếp lại các frame.

- Main profile (profile chính): có tầm ứng dụng khá rộng. Nó rất quan trọng vì đáp ứng được độ phân giải đối với truyền hình quy ước. Nó cho chất lượng ảnh tốt hơn với cùng một tốc độ bit so với low profile nhưng thời gian trì hoãn khi mã và giải mã tăng lên.

- SNR profile scalable (profile phân cấp theo SNR): có các công cụ của main profile và cho phép phân cấp theo tỉ số tín hiệu trên tạp âm. Tính phân cấp theo tỉ số tín hiệu trên tạp âm có nghĩa là chất lượng hình ảnh và tỉ số tín hiệu trên tạp âm có tính thõa hiệp. Chuỗi ảnh có thể chia thành hai phân lớp phân biệt nhau về chất lượng. Các lớp thấp bao gồm ảnh có chất lượng cơ sở, lớp cao bao gồm các lớp hoàn thiện hơn đối với lớp thấp hơn, cho phép khôi phục cùng ảnh đó nhưng chất lượng tốt hơn. Lớp thấp hơn, ví dụ chứa tín hiệu video theo chuẩn 4:2:0, còn lớp cao hơn với tín hiệu video trong chuẩn 4:2:2. Có thể mã hóa kênh khác nhau cho các lớp riêng. Trong trường hợp này, lớp dưới có tín hiệu video chất lượng thấp hơn (ví dụ lượng tử hóa với độ chính xác thấp). Còn lớp cao hơn thì lớp bảo vệ cho phép khôi phục lại tín hiệu video với độ chính xác đầy đủ thông số kênh truyền hoặc bộ mã hóa.

- Spatially Scalable profile (phân cấp theo không gian): tương tự với SNR profile nhưng thêm vào lớp cơ bản lớp nâng cao chất lượng độ phân giải ảnh (Picture Resolution Enhancement layer). Tính phân cấp theo không gian có nghĩa là có sự thõa hiệp đối 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 ví dụ như truyền hình tiêu chuẩn, lớp cao hơn bao gồm ảnh có độ phân giải cao hơn ví dụ như truyền hình độ phân giải cao (HDTV).

- High profile (profile cao): gồm các đặc điểm của spatial profile thêm vào cấu trúc lấy mẫu 4:2:2. Nó bao gồm toàn bộ các công cụ của spatially scalable profile cộng thêm khả năng mã hóa các tín hiệu màu khác nhau cùng một lúc. Nó được dự định dùng cho HDTV, cho phép các bộ thu HDTV giải mã cả hai lớp để hiển thị một ảnh HDTV. “High profile” là một hệ thống hoàn hảo được thiết kế cho toàn bộ các ứng dụng mà không hạn chế tốc độ bit.

- 4:2:2 profile: tương tự MP, nhưng cho phép một tốc độ bit cao hơn. Nó gia tăng kích thước ảnh dọc lên 576 lines với chuẩn quét 625/50 và 512 lines với chuẩn quét 525/60.

Vấn đề hạn chế các mức có liên quan đến độ phân giải cực đại của ảnh. Có 4 mức hạn chế sau :

- Low level (mức thấp): ứng với độ phân giải của MPEG-1, có nghĩa là bằng ¼ độ phân giải truyền hình tiêu chuẩn.
- Main level (mức chính): độ phân giải của truyền hình tiêu chuẩn.
- High – 1440 level (mức cao 1440): độ phân giải của HDTV với 1440 mẫu/dòng.
- High level (mức cao): độ phân giải HDTV với 1920 mẫu/dòng.

Bảng 2.3.6. Bảng thông số chính profile và level của tín hiệu chuẩn MPEG-2.

Profile / Level	Đơn giản (Simple)	Chính (Main)	Phân cấp theo SNR	Phân cấp theo không gian	Cao (High)
Thấp (Low)		4:2:0 352x288 4 Mbps	4:2:0 352x288 4Mbps I, P, B		
Chính (Main)	4:2:0 720x576 15 Mbps I, P	4:2:0 720x576 15 Mbps I, P, B	4:2:0 720x576 15 Mbps I, P, B		4:2:0 720x576 20 Mbps I, P, B
Cao – 1440 (High–1440)		4:2:0 1440x1152 60 Mbps I, P, B		4:2:0 1440x1152 60 Mbps I, P, B	4:2:0,4:2:2 1440x1152 80 Mbps I, P, B
Cao (High)		4:2:0 1920x1152 80 Mbps I, P, B			4:2:0,4:2:2 1920x1152 100 Mbps I, P, B

Kết hợp 4 level và 5 profile ta được tổ hợp 20 khả năng và hiện nay đã có 11 khả năng được ứng dụng như Bảng 2.3.6 (theo tài liệu của Tektrronic). Trong các ô của Bảng 2.4.6, lần lượt từ trên xuống là: tỷ lệ lấy mẫu (4:2:0 hoặc 4:2:2); dòng dưới ghi điểm ảnh theo chiều

ngang x theo chiều dọc; dòng dưới nữa là vận tốc cao nhất của dòng dữ liệu sau khi nén; dòng cuối cùng là các loại ảnh sử dụng để nén.

2.3.3.10.4 Ứng dụng MPEG-2 trong nén tín hiệu video

❖ Các tính chất nén tín hiệu video

Tính chất nén tín hiệu video (hoặc giảm tốc độ bit của video BRR – Bit rate reduction) là sự kết hợp nhiều yếu tố khác nhau :

- Tỉ lệ nén : tỉ lệ nén từ 2:1 đến 150:1, tùy thuộc vào chất lượng ảnh yêu cầu cho từng ứng dụng.

- Chất lượng ảnh : chất lượng ảnh cao thường dùng cho khâu xử lý ảnh, trong khâu hậu kỳ (dựng hình); giảm hơn trong khâu lấy tin (news), truyền dẫn phát sóng.

- Khả năng tạo nhiều lần : Trong quá trình sản xuất hậu kỳ, truyền dẫn phát sóng; tín hiệu video gốc phải đi qua nhiều công đoạn, nén và giải nén.

- Đối xứng/ không đối xứng : với sơ đồ nén đối xứng, số lượng xử lý ở phần mã hóa và giải mã giống nhau. Sơ đồ MPEG-2 là không đối xứng vì các công đoạn giải mã ít hơn so với mã hóa.

- Trễ giữa mã hóa và giải mã : độ trễ này phụ thuộc vào cấu trúc và độ phức tạp của bộ mã hóa, kích thước GOP và chuỗi GOP. Trong truyền hình, độ trễ tổng cộng có thể chấp nhận được là <1ms cho trường hợp phỏng vấn trực tiếp. Trong truyền dẫn phát sóng thì vấn đề này ít khắc khe hơn.

- Khả năng dựng hình : dựng hình với độ chính xác 1 frame là yêu cầu cao trong khâu hậu kỳ. Hiện tại, trong khâu hậu kỳ phải giải mã nhiều frames (I, B, P) và mã hóa lại sau khi cấy một đoạn mới vào. Do có thể thay đổi chiều dài GOP xuống còn ảnh I, cho nên MPEG-2 cho phép dựng hình với độ chính xác từng frame.

- Độ phức tạp và giá thành : có một sự thỏa hiệp giữa kỹ thuật xấp xỉ chuyển động có hiệu quả (nâng cao hiệu quả nén cao) và giảm độ phức tạp và giá thành của các chip xử lý.

2.3.3.11 Tiêu chuẩn MPEG-4

MPEG-4 bao gồm 2 phần là version 1 và version 2. Bắt đầu từ năm 1993 và hình thành các đề nghị vào tháng 7 năm 1995. Các đề nghị về audio và video được đánh giá bởi các chuyên gia và đưa ra bản thảo vào tháng 11 năm 1997 và trở thành tiêu chuẩn quốc tế ISO/IEC vào năm 1999. Năm 2000 MPEG-4 được bổ xung và nâng cấp lên thành các version 3 và 4.

Đặc điểm chính của MPEG-4 là mã hóa video và audio với tốc độ bit rất thấp. Thực tế tiêu chuẩn đưa ra với 3 dãy tốc độ bit

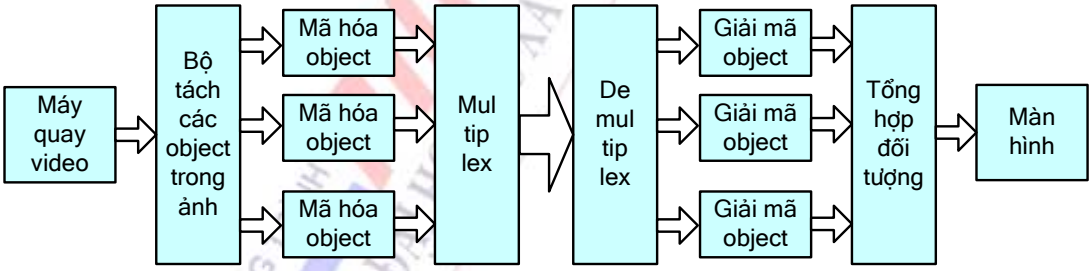
- dưới 64 kbps
- 64 đến 384 kbps
- 384 Kbps đến 4 Mbps

Đặc điểm quan trọng của chuẩn MPEG-4 là cho phép khôi phục lỗi tại phía thu, vì vậy chuẩn nén này đặc biệt thích hợp đối với môi trường dễ xãy ra lỗi như truyền dữ liệu qua các thiết bị cầm tay. Những profile và level khác trong MPEG-4 cho phép sử dụng tốc độ bit lên đến 38.4 Mbps và việc xử lý chất lương studio cần các profile và level lên đến 1.2Gbps.

MPEG-4 là chuẩn quốc tế đầu tiên dành cho mã hoá các đối tượng (object) video. Với độ linh động và hiệu quả do mã hoá từng đối tượng video, MPEG-4 đạt yêu cầu ứng dụng cho các dịch vụ nội dung video có tính tương tác và các dịch vụ truyền thông video trực tiếp hay lưu trữ. Trong MPEG-4, khung ảnh của một đối tượng video (hay còn gọi là phẳng đối tượng video) được mã hoá riêng lẽ. Sự cách ly các đối tượng video như vậy mang đến độ mềm dẻo hơn cho việc thực hiện mã hoá thích nghi làm tăng hiệu quả nén tính hiệu. Mặc dù tập trung vào những ứng dụng tốc độ bit thấp nhưng MPEG-4 cũng bao gồm cả studio chất lượng cao và HDTV

Các đối tượng khác nhau trong một cảnh gốc có thể được mã hóa và truyền đi riêng biệt như là video object và audio object và được kết hợp trở lại tại bộ giải mã. Các loại object khác nhau sẽ được mã hóa với những kỹ thuật khác nhau và với các công cụ phù hợp nhất. Những object khác nhau có thể được tạo ra một cách độc lập và trong một vài trường hợp một cảnh có thể phân tích riêng thành object nền (background) và object cận cảnh. Ví dụ: đoạn video quay trận bóng đá được xử lý để tách riêng quả bóng ra khỏi cảnh sân cỏ. Background (cảnh không có quả bóng) sẽ được truyền đi và ai cũng có thể thấy game để thu hút khan giả nhưng chỉ những người có trả tiền mới thấy quả bóng.

Hình 2.3.24 cho thấy các khái niệm chung về quá trình mã hóa và giải mã các đối tượng độc lập trong chuẩn nén MPEG-4.



Hình 2.3.24 Nguyên lý mã hóa tín hiệu hình ảnh trong MPEG-4

Như đã biết không có phương pháp mã hóa nào có thể gọi là tối ưu hoàn toàn. DCT và phép lương tử chỉ tối ưu đối với các ảnh có băng thông giới hạn và các ảnh có mức chói thay đổi chậm nhưng sẽ không tối ưu với nhiều chi tiết ảnh có kích thước nhỏ thường xuất hiên trên đoạn video. Một ví dụ đơn giản nếu một cảnh có xen các dòng chữ (text) thì hệ thống thông thường sẽ xem các chữ như là các chi tiết ảnh thường, do đó sau khi mã hóa bằng MPEG-1 hay MPEG2,các chi tiết nhỏ được thêm vào với cách thức như trên sẽ tạo ra tín hiệu năng lượng có tần số cao và các cạnh của dòng chữ sẽ không được mã hóa tốt bởi DCT

Do đó việc thêm chữ vào ảnh hưởng rất lớn đến hiệu suất mã hóa video. Tuy nhiên có thể mã hóa các chữ theo một cách đơn giản như các ký tự thuộc mã ASCII, vị trí, font, kích thước, màu, thông tin về không gian có thể được thêm vào với số bit tương đối nhỏ. Nhưng để làm điều này bộ giải mã cần phải có khả năng tạo ra các title từ những thông tin được cung

cấp và khóa các title này khi qua bộ giải mã video trước khi hiển thị. Việc giải mã luồng bit video MPEG-4 yêu cầu bộ giải mã có nhiều cơ chế giải mã và khả năng thực hiện các hoạt động đa hợp. Trong MPEG-4 có thể truyền nhiều luồng text và việc chọn ở bộ giải mã luồng nào trong số những luồng trên kết hợp với video. Việc lựa chọn này có thể do người xem quyết định hoặc do các thông tin khác được truyền trong luồng bit.

Ba đặc tính rất quan trọng của MPEG-4 là:

- Nhiều object có thể được mã hóa với các kỹ thuật khác nhau và kết hợp lại ở bộ giải mã
- Các object có thể là các cảnh có được từ camera hay tự tạo như text
- Các thông tin trong luồng bit có thể hiển thị nhiều dạng khác nhau từ cùng một luồng bit (tùy theo lựa chọn người xem chẳng hạn như ngôn ngữ)

MPEG-4 cho khả năng mã hóa video và audio hơn hẳn MPEG-2 cũng như khả năng khôi phục lỗi. Tuy nhiên sức mạnh thật sự của MPEG-4 là các ứng dụng mới mà có thể xây dựng dựa vào việc mã hóa độc lập các object cho hiệu suất mã cao hơn, và việc tách riêng các object cho phép tương tác các object với nhau đặc biệt là các chương trình giáo dục và các trò chơi. Và cũng do khả năng tách biệt các object mà có thể thay đổi tỷ lệ tạm thời chẳng hạn như vẫn duy trì độ phân giải của các object cận cảnh quan trọng nhưng giảm ảnh phong xuống tốc độ thấp hơn nếu hệ thống sử dụng có băng thông bị hạn chế hoặc thiếu tài nguyên (bộ nhớ, tốc độ tính)

Tuy nhiên cũng có một số nhược điểm là bộ giải mã phải có khả năng giải mã hết tất cả các luồng bit mà nó hỗ trợ và có khả năng kết hợp. Do đó phần cứng của bộ giải mã MPEG-4 phức tạp hơn so với bộ giải mã MPEG-2. Và ngày nay thì càng có nhiều bộ mã thực hiện giải mã bằng phần mềm nhưng bộ giải mã bằng phần cứng có thể bị hạn chế về khả năng linh hoạt

2.3.3.11.1 Video trong MPEG-4

Trước khi tìm hiểu kỹ thuật nén video trong MPEG-4 cần tìm hiểu cấu trúc của một cảnh video được MPEG-4 định nghĩa. Một cảnh tiêu biểu bao gồm phong ảnh (background) một hoặc nhiều đối tượng cận ảnh (foreground) chẳng hạn như đồ vật, một hoặc nhiều người và một vài phần tử đồ họa. Trong MPEG-1 và 2 một cảnh được lấy mẫu một lần cho một khung và tạo ra các bitmap sẽ được mã hóa. MPEG-4 cũng làm việc giống như vậy nhưng nó có thể giải quyết từng đối tượng riêng rẽ. Để đơn giản hơn có thể không xét đến các đồ vật như vậy ngoài các thành phần đồ họa cảnh bao gồm background, một người được xem là foreground. Ví dụ: người dự báo thời tiết đứng trước nền màu xanh biển hay xanh lá cây và một nền (background) khác chẳng hạn như bản đồ thời tiết gọi là "chroma keyed". Trong studio ảnh một người đứng trước nền màu sẽ được xử lý để loại bỏ nền màu và tạo thành "key signal" hay alpha channel diễn tả hình dạng của người cận cảnh. Thông tin về hình dạng người sẽ được kết hợp với thành phần cảnh. Nơi người đứng thì cảnh nền được thay thế bằng ảnh người và những nơi khác của ảnh nền thì không thay đổi. Trong thuật ngữ của MPEG-4 thì người cận ảnh được xem là đối tượng video (video object) được tương trưng bởi hai phần tử là ảnh video của người gọi là "texture" và key signal hay alpha channel được xem là shape.

2.3.3.11.2 Cấp độ của video MPEG-4

Trước tiên object phải được lấy mẫu. Hầu hết các object được lấy mẫu trong khoảng thời gian không đổi (gọi là frame) và mỗi thời gian lấy mẫu đươc gọi là video object plane (VOP). Như vậy mỗi object trong một cảnh được tượng trưng bởi 1 chuỗi các VOP ngoại trừ các object tĩnh có thể dùng một VOP.

VOP bao gồm dữ liệu texture và thông tin về đường nét (shape) có dạng chữ nhật hoặc dữ liệu đường nét phức tạp kết hợp với object. VOP cũng giống như các frame của các version trước của MPEG có thể được mã hóa với intradata hoặc sử dụng bù chuyển động.

Tiếp theo là nhóm các VOP với nhau thành GOV (Group of video object planes). GOV tương tự như GOP (group of pictures) của MEPG trước và cung cấp điểm trong luồng bit mà VOP được mã hóa độc lập với các VOP khác và như thế nó cung cấp các điểm truy xuất ngẫu nhiên trong luồng bit

VOL (Video object layer) cho phép thay đổi tỷ lệ mã hóa chuỗi các VOP hoặc GOV. Nhiều VOL tương ứng với nhiều tỷ lệ của chuỗi (VOP hoặc GOV) và mỗi tỷ lệ phù hợp với một tập các tài nguyên có thể thông thường giới hạn băng thông hoặc giới hạn khả năng tính toán. Mức video object (VO) bao gồm mọi thành phần trong luồng bit mô tả đối tương video đặc biệt.

Cuối cùng là Video session (VS) là mức video cao nhất của cảnh MPEG-4 bao gồm tất cả đối tượng video cả tự nhiên và tự tạo trong một cảnh.

2.3.3.11.3 Mã hóa đường nét (shape)

Có hai loại đường nét với đối tượng video trong MPEG-4 là chữ nhật và tùy ý. Dạng chữ nhật chỉ đơn thuần là chỉ phạm vi của ảnh nên ít quan trọng. Tuy nhiên nó vẫn được dùng để tăng tính linh hoạt trong các chuẩn trước. Chẳng hạn trong MPEG-2 phạm vị của ảnh được mã hóa trong phần header của luồng bit. Trong MPEG-4 kích thước chữ nhật của đối tượng video nền đơn giản là có thể so sánh nhưng cũng có thể có các đối tượng chữ nhật khác trong cùng một session như ảnh trong ảnh (picture in picture).

Đường nét cũng tượng trưng cho đối tượng video và ở bất kỳ điểm nào trong mặt phẳng ảnh nó xác định có đối tượng nào được kết hợp với nó thì có thể nhìn thấy được. Đường nét dạng chữ nhật được gọi là mask và có kích thước có thể thay đổi theo kích thước ngang và dọc lớn nhất của đối tượng. Cả hai kích thước ngang và dọc của mask là bội số của 16 pixel.

Đường nét tùy ý có thể được mã hóa như dữ liệu nhị phân hoặc dữ liệu xám. Đường nét nhị phân là dạng đơn giản nhất chỉ ra đối tượng là rõ ràng hay không rõ ràng (thấy được hoặc không thấy) ở bất kỳ điểm đã cho.

2.3.3.11.4 Mã hóa texture

Mã hóa texture, là thuật ngữ trong MPEG-4 tương ứng với việc mã hóa dữ liệu ảnh chuyển động, dựa vào mã hóa MPEG-2 có mở rộng và cải tiến. Các đối tượng video có thể được mã hóa với I-VOP, P-VOP, B-VOP. Hầu hết các profile MPEG-4 đều sử dụng tiêu chuẩn 4:2:0 và YUV để mô tả đối tượng video texture.

Trong MPEG-4 không phải tất cả các đối tượng video có cùng kích thước và việc mã hóa texture chỉ cần thiết ở những khu vực là một phần của đối tượng. Đối với những đối tượng chữ nhật thì đơn giản chọn kích thước là bội số của 16 pixel (một macroblock) theo mỗi hướng và tất cả các macroblock sẽ được xử lý. Đối với các đối tượng có đường nét phức tạp thì đường biên (boundary) được định nghĩa là tín hiệu đường nét (shape signal). Phạm vi của đối tượng cũng được định nghĩa bởi dãy hình chữ nhật các macroblock nhưng mã hóa texture được thực hiện đối với toàn bộ các macroblock trên đường biên hoặc phần bên trong đường biên đối tượng

I-VOP được mã hóa như khung I trong MPEG-2. MPEG-4 sử dụng bộ dự đoán thích ứng đối với các giá trị DC. Bộ dự đoán cũng xác định gradient độ sáng ngang và dọc và dự đoán giá trị DC từ các khối ở trên và bên trái theo hướng của gradient nhỏ hơn

Sự tương quan của các ảnh ngoài việc có lợi cho dự đoán hệ số DC còn giúp việc mã hóa các hệ số AC. Những vùng texture giống nhau sẽ tạo ra một dãy các hệ số AC giống nhau sau khi biến đổi DCT. Các hệ số AC quan trọng nhất tương trưng cho năng lượng lớn nhất của texture giống nhau rất nhiều (có lợi cho quá trình mã hóa). Các hệ số này thông thường là các hệ số khác zero trong hàng đầu tiên hoặc cột đầu tiên, chúng thường được lượng tử hóa với mức độ chính xác cao nhất. Trong MPEG-4 các hệ số AC của hàng đầu tiên hoặc cột đầu tiên được dự đoán từ các khối ngay ở trên và bên trái.

Việc lượng tử các hệ số cũng tương tự như phương pháp sử dụng trong MPEG-2 nhưng cơ chế quét các hệ số và mã hóa với chiều dài từ mã thay đổi thì được cải tiến hơn.

Các phương pháp được chọn để đọc hệ số ra được xác định dựa vào quá trình dự đoán DC. Khi không có dự đoán DC thì quét zigzag như trong MPEG-2 được sử dụng. Nếu hệ số DC được dự đoán từ khối phía bên trái thì sử dụng quét dọc luân phiên (Alternate-vertical scanning) là hệ thống quét sẽ đọc theo chiều dọc trước tiên. Tuy nhiên nếu hệ số DC được dự đóan từ các khối ở trên thì chọn quét ngang luân phiên (Altrenate Horziontal scan).

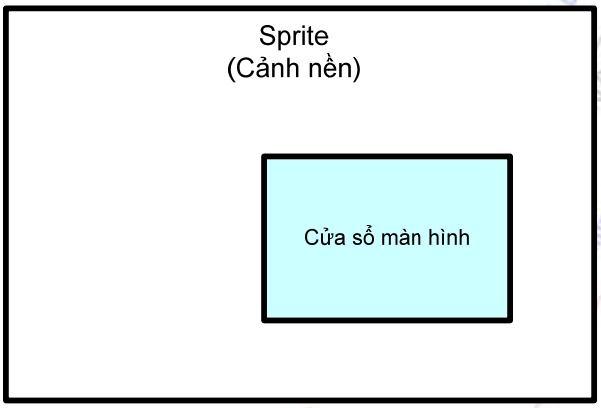
Để cải thiện hiệu quả của bộ mã hóa có chiều dài thay đổi (VLC) trong MPEG-4 dùng hai bảng VLC khác nhau.

2.3.3.11.5 Mã hóa đường biên

Khả năng mã hóa những đối tượng có đường nét tùy ý thường gặp những tình huống đặc biệt ở đường biên của đối tượng. Khi đó, các khối nằm bên ngoài đường biên thì không cần mã hóa texture. Nhưng tất cả những khối bên trong đường biên được mã hóa theo các kỹ thuật đã trình bày. Mã hóa texture cần cho các khối đường biên nhưng trong các khối đó chỉ có một phần thuộc đối tượng. Trước khi mã hóa đường biên, người ta thực hiện biện pháp "đệm" (padding) các khối này. Những pixel không phải là thành phần của đối tượng được gán cho những giá trị bằng nhau và bằng với giá trị trung bình của tất cả các pixel trong phần đối tượng. Giá trị của các pixel bên trong đối tượng không thay đổi. Việc thay đổi giá trị bên ngoài đối tượng không ảnh hưởng đến kết quả sau cùng bởi vì các pixel này không được hiển thị. Quá trình trên được xem là làm giảm thiểu năng lượng của các hệ số khi biến đổi DCT

2.3.3.11.6 Sprites

MPEG-4 có một loại đối tượng khác thường được dùng làm cảnh nền là sprite. Sprite là đối tượng video thường có kích thước lớn hơn màn hình hiển thị. Sprite là đối tượng được sử dụng liên tục trong một cảnh (tương tự như cảnh nền tĩnh). Thông thường một cảnh của game bao gồm cảnh nền và một số đối tượng nhân tạo di chuyển theo kịch bản của game và hành động của người chơi. Trong quá trình hành động cảnh được nhìn thấy chỉ là một vùng nhỏ trong cảnh nền, vùng này là thành phần của cùng một ảnh tĩnh (Hình 2.3.25). MPEG-4 cung cấp khả năng truyền toàn bộ cảnh nền như sprite và khả năng tạo cảnh khác nhau bằng cách truyền các thông tin cropping và wraping để xác định phần sprite sẽ được hiển thị ở một thời điểm nhất định. Sau khi sprite được truyền đi thì chỉ có thông tin cropping/wraping cho sprite và các đối tượng cận ảnh (foreground) cần được truyền. Trong game điển hình mỗi phần của sprite có thể được sử dụng nhiều lần vì thế lượng dữ liệu cần truyền sẽ giảm đáng kể.



Hình 2.3.25 Cảnh nền (sprite) được truyền đi có kích thước lớn hơn khả năng hiển thị của màn hình

Việc truyền toàn bộ sprite ngay khi bắt đầu chương trình có thể rất hiệu quả nhưng sẽ làm tăng băng thông và thời gian truyền trước khi hoạt động có thể bắt đầu. MPEG-4 sử dụng phương pháp sau để tránh vấn đề này. Sprite có thể truyền từng phần khi cần. Một phần sprite cần thiết tại thời điểm tức thời sẽ được truyền đi. Tất cả các cảnh sprite sẽ được lưu trữ ở bộ giải mã như là một phần của sprite. Theo phương pháp khác, sprite có thể được mã hóa liên tiếp và truyền đi toàn bộ với độ phân giải thấp và độ phân giải cao hơn sẽ được truyền sau.

Sprite được mã hóa như tín hiệu chói với hai thành phần màu như trong MPEG trước và luôn được mã Intra bởi vì bản chất của ảnh là tĩnh.

2.3.3.11.7 Animations

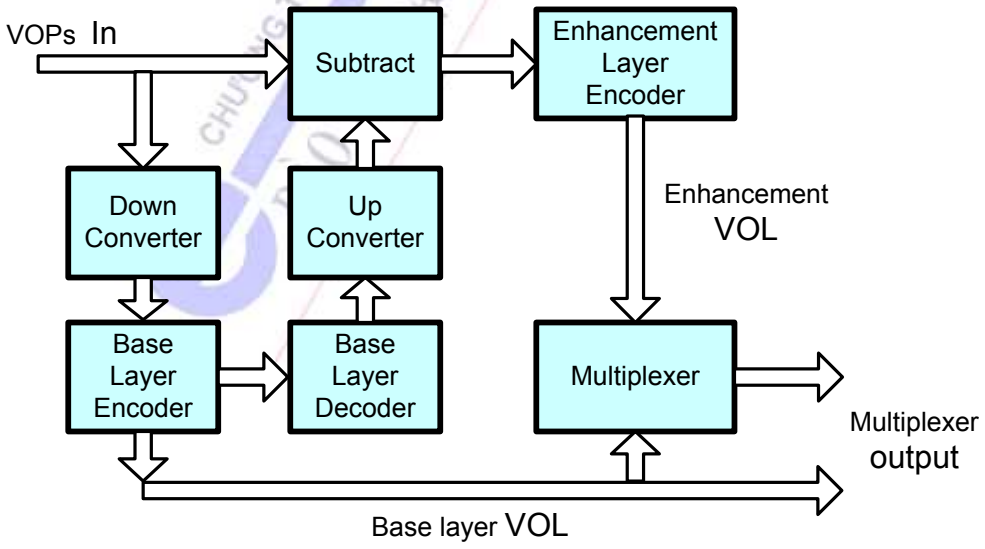
Như đã biết một trong số những điểm mạnh của MPEG-4 là khả năng truyền cả đối tượng tự nhiên cũng như các đối tượng được tự tạo ra (vì dụ hình animation) và kết hợp chúng ở bộ giải mã. Một trong nhiều khả năng thú vị sử dụng đối tượng nhân tạo là mặt người hoạt hình (facial animation). Đây cũng là một ví dụ khác của việc ánh xạ texture thành đường nét chuyển động nhưng trong trường hợp này đường nét được chỉ định bởi mô hình lưới hay mô hình 3D được hình thành bởi các node. Vị trí của mỗi node được mã hóa sử dụng mã hóa dự đoán trước để tăng hiệu suất mã khi đường nét của khuôn mặt thay đổi.

Version 2 của MPEG-4 có thể thêm vào thân hình (body) động. Thân hình là một đối tượng có thể tạo ra các mô hình thân hình ảo và động dưới dạng tập hợp lưới 3D nhiều cạnh. Hai tập hợp các thông số định nghĩa cho body: Tập hợp các tham số định nghĩa body (BDP-body definition parameter) và tập hợp các thông số động body (BAP - body animation parameter). Tập BDP định nghĩa tập các thông số để biến đổi body mặc định thành các body theo yêu cầu khách hàng với bề mặt body, kích thước body và texture. Các tham số động body (BAP) cho phép tạo ra chuyển động với các mô hình body khác nhau. Như vậy, có thể ngay lập tức nhận BAP từ luồng bit thu mà tạo ra sự linh hoạt của body. Khi thu được, BDP được dùng để biến đổi body chung (body một người chuẩn chẳng hạn) thành các body riêng biệt dựa vào giá trị của các thông số. Bất cứ thành phần nào cũng có thể trống. Một thành phần trống có thể được thay thế bằng thành phần mặc định tương ứng khi body được biểu hiện. Các đặc điểm mặc định được xem như các đặc điểm chuẩn. Các đặc điểm này được định nghĩa như sau: bàn chân chỉ đến hướng phía trước, hai cánh tay đặt ở hai bên của body với lòng bàn tay hướng vào trong body. Các đặc điểm này cũng ngầm định trong tất cả BAP có giá trị mặc định.

Mô hình body người có thể hỗ trợ những ứng dụng khác nhau từ mô phỏng chuyển động thực của người đến công nghệ game sử dụng mô hình giống như người.

2.3.3.11.8 Scalability

MPEG-4 cung cấp chế độ phân cấp chất lượng tức thời và cố định ở mức đối tượng. Trong cả hai trường hợp, kỹ thuật này được sử dụng để tạo ra lớp cơ sở (base layer) tượng trưng cho chất lượng thấp nhất từ luồng bit và một hoặc nhiều lớp tăng cường (enhancement layer). Những lớp này có thể được tạo ra bằng việc mã hóa đơn giản. Chất lượng hình ảnh có thể được điều chỉnh thể bằng hai cách khác nhau. Nếu băng thông bị giới hạn thì luồng bit truyền đi sẽ chỉ bao gồm chỉ lớp cơ sở hoặc lớp cơ sở và một vài lớp tăng cường bấc thấp. Một cách khác tất cả các lớp có thể được truyền đi đến bộ giải mã, tại đây sẽ quyết định những lớp nào sẽ được sử dụng để giải mã. Nếu thiết bị hiển thị có độ phân giải thấp hoặc tài nguyên tính toán không đủ thì lớp tăng cường có thể bỏ qua.



Hình 2.3.26 Bộ mã hóa phân cấp chất lượng MPEG-4

Hình 2.3.26 mô tả sơ đồ khối của bộ mã hóa thực hiện quá trình phân cấp chất lượng với 2 mức cố định. Các VOP ngõ vào được biến đổi xuống còn độ phân giải thấp và mã hóa để tạo ra lớp cơ sở. Lớp cơ sở được đưa tới ngõ ra và tới bộ tổng hợp Multiplexer. Lớp cơ sở còn được giải mã tại chỗ và đưa tới bộ up-converted để có cùng độ phân giải như ngõ vào. Tín hiệu có độ phân giải cao này sẽ được so sánh với tín hiệu ngõ vào tại bộ trừ (Subtract), ảnh sai biệt ở đầu ra bộ trừ được mã hóa riêng ở bộ mã hóa lớp tăng cường. Chú ý rằng mỗi luồng VOP sau khi mã hóa tạo thành lớp đối tượng video. Lớp cơ sở sử dụng cả mã hóa Intra và Inter trong khi lớp tăng cường chỉ sử dụng mã hóa dự đoán.

Việc phân cấp chất lượng tạm thời thì đơn giản hơn. Luồng bit đến của VOP được chia thành các luồng nhỏ. Số VOP được yêu cầu sẽ được gửi đến bộ mã hóa lớp cơ sở, phần còn lại được gửi đến một hoặc nhiều bộ mã hóa tăng cường.

QUa strình phân cấp trong MPEG-4 có thể được áp dụng riêng rẽ đối với mỗi đối tượng cơ sở. Điều này làm cho quá trình mã hóa và giải mã trửo nên linh động hơn. Chẳng hạn một bộ giải mã trong hệ thống game không có đủ khả năng để giải mã tất cả các đối tượng ở tốc độ cao nhất có thể nên nó chỉ chọn giải mã đối với cảnh nền tốc độ thấp và chấp nhận mức độ chuyển động nền bị giựt, đồng thời giải mã với tốc độ cao đối với các đối tượng cận ảnh làm cho chuyển động của chúng mịn hơn

2.3.3.11.9 Mã hóa mở rộng (ACE: Advanced Coding Extension)

Version 2 của MPEG-4 đưa ra ba công cụ mới để cải thiện hiệu suất mã hóa đối với đối tượng video. Bao gồm: Bù chuyển động toàn cục (GMC: global motion compensation), bù chuyển động phần tư (quarter pel motion compensation) và DCT hình dáng thích ứng (shape-adative DCT). Các công cụ này cải thiện hiệu suất mã đến hơn 50% so với version 1 tùy theo loại ảnh và tốc độ bit.

GMC cho phép mã hóa toàn bộ chuyển động của đối tượng với một vài thông số và cải thiện độ phân giải của vector chuyển động bằng cách giảm sai số do dự đoán và sử dụng độ dư

Shape-adaptive DCT có thể được dùng để cải thiện hiệu suất mã của những khối đường biên khi không phải tất cả các pixel đều là phần tử của ảnh. Thay vì sử dụng DCT hai chiều 8x8 thì dùng khối DCT một chiều đối với hàng dọc được trước sau đó đến hàng ngang và chỉ xét những pixel thuộc đối tượng, gọi là các active pixel.

2.3.3.12 Chuẩn H.261

Khuyến cáo H.261 của CCITT là chuẩn nén cho các dịch vụ hội nghị truyền hình và điện thoại truyền hình qua mạng số dịch vụ tích hợp ISDN ở tốc độ n×64Kbps. Chuẩn này có 2 đặc tính quan trọng là ngưỡng trễ mã hoá tối đa là 150ms vì trễ này phù hợp với truyền thông video hai chiều dựa vào cảm nhận của người xem về hình ảnh phản hồi trực tiếp và dễ dàng thực hiện mạch tích hợp VLSI chi phí thấp cho việc thương mại hoá sản phẩm rộng rãi.

2.3.3.13 Chuẩn H.263

H.263 là chuẩn dành cho video tốc độ thấp 46 Kbps dùng trong các ứng dụng hội nghị từ xa qua mạng PSTN. Chuẩn này có cả đặc tính của MPEG-1 và MPEG-2. Mã hoá video của H.263 dựa trên chuẩn H.261 và thực chất nó là phiên bản mở rộng của H.261 với phương

pháp mã hoá video kết hợp DPCM/DCT. Cả hai chuẩn này đều dùng kỹ thuật chính như DCT, bù chuyển động, mã hoá chiều dài từ mã thay đổi, lượng tử hoá vô hướng và xử lý trên cấp macroblock. Duy chỉ có khái niệm về khung PB trong H.263 là khá đặc biệt, tên PB có nguồn gốc từ P và B, là sự kết hợp của P và B.

2.3.4 Ứng dụng và đánh giá khả năng kinh tế của các tiêu chuẩn nén

Bảng 2.3.7 tập hợp các ứng dụng của các tiêu chuẩn nén. Định dạng MJPEG có hạn chế trong ứng dụng. Nó được dùng chủ yếu trong máy ghi hình bằng đĩa (quá trình mã hóa và giải mã trong cùng một đơn vị).

Bảng 2.3.7. Các ứng dụng của các tiêu chuẩn nén.

Ứng dụng	Tiêu chuẩn mã hóa	Độ phân giải cực đại	Tốc độ bit cực đại
Videophone	H.261	176x144	64 ÷128 Kbps
Hội nghị truyền hình	H.261	352x288	0.384 ÷1.554 Mbps
	MPEG-2	720x576	
Truyền hình cáp	DPCM	720x576	4…9 Mbps
Truyền hình	MPEG-2		< 50 Mbps
	MJPEG	720x576	
Studio / sản xuất	MPEG-2(@4:2:2)		< 50 Mbps
		1920x1280	
HDTV / sản xuất	MPEG-2	1920x1280	100 Mbps
HDTV / truyền dẫn	MPEG-2		20 Mbps

Việc ra đời chuẩn MPEG-2 đã khẳng định ưu thế của MPEG so với MJPEG khi nó đáp ứng được từ khâu sản xuất đến truyền dẫn và phát sóng. Với tiềm năng kinh tế mạnh thì việc đầu tư thiết bị theo chuẩn MPEG cho tất cả các khâu trong truyền hình là điều không bàn cãi. Nhưng khả năng đầu tư cho truyền hình Việt Nam còn hạn chế, chúng ta lại đang ở giai đoạn nghiên cứu và thử nghiệm, cho nên việc tìm giải pháp thích hợp đảm bảo điều kiện kinh tế, chất lượng hình ảnh phù hợp cho từng công đoạn là vấn đề cần thiết.

Trong khâu phát sóng, chuẩn MPEG-2 MP@ML (4:2:0) là sự lựa chọn tối ưu có nguyên nhân sâu sắc và có yếu tố khách quan. Sâu sắc là người ta phải tiết kiệm tối đa dải thông của đường truyền, tăng số lượng chương trình. MPEG-2 MP@ML sử dụng các ảnh I, P, B và nén với hệ số nén rất cao, giảm vận tốc của dòng chương trình còn rất thấp để phát quảng bá tới các máy thu. Ví dụ cụ thể, phát qua bộ phát đáp của vệ tinh có dải thông 36 MHz, sử dụng nén MPEG-2 ta truyền được 10÷12 chương trình. Yếu tố khách quan là mắt người chỉ cảm nhận về độ phân tích đến một chừng mực nào đó là đủ. Nếu truyền tín hiệu có vận tốc cao

hơn, mắt người cũng chỉ cảm nhận thêm không đáng là bao mà số lượng chương trình phải giảm đi đáng kể. Điều đó đồng nghĩa với hiệu quả phát sóng thấp. Tóm lại, mục đích ưu tiên của phát quảng bá là nén làm sao để phát được nhiều chương trình.

Như vậy, vấn đề ta quan tâm ở đây là lựa chọn chuẩn nào trong công đoạn sản xuất hậu kỳ. Đối với nén của máy ghi hình cần phải chú ý giữ chất lượng tín hiệu còn đủ cho gia công xử lý trong studio. Các phương pháp nén trong máy ghi hình số thực chất là xử lý nội tại trong máy để giảm vận tốc dòng số ghi lên băng còn vừa đủ thấp nhằm giải quyết khá nhiều vấn đề thực tế.

Đối với MPEG-2, có thể sử dụng MPEG-2 4:2:2P@ML trong khâu hậu kỳ. Về mặt chất lượng, nếu sử dụng máy ghi công nghệ nén MPEG dùng một loại ảnh I cũng thuận lợi cho việc dựng in tương đương M-JPEG. Nhưng cách giải quyết như thế là chưa kinh tế, vì trong hệ thống mạch theo công nghệ nén MPEG, mạch "dự đoán bù chuyển động" là mạch phức tạp nhất và có giá thành cao mà lại không sử dụng (chỉ dùng một ảnh I, không sử dụng ảnh P và B). Vì mục tiêu phát được nhiều chương trình nên phát quảng bá sử dụng nén MPEG-2 với nhóm ảnh đầy đủ I, P, B và hệ số nén rất cao.

Qua phân tích ở trên, ta có thể thấy rằng chuẩn M-JPEG sử dụng trong các thiết bị sản xuất chương trình tiện dụng cho sản xuất studio và dựng hậu kỳ, làm kỹ xảo với giá thành hệ thống phù hợp.

PHỤ LỤC: CÁC TIÊU CHUẨN MÃ HOÁ ÂM THANH VÀ HÌNH ẢNH TRONG TRUYỀN THÔNG ĐA PHƯƠNG TIỆN

1. Các tiêu chuẩn của ITU- T cho âm thanh

Các tiêu chuẩn ITU-T cho âm thanh bao gồm G.711 · **G.722** · G.722.1 · G.722.2 · G.723 · G.723.1 · G.726 · G.728 · G.729 · G.729.1 · G.729a

G.711 là chuẩn ITU-T dùng cho thoại chủ yếu trong các hệ thống tổng đài, được phát hành chính thức vào năm 1972.

G.711 trình bày các mẫu điều chế xung mũ logarit cho tín hiệu ở băng tần thoại, tần số lấy mẫu là 8000 mẫu trong một giây.

Có hai giải thuật chính được định nghĩa trong chuẩn này, giải thuật μ-law dùng ở khu vực Bắc Mỹ, Nhật và giải thuật A-law dùng ở khu vực Châu Âu và những nước còn lại. Cả hai giải thuật điều tính toán trên mũ logarit, nhưng giải thuật A-law được thiết kế đặc biệt cho mục đích thực hiện các phép tính trong quá trình tính toán sao cho đơn giản hơn, chuẩn này cũng định nghĩa một chuỗi các giá trị mã lặp lại có mức công suất là 0 dB.

Hai giải thuật μ-law được mã hóa ở dạng các mẫu PCM tuyến tính 14-bit và A-Law là 13-bit với mẫu 8-bit. Như vậy, bộ mã hóa G.711 sẽ tạo được luồng dữ liệu bit có tốc độ 64kbit/giây với tần số lấy mẫu là 8kHz.

G.722 là chuẩn ITU-T dùng cho mã hóa tiếng nói băng tần rộng hoạt động với tốc độ truyền 32-64 kbit/giây. Công nghệ mã hóa dựa trên việc phân chia băng tần ADPCM.

G.722.1 cung cấp được việc nén dữ liệu với tốt độ bit thấp. Một biến thể mới của G722.1 là G.722.2, được biết dưới tên là AMR-WB (Adaptive Multirate Wideband), cho phép việc nén với tốc độ thấp hơn nữa, có thể đáp ứng tốt với các kiểu nén khác nhau cũng như các thay đổi địa hình mạng. Trong trường hợp sau, băng thông được tự động bảo tồn khi có sự nghẽn mạch cao. Khi việc nghẽn quay trở về ở mức bình thường, thì chế độ tốc độ bit cao hơn và mức nén thấp hơn được phục hồi.

Chuẩn G.722 và dữ liệu mẫu âm thanh tại tốc độ 16kHz, gấp đôi tốc độ xử lý tại các giao tiếp thoại truyền thống, kết quả là chất lượng thoại tốt hơn.

Chuẩn G.722.1, được biết qua tên khác là "Siren™", là một chuẩn quốc tế cho mã hóa âm thanh băng rộng ở tốc độ 24 và 32 kbps (băng thông thoại 50Hz-7kHz, tần số lấy mẫu là 16 ksps)tốc độ 16kb/giây), sử dụng trong các hệ thống hội nghị truyền hình được phê chuẩn vào 30 tháng 09 năm 1999.

Chuẩn G.722.1 là bộ nén dựa trên sự biến đổi sao cho tối ưu hóa cả âm thoại lẫn nhạc. Độ phức tạp tính toán tương đối thấp đối với bộ nén chất lượng cao, độ trễ của giải thuật của hai điểm đầu cuối là 40ms.

Phiên bản G.722.1/Annex C, được phê chuẩn bởi ITU-T vào 14 tháng 05 năm 2005, còn được biết thông qua tên Siren14™, được phát triển bởi Polycom với dạng không cần bản quyền truyền với tần số 14kHz (32ksps).

Số lượng mã hóa âm thanh băng tần rộng ITU đôi khi không được hiểu chính xác. Thực tế, có ba loại mã hóa cơ bản phân biệt, nhưng điều có chung một tên là G.722. Đầu tiên, G.722 là mã hóa với tần số 7kHz, sử dụng ADPCM hoạt động với tốc độ truyền 48-64kbps. Một phiên bản khác G.722.1 hoạt động với tốc độ dữ liệu bằng một nửa nhưng có chất lượng tốt như G.722 với phương pháp mã hóa dựa vào nền tảng chuyển đổi. Và chuẩn G.722.2, hoạt động với âm thoại băng tần rộng với tốc độ bit truyền rất thấp, sử dụng giải thuật CELP-based.

Về vấn đề bản quuyền, đến thời điểm này, giấy đăng ký bản quyền cho G.722 đã hết hạn, cho nên hiện tại chuẩn này được xem như là chuẩn miễn phí. G.722.1 thuộc bản quyền của tập đoàn Polycom và chuẩn G.722.2 còn có tên là AMR-WB, thuộc quyền sở hữu của tập đoàn VoiceAge.

G.722.2 (GSM AMR WB)

Adaptive Multi Rate - WideBand hay **AMR-WB** là một chuẩn mã hóa tiếng nói được phát triển sau khi AMR sử dụng cùng công nghệ tương tự như ACELP. Mã cung cấp chất lượng âm thoại tuyệt vời bởi vì sử dụng băng tần thoại rộng hơn 50-7000 Hz khi so sánh với các mã âm thoại băng hẹp hiện đang dùng rộng rãi trong các POTS với 300-3400Hz. AMR-WB được hệ thống hóa thành G.722.2, là một chuẩn mã hóa âm thoại chuẩn ITU-T.

Các trạng thái hoạt động của ẢM: AMR-WB hoạt động tương tự AMR với nhiều tốc độ bit khác nhau gồm: 6.60; 8.85; 12.65; 14.25; 15.85; 18.25; 19.85; 23.05 và 23.85 kbps. Tín hiệu truyền với tốc độ thấp nhất cho chất lượng thoại tốt nhất ứng với môi trường không nhiễu là 12.65 kbps. Tốc độ bit cao rất hữu dụng trong môi trường có nhiễu và trong trường hợp tín hiệu truyền là âm nhạc. Tốc độ bit 6.60 à 8.85 cung cấp chát lượng chấp nhận được khi so sánh với mã hóa băng tần hẹp.

AMR-WB được chuẩn hóa cho việc sử dụng trong tương lai trong các hệ thống mạng như UMTS. Chuẩn này cung cấp chất lượng thoại tốt hơn rất nhiều và được chọn dùng cho nhiều mạng cũ hỗ trợ cho băng rộng. Tháng 10 năm 2006, kiểm nghiệm AMR-WB đầu tiên được thực hiện trên hệ thống mạng thực do T-Mobile và Ericssion phối hợp tại Đức.

G.723 là một chuẩn ITU-T mã hóa âm thoại băng tần rộng, là chuẩn mở rộng của G.721 điều chế xung sai phân tương thích với tốc độ truyền 24 và 40 kbps cho các ứng dụng thiết bị nhân mạch số, hiện nay G.723 được thay thế bởi chuẩn G.276, do đó hiện tại chuẩn này là lỗi thời.

Chuẩn G.723.1 là chuẩn mã hóa âm thanh cho thoại với tính năng nén thoại trong khung 30 mili giây, chu kỳ 7.5ms cũng được sử dụng. Nhạc hoặc âm tone như DTMF hoặc fax ton không thể truyền tin cậy với chuẩn mã hóa này, do đó một số các phương pháp khác nhưu G.711 hoặc phương pháp ngoài dãy băng tần dùng để truyền các tín hiệu này.

Chuẩn G.723.1 chủ yếu dùng trong các ứng dụng Voice over IP (VoIP) vì yêu cầu băng thông thấp. Nó trở thành chuẩn ITU-T vào năm 1995, điều phức tạp của giải thuật là yêu cầu là dưới 16MIPS với 2.2kByte về RAM.

Có hai tốc độ bit mà G.723.1 có thể hoạt động:

- 6.3 kbit/s (sử dụng khung 24 byte), dùng giải thuật MPC-MLQ (MOS 3.9)
- 5.3 kbit/s (sử dụng khung 20 byte) dùng giải thuật ACELP (MOS 3.62)

G.726 là chuẩn mã hóa tiếng nói ITU-T ADPCM truyền âm thanh với các tốc độ 16, 24, 32, và 40 kbps. Là chuẩn thay thế cho cả G.721 (ADPCM tốc độ 32kbps) và chuẩn G.723 (ADPCM với tốc độ 24 và 40 kbps). G.726 hoạt động với tần số là 16 kbps. Bốn tốc độ bit thường sử dụng cho chuẩn G.726 tương ứng với kích thước của một mẫu theo thứ tự là 2-bits, 3-bits, 4-bits, và 5-bits.

Tốc độ thường dùng là 32 kbps, bởi vì đây chính là tốc độ bằng một nửa so với chuẩn G.711, như thế làm gia tăng dung lượng củ mạng lên 50%. Thông thường được dùng trong các mạng điện thoại quốc tế cũng như hệ thống điện thoại không dây DECT.

G.721 được giới thiệu lần đầu tiên vào năm 1984, trong khi chuẩn G.723 được giới thiệu vào năm 1988. Cả hai được gộp chung thành chuẩn G.726 vào năm 1990.

G.727 được giới thiệu cùng thời điểm với G.726, cùng tốc độ bít nhưng tối ưu hơn cho môi trường PCME Packet Circuit Multiplex Equipment. Điều này đạt được bằng cách nhúng bộ lượng tử hóa 2 bit vào bộ lượng tử hóa 3 bit, cho phép hủy bỏ bit có trọng số nhỏ nhất trong chuỗi bit truyền mà không có ảnh hưởng xấu đến tín hiệu âm thoại.

G.728 là chuẩn ITU-T mã hóa âm thoại với tốc độ 16kbps. Công nghệ sử dụng là LD-CELP, Low Delay Code Excited Linear Prediction. Độ trễ của mã chỉ 5 mẫu (0.625 ms). Dự đoán tuyến tính được thực hiện tính toán với bộ lọc LPC ngược bậc 50. Ngõ vào kích thích được tạo ra để đảm bảo nhận được độ lợi VQ. Chuẩn được đưa ra vào năm 1992 dưới dạng giải thuật mã dấu chấm động. Năm 1994, bản dùng cho dấu chấm tĩnh được phát hành. G.728 có tốc độ lên đến 2400 bps. Độ phức tạp của bảng mã là 30 MIPS, với yêu càu 2.2kByte về RAM.

G.729 là một giải thuật nén dữ liệu âm thanh dùng cho tín hiệu thoại, nén tín hiệu âm thanh với khung 10 mili giấy. Các tone nhạc như DTMF hoặc fax không thể truyền với bộ mã hóa này, mà phải sử dụng G.711 hoặc phương pháp ngoại băng tần để truyền các tín hiệu này.

G.729 đa số dùng trong các ứng dụng Voice over IP (VoIP) với yêu cầu băng tần thấp. Chuẩn G.729 hoạt động ở tốc độ 8 kbps, nhưng các phiên bản mở rộng có thể hoạt động tại 6.4 kbps đối với môi trường truyền xấu và 11.8 kbps với yêu cầu chất lượng thoại tốt hơn. Trong thực tế, người ta thường dùng chuẩn G.729a, tương tự như G.729 nhưng có độ tính toán đơn giản hơn, tuy nhiên chuẩn này lại không cho chấtl lượng thoại tốt hơn.

Phiên bản G.729b là một chuẩn có bản quyền, sử dụng module VAD để phát hiện tín hiệu thoại hay phi thoại. Nó cũng bao gồm một module DTX dùng để quyết định nâng cấp các thông số nhiễu nền cho tín hiệu phi thoại (các khung nhiễu). Các khung này được truyền để thực hiện việc nâng cấp này được gọi là các khung SID. Một bộ tạo nhiễu (CNG) cũng được tích hợp trong chuẩn này, bởi vì trong một kênh truyền, nếu việc truyền bị dừng lại vì lý do tín hiệu là tín hiệu phi thoại, thì site còn lại sẽ xem như đường kết nối này bị đứt. Vì thế khi sử dụng chuẩn này cần phải thận trọng.

Những năm gần đây, chuẩn G.729 đã được nghiên cứu mở rộng để hỗ trợ cho tín hiệu âm thoại băng tần rộng và mã hóa âm thanh thành chuẩn G.729.1. Bộ mã hóa G.729.1 được thiết kế theo mô hình phân cấp, tốc độ bit và chất lượng điều hiệu chỉnh đơn giản bằng cách thức cắt giảm chuỗi bit truyền.

G.729.1 thêm chức năng băng tần rộng so với G.729 thông qua các lớp được nhúng vào. Lớp đầu tiên trên cùng G.729 (12kps) vẫn là dạng băng tần hẹp. 14 kbps thêm vào chất lượng băng tần rộng thông qua việc tái tạo phổ, sử dụng đóng gói thời gian và đóng gói tần số (có tốc độ truyền cộng thêm là 2kbps). Các lớp khác (ứng với từn bước 2 kbps) thêm nhiều thông tin về nội dung của phổ ở các tần số cao và như thế làm gia tăng chất lượng tín hiệu.

Các mã được phát triển bởi sự phối hợp của các tổ chức: France Telecom, tập đoàn Mitsubishi Electric, tập đoàn Nippon Telegraph và Telephone (NTT), và Université de Sherbrooke.

2. Các tiêu chuẩn của ITU- T cho hình ảnh và Video.

Chuẩn H.261 là chuẩn ITU mã hóa tín hiệu video năm 1990 được đưa ra để truyền trên hệ thống đường dây ISDN với các tốc độ dữ liệu là số nhân của 64 kbps. Tốc độ dữ liệu của giải thuật mã hóa được đưa ra để có thể hoạt động được giữa 40 kbps và 2 Mbps. Chuẩn hỗ trợ các khung video CIF và QCIF với độ phân giải 352x288 và 176x144 theo thứ tự tương ứng (và 4:2:0 mẫu với độ phân giải màu là 176x144 và 88x72 theo thứ tự tương ứng). Chuẩn cũng xét đến tình huống dự phòng cho việc truyền các hình với độ phân giải 704x576 (được hiệu chỉnh vào năm 1994).

Chuẩn H.261 là chuẩn mã hóa tín hiệu video số đầu tiên được áp dụng trong thực tế. Việc thiết kế chuẩn H.261 là một nỗ lực tiên phong, các chuẩn mã hóa video toàn cầu sau này (MPEG-1, MPEG-2/H.262, H.263, và ngay cả H.264) cũng chủ yếu dựa trên chuẩn này. Ngoài ra, cac phương pháp được sử dụng bởi hội đồng phát triển H.261 (đứng đầu là Sakae Okubo) cộng tác phát triển chuẩn vẫn được ứng dụng trong các công việc chuẩn hóa các chuẩn sau này trong lĩnh vực này. Giải thuật mã hóa sử dụng một hybrid của sự chuyển dộng của ước đoán hình ảnh nội tại và mã hóa truyền trong không gian với việc lượng tử vô hướng, phân hình theo kiểu zig-zac và mã hóa entropy.

2.1 Chuẩn H.261

Quá trình cơ bản của việc thiết kế được gọi là macroblock. Mỗi macroblock bao gồm 1 dãy 16x16 các mẫu luma và hay dãy mẫu chroma 8x8 dùng việc lấy mẫu 4:2:0 và không gian màuYCbCr.

Dự đoán hình ảnh nội tại thực hiện laọi bỏ các dư thừa tạm thơi, với các vector chuyển động được dùng để hỗ trợ cho việc bù mã hóa cho việc di động. Mã di chuyển sử dụng chuyển đổi cosin rời rạc 8x8 (DCT) dùng để loại bỏ các dư thừa thuộc không gian, và các hệ số biến đổi lượng tử được phân hình theo kiểu zig-zac và mã hóa entropy (dùng mã Run-Level variable-length) để loại bỏ các dư thừa đã thống kê.

Chuẩn H.261 thật sự chỉ định rõ bằng cách nào để giải mã video. Các nhà thiết kế bộ mã hóa được tự do trong việc đưa ra các giải thuật mã hóa của riêng họ, ngay cả với tín hiệu ngõ ra bộ mã hóa không được tự nhiên nhằm mục đích có thể được giải mã bằng bất kỳ bộ giải mã nào miễn là được thiết kế theo đúng chuẩn. Các bộ mã hóa cũng được thiết kế tuỳ ý

nhằm thực hiện quá trình tiền xử lý mà chúng muốn ngõ vào video ưu tiên mặc định thực hiện. Một kỹ thuật hiệu quả trong vấn đề hậu xử lý trở thành phần tử chính yếu của các hệ thống tốt nhất dựa trên chuẩn H.261 là lọc giải khóa. Nó thực hiện việc giảm sự xuất hiện của vật nhân tạo nhiễu có dạng hình khối gây ra bởi việc bù di động theo dạng khối và các phần chuyển đổi do việc thiết kế tạo ra. Việc lọc giải khóa đã trở thành một phần tích hợp trong hầu hết các chuẩn hiện nay, H.264 (ngay cả sử dụng chuẩn H.264, việc hậu xử lý vẫn cho phép thực hiện và có thể cho được chất lượng cao)

Việc lọc được đề cập trong việc chuẩn hóa có ảnh hưởng đến việc cải tiến quan trọng giữa khả năng nén và thiết kế H.261. Tuy nhiê, H.261 vẫn là định hướng lịch sử chính trong lãnh vực phát triển của mã hóa video.

2.2 Chuẩn H.262

Chuẩn H.262 là một chuẩn mã hóa video số ITU-T. Chuẩn này liên quan đến phần video của chuẩn ISO/IEC MPEG-2 (được biết dưới cái tên ISO/IEC 13818-2). Chuẩn này được phát triển do sự hợp tác của ITU-T và các tổ chức ISO/IEC JTC 1, và trở thành chuẩn chung cho cả hai tổ chức này. ITU-T Recommendation H.262 và ISO/IEC 13818-2 được phát triển và phát hành dưới dạng là chuẩn quốc tế. Hai tài liệu này mô tả hầu hết tất cả các khía cạnh của chuẩn.

2.3 Chuẩn H263

Chuẩn H.263 là chuẩn mã hóa ITU-T thiết kế vào năm 1995/1996 dùng cho giải pháp mã hóa nén tốc độ truyền thấp cho các dịch vụ hội nghị truyền hình.

Mã đầu tiên được thiết kế trong các hệ thống H.324 (PSTN hoặc các mạch chuyển mạch mạch khác truyền dịch vụ hội nghị truyền hình và điện thoại truyền hình), cũng như trong các hệ thốgn dùng mã H.323 (hội nghị truyền hình RTP/IP-based), H.320 (hội nghị truyền hình ISDN-based), RTSP (phương tiện truyền thông dạng streaming) và SIP (hội nghị Internet). Hầu hết nội dung Flash Video(dùng trên các site như YouTube, Google Video, MySpace, v.v….) được mã hóa dưới dạng định dạng này, tuy vẫn có site sử dụng mã hóa VP6, hỗ trợ phiên bản Flash 8. Tín hiệu video H.263 có thể được giải mã bằng thư viện phi bản quyền LGPL-licensed dùng trong các chương trình như ffdshow, VLC media player và MPlayer.

Chuẩn H.263 được phát triển như là một phiên bản nâng cấp dựa trên chuẩn H.261, và chuẩn MPEG-1, MPEG-2. Phiên bản đầu tiên được hoàn thành vào năm 1995 và hoàn toàn phù hợp trong việc thay thế cho H.261 với tất cả các tốc độ truyền. Hiện tại đã có các phiên bản H.263v2 (còn gọi là chuẩn H.263+ 1998) và chuẩn H.263v3 (H.263++ 2000).

Chuẩn mã hóa được ITU-T sau H.263 là H.264, còn có tên là AVC và MPEG-4 phần thứ 10. Hầu hết các sản phẩm hội nghị truyền hình công nghệ mới hiện nay luôn tích hợp cả ba chuẩn H.264, H.263 và H.261.

2.4 Chuẩn H264

Chuẩn H.264, MPEG-4 Part 10, hay AVC (dùng cho Advanced Video Coding), là một chuẩn mã hóa video số với độ nén cực cao, là kết quả của ITU-T Video Coding Experts Group (VCEG) kết hợp với ISO/IEC Moving Picture Experts Group (MPEG), được xem là sản phẩm thương mại Joint Video Team (JVT). Chuẩn ITU-T H.264 và ISO/IEC MPEG-4 Part 10(ISO/IEC 14496-10) ứng dụng các công nghệ lý tưởng. Phiên bản nháp đầu tiên được hoàn thành vào tháng 05 năm 2003.

Chuẩn H.264 được đặt tên theo cùng dòng ITU-T H.26x của các chuẩn video, trong khi tên AVC được đặt tên dựa theo tên dự án hợp tác, với tên của dự án là H.26L. Chuẩn còn được gọi bằng các tên khác H.264/AVC, AVC/H.264, H.264/MPEG-4 AVC, MPEG-4/H.264 AVC nhằm nhấn mạnh tính kế thừa. Đôi khi, còn được gọi là "mã hóa JVT" với lý do là tổ chức JVT phát triển.

Mục đích của dự án H.264/AVC là tạo ra một chuẩn có khả năng cung cấp tín hiệu video chất lượng cao với các tốc độ bit truyền thấp, nhỏ hơn hay bằng một nửa so với tốc độ của các chuẩn trước (như MPEG-2, H.263, hay MPEG-4 Part 2) với tính ứng dụng cao trong thực tế. Ngoài ra, chuẩn phải đáp ứng yêu cầu cung cấp cách thức linh động cho phép chuẩn được ứng dụng rộng rãi trong nhiều trình ứng dụng (ví dụ cho cả tốc độ bit cao và thấyp hoặc độ phân giải cao hoặc thấp, và chạy ổn định trong nhiều hệ thống cũng như mạng (cho việc broadcast, lưu trữ DVD, mạng gói RTP/IP, và các hệ thống tổng đài đang phương tiện ITU-T)

2.5 Chuẩn JVT

Chuẩn JVT đã hoàn thành việc nâng cấp, phát triển một số tính năng mở rộng so với chuẩn nguyên thuỷ, được biết dưới tên là Fidelity Range Extensions (FRExt). Các phiên bản mở rộng hỗ trợ mã hóa video với độ trung thực cao bằng cách thức gia tăng độ chính xác lấy mẫu (bao gồm mã hóa 10-bit và 12-bit) với thông tin màu độ phân gải cao (gồm các cấu trúc lấy mẫu như YUV 4:2:2 và YUV 4:4:4). Một số tính năng khác trong dự án Fidelity Range Extensions (như phép biến đổi số nguyên chuyển mạch tương thích 4×4 và 8×8, các ma trận trọng số lượng tử hóa dựa trên giác quan, mã hóa không mất mát hình nội tại hiệu quả, hỗ trợ các không gian màu cộng thêm và phép biến đổi màu số dư). Công việc thiết kế trong dự án được hoàn thành vào tháng 7 năm 2004 và phiên bản nháp được ra mắt vào tháng 09 năm 2004.

TÀI LIỆU THAM KHẢO

1. Anil K. Jain, Fundamentals of Digital Image Processing, Prentice Hall, 1988.
2. J. R. Parker, Algorithms for Image Processing and Computer Vision, Wiley, 1996.
3. Alan C. Bovik , Handbook of Image and Video Processing, Academic Press, 2000.
4. John R. Deller, John H. L. Hansen, John G. Proakis, Discrete-Time Processing of Speech Signals, Wiley-IEEE Press, 1999.
5. R. C. Gonzalez, R. E. Woods, Steven L. Eddins, Digital Image Processing Using MATLAB, Prentice Hall, 2003.
6. R. C. Gonzalez, R. E. Woods Digital Image Processing , Prentice Hall, 2002.
7. William K. Pratt, Digital Image Processing: PIKS Inside, Third Edition © 2001 John Wiley & Sons, Inc.
9. Michael Robin & Michel Poulin, Digital Television Fundamental, McCraw-Hill Companies. Inc.
10. Đỗ Hoàng Tiến, Dương Thanh Phương Truyền hình kỹ thuật số. NXB Khoa học và kỹ thuật, 2004.
11. Lương Mạnh Bá, Nguyễn Thanh Thủy, Nhập môn xử lý ảnh số, NXB Khoa học và kỹ thuật, 1999.

MỤC LỤC

XỬ LÝ ÂM THANH, HÌNH ẢNH

Mã số: 411XAH450

Chịu trách nhiệm bản thảo

TRUNG TÂM ĐÀO TẠO BƯU CHÍNH VIỄN THÔNG 1