

Học viện Công nghệ Bưu chính Viễn thông

Đỗ Trung Tuấn

Cơ sở dữ liệu đa phương tiện

Hà Nội, 2010

Mục lục

Mục lục.....	2
Giới thiệu.....	5
Chương I. Tổng quan về cơ sở dữ liệu đa phương tiện.....	6
1.1 Mở đầu.....	6
1.2 Khái niệm dữ liệu đa phương tiện.....	6
1.1.1. Kiểu dữ liệu và đa phương tiện.....	6
1.1.2. Cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu.....	7
1.1.3. Tìm kiếm thông tin tư liệu văn bản.....	7
1.1.4. Tìm kiếm và chỉ số hóa đa phương tiện.....	7
1.1.5. Trích đặc trưng, thể hiện nội dung và chỉ số hóa.....	8
1.3. Đặc trưng của các đối tượng đa phương tiện.....	8
1.3.1. Sự gia tăng dữ liệu đa phương tiện và các tính chất của chúng.....	8
1.3.2. Hệ quản trị cơ sở dữ liệu và vai trò quản lý dữ liệu đa phương tiện.....	9
1.3.3. Hệ thống tìm kiếm thông tin đối với dữ liệu đa phương tiện.....	11
1.3.4. Tiếp cận tích hợp để tìm kiếm và chỉ số hóa đa phương tiện.....	11
1.3.5. Tổng quan về hệ thống tìm kiếm và chỉ số hóa đa phương tiện.....	12
1.4. Cấu trúc lưu trữ cơ sở dữ liệu đa phương tiện.....	12
1.4.1. Giới thiệu.....	13
1.4.2. Cây k-D.....	13
1.4.3. Cây tứ phân.....	18
1.4.4. Cây tứ phân MX.....	21
1.4.5. Cây R.....	24
1.4.6. So sánh các cấu trúc dữ liệu đa phương tiện.....	26
1.5. Ngôn ngữ thao tác dữ liệu đa phương tiện.....	27
1.5.1. Giao diện người dùng.....	27
1.5.2. Khả năng của hệ thống tìm kiếm và chỉ số hóa và ứng dụng.....	27
1.6. Kết luận.....	28
Chương 2. Tư liệu đa phương tiện tương tác.....	29
2.1 Cơ sở dữ liệu đa phương tiện tương tác.....	29
2.1.1. Giới thiệu.....	29
2.1.2. Kiến trúc của MIRS.....	29
2.1.3. Các mô hình dữ liệu.....	31
2.1.4. Thiết kế giao diện người dùng.....	35
2.2. Mô hình hoá tư liệu đa phương tiện tương tác IMD.....	37
2.2.1. Mô hình hoá tương tác với các sự kiện.....	38
2.2.2. Tổ hợp không gian, thời gian và các nhân tố.....	40
2.2.3. Dữ liệu văn bản.....	42
2.2.4. Đồ họa vecto và hình động.....	44
2.2.5. Âm thanh.....	50
2.2.6. Hình ảnh số.....	57
2.2.7. Video số.....	64
2.3. Phân loại.....	69

2.3.1. Một số chuẩn.....	70
2.3.2. Các đặc tính và yêu cầu của dữ liệu và ứng dụng đa phương tiện.....	71
2.4. Mô hình kịch bản.....	74
2.4.1. Kịch bản trong IMD.....	74
2.4.2. Kịch bản đa phương tiện.....	75
2.5. Tìm kiếm tư liệu đa phương tiện tương tác.....	77
2.5.1. Tìm tư liệu đa phương tiện tương tác dựa trên cấu trúc không gian, thời gian.....	78
2.6. Kết luận.....	81
Chương 3. Thành tựu và xu hướng.....	82
3.1. Các thành tựu chính của công nghệ hệ quản trị cơ sở dữ liệu đa phương tiện.....	82
3.1.1. Mô hình hoá.....	82
3.1.2. Toàn vẹn.....	82
3.1.3. Tìm theo nội dung.....	82
3.2. Các sản phẩm thương mại và mẫu nghiên cứu.....	86
3.2.1. Một số sản phẩm.....	86
3.2.2. Quản lý đa phương tiện.....	86
3.2.3. Các vai trò trong dự án đa phương tiện.....	89
3.3. Hướng phát triển của cơ sở dữ liệu đa phương tiện.....	90
3.3.1. Một số hướng hiện tại và khuynh hướng.....	90
3.3.2. An toàn dữ liệu đa phương tiện.....	91
3.3.3. Yêu cầu về tổ chức dữ liệu đa phương tiện.....	93
3.4. Kết luận.....	95
Chương 4. Quản trị dữ liệu đa phương tiện.....	96
4.1. Khái niệm về quản trị cơ sở dữ liệu đa phương tiện.....	96
4.1.1. Dạng dữ liệu đa phương tiện.....	96
4.1.2. Ngôn ngữ hỏi dữ liệu đa phương tiện.....	97
4.1.3. Vấn đề khác.....	98
4.2. Kiến trúc hệ quản trị cơ sở dữ liệu đa phương tiện.....	98
4.2.1. Các kiến trúc về tổ chức nội dung.....	98
4.2.2. Nguyên tắc tự quản.....	98
4.2.3. Nguyên tắc đồng đều.....	98
4.2.4. Nguyên tắc tổ chức hỗn hợp.....	99
4.2.5. Một số nhận xét.....	99
4.2.6. Tổ chức cơ sở dữ liệu dựa trên nguyên tắc thống nhất.....	100
4.3. Các kỹ thuật mô hình hóa dữ liệu.....	100
4.3.1. Mô hình quan hệ.....	100
4.3.2. Cơ sở dữ liệu hướng đối tượng.....	101
4.3.3. Cơ sở dữ liệu đa phương tiện.....	107
4.4. Các kỹ thuật chỉ số hoá và trừu tượng hoá.....	108
4.4.1. Giới thiệu.....	108
4.4.2. Chỉ số hoá cơ sở dữ liệu đa phương tiện.....	109
4.4.3. Các chỉ số hiển hiện.....	109
4.4.4. Trừu tượng hoá video.....	110
4.4.5. Đồ thị chuyển cảnh.....	112

4.5. Tìm thông tin đa phương tiện dựa trên nội dung.....	112
4.5.1. Giới thiệu về tìm thông tin đa phương tiện.....	112
4.5.2. Lọc thông tin	113
4.5.3. Hồi dữ liệu đa phương tiện	113
4.5.4. Tìm theo nội dung, sử dụng từ khoá.....	114
4.6. Thí dụ về cơ sở dữ liệu đa phương tiện.....	114
4.6.1. Một số hệ thống.....	114
4.6.2. Tìm các đối tượng dựa trên hình dạng.....	117
4.6.3. Thể hiện hình dạng	118
4.6.4. Việc khớp các hình	118
4.6.5. Các liên kết video đa phương tiện.....	118
4.7. Các ứng dụng của đa phương tiện	119
4.7.1. Các hình ảnh thô.....	120
4.7.2. Thể hiện ảnh đã nén.....	122
4.7.3. Xử lý ảnh thông qua việc phân đoạn ảnh	124
4.7.4. Tìm kiếm dựa trên sự tương tự.....	126
4.7.5. Tổng quát về cơ sở dữ liệu ảnh	129
4.7.6. Thể hiện cơ sở dữ liệu ảnh nhờ mô hình quan hệ	129
4.7.7. Thể hiện cơ sở dữ liệu ảnh trên cây R	132
4.7.8. Kết luận về cơ sở dữ liệu ảnh.....	134
4.8. Nhận xét về dữ liệu đa phương tiện.....	134
4.8.1. Đảm bảo QoS trong hệ thống truyền thông, tại máy chủ và máy khách	134
4.8.2. Một số vấn đề khác.....	135
4.9. Kết luận.....	137
Hướng dẫn sử dụng tài liệu theo chương trình khung.....	138
Tài liệu tham khảo	142

Giới thiệu

Trong nhiều năm, nghiên cứu và phát triển đa phương tiện là cần thiết trong ứng dụng truyền thông và để thể hiện thông tin đa phương tiện. Ngày càng nhiều dữ liệu số đa phương tiện được thể hiện dưới dạng hình ảnh, video, âm thanh... đòi hỏi các kỹ thuật lưu trữ, tìm kiếm hiệu quả và mạnh. Người ta có thể so sánh yêu cầu này với yêu cầu thể hiện dữ liệu ký tự dưới dạng tính toán được ở những năm 70 của thế kỷ XX.

Do vậy phát triển về quản trị dữ liệu đa phương tiện là bình thường đối với các tổ chức. Trước hết do nhu cầu thực tế, tiếp theo là công nghệ hiện tại không đủ khả năng giải quyết vấn đề đối với dữ liệu đa phương tiện. Một trong những khó khăn là việc chỉ số hóa và tìm kiếm dữ liệu đa phương tiện.

Người ta thấy cần biết công nghệ hiện tại của quản lý dữ liệu đa phương tiện. Đầu tiên là các đặc tính của dữ liệu đa phương tiện và các khía cạnh về thiết kế cho phép hệ thống cơ sở dữ liệu đa phương tiện đáp ứng các yêu cầu về dữ liệu. Đối với từng loại dữ liệu đa phương tiện, như văn bản, hình ảnh, âm thanh và video, cần có kỹ thuật chỉ số hóa riêng, ứng với đặc tính chính của dữ liệu thô. Công cụ tìm kiếm dữ liệu đa phương tiện cần lộ được câu hỏi người dùng, dựa trên mức độ tương tự của mẫu và dữ liệu đã lưu trữ. Việc tìm kiếm và chỉ số hóa theo nội dung dữ liệu đa phương tiện là quan trọng và khó khăn, do các khía cạnh rút từ dữ liệu thô thường được thể hiện qua vectơ nhiều chiều, đòi hỏi nhiều thời gian xử lý.

Các kỹ thuật và các cấu trúc dữ liệu có vai trò liên quan đến hiệu quả tìm kiếm dữ liệu. Cơ sở dữ liệu đa phương tiện với truy cập từ xa, qua mạng máy tính, theo mô hình khách/ chủ... sẽ phải xử lý các tình huống liên quan đến truyền dữ liệu, mã hóa dữ liệu. Vậy kiến trúc máy tính, việc lưu trữ đa phương tiện, hệ thống điều hành, hạ tầng mạng cần được quan tâm.

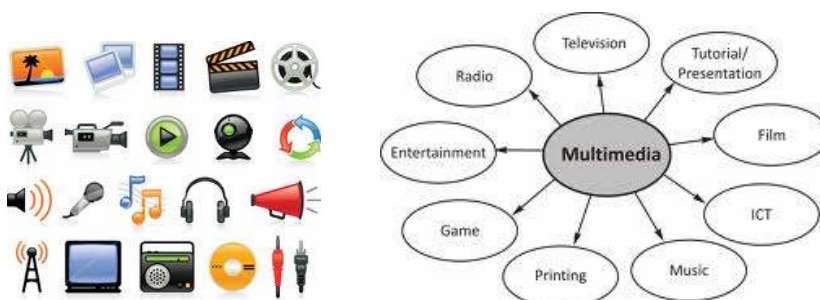
Trong hệ quản trị cơ sở dữ liệu truyền thống, hiệu năng liên quan đến tính hiệu quả, theo thời gian trả lời câu hỏi. Trong hệ thống đa phương tiện, hiệu quả cũng quan trọng, nhưng hiệu quả đối với tìm kiếm, đối với đối tượng đã có và phát hiện đối tượng tiềm ẩn, là có ý nghĩa. Người ta đề cập điều này do việc tìm kiếm ở đó theo so sánh tương tự, và các dữ liệu cũng không cho phép so sánh khớp. Do vậy độ đo hiệu quả là cần thiết đối với hệ quản trị cơ sở dữ liệu đa phương tiện. Một số khía cạnh khác, như an toàn dữ liệu, chuẩn... cũng đáng được quan tâm.

Chương I. Tổng quan về cơ sở dữ liệu đa phương tiện

1.1 Mở đầu

Các nghiên cứu và phát triển về đa phương tiện nhằm vào truyền thông và thể hiện dữ liệu đa phương tiện, xác định quyền tác giả. Hệ quản trị cơ sở dữ liệu đa phương tiện giữ vai trò như hệ quản trị truyền thống, khác là dữ liệu phức tạp và đa dạng. Để đảm bảo tính hiệu quả truy cập và tìm kiếm, hệ quản trị cần có kỹ thuật tìm kiếm và chỉ số hóa khác.

Vậy việc chỉ số hóa và tìm kiếm đa phương tiện là mục đích chính, trước khi xem xét các chức năng của hệ quản trị. Phần đầu sẽ thể hiện hệ thống chỉ số hóa và tìm kiếm MIRS¹ và một vài ứng dụng chung của nó.



Hình. Một số logo đa phương tiện

1.2 Khái niệm dữ liệu đa phương tiện

Cần thiết xác định từ đầu một số khái niệm, định nghĩa sử dụng trong suốt quá trình liên quan đến hệ thống đa phương tiện.

1.1.1. Kiểu dữ liệu và đa phương tiện

Định nghĩa: Phương tiện²: phương tiện nhằm đến các kiểu thông tin hay kiểu thể hiện thông tin, như dữ liệu số, chữ, hình ảnh, âm thanh, video.

Có nhiều cách xác định phương tiện. Phân loại thông thường dựa vào dạng vật lý và mối quan hệ phương tiện với thời gian. Ở đây xác định phương tiện không đề cập yếu tố thời gian. Thời gian cho phép xác định phương tiện tĩnh với phương tiện động, tức thời gian liên tục.

Định nghĩa: Phương tiện tĩnh³: phương tiện không có chiều thời gian, và nội dung và ý nghĩa của chúng không phụ thuộc vào thời gian thể hiện.

Các phương tiện tĩnh gồm dữ liệu số, chữ, đồ họa, hình tĩnh. Hình tĩnh được xem là sản phẩm được vẽ, quét hay chụp bằng máy chụp ảnh.

¹ multimedia indexing and retrieval systems

² media

³ static media

Định nghĩa: Phương tiện động¹: phương tiện có các chiều thời gian, với ý nghĩa và tính chính xác tùy theo tốc độ thể hiện.

Phương tiện động gồm hình động, âm thanh và video. Các phương tiện này có khoảng đơn vị bên trong hay tốc độ. Chẳng hạn video có 25 khung trong một giây. Việc thể hiện lại cần theo cách tổ chức trước đó. Do các phương tiện này thể hiện lại liên tục theo tốc độ cố định, chúng được gọi là phương tiện liên tục. Người ta cũng gọi chúng là phương tiện đẳng thời, tức chiếm thời gian như nhau, bởi quan hệ cố định giữa các đơn vị phương tiện và thời gian.

Đa phương tiện nhằm vào tập các kiểu phương tiện sử dụng cùng nhau. Nó cũng ngầm xác định có kiểu dữ liệu khác số, chữ. Do vậy thuật ngữ “đa phương tiện” cũng nhằm chỉ tính chất như tính từ.

Định nghĩa: Dữ liệu đa phương tiện²: dữ liệu hướng đến thể hiện máy đọc được của các kiểu phương tiện gộp.

Thông tin đa phương tiện hướng tới thông tin được truyền tải nhờ các kiểu phương tiện gộp. Đôi khi người ta dùng lẫn dữ liệu đa phương tiện với thông tin đa phương tiện. Người ta cũng sử dụng thuật ngữ đa phương tiện và phương tiện để chỉ thực thể tự trị trong MIRS, cho phép hỏi, tìm kiếm và thể hiện. Thuật ngữ “đối tượng” không hoàn toàn chính xác như trong tiếp cận hướng đối tượng.

1.1.2. Cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu

Trong tài liệu về cơ sở dữ liệu, hệ thống cơ sở dữ liệu đã phân biệt hệ quản trị cơ sở dữ liệu DBMS với cơ sở dữ liệu DB³. Giữa những vấn đề về đa phương tiện này, đôi khi lẫn lộn hai thuật ngữ.

Định nghĩa: Hệ quản trị cơ sở dữ liệu⁴: phần mềm cho phép mô tả, lưu trữ và xử lý các dữ liệu một cách khoa học.

1.1.3. Tìm kiếm thông tin tư liệu văn bản

Hệ thống tìm kiếm thông tin tự động IR⁵ được phát triển để xử lý khối lượng lớn các tài liệu khoa học từ 1940. Chức năng chính của hệ thống là lưu và quản lý số lớn các tư liệu văn bản theo cách tìm kiếm nhanh tư liệu hiện ra cho câu hỏi người dùng.

Định nghĩa: Hệ thống tìm thông tin tự động (IR): hệ thống là lưu và quản lý số lớn các tư liệu văn bản theo cách tìm kiếm nhanh tư liệu hiện ra đối với câu hỏi người dùng.

1.1.4. Tìm kiếm và chỉ số hóa đa phương tiện

Việc tìm kiếm trong DBMS dựa trên cấu trúc dữ liệu cho phép so sánh khớp. IR được xem là hệ thống tìm kiếm theo văn bản. Việc tìm kiếm theo nội dung dựa trên đặc trưng phương tiện thực tại như màu sắc, hình dáng, thay vì ghi chú, diễn giải văn bản của phương tiện.

¹ dynamic media

² multimedia data

³ DBMS Database Management Systems, DB Database

⁴ DBMS

⁵ information retrieval

Định nghĩa: Tìm kiếm theo nội dung¹: tìm kiếm dựa trên đặc trưng thực tại như màu sắc, hình dáng... của phương tiện.

Tìm kiếm theo nội dung thường dựa vào tính tương tự, thay vì so sánh khớp giữa câu hỏi và tập các mục tin của cơ sở dữ liệu. Ở hệ thống đa phương tiện, MIRS hướng đến tìm kiếm theo cả (i) DBMS; (ii) IR; (iii) theo nội dung.

1.1.5. Trích đặc trưng, thể hiện nội dung và chỉ số hóa

Định nghĩa: Đặc trưng là điểm nổi bật, giúp phân biệt cá thể đã cho với các cá thể khác mà ta có thể đem ra so sánh.

Trong MIRS, trích đặc trưng hay thể hiện nội dung, tức đặc trưng chính và nội dung trong mục tin, là việc quan trọng. Quá trình trích đặc trưng có thể thực hiện tự động hay bán tự động. Trong vài tư liệu về tìm theo nội dung, thuật ngữ “trích đặc trưng” cũng dùng cho việc chỉ số hóa.

Do vậy, khi sử dụng “chỉ số”, người ta nhằm đến cấu trúc dữ liệu hay tổ chức các đặc trưng được rút ra để tìm hiệu quả.

1.3. Đặc trưng của các đối tượng đa phương tiện

Đề cập nhu cầu về hệ thống chỉ số hóa và tìm kiếm đa phương tiện, người ta thấy ba việc sau đây cho thấy MIRS có ý nghĩa:

1. Ngày càng nhiều dữ liệu đa phương tiện được thu thập và lưu trữ;
2. Dữ liệu đa phương tiện khác với dữ liệu truyền thống ở tính chất riêng, có yêu cầu và ý nghĩa khác;
3. Cho dù các kỹ thuật IR có thể dùng để tìm kiếm đa phương tiện, nhưng một mình nó không đảm bảo tính hiệu quả đối với xử lý dữ liệu đa phương tiện.

Định nghĩa: Đối tượng là một vật, khái niệm hay thực thể.

1.3.1. Sự gia tăng dữ liệu đa phương tiện và các tính chất của chúng

Không thể không đối mặt với thông tin đa phương tiện. Người ta không thể tránh hết các dữ liệu âm thanh và tranh, ảnh. Xu hướng sử dụng dữ liệu đa phương tiện làm tăng công nghệ lưu trữ số. Dễ dàng đáp ứng nhu cầu đa phương tiện nhỏ, nhưng đối với yêu cầu toàn diện, đòi hỏi cả hệ thống tổ chức dữ liệu, tìm kiếm nhanh.



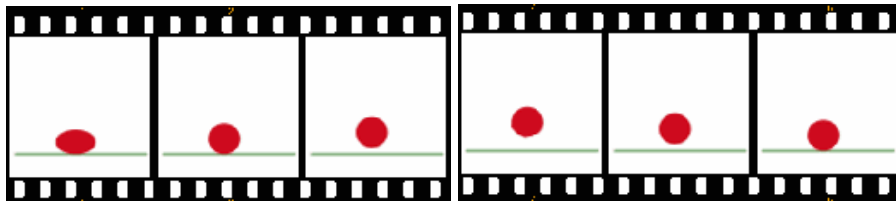
Hình. Sử dụng khái niệm đa phương tiện: máy bay vượt ngưỡng âm thanh

Người ta không chỉ chịu sức ép về khối lượng dữ liệu, mà còn các kiểu dữ liệu đa

¹ Content-based retrieval

dạng và các tính chất khác với dữ liệu số, chữ truyền thống. Các tính chất chính của dữ liệu đa phương tiện được liệt kê:

- Dữ liệu đa phương tiện, đặc biệt dữ liệu âm thanh, video là những dữ liệu được nén với tỉ lệ cao, khoảng 1 Gb chỉ chứa được khoảng 10 phút video;
- Dữ liệu âm thanh và video có chiều thời gian, đòi hỏi thể hiện theo tốc độ cố định để đạt hiệu quả dự định;
- Các dữ liệu âm thanh, hình ảnh và video số được thể hiện theo một loạt các mẫu riêng, do vậy khó tự động ghi nhận nội dung bởi không dễ xác định cấu trúc ngữ nghĩa của chúng;
- Nhiều ứng dụng đa phương tiện cần thể hiện đồng thời của nhiều dạng phương tiện theo cách tương ứng với không gian và thời gian;
- Ý nghĩa của dữ liệu đa phương tiện thường mờ và chủ quan, không tiện xác định rõ;
- Dữ liệu đa phương tiện mang nhiều thông tin, đòi hỏi nhiều tham số thể hiện nội dung.



Hình. Hình động tạo bởi các khung

1.3.2. Hệ quản trị cơ sở dữ liệu và vai trò quản lí dữ liệu đa phương tiện

Định nghĩa: Hệ quản trị cơ sở dữ liệu, DBMS, là phần mềm cho phép mô tả, lưu trữ và xử lí dữ liệu một cách khoa học.

Các DBMS truyền thống phù hợp với dữ liệu có cấu trúc. Mô hình dữ liệu quan hệ là mô hình thông dụng từ 1980 đến nay. Dữ liệu theo mô hình quan hệ được tổ chức trong các bảng quan hệ, có thuộc tính và các n_bộ. Có ba lớp ngôn ngữ hỏi dữ liệu, nhưng ngôn ngữ đại số quan hệ mà đại diện là SQL được sử dụng nhiều.

Thí dụ

```
create table STUDENT (
  stu# integer,
  name char (20),
  address char (100) );
```

và câu bổ sung dữ liệu

```
insert into STUDENT values (10, "Lew, Tom", "2 Main St.,
Churchill, Australia");
```

Bảng. Dữ liệu thí dụ

The Initial STUDENT Table

stu#	name	address
10	Lew, Tom	2 Main St., Churchill, Australia

The STUDENT Table After One Record Inserted

stu#	name	address
10	Lew, Tom	2 Main St., Churchill, Australia

Tìm kiếm dữ liệu

```
select name
from STUDENT
where stu# = 32
```

Thuộc tính trong DBMS là dữ liệu theo kiểu cố định, chẳng hạn số nguyên, chiếm 32 bit. Để trợ giúp các biến độ dài thay đổi, người ta đưa ra khái niệm đối tượng nhị phân lớn BLOB¹.

Định nghĩa: BLOB²: BLOB là xâu bit lớn có độ dài thay đổi.

Thí dụ cần lưu ảnh của sinh viên

```
create table STUDENT (
  stu# integer,
  name char (20),
  address char (100)
  picture BLOB);
```

BLOB chỉ là xâu bit, nên DBMS quan hệ không biết nội dung hay ngữ nghĩa của nó, mà chỉ xem như khối dữ liệu.

DBMS hướng đối tượng kết hợp nhiều khía cạnh của quản trị dữ liệu, như lưu trữ và tìm kiếm, với tiếp cận hướng đối tượng, với khía cạnh đóng gói, thừa kế, xác định đối tượng. Người ta cũng tự hỏi liệu tốt không nếu kết hợp mô hình quan hệ với mô hình hướng đối tượng, trong khi chưa có mô hình dữ liệu đa phương tiện. Các đối tượng đa phương tiện được nhận thức theo tiếp cận hướng đối tượng sẽ phù hợp và dễ thể hiện lại hơn. Hệ thống quan hệ đối tượng cho phép (i) xác định đối tượng theo nghĩa hướng đối tượng; (ii) sử dụng chức năng xử lý dữ liệu quan hệ đã có.

Thí dụ: xác định kiểu dữ liệu hình ảnh:

```
create type IMAGE (
  private
  size integer,
  resolution integer,
  content float[],
  public
  ...
);
```

Và sử dụng nó trong

```
create table STUDENT (
  stu# integer,
  name char (20),
```

¹ binary large object
² binary large object

address char (100)
picture IMAGE);

Khác nhau chính giữa BLOB và các đối tượng là các đối tượng được xác định phù hợp, với thuộc tính và chịu các phép xử lý trên thuộc tính, mà BLOB không thể. Khái niệm về BLOB và các đối tượng cũng đánh dấu một bước quản lý dữ liệu đa phương tiện, nhưng BLOB chỉ dùng để lưu trữ dữ liệu, trong lúc các đối tượng cần nhiều thuộc tính để thể hiện nội dung dữ liệu. Các tính năng cần phát triển cho dữ liệu đa phương tiện là:

- Công cụ trích rút nội dung và đặc trưng đa phương tiện, tự động hay bán tự động, trong các dữ liệu đa phương tiện;
- Cấu trúc chỉ số để quản lý các vectơ đặc trưng đa phương tiện;
- Độ đo tính tương tự, trong tìm kiếm thay cho so sánh khớp;
- Hệ thống lưu trữ phân tán đối với dữ liệu lớn và hạ tầng truyền thông đối với yêu cầu thời gian thực;
- Giao diện người dùng, thiết kế cho kiểu dữ liệu đa phương tiện khác nhau và cho phép thể hiện đa dạng.

1.3.3. Hệ thống tìm kiếm thông tin đối với dữ liệu đa phương tiện

Định nghĩa: Công cụ tìm kiếm là một phần mềm nhằm cho phép người dùng tìm kiếm và đọc các thông tin có trong phần mềm đó, trên một trang Web, hay trên toàn bộ Internet.

Hệ thống IR dùng cho dữ liệu văn bản. Các kỹ thuật sử dụng trong IR quan trọng nhờ (i) nhiều tư liệu văn bản trong tổ chức; (ii) văn bản có thể minh họa cho phương tiện khác, như âm thanh, hình ảnh, video. Người ta có thể sử dụng kỹ thuật IR đối với dữ liệu đa phương tiện. Tuy nhiên chúng có vài hạn chế:

- Minh họa, ghi chú cho phương tiện khác thường thực hiện thủ công, tốn thời gian;
- Minh họa bằng văn bản không thể đủ và chủ quan;
- Các kỹ thuật IR chỉ cho phép hỏi theo văn bản;
- Một vài đặc trưng đa phương tiện như bề mặt ảnh, hình dáng đối tượng là phức tạp, không dùng văn bản mà mô tả hết.

1.3.4. Tiếp cận tích hợp để tìm kiếm và chỉ số hóa đa phương tiện

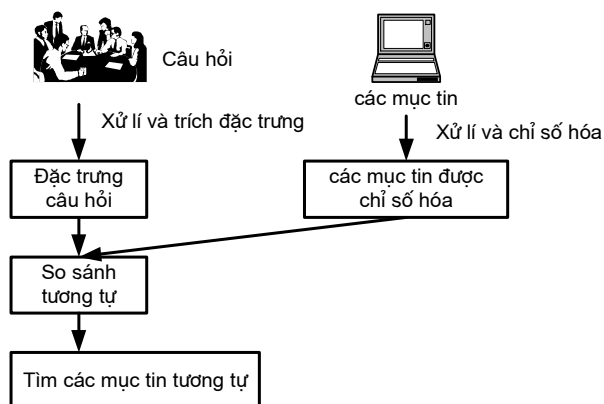
Dù hệ thống tìm kiếm thông tin IR có ích trong khung cảnh xử lý dữ liệu đa phương tiện, đối với dữ liệu có cấu trúc, người ta cần kỹ thuật mới để quản lý các tính chất riêng của dữ liệu đa phương tiện.

Ngoài các kỹ thuật xử lý dữ liệu của DBMS truyền thống đối với dữ liệu có cấu trúc, việc tích hợp thêm các kỹ thuật cho dữ liệu hình ảnh, âm thanh, video sẽ cho phép hệ thống tìm kiếm và chỉ số hóa đa phương tiện MIRS.

Định nghĩa: Chỉ số là hệ thống dùng để tìm kiếm thông tin nhanh và dễ hơn.

1.3.5. Tổng quan về hệ thống tìm kiếm và chỉ số hóa đa phương tiện

Trong hệ thống MIRS, mục tin trong cơ sở dữ liệu được tiền xử lí để rút các đặc trưng và nội dung ngữ nghĩa. Các mục tin này được chỉ số hóa nhằm tăng tốc tìm kiếm. Các câu hỏi người dùng được trích đặc trưng, so sánh tương tự với đặc trưng đã được chỉ số.



Hình. Mô hình tìm kiếm thông tin tổng quát

Còn một số vấn đề: (i) mục tin có chứa kiểu phương tiện ?; (ii) cách rút đặc trưng từ các mục tin phương tiện; (iii) độ đo tương tự, với khung nhìn đa dạng của người dùng; (iv) đánh giá hiệu quả tìm kiếm.

Định nghĩa: Thị giác là khả năng nhận và diễn giải thông tin từ ánh sáng đi vào mắt. Việc tri giác này còn được gọi là thị lực, sự nhìn.



Hình. Giác quan của con người

1.4. Cấu trúc lưu trữ cơ sở dữ liệu đa phương tiện

Khi đề cập về cấu trúc dữ liệu đa phương tiện, một điểm cần lưu ý trước tiên là dữ liệu đa phương tiện luôn mang khía cạnh (i) không gian và (ii) thời gian. Không gian đối với dữ liệu đa phương tiện thường nhiều chiều.

Định nghĩa: Cấu trúc dữ liệu là cách lưu trữ và tổ chức dữ liệu cụ thể trong máy tính để các dữ liệu được sử dụng hiệu quả.

Người ta quan tâm việc nhận thức các dữ liệu đa phương tiện, cách hình thức hoá chúng. Các khái niệm đa phương tiện cần được áp dụng phương pháp hình thức để có thể phát triển ứng dụng. Hầu hết các dữ liệu n chiều đều dùng cách thể hiện phân rã theo cây thông tin.

Để hình thức hoá, trừu tượng hoá dữ liệu đa phương tiện, người ta cần đến các kĩ thuật như cây k-chiều, cây tứ phân theo điểm, cây tứ phân, cây R. Đồng thời còn phải quan tâm đến các phương pháp cài đặt các kĩ thuật này.

Định nghĩa: Không gian, thời gian: Không gian là mở rộng ba chiều, không giới hạn, trong đó các đối tượng và các sự kiện xảy ra và có vị trí và hướng tương đối. Thời gian là một phần của hệ thống đo, dùng cho các sự kiện tuần tự, để so sánh thời lượng của các sự kiện và khoảng thời gian giữa chúng, và để lượng hóa tỉ lệ thay đổi chuyển động của các đối tượng.

1.4.1. Giới thiệu

Dạng tổ chức dữ liệu địa lí thông dụng như GIS, thể hiện ở dạng bản đồ, tức các điểm. Mỗi điểm không gian được thể hiện qua cấu trúc dữ liệu.

Dữ liệu địa lí đa dạng, có thể thấy như điểm, đường thẳng, các đường cong, các hình, các bản đồ, các dữ liệu gắn với bản đồ theo các tầng địa lí khác nhau. Các thông tin này thay đổi theo thời gian...

1.4.2. Cây k-D

Sử dụng cây k-d để lưu trữ các điểm của bản đồ.

Định nghĩa: Cây là cấu trúc dữ liệu thông dụng, thể hiện cấu trúc dữ liệu phân cấp với tập các nút liên kết nhau. Về toán học, cây là đồ thị không xoay vòng của các nút; mỗi nút có các nút con, và thuộc về một nút cha.

1.4.2.1. Cấu trúc nút

Nút của cây k-d có cấu trúc

Kiểu dữ liệu nút = record

INFO: kiểu thông tin;

XVAL: real;

YVAL: real;

LLINK: ↑ kiểu nút;

RLINK: ↑ kiểu nút;

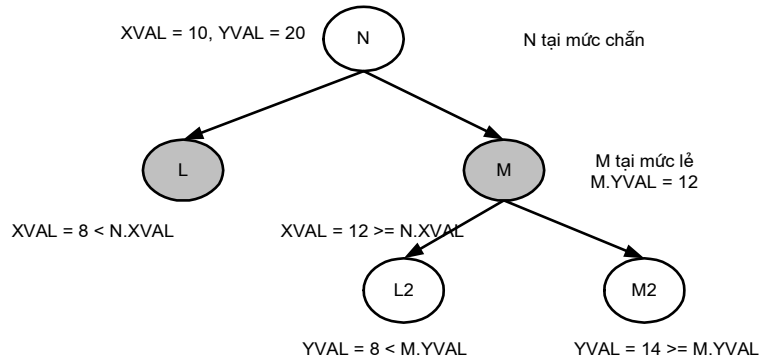
End;

Các mối nối RLINK và LLINK cho phép trở tới các nút con phải và trái của một nút.

Giả sử T là trở đến gốc cây k-d. Nếu N là nút cây thì người ta xác định thêm mức¹ của nút N: Mức (N) = 0 nếu N là gốc cây; mức (P) + 1 khi N là nút con của nút cha P.

Định nghĩa: Cây 2-d là cây nhị phân thoả mãn (i) Khi N là nút cây mức chẵn thì nút cây con trái M có $M.XVAL < N.XVAL$ và nút L bên phải có $L.XVAL \geq N.XVAL$; (ii) Khi N là nút cây mức lẻ thì nút cây con trái M có $M.YVAL < N.YVAL$ và nút L bên phải có $L.YVAL \geq N.YVAL$.

¹ level



Hình. Qui định các giá trị X, Y của các nút con

1.4.2.2. Bổ sung và tìm kiếm trên cây 2-d

Yêu cầu thêm nút N vào cây đã có. Cây này có trục T , ngầm hiểu con trục trục đến nút T . Người ta thấy có các khả năng:

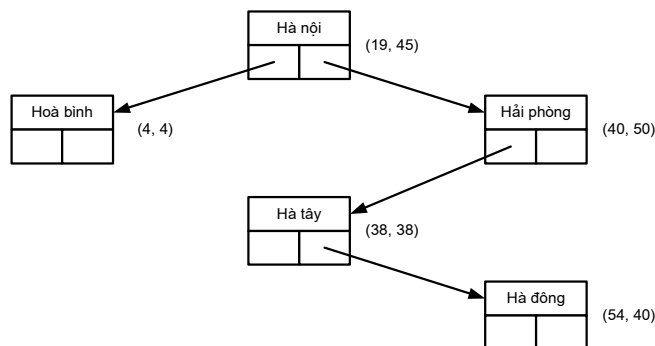
1. Nếu $N \equiv T$ thì hiển nhiên đã có nút N trên cây;
2. Ngược lại, người ta sang trái nếu $N.XVAL < T.XVAL$, do nút T là nút chẵn (0); và sang phải nếu $N.XVAL$ không nhỏ hơn;
3. Tại nút vừa đến P , nếu $P \equiv N$ thì đã xong, ngược lại do P là nút lẻ, nên căn cứ vào giá trị $YVAL$ để quyết định sang trái hay phải. Nếu $N.YVAL < p.YVAL$, người ta sang trái; ngược lại sang phải.
4. Tiếp tục từ bước 2, cho đến khi dừng hay đến nút là thì bổ sung nút N .

Việc tìm kiếm nút N với giá trị $XVAL$ và $YVAL$ được thực hiện như dò tìm để bổ sung nút mới.

Thí dụ:

Người ta có bản đồ với các điểm Hà nội (19, 45), Hải phòng (40, 50), Hà tây (38, 38), Hà đông (54, 40), và Hoà bình (4, 4). Nếu xuất phát từ Hà nội, người ta xây dựng được cây thông tin.

Dựa trên cây 2-d đã có, người ta có thể tiến hành (i) tìm kiếm; (ii) bổ sung nút mới theo cách trên.

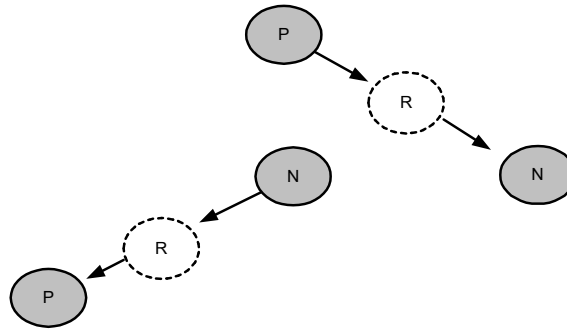


Hình. Các nút trên cây 2-d

1.4.2.3. Xóa nút trên cây 2-d

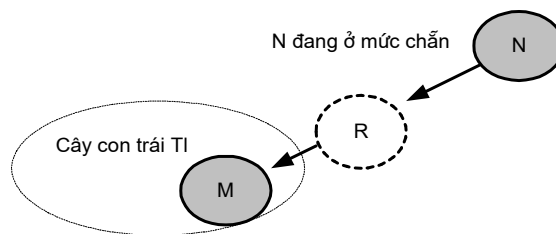
Vấn đề đặt ra là: giả sử cây cần xóa nút có trục T ; nút cần xóa có tọa độ (x, y) .

1. Bước đầu tiên cần tìm ra nút (x, y) , thoả mãn $N.XVAL = x$ và $N.YVAL = y$. Bước này khẳng định có nút N cần xoá, được thực hiện theo cách đã nêu trong mục trước;
2. Nếu nút N là nút lá, người ta huỷ nút N bằng cách thay đổi con trỏ của các nút cha trỏ đến nút con N . Các trỏ của nút cha sẽ là NIL ;
3. Nếu N là nút trung gian, công việc phức tạp hơn.



Hình. Chọn nút R thay thế N

- i. Khi $N.LLINK \neq NIL$, tức N có cây con trái T_l , và $N.RLINK \neq NIL$, tức có cây con phải T_r : cần tìm nút R trong T_l hay T_r để thay thế N , và R bị xoá khỏi cây $\{T_l, T_r\}$. Do vậy người ta thực hiện ba bước nhỏ sau:
 - Bước 1. Tìm nút R để thay thế $\{T_l, T_r\}$;
 - Bước 2. Thay trường nối $LLINK$, $RLINK$ không rỗng của N vào các trường nối của R ;
 - Bước 3. Xoá R ra khỏi $\{T_l, T_r\}$;
 - Quá trình này sẽ kết thúc do có $T_i \in \{T_l, T_r\}$ với mức cao hơn;
- ii. Vấn đề đặt ra là bước thứ nhất tìm R thay thế $\{T_l, T_r\}$ cần có quan hệ không gian đối với tất cả các nút P trong $\{T_l, T_r\}$, mà N sinh ra, cho đến nút P này.
 - Nếu P thuộc góc tây nam của N thì P là tây nam của R ;
 - Nếu P là tây bắc của N thì P là tây bắc của R ;
 - Điều này có nghĩa R thay thế có tính chất
 - a. Mỗi M trong $T_l: M.XVAL < R.XVAL$ nếu N mức chẵn; $M.YVAL < R.YVAL$ nếu N mức lẻ;

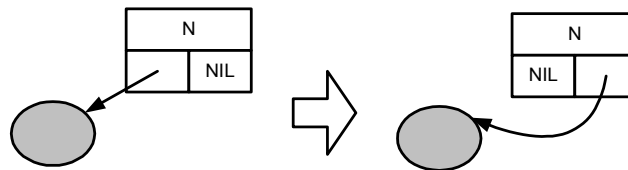


Hình. Vị trí nút thay thế R khi N tại mức chẵn

- b. Mỗi M trong T_r thoả mãn: $M.XVAL \geq R.XVAL$ nếu N mức chẵn; ngược lại khi N mức lẻ.
- iii. Khi $T_r \neq NIL$, xét trường hợp mức chẵn lẻ của N :

- Nếu N mức chẵn thì bất kì nút nào trong T_r có giá trị XVAL nhỏ nhất là nút R thay thế. Chẳng hạn nút Hà nội bị loại thì nút Hà tây thay thế;
 - Nếu N mức lẻ, nút thay thế trong T_r có YVAL bé nhất.
- iv. Nhìn chung người ta thực hiện “tìm nút thay thế trong cây con trái chỉ với một vài điều kiện”.
- Nếu N ở mức chẵn, nút phù hợp trong T_l là nút có XVAL lớn nhất;
 - Nếu N mức lẻ, người ta chọn nút có YVAL đạt cực đại;
- v. Vấn đề đặt ra khi trong T_l có nhiều nút đạt giá trị cực đại, tại XVAL hay YVAL, thì điều (2) trong định nghĩa cây 2-d sẽ bị vi phạm khi áp dụng ba bước thực hiện trên. Do vậy khi cần xoá nút N trung gian, nên tìm nút thay thế R trên cây con phải, bởi vì việc tác động vào cây con trái hay gây mất bền vững cấu trúc cây;
- vi. Khi cây con phải $T_r = \text{NIL}$, người ta chọn R thay thế trong T_l với giá trị x nhỏ nhất trong T_l khi N là nút tại mức chẵn, hay chọn R thay thế có y nhỏ nhất khi N mức lẻ; và trong ba bước thực hiện nhỏ trên, thay

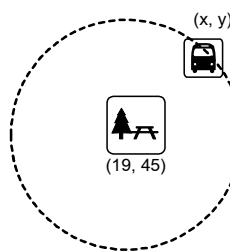
Bước 2. Thay tất cả trường không rỗng của N bằng các trường của R. Đặt $N.\text{RLINK} = N.\text{LLINK}$, $N.\text{LLINK} = \text{NIL}$;



Hình. Cấu trúc nút

1.4.2.4. Câu hỏi về phạm vi trong cấu trúc cây 2-d

Vấn đề là tìm các nút (x, y) trong phạm vi bán kính r từ (x_0, y_0) . Sẽ có vòng tròn bán kính r với tâm là điểm (x_0, y_0) .



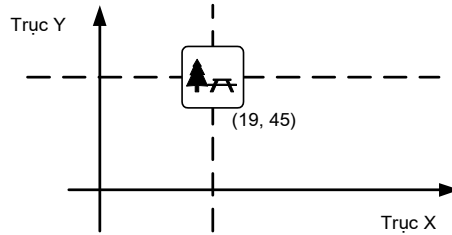
Hình. Miền cần tìm khi có nút (19, 45)

Liên quan đến câu hỏi này, người ta thấy mỗi nút N có miền RN với bán kính r . Vòng tròn tạo ra từ (x_0, y_0) không giao với vòng tròn RN thì không xét cây con của N.

Khái niệm về miền liên quan đến một nút, có 5 loại:

1. Xét nút Hà nội (19, 45), người ta tạo được miền Hà nội = $\{(x, y)\}$;
2. Nút Hải phòng cho biết miền Hải phòng = $\{(x, y) \mid x \geq 19\}$;

3. Nút Hà tây cho biết miền Hải tây = $\{(x, y) \mid x \geq 19, y < 50\}$;
4. Nút Hà đông cho biết miền Hà đông = $\{(x, y) \mid x \geq 38, y < 50\}$;
5. Nút Hoà bình cho biết miền Hoà bình = $\{(x, y) \mid x < 19\}$.



Hình. Xác định các biên

Nhìn chung mỗi nút N có nhiều nhất 4 ràng buộc liên quan, gắn với việc xác định vùng:

1. Xác định biên dưới về X, XLB^1 , có dạng $x \geq c_1$;
2. Xác định biên trên về X, XUB^2 , có dạng $x < c_2$;
3. Xác định biên dưới về Y, YLB^3 , có dạng $y \geq c_3$;
4. Xác định biên trên về Y, YUB^4 , có dạng $y < c_4$.
5. Do vậy cần mở rộng kiểu dữ liệu về nút, với tên là kiểu nút dữ liệu mới:

Kiểu nút mới = record

INFO: kiểu thông tin;

XVAL, YVAL: real;

XLB, XUB, YLB, YUB: $\text{real} \cup \{-\infty, +\infty\}$;

LLINK, RLLINK: ↑ kiểu nút mới;

End;

Việc bổ sung các nút cần thực hiện:

1. Góc cây nhận $XLB: = -\infty$, $YLB: = -\infty$ và $XUB: = +\infty$, $YUB: = +\infty$;
2. Nếu nút N có cha P và P tại mức chẵn, thì các giá trị phạm vi của N được xác định theo P:
 - $N.XLB = P.XLB$ nếu N là con trái của P; ngược lại N là con phải, $N.XLB = P.XVAL$;
 - $N.XUB = P.XVAL$ nếu N là con trái của P; ngược lại N là con phải, $N.XUB = P.XUB$;
 - $N.YLB = P.YLB$;
 - $N.YUB = P.YUB$;
3. Nếu N có cha là P và mức P là lẻ, thì:
 - $N.YLB = P.YLB$ nếu N là con trái của P; ngược lại N là con phải, $N.YLB =$

¹ XLB lower bound on x

² XUB upper bound on x

³ YLB lower bound on y

⁴ YUB upper bound on y

P.YVAL;

- N.YUB = P.YVAL nếu N là con trái của P; ngược lại N là con phải, N.YLB = P.YUB;
- N.XLB = P.XLB;
- N.XUB = P.XUB;

Đối với việc tìm kiếm vùng trên cây, người ta có nhận xét:

- Khi gốc T không giao, thì xét cây con trái, cây con phải;
- Với cây con không giao, có thể loại bỏ trong quá trình tìm kiếm.

1.4.2.5. Cây k-d tổng quát

Cây k-d tổng quát có $k \geq 2$. Cây 2-d được dùng để thể hiện các nút trong không gian 2 chiều. Cây k-d với $k \geq 2$ được dùng cho không gian k chiều, chẳng hạn các điểm (x, y, z) sử dụng cây 3-d. Đối với cây tổng quát, không thể sử dụng XVAL, YVAL để trở đến hai cây con, mà dùng vecto VAL k chiều, ứng với bộ các con trở đến các cây con.

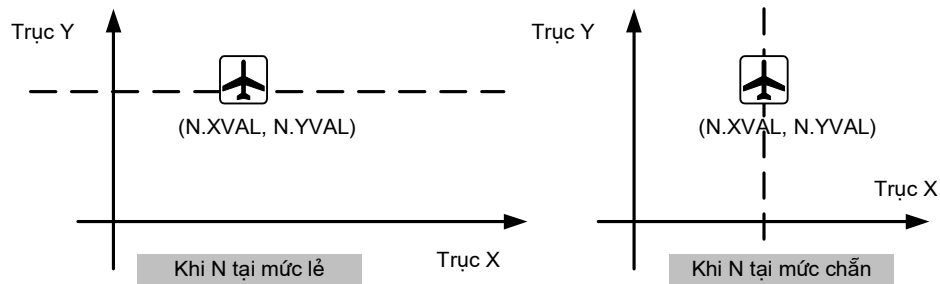
Định nghĩa: Cây T có cấu trúc nút cây k-d nếu đối với nút N trên cây T, thì:

- Nút N có mức i. Mức này được tính $i = \text{"mức thực sự"} \bmod k$;
- nút M trên cây con trái của N, $M.VAL[i] < N.VAL[i]$;
- nút P trên cây con phải của N, $P.VAL[i] \geq N.VAL[i]$.

Các thuật toán đã sử dụng cho cây 2-d được phát triển dùng cho cây k-d. Khi $k = 1$, người ta làm việc với cây nhị phân quen thuộc.

1.4.3. Cây tứ phân

Cây tứ phân¹, hay cây bốn phần, được dùng để thể hiện điểm không gian 2 chiều. Cây 2-d cũng có tác dụng như vậy. Tuy nhiên điểm khác biệt là cây tứ phân chia miền thành 4 phần, trong khi cây 2-d cho phép tách miền ra hai phần.



Hình. Vai trò của trục ngang và dọc khi chia các miền

Bốn phần ứng với bốn cây con trên cây tứ phân được gọi tên theo hướng đối với nút N: đông bắc NE, tây bắc NW, tây nam SW, và đông nam SE².

Kiểu dữ liệu được mô tả là:

Kiểu nút tứ phân = record

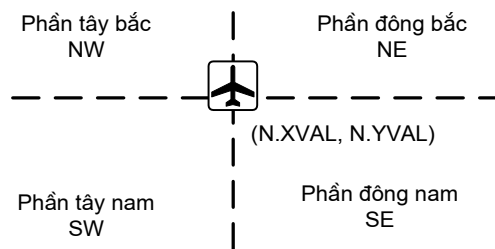
INFO: kiểu thông tin;

XVAL: real;

¹ point quadtree

² north east, north west, south west, south east

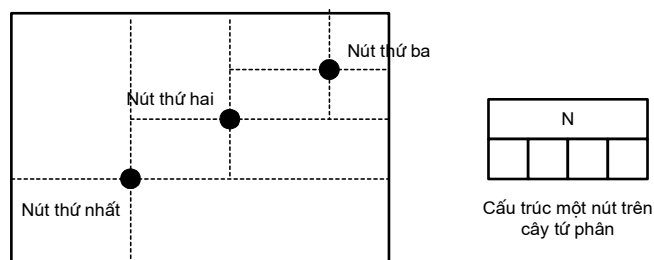
YVAL: real;
 NW, SW, NE, SE: ↑ kiểu nút tứ phân;
 End;



Hình. Bốn phần so với một nút

1.4.3.1. Tìm kiếm thông tin và bổ sung nút trên cây tứ phân

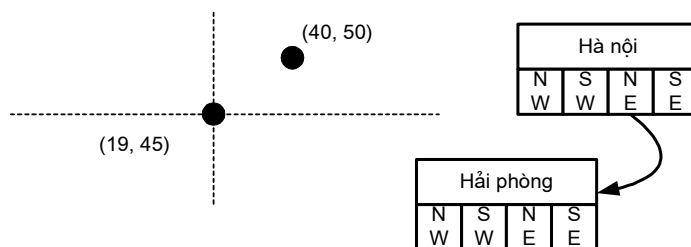
Khi có các giá trị điểm không gian, như trong thí dụ trước với bản đồ các điểm Hà nội (19, 45), Hải phòng (40, 50), Hà tây (38, 38), Hà đông (54, 40), và Hoà bình (4, 4), người ta có thể xây dựng được cây bằng cách phát triển dần các nút. Chẳng hạn từ nút Hà nội, người ta dùng 4 con trỏ đến các nút khác, theo vị trí tương đối so với nút Hà nội.



Hình. Nguyên tắc mô tả nút cây tứ phân

Việc xây dựng cây được tiến hành theo:

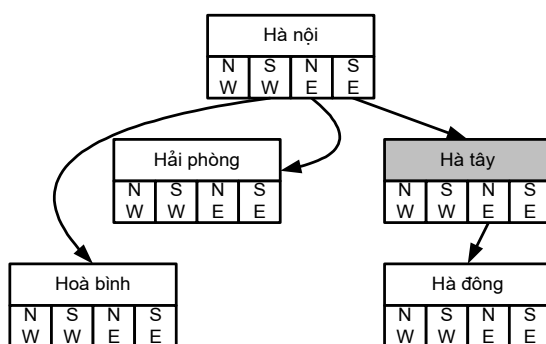
1. Cây rỗng, người ta xác định nút đầu là Hà nội, toạ độ (19, 45);
2. Chia ra 4 phần không gian;



Hình. Thêm nút Hà nội và xác định được vị trí tương đối của Hải phòng so với Hà nội

3. Bổ sung nút Hà tây. Do phần này chưa có nút nào, con trỏ từ nút Hà nội trỏ trực tiếp đến nút Hà tây;
4. Với nút Hà đông: do đã có nút Hà tây tại SE của nút Hà nội, người ta nhận thấy chính nút Hà tây cho phép tách không gian ra 4 phần, và nút Hà đông là NE của nút Hà tây;

5. Việc bổ sung nút Hoà bình đơn giản.



Hình. Các nút cây tứ phân

Nhìn chung, trong trường hợp xấu nhất, n nút tạo nên độ cao $n-1$ mức. Lưu ý rằng thời gian tìm kiếm nút, bổ sung nút được biểu diễn tuyến tính theo số nút n .

1.4.3.2. Xóa một nút trên cây tứ phân

Khi nút N cần xóa là nút lá thì công việc đơn giản. Nhưng khi N là nút trung gian, việc tìm nút thay thế rất phức tạp, do:

- Mỗi nút gắn với miền, vùng của bản đồ, có cách xác định khác cây 2-d. Trong cây 2-d, người ta đã dùng các ràng buộc $x \geq c_1$, $x < c_2$, $y \geq c_3$, và $y < c_4$;
- Trong cây tứ phân, mô tả kiểu dữ liệu là

Kiểu nút tứ phân = record

INFO: kiểu thông tin;

XVAL, YVAL: real;

XLB, YLB, XUB, YUB: $\text{real} \cup \{-\infty, +\infty\}$

NW, SW, NE, SE: ↑ kiểu nút tứ phân;

End;

Việc xóa nút cần lưu ý đến việc bổ sung nút. Việc bổ sung một nút N vào cây có gốc T được thực hiện:

- Nếu N là gốc cây T , thì $N.XLB = -\infty$, $N.YLB = -\infty$, $N.XUB = +\infty$, $N.YUB = +\infty$;
- Nếu P là nút cha của con n , thì việc gán giá trị theo tham chiếu đến bảng sau, tùy theo vị trí tương đối của con N đối với P . Trong bảng người ta sử dụng kí hiệu độ rộng của hình bao $w = P.XUB - P.XLB$, và chiều cao của hình bao $h = YUB - YLB$.

Trường hợp	N.XLB	N.XUB	N.YLB	N.YUB
N là con tây bắc	P.XLB	$P.XLB + w / 2$	$P.YLB + h / 2$	P.YUB
N là con tây nam	P.XLB	$P.XLB + w / 2$	P.YLB	$P.YLB + h / 2$
N là con đông bắc	$P.XLB + w / 2$	P.XUB	$P.YLB + h / 2$	P.YUB
N là con đông nam	$P.XLB + w / 2$	P.XUB	P.YLB	$P.YLB + h / 2$

Hình. Bảng tra cứu giá trị cận trên/ dưới cho các nút con

Trong quá trình tìm nút R thay thế N trong số các cây con, để xóa N , người ta nhận

thấy trường hợp xấu nhất là phải tạo lại một số nút, mà N đã trở đến. Việc này sẽ tránh công sức tạo lại các nút con của N.

Với cấu trúc dữ liệu cây tứ phân MX, người ta có thể thực hiện việc xoá một nút dễ dàng hơn.

1.4.3.3. Câu hỏi về phạm vi trên cây tứ phân

Việc hỏi dữ liệu trên cây tứ phân được thực hiện qua việc tìm dữ liệu trên cây, với qui trình như đối với cây 2-d.

Có thể thể hiện việc hỏi dữ liệu này qua chương trình giả lập như sau:

```
Proc Miền (T: kiểu dữ liệu nút tứ phân, C: hình tròn);
Nếu miền (T)  $\cap$  C = rỗng thì dừng;
Ngược lại:
    Nếu (T.XVAL, T.YVAL)  $\in$  C thì in (T.XVAL, T.YVAL);
    Ngược lại
        Miền (T, NW, C);
        Miền (T, SE, C);
        Miền (T, NE, C);
        Miền (T, SW, C);
End Proc;
```

1.4.4. Cây tứ phân MX

1.4.4.1. Giới thiệu

Trong hai cấu trúc dữ liệu cây vừa giới thiệu, là cây k-d và cây tứ phân, người ta thấy hình dáng của cây lệ thuộc vào việc (i) chọn nút nào là nút gốc; (ii) quá trình bổ sung các nút. Do vậy đôi khi cây có hình dạng tự do, và không cân đối, không tiện cho việc tổ chức dữ liệu cũng như tìm kiếm dữ liệu trên cây.

Việc chia miền với cấu trúc cây trên đã tuân theo số chặn miền, tức chia một miền thành 2 miền, thành 4 miền. Tuy nhiên, người ta thấy tùy theo vị trí của nút N, có thể chia miền thành số lẻ miền con. Đó là tiếp cận của cây tứ phân MX.

Mục đích sử dụng cây tứ phân MX là:

- Đảm bảo hình dáng cây, hay độ cao, không lệ thuộc vào các nút có mặt trên cây, cũng như thứ tự bổ sung, phát triển cây;
- Đảm bảo xoá và tìm kiếm dữ liệu hiệu quả hơn.

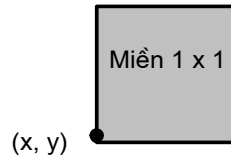
Người ta làm việc với cây tứ phân MX theo:

1. Giả sử bản đồ đã được chia thành lưới $2^k \times 2^k$, với k là số nguyên. Người dùng tự do chọn k để tiện lợi cho họ rồi sau này sẽ cố định giá trị k;
2. Cây có cấu trúc nút như cây tứ phân, khác ở chỗ gốc cây mang các giá trị
3. $XLB = 0$; $XUB = 2^k$; $YLB = 0$; $YUB = 2^k$;
4. Chia miền tại giữa miền: nếu N là nút thì 4 nút con của N có thông số, trong đó $w = N.XUB - N.XLB$. ở đây người ta không phân biệt độ rộng và độ cao;

Trường hợp	XLB	XUB	YLB	YUB
N là con tây bắc	N.XLB	$N.XLB + w / 2$	$N.YLB + w / 2$	$N.YLB + w$

N là con tây nam	N.XLB	$N.XLB + w / 2$	N.YLB	$N.YLB + w / 2$
N là con đông bắc	$N.XLB + w / 2$	$N.XLB + w$	$N.YLB + w / 2$	$N.YLB + w$
N là con đông nam	$N.XLB + w / 2$	$N.XLB + w$	N.YLB	$N.YLB + w / 2$

Hình. Bảng tra cứu giá trị cận trên/ dưới cho các nút con



Hình. Miền 1 x 1

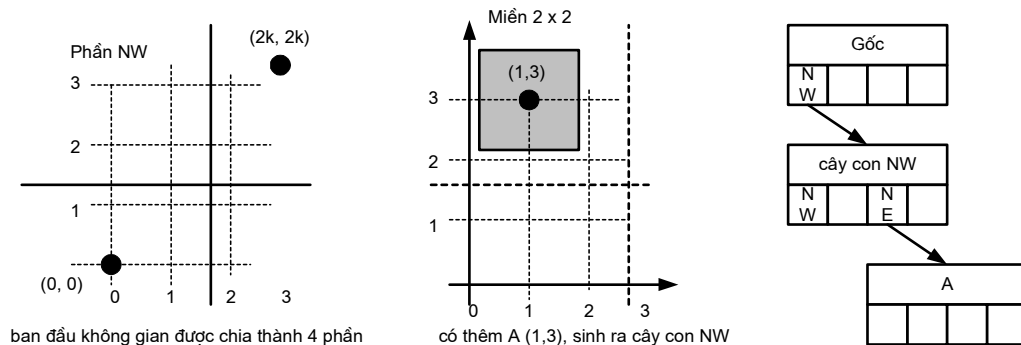
1.4.4.2. Tìm kiếm và bổ sung trên cây tứ phân MX

Nguyên tắc tìm kiếm dữ liệu và bổ sung nút cây được thực hiện theo nguyên tắc: mỗi điểm (x, y) trên cây thể hiện miền 1 x 1, có góc trái dưới là (x, y) . Mỗi điểm được bổ sung vào nút được kèm theo miền 1 x 1 ứng với điểm đó.

Thí dụ bổ sung các điểm A, B, C, và D. Toạ độ của chúng: A (1, 3), B (3, 3), C (3, 1), và D (3, 0.5).

Đầu tiên, bổ sung A vào toạ độ (1, 3).

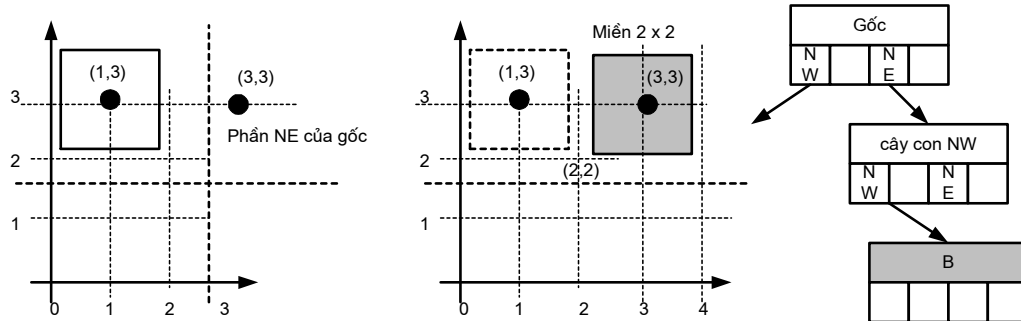
- Gốc cây thể hiện cả vùng; A thuộc vào miền NW của nó;
- Do vậy gốc cây con NW ứng với miền 2 x 2, với góc trái thấp là điểm (0, 2);
- A thuộc vào miền NE của miền này;
- A được bổ sung vào mức 2 của cây, tức là mức k. Nhìn chung các điểm luôn được bổ sung vào lớp k của cây;



Hình. Hình dáng miền cần phân chia khi bắt đầu bổ sung nút đầu tiên

Thứ hai, bổ sung nút B (3, 3).

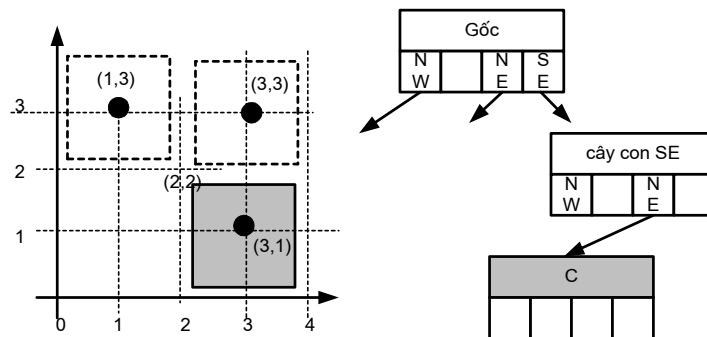
- Xét trên cây vừa được tạo sau khi bổ sung nút A: từ nút gốc người ta thấy B ở vị trí NE của gốc, cần sinh ra nhánh NE;
- Gốc cây con NE tạo ra miền 2 x 2 với góc trái thấp là điểm toạ độ (2, 2);
- Nút B thuộc miền NE của miền vừa tạo;



Hình. Các miền gắn với các nút và cách chia các miền khi có các nút mới

Thứ ba, bổ sung nút C (3, 1). Khi đã có hai miền NW và NE của gốc, người ta thấy C thuộc miền SE của gốc.

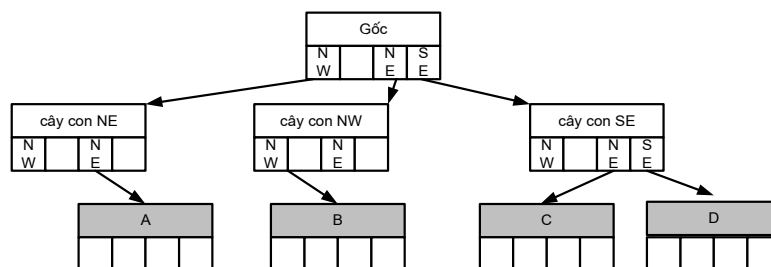
- Do chưa có cây con SE, cần tạo cây con SE của gốc;
- Trong miền mới này, điểm C thuộc phần NE;



Hình. Bổ sung nút C khi đã có A, B chia không gian ra 4 phần một cách tương đối

Thứ tư, nút D (3, 0.5) có vị trí SE của gốc, thuộc cây con SE cùng với C.

- Cây con SE của gốc đã có;
- Nút D nằm tại miền SE của cây con này.



Hình. Kết quả cuối cùng trên cây tứ phân MX khi bổ sung A, B, C và D

1.4.4.3. Xóa một nút trên cây tứ phân MX

Người ta nhận thấy các nút dữ liệu đều là nút lá của cây. Do vậy việc xóa một nút trên cây tứ phân MX đơn giản.

Lưu ý rằng nếu nút N là nút trung gian trong cây, có trở T trở đến gốc thì miền do N thể hiện chứa ít nhất một điểm có nội dung trên cây. Khi muốn xóa N (x, y) khỏi cây T, cần đảm bảo tính chất các con trở đến N từ các nút cha sẽ thay bằng NIL.

1.4.4.4. Câu hỏi về miền trên cây tứ phân MX

Câu hỏi về miền đối với cây tứ phân MX được thể hiện như đối với cây tứ phân. Một chút khác nhau đối với cây tứ phân MX là:

- Nội dung của XLB, XUB, YLB, YUB khác;
- Do các điểm được lưu trữ tại mức lá, nên việc kiểm tra điểm nằm trong hình tròn không được xác định theo nhu cầu hỏi đối với nút lá.

1.4.4.5. Cây tứ phân PR

Cây tứ phân PR là một dạng khác của cây tứ phân MX.

Với cây tứ phân MX, người ta nhận thấy các điểm dữ liệu đều được lưu tại mức lá. Do vậy với cây có số miền là $2^k \times 2^k$, quá trình tìm kiếm và bổ sung có độ phức tạp theo $O(k)$.

Với mỗi nút N trên cây tứ phân MX, người ta có một miền, chẳng hạn gọi là m (N). Khi phát triển cấu trúc dữ liệu để lưu trữ các điểm, cần thay đổi qui luật tách cây tứ phân MX để nút N được tách khi và chỉ khi m (N) chứa hai hay nhiều điểm. Nếu N chỉ có một điểm thì tự nút đó là nút lá và đại diện cho điểm. Thay đổi này đối với cây tứ phân PR nhằm giảm thời gian tìm kiếm và bổ sung.

1.4.5. Cây R

1.4.5.1. Giới thiệu

Cây miền chữ nhật được sinh ra nhằm nhằm cho phép sử dụng cấu trúc thể hiện dữ liệu là các miền chữ nhật. Người ta gọi nó là cây R. Ngoài tác dụng trực quan trên, cây R có ưu điểm cho phép tổ chức thông tin trên đĩa từ, ít ra là ưu điểm trong việc giảm số lần truy cập đĩa từ..

Mỗi cây R có một thứ tự đi kèm, được thể hiện qua số nguyên K. Mỗi nút không phải nút lá trên cây chứa một tập, nhiều nhất có K hình chữ nhật, ít nhất $\lceil K/2 \rceil$ hình. Điều này dẫn đến các nút trung gian, không là nút lá hay nút gốc, chứa ít ra là một nửa số nút.

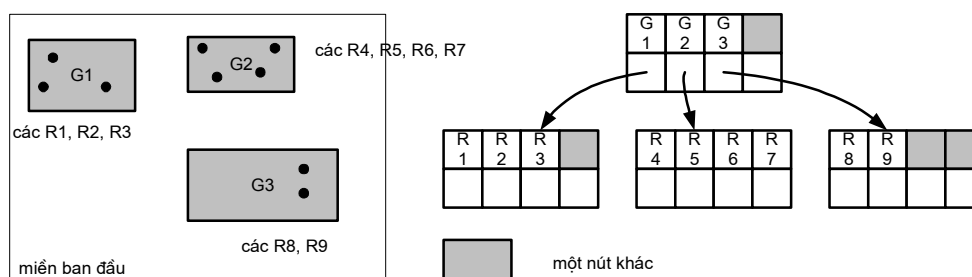
Hình chữ nhật vừa nêu là hình chữ nhật thực, hay nhóm các hình chữ nhật. Riêng nút lá là nút với một hình chữ nhật thực; còn nút trung gian chứa một nhóm các hình. Cấu trúc dữ liệu để mô tả một nút trên cây R:

Kiểu nút cây R = record

REC_1, \dots, REC_k : hình chữ nhật;

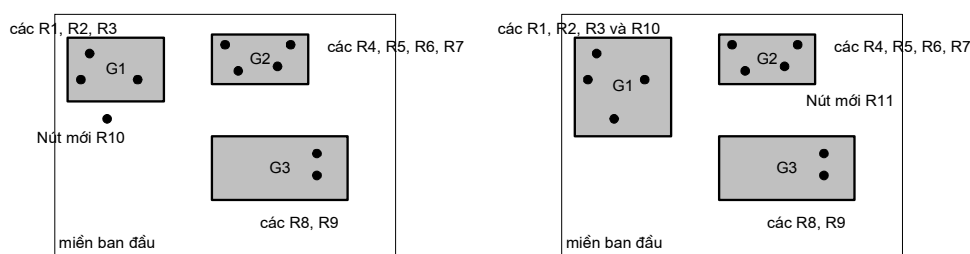
P_1, \dots, P_k : ↑ kiểu nút cây R;

End;



Hình. Các hình chữ nhật được tạo nhằm chứa các điểm

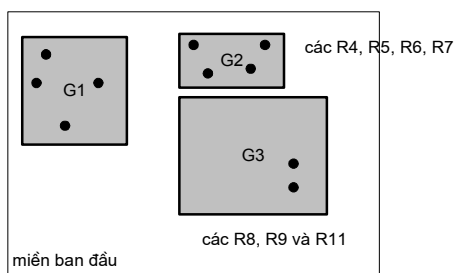
1.4.5.2. Bổ sung, tìm kiếm nút trên cây R



Hình. Các hình chữ nhật thay đổi kích thước khi có điểm mới

Để bổ sung một nút mới vào cây R, chẳng hạn nút R_{10} vào cây hình trên, người ta thực hiện các bước sau:

1. Xem xét các hình chữ nhật đang gắn với gốc, chọn hình chữ nhật cần nói rộng ra ít nhất, tức diện tích miền nói ra ít, mà đảm bảo các hình có vị trí tương đối tách biệt, rõ ràng với nhau. Trong trường hợp hình trên, việc nói G_1 là thuận lợi, do số diện tích phát triển để bao trùm R_{10} từ G_1 ít hơn từ G_2 hay G_3 ;
2. Bổ sung R_{10} ;
3. Giả sử cần bổ sung thêm R_{11} vào cây. Do vị trí của R_{11} , nói G_2 là hợp lí. Tuy nhiên G_2 đã đầy;
4. Cần xác định trong số các hình chữ nhật chưa đầy để bổ sung R_{11} . Trong trường hợp này, cần nói G_3 .

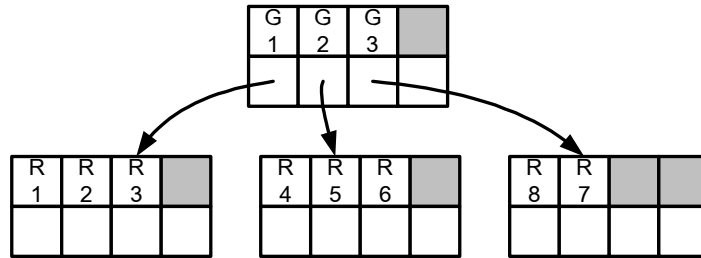


Hình. Các trường hợp của miền chung khi bổ sung nút mới R_{10} và R_{11}

1.4.5.3. Xoá nút trên cây

Xoá đối tượng, gắn với hình chữ nhật, trên cây R có thể làm mất nút trong đối tượng đó. Khi xoá một nút trong một hình chữ nhật cần đảm bảo điều kiện tối thiểu của cây R là hình chữ nhật chứa ít nhất $[K / 2]$ hình thực hay nhóm hình.

Chẳng hạn cần xoá nút R_9 trong cây R trên. Nếu huỷ R_9 trong hình G_3 , G_3 chỉ còn R_8 , vào tình trạng dưới tải. Điều này khiến người ta cần tổ chức lại cây R, tức tạo ra các nhóm logic mới. Có thể đề xuất phương án R_7 từ G_2 chuyển sang G_3 .



Hình. Cây chỉnh lý sau khi xoá nút R_3 trong G_3

1.4.6. So sánh các cấu trúc dữ liệu đa phương tiện

Để lựa chọn cấu trúc dữ liệu phù hợp cho các đối tượng đa phương tiện, người ta cần xem xét nhiều mặt. Tuy nhiên trong số một vài cấu trúc, người ta có thể so sánh để chọn ra cấu trúc phù hợp với các thông số đã xem xét khác.

Trong chương này, 4 kiểu cấu trúc đã được trình bày là (i) cây k-d; (ii) cây tứ phân; (iii) cây tứ phân MX; và cây R. Phân tích ưu nhược điểm các cấu trúc, người ta thấy:

1. Cây tứ phân dễ cài đặt. Nhìn chung một điểm gồm k nút có độ cao k. Điều này, như đã nhận xét, làm tăng độ phức tạp tìm kiếm và bổ sung điểm mới, tới $O(k)$.
 - Mỗi so sánh cần được thực hiện với hai toạ độ, chứ không chỉ một toạ độ;
 - Việc xoá trên cây tứ phân thường là khó, do việc tìm nút R thay thế;
 - Câu hỏi về phạm vi cần thời gian $O(2\sqrt{n})$, trong đó n là số bản ghi trên cây;
2. Cây k-d cũng dễ cài đặt. Tuy nhiên khi k lớn, việc cài đặt sẽ phức tạp hơn.
 - Thực tế cho thấy đường đi từ gốc trong cây k-d dài hơn trên cây tứ phân, do cây k-d dựa theo cây nhị phân;
 - Việc tìm kiếm trên cây k-d có độ phức tạp $O(k.n^{1-\frac{1}{k}})$. Khi $k = 2$ thì độ phức tạp là $O(2\sqrt{n})$, giống như đối với cây tứ phân;
3. Cây tứ phân MX sẽ có độ cao theo định nghĩa, nhiều nhất là $O(n)$. Miền được thể hiện qua $2^n \times 2^n$ phần tử;
 - Việc thêm, xoá hay tìm kiếm trên cây tứ phân MX có độ phức tạp $O(n)$;
 - Việc tìm miền dữ liệu trên cây tứ phân MX tỏ ra hiệu quả, có độ phức tạp $O(N + 2^h)$, trong đó N là số điểm cần trả lời, h là độ cao của cây;
4. Cây R có nhiều hình chữ nhật gắn với nút. Do vậy số lần truy cập đĩa từ bên ngoài sẽ tỉ lệ với độ cao của cây. Cây R là cấu trúc dữ liệu đa phương tiện được sử dụng nhiều, do tính phổ dụng của nó;
5. Cây R có nhược điểm về cấu trúc. Các biên của miền chữ nhật có thể trùm lên nhau. Điều này khiến việc tìm kiếm có thể lẫn sang miền khác, tức người ta có thể tìm kiếm theo nhiều đường khác nhau, nhất là các dữ liệu sát biên, hay dữ liệu nằm tại giao của các miền láng giềng;

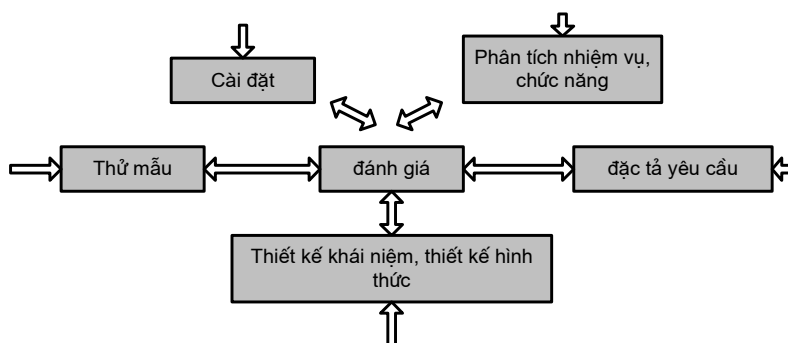
Cây R tỏ ra tốt hơn cây k-d và cây tứ phân trong các ứng dụng đa phương tiện. Ưu điểm này do cấu trúc dữ liệu cây R phù hợp với tổ chức thông tin trên bộ nhớ. Tuy

nhiên khi dùng chỉ số, cây tứ phân MX tỏ ra tốt hơn cả.

1.5. Ngôn ngữ thao tác dữ liệu đa phương tiện

1.5.1. Giao diện người dùng

Giao diện người dùng cần là giao diện tương tác, thân thiện. Ngoài khả năng tương tác người-máy, giao diện cho phép người dùng dễ dàng thao tác, điều khiển các chức năng.



Hình. Các pha thiết kế giao diện hướng người dùng, theo mô hình hình sao

Các thiết bị trong giao diện đáp ứng một số yêu cầu. Các yêu cầu này được nêu khi thiết kế giao diện.

- Biểu tượng, thể hiện sản phẩm đa phương tiện;
- Đòi nét về ý đồ thiết kế biểu tượng;
- Hiểu biết về khán giả đa phương tiện;
- Thử nghiệm tính đúng đắn của thiết kế giao diện;
- Xác định môi trường tương tác;
- Khả năng dùng âm thanh;
- Việc dùng màu sắc;
- Việc đóng gói các tư liệu liên quan.

1.5.2. Khả năng của hệ thống tìm kiếm và chỉ số hóa và ứng dụng

Qua các câu hỏi trên kiểu phương tiện, người ta thấy MIRS hứa hẹn nhiều năng lực và mềm dẻo. Các câu hỏi được phát biểu dưới dạng:

1. Câu hỏi dựa trên dữ liệu meta. Dữ liệu meta hướng đến các thuộc tính hình thức đối với mục tin cơ sở dữ liệu. Chẳng hạn “liệt kê các phim do đạo diễn Nguyễn làm, sau 1970”. Loại câu hỏi này nhờ DBMS xử lý.
2. Các câu hỏi dựa trên chú giải. Chú giải nhằm vào mô tả văn bản của nội dung mục tin. Các câu hỏi được đặt theo từ khóa hay dạng văn bản tự do và hệ thống tìm kiếm theo tương tự. Chẳng hạn “chỉ ra đoạn video mà diễn viên Trần đang chạy”. Loại câu hỏi này yêu cầu mục tin được chú thích phù hợp với kỹ thuật IR.
3. Các câu hỏi dựa trên mẫu hay đặc trưng dữ liệu. Mẫu dữ liệu hướng đến thông tin

thống kê về dữ liệu đa phương tiện, như mức ồn, phân bố màu, mô tả bề mặt. Chẳng hạn “cho xem đoạn phim có âm thanh có giai điệu dân ca Quan họ”. Do vậy, thông tin thống kê về mục tin cần được sưu tập trước.

4. Câu hỏi theo thí dụ. Các câu hỏi theo dữ liệu đa phương tiện, như các hình ảnh, mẫu âm thanh thuộc loại này. Chẳng hạn “cho xem phim có cảnh con sông quê hương”. Đối với loại câu hỏi này, cần có mối quan hệ không gian, thời gian giữa các đối tượng.
5. Các câu hỏi chuyên dụng. Nhiều ứng dụng chuyên sâu có câu hỏi riêng. Chẳng hạn “tìm chỉ tiết máy, với kích thước rộng 2.5 cm”

Liên quan đến ứng dụng của MIRS mà các câu hỏi nêu trên đáp ứng nhu cầu xử lý đa phương tiện.

- i. *Y học*. Người ta sử dụng ảnh siêu âm và chẩn đoán theo dấu hiệu trên ảnh.
- ii. *An ninh*. Theo các băng ghi tự động, người ta có thể phát hiện đối tượng bị theo dõi.
- iii. *Giáo dục*. Người dùng lựa chọn được phương tiện ưu dùng và tìm kiếm tri thức cần thiết. Tương tác người dùng cần thông minh trong hệ thống học.
- iv. *Báo chí*. Phương tiện quan trọng đối với báo chí, trợ giúp lấy lại tư liệu cũ, và lưu trữ các dữ liệu đa phương tiện.
- v. *Giải trí*. Không thể phủ nhận vai trò của đa phương tiện trong trò chơi, giải trí.
- vi. *Thương mại*. Việc kiểm định chất lượng, nhãn mác hàng hóa sẽ sử dụng kỹ thuật tìm kiếm đa phương tiện.

Về sau, MIRS tập trung vào chính thông tin, thay vì các kiểu phương tiện, và việc thể hiện thông tin có thể khớp hay chuyển đổi kiểu phương tiện này sang kiểu khác. Người ta xây dựng máy tìm kiếm để khớp mục tin trong cơ sở dữ liệu.

1.6. Kết luận

Chương này đã đề cập một số khái niệm chung về cơ sở dữ liệu đa phương tiện. Một số khái niệm sẽ được chi tiết hóa trong các chương sau.

Một số vấn đề được giới thiệu trong chương là:

- Khái niệm về đa phương tiện và dữ liệu đa phương tiện;
- Vai trò của hệ quản trị cơ sở dữ liệu đa phương tiện;
- Nhu cầu về xử lý dữ liệu đa phương tiện;
- Quan niệm về hệ thống xử lý thông tin đa phương tiện.

Chương 2. Tư liệu đa phương tiện tương tác

2.1 Cơ sở dữ liệu đa phương tiện tương tác

2.1.1. Giới thiệu

Trước phần này, người ta đã biết được lí do hệ quản trị cơ sở dữ liệu truyền thống không thể quản lí tốt các dữ liệu đa phương tiện, đã biết nhu cầu về ứng dụng của hệ thống MIRS và yêu cầu đối với dữ liệu đa phương tiện. Trong phần này, cần thiết mô tả khía cạnh về thiết kế và yêu cầu đối với MIRS để đáp ứng nhu cầu ứng dụng. Các khối trong hệ thống được thiết kế sẽ đảm nhiệm các dạng chính của phép toán trong MIRS về xử lí thông tin.

Định nghĩa: Tương tác: tương tác là một loại hành động xảy ra khi hai hay nhiều đối tượng có sự tác động lẫn nhau; ý tưởng tác động hai chiều là đặc biệt theo khái niệm tương tác, khác với hiệu ứng gây nên bởi một bên.



Hình. Thiết kế là sáng tạo

Mô hình dữ liệu của MIRS tác động đến mọi khía cạnh của thiết kế và thao tác, xác định cách tổ chức và lưu trữ thông tin, và dạng câu hỏi cần thiết. Mô hình dữ liệu trong MIRS sẽ được mô tả như phân cấp mô hình dữ liệu đa phương tiện và một số mô hình dữ liệu riêng. Kiến thức về thiết kế hướng đối tượng cần thiết đối với thiết kế cơ sở dữ liệu đa phương tiện. Giao diện người dùng cần được thiết kế phù hợp với dạng phương tiện và đặc trưng chủ quan, mờ của câu hỏi. Trong MIRS, các đặc trưng và các thuộc tính của các mục tin dữ liệu đa phương tiện được trích và lưu trữ theo chính các mục tin. Các đặc trưng này được tổ chức theo cấu trúc chỉ số để tìm kiếm hiệu quả; việc tìm kiếm dựa trên độ đo tương tự giữa đặc trưng hỏi và đặc trưng của mục tin. Các khía cạnh như (i) trích dữ liệu; (ii) chỉ số hóa; (iii) độ đo tương tự cũng được đề cập. Tìm kiếm và thể hiện đa phương tiện sẽ chịu ràng buộc không gian và thời gian. Các ràng buộc này được mô tả trong chất lượng dịch vụ QoS¹ về các đối tượng và ứng dụng đa phương tiện. Vậy nên cần thể hiện được các khái niệm QoS và mô tả cách đảm bảo chất lượng trong toàn hệ thống, kể cả hệ thống điều hành, hệ thống truyền thông, quản trị lưu trữ... Việc nén dữ liệu, chuẩn thể hiện và xử lí, trả lời câu hỏi cũng sẽ được trình bày.

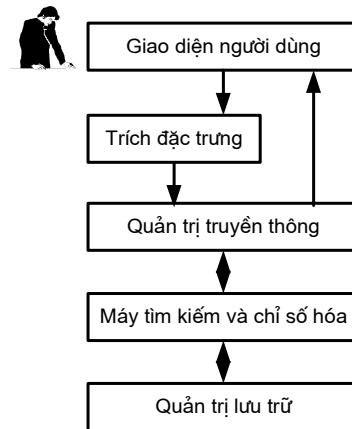
2.1.2. Kiến trúc của MIRS

Kiến trúc MIRS là mềm dẻo và mở rộng được, để dùng cho nhiều ứng dụng, nhiều

¹ quality of service

loại câu hỏi và nội dung. Để đáp ứng các yêu cầu, MIRS thông dụng có một số khối chức năng, được gọi là các hàm thư viện, hay các khối quản trị. Người ta có thể bổ sung các bộ phận quản trị để đáp ứng các chức năng cần có, cũng như có thể hủy bỏ vài thành phần.

Các đặc tính khác của MIRS như thường là phân tán, theo mô hình khách/ chủ. Đặc tính này đáp ứng (i) kích thước lớn của dữ liệu đa phương tiện, việc dùng bản sao dữ liệu là bình thường; (ii) cách sử dụng dữ liệu đa phương tiện, thường có nhiều người dùng cùng truy cập, như video theo yêu cầu.



Hình. Kiến trúc của MIRS

Trong hình là kiến trúc của MIRS, với các khối chính (i) giao diện; (ii) trích rút đặc trưng; (iii) quản trị truyền thông, tại hai đầu máy chủ và máy khách; (iv) máy chỉ số và tìm kiếm; (v) quản trị lưu trữ.

Định nghĩa: Kiến trúc : Kiến trúc là khoa học về thiết kế và dựng các ngôi nhà và các cấu trúc vật lý. Kiến trúc có nghĩa cung cấp dịch vụ kết nối chuyên nghiệp các kiến thiết công trình, nhóm không gian của các công trình.

Các chức năng của các khối được thể hiện qua kịch bản thao tác của MIRS. Hai phép chính, hay hai pha, là (i) bổ sung một mục tin đa phương tiện mới; (ii) tìm kiếm. Trong pha bổ sung, người dùng mô tả một hay nhiều nhóm mục thông qua giao diện. Các mục là các tệp được lưu trữ hay được nhập vào qua thiết bị như micro, CD, VCR hay máy quay. Người dùng có thể vẽ tranh làm đầu vào hệ thống. Nội dung hay các đặc trưng của các mục đa phương tiện được trích hoặc tự động hoặc bán tự động thông qua công cụ do bộ trích rút cung cấp. Các đặc trưng này và các mục ban đầu được gửi đến máy chủ qua bộ truyền thông. Tại các máy chủ, các đặc trưng được tổ chức, bổ sung, tùy theo lược đồ chỉ số để tìm kiếm hiệu quả nhờ máy chỉ số và tìm kiếm. Thông tin chỉ số hóa và các mục ban đầu được lưu trữ phù hợp với bộ lưu trữ.

Khi tìm kiếm thông tin, người dùng mô tả câu hỏi qua giao diện. Câu hỏi có thể là tệp đã lưu trên đĩa, hay được nhập qua thiết bị nhập. Giao diện người dùng cho phép duyệt các mục trong cơ sở dữ liệu và dùng các mục đã duyệt để hỏi. Nếu câu hỏi không là mục của cơ sở dữ liệu, đặc trưng chính của câu hỏi được lấy ra theo cùng cách với việc bổ sung mục. Các đặc trưng này được chuyển đến máy chủ qua bộ truyền thông. Máy chỉ số và tìm kiếm sẽ tìm trong cơ sở dữ liệu để thấy các mục trong cơ sở dữ liệu

khớp nhất với câu hỏi. Các mục được tìm theo nghĩa của bộ lưu trữ và được chuyển đến giao diện người dùng theo bộ truyền thông. Giao diện hiển thị danh sách các mục tới người dùng.

Trên là kiến trúc cơ bản. Các khối chức năng hay bộ quản trị cần thiết trong thực tế là:

1. Từ điển chuyên đề, cho biết các từ đồng nghĩa và mối quan hệ giữa các mục;
2. Cơ sở các luật toàn vẹn, cho phép kiểm tra toàn vẹn, ràng buộc đối với ứng dụng;
3. Quản trị ngữ cảnh, lưu các ngữ cảnh của ứng dụng.

2.1.3. Các mô hình dữ liệu

2.1.3.1. Nhu cầu về mô hình dữ liệu

Định nghĩa: Mô hình dữ liệu¹: tập các mô tả và qui luật mô tả tập dữ liệu

Trong DBMS, mô hình dữ liệu có vai trò cung cấp khung, hay ngôn ngữ, diễn tả các tính chất của mục tin được lưu trữ và tìm kiếm trong hệ thống. Khung này cho phép người thiết kế và người dùng xác định, thêm, bớt, sửa, và tìm kiếm các mục cơ sở dữ liệu và tính chất của chúng. Trong MIRS và hệ thống đa phương tiện, mô hình dữ liệu giả sử vai trò bổ sung là (i) mô tả; (ii) tính toán các mức trừu tượng khác nhau từ dữ liệu đa phương tiện.

Các mô hình dữ liệu đa phương tiện nắm bắt các tính chất tĩnh và động của các mục cơ sở dữ liệu, vậy nên cung cấp cơ sở hình thức để phát triển các công cụ phù hợp dùng với dữ liệu đa phương tiện.. Các tính chất tĩnh gồm (i) các đối tượng tạo nên dữ liệu đa phương tiện; (ii) mối quan hệ giữa các đối tượng; (iii) thuộc tính đối tượng. Các tính chất động gồm (i) tính chất liên quan đến tương tác giữa các đối tượng; (ii) phép toán trên các đối tượng; (iii) tương tác người dùng... Sự phong phú của mô hình dữ liệu có vai trò tăng tính sử dụng của MIRS. Dù các kiểu dữ liệu đa phương tiện cơ bản đã có, nhưng đó chỉ cung cấp nền để xây dựng các đặc trưng.

Không gian đặc trưng nhiều chiều là tính chất của chỉ số hóa đa phương tiện. Một mô hình dữ liệu trợ giúp thể hiện không gian nhiều chiều, đặc biệt độ đo khoảng cách trong không gian đó. Nhìn chung, mô hình dữ liệu MIRS cần đạt yêu cầu sau:

1. Mô hình dữ liệu cần mở rộng được, để bổ sung kiểu dữ liệu mới;
2. Mô hình dữ liệu cần thể hiện các kiểu phương tiện cơ bản và các đối tượng hỗn hợp với các mối quan hệ không gian và thời gian.
3. Mô hình dữ liệu cần mềm dẻo để có thể mô tả, hỏi và tìm kiếm các mục theo các mức trừu tượng;
4. Mô hình dữ liệu cho phép lưu trữ và tìm kiếm hiệu quả.

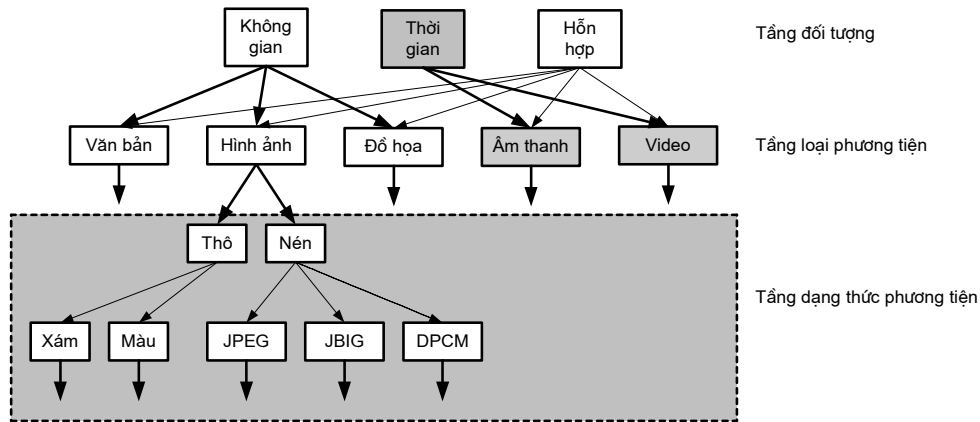
2.1.3.2. Mô hình dữ liệu đa phương tiện tổng quát

Nhiều người cho rằng mô hình dữ liệu MIRS cần là mô hình hướng đối tượng và phân cấp nhiều tầng. Thiết kế OO² đảm bảo đóng gói mã và dữ liệu vào cùng đơn vị,

¹ data model

² hướng đối tượng, Object Oriented

gọi là đối tượng. Mã này xác định các phép toán thực hiện trên dữ liệu. Việc đóng gói làm tăng tính khối và che đi chi tiết về phương tiện cụ thể và việc xử lý. Điều quan trọng hơn, tiếp cận OO cung cấp khả năng mở rộng do cơ chế tăng cường và mở rộng các đối tượng đã có.



Hình. Mô hình dữ liệu đa phương tiện tổng quát

2.1.3.2.1. Tầng đối tượng

Định nghĩa: Đối tượng đa phương tiện: Một đối tượng gồm một hay nhiều mục phương tiện với các mối quan hệ không gian, thời gian được mô tả.

Các đối tượng tập trung quanh chủ đề chính. Thí dụ về đối tượng đa phương tiện là trình diễn gồm các chiếu hình, trong đó có các hình ảnh và âm thanh đi kèm. Khía cạnh mấu chốt là cách mô tả các mối quan hệ không gian và thời gian. Các quan hệ không gian được mô tả nhờ hiển thị kích thước cửa sổ và vị trí của mỗi mục. Phương pháp chung đối với đặc tả thời gian là đặc tả dựa trên trục thời gian, có thời gian bắt đầu, thời lượng của mỗi mục theo đồng hồ chung. Các phương pháp khác gồm các mô hình theo sự kiện, theo kịch bản. Khi thể hiện, ngữ nghĩa của một đối tượng hiện lên khi hiện các mối quan hệ không gian và thời gian.

2.1.3.2.2. Tầng kiểu phương tiện

Tầng này mang các kiểu dữ liệu chung, như văn bản, đồ họa, hình ảnh, âm thanh và video. Các kiểu phương tiện này được suy ra từ lớp phương tiện trừu tượng. Mức này mô tả các đặc trưng hay thuộc tính. Chẳng hạn kiểu dữ liệu hình ảnh: có kích thước ảnh, lược đồ màu, các đối tượng chính bao gồm trong hình ảnh. Các đặc trưng này được dùng trực tiếp cho tính toán khoảng cách và tìm kiếm.

2.1.3.2.3. Tầng dạng thức phương tiện

Tầng dạng thức đa phương tiện mô tả các dạng phương tiện mà dữ liệu được lưu trữ. Một phương tiện thường có nhiều dạng thức có thể. Chẳng hạn một hình ảnh có thể ở dạng bitmap thô hay dạng nén. Cũng có nhiều kỹ thuật và chuẩn nén. Thông tin trên tầng này được dùng cho mã hóa, phân tích và thể hiện.

2.1.3.2.4. Khía cạnh khác

Ứng dụng khác nhau yêu cầu mô hình dữ liệu khác nhau, tùy theo các đặc trưng và các đối tượng trong các ứng dụng đó. Nhưng nhiều ứng dụng có thể chia sẻ mô hình dữ liệu chung nếu nó được thiết kế phù hợp và các đặc trưng, đối tượng mới có thể được bổ sung hay suy diễn từ mô hình cơ sở để đáp ứng yêu cầu của ứng dụng.

Lúc này, mỗi tầng trên của mô hình dữ liệu chưa được thiết kế đầy đủ. Chẳng có chuẩn cho việc này, mà tùy vào qui mô phát triển MIRS. Hầu hết MIRS là ứng dụng chuyên dụng, hướng về số có hạn các đặc trưng và sử dụng số phương tiện có hạn. Các mô hình dữ liệu được thiết kế như sáng tạo ngoài lề. Rất cần công sức để mô hình hóa dữ liệu đa phương tiện cho MIRS và hệ thống đa phương tiện.

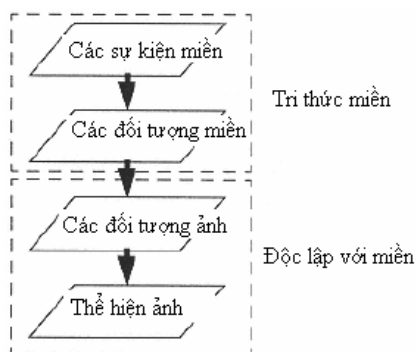
2.1.3.3. Các mô hình thí dụ

Mô hình dữ liệu nêu trên là mô hình tổng quát. Hiện thời, các mô hình dữ liệu riêng được dùng trong các ứng dụng khác nhau. Một số ý tưởng trong các mô hình riêng cũng được dùng trong mô hình dữ liệu đa phương tiện tổng quát.

2.1.3.3.1. Mô hình dữ liệu VIMSYS

Mô hình này nhằm quản lí thông tin hiển thị, tức hình ảnh và video. Nó gồm bốn tầng; mỗi đối tượng trên các tầng có tập các thuộc tính và phương pháp.

1. Tầng thể hiện ảnh. Tầng này chứa dữ liệu ảnh và bất kì chuyển hóa tác động đến thể hiện ảnh. Chẳng hạn chuyển hóa nén, đổi màu, tăng cường ảnh. Tầng này không mạnh để xử lí câu hỏi, mà chỉ quản lí câu hỏi cần đến thông tin dựa trên pixel. Tuy nhiên tầng thể hiện ảnh cung cấp dữ liệu thô cho tầng trên để trích và xác định đặc trưng mức cao.
2. Tầng đối tượng ảnh. Tầng đối tượng ảnh có hai tầng con (i) tầng phân đoạn; (ii) tầng đặc trưng. Tầng phân đoạn cô đặc thông tin về toàn bộ ảnh hay video vào các cụm không gian hay thời gian của các thuộc tính tóm tắt. Các thuộc tính cục bộ này có thể được tìm kiếm trực tiếp. Tầng con này liên quan chặt với kĩ thuật phân đoạn ảnh, video. Tầng con đặc trưng chứa các đặc trưng để tính toán nhanh, được tổ chức theo cấu trúc dữ liệu, tuân theo tính toán khoảng cách để gán tỉ số xếp hạng. Đặc trưng chung là phân bố màu, dạng đối tượng, và bề mặt.



Hình. Mô hình dữ liệu VIMSYS

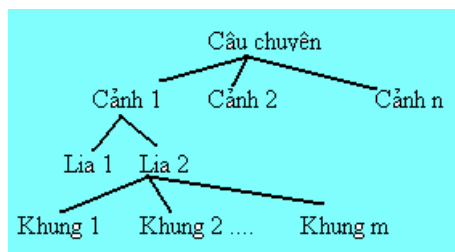
3. Tầng đối tượng miền. Đối tượng miền là thực thể do người dùng xác định để thể hiện thực thể vật lí hay khái niệm được suy từ một hay nhiều đặc trưng trong tầng thấp hơn. Tri thức về miền cần thiết để suy ra đối tượng miền.
4. Tầng sự kiện miền. Tầng này xác định các sự kiện mà người dùng có thể hỏi. Các sự kiện được xác định theo chuyển động của đối tượng, quan hệ thời gian, không gian giữa các đối tượng, việc hiện ra và biến đi của đối

tượng... Phát hiện sự kiện và cơ chế tổ chức cần dùng trong tầng này.

2.1.3.3.2. Mô hình video tổng quát

Mô hình video thường được dùng gồm bốn mức (i) khung; (ii) lia¹; (iii) cảnh, hay chuỗi; (iv) đoạn, tức tư liệu video. Một đoạn là đơn vị hoàn chỉnh, như một đoạn phim TV.

Định nghĩa: Khung là hình ảnh đơn trong video; Một lia là tập các khung được ghi bằng máy quay trong lần lia máy; Một cảnh là tập các lia với cùng ngữ nghĩa; Một đoạn là tập các cảnh.



Hình. Mô hình video tổng quát

Các thuộc tính có thể gán cho mỗi mức video. Tại mức đoạn video, người ta có thể gán dữ liệu sự kiện như (i) nhan đề; (ii) tác giả và ngày tháng; (iii) thông tin liên quan đến loại video. Một cảnh mang ngữ nghĩa chung mà các lia cùng chia sẻ. Các lia được đặc trưng bằng các khung chính, mang tính thể hiện, và các dữ liệu khác như các đối tượng chính và thời gian lia, vị trí lia. Các khung đơn chứa dữ liệu ảnh thô cũng như thống kê ảnh, như phân bố màu.

2.1.3.3.3. Cấu trúc lược đồ hình ảnh Virage

Định nghĩa: Máy tìm kiếm Virage cung cấp khung mở để xây dựng hệ thống tìm kiếm theo nội dung

Cấu trúc lược đồ của máy Virage gồm ba mức (i) lược đồ; (ii) các sơ khởi; (iii) các kiểu dữ liệu. Lưu ý rằng các hệ thống tìm kiếm thông tin dựa trên nội dung cần tạo trừu tượng về thông tin thô dưới dạng các đặc trưng, rồi làm việc chỉ với các đặc trưng đó.

1. Trong hệ thống tìm kiếm hình ảnh, người ta xác định năm kiểu dữ liệu (i) các giá trị; (ii) các phân bố; (iii) các giá trị chỉ số; và (v) các đồ thị. Giá trị là tập các vectơ thể hiện tính chất bao trùm của hình ảnh, như màu chủ đạo. Phân bố thể hiện tính chất thống kê của hình ảnh, như phân bố màu sắc. Giá trị chỉ số định vị vùng hình ảnh và được thể hiện như tập chỉ số hóa của các vectơ. Một phân bố chỉ số là mẫu cục bộ, hay một phân bố, như dạng được nhấn mạnh của một vùng quan tâm. Một đồ thị thể hiện thông tin quan hệ như quan hệ giữa các vị trí không gian của hai vùng quan tâm trong ảnh, và có thể được cài đặt theo thuật ngữ của bốn kiểu dữ liệu kia. Do vậy các vectơ tạo nên kiểu cơ sở đồng nhất, cho phép thể hiện đặc trưng nội dung ảnh. Kiểu dữ liệu sơ cấp trong máy Virage là tổ hợp các vectơ. Phép toán chính trên kiểu dữ liệu này là:

¹ shot

- Tạo tổ hợp: tạo nên tổ hợp rỗng của các vecto;
 - tạo vecto: trích và tạo vecto đặc trưng riêng từ hình ảnh, nhờ hàm trích đặc trưng riêng;
 - Trích rút: truy cập phần tử trong tổ hợp;
 - Tính khoảng cách: so sánh hai vecto và cho biết độ đo khoảng cách giữa chúng theo hàm khoảng cách mô tả đối với kiểu vecto;
 - Kết hợp: tạo vecto mới bằng cách kết hợp hai vecto đã biết;
 - Hủy vecto; hủy tổ hợp: giải phóng bộ nhớ liên kết với vecto cụ thể; xóa tổ hợp khỏi bộ nhớ.
2. Mức tiếp của cấu trúc lược đồ của máy Virage là sơ khởi. Một sơ khởi được xác định như tổ hợp các vecto, cho phép thể hiện phạm trù đơn về thông tin ảnh. Do vậy một sơ khởi là một đặc trưng có ý nghĩa ngữ nghĩa của hình ảnh, như màu, bề mặt, hình dạng. Một sơ khởi được mô tả theo tên duy nhất, tên phạm trù, các chức năng quản trị và tìm kiếm dữ liệu.
 3. Mức cao nhất là cấu trúc lược đồ, xác định bởi tên lược đồ và tập các sơ khởi. Một ứng dụng sử dụng và dùng nhiều lược đồ. Mức này cung cấp các chức năng bổ sung ảnh và tìm kiếm theo nội dung.

2.1.4. Thiết kế giao diện người dùng

Giao diện cho phép người dùng sử dụng, tương tác và truyền thông với MIRS. Vậy giao diện phần nào xác định tính sử dụng của hệ thống. Chức năng chính của giao diện người dùng là cho phép bổ sung các mục vào cơ sở dữ liệu và đưa câu hỏi, thể hiện trả lời. Vậy giao diện tốt cần đạt các yêu cầu:

1. Cung cấp công cụ để người dùng bổ sung các mục cơ sở dữ liệu dễ dàng;
2. Cung cấp các công cụ để người dùng nhập câu hỏi và thông tin cần thiết vào hệ thống một cách hiệu quả;
3. Thể hiện kết quả câu hỏi cho người dùng hiệu quả;
4. Thân thiện.

Nhiều khía cạnh đáp ứng ba yêu cầu đầu tiên do đặc tính của dữ liệu và ứng dụng đa phương tiện.



Hình. Giao diện người dùng

2.1.4.1. Lực lượng cơ sở dữ liệu

Trong BDMS truyền thống, mỗi bản ghi có cấu trúc tĩnh với số cố định các thuộc tính. Mỗi bản ghi được đưa vào cơ sở dữ liệu thủ công bằng cách mô tả thuộc tính trong ngôn ngữ hỏi, như SQL. Trong MIRS hay hy đa phương tiện. các mục cơ sở dữ liệu

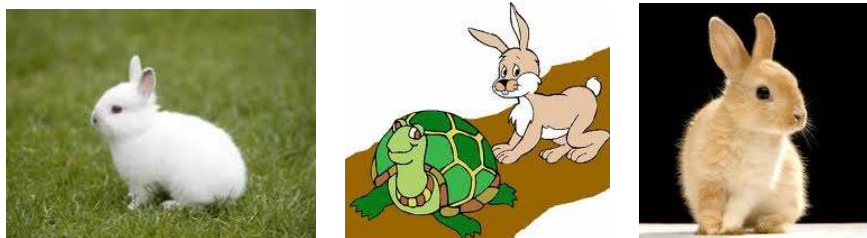
theo kiểu dữ liệu hay kiểu hỗn hợp. Chúng không có cấu trúc, thuộc tính cố định. Vậy nên giao diện người dùng sẽ cho phép người ta xác định các kiểu nhập vào, các đối tượng gộp đa phương tiện, và mô tả kiểu thuộc tính để trích và chỉ số hóa. Việc xử lý thông tin đa phương tiện và trích đặc trưng cần thời gian, như với đối tượng phim phức tạp. Người ta cần xác định phép toán trên nhóm các đối tượng, khi có thể tự động hóa việc trích các đặc trưng.

2.1.4.2. Trợ giúp hỏi

Các câu hỏi đa phương tiện là đa dạng và mờ. Chúng đa dạng bởi lẽ người dùng có thể mô tả các câu hỏi theo nhiều cách khác nhau, trên các phương tiện. Chúng mờ do người dùng có thể hiểu cái cần tìm, nhưng không mô tả chính xác được, hay nhu cầu thông tin cần tìm không chính xác như thông tin đã mô tả. Để khắc phục việc này, cần cung cấp các công cụ tìm kiếm, quét, và làm tinh câu hỏi.

2.1.4.2.1. Tìm kiếm

Tìm kiếm là nhiệm vụ cơ sở của tất cả các hệ thống cơ sở dữ liệu. Trong ngữ cảnh MIRS, có hai dạng tìm kiếm (i) tìm theo mô tả; (ii) tìm theo thí dụ. Thông thường, người dùng sử dụng từ khóa và tham số để mô tả đặc trưng chính hay thuộc tính của nhu cầu thông tin. Một khó khăn là cách khớp mô tả mức cao của người dùng theo ngôn ngữ tự nhiên thành các mẫu dữ liệu đa phương tiện đo được. Chẳng hạn người dùng tìm “thỏ trắng”, không có y hệt trong cơ sở dữ liệu, nhưng hệ thống giữ lược đồ màu và dạng đối tượng của hình ảnh. Câu hỏi được phân rã “thỏ” và “trắng” và tìm theo hình dạng, tách khỏi tìm theo lược đồ màu. Câu hỏi thành đối tượng thí dụ và hệ thống tìm mục tương tự với đối tượng thí dụ. Người ta vẽ hình dạng con thỏ rồi tô màu yêu cầu. Hệ thống tính lược đồ màu và tham số hình dạng để tìm.



Hình. Thỏ trắng dùng làm câu hỏi

Để trợ giúp tìm kiếm, giao diện cung cấp nhiều công cụ nhập như micrro, công cụ đồ họa, máy quay video, máy quét và các công cụ đa phương tiện. Người dùng sử dụng các mục đã có trong câu hỏi, thực hiện việc duyệt cơ sở dữ liệu.

2.1.4.2.2. Duyệt

Đôi khi người dùng không biết đích xác họ muốn gì nhưng có thể nhận dạng điều họ muốn tại thời điểm trong thấy. Thông tin cần thiết này có thể thấy nhờ duyệt. Duyệt cũng trợ giúp khả năng tìm theo thí dụ. Có ba phương pháp khởi động việc duyệt.

1. Bắt đầu với câu hỏi rộng, người dùng duyệt kết quả thu được.
2. Thông tin trong cơ sở dữ liệu cần được tổ chức theo một vài tiêu chuẩn, chẳng hạn chủ đề, ngày tháng, để người dùng duyệt theo tổ chức này.
3. Cơ sở dữ liệu thể hiện ngẫu nhiên một số mục cơ sở dữ liệu, người

dùng bắt đầu từ đó.

Nếu người dùng không thấy mục ưng ý nào, người ta yêu cầu tập mục ngẫu nhiên khác. Để tăng hiệu quả duyệt, người ta gắn biểu tượng với nhóm các mục cơ sở dữ liệu. Chẳng hạn với dữ liệu video, người ta sử dụng khung chính và biểu tượng hình 3D.

Các siêu liên kết giúp người ta duyệt dữ liệu. Thực tế người ta tin cậy vào duyệt và tìm theo các siêu liên kết để định vị thông tin hữu ích, như trên WWW.

2.1.4.2.3. Tinh chỉnh câu hỏi

Ban đầu hầu hết các câu hỏi đa phương tiện là không chính xác. Giao diện người dùng cần cung cấp các công cụ để làm tinh câu hỏi dựa theo kết quả đầu tiên. Làm tinh câu hỏi thường dựa vào phản hồi của câu hỏi đầu. Khi người dùng thấy mục sát với điều họ tìm, họ có thể tích hợp đặc trưng của mục đó vào câu hỏi mới. Sau vài lần lặp, người ta thấy mục lộ ra, nếu nó có trong cơ sở dữ liệu. Tri thức lĩnh vực và kiến thức người dùng có thể giúp chỉnh lí câu hỏi.

Phản hồi là đặc biệt hữu ích trong ứng dụng đa phương tiện, vì người dùng có thể cho biết hình ảnh hay âm thanh có đúng như họ muốn không. Thực tế, định vị thông tin đa phương tiện là kết hợp của tìm, duyệt, và làm tinh câu hỏi.

2.1.4.3. Thể hiện kết quả

Câu hỏi được MIRS trả lời quan giao diện. Có nhiều khía cạnh thiết kế đối với việc thể hiện kết quả.

1. Giao diện có khả năng thể hiện các kiểu phương tiện và quản trị mối quan hệ về thời gian, không gian của chúng. Quản trị quan hệ không gian, thời gian là khía cạnh chính của đảm bảo QoS.
2. Thông tin kết quả được chứa trong nhiều đoạn âm thanh, hình ảnh lớn, video dài. Cần có cách trích, hay kiến thiết, và thể hiện các thông tin cốt lõi cho người dùng, để duyệt và lựa chọn. Nói cách khác, cần có các kỹ thuật để kiến thiết cấu trúc thông tin để người ta có thể nhanh chóng biết được năng lực. Các biểu tượng, các ảnh nhấn mạnh là các công cụ quan trọng.
3. Cần trả lời nhanh. Thời gian trả lời được xác định bằng cả hệ thống con truyền thông và hiệu quả của tìm kiếm cơ sở dữ liệu. Một kỹ thuật đầy nhanh việc trả lời là kỹ thuật tăng cường giải mã và hiển thị trong khi đang truyền thông dữ liệu.
4. Thể hiện kết quả cần làm lộ ra phản hồi và điều chỉnh câu hỏi.

2.2. Mô hình hoá tư liệu đa phương tiện tương tác IMD

Theo mô hình dữ liệu, người ta quan tâm đến đặc tính của thông tin đa phương tiện: (i) tương tác; (ii) đa chiều; (iii) phụ thuộc về không gian, thời gian; (iv) độ phức tạp; (v) toàn vẹn thời gian.

Một IMD cần đến một loạt các đối tượng đa phương tiện cá thể, được thể hiện theo tập các đặc tả (là các kịch bản IMD).

Các đối tượng đa phương tiện tham gia trong IMD được chuyển hoá (i) về không gian, (ii) thời gian, theo yêu cầu của người dùng. Ngoài ra, tác giả cần xác định trật tự không gian, thời gian cho các đối tượng theo ngữ cảnh tư liệu và mối quan hệ giữa

chúng.

Cách mà người dùng tương tác với phần thể hiện, cũng như cách mà ứng dụng đối xử với các sự kiện ứng dụng và hệ thống cũng được xác định. Các ứng dụng gồm: (i) TV tương tác; (ii) phim số hoá; (iii) ứng dụng hiện thực ảo.

Theo IMD, người ta cần xem xét các (i) sự kiện; tổ hợp không gian, thời gian; (iii) kịch bản trong mô hình.

- Các *sự kiện* là cái cơ bản trong tương tác theo ngữ cảnh của IMD, được thể hiện qua hành động của người dùng, qua các đối tượng tham gia trong IMD, hay qua hệ thống. Sự kiện có thể là đơn giản hay phức tạp, gắn với điểm mốc không gian, thời gian;
- *Tổ hợp không gian, thời gian*: phần đặc biệt của IMD, cho biết trật tự không gian, thời gian của các đối tượng đa phương tiện trong lĩnh vực đang xét. Vấn đề là quan hệ giữa các đối tượng theo không gian, thời gian;
- *Kịch bản*: dùng cho nội dung về hành vi được tích hợp của IMD, tức là một loại sự kiện mà IMD thể hiện và là cách mà các hành động được kích hoạt như là kết quả. Vậy kịch bản gồm tập hợp đơn vị chức năng (các bộ (tuple) kịch bản), trong đó có (i) các sự kiện (bắt đầu và dừng), (ii) các hành động thể hiện (theo cách không gian, thời gian), (iii) được thực hiện theo ngữ cảnh của các bộ kịch bản, liên quan với sự kiện đồng bộ hoá (tức là các sự kiện có vai trò kích hoạt, khi bộ kịch bản bắt đầu hay dừng).

Để trợ giúp IMD phức tạp, mỗi hệ thống đưa ra (i) mô hình mức cao cho IMD; (ii) thể hiện đa phương tiện tương tác chấp nhận được. Mô hình hoá sẽ đảm bảo (i) tổ hợp không gian, thời gian trên phương tiện tham gia; (ii) xác định tương tác giữa người dùng và IMD; (iii) đặc tả về đồng bộ hoá phương tiện.

Việc mô hình hoá IMD sẽ nhấn mạnh vào phần tương tác của ứng dụng. Về nguyên tắc, mô hình hoá tương tác sẽ bao trùm các thủ tục, và người dùng lẫn thiết bị.

2.2.1. Mô hình hoá tương tác với các sự kiện

Khái niệm các sự kiện trong cơ sở dữ liệu được xác định như “sự xảy ra điều đang quan tâm”. Sự kiện sinh ra do hành động, xảy ra tại điểm đặc biệt, là đơn hay phức hợp.

- Các hệ thống thông tin đa phương tiện trải rộng trên nhiều ngữ cảnh và sự kiện, tạo nên miền hành động của cơ sở dữ liệu.
- Khía cạnh thời gian được thể hiện qua điểm thời gian;
- Các sự kiện IMD mang thông tin không gian.

Định nghĩa: Sự kiện: sự kiện lộ ra khi xảy ra hành động và gắn với thể hiện không gian, thời gian. Sự kiện được ghi nhận thông qua người hay quá trình.

Một phần của thể hiện thời gian cũng được gán thể hiện không gian trong trường hợp sự kiện liên quan đến đối tượng phương tiện hiển thị. thể hiện không gian này được hình thức hoá qua hình chữ nhật bao lấy sự kiện, tức vùng màn hình mà hình ảnh cần để thể hiện). Vậy, cần tích hợp hai khái niệm (i) thời gian và (ii) không gian trong định nghĩa của sự kiện. Do vậy mà thuật ngữ thể *hiện không gian, thời gian* được thể hiện

dưới dạng (sp_inst, temp_inst), trong đó sp_inst là thể hiện không gian, và temp_inst là thể hiện thời gian.

2.2.1.1. Phân loại

Cần có thư mục cho mỗi loại sự kiện, cần phân loại các sự kiện theo phạm trù. Việc phân loại các sự kiện được thực hiện trên cơ sở các thực thể sinh ra các sự kiện.

Các phạm trù gồm:

- Tương tác người dùng: là sự kiện được tạo thực sự bởi người dùng trong tương tác với ngữ cảnh IMD. Chúng là các sự kiện nhập vào chính, như tương tác người dùng với hệ thống qua các thiết bị nhập (chuột, bàn phím...). Các sự kiện điều khiển thời gian được biết như hành động: bắt đầu, dừng, đi quá, quay lại, dừng ngẫu nhiên, và liên quan đến việc khai thác của một hay nhiều nhóm đối tượng đa phương tiện;
- Sự kiện bên trong đối tượng: gồm các sự kiện liên quan đến chức năng bên trong của một đối tượng được thể hiện trong IDM. Chức năng này được thực hiện theo tiếp cận hướng đối tượng;
- Các sự kiện giữa các đối tượng: sự kiện này xảy ra khi hai hay nhiều đối tượng cần đến khi một hành động quan tâm xảy ra. Các sự kiện này lộ ra nếu có mối quan hệ thời gian giữa hai hay nhiều đối tượng. Sự kiện cũng có thể xảy ra khi sự đồng bộ của hai đối tượng liên tiếp bị quá ngưỡng hạn chế;
- Các sự kiện do người dùng xác định: Các sự kiện này do người thiết kế IMD xác định. Chúng liên quan đến nội dung của khai thác IMD. Người dùng có thể tham chiếu đến nội dung của đối tượng phương tiện.

2.2.1.2. Mô hình hoá và tổ hợp các sự kiện

Mô hình cho các sự kiện đơn giản và phức tạp trong IMD, dựa trên các khái niệm về sự kiện và phân loại.

Theo định nghĩa sự kiện, để thể hiện sự kiện, cần thuộc tính sau: (i) chủ đề; (ii) các thuộc tính về đối tượng, tức là (i) loại đối tượng OList, và (ii) cái thể hiện các đối tượng gây ra hay chịu tác động của sự kiện. Thuộc tính nhãn_không gian_thời gian được gắn với sự kiện.

Vậy cấu trúc của lớp sự kiện theo ngôn ngữ giả hướng đối tượng:

```
CLASS sự_kiện thừa kế từ đối_tượng
Thuộc_tính
    Chủ_đề OList;
    Hành_động ActionList;
    đối_tượng Olist;
    nhãn_không gian_thời gian thể hiện không gian, thời gian;
END.
```

Cần đảm bảo các công cụ cho phép xác định tổ hợp các sự kiện. Tổ hợp các sự kiện theo ngữ cảnh của IMD có hai khía cạnh:

1. Tổ hợp đại số: tổ hợp các sự kiện theo các phép đại số, theo nhu cầu của IMD;

2. Tổ hợp không gian, thời gian: phản ánh mối quan hệ không gian, thời gian giữa các sự kiện.

Ở đây chỉ đề cập khía cạnh thứ nhất trong hai khía cạnh trên. Một vài khái niệm cơ bản là (i) điểm tham chiếu không gian, thời gian θ , là điểm bắt đầu của kịch bản IMD, được dùng như tham chiếu đối với bất kì sự kiện không gian, thời gian nào trong IMD; (ii) khoảng thời gian: là khoảng giữa 2 sự kiện (e_1, e_2) hai đầu khoảng $t_int := (e_1, e_2)$. Các sự kiện e_i gắn với điểm thời gian xác định trước hay xảy ra dị bộ.

2.2.1.3. Tổ hợp đại số của các sự kiện

Người ta phân biệt các trường hợp (i) hợp; (ii) giao; (iii) bao hàm; (iv) phủ định.

1. Giao: $e := \text{OR}(e_1, \dots, e_n)$, xảy ra khi ít nhất một sự kiện e_i xảy ra. Thí dụ đèn sáng khi nút A (e_1) hoặc nút B (e_2) được nhấn;
2. Hợp:
 - $e := \text{ANY}(k, e_1, \dots, e_n)$, xảy ra khi ít nhất k sự kiện e_i xảy ra. Thứ tự các e_i có thể bất kì.
 - $e := \text{SEQ}(e_1, \dots, e_n)$ xảy ra khi tất cả e_i xảy ra theo thứ tự trong danh sách;
 - $e := \text{TIMES}(n, e_1)$ xảy ra khi có n lần liên tiếp e_1 xảy ra. Có thể ngầm định cho phép sự kiện e_2 xảy ra giữa các e_1 .

Trong nhiều trường hợp, người ta muốn sử dụng ràng buộc đối với sự kiện xảy ra trong một khoản thời gian.

3. bao hàm: $e := \text{IN}(e_1, t_int)$ xảy ra khi sự kiện e_1 xảy ra trong khoảng t_int , thí dụ sự kiện $e = \text{IN}(\text{TIMES}(3, \text{mouse, click}), (e_2, e_3))$ cho biết 3 lần nhấp chuột trong khoảng $e_2 \rightarrow e_3$;
4. Phủ định: $e := \text{NOT}(e_1, t_int)$ xảy ra khi e_1 không xảy ra trong t_int .

2.2.2. Tổ hợp không gian, thời gian và các nhân tố

Trong IMD có tập các đối tượng đơn phương tiện (được coi là các nhân tố), nhiều khi có thể hiện liên quan đến không gian và/hoặc thời gian. Tập các hành động liên quan đến nhân tố như bắt đầu, kết thúc, thể hiện, ẩn.

Thuật ngữ *đồng bộ* được dùng rộng rãi để mô tả thứ tự thời gian của nhân tố trong ứng dụng đa phương tiện. Tuy nhiên do có nhân tố không gian nên thuật ngữ này trở nên nghèo nàn. Do vậy người ta dùng *tổ hợp*¹ để thể hiện nhân tố cả không gian và thời gian.

2.2.2.1. Mối quan hệ không gian và thời gian và các phép toán

Cần thiết đề cập khía cạnh thời gian của tổ hợp các nhân tố, cho phép có lược đồ thu giữ được các mối quan hệ về thời gian:

- Các điểm bắt đầu và kết thúc của thể hiện đa phương tiện được dùng là sự kiện;

¹ composition

- X:== integer
- Y:== integer
- Temp_rel:== t_event t_khoảng phép_toán_TAC

2.2.3. Dữ liệu văn bản

MIRS đòi hỏi xử lý, chỉ số hóa, lưu trữ, truyền và thể hiện dữ liệu đa phương tiện. Dữ liệu gồm văn bản, âm thanh, hình ảnh, hình động và dạng tổ hợp của chúng. MIRS khác với DBMS ở chỗ tính chất và yêu cầu khác nhau đối với dữ liệu đa phương tiện. Vậy hiểu về các tính chất và yêu cầu đối với dữ liệu đa phương tiện là cần để thiết kế, cài đặt MIRS.

Đối với mỗi loại dữ liệu, người ta cần biết dạng thức thô và dạng thức nén thông dụng của chúng. Hầu hết dữ liệu đa phương tiện đều được nén. Với mỗi loại, cần biết (i) cách nhìn nhận và lưu dữ liệu; (ii) kỹ thuật nén; (iii) cách trích đặc trưng và chỉ số hóa; (iv) yêu cầu lưu trữ; (v) thể hiện dữ liệu.

Một số khía cạnh trong phần này:

1. Dữ liệu văn bản là thông dụng và còn được dùng để chú giải dữ liệu đa phương tiện khác. Dạng thức hỗn hợp của văn bản, như tư liệu Acrobat, siêu văn bản, được đề cập.
2. Hai dạng dữ liệu đồ họa và hình động là (i) vecto; (ii) bitmap. Đồ họa bitmap tương tự như hình ảnh số và hình động bitmap như video số.
3. Đối với dữ liệu âm thanh, cần biết thuộc tính của tín hiệu âm thanh tương tự, quá trình chuyển tín hiệu tương tự sang tín hiệu số, thể hiện số của ứng dụng âm thanh chung, kỹ thuật nén âm thanh. Bộ chuyển tương tự sang số ADC¹ sẽ được sử dụng.
4. Hình ảnh số và các kỹ thuật nén ảnh được chi tiết với thể hiện ảnh số, yêu cầu về lưu trữ, lý do nén dữ liệu, lấy mẫu, lượng hóa vecto, các chuẩn nén JPEG² và JPEG2000.
5. Các kỹ thuật và chuẩn nén video là cần thiết. Video số được thể hiện qua chuỗi các khung theo tốc độ cố định. Các chuẩn chung là MPEG³, với MPEG-1, MPEG-2, MPEG-4, và MPEG-7.

Định nghĩa: Tín hiệu là bất kỳ lượng nào có thay đổi theo thời gian, hay không gian. Trong lý thuyết thông tin, tín hiệu là thông báo được mã hóa, tức dãy các trạng thái trong kênh truyền thông để mã hóa thông báo. Theo nghĩa của xử lý tín hiệu, người ta không xem dòng dữ liệu nhị phân tùy ý là tín hiệu; mà tín hiệu là dạng tương tự hay dạng số cho phép thể hiện lượng vật lý tương tự.

2.2.3.1. Văn bản đầy đủ

Định nghĩa: Văn bản⁴: Văn bản gồm các ký tự số, chữ.

Thể hiện chung nhất của ký tự là theo mã ASCII¹. Người ta dùng 7 bit cho mỗi mã,

¹ analog-to-digital conversion
² Joint Photographic Expert Group
³ Motion Picture Expert Group
⁴ text

nhưng sử dụng chung là 8 bit, thêm bit chẵn lẻ. Người ta sử dụng bộ mã khác đối với thể hiện phi tiếng Anh. Bộ nhớ dành cho văn bản được tính theo số kí tự, số trang.

Trong hệ thống IR, thuộc tính ứng với văn bản độ dài thay đổi không phù hợp với DBMS, nên khó xác định hiệu suất xử lí.

2.2.3.2. Văn bản có cấu trúc

Hầu hết tư liệu văn bản có cấu trúc, gồm (i) nhan đề; (ii) đoạn; (iii) mục... Cấu trúc thể hiện ra khi in. có nhiều dạng thức và chuẩn mã hóa văn bản cấu trúc, như SGML², ODA³, LaTeX và PDF⁴.

Bình thường đầu tệp dùng để thông tin về dạng thức tư liệu. Khi biết dạng thức tệp, thông tin cấu trúc được trích rút dùng cho tìm kiếm. Liên quan của tư liệu có thể được xác định qua liên kết với tư liệu khác, nếu người ta sử dụng tập các siêu liên kết.

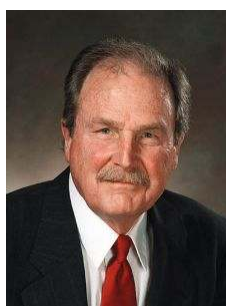
2.2.3.3. Nén văn bản

Dù yêu cầu lưu trữ văn bản là bình thường so với các dữ liệu đa phương tiện khác, người ta vẫn nén văn bản. Văn bản có thể được nén dựa trên nhiều kí tự giống nhau xuất hiện.

Định nghĩa: Đặc trưng chính của nén văn bản là nén không mất thông tin, tức văn bản có thể được khôi phục chính xác khi giải nén.

2.2.3.3.1. Mã Huffman

Mã không tổn thất thông dụng là mã Huffman. Nó gán ít bit cho các kí hiệu xuất hiện nhiều hơn, và gán bit dài hơn cho các kí hiệu xuất hiện ít. Cách này hiệu quả bởi xác suất xuất hiện các kí hiệu thay đổi. Nó được dùng với cách mã khác.



Hình. David Albert Huffman (1925 –1999) là người tiên phong trong lĩnh vực công nghệ thông tin

Bảng. Sách mã thí dụ

<i>Symbol</i>	<i>Probability</i>	<i>Code</i>
<i>e</i>	0.8	1
<i>t</i>	0.16	01
<i>x</i>	0.02	001
<i>z</i>	0.02	000

¹ American Standard Code for Information Interchange

² Standard General Markup Language

³ Office Document Architecture

⁴ Portable Document Format

Định nghĩa: Mã Huffman là một thuật toán mã hóa dùng để nén dữ liệu. Nó dựa trên bảng tần suất xuất hiện các ký tự cần mã hóa để xây dựng một bộ mã nhị phân cho các ký tự đó sao cho dung lượng, theo số bit, sau khi mã hóa là nhỏ nhất.

Thí dụ: Có tập 1000 ký tự, với bốn ký tự: (i) e, với xác suất xảy ra $p(e) = 0.80$; (ii) t, $p(t) = 0.16$; (iii) x, $p(x) = 0.02$; và (iv) z, $p(z) = 0.02$. Người ta dùng hai bit thể hiện các ký tự, vậy cần 2000 bit cho cả tập. Sử dụng mã Huffman, người ta dùng bit 1 cho e, 01 cho t, 001 cho x và 000 cho z. Vậy tổng cộng $1000 * (1 * 0.8 + 2 * 0.16 + 3 * 0.02 + 3 * 0.02) = 1240$ bit.

2.2.3.3.2. Mã độ dài chạy

Định nghĩa: Mã độ dài chạy là phương pháp nén dữ liệu cho phép giảm về vật lý bất kỳ chuỗi ký tự lặp lại.

Người ta sử dụng xuất hiện lặp của cùng ký tự là “chạy” và số lần lặp là “độ dài”. Thực tế người ta sử dụng ký tự phụ để biết việc dùng mã độ dài chạy.

S_c	X	C
-------	-----	-----

Hình. S_c cho biết mã được sử dụng, X là ký tự lặp, C là số đếm.

Thí dụ: Có "BBBBBBBAnnnnnnnn.". Nếu dùng @ để chỉ ký tự sau đó là độ dài chạy được mã, thì thu được @B7A@n8. Như vậy người ta rút từ 16 ký tự xuống 7 ký tự.

2.2.3.3.3. Mã LZW

Định nghĩa: Mã LZW¹ thay thế nhóm lặp của các ký tự bằng ký tự từ điển.

Trong tư liệu thường xuất hiện đoạn ký tự lặp lại, nên người ta xây dựng từ điển các đoạn có tần suất xuất hiện cao để gán thẻ. Thẻ thay thế đoạn ký tự trong tư liệu.

Thí dụ người ta sử dụng ký tự gõ tắt khi soạn văn bản hay dùng thẻ gán cho các đoạn điệp khúc trong ca từ.

2.2.4. Đồ họa vecto và hình động

Có hai loại đồ họa (i) vecto; (ii) bitmap. Đồ họa bitmap chia nhỏ thành các pixel, mỗi pixel ứng với chấm trên màn hình. Cường độ màu của pixel được lưu trong tệp đồ họa pixel. Theo dạng này, đồ họa có tính chất như hình ảnh.

Định nghĩa: Đồ họa²: Đồ họa là một lĩnh vực truyền thông trong đó thông điệp được tiếp nhận qua con đường thị giác. Thiết kế đồ họa là tạo ra các giải pháp bằng hình ảnh cho các vấn đề truyền thông.

Ở Việt Nam, việc dùng các thuật ngữ đồ họa chỉ là tương đối, bởi việc đặt tên các thuật ngữ đồ họa là dựa trên ý nghĩa sử dụng của nó, chẳng hạn đồ họa thương nghiệp, đồ họa vẽ tem, đồ họa quảng cáo... Trong đồ họa vecto, phần tử đồ họa được thể hiện theo các mô hình xác định trước hay theo công thức toán học. Việc lưu trữ dữ liệu vecto

¹ Lempel-Ziv-Welch
² graphic

đơn giản, do người ta lưu trữ chỉ dẫn cơ bản cho phép sinh ra đồ họa. Đối với đồ họa vecto, dễ trích nội dung. Thuộc tính của mỗi phần tử, như hình dáng, kích thước, có thể lấy ra từ tệp đồ họa và được sử dụng để chỉ số hóa và tìm kiếm.

Định nghĩa: Hình động¹: Hình động là thể hiện nhanh chuỗi các hình ảnh của tác phẩm 2D hay 3D hay các vị trí mô hình, để tạo nên ảo ảnh về chuyển động.

Hình động được sinh ra do làm tình, biểu hiện tuần tự các khung đồ họa. Nếu đồ họa dùng bitmap, hình động như đoạn video. Đối với hình động theo vecto, việc chỉ số hóa và tìm kiếm được thực hiện như với đồ họa vecto, trừ việc xử lý yếu tố thời gian.

Định nghĩa: Video là công nghệ điện tử để thu, ghi lại, xử lý, lưu trữ, truyền và tái tạo chuỗi các hình ảnh tĩnh nhằm thể hiện cảnh chuyển động.

2.2.4.1. Dữ liệu hình ảnh

2.2.4.1.1. Ảnh bitmap và ảnh vecto

Ảnh bitmap thể hiện ảnh trong ma trận các điểm ảnh. Sử dụng nhiều điểm ảnh sẽ làm hình ảnh mịn hơn.

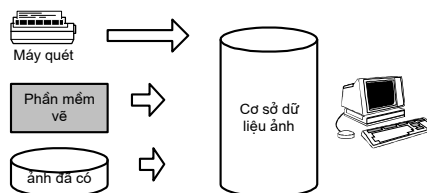
Ảnh vecto không dùng ma trận điểm ảnh, mà thể hiện nội dung ảnh qua phương trình thể hiện của hình ảnh, sau đó thiết bị hiển thị sẽ thể hiện lại hình ảnh, dựa trên phương trình của ảnh.

2.2.4.1.2. Thể hiện ảnh

Người ta tạo và số hoá các đồ họa để chuẩn bị tư liệu ảnh tĩnh cho việc tích hợp dữ liệu, tạo sản phẩm đa phương tiện.

- Việc số hoá sẽ chuyển các đồ họa sang dạng điểm ảnh hai chiều, đồ họa Bitmap, gọi là ảnh RASTER;
- Các đồ họa được tạo ra sử dụng các đường cong toán học thay vì dùng điểm ảnh được gọi là đồ họa vectơ.

Các cách số hoá ảnh đồ họa gồm (i) quét từ máy quét; (ii) nhập vào từ đĩa CD; (iii) nhập vào từ nguồn khác.

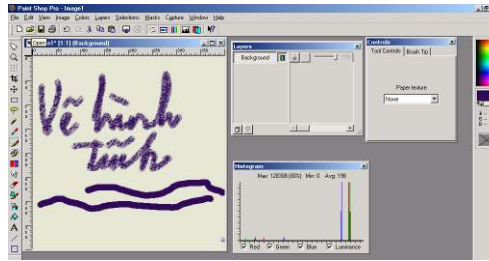


Hình. Nguồn dữ liệu về ảnh tĩnh

Một số vấn đề liên quan đến sản phẩm đồ họa:

- Vai trò của hình ảnh so với từ ngữ;
- Khái niệm về màu sắc, về hình vẽ;

¹ animation



Hình. Tạo hình tĩnh trong PaintShop PRO

- Ảnh bitmap, hay bức tranh;
- Các hình 3D;
- DPI và độ sâu bit;
- Khả năng xem trước ảnh trước khi sử dụng kỹ xảo ảnh;
- Liên kết đồ họa máy tính và đồ họa video;
- Đồ họa hoạt hình trên Web;
- Việc quản lý tài sản về hình ảnh tĩnh;
- Sự tương thích về hình và chất lượng ảnh.

2.2.4.2. Dữ liệu hình động

Việc mô phỏng chuyển động được tạo bởi hiện một loạt các hình, tức khung hình, được gọi là hình động. Phim hoạt hình trên TV là thí dụ về hình động. Hình động trên máy tính có vai trò chính trong thể hiện đa phương tiện. Có nhiều ứng dụng phần mềm cho phép người ta tạo hình động, rồi thể hiện trên màn hình máy tính.

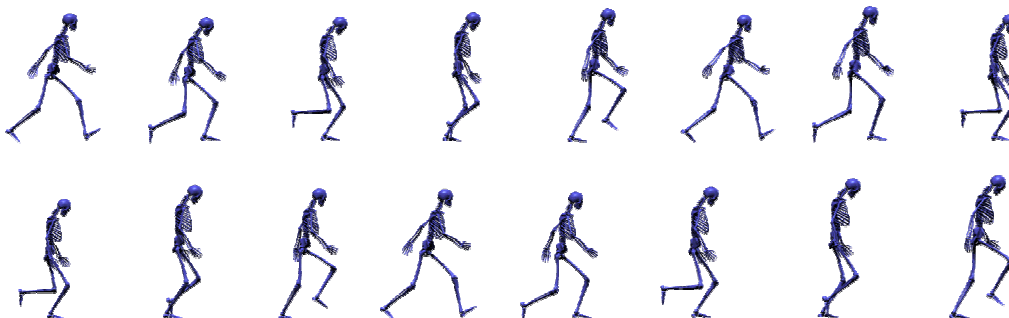
Cần phân biệt hình động với video. Trong khi video hiện các chuyển động liên tục, sử dụng các khung rời rạc, hình động bắt đầu bằng các ảnh độc lập, rồi đặt chúng với nhau để tạo nên ảo ảnh về chuyển động liên tục.

Rộng, cao	FPS	CODEC	Độ sâu của màu	Audio CODEC	Tỉ lệ lấy mẫu âm thanh	Kích thước mẫu âm thanh	Mono Stereo	Kích thước tệp tổng cộng
W = 320 H = 240	15	Sorenson	32. 10 ⁶ Bit	IMA 4:1	22.05 kHz	16 bits	mono	2.8 MB
W = 240 H = 180	15	Sorenson	32. 10 ⁶ Bit	IMA 4:1	22.05 kHz	16 bits	mono	1.7 MB
W = 240 H = 180	.53	Sorenson	32. 10 ⁶ Bit	Qdesign Music 2	8 kHz	16 bits	mono	108 KB
W = 160 H = 120	15	Sorenson	32. 10 ⁶ Bit	Qdesign Music 2	22.05 kHz	16 bits	mono	536 KB
W = 320 H = 240	15	Cinepak	32. 10 ⁶ Bit	DVI IMA	22.05 kHz	16bits	mono	2.7 MB
W = 160 H = 120	15	Cinepak	32. 10 ⁶ Bit	DVI IMA	22.05 kHz	16 bits	mono	1.7 MB
W = 320 H = 240	15	MPEG-4	32. 10 ⁶ Bit	Microsoft IMA ADPCM	22.05 kHz	16 bits	mono	4.6 MB
W = 320 H = 240	30	MPEG-4	32. 10 ⁶ Bit	Windows Media	22.05 kHz	16.bits	stereo	1.8 MB

Hình. Đoạn phim 30 giây ghi lại bằng máy quay VHS, được ghi lại theo các dạng mã khác nhau



Hình. Chuyển động con thỏ nhảy trong 1 giây, sử dụng 66 KByte



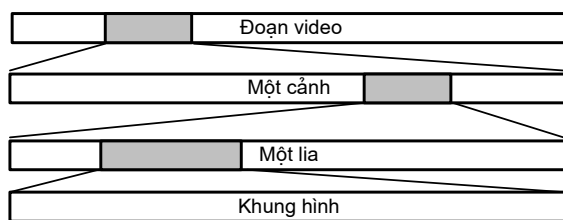
Hình. Tạo người chuyển 24 – 30 hình trong 1 giây

2.2.4.2.1. Các loại hình động

Quá trình đa phương tiện dùng khái niệm hình động để chỉ (i) các hình tạo nên bằng phần mềm tạo hình động; (ii) đoạn video. Tín hiệu từ các khung hình động đã ở dạng số hóa; tín hiệu từ video có thể là tương tự hay số hóa.

Phim hoạt hình có nhiều loại (i) hoạt hình; (ii) cắt giấy; (iii) búp bê, rồi... khi đưa lên sản phẩm dưới dạng phim nhựa, người ta được các tín hiệu tương tự. Được chuyển sang dạng số, các nội dung trên video mới tích hợp vào sản phẩm đa phương tiện.

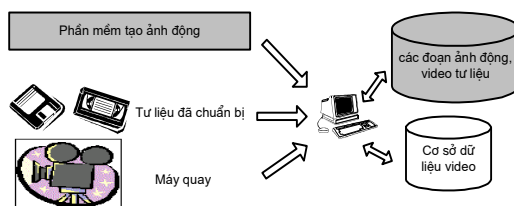
2.2.4.2.2. Xử lý dữ liệu video



Hình. Các mức trừu tượng video

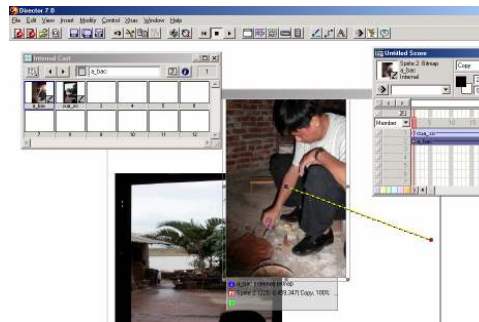
2.2.4.2.3. Quản lý sản phẩm video

Dữ liệu hình động được tạo ra do máy quay video hay phần mềm tạo ảnh động. Người ta có thể sử dụng kỹ thuật làm phim hoạt hình để xây dựng đoạn phim. Rồi ghi lại dưới dạng ảnh động.



Hình. Các nguồn thu thập dữ liệu hình động

Trước khi tích hợp dữ liệu trên máy tính, người ta cần chuyển đổi các dạng dữ liệu sang dạng số hoá. Với dữ liệu hình động như dữ liệu video, cần xem xét chuẩn video, kích thước khuôn hình, số hình được thể hiện trong một giây...



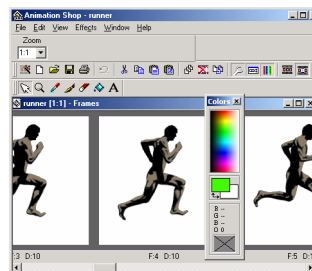
Hình. Phần mềm Macromedia Director cho phép tạo hình động

Phần mềm Macromedia Director của công ty hãng Macromedia cho phép

- Tạo ra chuyển động định hướng, tạo nên thành phần đa phương tiện, gồm có hoạt hình, âm, video số hay bất kì phương tiện khác. Ví dụ người ta có thể tạo ra logo, trò chơi điện tử;
- Tạo nên sản phẩm hoàn chỉnh, dưới dạng tệp *.exe, có thể khai thác trực tiếp không qua phần mềm của công tin tin học.
- Việc tạo và số hoá ảnh động đòi hỏi người dùng thực hiện nhiều bước, trang bị nhiều kĩ năng. Công việc này bao gồm cả mô hình hoá 3 chiều và xây dựng hoạt hình. Hoạt hình 2 chiều biểu diễn các đối tượng theo trục X và Y; còn hình 3 chiều dùng thêm trục thứ ba Z.

Các bước tạo và số hoá hình 3 chiều:

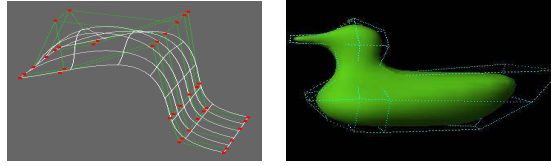
- Mô hình hoá đối tượng: tạo và dựng đối tượng 3 chiều;
- Hoạt hình: di chuyển nhiều hình theo dãy các khung hình;
- Hoàn thiện, dựng, tạo lại¹: thiết lập nền và ánh sáng phù hợp.



Hình. Tạo hình động trong phần mềm vẽ PaintShop PRO

Hoạt hình thường là sản phẩm sáng tạo của một nhóm, không phải là sản phẩm của cá nhân.

¹ rendering



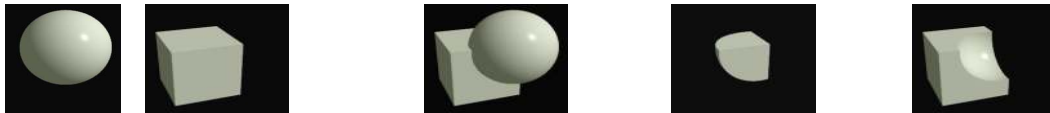
Hình. Đối tượng được quản lí số hoá

2.2.4.2.4. Nguyên tắc cơ bản của video

Video là một dạng hình động trong số các dữ liệu đa phương tiện. Có thể xem hình động và video như một loại dữ liệu. Nguồn dữ liệu video nhập vào máy tính để tích hợp sản phẩm đa phương tiện từ máy quay video. Liên quan đến việc quay và số hoá dữ liệu video, có khái niệm:

- Video là hình ảnh thực được ghi lại và thể hiện nhờ máy ghi video, phần mềm video và các công cụ phần cứng;
- Tập video gồm một loạt các hình tĩnh được biểu diễn nhanh;
- Dạng nén thông dụng cho video là MPEG.

Về tạo hình, cần quan tâm (i) tạo hình tĩnh; (ii) tạo hình động, với cấu trúc động; (ii) nên có thể hiện ý tưởng sáng tác trong tạo hình, tuy mất đi tính thực của đối tượng. Về bố cục hình, không thể không nắm các qui luật hình hoạ; vậy nên luôn lưu ý đến kiến thức về (i) phối cảnh; (ii) đường chân trời; (iii) cấu trúc hình. Chẳng hạn khi vẽ người phải chú ý về tỷ lệ của bộ xương, và phải nhớ một số đặc điểm riêng, như là người châu Á chân ngắn hơn người châu Âu.



Hình. Vị trí tương đối giữa các đối tượng 3D



Hình. Quan niệm về hình vững chãi

2.2.4.2.5. Các chuẩn truyền hình video

Với dữ liệu video, một số công việc được quan tâm là:

- Nguyên tắc quản lí sản phẩm video;
- Hiểu biết các nguyên tắc cơ bản của video;
- Các chuẩn truyền hình video (i) NTSC; (ii) PAL; (iii) SECAM;
- Các thành phần video;

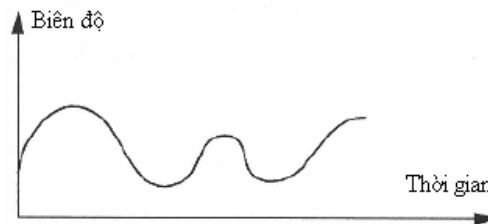
- Chuẩn bị phòng quay và ghi phòng vắn;
- Phân loại thời lượng ghi: (i) ghi dài; (ii) ghi trung bình; (iii) ghi ngắn;
- Phân loại cảnh ghi (i) cận cảnh; (ii) toàn cảnh;
- Phương tiện cho phép sửa dữ liệu;
- Máy quay video; băng ghi hình; đĩa hình;
- Lựa chọn phương pháp số hoá dữ liệu video;
- Băng sao lưu dùng cho quá trình đa phương tiện;
- Các kĩ thuật nén;
- Kiểm tra chất lượng.

2.2.5. Âm thanh

2.2.5.1. Tính chất cơ bản của tín hiệu âm thanh

Định nghĩa: Âm thanh¹: Âm thanh là các dao động cơ học (biến đổi vị trí qua lại) của các phân tử, nguyên tử hay các hạt làm nên vật chất và lan truyền trong vật chất như các sóng. Âm thanh, giống như nhiều sóng, được đặc trưng bởi tần số, bước sóng, chu kỳ, biên độ và vận tốc lan truyền.

Âm thanh là tác động của áp suất không khí lên trống của tai con người. Con người nghe được với ngưỡng âm thanh tần số 20 Hz đến 20.000 Hz. Tham số khác đo âm thanh là biên độ, cho phép âm thanh mềm hay thô.



Hình. Thí dụ về sóng âm

Tai người có ngưỡng nghe rộng, từ mức nghe thấy đến mức gây đau tai. Ngưỡng nghe được đối với sóng âm dạng hình sin 1 kHz thiết lập 0.000283 dyne² trên cm². Biên độ sóng âm có thể tăng từ ngưỡng nghe thấy với hệ số giữa 100.000 đến 1.000.000 trước khi tai đau. Vậy biên độ âm thanh thường được thể hiện qua dB. Cho hai dạng sóng có biên độ đỉnh là X và Y, đo dB về sai khác giữa hai biên độ được xác định theo $dB = 20 \cdot \log_{10}(X / Y)$. Nếu ngưỡng nghe thấy là 0.000283 dyne trên cm² đối với tín hiệu 1 kHz, như 0 dB, thì ngưỡng đau tai của đa số người là 100 đến 120 dB. Sóng âm thanh là liên tục theo biên độ và thời gian.

2.2.5.2. Thể hiện số của âm thanh

Dạng sóng âm thanh liên tục được micro chuyển thành tín hiệu điện liên tục. Trong hình là tín hiệu điện có cùng dạng, dù biên độ đã nhân với số dương. Tín hiệu điện này

¹ audio

² 1 dyn = 1 g·cm/s² = 10⁻⁵ kg·m/s² = 10⁻⁵ N

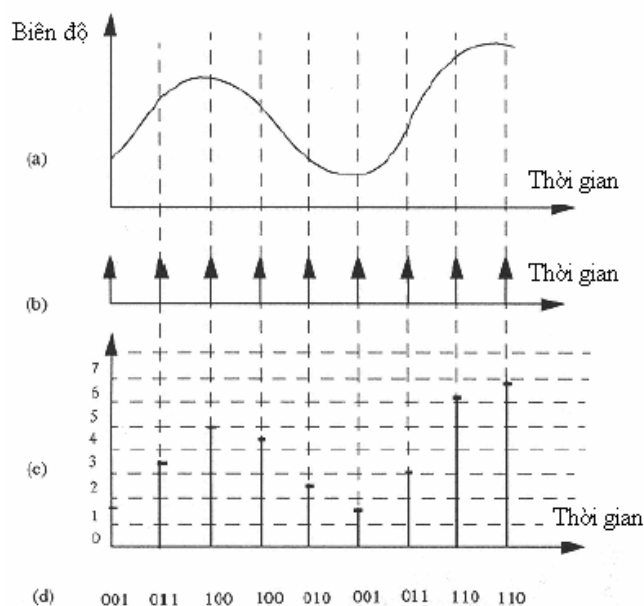
được đo bằng vôn, điện thế. Loại tín hiệu với thời gian và biên độ liên tục được gọi là tín hiệu tương tự. Lúc được xử lý trên máy tính, tín hiệu điện liên tục này được chuyển sang tín hiệu số. Ba giai đoạn ADC: (i) lấy mẫu; (ii) lượng hóa; (iii) mã hóa.

Định nghĩa: Tín hiệu tương tự là tín hiệu với thời gian và biên độ liên tục.

2.2.5.2.1. Lấy mẫu

Định nghĩa: lấy mẫu¹: Quá trình chuyển thời gian liên tục thành các giá trị rời rạc.

Trong hình thí dụ, trục thời gian được chia thành các khoảng cố định. Việc đọc giá trị xảy ra của tín hiệu tương tự lấy tại nơi bắt đầu của mỗi khoảng. Khoảng này được xác định bằng xung đồng hồ. Tần số của đồng hồ được gọi là tần số lấy mẫu, hay tần số mẫu. Giá trị mẫu giữ không đổi cho khoảng thời gian tiếp theo. Chu trình lặp đó gọi là lấy mẫu. Mỗi một mẫu vẫn là tương tự về biên độ, tức vẫn đảm bảo có giá trị bất kì trong dải liên tục. Tuy nhiên, nó rời rạc theo thời gian, tức chỉ có một giá trị trong mỗi khoảng thời gian.



Hình. Quá trình chuyển tương tự sang số: (a) tín hiệu tương tự gốc; (b) lấy mẫu xung; (c) giá trị lấy mẫu và khoảng lượng tử hóa; và (d) chuỗi số hóa.

2.2.5.2.2. Lượng tử hóa

Định nghĩa: Lượng tử hóa² là quá trình chuyển các giá trị mẫu liên tục thành các giá trị rời rạc.

Trong quá trình này, dải tín hiệu được chia thành các khoảng cố định. Mỗi khoảng có cùng kích thước và được gán bằng một con số. Trong hình, các số từ 0 đến 7. Mỗi mẫu rơi vào một trong các khoảng, được gán bằng con số của khoảng đó. Thí dụ hai mẫu cuối cùng mang giá trị khác nhau trước khi lượng tử hóa, sẽ mang cùng giá trị 6. Kích thước của khoảng lượng tử hóa được gọi là bước lượng tử hóa.

¹ sampling
² quantization

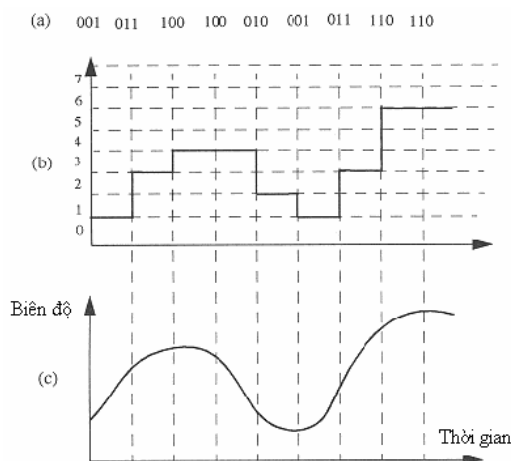
2.2.5.2.3. Mã hóa

Định nghĩa: Mã hóa¹ là quá trình thể hiện các giá trị lượng tử theo cách số hóa

Theo wiki 2010, trong mật mã học, một ngành toán học ứng dụng cho công nghệ thông tin, mã hóa là phương pháp để biến thông tin (phim ảnh, văn bản, hình ảnh...) từ định dạng bình thường sang dạng thông tin không thể hiểu được nếu không có phương tiện giải mã. Giải mã là phương pháp để đưa từ dạng thông tin đã được mã hóa về dạng thông tin ban đầu, quá trình ngược của mã hóa.

Trong thí dụ trên, có 8 mức lượng tử. Các mức này có thể được mã bằng 3 bit, nên mỗi mẫu được thể hiện theo 3 bit. Các số là 001, 011, 100, 100, 010, 001, 011, 110, và 110.

Từ quá trình trên, người ta thấy nếu tần số lấy mẫu và số các mức lượng tử đủ lớn, tín hiệu số sẽ thể hiện sát với tín hiệu tương tự. Khi cần tạo lại tín hiệu tương tự ban đầu, bộ chuyển số sang tương tự DAC² được dùng. Trong hình mô tả quá trình này, các giá trị lượng tử được xác định theo thể hiện số và các bước lượng tử. Mỗi giá trị được lấy đối với khoảng thời gian bằng khoảng lấy mẫu, tạo nên một loạt các tín hiệu bước. Các tín hiệu bước này chuyển sang bộ lọc duyệt thấp³ để tạo lại xấp xỉ của tín hiệu gốc. Người ta nói xấp xỉ do tín hiệu được tạo lại không y hệt tín hiệu gốc do sai số của việc lượng tử. Nguyên tắc ADC và DAC mô tả ở đây cũng áp dụng cho video và tín hiệu khác. Trong quá trình ADC, khía cạnh quan trọng nhất là cách chọn tần số lấy mẫu và số mức lượng tử đối với các tín hiệu tương tự khác nhau, ứng dụng khác nhau.



Hình. Quá trình chuyển số sang tương tự: (a) chuỗi số; (b) các tín hiệu bước; (iii) tín hiệu lấy ra sau khi qua bộ lọc qua thấp.

2.2.5.2.4. Xác định tỉ lệ lấy mẫu

Tỉ lệ lấy mẫu tùy theo tần số cực đại của tín hiệu tương tự được chuyển. Theo định lý Nyquist, nếu tín hiệu tương tự chứa các thành phần tần số đến f Hz, thì tỉ lệ lấy mẫu cần là $2f$ Hz. Nếu tỉ lệ lấy mẫu chính xác là $2f$ Hz, người ta gọi nó là lấy mẫu căng. Thực tế, tỉ lệ lấy mẫu lớn hơn $2f$ Hz một chút. Chẳng hạn tỉ lệ lấy mẫu của CD âm thanh

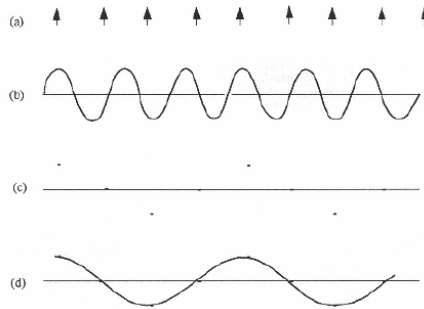
¹ coding

² digital-to-analog converter

³ low-pass filter

là 44.1 kHz, và tỉ lệ lấy mẫu của băng âm thanh số DAT¹ là 48 kHz, phủ dải tần nghe được 20 kHz. Các thành phần tần số của tiếng người là trong vòng 3.1 kHz. Vậy hệ thống điện thoại tương tự hạn chế tín hiệu dưới 3.1 kHz. Để chuyển tín hiệu tiếng người sang dạng số, thường dùng tỉ lệ lấy mẫu 8 kHz.

Nếu băng thông, tức dải tần số, của tín hiệu tương tự lớn hơn nửa tỉ lệ mẫu, băng thông cần giảm đi nhờ bộ lọc qua thấp, để nó thấp hơn hay bằng nửa tỉ lệ mẫu. Ngược lại sẽ có hiệu ứng trù phổ. Trong hình cho thấy đồng hồ lấy mẫu 8 kHz (a). Một thành phần tần số của tín hiệu được lấy mẫu là 6 kHz (b). Các giá trị mẫu thu được do thành phần 6 kHz (c). Nếu các mẫu này được lưu và chuyển ngược sang tương tự nhờ DAC, tín hiệu khác đi (d). Do vậy 2 kHz là trù phổ của tín hiệu 6 kHz, do 2 kHz trong dải nghe thấy, hiện ra như ồn trên đầu của âm thanh gốc. Trù phổ là vấn đề nghiêm trọng cho mọi hệ thống sử dụng cơ chế lấy mẫu khi tín hiệu lấy mẫu có các thành phần tần số lớn hơn một nửa tỉ lệ lấy mẫu.



Hình. Tín hiệu đầu vào với tần số lớn hơn một nửa tần số mẫu đã tạo nên tín hiệu trù phổ cho tần số khác, thấp hơn: (a) đồng hồ mẫu 8 kHz; (b) tín hiệu tương tự 6 kHz; (c) dãy giá trị mẫu; (d) tín hiệu tái tạo.

2.2.5.2.5. Xác định số mức lượng tử

Số các mức lượng tử quyết định sự trung thực về biên độ của tín hiệu số so với tín hiệu tương tự gốc. Sai khác cực đại giữa các giá trị mẫu lượng tử và các giá trị tín hiệu tương tự tương ứng là bước lượng tử. sai khác này được gọi là sai số lượng tử hóa hay ồn lượng tử hóa.

Số mức lượng tử càng lớn thì bước lượng tử càng nhỏ, và ồn lượng tử càng nhỏ. Số mức lượng tử quyết định cần bao nhiêu bit để thể hiện mỗi mẫu. Mối quan hệ này được quyết định bởi $b - \log_2 Q$, trong đó b là số các bit cần để thể hiện mỗi mẫu, Q là số mức lượng tử. Thực tế người ta muốn làm việc quá mức Q khi có số các bit dùng cho mỗi mẫu. Người ta có $Q - 2^b$.

Chất lượng tín hiệu số ứng với tín hiệu tương tự gốc được đo bằng tỉ lệ tín hiệu so với ồn SNR^1 , tính bằng dB. Nó được xác định $SNR = 20 \cdot \log_{10}(S / N)$, S là biên độ tín hiệu cực đại, N là ồn lượng tử. Giả sử bước lượng tử là q , thì $N = q$ và $S = 2bq$; vậy $SNR = 20b \cdot \log_{10} 2 - 6b$. Nó cho thấy sử dụng quá 1 bit để thể hiện mẫu làm tăng SNR 6

¹ digital audio tape

dB.



Hình. Tai người

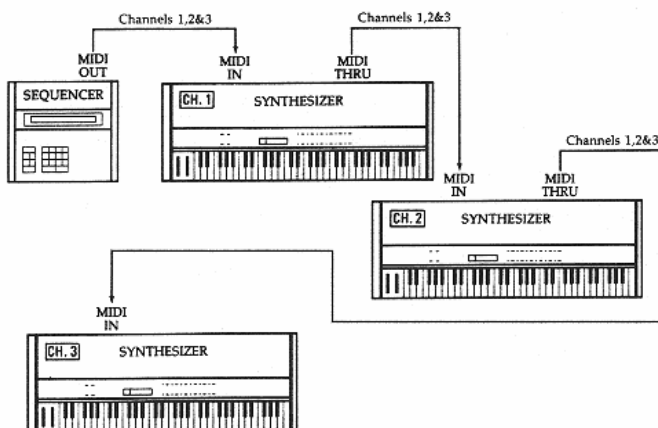
Nếu để ý đến ngưỡng nghe thấy và ngưỡng đau tai, ồn lượng tử sẽ nhỏ hơn ngưỡng nghe thấy. Nếu nó lớn hơn, người ta sẽ thấy tiếng ồn. Ngưỡng đau tai là 100 đến 120 dB, là dải cực đại của tín hiệu. Do vậy SNR của tín hiệu âm thanh số hóa cần nhỏ hơn 100 dB để không nghe thấy ồn lượng tử. Âm thanh CD dùng 16 bit cho mẫu, vậy người ta có $SNR = 96$ dB. Nó thấp hơn ngưỡng 100, 120 dB, nhưng vì $16 = 2^4$, tiện cho xử lý trong hệ thống số, nên 16 bit được dùng thay cho 17 bit cho mẫu.

Tóm lại, âm thanh số được lấy mẫu liên tục với tỉ lệ cố định. Mỗi mẫu được thể hiện bằng số cố định các bit. Bảng cho thấy một số trường hợp lấy mẫu.

Bảng. Các đặc tính của âm thanh số thông thường

ứng dụng	số các kênh	tỉ lệ lấy mẫu	Bit/ mẫu
CD-audio	2	44,100	16
DAT	2	48,000	16
Digital telephone	1	8,000	8
Digital radio, long-play DAT	2	32,000	16

2.2.5.3. Giao diện số cho nhạc cụ



Hình. MIDI

¹ signal-to-noise ratio

Thay vì thể hiện âm nhạc bằng xâu các mẫu, giao diện số cho nhạc cụ MIDI¹ xác định tập các thông báo hay mệnh lệnh. Thiết bị âm nhạc với năng lực MIDI, được gọi là bộ nhạc điện tử, sẽ tạo ra âm nhạc dựa trên thông báo MIDI. Các tệp âm nhạc MIDI nhỏ hơn tệp âm nhạc theo mẫu. Hơn nữa, các tệp MIDI có cấu trúc với tập các lệnh xác định trước, gồm một số trường xác định. Do vậy người ta có thể dùng thông tin có cấu trúc để chỉ số hóa và tìm kiếm âm nhạc MIDI.

Hạn chế của MIDI là (i) khó thể hiện và tạo ra đoạn nhạc phức tạp của nhiều nhạc cụ; (ii) bộ tổng hợp khác nhau có thể cho ra âm thanh khác nhau với cùng dữ liệu MIDI.

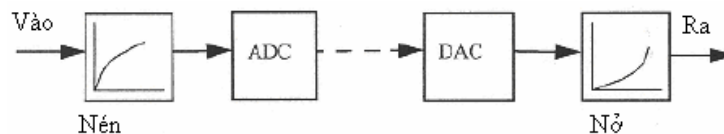
2.2.5.4. Nén âm thanh

2.2.5.4.1. Lượng tử phi tuyến

Khi lượng tử đều, cùng kích thước bước lượng tử được dùng cho quá trình ADC, mà không tính đến biên độ tín hiệu. Quá trình ADC như nhau ấy thường dùng như biến điệu mã xung PCM². Nó đơn giản, nhưng không hiệu quả về sử dụng số bit dữ liệu, do lượng tử tuyến tính sẽ gây SNR cao hơn tại nơi có biên độ tín hiệu cao hơn vùng có biên độ thấp. SNR tăng thêm tại biên độ tín hiệu cao hơn không tăng chất lượng âm thanh nhận biết vì người ta nhạy hơn với các thành phần biên độ thấp hơn.

Theo nhân tố này, kích thước bước lượng tử tăng theo hàm logarit của biên độ tín hiệu được dùng rộng rãi trong lượng tử hóa tín hiệu tiếng nói. Theo lược đồ lượng tử này, các bước lượng tử là càng nhỏ khi biên độ tín hiệu càng thấp và bước lượng tử càng lớn khi biên độ tín hiệu càng cao.

Thực tế, lượng tử đều được dùng để chuyển hóa các tín hiệu phi tuyến, thay vì sử dụng lượng tử phi tuyến cho tín hiệu tuyến tính. Các kết quả của hai tiếp cận này là như nhau. Quá trình chuyển tín hiệu tuyến tính sang tín hiệu phi tuyến được gọi là nén nờ³. Số hóa đều của tín hiệu nén nờ được gọi là PCM nén nờ. Các hàm chuyển thông dụng là hàm luật A và luật μ . Hàm luật A là $y = Ax / (1 + \ln A)$, $0 \leq x \leq 1/A$; $y = (1 + \ln Ax)/(1 + \ln A)$, $1/A \leq x \leq 1$, trong đó x là biên độ tín hiệu vào gốc, y là biên độ tín hiệu được chuyển, và A là hằng. Đối với mạng điện thoại, $A = 87.6$. Luật μ có dạng $y = \ln(1 + \mu x) / \ln(1 + \mu)$, trong đó x, y có nghĩa như trong luật A, và μ là hằng. Với mạng điện thoại, $\mu = 255$. Nén nờ thực sự là kỹ thuật nén tương tự, do tín hiệu tương tự được nén trước ADC và nở ra sau DAC.



Hình. Hệ thống với kỹ thuật nén nở

Sử dụng kỹ thuật nén nở, tín hiệu 8 bit có thể cho chất lượng tín hiệu tương đương tín hiệu mã theo PCM 12 bit. PCM nén nở đôi khi được gọi là PCM log do hàm logarit được dùng trong hàm chuyển.

¹ Musical Instrument Digital Interface

² pulse-coded modulation

³ companding

2.2.5.4.2. Mã hóa đoán trước

Người ta có thể mã hóa giá trị mẫu để truyền, nhưng mã đoán trước lại mã hóa sự khác nhau giữa giá trị đoán trước về mẫu và giá trị mẫu thực sự. Nếu sự khác nhau này hay lỗi đoán trước được lượng tử và mã hóa, lược đồ mã đoán trước được gọi là biến điệu mã xung khác nhau DPCM¹. Bình thường người ta dùng các từ mã hóa với độ dài cố định để mã lỗi dự đoán lượng tử. Do giá trị dự đoán đối với mỗi mẫu được tính chỉ từ thông tin quá khứ về tín hiệu được mã, giá trị dự đoán cũng được mã và giải mã. Để khôi phục mẫu lúc giải mã, sai số dự báo được truyền tương ứng được bổ sung vào giá trị dự đoán của mẫu này. Hiệu quả của DPCM đạt được qua việc các giá trị mẫu kề bên có liên quan, và sai số dự báo thường nhỏ hơn giá trị mẫu gốc. Do sai số dự đoán có dải biên độ nhỏ, chỉ một ít bit dành để mã hóa nó.

Thí dụ nếu người ta nén dòng âm thanh với 16 bit một mẫu, và dùng 4 bit để thể hiện sai số dự báo, chất lượng âm thanh được quản lý tại cùng mức 16 bit một mẫu nếu sai số dự đoán luôn nhỏ hơn 16 bước lượng tử gốc. Nếu sai số lớn hơn 16 bước lượng tử, chất lượng âm thanh gốc sẽ không giữ được. Tín hiệu tạo lại khi giải mã sẽ tiệm cận tín hiệu gốc khi biên độ tín hiệu thay đổi nhỏ hơn. Để cải thiện hiệu năng DPCM, người ta dùng bộ DPCM thích hợp, hay ADPCM. Để nắm được cả hai tín hiệu thay đổi nhanh và chậm, kích thước bước lượng tử giữa các mẫu kề nhau thay đổi theo chính tín hiệu, tức nếu dạng sóng thay đổi nhanh, người ta dùng các bước lượng tử lớn, và ngược lại.

Ban viễn thông quốc tế ITU-TS² khuyến cáo một loạt lược đồ nén tiếng nói. Khuyến cáo này như trong bảng, dùng cho truyền tiếng nói. Do vậy mà băng thông hạn chế 3.4 – 7 kHz. Riêng G.722 sử dụng 8 bit cho mẫu trong quá trình ADC. Hầu hết khuyến cáo dựa trên kỹ thuật ADPCM, trừ G.722 dùng ADPCM băng con và G.728 dùng kỹ thuật lượng tử vectơ.

Bảng. Tóm tắt về các khuyến cáo về mã hóa tiếng nói ITU-TS

<i>Recommendation</i>	<i>Compression technique</i>	<i>Speech bandwidth (kHz)</i>	<i>Sampling rate (kHz)</i>	<i>Compressed bit rate (kbps)</i>
G.711	PCM (no compression)	3.4	8	64
G.721	ADPCM	3.4	8	32
G.722	Subband ADPCM	7	16 (14 bits per sample)	48,56,64
G.723	ADPCM	3.4	8	24
G.728	Vector quantization	3.4	8	16

Trong ADPCM băng con, tín hiệu tiếng nói được tách thành hai băng con, băng tần số cao và thấp; đều được mã nhờ ADPCM. Mục đích của việc tách này: băng con tần số thấp là quan trọng hơn, nên cần được mã chính xác. Băng con tần số cao được mã thô sẽ cung cấp thông tin bổ sung để cải thiện chất lượng tiếng nói trên băng tần số thấp kia.

Kỹ thuật lượng tử vectơ sử dụng trong G.728 có nguyên lý tương tự kỹ thuật nén

¹ differential pulse-coded modulation

² International Telecommunication Union-Telecommunications Sector. ITU-T là cụm từ viết tắt của International Telecommunication Union - Telecommunication Standardization Sector là lĩnh vực Tiêu chuẩn viễn thông - thuộc Tổ chức Viễn thông quốc tế.

LZW. Một số mẫu tiếng nói được nhóm lại vào vecto. Mỗi vecto được thể hiện bằng chỉ số đầu vào từ điển. Khác nhau duy nhất là việc khớp giữa vecto và mẫu trong từ điển, tức sách để mã, không cần chính xác; tức là lượng từ vecto giảm nhẹ.

2.2.5.4.3. Kỹ thuật nén MPEG cho âm thanh, dùng thuộc tính mặt nạ

Việc nén âm thanh và chi tiết hơn là nén tiếng nói đã được nêu trên, với giả thiết băng thông âm thanh trong 3.4 – 7 kHz. Tiếp sau đây là kỹ thuật nén âm thanh dài nghe thấy 20 kHz.

Có âm thanh nhỏ hay quá lớn mà người khác không nghe được. Đó được gọi là *mặt nạ*. Mặt nạ của một âm có thể là một phần hay toàn bộ âm. Các phần của âm có thể làm mặt nạ các phần của âm khác.

Định nghĩa: Mặt nạ¹: đoạn của âm thanh nhỏ hay quá lớn mà người khác không nghe được

Do âm thanh mặt nạ không nghe được, có thể bỏ chúng mà không tác động đến chất lượng âm thanh. Người ta dùng điều này trong chuẩn MPEG cho âm thanh. Nó là chuẩn nén âm thanh tổng quát. Không như kiểu nén chuyên dụng, MPEG âm thanh không quan tâm đến bản chất âm thanh nguồn, nhưng bộ mã hạn chế mức nhận biết của hệ thống nghe của con người. Do phần bỏ đi không gây nhiễu loạn mức nghe, MPEG âm thanh có thể được dùng cho âm thanh con người nghe.

MPEG âm thanh cho phép ba mức lấy mẫu 32, 44.1, hay 48 kHz. Dòng bit được nén có thể trợ giúp một hay hai kênh âm thanh. Dòng được nén có thể có một trong vài lưu lượng cố định xác định trước, trong dải 32- 244 kbps trên kênh. Tùy theo tỉ lệ lấy mẫu âm thanh, người ta có tỉ lệ nén từ 2.7 đến 24.

Do MPEG âm thanh chỉ loại bỏ phần âm không nghe thấy, nó là nén tổn thất, dù rằng dữ liệu âm thanh gốc được khôi phục từng bit sau khi nén. Người ta thử nghiệm và thấy tỉ lệ nén 6:1, điều kiện nghe lí tưởng, chuyên gia cũng không phát hiện được sai khác giữa đoạn âm thanh gốc với đoạn nén. MPEG âm thanh không chỉ dùng thuật toán nén đơn, mà ba lược đồ nén, gọi là MPEG âm thanh tầng 1, tầng 2, và tầng 3, với mức độ phức tạp tăng.

2.2.6. Hình ảnh số

Hình ảnh số có được do (i) số hóa ảnh và tranh in nhờ máy quét; nguyên lí số hóa như với dữ liệu âm thanh (ii) một phần, khung của video số hóa; (iii) hình ảnh số được tạo nhờ phần mềm đồ họa. Ở dạng thô, bitmap, tất cả các hình ảnh số được thể hiện theo cùng cách, không kể nguồn khác nhau.

2.2.6.1. Thể hiện ảnh số

Hình ảnh có thể là xám hay có màu. Đối với ảnh xám, dễ thể hiện. Vấn đề phức tạp chỉ với ảnh màu.

2.2.6.1.1. Thể hiện các ảnh xám

Hình ảnh trên màn máy tính gồm nhiều chấm. Các chấm nhỏ được gọi là pixel. Một hình ảnh được tạo bởi ma trận các pixel, với số các chấm ngang, dọc cố định. Chẳng hạn dùng 8 bit để thể hiện pixel, có 256 mức cường độ sáng khác nhau, tối nhất ứng với

¹ mask

mức 0, sáng nhất 255. Vậy ảnh mức xám có thể được thể hiện như ma trận hai chiều, các phần tử là pixel. Phân bố các giá trị pixel trong vùng liên thông trong bộ nhớ được gọi là bitmap. Bộ nhớ lưu dữ liệu ảnh bitmap được gọi là kho khung, hay vùng đệm khung.

Định nghĩa: Bitmap: Phân bố các giá trị pixel trong vùng liên thông trong bộ nhớ.

2.2.6.1.2. Thể hiện các ảnh màu

Ảnh màu cũng được tạo nên nhờ các ma trận pixel. Điểm khác với ảnh xám là người ta sử dụng ba giá trị cho mỗi pixel. Thể hiện này dựa trên lý thuyết về màu, cho biết màu bất kỳ được tạo nên bởi ba màu sơ cấp. Trong TV, người ta dùng ba màu: đỏ, xanh lá cây, và xanh dương. Vậy nên có thể sử dụng ba ma trận hai chiều đối với ảnh màu, ứng với các màu sơ cấp.

2.2.6.2. Tham số chính của ảnh số

Chất lượng ảnh tùy vào một số tham số.

1. Kích thước ảnh cho phép xác định số pixel theo dòng, cột. Do vậy người ta thường xem kích thước một ảnh như x.y. Để đảm bảo chất lượng hình ảnh, tức ảnh chụp, tranh, khung video, số pixel trên dòng và số dòng không thấp hơn số tối thiểu mà định lý Nyquist đã nêu; nếu không ảnh sẽ khác với ảnh gốc.
2. Độ sâu của pixel xác định số bit dùng cho mỗi pixel. Người ta dùng 8 bit cho pixel đối với ảnh xám tạm được. Đối với ảnh màu, mỗi pixel cần 24 bit.



Hình. Harry Nyquist (1889 - 1976)

Khi biết kích thước ảnh và độ sâu pixel, người ta tính được lượng dữ liệu sử dụng cho ảnh, chẳng hạn khối lượng $D = 512 \text{ pixel} * 512 \text{ dòng} * 24 \text{ bit} = 768 \text{ KB}$.

2.2.6.3. Nén ảnh

2.2.6.3.1. Giới thiệu

Hình ảnh số thô, và dữ liệu đa phương tiện khác, cần nhiều khối lượng bộ nhớ lưu trữ. Vậy nén ảnh là vấn đề cần giải quyết trong ứng dụng đa phương tiện. Trước khi trình bày các chuẩn nén ảnh, người ta xem xét lý do ảnh có thể nén.

1. Người ta chỉ cần ảnh đủ cho con người nhận biết. Đối với ảnh số, các mẫu kẻ tương tự nhau, có thể được loại bỏ. Sự tương tự này gây thừa về không gian, có thể được loại bỏ nhờ kỹ thuật mã dự đoán và kỹ thuật mã chuyển.
2. Trong truyền thông, người dùng có thể bỏ qua lỗi nhỏ miễn là không ảnh hưởng đến hiệu năng truyền. Điều này có nghĩa chấp nhận thông tin khôi phục lại không cần chính xác như bản gốc, khác với xử lý văn bản không được mắc lỗi.
3. Người ta có thể sử dụng đặc trưng khác để đạt độ nén cao: nhạy cảm về nhận thức là

khác nhau đối với các mẫu tín hiệu khác nhau.

Do vậy kỹ thuật nén cố gắng giữ thông tin quan trọng và bỏ thông tin không cần thiết.

2.2.6.3.2. Lấy mẫu con không gian

Khi có dữ liệu thừa trong ảnh, việc bỏ đi là bình thường. Trong kỹ thuật lấy mẫu con không gian, lúc tái tạo ảnh sau nén, các pixel bị mất được nội suy dựa vào pixel kề, hoặc ảnh tái tạo sử dụng độ phân giải thấp hơn trước. Ngoài ra, có thể mã hóa và hiển thị các hình ảnh lấy mẫu con nhỏ hơn. Nếu pixel có thành phần độ chói và màu vàng, thành phần màu vàng được lấy mẫu con tỉ lệ cao hơn, lượng tử thô hơn, do mắt của người ít nhạy cảm với màu vàng.

Kỹ thuật này là đơn giản, nhưng hiệu quả. Khi một trong bốn pixel được truyền, người ta có tỉ lệ nén 4:1. Kỹ thuật này được dùng trước khi dùng các kỹ thuật khác, tức ảnh sẽ được nén nhiều hơn.

2.2.6.3.3. Mã đoán trước

Nguyên tắc mã đoán trước cũng như với dữ liệu âm thanh. Các giá trị mẫu của pixel kề có liên quan. Sự phụ thuộc thống kê tuyến tính cho phép dự đoán tuyến tính các giá trị mẫu. Một thuật toán dự đoán một chiều sử dụng tương quan của các phần tử ảnh liền kề trong dòng quét. Lược đồ phức tạp khác sử dụng tương quan dòng với dòng, khung với khung, cho phép thuật toán dự đoán hai hay ba chiều. Trong hệ thống mã dự báo, sự thay đổi càng nhỏ của tín hiệu đem lượng tử hóa, mã hóa, hay truyền có thể làm giảm dải biên độ của bộ lượng tử, giảm số mức lượng tử, và càng ít bit cần thiết dùng cho pixel.

2.2.6.3.4. Mã chuyển hóa

Ý tưởng của việc mã chuyển là hủy tương quan giữa các pixel, tức chuyển các phần tử ảnh phụ thuộc thống kê thành các hệ số độc lập, và tập trung nội dung ảnh chỉ vào một ít hệ số, để có thể loại bỏ dư thừa trong ảnh. Trong mã chuyển hóa, ảnh gốc được chia thành các ảnh con, hay các phần nhỏ. Phép chuyển toán học một ngôi áp dụng cho ảnh nhỏ, từ miền không gian sang miền tần số. Như vậy, phân tích cực của ảnh tập trung vào một ít mẫu ở miền tần số thấp. Lưu ý rằng việc chuyển hóa không làm giảm dữ liệu, chỉ chuyển dữ liệu ảnh từ miền không gian sang miền tần số. Nếu dữ liệu trong miền không gian có tương quan cao, kết quả trong miền tần số sẽ có dạng phù hợp để giảm dữ liệu như kỹ thuật Huffman và mã độ dài chạy. Mã chuyển có ưu điểm giảm dư thừa dựa trên phụ thuộc thống kê. Nhiều loại chuyển một ngôi đã có, thông dụng như chuyển KLT¹, biến đổi cosin rời rạc DCT², chuyển WHT³, biến đổi Fourier rời rạc DFT⁴. Trong số đó, KLT là hiệu quả nhất đối với xử lý tương quan. Do có thuật toán tính nhanh cho DCT, kỹ thuật này được dùng nhiều để mã hóa ảnh.

Định nghĩa: Phép biến đổi cosin mô tả dãy các điểm dữ liệu hữu hạn theo thuật ngữ tổng các hàm cosin cho phép hiện tại các tần số khác nhau.

¹ Karhunen-Loeve transform
² discrete cosine transform
³ Walsh-Hadamard transform
⁴ discrete Fourier transform



Hình. Nhà toán học Joseph Fourier (1768-1830), phân tích điều hòa

Việc cài đặt hệ thống chuyển hóa, người ta thấy có các bước chính sau:

1. Lựa chọn loại chuyển hóa. Loại tốt sẽ thể hiện hài hòa giữa việc (i) bỏ tương quan và khả năng tập trung năng lượng ảnh, với (ii) thực hiện khả thi. Trong nhiều trường hợp, DCT có ưu điểm.
2. Chọn kích thước khối và áp phép chuyển lên khối. Nếu kích thước quá nhỏ, tương quan giữa các pixel lân cận không lộ rõ. Khối quá lớn sẽ làm tăng độ phức tạp của phần cứng và làm cho các biên lộ rõ. Kích thước 8x8 là phù hợp.
3. Lựa chọn và lượng tử hóa hiệu quả các hệ số bị loại bỏ để truyền hay lưu trữ.
4. Bước cuối phân bố bit cho các hệ số lượng tử. Có thể sử dụng mã Huffman hay mã độ dài chạy.

Định nghĩa: Phép biến đổi Fourier: trong toán học là phép toán chuyển một hàm với giá trị phức của các biến thực thành hàm khác. Trong ứng dụng như xử lý tín hiệu, miền xác định của hàm điển hình là thời gian, gọi là miền thời gian. Miền mới của hàm là miền tần số; hàm mới được gọi là thể hiện miền tần số của hàm ban đầu. Nó mô tả các tần số được thể hiện trong hàm ban đầu.

2.2.6.3.5. Lượng tử vecto

Trong các lược đồ mã hóa nêu trên, lượng tử hay mã thực sự được thực hiện trên các đại lượng vô hướng, ở đây là các mẫu giá trị thực về dạng sóng hay pixel, hình ảnh.

Mã theo cách chuyển hóa (i) chuyển các khối pixel; rồi (ii) mã riêng biệt các hệ số chuyển. Mã dự đoán (i) lượng tử hóa thuật ngữ lỗi tạo nên như sự sai khác giữa mẫu mới và dự đoán của mẫu mới, theo các đầu ra đã có.

Kết quả cơ bản của lý thuyết Shannon về tỉ lệ biến dạng, liên quan đến việc nén dữ liệu, cho biết có thể đạt được hiệu năng tốt hơn nhờ mã hóa theo vecto, tức nhóm các giá trị, thay vì mã hóa đại lượng vô hướng, tức các giá trị đơn.



Hình. Claude Shannon (1916-2001)

- Trước kia, lý thuyết này tác động có hạn đến thiết kế hệ thống thực sự, do (i) lý thuyết Shannon không đảm bảo các kỹ thuật thiết kế kiến thiết để mã hóa theo vectơ; (ii) mã theo vô hướng thường đạt được hiệu năng mong muốn khi tinh chỉnh phù hợp.
- Trên chục năm qua, lượng tử vectơ VQ^1 được xem xét. Lượng tử vectơ được xác định như ánh xạ Q từ không gian Oclit K chiều R^K sang tập con hữu hạn Y của R^K , $Y = (x_i)$, $i = 1..n$, x_i là vectơ thứ i của Y .

Y là tập các vectơ được sinh ra, được gọi là sách mã VQ , hay bảng VQ ; N là số các vectơ trong Y . Trong bộ tạo mã, mỗi vectơ dữ liệu x trong R^K được khớp hay xấp xỉ với từ mã trong sách mã, và địa chỉ, hay chỉ số, của từ mã đó được chuyển, thay vì chuyển vectơ dữ liệu. Trong bộ giải mã, chỉ số này được ánh xạ với từ mã sách, và từ mã này được dùng để thể hiện vectơ dữ liệu gốc. Trong bộ mã hay giải mã, có mã định danh với các lỗi vào chứa tổ hợp các pixel trong khối. Giả sử kích thước khối ảnh là $n \times n$ pixel và mỗi pixel được thể hiện qua m bit, về lý thuyết có thể có $(2.m).n.n$ dạng khối. Thực tế chỉ có số giới hạn các tổ hợp thường xảy ra, nên giảm được đáng kể kích thước của sách mã. Đó là cơ sở cho lượng tử vectơ. Nếu sử dụng thuộc tính của hệ thống nhìn của con người, kích thước sách mã có thể giảm nữa và vài bit là đủ để thể hiện chỉ số cho các lỗi vào sách mã. Nguyên lý lượng tử hóa vectơ là tương tự như thuật toán LZW, sử dụng một chỉ số hay thẻ bài để thể hiện khối dữ liệu.

2.2.6.3.6. Mã hình ảnh Fractal

Các kỹ thuật nén ảnh chia nhỏ là bước phát triển đáng kể để nén ảnh.

Định nghĩa: Một phần chia nhỏ²: là một hình ảnh về bề mặt hay hình dạng được thể hiện bằng một hay nhiều công thức toán

Theo hình học chia nhỏ, một phần nhỏ là dạng hình học có chi tiết khác thường lặp lại theo góc và thang độ khác, mà điều này có thể mô tả qua phép chuyển, tức công thức, chia nhỏ hay có liên quan. Các phần nhỏ được dùng để tạo ảnh trong ứng dụng như các cảnh mô phỏng bay, hiệu ứng đặc biệt trong các hình động. Các công thức chia nhỏ giờ đây có thể dùng cho hầu hết hình ảnh thế giới thực.



Hình. Thí dụ về chia nhỏ fractal

Nén ảnh chia nhỏ ngược với tạo ảnh chia nhỏ: thay vì tạo một hình ảnh theo công thức, nén ảnh chia nhỏ cần tìm tập các phần nhỏ trong hình ảnh số hóa mà mô tả toàn thể hình ảnh. Một khi các tập phù hợp của các phần nhỏ được xác định, chúng được nén, tức làm giảm, đến mã hay công thức của phần rất nhỏ. Các mã này là các qui tắc cho phép sinh lại các tập đa dạng của các phần nhỏ; các phần này tạo lại toàn thể hình

¹ vector quantization
² fractal

ảnh. Do các mã biến đổi chia nhỏ cần lượng rất nhỏ dữ liệu để thể hiện và lưu trữ công thức, nén chia nhỏ có tỉ lệ nén cao. Mã chia nhỏ có tính bất đối xứng cao, do các xử lý có ý nghĩa nhất là tìm kiếm và mã hóa, chứ không trong quá trình giải mã. Cần nhiều biến đổi và so sánh để tìm tập các phần nhỏ, trong khi quá trình giải mã chỉ đơn giản tạo hình ảnh theo công thức chia nhỏ.

2.2.6.3.7. Nén sóng nhỏ

Biến đổi sóng nhỏ cũng được dùng cho việc nén ảnh. Nguyên lý cơ bản của nén sóng nhỏ tương tự như nén theo biến đổi DCT, tức chuyển các tín hiệu từ miền thời gian sang miền mới, mà các bộ phận cốt lõi của tín hiệu được giữ và mã hóa để nén dữ liệu.

2.2.6.3.8. Hệ thống mã thực tế

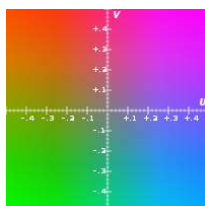
Trong hệ thống ứng dụng, người ta kết hợp vài kĩ thuật mã hóa nêu trên để có tỉ lệ nén cao, mà chất lượng ảnh chấp nhận được. Có nhiều tổ hợp đối với ứng dụng khác nhau. Lược đồ thường dùng để mã hóa dữ liệu video tuân theo các bước sau: (i) lấy mẫu con thời gian và không gian; (ii) dùng DPCM theo ước lượng chuyển động và có bù; (iii) dùng DCT hai chiều; (iv) mã Huffman; (v) mã độ dài chạy. Lược đồ lai này là cơ sở của nhiều chuẩn quốc tế hiện dùng.

2.2.6.3.9. Chuẩn JPEG

Định nghĩa: Chuẩn JPEG: JPEG là chuẩn nén ảnh số quốc tế đầu tiên đối với các hình ảnh tĩnh với mức độ, sắc thái liên tục, cho cả ảnh xám lẫn ảnh màu.

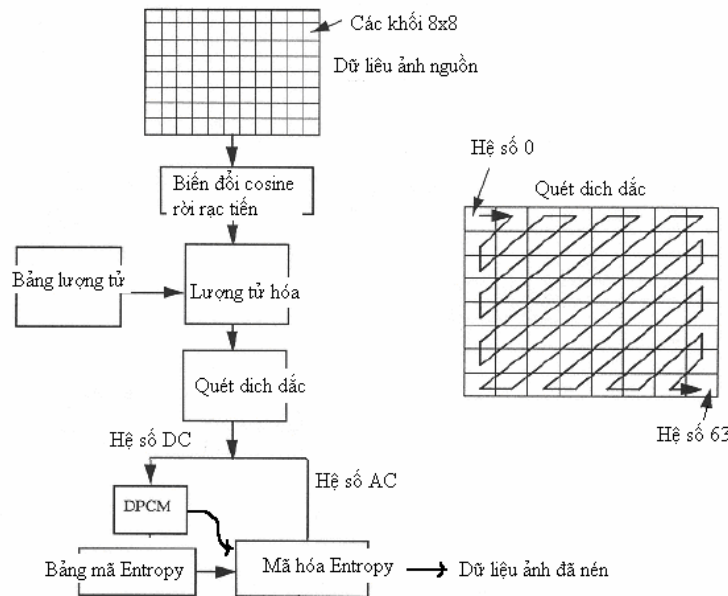
Chuẩn này được dùng rộng rãi trong ứng dụng cần đến hình ảnh, được cài đặt trên cả phần cứng lẫn phần mềm. Tuy chuẩn sử dụng cho ảnh tĩnh, đôi khi nó được dùng cho nén video. Ứng dụng này được gọi là JPEG chuyển động hay MJPEG. Chuẩn JPEG xác định bốn kiểu thao tác:

1. Mã DCT tuần tự giảm, trong đó mỗi thành phần hình ảnh được mã theo việc quét trên xuống, trái sang phải.
2. Mã DCT giảm khai triển, đảm bảo tăng cường cho phép toán đường ranh giới. Một kiểu tăng cường được biết đến là mã không ngừng, trong đó hình ảnh được mã theo quét nhiều lần để sinh được hình ảnh giải mã thô, nhanh, khi băng thông chậm.
3. Mã không tổn thất, mà hình ảnh được mã với đảm bảo tái tạo chính xác.
4. Mã phân cấp, trong đó hình ảnh được mã theo nhiều độ phân giải. Kiểu vạch ranh giới thường được sử dụng. Dữ liệu ảnh nguồn được chuyển thành thành phần sáng Y và hai thành phần màu U và V. U và V được lấy mẫu không gian chuẩn theo chiều ngang và dọc theo hệ số 2 để lợi dụng mức nhạy thấp của nhận thức con người đối với tín hiệu ánh sáng vàng.



Hình. Màu U-V, giá trị Y = 0.5

Giả sử các giá trị mẫu gốc trong dải $[0, 2^{b-1}]$, mẫu mẫu sử dụng b bit. Các giá trị này trượt trong dải $[-2^{b-1}, 2^{b-1}-1]$. Dải giá trị mới này cho phép tính chính xác mức thấp trong DCT. Đối với đường ranh giới, mỗi mẫu dùng 8 bit. Vậy dải giá trị gốc $[0, 255]$ trượt thành $[-128, 127]$.



Hình. Các bước mã theo DCT tuần tự

Do vậy mỗi thành phần được chia thành khối 8x8 pixel. Phép chuyển DCT 2 chiều tiên, FDCT, áp dụng lên mỗi khối dữ liệu. Kết quả là ma trận 64 hệ số với năng lượng tập trung vào một số ít hệ số đầu tiên. Các hệ số này được lượng tử hóa với giá trị lượng tử được xác định trong bảng lượng tử. Các giá trị lượng tử được dùng để điều chỉnh giữa tỉ lệ nén và chất lượng nén. Việc lượng tử hóa chia hệ số DCT cho giá trị lượng tử đã xác định. Các hệ số được lượng tử hóa được quét dịch dắc, thu được dãy dữ liệu một chiều dùng cho mã Huffman. Hệ số đầu tiên được gọi là hệ số DCT, cho biết cường độ trung bình của khối. Phần còn lại của các hệ số này, 1 đến 63, được gọi là hệ số AC. Mục đích của quét dịch dắc là sắp xếp tăng dần các hệ số theo tần số phổ. Do hệ số tần số cao, ở góc phải dưới của ma trận hệ số, việc quét dịch dắc sẽ cơ sở nhiều số 0 tại cuối vùng quét, cho phép hiệu quả khi dùng mã Huffman và mã độ dài chạy.

Hệ số DC liên quan theo mã DPCM với hệ số DC của khối trước. Các hệ số AC được mã theo độ dài chạy. Các hệ số DC mã theo DPCM và AC mã theo độ dài chạy được nén theo mã Huffman. Đầu ra của quá trình là dữ liệu ảnh đã nén.

2.2.6.3.10. Chuẩn JBIG

JBIG là chuẩn ISO, xác định thuật toán nén không tổn thất đối với hình ảnh nhị phân, tức ảnh dùng 1 bit cho pixel. Ý của JBIG là thay thế thuật toán gửi bản sao thuộc nhóm 3, 4 kém hiệu quả đang dùng. Nén JBIG dựa trên tổ hợp mã đoán trước và mã số học.

Chuẩn này có thể dùng cho ảnh xám hay ảnh màu mà không tổn thất, bằng cách dùng thuật toán mặt 1 bit. Việc này thực hiện tốt với 6 bit cho mỗi pixel; nhưng khi đó chuẩn JPEG không tổn thất lại tốt hơn.

2.2.6.3.11. Chuẩn JPEG-2000

Chuẩn JPEG-2000 là chuẩn mới, 2009, có các đặc trưng chính sau:

- Cung cấp kiểu thao tác bit thấp so với chuẩn JPEG, JBIG, cho chất lượng ảnh và mức biến dạng tốt hơn;
- Cho phép xử lý ảnh lớn hơn 64x64 K pixel mà không cần dùng kỹ thuật lát gạch vuông¹, cho phép dùng thuật toán giải nén đơn đối với tất cả các ảnh để tránh vấn đề về không tương thích về kiểu chuẩn;
- JPEG-2000 dễ truyền trong môi trường ồn, cũng hiệu quả để nén hình ảnh do máy tính tạo ra;
- Nó được dùng cho các hình ảnh sắc thái liên tục hay hình ảnh hai mức, để trao đổi, lưu trữ các tư liệu tổng hợp trong hệ thống ảnh tư liệu, với khuôn dạng tổn thất hay không tổn thất;
- Cho phép truyền tăng dần các ảnh, nhằm các hình ảnh được tái tạo với chính xác pixel, độ phân giải không gian tăng. Các hình ảnh sẽ được truyền thời gian thực trên băng thấp sang không gian bộ nhớ hạn hẹp;
- Chuẩn cho phép xác định vùng ảnh được nén với mức biến dạng thấp hơn vùng khác;
- Cung cấp tiện nghi mô tả theo nội dung ảnh, cơ chế thủy vân, đánh nhãn, dán tem, và ảnh mật;
- Các ảnh tĩnh tương thích với các ảnh động MPEG-4.

2.2.7. Video số

2.2.7.1. Thể hiện video số

Video số thường thu được do số hóa video tương tự. Xâu video số gồm nhiều khung hay hình ảnh thể hiện với tốc độ cố định. Tỷ lệ khung được xác định theo ba yếu tố: (i) tỷ lệ khung đủ lớn để chuyển động mềm mại, thường trên 25 fps; (ii) tỷ lệ khung cao đòi hỏi băng thông lớn, để truyền kịp thời. Vậy 25 fps là hợp lý; (iii) cần tránh thể hiện hình nhấp nháy. Do màn hình sáng do điện tử bắn vào, tắt ngay khi không bị bắn, nên màn hình cần làm tươi trước khi có điện tử khác bắn, để khoảng thời gian không quá lớn. Nghiên cứu cho thấy làm tươi 50 lần trong một giây là đủ. Trường hợp sử dụng 50 fps, băng thông cần tăng đáng kể. Để giải quyết vấn đề này, kỹ thuật hiển thị, gọi là đan chéo², được dùng trong các TV, sử dụng quá một bộ quét cột để tạo lại mỗi khung hoàn chỉnh. Mỗi quét cột được gọi là màn hình³. TV phát theo đan chéo 2:1, tức hai màn hình đứng cho một khung hoàn chỉnh. Trong đan chéo 2:1, một màn hình, gọi là màn hình lẻ, hiển thị tất cả các dòng lẻ của của khung, và quét cột thứ hai, gọi là màn hình chẵn, hiển thị tất cả các dòng chẵn của khung. Do vậy với 25 fps, tỷ lệ màn hình là 50 màn hình trên giây. Do mắt người không phân biệt được các đối tượng nhấp nháy nhỏ, tỷ lệ nhận biết lặp 25 lần trong 1 giây của dòng sẽ không nhấp nháy. Theo các yếu tố này, hai tỷ lệ khung thông dụng được dùng trong phát TV: (i) loại 25 fps trong hệ thống PAL tại các nước châu Âu, Trung quốc và Úc; (ii) 30 fps dùng trong hệ thống NTSC tại Bắc Mỹ và Nhật bản.

¹ tiling

² Interlace

³ field

Số mẫu 50, 60 được chọn để khớp với tần số mạng phân phối điện tại các nước trên, để giảm tác động đến chất lượng ảnh thể hiện.

Hai tính chất chính của video là (i) có chiều thời gian; (ii) khối lượng dữ liệu lớn cần thể hiện. Thường thì 10 phút video, với 30 fps, hình ảnh kích thước 512 pixel/ dòng, 512 dòng, độ sâu pixel là 24 bit/pixel, cần đến bộ nhớ $600 \times 30 \times 512 \times 512 \times 3 = 13.8$ GB. Vậy nén là việc không tránh được.

2.2.7.2. Nén video

Việc nén video nhằm giảm dư thừa và phù hợp với tính chất nhìn của con người. Video số là chuỗi các hình ảnh, nên sẽ dư thừa không gian. Ngoài ra, các hình kề nhau là tương tự. Tương tự này được gọi là dư thừa thời gian, được loại bỏ nhờ mã dự báo đối với các hình ảnh. Các kỹ thuật nén mô tả trong phần trước được sử dụng trong nén video. Sau đây là kỹ thuật nén loại bỏ dư thừa thời gian và một số chuẩn nén.

2.2.7.2.1. Ước lượng chuyển động và bù

Ước lượng chuyển động và bù xử lý thừa thời gian trong video. Chuyển động trong các hình ảnh ngầm cho biết các pixel trong hình ảnh trước ở vị trí khác với vị trí trong hình ảnh hiện tại. Trong các kỹ thuật ước lượng và bù chuyển động, mỗi hình ảnh được tách thành hai khối kích thước cố định. Trong hình ảnh trước có phần khớp nhất đối với mỗi khối. Trượt vị trí giữa hai khối được gọi là vectơ chuyển động. Khối chênh lệch thu được bằng độ lệch theo từng pixel. Vectơ chuyển động và khối lệch theo pixel sẽ được mã hóa và truyền. Sự lệch pixel thường nhỏ và cho phép hiệu quả để truyền vectơ chuyển động cùng với độ lệch, thay vì truyền mô tả của bản thân khối hiện tại.

2.2.7.2.2. Chuẩn MPEG

Năm 1988, chuẩn MPEG được lập tại JTC 1¹ về công nghệ thông tin. Nhiệm vụ của nhóm chuẩn này là thể hiện mã hóa các hình động, âm thanh liên kết, và tổ hợp của chúng trong lưu trữ và tìm kiếm trên các phương tiện số DSM². Các DSM được xem là thiết bị lưu trữ thông thường, như CD-ROM, thiết bị băng, đĩa quang, các kênh truyền ISDN, mạng LAN. Ban đầu có ba nhóm, để mã hình động và âm thanh liên kết với (i) 1.5 Mbps; (ii) 10 Mbps; và (iii) 40 Mbps, ứng với MPEG-1, 2, 3.

- Tập trung của MPEG-1 là mã video chất lượng VHS, 360x280 pixel, 30 fps, lưu lượng khoảng 1.5 Mbps. Tỷ lệ 1.5 Mbps được chọn do thông lượng thiết bị CD-ROM vào lúc đó.
- Ban đầu MPEG-2 mã hóa video chất lượng TV số CCIR 601, 720x480 pixel, 30 fps, lưu lượng 2-10 Mbps.
- Lúc đầu MPEG-3 dùng để mã video chất lượng HDTV với lưu lượng 40 Mbps. Về sau nó thực hiện chức năng theo MPEG-2, với yêu cầu mới, nên nó dừng vào 1992. Tuy nhiên nó đưa ra vấn đề về mã hóa theo đối tượng và lưu lượng thấp đối với thông tin nghe nhìn.
- MPEG-4 bắt đầu 1991, được chấp thuận 1993. Tới nay, MPEG đã có ba chuẩn, là MPEG-1, 1993, MPEG-2, 1994. MPEG-4, 1999.

Các hoạt động của nhóm MPEG không chỉ cho nén video, vì nén âm thanh liên kết

¹ Joint ISO/IEC Technical Committee
² digital storage media

và đồng bộ nghe nhìn không thể tách rời nén video. Các chuẩn MPEG có ba phần chính (i) MPEG video; (ii) MPEG âm thanh; (iii) MPEG hệ thống. MPEG hệ thống xử lý khía cạnh đồng bộ hóa và dồn kênh các dòng bit nén hỗn hợp âm thanh và video. Ngoài ra có phần 4, gọi là phần thích nghi¹, mô tả các thủ tục cho phép quyết định đặc tính của dòng bit được nén và cho phép thử đáp ứng các yêu cầu do ba phần kia thiết lập.

Lưu ý rằng điều quan trọng nhất của chuẩn MPEG là chúng chỉ mô tả cú pháp dòng bit được mã, nên bộ giải mã theo chuẩn này có thể giải mã dòng bit. Các chuẩn không mô tả cách tạo dòng bit, mà dành cho việc thiết kế, cài đặt..

Các chuẩn MPEG-1, MPEG-2 được dùng rộng rãi trong ứng dụng đa phương tiện. Tương lai chuẩn này với ứng dụng nghe nhìn không tách rời. Người ta đề xuất dùng MPEG-2 như chuẩn mã HDTV tại châu Âu và Hoa kì.

2.2.7.2.3. MPEG-1

Chuẩn MPEG-1 nhằm mã hóa video và âm thanh liên kết với lưu lượng 1.5 Mbps. Dù hoạt động tối ưu với lưu lượng đó, chuẩn này không hạn chế lưu lượng lớn hay nhỏ hơn. MPEG-1 sử dụng tổ hợp nén chuyển động và các kĩ thuật mã theo DCT. Các bước mã hóa chính gồm:

1. Các hình ảnh nhập vào được xử lý sơ bộ, chuyển không gian màu, bổ sung phân giải không gian, để được dạng thức riêng.
2. Các kiểu ảnh được quyết định đối với mỗi hình ảnh, tức các kiểu I, P, B hay D. Thách thức đối với thiết kế thuật toán MPEG là (i) để đạt tỉ lệ nén cao, cần có kĩ thuật mã giữa trong khung², giữa các khung, liên khung³; (ii) mã thuận tủy trong khung cần đáp ứng các yêu cầu tìm kiếm ngẫu nhiên. Để đạt được hai đích này, vài hình ảnh là khung mã hóa và các hình ảnh khác là liên khung được mã hóa. MPEG xác định 4 loại ảnh sau:
 - Ảnh trong, I⁴, được mã không kèm tham chiếu ảnh khác. Chúng cung cấp các điểm truy cập đến xâu mã hóa, để việc giải mã bắt đầu, nhưng được mã với mức nén bình thường. Chúng được dùng như hình ảnh tham chiếu cho các ảnh dự báo.
 - Ảnh dự báo P nhằm mã hóa hiệu quả hơn bằng cách sử dụng dự đoán bù chuyển động của ảnh I, ảnh P trước đó. Ảnh P cũng được dùng cho dự đoán ảnh kế tiếp.
 - Ảnh dự đoán theo hai hướng B, cung cấp bù chuyển động mức cao dựa trên các ảnh trước và sau. Người ta không dùng ảnh B trong dự đoán ảnh khác.
 - Ảnh theo mã DC D, được mã không với tham chiếu ảnh khác. Vậy chỉ hệ số DC trong DCT được sử dụng. Ảnh D chỉ nhằm tìm kiếm nhanh về trước.

Các ảnh I và D được gọi chung là các ảnh tham chiếu, ảnh P và B được gọi

¹ Conformance

² intraframe

³ Interframe

⁴ I, B, P, D: intracoded pictures, predictive-coded pictures, bidirectionally predictive-coded pictures, bidirectionally predictive-coded pictures

chung là ảnh dự đoán. Việc tổ chức ba loại ảnh I, P, B trong xâu video là linh hoạt. Việc chọn thích hợp tùy theo cài đặt trong ứng dụng. Diễn hình là dùng ảnh I sau mỗi nửa giây, với hai ảnh B giữa cặp ảnh I và P.

3. Mỗi hình ảnh được chia thành các khối gộp¹ 16x16 pixel. Khối gộp theo ảnh I được mã thực sự. Khối gộp theo ảnh P hoặc được mã thực, hoặc được dự đoán trước theo ảnh I, P trước đó, tùy theo tính hiệu quả của việc mã hóa. Khối gộp theo ảnh B được hoặc là mã thực sự, dự đoán trước, dự đoán sau, mã theo dự đoán hai phía. Đối với các khối gộp dự đoán, gồm BP, FP và khối gộp B, có các vecto chuyển động cùng với sai số dự báo.
4. Các khối gộp theo mã hóa trong và các lỗi dự báo của các khối gộp theo mã dự báo được chia thành các khối 8x8 pixel, có bốn thành phần sáng và hai màu vàng. DCT hai chiều sử dụng cho mỗi khối, thu được các hệ số biến đổi, rồi được lượng tử hóa và quét dịch đặc.
5. Các hệ số biến đổi lượng tử hóa và thông tin bên cạnh, như loại ảnh, địa chỉ khối gộp, vecto chuyển động, được mã hóa bằng tổ hợp Huffman và độ dài chạy.

Chuẩn này không mô tả quá trình mã, dù chúng là các hàm khuyến cáo dùng cho bộ mã hóa. Chuẩn mô tả dạng thức của dòng bit ra từ bộ mã hóa, để bộ giải mã có thể phân tích thông tin được mã trong dòng bit. Chuẩn không cho biết cách cài đặt bộ giải mã, miễn là dòng bit ra phù hợp với dạng thức đã mô tả.

2.2.7.2.4. MPEG-2

MPEG-2 là mở rộng của MPEG-1 để cung cấp mã tín hiệu nghe nhìn chất lượng cao. Dù có khái niệm như nhau, nhưng nó được cải tiến nhiều so với MPEG-1. Sau đây là một vài khác biệt theo ba phần, hệ thống, video và âm thanh.

Chức năng của MPEG-2 hệ thống như là MPEG-1 hệ thống: mô tả dạng thức mã để dồn kênh âm thanh, video, và dữ liệu khác thành dạng thích hợp để truyền hay lưu trữ. Mỗi quan hệ thời gian trong số phương tiện có liên quan được mô tả để mã hóa đồng bộ và thể hiện trong bộ giải mã. MPEG-2 mô tả hai dạng thức dòng dữ liệu (i) dòng chương trình: Dòng chương trình tương tự MPEG hệ thống, tương thích với dòng MPEG-1. Dòng chương trình được tối ưu đối với ứng dụng tìm kiếm đa phương tiện, đối với xử lý hệ thống trong phần mềm; (ii) dòng vận chuyển có thể chuyển đồng thời nhiều chương trình và được tối ưu để dùng trong ứng dụng mà có thể mất dữ liệu. Dòng vận chuyển gồm các gói độ dài cố định. Khái niệm này tương tự như trong tế bào ATM, nhưng dài 188 byte, có 4 byte đầu, thay vì 53 byte. Dòng vận chuyển thích hợp để truyền thoại video, TV số trên đường cáp quang, vệ tinh, mạng ATM và mạng khác.

Nguyên lý nén MPEG-2 video như nén MPEG-1 video, với vài cải tiến để đạt chất lượng cao. Các mở rộng chính:

1. Trợ giúp cả video đan chéo hay không đan chéo. Nó trợ giúp video đan chéo CCIR 601.
2. Cú pháp dòng video cho phép kích thước hình ảnh lớn 16.383x16.383 pixel.

¹ Macro block

3. Cú pháp dòng video cung cấp chỉ thị về đặc tính video hỗn hợp nguồn, như PAL, NTSC, để tiện cho phép hậu xử lí.
4. MPEG-1 cho phép video vô hướng: bốn loại vô hướng mã dòng video MPEG-2 thành các tầng khác nhau, tức các tầng cơ sở, đệm và tầng cao, để ưu tiên cho dữ liệu video. Hai lí do chính (i) dữ liệu video quan trọng được ưu tiên để ít sai sót; khi nghẽn dữ liệu, các dữ liệu kém ưu tiên bị rớt, cho phép vẫn tái tạo dữ liệu; được gọi là suy giảm tốt; (ii) video vô hướng cho phép bộ giải mã phần video đã mã theo lựa chọn. Chẳng hạn nếu video HDTV được mã theo các tầng khác nhau, một tầng ứng với phân giải TV chuẩn, thì TV chuẩn cần cho việc giải mã chỉ tầng đó để đạt được chương trình TV bình thường.
5. Trong MPEG-1 video, các tham số ràng buộc xác định tập con các tham số dùng cho hầu hết ứng dụng MPEG-1. Trong MPEG-2, khái niệm tương tự xác định tập con các tham số để hạn chế khả năng cài đặt cụ thể. Đặc tả MPEG-2 video nhắm vào tổng quan trong cảnh, dùng trong diện rộng các ứng dụng, lưu lượng, độ phân giải, chất lượng và dịch vụ. Các ứng dụng bao trùm phương tiện lưu trữ số, truyền TV, truyền thông.

Xem xét các cài đặt thực tế với đầy đủ cú pháp của đặc tả, một số hạn chế tập con các cú pháp đã định ra *hình bóng* và *mức*¹. Hình xác định tập con của toàn cú pháp dòng bit. Trong biên mô tả, người ta vẫn có thể thay đổi để đạt hiệu quả về nén/ giải nén tùy theo các tham số lựa chọn. Trong hình bóng, có các mức.

2.2.7.2.5. MPEG-4

Định nghĩa: MPEG-4 cung cấp các công nghệ lõi chuẩn hóa, để lưu trữ hiệu quả, truyền, xử lí dữ liệu đa phương tiện.

Đặc trưng chính của phần đồ họa của MPEG-4 là các công cụ để giải nén và thể hiện các phần nhân của nội dung hình ảnh, video, gọi là các đối tượng video VO². Chẳng hạn VO là người đang vẽ, có thể tích hợp với đối tượng nghe nhìn khác để tạo nên các khung hay hình ảnh.

MPEG-4 mô tả mã và thể hiện của VO tự nhiên hay tổng hợp. VO tổng hợp được coi như đồ họa máy tính thích hợp để diễn hay cho hình động. Đối với VO tự nhiên, dạng hay biên của các đối tượng cần được xác định trước tiên. Thông tin về dạng và bề mặt đối tượng được mã tách biệt. Thông tin bề mặt được mã theo cách tương tự MPEG-1 video. Lưu ý MPEG-4 không mô tả cách quyết định biên đối tượng. Trong chuẩn này, người ta đã phát triển kĩ thuật phân đoạn đối tượng.

Với mã theo đối tượng, có thể dùng nén hiệu quả và tính vô hướng theo nội dung. MPEG-4 là phát triển quan trọng trong công tác mã hóa nghe nhìn, từ mã dựa trên pixel đến mã dựa trên đối tượng. Việc mã theo đối tượng cần đến chỉ số hóa và tìm kiếm theo nội dung trên dữ liệu đa phương tiện.

2.2.7.2.6. MPEG-7

Thành phần mới trong MPEG này được gọi là “giao diện mô tả nội dung đa phương

¹ Profile, level

² Video object

tiện”, xác định tập chuẩn các mô tả cho các loại thông tin đa phương tiện khác nhau. MPEG-7 chuẩn hóa cách xác định các mô tả khác, như cấu trúc, tức lược đồ mô tả, đối với các mô tả và mối quan hệ giữa chúng. Mô tả này, tức tổ hợp cả mô tả lẫn lược đồ mô tả, được gắn với nội dung, cho phép tìm kiếm hiệu quả và nhanh trên vật liệu¹ của người dùng. Các vật liệu AV² này cũng được chỉ số hóa để tiện tìm kiếm. Vật liệu bao gồm ảnh tĩnh, đồ họa, mô hình 3D, âm thanh, tiếng nói, video, hình động, và thông tin về cách các phần tử này kết hợp trong thể hiện đa phương tiện, theo kịch bản đa phương tiện. Chuẩn MPEG-7 xây dựng trên thể hiện khác như tương tự, PCM, MPEG-1, 2, 4. Một chức năng của chuẩn là cung cấp các tham chiếu đến phần thích hợp của chúng. Chẳng hạn cả mô tả dạng dùng trong MPEG-4 là hữu dụng trong ngữ cảnh MPEG-7, và cùng cách áp dụng cho các dải vectơ chuyển động dùng trong MPEG-1 và MPEG-2.

Đương nhiên các mô tả MPEG-7 không phụ thuộc vào cách mã hóa hay lưu trữ nội dung được mô tả. Có thể gắn mô tả MPEG-7 với phim tương tự hay với hình ảnh in trên giấy. Dù mô tả MPEG-7 không lệ thuộc vào thể hiện mã hóa của vật liệu, chuẩn theo cách của MPEG-4, cung cấp phương thức mã hóa vật liệu nghe nhìn như các đối tượng có quan hệ theo thời gian, tức đồng bộ, và không gian, tức trên cùng màn video, hay trong phòng âm thanh. Sử dụng mã MPEG-4, có thể gắn mô tả với phần tử, hay các đối tượng, trong cảnh, như các đối tượng nghe nhìn. Chuẩn MPEG-7 cho phép mức nhỏ, tức mức hạt, khác nhau trong mô tả, cho phép phân biệt chi tiết.

Do đặc trưng mô tả cần có nghĩa theo ngữ cảnh ứng dụng, đặc trưng là khác nhau đối với ứng dụng, lĩnh vực người dùng khác nhau. Điều này ngầm cho biết cùng một vật liệu có thể được mô tả theo các loại đặc trưng khác nhau, hướng về miền ứng dụng. Tất cả các mô tả này sẽ được mã hóa theo cách hiệu quả, tức hiệu quả tìm kiếm.

Mức trừu tượng có liên quan đến cách trích rút đặc trưng. Nhiều đặc trưng mức thấp có thể được rút ra theo cách tự động, trong khi các đặc trưng mức cao cần đến tương tác người-máy.

Dữ liệu MPEG-7 có thể được định vị với vật liệu AV liên kết, trong cùng dòng dữ liệu hay trên cùng vật lưu trữ, nhưng các mô tả vẫn là tổng quát. Khi nội dung và các mô tả không được định vị cùng nơi, các cơ chế liên kết vật liệu AV với mô tả MPEG-7 được sử dụng, cho phép làm việc theo hai hướng.

Lưu ý rằng MPEG-7 sẽ chuẩn hóa mô tả nội dung và đặc trưng đa phương tiện, nhưng không chuẩn hóa các công cụ trích đặc trưng và công cụ dùng để mô tả. MPEG-7 là chuẩn 2001.

2.2.7.2.7. Các chuẩn khác

Ngoài MPEG, các chuẩn khác là ITU-T H.261 và H.263. Nguyên lý nén là tương tự với MPEG. Chúng được phát triển cho hội nghị truyền hình và thoại video.

2.3. Phân loại

Sau một số thông tin chi tiết về các loại dữ liệu đa phương tiện, cần đề cập các chuẩn cho tư liệu đa phương tiện hỗn hợp. Phần trên đã trình bày các loại phương tiện đơn và các kỹ thuật nén. Đối với tư liệu hỗn hợp nhiều phương tiện, cần có chuẩn khác,

¹ Material
² AV

với các khái niệm và chức năng mới. Hiệu quả chính của các chuẩn này đối với tìm kiếm thông tin đa phương tiện là thông tin cấu trúc sẽ giúp cải thiện độ chính xác của việc chỉ số hóa và tìm kiếm.

2.3.1. Một số chuẩn

2.3.1.1. Chuẩn SGML

SGML¹ là chuẩn quốc tế, cho phép mô tả thông tin cấu trúc độc lập với cách xử lý thông tin. Nó cung cấp cú pháp chuẩn để xác định định nghĩa kiểu tư liệu DTD² cho lớp thông tin cấu trúc.



Hình. Trang SGML

Thông tin được tạo theo DTD, để cấu trúc của nó là hiện và truy cập được. Quá trình tạo này sẽ kiểm tra DTD và đảm bảo nó đúng, và cấu trúc thông tin phù hợp với lớp đã mô tả. HTML là một lớp của SGML. SGML và DTD không mô tả cách xử lý hay tạo dạng tư liệu, mà làm cấu trúc tư liệu hiện, với các phép như tạo dạng, hiển thị, tìm kiếm.

Định nghĩa: Ngôn ngữ đánh dấu chuẩn tổng quát SGML³ là công nghệ chuẩn ISO để xác định các ngôn ngữ đánh dấu cho tư liệu.

2.3.1.2. Chuẩn ODA

Đây là chuẩn quốc tế cho phép mô tả cách các thành phần logic và thể hiện của tư liệu có thể kết hợp và liên quan với nhau. ODA cũng mô tả dạng thức chuyển hóa để cấu trúc logic và cấu trúc thể hiện được nhúng trong dòng byte với nội dung tư liệu. Mục tiêu chính của chuẩn là thuận tiện việc trao đổi mở của các tư liệu nội dung hỗn hợp. Thông tin logic và thông tin thể hiện có thể trợ giúp quá trình tìm kiếm.

2.3.1.3. Chuẩn PDF

Hãng Adobe đã đưa ra chuẩn PDF⁴, như chuẩn sự kiện trong việc thể hiện tư liệu. PDF dựa trên ngôn ngữ mô tả trang Postscript, mô tả trang in, gồm cấu trúc logic và cấu trúc tạo dạng cho tư liệu. Trong tư liệu có đồ họa, hình ảnh lẫn văn bản. Không như Postscript, PDF cho phép dữ liệu được trích từ tư liệu, như dạng ngược lại. Nó cũng trợ giúp các kí pháp, liên kết siêu văn bản, tư liệu có cấu trúc trong ngôn ngữ đánh dấu như SGML, và có chức năng tương tự ODA.

¹ SGML:

² document type definition

³ Standard Generalized Markup Language

⁴ Acrobat Portable Document Format

2.3.1.4. Chuẩn HyTime

HyTime là chuẩn ISO để thể hiện các tư liệu siêu phương tiện theo SGML. Nó mô tả cấu trúc tư liệu, gồm phần tử tư liệu với siêu liên kết, mối liên quan giữa các phần tử tư liệu theo hệ thống tọa độ không gian, thời gian, và chiều đo khác. Tuy nhiên nó không cung cấp mô hình về dạng thức tương tác và thể hiện.

2.3.1.5. Chuẩn MHEG

Đây là chuẩn ISO, xác định mô hình hướng đối tượng cho trao đổi đa phương tiện. Nó mô tả dạng thức tư liệu để trao đổi thông tin giữa những hệ thống độc lập. Trước khi tư liệu được chuyển đến hệ thống khác, hay đến máy khách, nó được chuyển sang dạng MHEG. Tại máy khách, tư liệu theo dạng MHEG được chuyển, nhờ máy MHEG, sang dạng yêu cầu tại hệ thống cục bộ để hiển thị hay xử lý. Theo cách này, ứng dụng hay tư liệu đa phương tiện có thể có nhiều hệ thống khách.

2.3.2. Các đặc tính và yêu cầu của dữ liệu và ứng dụng đa phương tiện

Phần trên đã trình bày các dạng phương tiện. Tiếp theo cần thiết nêu các đặc tính của các phương tiện, nhằm phục vụ công tác thiết kế hệ thống MIRS hay hệ thống cơ sở dữ liệu đa phương tiện.

2.3.2.1. Yêu cầu về băng thông và lưu trữ

Yêu cầu lưu trữ được đo theo byte. Trong lĩnh vực số, băng thông được đo bằng lưu lượng, bps. Đơn vị cơ bản để lưu trữ là byte, truyền thông là bit.

Các hình ảnh đơn không có chiều thời gian. Nếu hạn chế thời gian truyền đối với hình ảnh, yêu cầu băng thông được tính theo yêu cầu lưu trữ. Trong nhiều ứng dụng, hình ảnh cần được hiển thị đồng bộ với phương tiện liên tục, như âm thanh. Khi đó, truyền hình ảnh thiết lập thời gian chính xác và yêu cầu về băng thông.

Cả âm thanh và video đều có thời gian liên tục. Người ta đặc trưng chúng theo bps. Đối với âm thanh, số này được tính theo tỉ lệ lấy mẫu và số bit cho mẫu. Lưu lượng đối với video được tính theo cùng cách, nhưng người ta tính theo khối lượng dữ liệu trong mỗi hình ảnh, tức mỗi khung, và số khung trong một giây. Kết quả sẽ mô tả lưu lượng băng truyền. Nếu lưu và tìm kiếm video và âm thanh số, kết quả này xác định tỉ lệ truyền theo yêu cầu lưu trữ. Nếu người ta biết thời lượng video, âm thanh, khối lượng bộ nhớ có thể được tính.

Định nghĩa: Khung chính¹: trong hình động và trong làm phim, khung chính xác định các điểm bắt đầu và kết thúc của một chuyển cảnh mềm. Chúng được gọi là “khung” do vị trí thời gian của chúng được đo theo khung trên kịch bản phim.

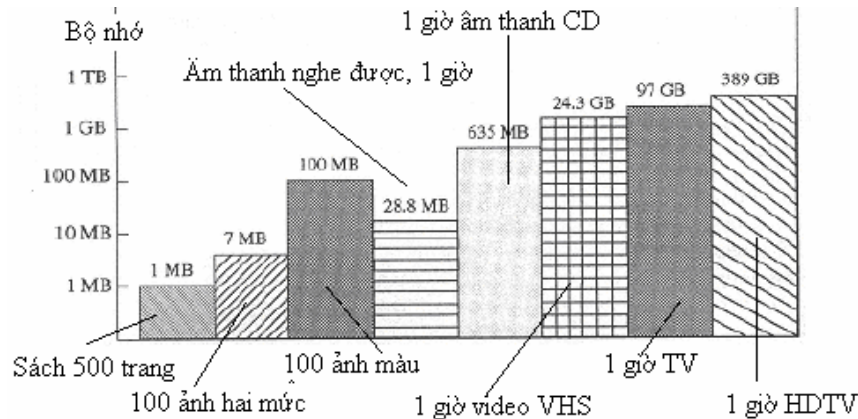
Bảng. Tỉ lệ bit thô của âm thanh thông thường và của các ứng dụng video

Các ứng dụng	Tỉ lệ dữ liệu (kbps)
CD âm thanh	1411.2
DAT	1536
Điện thoại số	64
Radio số, DAT dài	1024

¹ key frame

Video chất lượng TV	216000
Video VHS	54000
HDTV	864000

Bảng cho biết yêu cầu về băng thông của dữ liệu video, âm thanh liên tục, với chất lượng khác nhau, chưa nén. Hình cũng cho thấy yêu cầu lưu trữ đối với phương tiện thông dụng và phương tiện liên tục theo thời lượng khác nhau.



Hình. Yêu cầu bộ nhớ của phương tiện thông dụng

2.3.2.2. Cấu trúc ngữ nghĩa của thông tin đa phương tiện

Video, âm thanh và hình ảnh số là một loạt giá trị lấy mẫu, không có cấu trúc ngữ nghĩa rõ ràng. Điều này khác với thông tin số, chữ, được mô tả và định danh theo bảng mã ASCII hay mã tương đương. Theo mã này, các máy tính có thể tìm kiếm mục tin số, chữ trong cơ sở dữ liệu về tư liệu.

Để tìm kiếm thông tin thuận tiện, đặc trưng hay nội dung được trích từ dữ liệu đa phương tiện thô. Các đặc trưng, tổ chức của đặc trưng và so sánh tương tự giữa các mô tả đặc trưng là khía cạnh thiết kế chính của MIRS.

2.3.2.3. Yêu cầu về trễ và trễ nhảy

Yêu cầu về trễ do yếu tố thời gian và tính tương tác. Dữ liệu âm thanh và video phụ thuộc vào thời gian liên tục. Điều này có nghĩa chất lượng thể hiện dữ liệu đa phương tiện, các mẫu video, âm thanh phải tuân theo chuẩn. Chẳng hạn nếu âm thanh lấy mẫu 8 kHz, nó cần được thể hiện với 8.000 mẫu trên một giây. Phương tiện liên tục có chiều thời gian; tính chính xác của chúng phụ thuộc vào các giá trị mẫu, nhưng vào cả các mẫu thời gian thể hiện. Điều này trái với phương tiện tĩnh, như chương trình máy tính, tệp dữ liệu, mà tính chính xác chỉ tùy thuộc vào nội dung. Cho dù muốn nhận được phương tiện nhanh có thể, chúng cần đúng về nội dung. Ngoài ra, MIRS là hệ thống tương tác, dễ bị trễ.

2.3.2.4. Mối quan hệ không gian và thời gian của các phương tiện liên quan

Trong ứng dụng đa phương tiện, các loại phương tiện như phương tiện có cả tính lẫn động cần được lưu ý thể hiện. Để đạt được hiệu ứng mong muốn, tìm kiếm và truyền phương tiện liên quan này cần thể hiện được mối quan hệ thời gian. Theo ngữ cảnh này, (i) sự đúng đắn; hay (ii) hiện diện thời gian; hay (iii) thể hiện của mục tin đa phương tiện được gọi là đồng bộ hóa. Lược đồ đồng bộ xác định cơ chế dùng để đạt đến mức độ yêu cầu về đồng bộ. Công việc trong đồng bộ hóa đa phương tiện là đặt đúng

miền giữa hai miền (i) miền đầu tiên nhằm phát triển cơ chế và công cụ để tác giả hay người dùng mô tả quan hệ thời gian một cách dễ dàng; (ii) miền thứ hai nhằm đảm bảo các mối quan hệ thời gian đã định nhờ vượt qua bản chất không tất định của các hệ thống truyền thông.

2.3.2.5. Tính chủ quan và tính mờ về nghĩa của dữ liệu đa phương tiện

Ngoài các kiểu phương tiện và đặc trưng, dữ liệu đa phương tiện có đặc tính khác, trong đó ý nghĩa là chủ quan và mờ, gây phức tạp cho thiết kế MIRS.

2.3.2.6. Dữ liệu đa phương tiện trên mạng

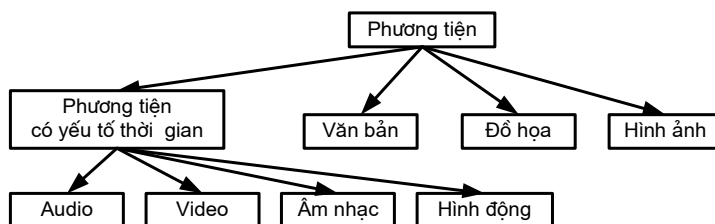
Người ta không phủ nhận được vai trò của mạng máy tính trong việc chia sẻ tài nguyên thông tin chung. Đa số dữ liệu trên mạng Internet là dữ liệu đa phương tiện. Người ta có thể tải và đưa dữ liệu với mạng này.

2.3.2.6.1. Đối tượng đa phương tiện

Liên quan đến các đối tượng đa phương tiện trong quá trình đa phương tiện, người ta đề cập các vấn đề (i) lập trình hướng đối tượng; (ii) phân cấp các đối tượng đa phương tiện; (iii) mô hình và dạng thức thể hiện dữ liệu đa phương tiện; (iv) lớp dữ liệu theo thời gian; (v) một số chuẩn đa phương tiện, siêu phương tiện.

Tiếp cận hướng đối tượng là cần đối với nghiên cứu dữ liệu đa phương tiện. Trong phần đề cập tổ chức dữ liệu đa phương tiện, nhất là xây dựng cơ sở dữ liệu, người ta sử dụng tiếp cận hướng đối tượng.

Từ năm 1995, các đối tượng đa phương tiện được xếp theo các lớp (i) dữ liệu liên tục; (ii) văn bản; (iii) hình ảnh; (iv) đồ họa; (v) cửa sổ; (vi) các ô nhân; (vi) các điều khiển gạt.



Hình. Phân cấp các lớp đa phương tiện

Bảng. Chất lượng dịch vụ truyền thông

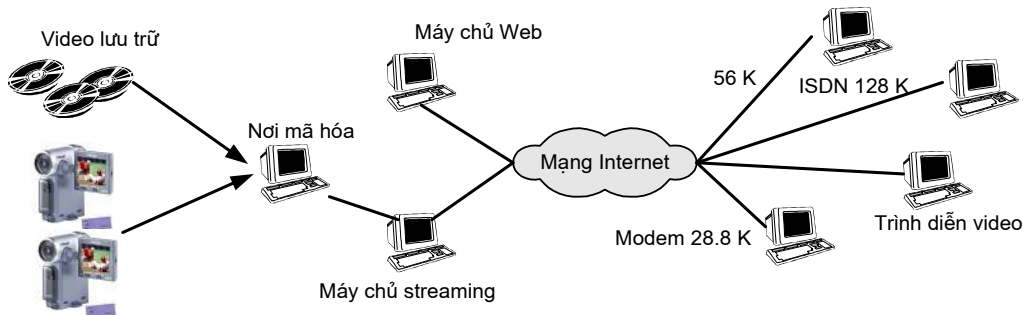
Tham số yêu cầu dịch vụ	Băng truyền (Mbit/s)	Trễ (s)	Tỉ lệ bit sai	Nhân tố bùng nổ (tỉ lệ đỉnh/tỉ lệ trung bình)
Âm thanh không nén, ISDN	0.064	<0.25	10^{-3}	1
Âm thanh nén, VoIP, GSM, UMTS	0.002 - 0.016	<0.25	10^{-4}	2-3
Video không nén	1-5000		10^{-3}	1
Video nén	0.1-500		$<10^{-5}$	2-30
Video theo yêu cầu		<10		3-30
Hội nghị video		<0.25		2-10
Truyền dữ liệu điểm-điểm		10-10000	0	lớn
Truyền dữ liệu Email		1-1000	0	
Truyền dữ liệu WWW		<5	0	

Các lớp phương tiện rời rạc được kể đến là (i) văn bản; (ii) đồ họa ma trận điểm; (iii) đồ họa vectơ. Với giao diện người dùng đồ họa, người ta đề cập các cửa sổ, các ô nhấn, các thực đơn, các trường hội thoại. Lớp các phương tiện có yếu tố thời gian gồm (i) lớp video; (ii) lớp âm thanh; (iii) lớp các dữ liệu hỗn hợp; (iv) thu thập video; (v) ghi âm thanh. Do có nhiều lớp dữ liệu phức tạp, xét trên khía cạnh không gian, thời gian... người ta sử dụng công cụ mô hình để hình thức hóa thế giới thực. Ngôn ngữ sẽ hỗ trợ trong mô hình như vậy.

2.3.2.6.2 Mạng đa phương tiện

Quá trình đa phương tiện sử dụng hạ tầng mạng. Cần có các giao thức phù hợp với dữ liệu đa phương tiện, đặc biệt đối với dữ liệu như video. Những giao thức truyền thông sẽ được làm phù hợp đối với dữ liệu đa phương tiện.

Các thuật toán truyền thông truyền thống không phù hợp đối với dòng liên tục các gói đa phương tiện. Chúng làm rớt gói, dạng khác của trễ. Điều này xảy ra trên (i) các máy MAC trong mạng LAN; (ii) điều khiển truyền lại bị lỗi; (iii) điều khiển dòng dữ liệu theo cửa sổ trượt. Các thuật toán bình thường không đảm bảo chất lượng dịch vụ. Công nghệ streaming cho phép tăng cường đáng kể việc tải và thể hiện các sản phẩm đa phương tiện. Chúng cho phép dữ liệu đến nơi cần được phân phối một cách liên tục, trễ ít. Dữ liệu đa phương tiện đến sẽ được đặt trong vùng nhớ đệm trước khi phân phối. Có máy chủ làm nhiệm vụ này.



Hình. Video streaming

2.4 Mô hình kịch bản

2.4.1. Kịch bản trong IMD

Thuật ngữ *kịch bản* trong ngữ cảnh IMD dùng với nghĩa nội dung hành vi được tích hợp, tức là loại sự kiện mà IMD dùng và loại hành động được kích hoạt như kết quả.

Kịch bản gồm tập các đơn vị chức năng, tức bộ kịch bản, có (i) các sự kiện kích hoạt (bắt đầu và dừng kịch bản), (ii) hành động thể hiện để chuyển tải ngữ cảnh đối với bộ kịch bản, (iii) các sự kiện đồng bộ liên quan; (iv) ràng buộc nếu có.

Chi tiết về bộ kịch bản, có các thuộc tính sau:

- Sự kiện bắt đầu: thể hiện sự kiện kích hoạt khai thác các hành động trong danh sách các hành động hành động_list;
- Sự kiện dừng: thể hiện sự kiện cho biết chấm dứt khai thác bộ này, tức khai thác hành động mô tả trong danh sách hành động trước khi chấm dứt;

- Danh sách hành động: thể hiện danh sách các hành động thể hiện đa phương tiện một cách đồng bộ, khi bộ kịch bản được kích hoạt;
- Sự kiện đồng bộ: tham chiếu đến sự kiện được tạo nên tại lúc bắt đầu và chấm dứt việc khai thác bộ hiện tại, dùng để đồng bộ hoá.

Bộ kịch bản được xác định hình thức như sau:

- Kịch bản:== bộ kịch bản [{, bộ kịch bản}]
- Bộ kịch bản:== sự kiện đầu “, ” sự kiện dừng “, ” danh sách sự kiện “, ” sự kiện đồng bộ
- Sự kiện đầu:== sự kiện
- Sự kiện dừng:== sự kiện
- Danh sách hành động:== tổ hợp
- Sự kiện đồng bộ:== “(“ bắt đầu, chấm dứt “)”
- Bắt đầu:== sự kiện | “”
- Chấm dứt:== sự kiện | “”

2.4.2. Kịch bản đa phương tiện

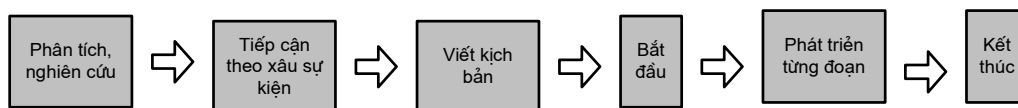
2.4.2.1. Định nghĩa kịch bản

Viết kịch bản là sáng tác. Công tác sáng tác là hoạt động trí tuệ. Một sản phẩm nghệ thuật là tách được cái riêng trong những cái chung.

Định nghĩa: Kịch bản là câu chuyện viễn tưởng, cá nhân với các đặc tính, sự kiện, sản phẩm và môi trường. Nó giúp người thiết kế khai thác ý tưởng và chia quyết định thiết kế ra các tình huống cụ thể.

Một hình là thể hiện đơn, như hình hoạt hình, mang một ý nghĩa trong tương tác. Câu chuyện là dãy các hình đơn, tập trung vào các hành động chính trong một hoàn cảnh. Bằng cách sử dụng kỹ thuật này, người thiết kế có thể chuyển từ tương tác này sang tương tác khác. Người dùng trong hệ thống cần đáp ứng các cảnh “đúng”, tuân theo mẫu thử với nhiều nhân tố.

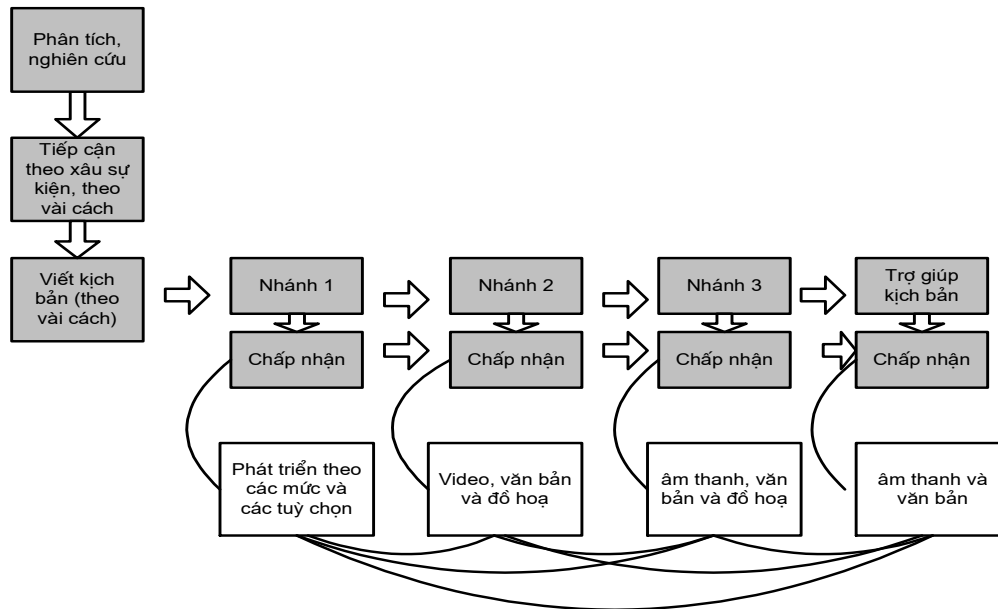
Có nhiều dạng kịch bản sử dụng trong đề án đa phương tiện, từ kịch bản văn học, kịch bản phù hợp với đa phương tiện, kịch bản chi tiết, kịch bản phân cảnh...



Hình. Viết kịch bản đối với phương tiện sử dụng tuần tự, hay tuyến tính

Chủ đề trong đề án hay chủ đề của kịch bản quán triệt một số ý (i) xác định được khái niệm, chủ đề của kịch bản đa phương tiện; (ii) sản phẩm đa phương tiện có câu chuyện xuyên suốt; (iii) chủ đề cho sản phẩm với qui mô lớn cần có thông tin điều tra nhu cầu; (iv) chủ đề tốt có thể dùng nhiều phương tiện, nhiều cách thể hiện với nhiều loại hình trình diễn.

Định nghĩa: Kịch bản¹ là phương tiện liên kết các văn bản, ảnh... theo chủ đề, nhan đề của sản phẩm đa phương tiện.



Hình. Viết kịch bản cho đa phương tiện

Những nét chính của kịch bản gồm:

1. Chi tiết về câu chuyện và cấu trúc câu chuyện;
2. Chi tiết về các sự kiện và những gì xảy ra khi có sự kiện;
3. Khả năng tính tương tác giữa người dùng và hệ thống đa phương tiện;
4. Khả năng thao tác cho phép đối với người dùng;
5. Theo phương châm chia sản phẩm ra nhiều đoạn khác nhau, một câu chuyện chia ra làm nhiều cảnh². Danh sách các cảnh được tổ chức theo bảng, ghi rõ (i) số cảnh; (ii) tên cảnh; (iii) tình huống của cảnh.

2.4.2.2. Kịch bản văn học

Việc viết kịch bản cần thu được bản viết, thành lời văn, có câu chuyện. Câu chuyện này có thể được thể hiện theo văn bản ở dạng (i) văn bản hiển thị; (ii) văn bản trên phím bấm; (iii) trong thân câu chuyện; (iv) dưới dạng âm thanh xen kẽ. Cần lưu ý để kịch bản phải thể hiện đúng ý người đặt hàng.

Sự kiện xảy ra trong kịch bản được chi tiết hoá. Các sự kiện được đặt trong một danh sách, tiến theo dõi, được gọi là danh sách các sự kiện. Danh sách các sự kiện là danh sách tổng hợp các nội dung theo các tình huống, trong đó chưa có giả thiết phục vụ người dùng. Các thông tin về các sự kiện gồm (i) chỉ số của cảnh chứa sự kiện; (ii) tên sự kiện; (iii) đích con của sự kiện; (iv) yếu tố cho phép kích hoạt sự kiện; (v) mô tả sự kiện và kịch tính trong sự kiện.

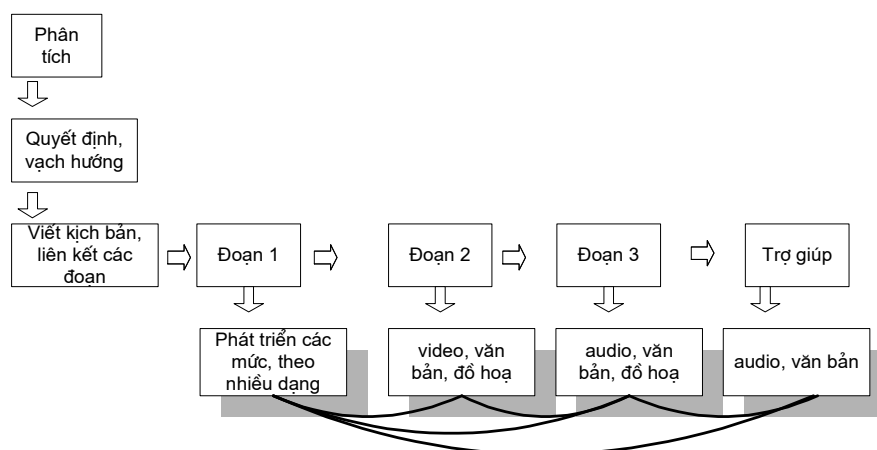
¹ script
² scene

2.4.2.3. Kịch bản phân cảnh

Dựa trên kịch bản văn học, người ta biên soạn kịch bản đa phương tiện, như kịch bản điện ảnh. Kịch bản này phù hợp với quá trình đa phương tiện. Những đoạn “ngày xưa ngày xưa...” trong văn học cần được thể hiện bằng ngôn ngữ khác trong kịch bản đa phương tiện.

Kịch bản đa phương tiện gồm nhiều cảnh. Mỗi cảnh được xây dựng như sản phẩm riêng, do dữ liệu đa phương tiện phức tạp, và mỗi cảnh có bối cảnh dựng sản phẩm có thể khác nhau. Người ta sẽ lắp ghép các cảnh thành sản phẩm chung.

Các cảnh sẽ yêu cầu về (i) thời lượng; (ii) hạ tầng kỹ thuật, thiết bị đa phương tiện; (iii) phần mềm đa phương tiện. Trước khi thực hiện quá trình đa phương tiện, tức sản xuất, người ta sử dụng phần mềm StoryBoard để phác thảo một số hình chính trong cảnh đó. Người lập trình dựa vào ý tưởng trong các hình liên hoàn để có cái nhìn tổng thể về các nhân vật trong sản phẩm đa phương tiện cuối cùng, để tạo ra nhân vật phù hợp.



Hình. Kịch bản đa phương tiện

2.4.2.4. Kịch bản cho đa phương tiện

Xây dựng kịch bản theo từng mức, đảm bảo các ý:

- Quyết định về vật liệu;
- Loại phương tiện và mục đích thể hiện;
- Tương hợp với lược đồ, thể loại, cấu trúc;
- Hiểu nhu cầu người dùng, người lập trình;
- Giữ cái nhìn tổng thể và chi tiết.

2.5. Tìm kiếm tư liệu đa phương tiện tương tác

Liên quan đến việc tìm kiếm tư liệu đa phương tiện tương tác, người ta thấy có hai vấn đề quan tâm đối với IMD là (i) tìm kiếm; (ii) thể hiện, cần xử lý khác với đối tượng đơn phương tiện.

- Đồng bộ hoá và thể hiện: tìm và thể hiện các đối tượng đa phương tiện trong hệ quản trị cơ sở dữ liệu đa phương tiện cần các khía cạnh riêng do khía

cạnh phụ thuộc thời gian của hầu hết loại dữ liệu đa phương tiện. Có nhiều tham số cần đến công nghệ như: mạng truyền thông; bộ nhớ phụ, thuật toán nén...

- Ngôn ngữ hỏi, tìm theo nội dung và chỉ số hoá: Tìm theo nội dung cần kết quả của nghiên cứu và hỗ trợ của công nghiệp. Nghiên cứu đã tập trung vào: chỉ số hoá theo nội dung ảnh, tức là tìm nhanh các đối tượng theo đặc trưng về nội dung (màu, đường viền, hình dạng).

2.5.1. Tìm tư liệu đa phương tiện tương tác dựa trên cấu trúc không gian, thời gian

Đề cập chỉ số hoá dựa trên cấu trúc không gian, thời gian trong IMD. Các câu hỏi liên quan, phụ thuộc vào mối quan hệ không gian, thời gian, có thể được phân loại theo các phạm trù sau:

1. Thuần túy không gian, thời gian. Chỉ mối quan hệ không gian, thời gian được dùng. Chẳng hạn “đối tượng nào hiện đang đề lên thể hiện logo X?”
2. Tính không gian, thời gian. Chỉ định nơi có mối quan hệ không gian, thời gian, chẳng hạn “đối tượng nào thể hiện trong khoảng thời gian thể hiện của X?”
3. Phần hiện, liên quan đến phần hiện không gian và thời gian của ứng dụng. Thí dụ “phần hiện trên màn tại giây 11 của ứng dụng?”

Một lược đồ lưu trữ tuần tự gồm cả tọa độ không gian và thời gian là không hiệu quả, do IMD điển hình có hàng vạn đối tượng. Do vậy, các kỹ thuật chỉ số hoá cho phép nắm được đặc tính không gian, thời gian là cần thiết.

2.5.1.1. Các kỹ thuật chỉ số hoá đối với tư liệu đa phương tiện tương tác lớn

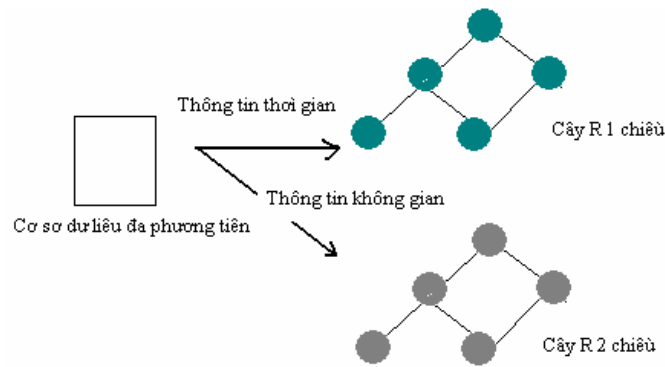
Do IMD cần khối lượng lớn bộ nhớ cho các đối tượng đa phương tiện (âm, hình, video, văn bản), người ta đưa ra câu hỏi theo mối quan hệ không gian, thời gian để kiến thiết hiệu quả IMD. Các khía cạnh không gian, thời gian của các đối tượng được xác định tên bằng 6 tọa độ chiếu trên trục X (x_1, x_2), trục Y (y_1, y_2) và trục t (t_1, t_2). Lược đồ lưu trữ tuần tự quản lý các đặc tính của các đối tượng thông qua tập của các bộ 7 giá trị (tên id, $x_1, x_2, y_1, y_2, t_1, t_2$) và tổ chức chúng trên đĩa.

Một kỹ thuật lưu trữ hiệu quả mà vẫn đơn giản dựa trên 3 ma trận đĩa, chứa tọa độ thấp của các đối tượng, tức x_1, y_1, t_1 tách biệt với trật tự được lưu trữ. Một vài câu hỏi bằng các toán tử không gian, thời gian cần tìm chỉ trên một ma trận, thông qua kỹ thuật chia-đề-trị.

Sau đây là hai kỹ thuật chỉ số hoá và các thủ tục tìm tương ứng.

2.5.1.2. Lược đồ chỉ số hoá theo thời gian và không gian

Lược đồ đơn giản giữ được đặc tính không gian, thời gian gồm hai chỉ số (i) chỉ số không gian (2 chiều) đối với các đặc tính không gian, tức id và x_1, x_2, y_1, y_2 của đối tượng; (ii) chỉ số thời gian đối với đặc tính thời gian, tức id và t_1, t_2 của đối tượng.

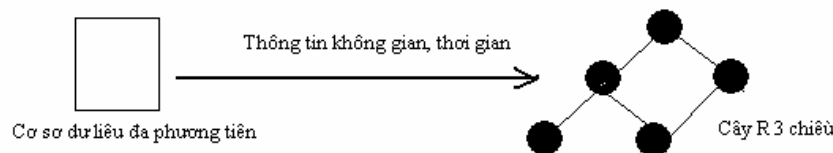


Hình. Lược đồ chỉ số đơn giản (không gian, thời gian)

Người ta nhất trí chấp nhận lược đồ này, tăng cường các phép tìm không gian, thời gian, so với lược đồ ma trận được sắp xếp. Dù có phép toán phức tạp cần truy cập cả hai cây chỉ số, giá trả lời trên cả hai chỉ số vẫn thấp hơn giá trên cả 3 ma trận.

2.5.1.2. Lược đồ chỉ số hoá không gian, thời gian thống nhất

Lược đồ này làm tăng hiệu quả so với lược đồ trước và tăng cường hiệu năng công cụ IMD. Lược đồ này chỉ một chỉ số: chỉ số không gian (3 chiều) dành cho thông tin không gian, thời gian phức hợp. Thông tin được gán tọa độ không gian, thời gian của đối tượng.



Hình. Lược đồ chỉ số thống nhất.

Nếu giả sử cây R là cơ chế chỉ số hoá không gian hiệu quả, thì lược đồ thống nhất được mô tả như hình, với tiến bộ là:

- Cơ chế chỉ số dựa trên khung thống nhất, Chỉ cần một cấu trúc dữ liệu, tức cây R, để cài đặt và bảo trì;
- Các phép toán không gian, thời gian được hỗ trợ nhiều hơn. Nhờ các định nghĩa phù hợp, các phép toán không gian, thời gian được thực hiện như câu hỏi 3 chiều, tìm trên chỉ số 3 chiều, vậy nên không cần phép giao không gian.

2.5.1.3. Các phép toán tìm không gian, thời gian trên cây R

Nét chính của cấu trúc dữ liệu đa chiều đã được thiết kế dựa trên việc chỉ số hoá theo abc cổ điển, trên cây R. Người ta thường chia mặt phẳng thành các miền con phù hợp và lưu trữ các miền con này trong cấu trúc cây phân cấp.

Các đối tượng được thể hiện trong cấu trúc cây theo cách xấp xỉ, sử dụng hình chữ nhật bao tối thiểu MBR¹ để đơn giản và hiệu quả.

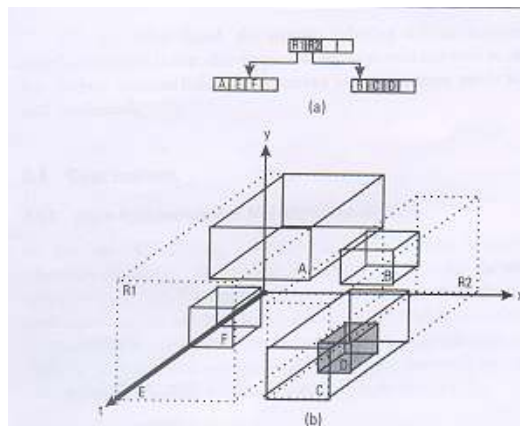
¹ minimum bounding rectangle

Tuy nhiên vị trí tương đối của hai hình liên kế không phủ toàn bộ thông tin về quan hệ không gian (topo, hướng và khoảng cách) giữa các đối tượng thực. Chính vì vậy mà câu hỏi không gian cần được thực hiện theo chiến lược hai bước:

1. *Bước lọc*: cấu trúc cây được dùng để loại bỏ nhanh các đối tượng không có khả năng thoả mãn câu hỏi. Kết quả của bước này là tập các ứng cử chứa các kết quả và có thể có vài lỗi;
2. *Bước làm mịn*: mỗi ứng cử được rà xét, nhờ kĩ thuật hình học tính toán. Lỗi sai sẽ được loại bỏ,

Cây R là cấu trúc dữ liệu đa chiều phân cấp hiệu quả. Một cây cân đối cao, gồm nút, lá (lưu trong bộ nhớ phụ).

- Hình chữ nhật MBR của các đối tượng dữ liệu thực giả thiết được lưu trong lá cây;
- Các nút trung gian được xây dựng bởi nhóm các hình chữ nhật, hay siêu-chữ nhật, là các nút mức thấp hơn. Nút trung gian gắn với hình chữ nhật nào đó, mà bao các hình ứng với nút này;
- Để tìm các đối tượng liên quan đến câu trả lời cho phép toán không gian, thời gian, người ta cần xác định hình chữ nhật MBR bao các đối tượng, rồi tìm các nút trung gian có MBR này. Kĩ thuật này được cài đặt để trợ giúp các phép toán với độ phân giải cao, thông dụng trong GIS.

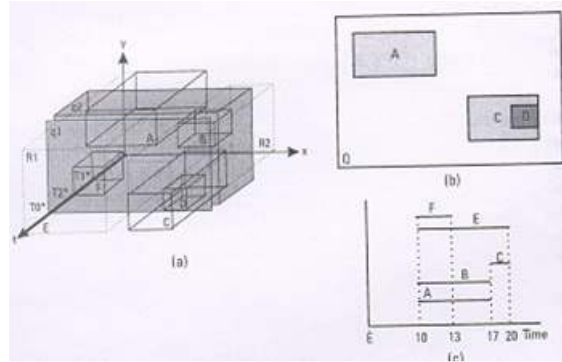


Hình. Dùng cây R 3D với toán tử tìm overlap-during

Trong hình là một số thao tác: (i) thể hiện đối tượng trên cây R; (ii) câu hỏi không gian, thí dụ đè nhau¹.

Một loại hỏi đặc biệt khác đòi hỏi tìm thể hiện không gian và thời gian. thí dụ “tìm các đối tượng và vị trí của chúng trên màn hình tại giây t_0 ?” (thể hiện không gian), hay “tìm đối tượng xuất hiện trong ứng dụng trong đoạn (t_1, t_2) và đoạn thời gian của nó?” (thể hiện thời gian).

¹ overlap



Hình. Tìm thể hiện không gian, thời gian (a) của số câu hỏi thể hiện không gian, thời gian; (b) thể hiện không gian; (c) thể hiện thời gian

2.6. Kết luận

Chương này tập trung vào tính tương tác của cơ sở dữ liệu đa phương tiện. Vấn đề (i) mô hình hóa dữ liệu; (ii) kiến trúc của hệ thống xử lý đa phương tiện; (iii) yếu tố thời gian, không gian của dữ liệu đa phương tiện... và một số loại dữ liệu đa phương tiện được trình bày.

Các vấn đề chương này trình bày gồm:

- Cơ sở dữ liệu đa phương tiện tương tác;
- Mô hình hoá tư liệu đa phương tiện tương tác IMD;
- Một số loại dữ liệu đa phương tiện và nhu cầu về chuẩn dữ liệu;
- Tìm kiếm tư liệu đa phương tiện tương tác

Chương 3. Thành tựu và xu hướng

3.1 Các thành tựu chính của công nghệ hệ quản trị cơ sở dữ liệu đa phương tiện

Nói về các thành tựu chính của công nghệ hệ quản trị cơ sở dữ liệu đa phương tiện, dễ thấy chúng được liệt kê tập trung vào (i) mô hình hoá; (ii) tìm kiếm IMD.

3.1.1. Mô hình hoá

- Theo Zdonik, người ta đã xác định các vai trò đa dạng cho phép cơ sở dữ liệu có thể đóng góp cho hệ thống đa phương tiện, trong đó có việc tích hợp logic của dữ liệu trên nhiều phương tiện;
- Theo Kim, người ta dùng cơ sở dữ liệu hướng đối tượng và mở rộng để dùng cho ứng dụng đa phương tiện;
- Little và Ghafoor đã phát triển phương pháp cho phép thể hiện ràng buộc theo thời gian trong hệ thống đa phương tiện.

3.1.2. Toàn vẹn

- Có các nghiên cứu về kiểm chứng tư liệu đa phương tiện và tính toàn vẹn. Người ta đề xuất *mô hình đồng bộ* để mô tả hình thức tư liệu đa phương tiện, hay tạo ra lịch thể hiện bền vững thông qua việc tạo ôtomat.
- Người ta còn kiểm chứng tư liệu đa phương tiện, nhằm định tên sự mất bền vững tiềm năng theo thời gian.
- Một tiếp cận khác nhằm thể hiện mềm dẻo đa phương tiện với sự tham gia của người dùng và đề xuất mô hình đồng bộ, cho phép xác định tham gia của người dùng trong khi thể hiện.

3.1.3. Tìm theo nội dung

Tìm kiếm đa phương tiện thách thức công tác nghiên cứu và công nghiệp. Người ta tập trung vào điểm quan trọng là tìm ảnh và video theo nội dung.

3.1.3.1. Tìm ảnh

Các phạm trù câu hỏi có thể yêu cầu một hay nhiều khía cạnh:

1. Câu hỏi về khía cạnh hiển thị đơn; người dùng xác định giá trị nào đó với phần trăm của khía cạnh, như tìm ảnh 70% xanh...;
2. Câu hỏi tổ hợp các khía cạnh;
3. Câu về khía cạnh cục bộ. Người dùng xác định giá trị và vị trí trên nền, chẳng hạn tìm ảnh mà nền trời xanh ở phía nửa trên...;
4. Câu hỏi theo thí dụ;
5. Đối tượng so với hình ảnh. Người dùng có thể mô tả khía cạnh của đối tượng trong hình, đối lập với việc mô tả toàn bộ ảnh. Chẳng hạn “tìm hình có oto xanh gần giữa ảnh?”;

6. Câu hỏi về thuộc tính do người dùng xác định. Người dùng xác định giá trị của thuộc tính người ta xác định, thí dụ “tìm ảnh trong đó vị trí là Hanoi, ngày 19/5, có độ phân giải ít hơn 300 dpi?”;
7. Câu hỏi về mối quan hệ đối tượng. Người dùng xác định các đối tượng, thuộc tính và mối quan hệ giữa chúng. Thí dụ “tìm ảnh mà có người già cầm tay đứa trẻ?”;
8. Câu hỏi về khái niệm. Một vài hệ thống cho phép người dùng xác định khái niệm đơn giản dựa trên khía cạnh rút ra từ hệ thống. Thí dụ người dùng định nghĩa vùng núi là tam giác và có hình tròn mặt trời, và “tìm ảnh có vùng núi?”.

Nội dung hiển thị có thể được mô hình hoá theo phân cấp các trừu tượng. (i) Tại mức đầu có các điểm ảnh thô, với thông tin về màu và độ sáng. (ii) Tiếp theo việc xử lý thu được các khía cạnh về góc, đường, vùng màu... (iii) mức trừu tượng cao hơn có thể tổ hợp và diễn giải các khía cạnh như các đối tượng và các thuộc tính. (iv) Tại mức cao nhất là các khái niệm mức con người, cần một hay nhiều đối tượng và mối quan hệ giữa chúng. Các phương pháp nửa tự động hay các từ khoá do người dùng cung cấp cũng được dùng.

Còn một vài khía cạnh về hình ảnh:

- Màu sắc cũng giữ vai trò có ý nghĩa trong việc tìm ảnh. Màu khác nhau được thể hiện theo lược đồ RGB¹, hệ thống sáng và phản quang theo CIE², độ bão hoà HSI³;
- Mẫu, bề mặt⁴ là phần hiển thị mà một số lớn các phần tử hiển thị được cô đọng và được xếp. Mẫu là miền có độ nhẵn thống nhất, theo hình đơn giản và được lặp lại;
- Hình dạng⁵ là vấn đề khó trong việc tìm ảnh tổng quát, do khó phân loại các đoạn đối tượng cần quan tâm trong ảnh.

3.1.3.2. Tìm video

Về bản chất, video có các đơn vị phân cấp với các khung riêng tại mức cơ sở và mức cao hơn, như lia cảnh, cảnh, đoạn phim. Nhiệm vụ quan trọng khi phân tích nội dung video là phát hiện biên của đoạn video.

- Một *lia cảnh*⁶ là tập hợp được ghi tuần tự các khung, thể hiện hành động liên tục trong thời gian và khoảng không của máy quay đơn; Có một vài kỹ thuật cho phép phát hiện thay đổi lia cảnh;
- Một dãy các lia tập trung vào một điểm hay cùng vị trí quan tâm được gọi là *cảnh*⁷;
- Tập hợp các cảnh liên quan tạo nên đoạn phim⁸.

Điều quan trọng là: phát hiện và theo vết đối tượng. Trong video, hai nguồn thông tin có thể được dùng để phát hiện và theo vết các đối tượng: (i) khía cạnh hiển thị, như màu, mẫu, và (ii) thông tin chuyển động.

¹ red green blue

² international commission on illumination

³ hue saturation intensity

⁴ texture

⁵ shape

⁶ shot

⁷ scene

⁸ episode

Một chiến lược diễn hình là chia miền khởi đầu dựa trên thông tin về màu sắc và mẫu. Sau đó các miền với các vectơ chuyển động tương tự có thể được hợp nhất lại theo điều kiện nào đó. Hệ thống dùng để phát hiện các chuyển động cụ thể như vào, ra cảnh, thay thế chuyển động... sẽ được phát triển.

Một khi các khía cạnh được phát hiện, các kỹ thuật chỉ số hoá và tìm kiếm cần hỗ trợ cho câu hỏi. Bản chất thời gian và kích thước sở lớn tương đối của dữ liệu video cần đến chức năng rà soát và hỏi chuyên dụng. Tiếp cận thông dụng cho phép duyệt nhanh là phát hiện thay đổi lia cảnh và liên kết một biểu tượng nhỏ của khung chính đối với mỗi lia cảnh. Việc tìm kiếm các biểu tượng... dùng cấu trúc phân cấp video.

3.1.4. Trích đặc trưng, chỉ số hóa và đo tương tự

Đặc trưng và thuộc tính của mục thông tin trong MIRS được trích, tham số hóa, lưu trữ với bản thân mục tin. Các đặc trưng và thuộc tính của câu hỏi cũng được trích theo cách tương tự nếu chúng không hiển hiện. Hệ thống tìm trong cơ sở dữ liệu các mục với đặc trưng và thuộc tính tương tự theo độ đo tương tự nào đó. Để xúc tiến quá trình tìm kiếm, các đặc trưng và thuộc tính được tổ chức thành cấu trúc chỉ số.

3.1.4.1. Trích đặc trưng

Các mục thông tin đa phương tiện trong cơ sở dữ liệu thường được tiền xử lý để trích đặc trưng và thuộc tính. Trong quá trình tìm kiếm, người ta tìm và so sánh theo đặc trưng và thuộc tính này, thay vì theo bản thân thông tin. Do vậy, chất lượng của việc trích đặc trưng sẽ quyết định hiệu quả tìm kiếm. Nếu không trích được đặc trưng thì không đáp ứng được các câu hỏi trên các đặc trưng đó. Đây là khác nhau chính giữa DBMS và MIRS (i) trong DBMS có tất cả các thuộc tính và đầy đủ; (ii) trong MIRS cần trích đặc trưng và thuộc tính theo các kiểu câu hỏi sẽ đến, và không đầy đủ. Việc trích đặc trưng cần đáp ứng yêu cầu sau:

1. Các đặc trưng và thuộc tính trích ra cần đầy đủ để thể hiện nội dung của mục thông tin;
2. Các đặc trưng cần được thể hiện và lưu trữ cô đặc. Các đặc trưng phức tạp và lớn không phù hợp với việc trích đặc trưng;
3. Việc tính khoảng cách giữa các đặc trưng cần hiệu quả, nếu không sẽ làm trễ trả lời.

Nhìn chung có bốn mức đặc trưng và thuộc tính (i) dữ liệu meta; (ii) ghi chú văn bản; (iii) đặc trưng nội dung mức thấp; (iv) đặc trưng nội dung mức cao.

1. Dữ liệu meta chứa thuộc tính sự kiện và hình thức về các dữ liệu đa phương tiện như tên tác giả, ngày tạo, nhan đề đối tượng. Lưu ý rằng dữ liệu meta không mô tả hay diễn giải nội dung đối tượng. Các thuộc tính do các kỹ thuật DBMS quản lý.
2. Ghi chú văn bản là mô tả bằng văn bản về nội dung đối tượng. Ghi chú có thể ở dạng một số các từ khóa hay mô tả bằng văn bản tự do. Việc chỉ số hóa và tìm kiếm dựa trên ghi chú có thể được kỹ thuật IR thực hiện. Dù ghi chú sẽ hạn chế về chủ quan và chưa đầy đủ, nó vẫn là phương pháp thông dụng và hiệu lực. Người ta sử dụng ghi chú văn bản với đặc trưng khác trong các ứng dụng đa phương tiện. Các hạn chế có thể khắc phục nhờ thông tin phản hồi sau câu hỏi. Hiện tại, ghi chú văn bản là quá trình thủ công. Các công cụ nửa tự động cần được phát triển để giúp quá

trình này. Tri thức lĩnh vực và từ điển là có ích để đạt hiệu năng IR cao.

3. Đặc trưng nội dung mức thấp nắm các mẫu dữ liệu và các thông kê về đối tượng đa phương tiện, và quan hệ không gian, thời gian giữa chúng. Phương tiện khác nhau có đặc trưng nội dung mức thấp khác nhau. Với âm thanh, đặc trưng mức thấp là ồn trung bình, phân bố tần số, tỉ lệ im lặng. Với dữ liệu ảnh, đặc trưng mức thấp gồm phân bố màu, bề mặt, dạng đối tượng và cấu trúc không gian. Với dữ liệu video, đặc trưng mức thấp gồm cấu trúc thời gian và các đặc trưng như đối với ảnh. Ưu điểm chính của đặc trưng mức thấp là cho phép trích tự động.
4. Trích đặc trưng mức cao nhằm nhận dạng và hiểu đối tượng. Trừ việc nhận dạng văn bản và tiếng nói, nhìn chung khó nhận dạng và hiểu đoạn âm thanh hay các đối tượng video. Tuy nhiên người ta cũng có ý tưởng giải quyết. Trong ứng dụng đặc biệt với số đối tượng hạn chế, việc mô tả và nhận dạng đối tượng chung là có ích và có thể được. Chẳng hạn trên 95% video có chủ đề con người, nhóm người. Vậy hệ thống nhận ra và diễn giải con người là có ý nghĩa. Ban đầu công việc này là bán tự động.



Hình. Đặc trưng, phong cách

Định nghĩa: Tìm theo nội dung là việc tìm kiếm theo các đặc trưng mức thấp và đặc trưng mức cao

Hệ thống sẽ sử dụng tất cả bốn đặc trưng để câu hỏi linh hoạt của người dùng chấp nhận được. Các đặc trưng sẽ bổ trợ lẫn nhau và hoàn thiện mô tả về đối tượng. Chẳng hạn ghi chú văn bản là tốt khi nắm bắt các khái niệm trừu tượng như cảm xúc, hạnh phúc, buồn, nhưng không thể mô tả mẫu dữ liệu phức tạp như hình dạng phi chuẩn hay bề mặt. Mặt khác, đặc trưng nội dung mức thấp có thể nắm được các mẫu dữ liệu này, nhưng không diễn tả được khái niệm trừu tượng. Khi đa phương tiện dùng nhiều kiểu phương tiện, mối quan hệ và tương tác giữa các phương tiện được dùng để trích đặc trưng, diễn giải và tìm kiếm. Một vài kiểu phương tiện là dễ hiểu và diễn giải hơn phương tiện khác, người ta cũng có thể dùng hiểu biết này để trích đặc trưng của phương tiện khác. Chẳng hạn nếu có dòng video và rãnh tiếng nói, người ta nhận dạng tiếng nói để có tri thức về đối tượng và dùng nó để phân đoạn và trích đặc trưng và các đối tượng từ dòng video.

Lưu ý MPEG-7 sẽ chuẩn hóa các mô tả về đặc trưng, nhưng sẽ không mô tả cách rút đặc trưng đó.

3.1.4.2. Cấu trúc chỉ số

Sau khi trích đặc trưng, cần sử dụng các cấu trúc chỉ số để tổ chức các đặc trưng

này để việc tìm kiếm hiệu quả. Nhiều đặc trưng cần thiết cho việc thể hiện một đối tượng và mỗi đặc trưng cần thể hiện quan nhiều tham số. Chẳng hạn phân bố màu thường được thể hiện qua lược đồ với nhiều màu. Các cấu trúc chỉ số dùng trong DBMS là không phù hợp cho các đặc trưng theo nội dung. Việc chỉ số hóa trong MIRS sẽ là phân cấp và có nhiều mức: (i) mức cao nhất là phân loại ứng dụng; (ii) mức chỉ số thứ hai khác với mức đặc trưng; chỉ số hóa khác nhau cần cho nhiều đặc trưng; (iii) mức chỉ số hóa thứ ba theo mỗi quan hệ thời gian và không gian giữa các đối tượng.

3.1.4.3. Đo sự tương tự

Việc tìm kiếm đa phương tiện dựa trên tương tự thay vì khớp giữa câu hỏi và mục cơ sở dữ liệu. Tương tự được tính theo thuộc tính, đặc trưng trích ra và được thể hiện ở dạng một hay nhiều giá trị. Tất nhiên kết quả lộ ra cần được con người giám định. Do vậy yêu cầu chính của độ đo tương tự là các giá trị tương tự tính được cần hợp với kiểm chứng của người dùng. Các kiểu đặc trưng dùng để mô tả các đối tượng cũng có vai trò trong việc thể hiện kết quả này. Độ đo tương tự phức tạp do kiểm định của con người là chủ quan và tùy thuộc vào ngữ cảnh. Tất cả những nhân tố này khiến việc đánh giá phức tạp nhưng cần thiết.

3.2 Các sản phẩm thương mại và mẫu nghiên cứu

3.2.1. Một số sản phẩm

Để thể hiện sản phẩm thương mại và mẫu nghiên cứu về đa phương tiện, người ta trình bày một vài hệ thống nghiên cứu và thương mại đảm bảo tìm kiếm và chỉ số hoá dựa trên khía cạnh hiển thị như màu và mẫu.

3.2.1.1. Các hệ thống nghiên cứu

- Hệ thống Photobook;
- VisualSEEK;
- Hệ thống VideoQ.

3.2.1.2. Các hệ thống thương mại

- DB2 của IBM;
- ORACLE;
- INFORMIX;
- QBIC.

3.2.1.3. Các hệ thống cho WWW

Hệ thống mạng toàn cầu với WWW là thông dụng. Người ta có thể trao đổi, thảo luận mở về chủ đề này.

3.2.2. Quản lý đa phương tiện

3.2.2.1. Một vài loại hình sản xuất đa phương tiện

Việc sản xuất ra sản phẩm đa phương tiện qua nhiều pha, hay nhiều bước. Người ta thực hiện quá trình sản xuất theo dạng đề án công nghệ thông tin, hay được gọi là đề án đa phương tiện. Bảng nêu bên cho thấy một số loại cán bộ, giữ các vai trò khác nhau trong đề án và một số pha sản xuất.

Bảng. Các loại hình sản xuất đa phương tiện

	Websize trên mạng truyền thông	Sản phẩm video	Tính toán	Xuất bản	Giáo dục, đào tạo có tương tác
Vai trò	Quản trị Web Soạn thảo Web Thiết kế Web Lập trình Web	Người sản xuất Giám đốc Người viết kịch Nghệ sĩ đồ hoạ	Quản lý đề án Phân tích Lập trình Nghệ sĩ đồ hoạ	Tác giả Người biên tập Ban biên tập	Phân tích đào tạo Thiết kế tương tác
Pha sản xuất	Phân tích Thiết kế Thể hiện Sản xuất	Kịch bản nghiên cứu Phân phối sản phẩm Quay	Phân tích Thử nghiệm Lập trình	Thể hiện Phát hành Sản xuất	Phân tích Thiết kế Thử nghiệm
Tư liệu (Khách hàng kết quả)	Đề xuất Kịch bản Thể hiện	Story Board Câu chuyện Thể hiện	Đặc tả chức năng Đặc tả kỹ thuật, thủ tục quản lý và thay đổi	Bản nháp Các đề xuất	Thiết kế tổng thể Thiết kế chi tiết Những qui định

Đề án đa phương tiện được thông tin qua bảng cần chú trọng vào (i) tính đa dạng của đề án; (ii) tính đa dạng của khách hàng; và (iii) tổng quan về vòng đời của đề án đa phương tiện tập trung vào lợi ích của khách hàng.

3.2.2.2. Xác định phạm vi

Xác định phạm vi của đề án đa phương tiện tạo điều kiện xác định đúng yêu cầu cần thực hiện. Do công việc đa phương tiện có nhiều nhiệm vụ và để phát sinh nhiệm vụ mới theo yêu cầu của người đặt hàng, đề án đa phương tiện cần liệt kê đầy đủ các nhiệm vụ và đích cần đạt được.

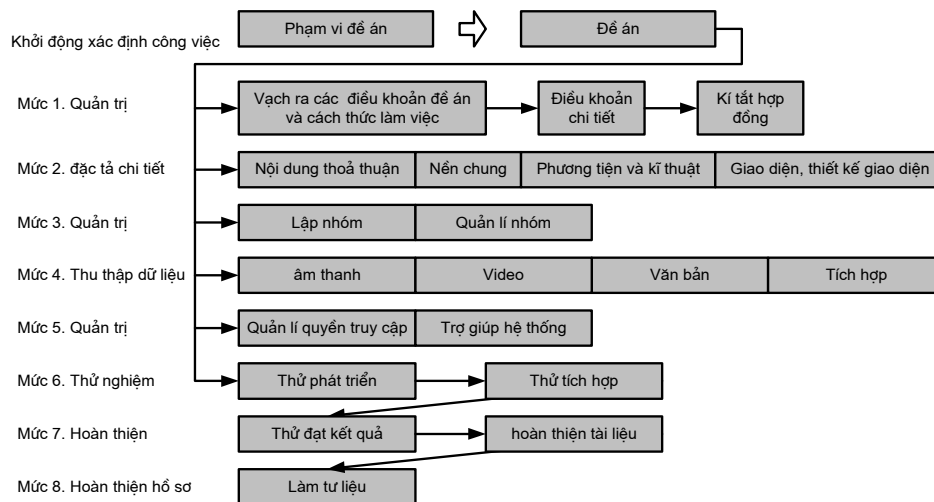
Việc xác định đích không hẳn là dễ dàng, do quá trình đa phương tiện đa số là quá trình sáng tạo; người ta chưa xác định rõ ràng đích.

3.2.2.3. Bước khởi động

Triển khai đề án đa phương tiện cũng như đề án công nghệ thông tin khác: có (i) nhóm công tác; (ii) trưởng nhóm; (iii) điều kiện thiết bị; phần mềm... (iv) các ràng buộc đề án...

Mặt khác, để có sản phẩm như sản phẩm công nghệ thông tin, người ta cần thực hiện bước phân tích, tìm hiểu thực tế, nhằm xác định nhu cầu người dùng cũng như chỉnh lí các ràng buộc đề án.

Việc tiếp xúc cơ sở thực tế, lấy thông tin liên quan đến đề án và chỉnh lí nhiệm vụ... sẽ được thực hiện theo qui trình. Qui trình này đã trở thành chuẩn.



Hình. Phân loại mức sản xuất trong đề án đa phương tiện

3.2.2.4. Quản trị một đề án đa phương tiện

Quản trị đề án đa phương tiện trong tài liệu này không chỉ nhằm vào công tác quản trị đề án công nghệ thông tin, liên quan đến quản trị tài nguyên đề án; quản trị kinh phí; quản trị rủi ro; quản trị về thời hạn, nhân lực; quản trị về chương trình và kiểm tra sai sót chương trình...

Việc quản trị đề án đa phương tiện thực hiện theo các bước, bắt đầu từ phân tích nhu cầu đề án, kết thúc bằng sản phẩm trên đĩa CD ROM, cho phép phân phối, trình chiếu.

Trong phần tiếp theo, từng bước được trình bày. Tuy nội dung không quá phức tạp, nhưng thể hiện các yêu cầu của các bước trong đề án cụ thể không dễ dàng. Các đề án đa phương tiện thường phát sinh kinh phí và không hoàn thành đúng thời hạn yêu cầu.

3.2.2.5. Mô tả tên sản phẩm

Khi chuẩn bị sản xuất, chọn tên sản phẩm đa phương tiện là vấn đề cần đầu tư, suy nghĩ. Một sản phẩm có thể mang nhiều tên, tùy theo đối tượng người ta cần giới thiệu sản phẩm. Thông thường một sản phẩm đa phương tiện có các tên (i) Tên sản phẩm; (ii) Tên mang thông tin; (iii) Tên giải trí; (iv) Tên sáng tác; và (v) Tên giáo dục.

3.2.2.6. Khán giả của sản phẩm

Không như các sản phẩm công nghệ khác, sản phẩm đa phương tiện không thể sản xuất ra mà không có người tiếp nhận. Khâu xác định số lượng người dùng đối với sản phẩm đa phương tiện quyết định số lượng sản phẩm cần sản xuất. Vậy nên việc xác định khán giả quyết định thành công và gợi ý sáng tác cho sản phẩm sẽ được phát triển.

Khán giả là một khía cạnh quan trọng cần chú ý vì khán giả là đối tượng mà đa phương tiện nhằm vào. Phải luôn luôn chú ý vì yếu tố này là động, thay đổi. Khi sử dụng các dữ liệu điều tra nên dùng dữ liệu hiện tại, không nên dựa trên dữ liệu điều tra trong quá khứ. Thí dụ một chiếc máy không bán được ngày nay có thể sẽ bán được ngày hôm sau, nhưng không thể bán tờ báo hôm nay cho ông khách ngày mai.

3.2.2.7. Lưu ý

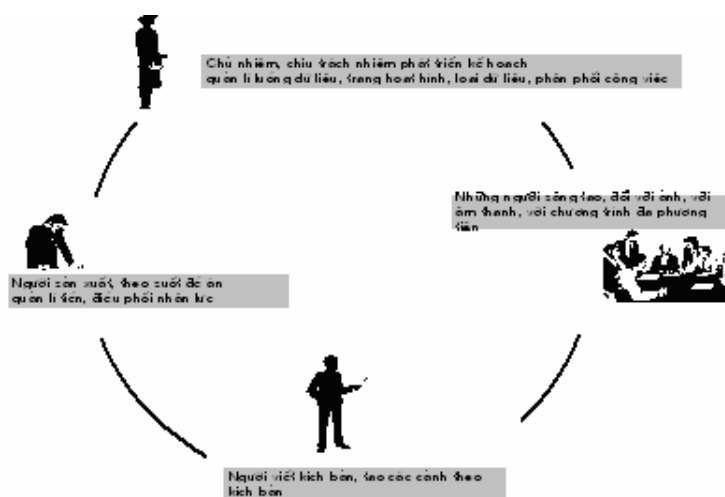
- Cần thiết nêu lên các đề xuất và giải pháp thực hiện đề án đa phương tiện.

- Những điều khoản trong đề xuất sẽ được chi tiết hóa đối với từng loại dữ liệu đa phương tiện, trong từng công việc cụ thể của quá trình thực hiện đề án đa phương tiện.

Trách nhiệm của người quản lý đề án, tức lãnh đạo đề án, và trách nhiệm của khách hàng cần được liệt kê. Trách nhiệm của quản lý đề án là:

- Làm việc khách hàng, thực hiện điều khoản đề án;
- Lên lịch chi tiết;
- Giám sát và dành thời gian cho đề án;
- Thông tin cho khách hàng về quá trình đề án, hoạt động, đề xuất thay đổi, nhân tố...
- Đảm bảo kỹ thuật;
- Tuân theo thiết kế kỹ thuật;
- Đảm bảo nội dung đã ký với khách hàng;
- Theo thời hạn đã xác định với khách hàng, đặc biệt thời hạn cuối;
- Hoàn thành đề án đa phương tiện.
- Trách nhiệm của khách hàng là:
- Chuẩn bị mô tả ngắn gọn với người phát triển;
- Làm việc với các đặc tả;
- Làm việc với các chuyên gia để đảm bảo thời hạn theo lịch;
- Thoả thuận về thay đổi;
- Giúp người phát triển truy cập tài nguyên.

3.2.3. Các vai trò trong dự án đa phương tiện



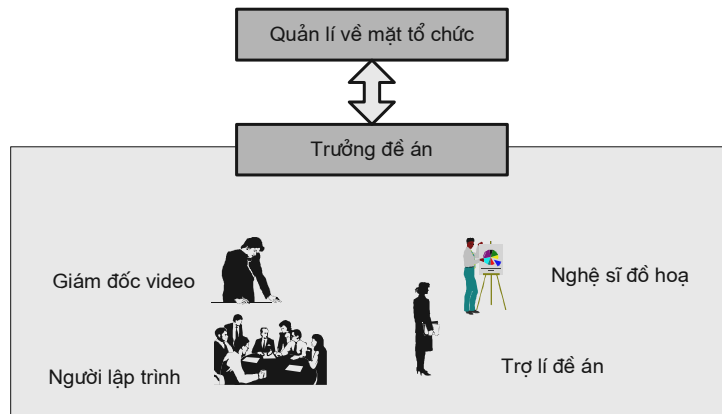
Hình. Các vai trò trong sản xuất đa phương tiện

Khâu tổ chức thực hiện đề án công nghệ thông tin cũng quan trọng như hạ tầng sản

xuất, nhân lực đề án. Người ta có thể đặt vấn đề về (i) quản lý theo thủ công, kinh nghiệm; (ii) quản lý theo qui trình công nghiệp. Dẫu sao vẫn phải xác định (i) Xuất xứ của đa phương tiện; (ii) Quản trị đề án đa phương tiện thích hợp với những nơi đâu ?

Việc so sánh quản trị đề án đa phương tiện đang được thực hiện với các quá trình quản trị các đề án khác cũng cần thiết, để thấy được:

- Các hạn chế của các phương pháp quản trị đề án;
- Chất lượng của phương tiện tương tác;
- Giá sản xuất, thời gian, chất lượng của sản phẩm đa phương tiện.



Hình. Các vai trò trong đề án đa phương tiện

Khi tiến hành đề tài đa phương tiện, người ta thường cho rằng có thể dễ dàng thay đổi thiết kế ban đầu. Điều này không đúng với quản trị đề án đa phương tiện. Lí do không nên thay đổi các thiết kế ban đầu của đề án đa phương tiện dựa trên các hiểu biết:

- Thay đổi đồng nghĩa với việc tăng giá quản trị và sản xuất;
- Thay đổi làm tăng thời gian làm lại;
- Thay đổi đôi khi thành công nếu điều khiển được; ngược lại việc thay đổi thường làm chậm quá trình;
- Nếu việc thay đổi mà không gây kéo dài thời gian thì chi phí, giá cả sẽ tăng;
- Chất lượng sản phẩm có thể thay đổi để phù hợp với ràng buộc thời gian;
- Một số quan niệm cho rằng “chỉ thay chút ít về thiết kế” không gây ảnh hưởng; điều này đôi khi dẫn đến các thay đổi khác, và khái niệm “không lớn” là khái niệm mờ.

3.3. Hướng phát triển của cơ sở dữ liệu đa phương tiện

3.3.1. Một số hướng hiện tại và khuynh hướng

Người ta quan tâm nhiều đến hệ thống đa phương tiện, từ khoa học máy tính và cả mạng máy tính, cơ sở dữ liệu, tính toán phân tán, nén dữ liệu, xử lý văn bản, giao diện người dùng, đồ hoạ máy tính, nhận dạng, và trí tuệ nhân tạo.

3.3.1.1. Mô hình hoá, tính toán vẹn

Liên quan đến mô hình hóa dữ liệu và tính toán vẹn của dữ liệu đa phương tiện, các

phần trên đã đề cập. Có thể truy cập tài liệu gốc về cơ sở dữ liệu đa phương tiện và nguồn thông tin trên Internet để trao đổi, thảo luận về vấn đề này.

3.3.1.2. Tìm theo nội dung

Cơ sở dữ liệu đa phương tiện quan tâm đến khía cạnh mới so với cơ sở dữ liệu truyền thống:

1. *Quản lý tính không chắc chắn trong câu hỏi theo hướng phương tiện và/ hoặc thay đổi thời gian trong dữ liệu.* Thay đổi này cần được tích hợp vào ngôn ngữ hỏi, do chúng có nhiều ứng dụng;
2. *Quản lý tổ hợp logic của các câu hỏi nguyên tố.* Trước tiên người ta cho ngữ nghĩa có lý, cần đến chức năng gộp, rồi tăng hiệu quả của thuật toán qua việc nối các câu hỏi nguyên tố, là cách tối ưu;
3. *Vai trò của cấu trúc và mối quan hệ không gian, thời gian.* cấu trúc không gian, thời gian là quan trọng.

3.3.1.3. Khía cạnh chất lượng dịch vụ đối với tìm trên WWW

Do tiến nhanh của WWW, nội dung WWW yêu cầu nội dung đa phương tiện, cần đến hệ quản trị cơ sở dữ liệu đa phương tiện. Một vài hướng nghiên cứu:

- *Thực hiện (mịn) IMD trên WWW.* Thể hiện của IMD phức tạp ngầm quản lý tương tự bên trong, bên ngoài và đảm bảo đặc tả thể hiện không gian, thời gian
- *Chất lượng dịch vụ (QoS).* Cần có đề xuất đảm bảo chất lượng. Do trong hệ thống phân tán không thể thực hiện điều khiển tập trung về chất lượng, cần có phương pháp phù hợp.

3.3.2. An toàn dữ liệu đa phương tiện

3.3.2.1. Khía cạnh an toàn

Định nghĩa: An toàn thông tin có nghĩa bảo vệ thông tin và hệ thống thông tin khỏi các truy cập, sử dụng, phá hoại, ngăn cản trái phép.

3.3.2.1.1. Nhu cầu về an toàn dữ liệu

Cần thiết nghiên cứu về an toàn đa phương tiện (i) bảo vệ dữ liệu trước truy cập trái phép, cho phép truy cập tin cậy; (ii) giữ toàn vẹn dữ liệu đa phương tiện; (iii) đảm bảo dữ liệu thực, là bản gốc. Có nhiều thuật toán mã hóa dữ liệu đa phương tiện.

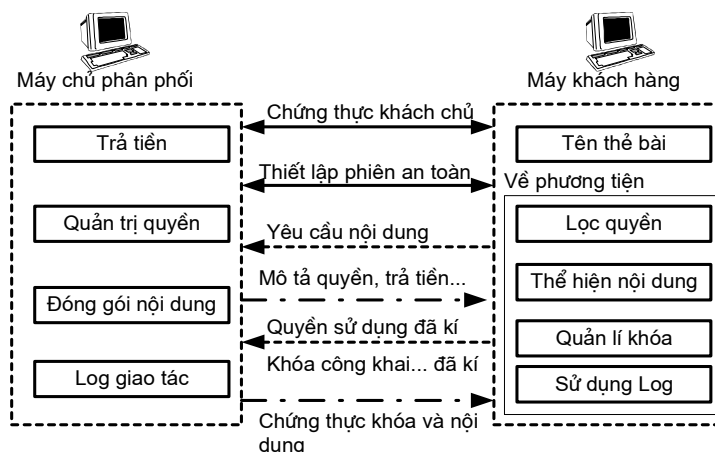
3.3.2.1.2. Xử lý nội dung đa phương tiện

Nội dung đa phương tiện có thể bị xử lý, qua các hành động (i) thay đổi; (ii) bị xử lý; (iii) tái tạo; (iv) lấy đi khỏi. Nhìn nhận về các tấn công trên máy tính hay trên mạng, có (i) kẻ tấn công: gián điệp, tin tặc, tội phạm chuyên nghiệp; (ii) các công cụ kẻ tấn công sử dụng: câu lệnh người dùng, đoạn chương trình, tác tử tự trị, các bộ công cụ...; (iii) tác động của truy cập trái phép: làm hỏng thiết kế, làm hỏng cấu hình, truy cập trái phép, tác động đến quá trình, làm hỏng dữ liệu; (iv) tác hại: làm mất thông tin, ăn cắp dịch vụ, làm mất khả năng sẵn sàng phục vụ; (v) giải pháp: chính sách an toàn, các thách thức.

3.3.2.1.3. Vi phạm an toàn liên quan đến quyền tác giả

Vi phạm an toàn liên quan đến quyền tác giả. Người ta đưa ra các qui định về bản

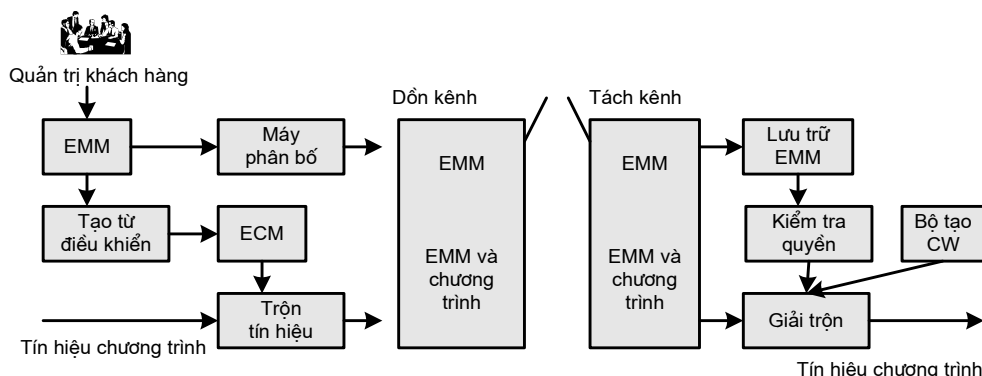
quyền, xác định khía cạnh tội phạm đối với đa số phần mềm thương mại. Không được bẻ khóa phần mềm, tác động đến thiết bị mã hóa. Thực chất công việc này nhằm đảm bảo môi trường thông tin chung, nhất là trên Internet.



Hình. Yêu cầu về an toàn cho hệ thống DRM

Người ta có khái niệm quản trị quyền số hóa DRM¹, yêu cầu mô tả, xác định, đảm bảo, giám sát các dạng sử dụng có quyền, đối với cả truy cập hiển thị hay ngấm định, trên dạng dữ liệu số hóa hay dạng vật lí. DRM sinh ra nhằm bảo vệ quyền tác giả trên Internet hay phương tiện số qua kênh phân phối hợp pháp hay không. Các dữ liệu cần được mã hóa.

Bảo vệ quyền tác giả trong hệ thống truyền thông công cộng có ý nghĩa hơn là sản phẩm đa phương tiện cá nhân. Người ta dùng (i) TV trả tiền; (ii) bảo vệ đĩa DVD; (iii) có chính sách về an toàn thông tin đa phương tiện.



Hình. Hệ thống TV trả tiền

3.3.2.2. Mật mã

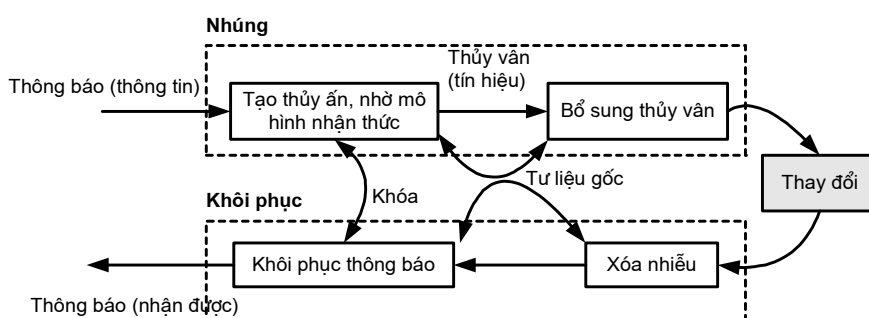
Mật mã là khoa học phân tích, tạo mã và giải mã. Công việc mật mã là viết bằng mã hiệu (i) thể hiện lại dưới dạng mã; (ii) chuyển đổi mã; (iii) ghi lại dưới dạng khác; (iv) thực hiện quá trình ngược, tức giải mã.

¹ Digital Rights Management

3.3.2.3. Thủy vân

Định nghĩa: Thủy vân số là quá trình nhúng thông tin vào tín hiệu số theo cách khó tháo gỡ ra.

Người ta sử dụng mã hiệu, kí tự đặc biệt để không cho phép người bình thường thấy, trừ phi người dùng có khóa mở. Theo cách này, một nhánh của mật mã là giấu trong thông báo một lượng tin. Lượng tin giấu này đảm bảo quyền của tác giả dữ liệu đa phương tiện, tức góp phần bảo vệ quyền tác giả. Mặt khác, việc giấu tin cũng cho phép tải lượng thông tin bí mật qua thông tin hiển hiện. Việc giấu tin cho phép (i) giấu thông tin bằng cách mã hóa; (ii) nhúng thông tin mật trong thông tin, tức thủy vân. Thủy vân có thể là (i) không bị phát hiện; (ii) phần nào tác động để toàn vẹn dữ liệu; (iii) bền vững trước các tấn công nhằm tách thông tin mật.



Hình. Mô hình truyền thông thủy vân, 2002

Thủy vân số nhằm (i) chứng thực quyền tác giả; (ii) để lại dấu ấn; (iii) đảm bảo toàn vẹn dữ liệu; (iv) điều khiển bản sao; (v) có một số thông số: bền vững, phức tạp, kiểm tra được, nhận biết được thủy vân.

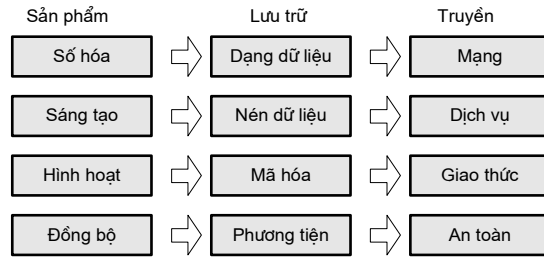
Một số nguyên tắc thủy vân đối với từng loại dữ liệu đa phương tiện:

- Thủy vân âm thanh yêu cầu nhúng thông tin mật vào đoạn giữa dữ liệu, không thêm vào đầu hay cuối. Tai người không nghe thấy thông tin mật. Thông tin này không bị tin tặc phá hoại, hay không làm hỏng dữ liệu âm thanh.
- Thủy vân hình ảnh được thực hiện với nhiều thuật toán. Thông tin giấu vào cần là các bit không nhiều ý nghĩa nghệ thuật, nội dung. Đích của việc thủy vân là giấu ảnh này vào tấm ảnh kia. Việc xử lí ảnh gốc không tác động đến ảnh giấu vào. Thủy vân cần là bền vững (i) áp dụng biến đổi DCT 2D; (ii) tránh bị phát hiện; (iii) thông tin giấu vào không bị thay đổi qua các phép xử lí ảnh gốc.
- Thủy vân văn bản cũng được tính đến. Thông tin thêm vào không làm ảnh hưởng đến nội dung văn bản.

3.3.3. Yêu cầu về tổ chức dữ liệu đa phương tiện

3.3.3.1. Một số khía cạnh quan tâm

Quan tâm đến chủ đề đa phương tiện, người ta thấy



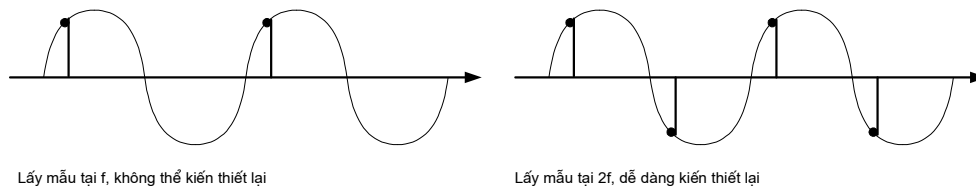
Hình. Một số khía cạnh liên quan đến đa phương tiện

3.3.3.2. Về âm thanh

Về âm thanh, có một số nhận xét: (i) âm thanh uyển chuyển; (ii) tai người nghe tần số âm thanh 20 Hz - 20 KHz; (iii) tiếng nói của con người có tần số 50 – 10 KHz. Âm nhạc là dạng âm thanh không có tiếng nói; (iv) có nhiều phương tiện chứa âm thanh: các đĩa CD, các băng cassette, các băng lớn, các đĩa MIDI của Sony; (v) âm thanh được số hóa theo điều biến xung PCM, điều biến xung sai phân DPCM, điều biến mã xung phù hợp ADPCM, MPEG...

Liên quan đến tiếng nói, khi xử lý người ta có thể dùng nhiều kỹ thuật thể hiện. Xét kỹ thuật PCM, có 3 bước (i) lấy mẫu; (ii) lượng hóa; (iii) mã hóa. Tần số lấy mẫu là (i) 8 KHz đối với điện thoại; (ii) 44.1 KHz đối với đĩa âm thanh CD; (iii) 48 KHz đối với băng âm thanh số DAT.

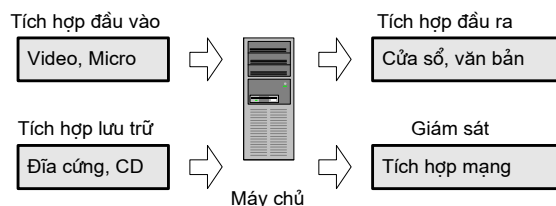
Theo định lý Nyquist
 1. Thể hiện trung thực tín hiệu tương tự :
 Nếu tần số cực đại là f , tỉ lệ lấy mẫu ít nhất là $2f$
 2. Ứng dụng cho âm thanh :
 Nếu tỉ lệ lấy mẫu là 8 KHz, độ rộng băng là 3.4. KHz



Hình. Lấy mẫu điều biến xung PCM

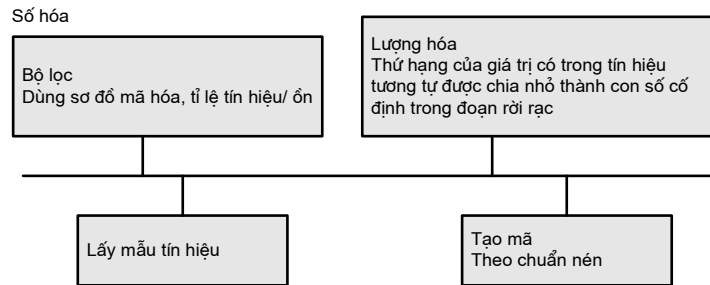
Nguyên lý tổng hợp âm nhạc điện tử (i) năm 1960, người ta tổng hợp tương tự; (ii) cuối những năm 70, người ta biến điệu tần số FM; (iii) khoảng giữa những năm 80, tổng hợp sóng; (iv) giữa những năm 80, người ta tổng hợp mẫu; (v) đến những năm 90, người ta có thể tổng hợp theo mô hình điện tử.

3.3.3.3. Số hóa thông tin



Hình. Thông tin số được tích hợp

Cần thiết số hóa thông tin. Lí do (i) dễ tiện cho lưu trữ, về băng thông, về năng lực kho chứa; (ii) thuận tiện cho truyền thông: tích hợp trên mạng máy tính, tiện cho giữ bí mật, dễ phát hiện sai sót dữ liệu; (iii) tạo điều kiện xử lí tốt dữ liệu đa phương tiện: dữ liệu ghi lại theo mẫu, đã được lọc, đã được xử lí sơ bộ.



Hình. Số hóa thông tin

3.4. Kết luận

Chương trên đã cho thấy một số hệ thống ứng dụng, cho phép xử lí dữ liệu đa phương tiện. Một số sản phẩm thương mại có thông tin quảng cáo trên Internet.

Một khía cạnh trình bày trong chương là qua trình sản xuất đa phương tiện, tức thực hiện dự án về đa phương tiện.

Khía cạnh về an toàn dữ liệu đa phương tiện, tổ chức dữ liệu đa phương tiện được trình bày như xu hướng được quan tâm.

Chương 4. Quản trị dữ liệu đa phương tiện

4.1. Khái niệm về quản trị cơ sở dữ liệu đa phương tiện

Trước hết xét việc mô tả trừu tượng các đối tượng đa phương tiện. Trừu tượng về phương tiện nhằm vào cấu trúc hình thức, cho phép thu nhận nội dung của phương tiện. Theo cách trực quan, trừu tượng về phương tiện xác định cấu trúc dữ liệu để mô tả thông tin về nguồn phương tiện.

Định nghĩa: Trừu tượng là tóm tắt ngắn của nghiên cứu, bài báo hay là phân tích sâu về chủ đề, lĩnh vực cụ thể, để người ta dễ dàng thấy được mục tiêu của vấn đề chính. Việc trừu tượng hóa là quá trình rút ra khái niệm cao hơn, nhờ sử dụng phân loại hay khái niệm cụ thể, hay các nguyên tắc ban đầu; do vậy mà trừu tượng là khái niệm liên quan đến phạm trù cao hơn của các phạm trù nhỏ, và liên kết các khái niệm, lĩnh vực, hay phạm trù khác.

4.1.1. Dạng dữ liệu đa phương tiện

Trước khi xác định được trừu tượng về phương tiện, người ta thường tự đặt câu hỏi về dạng dữ liệu chung khi có nhiều dạng dữ liệu? Khi đó người ta cần xác định loại khác nhau của dữ liệu, nhằm thu được nhìn nhận các dữ liệu về đối tượng theo kiểu chung. Do vậy người ta xây dựng công cụ chung, được gọi là Shell hay cấu trúc xương, dùng cho mục đích chung. Đối với cấu trúc chung này, cần có các đặc tính để phân biệt các dạng dữ liệu về bản chất đã khác nhau.

Các đặc tính chung của nội dung phương tiện có (i) các đối tượng riêng biệt, với nội dung được mô tả là hình ảnh, video, hay văn bản; (ii) trong mỗi đối tượng đều có các phần tử nhỏ, chi tiết hoá các hành động; mỗi hành động nhỏ cũng có các thuộc tính.

Định nghĩa: Trừu tượng về phương tiện là bộ 8 ($S, fe, att, \lambda, R, \phi$, biến 1, biến 2); trong đó S là tập các đối tượng, ứng với các trạng thái¹; fe là tập các đối tượng, ứng với các khía cạnh của đối tượng²; att là tập các đối tượng, gọi là các giá trị thuộc tính¹; λ là ánh xạ $S \rightarrow 2^{fe}$, từ các trạng thái sang tập các khía cạnh; R là tập các quan hệ trên $fe^i \times att^j \times S$ đối với $i, j \geq 0$; ϕ là tập các quan hệ của S ; biến 1 là tập các đối tượng, là biến trên S ; và biến 2 là tập các biến trên fe .

Để mô tả bộ 8 trên, có thể thuyết minh thêm về một số thành phần:

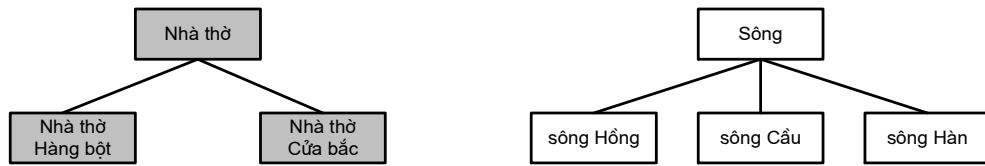
1. Trạng thái là phần tử nhỏ nhất của dữ liệu về phương tiện. Chẳng hạn mỗi hình ảnh trong cơ sở dữ liệu đa phương tiện là một trạng thái. Khi xét hình, người ta xem hình ứng với các đối tượng cá thể, với các thuộc tính được quan tâm. Chẳng hạn video = {khung hình};
2. Khía cạnh là bất kì đối tượng nào trong trạng thái, nhằm vào ứng dụng đa phương tiện được tạo nên. Chẳng hạn có ảnh photo.gif của Toto, thì khía cạnh được quan tâm là tên người trong ảnh, như Mơ, Mận,... Khía cạnh cũng có vài

¹ state
² feature

thuộc tính, chẳng hạn cô Mơ xinh, ông Mận lái xe...;

3. Hàm λ trích ra khía cạnh, cho biết khía cạnh xảy ra trong trạng thái nào;
4. R là tập các quan hệ $fe^i X att^j X S$. các quan hệ là phụ thuộc trạng thái². Chẳng hạn trong photo.gif, quan hệ phụ thuộc trạng thái là (Mơ, Mận, photo.gif), và có thể có dạng (Mận, 38 tuổi...);
5. ϕ xác định quan hệ của các đối tượng. Chẳng hạn ϕ xác định (photo₁.gif, photo₂.gif) \in quan hệ “trước” thì người ta có photo₁ được chụp trước photo₂.

Định nghĩa: cơ sở dữ liệu đa phương tiện đơn giản là tập hữu hạn các triểu tượng về phương tiện.



Hình. Quan hệ \leq trong hệ thống cơ sở dữ liệu đa phương tiện có cấu trúc

Định nghĩa: hệ thống cơ sở dữ liệu đa phương tiện có cấu trúc SMDS³ là bộ 5 $(\{M_1, \dots, M_n\}, \equiv, \leq, F_1, F_2)$; trong đó $M_i = (S^i, fe^i, att^i, \lambda^i, R^i, \phi, \text{biến } I^i, \text{biến } 2^i)$, là bộ 8 triểu tượng về phương tiện; \equiv là quan hệ tương đương trên $\phi = \cup fe^i$, $i = 1..n$; \leq là thứ tự bộ phận trên tập ϕ của các lớp tương đương trên ϕ ; F_1 là ánh xạ $\phi \rightarrow 2^{\phi \equiv}$; chẳng hạn $[f_1] F_1 ([f_2])$ ngầm định $[f_1] \leq [f_2]$, tức là ánh xạ liên kết mỗi khía cạnh f sang tập các khía cạnh “dưới” f theo quan hệ \leq ; F_2 là ánh xạ $\cup att^i (i = 1..n) \rightarrow 2^{att^i}$, chẳng hạn quan hệ \leq trong SMDS.

4.1.2. Ngôn ngữ hỏi dữ liệu đa phương tiện

Định nghĩa: Ngôn ngữ hỏi dữ liệu là ngôn ngữ máy tính, tức tập các động từ cho phép đặt câu hỏi xử lý dữ liệu, để hỏi cơ sở dữ liệu hay hệ thống thông tin.

Ngôn ngữ hỏi dữ liệu cho phép tìm kiếm, xử lý dữ liệu. Với dữ liệu quan hệ, người ta đã dùng ngôn ngữ SQL. Với dữ liệu đa phương tiện được mô hình theo tiếp cận mô hình dữ liệu quan hệ, ngôn ngữ giả như SQL được đề xuất. Ngôn ngữ này (i) Mở rộng SQL cho phép hỏi dữ liệu theo kiến trúc thể hiện thống nhất; (ii) Mở rộng SQL cho kiến trúc lai, kết hợp tiếp cận thống nhất và tự trị.

Người ta có thể đặt câu hỏi theo:

1. Hỏi về hệ thống SMDS: có các hàm mẫu xây dựng sẵn, cho phép tìm đối tượng, tìm khía cạnh của đối tượng, tìm các thuộc tính. Câu select... from của SQL cần được thay đổi phù hợp;
2. Hỏi trên các thể hiện hỗ trợ của các dữ liệu đa phương tiện.

¹ attribute value
² state dependancy
³ structured multimedia database systems

4.1.3. Vấn đề khác

Ngoài một số vấn đề liên quan đến mô tả hình thức hệ thống cơ sở dữ liệu đa phương tiện, cần thiết xem xét đến:

1. Chỉ số hoá trong SMDS bằng hệ thống chỉ số ngược;
2. Mở rộng và tối ưu các câu hỏi trên dữ liệu đa phương tiện.

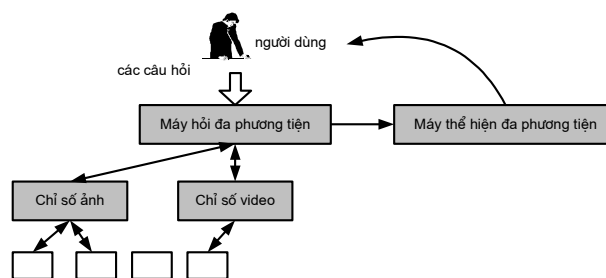
Đối với việc xây dựng cơ sở dữ liệu đa phương tiện, trước hết cần quan tâm (i) loại dữ liệu và phương tiện nào mà cơ sở dữ liệu đa phương tiện cần; (ii) thuật toán cho phép chỉ số hoá các dữ liệu một cách tin cậy và chính xác theo phương pháp chỉ số hoá theo nội dung. Người ta cần dùng lại các phần mềm đã có, và cũng kiểm tra đã có sẵn các phần mềm nào.

4.2. Kiến trúc hệ quản trị cơ sở dữ liệu đa phương tiện

Liên quan đến việc thiết kế và kiến trúc của các cơ sở dữ liệu đa phương tiện, người ta thấy: việc thiết kế và xây dựng mô hình của các cơ sở dữ liệu đa phương tiện liên quan đến việc tổ chức nội dung dữ liệu đa phương tiện và việc thể hiện vật lý của các dữ liệu.

4.2.1. Các kiến trúc về tổ chức nội dung

Người ta đưa ra ba kiến trúc ứng với các tiếp cận khi tổ chức nội dung chỉ số hoá trong hệ thống cơ sở dữ liệu đa phương tiện.



Hình. Kiến trúc các khối chức năng cho hệ thống xử lý dữ liệu đa phương tiện

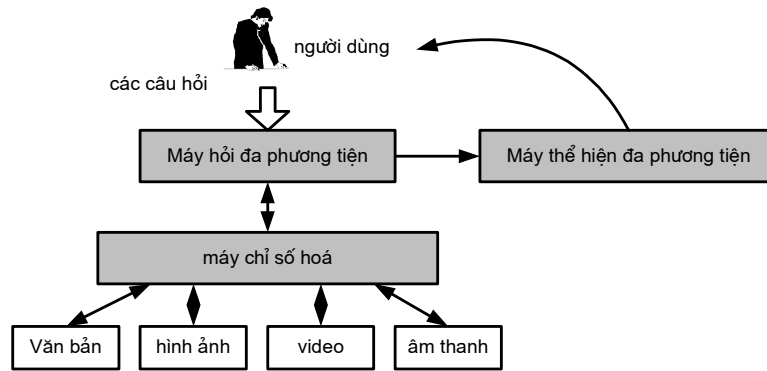
4.2.2. Nguyên tắc tự quản

Theo tiếp cận tự quản, mỗi loại dữ liệu được tổ chức theo cách tự phù hợp. Sự phù hợp được hiểu tùy thuộc vào loại phương tiện cụ thể.

4.2.3. Nguyên tắc đồng đều

Một cách tiếp cận đề xuất kiến trúc hệ thống là đảm bảo tính đồng đều¹, cho phép chỉ số hoá tất cả các dữ liệu đa phương tiện. Việc “chỉ số hoá một cách thống nhất” đồng nghĩa với việc người ta sử dụng một cấu trúc thống nhất cho mọi loại dữ liệu đa phương tiện.

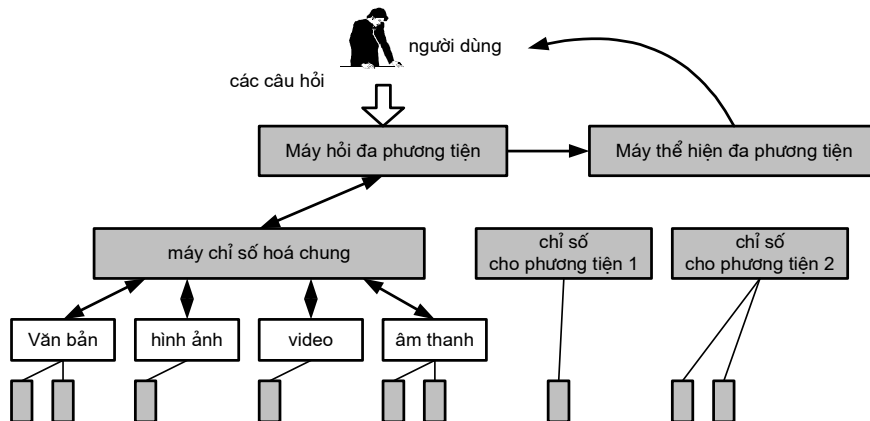
¹ uniformity



Hình. Kiến trúc đảm bảo tính thống nhất

4.2.4. Nguyên tắc tổ chức hỗn hợp

Tiếp cận theo nguyên tắc hỗn hợp kết hợp cả hai tiếp cận trên. Kiến trúc này có dữ liệu đa phương tiện có loại chỉ số riêng, và số khác tuân theo chỉ số chung.



Hình. Kiến trúc chỉ số hoá hỗn hợp

4.2.5. Một số nhận xét

Một số nhận xét về các kiến trúc chỉ số hoá:

1. Nếu sử dụng chỉ số theo tiếp cận tự trị, mỗi loại dữ liệu có cách chỉ số hoá riêng sẽ không thuận lợi cho công tác bảo trì các mối nối dữ liệu. Đối với chương trình xử lý nhiều loại dữ liệu, việc truy cập đến nhiều loại chỉ số hoá sẽ khó khăn, vì phải có các thao tác riêng đối với mỗi cách chỉ số. Vậy việc tạo cơ sở dữ liệu chung sẽ gặp không ít khó khăn;
2. Phương pháp thống nhất cách chỉ số hoá được thể hiện qua thiết bị về thông báo, dữ liệu meta, mà thông tin về thiết bị chung có trong ngôn ngữ xử lý dữ liệu. Tuy nhiên kiến trúc theo tiếp cận này cần xác định được hình thức trừu tượng áp dụng cho tất cả các loại dữ liệu đa phương tiện;
3. Theo tiếp cận hỗn hợp, người ta tránh được nhiều nhược điểm mà hai tiếp cận trên phải gặp.
 - Giả sử cần tạo cơ sở dữ liệu đa phương tiện với các phương tiện M_1, \dots, M_n ;

- Cần tách các M_i ra (i) các phương tiện có xuất xứ hợp lệ, kèm với chỉ số và thuật toán xử lý chỉ số. Do vậy, cần dùng các ưu điểm của thuật toán và chỉ số; (ii) các phương tiện không có nguồn gốc hợp lệ, không có chỉ số. Khi đó người ta khuyến cáo sử dụng tiếp cận kiến trúc thống nhất;
- Cần tạo các mã cần thiết để liên kết chéo các nguồn dữ liệu theo các chỉ số. Điều này cho phép làm tăng các điểm chung, và giảm các cái riêng.

4.2.6. Tổ chức cơ sở dữ liệu dựa trên nguyên tắc thống nhất

Tư tưởng cơ bản của nguyên tắc thống nhất là “theo quan điểm về ngữ nghĩa, nội dung của dữ liệu đa phương tiện thường độc lập với nhau”. Khi có càng ít dữ liệu meta, người dùng càng cần tiến hành các tương tác theo các câu hỏi chi tiết.

Tiếp cận thống nhất việc chỉ số hoá, có sử dụng dữ liệu meta để hướng dẫn thiết kế, nhưng không có đối với tất cả các đối tượng đa phương tiện trong cơ sở dữ liệu, cũng có những khó khăn cần được giải quyết.

Nhìn chung tiếp cận thống nhất có nhiều ưu điểm. Có thể liệt kê:

1. Dữ liệu meta thường được lưu trữ trong các cấu trúc quan hệ hay hướng đối tượng. Người ta có thể dùng ngôn ngữ SQL để hỏi dữ liệu đa phương tiện;
2. Viết chương trình xử lý các dữ liệu meta không khó;
3. Viết chương trình với một phần dữ liệu, hay phần dữ liệu trích ra, là công việc mà người lập trình quen thuộc.

Do vậy có thể sử dụng kiến trúc thống nhất cách chỉ số và sử dụng dữ liệu meta để hướng dẫn việc thống nhất hoá.

4.3. Các kỹ thuật mô hình hóa dữ liệu

Đề cập vai trò của mô hình quan hệ và mô hình hướng đối tượng đối với tổ chức dữ liệu đa phương tiện, người ta thấy việc sử dụng cơ sở dữ liệu quan hệ chủ yếu là do thói quen của người dùng và bản thân ưu điểm của mô hình quan hệ. Theo các chuyên gia về đa phương tiện, mô hình hướng đối tượng tỏ ra tốt hơn mô hình quan hệ trong việc mô tả các đối tượng của thế giới thực. Trong bài báo của Newman, những so sánh giữa hai mô hình đối với vấn đề đa phương tiện đã thiên về sử dụng mô hình hướng đối tượng.

4.3.1. Mô hình quan hệ

Mô hình quan hệ là mô hình dữ liệu quen thuộc. Tuy nhiên trong tài liệu này cần điểm lại một số nét chính của nó, nhằm mô tả các chức năng xử lý dữ liệu trong hệ thống đa phương tiện. Yêu cầu người dùng xác định được

- Các công nghệ đi với mô hình quan hệ;
- Hai môi trường có công nghệ cơ sở dữ liệu cần được nghiên cứu là UNIX và WINDOWS;
- Về hệ thống cơ sở dữ liệu, cần quan tâm đến (i) hệ thống quản lý các bảng dữ liệu; (ii) lược đồ quan hệ (A_1, A_2, \dots, A_n), trong đó A_i là tập dữ liệu, là cột thứ i trong bảng dữ liệu; (iii) thí dụ về cơ sở dữ liệu quan hệ.

4.3.1.1. Đại số quan hệ

Đại số quan hệ là một trong ba ngôn ngữ hỏi dữ liệu trong mô hình quan hệ. Về đại

số quan hệ, lưu ý có năm phép đại số quan hệ là các phép nhân, chiếu, hạn chế, hợp và trừ; các phép bổ sung là phép nối, chia và giao; tổng cộng có tám phép đại số quan hệ.

- Phép chiếu quan hệ $R(A_1, \dots, A_n)$ trên danh sách các thuộc tính A_i, \dots, A_j được quan hệ $S(A_i, \dots, A_j)$ và chứa các bộ là các bộ của quan hệ R , chỉ sử dụng các thuộc tính chiếu và bỏ đi các bộ trùng.
- Phép hạn chế quan hệ $R(A_1, \dots, A_n)$ với điều kiện Q có dạng $A_i \theta c_i$, trong đó A_i là thuộc tính, c_i là hằng số, được quan hệ có cùng lược đồ $S(A_1, \dots, A_n)$, chứa các bộ thỏa mãn điều kiện Q .
- Hợp của hai quan hệ có cùng lược đồ là quan hệ có cùng lược đồ và chứa các bộ của hai quan hệ đầu.
- Phép nhân hai quan hệ $R(A_1, \dots, A_n)$ và $S(B_1, \dots, B_m)$ là quan hệ có lược đồ $(A_1, \dots, A_n, B_1, \dots, B_m)$ và có các bộ là ghép các bộ của quan hệ R với các bộ của quan hệ S .
- Trừ hai quan hệ có cùng lược đồ R và S được quan hệ có cùng lược đồ, có các bộ thuộc quan hệ R mà không thuộc quan hệ S .
- Nối hai quan hệ $R(A_1, \dots, A_n)$ và $S(B_1, \dots, B_m)$ với điều kiện đa thuộc tính $A_i \theta B_j$ là quan hệ có lược đồ $(A_1, \dots, A_n, B_1, \dots, B_m)$ và có các bộ là các bộ của tích R và S , thỏa mãn điều kiện.
- Giao của hai quan hệ có cùng lược đồ là quan hệ có cùng lược đồ và chứa các bộ thuộc cả hai quan hệ.
- Thương của quan hệ $R(A_1, \dots, A_m, A_{m+1}, \dots, A_n)$ cho quan hệ $S(A_1, \dots, A_m)$ là quan hệ có lược đồ (A_{m+1}, \dots, A_n) , chứa các bộ sao cho ghép bộ này với các bộ thuộc S đều được bộ thuộc R .

4.3.1.2. Tính toán quan hệ

Tính toán quan hệ, hay phép toán quan hệ được xác định trên phép toán vị từ bậc một. Người ta tách phép toán quan hệ trên miền và phép toán quan hệ trên bộ.

Đại diện của đại số quan hệ là ngôn ngữ SQL, còn đại diện của phép toán quan hệ là QUEL, loại phép toán quan hệ trên bộ, và QBE, loại phép toán quan hệ trên miền.

4.3.2. Cơ sở dữ liệu hướng đối tượng

Cần thiết hiểu biết sâu về cơ sở dữ liệu hướng đối tượng và cơ sở dữ liệu quan hệ hướng đối tượng. Việc sử dụng mô hình quan hệ hướng đối tượng là cách phù hợp để sử dụng được các kinh nghiệm về cơ sở dữ liệu quan hệ và nhìn nhận thế giới thông qua các đối tượng.

Các khía cạnh liên quan đến mô hình hướng đối tượng được quan tâm là (i) thông báo; (ii) phương pháp; (iii) thừa kế. Các kiểu dữ liệu hướng đối tượng hiện đang thông dụng và các kiểu dữ liệu trong thời kì đầu của mô hình dữ liệu hướng đối tượng sẽ có ích trong việc mô tả dữ liệu đa phương tiện. Ngoài ra, các kĩ thuật cài đặt mô hình hướng đối tượng cho phép thể hiện các khái niệm mới.

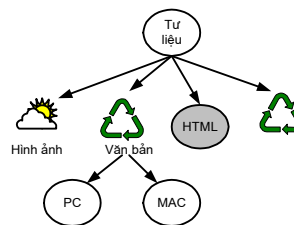
Lưu ý một số vấn đề nảy sinh khi phát triển cơ sở dữ liệu quan hệ: (i) dữ liệu ở

dạng phẳng, ngay cả khi cần thiết mô tả dữ liệu phức tạp; (ii) lược đồ quan hệ là tĩnh một cách tương đối; người ta không thể dễ dàng thay đổi tạm thời các cột, các dòng; không có các phương tiện trợ giúp hữu hiệu; (iii) khi cần mô tả mối quan hệ, trong một hay nhiều bảng, người ta dùng đến cả điều kiện toàn vẹn, và người dùng không được hỗ trợ tốt về khả năng này.

Do nhu cầu phát triển của công nghệ thông tin và đảm bảo tính kế thừa, việc chương trình hoá hướng đối tượng được đưa ra. Tư tưởng chính của việc này nằm ở (i) các đối tượng; (ii) các lớp; (iii) tính phân cấp. Việc phân cấp sẽ cho phép các cấu trúc đồ thị không chu trình trên tập các lớp đối tượng. Khi phát triển các lớp, người ta đề cập vấn đề thực thể độc lập, như tập các chương trình, tương tác với nhau qua việc truyền các thông báo.

4.3.2.1. Các đối tượng và giá trị

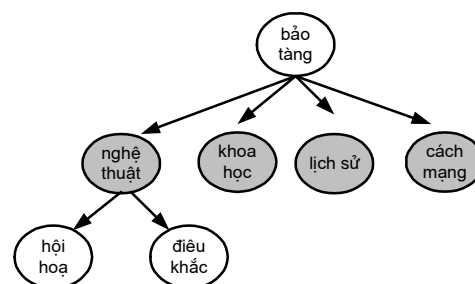
Thí dụ về các đối tượng và giá trị là HTML. Thí dụ khác về kịch bản: giả sử d_1, d_2, \dots điều kiện là tập các tư liệu mà người ta cần truy cập. Mỗi tư liệu rơi vào một lớp; có tư liệu trên máy tính chủng loại này, cái trên máy tính chủng loại khác, như máy trạm, PC hay MAC.



Hình. Phân loại dữ liệu đa phương tiện

Thí dụ về bảo tàng:

- Mỗi bảo tàng được xem như một đối tượng;
- Bảo tàng có thể được nhóm lại theo (i) bảo tàng khoa học; (ii) bảo tàng lịch sử...
- Mỗi đối tượng mô tả theo thuộc tính, như tên tác giả,...



Hình. Phân loại đối tượng đa phương tiện

Định nghĩa: Đối tượng văn bản¹ bao gồm:

1. Một tập các phần tử O_{id} (object id);
2. Tập các phần tử, C_{id} (class id);

¹ object alphabet

3. Tập các thuộc tính. Liên quan đến các thuộc tính a thuộc tập thuộc tính Att là miền xác định, kí hiệu $dom(a)$.

Định nghĩa: giả sử $\Sigma = (tập\ O_{id},\ Cid,\ Att)$ là bộ đối tượng và $Att-core \subseteq Att$ là một số thuộc tính. Tập các giá trị được Σ tạo và $att-core$ gọi là $Value(\Sigma, Att-core)$, tức không gian giá trị do Σ tạo ra, được xác định:

1. Mỗi phần tử thuộc $O_{id} \cup (\cup_{a \in A} dom(a))$ là một giá trị;
2. NIL là giá trị đặc biệt;
3. Nếu $A_1, A_2, \dots, A_n \in Att-core, c_i \in dom(A_i) \forall i = 1..n$ thì $[a_1 = c_1, \dots, a_n = c_n]$ là một giá trị;
4. Nếu v_1, \dots, v_m là giá trị thì $\langle v_1, \dots, v_m \rangle$ là một giá trị, gọi là giá trị bộ¹;
5. Nếu v_1, \dots, v_m là giá trị thì $\{v_1, \dots, v_m\}$ là một giá trị, gọi là giá trị tập².

Trong định nghĩa trên, $Att-core$ được gọi là thuộc tính nhân.

Thí dụ trong kịch bản đa phương tiện $O_{id} = \{b_1, \dots, b_6\}; c_{id} = \{\#html, \#ảnh, \#văn\ bản\ gốc, \#văn\ bản\ MAC, \#văn\ bản\ PC\}; Att-core$ được xác định:

- $Att-core: \{real, bool, int, string\};$
- $Att-Ncore: \{tác\ giả, ngày\ tạo, ngày\ thay\ đổi, tư\ liệu\ liên\ quan\};$

Không gian các giá trị $value(\Sigma)$ liên quan là

- $[tác\ giả = tutu];$
- $[tác\ giả = titi; ngày\ tạo = (20/03/05)];$
- $[tác\ giả = toto, tư\ liệu\ liên\ quan = \{b_2, b_6\}]$

Định nghĩa: cho bộ Σ và nhân $Att-core$. Một đối tượng O là cặp (id_0, val_0) , trong đó $id_0 \in O_{id}$, là tên; val_0 là giá trị của O .

Trong cơ sở dữ liệu hướng đối tượng, người ta có thể xác định các đối tượng theo cú pháp

```
Declare tên
Values giá trị
Declare b2
Values [tác giả = toto;
        url = http://www.vnn.vn;
        ngày tạo = (20/ 03/ 05);...
```

4.3.2.2. Kiểu dữ liệu và lớp

Kiểu dữ liệu³ xác định các kiểu, trạng thái dữ liệu, như kiểu logic, nguyên, kí tự... Người ta sử dụng các từ khoá tiếng Anh quen dùng như bool, int, char...

- Mỗi $Att-core$ là một kiểu dữ liệu;
- Mỗi Cid cũng là kiểu dữ liệu;
- Kiểu bản ghi có các trường. Mỗi trường có các kiểu dữ liệu tương xứng. Do vậy người ta sử dụng kí hiệu trường là f_1, f_2, \dots, f_n , và kiểu dữ

¹ tuple value
² set value
³ type

- liệu trường là J_1, J_2, \dots, J_n , hay $[f_1:J_1, \dots, f_n:J_n]$;
- Kiểu tập được viết là $[J]$, kí hiệu tập các mục tin dữ liệu;
- Danh sách có dạng $\langle J \rangle$, kí hiệu danh sách các mục tin dữ liệu.

Thí dụ người ta dùng

[tác giả: string

url: kiểu url...]

Định nghĩa: Một phân cấp lớp¹ là bộ ba $(G, \leq, \text{kiểu dữ liệu})$, trong đó G là tập các đối tượng và lớp; \leq là thứ tự bộ phận trên G ; kiểu dữ liệu là ánh xạ một kiểu dữ liệu sang $g \in G$.

$\forall g_1, g_2 \in G, g_1 \leq g_2 \rightarrow \text{kiểu}(g_1)$ kiểu con $(\text{kiểu}(g_2))$, trong đó quan hệ kiểu con được xác định:

- $[f_1:J_1, \dots, f_{n+k}:J_{n+k}]$ là kiểu con của $[f_1:J_1, \dots, f_n:J_n]$;
- Nếu J_1 là kiểu con của J_2 , thì $\{J_1\}$ là kiểu con của $\{J_2\}$;
- Nếu J_1 là kiểu con của J_2 , thì $\langle J_1 \rangle$ là kiểu con của $\langle J_2 \rangle$;

Lưu ý rằng

- trong (1), tuy $n+k$ có nhiều thuộc tính hơn nhưng mang thông tin về gốc là $1..n$;
- $g_1 \leq g_2 \equiv$ liên kết $g_1, g_2 \equiv g_2$ ở trên² g_1 ;

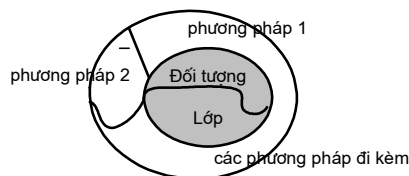
Định nghĩa: giả sử $(G, \leq, \text{kiểu})$ là phân cấp lớp và $g \in G$. Thành viên của lớp G gồm các phần tử trong tập $\{g' \in G \mid g' < g\}$

4.3.2.3. Phương pháp

Phương pháp nhằm các thủ tục, thao tác thực hiện trên các đối tượng. Có thể xem phương pháp là khía cạnh động, liên quan đến hành vi của đối tượng. Vậy phương pháp là nét quan trọng trong tiếp cận hướng đối tượng.

Khi cho phân cấp lớp $(G, \leq, \text{kiểu dữ liệu})$, mỗi lớp $g \in G$ có kiểu dữ liệu riêng.

- Các kiểu dữ liệu này là kiểu con của g' (g' là cha, ông của g);
- Có thể có trường mới mà cha ông nó không có;
- Phương pháp là chương trình gắn với lớp/ đối tượng $g \in G$ thực hiện các cấu trúc của mô tả trong định nghĩa của g' . Tức là trong cơ sở dữ liệu hướng đối tượng, mỗi lớp g được gắn với tập các chương trình phương pháp (g) dùng cho lớp này. Do vậy khái niệm “nắm được, đóng gói³” được sử dụng;
- Các đối tượng trong lớp được phương pháp xử lí; tức phương pháp thuộc lớp đó, chứ không nhằm vào các phương pháp khác.



Hình. Các đối tượng trong tiếp cận hướng đối tượng

¹ class hierarchy
² above
³ encapsulation

4.3.2.3.1. Các kiểu vào / ra

Định nghĩa: Nếu m là phương pháp gắn với lớp g , thì m có kiểu vào/ ra, được gọi là chữ kí¹, cho phép xác định cấu trúc kiểu dữ liệu nhập vào mà g mong muốn, và có cả dữ liệu ra.

Trong trường hợp tổng quát, người ta thấy:

- Kiểu vào của m có dạng $J_1 \times \dots \times J_k$, trong đó $J_i, i = 1..n$, là một kiểu;
- Kiểu ra là J_0 với kiểu J_0 nào đó;

Tức là phương pháp m hi vọng k đầu vào J_1, \dots, J_k và một đầu ra J_0 .

4.3.2.3.2. Lưu ý. Một số lưu ý

- Cùng một phương pháp có thể được xác định trong các lớp khác nhau;
- Nếu lớp g , mà m xác định, được xác định không rõ ràng theo ngữ cảnh, chẳng hạn như m được xác định trong nhiều lớp, thì người ta sẽ dùng kí pháp m_g để chỉ phương pháp m cho lớp g .

4.3.2.3.3. Thừa kế

Tính thừa kế nhằm chuyển giao các thuộc tính của lớp trên cho lớp sau. Khi m xác định cho lớp g thì m xác định cho cả lớp con của g . Đó là kế thừa m . Có hai vấn đề cần đề cập (i) tương tự về cấu trúc²; (ii) giải pháp xung đột³.

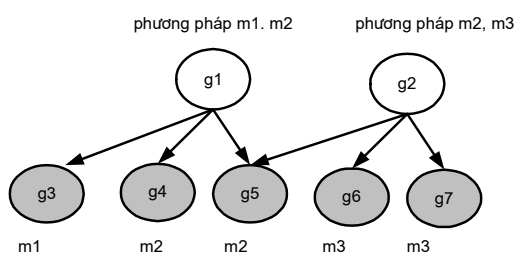
Vấn đề 1. Tương tự về cấu trúc

Giả sử có hai lớp g_1 và g_2 . $g_1 \leq g_2$. Phương pháp m được xác định cho cả hai:

- $M = \text{FindDoc}$;
- g_1 = văn bản;
- g_2 = tư liệu.

Người ta muốn m thể hiện sự tương tự về cấu trúc qua hai lớp, tức là nếu kiểu dữ liệu nhập của m trong lớp con g_1 là $J_1 \times \dots \times J_n$ và nếu kiểu dữ liệu nhập của m trong lớp cha g_2 là $J'_1 \times \dots \times J'_n$ thì $\forall i = 1..n, J_i$ là kiểu con của J'_i .

Thí dụ về thừa kế của phương pháp



Hình. Thừa kế trong tiếp cận hướng đối tượng

Nhìn chung, khi g là lớp thì $\uparrow g$ là tập, $\{g' \in G \mid g \leq g'\}$, người ta nói lớp g thừa kế tiềm năng phương pháp m từ lớp g^* nếu (i) $g^* \in \uparrow g$ và m xác định cho g^* ; và (ii) không có lớp g_0 mà $g \leq g_0 < g^*$ và phương pháp m xác định cho g_0 .

Vấn đề 2. Giải pháp xung đột trong việc thừa kế phương pháp

¹ signature
² Structure similarity
³ Conflict resolution

Có thể xảy ra trường hợp g thừa kế phương pháp từ hai lớp cha, mà hai lớp cha không tương hợp, tức không sắp thứ tự đối với g_1, g_2 được: $g_1 \sim \leq g_2 / g_2 \sim \leq g_1$, thì xảy ra xung đột về việc thừa kế phương pháp. Trường hợp điển hình: tập các lớp mà g thừa kế có nhiều phần tử.

Chính sách giải quyết xung đột là dùng ánh xạ crp^1 : $Crp(\{g_1, g_2\}) = g_1$.

Người ta có thể thực hiện điều này theo nhiều cách:

- Dùng trật tự từ vựng, $crp(X)$ là phần tử đầu tiên trong thứ tự tổng cộng;
- Gần nhất², $crp(X) = g'$, trong đó g' vừa mới được phát triển. Lúc này cần có đánh dấu thời gian;
- Liên kết với con số, như mức độ ưu tiên, dùng $pr(g)$ đối với g . Có thể độ ưu tiên cao ứng với giá trị số lớn.

Nhìn chung người phát triển có thể sử dụng chiến lược giải quyết riêng.

4.3.2.4. Xác định đối tượng và ngôn ngữ hỏi dữ liệu

Người ta có thể thực hiện nhiều cách trên các dạng cú pháp. Do vậy cần (i) phân tán các ngôn ngữ quản trị dữ liệu đối tượng; (ii) chuẩn hoá các khuynh hướng sử dụng cú pháp.

Thị trường có ODMG³ với nhiều cố gắng về chuẩn hoá và xử lý dữ liệu đối tượng. Hệ quản trị này gắn với hai khái niệm quan trọng:

- Ngôn ngữ xác định đối tượng (ODL) cho phép xác định các đối tượng;
- Ngôn ngữ hỏi đối tượng (OQL) cho phép hỏi về định nghĩa đối tượng.

4.3.2.4.1. Ngôn ngữ xác định đối tượng

Mục tiêu đối với ngôn ngữ xác định các đối tượng là đảm bảo ngôn ngữ đơn giản mà xác định được cả đối tượng lẫn giao diện đối tượng. Một số nguyên tắc:

- Mỗi đối tượng gắn với lớp các phương pháp. Do vậy chương trình ngoài muốn truy cập hay xử lý đối tượng cần có cách gọi các phương pháp. Vậy cần truy cập các kiểu vào/ ra, tức đánh dấu⁴;
- ODL cung cấp cú pháp hình thức;
- ODL không xác định cách thức phương pháp được cài đặt, mà chỉ kiểu vào/ ra của phương pháp này.

Thí dụ

```
Interface html: tư liệu
  (external html_tư liệu
    keys url: persistent
    {<thuộc tính>
    <các mối quan hệ>}
  );
```

4.3.2.4.2. Ngôn ngữ hỏi đối tượng

Ngôn ngữ OQL là dạng mở rộng của ngôn ngữ SQL dùng cho hỏi đối tượng. SQL

¹ Conflict resolution policy

² Recent

³ object database management group

⁴ signature

chỉ thao tác trên các bảng quan hệ, được coi là phẳng. Các dữ liệu mà OQL xử lý có cấu trúc lồng, có các kiểu (i) Kiểu tập hợp¹; (ii) Tập; (iii) Danh sách; (iv) Túi².

Ngôn ngữ OQL cho phép truy cập các kiểu dữ liệu trên. Thí dụ:

```
select struct (trường 1: X.url, trường 2: X.liên kết)
```

```
From văn bản X
```

```
Where X. tác giả = "toto"
```

4.3.2.5. Các hệ thống quan hệ đối tượng

Bản thân mô hình dữ liệu hướng đối tượng chưa khác biệt mô hình dữ liệu quan hệ, chưa cho phép người ta có cách nhận thức mới về dữ liệu thế giới thực. Mặt khác, mô hình dữ liệu quan hệ đang được sử dụng, tuy có một số hạn chế, giải quyết đa số các vấn đề và làm nền cho các mô hình dữ liệu tiên tiến.

Hệ quản trị cơ sở dữ liệu theo mô hình hướng đối tượng có thể sử dụng nhiều khía cạnh của hệ quản trị cơ sở dữ liệu quan hệ. Mô hình dữ liệu quan hệ xử lý các bảng phẳng có ưu điểm tiện theo dõi.

Để mở rộng hệ quản trị cơ sở dữ liệu quan hệ cho mục đích hướng đối tượng, người ta có thể làm phức tạp dữ liệu quan hệ, tiến đến loại dữ liệu quan hệ-đối tượng. Chẳng hạn thực thể người (tên, tuổi, địa chỉ, photo), trong đó thuộc tính photo gắn với đối tượng không theo các kiểu dữ liệu quan hệ thường dùng.

Về ngôn ngữ hỏi đối tượng: người ta không sử dụng trực tiếp ngôn ngữ SQL cho các đối tượng. Thao tác đơn giản như so sánh khớp mà cơ sở dữ liệu hướng đối tượng cần là thao tác không đặc thù trong cơ sở dữ liệu quan hệ.

Giả sử có tập các đối tượng), gắn với các thuộc tính và phương pháp. Lược đồ quan hệ đối tượng có dạng $(A_1:T_1, \dots, A_n:T_n)$, trong đó A_i : tên thuộc tính, T_i : tên đối tượng.

Lưu ý rằng các xâu và số nguyên, số thực cũng được xem như các lớp đối tượng, với thao tác cộng, trừ, so sánh logic.

Thí dụ: (tên: str,... photo: image)

Việc khớp các ảnh cần thiết so sánh kích thước, các điểm ảnh...

4.3.3. Cơ sở dữ liệu đa phương tiện

Mô hình dữ liệu đa phương tiện không được xếp vào các cơ sở dữ liệu truyền thống. Các cơ sở dữ liệu thuộc hệ quản trị cơ sở dữ liệu thế hệ thứ nhất, thế hệ hai và các cơ sở dữ liệu theo mô hình tiên tiến, là mô hình dữ liệu phân cấp, mô hình dữ liệu suy diễn và mô hình dữ liệu hướng đối tượng, hay được kể đến. Tuy nhiên theo khía cạnh ứng dụng, không thể không kể đến cơ sở dữ liệu đa phương tiện.

Cần đưa ra các khái niệm toán học hình thức, thường được gọi là các phương tiện trừu tượng để mô tả các kiểu dữ liệu đa phương tiện và các thủ tục xử lý dữ liệu đa phương tiện. Người ta cần xét các khía cạnh đặc biệt hơn so với các mô hình dữ liệu truyền thống, như vấn đề chỉ số hoá, vấn đề nén dữ liệu đa phương tiện.

Trong phần mở đầu, các loại dữ liệu đa phương tiện được nêu là (i) văn bản; (ii)

¹ collection type

² Set, list, bag

hình ảnh; (iii) hình động; và (iv) âm thanh. Tiếp theo người ta cần quan tâm đến cả các phương tiện cho phép lưu trữ các loại dữ liệu đa phương tiện này. Nhìn chung về mặt vật lý, cơ sở dữ liệu đa phương tiện là tập các dữ liệu đa phương tiện hay các dữ liệu truyền thống.

4.4 Các kĩ thuật chỉ số hoá và trừu tượng hoá

4.4.1. Giới thiệu

Việc tìm kiếm các đối tượng đa phương tiện trong cơ sở dữ liệu đa phương tiện, hay trong hệ thống thông tin, có thể cấu trúc hay không cấu trúc, có thể dùng (i) ngôn ngữ hỏi; (ii) bộ lọc; (iii) các liên kết kiểu siêu văn bản. Người ta có thể vạch ra sự tương tự giữa (i) các hệ thống cơ sở dữ liệu đa phương tiện, với (ii) các hệ thống tìm thông tin. Điểm chung là các yêu cầu đối với các đối tượng cơ sở dữ liệu lớn dựa trên *nội dung đối tượng*:

- Cách người ta hỏi dữ liệu tương tự nhau;
- Với đa phương tiện, người ta duyệt dữ liệu thông qua câu hỏi cấu trúc hơn về nội dung;
- Trong thư viện video, như kho tư liệu băng của IBM, người ta duyệt qua, duyệt lại...

Để tìm các đối tượng như file âm thanh, đoạn video, ảnh đồ hoạ, các hệ thống thường trả lời theo *mô tả văn bản của hình ảnh*. Do vậy phải có thuyết minh thông tin bằng văn bản. Với ảnh 2D cũng vậy. Do đó:

- i. các chỉ số bình thường được sử dụng, theo mô tả văn bản đó;
- ii. các hệ thống phân loại chỉ số theo văn bản, theo cách thông thường, sẽ dựa trên số hạn chế các từ khoá (như tạp chí ACM đã dùng);
- iii. ứng dụng chỉ số theo nội dung là mở rộng việc tìm văn bản, như là tìm theo từ khoá mà người ta quên thấy trong soạn văn bản;
- iv. tiếp cận truyền thống trong hệ thống tìm thông tin áp dụng cho việc chỉ số theo nội dung của các tư liệu có số lượng các từ khoá nhỏ; việc tìm được chuyển sang tìm tư liệu liên kết với khoá cụ thể. (a) Việc chỉ số hoá thủ công yêu cầu bằng chỉ số đọc mọi văn bản để tìm ra chỉ số; (b) chỉ số hoá tự động tự sinh ra các chỉ số, đáp ứng được văn bản đang dùng và có thể sẽ dùng;
- v. đối với hệ thống đa phương tiện, việc chỉ số hoá chưa rõ ràng, cũng như cách thức chỉ số hoá các đối tượng đa phương tiện chưa rõ. Vậy các (i) từ khoá văn bản và (ii) các trừu tượng liên kết với đối tượng video nhằm lập chỉ số và tìm kiếm; còn việc chỉ số đối với các đối tượng hình hiển thị đang được phát triển.

Các hệ thống trừu tượng hoá và chỉ số hoá dựa vào việc nhận biết các mức video khác nhau của các trừu tượng. Mức thấp nhất của video gồm các khung, tức là cái ứng với một ảnh tĩnh. Các khung được nhóm lại thành một lia, được ghi liên tục bằng cùng một máy quay.

Biên xác định khởi đầu và kết thúc một lia căn cứ vào (i) phân bố màu sắc của các khung liên tiếp; (ii) hiệu ứng xử lý (hay soạn thảo) như chuyển cảnh theo cách trộn lẫn,

đối chỗ, cắt...(iii) xuất hiện hay biến đi của đối tượng. Người ta đã:

- i. Đưa ra nhiều phương pháp đáp ứng yêu cầu này;
- ii. Có một số kỹ thuật phát triển tốt cho phép định biên, như báo cáo 1995 của Flickner: khi thử các video có 2.000 đến 5.000 khung không cắt, có hiện tượng mất và ít khi phát hiện được biên sai về các lia cảnh.

Dãy các lia cảnh có liên quan nhau tạo nên cảnh. Việc nhóm các lia để tạo nên cảnh dựa vào chủ đề. Các nhân tố được tính đến khi xác định biên giữa các cảnh:

1. *Nội dung về màu* của các khung liên tiếp, giống như sự tương tự (i) của các khung trong cảnh; (ii) lia trong cảnh cùng chia sẻ nền chung, cùng độ sáng (ngày/ đêm...);
2. *Hội thoại giữa hai nhân vật* sẽ tạo nên các phần của cùng một cảnh, cho dù nội dung về màu và khung có thể khác nhau. Chẳng hạn như hai người đang gọi điện thoại. Yeung, Yeo và Liu, 1996 đã phát triển thuật toán tính khoảng cách giữa hai lia tương tự nhau;
3. *Nội dung về âm* thường thay đổi nhiều mỗi khi thay cảnh so với thay đổi trong một cảnh. Khi hội thoại, hai nhân vật thường không đồng bộ hoá một cách chính xác với nút chuyển của máy quay gắn với hai người; vậy nên nếu cắt video không gắn với cắt âm thì không thể hiện được cắt cảnh.

4.4.2. Chỉ số hoá cơ sở dữ liệu đa phương tiện

Chỉ số đảm bảo cơ chế để định chỗ các đối tượng dữ liệu riêng lẻ, hơn là tìm kiếm tuần tự trong cơ sở dữ liệu hay trên băng video cá nhân. Chỉ số là cấu trúc phân cấp, mỗi nút có khoá, mà mục tin được xác định, và trở đến chỉ số mức tiếp theo hay đến đối tượng dữ liệu.

Các kỹ thuật chỉ số đối với cơ sở dữ liệu đa phương tiện cần có khả năng định chỗ nhiều phương tiện theo các từ khoá hay hình ảnh liên kết tự động hay thủ công.

Nhiều nhà nghiên cứu đã kiểm tra phương pháp để ghi nhận tự động các hình hiển thị, có âm.

1. *Về chỉ số văn bản*: Các tư liệu văn bản có thể được chỉ số hoá tự động, căn cứ vào tần suất sử dụng từ. Các từ xuất hiện với tần suất cao sẽ được chọn làm từ khoá cho việc chỉ số hoá;
2. *Về chỉ số hoá tiếng nói*: Chang và đồng nghiệp đã quan sát, nhận thấy phân tích tiếng nói có ý nghĩa để xác định nội dung hơn là để nhận ra hình trong video. Họ đã phát triển cơ sở dữ liệu video demo để ghi nhận các sự kiện thể thao từ băng bình luận. Người ta thấy: kỹ thuật chỉ số hoá văn bản cũng trợ giúp cho việc chuyển đổi tiếng nói, như là cho tư liệu văn bản.

4.4.3. Các chỉ số hiển hiện

Zhang, Wu và Smolier năm 1995 sử dụng tổ hợp các tiếp cận để chỉ số video. Họ cho phép người dùng *cung cấp mô tả video bằng văn bản*, rồi phát triển hệ thống chỉ số hoá tự động theo nội dung. Video được tự động (i) chia thành lia cảnh; (ii) khung chính được trích ra từ mỗi lia; và (iii) nội dung của khung chính trở thành chỉ số. Khung chính

được thể hiện qua (i) màu; (ii) mẫu; (iii) hình dạng; (iv) khía cạnh lẽ.

Đa số các nhóm chỉ số hoá đối tượng đa phương tiện theo nội dung. Ngoài ra có kết quả của Theodoridis và đồng nghiệp, năm 1996, chỉ số hoá theo các đặc trưng không gian, thời gian của đối tượng xuất hiện trong ứng dụng đa phương tiện. Khi khai thác ứng dụng đa phương tiện, các đối tượng đa phương tiện được thể hiện hoặc theo tiếng hoặc theo hình. Hệ thống chỉ số hoá cho phép hỏi theo:

1. Đối tượng dữ liệu nào xuất hiện trên màn tại thời điểm xác định (mối quan hệ về thời gian)?
2. Các đối tượng dữ liệu nào xuất hiện trên màn liên quan đến đối tượng xác định (quan hệ về không gian)?
3. Nhân vật nào xuất hiện chen vào đối tượng xác định khi nó xuất hiện trên màn (quan hệ không gian, thời gian)?

Nhóm Theodoridis và đồng nghiệp cho thấy: (i) chỉ số này sẽ tăng tốc tìm kiếm các đối tượng để các ứng dụng đa phương tiện thực hiện mịn hơn; (ii) sử dụng cây R để nắm bắt mối quan hệ cả không gian và thời gian. Các vị trí không gian được thể hiện 2D và thời gian 1D. Chỉ số cây R chấp nhận quản lý đối tượng 3D để đáp ứng bản chất không gian, thời gian của đối tượng đa phương tiện.

Taniguchi đã phát triển hệ thống tự động chỉ số các video nhận được theo thời gian thực. Nguồn video từ (i) sóng TV, (ii) máy quay an ninh.

- Các đối tượng được thu trong 24 giờ và chỉ số hoá cho phép truy cập trực tiếp.
- Dòng dữ liệu video đến được chuyển sang khối phát hiện tách (i) các thay đổi cảnh, (ii) thay đổi âm thanh có ý nghĩa, (iii) phép xử lý máy quay, (iv) các đối tượng chuyển động trong video; các biên của các sự kiện cần phù hợp với ứng dụng: chuyển cảnh trên màn ảnh TV hay phát hiện chuyển đổi tượng tại nơi đang theo dõi.
- Khung đầu tiên của một sự kiện được dùng như khung chính. Khung chính, tức sự kiện, gồm cả video và âm, chỉ số theo thời gian, được lưu trong 24 giờ. Người ta dùng chuẩn nén MPEG1, và 24 giờ cần 1.6 Gbyte không gian lưu trữ.
- Cách thể hiện chỉ số cũng có thể thay đổi để thể hiện tổng quan. Trong 24 giờ, có trên 19.000 khung chính, nên cần đảm bảo khả năng duyệt hiệu quả.

4.4.4. Trừu tượng hoá video

Trong đề án MoCA¹, Lienhart, Pfeiffer và Effelsberg, 1997, đã phát triển kỹ thuật để kiến thiết trừu tượng video. *Trừu tượng video* là thể hiện hình, âm của video. Trừu tượng video đảm bảo nội dung, nhưng ngắn hơn video.

- Ban đầu, video được chia ra các lia và cảnh. Một khi video được tách, trừu tượng video sẽ bao gồm dãy các khung. Trừu tượng video được dùng cho nhiều mục đích: phim, tư liệu trong giáo dục, tin tức;
- Các ứng dụng khác nhau sẽ có các *chỉ tiêu* khác nhau cho phép tạo nên trừu tượng video. Vấn đề là tạo trừu tượng một cách tự động, và nhóm nghiên cứu của MoCA

¹ Movie Content analysis video abstracting system

tập trung vào;

- Tiếp cận của nhóm Lienhart, Pfeiffer và Effelsberg đã định tên được bốn thành phần cơ bản của trừu tượng video. Những mặt được định tên như các đối tượng trong video và các mặt ít xuất hiện sẽ bị loại bỏ. Kỹ thuật nhóm nghiên cứu sử dụng dựa trên thuật toán của nhóm nghiên cứu tại đại học Carnegie-Melo, Mỹ, dựa trên mạng nơron với hàng nghìn mặt, và mắt và mũi được định tên thủ công.
- Các kỹ thuật nhận mặt người được phát triển cho hình tĩnh, và mặt cần
- xuất hiện chính diện; kỹ thuật cần mở rộng cho sự xuất hiện đa dạng, vì trên băng video, ít khi mặt hiện ra trực tiếp.
- Để tăng tốc xử lý, người ta xử lý cùng lúc ba khung, và chỉ các khung có các pixel với màu sắc gần đúng với màu mặt được chuyển đến thuật toán phát hiện. Thuật toán phát hiện mặt người định tên các khung có hình ảnh mặt. Với mục đích tạo trừu tượng video, chỉ những mặt chiếm trên 30% của khung mới được coi là nhân vật chính.
- Các khung chứa các ảnh khác nhau của cùng khuôn mặt cần được nhóm lại thành *lớp dựa trên mặt*. Các hình vẽ mặt trong một lia cảnh cần có liên quan, trên cơ sở kích thước và vị trí trong các khung kề nhau. Hội thoại giữa hai nhân vật có lúc máy quay cắt cũng phản ánh thay đổi về kích thước và vị trí. Thuật toán nhận dạng khuôn mặt có thể được dùng để nhóm các hình tương tự vào lớp mặt. Lớp mặt của cùng một nhân vật được hợp nhất, tạo nên *tập hợp dựa trên mặt*. Tập này dùng để định tên mỗi nhân vật, theo kích thước và vị trí trong video.
- Thông tin quan trọng về phim, như nhân đề, tên diễn viên chính... xuất hiện trong dây mở đầu. Vị trí của văn bản được xác định theo độ cách dòng, được tạo tính hay chuyển động trong các khung; có thể xác định theo chuẩn OCR.
- Dây các hoạt động nhìn chung được định tên từ rãnh âm thanh liên quan. Dây này có đặc tính audio đặc trưng (i) độ âm thanh chính; (ii) tần số; (iii) khoảng thời gian cho phép đưa vào/ làm mất dần dần. Các rãnh âm đặc trưng cho các kiểu khác nhau của các sự kiện được lưu trữ và có thể được dùng để so sánh khi xác định bản chất của dây hoạt động.

Lựa chọn đoạn video dùng cho trừu tượng tuân theo một số tiêu chuẩn: (i) văn bản nhan đề luôn có mặt trong trừu tượng cuối cùng; (ii) Tỷ lệ trừu tượng dùng cho dây hoạt động, hội thoại, và việc kết thúc được xác định theo tham số; nếu dài quá thì chọn theo các sự kiện; (iii) các sự kiện được trích ra từ 80% phim được dùng để đảm bảo nội dung.

Trừu tượng được thêm dần dần với mẫu video. Phim được chia thành các đoạn. Các mẫu video được chọn từ các đoạn với ít mẫu nhất, cho đến khi đạt được trừu tượng đủ dài.

Chuyển dịch giữa các các đoạn trừu tượng phụ thuộc vào bản chất của sự kiện. Người ta dùng: cắt cứng (hard cut) được dùng giữa các dây hoạt động và đoạn video; (ii) chuyển mềm được dùng với đoạn nhẹ nhàng, như đoạn hội thoại.

4.4.5. Đồ thị chuyển cảnh

4.4.5.1. Đồ thị

Liên kết về hướng đơn giản giữa hai nhóm (cluster) thể hiện *bờ cắt* (cut edge). Nếu liên kết này bị huỷ thì đồ thị chuyển cảnh sẽ được tách làm hai. Do vậy bờ cắt hướng đơn xác định biên giữa hai cảnh.

Khung đầu tiên của lia đầu tiên trong nhóm được dùng để thể hiện nhóm một cách có lược đồ trong đồ thị chuyển cảnh. Khung được thể hiện qua ô kích thước 40x30 điểm ảnh.

4.4.5.2. Duyệt các hình

Người dùng có thể lia vào bất kì ô nào để xem đoạn phim. Thuật toán dùng để nhận biết sự tương tự giữa các khung kề nhau sẽ nén và giải nén dữ liệu.

Các kĩ thuật trừu tượng hoá và chỉ số hoá trong hệ thống cơ sở dữ liệu đa phương tiện là các khía cạnh tách biệt, nhưng chung mục đích.

- Chỉ số là cấu trúc lưu trữ cho phép truy cập trực tiếp đến dữ liệu cần thiết;
- Trừu tượng là công cụ hiển thị, đảm bảo người dùng cơ sở dữ liệu định tên được video hay hình yêu cầu, hay định tên cảnh trong video... cần xem, thông qua phương tiện do hệ thống cung cấp.

4.5. Tìm thông tin đa phương tiện dựa trên nội dung

4.5.1. Giới thiệu về tìm thông tin đa phương tiện

Nghiên cứu về vấn đề tìm thông tin cơ sở dữ liệu trong hệ thống đa phương tiện bắt đầu từ đầu những năm 80 của thế kỉ XX, để tìm ảnh tĩnh; cần đến nhận dạng ảnh. Việc nghiên cứu tìm ảnh từ video đạt được nhiều kết quả vào những năm 90. Về việc nhận dạng tiếng nói và định danh âm nhạc từ các rãnh âm thanh còn đang tiến triển.

Tiếp cận bổ sung việc tìm các hình hiển thị từ các đồ hoạ và video có thể được phân loại thành ba phạm trù:

1. Dựa vào từ khoá, trong đó nội dung của hình ảnh được mô tả bằng chỉ số theo (i) từ khoá và (ii) trừu tượng. Phạm trù này thể hiện hiện trạng mà hầu hết các hệ thống quản trị cơ sở dữ liệu bình thường liên quan. Trong các hệ quản trị bình thường, các đối tượng dữ liệu đa phương tiện được coi như mục tin nguyên khối, có thể được mô tả theo nghĩa thuộc tính;
2. Dựa vào khía cạnh được rút ra tự động từ hình ảnh. Việc tìm hình ảnh dựa trên nội dung đạt được do nhận dạng các đối tượng tùy theo chỉ tiêu xác định nào đó, như so sánh các màu, mẫu, hình dạng; cho phép người dùng vạch ra hình ảnh rồi tìm các hình tương tự trong cơ sở dữ liệu thông qua (i) nhận thức được hướng chuyển động của đối tượng; (ii) định tên một đối tượng so với vị trí của đối tượng khác;
3. Dựa vào nội dung, trong đó việc diễn giải ngữ nghĩa của đối tượng được dùng. Chẳng hạn định tên đối tượng theo (tên, loại xe cộ...). Diễn giải ngữ nghĩa này của đối tượng thường được đưa vào thông qua việc chỉ số hoá.

Một số kết quả trong các hệ thống OVID, CORE, SCORE và infoscope. Hệ thống WebSEEK của Chang, 1997 là hệ thống nửa tự động, cho phép tìm, phân tích, phân loại và chỉ số hoá các hình hiển thị trên WWW.

Có nhiều nghiên cứu về cách tìm thông tin trên Internet. Tính tương tự giữa hệ thống tìm thông tin và hệ thống cơ sở dữ liệu đa phương tiện được ghi nhận và cách người dùng tìm thông tin đa phương tiện cũng như nhau. Cách tìm phụ thuộc vào (i) kinh nghiệm của người dùng; và (ii) bản chất của việc tìm. Người ta có khuynh hướng sử dụng một hay vài cách.

- *Tìm biên*, trong đó người dùng khai phá phần mở rộng của nội dung cơ sở dữ liệu, không sử dụng mẫu tìm kiếm; chính người dùng tạo nội dung, tức nêu ra bản chất của nội dung, để làm cơ sở cho lần tìm sau. Việc tìm kiếm không mang tính hệ thống và không có đích;
- *Duyệt*, trong đó người dùng có thông tin yếu về đích, tìm thông tin theo kinh nghiệm;
- *Tìm kiếm*, khi đó người ta đã có đích rõ ràng và xác định.

4.5.2. Lọc thông tin

Các lọc thông tin là cầu nối giữa người dùng và nguồn thông tin. Các bộ lọc trợ giúp tìm thông tin, qua được đồng thông tin từ nhiều hệ thống khác nhau.

Cơ chế lọc thông tin chọn các đối tượng được phân phát đến người dùng theo tiêu chuẩn đáng quan tâm. Cơ chế ra quyết định dựa vào tham số ban đầu do người dùng cung cấp để lọc và thông tin phản hồi từ người dùng của lần tìm kiếm trước. Phản hồi có thể ở dạng ẩn hay hiện. (i) Phản hồi hiện trả lời câu hỏi người dùng; (ii) phản hồi ẩn được suy ra từ các đối tượng mà người dùng thấy, và thời gian dùng để xem đối tượng.

Shoshana Loeb đã xác định các ý chính mà người phát triển cần biết khi lọc thông tin trong cơ sở dữ liệu đa phương tiện.

1. Hệ thống lọc được phát triển cho nguồn thông tin chuyên dụng và lớp người dùng chuyên, do không thể có lọc tổng quát;
2. Thời gian thiết kế và phát triển bộ lọc tùy thuộc vào thời gian sống của cơ sở dữ liệu, các báo cáo kĩ thuật.
3. Việc phân phát thông tin cũng là đặc tính của phương tiện và liên quan đến khả năng vùng đệm lưu trữ thông tin.
4. Thông tin lọc cần được kiểm tra để xác định loại thông tin đưa đến người dùng.

Người dùng có thể sử dụng hệ thống thường xuyên hay thi thoảng.

4.5.3. Hỏi dữ liệu đa phương tiện

Để tạo nên câu hỏi cơ sở dữ liệu, người dùng cần xác định (i) đối tượng dữ liệu cần tìm; (ii) bảng cơ sở dữ liệu cần trích dữ liệu; (iii) các vị từ tạo câu hỏi. Các câu hỏi cơ sở dữ liệu truyền thống được diễn tả ở dạng văn bản, thông qua ngôn ngữ hỏi, như ngôn ngữ chuẩn công nghiệp SQL. Các câu hỏi cơ sở dữ liệu đa phương tiện cần đến chức năng phụ trợ để tìm theo nội dung. Đề xuất mở rộng SQL, các ngôn ngữ hỏi theo nội

dung mới và ngôn ngữ hiển thị đang được đưa ra.

4.5.4. Tìm theo nội dung, sử dụng từ khoá

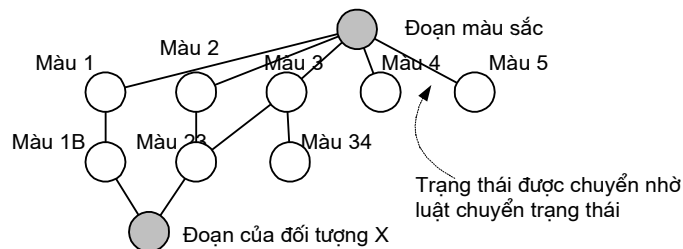
Nếu tìm hình ảnh dựa vào từ khoá thì hoặc (i) hình cần được chỉ số hoá; (ii) hay các đối tượng trong hình cần được tự động ghi nhận và có từ khoá tương ứng trong chỉ số. Tuy nhiên khi dùng chỉ số hoá thủ công hay tự động, các từ mô tả được thêm vào chỉ số có thể phù hợp hay không phù hợp với các từ khoá đã được người dùng sử dụng.

ONO đã đề nghị hệ thống dùng hai hệ thống chỉ số khác nhau để tìm hình trong cơ sở dữ liệu. Đó là chỉ số theo (i) khoá khái niệm và (ii) khoá miêu tả cảnh.

Chỉ số theo từ khoá khái niệm được tạo tự động với mô hình chuyển trạng thái. Bắt đầu từ màu cơ bản của đối tượng, các luật dựa trên đường viền ngoài, hướng trục chính...

- Nếu đoạn của hình đối tượng thoả mãn các chỉ tiêu về luật chuyển trạng thái, có thể chuyển sang trạng thái tiếp trên cây phân cấp.
- Nếu luật đối với hai đoạn được tích hợp thì tất cả đều được chuyển trên cây phân cấp.
- Khi không còn luật có thể sử dụng, người ta đã đến mức mịn nhất.

Theo mô hình trạng thái: (i) Trạng thái hiện tại trên cây trở thành điểm truy cập theo khoá khái niệm đối với cây chỉ số; mức ghi nhận tại thời điểm này chỉ là mức khái niệm cao nhất có thể đạt được; (ii) Nếu lỗi truy nhập được phát hiện tự động mà không phải là khoá người dùng đã chọn để tìm đối tượng, thì không tìm ra hình.



Hình. Tìm kiếm theo nội dung, dựa vào từ khóa, trên mô hình chuyển trạng thái

Với cách chỉ số dùng từ khoá miêu tả cảnh, các đặc tính của cảnh được xác định. Người ta có thể dùng đến 6 từ khoá, xác định vị trí ngang, dọc của đối tượng, vùng mà đối tượng chiếm, màu chính, hướng trục chính và hình dạng của đối tượng.

Các câu hỏi dựa trên loại từ khoá miêu tả gần giống với việc tìm kiếm các đối tượng, nhưng số phần trăm tìm được là thấp hơn.

4.6. Thí dụ về cơ sở dữ liệu đa phương tiện

4.6.1. Một số hệ thống

4.6.1.1. Hệ thống PICTON:

Tại trung tâm cao cấp để phân tích và nhận dạng tư liệu CEDAR tại đại học Newyork, các nghiên cứu viên đã phát triển hệ thống ghi nhận và chỉ số hoá ảnh người xuất hiện trong ảnh chụp tạp chí, như Srihari, 1995. Người ta dùng hai khía cạnh chính đối với xử lý (i) diễn giải việc nắm bắt và (ii) định vị hình ảnh mặt người trong ảnh.

Bước đầu tiên cần xác định tên tác giả, tức tên riêng PNC¹, như tên người... Sau đó xác định người đó là ai, chính xác hoá. Một vài hình có thể bị nhận dạng nhập nhằng. Tiếp sau, hình mặt người được định vị, hầu hết dựa trên nét đặc trưng, mạnh như mắt, mũi. Hệ thống xác định trên khuôn mặt chính diện.



Hình. Tìm kiếm trong PICTION



Hình. Tìm kiếm trong QBIC

4.6.1.2. Hệ thống QBIC

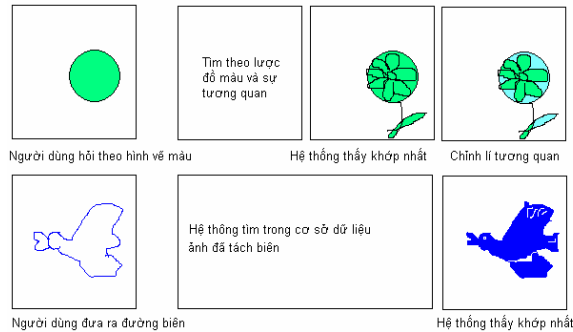
QBIC² là hệ thống do IBM phát triển để thể hiện một loạt các phương pháp tìm theo nội dung. Câu hỏi có thể dựa trên mẫu về màu và mẫu, khung xương, hình thí dụ, thông tin về chuyển động. Cơ sở dữ liệu QBIC chứa cả ảnh tĩnh lẫn video. Bình thường video được chia thành các lia, như giới thiệu trước đây.

Các đối tượng được tách khỏi nền nhờ kỹ thuật nửa tự động “Magic wand”, là kỹ thuật thông dụng trong bộ đồ hoạ. Người dùng chọn điểm ảnh riêng, sang điểm ảnh liền kề theo màu nằm trong ngưỡng, rồi mở rộng biên của vùng được chọn. Giá ngưỡng thay đổi theo các hình, theo môi trường cho phép chụp ảnh đối tượng.

Các đối tượng cũng có thể được xác định bằng cách đi theo lề; người dùng vẽ quanh đối tượng và đường cong sẽ tạo nên lề của đối tượng. Việc này thực hiện thời gian thực.

¹ proper noun complexe

² Query by Image content



Hình. Hỏi theo nội dung thông tin với ngôn ngữ QBIC

Các câu hỏi được thực hiện theo cách khớp các chuyển động của đối tượng, về hình dạng...

4.6.1.3. Hệ thống CVQL

CVQL là ngôn ngữ hỏi cho phép xác định câu hỏi theo thuật ngữ không gian, thời gian của các đối tượng trong video, thông qua các hàm xác định trước. 1996, Kuo và Chen sử dụng nó để tìm hình trong cơ sở dữ liệu video. Họ dựa trên việc tìm hình tùy theo các đặc tính không gian, thời gian. Video được nhìn theo cấu trúc phân cấp với phân loại video = {thể thao, xổ số, kinh tế}; thể thao = {bóng đá, tennis}...



Hình. Tìm kiếm trong CVQL

Một số hàm trong CVQL như (i) EXIST () cho phép xác định một đối tượng kí hiệu có xuất hiện trong khung cụ thể hay trong dãy các khung không; (ii) FP () cho biết vị trí của đối tượng kí hiệu trong khung; (iii) OM () xác định một đối tượng kí hiệu là tĩnh hay thay đổi vị trí trong dãy các khung; (iv) OOM () tính biên của miền chuyển động của đối tượng trong dãy các khung; (v) RP () cho biết vị trí tương đối giữa hai đối tượng; (vi) DISTANCE () có một số miêu tả, tùy vào ngữ cảnh được dùng, có thể dùng để xác định vị trí tương đối.

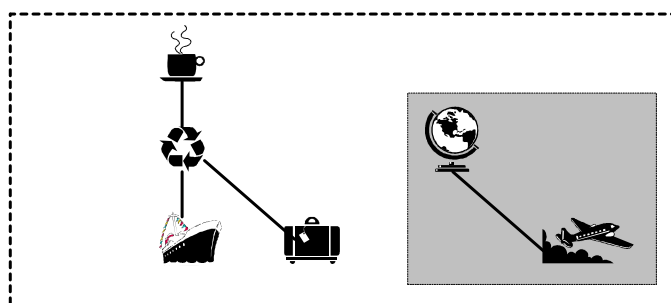
Các câu hỏi được thực hiện theo cách loại bỏ:

1. Trước tiên “bảng kí hiệu video” được dùng để loại bỏ đoạn video không xuất hiện các đối tượng kí hiệu.
2. Sau đó “bảng thời gian sống của đối tượng kí hiệu” được dùng để loại bỏ khung (trong video) không thấy đối tượng xuất hiện.
3. Cuối cùng, “bảng thông tin không gian, thời gian của đối tượng kí hiệu” được dùng để xác định khung đáp ứng yêu cầu, hay đưa ra câu trả lời.

4.6.1.4. Hệ thống VEVA

VEVA¹ do Golshani và Dimitrova, 1998 đề xuất như ngôn ngữ hỏi cho giao diện theo kí tự và hiển thị. Giao diện kí tự của VEVA là ngôn ngữ hỏi đại số; dữ liệu âm thanh được coi như kí hiệu một chiều, hình đồ hoạ được xem là hai chiều và video như đối tượng ba chiều. Ngôn ngữ xác định các phép sơ khởi để soạn thảo video, tìm thuộc tính video và thể hiện đoạn video. Ngoài ra còn có các phép toán không gian, thời gian. Họ đã nghiên cứu phương pháp phân tích thuộc tính và đường chuyển động của đối tượng trong video.

Giao diện hiển thị được xác định theo ngôn ngữ đại số. Các câu lệnh được cấu trúc theo các thể và biểu tượng. Người dùng được phép chọn các biểu tượng, thể hiện, các hành động và các liên kết.

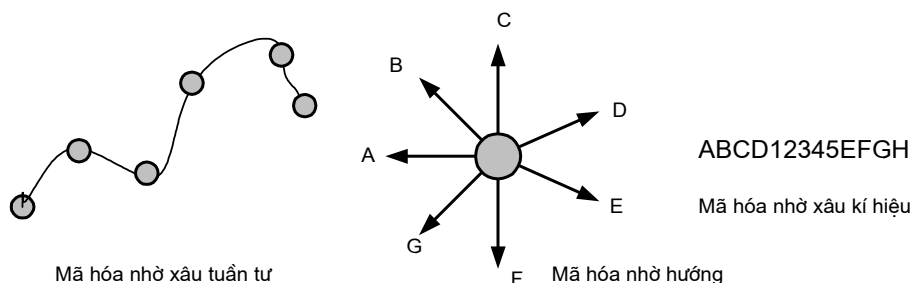


Hình. Câu hỏi hiển thị trong VEVA

4.6.1.5. Hệ thống V-QBE

Một tiếp cận tìm trên video, dựa theo chuyển động của các đối tượng trong video, được phát triển tại đại học Hiroshima. V-QBE cho phép người dùng tìm đoạn video theo khung xương hay đường thí dụ của chuyển động. Người dùng có thể xác định thông tốc độ và/hoặc kích thước của đối tượng.

Quĩ đạo của đối tượng được xác định theo tâm của đối tượng. Biên của đối tượng được xác định theo tính tương tự của giá trị màu của điểm ảnh liền kề. Vị trí của tâm đối tượng có thể vạch thành đường chuyển động của đối tượng. Hướng chuyển động được thể hiện như dây mã hoá.



Hình. thể hiện quỹ đạo

4.6.2. Tìm các đối tượng dựa trên hình dạng

Có một số điều liên quan tới việc định tên các đối tượng theo hình dạng. (i) Trước

¹ Visual extension to VARQA

tiên cách thể hiện hình dạng và cách thể hiện hình dạng được trích ra từ hình ảnh. (ii) Thứ hai, cách so sánh hai hình, và điều tạo nên tính tương tự. (iii) Thứ ba, cách tổ chức cơ sở dữ liệu các hình dạng để tìm kiếm hiệu quả.

- Các đối tượng có thể cố định theo hình dạng hay động dầy. Thí dụ căn nhà là tĩnh và hai đồ vật trong nhà có thể di chuyển.
- Các đối tượng trong hình có thể chiếm một phần hay toàn bộ miền thấy được.
- Việc trích rút hình của các đối tượng bị tác động của nguồn sáng, độ tương phản khác nhau giữa đối tượng và nền.
- Khía cạnh hiển thị ảnh hưởng đến việc nhận dạng đối tượng.

4.6.3. Thể hiện hình dạng

Trong FIBSSR¹, do Mehrotra và Gary đưa ra 1995, mỗi đối tượng được lưu trữ trong cơ sở dữ liệu sẽ được xử lí để xác định hình dạng. Kích thước của đối tượng là không quan trọng, do người ta dùng hình dạng để so sánh khớp.

- Hình dạng của đối tượng được thể hiện qua tập có thứ tự của biên điểm tọa độ.
- Điểm biên cũng là “điểm đáng quan tâm”;
- Hình của đối tượng được vạch ra bởi tập các điểm đáng quan tâm, xác định các điểm trên bờ của hình.
- Hình dạng của đối tượng được thể hiện theo “các tham số chuyển hoá”, được lưu trong chỉ số.

Việc tìm kiếm không cần dựa vào toàn bộ khung của hình ảnh, nhưng có thể dựa trên bộ khung về các khía cạnh của đối tượng. Việc chỉ số để tìm tất cả các hình thoả mãn các khía cạnh yêu cầu.

4.6.4. Việc khớp các hình

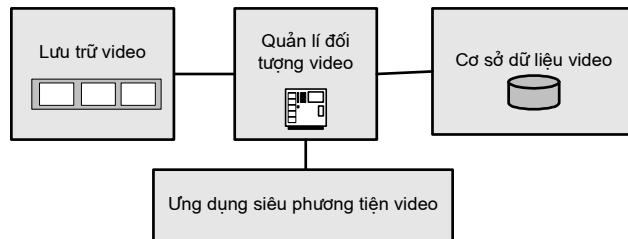
Năm 1996 Del Bimbo đã phát triển hệ thống khớp hình, cho phép người dùng vạch ra khung xương thể hiện hình dạng đối tượng muốn tìm. Rồi hệ thống tìm hình ảnh trong cơ sở dữ liệu khớp với hình được vạch. Người dùng có thể làm tinh việc tìm kiếm bằng cách lựa chọn hình ảnh khớp nhất với các đối tượng mong muốn.

Người dùng có thể sửa hình ảnh, huỷ các đối tượng không cần thiết, chọn một phần hình để tìm kiếm tiếp theo. Việc tìm có thể được lặp lại trên cơ sở khớp các hình ảnh các hình được lưu trữ.

4.6.5. Các liên kết video đa phương tiện

Năm 1996 Abe và Wakimoto đã phát triển hệ thống siêu phương tiện cho phép tìm theo nội dung thông tin trong cơ sở dữ liệu video. Để kiểm tra hệ thống, họ đã tạo “hệ thống học với trợ giúp của máy tính trên đa phương tiện” đối với khoa học không gian.

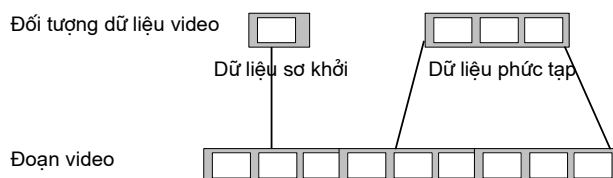
¹ feature index based similar shape retrieval



Hình. Môi trường đa phương tiện

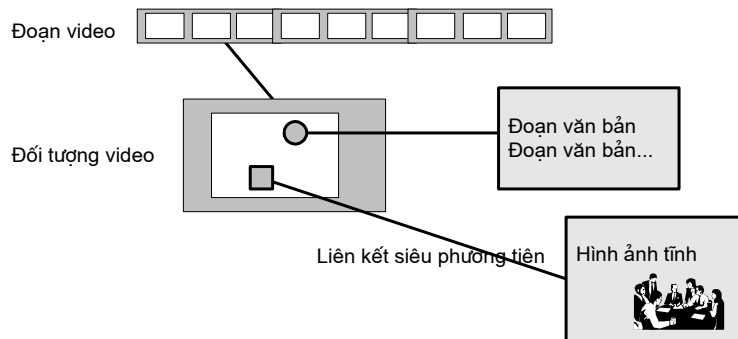
Trong hệ thống này, dữ liệu video thô được lưu trong “bộ nhớ video”. Các đoạn video có thể được trích từ bộ lưu trữ, lưu thành “cơ sở dữ liệu đối tượng video” với “quản trị đối tượng video”. Người phát triển ứng dụng siêu phương tiện có thể dùng “quản trị đối tượng video” để tìm và thực hiện các đối tượng dữ liệu video.

Cấu trúc dữ liệu video trong môi trường đa phương tiện này cho phép người ta trích thủ công các đoạn video để tạo nên các đối tượng dữ liệu video. Các thuộc tính, như thể hiện kích thước và tần suất khung, thừa kế... gắn với đoạn video do người phát triển xác định.



Hình. Các đối tượng dữ liệu video sơ khởi và phức tạp

Để tạo liên kết siêu phương tiện từ một video, người ta chọn một khung từ đoạn video. Các đối tượng trong khung có thể là mốc về hình dạng để liên kết với đối tượng khác. Người ta dùng hình chữ nhật hay hình tròn để tạo nên miền chú ý.



Hình. Các liên kết siêu phương tiện

4.7. Các ứng dụng của đa phương tiện

Với nhan đề ứng dụng của cơ sở dữ liệu đa phương tiện, phần này tập trung vào cơ sở dữ liệu ảnh. Nội dung về cơ sở dữ liệu đa phương tiện khác, như cơ sở dữ liệu video, cơ sở dữ liệu âm thanh sẽ được đề cập sau.

Trước tiên câu hỏi về các ảnh, hay hình ảnh, khác với câu hỏi trên văn bản. Cơ sở dữ liệu ảnh khác với cơ sở dữ liệu văn bản. Một ảnh có thể được mô tả tự động hay thủ

công, trước khi được lưu trong cơ sở dữ liệu.

Cũng như các dữ liệu đa phương tiện khác, người ta không thể dùng các mô hình dữ liệu truyền thống để mô tả và xử lý dữ liệu ảnh. Tuy nhiên người ta sử dụng 3 kỹ thuật cài đặt, cho phép xây dựng cơ sở dữ liệu ảnh (i) mở rộng mô hình quan hệ; (ii) cấu trúc dữ liệu n-chiều; (iii) chuyển hoá ảnh.

Khi làm việc với môi trường hệ thống máy tính, ngoài bộ xử lý trung tâm, hệ thống điều hành và môi trường cần được lưu ý. Môi trường trên máy vi tính với Intel thông dụng dùng Windows và UNIX.

Nhận xét về ứng dụng của cơ sở dữ liệu ảnh, người ta thấy (i) Trong nhiều năm qua, các cơ quan chính phủ và tổ chức công nghiệp đã có nhu cầu về lưu trữ và xử lý hình ảnh; (ii) Mô hình dữ liệu quan hệ cũng cho phép hỏi dữ liệu hình ảnh, theo trực tiếp ảnh, hay theo mức độ tương tự.

Phần này sẽ đề cập:

1. Các dữ liệu ảnh thô;
2. Chuyển hoá ảnh;
3. Kỹ thuật phân đoạn;
4. Tìm thông tin theo mức độ tương tự;
5. Chuyển hoá ảnh.



Hình. ảnh lưu trong cơ sở dữ liệu, tệp *.gif

4.7.1. Các hình ảnh thô

Nội dung của một ảnh gồm {các đối tượng trong ảnh | đối tượng phù hợp, hay tốt đối với ứng dụng}.

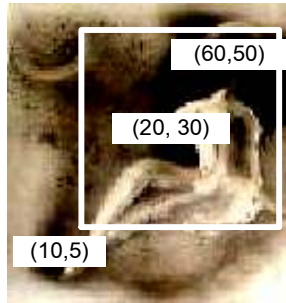
Các đối tượng có các thuộc tính¹:

1. Mô tả hình dạng: gồm hình dạng, vị trí của miền lưu đối tượng;
2. Mô tả thuộc tính: thuộc tính của điểm ảnh, pixel, đơn, hay nhóm các điểm ảnh. Chẳng hạn thuộc tính RGB² của điểm ảnh; độ xám của ảnh đen trắng. Gắn với điểm ảnh cụ thể là một tế bào³, gồm các hình chữ nhật ứng với các điểm ảnh.

Thí dụ xem ảnh, người ta thấy hình chữ nhật cho phép mô tả hình dạng ảnh. $XLB =$

¹ property
² red green blue
³ cell

10, XUB = 60, YLB = 5, YUB = 50. Mô tả thuộc tính bằng mức độ màu, chẳng hạn điểm ảnh (20, 30) có giá trị RGB là đỏ = 5, lục = 7, xanh = 6.



Hình. ảnh với các toạ độ

Nhìn chung, thay vì xét ảnh có $a \times b$ điểm ảnh, người ta có thể xét ảnh với $m \times n$ tế bào, trong đó $a \bmod m = 0$; $b \bmod n = 0$; $m < a$; $n < b$. Điều này có nghĩa mỗi tế bào thể hiện $a \div m \times b \div n$ điểm ảnh. Do vậy tế bào = {điểm ảnh}.

Định nghĩa: Mỗi ảnh I có cặp số (m, n) đi kèm gọi là độ phân giải¹ của ảnh. Việc chia ảnh thành $m \times n$ tế bào kích thước như nhau gọi là chia lưới ảnh².

Mỗi tế bào trong ảnh I chia lưới $m \times n$ sẽ gồm nhiều điểm ảnh.

Định nghĩa: thuộc tính của tế bào là bộ ba (tên, giá trị, phương pháp); trong đó tên là xâu kí tự cho biết tên của thuộc tính; giá trị là tập các giá trị mà tế bào có thể nhận; phương pháp là thuật toán, hay thủ tục cho phép tính toán theo yêu cầu của thuộc tính.

Thí dụ:

- Hình ảnh đen trắng, tế bào có thuộc tính tế bào (màu đen trắng, {đ, t}, f). Phương pháp f cho biết khi nhập tế bào, cho ra đ (đen) hay t (trắng) theo cách kết hợp đen/ trắng của điểm ảnh trong tế bào;
- Hình ảnh xám: nếu qui định trắng ~ 0 , đen ~ 1 thì tế bào (mức xám, $[0, 1]$, f); f cho biết mức xám.

Tế bào có thể được đặc trưng theo đường biên của tế bào, như cây R, thông qua XLB, XUB, YLB và YUB.

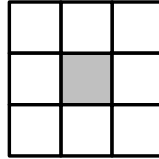
Thí dụ tế bào được mô tả qua tập các điểm ảnh $\{(i, j) \mid XLB \leq i \leq XUB, YLB \leq j \leq YUB\}$. Một trong 4 phương pháp tính theo trung bình $f(\text{tế bào}) = (\sum \sum \text{xám}(i, j), XLB \leq i \leq XUB, YLB \leq j \leq YUB) / ((XUB - XLB) \cdot (YUB - YLB))$, trong đó xám (i, j) = mức xám của điểm ảnh (i, j) .

Nhiệm vụ của kĩ sư phần mềm là:

- Xác định thuộc tính tế bào;
- Các phương pháp để thoả mãn thuộc tính.

¹ grid resolution
² image grid

Định nghĩa: Hình dạng đối tượng¹ là tập P bất kì các điểm, mà nếu $p, q \in P$ thì tồn tại dãy điểm p_1, \dots, p_n trong P mà: (i) $p = p_1, q = p_n$; (ii) $\forall i: 1 \leq i < n, p_{i+1}$ là láng giềng p_i , tức nếu $p_i = (x_i, y_i), p_{i+1} = (x_{i+1}, y_{i+1})$ thì (x_{i+1}, y_{i+1}) là 8 trường hợp $\{(x_i \pm 1, y_i), (x_i, y_i \pm 1), (x_i \pm 1, y_i \pm 1)\}$. Trong trường hợp này, mỗi đối tượng khớp với hình chữ nhật.



Hình. Các giá trị (x_{i+1}, y_{i+1}) kề (x_i, y_i)

Định nghĩa: Hình chữ nhật (rectangle) là dạng đối tượng P mà tồn tại XLB, XUB, YLB, YUB để $P = \{(x, y) \mid XLB \leq x < XUB, YLB \leq y < YUB\}$.

Định nghĩa: cơ sở dữ liệu ảnh IDB^2 là bộ ba $(GI, \text{thuộc tính}, Rec)$; trong đó GI là tập các hình chia ô có dạng (hình, m, n), tức tên hình và giá trị chia ô m, n ; thuộc tính là tập các thuộc tính tế bào; Rec là ánh xạ, từ mỗi ảnh sang tập các đối tượng theo hình chữ nhật.

Khi thể hiện dữ liệu ảnh, người ta cần tính đến hai nhân tố:

1. Các hình ảnh thường là các đối tượng rất lớn, $p_1 \times p_2$ ma trận điểm ảnh. Việc lưu thuộc tính theo điểm ảnh là không khả thi. Do vậy cần có thuật toán nén ảnh để giảm số các điểm ảnh;
2. Cho hình I , ở dạng nén hay dạng thô. Người ta cần xác định có khía cạnh nào trong ảnh. Do vậy cần chuyển ảnh thành tập các vùng chữ nhật đồng nhất, gọi là đoạn³. Quá trình này gọi là phân đoạn.

Khi ảnh được phân đoạn, người ta có nhu cầu khớp một đoạn hay nhiều đoạn, tức toàn bộ ảnh.

4.7.2. Thể hiện ảnh đã nén

4.7.2.1. Tạo ảnh nén

Giả sử I là ảnh 2 chiều, $2D^4$, có $p_1 \times p_2$ điểm ảnh. Vậy con số $I(x, y)$ cho biết thuộc tính về điểm ảnh tại (x, y) . Chẳng hạn $I(x, y) = 0..255$. Tuy nhiên p_1, p_2 nhận giá trị đến 1024 độ phân giải cao, tức mịn hơn, và có hàng triệu lổ truy cập ma trận I . Người ta thể hiện I dưới dạng nén $cr(I)^5$.

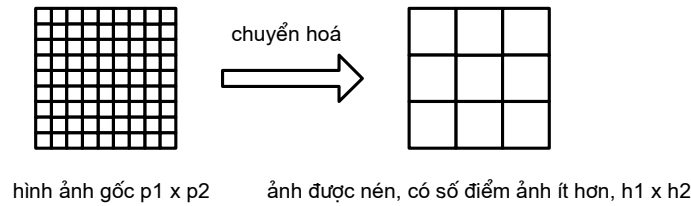
¹ object shape

² image database

³ segment

⁴ two dimension

⁵ compress representation



Hình. Nén ảnh

Việc tạo ảnh nén được thực hiện theo hai bước:

1. Chọn kích thước:

- Kích thước h của ảnh nén được người thiết kế cơ sở dữ liệu lựa chọn;
- Kích thước càng lớn, độ tin cậy càng cao, nhưng độ phức tạp tăng lên;
- $Cr(I) = (h_1, h_2)$;

2. Chọn phương pháp chuyển hoá:

- Cần cho biết ánh xạ $(I, i, j) \rightarrow Cr(i, j)$; $i \in [1, h_1]$; $j \in [1, h_2]$;
- Người ta sử dụng hai cách biến đổi phổ biến là (i) biến đổi Fourier DFT^1 ; và biến đổi cosin DCT^2 .

4.7.2.2. Phương pháp biến đổi Fourier

Phương pháp chuyển Fourier là cách chuyển ảnh thông dụng, theo công thức

$$DFT(x, y) = \frac{1}{p_1 \cdot p_2} \sum_{a=0}^{p_1-1} \sum_{b=0}^{p_2-1} I(a, b) \cdot e^{-j(\frac{2\pi \cdot x \cdot a}{p_1} + \frac{2\pi \cdot y \cdot b}{p_2})}, \text{ trong đó } j \text{ là số phức, } j = \sqrt{-1}.$$

Thuộc tính của DFT là:

- Cho phép tìm lại được ảnh I ban đầu, tức $I(p_1, p_2)$, thông qua $DFT^{-1}(I)$. Điều này rất tiện lợi;
- Không phải tất cả các ảnh nén đều cho phép quay về ảnh cũ, do quá trình nén sẽ lược bỏ một số thông tin không quan trọng;
- Có thể áp dụng DFT với một số phép toán không cần thiết quay trở về ảnh ban đầu.

4.7.2.3. Phương pháp biến đổi cosine

Thuật toán biến đổi cosine được nhiều người biết, cho phép tạo ra ảnh chuẩn JPEG³. Tỷ lệ nén ảnh theo phương pháp này cao, mà không làm mất nhiều thông tin về ảnh. Công thức chuyển của phương pháp này:

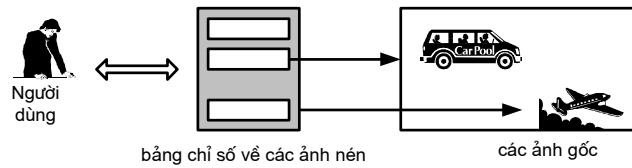
$$DCT(i, j) = \frac{1}{\sqrt{p_1 \cdot p_2}} \alpha(i) \cdot \beta(j) \cdot \sum_{r=0}^{p_1-1} \sum_{s=0}^{p_2-1} \cos \frac{(2r+1) \cdot \pi \cdot i}{2r} \cdot \cos \frac{(2s+1) \cdot \pi \cdot j}{2s}; \text{ trong đó}$$

$$\alpha(i), \beta(j) = \frac{1}{\sqrt{2}} \text{ hay } 1.$$

¹ discrete Fourier transform

² discrete cosine transform

³ joint photographic experts group



Hình. Chỉ số về ảnh nén

Người ta nhận xét cả hai phương pháp DFT và DCT đều có công thức cải tiến, cho phép trợ giúp các yêu cầu riêng, nhằm tăng tốc độ nén ảnh hoặc tăng chất lượng ảnh. DCT có vài ưu điểm trội so với DFT.

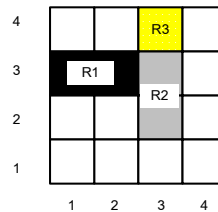
Khi có nhiều ảnh nén $\{cr(r_1), \dots, cr(r_n)\}$, trong đó r_i là miền đang quan tâm, người ta dùng kĩ thuật chỉ số, chẳng hạn cây R, để tìm lại ảnh gốc.

4.7.3. Xử lí ảnh thông qua việc phân đoạn ảnh

Giả sử trong hoàn cảnh lí tưởng, người ta có miền đang xem xét là một ảnh; người ta có nhu cầu xét chi tiết, phân đoạn miền này. I là ảnh $m \times n$ tế bào. Trong trường hợp không thuận lợi, có thể tế bào chính là điểm ảnh, nhưng thông thường tế bào là hình chữ nhật gồm các điểm ảnh.

Định nghĩa: Miền liên thông R trong ảnh I là tập các tế bào: nếu $(x_1, y_1), (x_2, y_2) \in R$ thì tồn tại dãy các tế bào c_1, \dots, c_n trong R để (i) $c_1 = (x_1, y_1)$; (ii) $c_n = (x_2, y_2)$; (iii) Khoảng cách Oclit giữa tế bào c_i, c_{i+1} là 1, $\forall i = 1..n$.

Thí dụ ba miền R_1, R_2, R_3 , là ba miền liên thông. Ngoài ra các miền liên thông khác là $R_1 \cup R_2, R_2 \cup R_3, R_1 \cup R_2 \cup R_3$. R_1 không liên thông R_3 , do khoảng cách giữa tế bào (2, 3) đến tế bào (3, 4) là $\sqrt{2} > 1$.



Hình. Ba miền

Định nghĩa: vị từ đồng nhất¹ gắn với hình I là hàm H , cho phép gắn miền R với hình I , cho giá trị true/false.

Thí dụ về vị từ 1:

Cho δ là số thực $[0, 1]$. Xét hình đen trắng. Trên hình, người ta xác định vị từ đồng nhất:

$H_\delta^{\text{đen trắng}} = \text{true}$ nếu $(100 \cdot \delta) \%$ các tế bào R cùng màu; ngược lại false.

Miền	Số các điểm ảnh đen	Số các điểm ảnh trắng
------	---------------------	-----------------------

¹ homogeneity predicate

R_1	800	200
R_2	900	100
R_3	100	900

Hình. Mô tả ba miền

Miền	Vị từ $H^{\text{đen trắng}}, \delta = 0.8$	Vị từ $H^{\text{đen trắng}}, \delta = 0.89$	Vị từ $H^{\text{đen trắng}}, \delta = 0.92$
R_1	True	False	False
R_2	True	True	False
R_3	True	True	false

Hình. Xét sự cùng màu đối với các miền, theo các một vài vị từ

Thí dụ về vị từ 2: giả sử ảnh là toàn bộ miền, mỗi điểm ảnh gắn với giá trị thực thuộc $[0, 1]$. Giá trị này được gọi là mức đen trắng, gần 0 là trắng, gần 1 là đen. Người ta xác định vị từ f : tế bào $\rightarrow [0, 1]$. Hàm H sử dụng hệ số ồn η , $0 \leq \eta \leq 1$, và sử dụng δ như trong thí dụ trên. Vị từ đồng nhất:

$H^{f, \eta, \delta}(R) = 1$ nếu số các (x, y) mà $|\text{mức đen trắng}(x, y) - f(x, y)| < \eta$ là lớn hơn $\delta X(n.m)$, tức là đường cơ bản f^1 mà vị từ đồng nhất sử dụng cho phép ồn cực đại. Sự tương tự được xác định qua f , tức $|\text{mức đen trắng}(x, y) - f(x, y)| < \eta$.

Khi có đủ nhiều các tế bào, do δ xác định, trong miền ảnh khớp với f tiên đoán, thì miền R được xem là đồng nhất, tức hàm H nhận giá trị true.

Thí dụ về vị từ 3: Trước tiên người ta phân loại các mức đen trắng, chẳng hạn mức với giá trị đen trắng $[0, 0.1]$ ứng với lớp 1, $[0.1, 0.2]$ ứng với 2,... đến 10. Cho $\delta \in [0, 1]$. Xác định hàm $H^{\text{lớp}}(R) = \text{true}$ nếu quá $(100 * \delta) \%$ các tế bào trong R rơi cùng vào một lớp.

Thí dụ về vị từ 4: Phát triển vị từ trong thí dụ trên, $H^{\text{lớp}}$, người ta xác định vị từ động, kí hiệu là $H^{\text{động}}$, do vị từ $H^{\text{lớp}}$ chỉ xác định các lớp theo giá trị ưu tiên, không linh động. Vị từ động cho rằng R là đồng nhất là tùy vào vị từ $H^{\text{động}, \eta, \delta} = \text{true}$. Vị từ là đúng khi trên $(100 * \delta) \%$ các tế bào trong miền R nằm trong phạm vi $[\eta, r]$, trong đó r là số thực được xác định khi thực hiện.

Các vị từ đồng nhất không nhất thiết dựa trên các mức đen trắng, hay mức xám. Người ta xác định tùy ý vị từ theo tính toán thu được. Ngoài các nhân tố quyết định H , còn một vài nhân tố, như độ sâu, độ đậm, lược đồ màu, nền... của ảnh.

Định nghĩa: Phân đoạn². Cho ảnh I , là tập các điểm ảnh $m \times n$. Phân đoạn của ảnh I ứng với vị trí đồng nhất, là tập các miền R_1, \dots, R_k :

- $R_i \cap R_j = \text{rỗng}, 1 \leq i \neq j \leq k$;
- $I = \bigcup R_i$ theo i ;
- $H(R_i) = \text{true} \forall i \neq 1..k$;
- $\forall i \neq j, i, j = 1..n: R_i \cup R_j$ là liên thông thì $H(R_i \cup R_j) = \text{false}$.

Thí dụ theo vị từ đồng nhất động $H^{\text{động}, 0.03, 1}$, miền R trong ảnh ma trận trên là đồng

¹ base line
² segmentation

nhất nếu $\exists r: \forall$ tế bào trong R có mức đen trắng v: $|v-r| \leq 0.03$. Tùy theo việc phân loại, người ta thấy có 5 miền ứng với các đoạn đúng trong ảnh trên.

$$R_1 = \{(1, 1), (1, 2)\}$$

$$R_2 = \{(1, 3), (2, 1), (2, 2), (2, 3)\}$$

$$R_3 = \{(3, 1), (3, 2), (3, 3), (4, 1), (4, 2)\}$$

$$R_4 = \{(3, 4), (4, 3), (4, 4)\}$$

$$R_5 = \{(1, 4), (2, 4)\}$$

4	0.60	0.63	0.85	0.90
3	0.35	0.30	0.55	0.80
2	0.05	0.30	0.60	0.60
1	0.10	0.25	0.50	0.50
	1	2	3	4

Hình. Ảnh ma trận thí dụ

Sử dụng vị từ đồng nhất, người ta có thể phát hiện được các đoạn của một ảnh, và thực hiện một số công việc đơn giản, liệt kê dưới đây:

1. Tách¹: từ một ảnh chiếm toàn miền, nếu hàm đồng nhất thấy miền đồng nhất, tức giá trị true, thì cả ảnh là một đoạn. Ngược lại một ảnh được phân đoạn ra các R_1, \dots, R_n , là các miền đồng nhất và thỏa mãn các điều kiện của miền đồng nhất, trừ điều kiện thứ tư. Như vậy người ta tách được các đoạn trong ảnh;
2. Hợp nhất²: sau khi ảnh có các đoạn $\{R_i\}$, người ta có thể kiểm tra để cho phép R_i nhập lại với R_j nào. Sau thao tác nhập một số miền có thuộc tính gần nhau, người ta thu được các R'_1, \dots, R'_k của ảnh, $k \leq n$. Miền $R'_i = \cup R_j$. Công việc này đơn giản, tùy theo người dùng xác định sự gần nhau; có thể được chương trình hoá;

4.7.4. Tìm kiếm dựa trên sự tương tự

Từ một ảnh, người ta có thể thu được ảnh nén nhờ phương pháp DFT hay DCT. Ngoài ra phương pháp sóng nhỏ³ cũng cho phép thực hiện nén ảnh. Sau khi được phiên bản ảnh nén, việc xác định sự giống nhau giữa các ảnh vẫn là nhu cầu thông thường. Người ta đề xuất hai tiếp cận, cho phép tìm kiếm theo mức độ tương tự. Đó là (i) tiếp cận độ đo⁴ cho phép hai ảnh gần nhau nếu có độ đo như nhau; (ii) tiếp cận chuyển hoá. Tiếp cận độ đo sử dụng sự tương tự như đối tượng không đổi, nên có thể dùng nó làm chỉ số tìm kiếm ảnh. Do vậy các ứng dụng khác nhau cần có khái niệm tương tự khác nhau, tùy theo người dùng.

4.7.4.1. Tiếp cận độ đo

Giả sử có tập O các đối tượng, với các thuộc tính điểm ảnh p_1, \dots, p_n . Một đối tượng o có thể được xem như tập S (o) của $n+2$ bộ, có dạng (toạ độ x, toạ độ y, v_1, \dots, v_n); trong đó v_i là giá trị thuộc tính p_i , gắn với điểm ảnh tại (x, y).

¹ split
² merge
³ wavelet
⁴ metric

Một đối tượng là toàn hình hay là một đoạn trong ảnh, chẳng hạn đối tượng là cả người được chụp hay chỉ là khuôn mặt trong ảnh cả người.

Theo tiếp cận độ đo, hai đối tượng là xa nhau được gắn với khoảng cách hai đối tượng. Khoảng cách d (tập X đến $[0, 1]$) là hàm khoảng cách nếu thỏa mãn các tiên đề $\forall x, y, z \in X$:

1. $d(x, y) = d(y, x)$;
2. $d(x, y) \leq d(x, z) + d(z, y)$;
3. $d(x, x) = 0$.

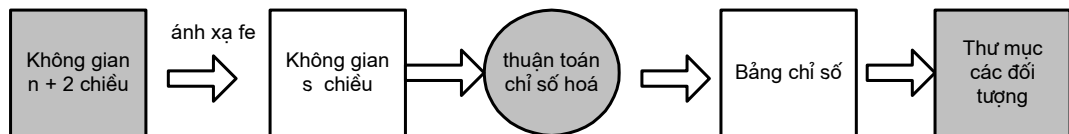
Một số kí hiệu: d_O là hàm khoảng cách trong không gian của tất cả các đối tượng. Một đối tượng ứng với một điểm trong không gian với chiều $k = n + 2$. Việc tính d_O phức tạp hơn tính khoảng cách thường.

Thí dụ: O gồm 256×256 hình với 3 thuộc tính (đỏ, lục, xanh). Mỗi thuộc tính có giá trị 0..7. $D(O_1, O_2) = \sqrt{\sum_{i=1}^{255} \sum_{j=1}^{255} (d_1(i, j) + d_2(i, j) + d_3(i, j))}$, trong đó khác nhau về đỏ d_1

$(i, j) = (O_1[i, j].\text{đỏ} - O_2[i, j].\text{đỏ})^2$; khác nhau về xanh $d_2(i, j) = (O_1[i, j].\text{xanh} - O_2[i, j].\text{xanh})^2$; khác nhau về lục $d_3(i, j) = (O_1[i, j].\text{lục} - O_2[i, j].\text{lục})^2$;

Số lượng theo các điểm ảnh sẽ rất lớn, nên người ta sẽ gặp khó khăn khi dùng cấu trúc dữ liệu đa phương tiện như cây tứ phân hay cây R. Do vậy người ta thay đổi dạng hàm f với các khía cạnh tốt để được fe, chẳng hạn sử dụng hàm có cả DFT lẫn DCT. Trên không gian s chiều, $s < n + 2$, người ta đạt được các tiết kiệm sau:

- Khi đối tượng o là điểm trong không gian $n+2$ chiều, thì fe (o) là điểm đơn;
- Fe (o) là điểm trong không gian s chiều $s \ll n + 2$.



Hình. Dữ liệu trong qui trình xử lí dữ liệu đa phương tiện

4.7.4.2. Tiếp cận chuyển hoá

Tiếp cận chuyển hoá để tìm các đối tượng tương tự được xem là tổng quát hơn tiếp cận độ đo. Nguyên tắc của tiếp cận chuyển hoá là sự khác nhau giữa hai đối tượng o_1, o_2 tỉ lệ thuận với chi phí, gọi là giá, để chuyển hoá đối tượng o_1 trở thành o_2 và ngược lại.



Hình. Một bức ảnh được (i) làm nhoè; (ii) làm rõ đường viền; (iii) biến dạng

Người ta xây dựng tập các toán tử chuyển hoá to_1, \dots, to_r . Chẳng hạn một ảnh được di chuyển, quay với góc α , co giãn với tỉ lệ β . Toán tử mở rộng được xác định qua việc bổ sung các chi tiết mới vào đối tượng; hay toán tử cắt xén cho phép lấy ra một ảnh từ một ảnh gốc.

Thí dụ sử dụng phần mềm vẽ, có toán tử vẽ (màu, giá trị 1, màu 2, giá trị 2). Toán tử này được dùng trong thao tác, có giá là giá (vẽ) = sai khác (mau1, mau2)³ + gt₁ - gt₂)². chẳng hạn có sai khác (đỏ, lục) = 3; sai khác (lục, đỏ) = 3...

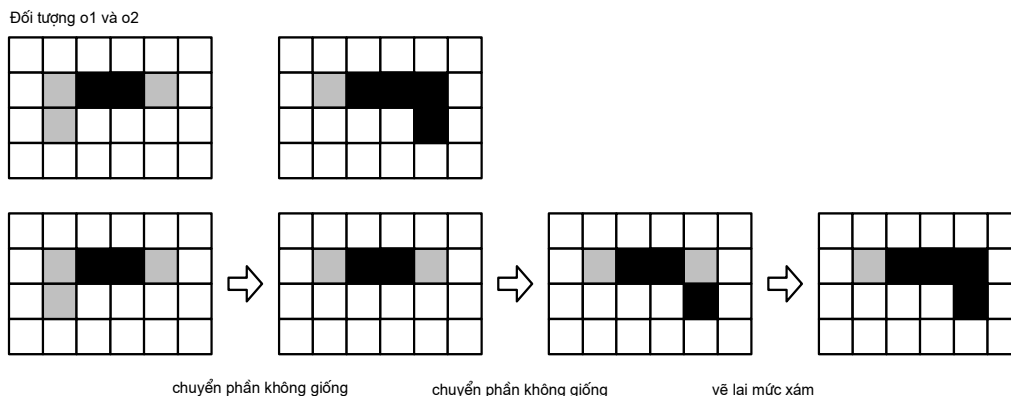
Việc chuyển đổi tượng o_1 sang o_2 được thực hiện qua một dãy các phép chuyển, tác động lên nhiều đối tượng:

- $To_1(o) = o_1$;
- $To_i(o_{i-1}) = o_i$;
- $To(o_r) = o_2$.
- Giá chung = giá (một loạt các phép chuyển TS) = $\sum \text{giá}(to_i)$; $i = 1..n$.

Giả sử có dãy các phép chuyển TS¹, TS (o_1, o_2) cho phép chuyển đổi tượng o_1 sang o_2 . Khác nhau giữa o_1 và o_2 , kí hiệu d (o_1, o_2), ứng với tập các toán tử chuyển hoá và các hàm áp giá:

$$D(o_1, o_2) = \min \{ \text{giá}(TS) \mid TS \in TS(o_1, o_2) \cup TS(o_2, o_1) \}.$$

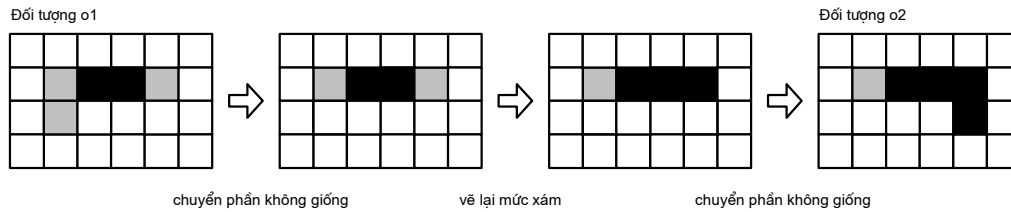
Thí dụ: Có hai đối tượng o_1, o_2 . Mỗi đối tượng có hai mức xám khác nhau, hoặc hai màu. Yêu cầu chuyển hoá o_1 để có hình dạng giống o_2 . Sau khi chuyển các ô chiều dọc, chiều ngang không đổi, người ta cần vẽ lại mức xám cho ô phát triển.



Hình. Chuyển đổi tượng o_1 thành o_2

¹ Transform sequence

Tuy nhiên, tùy theo thói quen, hay giá thực hiện toán tử mà người ta có thể thực hiện các phép khác để chuyển o_1 thành o_2 .



Hình. Phương án khác

Một số nhận xét:

- Mô hình chuyển hoá cho phép đạt được độ mềm dẻo so với mô hình độ đo ở chỗ:
 - Người dùng có thể thiết lập kí pháp riêng về sự tương tự bằng cách xác định toán tử chuyển hoá, hay không dùng toán tử chuyển hoá;
 - Người ta có thể liên kết hàm giá với toán tử chuyển, dựa trên loại toán tử và các tham số;
- Mô hình độ đo có hai ưu điểm sau:
 - Chỉ cần dùng một độ đo. Người ta có thể dùng chỉ số để tối ưu, và sử dụng khái niệm gần hay xa nhất;
 - Dễ tính được hiệu quả của quá trình.

4.7.5. Tổng quát về cơ sở dữ liệu ảnh

Cho đến lúc này, người dùng vẫn chưa giải đáp được câu hỏi về nhu cầu thể hiện ảnh và lưu trữ các ảnh, rồi mới có thể thực hiện các phép tìm kiếm ảnh.

Có nhiều cách thể hiện cơ sở dữ liệu ảnh:

- Thể hiện cơ sở dữ liệu ảnh với các thuộc tính trong bảng quan hệ của mô hình cơ sở dữ liệu quan hệ;
- Thể hiện cơ sở dữ liệu ảnh với các cấu trúc dữ liệu không gian. đây là ý tưởng đơn giản, sử dụng các khái niệm về đường, điểm, hình...;
- Thể hiện qua các phép chuyển hoá. Với ý tưởng này, người ta sử dụng ảnh độ phân giải 1024×1024 điểm ảnh. Do số điểm ảnh lớn, các kĩ thuật nén là cần thiết; việc nén được đặc trưng bằng vecto kích thước k , k không quá 100.

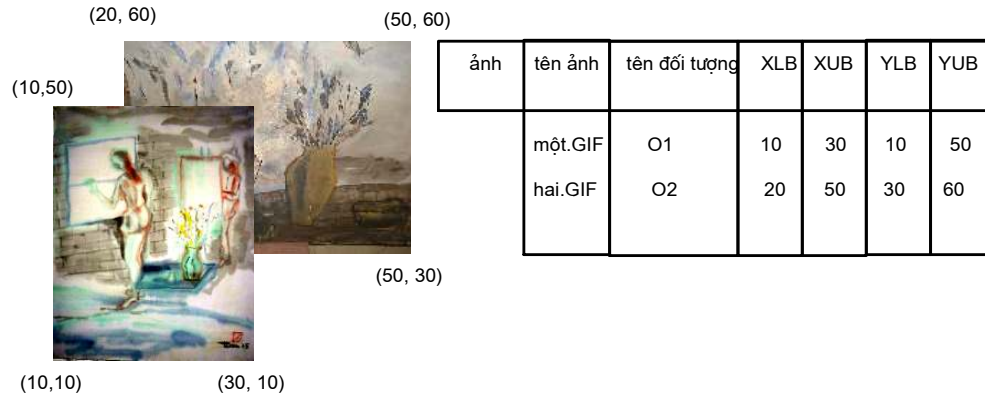
Khi tìm kiếm ảnh, các giá trị đặc tính về ảnh phụ thuộc vào các nhân tố khách quan như ánh sáng, vị trí đặt máy chụp, máy quay, chủ đề quan tâm...

Một số phần mềm xử lí ảnh đã thông dụng. Người ta có thể truy cập mạng máy tính để tham khảo và thử nghiệm.

4.7.6. Thể hiện cơ sở dữ liệu ảnh nhờ mô hình quan hệ

4.7.6.1. Lược đồ

Phân trên đã xác định cơ sở dữ liệu ảnh, là bộ IDB = (tập ảnh, thuộc tính, ánh xạ Rec ảnh sang tập hình chữ nhật). Bất kì bộ cơ sở dữ liệu nào đều được thể hiện qua mô hình quan hệ.



Hình. ảnh thí dụ và bộ dữ liệu trong bảng quan hệ "ảnh"

Trước tiên, thiết lập bảng quan hệ ảnh, với lược đồ quan hệ *ảnh* (*tên hình*, *tên đối tượng*, *XLB*, *XUB*, *YLB*, *YUB*); trong đó các giá trị cận thể hiện hình chữ nhật bao, R (*XLB*, *XUB*, *YLB*, *YUB*). Nếu $R \in \text{Rec}(I)$ thì tồn tại bộ (*I*, *tên mới*, *XLB*, *XUB*, *YLB*, *YUB*) trong quan hệ ảnh. Việc tạo ra quan hệ ảnh xuất phát từ phép tạo các đoạn, phân đoạn ảnh.

Đối với các thuộc tính $p \in \text{Thuộc tính}$, người ta tạo quan hệ R_p với lược đồ

R_p (*tên tệp ảnh*, *XLB*, *XUB*, *YLB*, *YUB*, giá trị thuộc tính p);

Trong đó các giá trị biên hình bao mô tả tế bào hình chữ nhật. Thuộc tính có ba dạng:

1. Thuộc tính mức điểm ảnh: thuộc tính như màu RGB của các điểm ảnh độc lập;
2. Thuộc tính mức đối tượng, hay mức miền: một vài đối tượng trong hình có thuộc tính riêng. Chẳng hạn O_1 của ảnh một.gif có thuộc tính tên, tuổi, chiều cao...;
3. Thuộc tính mức ảnh: thuộc tính khi đã nắm được, các quyền truy cập cơ sở dữ liệu...

Cho thuộc tính $p \in \text{Thuộc tính}$, người ta có thể thể hiện tường minh thuộc tính này như quan hệ R_p . Thực tế cho thấy việc thể hiện tường minh mỗi giá trị điểm ảnh như một bộ của quan hệ gây khó hiểu. Do vậy người ta chỉ thể hiện tường minh đối với (i) đối tượng, hay miền; (ii) mức ảnh, như dữ liệu quan hệ.

4.7.6.2. Hỏi dữ liệu trên cơ sở dữ liệu ảnh theo mô hình quan hệ

Do đã lưu trữ các đối tượng trong các bảng quan hệ, việc hỏi dữ liệu theo ngôn ngữ SQL không mấy khó khăn. Chẳng hạn tìm người và ảnh theo tên:

Select tên ảnh

From quan hệ ảnh I; quan hệ tên N

Where I.tên đối tượng = N.tên đối tượng and N. tên = "toto"

Lưu ý rằng đối với các dữ liệu đa phương tiện, so sánh khớp đúng có thể được xem như gần đúng, xấp xỉ. Để đạt độ chính xác cao, việc so sánh cần thực hiện theo thuật toán đặc biệt, được nhúng vào hệ thống chung.

Giả sử mỗi đối tượng có các giá trị thuộc tính $p \in \text{Thuộc tính}$. Vậy mỗi quan hệ R

liên kết với các thuộc tính theo mức đối tượng, hay miền, ảnh, có 2 cột (i) cột tên, cho biết tên đối tượng hay tên miền, ảnh; (ii) cột giá trị thuộc tính. Chẳng hạn người ta có quan hệ R có dạng:

R	Tên đối tượng	Tên
	O ₁	Toto
	O ₂	Tutu
	O ₃	Titi

Hình. Bảng quan hệ thí dụ

1. Vai trò của xác suất. Để xác định sự giống nhau, hai đối tượng ảnh sẽ mang thêm thuộc tính có giá trị xác suất, dùng trong việc so sánh khớp.

R	Tên đối tượng	Tên	Xác suất
	O ₁	Toto	0.8
	O ₂	Tutu	0.7
	O ₃	Titi	0.7

Hình. Cột Xác suất cho phép đánh giá tin cậy khi so sánh

Trong quan hệ có giá trị xác suất này, người ta đặt câu hỏi “tìm ảnh giống Mơ và Mận” sẽ thu được nhiều tệp ảnh *.gif.

Việc dùng giá trị xác suất theo lý thuyết xác suất cũng như dùng độ tin cậy trong lý thuyết chắc chắn, hay dùng giá trị mờ... đều nhằm xử lý các đối tượng không tất định. Đối với dữ liệu đa phương tiện, việc tìm theo xấp xỉ có vai trò quan trọng hơn đối với dữ liệu truyền thống, thuần nhất.

2. Sử dụng khoảng xác suất. Giả sử cần lưu trữ thông tin về dạng “đối tượng o₁ xảy ra trong ảnh i là Đào, với xác suất p = 0, 6”, người ta dùng xác suất khoảng [l, u] thay cho xác suất điểm.

Việc phát triển này hoàn toàn tự nhiên, nhằm mô tả các giá trị có sai số ε của p, tức thay vì sử dụng p, người ta sử dụng $[p-\varepsilon, p+\varepsilon]$. Do vậy nếu xác định sai số $\pm x$, chẳng hạn $x = 0.03 \%$, thì xác suất p được thay thế bằng $[\max(0, p - 0.03 * p), \min(1, p + 0.03 * p)]$.

Do vậy trong bảng dữ liệu quan hệ, bộ (O₁, Toto, 0.8) được thay bằng (O₁, Toto, 0.77, 0.83). Việc thay thế này yêu cầu quá trình tìm kiếm thực hiện theo cặp giá trị khoảng.

3. Tiếp cận tổng quát. Một số khái niệm:

- i. Xác định quan hệ xác suất trên lược đồ (A₁,... A_n) là quan hệ bình thường (A₁,... A_n, LB, UB), trong đó LB, UB là khoảng [0, 1] số thực;

Thực tế cho thấy quan hệ R là quan hệ có 3 thuộc tính (tên ảnh, tên đối tượng, tên), với điều kiện toàn vẹn:

- $\forall t_1, t_2: t_1. \text{tên đối tượng} = t_2. \text{tên đối tượng} \rightarrow t_1. \text{tên ảnh} = t_2. \text{tên ảnh}$, cho

- phép khẳng định tên đối tượng chỉ liên kết với một ảnh;
- $\forall t: t.LB \leq t.UB$;
- ii. Cơ sở dữ liệu ảnh gồm quan hệ xác suất R như trên, với các quan hệ bình thường R_1, \dots, R_k ứng với các thuộc tính ảnh, tức chưa tính đến giá trị xác suất. Lí do sử dụng hai dạng quan hệ là đối với các phép toán thường gây ra sự không chắc chắn sẽ được áp dụng với các đối tượng trong quan hệ R;
- iii. Câu hỏi về mối quan hệ¹ là câu có dạng “tìm các ảnh trong cơ sở dữ liệu ảnh có các đối tượng mang tên là s_1, \dots, s_n ”. Câu hỏi viết bằng ngôn ngữ SQL là:


```
select tên ảnh
from R, R1,... Rn
where R1. tên đối tượng = s1 and... and Rn. Tên đối tượng = sn and
R1. tên ảnh = R2. tên ảnh and... and R1. tên ảnh = Rn. tên ảnh
```
- iv. Kết quả của câu hỏi về mối quan hệ là bảng có 2 cột (i) cột tên ảnh; (ii) cột LB, UB, tức là bộ (tên ảnh, l, u) $\leftrightarrow \forall j = 1..n, \exists$ bộ $t_j \in R$:
 - t. tên ảnh = tên ảnh;
 - t. LB = l_i , t, UB = u_i ;
 - $[l, u] = [l_1, u_1] \otimes \dots \otimes [l_n, u_n]$; trong đó toán tử \otimes được xác định $[x, y] \otimes [x_2, y_2] = [\max(0, x + x_2 - 1), \min(y, y_2)]$, và người ta thấy việc sử dụng toán tử \otimes sẽ đạt hiệu quả tìm kiếm.

4.7.7. Thể hiện cơ sở dữ liệu ảnh trên cây R

Việc thể hiện trên cây R định trước các đối tượng đa phương tiện được thể hiện như các hình chữ nhật. Việc xây dựng cơ sở dữ liệu ảnh được thực hiện theo:

- Tạo quan hệ R với hai thuộc tính R (tên ảnh, tên đối tượng). Xác định các đối tượng xuất hiện trong ảnh, tức quan hệ R cũng như quan hệ “hiện trong”;
- Tạo cây R lưu trữ tất cả các ảnh. Nếu có cùng một hình chữ nhật xuất hiện trong cả hai ảnh, thì người ta sử dụng danh sách “trùng nhau” gắn với nút trên cây R;
- Mỗi hình chữ nhật có tập các trường xác định thuộc tính đối tượng, hay vùng cho ảnh. Đó là thông tin mở rộng về nội dung.

Thí dụ: cơ sở dữ liệu về khuôn mặt

Các ảnh một.gif, hai.gif, và ba.gif cho phép thể hiện các hình chữ nhật gắn với các đối tượng Toto, Titi,... Nút cây R liên kết với cơ sở dữ liệu có cấu trúc dữ liệu:

¹ relationship



	tên đối tượng	ảnh	XLB	XUB	YLB	YUB
	Toto	một.gif				
	Titi	một.gif				
	Tutu	hai.gif				
	Tata	ba.gif				

Hình. Các hình chữ nhật trùm nhau, gắn với các đối tượng về khuôn mặt

Nút khuôn mặt = record

Rec_1, rec_2, rec_3 : chữ nhật;

P_1, p_2, p_3 : ↑ kiểu nút;

End;

Kiểu chữ nhật = record

XLB, XUB, YLB, YUB: integer;

Danh sách đối tượng: ↑ kiểu nút đối tượng;

Ngày, tháng, năm: integer;

Loại máy chụp: string;

Thông tin khác: kiểu bản ghi;

Vị trí: string;

End;

Kiểu nút đối tượng = record

Tên đối tượng: string;

Tên ảnh: string;

Data: ↑ kiểu thông tin;

Tiếp: ↑ kiểu nút đối tượng;

End;

Kiểu thông tin = record

Tên đối tượng: string;

Dữ liệu về đối tượng: bản ghi dữ liệu đối tượng;

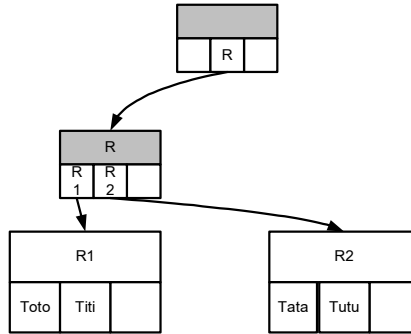
L_p, up : real; % biên xác suất thấp/ cao

Tiếp: ↑ kiểu thông tin;

End;

Giả sử các ảnh một.gif, hai.gif, và ba.gif thể hiện về đối tượng Toto, Tutu, Titi, Tata với xác suất khoảng, nêu trong bảng quan hệ.

Cây R thể hiện 3 ảnh và 4 đối tượng. Tiếp theo người ta hoàn thiện cây R bằng cách bổ sung các thông tin liên quan đến các đối tượng trong bảng dữ liệu. Người ta có thể bổ sung một đối tượng mới, với các giá trị biên của hình chữ nhật tương ứng.

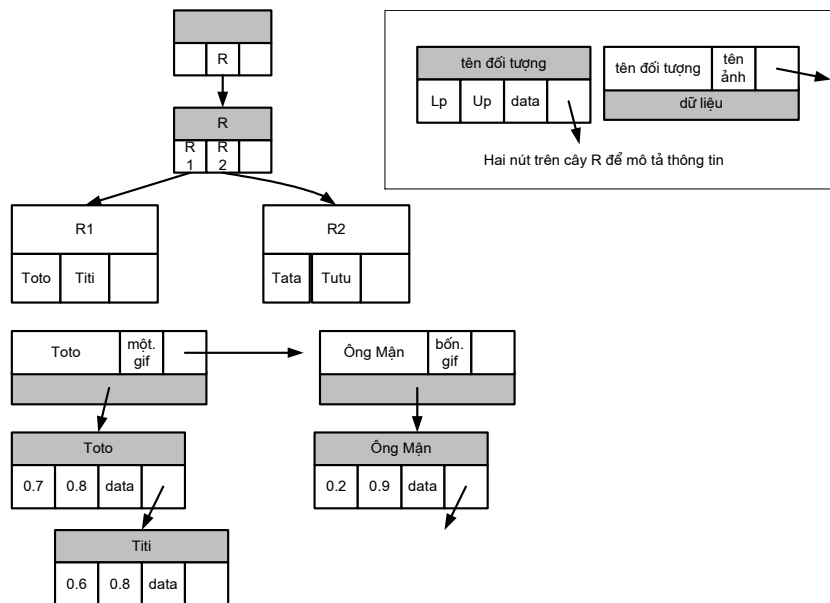


Hình. Cây R ứng với 3 ảnh

4.7.8. Kết luận về cơ sở dữ liệu ảnh

Người ta dùng cây R để thể hiện các ảnh. Khi đã xây dựng được cây với một số ảnh, người ta có thể thể hiện tiếp các ảnh được bằng quan hệ mô tả lên cây.

Về cơ sở dữ liệu, còn một số khía cạnh khác cần được xét đến. Đó là (i) tìm kiếm ảnh trên mặt của không gian; (ii) một số vấn đề về cài đặt hệ thống cơ sở dữ liệu ảnh.



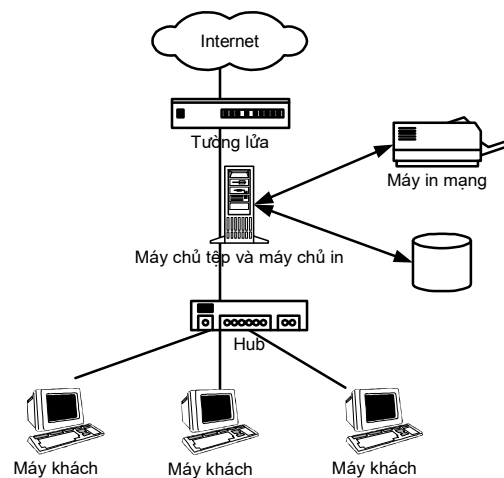
Hình. Cây R với dữ liệu đối tượng cũ và bổ sung đối tượng mới

4.8. Nhận xét về dữ liệu đa phương tiện

4.8.1. Đảm bảo QoS trong hệ thống truyền thông, tại máy chủ và máy khách

MIRS thường là phân tán. Các đối tượng đa phương tiện tìm thấy trên máy chủ được chuyển sang máy khách để thể hiện. Dữ liệu đa phương tiện có yêu cầu riêng. Chúng cần băng thông lớn, không gian lưu trữ rộng, lưu lượng nhiều, chịu trễ, đồng bộ về thời gian, không gian. Phương tiện khác nhau có yêu cầu khác nhau. Các yêu cầu này cần được thỏa mãn để truyền thông và thể hiện trong toàn hệ thống. Để đảm bảo khung thống nhất cho việc mô tả và đảm bảo các yêu cầu đa dạng, người ta đề xuất QoS, là tập các tham số. Không có tập tham số nhất trí nào về QoS. Tuy nhiên các tham số cơ bản chiếu theo các vấn đề trao đổi trước đây, gồm băng thông, lưu lượng truyền, lưu trữ,

đồng bộ... Các tham số này thường được mô tả theo hai bậc (i) chất lượng mong muốn; (ii) chất lượng chấp nhận được. QoS là cam kết được trao đổi và nhất trí trong các ứng dụng đa phương tiện và hệ thống đa phương tiện. Khi ứng dụng bắt đầu phiên mới, nó trình bày yêu cầu QoS tới hệ thống; hệ thống có thể hoặc từ chối hoặc chấp nhận yêu cầu, có thể cần thương thuyết lại. Nếu hệ thống chấp thuận, cam kết giữa hệ thống và ứng dụng được ký và hệ thống cần đảm bảo các yêu cầu. Đảm bảo này có thể ở ba dạng (i) về phần cứng và tất định; (ii) về phần mềm và đảm bảo thống kê; (iii) nỗ lực đảm bảo. Đối với đảm bảo tất định, yêu cầu QoS được hoàn toàn đáp ứng. Đảm bảo thống kê cung cấp tin cậy xác suất. Đối với chính sách đảm bảo nỗ lực, không có đảm bảo cụ thể, tùy theo khai thác của ứng dụng. Đó là chính sách chia sẻ hệ thống truyền thông. Đảm bảo cần có tại các đầu nút và toàn hệ thống. Hệ thống đa phương tiện điển hình có ba thành phần chính; (i) trạm chủ, gồm máy chủ và máy khách chịu điều khiển của hệ thống điều hành; (ii) quản trị bộ nhớ; (iii) hệ thống truyền thông. Hệ thống vận chuyển gồm (i) giao thức truyền thông; (ii) kiến trúc mạng vật lý.



Hình. Kiến trúc khách/ chủ

QoS có thể chỉ đảm bảo khi tài nguyên hệ thống cần có được quản lý thích hợp. Các tài nguyên hệ thống gồm (i) chu kỳ CPU; (ii) bộ nhớ; (iii) băng thông... Mỗi thành phần hệ thống có bộ quản trị, với màn giám sát hiện trạng sử dụng và tài nguyên còn dư. Khi nhận được yêu cầu của phiên, hệ thống tiến hành thử quản trị. Nếu tài nguyên còn đủ đối với yêu cầu của phiên, hoạt động của phiên này không can thiệp vào các phiên khác, yêu cầu được chấp thuận. Ngược lại, tập tham số QoS khác đề ra cho ứng dụng theo tài nguyên có thể. Phiên làm việc sẽ tiếp tục được thực hiện hoặc bị hủy.

Nghiên cứu về QoS còn rộng mở. Một số vấn đề (i) cách chuyển tham số QoS sang yêu cầu tài nguyên; (ii) cách lập lịch các phiên để dùng chung tài nguyên. Đặc biệt đối với ứng dụng đa phương tiện, truyền thông đa phương tiện cần được quan tâm.

4.8.2. Một số vấn đề khác

Phần trước đề cập vài khía cạnh thiết kế chính của MIRS. Một số vấn đề khác như (i) nén dữ liệu đa phương tiện; (ii) chuẩn hóa thể hiện dữ liệu đa phương tiện; (iii) xử lý câu hỏi và tìm kiếm sẽ được đề cập dưới đây.

4.8.2.1. Nén dữ liệu đa phương tiện

Đích chính của việc nén dữ liệu đa phương tiện là nén càng nhiều càng tốt, không quan tâm khía cạnh thể hiện và tìm kiếm. Hầu hết tệp âm thanh, hình ảnh, video sử dụng chuẩn và kỹ thuật chuyên dụng. Để trích đặc trưng từ các tệp, cần giải nén tệp. Tiếp cận này không hiệu quả về tính toán, vì nén và giải nén được thực hiện vài lần trong quá trình trích đặc trưng; về lưu trữ cũng không hiệu quả do tệp nén và tệp đặc trưng tách biệt. Do vậy, (i) cần có kỹ thuật nén cho phép trích đặc trưng trực tiếp từ dữ liệu nén; và (ii) nén dựa trên đối tượng chứ không theo giá trị mẫu riêng. Người ta đã đề cập chỉ số hóa hình ảnh từ DCT, lượng tử vectơ, dữ liệu nén sóng nhỏ, nhưng kết quả chưa khả quan. Về sau, nếu người ta chuyển các ảnh, hay đồ họa, bitmap sang đồ họa vectơ, người ta sẽ đạt được tỉ lệ nén cao và dễ tìm kiếm. Mô hình dựa trên nén hình ảnh và video được thực hiện từng bước. Chuẩn MPEG-4, MPEG-2000 là nỗ lực kết hợp nén và tìm kiếm.

Khía cạnh khác của nén dữ liệu là tính phù hợp với thể hiện và truyền thông. Trong nhiều ứng dụng, một số hình biểu tượng hiện ra trên màn hình để người dùng duyệt. Nếu họ quan tâm hình ảnh cụ thể, họ sẽ chọn và hình ảnh với độ phân giải cao hơn được tìm kiếm và hiển thị. Bình thường người ta thực hiện yêu cầu theo hai phương pháp sau:

1. Các hình ảnh với kích thước khác nhau suy ra từ cùng ảnh được tạo, được nén tuần tự, và lưu trong máy chủ một cách độc lập. Các hình ảnh theo kích thước yêu cầu được gửi đến máy khách để hiển thị. Chẳng hạn dữ liệu hình biểu tượng được truyền cho lần duyệt đầu, hình ảnh kích thước đầy đủ sẽ được gửi theo các yêu cầu lần lượt;
2. Các hình ảnh được nén và lưu tuần tự theo kích thước gốc. Theo yêu cầu, dữ liệu hình ảnh đầy đủ được chuyển đến máy khách mà không quan tâm đến yêu cầu của họ. Đa số trường hợp, người dùng sẽ giảm kích thước dữ liệu ảnh đã nhận để hiển thị, rồi khi muốn chọn ảnh nào, hệ thống truyền lại ảnh đầy đủ.

Hai phương pháp không đủ hiệu quả theo nghĩa không gian lưu trữ và băng thông. Phương pháp đầu lãng phí bộ nhớ vì các ảnh kích thước khác nhau có liên quan; băng thông cũng lãng phí do máy chủ truyền ảnh đầy đủ. Phương pháp thứ hai tiết kiệm không gian nhớ, nhưng cũng lãng phí băng thông khi người dùng chỉ cần ảnh biểu tượng. Việc truyền ảnh với kích thước thực sẽ gây trễ truyền.

Để giải quyết vấn đề trên, các kỹ thuật nén phân cấp liên tiếp, vô hướng, như trong MPEG và MPEG-4 được dùng. Chúng tiết kiệm không gian và đường truyền, và cải thiện thời gian trả lời khi truyền ảnh, và thể hiện dần dần thay vì đợi giải nén xong.

4.8.2.2. Chuẩn hóa thể hiện dữ liệu

Trong quá trình trích đặc trưng, giả sử có các giá trị mẫu thô của mỗi phương tiện theo cùng cách và cùng ý nghĩa. Thực tế giả thiết này không hợp lý; chẳng hạn các đoạn âm thanh có thể được ghi theo các mức khuếch đại khác nhau, nên việc so sánh các giá trị mẫu của các đoạn khác nhau là không có nghĩa. Tương tự, các giá trị pixel ảnh từ các ảnh khác nhau có thể có nghĩa khác nhau tùy theo giá trị chỉnh gamma và hệ thống màu khác nhau. Bởi vậy, thông tin tác động đến giá trị mẫu cần được ghi trong đầu tệp và các giá trị mẫu hỗ trợ một cách phù hợp, khi trích đặc trưng. Hiện giờ, các dạng thức âm thanh và hình ảnh thông thường không mang thông tin này. Vậy cần có thể hiện chuẩn đối với mỗi phương tiện. Điều này không chỉ làm các giá trị mẫu có cùng ý nghĩa, mà còn đơn giản hóa quá trình giải nén và thể hiện, vì tránh được việc giải nén và quản lý

nhiều dạng thức và thể hiện khác nhau.

4.8.2.3. Xử lý câu hỏi và tìm kiếm

Nhiều thứ liên quan đến quá trình hỏi và tìm kiếm thông tin: (i) nhu cầu về ngôn ngữ hỏi đa phương tiện hình thức; người ta có SQL/MM, tuy khả năng hạn chế; (ii) câu hỏi đa phương tiện có thể sử dụng vài đặc trưng. Mỗi đặc trưng sinh ra danh sách xếp hạng, nên vấn đề là cách thu được danh sách kết hợp từ nhiều cái riêng lẻ, và đặc trưng nào là quan trọng hơn; (iii) về tích hợp hệ thống: thông tin thể hiện trên một số phương tiện; người dùng quan tâm đến thông tin, chứ không rõ về phương tiện. Vấn đề là cách để các kiểu phương tiện khác nhau và các kỹ thuật quản lý có thể được tích hợp để thể hiện thông tin hiệu quả đến người dùng.

4.9. Kết luận

Chương này với nhan đề Quản trị dữ liệu đa phương tiện đã đề cập các khái niệm:

- Khái niệm về quản trị cơ sở dữ liệu đa phương tiện, về (i) dạng dữ liệu đa phương tiện; (ii) ngôn ngữ hỏi dữ liệu đa phương tiện;
- Kiến trúc hệ quản trị cơ sở dữ liệu đa phương tiện, về (i) nguyên tắc thiết kế; (ii) kiến trúc hệ thống đa phương tiện;
- Các kỹ thuật mô hình hóa dữ liệu, tập trung vào ưu điểm của mô hình dữ liệu hướng đối tượng;
- Kỹ thuật chỉ số hoá và trừu tượng hoá, cho thấy vai trò của mô hình khái niệm và nhu cầu của chỉ số đối với tìm kiếm dữ liệu đa phương tiện;
- Tìm thông tin đa phương tiện dựa trên nội dung, về (i) lọc dữ liệu; (ii) sử dụng từ khóa khi tìm theo nội dung;
- Một vài hệ thống thí dụ về cơ sở dữ liệu đa phương tiện và ứng dụng của hệ thống đa phương tiện;
- Khía cạnh chất lượng dịch vụ đa phương tiện và an toàn dữ liệu đa phương tiện.

Hướng dẫn sử dụng tài liệu theo chương trình khung



ĐỀ CƯƠNG CHI TIẾT HỆ CƠ SỞ DỮ LIỆU ĐA PHƯƠNG TIỆN

1. Thông tin chung về môn học

- Tên môn học: **HỆ CƠ SỞ DỮ LIỆU ĐA PHƯƠNG TIỆN**
MULTIMEDIA DATABASE SYSTEMS
- Mã môn học:
- Số đvht: 4
- Loại môn học: bắt buộc
- Các môn học tiên quyết: Cơ sở dữ liệu, Multimedia, Hệ quản trị cơ sở dữ liệu quan hệ
- Phân bổ giờ đối với các hoạt động:
 - Giảng lý thuyết : 48 tiết
 - Thực hành, thí nghiệm : 12 tiết
 - Tự học : 60 giờ

Lưu ý: các giờ Thảo luận, Thực hành, Thí nghiệm là số tiết đã được quy đổi theo Quy chế)

- Khoa/Bộ môn phụ trách môn học: Khoa học máy tính

2. Mục tiêu của môn học

- Kiến thức: Trang bị cho sinh viên các kiến thức về tổ chức dữ liệu đa phương tiện và xử lý dữ liệu đa phương tiện.
- Kỹ năng: Sau khi học xong, sinh viên nắm vững các kiến thức về tổ chức dữ liệu đa phương tiện và xử lý dữ liệu đa phương tiện và có khả năng áp dụng và phân tích thiết kế các hệ thống CSDL đa phương tiện trong thực tế.
- Thái độ, chuyên cần: đảm bảo số giờ học trên lớp và tự học.

3. Tóm tắt nội dung môn học

Môn học cung cấp cho sinh viên các kiến thức về tổ chức dữ liệu đa phương tiện và xử lý dữ liệu đa phương tiện. Thực hành trên hệ quản trị cơ sở dữ liệu để thể hiện chức năng xử lý dữ liệu đa phương tiện.

4. Nội dung chi tiết môn học

Chương I. Tổng quan về cơ sở dữ liệu đa phương tiện

- 1.1 Mở đầu
- 1.2 Khái niệm dữ liệu đa phương tiện
- 1.3 Đặc trưng của các đối tượng đa phương tiện
- 1.3 Cấu trúc lưu trữ cơ sở dữ liệu đa phương tiện
- 1.4 Ngôn ngữ thao tác dữ liệu đa phương tiện.

Chương II. Tư liệu đa phương tiện tương tác

- 2.1 Cơ sở dữ liệu đa phương tiện tương tác
- 2.2 Mô hình hoá tư liệu đa phương tiện tương tác IMD
- 2.3 Phân loại
- 2.4 Mô hình kịch bản
- 2.5 Tìm kiếm tư liệu đa phương tiện tương tác

Chương III. Thành tựu và xu hướng

- 3.1 Các thành tựu chính của công nghệ hệ quản trị cơ sở dữ liệu đa phương tiện
- 3.2 Các sản phẩm thương mại và mẫu nghiên cứu
- 3.3 Hướng phát triển của cơ sở dữ liệu đa phương tiện

Chương IV. Quản trị dữ liệu đa phương tiện

- 4.1 Định nghĩa quản trị cơ sở dữ liệu đa phương tiện
- 4.2 Kiến trúc hệ quản trị cơ sở dữ liệu đa phương tiện
- 4.3 Các kỹ thuật mô hình hóa dữ liệu
- 4.4 Các kỹ thuật chỉ số hoá và trừu tượng hoá
- 4.5 Tìm thông tin đa phương tiện dựa trên nội dung
- 4.6 Thí dụ về cơ sở dữ liệu đa phương tiện
- 4.7 Các ứng dụng của đa phương tiện
- 4.8 Nhận xét về dữ liệu đa phương tiện

5. Học liệu

- Học liệu bắt buộc:

- [1] Subrahmania V. S. *Principles of Multimedia Database Systems*, Ed. The Morgan Kaufmann, 1998.

- Học liệu tham khảo

- [2] Mario Piattini, *Advanced databases Technology and Design*, Artech House Publishers, 2000.
- [3] Guojun Lu, *Multimedia Database Management Systems*, Artech House Publishers, 1999.

6. Hình thức tổ chức dạy học:

Đây là nội dung rất quan trọng đối với giáo viên, sinh viên và các đơn vị quản lý. Mỗi nội dung kiến thức đều được tổ chức dạy dưới các hình thức chủ yếu như: lý thuyết, bài tập, thảo luận, thực hành, hoạt động theo nhóm và tự học, tự nghiên cứu... Giáo viên có trách nhiệm cập nhật nội dung này hàng năm. Mỗi nội dung trong lịch trình dạy-học, phải xác định được **số tiết** sẽ thực hiện ở từng hình thức trên. Lưu ý rằng để chuẩn bị cho một tiết lý thuyết, thực hành, thí nghiệm sinh viên cần 1 giờ chuẩn bị cá nhân.

Lịch trình dạy-học (thiết kế cho cả tiến trình, tối thiểu là 15 tuần, mỗi tuần ... tiết. Ví dụ: mỗi tuần 03 tiết đối với môn học 3 đvht; mỗi tuần 04 tiết đối với môn học 4 đvht)

Thời gian	Nội dung	Hình thức tổ chức dạy-học					Yêu cầu sinh viên chuẩn bị trước khi lên lớp
		Giờ lên lớp			Thực hành	Tự học	
		Lý thuyết	Hdẫn Btập	Thảo luận			
Tuần 1	Chương 1: 1.1, 1.2, 1.3	4					Quyển 1: Phần I, Quyển 2: Chương 8
Tuần 2	Chương 1: 1.4, 1.5	4					Quyển 1: Phần I; Quyển 2: Chương 8
Tuần 3	Chương 2: 2.1, 2.2	4					Quyển 1: Phần II
Tuần 4	Chương 2: 2.3, 2.4	4					Quyển 1: Phần II; Quyển 2: Chương 8
Tuần 5	Chương 2: 2.5	4					Quyển 1: Phần II; Quyển 2: Chương 8
Tuần 6	Chương 3: 3.1, 3.2	4					Quyển 1: Chương 15; Quyển 2: Chương 8
Tuần 7	Chương 3: 3.2, 3.3	4					Quyển 1: Chương 15; Quyển 2: Chương 8
Tuần 8	Chương 4: 4.1, 4.2	4					Quyển 1: Phần IV; Quyển 3: Chương 12
Tuần 9	Chương 4: 4.2, 4.3	4					Quyển 1: Phần IV; Quyển 3: Chương 12
Tuần 10	Chương 4: 4.4	4					Quyển 1: Phần IV; Quyển 3: Chương 12
Tuần 11	Chương 4: 4.5, 4.6	4					Quyển 1: Phần IV; Quyển 3: Chương 12
Tuần 12	Chương 4: 4.7, 4.8	4					Quyển 1: Phần IV; Quyển 3: Chương 12
Tuần 13	Thực hành				4		
Tuần 14	Thực hành				4		
Tuần 15	Thực hành				4		

Ghi chú: Thống nhất toàn bộ các môn học sẽ thực hiện kiểm tra giữa kỳ vào tuần thứ 8

7. **Thang điểm đánh giá:** từ 0 đến 10

8. Phương pháp, hình thức kiểm tra - đánh giá kết quả học tập môn học: bao gồm các phần sau (hình thức kiểm tra cuối kỳ, trọng số các nội dung cần phải thông qua Trường Bộ môn).

8.1 Các loại điểm kiểm tra và hình thức đánh giá:

- Tham gia học tập trên lớp: *(đi học đầy đủ, chuẩn bị bài tốt và tích cực thảo luận,...)*;
- Phần tự học, tự nghiên cứu có hướng dẫn của giảng viên: *(hoàn thành tốt nội dung, nhiệm vụ mà giảng viên giao cho cá nhân: thực hành; thí nghiệm; bài tập nhóm/tháng; bài tập cá nhân/học kì,...)*;
- Thí nghiệm, thực hành: Tham gia đầy đủ
- Kiểm tra - đánh giá giữa kì: Kiểm tra viết
- Kiểm tra - đánh giá cuối kì: Thi viết

8.2 Trọng số các loại điểm kiểm tra:

- Tham gia học tập trên lớp: 10 % *(đánh giá chuyên cần và thái độ học tập)*
- Thực hành/Thí nghiệm/Bài tập/Thảo luận: 10 %
- Kiểm tra giữa kỳ: 10 %
- Kiểm tra cuối kỳ: 70 % *(không được thấp hơn 50%)*

9. Ngày phê duyệt:

Giảng viên
(Ký và ghi rõ họ tên)
Hoàng Xuân Dậu

Tổ trưởng Bộ môn
(Ký và ghi rõ họ tên)
Hoàng Xuân Dậu

Trưởng khoa
(Ký và ghi rõ họ tên)
Từ Minh Phương

Tài liệu tham khảo

1. Arjen P. De Vries, Content and multimedia database management systems, Luận văn Centre for Telematics and Information Technology, University of Twente, The Netherlands, 1999
2. Đặng Văn Đức, Nguyễn Thị Phương Trà, Giao diện người-máy, Nxb. Khoa học Tự nhiên và Công nghệ, 2009
3. Đỗ Trung Tuấn, Giao diện người-máy, Nxb. Khoa học và Kỹ thuật, 2007
4. Elaine England, Andy Finney, Managing Multimedia, Addison Wesley Ed., 2 ed., 1999
5. Guojun Lu, Multimedia Database Management Systems, Artech House Publishers, 1999.
6. Harald Kosch, Mario Doller, Multimedia Database Systems: where are we now ? trong “Distributed Multimedia database technologies supported by MPEG-7 and MPEG-21”, 2003
7. <http://borkweb.com/story/an-introduction-to-multimedia>, 2010
8. <http://www.adobe.com/products/director/>, 2010
9. <http://www.adobe.com/shockwave/download>, 2010
10. <http://www.corel.com/servlet/Satellite/us/en/Product>, 2010
11. John L. Newman, Multimedia Database Systems
12. John Villamil Casanova, Louis Molina, An interactive guide to Multimedia, QUE E&T Ed., 1998
13. Mario Piattini, Advanced databases Technology and Design, Artech House Publishers, 2000.
14. Nguyễn Nhung, Nhiếp ảnh và nghệ thuật, Nxb. Văn hoá, 1992
15. Nguyễn Quân, Ngôn ngữ của hình và màu sắc, Nxb. Văn hóa Thông tin, 2006
16. Nguyễn Thanh Bình, Võ Nguyễn Quốc Bảo, Xử lý âm thanh, hình ảnh, Học viện Công nghệ Bưu chính Viễn thông, 2007
17. Subrahmania V. S., Principles of Multimedia Database Systems, Ed. The Morgan Kaufmann, 1998
18. Tay Vaughan, Multimedia. Making it work, Osborne MacGrawHill Ed., 1998
19. Wikipedia, the free encyclopedia. Wikimedia. Internet: [http:// en.wikipedia.org/wiki/](http://en.wikipedia.org/wiki/), 2010
20. Yamaha. The Sound Reinforcement Handbook. Written for Yamaha by G. Davis & R. Jones. Hal Leonard Publishing Corporation, Milwaukee, WI, USA, 1990