

Một số kỹ thuật tìm kiếm ảnh theo nội dung trong cơ sở dữ liệu đa phương tiện / Dương Hoàng Huyền ; Nghd. : PGS.TS. Đặng Văn Đức

Ngay từ ban đầu máy tính đã được coi là các thiết bị xử lý các kí tự, đầu vào có dạng *alphabet* và đầu ra là các biểu tượng có cùng dạng trên. Điều này đã trở thành mô hình tính toán chuẩn dựa trên cơ sở máy *Turing*.

Tuy nhiên, trong những năm gần đây xuất hiện nhu cầu vô cùng lớn về khả năng khai thác và xử lý dữ liệu với số lượng khổng lồ mà các dữ liệu này thì không dễ dàng diễn tả bằng các kí hiệu, các loại dữ liệu đó là hình ảnh, video, audio, tài liệu và dữ liệu viết tay.

Các kiểu dữ liệu liệt kê trên đây chỉ là một phần trong rất nhiều hình thức biểu hiện của dữ liệu phát sinh tự nhiên trong các ứng dụng khác nhau. Từ những dữ liệu mới phát sinh trên, ta thấy rằng các *CSDL* truyền thống không thể quản lý các kiểu dữ liệu này được. Do đó cần phải có một hệ thống quản lý tất cả các loại dữ liệu này và *CSDL* đa phương tiện được hình thành.

Luận văn này trình bày một số vấn đề cơ bản của một hệ quản trị *CSDL* đa phương tiện chẳng hạn các khái niệm, kiến trúc hệ thống và mô hình dữ liệu đa phương tiện. Trong đó, tập trung nghiên cứu và cài đặt thử nghiệm một số phương pháp tìm kiếm theo nội dung trên dữ liệu đa phương tiện, cụ thể là dữ liệu ảnh.

Cấu trúc của luận văn bao gồm phần mở đầu, 3 chương nội dung, kết luận và danh mục 8 tài liệu tham khảo.

Chương 1 trình bày các khái niệm cơ bản về dữ liệu và mô hình dữ liệu đa phương tiện tổng quát, nguyên lý thiết kế, kiến trúc hệ thống của một hệ quản trị cơ sở dữ liệu đa phương tiện. Giới thiệu khái quát cơ sở dữ liệu ảnh và một hệ thống tìm kiếm ảnh theo nội dung tiêu biểu.

Chương 2 trình bày các đặc trưng trực quan của nội dung ảnh bao gồm các đặc trưng cơ bản đó là màu (color), kết cấu (texture) và đặc trưng hình dạng (shape). Trong chương này cũng trình bày một số kỹ thuật tìm kiếm ảnh theo nội dung dựa vào 3 đặc trưng quan đã nêu.

Chương 3 cài đặt một số phương pháp đã trình bày trong chương 2 bao gồm các phương pháp tìm kiếm dựa vào đặc trưng màu (dùng biểu đồ màu và moment màu), dựa vào đặc trưng kết cấu (dùng ma trận đồng hiện) và đặc trưng hình dạng (dùng các moment bất biến).

CHƯƠNG 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU ĐA PHƯƠNG TIỆN

1.1 Các khái niệm cơ bản

Dữ liệu *Media* là các kiểu thông tin hoặc biểu diễn của các kiểu thông tin như các ký tự, ảnh, *audio* và *video*. *Multimedia* là tập hợp các kiểu *media* được sử dụng với nhau. Hệ quản trị cơ sở dữ liệu đa phương tiện (*Multimedia DataBase Management System - MMDBMS*) là một hệ thống giúp người dùng thao tác trên các dữ liệu *media* một cách thích hợp và hiệu quả. Các thao tác cơ bản của hệ

quản trị *CSDL* thường là chèn, cập nhật, xóa và tìm kiếm một đối tượng trong *CSDL*. Các dữ liệu *media* có những đặc điểm chính như sau:

- Dữ liệu *media*, đặc biệt là dữ liệu *audio* và *video* có dung lượng rất lớn.
- Dữ liệu *audio* và *video* có tham số thời gian.
- Dữ liệu *media* được biểu diễn ở dạng đặc thù, không có cấu trúc ngữ nghĩa rõ ràng để máy tính có thể tự động nhận biết nội dung.
- Ý nghĩa dữ liệu đa phương tiện đôi khi không rõ ràng phụ thuộc vào cách quan sát và ý kiến chủ quan của con người.
- Dữ liệu đa phương tiện rất giàu thông tin, để biểu diễn đầy đủ nội dung cần phải có rất nhiều tham số.

1.2 nguyên lý thiết kế *CSDL* đa phương tiện

Trong một *MMDBMS* ta thường sử dụng ba nguyên lý sau để tổ chức nội dung:

1. Nguyên lý tự trị (*principle of autonomy*): ta phải lựa chọn để nhóm chung *media* cùng loại. Chỉ số hóa mỗi loại *media* này theo cách riêng và hiệu quả nhất để thuận tiện khi truy cập vào các đối tượng. Ta gọi là tự trị vì mỗi loại *media* được tổ chức riêng và phù hợp với từng loại.

2. Nguyên lý đồng nhất (*Principle of uniformity*): cố gắng tìm ra một cấu trúc trừu tượng hơn để chỉ số hóa mọi kiểu dữ liệu, khi truy cập vào các *media* khác nhau đều thông qua chỉ số này. Nói cách khác, chúng ta biểu diễn nội dung của các đối tượng *media* khác nhau

(ảnh, tài liệu, *audio*, *video*...) trong cùng một cấu trúc dữ liệu, sau đó phát triển thuật toán để thao tác cấu trúc dữ liệu đó.

3. Nguyên lý tổ chức lai (*Principle of hybrid organization*): là sự kết hợp hai nguyên lý trên. Theo nguyên lý này, một số dữ liệu *media* sử dụng chỉ số riêng và những dữ liệu còn lại sẽ sử dụng cùng một chỉ số.

1.3 Kiến trúc hệ thống và mô hình dữ liệu *CSDL* đa phương tiện

1.3.1 Kiến trúc tổng quát

Kiến trúc tổng quát bao gồm các khối chức năng chính là giao diện người dùng, khối trích chọn đặc trưng, bộ phận truyền thông, công cụ tìm kiếm và chỉ số hóa, và cuối cùng là bộ phận quản lý lưu trữ. Các khối chức năng này được mô tả thông qua các hoạt động của nó.

1.3.2 Mô hình dữ liệu

Một mô hình dữ liệu bao gồm các tầng chính sau. Tầng đối tượng chỉ rõ các mối quan hệ về không gian, thời gian và sự tích hợp giữa các đối tượng. Tầng các kiểu *media* chứa các kiểu *media* chung như *text*, *image*, *audio* và *video*. Tại mức này, các đặc trưng và thuộc tính của đối tượng được chỉ rõ. Ví dụ, đối với ảnh có thể có các tính chất như kích thước ảnh, biểu đồ màu, các đối tượng chính chứa trong ảnh. Tầng khuôn dạng *media* chỉ rõ khuôn dạng lưu trữ dữ liệu. Một kiểu *media* thông thường có nhiều khuôn dạng ví dụ như ảnh có thể có dạng thô hoặc dạng nén. Kiến trúc vừa nêu chỉ là một kiến trúc tổng quát, tùy theo mỗi ứng dụng khác nhau sẽ có một mô hình dữ liệu khác nhau, không có một tiêu chuẩn chung cho các tầng.

màu và *texture* rất khả quan, tuy nhiên kết quả tìm kiếm dùng đặc trưng hình dạng còn nhiều hạn chế và kém chính xác bởi vì vấn đề tìm kiếm dựa vào hình dạng cho đến nay vẫn còn là một vấn đề khó. Một vấn đề còn tồn tại trong luận văn này là chưa đánh giá được hiệu suất tìm kiếm của các phương pháp đã giới thiệu ở trên.

Hướng nghiên cứu tiếp theo sẽ cài đặt các kỹ thuật còn lại đã trình bày trong luận văn. Tích hợp các đặc trưng trực quan này với nhau để cải thiện hiệu quả tìm kiếm. Tổ chức các đặc trưng của ảnh thành một *CSDL* đặc trưng để giảm thời gian tìm kiếm.

Output: M_{pq} là *moment* cấp p, q của đối tượng

Algorithm:

```

Begin
   $M_{pq} = 0;$ 
  For ( $x=0; x<I.width; x++$ )
    For ( $y=0; y<I.height; y++$ )
      {
         $Color = I.pixel(x, y);$ 
        If ( $Color = black$ )
           $M_{pq} = M_{pq} + (x-x_c)^p \times (y-y_c)^q;$ 
      }
  Return  $M_{pq};$ 
End.

```

KẾT LUẬN

Luận văn đã giới thiệu một số kỹ thuật cơ bản tìm kiếm ảnh theo nội dung sử dụng các đặc trưng trực quan. Các đặc trưng trực quan thường được sử dụng trong các hệ thống *CBIR* là màu, kết cấu và hình dạng. Đặc trưng màu được biểu diễn bởi biểu đồ màu và *moment* màu trong một không gian màu nào đó, chẳng hạn như *RGB*, *HSV*, v.v..., đặc trưng *texture* được biểu diễn bởi các đặc trưng được trích chọn từ ma trận đồng hiện hoặc các đặc trưng của Tamura. Hình dạng có thể được biểu diễn dựa vào các *moment* bất biến, hoặc các kí hiệu mô tả *Fourier*.

Luận văn đã cài đặt thành công một số kỹ thuật tìm kiếm dựa vào ba đặc trưng trực quan đã nêu trên. Đối với đặc trưng màu, cài đặt ba phương pháp cơ bản đó là *GCH*, *LCH* và *moment*. Đối với đặc trưng *texture*, cài đặt kỹ thuật sử dụng ma trận đồng hiện để trích chọn các đặc trưng. Đối với đặc trưng hình dạng, chỉ cài đặt phương pháp dùng các *moment* bất biến. Kết quả tìm kiếm dùng hai đặc trưng

1.3.3 Giao diện người dùng

Người dùng tương tác với *MMDBMS* thông qua một giao diện.

Chức năng chính của giao diện là cho phép người dùng chèn, xóa, cập nhật các đối tượng mới vào *CSDL*, nhập vào các truy vấn và biểu diễn các kết quả truy vấn. Một giao diện người dùng tốt phải hỗ trợ các chức năng sau:

- Cung cấp công cụ cho người dùng chèn, xóa, cập nhật các đối tượng mới vào *CSDL* một cách dễ dàng.
- Cung cấp công cụ cho người dùng nhập truy vấn một cách hiệu quả hoặc cho hệ thống biết các thông tin cần thiết của truy vấn.
- Biểu diễn kết quả truy vấn một cách hiệu quả.
- Giao diện thân thiện với người dùng.

1.3.4 Trích chọn đặc trưng, chỉ số hóa và độ đo tương tự

Thao tác trích chọn đặc trưng phải giải quyết được các vấn đề sau:

- Đặc trưng và thuộc tính đã trích chọn phải đầy đủ để có thể biểu diễn nội dung của các mục thông tin.
- Đặc trưng và thuộc tính đã được trích chọn phải cô đọng, súc tích. Nếu đặc trưng quá phức tạp và lớn sẽ ảnh hưởng đến hiệu quả tìm kiếm.
- Tính toán khoảng cách của đặc trưng phải hiệu quả.

Tổng quát, có bốn loại đặc trưng cơ bản sau: *metadata*, chú thích văn bản, đặc trưng nội dung mức thấp, và đặc trưng nội dung mức cao. *Metadata* bao gồm các dạng hoặc các thuộc tính thực của đối tượng ví dụ như tên tác giả, ngày tạo và tiêu đề của đối tượng.

Metadata không mô tả hoặc giải thích nội dung của đối tượng.

Chú thích văn bản là dùng văn bản để mô tả nội dung đối tượng.

Đặc trưng mức thấp nắm bắt các thông số của đối tượng đa phương tiện và các mối quan hệ không gian, thời gian của các cặp đối tượng. Các loại *media* khác nhau sẽ có những đặc trưng mức thấp khác nhau. Trong *audio*, đặc trưng mức thấp bao gồm âm sắc, phân phối tần số, sóng siêu âm. Các đặc trưng mức thấp của ảnh bao gồm phân phối màu, kết cấu bề mặt, hình dạng đối tượng và các mối quan hệ không gian. Các đặc trưng mức thấp của *video* bao gồm tham số thời gian và một số đặc trưng giống như ảnh. Điểm thuận lợi của các đặc trưng mức thấp là nó có thể được trích chọn một cách tự động.

Đặc trưng mức cao nhận dạng và hiểu các đối tượng. Ngoại trừ nhận dạng văn bản và tiếng nói, rất khó để có thể nhận dạng được các mẫu *audio* và các đối tượng trực quan khác.

Sau khi đã trích chọn đặc trưng, các kỹ thuật chỉ số hóa giúp tổ chức các đặc trưng để tìm kiếm hiệu quả. Mỗi đối tượng có thể có nhiều đặc trưng để biểu diễn và mỗi đặc trưng lại có nhiều tham số, do đó cần phải có cơ chế chỉ số hóa tốt để tổ chức các đặc trưng này.

Thao tác tìm kiếm trong dữ liệu đa phương tiện thường dựa vào tính tương tự thay vì tìm kiếm chính xác giữa các mục trong *CSDL*. Tính tương tự được tính toán dựa vào các đặc trưng và thuộc tính đã trích chọn.

```
For (y=0; y<I.Height-dy; y++)
{
    i = I.Pixel(x,y);
    j = I.Pixel(x+dx,y+dy);
    N[i,j] = N[i,j] + 1;
}
Return N;
```

End.

3.3 Tìm kiếm dựa vào đặc trưng hình dạng sử dụng các moment bất biến

Đầu tiên, tính trọng tâm của đối tượng, chi tiết như sau:

Input: ảnh nhị phân I có chứa một đối tượng

Output: x_c, y_c là tọa độ trọng tâm của đối tượng

Algorithm:

```
Begin
    xc = 0; yc = 0; N = 0;
    For (x=0; x<I.width; x++)
        For (y=0; y<I.height; y++)
        {
            Color = I.pixel(x,y);
            if (Color = black)
            {
                xc = xc + x;
                yc = yc + y;
                N = N + 1;
            }
        }
    xc = xc / N;
    yc = yc / N;
    return {xc, yc};
End.
```

Sau khi tính trọng tâm, ta tính moment cấp $p+q$ của đối tượng, chi tiết như sau:

Input: ảnh nhị phân I chứa đối tượng, cấp p, q của *moment*.

```

EB = Blue / N;
σR = 0; σG = 0; σB = 0;
SR = 0; SG = 0; SB = 0;
For (x=0; x<I.Width; x++)
    For (y=0; y<I.Height; y++)
    {
        color = I.pixel(x, y);
        σR = σR + (color.Red-ER)2;
        σG = σG + (color.Green-EG)2;
        σB = σB + (color.Blue-EB)2;
        SR = SR + (color.Red-ER)3;
        SG = SG + (color.Green-EG)3;
        SB = SB + (color.Blue-EB)3;
    }
σR =  $\sqrt{\frac{\sigma_R}{N}}$ ; σG =  $\sqrt{\frac{\sigma_G}{N}}$ ; σB =  $\sqrt{\frac{\sigma_B}{N}}$ ;
SR =  $\sqrt[3]{\frac{S_R}{N}}$ ; SG =  $\sqrt[3]{\frac{S_G}{N}}$ ; SB =  $\sqrt[3]{\frac{S_B}{N}}$ ;
Return {ER, EG, EB, σR, σG, σB, SR, SG, SB};

```

End.

3.2 Tìm kiếm dựa vào đặc trưng texture sử dụng ma trận đồng hiện

Đầu tiên tính ma trận đồng hiện của ảnh, sau đó tính bảy đặc trưng được trích chọn từ ma trận đồng hiện đó. Thuật toán tính ma trận đồng hiện được mô tả như sau:

Input: ảnh I

Output: ma trận đồng hiện N

Algorithm:

```

Begin
    For (i=1; i<=b; i++)
        For (j=1; j<=b; j++)
            N[i, j] = 0;
    For (x=0; x<I.Width-dx; x++)

```

1.4 Cơ sở dữ liệu ảnh

1.4.1 Các cơ sở dữ liệu truyền thống

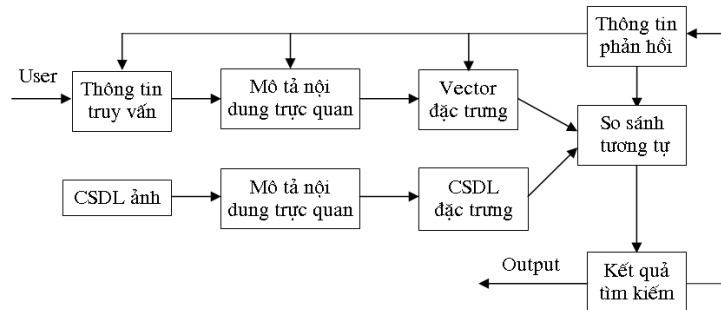
Định nghĩa không hình thức *CSDL* quan hệ là một bảng gồm các cột và các dòng, mỗi dòng được gọi là một bộ, mỗi cột gọi là một thuộc tính, mỗi thuộc tính phải có một kiểu dữ liệu duy nhất. Một cơ sở dữ liệu khác, đó là cơ sở dữ liệu hướng đối tượng, trong *CSDL* này, các thuộc tính sẽ là một đối tượng cụ thể nào đó. Một kết hợp của hai cơ sở dữ liệu trên đó là *CSDL quan hệ-đối tượng*, có dạng tổng quát như sau: $(A_1:T_1, A_2:T_2, \dots, A_N:T_N)$ trong đó A_i là một thuộc tính của quan hệ và T_i là một đối tượng nào đó với các thuộc tính và phương thức riêng.

1.4.2 Cơ sở dữ liệu ảnh

Cơ sở dữ liệu ảnh là một bộ ba $(GI, Prop, Rec)$ trong đó GI là tập các lưới ảnh mỗi tế bào có dạng $(Image, m, n)$, $Prop$ là tập các đặc tính tế bào, Rec là tập các số nguyên mô tả hình dạng của đối tượng trong mỗi ảnh. Thông thường ta dùng cơ sở dữ liệu *quan hệ-đối tượng* để biểu diễn *CSDL* ảnh.

1.4.3 Tìm kiếm ảnh theo nội dung

Tìm kiếm ảnh theo nội dung (*Content-Based Image Retrieval-CBIR*) là kỹ thuật sử dụng các nội dung trực quan (nội dung mức thấp) để tìm kiếm ảnh từ *CSDL* ảnh rộng lớn. Hình sau đây minh họa một hệ thống tìm kiếm ảnh theo nội dung tiêu biểu.



Hình 1.6 Hệ thống tìm kiếm ảnh theo nội dung tiêu biểu

CHƯƠNG 2: MỘT SỐ KỸ THUẬT TÌM KIẾM THEO NỘI DUNG TRONG CƠ SỞ DỮ LIỆU ẢNH

2.1 Kí hiệu mô tả nội dung ảnh

Kí hiệu mô tả nội dung ảnh dùng để biểu diễn nội dung của một ảnh. Các loại đặc trưng mô tả nội dung ảnh đó là màu, kết cấu, hình dạng và mối quan hệ không gian. Một kí hiệu mô tả nội dung ảnh tốt phải bất biến với các phép đổi như xoay, co giãn và tịnh tiến. Kí hiệu mô tả nội dung ảnh có thể là toàn bộ ảnh hay một vùng ảnh.

2.2 Tìm kiếm ảnh dựa vào đặc trưng màu

2.2.1 Không gian màu

Không gian màu được định nghĩa trong không gian ba chiều, các loại ảnh hiện nay đều sử dụng không gian màu *RGB* để biểu diễn ảnh trong máy tính. Tuy nhiên trong các hệ thống tìm kiếm ảnh theo nội dung, thường sử dụng hai không gian màu *RGB* và *HSV*.

2.2.2 Lượng tử hóa màu

Lượng tử hóa là quá trình giảm số màu biểu diễn ảnh. Trong

```

h = dy * (r-1);
For (x=w; x<w+dx; x++)
    For (y=h; y<h+dy; y++)
    {
        i = I.Pixel(x,y);
        CH[i] = CH[i] + 1;
    }
return CH;

```

End.

3.1.2 Phương pháp dùng biểu đồ màu toàn cục

Phương pháp tìm kiếm ảnh dựa trên biểu đồ màu toàn cục là một trường hợp đặc biệt của phương pháp biểu đồ màu cục bộ (khi lưới chia của ảnh trong phương pháp biểu đồ màu cục bộ là 1×1).

3.1.3 Phương pháp dùng moment màu

Để cài đặt thuật toán này trước hết phải tính trung bình từng kênh màu của ảnh, sau đó tính độ lệch chuẩn và căn bậc ba của phương sai. Chi tiết của thuật toán như sau:

Input : ảnh *I*

Output: 9 giá trị *moment* màu của ảnh

Algorithm:

Begin

```

N = (I.Width * I.Height);
Red = 0; Green = 0; Blue = 0;
For (x=0; x<I.Width; x++)
    For (y=0; y<I.Height; y++)
    {
        color = I.pixel(x,y);
        Red = Red + color.Red;
        Green = Green + color.Green;
        Blue = Blue + color.Blue;
    }
ER = Red / N;
EG = Green / N;

```

3.1.1 Phương pháp dùng biểu đồ màu cục bộ

Bước đầu tiên trong phương pháp này là lượng tử hóa ảnh.

Thuật toán như sau:

Input: ảnh I

Output: ảnh H là ảnh I sau khi được lượng tử hóa

Algorithm:

```

Begin
  For (x=0; x<I.width; x++)
    For (y=0; y<I.height; y++)
    {
      color = I.pixel(x, y);
      Red   = (color.Red div 51)×51 + 25;
      Green = (color.Green div 51)×51 + 25;
      Blue  = (color.Blue div 51)×51 + 25;
      color = (Red, Green, Blue);
      H.pixel(x, y) = color;
    }
  Return H;
End.

```

Bước tiếp theo chia ảnh thành các vùng và tính biểu đồ màu cho từng vùng. Thuật toán tính biểu đồ màu từng vùng như sau:

Input: ảnh I đã lượng tử hóa thành 125 màu; m, n số cột, hàng của lưới, c, r cho biết vùng (c, r) của ảnh.

Output: biểu đồ màu của vùng ảnh $CH[i], i=1..125$.

Algorithm:

```

Begin
  For (i=1; i<125; i++)
    CH[i] = 0;
    dx = I.width div m;
    dy = I.height div n;
    w = dx × (c-1);

```

luận văn này sử dụng công thức sau để lượng tử hóa màu:

$$r' = \left\lfloor \frac{r}{51} \right\rfloor \times 51 + 25, \quad g' = \left\lfloor \frac{g}{51} \right\rfloor \times 51 + 25, \quad b' = \left\lfloor \frac{b}{51} \right\rfloor \times 51 + 25,$$

trong đó r, g, b là giá trị ba thành phần của màu trước khi lượng tử hóa, r', g', b' là giá trị ba thành phần của màu sau khi đã lượng tử hóa.

2.2.3 Biểu đồ màu

Giả sử ảnh I có kích thước là $W \times H$ và $I(i, j)$ là giá trị màu (sau khi ảnh đã được lượng tử hóa) của điểm ảnh tại vị trí (i, j) , $C = (y_1, y_2, \dots, y_M)$ là tập hợp các màu biểu diễn ảnh sau khi lượng tử hóa, y_m là màu thứ m của tập các màu lượng tử hóa. Khi đó biểu đồ màu của ảnh I kí hiệu là $H_I = \{H[1], H[2], \dots, H[M]\}$ định nghĩa như sau:

$$H_I[m] = \sum_{i=1}^{W-1} \sum_{j=1}^{H-1} \alpha(I(i, j), y_m), \quad m=1, \dots, M.$$

Trong đó $\alpha(I(i, j), y_m) = 1$ nếu $I(i, j) = y_m$ và $= 0$ trong trường hợp còn lại.

2.2.4 Độ đo khoảng cách của biểu đồ màu

Có nhiều hàm đo khoảng cách của hai biểu đồ màu, trong đó hai dạng dạng hàm cơ bản nhất đó là dạng Minkowski và dạng toàn phương. Công thức cụ thể sau đây với H là biểu đồ màu:

$$\text{Dạng toàn phương: } d(Q, I) = \sum_{i=1}^N |H_Q[i] - H_I[i]|^r.$$

$$\text{Dạng Minkowski: } d(Q, I) = (H_Q - H_I)^T A (H_Q - H_I).$$

2.2.5 Kỹ thuật tìm kiếm ảnh dùng biểu đồ màu

2.2.5.1 Kỹ thuật dùng biểu đồ màu toàn cục

Theo kỹ thuật này, ta sẽ tính biểu đồ màu cho toàn bộ ảnh, sau đó tính khoảng cách của hai ảnh thông qua biểu đồ này.

2.2.5.2 Kỹ thuật dùng biểu đồ màu cục bộ

Theo kỹ thuật này, ta chia ảnh thành các vùng, sau đó tính biểu đồ màu cho các vùng có cùng vị trí. Khoảng cách giữa hai vùng chính là khoảng cách giữa hai biểu đồ màu. Khoảng cách của hai ảnh là tổng khoảng cách giữa các vùng.

2.2.5.3 Phương pháp Harbin

Phương pháp này được phát triển dựa vào kỹ thuật biểu đồ cục bộ. Việc so sánh không phải thực hiện tại các vùng có cùng vị trí, mà so sánh vùng của ảnh này đến tất cả các vùng còn lại của ảnh khác, sau đó chọn ra một khoảng cách gần nhất giữa hai vùng.

2.2.6 Kỹ thuật tìm kiếm ảnh dùng moment màu

Trong kỹ thuật này, đặc trưng phân bố màu được thể hiện qua 3 tham số. Đó là: giá trị trung bình (*average*), độ biến đổi (*variance*), độ lệch (*knewness*) của các kênh màu. Được định nghĩa tương ứng như sau:

$$E_i = \frac{1}{F} \sum_{j=1}^F P_{ij}, \sigma_i = \left(\frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^2 \right)^{\frac{1}{2}}, s_i = \left(\frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^3 \right)^{\frac{1}{3}}.$$

trong đó P_{ij} là giá trị kênh màu thứ i tại điểm ảnh j ; E_i , σ_i , s_i tương ứng là giá trị trung bình, độ lệch chuẩn, căn bậc ba của kênh màu thứ i , F là tổng số điểm ảnh. Trong không gian RGB có ba kênh màu tương ứng là Red, Blue, Green.

$$f = \left(\frac{|FD_2|}{|FD_1|}, \frac{|FD_3|}{|FD_1|}, \dots, \frac{|FD_{N-1}|}{|FD_1|} \right).$$

Vector đặc trưng biểu diễn kí hiệu khoảng cách trọng tâm và độ cong:

$$f = \left(\frac{|FD_2|}{|FD_0|}, \frac{|FD_3|}{|FD_0|}, \dots, \frac{|FD_{N/2}|}{|FD_0|} \right).$$

Bây giờ, giả sử có hai hình dạng A và B được chỉ số hóa thành các *vector* đặc trưng $f_A = (f_A^1, f_A^2, \dots, f_A^N)$ và $f_B = (f_B^1, f_B^2, \dots, f_B^N)$. Ta có thể sử dụng khoảng cách *Euclidean* để tính khoảng cách hai *vector* này.

2.5 Kết luận

Dữ liệu ảnh được sử dụng rộng rãi trong các lĩnh vực ứng dụng, cụ thể trong các lĩnh vực như giải trí, thương mại, y tế, viễn thám, quản lý..., do đó dữ liệu ảnh thuộc ứng dụng nào sẽ có các đặc điểm riêng của nó, và tùy theo đặc điểm riêng đó ta sẽ có một phương pháp tìm kiếm thích hợp.

CHƯƠNG 3: CÀI ĐẶT VÀ THỬ NGHIỆM

3.1 Tìm kiếm ảnh dựa vào đặc trưng màu

Trong phần này sẽ cài đặt ba phương pháp tìm kiếm dựa màu, đó là phương pháp dùng biểu đồ màu cục bộ, biểu đồ màu toàn cục và *moment*. *CSDL* khoảng 180 ảnh màu *JPEG* các loại khác nhau có kích thước khoảng 400×400 *pixel* và 24-bit màu. Ảnh truy vấn là một ảnh màu bất kì.

2.4.3 Ký hiệu mô tả Fourier

1. Tọa độ phức (*Complex coordinate*): là một hàm được sinh ra từ các điểm biên của hình dạng được xác định như sau: $z(t)=[x(t)-x_c]+i[y(t)-y_c]$, với (x_c, y_c) là tọa độ trọng tâm của đối tượng và được xác định như sau: $x_c = \frac{1}{L} \sum_{t=0}^{L-1} x(t)$, $y_c = \frac{1}{L} \sum_{t=0}^{L-1} y(t)$. Các biểu diễn này là bất biến đối với phép tịnh tiến.

2. Khoảng cách trọng tâm (*Centroid distance*): hàm khoảng cách trọng tâm được mô tả bởi khoảng cách của các điểm trên biên đến tâm (x_c, y_c) của hình dạng. Hàm được định nghĩa như sau:

$$r(t) = ([x(t)-x_c]^2 + [y(t)-y_c]^2)^{1/2}.$$

3. Độ cong (*curvature*): biểu diễn đạo hàm cấp hai và đạo hàm cấp một của biên. Hàm độ cong (*curvature function*) được định nghĩa như là hiệu của các góc ở biên với trục x và được tính trong cửa sổ w . Định nghĩa hình thức của hàm độ cong như sau:

$$k(t) = \theta(t) - \theta(t-1), \text{ với } \theta(t) = \arctan \frac{y(t) - y(t-w)}{x(t) - x(t-w)}.$$

Khi đó biến đổi *Fourier* rời rạc của $s(t)$ (một trong 3 ký hiệu trên) được định nghĩa như sau:

$$u_n = \frac{1}{N} \sum_{t=0}^{N-1} s(t) \exp\left(\frac{-i2\pi nt}{N}\right), \text{ với } n = 0, 1, \dots, N-1. \text{ Các hệ số}$$

u_n được gọi là các ký hiệu mô tả *Fourier* (*Fourier Descriptor*) của hình dạng viết tắt là FD_n , $n = 0, 1, \dots, N-1$. Khi đó, *vector* đặc trưng để biểu diễn ký hiệu tọa độ phức như sau:

Khoảng cách của hai ảnh được tính như sau:

$$d(Q, I) = \sum_{i=1}^r \left(|E_i^Q - E_i^I| + |\sigma_i^Q - \sigma_i^I| + |s_i^Q - s_i^I| \right).$$

2.3 Tìm kiếm ảnh dựa vào đặc trưng texture

2.3.1 Một số khái niệm về texture

Texture là tập những *texel* (*texture element*), được sắp xếp theo một quy luật nào đó, hay có cấu trúc lặp đi lặp lại. Ví dụ, một viên gạch trong bức tường gạch. Trong ví dụ này, ta có thể coi viên gạch là *texel* và sự sắp xếp của các viên gạch trong bức tường sẽ tạo nên kết cấu bề mặt của bức tường.

2.3.1 Ma trận đồng hiện

Cho ma trận P_d có cấp $G \times G$ và một *vector* $d(dx, dy)$. Khi đó, ma trận đồng hiện được định nghĩa như sau: mỗi phần tử (i, j) của ma trận P_d là số lần xuất hiện của hai mức xám i, j với khoảng cách không gian là $d(dx, dy)$. Định nghĩa hình thức như sau:

$P_d(i, j) = \left| \{(c, r): I(c, r) = i \text{ và } I(c+dx, r+dy) = j\} \right|$ trong đó $(c, r), (c+dx, r+dy) \in M \times N$ với M, N là kích thước của ảnh I ; $i, j = 1..G$ là các giá trị mức xám có thể xuất hiện trong ảnh I . Lưu ý, G là số lượng mức xám trong ảnh I sau khi đã lượng tử hóa.

Ma trận đồng hiện chuẩn hóa:

$$N_d(i, j) = \frac{P_d(i, j)}{\sum_{k=1}^G \sum_{h=1}^G P_d(k, h)}.$$

Một số đặc trưng được trích chọn từ ma trận đồng hiện chuẩn hóa sau đây:

1. $Energy = \sum_{i=1}^G \sum_{j=1}^G N_d^2(i, j)$
2. $Entropy = - \sum_{i=1}^G \sum_{j=1}^G N_d(i, j) \log N_d(i, j)$
3. $Homogeneity = \sum_{i=1}^G \sum_{j=1}^G \frac{1}{1 + (i - j)^2} N_d(i, j)$
4. $Contrast = \sum_{i=1}^G \sum_{j=1}^G (i - j)^2 N_d(i, j)$
5. $Cluster\ Shade = \sum_{i=1}^G \sum_{j=1}^G ((i - \mu_i) + (j - \mu_j))^3 N_d(i, j)$
6. $Cluster\ Prominence = \sum_{i=1}^G \sum_{j=1}^G ((i - \mu_i) + (j - \mu_j))^4 N_d(i, j)$
7. $Correlation = \sum_{i=1}^G \sum_{j=1}^G \frac{(i - \mu_i)(j - \mu_j) N_d(i, j)}{\sigma_i \sigma_j}$

Trong đó $\mu_i, \mu_j, \sigma_i, \sigma_j$ được định nghĩa như sau:

$$\mu_i = \sum_{i=1}^G i \sum_{j=1}^G N_d(i, j), \quad \mu_j = \sum_{j=1}^G j \sum_{i=1}^G N_d(i, j)$$

$$\sigma_i = \sum_{i=1}^G (i - \mu_i)^2 \sum_{j=1}^G N_d(i, j), \quad \sigma_j = \sum_{j=1}^G (j - \mu_j)^2 \sum_{i=1}^G N_d(i, j)$$

Các phần tử của ma trận đồng hiện phụ thuộc vào vector d , vấn đề đặt ra là làm sao chọn d để có ma trận đồng hiện thể hiện được tính cấu trúc *texture* nhất. Zuker và Tezopoloulos đề xuất một phương

cũng bất biến đối với phép biến đổi, xoay và co giãn:

$$\phi_1 = \mu_{2,0} + \mu_{0,2}$$

$$\phi_2 = (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2$$

$$\phi_3 = (\mu_{3,0} - 3\mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2$$

$$\phi_4 = (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} + \mu_{2,1})^2$$

$$\phi_5 = (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] + (\mu_{0,3} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1})[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2]$$

$$\phi_6 = (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1})^2] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1})$$

$$\phi_7 = (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2]$$

2.4.2 Biểu diễn hình dạng dựa vào vùng

Ý tưởng của phương pháp như sau: cho một vùng có chứa hình dạng cần quan tâm, chia vùng thành lưới các ô vuông có kích thước cố định. Xây dựng chuỗi nhị phân như sau: ta đặt 1 nếu ô vuông đó có ít nhất là 25% số điểm nằm trong hình dạng, và đặt 0 trong trường hợp còn lại. Khi đó ta nhận được một chuỗi nhị phân thứ tự từ trái sang phải biểu diễn đặc trưng của hình dạng.

Giả sử R, R' là biểu diễn *vector* dòng của hai vùng hình dạng. C, C' là *vector* cột biểu diễn hai vùng hình dạng tương ứng. Tính $R_d = \sum_i |R_i - R'_i|$, $C_d = \sum_i |C_i - C'_i|$, với R_i, R'_i là dòng thứ i của hai *vector* chỉ số hóa của hai vùng ảnh, tương tự C_i, C'_i là cột thứ i của hai vùng ảnh. Nếu $R_d + C_d < t$ (ngưỡng) thì hai hình dạng đó tương tự nhau.

và *texture*, đặc trưng hình dạng mô tả ảnh sau khi được phân đoạn thành các vùng hoặc các đối tượng cụ thể. Việc phân đoạn ảnh một cách chính xác và rõ ràng là rất khó, do đó khi sử dụng đặc trưng hình dạng để tìm kiếm ảnh thường bị giới hạn trong các ứng dụng đặc biệt, nơi mà các đối tượng và vùng chứa các đối tượng đã có sẵn. Các phương pháp biểu diễn hình dạng thường được chia thành hai loại: biểu diễn dựa vào biên và biểu diễn dựa vào vùng. Một phương pháp biểu diễn hình dạng tốt phải bất biến đối với phép xoay, co giãn, tịnh tiến. Trong phần này chỉ mô tả khái quát một số phương pháp biểu diễn hình dạng thường dùng.

2.4.1 Moment bất biến

Một phương pháp biểu diễn hình dạng cổ điển là sử dụng tập các *moment* bất biến. Nếu đối tượng R được biểu diễn như là ảnh nhị phân, tức là: $I(x,y) = 1$ nếu $(x,y) \in R$ và $I(x,y) = 0$ nếu $(x,y) \notin R$. *Moment* trọng tâm cấp $p+q$ biểu diễn hình dạng được định nghĩa như sau:

$$\mu_{p,q} = \sum_{(x,y) \in R} (x - x_c)^p (y - y_c)^q, \text{ trong đó } x_c, y_c \text{ là tâm của đối}$$

$$\text{tượng được xác định theo công thức: } x_c = \frac{\sum_{(x,y) \in R} x}{|R|}, y_c = \frac{\sum_{(x,y) \in R} y}{|R|}.$$

Moment trọng tâm có thể chuẩn hóa để bất biến đối với phép co giãn:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma}, \text{ với } \gamma = \frac{p+q+2}{2}.$$

Dựa vào *moment* được định nghĩa trên, ta có các *moment* sau

pháp để tìm d như sau: tìm d để biểu thức sau đạt giá trị cực đại

$$\chi^2(d) = \sum_{i=1}^G \sum_{j=1}^G \frac{N_d^2(i,j)}{N_d(i)N_d(j)} - 1.$$

2.3.3 Phương pháp chuỗi *texture*

Ojala và các đồng nghiệp đã đề xuất phương pháp như sau: một đơn vị *texture* được biểu diễn bởi tám phần tử, mỗi phần tử có nhận một trong hai giá trị $\{0,1\}$ và nhận được từ “tám láng giềng” của điểm đang xét. Các đơn vị *texture* này được gọi là mẫu nhị phân cục bộ (*Local Binary Pattern- LBP*) và sự xuất hiện phân phối trên vùng tạo ra chuỗi *texture*. *LBP* được tính bằng việc tính ngưỡng của mỗi điểm ảnh láng giềng so với giá trị của điểm ảnh trung tâm cho kết quả 256 mẫu nhị phân khác nhau. Phương pháp *LBP* có thể kết hợp với độ đo tương phản đơn giản. Chi tiết của thuật toán tính mẫu nhị phân và độ đo tương phản được mô tả như sau: cho trước một điểm trung tâm có giá trị là P_0 và “tám láng giềng” có giá trị là $P_1, P_2, \dots, P_8$.

1. Tính ngưỡng của điểm P_i so với giá trị trung tâm:

$$P_i' = \begin{cases} 0 & \text{nếu } P_i < P_0 \\ 1 & \text{nếu } P_i \geq P_0 \end{cases}$$

2. Đếm số n : $n = \sum_{i=1}^8 P_i'$

3. Tính mẫu nhị phân cục bộ: $LBP = \sum_{i=1}^8 P_i' 2^{i-1}$

4. Tính độ tương phản cục bộ

$$Contrast = \begin{cases} 0 & \text{nếu } n=0 \text{ hoặc } n=8 \\ \frac{1}{n} \sum_{i=1}^8 P_i P_i - \frac{1}{8-n} \sum_{i=1}^8 (1-P_i) P_i & \text{còn lại} \end{cases}$$

2.3.4 Đặc trưng tương quan tự động

Định nghĩa hình thức hàm tương quan tự động p của ảnh I như sau:

$$p(x,y) = \frac{\sum_{u=0}^N \sum_{v=0}^N I(u,v) I(u+x, v+y)}{\sum_{u=0}^N \sum_{v=0}^N I^2(u,v)}.$$

Hàm tương quan tự động của các *texture* không chu kỳ được thể hiện bởi các điểm trội. Bề ngang và bề dài của điểm này sẽ xác định độ thô và hướng của *texture*.

2.3.5 Các đặc trưng của Tamura

Coarseness

Coarseness là một độ đo tính chất hạt (*granularity*) của *texture*. Đầu tiên tính trung bình động $A_k(x,y)$ bằng cách sử dụng của sổ có kích thước $2^k \times 2^k$ ($k=0, 1, \dots, 5$) tại mỗi điểm, tức là:

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}} \sum_{j=y-2^{k-1}}^{y+2^{k-1}} \frac{g(i,j)}{2^{2k}} \text{ với } g(i,j) \text{ là cường độ điểm ảnh}$$

tại (i,j)

Sau đó tính sự khác nhau giữa các cặp trung bình động không trùng lặp theo chiều ngang và chiều dọc của mỗi điểm ảnh, tức là:

$$E_{k,h}(x,y) = |A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y)|$$

$$E_{k,v}(x,y) = |A_k(x,y+2^{k-1}) - A_k(x,y-2^{k-1})|$$

Tiếp tục tính giá trị của k để giá trị E đạt cực đại theo hướng ngang hoặc dọc, đặt giá trị k là kích thước tốt nhất cho mỗi điểm ảnh, tức là:

$$S_{best}(x,y) = 2^k$$

Độ thô của một *texture* sẽ được tính là trung bình của S_{best} trên toàn bộ ảnh, tức là:

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i,j)$$

Contrast

Công thức của *Contrast* như sau:

$$F_{con} = \frac{\sigma}{\alpha^{1/4}}, \text{ với } \alpha = \frac{\mu_4}{\sigma^4}, \mu_4 \text{ là moment cấp 4, } \sigma^2 \text{ là phương sai.}$$

Công thức này được tính cho toàn bộ ảnh hoặc cho từng vùng ảnh.

2.3.6 Độ đo tương tự của texture

$$D(I, J) = \|f_I - f_J\| = (f_I - f_J)^T (f_I - f_J).$$

$$D(I, J) = |f_I - f_J| = \sum_{k=1}^n |f_{k,I} - f_{k,J}|$$

$$D(I, J) = \max_k |f_{k,I} - f_{k,J}|$$

2.4 Tìm kiếm ảnh dựa vào đặc trưng hình dạng

Đặc trưng hình dạng của một đối tượng đã được sử dụng nhiều trong các hệ thống tìm kiếm ảnh theo nội dung. So với đặc trưng màu