

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG



ISO 9001:2015

ĐỒ ÁN TỐT NGHIỆP

NGÀNH: CÔNG NGHỆ THÔNG TIN

Sinh viên : Vũ Lê Minh Hoàng
Giảng viên hướng dẫn: TS. Ngô Trường Giang

HẢI PHÒNG - 2018

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG**

**ƯỚC LƯỢNG TẬP ĐIỂM TƯƠNG ĐỒNG GIỮA HAI ẢNH
DỰA TRÊN ĐỐI SÁNH ĐẶC TRƯNG SIFT**

**ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
NGÀNH: CÔNG NGHỆ THÔNG TIN**

**Sinh viên : Vũ Lê Minh Hoàng
Giảng viên hướng dẫn: TS. Lê Trường Giang**

HẢI PHÒNG - 2018

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

NHIỆM VỤ ĐỀ TÀI TỐT NGHIỆP

Sinh viên: Vũ Lê Minh Hoàng

Mã SV: 1412101034

Lớp: CT1802

Ngành: Công nghệ thông tin

Tên đề tài: Ước lượng tập điểm tương đồng giữa hai ảnh dựa trên đối sánh đặc trưng SIFT

MỤC LỤC

LỜI CẢM ƠN	3
LỜI MỞ ĐẦU	4
DANH MỤC HÌNH VẼ	5
CHƯƠNG 1: PHÁT HIỆN VÀ MÔ TẢ ĐẶC TRƯNG ẢNH	6
1.1 Giới thiệu	6
1.2 Các loại đặc trưng ảnh	8
1.2.1 Đặc trưng toàn cục và cục bộ	8
1.2.2 Đặc điểm của phát hiện đặc trưng	9
1.2.3 Bất biến với tỷ lệ và biến đổi Affine	10
1.3 Phát hiện đặc trưng ảnh	11
1.3.1 Phát hiện đơn tỉ lệ.....	12
1.3.2 Phát hiện đa tỉ lệ	19
1.4 Mô tả đặc trưng ảnh	23
1.4.1 Scale Invariant Feature Transform (SIFT)	23
1.4.2 Gradient Location-Orientation Histogram (GLOH)	24
1.4.3 Speeded-Up Robust Features Descriptor (SURF)	25
CHƯƠNG 2: ĐỐI SÁNH ẢNH DỰA TRÊN ĐẶC TRƯNG SIFT	28
2.1 Giới thiệu về đối sánh ảnh	28
2.2 Các phương pháp đối sánh ảnh	29
2.2.1 Phương pháp dựa trên vùng (Area based methods)	29
2.2.2 Phương pháp dựa theo đặc trưng (Feature based methods)	40
2.3 Đối sánh ảnh dựa trên đặc trưng SIFT	44
2.3.1 Trích chọn đặc trưng SIFT	44
2.3.2 Đối sánh SIFT	46
CHƯƠNG 3: CHƯƠNG TRÌNH THỰC NGHIỆM	48
3.1 Môi trường thực nghiệm	48
3.1.1 Cấu hình phần cứng	48
3.1.2 Môi trường cài đặt	48

3.1.3 Thư viện OpenCV (Open Source Computer Vision Library	48
3.2 Trích chọn đặc trưng SIFT	50
3.3 Ước lượng tập điểm tương đồng.....	52
3.4 Một số kết quả thực nghiệm	55
KẾT LUẬN	57
TÀI LIỆU THAM KHẢO.....	58

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành nhất đến quý thầy cô Trường Đại Học Dân Lập Hải Phòng, những người đã dìu dắt em tận tình, đã truyền đạt cho em những kiến thức và bài học quý báu trong suốt thời gian em theo học tại trường.

Em xin trân trọng gửi lời cảm ơn đến tất cả các thầy cô trong khoa Công Nghệ Thông Tin, đặc biệt là thầy giáo TS. Ngô Trường Giang, thầy đã tận tình hướng dẫn và giúp đỡ em trong suốt quá trình làm tốt nghiệp. Với sự chỉ bảo của Thầy, em đã có những định hướng tốt trong việc triển khai và thực hiện các yêu cầu trong quá trình làm đồ án tốt nghiệp.

Ngoài ra, em cũng xin gửi lời cảm ơn tới tất cả bạn bè, đặc biệt là các bạn trong lớp CT1802 đã luôn gắn bó, cùng học tập và giúp đỡ em trong những năm qua và trong suốt quá trình thực hiện đồ án này.

Em xin chân thành cảm ơn!

*Hải Phòng, ngày 03 tháng 11 năm
2018*

Sinh viên

Vũ Lê Minh Hoàng

LỜI MỞ ĐẦU

Trong lĩnh vực nhận dạng đối tượng ngày nay, hướng nghiên cứu phổ biến trên thế giới là việc sử dụng các điểm bất biến (Invariant Feature) trong ảnh làm đặc trưng để nhận dạng. Tiêu biểu nhất trong các thuật toán đối sánh sử dụng đặc trưng dạng này là thuật toán SIFT (Scale-Invariant Feature Transform, David Lowe 1999 và 2004), SIFT có thể coi là thuật toán tiền đề cho các ứng dụng cũng như giải thuật khác về trích chọn biến đổi đặc trưng bất biến trong ảnh. Các giải thuật đang ứng dụng trong thực tế khác đều dựa trên hay phát triển theo các nhánh riêng của SIFT.

Các đặc trưng trong SIFT không phụ thuộc vào các phép biến đổi ảnh cơ bản như xoay, co giãn, thay đổi độ sáng, v.v. nên có thể xem tập các đặc trưng của một ảnh là thể hiện cho nội dung của ảnh đó. Vì vậy kết quả của việc nhận dạng sẽ có độ chính xác rất cao và thậm chí có thể khôi phục được đối tượng bị che khuất trong ảnh. Tuy nhiên giải thuật SIFT rất phức tạp trong cài đặt, đòi hỏi thời gian nghiên cứu và am hiểu nhiều thuật toán thành phần.

Trong phạm vi đề tài này, em sẽ sử dụng đặc trưng SIFT để ước lượng tập điểm tương đồng giữa hai ảnh nhằm mục đích hiểu biết thêm về phương pháp trích chọn đặc trưng trong việc triển khai các ứng dụng trong thực tế về lĩnh vực tầm nhìn máy tính. Ngoài ra đề tài này còn giúp em nắm chắc kiến thức hơn vì nó kết hợp kiến thức các môn em đã được học trong nhà trường.

DANH MỤC HÌNH VẼ

Hình 1-1: Biểu diễn đặc trưng ảnh toàn cục và cục bộ	9
Hình 1-2: Thực hiện loại bỏ điểm không cực đại	13
Hình 1-3: Phân loại điểm ảnh dựa trên giá trị riêng của ma trận tự tương quan M	14
Hình 1-4: Phát hiện đặc trưng trong một phần ảnh bằng FAST.....	17
Hình 1-5: Tìm kiếm cực trị không gian tỷ lệ 3D của hàm LoG	21
Hình 1-6: Tìm kiếm cực trị không gian tỷ lệ 3D trong hàm DoG.....	22
Hình 1-7: Sơ đồ biểu diễn của bộ mô tả SIFT cho một phần ảnh 16×16 điểm ảnh và một mảng mô tả 4×4	24
Hình 1-8: Sơ đồ của thuật toán GLOH	25
Hình 1-9: Chia vùng đặc trưng thành 4 x 4 vùng con để tính bộ mô tả SURF 26	
Hình 2-1: Nguyên lý hình học epipolar. Một mặt phẳng epipolar được xác định bởi các trung tâm chiếu O1 và O2 và một điểm đối tượng P. Các đường epipolar e' và e'' là các giao điểm của mặt phẳng epipolar với các mặt phẳng ảnh. (được chuyển thể từ Schenk, 1999).	31
Hình 2-2: Giải thích hình học của hệ số tương quan $r = \cos \theta = \frac{v_T \cdot v_S}{ v_T \cdot v_S }$ 32	
Hình 2-3: [Tài liệu “Image matching and its applications”] Nguyên lý đối sánh hình ảnh dựa trên việc tìm hệ số tương quan cực đại r.	33
Hình 2-4: Biểu diễn hình học của khoảng cách hình ảnh $D = v - v_s = v_T - v_S $	38
Hình 2-5: Ảnh gốc	44
Hình 2-6: Phát hiện cạnh	44
Hình 2-7: Đối sánh hai ảnh quay về đối sánh hai tập điểm đặc trưng trong không gian đặc trưng	46
Hình 3-1: Ảnh được hiển thị.....	51
Hình 3-2: Ảnh đã được phát hiện đặc trưng	52
Hình 3-3: Các cặp đối sánh giữa ảnh a1.png và a2.png	55
Hình 3-4: Các cặp đối sánh giữa ảnh b1.png và b2.png.....	56
Hình 3-5: Các cặp đối sánh giữa ảnh c1.png và c2.png	56

CHƯƠNG 1: PHÁT HIỆN VÀ MÔ TẢ ĐẶC TRƯNG ẢNH

1.1 Giới thiệu

Trong những thập kỷ qua, phát hiện đặc trưng và mô tả ảnh đã trở thành công cụ phổ biến trong cộng đồng thị giác máy tính. Các phương pháp đã có đang được áp dụng rộng rãi trong một lượng lớn các ứng dụng như: khôi phục ảnh, phân lớp và tra cứu ảnh, nhận dạng và đối sánh đối tượng, khôi phục cảnh 3D, theo dõi chuyển động đối tượng, phân lớp cấu trúc ảnh, định vị robot và hệ thống sinh trắc học, tất cả đều dựa trên biểu diễn sự ổn định và các đặc trưng đại diện trong ảnh. Do đó, phát hiện và trích chọn đặc trưng ảnh là bước quan trọng cho các ứng dụng này.

Để thiết lập cặp điểm tương ứng giữa hai ảnh thì việc tìm ra điểm nổi bật trên ảnh là rất cần thiết. Trong nhiệm vụ phân lớp, đặc trưng của ảnh truy vấn sẽ được đem đi đối sánh với đặc trưng của các ảnh đã được huấn luyện, cặp ảnh nào có nhiều đối sánh nhất thì được coi là đối sánh tốt nhất. Trong trường hợp này, đối sánh đặc trưng có thể dựa trên các độ đo khoảng cách như Euclidean hoặc Mahalanobis. Trong khôi phục ảnh, việc hiệu chỉnh những ảnh được thu nhận ở trong những điều kiện và thời điểm khác nhau là rất cần thiết. Các bước chính để thực hiện khôi phục ảnh hoặc hiệu chỉnh là: mô tả đặc trưng, đối sánh đặc trưng, xác định các hàm biến đổi dựa trên các đặc trưng tương ứng giữa hai ảnh ảnh và khôi phục hình ảnh dựa trên các hàm biến đổi. Trong đối sánh và nhận dạng, bước đầu tiên là phát hiện các điểm đặc trưng trong các ảnh và mô tả chúng. Khi các bộ mô tả được tính toán, chúng có thể được so sánh với nhau để tìm ra mối quan hệ giữa các ảnh để thực hiện nhiệm vụ đối sánh / nhận dạng.

Phát hiện đặc trưng là một bước quan trọng trong đối sánh ảnh. Ý tưởng chính của phát hiện đặc trưng là phát hiện ra các điểm chính, các điểm này bất biến với một lớp biến đổi nào đó. Sau đó, đối với mỗi điểm chính được phát hiện,

xây dựng một véc tơ để biểu diễn đặc trưng bất biến cho những điểm ảnh xung quanh điểm được phát hiện. Các mô tả đặc trưng được trích chọn từ ảnh có thể dựa trên thống kê bậc hai, các mô hình tham số, các hệ số thu được từ một phép biến đổi ảnh, hoặc thậm chí là kết hợp của các độ đo này. Hai kiểu đặc trưng ảnh có thể trích chọn từ mô tả nội dung ảnh là các đặc trưng toàn cục và các đặc trưng cục bộ. Đặc trưng toàn cục (ví dụ: màu sắc và kết cấu) nhằm mục đích mô tả toàn bộ ảnh và có thể được giải thích như là các thuộc tính đặc biệt của tất cả các điểm ảnh trong ảnh. Trong khi đó, các đặc trưng cục bộ nhằm mục đích phát hiện các điểm chính hoặc các vùng nổi bật trong ảnh và mô tả chúng.

Việc sử dụng các đặc trưng toàn cục đã được chứng minh thành công cho việc tìm các hình ảnh tương tự trong cơ sở dữ liệu, trong khi các đặc trưng có hướng cấu trúc cục bộ được xem như là phù hợp cho việc phân lớp đối tượng hoặc tìm những phần xuất hiện khác của đối tượng hoặc một cảnh tương tự. Mặt khác, trên các ứng dụng thời gian thực phải xử lý nhiều dữ liệu hơn hoặc chạy trên các thiết bị di động với khả năng tính toán hạn chế rất cần thiết phải mô tả cục bộ để tính toán nhanh, đối sánh nhanh, sử dụng bộ nhớ hiệu quả và đưa ra độ chính xác tốt. Hơn nữa, các mô tả đặc trưng cục bộ được chứng minh là một lựa chọn tốt cho đối sánh ảnh trên nền tảng di động, nơi xuất hiện những đối tượng khuyết cần phải được xác định. Với những ứng dụng như phân lớp ảnh, truy xuất ảnh, theo dõi hoặc nhận dạng đối tượng, mô tả đặc trưng cục bộ rất quan trọng để khắc phục các ảnh hưởng từ những thay đổi về độ sáng, thay đổi góc nhìn hoặc biến dạng ảnh (ví dụ: nhiễu, mờ, hoặc chiếu sáng). Trong khi đó, những nhiệm vụ tra cứu đặc biệt khác chẳng hạn như phát hiện khuôn mặt hoặc nhận dạng, cần phải sử dụng các kỹ thuật phát hiện và mô tả đặc biệt.

Có rất nhiều phương pháp được đề xuất để mô tả hình dạng, điển hình như SIFT (Scale Invariant Feature Transform), SURF (Speeded-up Robust Features). Mô tả SIFT sử dụng cực trị cục bộ ở trong đạo hàm DoG

(Difference Of Gaussian) để trích chọn đặc trưng. Trong khi đó, mô tả SURF được phát triển dựa trên ý tưởng của SIFT để tính toán đặc trưng cục bộ bất biến. Mô tả này được tính toán nhanh hơn SIFT và là mô tả được sử dụng rộng rãi và phổ biến trong rất nhiều ứng dụng.

1.2 Các loại đặc trưng ảnh

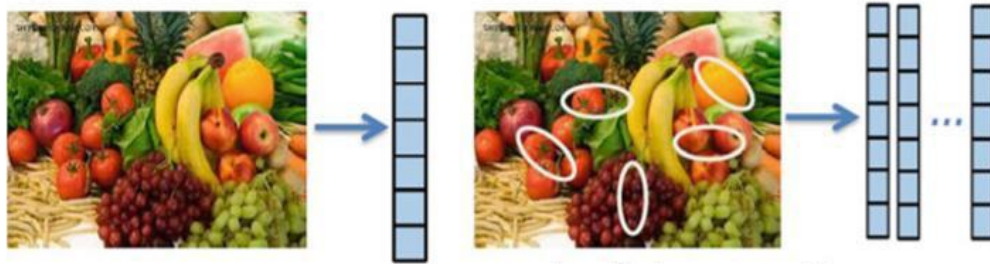
1.2.1 Đặc trưng toàn cục và cục bộ

Trong xử lý ảnh và thị giác máy tính, trích chọn đặc trưng và biểu diễn ảnh là rất cần thiết. Con người có thể dễ dàng trích chọn thông tin từ ảnh thô, tuy nhiên nó không đúng cho trường hợp máy tính. Nói chung có hai kiểu đặc trưng ảnh có thể trích chọn từ mô tả nội dung ảnh gọi là các đặc trưng toàn cục và các đặc trưng cục bộ.

Trong mô tả đặc trưng toàn cục, ảnh được biểu diễn bằng một véc tơ đặc trưng mô tả thông tin trong toàn bộ ảnh. Nói cách khác, phương pháp biểu diễn toàn cục tạo ra một véc tơ đơn với các giá trị độ đo các khía cạnh khác nhau của ảnh như màu sắc, kết cấu hoặc hình dạng. Thực tế, mỗi ảnh được biểu diễn một véc tơ đơn, sau đó 2 ảnh có thể được so sánh với nhau bằng việc so sánh các véc tơ đặc trưng của chúng. Ví dụ, khi ta muốn phân biệt các ảnh biển (xanh dương) và ảnh rừng (màu xanh lục), thì một bộ mô tả toàn cục về màu sắc sẽ tạo ra các véc tơ hoàn toàn khác nhau cho mỗi loại. Trong ngữ cảnh này, các đặc trưng toàn cục có thể được hiểu là một thuộc tính đặc biệt của ảnh liên quan đến tất cả các điểm ảnh. Thuộc tính này có thể là các biểu đồ màu, kết cấu, cạnh hoặc thậm chí một bộ mô tả đặc biệt được trích chọn từ một vài bộ lọc được áp dụng cho ảnh.

Mặt khác, mục đích chính của biểu diễn đặc trưng cục bộ là biểu thị rõ ràng hình ảnh dựa trên một số vùng nổi bật trong khi vẫn bất biến với các thay đổi về góc nhìn và chiếu sáng. Do đó, ảnh được biểu diễn dựa trên cấu trúc cục bộ của nó bởi một tập các mô tả đặc trưng cục bộ được trích chọn từ một tập hợp

các vùng hình ảnh được gọi là các vùng đặc trưng (các điểm chính). Mô tả đặc trưng toàn cục và cục bộ của ảnh được minh họa như trong Hình 1-1.



Hình 1-1: Biểu diễn đặc trưng ảnh toàn cục và cục bộ

Nói chung, việc sử dụng loại đặc trưng nào là tùy thuộc vào các ứng dụng cụ thể. Ví dụ, một người có mũi lớn hơn và mắt nhỏ hơn, và một người có mũi nhỏ hơn và mắt to hơn có thể có biểu đồ màu hoặc biểu đồ phân bố cường độ tương tự nhau. Vì vậy, các đặc trưng cục bộ hoặc mẫu toàn cục được trích ra từ các cụm đặc trưng cục bộ được xem như phù hợp hơn. Trái lại, đối với các tập dữ liệu rất lớn trong ứng dụng đánh chỉ số hình ảnh trên web thì những đặc trưng toàn cục được xem như là thích hợp.

Ưu điểm của các đặc trưng toàn cục là chúng được tính toán nhanh hơn và cô đọng hơn trong khi đó dễ tính toán và thường yêu cầu một lượng nhỏ bộ nhớ. Tuy nhiên, các biểu diễn toàn cục có một số hạn chế như: không bất biến với các biến đổi, nhạy cảm với nhiễu và một phần bị che khuất. Ngược lại, lợi thế của các đặc trưng cục bộ là có hiệu suất vượt trội. Sử dụng các đặc trưng cục bộ để tìm kiếm ảnh có hiệu suất cao hơn nhiều so với các đặc trưng toàn cục. Hơn nữa, vì các cấu trúc cục bộ ổn định hơn các cấu trúc khác ở các vùng ảnh mịn, nên nó được kỳ vọng sẽ hữu ích hơn cho việc đối sánh ảnh và nhận dạng đối tượng. Tuy nhiên, chúng thường đòi hỏi một lượng đáng kể về bộ nhớ vì ảnh có thể có hàng trăm đặc trưng cục bộ.

1.2.2 Đặc điểm của phát hiện đặc trưng

Tuytelaars và Mikolajczyk [3] đã định nghĩa đặc trưng cục bộ là một phần hình ảnh khác với vùng lân cận của nó. Vì vậy, họ coi mục đích của các đặc trưng bất biến cục bộ là cung cấp một biểu diễn cho phép đối sánh hiệu

qua các cấu trúc cục bộ giữa các ảnh. Để đáp ứng mục tiêu này, phát hiện và trích chọn đặc trưng phải có các thuộc tính rõ ràng phụ thuộc vào việc thiết lập những ứng dụng thực trong quá trình thực hiện. Các thuộc tính sau đây rất quan trọng để sử dụng bộ phát hiện đặc trưng trong các ứng dụng thị giác máy:

- Tính mạnh mẽ: Thuật toán có thể phát hiện ra những vị trí đặc trưng giống nhau độc lập với các biến đổi, co giãn, tỷ lệ, xoay, dịch chuyển, nén và nhiễu.
- Tính lặp lại: Thuật toán phát hiện ra các đặc trưng tương tự của cảnh hoặc đối tượng tương tự được lặp lại dưới một loạt những góc nhìn khác nhau.
- Tính chính xác: Xác định được chính xác vị trí các đặc trưng ảnh (vị trí các điểm ảnh tương tự).
- Tính tổng quát: Thuật toán phát hiện đặc trưng có thể phát hiện các đặc trưng được sử dụng trong các ứng dụng khác nhau.
- Tính hiệu quả: Thuật toán phát hiện đặc trưng có thể phát hiện các đặc trưng trong ảnh mới một cách nhanh chóng để hỗ trợ các ứng dụng thời gian thực.
- Tính chất lượng: Thuật toán phát hiện đặc trưng sẽ có thể phát hiện tất cả hoặc hầu hết các đặc trưng trong ảnh. Trong đó, mật độ của các đặc trưng được phát hiện cần phải phản ánh thông tin nội dung ảnh để cung cấp biểu diễn ảnh cô đọng.

1.2.3 Bất biến với tỷ lệ và biến đổi Affine

Trên thực tế, việc tìm các cặp tương ứng dựa trên việc so sánh các vùng có hình dạng cố định như hình chữ nhật hoặc hình tròn không đáng tin cậy khi có sự xuất hiện của một số biến dạng. Ngoài ra, các đối tượng trong ảnh kỹ thuật số xuất hiện theo những cách khác nhau tùy thuộc vào tỷ lệ quan sát. Do

đó, thay đổi tỷ lệ có ý nghĩa quan trọng khi phân tích nội dung ảnh. Có nhiều kỹ thuật khác nhau đã được đề xuất để giải quyết vấn đề phát hiện và trích chọn các đặc trưng ảnh bất biến trong các điều kiện này. Một số được thiết kế để xử lý các vấn đề thay đổi tỷ lệ, trong khi một số khác hướng tới các biến đổi affine. Để giải quyết các thay đổi về tỷ lệ, các kỹ thuật này giả định rằng sự thay đổi về tỷ lệ là giống nhau theo mọi hướng (tức là thống nhất) và chúng tìm kiếm các đặc trưng ổn định trên tất cả các tỷ lệ có thể có bằng cách sử dụng hàm nhân (kernel) liên tục của tỷ lệ được gọi là không gian tỷ lệ.

Mặt khác, trong trường hợp của phép biến đổi affine, tỷ lệ có thể khác nhau theo từng hướng. Tỷ lệ không đồng đều có ảnh hưởng đến vị trí, tỷ lệ và hình dạng của cấu trúc cục bộ. Do đó, các phát hiện bất biến tỷ lệ thường không thành công trong trường hợp có biến đổi Affine tỷ lệ lớn. Vì vậy, các bộ phát hiện được thiết kế để phát hiện các đặc trưng ảnh dưới tỷ lệ đều cần phải được mở rộng thành các bộ phát hiện bất biến với biến đổi Affine. Do đó, các bộ phát hiện bất biến affine có thể được xem như là một sự tổng quát hóa mô tả bất biến tỷ lệ.

Nói chung, các phép biến đổi Affine được tạo ra bằng cách sử dụng chuỗi các biến đổi tịnh tiến, co giãn, lật, quay và cắt. Biến đổi Affine (Affinity) là ánh xạ tuyến tính duy trì tính cộng tuyến và tỷ lệ với khoảng cách. Một hình thức, trong trường hợp hữu hạn chiều, nếu ánh xạ tuyến tính được biểu diễn bởi một phép nhân với ma trận A và phép tịnh tiến được biểu diễn như một phép cộng với véc tơ b thì một ánh xạ Affine f trên véc tơ x được biểu diễn như sau:

$$\bar{y} = f(\bar{x}) = A\bar{x} + b \quad [1.1]$$

1.3 Phát hiện đặc trưng ảnh

Phát hiện đặc trưng có thể được phân loại rộng thành ba loại: phát hiện đơn tỷ lệ, phát hiện đa tỷ lệ, và phát hiện bất biến affine. Trong đó, đơn tỷ lệ có nghĩa là chỉ có một biểu diễn đặc trưng hoặc cho các đường bao của đối

tượng sử dụng các tham số của bộ phát hiện. Phát hiện đơn tỷ lệ là bất biến đối với các phép biến đổi ảnh như xoay, dịch chuyển, thay đổi độ sáng và nhiễu. Tuy nhiên, chúng không có khả năng để giải quyết vấn đề co giãn. Với hai hình ảnh của cùng một cảnh giống nhau từ việc thay đổi tỷ lệ, ta muốn xác định nơi điểm đặc trưng tương tự có thể được phát hiện hoặc không. Do đó, nó là cần thiết để xây dựng bộ phát hiện đa tỷ lệ có khả năng trích chọn các đặc trưng đặc biệt đáng tin cậy dưới sự thay đổi tỷ lệ.

1.3.1 Phát hiện đơn tỷ lệ

1.3.1.1 Phát hiện của Moravec

Kỹ thuật của Moravec [4] được chỉ định để tìm các vùng riêng biệt trong hình ảnh có thể được sử dụng để khôi phục các khung ảnh liên tiếp. Nó đã được sử dụng như một thuật toán phát hiện góc trong đó một góc là một điểm có độ tương đồng thấp. Bộ phát hiện kiểm tra từng ảnh trong mỗi ảnh nhất định để xem có góc nào không. Nó xem xét một phần ảnh cục bộ tập trung vào điểm ảnh và sau đó xác định sự tương tự giữa phần ảnh và các phần ảnh chông chéo gần đó. Độ tương tự được đo bằng cách lấy tổng của bình phương khác biệt (SSD) giữa các phần ảnh trung tâm và các phần ảnh khác. Dựa trên giá trị của SSD, ba trường hợp cần được xem xét như sau:

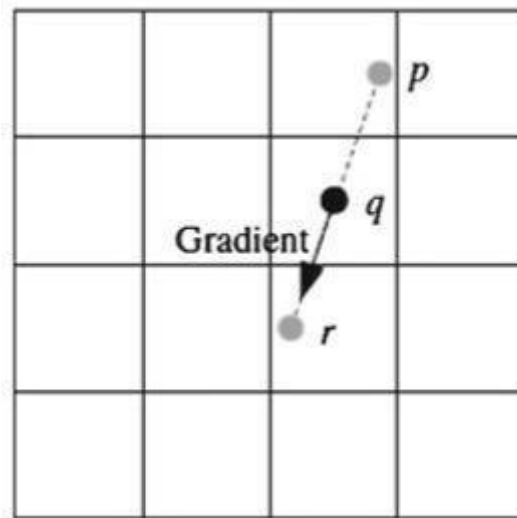
- Nếu điểm ảnh trong vùng có cường độ đồng nhất thì các phần ảnh gần đó sẽ trông giống nhau hoặc có xuất hiện thay đổi nhỏ.
- Nếu điểm ảnh nằm trên một cạnh thì các phần ảnh gần đó theo hướng song song với cạnh sẽ dẫn đến một thay đổi nhỏ và các phần ảnh theo hướng vuông góc với cạnh sẽ dẫn đến thay đổi lớn.
- Nếu điểm ảnh nằm trên một vị trí có sự thay đổi lớn về mọi hướng, thì không có phần ảnh nào gần đó trông giống nhau và góc có thể được phát hiện khi thay đổi được tạo ra với bất kỳ sự dịch chuyển nào lớn.

SSD nhỏ nhất giữa phần ảnh và các láng giềng của nó (ngang, dọc và hai đường chéo) được sử dụng làm độ đo cho góc. Một góc hoặc một điểm

đặc trưng được phát hiện khi SSD đạt đến cực đại cục bộ. Phát hiện Moravec được thực hiện theo những bước sau:

- Đầu vào: ảnh đa mức xám, cỡ cửa sổ, ngưỡng T
- Đối với mỗi điểm ảnh (x, y) trong hình ảnh tính toán thay đổi cường độ V từ một dịch chuyển (u, v) theo:

$$V_{u,v}(x, y) = \frac{I(x+u+a, y+v+v) - I(x+a, y+b)}{a, b \text{ window}}^2 \quad [1.2]$$



Hình 1-2: Thực hiện loại bỏ điểm không cực đại

- Xây dựng bản đồ góc bằng cách tính toán độ đo góc $C(x, y)$ cho mỗi điểm ảnh (x, y)

$$C(x, y) = \min_{u,v} \left(\frac{V(x, y)}{V(x+u, y+v)} \right) \quad [1.3]$$

- Phân ngưỡng bản đồ góc bằng việc thiết lập tất cả các giá trị $C(x, y)$ bên dưới ngưỡng T sẽ bằng 0.
- Loại bỏ các điểm không cực đại để tìm các điểm cực đại cục bộ. Tất cả các điểm khác 0 còn lại trong bản đồ là góc.

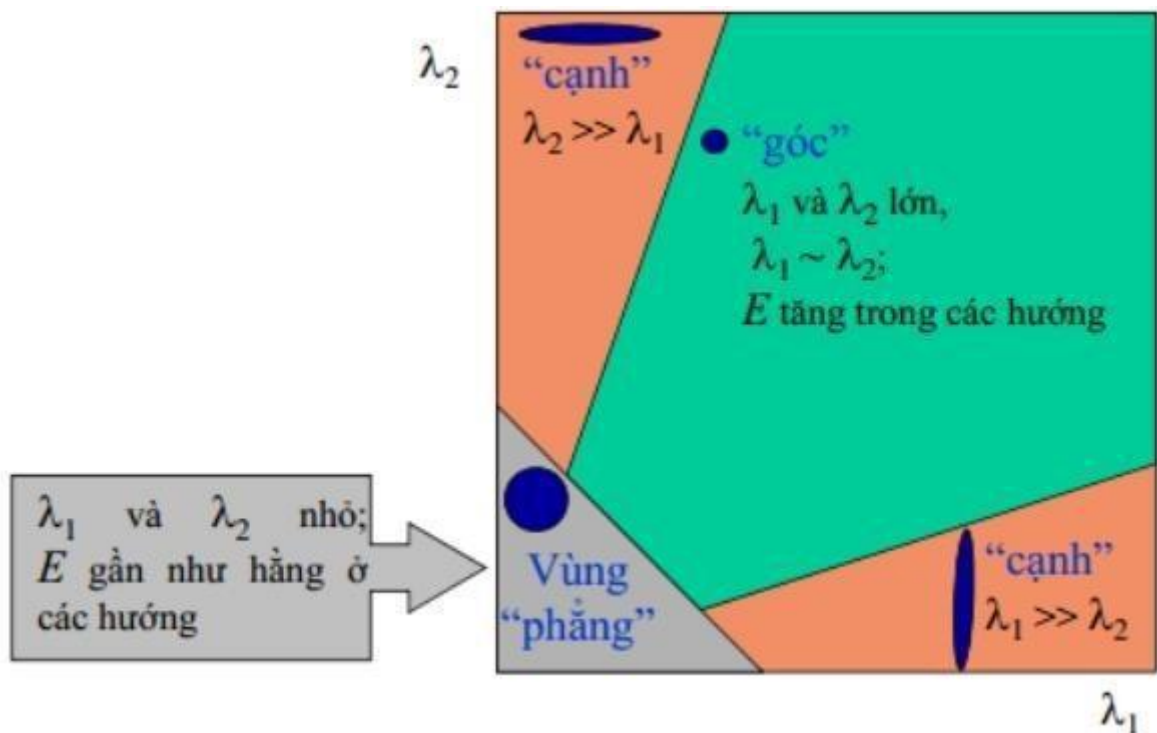
Để loại bỏ các điểm cực đại, ảnh được quét dọc theo hướng gradient của nó, sao cho vuông góc với cạnh. Bất kỳ điểm ảnh không phải là cực đại cục bộ được loại bỏ và được đặt thành 0. Như minh họa trong Hình 1-2, p và r là

hai điểm láng giềng theo hướng gradient của q . Nếu giá trị điểm ảnh của q không lớn hơn giá trị điểm ảnh của cả p và r , thì giá trị này bị loại bỏ. Ưu điểm của kỹ thuật Moravec là có thể phát hiện phần lớn các góc. Tuy nhiên, nó không đẳng hướng; những thay đổi cường độ chỉ được tính toán ở một tập rời rạc của sự dịch chuyển (tám hướng cơ bản) và bất kỳ cạnh nào không nằm trong một trong tám hướng thì được gán một độ đo góc lớn. Vì vậy, nó không phải là bất biến để xoay.

1.3.1.2 Phát hiện Harris

Kỹ thuật phát hiện Harris [5] là sự kết hợp phát hiện cạnh và góc để giải quyết hạn chế của kỹ thuật Moravec. Kết quả của kỹ thuật này dựa trên một công nghệ được sử dụng rộng rãi đó là ma trận tương quan tự động. Ma trận tương quan tự động đối xứng 2×2 được sử dụng để phát hiện các đặc trưng ảnh và mô tả cấu trúc cục bộ của chúng có thể được biểu diễn như sau:

$$M(x, y) = \begin{matrix} w(u, v)^* \\ u, v \end{matrix} \begin{matrix} I_x^2(x, y) & I_x I_y(x, y) \\ I_x I_y(x, y) & I_y^2(x, y) \end{matrix} \quad [1.4]$$



Hình 1-3: Phân loại điểm ảnh dựa trên giá trị riêng của ma trận tự tương quan M

Trong đó I_x và I_y là các đạo hàm riêng ảnh cục bộ xác định theo hướng x và y tương ứng, và $w(u, v)$ ký hiệu một cửa sổ trọng số trên vùng (u, v) . Nếu một cửa sổ tròn như Gaussian được sử dụng, thì kết quả sẽ là đẳng hướng và các giá trị sẽ có trọng số gần hơn với tâm. Để tìm các điểm đặc trưng, các giá trị riêng của ma trận M được tính cho mỗi điểm ảnh. Nếu cả hai giá trị riêng đều lớn, thì sẽ cho biết sự tồn tại của góc tại vị trí đó. Một sơ đồ minh họa cho phân loại các điểm phát hiện được thể hiện trong Hình 1-3. Xây dựng bản đồ kết quả có thể được thực hiện bằng cách tính toán độ đo góc $C(x, y)$ cho mỗi điểm ảnh (x, y) sử dụng

$$C(x, y) = \det(M) - K(\text{trace}(M))^2 \quad [1.5]$$

với

$$\det(M) = \lambda_1 * \lambda_2, \text{ và } \text{trace}(M) = \lambda_1 + \lambda_2 \quad [1.6]$$

K là một tham số điều chỉnh và λ_1, λ_2 là các giá trị riêng của ma trận tương quan tự động. Tính toán chính xác của các giá trị riêng là có độ phức tạp lớn. Do đó Harris đã gợi ý sử dụng độ đo góc bằng cách kết hợp 2 giá trị riêng ở trong một độ đo đơn.

1.3.1.3 Phát hiện SUSAN

Thay vì sử dụng đạo hàm ảnh để tính toán góc, Smith và Brady [6] đã giới thiệu một công nghệ xử lý hình ảnh mức thấp được gọi là SUSAN (Smallest Univalued Segment Assimilating Nucleus). Ngoài việc phát hiện góc, nó còn được sử dụng để phát hiện cạnh và giảm nhiễu ảnh. Góc được phát hiện bằng cách đặt một mặt nạ tròn bán kính cố định cho mỗi điểm ảnh ở trong ảnh. Điểm ảnh tâm được gọi là nhân, tất cả các điểm ảnh trong vùng mặt nạ được so sánh với nhân để kiểm tra xem các giá trị cường độ của nó là tương tự hay là khác. Các điểm ảnh có độ sáng gần giống với nhân được nhóm lại với nhau và vùng đó được gọi là USAN (Univalued Assimilating Nucleus). Một góc được tìm thấy tại nơi mà số điểm ảnh trong USAN đạt giá

trị cực tiểu cục bộ và thấp hơn một ngưỡng chỉ định. Để phát hiện các góc, hàm so sánh tương tự $C(r, r_0)$ giữa mỗi điểm ảnh trong mặt nạ và nhân mặt nạ sẽ được tính toán bởi

$$C(r, r_0) = \begin{cases} 1, & \text{if } |I(r) - I(r_0)| \geq T, \\ 0, & \text{otherwise,} \end{cases} \quad [1.7]$$

Và cỡ của vùng USAN là

$$n(r) = \sum_{r_c(r_0)} C(r, r) \quad [1.8]$$

trong đó r_0 và r là tọa độ của nhân và tọa độ của các điểm khác trong mặt nạ tương ứng. Việc thực thi của bộ phát hiện góc SUSAN chủ yếu dựa trên hàm so sánh tương tự $C(r, r_0)$.

Phát hiện SUSAN có một số ưu điểm như: (i) không sử dụng đạo hàm, do đó không cần giảm nhiễu và không phải tính toán phức tạp; (ii) Độ lặp lại cao để phát hiện các đặc trưng; và (iii) bất biến đối với việc dịch chuyển và phép quay. Tuy nhiên, nó không bất biến với tỷ lệ và biến đổi khác, và một ngưỡng cục bộ không phù hợp cho trường hợp tổng quát. Phát hiện góc cần một ngưỡng thích ứng và hình dạng của mặt nạ phải được thay đổi.

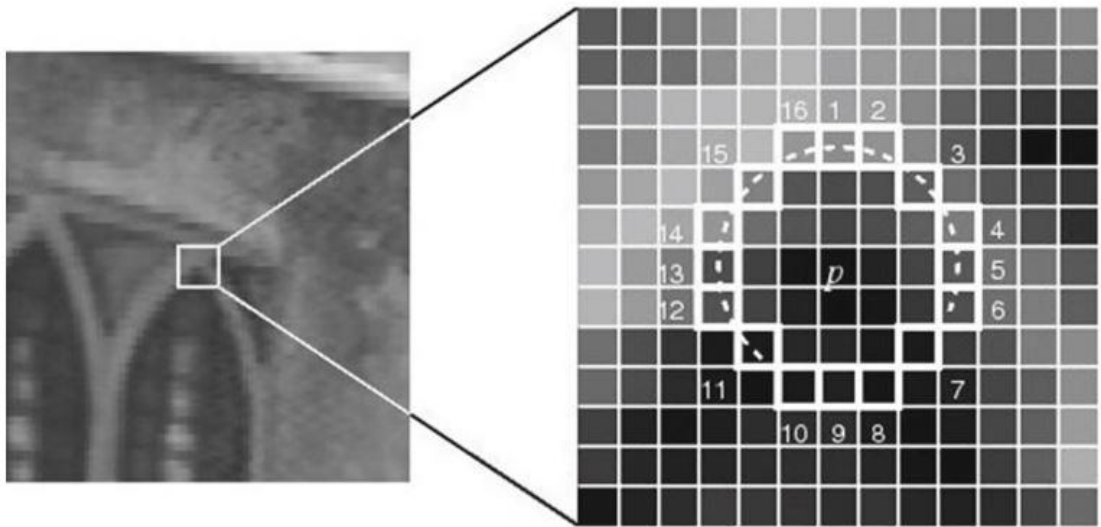
1.3.1.4 Phát hiện FAST

FAST (Features from Accelerated Segment Test) là một bộ phát hiện góc ban đầu được phát triển bởi Rosten và Drummond. Trong lược đồ phát hiện này, các điểm ứng viên được phát hiện bằng cách áp dụng kiểm tra phân đoạn cho mỗi điểm ảnh. Việc kiểm tra này được tính toán dựa trên 16 điểm ảnh xung quanh điểm ảnh ứng viên góc. Nếu một tập n điểm kề nhau trong vòng tròn Bresenham với bán kính r đều sáng hơn cường độ của điểm ảnh ứng

$$t, I_p + t$$

điểm tối hơn cường độ của điểm ảnh ứng viên trừ đi giá trị ngưỡng $I_p - t$ thì p được xem như là một góc. Một bước kiểm tra có thể được sử dụng để loại trừ một số lượng lớn các điểm không phải góc; bộ kiểm tra này chỉ kiểm tra

bốn điểm ảnh 1,5,9 và 13 như trong Hình 1-4. Một góc chỉ tồn tại nếu ba trong số các điểm ảnh kiểm tra này sáng hơn $I_p + t$ hoặc tối hơn so với $I_p - t$ và các điểm ảnh còn lại sau đó được kiểm tra để xem kết quả cuối cùng. Hình 1-4 minh họa quá trình xử lý, trong đó các hình vuông được tô sáng là các điểm ảnh được sử dụng trong phát hiện góc. Điểm ảnh tại p là tâm của góc ứng viên. Vòng cung được biểu thị bằng đường đứt nét đi qua 12 điểm liền kề sáng hơn p theo ngưỡng.



Hình 1-4: Phát hiện đặc trưng trong một phần ảnh bằng FAST

Mặc dù các kiểm tra này mang lại hiệu năng cao nhưng nó gặp phải một số hạn chế và nhược điểm là tốc độ tính toán hạn chế do phải kiểm tra nhiều điểm. Tuy nhiên điều này có thể được khắc phục bằng cách sử dụng những phương pháp học máy. Thứ tự các câu hỏi được xử dụng để phân lớp một điểm ảnh được học bởi thuật toán cây quyết định (ID3). Việc áp dụng học máy cho bước này nhằm mục đích tăng tốc độ. Khi bước kiểm tra tạo ra rất nhiều kết quả liền kề xung quanh điểm đặc trưng thì một điều kiện được thêm vào để loại bỏ điểm không cực đại. Điều này cho phép các đặc trưng được định vị một cách chính xác. Độ đo góc được sử dụng tại bước này là:

$$C(x, y) = \max_{j \in S_{bright}} |I_{p \rightarrow j} - I_p| - t, \quad \max_{j \in S_{dark}} |I_p - I_{p \rightarrow j}| - t \quad [1.9]$$

trong đó $I_{p \rightarrow j}$ là ký hiệu điểm ảnh nằm trên vòng tròn Bresenham. Trong bước này, thời gian xử lý được rút ngắn vì bước kiểm tra thứ hai chỉ được thực hiện trên một phần nhỏ của điểm ảnh đã vượt qua bước kiểm tra đầu tiên.

Quá trình kiểm tra được thực hiện theo hai giai đoạn:

- Trước hết, phát hiện góc với việc kiểm tra đoạn của n điểm và một ngưỡng phù hợp được thực hiện trên một tập ảnh. Mỗi điểm ảnh của 16 vị trí trên vòng tròn được phân loại thành màu tối hơn, tương tự hoặc sáng hơn.
- Tiếp theo, sử dụng thuật toán ID3 trên 16 vị trí để lựa chọn ra những vị trí tại những điểm cho nhiều thông tin nhất. Triệt tiêu điểm không cực đại được áp dụng trên tổng chênh lệch tuyệt đối giữa các điểm ảnh trong cung liền kề và điểm ảnh trung tâm.

Ở đây các góc được phát hiện sử dụng thuật toán ID3 có thể hơi khác so với kết quả thu được với phát hiện kiểm tra phân đoạn do thực tế mô hình cây quyết định phụ thuộc vào dữ liệu huấn luyện, không thể bao quát tất cả các góc có thể. So với nhiều phương pháp phát hiện khác, bộ phát hiện góc FAST rất phù hợp cho các ứng dụng xử lý video thời gian thực vì hiệu năng tốc độ cao của nó. Tuy nhiên, nó không bất biến với thay đổi tỷ lệ và không uy lực với nhiễu, cũng như nó dựa trên vào một ngưỡng, mà việc lựa chọn ngưỡng không phải là một nhiệm vụ dễ dàng.

1.3.1.5 Phát hiện Hessian

Phát hiện Hessian dựa trên ma trận 2×2 đạo hàm bậc hai của cường độ ảnh $I(x, y)$, được gọi là ma trận Hessian. Ma trận này có thể được sử dụng để phân tích cấu trúc ảnh cục bộ và nó được giải thích dưới dạng:

$$H(x, y) = \begin{pmatrix} I_{xx}(x, y) & I_{xy}(x, y) \\ I_{xy}(x, y) & I_{yy}(x, y) \end{pmatrix} \quad [1.10]$$

Ở đây I_{xx} , I_{xy} và I_{yy} là các đạo hàm ảnh bậc hai được tính toán bằng cách sử dụng hàm Gaussian với độ lệch chuẩn σ . Để phát hiện các đặc trưng, nó tìm kiếm một tập con của các điểm mà kết quả của đạo hàm là cao theo hai hướng trực giao. Tức là, phát hiện tìm kiếm các điểm mà định thức của ma trận Hessian có cực đại cục bộ

$$\det(H) = I_{xx}I_{yy} - I_{xy}^2 \quad [1.11]$$

Bằng cách chọn các điểm cực đại định thức của Hessian, độ đo này sẽ loại các cấu trúc có các đạo hàm bậc hai nhỏ (tức là, các thay đổi tín hiệu) trong một hướng nào đó. Bằng việc áp dụng loại trừ điểm cực đại sử dụng một cửa sổ có cỡ 3×3 trên từng ảnh nên chỉ những điểm ảnh có giá trị lớn hơn giá trị của tất cả các điểm ảnh láng giềng trong cửa sổ mới được giữ lại. Sau đó, bộ phát hiện trả về tất cả các vị trí còn lại mà có giá trị lớn hơn ngưỡng cho trước. Trong khi, ma trận Hessian được sử dụng để mô tả cấu trúc cục bộ trong một vùng láng giềng xung quanh một điểm, thì định thức của nó được sử dụng để phát hiện các cấu trúc ảnh biểu thị sự thay đổi tín hiệu theo hai hướng. So sánh với các toán tử khác như Laplacian, định thức của Hessian chỉ đáp ứng nếu mẫu ảnh cục bộ chứa các thay đổi đáng kể dọc theo hai hướng trực giao. Tuy nhiên, sử dụng các đạo hàm bậc hai trong bộ phát hiện là nhạy cảm với nhiễu. Ngoài ra, cực đại cục bộ thường được tìm thấy gần đường bao hoặc cạnh thẳng, nơi tín hiệu chỉ thay đổi theo một hướng. Do đó, các cực đại cục bộ này ít ổn định hơn vì bị ảnh hưởng bởi nhiễu hoặc các thay đổi nhỏ trong mẫu lân cận.

1.3.2 Phát hiện đa tỉ lệ

1.3.2.1 Laplacian of Gaussian (LoG)

Laplacian-of-Gaussian (LoG) là một tổ hợp tuyến tính của các đạo hàm bậc hai được sử dụng phổ biến trong phát hiện các khối. Với một ảnh đầu vào $I(x, y)$, biểu diễn không gian tỷ lệ của ảnh được định nghĩa bởi $L(x, y, \sigma)$

được tính toán bằng cách nhân chập ảnh với một hàm nhân Gaussian như $G(x, y)$ và được định nghĩa sau:

$$L(x, y) = G(x, y) I(x, y) \quad [1.12]$$

Với

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \quad [1.13]$$

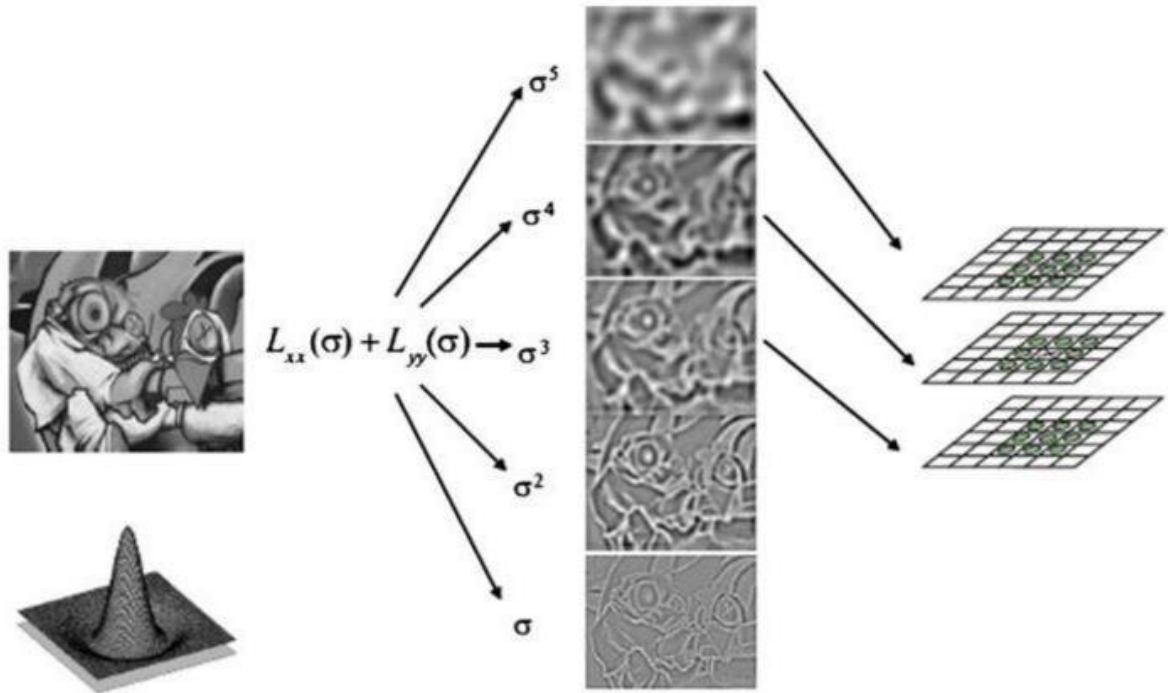
Công thức sau được sử dụng cho việc tính toán toán tử Laplacian:

$$\Delta L(x, y) = L_{xx}(x, y) + L_{yy}(x, y) \quad [1.14]$$

Điều này dẫn đến phản hồi ứng tích cực cho các khối màu tối và phản ứng tiêu cực với các khối sáng với cỡ $\sqrt{2}\sigma$. Tuy nhiên, phản hồi của toán tử phụ thuộc rất nhiều vào mối quan hệ giữa kích thước của các cấu trúc khối trong miền ảnh và kích thước của nhân làm mịn Gaussian. Độ lệch chuẩn của Gaussian được sử dụng để điều khiển tỷ lệ bằng việc thay giá trị độ mờ. Để thu nhận một cách tự động khối có kích thước khác nhau trong miền ảnh thì một phương pháp lựa chọn tỷ lệ tự động được sử dụng thông qua việc tìm kiếm cực trị không gian tỷ lệ của toán tử Laplacian chuẩn hóa theo:

$$\text{norm}^2 L(x, y) = \Delta^2 (L_{xx}(x, y) + L_{yy}(x, y)) \quad [1.15]$$

Việc chuẩn hóa này cũng có thể đồng thời phát hiện ra các điểm cực đại hoặc cực tiểu cục bộ đối với cả không gian và tỷ lệ. Toán tử LoG là đối xứng tròn; do đó nó không bất biến với phép quay. LoG không những thích hợp với phát hiện khối do đặc tính đối xứng tròn này, mà còn cung cấp một đánh giá tốt về đặc điểm tỷ lệ cho những cấu trúc cục bộ khác như là góc, cạnh, đường gờ và đa điểm nổi. Trong ngữ cảnh này, LoG có thể được áp dụng để tìm đặc tính tỷ lệ cho vị trí ảnh đưa vào hoặc cho vùng bất biến tỷ lệ bởi tìm kiếm cực trị 3D (vị trí + độ đo) của hàm LoG như minh họa trong Hình 1-5.



Hình 1-5: Tìm kiếm cực trị không gian tỷ lệ 3D của hàm LoG

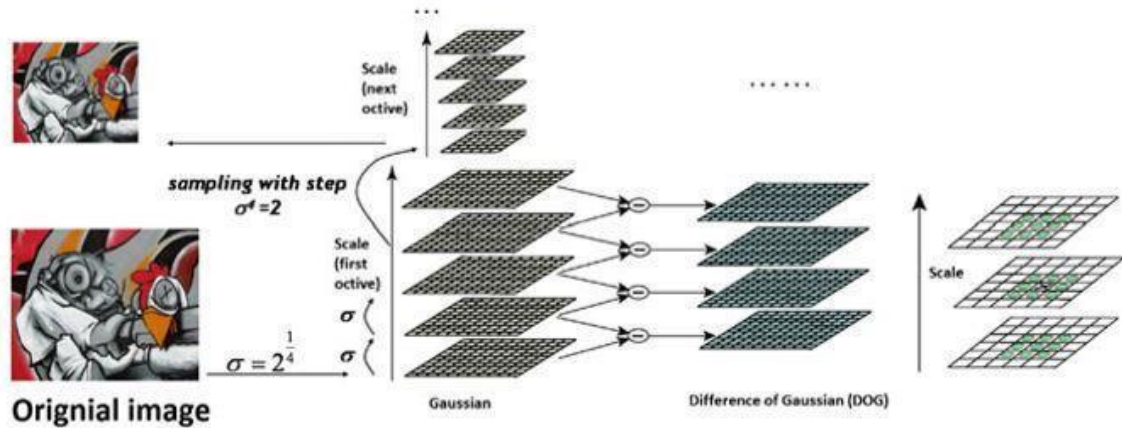
1.3.2.2 Difference of Gaussian (DoG)

Trong thực tế, việc tính toán các toán tử LoG tốn thời gian. Để tăng tốc độ tính toán, Lowe đã đề xuất một thuật toán hiệu quả dựa trên cực trị 3D cục bộ trong kim tự tháp không gian tỷ lệ được xây dựng với các bộ lọc Gaussian (DoG). Cách tiếp cận này được sử dụng trong thuật toán biến đổi đặc trưng bất biến tỷ lệ (SIFT). Trong đó, DoG cho xấp xỉ gần với Laplacian-of-Gaussian (LoG) và nó được sử dụng để phát hiện các đặc trưng ổn định từ cực trị không gian tỷ lệ. Hàm DoG $D(x, y, z)$ có thể được tính toán mà không cần nhân chập bằng cách trừ các mức tỷ lệ liên kế của một kim tự tháp Gaussian được phân cách bởi một hệ số k .

$$\begin{aligned}
 D(x, y, z) &= (G(x, y, z, k) - G(x, y, z)) * I(x, y, z) \\
 &= L(x, y, z, k) - L(x, y, z)
 \end{aligned}
 \tag{1.16}$$

Các loại đặc trưng được trích chọn bởi DoG có thể được phân loại tương tự như đối với toán tử LoG. Ngoài ra, bộ phát hiện vùng DoG tìm kiếm cực trị không gian tỷ lệ 3D của hàm DoG như trong Hình 1-6. Các nhược điểm

chung của cả hai biểu diễn LoG và DoG là cực đại cục bộ có thể được phát hiện trong các đường bao láng giềng của các cạnh thẳng, nơi tín hiệu chỉ thay đổi theo một hướng, do đó nó làm cho chúng ít ổn định hơn và nhạy cảm hơn với nhiễu hoặc những thay đổi nhỏ.



Hình 1-6: Tìm kiếm cực trị không gian tỷ lệ 3D trong hàm DoG

1.3.2.3 Phát hiện sóng Gabor

Gần đây, Yussof và Hitam [7] đã đề xuất phát hiện điểm đặc trưng đa tỷ lệ dựa trên nguyên tắc của sóng Gabor. Các sóng Gabor được tính toán như sau:

$$G_{u,v}(z) = \frac{\|K_{u,v}\|^2}{2} e^{-\frac{K_{u,v}^2 z^2}{2}} e^{iz, K_{u,v}} - e^{-\frac{K_{u,v}^2 z^2}{2}} \quad [1.17]$$

Trong đó $K_{u,v} = K_v e^{iu}$, $z = (x, y)$, u và v định nghĩa hướng và tỷ lệ của các sóng Gabor, $K_v = K_{\max} / f$ và $u = u / 8$, K_{\max} là tần số lớn nhất, và $f = 2 \sqrt{\quad}$

là hệ số không gian giữa các nhân trong miền tần số. Phương pháp này được tính toán bằng cách nhân chập ảnh I với sóng như sau:

$$G = I * \quad [1.18]$$

Ưu điểm của sóng Gabor là chúng cung cấp giải pháp tối ưu đồng thời trong cả miền tần số và không gian. Ngoài ra, các bộ sóng Gabor có khả năng tăng cường các đặc trưng cấp thấp như đỉnh, khe và đường gợn. Do đó, chúng được sử dụng để trích chọn các điểm từ ảnh ở các tỷ lệ khác nhau bằng cách

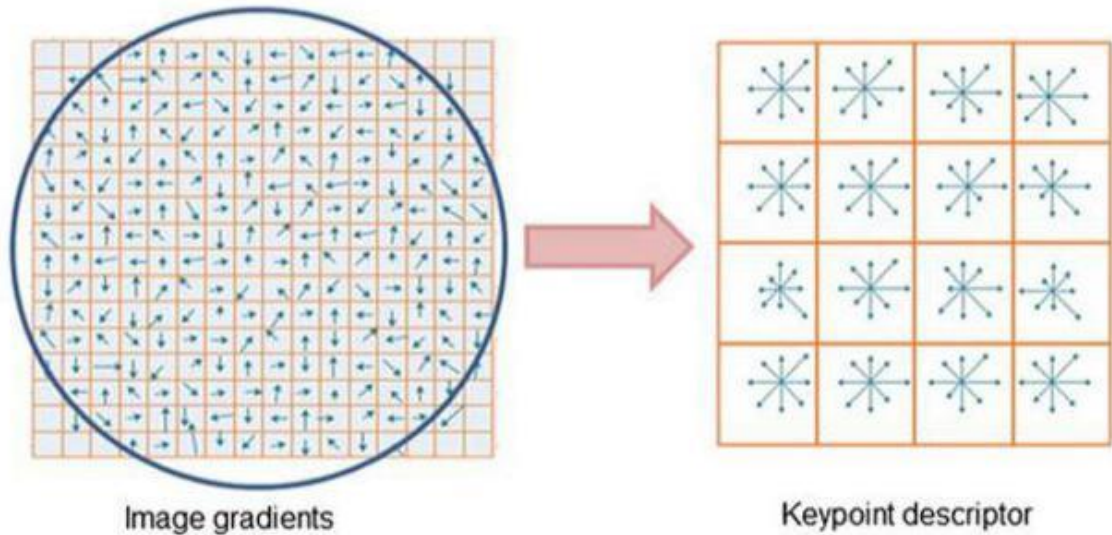
kết hợp nhiều hướng của ảnh. Các điểm đặc trưng được trích ra ở đa tỷ lệ với sự kết hợp của các hướng đều nhau. Họ đã chứng minh rằng các điểm thích chọn sử dụng phát hiện sóng Gabor có độ chính xác cao và khả năng thích ứng với các phép biến đổi hình học khác nhau.

1.4 Mô tả đặc trưng ảnh

Khi một tập hợp các điểm đặc trưng đã được phát hiện từ một hình ảnh tại vị trí $p(x, y)$, tỷ lệ s và hướng θ , nội dung hoặc cấu trúc hình ảnh của chúng trong vùng lân cận của p cần được mã hóa với bộ mô tả phù hợp để phân biệt và đối sánh với các biến dạng ảnh cục bộ. Có rất nhiều mô tả đặc trưng trong đó SIFT, SURF là những mô tả thường được sử dụng trong các ứng dụng.

1.4.1 Scale Invariant Feature Transform (SIFT)

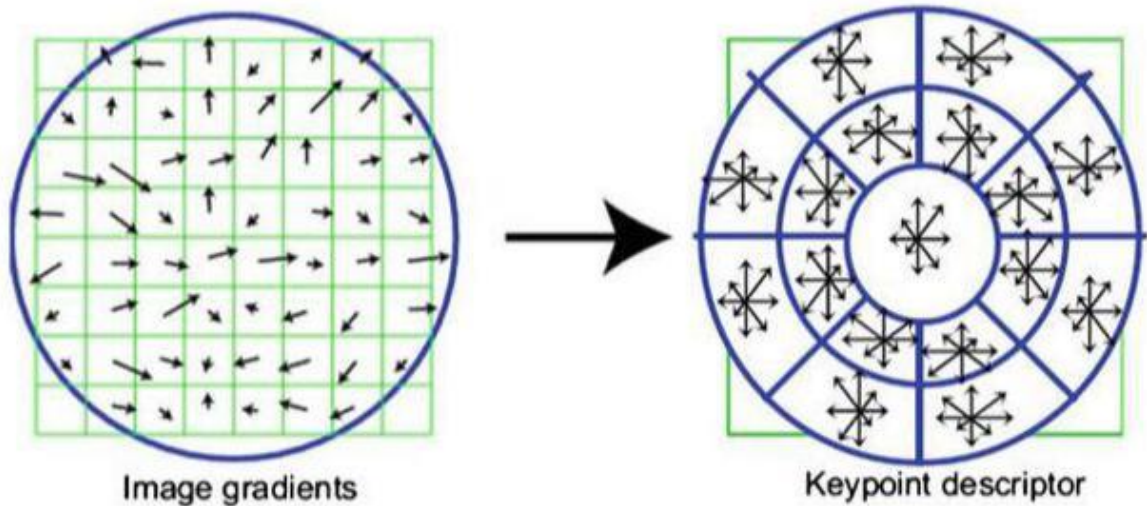
Thuật toán trung cục bộ bất biến SIFT (Scale Invariant Feature Transform) do Lowe (1999) đề xuất là phương pháp được áp dụng hiệu quả, với ưu điểm là không phụ thuộc việc thay đổi tỷ lệ, quay ảnh, góc nhìn, ảnh bị nhiễu, hoặc đôi khi là thay đổi độ sáng ảnh ở mức độ chấp nhận. Trong thuật toán này số điểm hấp dẫn sẽ được phát hiện bằng cách sử dụng toán tử DoG. Tại mỗi điểm hấp dẫn một véc tơ đặc trưng sẽ được trích chọn. Chiến lược mô tả SIFT bắt đầu bằng việc lấy mẫu độ lớn và hướng gradient của ảnh trong vùng cửa sổ 16×16 xung quanh mỗi điểm đặc trưng. Sau đó một tập các biểu đồ hướng sẽ được tạo ra, với mỗi một biểu đồ sẽ chứa các mẫu được hình thành từ các vùng con 4×4 của vùng láng giềng đã được lấy mẫu. Mỗi biểu đồ sẽ có 8 “bin” tương ứng với 8 hướng như Hình 1-7. Do có 4×4 biểu đồ và mỗi biểu đồ có 8 thành phần, nên véc tơ đặc trưng sẽ có $4 \times 4 \times 8 = 128$ phần tử cho mỗi điểm đặc trưng. Cuối cùng véc tơ đặc trưng được chuẩn hóa với chiều dài đơn vị để tạo ra sự bất biến với những thay đổi Affine.



Hình 1-7: Sơ đồ biểu diễn của bộ mô tả SIFT cho một phần ảnh 16×16 điểm ảnh và một mảng mô tả 4×4

1.4.2 Gradient Location-Orientation Histogram (GLOH)

Gradient Location-Orientation Histogram (GLOH) được phát triển bởi Mikolajczyk và Schmid [8] cũng là phần mở rộng của bộ mô tả SIFT. GLOH tương tự với bộ mô tả SIFT nhưng thay vì sử dụng các lưới vị trí Decac thì nó sử dụng một lưới vị trí log-polar và áp dụng thuật toán PCA (Principal Components Analysis) để giảm cỡ mô tả. GLOH sử dụng lưới vị trí log-polar với 3 “bin” trong mỗi hướng bán kính và 8 cho hướng góc. Kết quả sẽ có 17 bin vị trí như Hình 1-8. Mô tả này sẽ xây dựng một tập các biểu đồ sử dụng các hướng Gradient với 16 bin và kết quả véc tơ đặc trưng sẽ có $17 \times 16 = 272$ phần tử cho mỗi điểm đặc trưng. Đặc trưng này sau đó được giảm xuống 128 chiều bằng việc sử dụng thuật toán PCA. Phương pháp GLOH cho hiệu năng tốt hơn mô tả SIFT đặc biệt cho sự thay đổi chiếu sáng, tuy nhiên độ tính toán của nó phức tạp hơn.



Hình 1-8: Sơ đồ của thuật toán GLOH

1.4.3 Speeded-Up Robust Features Descriptor (SURF)

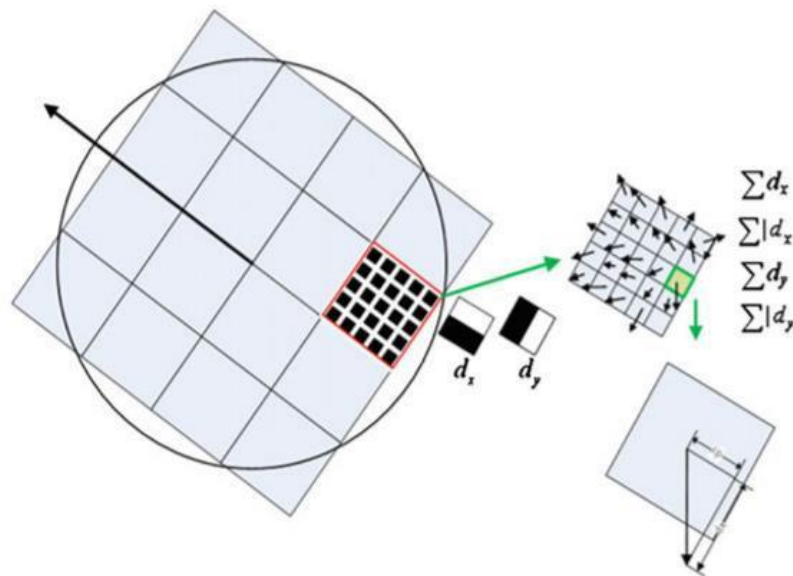
Sơ đồ mô tả trình phát hiện đặc trưng tăng tốc (SURF) được phát triển bởi Bay et al [9]. Nó nhanh hơn nhiều và mạnh mẽ hơn so với SIFT. Với chiến lược phát hiện các điểm đặc trưng, thay vì dựa vào đạo hàm Gaussian thì nó được tính toán dựa trên bộ lọc 2D đơn giản; trong đó, nó sử dụng bộ phát hiện khối bất biến dựa trên định thức của ma trận Hessian cho cả lựa chọn tỷ lệ và hướng. Ý tưởng cơ bản của nó là xấp xỉ các đạo hàm Gaussian bậc hai bằng một cách hiệu quả với sự trợ giúp của các ảnh tích phân bằng cách sử dụng một tập các bộ lọc. Các bộ lọc 9×9 được mô tả là xấp xỉ của Gaussian với $\sigma = 1.2$. Các xấp xỉ này được ký hiệu bởi D_{xx} , D_{yy} và D_{xy} . Do vậy, định thức xấp xỉ của Hessian có thể được biểu diễn dưới dạng

$$\det \left(H_{approx} \right) = D_{xx} D_{yy} - \left(w D_{xy} \right)^2 \quad [1.19]$$

Trong đó w là trọng số liên quan cho đáp ứng bộ lọc và nó được sử dụng để cân bằng biểu thức cho định thức của Hessian. Các định thức được xấp xỉ của Hessian biểu diễn cho đáp ứng khối trong ảnh. Các đáp ứng này được lưu trữ trong bản đồ đáp ứng khối và cực đại cục bộ được phát hiện sử dụng bằng phép nội suy bậc hai, như với DoG. Cuối cùng, những điểm không

cực đại sẽ được loại bỏ trong láng giềng $3 \times 3 \times 3$ để nhận được những điểm đặc trưng ổn định.

Mô tả SURF bắt đầu bằng cách xây dựng một vùng hình vuông xung quanh điểm đặc trưng được hướng theo và định hướng dọc theo hướng chính của nó. Cửa sổ của cửa sổ này là $20s$, trong đó s là tỷ lệ tại nơi điểm đặc trưng được phát hiện. Sau đó, vùng đặc trưng được chia thành vùng con 4×4 nhỏ hơn và cho mỗi vùng phản ứng sóng của Harris theo chiều dọc và ngang (biểu thị d_x và d_y , tương ứng) được tính toán tại 5×5 điểm được lấy mẫu như được hiển thị trong Hình 1-9.



Hình 1-9: Chia vùng đặc trưng thành 4×4 vùng con để tính bộ mô tả SURF

Những đáp ứng này được đánh trọng số với cửa sổ Gaussian có tâm tại điểm đặc trưng để tăng độ mạnh mẽ để có thể chống lại biến dạng hình học và lỗi vị trí. Các đáp ứng d_x và d_y được tổng hợp cho mỗi vùng con và được đưa vào véc tơ đặc trưng v , với

$$v = \begin{pmatrix} d_x & |d_x| \\ |d_y| & d_y \end{pmatrix} \quad [1.20]$$

Tính toán này cho tất cả các tiểu vùng 4×4 tạo ra một mô tả đặc trưng có chiều dài $4 \times 4 \times 4 = 64$ chiều. Cuối cùng, bộ mô tả đặc trưng được chuẩn hóa thành một véc tơ đơn vị để giảm ảnh hưởng chiếu sáng.

Ưu điểm chính của mô tả SURF so với SIFT là tốc độ xử lý vì nó sử dụng véc tơ đặc trưng 64 chiều để mô tả đặc trưng cục bộ, trong khi SIFT sử dụng 128 chiều. Tuy nhiên, mô tả SIFT phù hợp hơn để mô tả hình ảnh bị ảnh hưởng bởi dịch chuyển, xoay, và các biến dạng chiếu sáng khác. Mặc dù SURF cho thấy tiềm năng của nó trong một loạt các ứng dụng thị giác máy tính, nhưng nó cũng có một số hạn chế. Khi đối tượng 2D hoặc 3D được so sánh, nó không tốt nếu phép quay là nhiều hoặc góc nhìn quá khác nhau.

CHƯƠNG 2: ĐỐI SÁNH ẢNH DỰA TRÊN ĐẶC TRƯNG SIFT

2.1 Giới thiệu về đối sánh ảnh

Đối sánh ảnh là một bài toán đã và đang thu hút được sự quan tâm của các nhà nghiên cứu và phát triển. Mỗi khi bài toán này được giải quyết, nó mở ra rất nhiều các ứng dụng hữu ích như: tìm kiếm ảnh, nhận dạng, theo dõi và phát hiện đối tượng, ghép ảnh, v.v.

Đối sánh hai ảnh là tìm ra những vùng giống nhau trên hai ảnh. Thông thường, để đối sánh ảnh cần so sánh các phân tử cơ bản cấu thành nên nó. Giải pháp đầu tiên cho vấn đề đối sánh ảnh được đề xuất bởi Hobrough vào cuối những năm 1950. Hệ thống tự động tìm kiếm các điểm tương quan được giới thiệu lần đầu bởi công ty Wild Heerbrugg năm 1964 nhưng lại không được sử dụng phổ biến. Tuy nhiên, ý tưởng áp dụng mối tương quan chéo của Hobrough lại được nhiều người sử dụng. Từ những năm 1970, việc tập trung phát triển đối sánh ảnh và đối sánh tương quan gặt hái được nhiều thành công và được áp dụng trong hệ thống đo độ tương tự cho ảnh (Helava, 1978). Ngày nay, công nghệ đối sánh ảnh được tích hợp trong nhiều phần mềm xử lý ảnh được sử dụng như là một công cụ tính toán. Có rất nhiều nghiên cứu được thực hiện với mong muốn tìm cặp điểm tương đồng trên hai bức ảnh. Thuật toán tìm kiếm điểm tương đồng có thể thực hiện được trên ảnh 2D.

Vấn đề chính của việc đối sánh ảnh là việc chọn thực thể đối sánh (Một thực thể trong ảnh này được so sánh với một thực thể trong ảnh khác) và lựa chọn độ đo tương tự (Một độ đo định lượng đánh giá đối sánh của toàn bộ các thực thể). Đối sánh từng điểm ảnh sẽ không khả thi với những ảnh có kích thước lớn vì nó sẽ cần tính toán nhiều hơn, mất nhiều thời gian hơn, hoặc muốn rút ngắn thời gian thì cần có phần cứng xử lý mạnh hơn. Hơn nữa, nó thường dẫn đến sự nhập nhằng do các giá trị mức xám của ảnh xuất hiện lặp đi lặp lại và

do nhiều của ảnh. Do đó bài toán đối sánh ảnh thuộc về nhóm bài toán giả định yếu (ill-posed problems). Để chuyển đổi bài toán đối sánh ảnh thành bài toán giả định chặt (Well-posed problem) thì các thực thể đối sánh, các độ đo tương tự, các ràng buộc hình học và các giả thiết phải được định nghĩa trong một giới hạn phạm vi nhất định, nghĩa là không gian của tất cả các giải pháp sẽ bị hạn chế. Hai phương pháp cơ bản của đối sánh ảnh đã được phát triển và sử dụng trong quan trắc và thị giác máy là phương pháp dựa trên vùng và phương pháp dựa trên đặc trưng. Tổng quan của các phương pháp này được chỉ ra trong Bảng 2.1.

Bảng 2.1. Phương pháp đối sánh ảnh

Phương pháp đối sánh	Độ đo tương tự	Đối tượng đối sánh
Dựa trên vùng	Tương quan, đối sánh Giá trị mức xám hình vuông nhỏ nhất.	
Dựa theo đặc trưng	Hàm chi phí	Điểm đặc trưng, cạnh, vùng

2.2 Các phương pháp đối sánh ảnh

2.2.1 Phương pháp dựa trên vùng (Area based methods)

Các giá trị mức xám là những thực thể trong đối sánh dựa trên vùng. Đối sánh từng điểm ảnh dễ gặp phải vấn đề nhập nhằng, do vậy, các giá trị mức xám của một vài điểm ảnh lân cận sẽ được sử dụng. Một phần ảnh được cắt từ ảnh được gọi là mẫu, được sử dụng để tìm kiếm trong ảnh thứ hai. Mẫu gồm $m \times n$ điểm ảnh (thông thường là $m=n$). Vị trí của mẫu là điểm ảnh trung tâm, do vậy, m và n thường là lẻ. Mẫu sẽ được so sánh với phần ảnh có kích thước tương tự trong ảnh thứ hai. Việc so sánh được hạn chế với vùng được gọi là tìm kiếm dựa trên vùng hoặc là tìm kiếm cửa sổ. Giá trị độ đo tương tự được tính toán tại mỗi vị trí của mẫu trong vùng tìm kiếm. Dựa trên đặc tính của độ đo tương tự, mà các điểm tương ứng với tâm của mẫu sẽ là những

điểm có độ đo tương tự lớn nhất hoặc nhỏ nhất. Trong phép quan trắc thì tương quan chéo và đối sánh bình phương nhỏ nhất là những công nghệ được sử dụng nhiều cho đối sánh dựa trên vùng. Bên cạnh đó thông tin tương hỗ và khoảng cách ảnh cũng có thể được áp dụng.

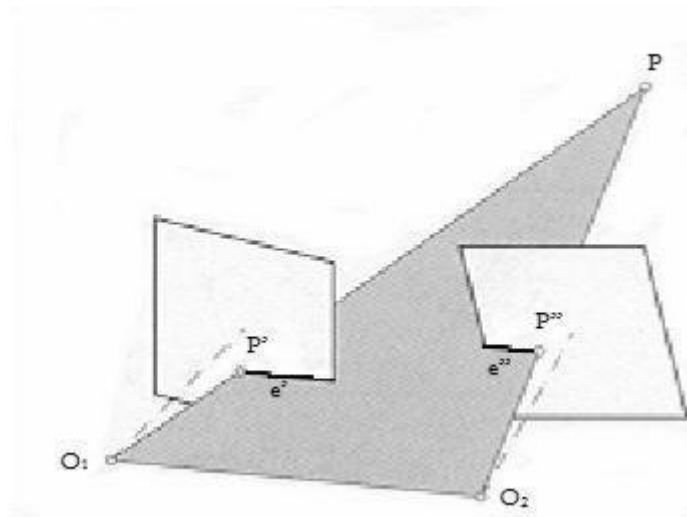
- Kích thước và vị trí của mẫu

Mẫu càng lớn thì yêu cầu tính duy nhất trong thực thể đối sánh càng lớn. Nói cách khác biến dạng hình học gây ra bởi sự thay đổi hướng của ảnh sẽ ảnh hưởng đến sự thay đổi các mẫu lớn. Yêu cầu về tính duy nhất không thể thực hiện được ở các khu vực có mẫu lặp lại hoặc cấu trúc và độ tương phản thấp. Nước và cát là những ví dụ điển hình cho trường hợp đối sánh ảnh thường gặp lỗi. Những phần bị che bởi các đối tượng cao nên được loại bỏ. Đối sánh dựa trên vùng là một quá trình tìm điểm liên hợp mặc dù 1 trong số chúng có thể bị ẩn trong ảnh khác. Các điểm tương quan trên phương tiện di chuyển hoặc trong vùng tối có thể dẫn đến việc xác định vị trí không đúng. Ở các khu vực sườn dốc, các phần ảnh tương ứng không giống nhau về mặt hình học. Để có được kết quả chấp nhận được trong đối sánh, kích thước của mẫu phải nhỏ hoặc hình dạng của nó phù hợp với biến dạng hình học (Ví dụ: Cửa sổ hình thang).

- Kích thước và vị trí của cửa sổ tìm kiếm

Để tránh các đối sánh sai, trong đối sánh dựa trên vùng, vị trí của cửa sổ tìm kiếm phải được xác định tương đối chính xác. Xấp xỉ các tham số được tính toán (Ví dụ: Tham số định hướng, DTM (Digital Terrain Model)) và các phương pháp phân cấp thường được sử dụng cho công việc xác định cửa sổ tìm kiếm. Cách tiếp cận phân cấp hoặc chiến lược Coarse-To-Fine có nghĩa là quá trình đối sánh bắt đầu ở các mức cao hơn của hình ảnh kim tự tháp (giảm kích thước điểm ảnh), nơi các chi tiết nhỏ bị bỏ qua. Các thông số được tính toán từ các phép đo ở mức cao hơn của hình ảnh kim tự tháp sau đó được sử dụng làm điểm bắt đầu cho đối sánh ở trong mức thấp.

Khi làm việc với một cặp đối sánh, các ràng buộc hình học bổ sung có thể được áp dụng như các đường epipolar. Hình 2-1 cho thấy khái niệm về ràng buộc đường epipolar. Các đường epipolar là các giao điểm của mặt phẳng epipolar và mặt phẳng ảnh. Mặt phẳng epipolar được định nghĩa bởi các tâm chiếu O_1 , O_2 và một điểm đối tượng P . Do đó các điểm liên hợp P' và P'' phải nằm trên các đường epipolar tương ứng e' và e'' . Để tạo ra một đối sánh dọc theo đường epipolar dễ dàng hơn, hình ảnh có thể được chuyển đổi được gọi là chuẩn hóa, tức là tất cả các đường epipolar trong hình ảnh đều song song.



Hình 2-1: Nguyên lý hình học epipolar. Một mặt phẳng epipolar được xác định bởi các trung tâm chiếu O_1 và O_2 và một điểm đối tượng P . Các đường epipolar e' và e'' là các giao điểm của mặt phẳng epipolar với các mặt phẳng ảnh. (được chuyển thể từ Schenk, 1999).

Độ chính xác phụ thuộc vào vị trí và độ biến dạng hình học của ảnh.

- Tiêu chí cho độ đo tương tự

Loại trừ sự không phù hợp là một trong những nhiệm vụ gắn liền với đối sánh ảnh. Một trong những giải pháp để vượt qua nhiều trong đối sánh ảnh chính là thiết lập ngưỡng cho độ đo tương tự. Các ngưỡng có thể ít khi được đặt cho tất cả các trường hợp. Sau khi tìm thấy vị trí phù hợp nhất, việc đánh giá tính chính xác và độ tin cậy của một kết quả đối sánh phải được thực hiện. Ngoài các biện pháp phân ngưỡng độ đo tương tự, những ràng buộc

hình học hoặc các công nghệ điều chỉnh sẽ được sử dụng trong các phép tính toán để loại trừ các kết quả đối sánh sai.

2.2.1.1 Tương quan (Correlation)

Hệ số tương quan chéo chuẩn hóa r là một trong những độ đo tương tự phổ biến được sử dụng trong phép quan trắc. Nó được tính theo công thức [2.1]:

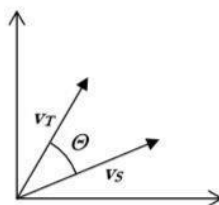
$$r = \frac{TS}{T \cdot S} = \frac{\sum_{i=1}^R \sum_{j=1}^C (g_T(i, j) - \bar{g}_T)(g_S(i, j) - \bar{g}_S)}{\left(\sum_{i=1}^R \sum_{j=1}^C (g_T(i, j) - \bar{g}_T)^2 \right)^{1/2} \left(\sum_{i=1}^R \sum_{j=1}^C (g_S(i, j) - \bar{g}_S)^2 \right)^{1/2}} \quad [2.1]$$

Trong đó:

- r : Hệ số tương quan – chéo được chuẩn hóa.
- τ, s : Độ lệch chuẩn của các giá trị mức xám trong mẫu và các phần ảnh tìm kiếm.
- TS : Hiệp phương sai của các giá trị mức xám trong các phần ảnh.
- g_T, g_S : Giá trị mức xám trong mẫu và phần ảnh tìm kiếm.
- \bar{g}_T, \bar{g}_S : Trung bình của các giá trị mức xám
- R, C : Số hàng và cột của các phần ảnh.

Nếu mẫu và phần ảnh được biểu diễn bởi v_T, v_S có kích thước $1 \times RC$ với (R, C là dòng, cột) thì hệ số tương quan được giải thích là: $r = \cos$

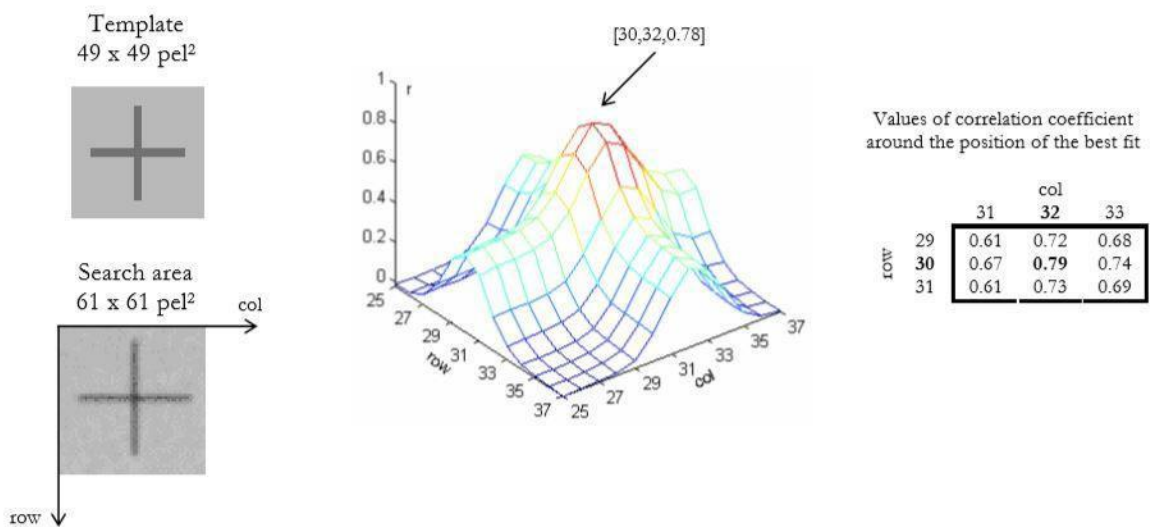
Trong đó θ là một góc giữa các véc tơ, như trong Hình 2-2.



Hình 2-2: Giải thích hình học của hệ số tương quan $r = \cos \theta = \frac{v_T \cdot v_S}{(|v_T| \cdot |v_S|)}$

Hệ số tương quan chuẩn hóa có các giá trị trong phạm vi $-1 \leq r \leq 1$. Giá trị 1 chỉ đạt được nếu các phần ảnh g_T và g_S được liên kết bởi một mối quan hệ tuyến tính $g_T = r_s g_S + r_t$, với $r_s \neq 0$, trong đó $r_s \neq 0$ tương ứng với hệ số tỉ lệ và r_t tương ứng với sự dịch chuyển giữa các giá trị mức xám trong g_T và g_S . Giá trị gần bằng 0 khi các phần đối sánh không tương tự và -1 khi các phần ảnh không giống nhau được đối sánh. Do đó, trong quá trình đối sánh ảnh, xử lý các giá trị dương gần 1 được đòi hỏi.

Một mẫu di chuyển trên từng điểm ảnh trên toàn bộ cửa sổ tìm kiếm và hệ số tương quan được tính theo từng vị trí. Vị trí nơi các hệ số tương quan đạt giá trị cao nhất được lựa chọn làm vị trí đối sánh tốt nhất. Ví dụ minh họa như trong Hình 2-3 đồ thị ở giữa cho thấy các giá trị của hệ số tương quan được tính trong các vị trí 13×13 của mẫu trong vùng tìm kiếm. Hệ số tương quan cực đại 0.79 tại vị trí hàng = 30 và cột = 32.



Hình 2-3: [Tài liệu “Image matching and its applications”] Nguyên lý đối sánh hình ảnh dựa trên việc tìm hệ số tương quan cực đại r .

Bản thân hệ số tương quan không phản ánh về độ chính xác của vị trí tìm thấy phù hợp nhất. Một số nghiên cứu chỉ ra liên hệ giữa phương sai của sự dịch chuyển được xác định từ tâm của cửa sổ tìm kiếm, dấu hiệu để tỷ lệ nhiễu, và cỡ của mẫu. Kết quả lý thuyết này đã không được áp dụng vào tính

toán thực tế cho đến nay. Tuy nhiên, nó chỉ ra rõ ràng rằng độ tin cậy của vị trí được xác định của đối sánh tốt nhất phụ thuộc vào thuộc tính phổ của các phần ảnh. Các thuộc tính này thường bị ảnh hưởng bởi sự chiếu sáng và góc nhìn khác nhau, những thay đổi về thời gian hoặc các phép chiếu của ảnh được đối sánh.

Độ lệch chuẩn của các giá trị mức xám và entropy là các độ đo độ tương phản và số lượng thông tin trong một phần ảnh và chúng có thể được sử dụng để đánh giá tính phù hợp của mẫu đã chọn cho việc đối sánh. Tương quan tự động (Autocorrelation) có thể được sử dụng cho những mục đích tương tự. Trong các thủ tục được tự động hóa các toán tử tìm biên và tìm điểm đặc trưng được áp dụng. Kết quả trong đối sánh chỉ được chấp nhận khi hệ số tương quan lớn nhất trội hơn ngưỡng đưa vào. Trong quá trình xác định đối tượng, đo ngưỡng là một trong những phương pháp thành công để loại bỏ hoặc giảm đáng kể số lượng các ngoại lệ. Nói chung, việc thiết lập một ngưỡng cho một hệ số tương quan không có nghĩa là tất cả các đối sánh lỗi được loại bỏ. Khi làm việc với yêu cầu thực tế, có thể gặp phải các trường hợp một số đối sánh tốt có hệ số tương quan thấp và một số đối sánh sai có hệ số tương quan cao. Với việc thiết lập ngưỡng, một số đối sánh thành công có thể sẽ bị loại bỏ trong khi một số đối sánh lỗi vẫn còn lại. Do đó các thuật toán để tính toán các tham số định hướng hoặc cho việc tạo DTM từ các điểm được đối sánh phải chứa các thủ tục để loại bỏ các ngoại lệ.

2.2.1.2 Đối sánh bình phương nhỏ nhất (Least squares matching)

Hệ số tương quan không phải là một độ đo lý tưởng của độ tương tự giữa hai phần ảnh do sự khác biệt giữa các thuộc tính hình học và phổ của chúng. Vào đầu thập niên tám mươi, rất nhiều phương pháp đối sánh được đề xuất để đối sánh những vùng ảnh. Trong đó phương pháp đối sánh bình phương nhỏ nhất (LSM) đã được nghiên cứu rộng rãi và đã được tích hợp trong nhiều gói phần mềm quan trắc. Ý tưởng của nó là cực tiểu hóa sự khác biệt các giá trị mức xám giữa mẫu và các vùng ảnh trong quá trình điều chỉnh

sự thay đổi thuộc tính hình học và thuộc tính phổ của cửa sổ đối sánh. Phương pháp này có thuộc tính thích nghi bởi vì nó cho khả năng tự động thay đổi số lượng tham số và trọng số quan sát dựa trên tầm quan trọng của nó và độ ổn định của giải pháp. Điều kiện quan trọng cho sự thành công của LSM là tìm vị trí xấp xỉ của vùng tìm kiếm tương đối chính xác. Mối quan hệ giữa các giá trị mức xám của hai vùng ảnh được thể hiện bằng công thức [2.2]:

$$g_1(r, c) + v(r, c) = g_2(r, c) \quad [2.2]$$

e là một véc tơ nhiễu gây ra bởi các hiệu ứng hình học và phổ khác nhau trong cả hai hình ảnh. Trong trường hợp lý tưởng của đối sánh hoàn hảo thì $e=0$. Mục đích là để tìm các tham số chuyển đổi hình học và phổ của một trong những cửa sổ sao cho véc tơ e là bé nhất. Có rất nhiều phương pháp để chọn một cửa sổ chính (Ví dụ: ổn định, không thay đổi) và cửa sổ phụ (tức là đã chuyển đổi) trong quá trình đối sánh.

LSM là một bài toán điều chỉnh phi tuyến. Do sự biến đổi hình học và phổ của một trong các phần ảnh nên công thức [2.2] phải được tuyến tính hóa. Một giải pháp cho vấn đề này được đề xuất bởi (Atkinson, 1996, Luhmann, 2000). Sự thích ứng giữa cửa sổ tìm kiếm g_s với cửa sổ mẫu g_t được phát biểu theo công thức [2.3] Việc tuyến tính hóa sau đó được thực hiện bởi trung bình chuỗi Taylor và được chỉ ra trong công thức [2.4]:

$$g_T(r, c) + v(r, c) = g_S(f_R(p_1, \dots, p_n, r, c), f_C(p_1, \dots, p_n, r, c))r_S + r_t = \bar{g}_S(r, c) \quad [2.3]$$

$$g_T(r, c) + v(r, c) = g_0(r, c) + g_{SR}(r, c)df_R + g_{SC}(r, c)df_C + g_0(r, c)dr_S + dr_t \quad [2.4]$$

$$g_0(r, c) = g_S(f_0(r, c), f_0(r, c))r_0 + r_t$$

$$df_R(r, c) = \frac{f_R(r, c)}{p_1} dp_1 + \dots + \frac{f_R(r, c)}{p_n} dp_n$$

$$df_C(r, c) = \frac{f_C(r, c)}{p_1} dp_1 + \dots + \frac{f_C(r, c)}{p_n} dp_n$$

Trong đó:

- r, c : hàng, cột
- $g_T(r, c)$: Các giá trị mức xám trong mẫu g_T
- $g_S(r, c)$: Các giá trị mức xám trong vùng tìm kiếm g_S
- $v(r, c)$: Các phần tử của véc tơ chênh lệch v
- $\frac{r}{g_S}(r, c)$: các giá trị mức xám đã được điều chỉnh trong vùng tìm kiếm
- $g_S^0(r, c)$: Các giá trị mức xám trong vùng tìm kiếm sau khi áp dụng xấp xỉ các tham số hình học và phổ
- f_R, f_C : Hàm biểu diễn cho phép biến đổi hình học giữa các phần ảnh.
- p_i : Tham số hình học
- n : Số tham số hình học
- r_s, r_t : Tỷ lệ phổ và dịch chuyển
- dp_i, dr_s, dr_t : Sự điều chỉnh các tham số hình học và phổ :
- g_{SR}, g_{SC} : Gradient trong giá trị mức xám theo hướng hàng và cột trong khu vực tìm kiếm

Một giải pháp hiệu quả hơn để thực hiện việc tuyến tính hóa đã được đề xuất bởi (Kraus, 2000) bằng việc thiết kế một ma trận của điều chỉnh bình phương nhỏ nhất sau đó giữ ổn định trong tất cả các lần lặp. Trong trường hợp này công thức [2.4] được sửa đổi theo:

$$g_S(r, c) + v(r, c) = g_T^0(r, c) + g_{TR}(r, c)df_R + g_{TC}(r, c)df_C + g_T^0(r, c)dr_s + dr_t \quad [2.5]$$

Số lượng các tham số hình học p_i phụ thuộc vào mô hình hình học được sử dụng trong điều chỉnh. Những phần ảnh bao phủ một vùng tương đối nhỏ trong không gian đối tượng. Giả sử vùng này là phẳng và phép chiếu

trung tâm cho thu nhận ảnh thì biến đổi chiếu là phù hợp nhất. Trong các thí nghiệm thực tế, phép biến đổi affine được xem như là một giải pháp xấp xỉ đủ do các phần ảnh để đối sánh là rất nhỏ so với toàn bộ ảnh. Bảng 2.2 đưa ra tổng quan về các mô hình và các tham số được sử dụng trong đối sánh bình phương nhỏ nhất.

Do sự chiếu sáng khác nhau tại thời điểm chụp ảnh, các góc nhìn khác nhau,...nên các thuộc tính phổ của các phần ảnh có kết quả không giống nhau. Do đó, hai tham số dịch chuyển độ sáng r_i và co giãn r_s được thêm vào trong công thức [2.4] và [2.5]. Một vấn đề với việc sử dụng cả hai tham số là ở tỷ lệ hội tụ thấp do mối tương quan cao của chúng. Theo (Schenk, 1999), để giảm số lượng các ẩn số và để tránh sự phụ thuộc tham số thì các giá trị tham số trong những phần ảnh có thể được điều chỉnh trước trong LMS. Cách tiếp cận này đặc biệt được khuyến nghị ở những vùng mà sự điều chỉnh tự động độ sáng và độ tương phản chỉ là tăng nhiều.

Bảng 2.2. Các mô hình hình học khác nhau được áp dụng trong đối sánh bình phương nhỏ nhất và các giá trị ban đầu của các tham số biến đổi không xác định.

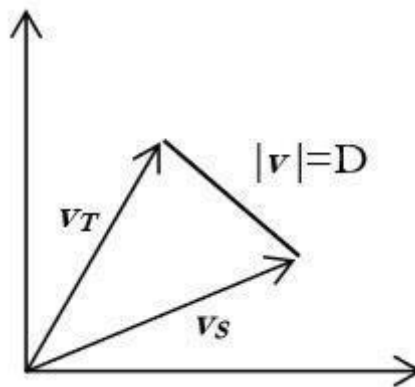
Least squares matching				First approximation of transformation parameters
Geometrical model		Transformation equation		
Conform	Translation (2 parameters)	$\begin{pmatrix} r_o \\ c_o \end{pmatrix} =$		$t_r^0 = t_c^0 = 0$
	Translation, Scale (3 parameters)		k	$t_r^0 = t_c^0 = 0$ $k^0 = 1$
	Translation, Scale, Rotation (4 parameters)		$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$	$\begin{pmatrix} r \\ c \end{pmatrix} + \begin{pmatrix} t_r \\ t_c \end{pmatrix}$
Affine (6 parameters)			$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$	$a_1^0 = b_2^0 = 1$ $a_2^0 = b_1^0 = t_r^0 = t_c^0 = 0$
Linear radiometric model		$g_{S_r}(r,c) = g_S(r,c) r_s + r_i$		$r_s^0 = 0$ $r_i^0 = 1$

2.2.1.3 Khoảng cách hình ảnh (Image distance)

Theo Yu và Jiang [1], thì khoảng cách thể hiện một phép đo đơn giản về sự giống nhau giữa hai phần ảnh g_T và g_S có cùng kích thước $R \times C$ và được biểu diễn theo công thức 2.6:

$$D(g_T, g_S) = \sqrt{\sum_{i=1}^R \sum_{j=1}^C (g_T(i, j) - g_S(i, j))^2} \quad [2.6]$$

Từ quan điểm hình học, khoảng cách D tương ứng với cỡ của véc tơ $v = v_T - v_S$, trong đó v_T và v_S là các véc tơ biểu diễn cho các mẫu và các phần ảnh tìm kiếm như trong Hình 2-4. Sự giống nhau giữa mẫu và phần ảnh tìm kiếm cao nếu $D(g_T, g_S) \rightarrow 0$.



Hình 2-4: Biểu diễn hình học của khoảng cách hình ảnh $D = |v| = |v_T - v_S|$

Công thức 2.6 có thể được chuẩn hóa để giảm ảnh hưởng của một sự dịch chuyển phổ giữa các vùng ảnh và cỡ của chúng theo các sửa đổi sau:

- Giảm giá trị mức xám theo giá trị trung bình \bar{g}_T và \bar{g}_S
- Chia cho tổng số điểm ảnh.

Khi đó khoảng cách ảnh được chuẩn hóa D_N được định nghĩa như sau:

$$D_N(g_T, g_S) = \sqrt{\frac{\sum_{i=1}^R \sum_{j=1}^C ((g_T(i, j) - \bar{g}_T) - (g_S(i, j) - \bar{g}_S))^2}{RC}} \quad [2.7]$$

Đối sánh ảnh bằng phương pháp hệ số tương quan lớn nhất và khoảng cách ảnh nhỏ nhất sẽ dẫn đến kết quả tương tự. Trong trường hợp các mục tiêu được xác định rõ ràng là không có hoặc chỉ có các biến dạng tối thiểu trong khu vực tìm kiếm như trong trường hợp của đường chéo trong Hình 2-3, các vị trí phù hợp nhất không khác nhau. Nếu tồn tại sự khác biệt về hình học hoặc phổ, kết quả thu được có thể khác nhau đáng kể.

2.2.1.4 Thông tin tương hỗ

Thông tin tương hỗ là một độ đo về sự phụ thuộc thống kê giữa hai biến ngẫu nhiên hoặc độ đo của lượng thông tin mà một biến chứa biến khác (Maes và cộng sự, 1997). Nó là cực đại nếu các phân ảnh được hiệu chỉnh theo hình học. Thông tin tương hỗ $I(g_T, g_S)$ của hai phân ảnh g_T và g_S chứa các giá trị mức xám trong phạm vi $t = 1, \dots, n_T$ và $s = 1, \dots, n_S$.

$$I(g_T, g_S) = \sum_{t=1}^{n_T} \sum_{s=1}^{n_S} P_{TS}(t, s) \log_2 \frac{P_{TS}(t, s)}{P_T(t) P_S(s)} \quad [2.8]$$

Trong đó

- $I(g_T, g_S)$: Thông tin tương hỗ
- P_T : Xác suất các giá trị mức xám $t = 1, \dots, n$ trong phân ảnh g_T
- P_S : Xác suất các giá trị mức xám $s = 1, \dots, n$ trong phân ảnh g_S
- P_{TS} : Xác suất hợp

Các giá trị p_T và p_S thu được bằng cách chuẩn hóa các biểu đồ phân ảnh. Các giá trị của xác suất hợp p_{TS} được tính toán từ các biểu đồ ảnh kết hợp. Các biểu đồ này được biểu diễn bởi ma trận $n_T * n_S$ với các phần tử bằng số điểm ảnh với $g_T = (i, j) = t$ và $g_S = (i, j) = s$ như được minh họa trong Bảng 2.3

Bảng 2.3. Xác suất của các giá trị mức xám p_T và p_S trong các phần ảnh g_T và g_S , p_{TS} xác suất chung và thông tin chung $I(g_T, g_S)$.

		PTS								
										p _T
t\s		1	2	3	4	5	6	7	8	
1	0.04	0.04								0.08
2	0.04	0.08								0.12
3			0.12							0.12
4			0.04	0.08	0.04					0.16
5				0.04	0.08					0.12
6					0.04	0.12				0.16
7							0.12	0.04		0.16
8							0.04	0.04		0.08
p _S	0.08	0.12	0.16	0.12	0.16	0.12	0.16	0.16	0.08	

g _T					g _S				
1	2	3	4	4	1	2	3	3	4
1	3	4	5	5	2	3	4	4	5
2	4	6	6	7	2	5	5	6	7
2	6	6	8	8	1	6	6	7	8
3	5	7	7	7	3	5	7	7	8

$I(g_T, g_S) = 1.96$

Mối quan hệ giữa thông tin tương hỗ $I(g_T, g_S)$ và entropy của các phần ảnh $H(g_T)$ và $H(g_S)$ được tính theo công thức [2.9] (Maes và cộng sự, 1997):

$$I(g_T, g_S) = H(g_T) - H(g_T | g_S) = H(g_S | g_T) = H(g_T) + H(g_S) - H(g_T, g_S) \quad [2.9]$$

Trong đó:

- $H(g_S | g_T), H(g_T | g_S)$: entropy có điều kiện của g_T cho g_S và g_S cho g_T tương ứng
- $H(g_T, g_S)$: entropy kết hợp

Nếu các phần ảnh g_T và g_S độc lập về mặt thống kê, thì $p_{TS}(t, s) = p_T(t) p_S(s)$ và $I(g_T, g_S) = 0$. Nếu g_T và g_S phụ thuộc tối đa, thì

$$I(g_T, g_S) = H(g_T) = H(g_S) = H(g_T, g_S).$$

2.2.2 Phương pháp dựa theo đặc trưng (Feature based methods)

Ngược lại với đối sánh dựa trên vùng sử dụng các toán tử trực tiếp trên các giá trị mức xám, các phương pháp dựa trên đặc trưng sẽ dựa trên việc đối sánh các đặc trưng được trích chọn như điểm, cạnh hoặc vùng. Các thủ tục đối sánh dựa trên đặc trưng bao gồm ba bước (được điều chỉnh từ Forstner, 1986):

- Chọn các đặc trưng riêng biệt (điểm, cạnh, góc) trong các ảnh riêng biệt.
- Xây dựng danh sách sơ bộ các cặp ứng viên của các đặc trưng tương ứng dựa trên độ đo tương tự được lựa chọn.
- Lấy danh sách cuối cùng các cặp đặc trưng phù hợp với mô hình đối tượng.

2.2.2.1 Điểm đặc trưng (Interest points)

Những điểm có độ chênh lệch cao về giá trị mức xám hoặc hàm tương quan tự động cao và có độ dốc gradient lớn thì gọi là những điểm đặc trưng. Các điểm đặc trưng nên có sự khác biệt, bất biến đối với các biến đổi hình học và biến đổi phổ, ổn định (một điểm nên xuất hiện trong tất cả các ảnh được đối sánh) (Forstner, 1986). Thủ tục để tìm kiếm những điểm đặc trưng trong ảnh đối sánh được thực hiện qua 2 bước:

- Tính toán các tham số đặc trưng ở mỗi cửa sổ của ảnh được chọn.
- So sánh giá trị của các tham số với một ngưỡng cho trước.

Các tham số đặc trưng là khác nhau cho mỗi toán tử nhưng về cơ bản dựa trên các giá trị mức xám (cấu trúc và kết cấu) bên trong cửa sổ được đánh giá. Chỉ có những cửa sổ mà có giá trị tham số lớn hơn hoặc nhỏ hơn ngưỡng mới được chấp nhận là điểm đặc trưng. Một danh sách các điểm đặc trưng của mỗi ảnh được đối sánh với tọa độ của nó (điểm trung tâm của mỗi cửa sổ trượt) và mô tả của nó là kết quả của quá trình này. Trong (Luhmann và Altrogge, 1986) đã đề xuất 3 toán tử điểm đặc trưng tên là Moravec, Forstner và Dreschler. Tuy nhiên Moravec và Forstner hoạt động tốt hơn với việc tìm kiếm các điểm đặc trưng theo các điều kiện hình học khác nhau.

Toán tử Moravec dựa trên giả sử điểm đặc trưng có sự biến thiên cao trên tất cả các hướng. Ví dụ: dòng, cột và cả đường chéo bên trong cửa sổ

$(2k+1) \times (2k+1)$ điểm ảnh (cửa sổ có kích thước lẻ). Điểm đặc trưng được chấp nhận nếu tham số đặc trưng m trong công thức sau lớn hơn một ngưỡng

$$M = \min_{\substack{r=n-k, c=m-k \\ n+k, m+k}} \left(\begin{matrix} g(r, c) - g(r+1, c) \\ g(r, c) - g(r+1, c+1) \end{matrix} \right)_2^2 \quad [2.10]$$

Trong đó:

- k : Định nghĩa cỡ của cửa sổ toán tử
- m, n : Tọa độ điểm ảnh trung tâm của cửa sổ toán tử trong một hình ảnh g
- $g(r, c)$: Giá trị màu xám tại vị trí (r, c)

Toán tử Forstner liên quan với phương pháp đối sánh bình phương nhỏ nhất. Ma trận hiệp phương sai của một dịch chuyển được ước lượng sigma có thể được viết như sau:

$$= \begin{matrix} 2 \\ 0 \end{matrix} \begin{matrix} g_R^2 \\ g_R g_C \\ g_C^2 \end{matrix} = \begin{matrix} 2q_{RR} \\ 0 \end{matrix} \begin{matrix} q_{RC} \\ q_{CC} \end{matrix} = \begin{matrix} 2 \\ 0 \end{matrix} Q = \begin{matrix} 2 \\ 0 \end{matrix} N^{-1} \quad [2.11]$$

Trong đó:

- σ : Độ lệch chuẩn của một trọng số đơn vị
- g_R, g_C : Gradient theo hướng hàng và cột
- Q : Ma trận đồng hệ số

2.2.2.2 Cạnh và vùng

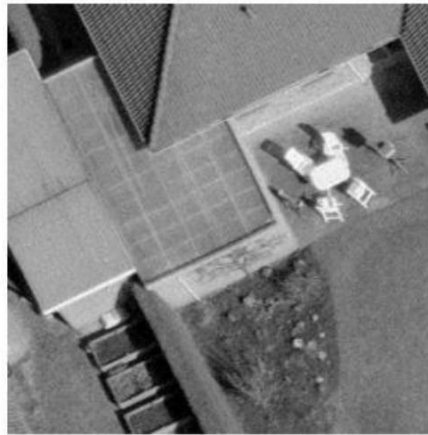
Các cạnh có thể được mô tả như các điểm gián đoạn trong hàm mức xám, ví dụ: các giá trị mức xám thay đổi nhanh chóng trong một khu vực nhỏ. Các cạnh thường tương ứng với đường bao của các đối tượng được hiển thị

trong hình ảnh. Quá trình trích chọn cạnh khá phức tạp và bao gồm các bước sau (Schenk, 1999):

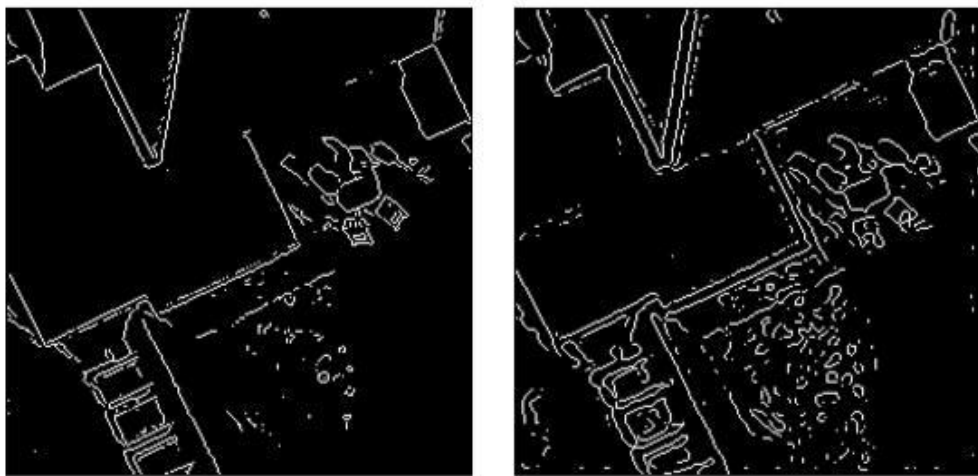
- Xác định các điểm ảnh nằm trên cạnh, giá trị mức xám không liên tục được phát hiện bởi một phép đo được gọi là toán tử cạnh. Một ngưỡng cho các giá trị mức xám khác nhau được thiết lập để quyết định các điểm là điểm cạnh.
- Liên kết các điểm cạnh thành các cạnh.
- Nhóm các cạnh: tức là xác định phân đoạn đường thẳng, đường đa giác, đường gấp khúc, đường song song, v.v.

Toán tử cạnh sẽ phát hiện ra sự thay đổi của giá trị mức xám trong ảnh, một trong số chúng dựa trên đạo hàm bậc nhất của các hàm mức xám để tìm ra cực trị và định vị điểm cạnh. Một số toán tử cạnh có thể dùng như toán tử Robert (Robert Cross), toán tử Sobel (Sobel Operator). Cả 2 toán tử đều dựa trên hướng, phát hiện theo chiều ngang, phát hiện theo chiều dọc. Toán tử Sobel sẽ ít bị ảnh hưởng bởi nhiễu của ảnh vì bao gồm cả những điểm ảnh lân cận.

Toán tử Laplacian thuộc về nhóm các toán tử đạo hàm bậc 2, nó độc lập về hướng để giảm ảnh hưởng của nhiễu trong ảnh. Nó kết hợp với toán tử Gaussian làm mịn ảnh. Sau khi áp dụng kết quả Laplacian của toán tử Gaussian trên ảnh gốc thì các điểm cạnh tương ứng với giá trị zero. Toán tử LoG giống toán tử đạo hàm bậc nhất được mô tả chi tiết ở Hình 2-5 và Hình 2-6:



Hình 2-5: Ảnh gốc



Hình 2-6: Phát hiện cạnh

2.3 Đối sánh ảnh dựa trên đặc trưng SIFT

2.3.1 Trích chọn đặc trưng SIFT

Phương pháp trích rút các đặc trưng bất biến SIFT được tiếp cận theo phương pháp lọc kim tự tháp, theo đó phương pháp được thực hiện lần lượt theo các bước sau:

- *Phát hiện các điểm cực trị (Scale-Space extrema detection)*: Bước đầu tiên này sẽ áp dụng đạo hàm của hàm Gaussian (DoG - Difference of Gaussian) để tìm ra các điểm có khả năng làm điểm đặc trưng tiềm năng, đó là những điểm rất ít phụ thuộc (bất biến) vào sự thu phóng ảnh và xoay ảnh.

- *Định vị các điểm đặc trưng (keypoint localization)*: Từ những điểm tiềm năng ở trên sẽ lọc và lấy ra tập các điểm đặc trưng tốt nhất (các điểm chính).
- *Gán hướng cho các điểm đặc trưng (Orientation assignment)*: Mỗi điểm đặc trưng sẽ được gán cho một hoặc nhiều hướng dựa trên hướng gradient của ảnh. Mọi phép toán xử lý ở các bước sau này sẽ được thực hiện trên những dữ liệu ảnh mà đã được biến đổi tương đối so với hướng đã gán, kích cỡ và vị trí của mỗi điểm đặc trưng. Nhờ đó, tạo ra một sự bất biến trong các phép xử lý này.
- *Mô tả các điểm đặc trưng (Keypoint descriptor)*: Các hướng gradient cục bộ được đo trong ảnh có kích cỡ cụ thể nào đó trong vùng lân cận với mỗi điểm đặc trưng. Sau đó, chúng sẽ được biểu diễn thành một dạng mà cho phép mô tả các tầng quan trọng của quá trình bóp méo hình dạng cục bộ và sự thay đổi về độ sáng.

Tập các điểm đặc biệt thu được thường phụ thuộc rất ít vào các phép biến đổi cơ bản như xoay, phóng to, thu nhỏ, tăng giảm cường độ sáng, vì vậy có thể xem đây là các đặc trưng mang tính cục bộ của ảnh. Để đối sánh và nhận dạng hai ảnh thì ta tìm tập điểm chính giống nhau trong hai ảnh, dựa vào hướng và tỉ lệ để có thể biết đối tượng trong ảnh gốc đã xoay, thu phóng bao nhiêu so với ảnh đem đối sánh. Cách tiếp cận của thuật toán này dựa vào điểm bất biến cục bộ của ảnh, chúng được trích xuất ra, được định hướng và mô tả sao cho hai điểm chính ở hai vùng khác nhau thì khác nhau. Tuy nhiên một yếu tố ảnh hưởng không nhỏ đến tốc độ thuật toán là số lượng các điểm chính được lấy ra là không nhỏ. Trung bình một ảnh kích thước 500×500 điểm ảnh thì sẽ trích xuất được khoảng 1000 điểm (số lượng điểm này phụ thuộc vào tùy từng ảnh và tham số lọc khác nhau). Số lượng các điểm đặc trưng có một tầm quan trọng trong vấn đề nhận dạng đối tượng, để nhận dạng một đối

tượng nhỏ trong một ảnh chứa tập hợp các đối tượng hỗn độn thì cần ít nhất 3 điểm đặc trưng giống nhau để phát hiện và bóc tách đối tượng.

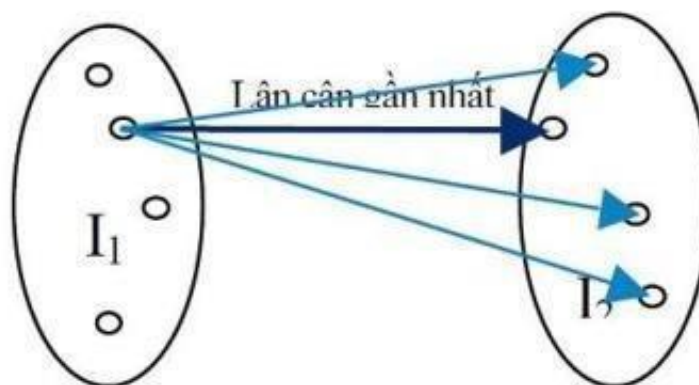
2.3.2 Đối sánh SIFT

2.3.2.1 Độ đo khoảng cách và độ đo tương tự

Độ đo tương tự là một trong những phương pháp tốt để máy tính phân biệt được các hình ảnh qua nội dung của chúng. Độ đo mang ý nghĩa quyết định kết quả tìm kiếm sẽ như thế nào, mức độ chính xác ra sao. Nhiều phép đo khoảng cách đã được sử dụng như: khoảng cách Euclide, khoảng cách Cosin, khoảng cách giao nhau của biểu đồ histogram, khoảng cách Minkowski...

2.3.2.2 Đối sánh đặc trưng cục bộ bất biến

Trước hết để đối sánh các ảnh với nhau thì cần trích xuất tập điểm chính tương ứng từ mỗi ảnh bằng các bước đã chỉ ra ở trên. Sau đó việc đối sánh sẽ thực hiện trên các tập điểm chính này. Bước chính trong kĩ thuật đối sánh sẽ thực hiện tìm tập con điểm chính so khớp nhau ở hai ảnh, để thực hiện việc này sẽ tìm các cặp điểm chính trùng nhau lần lượt ở hai ảnh. Tập con các điểm chính so khớp chính là vùng ảnh tương đồng. Việc đối sánh hai tập hợp điểm đặc trưng quy về bài toán tìm láng giềng gần nhất của mỗi điểm đặc trưng như trong Hình 2-7.



Hình 2-7: Đối sánh hai ảnh quay về đối sánh hai tập điểm đặc trưng trong không gian đặc trưng

Có 2 vấn đề cần được quan tâm:

- Tổ chức tập hợp điểm cho phép tìm kiếm láng giềng một cách hiệu quả
- Việc đối sánh phải đạt độ chính xác nhất định

Một phương pháp được đề xuất bởi D. Mount cho phép tìm kiếm nhanh các điểm lân cận được sử dụng, KNN là viết tắt của k-Nearest Neighbour. Nó cho phép tổ chức dữ liệu dưới dạng kd-tree, việc tìm kiếm láng giềng gần nhất mang tính xấp xỉ trên kd-tree. Cụ thể là hai điểm trong không gian đặc trưng được coi là giống nhau nếu khoảng cách Euclidean giữa hai điểm là nhỏ nhất và tỉ số giữa khoảng cách gần nhất với khoảng cách gần nhì phải nhỏ hơn 1 ngưỡng cho trước.

Giả sử cặp điểm chính có bộ mô tả lần lượt là:

$$A = (a_1, a_2, a_3, \dots, a_{128}) \text{ và } B = (b_1, b_2, b_3, \dots, b_{128})$$

Thì khoảng cách Euclidean giữa A và B được tính bằng công thức:

$$D(A, B) = \sqrt{\sum_{i=1}^{128} (a_i - b_i)^2} \quad [2.12]$$

2.3.2.3 Một số độ đo khoảng cách cho ảnh sử dụng đặc trưng SIFT

- Độ đo Cosin:

$$d(x, y) = \frac{x \cdot y}{\|x\| \cdot \|y\|} \quad [2.13]$$

- Khoảng cách góc:

$$d(x, y) = \cos^{-1}(x \cdot y) \quad [2.14]$$

- Độ đo Euclid:

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad [2.15]$$

- Độ đo Jensen-Shannon divergence:

$$d_{JSD}(H, H') = \frac{1}{2} \sum_{m=1}^n H_m \log \frac{2H_m}{H_m + H'_m} + \frac{1}{2} \sum_{m=1}^n H'_m \log \frac{2H'_m}{H'_m + H_m} \quad [2.16]$$

CHƯƠNG 3: CHƯƠNG TRÌNH THỰC NGHIỆM

3.1 Môi trường thực nghiệm

3.1.1 Cấu hình phần cứng

Bảng 3.1. Cấu hình phần cứng sử dụng trong thực nghiệm

Thành phần	Thông số
CPU	Core i7-6500U 2.50GHz
RAM	4GB
Hệ điều hành	Windows 10 Pro 64-bit
Bộ nhớ	1TB

3.1.2 Môi trường cài đặt

Bảng 3.3. Công cụ phần mềm sử dụng trong thực nghiệm

STT	Tên phần mềm	Nguồn
1	Visual Studio 2013	https://www.microsoft.com/en-us/download/details.aspx?id=40769

3.1.3 Thư viện OpenCV (Open Source Computer Vision Library)

Đây là một thư viện mã nguồn mở về thị giác máy tính và học máy. Thư viện được xây dựng để cung cấp nền tảng cho các ứng dụng thị giác máy tính nhằm đẩy mạnh sự phát triển về hàm lượng tri thức máy tính trong các sản phẩm thương mại. Nhờ giấy phép bản quyền BSD và được nhiều công ty lớn hàng đầu thế giới như Google, Yahoo, Microsoft, Intel, IBM,... cùng đóng góp xây dựng thư viện, OpenCV là một trong những công cụ mạnh và được sử dụng rộng rãi trong trường học cũng như các công ty khởi nghiệp.

OpenCV bao gồm nhiều giao diện dành cho C++, C, Python, Java, MATLAB và hỗ trợ cho các hệ điều hành khác nhau như Windows, Linux,

Android, MacOS. Trong phiên bản OpenCV 3.1, giao diện sử dụng cho CUDA và OpenCL cũng đã được phát triển hoàn thiện. OpenCV được viết nguyên bản bằng ngôn ngữ C++.

OpenCV có rất nhiều chức năng. Sau đây là những tóm tắt cơ bản về hệ thống các về chức năng của các hàm trong OpenCV

- Image and Video I/O.

Những giao diện này sẽ giúp ta đọc được dữ liệu ảnh từ file hoặc trực tiếp từ video. Ta cũng có thể tạo các file ảnh và video với giao diện này.

- Các thuật toán xử lý ảnh và thị giác máy (General computer-vision and image-processing algorithms (mid – and low level APIs)).

Sử dụng những giao diện này, ta có thể thực hành với rất nhiều chuẩn thị giác máy mà không cần phải có mã nguồn của chúng.

- Đồ họa.

Những giao diện này giúp ta viết chữ và vẽ trên hình ảnh. Thêm vào đó những chức năng này được sử dụng nhiều trong ghi nhãn và đánh dấu. Ví dụ nếu ta viết một chương trình cần nhận dạng nhiều đối tượng thì nó sẽ rất có ích cho tạo nhãn ảnh (label image) với kích thước và vị trí.

Với Window, khi cài đặt OpenCV, nó sẽ copy file OpenCV vào thư mục mà ta chọn. Cách thức lựa chọn trong đường dẫn hệ thống chứa mã nhị phân OpenCV, đăng kí một vài bộ lọc DirectX. Mặc định nó cài đặt đến C:/Program Files/OpenCV/<version>

Trong thư mục OpenCV có chứa một vài thư mục khác. Thư mục docs chứa file văn bản html cho toàn bộ các hàm OpenCV và kiểu dữ liệu. Từ file văn bản này ta có thể làm các ví dụ, ta cũng có thể muốn xem thư mục “samples”. Những file header sẽ cần thiết khi ta dịch chương trình sử dụng OpenCV. Ta có thể xác định file header bằng cách tìm kiếm trong thư mục cài đặt và những thư mục khác những file có dạng *.h, *.hpp. Header dùng cho

tất cả các module trừ HighGUI được tách riêng trong “include”. Ta có thể bỏ header trong thư mục “src”.

3.2 Trích chọn đặc trưng SIFT

Quá trình trích chọn đặc trưng SIFT được thực hiện theo các bước sau:

Bước 1: Đọc ảnh

Hàm đọc ảnh `imread(const string& filename, int flags)` sẽ đọc một ảnh từ một tệp được chỉ định và trả về một ma trận các giá trị mức xám. Nếu ảnh không được đọc thì hàm sẽ trả lại một ma trận rỗng. Các tham số được sử dụng trong hàm:

- Filename: Tên ảnh cần nạp
- Flags: Cờ chỉ định kiểu màu của ảnh được load, nếu không thì nó sẽ đọc theo đúng định dạng của ảnh.
 - o `CV_LOAD_IMAGE_ANYDEPTH`: Trả về ảnh 16 hoặc 32 bit khi ảnh đầu vào có chiều sâu tương ứng. Ngược lại ảnh được chuyển về 8 bit.
 - o `CV_LOAD_IMAGE_COLOR`: Chuyển đổi ảnh thành màu
 - o `CV_LOAD_IMAGE_GRAYSCALE`: Chuyển đổi thành ảnh đa mức xám.

Hàm hiển thị ảnh `imshow(const string& winname, InputArray mat)` sẽ hiển thị một ảnh trong cửa sổ được chỉ định. Nếu cửa sổ được tạo bằng cờ thì ảnh sẽ được hiển thị với kích thước ban đầu, tuy nhiên nó bị giới hạn bởi độ phân giải màn hình. Ngược lại thì ảnh sẽ thu nhỏ để vừa với cửa sổ. Các tham số của hàm có ý nghĩa như sau:

- Winname: Tên của cửa sổ
- mat: Ma trận ảnh được hiển thị

Ví dụ đọc và hiển thị ảnh

```
Mat img1 = cv::imread("D:/Nemo/a1.png");  
  
imshow("Ket qua", img1);  
waitKey(0);  
  
return 0;
```

Kết quả đọc và hiển thị ảnh



Hình 3-1: Ảnh được hiển thị

Bước 2: Phát hiện điểm đặc trưng

Trong OpenCV, phương thức sau cho phép phát hiện các điểm đặc trưng: `SiftFeatureDetector::detect(...)`. Ví dụ

```
cv::SiftFeatureDetector detector(0.0001, 1);  
  
vector<KeyPoint> keypoints1, keypoints2;  
detector.detect(img1, keypoints1);  
detector.detect(img2, keypoints2);
```

Trong đó:

- `img1, img2`: Ma trận điểm ảnh của ảnh 1 và ảnh 2 tương ứng.
- `keypoint 1, keypoint2`: là một véc tơ lưu trữ tập điểm đặc trưng cho ma trận ảnh `img1` và `img2` tương ứng.

Kết quả tính toán được hiển thị như trong hình 3-2:



Hình 3-2: Ảnh đã được phát hiện đặc trưng

Bước 3: Tính toán mô tả đặc trưng.

Chức năng này được thực hiện bằng cách gọi thủ tục

DescriptorExtractor::compute(...). Ví dụ

```
Mat descriptor1, descriptor2;  
extractor.compute(img1, keypoints1, descriptor1);  
extractor.compute(img2, keypoints2, descriptor2);
```

Trong đó:

- img1, img2: Ma trận điểm ảnh 1 và ảnh 2 tương ứng.
- keypoints1, keypoints2: là một véc tơ lưu trữ tập điểm đặc trưng cho ma trận ảnh img1 và img2 tương ứng.
- descriptors1, descriptors2: Các mô tả của ảnh img1 và ảnh img2 được tính toán.

3.3 Ước lượng tập điểm tương đồng

Ước lượng tập điểm tương đồng giữa hai ảnh được thực hiện dựa trên việc đối sánh các tập điểm đặc trưng. Công việc này được thực hiện bằng việc sử dụng thuật toán đối sánh Brute Force(BF) trong OpenCV. Đây là biến đổi

sánh đơn giản bằng cách lấy mô tả của một đặc trưng trong tập thứ nhất đối sánh với tất cả các đặc trưng khác trong tập thứ hai sử dụng một số độ đo khoảng cách. Cặp đối sánh nào có khoảng cách gần nhất sẽ được chọn.

Với đối sánh BF, trước hết cần tạo một đối tượng BFMatcher sử dụng phương thức `cv2.BFMatcher()`. Nó có hai tham số tùy chọn. Tham số thứ nhất `normType` là kiểu chuẩn, được sử dụng để chỉ định độ đo khoảng cách, Mặc định nó là chuẩn L2(`cv2.NORM_L2`). Tham số thứ hai `crossCheck` là một giá trị boolean, mặc định được đặt là `False`. Nếu được đặt là `True` thì kết quả chỉ trả về các đối sánh với giá trị (i,j) sao cho mô tả thứ i ở trong tập a đối sánh tốt nhất với mô tả thứ j trong tập b và ngược lại.

Khi đối tượng BFMatcher được tạo ra, thì hai phương thức quan trọng sẽ được sử dụng để đối sánh đó là `BFMatcher.match()` và `BFMather.knnMatch()`. Phương thức thứ nhất sẽ chỉ trả lại một đối sánh tốt nhất, trong khi đó phương thức thứ hai sẽ trả về k đối sánh tốt nhất. Với k được chỉ định bởi người dùng.

Để vẽ các điểm chính cần sử dụng phương thức `cv2.drawMatches()`. Phương thức này sẽ ghép hai ảnh theo chiều ngang và vẽ các đường là các đối sánh tốt nhất từ ảnh thứ nhất sang ảnh thứ hai. Khi cần vẽ tất cả k đối sánh tốt nhất thì phải sử dụng phương thức `cv2.drawMathesKnn()`. Nếu `k=2` nó sẽ vẽ hai đường từ mỗi điểm.

Ví dụ: Giả sử ta có hai tập mô tả đặc trưng sử dụng mô tả SIFT là `descriptors1`, `descriptors2` được tính toán ở bước 3 (mục 3.2).

Bước 1: Tìm đối sánh

- Trước hết cần phải tạo một đối tượng `mathcer` với độ đo khoảng cách chuẩn `norm_L2`

```
BFMatcher matcher(NORM_L2);  
std::vector<vector<DMatch> > knn_matches;
```


- Tiếp theo sử dụng phương thức `matcher.match()` để nhận các đối sánh tốt nhất giữa hai ảnh

```
matcher.match(descriptor1, descriptor2, matches);
```

Trong đó `matches` là các đối sánh thu được. Cỡ của các đối sánh có thể nhỏ hơn số lượng các mô tả. Nếu sử dụng phương thức `knn` thì thực hiện lệnh sau để tìm các đối sánh:

```
matcher.knnMatch(descriptor1, descriptor2, knn_matches, k);
```

Trong đó mỗi `knn_matches` là `k` hoặc ít hơn cho cùng một bộ mô tả, `k` là số đối sánh tốt nhất được tìm thấy cho mỗi mô tả hoặc ít hơn nếu mô tả có tổng số ít hơn `k` có thể đối sánh.

Bước 2: Tìm đối sánh tốt nhất

- Tính toán nhanh các khoảng cách tối đa và tối thiểu: Ban đầu, khoảng cách tối thiểu được gán bằng 0, khoảng cách tối đa được gán bằng 100.

Sau đó sẽ tìm khoảng cách tối thiểu và tối đa theo:

```
double dist = matches[i].distance;  
if (dist < min_dist) min_dist = dist;  
if (dist > max_dist) max_dist = dist;
```

- Chọn các đối sánh tốt nhất: là những đối sánh có khoảng cách nằm trong khoảng từ 0.03 tới $2 * \text{min_dist}$.

```
if (matches[i].distance <= max(2 * min_dist, 0.03))  
{  
    good_matches.push_back(matches[i]);  
}
```

Bước 3: Vẽ các đối sánh và hiển thị

Hàm `drawMatches(...)` sẽ được sử dụng để vẽ các đối sánh, sau đó được hiển thị bằng hàm `imshow()`. Ví dụ:

```
drawMatches(img1, keypoints1, img2, keypoints2, good_matches, img_matches,  
            Scalar::all(-1), Scalar::all(-1), vector<char>(), DrawMatchesFlags::NOT_DRAW_SINGLE_POINTS);
```

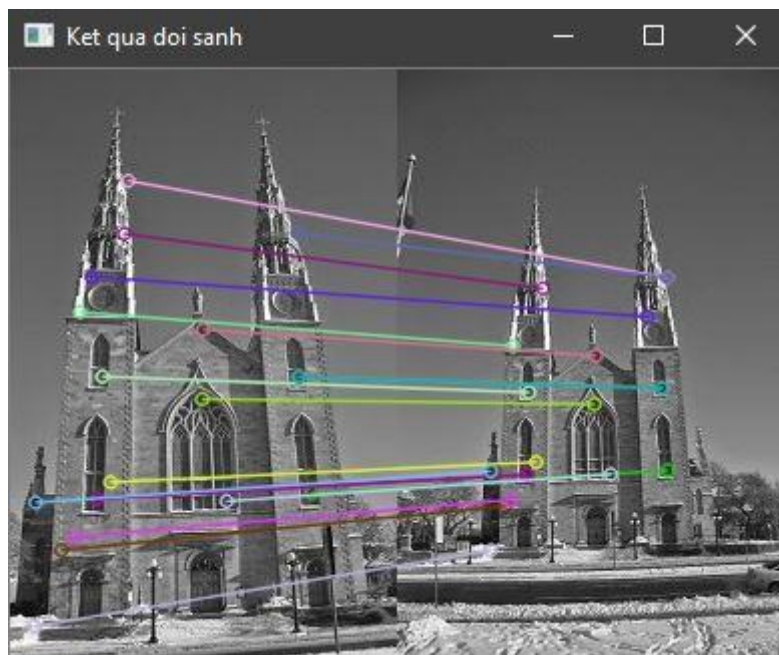
3.4 Một số kết quả thực nghiệm

Kết quả tìm các điểm tương đồng dựa trên đối sánh cho một số cặp ảnh được chỉ ra trong bảng 3.1. Các kết quả này được biểu diễn trực quan trên các hình: Hình 3-3 đến Hình 3-5.

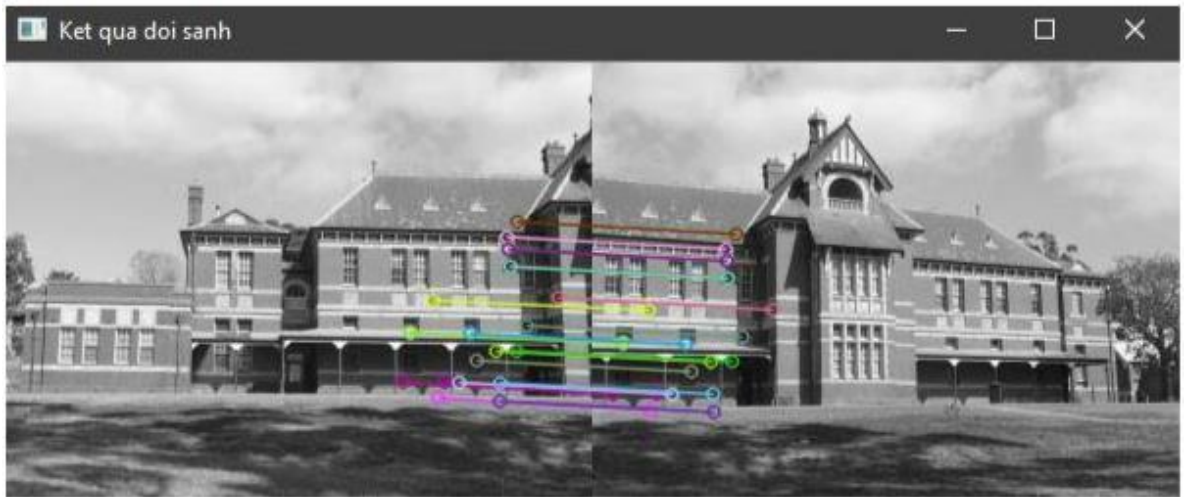
Bảng 3.1. Bảng thống kê kết quả đối sánh của một số cặp ảnh

Cặp ảnh đối sánh	Khoảng cách lớn nhất	Khoảng cách nhỏ nhất	Số điểm tương đồng
a1.png - a2.png	440.3	100	20
b1.png - b2.png	413.1	29.2	23
c1.png - c2.png	408.9	55.9	17

Ảnh kết quả thực nghiệm



Hình 3-3: Các cặp đối sánh giữa ảnh a1.png và a2.png



Hình 3-4: Các cặp đối sánh giữa ảnh b1.png và b2.png



Hình 3-5: Các cặp đối sánh giữa ảnh c1.png và c2.png

KẾT LUẬN

Sau một thời gian tìm hiểu và thực hiện đề tài, đồ án đã đạt được kết quả như sau:

- Nắm được nguyên lý của kỹ thuật đối sánh ảnh.
- Mô tả và giải thích các vấn đề cơ bản trong đối sánh ảnh
- Nắm được phương pháp tìm, trích chọn các điểm chính và biểu diễn đặc trưng SIFT.
- Áp dụng thành công kỹ thuật BFMatcher vào chương trình để ước lượng tập điểm tương đồng giữa hai ảnh.

Tuy nhiên, do thời gian và khả năng có hạn, nên em chưa đi sâu tìm hiểu được thêm về ứng dụng, vì vậy xử lý code trên OpenCV còn chưa tốt, chương trình vẫn còn nhiều thiết sót.

Trong thời gian tới, em sẽ tìm hiểu thêm các cải tiến của SIFT và xây dựng ứng dụng hoàn chỉnh bao gồm cả giao diện người dùng và CSDL đặc trưng một cách đầy đủ, hệ thống.

TÀI LIỆU THAM KHẢO

- [1] Markéta Potůčková, “Image matching and its applications in photogrammetry”, February 2006.
- [2] M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, “Image Features Detection, Description and Matching”.
- [3] Tuytelaars, T., Mikolajczyk, K., “Local invariant feature detectors: a survey”, (2007).
- [4] Moravec, H.P., “Towards automatic visual obstacle avoidance. In: 5th International Joint Conference on Artificial Intelligence, pp. 584–594”, (1977).
- [5] Harris, C., Stephens, M., “A combined corner and edge detector”, Manchester - UK, (1988).
- [6] Smith, S.M., Brady, J.M., “A new approach to low level image processing.”, J. Comput. Vis, (1997).
- [7] Yussof, W., Hitam, M., “Invariant Gabor-based interest points detector under geometric transformation. Digital Signal Process.”, (2014).
- [8] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L., “A comparison of affine region detectors.”, (2005).
- [9] Bay, H., Ess, A., Tuytelaars, T., Gool, L., “Speeded-up robust features (SURF).”, (2008).