

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC BÁCH KHOA TP. HỒ CHÍ MINH**  
**KHOA ĐIỆN – ĐIỆN TỬ**



**TIỂU LUẬN MÔN HỌC**  
**THỊ GIÁC MÁY TÍNH**

# **PANORAMA**

**GVHD: TS. TRƯƠNG QUANG VINH**

**HV: Tiến Hoàng Trí Nghĩa - 1870048**

**Nguyễn Thanh Tùng - 1870249**

*TP. Hồ Chí Minh, tháng 06 năm 2019*

# LỜI CẢM ƠN

Để hoàn thành đề tài tiểu luận này, chúng em xin gửi lời cảm ơn chân thành tới Thầy TS. Trương Quang Vinh đã tận tình hướng dẫn, giảng dạy, cung cấp tài liệu để chúng em có thể hoàn thành đề tài tiểu luận này.

Chúng em xin chân thành cảm ơn quý Thầy Cô của Bộ môn Điện Tử nói riêng, Khoa Điện điện tử nói chung đã truyền đạt cho chúng em những kiến thức nền tảng thiết yếu để chúng em có thể hoàn thành tiểu luận thuận lợi.

Một lần nữa, chúng em xin chân thành cảm ơn.

Tp. Hồ Chí Minh, ngày 22 tháng 06 năm 2019

# MỤC LỤC

<b>CHƯƠNG 1 - TỔNG QUAN VỀ PANORAMA .....</b>	<b>1</b>
<b>1.1. TỔNG QUAN VỀ GHÉP ẢNH.....</b>	<b>1</b>
1.1.1. Giới thiệu về ghép ảnh.....	1
1.1.2. Các kiểu ghép ảnh.....	1
1.1.3. Các kiểu ghép ảnh.....	4
<b>1.2. KỸ THUẬT GHÉP ẢNH PANORAMA DỰA TRÊN ĐẶC TRƯNG BẤT BIẾN CỦA ẢNH.....</b>	<b>10</b>
1.2.1. Thuật toán tìm kiếm góc Harris .....	10
1.2.2. Trích chọn đặc trưng cục bộ bất biến SIFT .....	12
1.2.3. Đối sánh các đặc trưng bất biến .....	20
1.2.4. Tính toán ma trận Homography .....	22
<b>CHƯƠNG 2 – CHƯƠNG TRÌNH MÔ PHỎNG .....</b>	<b>28</b>
<b>2.1. LƯU ĐỒ GIẢI THUẬT.....</b>	<b>28</b>
<b>2.2. CHƯƠNG TRÌNH THỰC HIỆN .....</b>	<b>28</b>
<b>CHƯƠNG 3 – KẾT LUẬN .....</b>	<b>32</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>33</b>

## DANH MỤC HÌNH ẢNH

Hình 1.1- Ví dụ về khảm ảnh	2
Hình 1.2 – Máy ảnh panorama	3
Hình 1.3 – Ảnh Panorama	3
Hình 1.4 - Máy ảnh được đặt trên một giá trượt	5
Hình 1.5 - Mô hình chụp ảnh có ván trượt	5
Hình 1.6 - Chụp ảnh panorama bằng điện thoại	6
Hình 1.7 - Ví dụ về Recognize panorama	8
Hình 1.8 - Ảnh panorama chưa được trộn màu	9
Hình 1.9 - Ảnh panorama sau khi được trộn màu	9
Hình 1.10 - Ảnh panorama chưa cắt	9
Hình 1.11 - Ảnh panorama sau khi được cắt	9
Hình 1.12 - Cửa sổ trượt phát hiện góc Harris	10
Hình 1.13 - Minh họa các trường hợp $\lambda_1$ và $\lambda_2$	12
Hình 1.14 - Mô tả sự không hoàn thiện của thuật toán Harris	13
Hình 1.15 - Quá trình tính không gian đo (L) và hàm sai khác D	15
Hình 1.16 - Quá trình tìm điểm cực trị trong các hàm sai khác DoG	16
Hình 1.17 - Bộ mô tả cục bộ	20
Hình 1.18 - Ví dụ về đối sánh hai tập đặc trưng	21
Hình 1.19 - Phép chiếu Homography	22
Hình 2.1 – Lưu đồ giải thuật chương trình	28
Hình 2.2 – Ảnh đầu vào và ảnh được chuyển đổi	29
Hình 2.3 – Các điểm matching giữa 2 ảnh	30
Hình 2.4 – Ảnh sâu khi được khâu từ các ảnh khác	31
Hình 2.5 – Ảnh kết quả cuối cùng	31

# CHƯƠNG 1 - TỔNG QUAN VỀ PANORAMA

## 1.1. TỔNG QUAN VỀ GHÉP ẢNH

### 1.1.1. Giới thiệu về ghép ảnh

Ghép ảnh là quá trình kết hợp nhiều hình ảnh nhỏ xếp chồng lên nhau để tạo ra một bức ảnh lớn có độ phân giải cao hơn. Thông thường việc ghép ảnh được thực hiện bằng việc sử dụng các phần mềm máy tính.

Ghép ảnh có rất nhiều ứng dụng khác nhau. Ứng dụng truyền thống nhất là tạo nên ảnh không gian rộng và ảnh vệ tinh từ một tập các ảnh, dùng để xây dựng bản đồ địa lý, ghép các tấm ảnh chụp được trên bề mặt của một ngôi sao thành một tấm ảnh có độ phân giải lớn hơn, vv...

Các vấn đề chính trong ghép ảnh là sắp xếp các ảnh thành phần, nắn chỉnh biến dạng, biến đổi màu sắc và làm mờ đường biên giữa các ảnh. Tất cả các thao tác này nhằm làm cho bức ảnh ghép trông giống như là một ảnh liền chứ không phải là được ghép từ nhiều ảnh nhỏ.

Việc ghép các thành phần của các đối tượng lại với nhau để thu được các ảnh tương ứng hoàn thiện hơn là một công việc khó khăn rất nhiều khi phải làm thủ công, mặt khác các ảnh khi thu nhận để ghép thường hay bị lệch và biến dạng đi một khoảng nào đấy. Yêu cầu đặt ra cần xác định độ sai lệch về thông tin giữa các phần ảnh định ghép, sau đó hiệu chỉnh độ sai lệch và cuối cùng là ghép chúng lại.

Nghiên cứu kỹ thuật ghép ảnh còn mở ra cho chúng ta một hướng phát triển mới trong tương lai đó là kỹ thuật giả lập 3D.

### 1.1.2. Các kiểu ghép ảnh

#### 1.1.2.1. Khảm ảnh

Khảm ảnh là việc tạo ra hình ảnh mới bằng cách ghép các ảnh nhỏ vào một ảnh lớn sao cho khi nhìn tổng thể vào ảnh lớn thì ta vẫn có thể nhìn thấy nội dung của bức ảnh lớn trước đó.

Từ bức tranh tổng thể ban đầu, bằng các kỹ thuật xử lý khác nhau thì các bức ảnh nhỏ được lồng ghép vào đó tạo nên bức ảnh mới. Tất nhiên là nếu nhìn một cách

tổng thể thì nó vẫn chính là bức tranh lớn ban đầu có điều nó khác đi một chút bởi những chi tiết bên trong đã được thay thế bởi các hình ảnh đơn lẻ.



Hình 1.1 - Ví dụ về khảm ảnh

#### 1.1.2.2. Ảnh toàn cảnh (Panorama)

Ảnh toàn cảnh là một cách nhìn rộng của một vật chất trong không gian. Nó cho phép biểu thị một góc nhìn rộng của các bức tranh, bản vẽ đồ họa, nghệ thuật nhiếp ảnh, phim hoặc video, hay mô hình 3 chiều.

Thuật ngữ panorama xuất hiện trước khi chúng ta có máy ảnh panorama. Nguyên gốc của từ panorama được xác định là do họa sĩ người Ireland – Robert Baker dùng để mô tả những bức tranh diện rộng ở Edinburgh (Đức). Những bức tranh panorama này được cuộn trong một ống hình trụ và người ta kéo ra từ từ để bức tranh dần được hiển thị.

Năm 1881, họa sĩ người Hà Lan – Hendrik Willem Mesdag đã tạo nên trường phái Panorama Mesdag với những ống hình trụ cuộn các bức tranh toàn cảnh với kích cỡ khổng lồ, cao 14m và dài có thể từ 40 – 120m. Thế kỷ XIX có hai bức tranh panorama được coi là lớn nhất thời kỳ này, đó là tranh mô tả trận chiến tại Atlanta với chiều cao gần 13m, dài 110m. Bức tranh được xác định lớn nhất là ở Wrocław (Ba Lan) với kích thước là 15m x 120m.

Do nhu cầu của con người và sự phát triển của khoa học công nghệ, người ta đã sáng tạo ra máy ảnh panorama. Nếu máy ảnh thường chỉ có thể chụp ảnh với một góc 90 độ thì máy ảnh panorama có thể chụp với một góc 175 độ, 180 độ hoặc 360

độ. Trước một không gian rộng lớn, máy ảnh thường bất lực trong việc ghi lại hình ảnh ở một góc rộng nhưng máy panorama lại phát huy được tác dụng của nó. Máy panorama thường được chụp bằng phim dương bản (còn được gọi là film slide). Chụp xong có thể xem film là biết được ảnh sẽ được in ra như thế nào.

Chính vì góc ảnh của panorama rộng nên máy ảnh panorama không có ống kính dài như máy thường. Ống kính của máy panorama có hình vòng cung. Khi chụp ống kính sẽ quét từ trái sang phải nên chúng ta phải sử dụng chân máy khi chụp.



Hình 1.2 - Máy ảnh panorama

Ảnh Panorama nói đơn giản là xem hình ảnh với góc nhìn rộng hơn ảnh bình thường, tức là những khung ảnh cực kỳ rộng lớn mà một khung hình chụp bằng máy ảnh không thể hiện hết. Ảnh được ghép từ những ảnh số chụp từng phần của một phong cảnh (những cảnh này có phần gối nhau) lại thành một ảnh toàn cảnh hoàn chỉnh.

Ta có thể hiểu một cách đơn giản panorama là chế độ chụp ảnh khổ rộng bằng cách chụp nhiều tấm ảnh liên tiếp, với thông tin tấm ảnh trước được thể hiện một phần trong tấm ảnh sau, nhằm hỗ trợ người dùng. Sau đó với sự trợ giúp của phần mềm xử lý ảnh thì ta sẽ có được một tấm ảnh khổ rộng.



Hình 1.3 - Ảnh Panorama

### 1.1.3. Các kiểu ghép ảnh

Quá trình ghép ảnh panorama trải qua bốn bước cơ bản

#### 1.1.3.1. Thu nhận ảnh

Giai đoạn đầu tiên của việc ghép ảnh yêu cầu lựa chọn vị trí chụp ảnh phù hợp sao cho ảnh ít bị biến đổi hình học nhất. Cần xác định rõ loại ảnh panorama cần ghép để lựa chọn cách chụp phù hợp.

Các phương pháp thu nhận ảnh khác nhau có thể được sử dụng để có thể thu được hình ảnh đầu vào khác nhau rồi từ đó sẽ tạo ra các loại ảnh panorama khác nhau. Có 3 cách để chụp các tấm ảnh đầu vào là:

- Máy ảnh được để trên chân máy và ta vừa quay máy ảnh vừa chụp để có thể thu được ảnh đầu vào.
- Máy ảnh được để trên một tấm ván trượt, hình ảnh đầu vào thu được bằng cách vừa di chuyển tấm ván và vừa chụp. Ưu điểm của phương pháp này là đảm bảo sự ổn định và chuẩn xác của ảnh đầu vào, đảm bảo không có hoặc ít sự biến đổi hình học của ảnh, các ảnh đầu vào nằm trên cùng một đường thẳng.
- Người chụp trực tiếp giữ máy ảnh trên tay và chụp ảnh bằng cách xoay hoặc đi bộ theo một hướng vuông góc với hướng chụp của máy ảnh. Nhược điểm của cách chụp này là có thể ảnh đầu vào bị biến dạng do tác động từ người chụp như bị rung, ảnh bị nghiêng, và không nằm trên một đường thẳng.

Cả ba cách chụp trên đều cần phải đảm bảo được việc ảnh sau phải chứa khoảng 15% hoặc lớn hơn nội dung của ảnh trước để đảm bảo việc xác định vị trí cần nối của hai ảnh và cần cố gắng hạn chế nhất có thể việc hình ảnh bị biến đổi dẫn đến kết quả ghép không còn chính xác nữa.

#### 1.1.3.2. Phương pháp thu nhận ảnh

##### A- Phương pháp thu nhận ảnh bằng camera dịch chuyển

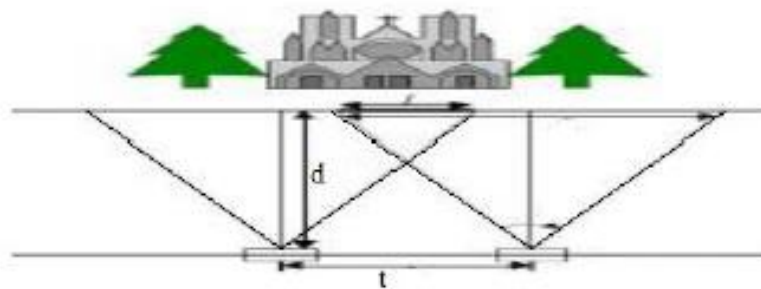
Sử dụng một máy ảnh để trên một tấm ván trượt và được di chuyển với hướng song song mặt phẳng cần chụp, máy ảnh được đặt trước đối tượng cần chụp và



hình ảnh được chụp bằng cách dịch chuyển tấm ván trượt và chụp đến một giới hạn mong muốn thì thôi.



Hình 1.4 - Máy ảnh được đặt trên một giá trượt



Hình 1.5 - Mô hình chụp ảnh có ván trượt

Trong đó  $t$  là khoảng trượt của camera giữa hai lần chụp,  $d$  là khoảng cách giữa máy ảnh và đối tượng được chụp.

Cần đảm bảo rằng hướng trượt của máy ảnh cần phải song song với mặt phẳng chứa đối tượng cần chụp, nếu không sẽ dẫn đến việc kích thước của đối tượng sẽ bị biến đổi giữa hai ảnh.

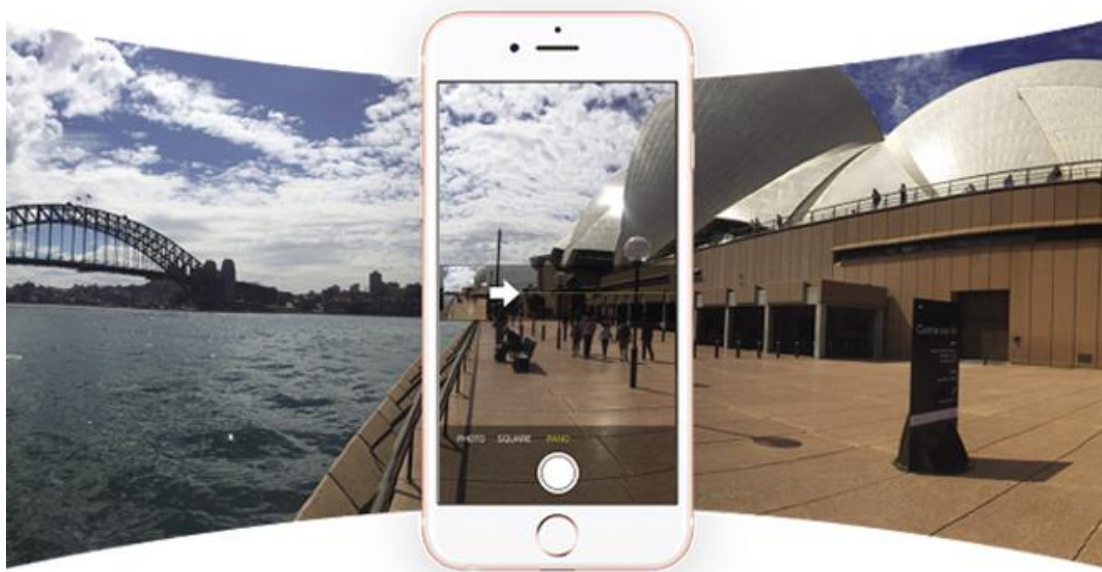
Tuy nhiên nhược điểm của phương pháp chụp ảnh này là ảnh sau khi được ghép sẽ không cho người xem cảm giác chân thật.

## **B- Phương pháp thu nhận ảnh bằng cách sử dụng máy ảnh cầm tay**

Người dùng chỉ cần giữ máy ảnh và vừa chụp vừa xoay hoặc di chuyển vuông góc với hướng chụp. Tuy nhiên hình ảnh thu được thì lại khó ghép hơn do tác động như ảnh bị nghiêng, rung ...

Trong trường hợp người dùng chụp ảnh bằng cách quay người thì lúc đó người chụp đóng vai trò như là một chân máy ảnh, nhưng vẫn sẽ có sự sai lệch do tác động không mong muốn.

Trong trường hợp người dùng chụp ảnh bằng cách di chuyển song song với mặt phẳng chứa đối tượng cần chụp thì lúc đó người chụp lại đóng vai trò như là một tấm trượt. Tuy nhiên khi đó sẽ khó đảm bảo được khoảng cách ổn định từ máy ảnh đến mặt phẳng chứa đối tượng được chụp.



Hình 1.6 - Chụp ảnh panorama bằng điện thoại

#### 1.1.3.3. Biến đổi ảnh

Là quá trình thay đổi về mặt hình học của ảnh để phù hợp với ảnh liền kề trước đó để tạo thành một ảnh panorama. Các ảnh cần được sắp xếp theo đúng thứ tự chụp trước sau để có thể đảm bảo được độ chính xác cao nhất có thể. Biến đổi ảnh là quá trình quan trọng nhất trong kỹ thuật ghép ảnh panorama. Quá trình bao gồm ba bước:

##### **Bước 1: Biến đổi ảnh về một không gian chiếu xác định**

Có nhiều không gian chiếu khác nhau như mặt cầu, hình trụ. Chiếu hình ảnh lên một “tấm màn chiếu” giúp biểu diễn được tấm ảnh ghép một cách chân thực hơn. Cũng có thể hiểu nó như một tấm bảng mà ta sẽ ghép các tấm ảnh lại thành ảnh panorama bằng cách dán các tấm ảnh chụp được này lên tấm bảng đó và sử

dụng các phép biến đổi ảnh để thay đổi ảnh sao cho ảnh này phù hợp hơn với ảnh liền kề trước đó.

## **Bước 2: Căn chỉnh ảnh**

Trong kỹ thuật ghép ảnh panorama thì căn chỉnh ảnh là một trong những phần quan trọng nhất. Để có thể căn chỉnh những tấm ảnh thì cần xác định được các điểm giống nhau giữa hai bức ảnh và làm cho khoảng cách giữa hai điểm đó là ngắn nhất, thậm chí là bằng không (trường hợp trùng nhau). Để có thể làm được việc đó ta cần phải trải qua nhiều bước khác nhau.

### **A- Xác định điểm đặc trưng của ảnh**

Điểm đặc trưng (hay còn gọi là key point) có thể coi là phần thông tin quan trọng được trích xuất từ ảnh và nổi bật, rõ ràng nhất của ảnh. Các điểm đặc trưng sẽ không bị thay đổi do sự biến dạng của ảnh như bị xoay hoặc do tịnh tiến ... Số lượng các điểm đặc trưng phải đủ lớn để có thể tính được một mô hình biến đổi giữa hai ảnh.

Tùy thuộc vào loại đặc trưng mà sử dụng phương pháp xác định phù hợp. Các điểm đặc trưng có thể xác định dựa vào các kỹ thuật như thuật toán phát hiện góc Harris, phát hiện cạnh...

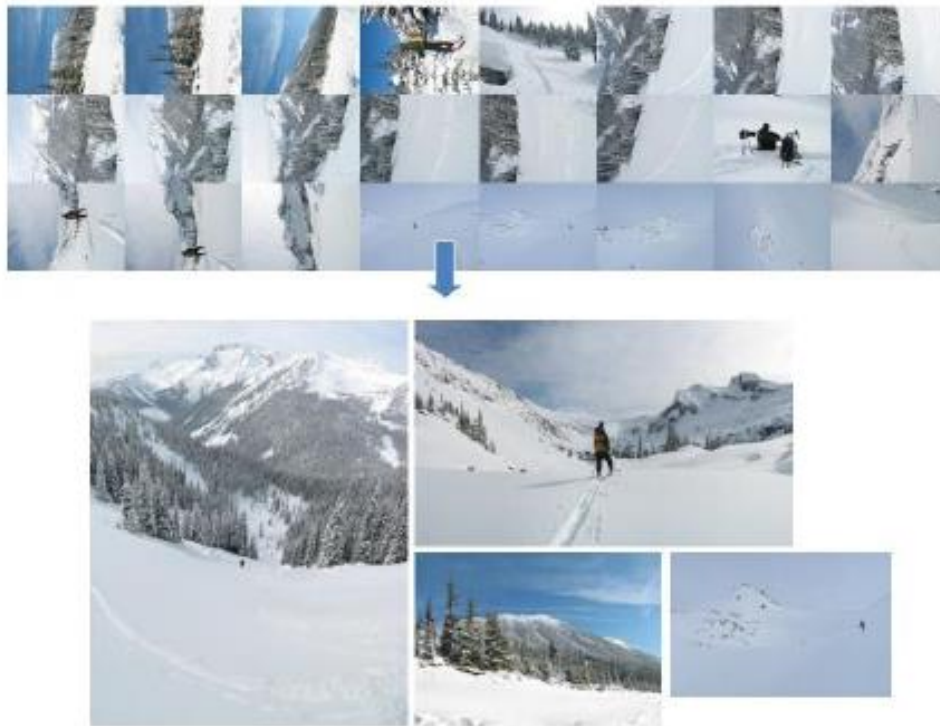
**Điểm đặc trưng bất biến SIFT:** Là thuật toán để mô tả các đặc trưng cục bộ dựa trên các điểm quan tâm và bất biến đối với việc co giãn hoặc xoay ảnh, không bị ảnh hưởng bởi độ sáng và nhiễu trong ảnh.

**Tính toán ma trận đồng nhất:** Ma trận đồng nhất giữa hai ảnh có thể được tính bằng thuật toán RANSAC. Ma trận đồng nhất là ma trận thể hiện sự biến đổi của ảnh này so với ảnh kia.

### **B- Nhận diện ảnh panorama (Recognizing panorama)**

Trong trường hợp ảnh đầu vào không theo thứ tự hoặc là ảnh thành phần thuộc nhiều ảnh panorama khác nhau thì việc xác định từng cặp ảnh để ghép với nhau là một việc quan trọng. Kết quả trả về có thể là nhiều ảnh panorama khác

nhau như hình 1.7 Brown và Lowe đã gọi đây là kỹ thuật nhận diện ảnh panorama (recognizing panorama).



Hình 1.7 - Ví dụ về Recognize panorama

### Bước 3: Chiếu ảnh

Hướng giải quyết là lựa chọn một ảnh làm tâm và biến đổi các ảnh khác theo ảnh đó. Có thể chiếu ảnh ghép lên một mặt phẳng, khi đó sẽ nhận được ảnh panorama phẳng. Cách khác là có thể sử dụng phép chiếu hình trụ (Szeliski 1994, Chen 1995) hoặc phép chiếu mặt cầu (Szeliski và Shum 1997).

#### 1.1.3.4. Trộn ảnh

Sau khi ghép ảnh, kết quả thu được là một ảnh panorama. Tuy nhiên do tác động bên ngoài như ánh sáng, độ phơi sáng nên khi chụp ảnh đầu vào sẽ dẫn đến sự chênh lệch độ sáng màu giữa hai khu vực ảnh giống nhau giữa hai ảnh nên khi ghép vào sẽ bị nhìn rõ phần bị ghép với nhau như hình 1.8. Vậy cần phải làm cân bằng độ sáng của phần được ghép của hai bức ảnh để giảm đi độ rõ của vùng được nối như hình 1.9

### Ghép ảnh panorama dựa trên đối sánh đặc trưng bất biến



Hình 1.8 - Ảnh panorama chưa được trộn màu



Hình 1.9 - Ảnh panorama sau khi được trộn màu

#### 1.1.3.5. Cắt ảnh

Cắt ảnh là kỹ thuật dùng để loại bỏ những đối tượng hoặc vùng ảnh bị thừa ra ngoài bức ảnh kết quả.



Hình 1.10 - Ảnh panorama chưa cắt



Hình 1.11 - Ảnh panorama sau khi được cắt



## 1.2. KỸ THẬT GHÉP ẢNH PANORAMA DỰA TRÊN ĐẶC TRƯNG BẤT BIẾN CỦA ẢNH

### 1.2.1. Thuật toán tìm kiếm góc Harris

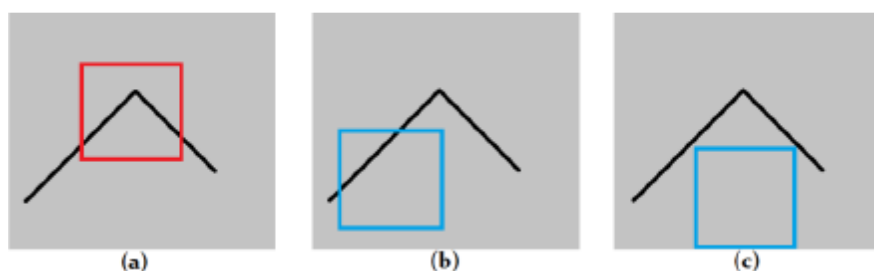
#### 1.2.1.1. Giới thiệu

Một trong những phương pháp tìm kiếm nổi bật cơ bản nhất chính là phương pháp tìm kiếm các điểm nổi của đường biên, các điểm nằm trên đường cong mà có độ cong tối đa hay còn gọi là các điểm góc.

Ban đầu các điểm này được phát hiện nhờ vào “độ nhọn” (sharp) của đường biên: biên của đối tượng được lưu dưới dạng mã xích, góc được phát hiện thông qua việc tìm kiếm những vị trí trên biên bị uốn một cách đáng kể. kỹ thuật phát hiện góc này rất phức tạp và triển khai trên nhiều bước.

Thuật toán Harris sử dụng một cửa sổ có thể trượt theo bất kỳ hướng nào bằng cách sử dụng hàm Gaussian và phép khai triển Taylor.

Về mặt ý tưởng, thuật toán Harris sẽ tìm kiếm sự thay đổi lớn về cường độ xám theo các hướng khác nhau bằng cách dùng một cửa sổ nhỏ để làm nhiệm vụ rà soát và phát hiện những điểm định nghĩa là góc.



Hình 1.12 - Cửa sổ trượt phát hiện góc Harris

Trong Hình 1-12 (a) cửa sổ trượt nằm trong vùng hình ảnh có chứa góc, khi di chuyển theo bất kỳ hướng nào đều có sự thay đổi về cường độ xám.

Trong Hình 1-12 (b) cửa sổ trượt nằm trên vùng hình ảnh có chứa cạnh, khi di chuyển cửa sổ trượt theo hai hướng của cạnh sẽ không có sự thay đổi về cường độ xám.

Trong Hình 1-12 (c) cửa sổ trượt nằm trên vùng hình ảnh không có góc cạnh, sau khi di chuyển cửa sổ trượt sẽ không có sự thay đổi về cường độ xám.

Dựa vào điều này ta có thể phát hiện ra điểm nào là điểm góc, điểm nào không phải.

### 1.2.1.2. Chi tiết thuật toán tìm kiếm góc Harris

Giả sử cho một ảnh xám ( $\mathbf{I}$ ), với mỗi điểm ( $\mathbf{u}, \mathbf{v}$ ) và độ dịch chuyển ( $\mathbf{x}, \mathbf{y}$ ) ta có thể tính toán được sự thay đổi trung bình cường độ xám bằng một cửa sổ dịch chuyển từ ( $\mathbf{u}, \mathbf{v}$ ) tới ( $\mathbf{u} + \mathbf{x}, \mathbf{v} + \mathbf{y}$ ) như sau:

$$S(\mathbf{x}, \mathbf{y}) = \sum_u \sum_v w(u, v) (\mathbf{I}(\mathbf{u} + \mathbf{x}, \mathbf{v} + \mathbf{y}) - \mathbf{I}(\mathbf{u}, \mathbf{v}))^2$$

Trong đó:

$\mathbf{S}(\mathbf{x}, \mathbf{y})$  là tổng bình phương giá trị độ lệch hay còn gọi là sự thay đổi cường độ xám tại ( $\mathbf{x}, \mathbf{y}$ )

$\mathbf{W}(\mathbf{u}, \mathbf{v})$  là cửa sổ tại ( $\mathbf{u}, \mathbf{v}$ )

$\mathbf{I}(\mathbf{u}, \mathbf{v})$  và  $\mathbf{I}(\mathbf{u} + \mathbf{x}, \mathbf{v} + \mathbf{y})$  là giá trị cường độ xám của pixel tại các vị trí ( $\mathbf{u}, \mathbf{v}$ ) và  $\mathbf{I}(\mathbf{u} + \mathbf{x}, \mathbf{v} + \mathbf{y})$

Giá trị  $\mathbf{I}(\mathbf{u} + \mathbf{x}, \mathbf{v} + \mathbf{y})$  có thể được khai triển theo công thức Taylor như sau:

$$\mathbf{I}(\mathbf{u} + \mathbf{x}, \mathbf{v} + \mathbf{y}) \approx \mathbf{I}(\mathbf{u}, \mathbf{v}) + \mathbf{I}_x(\mathbf{u}, \mathbf{v}) \mathbf{x} + \mathbf{I}_y(\mathbf{u}, \mathbf{v}) \mathbf{y}$$

Với  $\mathbf{I}_x, \mathbf{I}_y$  là đạo hàm theo thành phần  $\mathbf{x}, \mathbf{y}$ .

Từ đó, ta có thể được viết lại như sau:

$$S(\mathbf{x}, \mathbf{y}) = \sum_u \sum_v w(u, v) (\mathbf{I}_x(\mathbf{u}, \mathbf{v}) \mathbf{x} - \mathbf{I}_y(\mathbf{u}, \mathbf{v}) \mathbf{y})^2$$

Biểu diễn dưới dạng ma trận thì  $\mathbf{S}(\mathbf{x}, \mathbf{y})$  ta có:

$$S(\mathbf{x}, \mathbf{y}) \approx (\mathbf{x}, \mathbf{y}) \mathbf{A} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

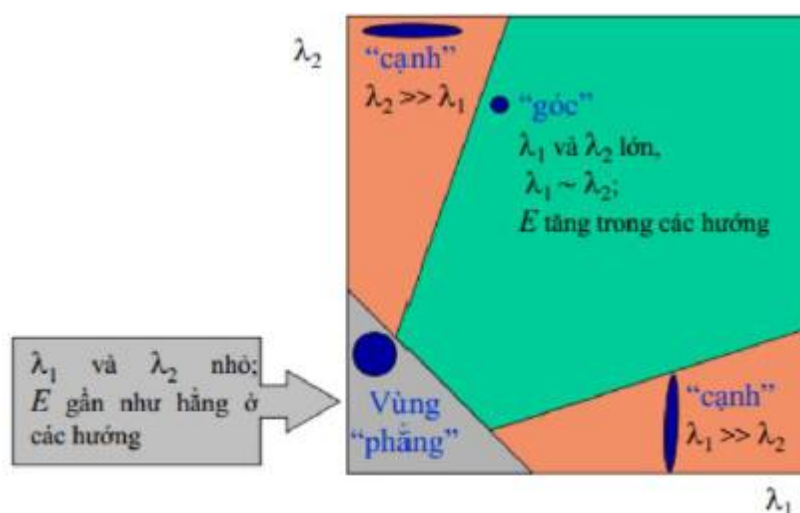
Trong đó,  $\mathbf{A}$  là một cấu trúc như sau:

$$S(\mathbf{x}, \mathbf{y}) = \sum_u \sum_v w(u, v) \begin{bmatrix} \mathbf{I}_x^2 & \mathbf{I}_x \mathbf{I}_y \\ \mathbf{I}_x \mathbf{I}_y & \mathbf{I}_y^2 \end{bmatrix} = \begin{bmatrix} \langle \mathbf{I}_x^2 \rangle & \langle \mathbf{I}_x \mathbf{I}_y \rangle \\ \langle \mathbf{I}_x \mathbf{I}_y \rangle & \langle \mathbf{I}_y^2 \rangle \end{bmatrix}$$

Gọi  $\lambda_1$  và  $\lambda_2$  là các giá trị riêng của A, k là hằng số được xác định thông qua thực nghiệm, thường có giá trị trong khoảng [0.04, ..., 0.15].

Khi đó, biểu thức dưới đây sẽ quyết định xem cửa sổ w có chứa góc hay là không:

$$M_c = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 = \det(A) - k(\text{trace}^2(A))$$



Hình 1.13 - Minh họa các trường hợp  $\lambda_1$  và  $\lambda_2$

Nếu cả  $\lambda_1$  và  $\lambda_2$  đều nhỏ. Có nghĩa là hàm  $S(x, y)$  gần như không thay đổi theo bất kỳ hướng nào. Khi đó vùng ảnh nằm trong cửa sổ gần như không có sự thay đổi về cường độ. Tức là trường hợp này không tìm thấy điểm góc.

Nếu  $\lambda_1$  là lớn và  $\lambda_2$  là nhỏ hoặc ngược lại. Có nghĩa là hàm  $S(x, y)$  có sự thay đổi nhỏ nếu cửa sổ trượt theo một hướng, và có sự thay đổi đáng kể nếu dịch chuyển theo hướng trực giao. Điều này cho thấy tồn tại một cạnh.

Nếu cả  $\lambda_1$  và  $\lambda_2$  đều lớn. Có nghĩa là hàm  $S(x, y)$  có sự thay đổi đáng kể về cường độ xám khi dịch chuyển cửa sổ trượt theo bất kỳ hướng nào. Điều này cho thấy có tồn tại một điểm góc.

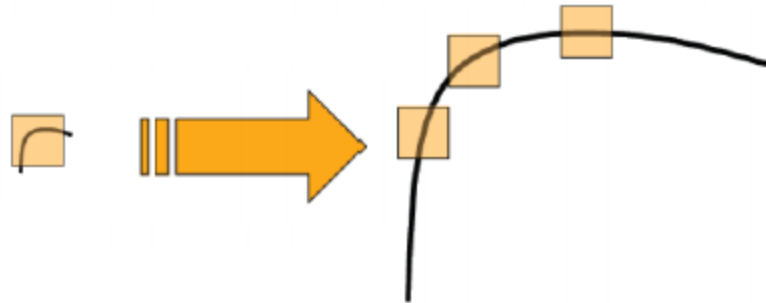
## 1.2.2. Trích chọn đặc trưng cục bộ bất biến SIFT

### 1.2.2.1. Giới thiệu và định nghĩa

Trong quá trình nghiên cứu việc so sánh đôi chiều các điểm đặc trưng của ảnh số, người ta thường gặp những vấn đề có liên quan đến việc các điểm đặc trưng



tương đồng giữa các ảnh không có được sự so sánh đối chiếu chính xác nếu như các điểm đó không có cùng tỉ lệ. Điều này có nghĩa là nếu ta cố gắng tìm ra các điểm đặc trưng tương đồng từ hai hình ảnh được chụp lại từ hai góc độ khác nhau của một khung hình, ta sẽ không có được kết quả tốt nhất.



Hình 1.14 - Mô tả sự không hoàn thiện của thuật toán Harris

Trong hình trên, ta dễ dàng nhận ra ở bên trái là một góc đã được thuật toán tìm kiếm góc Harris phát hiện. Thế nhưng nếu như ta xem xét góc này với một tỉ lệ lớn hơn nhiều thì đó không còn là một góc theo thuật toán Harris nữa, nhưng trên thực tế, đó hoàn toàn là một góc.

Để giải quyết vấn đề này, khái niệm “Điểm đặc trưng cục bộ bất biến SIFT” (Scale-Invariant Feature Transform): đã được David G. Lowe đưa ra hồi năm 2004 trong “Distinctive Image Features from Scale-Invariant Keypoints” với các tính chất sau:

- Là bất biến khi

- Thay đổi tỷ lệ ảnh
- Quay ảnh
- Thay đổi góc nhìn
- Thêm nhiễu ảnh
- Thay đổi cường độ chiếu sáng ảnh

- Các đặc trưng cục bộ bất biến được trích rút từ các điểm nổi bật trên ảnh.

#### 1.2.2.2. Thuật toán

Phương pháp trích chọn điểm đặc trưng cục bộ bất biến SIFT gồm các bước:

- Phát hiện các điểm cực trị Scale-Space: Tìm những khu vực chứa những điểm đặc trưng tiềm năng có tính chất không thay đổi dưới các phép phóng và xoay ảnh.
- Định vị các điểm nổi bật: Một hàm kiểm tra sẽ được đưa ra để quyết định xem các điểm nổi bật tiềm năng có được lựa chọn hay không.
- Xác định hướng cho các điểm nổi bật: Mỗi điểm nổi bật 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 nổi bật: 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.

### 1.2.2.3. Phát hiện các điểm cực trị Scale-Space

Định nghĩa không gian tỉ lệ của một hình ảnh được mô tả bằng hàm dưới đây:

$$L(x, y, \sigma) = G(x, y, k\sigma) * I(x, y)$$

Với  $G(x, y, k\sigma)$  là biến tỉ lệ Gaussian (Variable scale Gaussian)

$$G(x, y, k\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

$I(x, y)$  là ảnh đầu vào

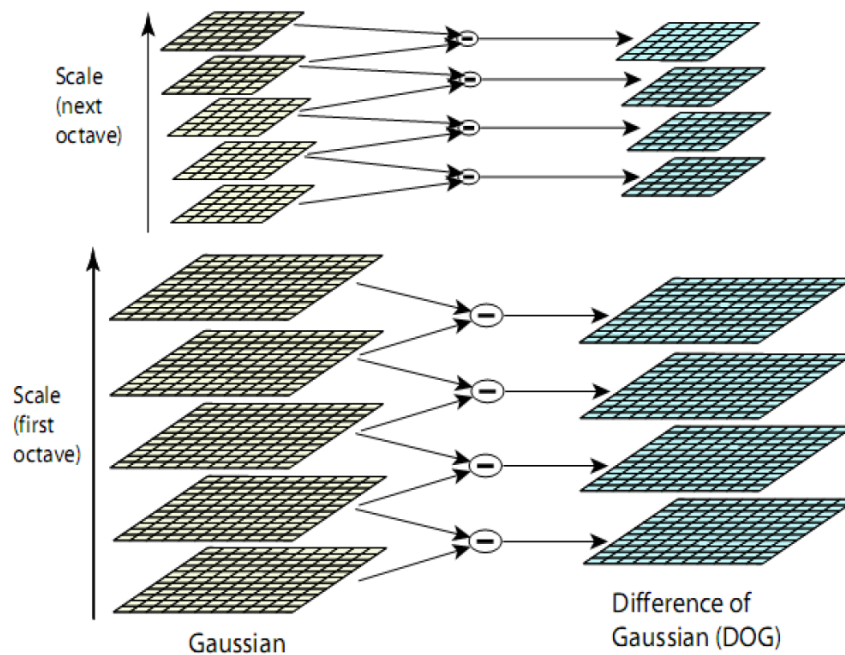
$L(x, y, \sigma)$  là hàm không gian tỉ lệ của ảnh

Để tìm những điểm nổi bật có tính bất biến cao, ta sẽ tìm cực trị cục bộ của hàm sai khác DoG (Difference of Gaussian), kí hiệu là  $D(x, y, \sigma)$ . Hàm này được tính toán từ sự sai khác giữa 2 không gian đo cạnh nhau của một ảnh với tham số đo lệch nhau một hằng số  $k$ .

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Với  $D(x, y, \sigma)$  được tính bằng phép trừ giữa Hàm không gian tỉ lệ của ảnh  $L(x, y, \sigma)$  và  $k$  lần của chính nó.

Lý do để lựa chọn hàm Gaussian là vì kỹ thuật này rất hiệu quả để tính toán  $L$  (cũng như làm tăng độ mịn của ảnh), mà  $L$  thì luôn phải được tính rất nhiều để mô tả đặc trưng trong không gian đo, và sau đó  $D$  sẽ được tính bằng phép trừ ma trận điểm ảnh, một phép tính toán đơn giản.



Hình 1.15 - Quá trình tính không gian đo (L) và hàm sai khác D

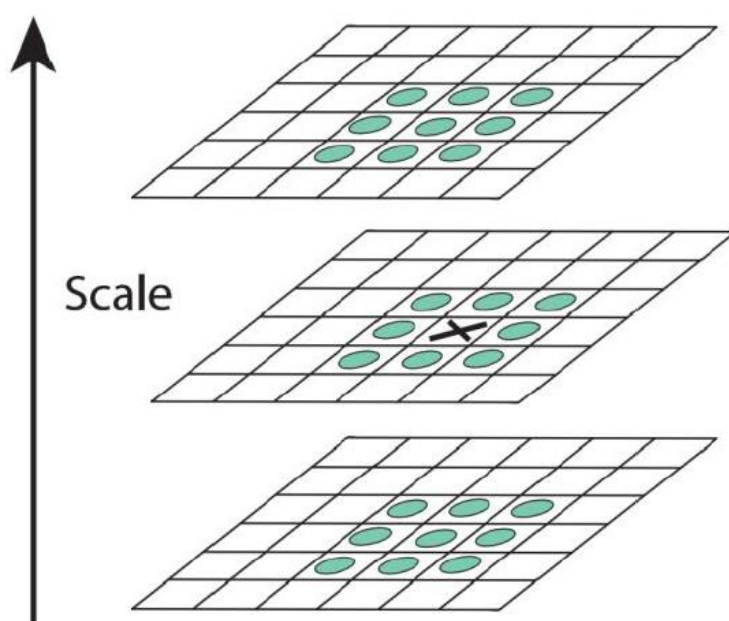
Hơn nữa, hàm sai khác DoG có thể được sử dụng để tạo ra một xấp xỉ gần với đạo hàm bậc hai Laplace có kích thước chuẩn của hàm Gaussian ( $\sigma^2 \nabla^2 G$ ) do tác giả Lindeberg đề xuất năm 1994. Ông đã chỉ ra rằng việc chuẩn hóa đạo hàm bậc hai với hệ số  $\sigma^2$  là cần thiết cho bất biến đo trở nên đúng. Cụ thể, ông đã công bố rằng các giá trị cực đại và cực tiểu của  $(\sigma^2 \nabla^2 G)$  chính là những giá trị có tính ổn định nhất (bất biến cao) so với một loạt các hàm đánh giá khác như: gradient, Hessian hay Harris.

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)(\sigma^2 \nabla^2 G)$$

Từ công thức này, ta thấy khi mà hàm sai khác DoG được tính toán tại các tham số đo lệch nhau một hằng số  $k$ , thì ta có thể sử dụng DoG để xấp xỉ đạo hàm bậc hai Laplace của Gaussian. Vì hệ số  $(k - 1)$  trong phương trình trên là hằng số trong mọi không gian đo nên nó sẽ không ảnh hưởng đến việc tìm các vị trí cực trị.

Sai số trong việc xấp xỉ đạo hàm bậc 2 tiến về 0 khi  $k$  gần với 1. Tuy nhiên, các kết quả thử nghiệm của tác giả cho thấy quá trình xấp xỉ đạo hàm không ảnh hưởng đến việc dò tìm các vị trí cực trị thậm chí ngay cả khi chọn  $k$  khá xa, ví dụ  $k = \sqrt{2}$ .

Sau khi áp dụng hàm DoG ta thu được các lớp kết quả khác nhau từ ảnh gốc, bước tiếp theo là tìm các cực trị trong các lớp kết quả theo từng miền cục bộ. Cụ thể là tại mỗi điểm trên các lớp kết quả sẽ được so sánh với 8 điểm lân cận trên cùng lớp và 9 điểm lân cận trên mỗi lớp khác.



Hình 1.16 - Quá trình tìm điểm cực trị trong các hàm sai khác DoG

Trong hình trên: điểm đánh dấu x sẽ được so sánh với 26 điểm lân cận (đánh dấu vòng tròn xanh). Điểm này sẽ được lấy làm điểm tiềm năng (điểm có thể làm điểm đặc biệt – candidate keypoint) nếu nó có giá trị lớn nhất hoặc nhỏ nhất so với 26 điểm lân cận như trên. Giải pháp cho việc tìm các điểm tiềm năng này là sử dụng thuật toán blob detection (dò tìm điểm) do Lindeberg đề xuất.

Vì số lượng các cực trị là rất lớn, vì vậy để tăng sự hiệu quả khi dò tìm các điểm cực trị (dò các điểm cực trị tốt nhất thay vì phải dò hết), ta cần xác định tần số lấy mẫu trong không gian đo và tần số lấy mẫu trong không gian quan sát (không gian ảnh). Thật không may là ta không thể xác định cả 2 loại tần số này một cách tự động trong mỗi tiến trình dò tìm.

Thay vì vậy, các tần số này sẽ được xác định thông qua phương pháp thử nghiệm. Sau khi thử nghiệm với nhiều nguồn dữ liệu ảnh khác nhau, tác giả đã chỉ ra tần số lấy mẫu trong không gian đo tốt nhất là 3 (giữ lại 3 lớp trong mỗi bộ 8 lớp), và tần số lấy mẫu  $\sigma = 16$ .

#### 1.2.2.4. Định vị điểm nổi bật

Sau bước một, ta sẽ thu được rất nhiều điểm tiềm năng có thể làm điểm đặc biệt, tuy nhiên một số điểm trong số đó là không cần thiết. Ở bước này, ta sẽ loại bỏ các điểm có độ tương phản kém (nhạy cảm với nhiễu) hoặc tính đặc trưng cục bộ ít hơn các điểm khác hoặc có xu hướng là đường biên đối tượng. Bước thực hiện này gồm 3 công đoạn:

##### a. Phép nội suy lân cận cho vị trí đúng của điểm tiềm năng

Phép nội suy lân cận sử dụng mở rộng Taylor (Taylor expansion) cho hàm Difference-of-Gaussian  $D(x, y, \sigma)$ :

$$D(x) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X$$

Với  $X = (x, y, \sigma)^T$  là độ dịch so với các điểm lân cận của điểm lấy mẫu.

Vùng chứa điểm nổi bật được xác định qua:

$$\hat{X} = \frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X}$$

Nếu  $\hat{X} > 0.5$ , điểm nổi bật nằm ở gần điểm lấy mẫu khác

Nếu  $\hat{X} < 0.5$ , những điểm này sẽ được đưa vào tập hợp mẫu tốt nhất để tiếp tục phân tích.

##### b. Loại trừ các điểm có tính tương phản kém

Các điểm nhạy cảm với độ sáng và nhiễu thì không được trở thành điểm đặc biệt và cần loại bỏ khỏi danh sách điểm tiềm năng.

Dùng  $D(\hat{X})$  để loại những điểm cực trị không ổn định (có độ tương phản thấp).

Thay  $\hat{X}$  vào  $D(x)$ , ta có:

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{X}$$

Nếu  $D(\hat{X}) < 0.03$  thì điểm lấy mẫu đó sẽ bị loại

### c. Loại bỏ các điểm dư thừa theo biên

Sử dụng hàm DoG sẽ cho tác động mạnh đến biên khi vị trí của biên là khó xác định và vì vậy các điểm tiềm năng trên biên sẽ không bất biến và bị nhiễu. Và để tăng sự ổn định cho các điểm sẽ được chọn làm điểm đặc biệt ta sẽ loại trừ các điểm tiềm năng khó định vị (tức là vị trí dễ thay đổi khi có nhiễu do nằm ở biên).

Sau khi áp dụng hàm DoG sẽ làm đường biên ảnh không rõ ràng và độ cong chính sẽ có giá trị lớn hơn nhiều so với độ cong dọc theo biên vì vậy cần loại bỏ bớt các điểm đặc biệt dọc theo cùng một biên. Giải pháp cho việc này là sử dụng giá trị của ma trận Hessian cấp 2:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$$

Các giá trị riêng của  $H$  tỉ lệ thuận với độ cong của  $D$ , các giá trị riêng  $\beta$  (giá trị nhỏ) và  $\alpha$  (giá trị lớn) có tỉ lệ  $r = \alpha/\beta$  sẽ được sử dụng. Các phần tử của  $H$  là  $D_{xx}$  và  $D_{yy}$

#### 1.2.2.5. Xác định hướng cho các điểm nổi bật

Bằng việc gán một hướng cho mỗi điểm nổi bật dựa vào thuộc tính ảnh cục bộ, bộ mô tả điểm nổi bật có thể được biểu diễn tương đối so với hướng này và do đó đạt được tính bất biến đối với các hiện tượng quay ảnh. Độ đo của các điểm nổi bật được sử dụng để tìm ra một ảnh đã lọc Gaussian  $L$  với kích thước gần nhất sao cho mọi tính toán sẽ được thực hiện trong cùng một bất biến về độ đo.

Tính toán hướng và độ lớn cho điểm nổi bật:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \left( \frac{(L(x, y+1) - L(x, y-1))}{(L(x, y+1) - L(x-1, y))} \right)$$

Với:

$m(x, y)$  là độ lớn của vector định hướng.

$\theta(x, y)$  là hướng của vector định hướng (biểu diễn qua góc  $\theta$ )

$L(x, y)$  là ảnh Gaussian ở tỷ lệ nhỏ nhất

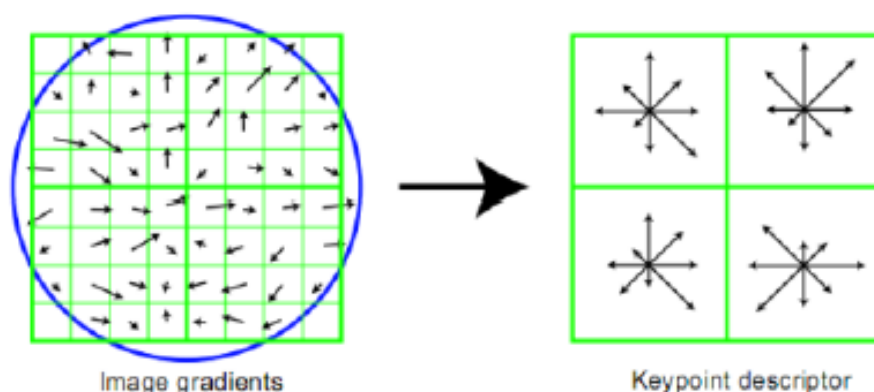
#### 1.2.2.6. Mô tả các điểm nổi bật

Các phép xử lý trên đã thực hiện dò tìm và gán tọa độ, kích thước và hướng cho mỗi điểm đặc trưng. Các tham số đó yêu cầu một hệ thống tọa độ cục bộ 2D có thể lập lại được để mô tả vùng ảnh cục bộ và nhờ vậy tạo ra sự bất biến đối với các tham số đó. Bước này sẽ tính toán một bộ mô tả cho một vùng ảnh cục bộ mà có tính đặc trưng cao (bất biến với các thay đổi khác nhau về độ sáng, thu – phóng ảnh, xoay).

Một cách tiếp cận đơn giản đó là lấy mẫu mật độ ảnh cục bộ lân cận điểm đặc trưng ở một độ đo thích hợp, và đối sánh các mật độ này sử dụng độ đo tương quan chuẩn.

Cách tiếp cận tốt hơn được đưa ra bởi Edelman, Intrator và Poggio (1997). Cách tiếp cận này dựa trên một mô hình thị giác sinh học, cụ thể là mô hình neuron phức tạp trong hệ thống não bộ. Các neuron sẽ tương ứng với một gradient tại một hướng và tần số không gian cụ thể, nhưng vị trí của gradient trên võng mạc được phép trượt trên một phạm vi nhỏ của khung nhìn.

Ảnh trái là mô phỏng biên độ gradient và hướng tại mỗi mẫu ảnh trong một vùng lân cận với điểm nổi bật. Các giá trị đó tập trung trong một cửa sổ Gaussian (nằm bên trong vòng tròn). Các mẫu này sau đó được gom lại thành một lược đồ hướng mô tả vắn tắt nội dung trong 4x4 vùng con như được mô tả bên phải với độ dài của mỗi hàng tương ứng với tổng biên độ gradient gần hướng đó bên trong một vùng.



Hình 1.17 - Bộ mô tả cục bộ

Điểm hấp dẫn sau khi được xác định hướng sẽ được biểu diễn dưới dạng các vector  $4 \times 4 \times 8 = 128$  chiều (Số chiều = 8 hướng x (4x4) điểm hấp dẫn = 128 chiều) bằng cách tổng hợp các vector định hướng của các điểm trong khu vực, các vector này có đặc điểm:

- Chung gốc.
- Độ dài mỗi vector tương ứng với độ lớn gradient  $m$  của nó.

### 1.2.3. Đối sánh các đặc trưng bất biến

#### 1.2.3.1. 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. Việc đối sánh theo nội dung sẽ truy vấn hình ảnh bằng phương pháp đo tương tự dựa trên các đặc trưng, việc xác định nó có thể dưới nhiều hình thức như phát hiện biên, màu sắc, vị trí điểm ảnh ..., các phương pháp như histogram, màu sắc và phân tích sử dụng biểu đồ để xác định độ tương tự.

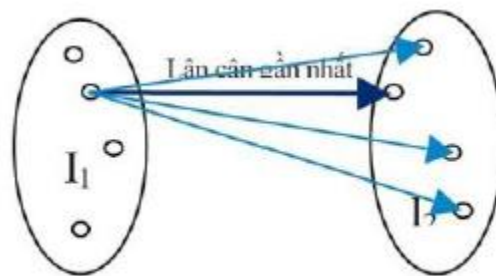
Do đó, độ đo có ý nghĩa quan trọng trong đối sánh ảnh dựa trên nội dung. Độ đo mang ý nghĩa quyết định kết quả đối sánh sẽ như thế nào, mức độ chính xác ra sao. Nhiều phép đo khoảng cách đã được khai thác trong việc đối sánh ảnh bao gồm:



khoảng cách Euclide, khoảng cách Cousin, khoảng cách giao nhau của biểu đồ histogram, khoảng cách Minkowski ....

### 1.2.3.2. Độ đo khoảng cách và độ đo tương tự

Việc đối sánh sẽ được thực hiện trên các tập keypoint tìm được. Bước chính trong kỹ thuật đối sánh sẽ thực hiện tìm tập con keypoint so khớp nhau ở hai ảnh, để thực hiện việc này sẽ tìm các cặp keypoint trùng nhau lần lượt ở hai ảnh. Tập con các keypoint so khớp chính là vùng ảnh tương đồng. Việc đối sánh hai tập đặc trưng quy về bài toán tìm láng giềng gần nhất của mỗi điểm đặc trưng.



Hình 1.18 - Ví dụ về đối sánh hai tập 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, ANN được viết tắt của Approximative Nearest Neighbour. Nó cho phép tổ chức dữ liệu dưới dạng kd-tree. Cụ thể là hai điểm trong không gian đặc trưng được coi là giống nhau nếu khoảng cách Euclide giữa hai điểm là nhỏ nhất và tỷ số giữa khoảng cách gần nhất với khoảng cách gần nhì phải nhỏ hơn một ngưỡng nào đó.

Giả sử cặp keypoint 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 Euclide giữa A và B được tính bằng công thức:

$$d(A, B) = \sum (a_i - b_i)^2$$

## 1.2.4. Tính toán ma trận Homography

### 1.2.4.1. Vài nét về Homography

Trong toán học, Homography là sự dịch chuyển sử dụng phép chiếu hình học, hay nói cách khác nó là sự kết hợp của cặp điểm trong phép chiếu phối cảnh. Ảnh thực trong không gian ba chiều có thể biến đổi về không gian ảnh bằng phép chiếu thông qua ma trận biến đổi Homography hay còn gọi là ma trận H. Các phép chiếu biến đổi thông qua ma trận Homography không đảm bảo về kích thước và góc của vật được chiếu nhưng lại đảm bảo về tỷ lệ.

Trong lĩnh vực thị giác máy, Homography là một ánh xạ từ mặt phẳng đối tượng đến mặt phẳng ảnh. Ma trận homography thường có liên quan đến các công việc xử lý giữa hai ảnh bất kỳ và có ứng dụng rất rộng rãi trong các công tác sửa ảnh, ghép ảnh, tính toán sự chuyển động, xoay hay dịch chuyển giữa hai ảnh.

Ta có công thức sau:

$$HX = sX'$$

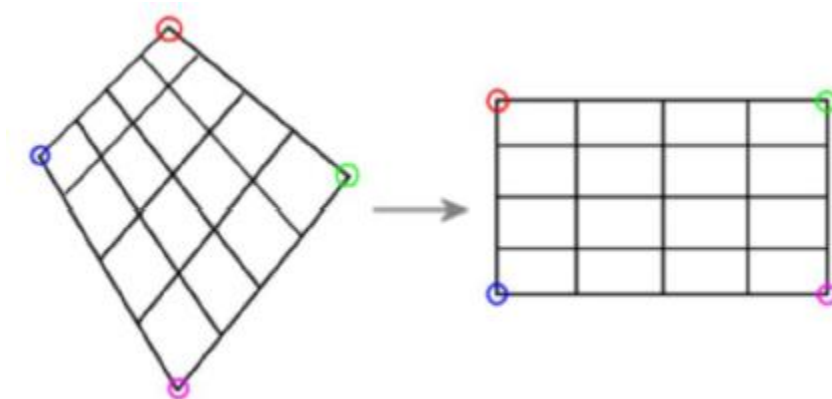
Trong đó:

s là hằng số tỉ lệ của phép chiếu và khác 0

X' là kết quả của phép ánh xạ

H là ma trận Homography, là một ma trận khả nghịch

Vì H là một ma trận khả nghịch, cho nên trong trường hợp muốn tái tạo ảnh X từ X', chỉ cần xác định được ma trận Homography là được.



Hình 1.19 - Phép chiếu Homography

#### 1.2.4.2. Tính toán Homography

Homography là một định nghĩa toán học. Đó là sự dịch chuyển sử dụng phép chiếu hình học, hay nói cách khác nó là sự kết hợp của cặp điểm trong phép chiếu phối cảnh. Ảnh thực trong không gian ba chiều có thể biến đổi về không gian ảnh bằng phép chiếu thông qua ma trận biến đổi Homography  $H$ . Các phép chiếu biến đổi thông qua ma trận Homography tuy không đảm bảo về kích thước và góc của vật được chiếu, nhưng đảm bảo về tỷ lệ.

Tính ma trận Homography bằng phương pháp Direct Linear Transform

Để tính ma trận Homography từ các cặp điểm tương ứng, người ta dùng phương pháp DLT (Direct Linear Transform), gồm có 2 bước: Đầu tiên, từ các cặp điểm tương ứng, ta chuyển về dạng ma trận  $A_h = 0$ . Sau đó, áp dụng phân rã SVD để tính ma trận  $H$ .

Giải thuật SVD hay được gọi là giải thuật phân rã giá trị đơn (Singular Value Decomposition) được Golub và Kahan công bố năm 1965, đó là một kỹ thuật phân rã được sử dụng để giảm hạng (hay số chiều) của ma trận.

SVD cho phép phân tích một ma trận phức tạp thành ba ma trận thành phần. Mục đích nhằm đưa ra việc giải quyết bài toán liên quan đến ma trận lớn, phức tạp về những bài toán nhỏ hơn.

$$A = USV^T$$

Trong đó:

$U$  là ma trận trực giao cấp  $m \times n$  ( $m$  số từ chỉ mục) các vector dòng của  $U$  là các vector từ chỉ mục.

$S$  là ma trận đường chéo cấp  $r \times r$  có giá trị suy biến  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$  với  $r = \text{rank}(A)$

$V$  là ma trận trực giao cấp  $r \times n$  – các vector cột của  $V$  là các vector văn bản

$$\begin{array}{c}
 \boxed{A_{m \times n}} = \boxed{U_{m \times m}} \times \boxed{\Sigma_{m \times n}} \times \boxed{V_{n \times n}^T} \\
 \text{(a) } (m < n)
 \end{array}$$

$$\begin{array}{c}
 \boxed{A_{m \times n}} = \boxed{U_{m \times m}} \times \boxed{\Sigma_{m \times n}} \times \boxed{V_{n \times n}^T} \\
 \text{(b) } (m > n)
 \end{array}$$

SVD cho ma trận A khi:  $m < n$  (hình trên), v  $m > n$  (hình dưới).  $\Sigma$  là một ma trận đường chéo với các phần tử trên đó giảm dần v không âm. Màu đỏ càng đậm thể hiện giá trị càng cao. Các ô màu trắng trên ma trận này thể hiện giá trị 0.

Hạng của ma trận A là các số dương trên đường chéo của ma trận S. Phương pháp DLT với các điểm nổi bật được tìm thấy từ thuật toán Harris:

Trong tọa độ không đồng nhất, ta có công thức:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Suy ra :

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cdot \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

Ta thấy rằng có hai phương trình độc lập tuyến tính, do đó lần lượt chia dòng thứ nhất của công thức trên cho dòng thứ ba và dòng thứ hai cho dòng thứ ba, ta có:

$$-h_1x - h_2y - h_3 + (h_7x + h_8y + h_9)u = 0$$

$$-h_4x - h_5y - h_6 + (h_7x + h_8y + h_9)v = 0$$

Viết dưới dạng ma trận ta có:

$$A_i h = \begin{bmatrix} -x & -y & -1 & 0 & 0 & 0 & ux & uy & u \\ 0 & 0 & 0 & -x & -y & -1 & vx & vy & v \end{bmatrix} (h_1 h_2 \dots h_9)^T$$

Áp dụng công thức phân rã SVD cho ma trận  $[A]$  ta có:

$$A = U \Sigma V^T = \sum_{i=1}^9 s_i u_i v_i^T$$

Với  $s_i$  là các giá trị đơn và được sắp xếp nhỏ dần, nên  $s_9$  là giá trị nhỏ nhất. Khi đó, giá trị của  $h_i$  bằng giá trị cuối cùng của cột  $v_i$ .

Trong thực tế các ảnh đầu vào có thể có gốc tọa độ ở góc trái của ảnh, cũng có thể gốc tọa độ nằm ở tâm ảnh. Nếu để tình trạng như vậy sẽ ảnh hưởng đến các kết quả biến đổi về sau như khi nhân ảnh với một hệ số hay các biến đổi tương tự affine. Các ảnh cần phải chuẩn hóa bằng phép biến đổi quay và dịch chuyển.

#### 1.2.4.3. Thuật toán RANSAC

**RANSAC (RANDOM SAMPLE CONSENSUS)** được công bố bởi Fischler và Bolles vào năm 1981. Ý tưởng chính của RANSAC như sau: Từ tập dữ liệu ban đầu, ta sẽ có hai loại dữ liệu nhiễu và không nhiễu (outlier và inlier), vì thế ta phải đi tính toán để tìm ra mô hình tốt nhất cho tập dữ liệu. Việc tính toán và chọn ra mô hình tốt nhất sẽ được lặp đi lặp lại  $k$  lần, với giá trị  $k$  được chọn sao cho đủ lớn để đảm bảo xác suất  $p$  (thường rơi vào giá trị 0.99) của tập dữ liệu mẫu ngẫu nhiên không chứa dữ liệu nhiễu.

Quá trình thực hiện thuật toán RANSAC được mô tả như dưới đây:

Từ tập dữ liệu đầu vào gồm có nhiễu và không có nhiễu ta chọn từ  $n$  dữ liệu ngẫu nhiên, tối thiểu để xây dựng mô hình:

- Tiến hành xây dựng mô hình với  $n$  dữ liệu đó, sau đó ra một ngưỡng dùng để kiểm chứng mô hình.

- Gọi tập dữ liệu ban đầu trừ đi tập  $n$  dữ liệu để xây dựng mô hình tập dữ liệu kiểm chứng. Sau đó, tiến hành kiểm chứng mô hình đã xây dựng bằng tập dữ liệu kiểm chứng. Nếu kết quả thu được từ mô hình vượt quá ngưỡng, thì điểm đó là nhiễu, còn không đó sẽ là ngược lại.

- Quá trình này sẽ được lặp lại trong  $k$  lần. Với  $k$  được tính theo công thức trên. Tại mỗi vòng lặp giá trị của  $k$  sẽ được tính lại.

- Kết quả là mô hình nào có số dữ liệu không nhiễu nhiều nhất sẽ được chọn là mô hình tốt nhất.

#### 1.2.4.4. Tính ma trận Homography bằng thuật toán RANSAC

Trong bài toán tạo ảnh Panorama, ma trận Homography được tính từ tập các cặp điểm nổi bật tương ứng của hai ảnh ban đầu đã được so sánh đối chiếu ở bước hai. Khi đó bốn cặp điểm nổi bật tương ứng không thẳng hàng, phương trình  $Ah = 0$  theo phương pháp DLT chuẩn hóa đã trình bày ở phần trên. Trong đó,  $A$  là ma trận có kích thước  $8 \times 9$ . Từ đó, ta xác định được ma trận  $h$ .

Với ma trận Homography được tính từ bốn cặp điểm ngẫu nhiên, ta có  $d$  là khoảng cách đo mức độ gần nhau của các cặp điểm đã được so sánh đối chiếu. Với cặp điểm nổi bật tương đồng  $(x, x')$  và  $d\left(\vec{a}, \vec{b}\right)$  là khoảng cách của hai vector, ta có công thức khoảng cách như sau:

$$d = d(\vec{x}, H\vec{x}') + d(\vec{x}', H\vec{x})$$

Thuật toán chi tiết:

Khởi tạo số vòng lặp  $k$ , ngưỡng distance,  $\max_{\text{inlier}}$ , và  $p = 0$ .

For ( $i = 1:k$ ), thực hiện các bước sau:

- Bước 1: Chọn 4 cặp điểm tương đồng ngẫu nhiên
- Bước 2: Kiểm tra xem các điểm có nằm trên cùng một đường thẳng hay không, nếu có, quay lại bước trên.

- Bước 3: Tính ma trận Homography  $H$  từ 4 điểm sử dụng phương pháp DLT chuẩn hóa.

- Bước 4: Tính khoảng cách  $d$  của các cặp điểm nổi bật tương đồng

$$d = d(\vec{x}, H_{curr}\vec{x}') + d(\vec{x}', H_{curr}\vec{x})$$

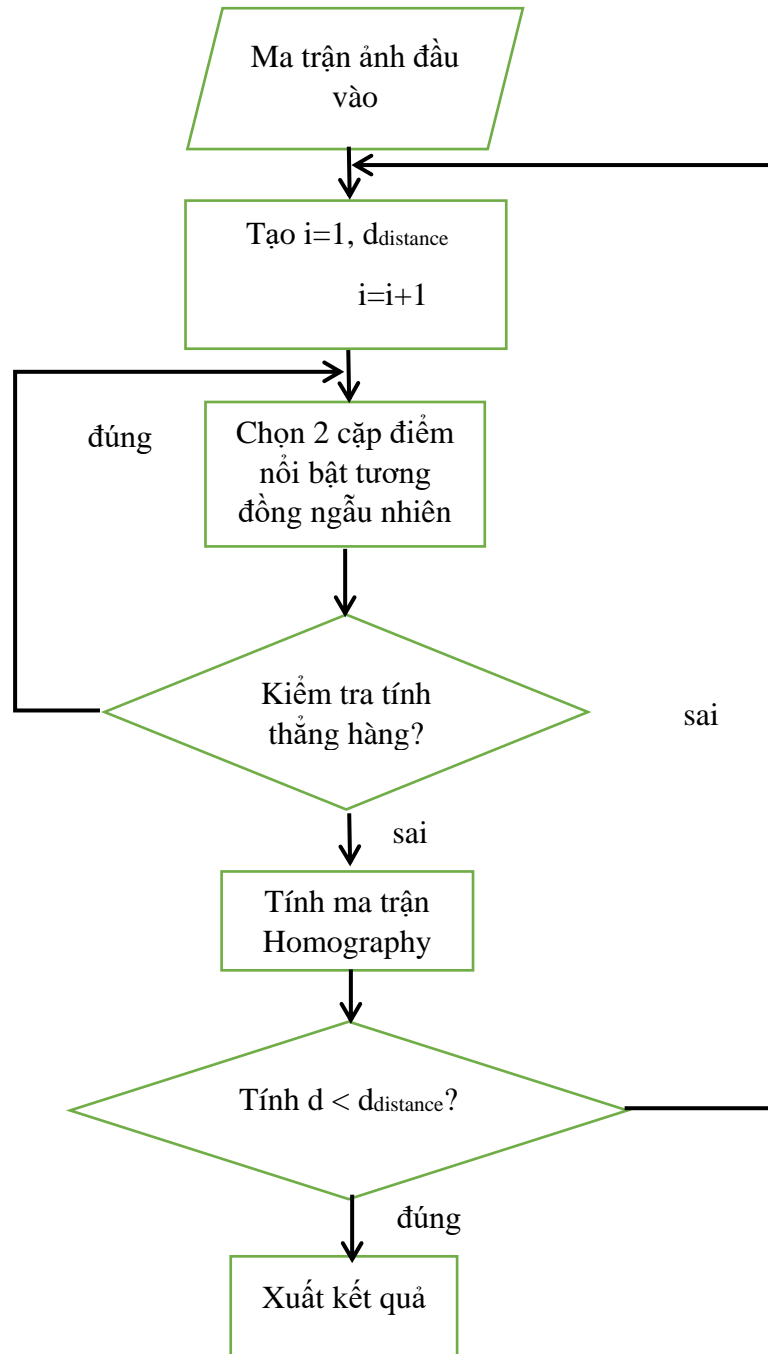
- Bước 5: Tính số lượng  $m$  các cặp điểm không ngẫu nhiên (inlier) thỏa điều kiện:  $d_i < \text{distance}$

- Bước 6: Nếu  $m > \max_{inlier}$  thì  $\max_{inlier} = m$ , ma trận Homography  $H = H_{curr}$

Tiếp tục tính lại ma trận  $H$  cho tất cả các cặp điểm tương đồng được coi là không nhiễu (inlier) bằng phương pháp DLT.

## CHƯƠNG 2 – CHƯƠNG TRÌNH MÔ PHỎNG

### 2.1. LƯU ĐỒ GIẢI THUẬT



Hình 2.1 – Lưu đồ giải thuật chương trình

### 2.2. CHƯƠNG TRÌNH THỰC HIỆN

#### Bước 1: Read Images and Warping



Đọc các file ảnh và chuyển đổi ảnh lần lượt theo toạ độ hình trụ



Hình 2.2 – Ảnh đầu vào và ảnh được chuyển đổi

**Bước 2: Feature Extraction**

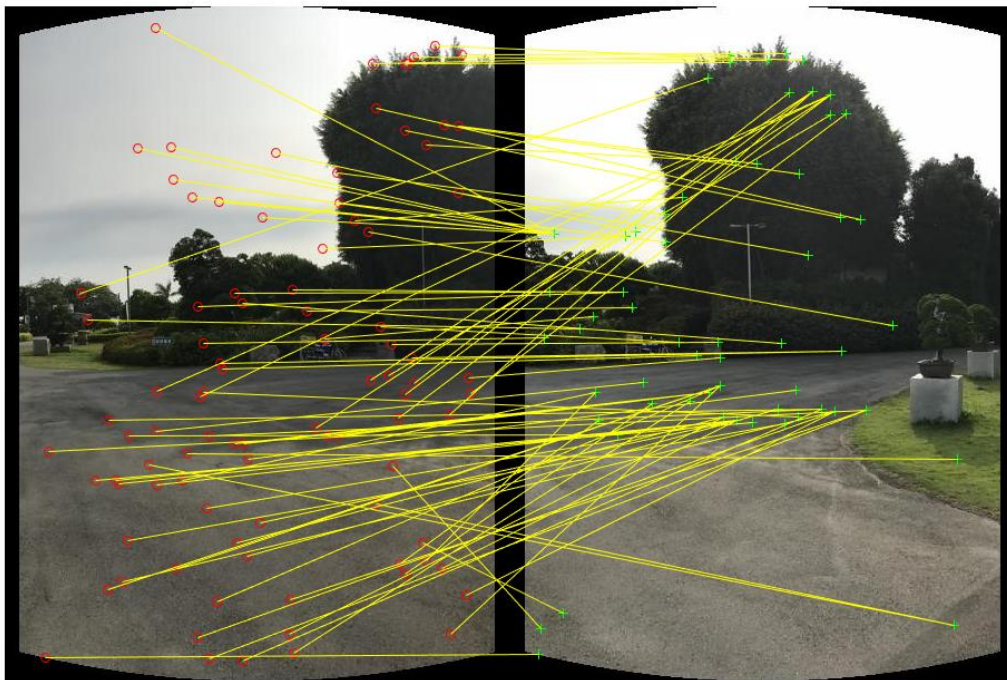
Trích xuất các điểm đặc trưng bất biến và mô tả của chúng và xác định các góc bằng Multi-Scale Harris.

**Bước 3: Feature Matching**

Tính toán khoảng cách Euclide giữa tất cả các feature trong hai hình ảnh.

Tìm 2 cặp có khoảng cách nhỏ nhất (bằng sorting), giả sử  $e1-nn$  và  $e2-nn$  là khoảng cách của hai cặp.

Xem cặp có khoảng cách nhỏ nhất là feature match nếu tỷ lệ  $e1-nn / e2-nn$  thấp hơn giá trị ngưỡng (threshold value).



Hình 2.3 – Các điểm matching giữa 2 ảnh

**Bước 4: Image Matching**

Kết quả trùng khớp được lưu trữ trong `match_pairs` với vị trí cả feature (hàng, col) trong cả hai hình ảnh.

**Bước 5: Combine Image and Blending**

Khâu hình ảnh từ trái sang phải để tạo thành 1 ảnh hoàn chỉnh.

Sử dụng linear blending trên các phần chồng chéo để tạo sự tương đồng về màu sắc.



Hình 2.4 – Ảnh sâu khi được khâu từ các ảnh khác

### **Bước 6: Cropping and Result**

Cắt các vùng không đều của ảnh ta được kết quả cuối cùng.



Hình 2.5 – Ảnh kết quả cuối cùng

## CHƯƠNG 3 – KẾT LUẬN

Những điểm đã đạt được:

- Tìm hiểu nguyên lý của kỹ thuật đối sánh ảnh.
- Tìm hiểu phương pháp tìm và trích chọn các điểm keypoint và biểu diễn đặc trưng SIFT.
- Hiểu được giải thuật RANSAC.
- Tìm hiểu quy trình ghép ảnh panorama dựa trên đối sánh đặc trưng.
- Viết được chương trình thử nghiệm ghép ảnh panorama.

Những mặt còn hạn chế:

- Trong trích chọn đặc trưng, mới chỉ tìm hiểu thuật toán Harris, chưa có sự so sánh với các thuật toán cũng như các phương pháp đối sánh khác.
- Thuật toán còn rườm rà và phải thử nghiệm nhiều lần để lấy được các thông số tốt nhất.
- Do phát triển trên môi trường MATLAB nên giao diện của chương trình không được đẹp, tốc độ xử lý chậm.

## **TÀI LIỆU THAM KHẢO**

- [1] Multi-Image Matching using Multi-Scale Oriented Patches, by Brown et al., 2005.
- [2] Random Sample Consensus: A Paradigm for Model Fitting with Applications Analysis and Automated Cartography, by Fischler et al., 1981.
- [3] Brown, Matthew; Lowe, David G. “Automatic Panoramic Image Stitching using Invariant Features”, University of British Columbia, Vancouver, Canada, 2007.
- [4] Szielski, Richard “Image Alignment and Stitching: A Tutorial”, Technical Report, December 2006.
- [5] Knight, Denis “The Perfect Panorama”, October 2008.
- [6] P. J. Rousseeuw and A. M. Leroy, Robust regression and outlier detection, Wiley, 1987.
- [7] S. Boyd and L. Vandenberghe, Convex optimization, sixth ed., Cambridge University Press, 2008.
- [8] O. Chum and J. Matas, Randomized RANSAC with Td,d test, 13th British Machine Vision Conference, September 2002.