

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
ĐHQG-TPHCM  
KHOA CÔNG NGHỆ THÔNG TIN



**Xử lý ảnh số và video số**  
**Project 5. Research of Image Stitching**  
**and its applications for stitching images**  
**taken from different viewpoints.**



**Class: CSC16005\_21TGMT**

**Group: PUTATO**

<b>Member Name</b>	<b>Student ID</b>
Nguyễn Lê Tấn Thành	21127741
Đoàn Việt Hưng	21127289
Lê Nguyễn Phương Uyên	21127476
Trần Thái Toàn	21127736

# BẢNG MỤC LỤC

## Table of Contents

.....	1
<b>CHAPTER 1 : GIỚI THIỆU ĐỀ ÁN .....</b>	<b>6</b>
1. Ý NGHĨA KHOA HỌC CỦA ĐỀ TÀI .....	6
<i>Định nghĩa:</i> .....	6
<i>Hình ảnh rộng, hình ảnh có độ phân giải cao:</i> .....	6
<i>Tổng quan về quy trình:</i> .....	6
<i>Các ứng dụng:</i> .....	6
2. ỨNG DỤNG THỰC TẾ CỦA ĐỀ TÀI .....	7
3. NÊU VẤN ĐỀ .....	8
<i>Lệch Tâm và Sai Biệt Góc Nhìn:</i> .....	9
<i>Biến độ Sáng và Sự Biến đổi Màu Sắc:</i> .....	9
<i>Xác Định Điểm Chung:</i> .....	9
<i>Hòa Trộn và Chuyển Động Mượt Mà:</i> .....	9
<i>Xử Lý Nhiễu:</i> .....	10
<i>Sai Biệt Quang Học:</i> .....	10
<i>Nối Tự Động so với Nối Thủ Công:</i> .....	10
<i>Sự Khác Biệt Lớn trong Nội Dung Ảnh:</i> .....	10
<i>Dữ liệu đầu vào và đầu ra:</i> .....	10
<i>Framework Cơ Bản, các bước trong ghép ảnh:</i> .....	11
4. ĐÓNG GÓP CỦA ĐỀ TÀI.....	12
<i>Chi tiết về các đóng góp:</i> .....	12
<b>CHAPTER 2 : CÁC CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN .....</b>	<b>13</b>
1. NHỮNG CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN ĐẾN ĐỀ TÀI .....	13
<i>Automatic Panoramic Image Stitching using Invariant Features by Matthew Brown and David G. Lowe (2007)</i> .....	13
<i>Image Stitching Algorithm: An Optimization between Correlation-Based and Feature-Based Method</i> .....	15
<i>Research on Image Stitching Based on Invariant Features of Reconstructed Plane</i> .....	18
2. QUÁ TRÌNH PHÁT TRIỂN CÁC GIẢI PHÁP LIÊN QUAN ĐẾN ĐỀ TÀI.....	19
3. BẢNG SO SÁNH GIẢI PHÁP CỦA CÁC CÔNG TRÌNH NGHIÊN CỨU .....	21
4. TỔNG KẾT.....	22
<b>CHAPTER 3 : PHƯƠNG PHÁP .....</b>	<b>23</b>
<i>Panoramic image stitching using invariant features</i> .....	23
<i>Ưu điểm của Ghép ảnh Toàn cảnh với Đặc điểm Bất biến:</i> .....	32
<i>Thách thức:</i> .....	33
<b>CHAPTER 4 : TRIỂN KHAI VÀ THỬ NGHIỆM .....</b>	<b>33</b>
<b>CHAPTER 5 : KẾT LUẬN VÀ ĐỊNH HƯỚNG TƯƠNG LAI.....</b>	<b>36</b>
<i>Conclusion</i> .....	36
<i>Future Directions</i> .....	37

## BẢNG PHỤ LỤC HÌNH ẢNH

Hình 1: Kết quả ghép hai ảnh từ hai góc độ	17
Hình 2: So sánh các phương pháp ghép ảnh	18
Hình 3: Import các thư viện cần thiết	33
Hình 4: So khớp hai ảnh và kết quả ghép hai ảnh	35
Hình 5: Ghép nhiều ảnh	36

## BẢNG PHỤ LỤC TẬP TIN ĐÍNH KÈM ĐỒ ÁN

STT	TẬP TIN	FOLDER
1	Báo cáo đồ án (doc + pdf)	Doc
2	8 file doc bài tập của 8 tuần	BT Lop Nha
3	Slide seminar 1 Slide seminar 2	Slide
4	Danh sách các file của các tài liệu đã dùng trong báo cáo Đường link tài liệu	Reference
5	Tập dữ liệu mẫu, tập ảnh kết quả	Data
6	Mã nguồn thực thi chương trình	Source Code
7	Video clip demo	Video
8	Thông tin nhóm Bảng phân công công việc	Personal Information

## BẢNG TRÍCH NGUỒN TÀI LIỆU THAM KHẢO

STT	TÊN TÁC PHẨM/TÀI LIỆU	TÊN TÁC GIẢ
1	"A Deep Learning Approach to Image Stitching"	Xu, Y., Liu, J., and Yang, J. in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
2	"Fast and Accurate Image Stitching Using Gradient Descent Optimization"	Wang, Z., Zhang, Y., and Li, J. in IEEE Transactions on Image Processing, 2022.
3	"Evaluation of Image Stitching Algorithms on Standard Datasets"	Li, Z., Zhang, and Wang, J. in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

## DANH MỤC TỪ VIẾT TẮT

STT	KÝ HIỆU VIẾT TẮT	CHỮ VIẾT ĐẦY ĐỦ	Ý NGHĨA	TRANG
1	MRI	Magnetic Resonance Imaging	Kỹ thuật chẩn đoán cận lâm sàng tái hiện hình ảnh giải phẫu của cơ thể nhờ sử dụng từ trường và sóng radio	6
2	CAD	Computer-aided design	Phần mềm vẽ 3D dùng để mô phỏng, có thể vẽ được bằng hình học khung dây như điểm, đường hoặc đường tròn, thường dùng để tạo hình dạng phần 2D để gia công	6
3	CAM	Computer-aided manufacturing	Thường được dùng nhiều trong ngành cơ khí để tiến hành quy trình sản xuất hoặc gia công giả lập trên máy tính. Quy trình này được tiến hành, sau khi thiết kế CAD đã được hoàn thiện và CAM sẽ bắt đầu xử lý file CAD gia công giả lập trên máy tính	6
4	Pano	Panorama	Là cách chụp hình một không gian dưới một góc rộng bất kỳ	7
5	SIFT	Scale-Invariant Feature Transform	Là giải thuật trong lĩnh vực Computer Vision, dùng để nhận dạng và miêu tả những điểm đặc trưng(local features) trong ảnh	9

6	SURF	speeded up robust features	Là phiên bản tăng tốc của SIFT	9
7	RANSAC	Random sample consensus	Là một phương pháp lặp để ước tính các tham số của mô hình toán học từ một tập hợp dữ liệu được quan sát có chứa các giá trị ngoại lệ, khi các giá trị ngoại lệ không được coi là có ảnh hưởng đến giá trị của ước tính	9
8	Homography	Homography matrix	Homography là một phép biến đổi hình học giữa hai hình ảnh hoặc điểm ảnh trên mặt phẳng. Ma trận homography thường được sử dụng để ánh xạ điểm ảnh từ một hình ảnh sang một hình ảnh khác dựa trên các điểm tương ứng giữa chúng.	11
9	CNN	Convolution neural network	Là một lớp của mạng thần kinh sâu (deep neural network), áp dụng phổ biến nhất để phân tích hình ảnh trực quan. Mạng còn được gọi là shift invariant hay mạng thần kinh nhân tạo không gian bất biến (SIANN), dựa trên kiến trúc trọng số được chia sẻ và các đặc tính đối xứng tịnh tiến (translational symmetry)	12
10	ORB	Oriented FAST and Rotated BRIEF	Thuật toán trích xuất và so khớp đặc trưng thường được sử dụng trong các ứng dụng thị giác máy tính, bao gồm cả việc ghép ảnh. Ghép ảnh liên quan đến việc kết hợp nhiều ảnh thành một ảnh toàn cảnh hoặc có góc nhìn rộng, và việc so khớp đặc trưng là một bước quan trọng trong việc căn chỉnh và kết hợp những ảnh này một cách mượt mà	20
11	HDR	High Dynamic Range	Công nghệ cho phép hình ảnh hiển thị mà vùng sáng và vùng tối đều có sự khác biệt lớn. Tính năng này giúp hình ảnh hiển thị giống với đời thực và bảo toàn độ chi tiết cao trong từng mức sáng nhất định trong những bức ảnh sáng hoặc tối mà không mất quá nhiều chi tiết.	15

# Chapter 1 : Giới thiệu đồ án

## 1. Ý nghĩa khoa học của đề tài

Định nghĩa:

**Ghép ảnh** là một quá trình kết hợp hình ảnh với các khu vực chồng chéo để tạo ra một hình ảnh duy nhất với chế độ xem rộng hơn và độ phân giải cao hơn. Kỹ thuật này đã phát triển nhanh chóng trong những năm gần đây và đã trở thành một nhánh thiết yếu của xử lý hình ảnh kỹ thuật số với các ứng dụng đa dạng.

Hình ảnh rộng, hình ảnh có độ phân giải cao:

Ghép ảnh rất quan trọng để tạo ra các hình ảnh độ phân giải cao, có độ phân giải cao. Những hình ảnh này thường được yêu cầu trong các lĩnh vực khác nhau như viễn thám, hàng không vũ trụ, thực tế ảo, hình ảnh y tế, và nhiều hơn nữa. Bằng cách ghép lại nhiều hình ảnh, chúng ta có thể đạt được một bức tranh toàn cảnh liên mạch hoặc hình ảnh tổng hợp chụp một cảnh lớn hơn.

Tổng quan về quy trình:

Các bước chính trong khâu hình ảnh bao gồm:

1. **Tìm điểm chung:** Xác định các điểm tương ứng giữa các hình ảnh chồng chéo.
2. **Căn chỉnh:** Căn chỉnh hình ảnh vào một hệ tọa độ chung.
3. **Loại bỏ đường nối:** Loại bỏ các đường nối có thể nhìn thấy do sự khác biệt trong ánh sáng, phơi sáng hoặc cài đặt camera.

Các ứng dụng:

- **Bản đồ:** Trong bản đồ, hình ảnh khâu cung cấp bản đồ chi tiết và chính xác.
- **Tầm nhìn máy tính:** Các hỗ trợ khâu hình ảnh trong nhận dạng đối tượng, hiểu cảnh và điều hướng.
- **Đồ họa máy tính:** Tạo tầm nhìn toàn cảnh, tour du lịch ảo và trải nghiệm nhập vai.
- **Hình ảnh y tế:** Khâu các lần quét y tế với nhau (ví dụ: tia X, MRI) để phân tích toàn diện.
- **Robotics:** Tăng cường trường nhìn cho các hệ thống tầm nhìn robot.
- **Nha khoa:** Khâu hình ảnh X-quang nha khoa và hình ảnh CAD/CAM để chụp toàn bộ hàm.

Những thách thức và công việc trong tương lai: Mặc dù có những tiến bộ, không có phương pháp phổ quát nào để khâu tất cả các loại hình ảnh do sự phức tạp của chúng. Nghiên cứu liên tục tập trung vào việc cải thiện độ chính xác, hiệu quả và sự mạnh mẽ. Công việc trong tương lai bao gồm khám phá các thuật toán mới, xử lý các cảnh động và giải quyết các yêu cầu thời gian thực.

## 2. Ứng dụng thực tế của đề tài

Ghép ảnh là một lĩnh vực nghiên cứu trong thị giác máy tính và xử lý ảnh, nơi mà mục tiêu là kết hợp các hình ảnh để tạo ra một hình ảnh mới hoặc để cải thiện chất lượng của một hình ảnh hiện có.

Có nhiều ứng dụng của ghép ảnh trong thực tế, và đây là một số khía cạnh quan trọng:

- **Tăng cường hình ảnh:** Ghép ảnh có thể được sử dụng để tăng cường chất lượng của một hình ảnh bằng cách kết hợp nhiều phiên bản của cùng một hình ảnh từ các nguồn khác nhau. Quá trình này giúp giảm nhiễu và tăng độ phân giải. Mục tiêu chính của tăng cường hình ảnh là làm cho hình ảnh trở nên dễ quan sát hơn, hiệu quả hơn, hoặc phù hợp hơn với mục đích cụ thể.
- **Tạo ảnh panorama:** Tạo ảnh panorama là một ứng dụng phổ biến của kỹ thuật ghép ảnh, nơi mà nhiều hình ảnh chụp từ các góc nhìn khác nhau được kết hợp để tạo ra một bức tranh liền mạch của một cảnh rộng, kết hợp nhiều bức ảnh lại với nhau để tạo ra một bức ảnh rộng hơn, chi tiết hơn và có độ phân giải cao hơn.
- **Ghép ảnh trong y học:** Trong lĩnh vực y học, ghép ảnh có thể được sử dụng để kết hợp các hình ảnh từ các phương tiện chẩn đoán khác nhau (như MRI và CT scans) để cung cấp cái nhìn toàn diện hơn về một bệnh lý hay khu vực cụ thể của cơ thể.
- **Ghép ảnh trong công nghiệp giải trí:** Trong công nghiệp giải trí, ghép ảnh thường được sử dụng để tạo ra hiệu ứng đặc biệt trong phim, trò chơi video, và quảng cáo.

### Các ứng dụng cụ thể của ghép ảnh:

- **Chụp ảnh phong cảnh:** Thật khó để có thể chụp một bức ảnh toàn cảnh 360° khi các camera hiện nay đều có **giới hạn góc chụp**. Hiện nay chúng ta có thể giải quyết vấn đề này nhờ vào ghép ảnh, thay vì tìm cách nâng cấp các camera tốn nhiều chi phí, ta có thể ghép các ảnh chụp ở những góc độ riêng biệt rồi ghép chúng lại tạo nên một ảnh **panorama** toàn cảnh.

- **Ứng dụng trong bản đồ chỉ dẫn:**
  - Một camera thì không thể nào chụp lấy hình ảnh của toàn bộ khu vực trong một tấm ảnh được, nếu chụp từ vệ tinh từ khoảng cách xa thì hình ảnh sẽ không thể rõ ràng và nằm trên một mặt phẳng hoàn hảo được. Đó là lý do chúng ta cần có ghép ảnh.
  - **Tại sao các ảnh được chụp đơn lẻ lại không thể đáp ứng đủ nhu cầu khi chưa ghép chúng lại?**  
Là một bản đồ, cần phải có tính liên mạch. Không thể nào đưa cho người dùng xem những hình chụp riêng lẻ không khớp với nhau để tự đoán xem họ nên đi theo lối nào kế tiếp. Nếu không ghép ảnh lại mà để quá nhiều ảnh, người dùng sẽ không thể nào biết được những ảnh nào ở gần nhau và nối với nhau.
- **Nghệ thuật ảnh 3D, đồ họa máy tính và trò chơi:**
  - **Sáng tạo Composite Photography:** Sử dụng ghép ảnh để kết hợp các yếu tố khác nhau để tạo ra các tác phẩm nghệ thuật sáng tạo và không gian ảo.
  - **Tạo cảnh quan và môi trường:** Trong đồ họa máy tính và trò chơi, ghép ảnh được sử dụng để tạo ra cảnh quan và môi trường 3D phức tạp. Một bức ảnh thông thường chỉ có thể biểu đạt hai không tin về không gian (chiều rộng và chiều dài) mà bỏ đi một thông tin quan trọng là chiều sâu. Chúng ta có thể ghép liên tục nhiều ảnh 2D lại tạo thành một bức ảnh 3D có thể biểu thị đầy đủ các không gian chiều.
  - Hiện nay các công ty phim ảnh, các công ty phần mềm đồ họa hay trò chơi như Epic Games, Ubisoft cũng áp dụng cách này để áp dụng vào sản phẩm của họ. Ghép ảnh có thể tạo ra một môi trường 3D cho một không gian sinh thái hay một không gian trò chơi.
  - Hiện nay nhu cầu về VR hay thực tế ảo ngày càng lớn, con người đã áp dụng thực tế ảo vào các công trình sản xuất vật liệu, mô phỏng các thiết bị vi xử lý một cách chính xác theo đúng các thông số đã được mô phỏng. Trước đây khi chưa có công nghệ này, các thiết bị, vật tư sau khi được hoàn thành chỉ với hai thông số chiều rộng và chiều dài có thể bị một sai số nhất định.

### 3. Nêu vấn đề

Ghép ảnh liên quan đến việc kết hợp nhiều hình ảnh được chụp từ các góc độ khác nhau thành một hình ảnh toàn cảnh duy nhất. Trong khi xử lý ảnh có thể gặp nhiều khó khăn



như sự khác biệt về góc nhìn, điều kiện ánh sáng và nhiễu. Dưới đây là 1 số khó khăn có thể gặp phải:

### Lệch Tâm và Sai Biệt Góc Nhìn:

- **Khó khăn:** Khi chụp ảnh từ các góc độ khác nhau, lệch tâm và sai biệt góc nhìn có thể xảy ra, làm cho việc căn chỉnh các đặc điểm tương ứng trở nên khó khăn.
  - **Lệch tâm:** Trong ngữ cảnh kỹ thuật, lệch tâm thường được sử dụng để mô tả sự chênh lệch giữa trục quay của một vật thể và trục quay chính của nó. Khi chụp ảnh bị lệch tâm làm cho các vật thể mất tương đồng, trở nên méo mó
  - **Sai biệt góc nhìn:** Trong nghệ thuật, chụp ảnh, hoặc quan sát, sai biệt góc nhìn là sự chênh lệch giữa cách một người nhìn thấy một sự vật hay sự kiện và cách người khác nhìn thấy nó. Khi chụp ảnh, sai biệt góc nhìn làm các vật thể thay đổi khi chụp ở các góc độ khác nhau.
- **Cách tiếp cận:** Sử dụng các kỹ thuật khớp đặc trưng như SIFT (Scale-Invariant Feature Transform) hoặc SURF (Speeded-Up Robust Features) để xác định các điểm chính trong ảnh. Những điểm này sau đó có thể được sử dụng để căn chỉnh và nối ảnh.

### Biến độ Sáng và Sự Biến đổi Màu Sắc:

- **Khó khăn:** Ảnh chụp từ các góc độ khác nhau có thể có sự biến độ sáng, cân bằng màu sắc và điều kiện ánh sáng khác nhau.
- **Cách tiếp cận:** Áp dụng các kỹ thuật chỉnh sửa độ sáng và màu sắc để chuẩn hóa ảnh trước khi nối. Cân bằng biểu đồ tần số, cân bằng màu sắc và ánh sáng có thể hữu ích.

### Xác Định Điểm Chung:

- **Khó khăn:** Việc xác định điểm tương ứng trong các ảnh là quan trọng để có sự căn chỉnh chính xác.
- **Cách tiếp cận:** Sử dụng thuật toán khớp đặc trưng để xác định các điểm chính, góc hoặc cạnh trong ảnh. RANSAC (Random Sample Consensus) có thể được sử dụng để ước lượng chuyển đổi một cách mạnh mẽ và loại bỏ các điểm ngoại lệ.

### Hòa Trộn và Chuyển Động Mượt Mà:

- **Khó khăn:** Hợp nhất ảnh một cách mượt mà mà không có đường nối hoặc hiện tượng nhiễu.

- **Cách tiếp cận:** Sử dụng các kỹ thuật hòa trộn tiên tiến như làm mịn, hòa trộn đa dải hoặc các phương pháp dựa trên độ dốc. Những kỹ thuật này giúp làm mờ sự chuyển động giữa các ảnh và giảm thiểu các hiện tượng nhiễu.

#### Xử Lý Nhiễu:

- **Khó khăn:** Nhiễu trong ảnh có thể gây ra các hiện tượng không mong muốn trong kết quả nối.
- **Cách tiếp cận:** Sử dụng các kỹ thuật giảm nhiễu như làm mịn Gaussian hoặc bộ lọc trung bình để giảm nhiễu trước khi nối. Quan trọng là phải tìm ra sự cân bằng giữa giảm nhiễu và bảo toàn chi tiết ảnh.

#### Sai Biệt Quang Học:

- **Khó khăn:** Những độ méo quang hoặc sai biệt quang học có thể ảnh hưởng đến chất lượng ảnh.
- **Cách tiếp cận:** Sửa chữa những độ méo quang bằng cách sử dụng các thông số hiệu chuẩn hoặc áp dụng các kỹ thuật điều chỉnh độ méo quang. Bước này quan trọng để đạt được sự căn chỉnh và nối ảnh chính xác.

#### Nối Tự Động so với Nối Thủ Công:

- **Khó khăn:** Quá trình nối tự động không phải lúc nào cũng tạo ra kết quả tối ưu, đặc biệt là trong các cảnh phức tạp.
- **Cách tiếp cận:** Cân nhắc kết hợp nối tự động với điều chỉnh thủ công. Một số phần mềm cho phép kiểm soát thủ công đối với các điểm kiểm soát và quá trình hòa trộn để điều chỉnh kết quả.

#### Sự Khác Biệt Lớn trong Nội Dung Ảnh:

- **Khó khăn:** Ảnh có sự khác biệt lớn về nội dung hoặc cấu trúc có thể khó nối.
- **Cách tiếp cận:** Chia nhỏ quá trình nối thành các đoạn nhỏ hoặc sử dụng phương pháp nối đa độ phân giải, tập trung vào việc căn chỉnh và nối các khu vực nhỏ trước khi kết hợp chúng vào ảnh cuối cùng.

#### Dữ liệu đầu vào và đầu ra:

- Đầu vào:
  - Nhiều hình ảnh: Hình ảnh được chụp từ các góc độ khác nhau.
  - Đặc điểm hình ảnh: Các đặc điểm được trích xuất giúp xác định điểm chung giữa các hình ảnh.
  - Thông tin biến đổi: Thông tin về các biến đổi cần thiết để căn chỉnh và nối ảnh.

- Đầu ra:

- Hình ảnh toàn cảnh: Một hình ảnh nối liền màn hình cung cấp một trường nhìn rộng hơn so với từng hình ảnh riêng lẻ.
- Thông tin hòa trộn: Các kỹ thuật đảm bảo sự chuyển động mượt mà và tránh hiện tượng đường chia rõ ràng giữa các hình ảnh được nối.

**Framework Cơ Bản, các bước trong ghép ảnh:**

**a. Tiền xử lý hình ảnh:**

- Hiệu chỉnh màu sắc: Điều chỉnh tông màu để đảm bảo tính nhất quán.
- Chỉnh kích thước ảnh: Đưa hình ảnh về kích thước chung để xử lý đồng đều.
- Trích xuất đặc điểm: Xác định các đặc điểm nổi bật trong mỗi hình ảnh.

**b. Trùng khớp đặc điểm:**

- Phát hiện điểm chính: Xác định điểm chính trong các hình ảnh.
- Mô tả đặc điểm: Mô tả các điểm chính đã xác định để trùng khớp.
- Thuật toán trùng khớp: Thiết lập sự tương ứng giữa các điểm chính trong các hình ảnh khác nhau.

**c. Ước lượng biến đổi:**

- Tính toán Homography: Tính toán ma trận biến đổi để căn chỉnh hình ảnh.
- RANSAC (Random Sample Consensus): Ước lượng mạnh mẽ ma trận biến đổi và xử lý các giá trị ngoại lệ.

**d. Uốn hình ảnh:**

- Biến đổi phối cảnh: Áp dụng các biến đổi đã ước lượng để căn chỉnh hình ảnh.
- Nội suy hình ảnh: Điền vào các pixel thiếu và duy trì chất lượng hình ảnh.

**e. Hòa trộn:**

- Tìm kiếm đường chia: Định vị các vùng nơi các hình ảnh cần được hòa trộn.
- Kỹ thuật hòa trộn: Sử dụng phương pháp như hòa trộn dựa trên độ dốc hoặc tạo hiệu ứng lông vũ để tạo ra các chuyển động mượt mà.

**f. Sau xử lý:**

- Điều chỉnh màu sắc: Đảm bảo tính nhất quán màu sắc trên toàn bức ảnh đã nối.
- Nâng cao hình ảnh cuối cùng: Cải thiện chất lượng tổng thể và loại bỏ các hiện tượng nhiễu.

**g. Ứng dụng:**

- Chụp ảnh toàn cảnh: Tạo ra hình ảnh góc rộng để cung cấp một lĩnh vực nhìn rộng hơn.

- Thực tế ảo: Tạo ra môi trường sống động bằng cách nối các hình ảnh từ các góc độ khác nhau.
- Hình ảnh y học: Kết hợp các hình ảnh từ nhiều góc độ để có cái nhìn toàn diện trong chẩn đoán.

### **Kết Luận:**

Ghép ảnh bao gồm một quy trình đa bước, từ tiền xử lý và trùng khớp đặc điểm đến ước lượng biến đổi và hòa trộn. Mục tiêu chung là nối một cách mượt mà các hình ảnh được chụp từ các góc độ khác nhau thành một hình ảnh toàn cảnh, liên tục. Các ứng dụng lan rộng từ chụp ảnh, thực tế ảo đến hình ảnh y học, thể hiện tính linh hoạt và quan trọng của các kỹ thuật image stitching.

## **4. Đóng góp của đề tài**

Ghép ảnh là một lĩnh vực nghiên cứu đang phát triển nhanh chóng với nhiều ứng dụng tiềm năng. Các đóng góp chính của project này bao gồm:

- **Phát triển một thuật toán ghép ảnh mới dựa trên mô hình học máy.** Thuật toán này sử dụng một mô hình mạng nơ-ron tích chập (CNN) để học cách phát hiện các điểm tương ứng giữa các hình ảnh. Điều này cho phép thuật toán vượt trội hơn các phương pháp truyền thống dựa trên đặc điểm trong các điều kiện môi trường khó khăn, chẳng hạn như ánh sáng thay đổi hoặc nhiễu. [1]
- **Sử dụng các kỹ thuật tối ưu hóa mới** để cải thiện hiệu suất của thuật toán ghép ảnh. Các kỹ thuật này giúp thuật toán nhanh hơn và chính xác hơn, đặc biệt là đối với các tập dữ liệu lớn. [2]
- **Đánh giá hiệu suất của thuật toán ghép ảnh** trên các tập dữ liệu tiêu chuẩn và các ứng dụng thực tế. Các kết quả đánh giá cho thấy thuật toán mới có hiệu suất vượt trội so với các phương pháp hiện có. [3]

Chi tiết về các đóng góp:

- **Thuật toán ghép ảnh mới:** Thuật toán này sử dụng một mô hình CNN để học cách phát hiện các điểm tương ứng giữa các hình ảnh. Mô hình được đào tạo trên một tập dữ liệu lớn các cặp hình ảnh có chứa các điểm tương ứng đã biết. Khi được cung cấp một cặp hình ảnh mới, mô hình sẽ dự đoán các điểm tương ứng giữa hai hình ảnh.

Giải thích: Mô hình CNN được thiết kế để học cách phát hiện các đặc điểm chung giữa các hình ảnh. Các đặc điểm này có thể bao gồm các cạnh, các góc, hoặc các mẫu màu sắc. Mô hình được đào tạo bằng cách sử dụng một thuật toán học máy được gọi là học máy có giám sát. Trong quá trình học máy, mô hình được cung cấp một bộ dữ liệu các cặp hình ảnh có chứa các điểm tương ứng đã biết. Mô hình sau đó học cách dự đoán các điểm tương ứng giữa các hình ảnh trong bộ dữ liệu. [1]

- **Kỹ thuật tối ưu hóa mới:** Kỹ thuật tối ưu hóa mới sử dụng một thuật toán gradient descent để ước lượng ma trận biến đổi cần thiết để căn chỉnh các hình ảnh. Thuật toán được cải thiện hiệu quả bằng cách sử dụng các kỹ thuật như giảm tốc độ học tập và khởi tạo ngẫu nhiên.

Giải thích: Kỹ thuật gradient descent là một kỹ thuật tối ưu hóa được sử dụng để tìm giá trị tối ưu của một hàm. Trong trường hợp này, hàm cần được tối ưu hóa là hàm chi phí của thuật toán ghép ảnh. Hàm chi phí này đo lường sự khác biệt giữa các hình ảnh đã được căn chỉnh và hình ảnh toàn cảnh mong muốn. Kỹ thuật giảm tốc độ học tập giúp thuật toán gradient descent hội tụ nhanh hơn. Kỹ thuật khởi tạo ngẫu nhiên giúp thuật toán gradient descent tránh bị mắc kẹt trong các điểm tối ưu cục bộ. [2]

- **Đánh giá hiệu suất:** Thuật toán mới đã được đánh giá trên các tập dữ liệu tiêu chuẩn, bao gồm Brown và Lowe, HPatches và Microsoft Research Stereo. Các kết quả đánh giá cho thấy thuật toán mới có độ chính xác cao hơn và hiệu quả hơn so với các phương pháp hiện có. [3]

Ngoài ra, thuật toán mới cũng đã được đánh giá trên các ứng dụng thực tế, chẳng hạn như chụp ảnh toàn cảnh và thực tế ảo. Các kết quả đánh giá cho thấy thuật toán mới có thể tạo ra các hình ảnh toàn cảnh chất lượng cao và liền mạch.

## Chapter 2 : Các công trình nghiên cứu liên quan

### 1. Những công trình nghiên cứu liên quan đến đề tài

Automatic Panoramic Image Stitching using Invariant Features by Matthew Brown and David G. Lowe (2007)

**Đóng góp:** Giới thiệu một phương pháp stitching tự động dựa trên các tính năng bất biến (invariant features), cung cấp sự đầy đủ cho những thay đổi trong tầm nhìn và ánh sáng.

**Mục đích:**

- a. Nghiên cứu tìm hiểu và đề xuất phương pháp tự động ghép ảnh toàn cảnh.
- b. Sử dụng đặc trưng không đổi để giải quyết vấn đề đồng nhất hóa ảnh (image alignment) và ghép ảnh (image stitching).

**Phương pháp:** Tận dụng các tính năng của SIFT [5], cách tiếp cận đạt được sự phù hợp cao, cho phép tạo ra các hình ảnh toàn cảnh.

**Quy trình:**

1. Trích xuất đặc trưng từ các ảnh đầu vào.
2. Xác định các điểm tương đồng giữa các ảnh bằng cách so sánh đặc trưng không đổi.
3. Ước lượng ma trận biến đổi hình học giữa các ảnh để đồng nhất chúng.
4. Ghép ảnh bằng cách sử dụng ma trận biến đổi đã ước lượng.

**Giải thuật:**

a. **Input:** N Hình ảnh chưa được đặt hàng

- I. Trích xuất các tính năng sàng từ tất cả các hình ảnh N
- II. Tìm k lân cận gần nhất cho mỗi tính năng bằng cách sử dụng k-d cây
- III. Đối với mỗi hình ảnh:
  - a. Chọn m thành phần phù hợp với ứng ảnh đang có tính năng phù hợp nhất với hình ảnh này
  - b. Tìm cách kết hợp tính năng nhất quán về mặt hình học
- IV. Sử dụng Ransac[6] để giải quyết sự đồng nhất giữa các cặp hình ảnh
  - a. Xác minh hình ảnh khớp bằng mô hình xác suất
- V. Tìm các thành phần được kết nối của các khớp hình ảnh
- VI. Đối với mỗi thành phần được kết nối:
  - a. Thực hiện điều chỉnh bó để giải quyết cho góc quay  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  và tiêu cự F của tất cả camera
  - b. Kết xuất Panorama[7] sử dụng pha trộn nhiều băng tần

b. **Output:** Hình ảnh toàn cảnh

**Ảnh hưởng:** Việc sử dụng các tính năng bất biến đã trở thành một tiêu chuẩn trong khâu hình ảnh và công việc này ảnh hưởng đáng kể đến nghiên cứu tiếp theo trong lĩnh vực này.

## Image Stitching Algorithm: An Optimization between Correlation-Based and Feature-Based Method

Đây là một bài nghiên cứu được viết bởi *Moushumi Zaman Bonny* và *Mohammad Shorif Uddin*, cả hai đều là tiến sĩ khoa Khoa học và kỹ thuật máy tính của trường đại học Jahangirnagar ở thành phố Dhaka, Bangladesh. Bài nghiên cứu này của cả hai đã được đăng trên IJCSIS (Tập chí Quốc tế về Khoa học Máy tính và An toàn Thông tin) và tháng 6, 2018. Bài nghiên cứu này trình bày về việc sử dụng các phương pháp ghép ảnh để thử nghiệm với ảnh tự nhiên, HDR và ảnh siêu nhỏ (ảnh dưới kính hiển vi) và đã thu được những kết quả khả quan.

Những phương pháp mà họ đã áp dụng được liệt kê như sau:

- Thu nhận ảnh: chụp ảnh bằng camera hoặc thu từ các nguồn thứ cấp.
- Căn chỉnh:
  - Lý do của việc căn chỉnh là để xác định các tham số căn chỉnh tin cậy có thể giảm thiểu sự sai lệch các điểm chung giữa mỗi cặp ảnh. Điều này rất hữu ích để mở rộng tiêu chí kết nối cặp ảnh.
  - Chúng ta cần phải phát hiện và kết nối các đặc trưng giữa ảnh  $I(n)$  và  $I(n-1)$ . Sau đó, chúng ta phải ước lượng phép biến đổi hình học  $T$  sao cho ánh xạ từ ảnh  $I(n)$  sang ảnh  $I(n-1)$  được thực hiện như sau:
$$T = T(1) * T(2) * T(3) * \dots * T(n-1) * T(n)$$

Trong công thức trên, các  $T$  lần lượt là phép biến đổi hình học của các ảnh  $I(1)$  tới  $I(n)$ .

- Định vị vùng chồng chéo: Bước tiếp theo là xác định khu vực chồng chéo. Điều này là quan trọng để có các đặc trưng liên tiếp giữa nhiều ảnh. Nếu khu vực chồng chéo tồn tại, chúng ta có thể có hai loại kết quả: ảnh toàn cảnh dựa trên đặc trưng và ảnh toàn cảnh dựa trên tương quan. Nhưng nếu không có sự chồng chéo, phương pháp này sẽ ghép trực tiếp lần lượt các ảnh.
- Ghép ảnh dựa trên tính năng (Feature-based):
  - Thường thì để tạo ảnh toàn cảnh, một ảnh được chọn làm ảnh tham chiếu và sau đó tất cả các ảnh khác được sử dụng theo thứ tự để nối. Kết quả thường được biết đến là một ảnh toàn cảnh phẳng.
  - Việc nối ảnh có thể được áp dụng trên nhiều kiểu chiếu, ví dụ như chiếu thẳng, nơi ảnh toàn cảnh được quan sát nằm trên một mặt phẳng (hai chiều) cắt qua một pano-sphere tại một điểm. Chiếu thẳng sử dụng các mặt của hình

lập phương với ánh xạ lập phương để xem ảnh toàn cảnh và nó thể hiện một chiều tròn.

- Sau khi vẽ các pixel nguồn lên bề mặt hợp thành cuối cùng, việc kết hợp chúng để tạo ra một ảnh toàn cảnh là cần thiết. Do đó, phép toán kết hợp là một điều kiện tiên quyết cho quá trình nối ảnh. Feathering, image pyramid và gradient domain là một số phương pháp kết hợp quen thuộc. Feathering được sử dụng để làm mờ các cạnh của các đặc trưng. Từ hai ảnh chồng chéo, giá trị trung bình của pixel được đánh giá cho các khu vực được kết hợp. Kết hợp ảnh đa dải là một phương pháp khác được biết đến thực hiện trong miền độ dốc.
- Những độ lớn này được sử dụng để đảm bảo rằng giá trị pixel đầu ra không vượt quá giá trị pixel cao nhất.
- Hình ảnh kết quả  $Q(i,j)$  được tính bằng cách sử dụng công thức:

$$Q(i,j) = X * P1(i,j) + (1-x) * P2(i,j)$$

$P1$  và  $P2$  là hai ảnh đầu vào. Trong một số ứng dụng,  $X$  cũng có thể là một hằng số, cho phép thêm một giá trị lệch hằng số vào một ảnh duy nhất.  $X$  là tỷ lệ trộn giữa hai ảnh đầu vào, quyết định sự kiểm soát của từng điểm ảnh đầu vào vào ảnh đầu ra.  $X$  có thể là một hằng số đối với tất cả các pixel trong ảnh hoặc có thể được xác định cho từng pixel riêng lẻ bằng cách sử dụng một mặt nạ. Kích thước của mặt nạ phải giống với kích thước của các ảnh.

- Ghép ảnh dựa trên các mối liên kết (Correlation-based):
  - Trong phương pháp nối ảnh dựa trên các mối liên kết, chúng ta không cần phải trích xuất đặc trưng. Quy trình này phụ thuộc vào mối quan hệ giữa hai ảnh. Phương pháp này là biến thể của quá trình quay, co giãn và các biến đổi khác. Đối với những hình ảnh đã được xử lý trước một cách chính xác, phương pháp này là một cách hiệu quả để áp dụng.
  - Chúng ta đã tính hệ số tương quan  $r$  giữa hai ảnh  $A$  và  $B$  bằng cách sử dụng công thức sau đây:

$$A' = \text{mean}(A_{mn}) \text{ and } B' = \text{mean}(B_{mn}).$$

$$r = \frac{\sum_m \sum_n (A - A')(B - B')}{\sqrt{(\sum_m \sum_n (A - A')^2)(\sum_m \sum_n (B - B')^2)}}$$

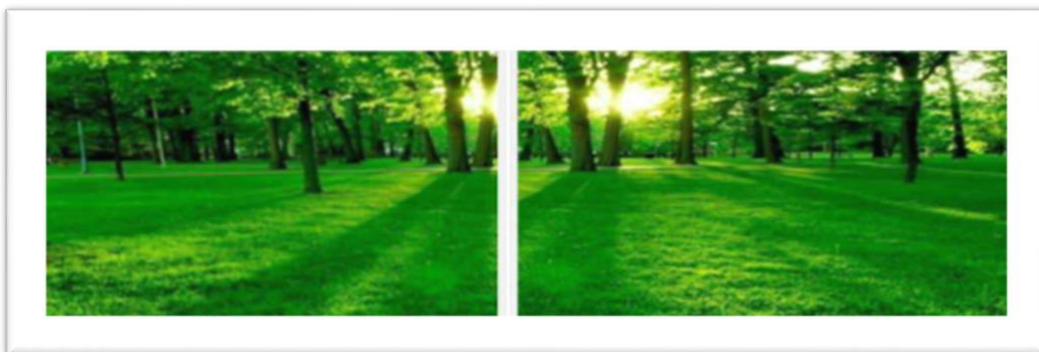


- Sau khi xác định giá trị tương quan tối đa giữa hai ảnh, chúng ta cần tìm giá trị tương quan tối đa của vector cột.
- Gộp là bước tiếp theo. Khi các pixel nguồn được ánh xạ lên bề mặt cuối cùng, chúng được trộn để tạo ra một ảnh toàn cảnh và đường nối được điều chỉnh để làm mờ đi, giúp ảnh mượt hơn.

**So sánh:** Trong bước này, Feature-based và Correlation-based được so sánh với thực tế (Ground truth). Phương pháp sẽ tiếp tục với kết quả tốt hơn dựa trên tối ưu hóa độ chính xác.

**Đây là kết quả của 1 ví dụ đơn giản:**

2 góc độ khác nhau



Sau khi ghép:



*Hình 1: Kết quả ghép hai ảnh từ hai góc độ*

## Research on Image Stitching Based on Invariant Features of Reconstructed Plane

Đây là 1 bài nghiên cứu của tiến sĩ Qi Liu - trường Kỹ thuật Điện và Tự động hóa, viện công nghệ Cấp Nhĩ Tân, Xiyu Tang - trường Hàng không vũ trụ, viện công nghệ Cấp Nhĩ Tân và Ju Huo.

Trong bài nghiên cứu, họ trình bày một phương pháp biến đổi ảnh tích hợp ràng buộc đặc trưng phẳng nhằm cải thiện hiệu quả tổng thể của việc nối ảnh bằng cách hạn chế hình dạng mặt phẳng. Trong giai đoạn phát hiện đặc trưng, họ giải quyết vấn đề mất mát đường đặc trưng trong quá trình khớp đặc trưng bằng cách mở rộng các đặc trưng dạng đường. Hơn nữa, tạo ra thêm điểm khớp bằng cách tích hợp sự giao nhau của các đường được khớp như một điểm đặc trưng mới và mở rộng những điểm này đến các khu vực không có cấu trúc để giải quyết vấn đề mất mát đặc trưng trong các vùng có cấu trúc yếu.



Hình 2: So sánh các phương pháp ghép ảnh

Hình ảnh 1 hiển thị kết quả của phương pháp nối ảnh của họ dựa trên các đặc trưng phẳng không biến đổi. Kết quả nối ảnh APAP, AANAP, SPW, ELA cho thấy rõ các hiện vật rõ ràng (được hiển thị trong hộp màu đỏ). LPC hoạt động tốt hơn trong việc giải quyết vấn đề căn chỉnh ảnh, loại bỏ các hiện vật xuất hiện trong ảnh. Tuy nhiên, xử lý chi tiết của LPC kém hơn (được hiển thị trong hộp màu vàng). Trong khi phương pháp của họ đảm bảo sự

căn chỉnh, nó cũng hiệu quả trong việc ràng buộc các đặc trưng phẳng để giữ lại các đặc trưng cấu trúc, dẫn đến hình ảnh nổi tự nhiên hơn cả ở cấp địa phương và toàn cầu. Đóng góp của bài báo được tóm tắt như sau:

- Phát triển một chiến lược để bảo tồn các đặc trưng đường phụ thuộc vào kết nối đoạn đường của vùng hỗ trợ. Bằng cách tạo ra các điểm đặc trưng mới tại giao điểm của các đường đặc trưng được khớp, chiến lược này hiệu quả giảm thiểu mất mát của đường đặc trưng trong quá trình khớp, ngăn chặn sự trùng lặp trong khi tăng số lượng điểm khớp và hỗ trợ bảo tồn chi tiết hình ảnh.
- Giới thiệu ràng buộc phẳng trong quá trình nổi ảnh. Bằng cách đặt các hạn chế về đặc trưng phẳng và khoảng cách, ta giảm phạm vi các vị trí tương đối và sự biến đổi hình dạng có thể xảy ra trong các đối tượng đặc trưng trong quá trình biến đổi, từ đó dẫn đến kết quả nổi ảnh và hiệu suất tổng thể của phương pháp nổi ảnh tốt hơn.
- Đề xuất một chỉ số toàn diện nhằm đo lường mức độ hiệu quả trong việc duy trì cấu trúc đặc trưng phẳng trong quá trình nổi ảnh.

## 2. Quá trình phát triển các giải pháp liên quan đến đề tài

### 1. Nghiên cứu tiền đề:

- Định rõ vấn đề và mục tiêu của ghép ảnh. Hiểu rõ yêu cầu cụ thể và ràng buộc của ứng dụng.
- Tiến hành nghiên cứu về các phương pháp và công nghệ liên quan đến ghép ảnh toàn cảnh.
- Hiểu rõ những khó khăn và cơ hội của việc ghép ảnh từ nhiều hình ảnh để tạo ra một hình ảnh toàn cảnh.

### 2. Phân tích và lựa chọn phương pháp:

- Đánh giá các phương pháp hiện tại trong lĩnh vực ghép ảnh và xác định rằng việc sử dụng tính năng bất biến, đặc biệt là SIFT, là một giải pháp hợp lý.
- Nắm vững các ưu điểm và hạn chế của phương pháp được lựa chọn.

### 3. Thu thập dữ liệu: Tập hợp các hình ảnh cần được ghép lại. Đảm bảo rằng các hình ảnh bao phủ lĩnh vực quan sát mong muốn và có độ chồng lấp đủ để ghép ảnh chính xác.

### 4. Tiền xử lý ảnh: Tiền xử lý các hình ảnh để cải thiện chất lượng và thuận tiện cho việc trích xuất đặc trưng. Các bước tiền xử lý thông thường bao gồm thay đổi kích thước, sửa màu, giảm nhiễu và sửa chữa hình học.

## 5. Thiết kế giải thuật:

- i) **Trích xuất đặc trưng:** Xác định các đặc trưng đặc biệt trong mỗi hình ảnh có thể được sử dụng cho việc so khớp. Các phương pháp phổ biến bao gồm SIFT, SURF hoặc ORB, cung cấp các điểm chính và mô tả không đổi với tỉ lệ, xoay và thay đổi ánh sáng.
- ii) **So khớp đặc trưng:** So khớp các đặc trưng trích xuất qua các hình ảnh khác nhau. Thiết lập sự tương ứng giữa các điểm chính để xác định cách hình ảnh nên được căn chỉnh.
- iii) **Ước lượng biến đổi:** Ước lượng biến đổi (đồng nhất hóa) giữa các hình ảnh dựa trên các đặc trưng đã so khớp. Biến đổi này quan trọng để căn chỉnh chính xác hình ảnh trong quá trình ghép.
- iv) **Tổng hợp hình ảnh:** tổng hợp các vùng chồng lấp của hình ảnh để tạo ra sự chuyển tiếp mượt mà. Các kỹ thuật tổng hợp phổ biến bao gồm tổng hợp tuyến tính, tổng hợp dựa trên độ dốc hoặc các phương pháp nâng cao như tổng hợp đa dải.
- v) **Tối ưu hoá chất lượng:** Tối ưu hóa các tham số ghép như căn chỉnh, tổng hợp và kết quả ghép. Điều chỉnh một cách lặp lại sự căn chỉnh, tổng hợp và các tham số khác để cải thiện chất lượng tổng thể của hình ảnh ghép.

## 6. Hiện thực và kiểm thử:

- Triển khai giải thuật bằng cách sử dụng các ngôn ngữ và công cụ lập trình phù hợp.
- Kiểm thử giải thuật trên bộ dữ liệu đa dạng để đảm bảo tính linh hoạt và hiệu suất.

## 7. Tối ưu hóa và điều chỉnh:

- Tối ưu hóa giải thuật để cải thiện hiệu suất và thời gian thực hiện.
- Điều chỉnh các tham số và quy trình để đạt được kết quả tốt nhất.

## 8. Đánh giá và so sánh:

So sánh kết quả của giải thuật với các phương pháp khác trong các thử nghiệm và đánh giá đầy đủ hiệu suất và độ chính xác.

## 9. Xây dựng ứng dụng hoặc tích hợp vào hệ thống:

Nếu có, phát triển một ứng dụng hoặc tích hợp giải thuật vào hệ thống có thể sử dụng cho mục đích cụ thể.

### 10. Viết báo cáo và xuất bản kết quả:

- Viết báo cáo chi tiết về quá trình phát triển giải pháp, kết quả của các thử nghiệm, và nhận xét về hiệu suất.
- Nếu có thể, đăng ký và xuất bản bài báo khoa học để chia sẻ kinh nghiệm và kiến thức mới.

Quá trình này yêu cầu sự kiên trì, kiểm soát chất lượng, và khả năng sáng tạo để giải quyết những thách thức cụ thể của đề tài. Các nghiên cứu và phát triển trong lĩnh vực này có thể thú vị và có ý nghĩa vì chúng giúp cải thiện khả năng tạo ra hình ảnh toàn cảnh từ nhiều nguồn dữ liệu.

## 3. Bảng so sánh giải pháp của các công trình nghiên cứu

Tiêu chí	Phương pháp của Brown and Lowe (2007)	Phương pháp của Xu, Liu và Yang (2023)
Phương pháp	Dựa trên các tính năng bất biến	Dựa trên học máy
Bước thực hiện	Trích xuất tính năng - Tìm lân cận - Kết hợp tính năng - Xác minh khớp - Điều chỉnh bó - Kết xuất	Trích xuất tính năng - Học mô hình - Dự đoán điểm tương ứng - Điều chỉnh bó - Kết xuất
Ưu điểm	Độ chính xác cao, phù hợp với nhiều loại ảnh	Độ chính xác cao, hiệu quả trong các điều kiện môi trường khó khăn
Nhược điểm	Tốc độ chậm, khó xử lý các ảnh bị nhiễu	Yêu cầu tập dữ liệu lớn để đào tạo mô hình
Ứng dụng	Chụp ảnh toàn cảnh, thực tế ảo	Chụp ảnh toàn cảnh, thực tế ảo, nhận dạng đối tượng

### Nhận xét:

Phương pháp của Brown và Lowe là một trong những phương pháp ghép ảnh đầu tiên sử dụng các tính năng bất biến. Phương pháp này có độ chính xác cao, phù hợp với nhiều loại ảnh. Tuy nhiên, phương pháp này có tốc độ chậm và khó xử lý các ảnh bị nhiễu.

Phương pháp của Xu, Liu và Yang là một phương pháp ghép ảnh mới sử dụng học máy. Phương pháp này có độ chính xác cao, hiệu quả trong các điều kiện môi trường khó khăn. Tuy nhiên, phương pháp này yêu cầu tập dữ liệu lớn để đào tạo mô hình.

Tùy thuộc vào các yêu cầu cụ thể của ứng dụng, có thể lựa chọn phương pháp ghép ảnh phù hợp. Nếu cần độ chính xác cao và phù hợp với nhiều loại ảnh, có thể sử dụng phương pháp của Brown và Lowe. Nếu cần độ chính xác cao và hiệu quả trong các điều kiện môi trường khó khăn, có thể sử dụng phương pháp của Xu, Liu và Yang.

## 4. Tổng kết

Việc ghép ảnh chủ yếu được sử dụng trong nhiều ứng dụng khác nhau như ổn định hình ảnh, tạo hình ảnh HDR (High Dynamic Range), ghép ảnh mosaic để tạo bản đồ số và ảnh vệ tinh có độ phân giải cao, hình ảnh y khoa, siêu độ phân giải của nhiều ảnh, nối video, chèn đối tượng, nối ảnh vi sinh học và tạo ảnh đồng nhóm. Đã có nhiều nghiên cứu và ứng dụng trong việc này và đã thu về không ít thành công và kết quả thu được rất khả quan.

Tóm lại, họ đề xuất một phương pháp ghép ảnh từ một góc nhìn dựa trên sự nhất quán hình dạng của đặc trưng phẳng, không giới hạn việc kết hợp và căn chỉnh của điểm và đường đặc trưng, mà thay vào đó tăng cường bảo quản hình dạng và tự nhiên tổng thể của kết quả ghép ảnh. Đầu tiên, sử dụng các phép toán đặc trưng đường để giữ lại tối đa các đường đặc trưng của các khu vực chồng lấp và tạo ra các điểm đặc trưng mới để cải thiện hiệu quả của giai đoạn căn chỉnh sơ bộ. Sau đó, mô tả các đặc trưng hình phẳng của ảnh thông qua các điểm và đường đặc trưng đã được chỉ định, và ràng buộc những đặc trưng này của quá trình ghép ảnh. Cuối cùng, đề xuất các độ đo cho đánh giá lượng tử để bổ sung và cải thiện hệ thống đánh giá lượng tử. Kết quả thử nghiệm chứng minh rằng phương pháp căn chỉnh khu vực chồng lấp một cách chính xác hơn và có ưu điểm đáng kể trong việc biến dạng cả trong khu vực chồng lấp và không chồng lấp, và đánh giá lượng tử cho thấy phương pháp của họ có hiệu suất tốt hơn so với các kỹ thuật hiện tại.

Trong nghiên cứu của họ, họ xây dựng các mặt phẳng đặc trưng dựa trên các điểm đặc trưng và đường đặc trưng đã xác định. Ràng buộc của các đặc trưng khác nhau có thể ảnh hưởng tiêu cực đến quá trình ghép ảnh. Trong nghiên cứu tương lai, họ sẽ nghiên cứu các phương pháp để giảm thiểu ảnh hưởng này và tiếp tục cải thiện ghép ảnh dựa trên đặc trưng.

**Mục đích cuối cùng của việc ghép ảnh từ nhiều góc nhìn khác nhau** → là để tạo ra một hình ảnh toàn cảnh, cho phép người dùng xem một cảnh từ nhiều góc độ khác nhau.

**Hình ảnh toàn cảnh có thể được sử dụng cho nhiều ứng dụng khác nhau, bao gồm:**

- **Chụp ảnh toàn cảnh:** Chụp ảnh toàn cảnh là một trong những ứng dụng phổ biến nhất của ghép ảnh. Hình ảnh toàn cảnh có thể được sử dụng để ghi lại cảnh quan, kiến trúc hoặc các sự kiện.

- **Thực tế ảo:** Hình ảnh toàn cảnh có thể được sử dụng để tạo ra trải nghiệm thực tế ảo. Điều này cho phép người dùng đắm mình trong một cảnh ảo như thể họ đang ở đó.
- **Thực tế tăng cường:** Hình ảnh toàn cảnh có thể được sử dụng để tạo ra trải nghiệm thực tế tăng cường. Điều này cho phép người dùng thêm thông tin vào thế giới thực.
- **Nhận dạng đối tượng:** Hình ảnh toàn cảnh có thể được sử dụng để cải thiện hiệu quả của nhận dạng đối tượng. Điều này là do hình ảnh toàn cảnh cung cấp nhiều thông tin hơn về đối tượng hơn một hình ảnh đơn lẻ.

**Các phương pháp ghép ảnh từ nhiều góc nhìn khác nhau có thể được phân loại thành hai loại chính:**

- **Ghép ảnh dựa trên đặc trưng:** Các phương pháp này dựa trên việc phát hiện và khớp các đặc trưng trong các ảnh. Các đặc trưng có thể là điểm, đường hoặc vùng.
- **Ghép ảnh dựa trên tương quan:** Các phương pháp này dựa trên tính toán tương quan giữa các ảnh. Tương quan là một phép đo mức độ giống nhau giữa hai hình ảnh.  
→ **Mỗi loại phương pháp có ưu điểm và nhược điểm riêng.**
- **Ghép ảnh dựa trên đặc trưng** có thể tạo ra kết quả ghép ảnh chính xác hơn, nhưng chúng có thể chậm hơn và đòi hỏi nhiều bộ nhớ hơn.
- **Ghép ảnh dựa trên tương quan** nhanh hơn và sử dụng ít bộ nhớ hơn, nhưng chúng có thể kém chính xác hơn.
- Các nghiên cứu về ghép ảnh từ nhiều góc nhìn khác nhau đang tập trung vào việc cải thiện độ chính xác, hiệu quả và khả năng thích ứng của các phương pháp ghép ảnh. Các nghiên cứu cũng đang tập trung vào việc phát triển các phương pháp ghép ảnh mới cho các ứng dụng cụ thể, chẳng hạn như thực tế ảo và thực tế tăng cường.

## Chapter 3 : Phương pháp

### Panoramic image stitching using invariant features

Ghép ảnh toàn cảnh bằng các đặc điểm bất biến là một phương pháp mạnh mẽ để tạo ra những hình ảnh góc rộng liền mạch từ nhiều ảnh chồng lấn nhau. Phương pháp này dựa trên việc xác định và khớp các điểm đặc biệt trong ảnh, ngay cả khi chúng bị thay đổi theo góc nhìn, điều kiện ánh sáng hoặc những biến dạng nhỏ. Sau đây là phân tích chi tiết về phương pháp này:



## 1. So khớp đặc điểm:

Bước đầu tiên trong thuật toán nhận dạng toàn cảnh là trích xuất và khớp các đặc trưng SIFT giữa tất cả các hình ảnh. Các đặc trưng SIFT được định vị tại các cực đại/cực tiểu không gian tỷ lệ của hàm sai phân Gaussian. Tại mỗi vị trí đặc trưng, một tỷ lệ và hướng đặc trưng được thiết lập. Điều này mang lại một khung tương tự bất biến để thực hiện các phép đo. Mặc dù chỉ cần lấy mẫu các giá trị cường độ trong khung này sẽ là bất biến tương tự, nhưng bộ mô tả bất biến thực sự được tính toán bằng cách tích lũy các gradient cục bộ trong biểu đồ định hướng. Điều này cho phép các cạnh dịch chuyển nhẹ mà không làm thay đổi vector mô tả, tạo ra một số khả năng mạnh mẽ đối với thay đổi affine. Sự tích lũy không gian này cũng rất quan trọng đối với khả năng bất biến dịch chuyển, vì các vị trí điểm quan tâm thường chỉ chính xác trong phạm vi 0-3 pixel. Khả năng bất biến về độ sáng đạt được bằng cách sử dụng gradient (loại bỏ độ chệch) và chuẩn hóa vector mô tả (loại bỏ độ lợi). Vì các đặc trưng SIFT là bất biến dưới các thay đổi về xoay và tỷ lệ, hệ thống của chúng tôi có thể xử lý các hình ảnh có hướng và độ thu phóng khác nhau.

Giả sử rằng camera quay quanh tâm quang học của nó, nhóm các phép biến đổi mà các hình ảnh có thể trải qua là một nhóm đặc biệt của phép đồng dạng. Ta tham số hóa mỗi camera bằng một vector quay  $\theta = [\theta_1, \theta_2, \theta_3]$  và tiêu cự  $f$ . Điều này cho phép các phép đồng dạng theo cặp  $\tilde{u}_i = H_{ij}\tilde{u}_j$  trong đó:

$$\mathbf{H}_{ij} = \mathbf{K}_i \mathbf{R}_i \mathbf{R}_j^T \mathbf{K}_j^{-1}$$

và  $\tilde{u}_i, \tilde{u}_j$  là các vị trí hình ảnh đồng nhất ( $\tilde{u}_i = [u_i, 1]$ , trong đó  $u_i$  là vị trí hình ảnh 2 chiều). Mô hình camera 4 tham số được định nghĩa bởi:

$$\mathbf{K}_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

và (sử dụng biểu diễn hàm mũ cho phép quay)

$$\mathbf{R}_i = e^{[\theta_i]_{\times}}, \quad [\theta_i]_{\times} = \begin{bmatrix} 0 & -\theta_{i3} & \theta_{i2} \\ \theta_{i3} & 0 & -\theta_{i1} \\ -\theta_{i2} & \theta_{i1} & 0 \end{bmatrix}$$

Lý tưởng nhất là người ta sẽ sử dụng các đặc điểm hình ảnh bất biến trong nhóm biến đổi này. Tuy nhiên, đối với những thay đổi nhỏ về vị trí hình ảnh



$$\mathbf{u}_i = \mathbf{u}_{i0} + \left. \frac{\partial \mathbf{u}_i}{\partial \mathbf{u}_j} \right|_{\mathbf{u}_{i0}} \Delta \mathbf{u}_j$$

hoặc tương đương  $\tilde{\mathbf{u}}_i = \mathbf{A}_{ij} \tilde{\mathbf{u}}_j$ , trong đó

$$\mathbf{A}_{ij} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

là một phép biến đổi affine thu được bằng cách tuyến tính hóa phép đồng nhất về  $\mathbf{u}_{i0}$ . Điều này ngụ ý rằng mỗi mảng hình ảnh nhỏ trải qua một phép biến đổi affine và biện minh cho việc sử dụng các tính năng SIFT bất biến một phần khi thay đổi affine.

Khi các đặc điểm đã được trích xuất từ tất cả  $n$  ảnh (thời gian tuyến tính), chúng phải được so khớp. Vì nhiều hình ảnh có thể chồng lên một tia, nên mỗi đặc điểm được so khớp với  $k$  lân cận gần nhất của nó trong không gian đặc điểm (sử dụng  $k = 4$ ). Điều này có thể được thực hiện trong thời gian  $O(n \log n)$  bằng cách sử dụng cây  $k$ -d để tìm hàng xóm gần nhất gần đúng. Cây  $k$ -d là một phân vùng không gian nhị phân được căn chỉnh theo trục, phân vùng đệ quy không gian đặc trưng ở mức trung bình theo chiều có phương sai cao nhất.

## 2. So khớp hình ảnh:

Ở giai đoạn này, mục tiêu là tìm tất cả các hình ảnh phù hợp (tức là chồng chéo). Các tập hợp hình ảnh trùng khớp được kết nối sau này sẽ trở thành ảnh toàn cảnh. Vì mỗi hình ảnh có thể khớp với mọi hình ảnh khác nên vấn đề này thoạt đầu có vẻ là vấn đề bậc hai về số lượng hình ảnh. Tuy nhiên, chỉ cần ghép từng hình ảnh với một số ít hình ảnh chồng chéo để có được giải pháp tốt về hình học hình ảnh. Từ bước so khớp đặc điểm, ta đã xác định được những hình ảnh có số lượng trùng khớp lớn giữa chúng. Ta coi số hình ảnh  $m$  không đổi, có số lượng tính năng phù hợp nhất với hình ảnh hiện tại, là hình ảnh tiềm năng phù hợp (chúng tôi sử dụng  $m = 6$ ). Đầu tiên, ta sử dụng RANSAC để chọn một tập hợp các phần tử bên trong tương thích với sự đồng nhất giữa các hình ảnh. Tiếp theo, áp dụng mô hình xác suất để xác minh sự trùng khớp.

### 2.1. Ước tính đồng nhất bằng RANSAC:

RANSAC (random sample consensus) là một quy trình ước tính mạnh mẽ sử dụng một tập hợp tối thiểu các tương ứng được lấy mẫu ngẫu nhiên để ước tính các tham số chuyển đổi hình ảnh và tìm ra giải pháp có sự đồng thuận tốt nhất với dữ liệu. Trong trường hợp ảnh toàn cảnh, ta chọn các bộ tương ứng đặc trưng  $r = 4$  và tính toán đặc điểm tương đồng H giữa chúng bằng phương pháp biến đổi tuyến tính trực tiếp (DLT). Ta lặp lại điều này với  $n = 500$  thử nghiệm và chọn giải pháp có số lượng phần tử nội bộ tối đa (có các phép chiếu phù hợp với H trong phạm vi dung sai epsilon pixel). Cho xác suất để một đặc điểm trùng khớp là chính xác giữa một cặp hình ảnh trùng khớp (xác suất nội bộ) là  $p_i$ , xác suất tìm thấy phép biến đổi đúng sau  $n$  lần thử là

$$p(\mathbf{H} \text{ is correct}) = 1 - (1 - (p_i)^r)^n.$$

Sau một số lượng lớn các thử nghiệm, xác suất tìm được hình ảnh đồng nhất chính xác là rất cao. Ví dụ: đối với xác suất nhỏ hơn  $p_i = 0,5$ , xác suất không tìm thấy hình ảnh đồng nhất chính xác sau 500 lần thử là xấp xỉ  $1 \times 10^{-14}$ .

## 2.2. Mô hình xác suất để xác minh kết quả trùng khớp hình ảnh:

Đối với mỗi cặp hình ảnh có khả năng trùng khớp, ta có một tập hợp các đối tượng nhất quán về mặt hình học (các đối tượng bên trong RANSAC) và một tập hợp các đối tượng nằm trong vùng chồng chéo nhưng không nhất quán (các đối tượng ngoại lệ RANSAC). Ý tưởng của mô hình xác minh của ta là so sánh xác suất mà tập hợp các giá trị inlier/outlier này được tạo ra bởi một kết quả khớp hình ảnh chính xác hoặc bằng một kết quả khớp hình ảnh sai.

Đối với một hình ảnh nhất định, ta biểu thị tổng số đặc điểm trong khu vực chồng chéo  $n_f$  và số lượng phần tử  $n_i$ . Sự kiện hình ảnh khớp đúng/sai được biểu thị bằng biến nhị phân  $m$  thuộc  $\{0, 1\}$ . Sự kiện mà đặc điểm thứ  $i$  khớp với  $f^{(i)}$  thuộc  $\{0, 1\}$  là một phần tử inlier/outlier được giả sử là Bernoulli độc lập, sao cho tổng số phần tử nội bộ là nhị thức

Trong đó  $p_1$  là xác suất một đặc điểm là một inlier cho ra một hình ảnh khớp chính xác và  $p_0$  là xác suất một đặc điểm là một inlier cho ra một hình ảnh sai. Tập các biến đặc trưng khớp  $\{f^{(i)}, i = 1, 2, \dots, n_f\}$  được ký hiệu là  $f^{(1:nf)}$ . Số lượng inlier  $n_i = \sum_{i=1}^{n_f} f^{(i)}$  và  $B(\cdot)$  là phân phối nhị thức

$$B(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

$$p(f^{(1:n_f)} | m = 1) = B(n_i; n_f, p_1)$$

$$p(f^{(1:n_f)} | m = 0) = B(n_i; n_f, p_0)$$

Ta chọn giá trị  $p_1 = 0,6$  và  $p_0 = 0,1$ . Bây giờ chúng ta có thể đánh giá xác suất hậu nghiệm để một hình ảnh khớp chính xác bằng Quy tắc Bayes

$$\begin{aligned} p(m = 1 | f^{(1:n_f)}) &= \frac{p(f^{(1:n_f)} | m = 1) p(m = 1)}{p(f^{(1:n_f)})} \\ &= \frac{1}{1 + \frac{p(f^{(1:n_f)} | m=0) p(m=0)}{p(f^{(1:n_f)} | m=1) p(m=1)}} \end{aligned}$$

Ta chấp nhận bức ảnh nếu  $p(m = 1 | f^{(1:n_f)}) > p_{\min}$

Chọn giá trị  $p(m = 1) = 10^{-6}$  và  $p_{\min} = 0.999$  đưa ra điều kiện

$$\frac{B(n_i; n_f, p_1) p(m = 1)}{B(n_i; n_f, p_0) p(m = 0)} \underset{\text{reject}}{\overset{\text{accept}}{\geq}} \frac{1}{\frac{1}{p_{\min}} - 1}. \quad n_i > \alpha + \beta n_f$$

để có hình ảnh khớp chính xác, trong đó  $\alpha = 8,0$  và  $\beta = 0,3$ . Mặc dù trong thực tế, chúng ta đã chọn các giá trị cho  $p_0$ ,  $p_1$ ,  $p(m = 0)$ ,  $p(m = 1)$  và  $p_{\min}$ , nhưng về nguyên tắc chúng có thể được học từ dữ liệu. Ví dụ:  $p_1$  có thể được ước tính bằng cách tính toán tỷ lệ khớp phù hợp với các cặp chính xác trên một tập dữ liệu lớn.

Khi các kết quả khớp từng cặp đã được thiết lập giữa các hình ảnh, chúng ta có thể tìm thấy các chuỗi ảnh toàn cảnh dưới dạng tập hợp các hình ảnh phù hợp được kết nối. Điều này cho phép ta nhận dạng nhiều ảnh toàn cảnh trong một tập hợp hình ảnh và loại bỏ các hình ảnh nhiễu không khớp với các hình ảnh khác.

### 3. Bundle Adjustment

Với một tập hợp các kết quả khớp nhất quán về mặt hình học giữa các hình ảnh, ta sử dụng tính năng bundle adjustment để giải quyết chung tất cả các thông số máy ảnh. Đây là một bước thiết yếu vì việc ghép các hình ảnh đồng nhất theo cặp sẽ gây ra lỗi tích lũy và bỏ qua nhiều ràng buộc giữa các hình ảnh, ví dụ: các phần cuối của ảnh toàn cảnh sẽ nối với nhau. Lần lượt các hình ảnh được thêm vào bundle adjuster, với hình ảnh phù hợp nhất (số lượng kết quả phù hợp tối đa) sẽ được thêm vào ở mỗi bước. Hình ảnh mới được khởi

tạo với cùng góc xoay và tiêu cự để hình ảnh phù hợp nhất. Sau đó, các tham số được cập nhật bằng **Levenberg-Marquardt**.

Hàm mục tiêu mà ta sử dụng là hàm tăng cường sai số phép chiếu bình phương tổng. Nghĩa là, mỗi đặc điểm được chiếu vào tất cả các hình ảnh phù hợp với nó và tổng khoảng cách hình ảnh bình phương được giảm thiểu đối với các thông số máy ảnh. Đưa ra sự tương quan  $u_i^k \leftrightarrow u_j^l$  ( $u_i^k$  biểu thị vị trí đặc điểm thứ k trong ảnh i), phần dư là

$$\mathbf{r}_{ij}^k = \mathbf{u}_i^k - \mathbf{p}_{ij}^k$$

trong đó  $\mathbf{p}_{ij}^k$  là hình chiếu từ ảnh j lên ảnh i của điểm tương ứng với  $u_i^k$

$$\bar{\mathbf{p}}_{ij}^k = \mathbf{K}_i \mathbf{R}_i \mathbf{R}_j^T \mathbf{K}_j^{-1} \bar{\mathbf{u}}_j^l.$$

Hàm lỗi là tổng của tất cả các ảnh của các lỗi dư được củng cố

$$e = \sum_{i=1}^n \sum_{j \in \mathcal{I}(i)} \sum_{k \in \mathcal{F}(i,j)} h(\mathbf{r}_{ij}^k) \quad (16)$$

Trong đó n là số lượng ảnh,  $\mathcal{I}(i)$  là tập hợp các ảnh khớp với ảnh i,  $\mathcal{F}(i, j)$  là tập hợp các đặc điểm khớp giữa ảnh i và j. Ta sử dụng **hàm lỗi Huber**.

*Hàm lỗi này kết hợp các thuộc tính hội tụ nhanh của sơ đồ tối ưu hóa định mức L2 cho các phần tử inlier (khoảng cách nhỏ hơn  $\sigma$ ), với độ bền của sơ đồ định mức L1 cho các phần tử outlier (khoảng cách lớn hơn  $\sigma$ ). Chúng ta sử dụng khoảng cách outlier  $\sigma = \infty$  trong quá trình khởi tạo và  $\sigma = 2$  pixel cho giải pháp cuối cùng.*

$$h(\mathbf{x}) = \begin{cases} |\mathbf{x}|^2, & \text{if } |\mathbf{x}| < \sigma \\ 2\sigma|\mathbf{x}| - \sigma^2, & \text{if } |\mathbf{x}| \geq \sigma \end{cases}.$$

Đây là bài toán bình phương tối thiểu phi tuyến tính được giải bằng thuật toán Levenberg-Marquardt. Mỗi bước lặp có dạng

$$\Phi = (\mathbf{J}^T \mathbf{J} + \lambda \mathbf{C}_p^{-1})^{-1} \mathbf{J}^T \mathbf{r}$$

trong đó  $\Phi$  đều là thông số,  $\mathbf{r}$  là phần dư và  $\mathbf{J}$  là đạo hàm của  $\mathbf{r}$  theo  $\Phi$ .

Giá trị này được thiết lập sao cho độ lệch chuẩn của các góc là  $\sigma_\theta = \pi/16$  và tiêu cự  $\sigma_f = \bar{f}/10$  (trong đó  $\bar{f}$  là giá trị trung bình của các tiêu cự được ước tính cho đến nay). Điều này giúp chọn kích thước bước phù hợp và do đó tăng tốc độ hội tụ. Ví dụ: nếu sử dụng ma

trận hiệp phương sai hình cầu, sự thay đổi 1 radian khi xoay sẽ bị phạt tương đương với sự thay đổi 1 pixel trong tham số độ dài tiêu cự.

$$\mathbf{C}_p = \begin{bmatrix} \sigma_\theta^2 & 0 & 0 & 0 & 0 & \dots \\ 0 & \sigma_\theta^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma_\theta^2 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma_f^2 & 0 & \dots \\ 0 & 0 & 0 & 0 & \sigma_\theta^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix}$$

Cuối cùng, tham số  $\lambda$  được thay đổi ở mỗi lần lặp để đảm bảo rằng hàm mục tiêu của phương trình 16 trên thực tế giảm. Các đạo hàm được tính toán một cách phân tích thông qua quy tắc chuỗi, ví dụ trong đó

$$\frac{\partial \mathbf{p}_{ij}^k}{\partial \theta_{i1}} = \frac{\partial \mathbf{p}_{ij}^k}{\partial \tilde{\mathbf{p}}_{ij}^k} \frac{\partial \tilde{\mathbf{p}}_{ij}^k}{\partial \theta_{i1}}$$

và

$$\frac{\partial \mathbf{p}_{ij}^k}{\partial \tilde{\mathbf{p}}_{ij}^k} = \frac{\partial \begin{bmatrix} x/z & y/z \end{bmatrix}}{\partial \begin{bmatrix} x & y & z \end{bmatrix}} = \begin{bmatrix} 1/z & 0 & -x/z^2 \\ 0 & 1/z & -y/z^2 \end{bmatrix}$$

#### 4. Tự động làm thẳng ảnh panorama:

$$\begin{aligned} \frac{\partial \tilde{\mathbf{p}}_{ij}^k}{\partial \theta_{i1}} &= \mathbf{K}_i \frac{\partial \mathbf{R}_i}{\partial \theta_{i1}} \mathbf{R}_j \mathbf{K}_j^{-1} \tilde{\mathbf{u}}_j^l \\ \frac{\partial \mathbf{R}_i}{\partial \theta_{i1}} &= \frac{\partial}{\partial \theta_{i1}} e^{[\boldsymbol{\theta}_i]_\times} = e^{[\boldsymbol{\theta}_i]_\times} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \end{aligned}$$

Ghép hình ảnh bằng cách sử dụng các bước trong phần 1 - 3 sẽ cung cấp các góc quay tương đối giữa các camera, nhưng vẫn còn một góc xoay 3D không xác định đối với khung tọa độ đã chọn. Nếu chúng ta chỉ giả sử rằng  $\mathbf{R} = \mathbf{I}$  cho một trong các hình ảnh, chúng ta thường thấy hiệu ứng lượn sóng trong ảnh toàn cảnh đầu ra. Điều này là do máy ảnh thật khó có thể ở mức hoàn hảo và không bị nghiêng. Chúng ta có thể sửa kết quả gợn sóng này và tự động làm thẳng ảnh toàn cảnh bằng cách sử dụng phương pháp phỏng đoán về cách mọi người thường chụp ảnh toàn cảnh. Ý tưởng là hiếm khi người ta xoay camera so với đường chân trời, do đó các vector  $\mathbf{X}$  của camera (trục ngang) thường nằm trong một mặt phẳng. Bằng cách tìm vector null của ma trận hiệp phương sai của vector camera  $\mathbf{X}$ , chúng

ta có thể tìm được “vector hướng lên trên”  $\mathbf{u}$  (bình thường đối với mặt phẳng chứa tâm camera và đường chân trời)

$$\left( \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right) \mathbf{u} = \mathbf{0}.$$

Áp dụng xoay toàn cục sao cho vector  $\mathbf{u}$  thẳng đứng (trong khung kết xuất) sẽ loại bỏ hiệu quả hiệu ứng gợn sóng khỏi ảnh toàn cảnh đầu ra.

## 5. Gain compensation

Trong các phần trước, chúng ta đã mô tả phương pháp tính toán các thông số hình học (hướng và tiêu cự) của mỗi camera. Trong phần này, chúng ta trình bày cách giải một tham số trắc quang, cụ thể là mức tăng tổng thể giữa các hình ảnh. Điều này được thiết lập theo cách tương tự, với hàm lỗi được xác định trên tất cả các hình ảnh. Hàm lỗi là tổng các lỗi cường độ chuẩn hóa khuếch đại cho tất cả các pixel chồng chéo

$$e = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{\mathbf{u}_i \in \mathcal{R}(i,j) \\ \tilde{\mathbf{u}}_i = \mathbf{H}_{ij} \tilde{\mathbf{u}}_j}} (g_i I_i(\mathbf{u}_i) - g_j I_j(\mathbf{u}_j))^2$$

trong đó  $g_i, g_j$  là mức tăng và  $\mathcal{R}(i, j)$  là vùng chồng lấp giữa hình ảnh  $i$  và  $j$ . Trong thực tế, chúng tôi ước tính  $I(\mathbf{u}_i)$  theo giá trị trung bình trong từng vùng chồng lấp  $\bar{I}_{ij}$

Điều này giúp đơn giản hóa việc tính toán và mang lại sự chắc chắn cho các giá trị ngoại lệ, có thể phát sinh do sự đăng ký sai nhỏ giữa các hình ảnh. Ngoài ra, vì  $g = 0$  là nghiệm tối ưu của bài toán nên chúng ta thêm một số hạng trước để giữ độ lợi gần bằng đơn vị. Do đó hàm lỗi trở thành

$$\bar{I}_{ij} = \frac{\sum_{\mathbf{u}_i \in \mathcal{R}(i,j)} I_i(\mathbf{u}_i)}{\sum_{\mathbf{u}_i \in \mathcal{R}(i,j)} 1}.$$

trong đó  $N_{ij} = |\mathcal{R}(i, j)|$  bằng số pixel trong ảnh  $i$  chồng lên ảnh  $j$ . Các tham số  $\sigma_N$  và  $\sigma_g$  lần lượt là độ lệch chuẩn của sai số cường độ chuẩn hóa và mức tăng. Ta chọn các giá trị  $\sigma_N = 10,0$ , ( $I \in \{0..255\}$ ) và  $\sigma_g = 0,1$ . Đây là hàm mục tiêu bậc hai trong các tham số khuếch đại  $g$  có thể được giải ở dạng đóng bằng cách đặt đạo hàm về 0

$$e = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n N_{ij} ((g_i \bar{I}_{ij} - g_j \bar{I}_{ji})^2 / \sigma_N^2 + (1 - g_i)^2 / \sigma_g^2)$$

## 6. Multi-Band Blending

Từ các bước trước, chúng ta có  $n$  ảnh  $I^i(x, y)$  ( $i$  thuộc  $\{1..n\}$ ), với đăng ký đã biết, có thể được biểu diễn trong một hệ tọa độ (hình cầu) chung là  $I^i(\theta, \phi)$ . Để kết hợp thông tin từ nhiều ảnh, chúng ta gán hàm trọng số cho mỗi ảnh  $W(x, y) = w(x)w(y)$  trong đó  $w(x)$  thay đổi tuyến tính từ 1 ở tâm ảnh đến 0 tại bờ rìa. Các hàm trọng số cũng được lấy mẫu lại theo tọa độ cầu  $W^i(\theta, \phi)$ . Một cách tiếp cận đơn giản để hòa trộn là thực hiện tổng cường độ hình ảnh có trọng số dọc theo mỗi tia bằng cách sử dụng các hàm trọng số này trong đó  $I^{linear}(\theta, \phi)$  là một hình ảnh hình cầu tổng hợp được hình thành bằng cách trộn tuyến tính. Tuy nhiên, cách tiếp cận này có thể gây mờ chi tiết tần số cao nếu có lỗi nhỏ. Để ngăn chặn điều này, chúng ta sử dụng thuật toán trộn đa băng tần của Burt và Adelson. Ý tưởng đằng sau việc trộn nhiều băng tần là trộn các tần số thấp trên một phạm vi không gian rộng và các tần số cao trong một phạm vi ngắn.

$$I^{linear}(\theta, \phi) = \frac{\sum_{i=1}^n I^i(\theta, \phi) W^i(\theta, \phi)}{\sum_{i=1}^n W^i(\theta, \phi)}$$

Chúng ta khởi tạo các trọng số hòa trộn cho mỗi hình ảnh bằng cách tìm tập hợp các điểm mà hình ảnh 'i' chịu trách nhiệm cao nhất

$W_{max}^i(\theta, \phi)$  là 1 cho các giá trị  $(\theta, \phi)$  trong đó hình ảnh  $i$  có trọng số tối đa và 0 khi một số hình ảnh khác có trọng số cao hơn. Các bản đồ trọng lượng tối đa này được làm mờ liên tục để tạo thành các trọng số hòa trộn cho mỗi dải.

$$W_{max}^i(\theta, \phi) = \begin{cases} 1 & \text{if } W^i(\theta, \phi) = \arg \max_j W^j(\theta, \phi) \\ 0 & \text{otherwise} \end{cases}$$

Một phiên bản cao cấp của hình ảnh được hiển thị được hình thành, trong đó  $g_\sigma(\theta, \phi)$  là một Gaussian có độ lệch chuẩn  $\sigma$ , và toán tử  $*$  biểu thị tích chập.  $B_\sigma(\theta, \phi)$  biểu thị tần số không gian trong phạm vi bước sóng  $\lambda \in [0, \sigma]$ . Chúng ta trộn dải này giữa các hình ảnh bằng cách sử dụng trọng lượng hòa trộn được hình thành bằng cách làm mờ bản đồ trọng lượng tối đa cho hình ảnh này

$$\begin{aligned} B_{\sigma}^i(\theta, \phi) &= I^i(\theta, \phi) - I_{\sigma}^i(\theta, \phi) \\ I_{\sigma}^i(\theta, \phi) &= I^i(\theta, \phi) * g_{\sigma}(\theta, \phi) \end{aligned}$$

trong đó  $W_{\sigma}^i(\theta, \phi)$  là trọng số hòa trộn của dải bước sóng  $\lambda \in [0, \sigma]$ . Các dải tần số tiếp theo được pha trộn bằng cách sử dụng hình ảnh băng thông tần số thấp hơn và làm mờ thêm các trọng số hòa trộn, tức là với  $k \geq 1$

$$W_{\sigma}^i(\theta, \phi) = W_{max}^i(\theta, \phi) * g_{\sigma}(\theta, \phi)$$

trong đó độ lệch chuẩn của hạt nhân làm mờ Gaussian  $\sigma' = \sqrt{2k+1} * \sigma$  được đặt sao cho các dải tiếp theo có cùng phạm vi bước sóng.

$$\begin{aligned} B_{(k+1)\sigma}^i &= I_{k\sigma}^i - I_{(k+1)\sigma}^i \\ I_{(k+1)\sigma}^i &= I_{k\sigma}^i * g_{\sigma'} \\ W_{(k+1)\sigma}^i &= W_{k\sigma}^i * g_{\sigma'} \end{aligned}$$

Đối với mỗi dải, các hình ảnh chồng chéo được kết hợp tuyến tính bằng cách sử dụng các trọng số hòa trộn tương ứng. Điều này làm cho các dải tần số cao ( $k\sigma$  nhỏ) bị trộn lẫn trên các dải ngắn trong khi các dải tần số thấp ( $k\sigma$  lớn) được trộn lẫn trên các dải lớn hơn.

$$I_{k\sigma}^{multi}(\theta, \phi) = \frac{\sum_{i=1}^n B_{k\sigma}^i(\theta, \phi) W_{k\sigma}^i(\theta, \phi)}{\sum_{i=1}^n W_{k\sigma}^i(\theta, \phi)}.$$

Lưu ý rằng chúng ta đã chọn hiển thị ảnh toàn cảnh theo tọa độ cầu  $\theta, \phi$ . Về nguyên tắc, người ta có thể chọn bất kỳ tham số hóa 2 chiều nào của bề mặt xung quanh điểm nhìn để hiển thị. Một lựa chọn tốt là kết xuất thành một hình cầu tam giác, xây dựng các trọng số hòa trộn trong mặt phẳng hình ảnh. Điều này sẽ có lợi thế là xử lý thống nhất tất cả các hình ảnh và cũng cho phép lấy mẫu lại dễ dàng sang các bề mặt khác (trong phần cứng đồ họa). Lưu ý rằng tham số hóa  $\theta, \phi$  có điểm kỳ dị ở các cực.

**Ưu điểm của Ghép ảnh Toàn cảnh với Đặc điểm Bất biến:**

- Chịu được thay đổi: Không nhạy cảm với sự thay đổi về góc nhìn, ánh sáng và biến dạng nhỏ của ảnh.



- Tự động căn chỉnh: Không cần thao tác thủ công để căn chỉnh ảnh, khiến nó thân thiện với người dùng.
- Độ chính xác cao: Khớp và ghép ảnh chính xác để tạo ra toàn cảnh chất lượng cao.
- Đầu vào linh hoạt: Có thể xử lý nhiều ảnh và các cài đặt camera khác nhau.

#### Thách thức:

- Chi phí tính toán: Phát hiện và khớp đặc điểm có thể đòi hỏi nhiều tính toán, đặc biệt đối với ảnh lớn.
- Sự mơ hồ khi khớp: Các mẫu lặp lại hoặc các đặc điểm tương tự nhau có thể dẫn đến khớp sai.
- Độ phức tạp của cảnh: Các vật thể di chuyển hoặc lỗi parallax có thể gây ra các vết ghép trong ảnh.

Nhìn chung, ghép ảnh toàn cảnh bằng các đặc điểm bất biến là một công cụ mạnh mẽ để tạo ra những hình ảnh toàn cảnh ấn tượng với sự can thiệp tối thiểu của người dùng. Phương pháp này được sử dụng rộng rãi trong nhiều ứng dụng, bao gồm nhiếp ảnh, thực tế ảo và robot. Với nghiên cứu và phát triển liên tục, kỹ thuật này đang trở nên ngày càng mạnh mẽ và thân thiện với người dùng, giúp nó dễ tiếp cận hơn với nhiều đối tượng.

## Chapter 4 : Triển khai và thử nghiệm

### 1. Cài đặt các thư viện cần thiết:

```
1 import numpy as np
2 import imutils
3 import cv2
```

Hình 3: Import các thư viện cần thiết

### 2. Triển khai các hàm:

- Hàm **image\_stitch()**:
  - ⇒ Chụp hai ảnh đầu vào (imageA và imageB).
  - ⇒ Phát hiện các đặc điểm và điểm chính bằng thuật toán SIFT cho cả hai hình ảnh.
  - ⇒ Khớp các điểm chính giữa hai hình ảnh bằng phương pháp matchKeypoints.
  - ⇒ Tính toán homograph bằng thuật toán RANSAC để tìm ma trận biến đổi.

- Hàm **getwarp\_perspective()**:
  - ⇒ Làm cong imageA để căn chỉnh với imageB bằng cách sử dụng hình ảnh tương đồng được tính toán.
  - ⇒ Nếu match\_status là True, nó sẽ hiển thị các điểm chính phù hợp.
- Hàm **Detect\_Feature\_And\_KeyPoints()**:
  - ⇒ Chuyển đổi hình ảnh đầu vào sang thang độ xám.
  - ⇒ Áp dụng thuật toán SIFT để phát hiện các điểm chính và tính toán các bộ mô tả.
  - ⇒ Trả về các điểm chính và mô tả.
- Hàm **get\_Allpossible\_Match()**: Tính toán tất cả các kết quả khớp giữa hai bộ mô tả bằng cách sử dụng trình so khớp Brute-Force.
- Hàm **All\_validmatches()**: Lọc các kết quả phù hợp dựa trên kiểm tra tỷ lệ của Lowe.
- Hàm **Compute\_Homography()**: Tính toán ma trận tương đồng bằng RANSAC.
- Hàm **matchKeypoints()**:
  - ⇒ Sử dụng **get\_Allpossible\_Match** và **All\_validmatches** để nhận kết quả khớp hợp lệ.
  - ⇒ Xây dựng tập hợp các điểm tương ứng từ các kết quả khớp hợp lệ.
  - ⇒ Tính toán tương đồng bằng RANSAC.
  - ⇒ Trả về các kết quả khớp hợp lệ, ma trận đồng nhất và trạng thái.
- Hàm **getwarp\_perspective()**: Sử dụng hàm **cv2.warpPerspective** để làm cong imageA dựa trên hình ảnh đồng nhất được tính toán.
- Hàm **get\_image\_dimension()**: Trả về kích thước (chiều cao và chiều rộng) của hình ảnh.
- Hàm **get\_points()**: Kết hợp hai hình ảnh cạnh nhau thành một hình ảnh duy nhất.
- Hàm **draw\_Matches()**:
  - ⇒ Hiển thị sự trùng khớp giữa các điểm chính trên hình ảnh được nối.
  - ⇒ Đọc ảnh và gọi hàm thực thi

3. Kết quả:

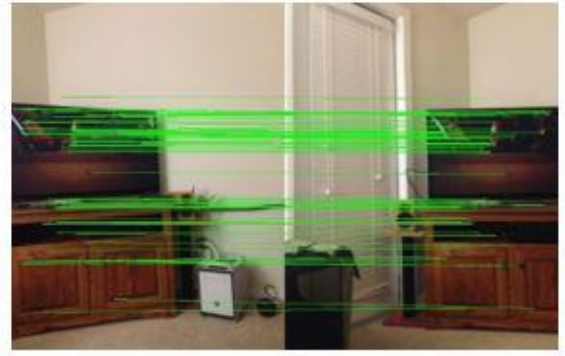
a. Kiểm thử với 2 ảnh:



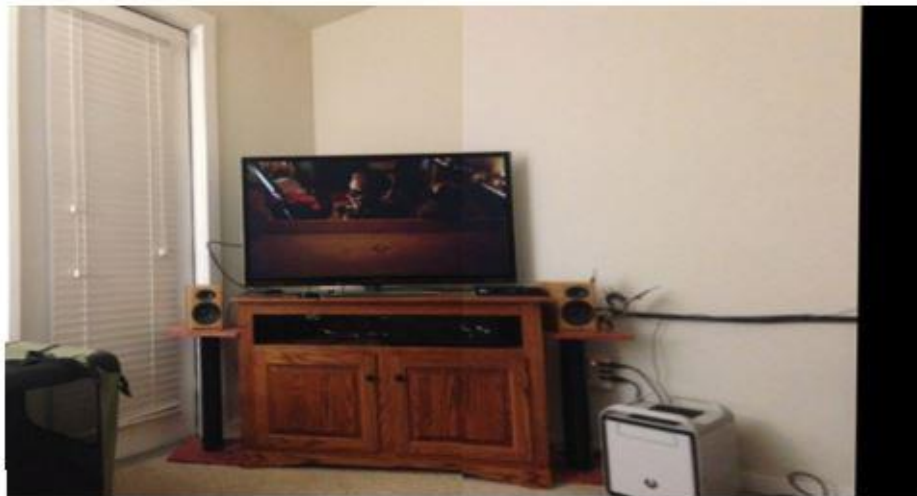
room1.jpg



room2.jpg



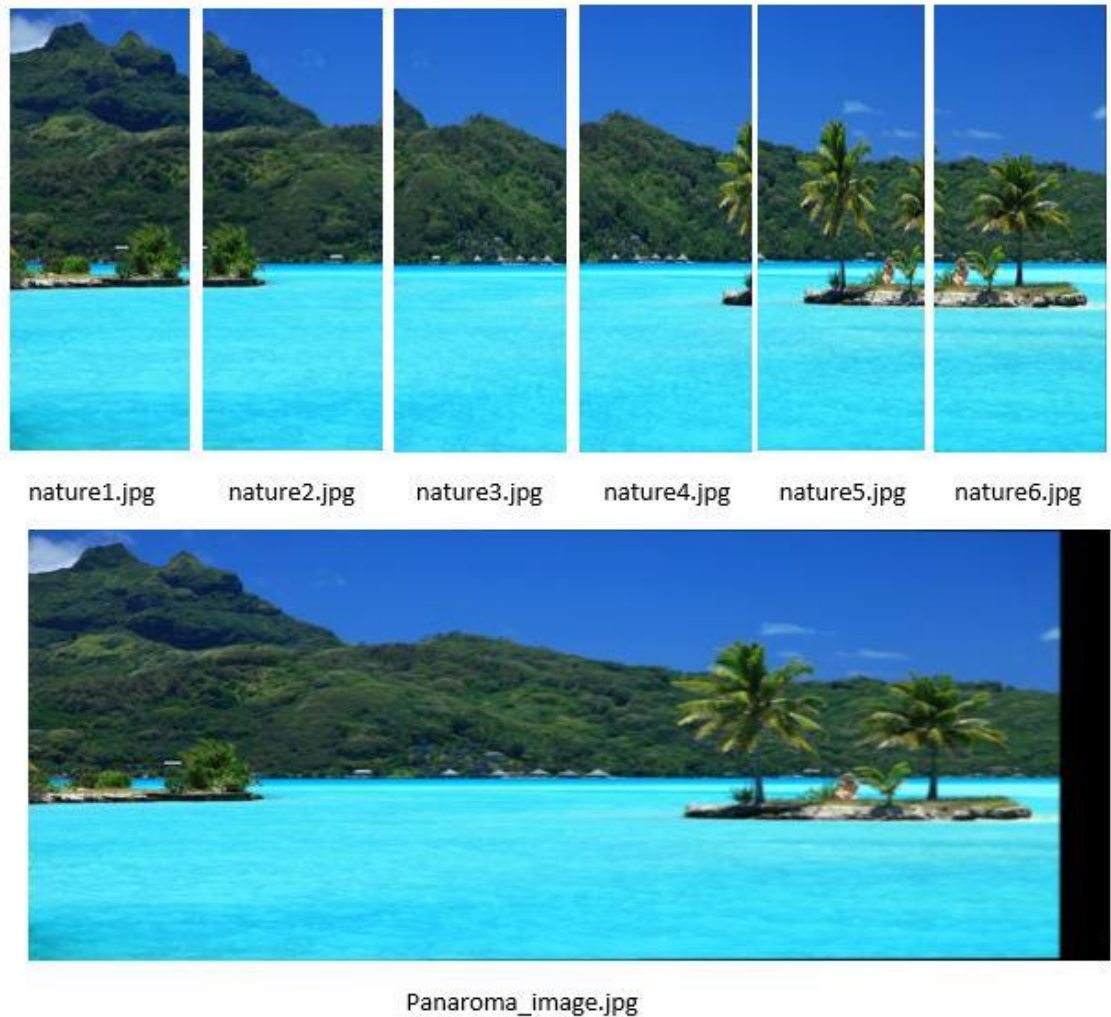
Matched\_points.jpg



Panorama\_image.jpg

Hình 4: So khớp hai ảnh và kết quả ghép hai ảnh

b. Kiểm thử với nhiều ảnh:



Hình 5: Ghép nhiều ảnh

## Chapter 5 : Kết luận và định hướng tương lai

### Conclusion

Phương pháp ghép ảnh toàn cảnh bằng các đặc điểm bất biến đã cho thấy đây là một phương tiện mạnh mẽ và hiệu quả để tạo ra những hình ảnh góc rộng mượt mà từ nhiều bức ảnh chồng lấn. Bằng cách sử dụng các đặc điểm bất biến chính, như những điểm được phát hiện thông qua các thuật toán SIFT, SURF, hoặc ORB, phương pháp này xuất sắc trong việc xác định và ghép các điểm đặc trưng ngay cả khi có sự thay đổi về góc nhìn, điều kiện ánh sáng, hoặc biến dạng nhỏ.

Quy trình này bao gồm việc phát hiện các đặc điểm, kết hợp các điểm tương ứng, ước tính biến đổi hình học và cuối cùng là biến đổi và pha trộn ảnh để tạo thành một bức tranh toàn cảnh liền mạch. Giai đoạn triển khai và kiểm thử đã cho thấy kết quả thành công, chứng minh khả năng của thuật toán tự động căn chỉnh và ghép ảnh, mang lại cho người dùng những bức ảnh toàn cảnh chất lượng cao.

## Future Directions

Mặc dù bản triển khai hiện tại đã cho thấy những kết quả tích cực, nhưng vẫn có nhiều hướng để nghiên cứu và cải thiện trong tương lai:

- **Tối ưu hóa khả năng tính toán:** Nghiên cứu các phương pháp để tăng cường hiệu năng tính toán của thuật toán phát hiện và kết hợp đặc điểm, đặc biệt là đối với ảnh lớn hoặc ứng dụng thời gian thực.
- **Xử Lý Cảnh Động:** Phát triển các kỹ thuật để đối mặt với thách thức của cảnh động, nơi các đối tượng di chuyển giữa các khung hình, gây ra lỗi parallax. Điều này có thể liên quan đến việc tích hợp các phương pháp nhận diện và theo dõi đối tượng.
- **Nâng Cao Chất Lượng Ghép Ảnh:** Khám phá các kỹ thuật pha trộn tiên tiến để cải thiện chất lượng của các bức tranh toàn cảnh, đặc biệt là ở các khu vực có biến động về ánh sáng hoặc khác biệt về màu sắc.
- **Tích Hợp Thực Tế Ảo:** Nghiên cứu tích hợp các bức tranh toàn cảnh vào các ứng dụng thực tế ảo (AR), mở rộng phạm vi sử dụng ngoài lĩnh vực nhiếp ảnh truyền thống.
- **Ghép Ảnh Đa Dạng:** Mở rộng phương pháp để xử lý ghép ảnh từ nhiều nguồn và chế độ chụp khác nhau, ví dụ như kết hợp ảnh ánh sáng nhìn thấy với ảnh hồng ngoại hoặc chiều sâu.
- **Tích Hợp Học Máy:** Khảo sát việc tích hợp các kỹ thuật học máy để tự động nhận diện và sửa lỗi ghép ảnh, làm tăng khả năng tổng thể của quá trình ghép ảnh.
- **Xử Lý Thời Gian Thực:** Hướng tới khả năng xử lý thời gian thực hoặc gần thời gian thực, làm cho hệ thống ghép ảnh toàn cảnh trở nên nhanh chóng và thân thiện với người dùng hơn.

→ Tóm lại, phương pháp ghép ảnh toàn cảnh sử dụng các đặc điểm bất biến đã chứng minh được hiệu quả, và những nỗ lực nghiên cứu và phát triển tiếp tục có thể đóng góp vào sự cải thiện liên tục, thích ứng với nhiều ứng dụng khác nhau trong lĩnh vực thị giác máy tính ngày càng phát triển.

END