

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA KHOA HỌC MÁY TÍNH



BÁO CÁO ĐỒ ÁN

MÔN NHẬP MÔN THỊ GIÁC MÁY TÍNH

ĐỀ TÀI

ỨNG DỤNG THỊ GIÁC MÁY TÍNH THỰC HIỆN FACE SWAPPING VIDEO

Giảng viên hướng dẫn:

TS. Nguyễn Vinh Tiệp

Sinh viên thực hiện:

Nguyễn Tấn Nghĩa – 19520774

Cùng các thành viên khác trong nhóm:

Thái Trung Hiếu – 17520060

Nguyễn Hoàng Tuấn – 20520344

Lê Viết Lâm Quang – 20520290

Tp. Hồ Chí Minh, ngày 19 tháng 01 năm 2023

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành và sự tri ân sâu sắc đối với Thầy Nguyễn Vinh Tiệp – giảng viên đứng lớp của lớp CS231.N11 – và cũng là giảng viên hướng dẫn của em trong môn học Nhập môn Thị giác máy tính nhiệt tình hướng dẫn em hoàn thành tốt đồ án môn học này.

Trong quá trình học tập và làm việc, cũng như là trong quá trình làm bài báo cáo đồ án, khó tránh khỏi sai sót, rất mong Thầy bỏ qua. Đồng thời do trình độ lý luận cũng như kinh nghiệm thực tiễn còn hạn chế nên bài báo cáo không thể tránh khỏi những thiếu sót, em rất mong nhận được ý kiến đóng góp của Thầy để em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn các bài báo cáo trong đoạn đường sắp tới.

Em xin chân thành cảm ơn!

Tp. Hồ Chí Minh, ngày 19 tháng 01 năm 2023

NHẬN XÉT CỦA GIẢNG VIÊN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

MỤC LỤC

LỜI CẢM ON.....	2
NHẬN XÉT CỦA GIẢNG VIÊN	3
MỤC LỤC.....	4
DANH MỤC BẢNG BIỂU, SƠ ĐỒ, HÌNH ẢNH	5
1. Giới thiệu	6
2. Phương pháp tiếp cận	6
2.1. Thư viện dlib và mô hình nhận dạng 68 landmarks khuôn mặt.....	6
2.2. Phương pháp chia lưới căn bản.....	6
2.2.1. Detect các điểm mốc dựa trên bộ tham số	7
2.2.2. Facing cropping.....	8
2.2.3. Matching facial	8
2.2.4. Face swapping.....	10
2.2.5. Face cloning	11
2.3. Áp dụng phương pháp trên cho video.....	12
3. Đánh giá kết quả, kết luận.....	12
3.1. Nhận xét ưu nhược điểm.....	12
3.1.1. Ưu điểm	12
3.1.2. Nhược điểm.....	12
3.2. So sánh với vài phương pháp khác của thành viên trong nhóm	13
3.3. Đề xuất hướng phát triển.....	13
4. Tài liệu tham khảo	14

DANH MỤC BẢNG BIỂU, SƠ ĐỒ, HÌNH ẢNH

Hình 1 - Danh sách vị trí 68 landmark	7
Hình 2 - Ảnh source khi được tạo mạng lưới các tam giác	8
Hình 3 - Ảnh dest khi được tạo mạng lưới các tam giác	8
Hình 4 - Mô tả biến đổi Affine	9
Hình 5 - Các ma trận biến đổi Affine	9
Hình 6 - Bao lồi khuôn mặt ảnh source ban đầu	10
Hình 7 - Bao lồi khuôn mặt ảnh dest	10
Hình 8 - Bao lồi khuôn mặt ảnh source sau khi biến đổi	10
Hình 9 - Ảnh minh họa quá trình swapping	10
Hình 10 - Kết quả thu được sau bước swapping	11
Hình 11 - Kết quả sau khi áp dụng cloning	12
 Bảng 1 - Hình ảnh detect 68 landmark ở ảnh source và ảnh dest	7

1. Giới thiệu

Với sự phát triển của thời đại 4.0, các kỹ thuật trong ngành công nghệ thông tin đã đạt được những tiến bộ vượt bậc mà con người cách đó vài chục năm không hề tưởng tượng ra, có thể nêu ra những thành tựu điển hình như: các phương tiện không người lái (xe hơi, máy bay, ...), nhận diện khuôn mặt cho các giao dịch, xác minh trong đời sống, điều phối đèn giao thông tự động,...

Và tất cả những ứng dụng kể trên chỉ là một trong những lĩnh vực của trụ cột trí tuệ nhân tạo trong thời đại 4.0: *Thị giác máy tính*. Và một lĩnh vực không kém cạnh trong việc ứng dụng thành tựu của Thị giác máy tính đó là điện ảnh. Gần như những bộ phim đình đám, bom tấn thị trường đều có mặt rất nhiều của các kỹ xảo điện ảnh ứng dụng Thị giác máy tính. Và có một thể loại phim cũng đang rất được ưa chuộng đó là những phim kể về cuộc đời của các nhân vật nổi tiếng có sức ảnh hưởng lớn. Tuy nhiên đa phần các phim những nhân vật được kể đều không còn nữa, và diễn viên thì được tuyển chọn hoá trang để đóng nên chúng ta vẫn có cảm giác không thân thuộc cho lắm. Và nhóm em nghĩ xu hướng ứng dụng kỹ xảo chuyển đổi khuôn mặt trong các bộ phim nói trên sẽ đem lại nhiều hiệu ứng cũng như kết quả tốt hơn trong sự phát triển thể loại phim này. Vì thế, trong đồ án môn *Nhập môn Thị giác máy tính* lần này, nhóm chúng em thực hiện đề tài chuyển đổi khuôn mặt trên video với mục đích tìm hiểu cũng như hiện thực những suy nghĩ của nhóm.

2. Phương pháp tiếp cận

2.1. Thư viện dlib và mô hình nhận dạng 68 landmarks khuôn mặt

Dlib là thư viện mã nguồn mở về Machine Learning, trong đó nổi tiếng nhất là Deep Learning. Về bản chất ban đầu nó là thư viện C++ cài đặt nhiều giải thuật xử lý ảnh cho các bài toán khác nhau, trong đó có bài toán nhận diện khuôn mặt. Và trong bài toán hoán đổi khuôn mặt này, đi kèm với thư viện hỗ trợ dlib, chúng ta sẽ sử dụng thêm model nhận diện 68 landmarks trên khuôn mặt. Model này sẽ kết hợp với thư viện dlib để thực hiện nhận diện các điểm landmarks và thao tác trên chúng.

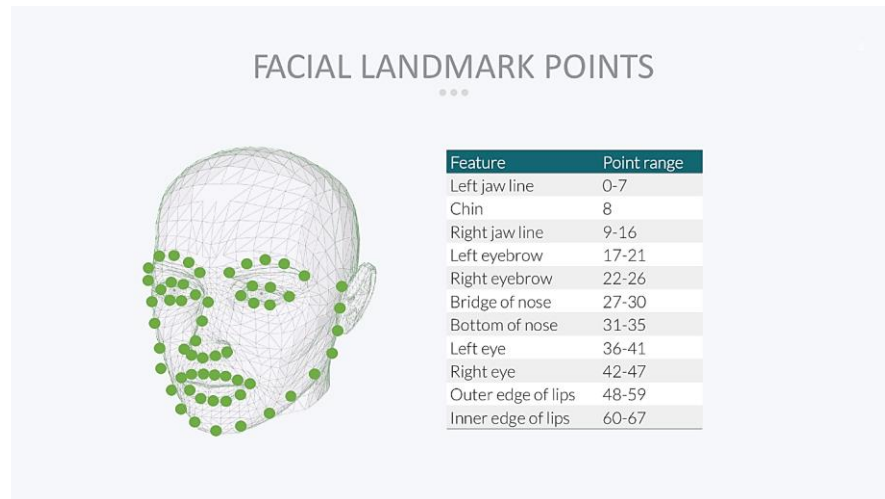
2.2. Phương pháp chia lưới căn bản

Trong phần trình bày, chúng ta gọi ảnh chứa khuôn mặt làm nguồn để thay thế là ảnh source và ảnh đích để ghép khuôn mặt từ ảnh source vào là ảnh dest (destination).

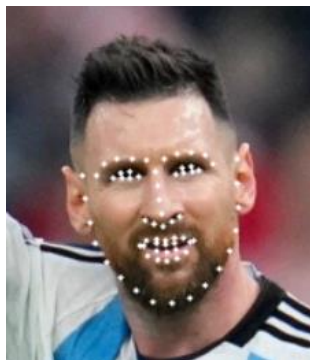
Phương pháp được tiến hành theo các bước sau đây:

2.2.1. Detect các điểm mốc dựa trên bộ tham số

- Sử dụng thư viện dlib để detect được khuôn mặt trong hình ảnh, sau đó tiến hành dự đoán các điểm landmark trên khuôn mặt vừa detect được.
- Việc xác định các điểm landmark cần đưa vào bộ tham số 68 điểm landmark để mô hình dự đoán. Việc xác định này được tiến hành trên ảnh source và ảnh dest.



Hình 1 - Danh sách vị trí 68 landmark



Bảng 1 - Hình ảnh detect 68 landmark ở ảnh source và ảnh dest

2.2.2. Facing cropping

- Từ các điểm mốc được xác định ta vẽ các đường bao lồi ở tập hợp điểm mỗi ảnh.
- Tạo lưới các tam giác từ mỗi 3 điểm mốc đã xác định (hàm danh sách các tam giác được hỗ trợ bởi hàm built-in của thư viện dlib, kết quả nhận được là danh sách vector 6 chiều chứa tọa độ của các điểm mỗi tam giác trong mạng lưới)
- Tiến hành cắt các đường bao lồi ở mỗi ảnh.



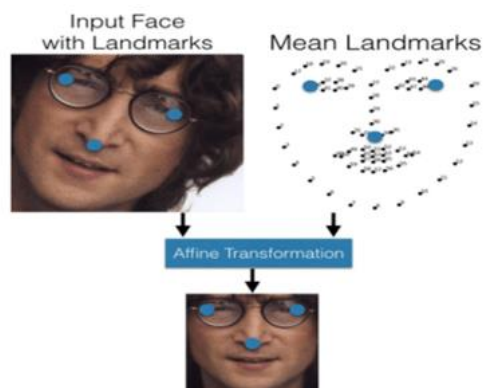
Hình 2 - Ảnh source khi được tạo mạng lưới các tam giác



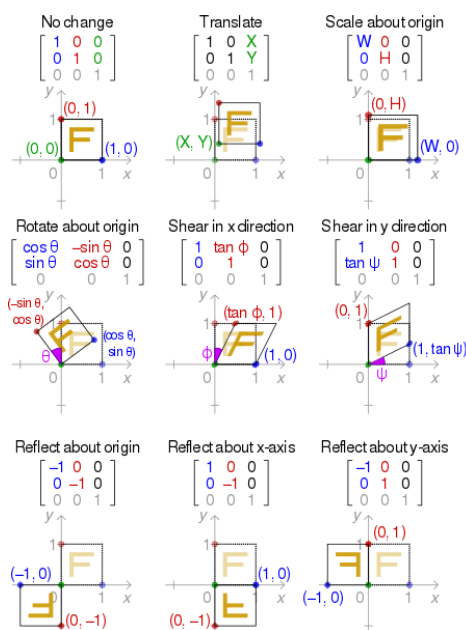
Hình 3 - Ảnh dest khi được tạo mạng lưới các tam giác

2.2.3. Matching facial

- So khớp mạng lưới 2 khuôn mặt rồi tiến hành biến đổi từng tam giác ở khuôn mặt ảnh source phù hợp với tam giác tương ứng ở ảnh dest. Ở đây, chúng ta sử dụng phép biến đổi Affine để thực hiện.
 - Một biến đổi affine tương ứng với 1 phép nhân một ma trận và cộng thêm 1 vector (hiểu đơn giản bao gồm 1 biến đổi tuyến tính đi kèm 1 phép tịnh tiến). Biến đổi affine là tổng thể của 3 phép biến đổi cơ bản: phép biến đổi tỷ lệ, phép quay và phép tịnh tiến.



Hình 4 - Mô tả biến đổi Affine



Hình 5 - Các ma trận biến đổi Affine

- Vì lý do biến đổi affine là tổng hoà của các phép biến đổi tuyến tính và tịnh tiến nên có thể xem đó là khái niệm cơ bản trong xử lý ảnh cũng như đồ hoạ máy tính, các tính chất của nó gồm có bảo toàn đường thẳng và bảo toàn khoảng cách – tỷ lệ nên lựa chọn biến đổi affine vừa cơ bản vừa giữ được cấu trúc của ảnh.
- Tổng hợp lại các tam giác sau khi biến đổi ta được khuôn mặt ảnh source có hình dạng khá khớp với hình dạng của khuôn mặt ảnh dest.



Hình 6 - Bao lồi khuôn mặt ảnh source ban đầu



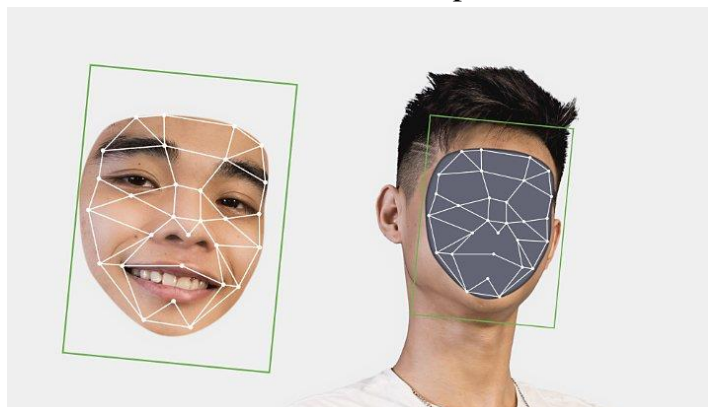
Hình 7 - Bao lồi khuôn mặt ảnh dest



Hình 8 - Bao lồi khuôn mặt ảnh source sau khi biến đổi

2.2.4. Face swapping

- Đầu tiên, cắt bỏ phần khuôn mặt ở ảnh dest.
- Thay thế khuôn mặt vừa biến đổi ở trên lên phần vừa cắt bỏ.



Hình 9 - Ảnh minh họa quá trình swapping

- Sau khi hoàn thành bước này, ta đã cơ bản hoàn thành việc thay đổi khuôn mặt, tuy nhiên kết quả không được đánh giá cao vì độ đồng màu không đều.



Hình 10 - Kết quả thu được sau bước swapping

2.2.5. Face cloning

- Tiến hành hoà trộn, cân bằng sắc độ, màu da, đường nét để giảm thiểu độ thô cứng cho ảnh vừa thay đổi khuôn mặt (phương pháp Seamless Cloning).
 - *Seamless Cloning (Poisson Cloning)* còn được gọi là xử lý hình ảnh miền gradient, là một kiểu xử lý ảnh kỹ thuật số hoạt động dựa trên sự khác biệt giữa các pixel lân cận, thay vì trực tiếp trên chính các giá trị pixel, chính vì tính hoạt động của kiểu xử lý này nên ta ứng dụng nó vào việc hoà trộn sắc độ sau khi ghép ảnh để hình ảnh được mượt mà và chân thực hơn (ngoài ra còn có nhiều phương pháp hoà trộn khác như Laplacian Pyramid Blending, Alpha Blending,...)



Hình 11 - Kết quả sau khi áp dụng cloning

2.3. Áp dụng phương pháp trên cho video

Khi tiến hành trên video, ta sẽ tiến hành biến đổi từng hình ảnh (frame) theo các bước đã cho, lặp qua tất cả frame có trong video và nối chúng lại với nhau, sau đó chọn FPS đầu ra tương ứng với FPS đầu vào và tiến hành export các frame theo số lượng FPS đã quy định. Ta sẽ thu được 1 video chuyển đổi khuôn mặt từ video gốc.

3. Đánh giá kết quả, kết luận

3.1. Nhận xét ưu nhược điểm

3.1.1. Ưu điểm

- Vì là phương pháp baseline cho các phương pháp khác trong nhóm nên phương pháp này được thực hiện dễ dàng.
- Việc kết hợp với thư viện dlib vô cùng mạnh mẽ đã giúp việc cài đặt dễ dàng hơn và hoạt động khá ổn định.
- Thu được kết quả khả quan khi ảnh dest có sai biệt góc độ khá lớn với ảnh source
- Không quá phức tạp cho quá trình kiểm lỗi cũng như nâng cấp.

3.1.2. Nhược điểm

- Vì thiết kế còn khá đơn giản nên phương pháp chưa mang đến hiệu năng cao.
- Hiệu quả thấp sẽ thấy rõ khi tiến hành trên ảnh source và ảnh dest có kích thước (độ phân giải) khác nhau
 - o Khi hai hình có kích thước khác nhau thì việc vỡ ảnh sẽ xảy ra làm giảm tính hiệu quả của phương pháp.

- Mất cân xứng nếu hai khuôn mặt có hình dạng và kích thước khác nhau.
 - o Việc 2 khuôn mặt có hình dạng khác nhau sẽ làm biến đổi sâu sắc ảnh source để phù hợp với hình dạng ảnh dest, do đó độ sai lệch khuôn mặt tỷ lệ thuận với độ sai lệch hình dạng 2 khuôn mặt (ví dụ khuôn mặt source là người mặt tròn, khuôn mặt dest là người mặt dài thì khi biến đổi ảnh source bị kéo dài cho phù hợp dẫn đến mất cân xứng tỷ lệ các bộ phận).
 - o Việc sai biệt về kích thước khuôn mặt tương tự việc sai biệt kích thước ảnh (ví dụ tỷ lệ khuôn mặt source là 30%, tỷ lệ khuôn mặt dest là 70% thì hiện tượng vỡ ảnh và việc phóng to ảnh source làm xuất hiện các đường răng cưa ở kết quả).

3.2. So sánh với vài phương pháp khác của thành viên trong nhóm

- Với phương pháp Resize Facial, phương pháp trên có lợi thế về mặt góc độ (khá thoải mái với sự sai lệch góc độ) và kích thước khuôn mặt dest (ở phương pháp trên có thể biến đổi thoải mái với mạng lưới tam giác riêng lẻ thay vì toàn bộ ma trận các điểm landmarks). Tuy nhiên, với phương pháp Resize Facial thì phương pháp trên cần kế thừa cách hoà trộn màu bằng cách kết hợp nhiều phương pháp phối trộn.
- Với phương pháp cải tiến Improving method, phương pháp trên có lợi thế về mặt tài nguyên khi tính toán góc độ (phương pháp cải tiến tốn nhiều tài nguyên để xác định góc độ và biến đổi, còn phương pháp căn bản thì biến đổi hình dạng từng tam giác tương ứng chứ không phải tốn tài nguyên tính toán góc độ). Tuy nhiên, ở đây phương pháp căn bản có thể phát triển bằng cách áp dụng phương pháp histogram matching và giữ các chi tiết đặc trưng của ảnh dest (mắt, miệng, chân mày,...) của phương pháp cải tiến.

3.3. Đề xuất hướng phát triển

- Nên sử dụng các ảnh thành phần có độ phân giải tương đối giống nhau.
- Sử dụng bộ tham số nhiều landmark hơn để giảm thiểu sai lệch cũng như giúp ảnh chi tiết hơn.
- Áp dụng các cải tiến hoà trộn cũng như việc giữ các chi tiết đặc trưng của ảnh.

4. Tài liệu tham khảo

- [1]. Luca96 (2018) “dlib-minified-models.” Available at:
<https://github.com/Luca96/dlib-minified-models> (Accessed: December 30, 2022).
- [2]. Mallick, S. (2016) “Face Swap using OpenCV (C++ / Python),”
LearnOpenCV, 5 April. Available at: <https://learnopencv.com/face-swap-using-opencv-c-python/> (Accessed: 2022).
- [3]. Mallick, S. (2015) “Seamless Cloning using OpenCV (Python , C++),”
LearnOpenCV, 2 March. Available at: <https://learnopencv.com/seamless-cloning-using-opencv-python-cpp/> (Accessed: January 3, 2023).
- [4]. MortezaNedaei (2019) “ Facial-Landmarks-Detection.” Available at:
<https://github.com/MortezaNedaei/Facial-Landmarks-Detection> (Accessed: December 30, 2022).
- [5]. Rosebrock, A. (2017) “Facial landmarks with dlib, OpenCV, and Python,”
pyimagesearch, 3 April. Available at: <https://pyimagesearch.com> (Accessed: December 25, 2022).