



**ĐỀ CƯƠNG KHÓA LUẬN TỐT NGHIỆP**  
**ĐỊNH DANH ĐỐI TƯỢNG TRONG VIDEO**  
**DÙNG ĐẶC TRƯNG KHUÔN MẶT 3D**  
*(Human identification in video using 3D facial features)*

## **1 THÔNG TIN CHUNG**

**Người hướng dẫn:**

– PGS. TS Lê Hoàng Thái (Khoa Công nghệ Thông tin)

**[Nhóm] Sinh viên thực hiện:**

1. Nguyễn Trọng Văn (MSSV: 1712202)

2. Lê Quang Vũ (MSSV: 1712209)

**Loại đề tài:** Nghiên cứu

**Thời gian thực hiện:** Từ 01/2021 đến 07/2021

## **2 NỘI DUNG THỰC HIỆN**

### **2.1 Giới thiệu về đề tài**

Định danh đối tượng luôn là một chủ đề nhận được sự quan tâm và nghiên cứu tích cực trong nhiều năm trở lại đây trên lĩnh vực về Khoa học máy tính, Thị giác máy tính và Sinh trắc học. Trong đó, nhận dạng khuôn mặt là một trong những

phương pháp định danh rất khổ biến bởi nhiều lợi ích về ứng dụng mà chúng mang lại thông qua việc có thể xác định, chứng thực được danh tính các đối tượng quan trọng với độ tin cậy và chính xác cao.

Tuy những giải pháp được đưa ra ngày càng cải thiện hơn về hiệu suất lẫn tốc độ thực thi của các hệ thống nhận dạng, nhưng vẫn tồn tại những rào cản về diện mạo, tư thế khuôn mặt, góc chụp, độ sáng, những vật chắn (kính, trang sức,...) vì đa phần các giải thuật và kiến trúc ngày nay đều nhận đầu vào là một ảnh chụp khuôn mặt 2D bình thường.

Chính vì thế, đề tài sẽ tập trung vào một cách tiếp cận mới trong định danh đối tượng thông qua nhận dạng khuôn mặt là thực hiện nhận dạng trên khuôn mặt được tái tạo 3D từ ảnh 2D trong video. Từ đó, xây dựng thực nghiệm, đánh giá độ chính xác và so sánh hiệu suất với các phương pháp tương tự.

## **2.2 Mục tiêu đề tài**

Đề tài tìm hiểu một giải pháp mới trong lĩnh vực nhận dạng. Cụ thể là nhận dạng mặt người từ khuôn mặt 3D được tái tạo từ các frame ảnh 2D trong video.

Việc tái tạo các khuôn mặt 3D sẽ giúp mang lại nhiều thông tin hơn cho quá trình học của các mô hình. Vượt qua được một số thách thức như: loại bỏ được ảnh hưởng do chiếu sáng, tập trung vào hình dáng khuôn mặt với các đặc trưng được bổ sung về chiều sâu (thay vì chỉ 2 chiều như trong ảnh 2D thông thường), giúp dự đoán các phần thông tin bị che khuất, hình dạng khuôn mặt từ những tư thế khác nhau từ đó cải thiện về độ chính xác trong quá trình chứng thực đối tượng.

Bên cạnh đó, đề tài sẽ tập trung xử lý định danh trên video với mục tiêu hướng đến là có thể nhận dạng trên thời gian thực vì phần lớn các ứng dụng ngày nay cần xử lý trên video nhiều hơn so với ảnh tĩnh, tiêu biểu như các hệ thống giám sát hành vi, đối tượng, các hệ thống điểm danh, chấm công, camera theo dõi,...

## **2.3 Phạm vi của đề tài**

Như đã trình bày bên trên, đề tài sẽ tập trung vào xử lý định danh đối tượng trong video. Chúng tôi giải quyết bài toán định danh này trong phạm vi về nhận dạng

mặt người. Các tập dữ liệu sẽ được lựa chọn để thực nghiệm bao gồm: FAMED<sup>1</sup> và YouTube Celebrities<sup>2</sup>. Đây là 2 tập dữ liệu video dành cho nhận dạng mặt người, cảm xúc và theo vết khuôn mặt. Các tập dữ liệu này cũng được các công trình nghiên cứu có liên quan sử dụng thực nghiệm, do đó có thể so sánh và đánh giá các kết quả cuối cùng đề tài thu được bởi các giải thuật khác.

Song song đó là các phương pháp Deep learning tham khảo dùng trong tái tạo 3D và phân lớp đang đạt được độ chính xác tốt và được công bố gần đây có thể dễ dàng thực hiện triển khai lại.

## 2.4 Cách tiếp cận dự kiến

Quá trình thực hiện đề tài dự kiến sẽ trải qua 3 giai đoạn (GD) chính:

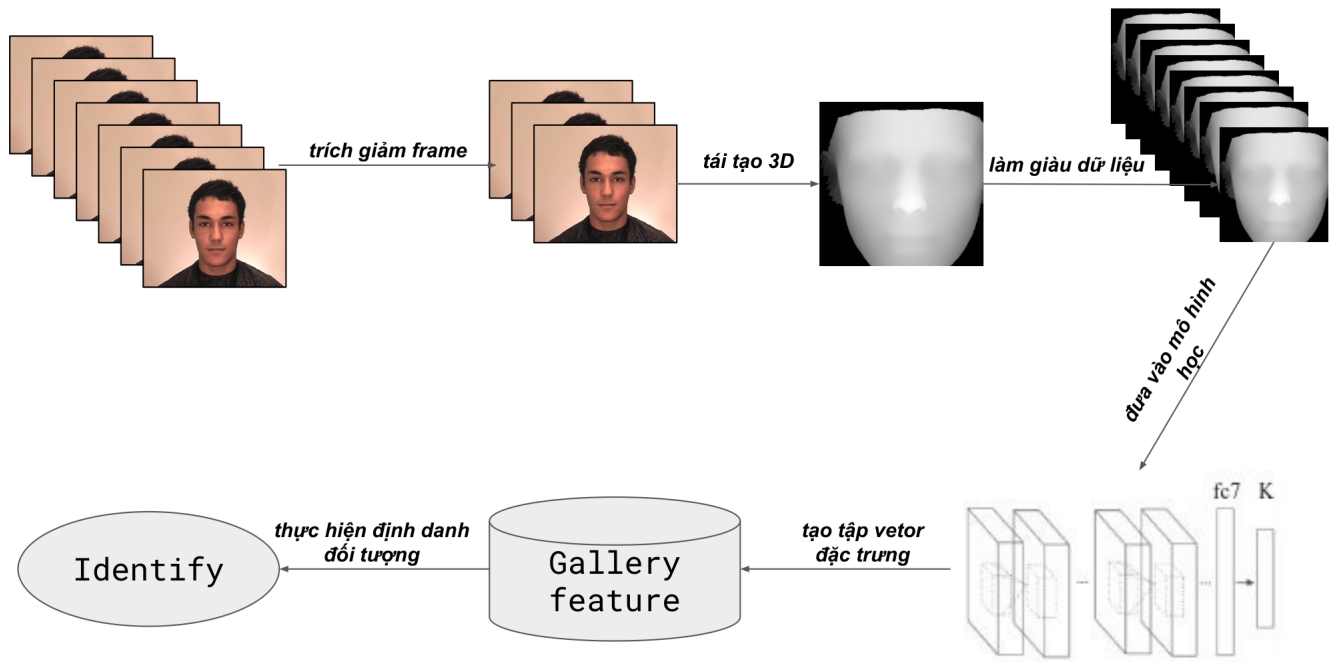
- GD1 - Chuẩn bị dữ liệu: Thực hiện rút trích giảm số lượng frame trên các video huấn luyện nhằm tránh làm dư thừa nhiều frame ảnh tương tự nhau, thực hiện tái tạo 3D khuôn mặt từ các frame ảnh, làm giàu dữ liệu trên các tư thế khuôn mặt và thêm nhiễu để làm phong phú tập dữ liệu. Ở giai đoạn này, chúng tôi sẽ thực hiện trên 2 tập dữ liệu video đã trình bày và kiến trúc mạng được lựa chọn để tái tạo 3D chúng tôi sẽ tham khảo kiến trúc của PRNet[1].
- GD2 - Huấn luyện mô hình: Đưa dữ liệu ở GD1 làm đầu vào huấn luyện mô hình học để lấy được trọng số tối ưu nhất trên tập dữ liệu ta có. Mô hình được lựa chọn để huấn luyện phân loại đối tượng là VGGFace16 như tham khảo ở công trình [2]. Sau khi có được trọng số tối ưu của mạng trên tập dữ liệu huấn luyện, chúng tôi thực hiện tạo tập vector đặc trưng (gallery feature) để lưu trữ đặc trưng khuôn mặt huấn luyện cho quá trình đối sánh, nhận dạng. Cụ thể, chạy lại dữ liệu huấn luyện trên trọng số có được, trích xuất vector đặc trưng ở layer cuối cùng trong kiến trúc mạng (fc7), thực hiện chuẩn hóa PCA vector đó để tạo thành tập vector đặc trưng của các lớp đưa vào huấn luyện. ([2])
- GD3 - Kiểm thử và đánh giá kết quả: Quá trình kiểm thử sẽ thực hiện dựa

---

<sup>1</sup><http://chrislongmore.co.uk/famed/>

<sup>2</sup><http://seqamlab.com/youtube-celebrities-face-tracking-and-recognition-dataset/>

trên độ đo tương đồng cosine giữa vector đầu vào đã được chuẩn hóa tương tự với quá trình tạo tập đặc trưng và các vector đặc trưng trong gallery feature. Nhận dạng đối tượng thông qua vector có độ tương đồng cao nhất với vector đầu vào. Kiểm thử trên tập dữ liệu kiểm tra và đánh giá hiệu suất từ thực nghiệm thu được, so sánh với các phương pháp khác.



Hình 1: Mô tả quy trình của phương pháp thực hiện đề tài.

## 2.5 Kết quả dự kiến của đề tài

Một hệ thống định danh đối tượng trong video với độ chính xác tốt hơn bằng phương pháp nhận dạng trên khuôn mặt 3D so với trên ảnh 2D tĩnh. Đồng thời, cũng mang lại kết quả khả quan hơn các phương pháp nhận dạng 3D khác trên cùng tập dữ liệu thực nghiệm.

Điều khác biệt so với các công trình đã được công bố và tham khảo là chúng tôi sẽ thực hiện việc tái tạo khuôn mặt dựa trên kiến trúc mạng học sâu PRNet thay vì các mô hình tĩnh đã quen thuộc (3DMM,...). Điều này cho ta sự tiếp cận mới mẻ hơn với các phương pháp tiên tiến hiện tại. Qua đó, phối hợp cùng với kiến

trúc mạng trong quá trình nhận dạng tạo thành một hệ thống nhận dạng sử dụng deep learning với độ tin cậy cao.

## 2.6 Kế hoạch thực hiện

Mỗi thành viên trong nhóm thực hiện đều phải hiểu rõ về đề tài công việc của bản thân cũng như của các thành viên khác. Công việc được chia làm 2 phần chính:

- Thực hiện rút trích frame ảnh, tái tạo khuôn mặt 3D và chuẩn bị dữ liệu đầu vào cho quá trình học:
  - Chịu trách nhiệm chính: Nguyễn Trọng Văn.
  - Tìm hiểu, hỗ trợ: Lê Quang Vũ.
- Xây dựng mô hình huấn luyện nhận dạng trên tập dữ liệu đầu vào:
  - Chịu trách nhiệm chính: Lê Quang Vũ.
  - Tìm hiểu, hỗ trợ: Nguyễn Trọng Văn.

Kế hoạch thực hiện:

Tháng 1/2021 - 2/2021	Tìm hiểu đề tài, tìm nguồn tài liệu tham khảo về kiến trúc mạng PRNet và VGG16.
Tháng 2/2021 - 4/2021	Thực hiện cài đặt tái tạo để chuẩn bị dữ liệu trên tập dữ liệu YouTube Celebrities, đưa vào quá trình học và kiểm tra đánh giá kết quả thực nghiệm, so sánh với các công trình liên quan.
Tháng 4/2021 - 6/2021	Thực hiện cài đặt tái tạo chuẩn bị dữ liệu trên tập dữ liệu FAMED, đưa vào quá trình học và kiểm tra đánh giá kết quả thực nghiệm, so sánh với các công trình liên quan.
Tháng 6/2021 - 7/2021	Viết báo cáo và chuẩn bị cho bảo vệ đề tài.

## Tài liệu

- [1] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, “Joint 3d face reconstruction and dense alignment with position map regression network,” *European Conference on Computer Vision (ECCV)*, pp. 534–551, 2018.
- [2] D. Kim, M. Hernandez, J. Choi, and GerardMedioni, “Deep 3d face identification,” *IEEE International Joint Conference on Biometrics (IJCB) 2017*, pp. 133–142, 2018.
- [3] K. T. Truong and T. H. Le, “Video-based face recognition using shape and texture information in 3d morphable model,” *JP Journal of Heat and Mass Transfer*, vol. 15, pp. 119–124, 2018.

XÁC NHẬN  
CỦA NGƯỜI HƯỚNG DẪN  
(Ký và ghi rõ họ tên)

TP. Hồ Chí Minh, ngày 05 tháng 03 năm 2021  
NHÓM SINH VIÊN THỰC HIỆN  
(Ký và ghi rõ họ tên)