

THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):

https://youtu.be/_TtGZYWLqMY

- Link slides (dạng .pdf đặt trên Github của nhóm):

<https://github.com/Ranisme/CS519.Q11.KHTN/blob/main/SLIDE.pdf>

- Họ và Tên: Nguyễn Thiện Nhân
- MSSV: 23521083



- Lớp: **CS519.Q11.KHTN**
- Tự đánh giá (điểm tổng kết môn): 9.5/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 9
- Số câu hỏi QT của cả nhóm: 18
- Link Github: <https://github.com/Ranisme>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
 - Đề xuất bài toán
 - Hội ý, hoàn thiện báo cáo, slide
 - Hoàn thiện làm video

- Họ và Tên: Bùi Ngọc Thiên Thanh
- MSSV: 23521436



- Lớp: **CS519.Q11.KHTN**
- Tự đánh giá (điểm tổng kết môn): 9.5/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 8
- Số câu hỏi QT của cả nhóm: 17
- Link Github: <https://github.com/Azure1709>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
 - Lên ý tưởng và hoàn thiện poster
 - Hội ý báo cáo, slide
 - Hỗ trợ làm video

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

NGHIÊN CỨU VÀ XÂY DỰNG HỆ THỐNG ẨN DANH HÓA ĐỐI TƯỢNG CÓ CHỌN LỌC TRONG VIDEO THỜI GIAN THỰC SỬ DỤNG MẶT NẠ KỸ THUẬT SỐ

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

RESEARCH AND DEVELOPMENT OF A SELECTIVE OBJECT ANONYMIZATION SYSTEM IN REAL-TIME VIDEO USING DIGITAL MASKS

TÓM TẮT

Hiện nay các buổi trực tiếp thực tế ngoài trời đang rất phát triển nhưng bên cạnh đó sự xuất hiện ngoài ý muốn của người lạ vào khung hình đang đặt ra những thách thức lớn về quyền riêng tư, tuy nhiên các giải pháp hiện hành vẫn chưa tối ưu. Những giải pháp, công cụ thiên về an ninh thường sẽ làm mờ (blur) hoặc pixel hóa toàn bộ khuôn mặt, gây mất thẩm mỹ và tạo cảm giác nặng nề làm giảm trải nghiệm người xem. Ngược lại, các filter giải trí làm đẹp thì lại quá tập trung vào chủ thể và bỏ qua việc xử lý các thông tin nền. Vì vậy nhóm nghiên cứu đề xuất giải pháp ẩn danh hóa người có chọn lọc, tập trung vào những người xung quanh. Thay thế khuôn mặt của những người đó bằng mặt nạ kỹ thuật số nhằm đảm bảo an toàn về thông tin cá nhân mà không làm giảm chất lượng của buổi phát sóng trực tiếp

Cụ thể, nghiên cứu bao gồm: (1) Nghiên cứu các giải pháp hiện đại như học sâu (Deep Learning) để giải quyết bài toán phát hiện đối tượng thời gian thực; (2) Nghiên cứu và đề xuất giải pháp định danh và theo dõi đa đối tượng (Multi-object Tracking) để phân loại tự động chủ thể và người lạ; (3) Xây dựng module để thay thế khuôn mặt của người lạ bằng mặt nạ kỹ thuật số.

Kết quả hướng tới một hệ thống có khả năng xử lý theo thời gian thực, đồng thời đảm bảo việc bảo vệ danh tính cho mọi người xung quanh mà không làm giảm trải nghiệm người xem.

GIỚI THIỆU (Tối đa 1 trang A4)

Bối cảnh và Tính cấp thiết: Lĩnh vực phát trực tiếp đang ngày càng phát triển, mở rộng cả về chủ đề, nội dung cũng như địa điểm quay. Đáng chú ý là xu hướng phát trực tuyến thực tế tại nơi công cộng (IRL Streaming). Đặc điểm của xu hướng này là sẽ được thực hiện ở ngoài trời, hàng quán, nơi công cộng; đặt ra những rủi ro vi phạm quyền riêng tư và gây khó chịu cho mọi người xung quanh [3]. Đặc biệt là trong bối cảnh cộng đồng đang ngày càng quan tâm hơn về việc dữ liệu sinh trắc học của mình sẽ được thu thập và sử dụng. Thực tế này đòi hỏi một giải pháp có thể giúp cân bằng giữa việc tự do sáng tạo và bảo vệ quyền riêng tư.

Hạn chế của những giải pháp hiện tại:

- Nhóm công cụ thiên về an ninh (Security-focused): Các công cụ hiện nay thường sử dụng các giải pháp biến dạng khuôn mặt như blur, pixel hóa [3]. Những giải pháp này tuy hiệu quả trong việc bảo mật nhưng lại làm giảm trải nghiệm người dùng nếu áp dụng vào nội dung phát trực tiếp. Điều này nghiêm trọng vì những nhà sáng tạo nội dung luôn muốn tối ưu trải nghiệm để gia tăng số lượng người xem.
- Nhóm các công cụ thiên về Giải trí (Entertainment-focused): Các lớp phủ (filter) này thường chỉ tập trung vào làm đẹp cho chủ thể nhưng lại bỏ qua những người khác ở hậu cảnh. Điều này không đảm bảo quyền riêng tư cho những người vô tình lọt vào ống kính.

Giải pháp đề xuất: Xây dựng một giải pháp có thể ẩn danh có chọn lọc: tập trung che người lạ và không làm ảnh hưởng đến chủ thể chính. Dựa trên các phương pháp học sâu hiện đại để giải quyết các bài toán như: phát hiện đối tượng thời gian thực [4], định danh và theo dõi đa đối tượng [5, 7]. Sau khi xác định được những người cần phải xử lý thì tiến hành phân tích thay thế bằng các mặt nạ kỹ thuật số sinh động, bám sát với biểu cảm của con người.

Mô tả bài toán:

- **Input:** Luồng video thời gian thực (Real-time video stream) chứa nhiều người.
- **Output:** Luồng video đã qua xử lý, trong đó khuôn mặt các đối tượng không

phải chủ thể được che phủ tự động. Hệ thống sử dụng các thuật toán phát hiện nhanh [4, 6] để định vị khuôn mặt và kỹ thuật so khớp đặc trưng [5] để phân loại đối tượng, đảm bảo mặt nạ bám sát theo chuyển động.

MỤC TIÊU (*Viết trong vòng 3 mục tiêu*)

1. Nghiên cứu cơ sở lý thuyết và công nghệ: Tìm hiểu và học cách áp dụng các phương pháp SOTA (State-of-the-art) của các bài toán liên quan như: phát hiện khuôn mặt, định danh, theo dõi đa đối tượng.
2. Xây dựng giải thuật và hệ thống: Đề xuất và xây dựng một hệ thống có thể tự động nhận diện và phân loại theo thời gian thực; đồng thời hoàn thiện module hiển thị mặt nạ kỹ thuật số tương ứng theo biểu cảm, chuyển động của khuôn mặt
3. Thực nghiệm và đánh giá: Đánh giá hiệu suất của hệ thống đề xuất trên các tiêu chí định lượng (tốc độ khung hình - FPS, độ trễ) và định tính (tính thẩm mỹ) so với các phương pháp làm mờ truyền thống.

NỘI DUNG VÀ PHƯƠNG PHÁP

Trong đề tài này, nhóm chúng tôi sẽ thực hiện nghiên cứu các nội dung chính sau:

- **Nội dung 1:** Tìm hiểu tổng quan đề tài
 - Phương pháp:
 - Tổng hợp và phân tích các tài liệu học thuật liên quan đến bảo vệ quyền riêng tư thị giác (Visual Privacy) .
 - Nghiên cứu và so sánh các hạn chế của các phương pháp ẩn danh hóa truyền thống (Làm mờ, Pixel hóa) so với các phương pháp hiện đại (Mask/Inpainting) [2, 3].
- **Nội dung 2:** Nghiên cứu cơ sở lý thuyết và các mô hình nền tảng
 - Phương pháp:
 - Khảo sát các phương pháp SOTA (State-of-the-art) từ các hội nghị hàng đầu (CVPR, ICCV,...) về các bài toán lõi: Phát hiện khuôn mặt (Face Detection) [4, 6], Trích xuất đặc trưng (Feature

Extraction) [5] và Theo dõi đa đối tượng (Multi-object Tracking).

- Tập trung phân tích các giải pháp hạng nhẹ, tối ưu [1] (MobileNetV3) . Để có thể chạy được ở nhiều thiết bị.

- **Nội dung 3:** Nghiên cứu phát triển các phương pháp mới

- Phương pháp:

- Dựa trên những gì đã nghiên cứu và tổng hợp. Tập trung nghiên cứu về phần hiệu năng, đề xuất các giải pháp để giải quyết đề theo dõi đa đối tượng và định danh. Tập trung vào hiệu năng xử lý để đáp ứng cho thời gian thực.
 - Xây dựng luồng xử lý (pipeline) tích hợp: Phát hiện – Theo dõi – Định danh.

- **Nội dung 4:** Xây dựng module hiển thị mặt nạ kỹ thuật số (Digital Masks):

- Phương pháp:

- Nghiên cứu và cài đặt thuật toán tính toán ma trận biến đổi (Homography/Affine transform) để xác định vị trí, kích thước và góc nghiêng của khuôn mặt.
 - Ứng dụng kỹ thuật ghép chồng hình ảnh (Image Blending) để thay thế khuôn mặt người lạ bằng các mặt nạ kỹ thuật số, đảm bảo độ bám sát theo chuyển động của đối tượng.

- **Nội dung 5:** Thực nghiệm và đánh giá hệ thống

- Phương pháp:

- Xây dựng tập dữ liệu kiểm thử (Test set): Tự thu thập dữ liệu video thực tế tại các môi trường công cộng (quán cà phê, công viên) với các điều kiện ánh sáng và mật độ người khác nhau.
 - Đánh giá định lượng: Đo lường hiệu năng hệ thống thông qua các chỉ số: Tốc độ khung hình (FPS), Độ trễ (Latency) và Độ chính xác của việc phân loại đối tượng.
 - Đánh giá định tính: So sánh tính thẩm mỹ và trải nghiệm người xem giữa phương pháp đề xuất và phương pháp làm mờ truyền

KẾT QUẢ MONG ĐỢI

1. Xây dựng thành công phần mềm có khả năng ẩn danh hóa đối tượng có chọn lọc. Hệ thống vận hành ổn định trên máy tính cá nhân và thực hiện chính xác chức năng phân loại chủ thể (streamer) và người lạ.
2. Đạt được hiệu năng xử lý tốt và đảm bảo tính thẩm mỹ: đạt độ chính xác cao, mặt nạ tạo cảm giác tự nhiên, thoải mái cho người xem.
3. Tập dữ liệu kiểm thử và báo cáo khoa học: xây dựng được tập dữ liệu kiểm thử hoạt động môi trường phát trực tiếp công cộng; Hoàn thiện báo cáo khoa học với đầy đủ cơ sở lý thuyết, quy trình xây dựng hệ thống và các kết quả đánh giá thực nghiệm định lượng, định tính.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

- [1]. Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, Hartwig Adam: Searching for MobileNetV3. ICCV 2019: 1314-1324
- [2]. Nicolai Hukkelås, Rudolf Mester, Frank Lindseth: DeepPrivacy: A Generative Adversarial Network for Face Anonymization. ISVC (2) 2019: 565-578
- [3]. José Ramón Padilla-López, Andréa A. Chaaraoui, Francisco Flórez-Revuelta: Visual privacy protection methods: A survey. Expert Syst. Appl. 42(9): 4177-4195 (2015)
- [4]. Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, Ali Farhadi: You Only Look Once: Unified, Real-Time Object Detection. CVPR 2016: 779-788
- [5]. Florian Schroff, Dmitry Kalenichenko, James Philbin: FaceNet: A Unified Embedding for Face Recognition and Clustering. CVPR 2015: 815-823
- [6]. Paul A. Viola, Michael J. Jones: Rapid Object Detection using a Boosted Cascade of Simple Features. CVPR (1) 2001: 511-518
- [7]. Nicolai Wojke, Alex Bewley, Dietrich Paulus: Simple Online and Realtime Tracking with a Deep Association Metric. ICIP 2017: 3645-3649

