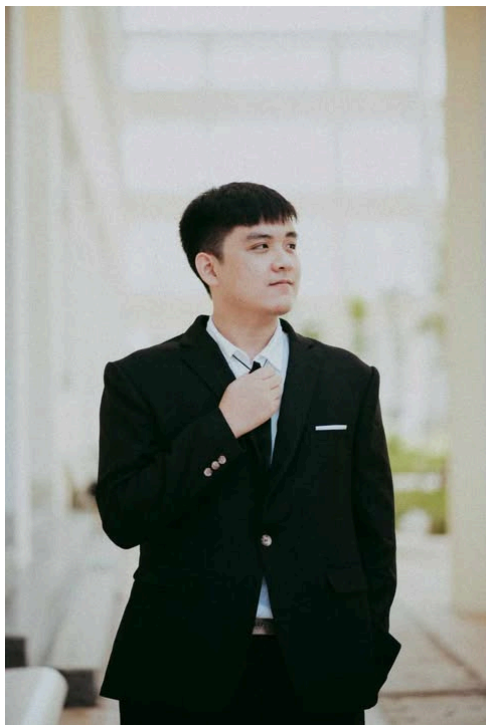


THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
<https://www.youtube.com/watch?v=6rFDN0k93BI>
- Link slides (dạng .pdf đặt trên Github của nhóm):
https://github.com/Truong99zvc/CS519.P11/blob/main/CS519_Proposal.pdf
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

<ul style="list-style-type: none">● Họ và Tên: Trương Phúc Trường● MSSV: 22521587 	<ul style="list-style-type: none">● Lớp: CS519.P11● Tự đánh giá (điểm tổng kết môn): 8.5/10● Số buổi vắng: 1● Số câu hỏi QT cá nhân: 6● Link Github: https://github.com/Truong99zvc/CS519.P11/● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none">○ Lên ý tưởng đề tài nghiên cứu○ Chỉnh sửa phần tóm tắt, giới thiệu.○ Viết phần nội dung và phương pháp, mục tiêu, kết quả mong đợi○ Làm và chỉnh sửa Poster○ Kiểm tra và chỉnh sửa Slide○ Kiểm tra và chỉnh sửa video YouTube
---	---

- Họ và Tên: Võ Đình Trung
- MSSV: 22521571



- Lớp: CS519.P11
- Tự đánh giá (điểm tổng kết môn): 8.5/10
- Số buổi vắng: 1
- Số câu hỏi QT cá nhân: 6
- Link Github:
<https://github.com/Truong99zvc/CS519.P11/>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
 - Lên ý tưởng đề tài nghiên cứu
 - Chỉnh sửa phần nội dung và phương pháp
 - Viết phần tóm tắt, giới thiệu
 - Làm Slide
 - Làm và chỉnh sửa Poster
 - Quay video YouTube

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

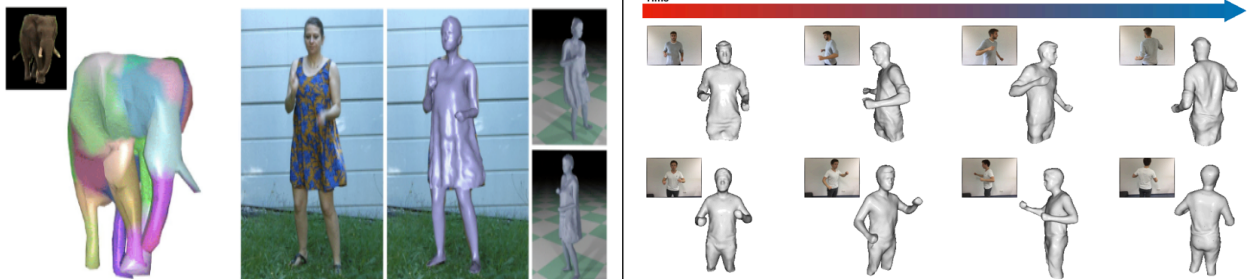
MỞ RỘNG KHẢ NĂNG TÁI TẠO 3D THỜI GIAN THỰC CỦA NEURALRECON TRONG MÔI TRƯỜNG ĐỘNG

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

EXTENDING NEURALRECON FOR REAL-TIME 3D RECONSTRUCTION IN DYNAMIC ENVIRONMENTS

TÓM TẮT (Tối đa 400 từ)

Nghiên cứu này đề xuất một phương pháp tái tạo 3D thời gian thực từ video đơn ảnh với khả năng xử lý cả môi trường tĩnh và các đối tượng động. Phương pháp này mở rộng “NeuralRecon” [1], vốn tập trung chủ yếu vào môi trường tĩnh, bằng cách tích hợp module phát hiện và theo dõi đối tượng (object tracking), kết hợp với tái tạo 3D chuyên biệt cho các đối tượng động. Các kết quả tái tạo này được hòa trộn vào mô hình môi trường tĩnh, đảm bảo sự chính xác và nhất quán trong cả môi trường động. Phương pháp sử dụng các mô hình học sâu được huấn luyện trước, sau đó tinh chỉnh để phù hợp với kiến trúc của “NeuralRecon”. Hiệu quả của giải pháp sẽ được đánh giá trên các bộ dữ liệu chứa đối tượng động, so sánh trực tiếp với “NeuralRecon” gốc. Kỳ vọng rằng phương pháp đề xuất sẽ cải thiện đáng kể khả năng tái tạo 3D trong môi trường động, mở ra tiềm năng ứng dụng trong các lĩnh vực AR/VR, robot và giám sát.



Hình 1. Tái tạo 3D từ video đơn ảnh ở môi trường tĩnh và môi trường động

GIỚI THIỆU (Tối đa 1 trang A4)

Công nghệ tái tạo 3D từ video đơn ảnh đã đạt được những bước tiến vượt bậc trong những năm gần đây, mở ra tiềm năng ứng dụng trong nhiều lĩnh vực như thực tế tăng cường (AR), thực tế ảo (VR), robot tự hành, thiết kế công nghiệp và giám sát. Trong số các phương pháp hiện đại, NeuralRecon đã nổi lên như một giải pháp tiên tiến, có khả năng tái tạo 3D chính xác và thời gian thực từ video đơn ảnh, bằng cách sử dụng mạng học sâu để học và hợp nhất các đặc trưng hình học. Tuy nhiên, NeuralRecon được thiết kế chủ yếu cho các môi trường tĩnh và gặp khó khăn trong việc xử lý các đối tượng động, điều này hạn chế đáng kể khả năng ứng dụng trong các bối cảnh thực tế.

Thách thức chính trong việc tái tạo 3D với sự xuất hiện của các đối tượng động nằm ở tính phức tạp vốn có của chúng. Các đối tượng động, như con người, phương tiện giao thông hoặc bất kỳ vật thể di chuyển nào, có thể gây nhiễu trong quá trình tái tạo 3D, che khuất các khu vực khác của bối cảnh (scene), và liên tục thay đổi hình dạng cũng như vị trí. Các phương pháp tái tạo 3D truyền thống thường không xử lý tốt các thách thức này, dẫn đến hiện tượng artifacts (các lỗi do quá trình xử lý gây ra), mô hình bị thiếu chính xác, hoặc không nhất quán trong kết quả tái tạo.

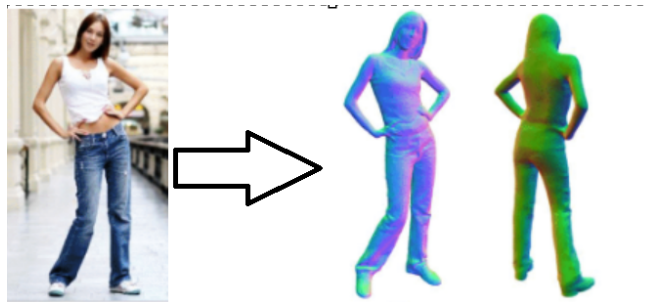
Đề tài này nhằm mục đích khắc phục những hạn chế trên bằng cách mở rộng NeuralRecon để có thể tái tạo 3D chính xác và nhất quán trong môi trường động. Cụ thể, chúng tôi đề xuất tích hợp các module phát hiện và theo dõi đối tượng (object detection [2] và tracking [3]) vào kiến trúc NeuralRecon, nhằm nhận diện và theo dõi các đối tượng động trong scene. Các đối tượng động sẽ được tái tạo riêng biệt bằng các kỹ thuật tái tạo 3D chuyên biệt, sau đó kết quả sẽ được hòa trộn một cách mượt mà với mô hình của môi trường tĩnh.

Phương pháp đề xuất sẽ sử dụng các mô hình học sâu đã được huấn luyện trước cho việc phát hiện và theo dõi đối tượng, sau đó tinh chỉnh để phù hợp với framework của NeuralRecon. Ngoài ra, chúng tôi sẽ nghiên cứu các kỹ thuật để tích hợp các đối tượng động và tĩnh một cách tự nhiên, giảm thiểu các hiện tượng artifacts. Hiệu quả của phương pháp sẽ được đánh giá trên các bộ dữ liệu chứa đối tượng động và so sánh

trực tiếp với phiên bản gốc của NeuralRecon.

Phát biểu bài toán:

- **Input:** Video đơn ảnh của một scene động, chứa cả các đối tượng tĩnh và động (ví dụ: con người, phương tiện giao thông).
- **Output:** Mô hình 3D chính xác và nhất quán của toàn bộ scene, bao gồm cả các phần tĩnh và động.



Hình 2. Ví dụ về kết quả của bài toán tái tạo 3D

MỤC TIÊU (Viết trong vòng 3 mục tiêu)

- ❖ **Phát triển một phương pháp mở rộng NeuralRecon:** Mục tiêu chính là xây dựng một hệ thống có khả năng tái tạo 3D thời gian thực trong môi trường có các đối tượng chuyển động, dựa trên nền tảng của NeuralRecon.
- ❖ **Tích hợp module object detection và tracking:** Mục tiêu thứ hai là tích hợp module object detection và tracking vào NeuralRecon, cho phép hệ thống phát hiện và theo dõi các đối tượng chuyển động trong scene.
- ❖ **Đánh giá phương pháp trên dataset có đối tượng động:** Mục tiêu cuối cùng là đánh giá hiệu quả của phương pháp đề xuất trên các dataset có chứa đối tượng động, so sánh kết quả với NeuralRecon gốc và các phương pháp khác.

NỘI DUNG VÀ PHƯƠNG PHÁP

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

- ❖ **Bước 1:** Nghiên cứu tổng quan
 - Mục tiêu: Tìm hiểu các công trình nghiên cứu liên quan đến tái tạo 3D

trong môi trường động, cũng như các phương pháp phát hiện và theo dõi đối tượng.

➤ Phương pháp:

- Tìm kiếm và tổng hợp các bài báo khoa học, tài liệu kỹ thuật về tái tạo 3D từ video đơn ảnh trong môi trường động [4], đặc biệt là các phương pháp sử dụng học sâu.
- Nghiên cứu các phương pháp object detection và tracking hiện đại, đánh giá ưu nhược điểm của từng phương pháp, có thể kể đến các phương pháp như Mask R-CNN [2], Siamese Networks [3] hay các phương pháp mới hơn như DETR [5].
- Phân tích kiến trúc và cơ chế hoạt động của NeuralRecon để xác định các điểm có thể cải tiến và mở rộng, chú trọng vào các module liên quan đến việc xử lý thông tin hình học.

❖ **Bước 2:** Thiết kế và xây dựng module tích hợp

➤ Mục tiêu: Thiết kế và xây dựng module tích hợp object detection và tracking vào kiến trúc của NeuralRecon.

➤ Phương pháp:

- Lựa chọn mô hình object detection phù hợp (ví dụ: Mask R-CNN) và mô hình tracking phù hợp (ví dụ: Siamese Networks), tùy thuộc vào sự phù hợp với hiệu suất và độ chính xác mà NeuralRecon yêu cầu.
- Tinh chỉnh các mô hình này trên dataset phù hợp, đảm bảo khả năng phát hiện và theo dõi đối tượng động chính xác, có thể sử dụng các dataset có chứa cả cảnh tĩnh và đối tượng động.
- Xây dựng giao diện kết nối giữa module phát hiện và theo dõi đối tượng với pipeline của NeuralRecon, có thể sử dụng các kỹ thuật message passing hoặc shared feature space.
- Thiết kế cơ chế truyền tải thông tin từ module phát hiện và theo

đổi đối tượng đến các module tái tạo 3D tiếp theo, ví dụ như truyền bounding box và mask của đối tượng.

❖ **Bước 3:** Tái tạo 3D cho đối tượng động

- Mục tiêu: Nghiên cứu và phát triển phương pháp tái tạo 3D riêng biệt cho các đối tượng động đã được phát hiện.
- Phương pháp:
 - Nghiên cứu các phương pháp tái tạo 3D object-level có thể áp dụng cho các đối tượng động, có thể kể đến các phương pháp shape completion [6] hoặc Neural Radiance Fields (NeRF) [7] hay các biến thể của chúng.
 - Xem xét khả năng sử dụng biến thể của NeuralRecon hoặc các phương pháp khác để tái tạo 3D cho đối tượng động, có thể bằng cách sử dụng các bounding box và mask từ module object detection như một guiding information.
 - Tối ưu hóa các phương pháp này để đạt được độ chính xác và hiệu suất thời gian thực, cân bằng giữa độ chính xác và tính toán.

❖ **Bước 4:** Hòa trộn đối tượng động và tĩnh

- Mục tiêu: Nghiên cứu và phát triển các kỹ thuật để hòa trộn mô hình 3D của đối tượng động và mô hình 3D của môi trường tĩnh một cách mượt mà.
- Phương pháp:
 - Nghiên cứu các phương pháp blending và ghép ảnh để giảm thiểu các artifacts và tạo ra sự chuyển tiếp tự nhiên, có thể sử dụng các kỹ thuật Poisson blending [8] hoặc feathering (làm mờ biên).
 - Sử dụng các thông tin về độ sâu và vị trí tương đối của các đối tượng để đảm bảo tính nhất quán của mô hình 3D tổng thể.
 - Phát triển thuật toán để hòa trộn một cách hiệu quả và tối ưu về

mặt thời gian thực, tận dụng các GPU để tăng tốc quá trình xử lý.

❖ **Bước 5:** Đánh giá và so sánh

- Mục tiêu: Đánh giá hiệu quả của phương pháp đề xuất trên các dataset có chứa đối tượng động.
- Phương pháp:
 - Sử dụng các dataset benchmark có chứa cả cảnh tĩnh và đối tượng động, ví dụ như KITTI dataset [9] hoặc các dataset tự tạo.
 - So sánh kết quả của phương pháp đề xuất với NeuralRecon gốc và các phương pháp khác bằng các metric đánh giá về độ chính xác, tính nhất quán, và hiệu suất thời gian thực.
 - Thực hiện phân tích định tính và định lượng để đánh giá ưu nhược điểm của phương pháp.

KẾT QUẢ MONG ĐỢI

- ❖ Một hệ thống mở rộng của NeuralRecon có khả năng tái tạo 3D thời gian thực trong môi trường động.
- ❖ Module object detection và tracking tích hợp vào NeuralRecon, có thể phát hiện và theo dõi các đối tượng động một cách chính xác.
- ❖ Phương pháp tái tạo 3D riêng biệt cho các đối tượng động, và các kỹ thuật hòa trộn đối tượng động và tĩnh một cách mượt mà.
- ❖ Báo cáo chi tiết về kết quả thực nghiệm, bao gồm cả các so sánh và đánh giá về hiệu quả của phương pháp.

TÀI LIỆU THAM KHẢO (*Định dạng DBLP*)

- [1]. Jiaming Sun, Yiming Xie, Linghao Chen, Xiaowei Zhou, Hujun Bao: NeuralRecon: Real-Time Coherent 3D Reconstruction from Monocular Video. CVPR 2021: 15598-15607
- [2]. Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick: Mask R-CNN. ICCV 2017: 2980-2988

- [3]. Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, Philip H.S. Torr: Fully-Convolutional Siamese Networks for Object Tracking. ECCV 2016: 850-865
- [4]. Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Andrew Fitzgibbon: KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera. ISMAR 2011: 127-136
- [5]. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko: End-to-End Object Detection with Transformers. ECCV 2020: 213-229
- [6]. Christian B. Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, Silvio Savarese: 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction. ECCV 2016: 628-644
- [7]. Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV 2020: 151-174
- [8]. Patrick Pérez, Michel Gangnet, Andrew Blake: Poisson Image Editing. SIGGRAPH 2003: 313-322
- [9]. Andreas Geiger, Philip Lenz, Raquel Urtasun: Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. CVPR 2012: 3354-3361