

TÁI TẠO BÀN TAY 3D DỰA TRÊN MÔ HÌNH THÔNG QUA HỌC TỰ GIÁM SÁT

Lưu Anh Dũng - 19521392

Nguyễn Trọng Doanh - 19521368

Tóm tắt

- Lớp: CS519.M11.KHCL
- Link Github của nhóm:
<https://github.com/anhdungbmt2001/CS519.M11.KHCL>
- Link YouTube video: <https://www.youtube.com/watch?v=e9PhturUY6A>
- Nhóm DDD:



Nguyễn Trọng Doanh



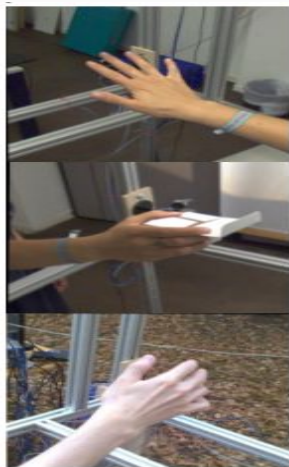
Lưu Anh Dũng

Giới thiệu

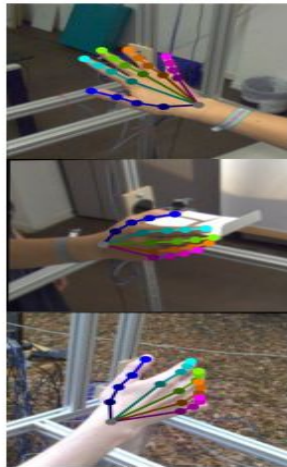
- Việc tái tạo 3D bàn tay người từ một hình ảnh là rất quan trọng đối với các tác vụ thị giác máy tính.
- Việc tái tạo 3D bàn tay người là một thách thức.
- Nhận thấy các tín hiệu 2D trong không gian hình ảnh có liên quan chặt chẽ với mô hình bàn tay 3D trong thế giới thực
- Ý tưởng ban đầu của mô hình là thu thập các đặc điểm hình học từ hình ảnh đầu vào thông qua các điểm khóa 2D mà có thể dễ dàng phát hiện và truy cập, sau đó sử dụng tính nhất quán giữa các phương pháp biểu diễn 2D và 3D để hợp lý hóa đầu ra của mạng neural.

Giới thiệu

- Input: 1 hình ảnh chụp một bàn tay ở tư thế tự do.
- Output: Hình biểu diễn các khớp của bàn tay và mô hình 3D của bàn tay ở tư thế giống như trong input.



Input Image



Output Joints



3D Reconstruction

Mục tiêu

- Nghiên cứu và xây dựng thành công mô hình mạng neural tái tạo bàn tay 3D thông qua học tự giám sát, cho ra kết quả chính xác các khớp và kết cấu 3D của bàn tay từ một hình ảnh duy nhất mà không sử dụng bất kỳ dữ liệu đã gắn nhãn nào.
- Tìm hiểu các mô hình tái tạo bàn tay 3D hiện có và phát triển mô hình của chúng tôi để đạt độ chính xác và tốc độ xử lý cao hơn các mô hình hiện có.
- Thử nghiệm mô hình trên một số bộ dữ liệu khó, qua đó cải thiện độ chính xác của mô hình.

Nội dung và Phương pháp

- Phương pháp mà chúng tôi nghiên cứu sẽ sử dụng bộ mã hoá tự động để nhận hình ảnh bàn tay làm đầu vào, và trả về hình dạng, kết cấu của tay và góc nhìn máy ảnh, sau đó tạo ra nhiều biểu diễn 2D trong không gian hình ảnh và thiết kế các hàm mất mát cho quá trình huấn luyện.
- Để hiện thực hoá phương pháp này, có một số vấn đề cần giải quyết
 - Vấn đề thứ nhất: làm thế nào để sử dụng hiệu quả các điểm khoá 2D để giám sát việc tái tạo bàn tay 3D có hình dáng lạ
 - Vấn đề thứ hai: làm cách nào để xử lý nhiễu (vì không sử dụng bất kỳ dữ liệu có gắn nhãn nào)

Nội dung và Phương pháp

- Để giải quyết vấn đề đầu tiên, một bộ mã hoá tự động sẽ được sử dụng để dự đoán các khớp và hình dạng 3D, trong đó các khớp 3D đầu ra sẽ được chiếu vào không gian hình ảnh và trong quá trình huấn luyện sẽ được căn chỉnh với các điểm khoá đã được phát hiện.
- Nếu chỉ căn chỉnh các điểm khoá trong không gian hình ảnh, có thể sẽ xảy ra nhiều tư thế tay không hợp lệ. Ngoài ra, các điểm khoá 2D không thể làm giảm mơ hồ về tỷ lệ của bàn tay 3D được dự đoán.
 - Cần nghiên cứu giải pháp để mạng neural trả về bàn tay 3D với tư thế và kích thước hợp lý.

Nội dung và Phương pháp

- Để giải quyết vấn đề thứ hai, cần có:
 - Một công cụ dự đoán điểm khoá 2D: trả về các điểm khoá 2D.
 - Một sự mất mát mới về tính nhất quán 2D-3D: liên kết công cụ dự đoán điểm khoá 2D và mạng tái tạo 3D để làm cho cả hai cùng giúp đỡ nhau trong quá trình huấn luyện.

Kết quả dự kiến

- Mô hình đạt độ chính xác hơn 80% khi thử nghiệm trên một số bộ dữ liệu lớn như FreiHAND, HO-3D,...
- Mô hình đạt độ chính xác cao hơn so với ít nhất 3 mô hình state-of-the-art.
- Mô hình được ứng dụng trong các ứng dụng thực tế ảo (VR), thực tế tăng cường (AR).

Tài liệu tham khảo

- [1]. Anil Armagan, Guillermo Garcia-Hernando, Seungryul Baek, Shreyas Hampali, Mahdi Rad, Zhaohui Zhang, Shipeng Xie, MingXiu Chen, Boshen Zhang, Fu Xiong, et al:
Measuring generalisation to unseen viewpoints, articulations, shapes and objects for 3d hand pose estimation under hand-object interaction.
European Conference on Computer Vision, 2020.
- [2]. Vassilis Athitsos, Stan Sclaroff:
Estimating 3d hand pose from a cluttered image.
CVPR 2003.
- [3]. Lihao Ge, Zhou Ren, Yuncheng Li, Zehao Xue, Yingying Wang, Jianfei Cai, Junsong Yuan:
3d hand shape and pose estimation from a single rgb image.
CVPR 2019.
- [4]. Yana Hasson, Bugra Tekin, Federica Bogo, Ivan Laptev, Marc Pollefeys, Cordelia Schmid:
Leveraging photometric consistency over time for sparsely supervised hand-object reconstruction.
CVPR 2020.
- [5]. Markus Holl, Markus Oberweger, Clemens Arth, Vincent Lepetit:
Efficient physics-based implementation for realistic hand-object interaction in virtual reality.
IEEE Conference on Virtual Reality and 3D User Interfaces, 2018.
- [6]. Maria Parelli, Katerina Papadimitriou, Gerasimos Potamianos, Georgios Pavlakos, Petros Maragos:
Exploiting 3d hand pose estimation in deep learning-based sign language recognition from rgb videos.
European Conference on Computer Vision. Springer, 2020.