

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

HOÀNG DANH TUYÊN

**NGHIÊN CỨU SỬ DỤNG DEEP LEARNING
ĐỂ ƯỚC LƯỢNG TỬ THỂ BÀN TAY 3D TỪ
DỮ LIỆU CỦA CẢM BIẾN MANG TRÊN NGƯỜI**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên – 2023

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

HOÀNG DANH TUYÊN

NGHIÊN CỨU SỬ DỤNG DEEP LEARNING
ĐỂ ƯỚC LƯỢNG TỬ THỂ BÀN TAY 3D TỪ
DỮ LIỆU CỦA CẢM BIẾN MANG TRÊN NGƯỜI

Ngành: Khoa học máy tính

Mã số: 8 48 0101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC:

TS. Lê Văn Hùng

Thái Nguyên - 2023

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn Thạc Sĩ với tiêu đề "Nghiên cứu sử dụng Deep Learning để ước lượng tư thế bàn tay 3D từ dữ liệu của cảm biến mang trên người" là kết quả nghiên cứu của tôi dưới sự hướng dẫn khoa học của TS. Lê Văn Hùng. Tất cả các số liệu, hình ảnh, bảng biểu, các bài báo được trình bày trong luận văn đều được trích dẫn nguồn trung thực và đầy đủ.

Thái Nguyên, tháng 07 năm 2023

Hoàng Danh Tuyên

LỜI CẢM ƠN

Lời cảm ơn đầu tiên, em xin bày tỏ lời cảm ơn sâu sắc tới Thầy TS Lê Văn Hùng đã tận tình hướng dẫn và giúp đỡ trong quá trình tìm hiểu, nghiên cứu và thực hiện luận văn. Luận văn này không thể hoàn thành nếu thiếu đi những góp ý chân thành và sự hướng dẫn tỉ mỉ từ Thầy Hùng. Dù thầy rất bận rộn với công việc giảng dạy, nghiên cứu khoa học nhưng Thầy luôn lắng nghe những vấn đề mà tôi gặp phải, luôn động viên kịp thời và đưa ra những định hướng để tôi hoàn thiện kết quả nghiên cứu. Tôi muốn bày tỏ lòng biết ơn đến Thầy và chúc Thầy sẽ đạt được nhiều thành công hơn nữa trong công việc, cũng như trong cuộc sống. Trong quá trình học Thạc sỹ tại trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên, tôi cũng đã được hỗ trợ và tạo điều kiện từ các thầy, cô trong trường. Quảng thời gian vừa qua là nhiều ngày vừa học, vừa làm, hàng ngày trong tuần tôi đều có mặt ở cơ quan, cuối tuần tôi phải có mặt ở trường. Hoàn thành cả hai công việc một lúc là rất nhiều nỗ lực của bản thân, nhưng sẽ không thể thành công nếu thiếu đi sự ủng hộ của mọi người. Cuối cùng, tôi xin dành lời cảm ơn tới gia đình tôi, điểm tựa để tôi vượt qua những giai đoạn khó khăn trong công việc và cuộc sống. Kết quả ngày hôm nay là lời tri ân chân thành nhất mà tôi có thể gửi tới gia đình mình. Cuối cùng tôi xin bày tỏ lời cảm ơn tới các đồng nghiệp, gia đình, bạn bè đã luôn động viên, chia sẻ, ủng hộ và giúp đỡ tôi vượt qua khó khăn để đạt được những kết quả nghiên cứu trong Luận văn này.

Học viên

Hoàng Danh Tuyên

Mục lục

MỞ ĐẦU	1
Chương 1. TỔNG QUAN ƯỚC LƯỢNG TƯ THỂ BÀN TAY 3D	7
1.1. Học sâu và mạng nơ ron tích chập	7
1.1.1. Định nghĩa học sâu	7
1.2. Ước lượng tư thể bàn tay 3D	10
1.2.1. Ước lượng tư thể bàn tay 3D từ ảnh màu	12
1.2.2. Ước lượng tư thể bàn tay 3D từ ảnh độ sâu	13
1.2.3. Ước lượng tư thể bàn tay 3D từ dữ liệu RGB-D và các loại dữ liệu khác	17
1.3. Cơ sở dữ liệu đánh giá ước lượng tư thể bàn tay 3D	17
1.3.1. Cơ sở dữ liệu thu từ góc nhìn thứ ba	18
1.3.2. Cơ sở dữ liệu thu từ góc nhìn thứ nhất	19
1.4. Kết quả ước lượng tư thể bàn tay 3D	23
1.5. Đánh giá những tồn tại	24
1.6. Kết luận Chương 1	24
Chương 2. MÔ HÌNH ƯỚC LƯỢNG TƯ THỂ BÀN TAY 3D	26
2.1. Bài toán ước lượng tự động tư thể bàn tay 3D	26
2.2. Phát hiện bàn tay trên ảnh màu	30
2.3. Xây dựng dữ liệu đám mây điểm của vùng dữ liệu bàn tay	32
2.4. Mạng HandFoldingNet cho ước lượng tư thể bàn tay 3D	34
2.5. Kết luận Chương 2	36
Chương 3. THỬ NGHIỆM MÔ HÌNH ƯỚC LƯỢNG TƯ THỂ BÀN TAY 3D....	37
3.1. Ước lượng tư thể bàn tay 3D trên cơ sở dữ liệu HOI4D	37
3.1.1. Cơ sở dữ liệu HOI4D	37
3.1.2. Chuẩn bị dữ liệu và thiết bị	38
3.1.3. Độ đo đánh giá	40

3.1.4. Kết quả phát hiện bàn tay và ước lượng tư thế bàn tay 3D trên cơ sở dữ liệu HOI4D	
42	
3.2. Minh họa kết quả ước lượng tư thế bàn tay 3D và xây dựng video demo....	47
3.3. Kết luận chương 3	52
KẾT LUẬN VÀ NGHIÊN CỨU TIẾP THEO	54
PHỤ LỤC - MỘT SỐ MÃ NGUỒN CHƯƠNG TRÌNH.....	60

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

VR	Virtual Reality
AR	Augmented Reality
HOI4D	Human-Object Interaction 4D
CNN, CNNs	Convolutional Neural Networks
GCN	Graph Convolutional Network
DNN	Deep Neural Network
DL, ML	Deep Learning, Machine Learning
ICVL	Institute of Computing and Informatics, Slovak University of Technology
NYU	New York University (Đại học New York)
MSRA	Microsoft Research Asia (Trung tâm Nghiên cứu Microsoft ở châu Á)
MSRC	Microsoft Research Cambridge
PSO	Particle swarm optimization
ICP	Iterative closest point
HCRNN	Hierarchical Convolutional Recurrent Neural Network
SE	Special Euclidean (Euclide đặc biệt)
VAE	Variational Autoencoder
GAN	Generative Adversarial Network
RGB	Red Green Blue
RGB-D	Red Green Blue - Depth
JMFC	Joint matrix factorization and completion
PEL	Permutation equivariant layer
A2J	Anchor-to-Joint
FC	Kiểu lớp (layer) Fully Connected
AUC	Area Under the Curve
PCK	Percentage of Correct Keypoints
STB	Stereo Hand Pose Tracking Benchmark
MS Kinect v1	Microsoft Kinect version 1
RDF	Randomized Decision Forest
DOF	Degrees of Freedom (số bậc tự do)
FPHAB	First-Person Hand Action Benchmark
WCVS	Wearable Computer Vision Systems
ELAN	Ensemble Learning Aggregation Network
E-ELAN	Efficient Extended Learning Aggregation Network
LFB	Local Feature-based
IOU	Intersection Over Union

Danh sách hình vẽ

1	<i>Minh họa cách thức cầm nắm đối tượng.</i>	1
2	<i>Minh họa điều khiển tivi trong nhà thông minh [8].</i>	2
3	<i>Minh họa sử dụng cử chỉ tay để chơi game (thực hiện tương tác với máy tính).</i>	2
4	<i>Minh họa cánh tay robot thực hiện cầm nắm, lấy đồ vật trên bàn.</i>	3
5	<i>Minh họa phân loại thiết lập camera.</i>	3
1.1	<i>Mối quan hệ giữa học sâu, học máy và trí tuệ nhân tạo.</i>	8
1.2	<i>Một mạng thần kinh điển hình.</i>	8
1.3	<i>Minh họa sự khác biệt về quá trình trích chọn đặc trưng để huấn luyện mô hình của ML và DL.</i>	9
1.4	<i>Minh họa sự khác biệt về quá trình trích chọn đặc trưng để huấn luyện mô hình của ML và DL.</i>	10
1.5	<i>Minh họa phát hiện bàn tay bằng YOLOv5.</i>	11
1.6	<i>Minh họa cấu trúc xương của bàn tay người [14].</i>	11
1.7	<i>Mô hình ước lượng tư thế bàn tay 3D từ ảnh màu [11].</i>	13
1.8	<i>Hai nhánh ước lượng tư thế bàn tay 3D từ ảnh độ sâu [11].</i>	15
1.9	<i>(a) 2D CNN lấy hình ảnh độ sâu làm bản đồ nhiệt đầu vào và đầu ra. (b) Các 2D CNN với nhiều hướng nhìn lấy các phép chiếu nhiều hướng nhìn làm đầu vào và tạo ra bản đồ nhiệt tương ứng với nhiều hướng nhìn. (c) 2D CNN lấy ảnh độ sâu làm đầu vào và hồi quy trực tiếp các vị trí khớp 3D. (d) sử dụng 3D CNN lấy biểu diễn thể tích làm đầu vào và hồi quy các vị trí khớp 3D [6].</i>	16
1.10	<i>Minh họa quá trình ước lượng tư thế bàn tay 3D từ ảnh độ sâu [11] từ bộ thư viện tư thế bàn tay 3D.</i>	16

1.11	Cấu trúc thư mục của cơ sở dữ liệu HOI4D [16].	21
1.12	Minh họa nhân của các hoạt động của cơ sở dữ liệu HOI4D.	22
2.1	Minh họa dữ liệu của các ngón tay bị che khuất trong cơ sở dữ liệu HOI4D.	27
2.3	Mô hình ước lượng khung xương/tư thế bàn tay 3D trên cơ sở dữ liệu HOI4D.	28
2.2	Minh họa kết quả phát hiện bàn tay trên cơ sở dữ liệu FPHAB.	28
2.4	Minh họa tọa độ (x,y,z) của một khung xương bàn tay trên cơ sở dữ liệu HOI4D.	29
2.5	Kiến trúc của E-ELAN [20].	31
2.6	Minh họa phương pháp này sử dụng dự đoán đầu khách hàng tiềm năng làm hướng dẫn để tạo nhãn phân cấp từ thô đến tinh [20].	32
2.7	Minh họa các lớp của Darknet-53.	32
2.8	Minh họa các cảm biến hình ảnh trên MS Kinect v2.	33
2.9	Quá trình chuyển và tiền xử lý đám mây điểm từ ảnh độ sâu.	34
2.10	Kiến trúc HandFoldingNet.	34
2.11	Kiến trúc của LFB của mạng HandFoldingNet.	35
3.1	Minh họa thư mục dữ liệu của 16 loại hành động của cơ sở dữ liệu HOI4D.	38
3.2	Minh họa dữ liệu của 16 loại hành động của cơ sở dữ liệu HOI4D.	38
3.3	Minh họa xác định đường bao bàn tay gốc cho đánh giá phát hiện bàn tay. .	39
3.4	Minh họa cấu hình mạng HandFoldingNet và cơ sở dữ liệu HOI4D trên máy chủ có GPU.	42
3.5	Minh họa một số trường hợp không phát hiện được bàn tay.	43
3.6	Minh họa khung xương bàn tay 3D gốc (khung xương màu xanh) và khung xương bàn tay 3D ước lượng được (khung xương màu đỏ).	44
3.7	Minh họa kết quả ước lượng tư thế bàn tay 3D trên một số hành động của cơ sở dữ liệu HOI4D.	45
3.8	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm ô tô đồ chơi.	47
3.9	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cốc. . .	47
3.11	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động kéo ngăn tủ.	48
3.10	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm labtop. .	48

3.12	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cái chai.	48
3.14	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cái bát. .	49
3.13	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm hộp an toàn.	49
3.15	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động sách xô nước.	49
3.16	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cái kéo.	50
3.17	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cái kìm.	50
3.18	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm ấm đun nước.	50
3.19	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm con dao.	51
3.20	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm thùng rác.	51
3.21	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cái đèn.	51
3.22	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cái ghim giấy.	52
3.23	Kết quả ước lượng tư thế bàn tay 3D khi thực hiện hành động cầm cái cái ghế.	52