K-SW Fall Program 2022

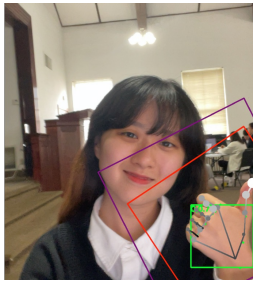
Drone Gesture Control Using Vision AI in Real-Time

Team Gangsture

Hyeongbin Park · Yujin Lee · Soeun Lee · Seunghwan Kim · Xinning Bai · Jake Hazelton

PURDUE UNIVERSITY | Polytechnic Institute

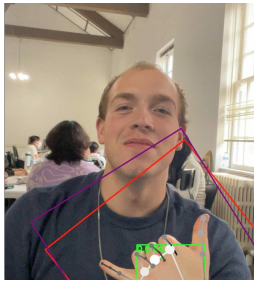
Members



Hyeongbin Park (Leader)
Vision Ai

Kwangwoon Univ

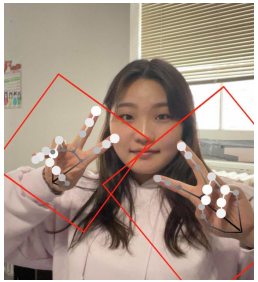
Information
Convergence



Jake Hazelton
Back-end

Purdue Univ

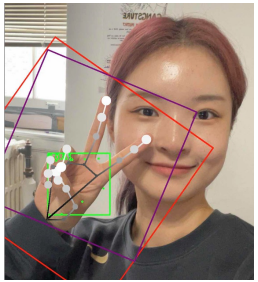
Computer &
Information
Technology



Soeun Lee
Networking

Jeju National Univ

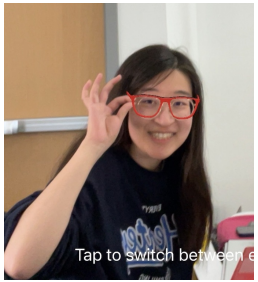
Computer &
Information
Technology



Yujin Lee
iOS

Kwangwoon Univ

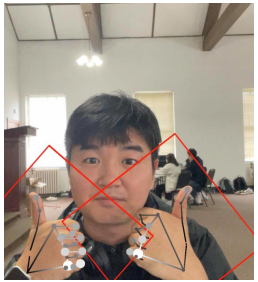
Software
Engineering



Xinning Bai
Researcher

Purdue Univ

Computer &
Information
Technology



Seunghwan Kim
Drone Computing

Jeju National Univ

Big Data
Science

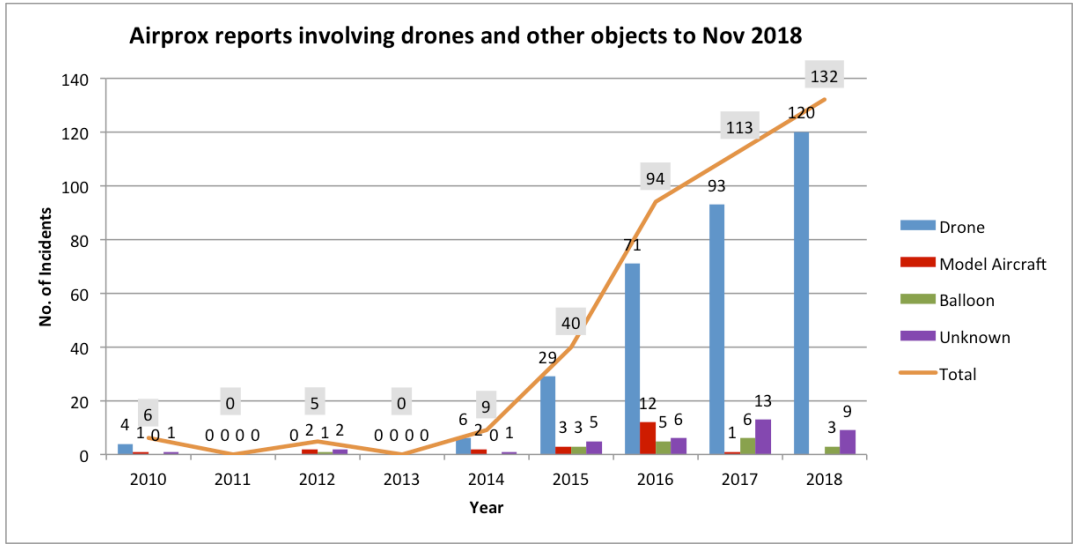
Index

- Introduction
- Literature Review
- Methodology
  - Dataset
  - Hand Recognition with ML
  - Face Recognition with Computer Vision and ML
- Implementation
- Conclusion

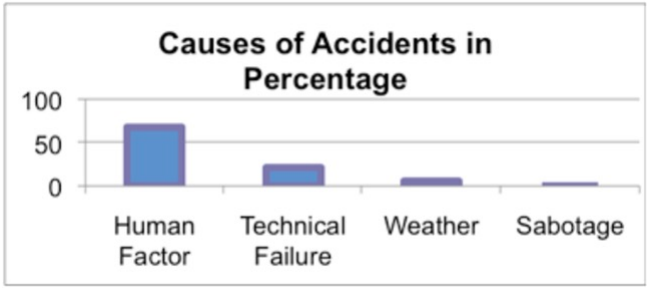
Background: The Use of Drone in Commercial and Private Industry

1. Introduction

Problem Statement: Drone Accidents with Operational Issues



"Fig. 3. Statistics on increasing drone accident rates"



"Fig 4. Statistics on the main causes of drone accidents"

1. Introduction

Problem Statement: Handheld Remote Controller

"...the external control method can be said to be a common problem in all drones. Unlike conventional fighter jets and cars, drones use external controls that allow users to observe and control the machine from the outside, rather than boarding the aircraft and controlling it by matching their own viewpoint and the machine's viewpoint"

- Youngjae Yoo et al. -

Project Purpose: Developing an Easier Drone Interface

"Develop the easier and efficient way to control the drone with hand gestures"

#machine learning #computer vision #intuitive #easier #low resource #mobile platform

Early Research: Extra Sensors and Equipment

Pros

- High accuracy in real time

Cons

- Expensive
- Another material to protect it

"Wind speed, GPS signal, and even brightness where the Leap Motion Controller is placed are all critical to the system performance"

"Fig. 5. The most representative equipment, Leap Motion Controller"

Recent Research : AI and Big Data

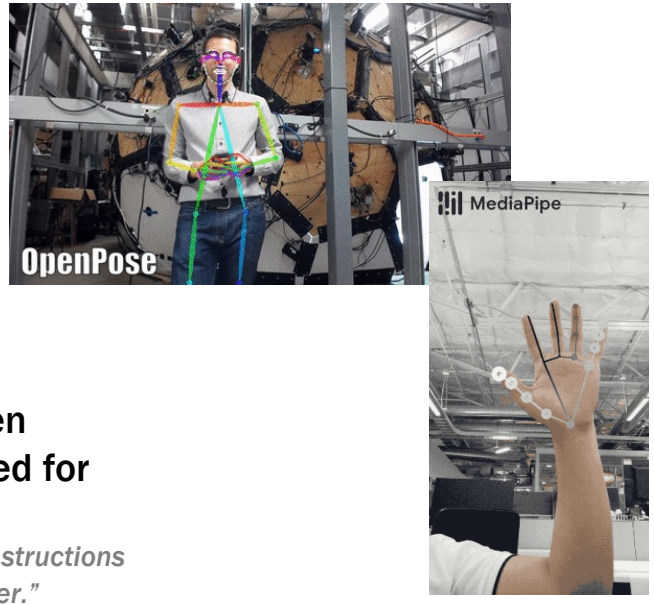
Pros

- Low cost of building a system
- Improved usability and reduced physical fatigue

Cons

- Delays and unstable accuracy
- High computational volume for multiple hands, poor recognition performance when the model faced unseen situation, and unintuitive open hand dataset specialized for specific task
- *“There is a delay of 3 seconds between the processing and execution of instructions sent by the hand gesture predictor and its execution on the drone controller.”*

"Fig. 6. Google Mediapipe Hand solution"



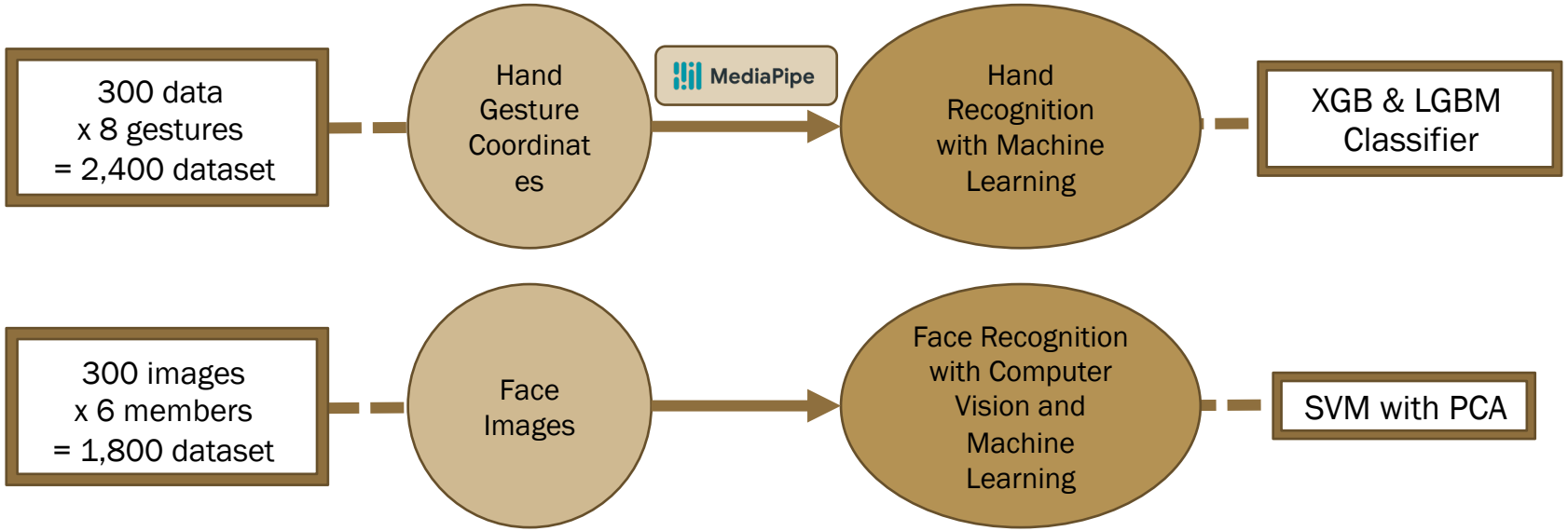
"Fig. 7. OpenPose developed by Carnegie Mellon University research team"

Gangsture's Research Plan

- Control the drone with hand gesture instead of the physical controller
- Replace additional equipment and sensors with AI
- To make the entire system use low resource, develop AI model as simply as possible
- Collect dataset and find out the optimal one to reinforce the generalization performance of the model (not affected by left or right hand, lightning, and background)
- Embed the entire system to the iOS app that users always carry
- With a single mobile camera, create two perspective screen of user and the drone
- Introduce face recognition function to grant the controller authority to the specific user

3. Methodology

Overview

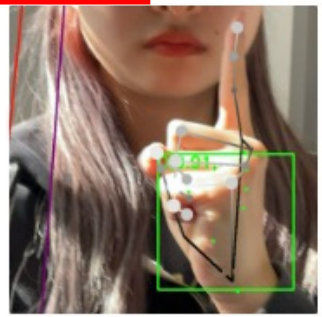


3. Methodology

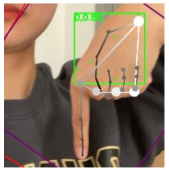
Dataset: Hand Gesture Coordinates System

0 Up

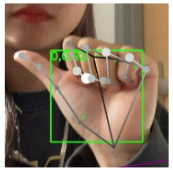
Gesture Label and Corresponding Command



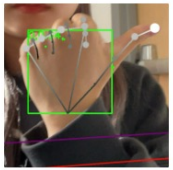
1 Down



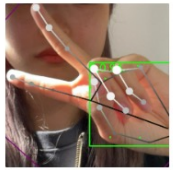
2 Left



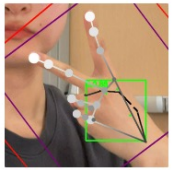
3 Right



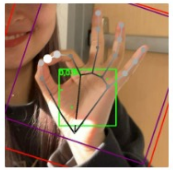
4 Go forward



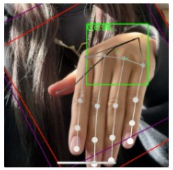
5 Back



6 Take off

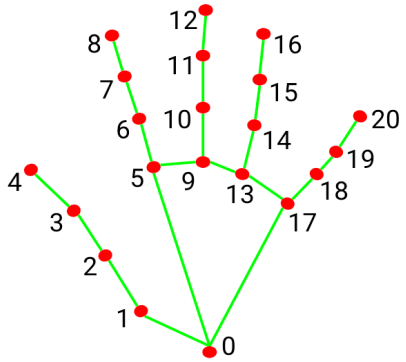


7 Land



300 data x 8 gestures = 2,400 dataset

MediaPipe



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

3. Methodology

Dataset: Face Images

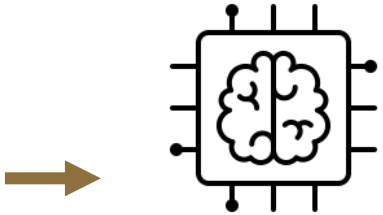
300 images x 4 members = 1,200 dataset

Dataset: Comparison Data of Hand Gestures

Which one is the best for the generalization of AI?

Comparison Factor	Common Factor	Dataset
The number of people	various hand size mixed hand type	Only One Person
		Only Multiple Person
Hand type	various hand size multiple people	Only Right Hand
		Only Left Hand
Hand size	mixed hand type multiple people	Only Small Size
		Only Big Size

"6 types of hand gesture dataset"



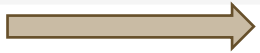
6 types of hand recognition models

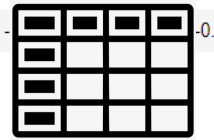
3. Methodology

Hand Recognition with ML: Preprocessing

	0	hand_type	1_x	1_y	1_z	1_rx	1_ry	1_rz
0	HandLandmark.WRIST	Right	0.649742	0.319549	4.393590e-07	0.000000	0.000000	0.000000
1	HandLandmark.THUMB_CMC	Right	0.573171	0.361034	-1.980279e-02	-0.076571	0.041485	-0.019803
2	HandLandmark.THUMB_MCP	Right	0.529059	0.449961	-3.892376e-02	-0.120683	0.130412	-0.038924
3	HandLandmark.THUMB_IP	Right	0.522001	0.541655	-5.097779e-02	-0.127741	0.222107	-0.050978
4	HandLandmark.THUMB_TIP	Right	0.499557	0.603344	-6.559844e-02	-0.150185	0.283795	-0.065599
...	...	...	...	...	...	...	...	...
6295	HandLandmark.RING_FINGER_TIP	Left	0.241911	0.694926	-1.355357e-01	-0.009612	0.369049	-0.135537
6296	HandLandmark.PINKY_MCP	Left	0.198408	0.431762	-9.598525e-02	-0.053114	0.105885	-0.095986
6297	HandLandmark.PINKY_PIP	Left	0.206384	0.536332	-1.148420e-01	-0.045139	0.210454	-0.114843
6298	HandLandmark.PINKY_DIP	Left	0.212087	0.600043	-1.198477e-01	-0.039436	0.274165	-0.119849
6299	HandLandmark.PINKY_TIP	Left	0.215907	0.650461	-1.236072e-01	-		-0.123608

6300 rows × 8 columns





2,400 rows x 128 columns

"Fig. 8. Original one hand coordinates system data"

1. Removing unnecessary column -> “landmark_name”

2. Converting “hand_type”
hand_type value is consisted of “right” and “left” -> replace “right” as 0, and “left” as 1

3. Calculating relative coordinates from original one
to add a relationship between each landmark -> create new relative coordinates system from landmark 0 (wrist)

4. Reshaping as single input data
all of data are aggregated as 1 x 2400 x 128 from 8 x 6300 x 8 with y label

Hand Recognition with ML: Training and Evaluation

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	60
1.0	1.00	1.00	1.00	60
2.0	1.00	1.00	1.00	60
3.0	1.00	1.00	1.00	60
4.0	0.98	1.00	0.99	60
5.0	1.00	0.98	0.99	60
6.0	1.00	1.00	1.00	60
7.0	1.00	1.00	1.00	60
accuracy			1.00	480
macro avg	1.00	1.00	1.00	480
weighted avg	1.00	1.00	1.00	480

Results of XGB Classifier

- Multi-Class Classification (8 class)
- 2 Types of Gradient Boosting model
  XGB / LGBM

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	60
1.0	1.00	1.00	1.00	60
2.0	1.00	1.00	1.00	60
3.0	1.00	1.00	1.00	60
4.0	0.98	1.00	0.99	60
5.0	1.00	0.98	0.99	60
6.0	1.00	1.00	1.00	60
7.0	1.00	1.00	1.00	60
accuracy			1.00	480
macro avg	1.00	1.00	1.00	480
weighted avg	1.00	1.00	1.00	480

Results of LGBM Classifier

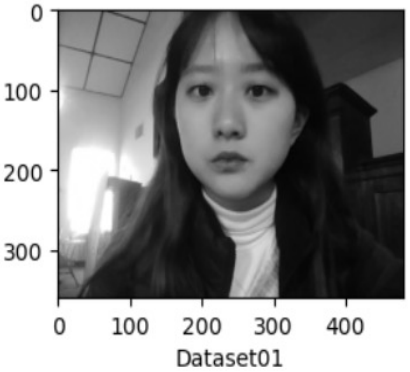
- Evaluation Metrics
  Precision / Recall / F1 Score / Accuracy
- Train/Test split as 8:2
- Train/Valid split as 8:2

Hand Recognition with ML: Real Test with Comparison Dataset

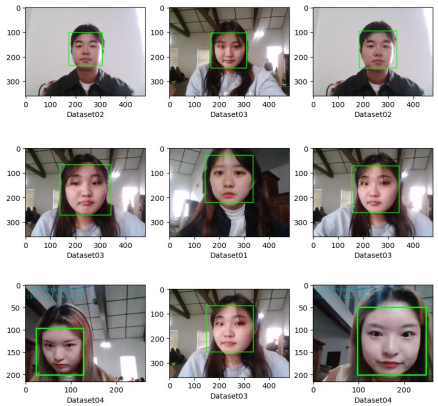
Comparison Factor	Dataset	Results
The number of people	Only One Person	No error on both hands When it is far, it can't detect back gestures of left hand
	Only Multiple Person	No error on both hands When it is far, the landing gesture is not working **The final dataset**
Hand type	Only Right Hand	No error on right hand Only three gestures can be detected on left hand When it is far, both hands have bad accuracy and unstable, but the landing gesture is always working well
	Only Left Hand	Only three gestures can be detected on both hands When it is far, only some gestures can be detected on both hands
Hand size	Only Small Size	No error when it is far When it close, it becomes unstable and get low accuracy
	Only Big Size	No error when it is close When it is far, it becomes unstable and get low accuracy

3. Methodology

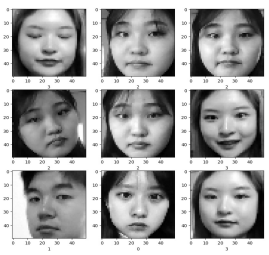
Face Recognition with Computer Vision and ML: Preprocessing



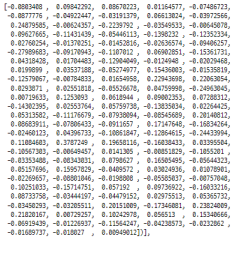
1. Read the face image dataset as a gray scale.



2. Extract the face from the gray scale image using the **Haar Cascade Classifier.**



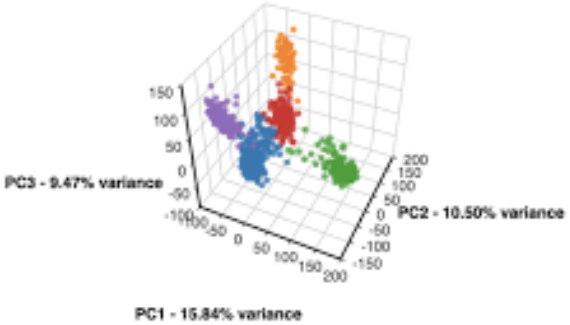
3. Crop and save the face based on the returned coordinates.

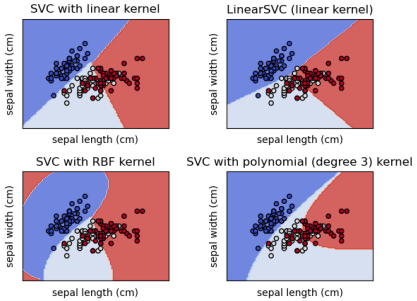


4. Encode the saved face using the Python face recognition library

* Haar Cascade Classifier:
A machine learning-based object detection algorithm.
It is used to detect objects in a video or image based on a feature. It returns coordinate value of face position.

Face Recognition with Computer Vision and ML: Training and Evaluation





5. Compress the encoded high-dimensional values with PCA.

6. Process face classification with SVM trained on the face feature vectors.

* PCA, Principle Component Analysis: a representative dimension reduction algorithm for extracting core features.

* SVM, Support Vector Machine: a supervised learning model, it is used for pattern recognition and data anlysis, classification and regression analysis.

Result

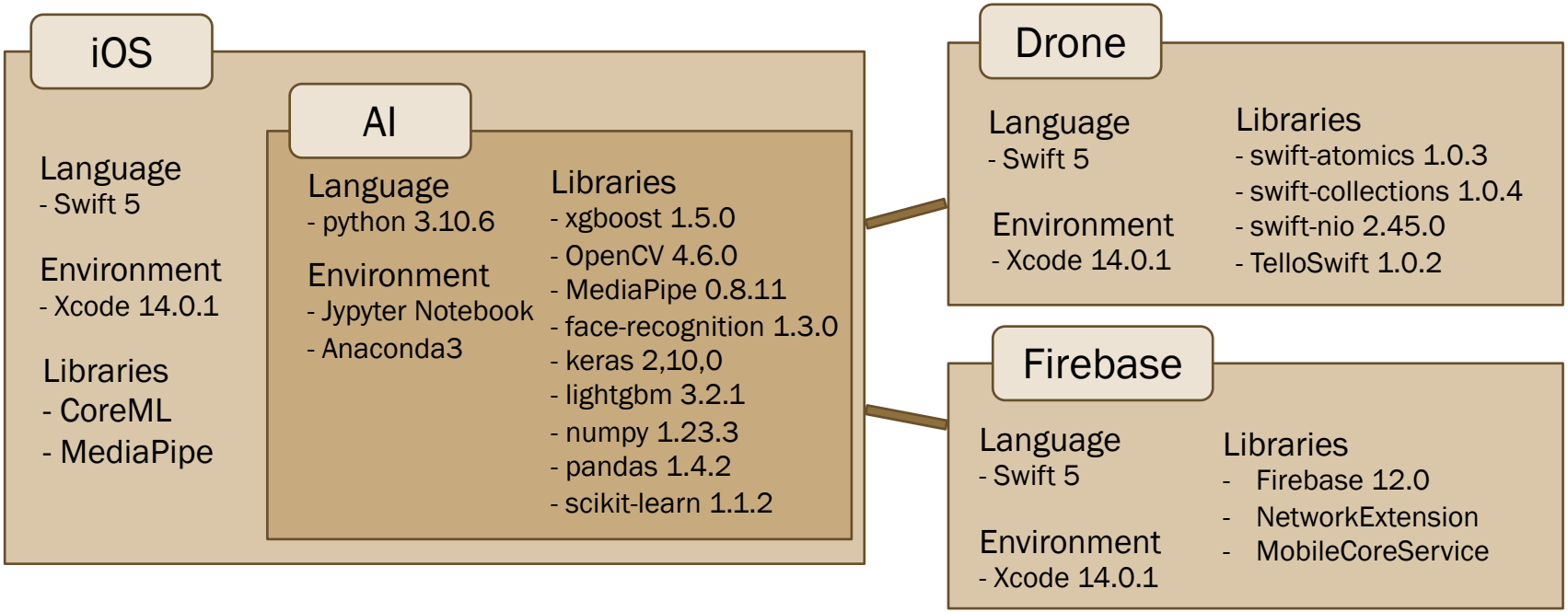
	precision	recall	f1-score	support
0	1.00	1.00	1.00	58
1	1.00	1.00	1.00	69
2	1.00	1.00	1.00	57
3	1.00	1.00	1.00	49
accuracy			1.00	233
macro avg	1.00	1.00	1.00	233
weighted avg	1.00	1.00	1.00	233

accuracy score: 1.0

- **Multi-Class Classification (4 class)**
- **SVM with radial basis function kernel**
- **Evaluation Metrics**
  Precision / Recall / F1 Score / Accuracy
- **Train/Test and Train/Valid as 8:2**

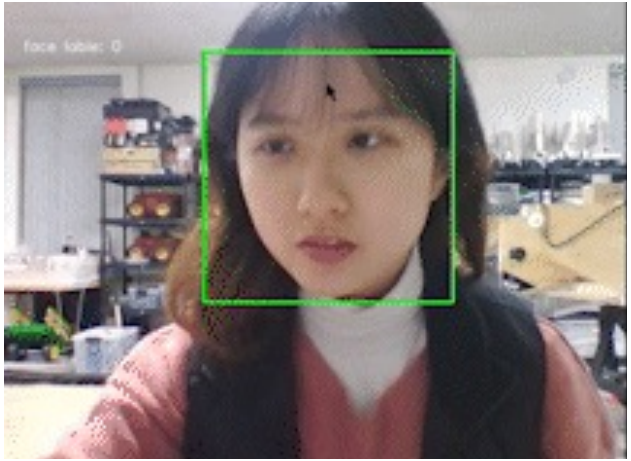
4. Implementation

System Architecture



4. Implementation

System flow 1. Face recognition in iOS application



Detect facial features through edge detector

The output is compressed to list and entered to the AI model.

4. Implementation

System flow 2. If it is true, do hand detection in iOS app

- The front camera's screen is delivered to the MediaPipe every moment and the MediaPipe provides 21 normalized landmarks.
- The AI model using Python was packed in CoreML format , iOS's artificial intelligence framework, and added to the iOS.
- CoreML allows you to obtain the result value of artificial intelligence embedded in the application

4. Implementation

System flow 3. Send command using UDP socket

- Set up a UDP client on the mobile device to send and receive messages from the Tello via the same port.
- Before sending any other commands, send “command” to the Tello via UDP PORT 8889 to initiate SDK mode.

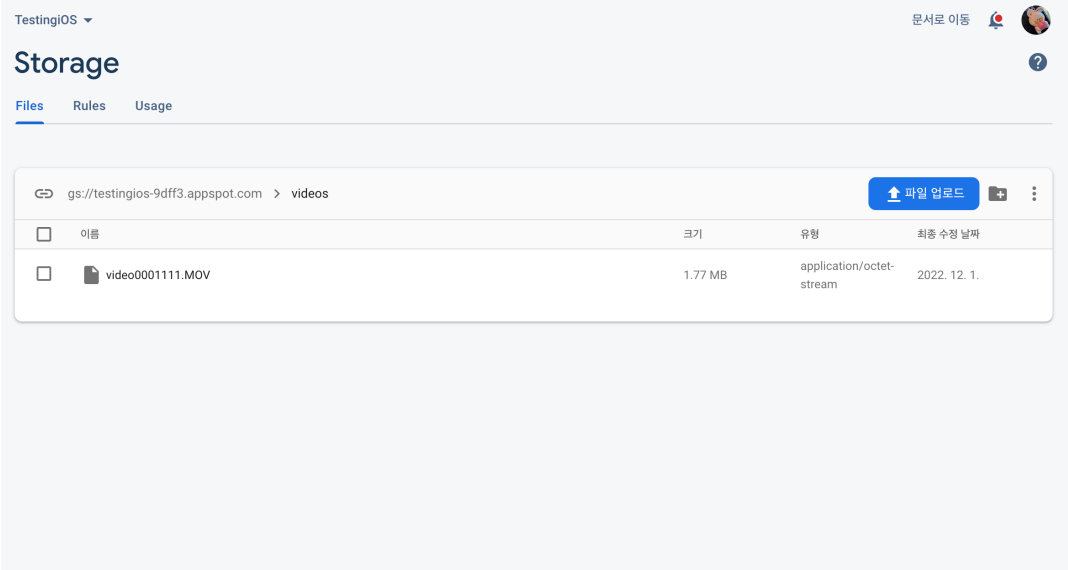
4. Implementation

System flow 4. Streaming and Recording while controlling

- Translated to h264 format through AVFoundation and transmitted to the iOS screen in the real-time
- Automatically stored when the flight is over
- Video recordings take place in separate threads.

4. Implementation

System flow 5. Upload to Firebase



"Fig. N. Result screenshot of uploading to Firebase"

Experiment Design with Drone Gesture Controller

6 Subjects

- People who haven't used a drone before
- Three women and three men
- Over 22 ages

2 types of tasks

- One is to check up and down movement
- One is to check left and right movement
- Common task is reading over obstacles
- Each three people is assigned a task

Environment

- KSW Square building
- With 4 assistant for safe experiment
- One obstacle
- One sign over the obstacle

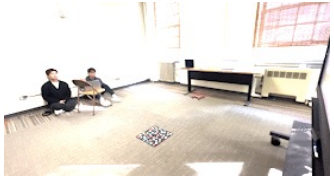
Evaluation

- Total experiment time
- Time to pass each task during the experiment
- Total number of trials
- Number of successes and failure
- Number of hits

5. Conclusion

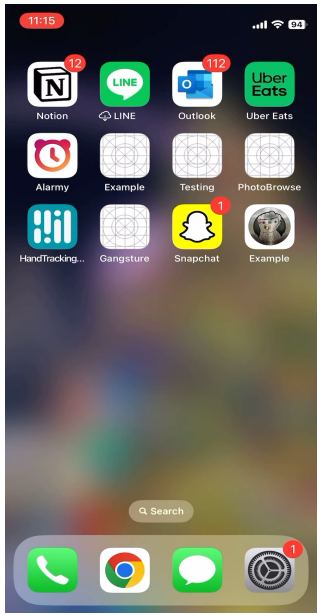
Experiment Results

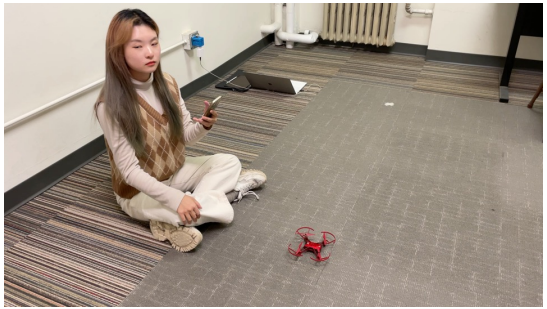
Subject Number	Trial	Success	Fail	Hits	Total Time	T1	T2	T3	T4	Note
1	2	1	1	1	103	29.5	35	25	29	hit
2	3	1	2	1	132.67	52	44	18	50	hit and landing miss
3	1	1	0	0	136	35	43	15	53	
4	1	1	0	0	197	38	26	20	113	
5	1	1	0	0	116	29	22	9	56	
6	2	1	1	1	123.5	53.5	41	25	74	hit
Total	1.67	1.00	0.67	0.50	134.69	39.50	35.17	18.67	62.50	

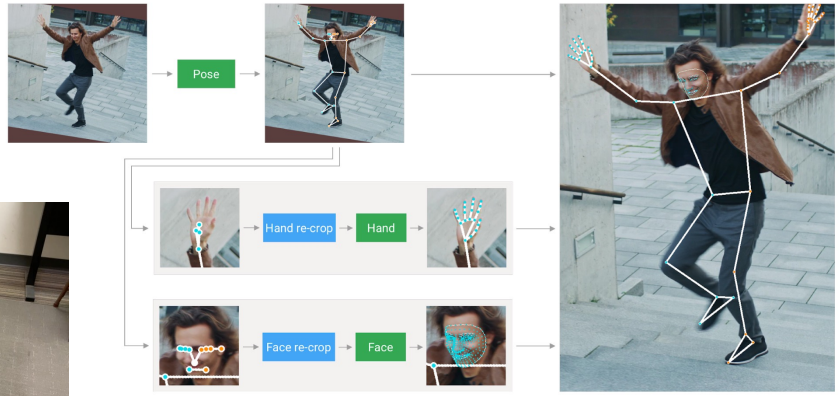


5. Conclusion

Demo and Future Plan







It is impossible to control the drone with gestures at a moment when the user's face is invisible.
If the body estimation, such as Holistic Solutions, would be used, control will be supported.

Reference

[1] A. Staff, “Amazon prime air prepares for drone deliveries,” *US About Amazon*, 13-Jun-2022. [Online]. Available:[Accessed: 13-Dec-2022].
[2] M. Murison, “13 ways commercial drones are changing work in 2022,” *DJI*. [Online]. Available: [Accessed: 13-Dec-2022].
[3] T. Editor, T. Editor, A. white says: and T. E. says: “UK Airprox board reports 30% rise in drone incidents,” *UAS VISION*. [Online]. Available: [Accessed: 13-Dec-2022].
[4] M. Asim, D. R. N. Ehsan, and K. Rafique, “Table 1 from probable causal factors in UAV accidents based on Human Factor Analysis and classification system: Semantic scholar,” *Table 1 from PROBABLE CAUSAL FACTORS IN UAV ACCIDENTS BASED ON HUMAN FACTOR ANALYSIS AND CLASSIFICATION SYSTEM | Semantic Scholar*, 01-Jan-1970. [Online]. Available [Accessed: 13-Dec-2022].
[5] Youngjae Yoo, Yunjung Kim, and Kwang-Soo Cho, “A Study of Gesture Vocabularies for User Centered Drone Control Interface,” *dbpia*. [Online]. Available: [Accessed: 13-Dec-2022].
[6] C. -C. Tsai, C. -C. Kuo and Y. -L. Chen, "3D Hand Gesture Recognition for Drone Control in Unity," 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), 2020, pp. 985-988, doi: 10.1109/CASE48305.2020.9216807.
[7] J. Jun, B. Zhao, P. Zhang, F. Jia and P. Fang, "Interaction with UAV Formation based on Gesture Recognition with Leap Motion," 2021 33rd Chinese Control and Decision Conference (CCDC), 2021, pp. 3275-3280, doi: 10.1109/CCDC52312.2021.9602101.
[8] B. Hu and J. Wang, "Deep Learning Based Hand Gesture Recognition and UAV Flight Controls," 2018 24th International Conference on Automation and Computing (ICAC), 2018, pp. 1-6, doi: 10.23919/IConAC.2018.8748953.
[9] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, doi: 10.1109/TPAMI.2019.2929257.
[10] M. Bhat, G. Mahto, S. Kesaria, V. Femandes and K. Arya, "Real-time gesture control UAV with a low resource framework," 2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA), 2021, pp. 19-24, doi: 10.1109/IRIA53009.2021.9588681.
[11] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C., Yong, M.G., Lee, J., Chang, W., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. *ArXiv, abs/1906.08172.*
[12] A. Osipov and M. Ostanin, "Real-time static custom gestures recognition based on skeleton hand," 2021 International Conference "Nonlinearity, Information and Robotics" (NIR), 2021, pp. 1-4, doi: 10.1109/NIR52917.2021.9665809.
[13] S. Sreenath, D. I. Daniels, A. S. D. Ganesh, Y. S. Kuruganti and R. G. Chittawadigi, "Monocular Tracking of Human Hand on a Smart Phone Camera using MediaPipe and its Application in Robotics," 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), 2021, pp. 1-6, doi: 10.1109/R10-HTC53172.2021.9641542.

Reference

[14] F. Naseer, G. Ullah, M. A. Siddiqui, M. Jawad Khan, K. -S. Hong and N. Naseer, "Deep Learning-Based Unmanned Aerial Vehicle Control with Hand Gesture and Computer Vision," 2022 13th Asian Control Conference (ASCC), 2022, pp. 1-6, doi: 10.23919/ASCC56756.2022.9828347.

[15] F. Patrona, I. Mademlis and I. Pitas, "An Overview of Hand Gesture Languages for Autonomous UAV Handling," 2021 Aerial Robotic Systems Physically Interacting with the Environment (AIRPHARO), 2021, pp. 1-7, doi: 10.1109/AIRPHARO52252.2021.9571027.

[16] “Hands,” *mediapipe*. [Online]. Available: https://google.github.io/mediapipe/solutions/hands.html. [Accessed: 13-Dec-2022].

[17] “Face-recognition,” *PyPI*. [Online]. Available: https://pypi.org/project/face-recognition/. [Accessed: 13-Dec-2022].

[18] “Cascade classifier,” *OpenCV*. [Online]. Available: https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html. [Accessed: 13-Dec-2022].

[19] “SKLEARN.DECOMPOSITION.PCA,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html. [Accessed: 13-Dec-2022].

[20] “1.4. Support Vector Machines,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html. [Accessed: 13-Dec-2022].

[21] “MediaPipe on IOS,” *mediapipe* . [Online]. Available: https://google.github.io/mediapipe/getting_started/ios.html#prerequisite. [Accessed: 02-Dec-2022].

[22] Szotp-Lc, “SZOTP-LC/Handtracker: Detect 3D hand landmark from video stream.,” *GitHub* . [Online]. Available: https://github.com/szotp-lc/HandTracker. [Accessed: 02-Dec-2022].

[23] J. Brownlee, “Python threading: The Complete Guide,” *Super Fast Python*, 19-Oct-2022. [Online]. Available: https://superfastpython.com/threading-in-python/. [Accessed: 09-Dec-2022].

[24] 911 security, “Data link,” *Drone Communication*. [Online]. Available: https://www.911security.com/learn/airspace-security/drone-fundamentals/drone-communication-data-link. [Accessed: 09-Dec-2022].

[25] N. Cast, “How drones communicate with the controller?,” *Remoteflyer*, 26-Oct-2021. [Online]. Available: https://www.remoteflyer.com/how-drones-communi with-the-controller/#h-wi-fi-controls. [Accessed: 09-Dec-2022].

[26] “Tello SDK 2.0 user guide - dl-cdn.ryzerobotics.com.” [Online]. Available: https://dl-cdn.ryzerobotics.com/downloads/Tello/Tello SDK 2.0 User Guide.pdf. [Accessed: 09-Dec-2022].

THANK YOU!

K-SW Fall Program 2022

Drone Gesture Control Using Vision AI in Real-Time

Team Gangsture

Hyeongbin Park · Yujin Lee · Soeun Lee · Seunghwan Kim · Xinning Bai · Jake Hazelton

PURDUE UNIVERSITY | Polytechnic Institute