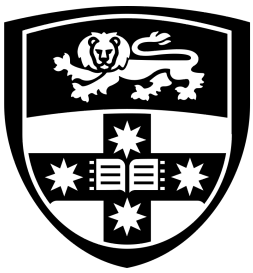

THE UNIVERSITY OF
SYDNEY

School of Aeronautical, Mechanical and Mechatronic Engineering

Hand Gesture Controlled Media Player

Name	Email	SID
Siwon Kang	skan6221@uni.sydney.edu.au	510440895
Varunvarshan Sideshkumar	vsid0095@uni.sydney.edu.au	520534445
Arin Adurkar	aadu5646@uni.sydney.edu.au	520587980

A project proposal submitted for the UoS:

AMME5710: Computer Vision and Image Processing

October 9, 2025

1 Project Description

The proposed project is a **real-time hand gesture–controlled media player** that allows users to manage playback functions using only hand movements. A webcam captures live video, detects hand landmarks, and classifies gestures into commands such as play, pause, next, and volume control. The system integrates computer vision and machine-learning techniques to achieve accurate, low-latency, and lighting-robust gesture recognition.

2 Proposed Methodology

The project applies fundamental computer vision and image processing concepts taught in AMME5710 to design a real-time **Hand Gesture–Controlled Media Player**. The system is structured into five main stages:

- **Segmentation:** Skin regions are detected using colour thresholding in HSV. This isolates the hand area under varying lighting conditions, improving robustness against background interference.
- **Feature Extraction:** Key shape descriptors such as contour outlines, convex-hull defects, fingertip positions, and centroid coordinates are computed from the segmented hand mask. These geometric features represent the underlying gesture structure.
- **Classification:** Extracted features are passed to a trained model for gesture recognition. Algorithms under consideration include *k*-Nearest Neighbour (k-NN), Support Vector Machine (SVM), Decision Tree, and compact CNN architectures like MobileNetV2-tiny for lightweight inference.
- **Decision and Control:** Gesture predictions are stabilised through short temporal voting windows (300–500 ms). Each recognised gesture is mapped to a corresponding media function—such as play, pause, volume up/down, or track change—executed via automation libraries like `pyautogui`.

3 Evaluation Plan

3.1 Datasets and Validation

The evaluation will employ both public and in-house datasets. Public datasets such as *HaGRID* (for static hand gestures) and *Jester* (for dynamic motion gestures) will be used for benchmarking and transfer learning. To ensure dataset relevance, an additional in-house dataset will be collected from 8–12 participants under five distinct lighting conditions, with roughly 200 gesture samples per participant. Cross-validation will be performed using a leave-one-subject-out strategy to evaluate generalisation performance and inter-user robustness.

3.2 Performance Metrics

System performance will be measured across several key metrics. Overall classification **accuracy** and class-wise **F1-scores** will assess recognition reliability, while **latency** will quantify frame-to-command delay, targeted to remain below 100 ms for real-time usability. **Robustness** will be examined under varying illumination and background conditions, drawing on radiometric principles of lighting variation and image normalisation covered in the AMME5710 lectures:contentReference[oaicite:1]index=1. A **confusion matrix** will further analyse inter-class errors and identify gesture ambiguities. Together, these measures will ensure the proposed system performs efficiently, accurately, and consistently across users and environmental conditions.

3.3 Expected Outcome

The expected outcome is a fully functional prototype capable of accurate and low-latency gesture recognition for real-time media control, evaluated against established computer-vision benchmarks and validated with experimental data.