iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68160

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Landmark-based Dataset Generation using Mediapipe

Dheeraj P[1], Ananya C[2], Afreena A[3], Raniya Rahoof E[4]
Department of Information Technology, University of Calicut

Abstract: ***This paper focuses on creating structured landmarks-based feature extraction using MediPpipe. MediaPipe is an open-source framework for building pipelines to perform computer vision inference over arbitrary sensory data such as video or audio. Hand and facial expression recognition play a significant role in various domains like Human-computer interaction, assistive technology and emotion analysis. Traditional datasets primarily rely on raw images, which pose challenges in terms of computational complexity and privacy concerns. This paper represents a alternative approach for dataset creation by extracting structured landmarks-based representation for hand gestures and facial expressions using MediaPipe***
Keywords: Mediapipe, dataset creation, landmark-based recognition, hand gesture recognition, face landmark detection

I. INTRODUCTION

Communication between Humans is not limited to verbal interaction; gestures and facial expressions play a significant role in expressing emotions and intentions. Proper identification of these non-verbal cues can significantly improve the quality of AI/ML applications such as sign language interpretation, virtual reality interactions, and mental health assessments. Such applications require a large number of high-quality datasets. Reliance on full image dataset can be computationally intensive and pose great privacy concerns. Our paper introduces a dataset generation methodology based on landmark extraction. By leveraging MediaPipe's hand and face tracking capabilities, we generate datasets containing only the necessary key points, making it ideal for training deep learning models.

II. RELATED WORK

Several datasets have been created for Hand and face gesture recognition using landmark extraction methods. However, a wide range of approaches exists for extracting and storing landmark-based data, each one with its own advantages and disadvantages. This paper specifically focuses on a structured methodology for creating a landmark-based dataset using MediaPipe framework, highlighting its efficiency, scalability and application in various tasks.

III.METHODOLOGY

A. Data Collection Approach

MediaPipe Hands and Face Mesh for extracting 3D landmark points, along with python for processing data and storing the extracted landmarks into a CSV file. The CSV file contains 3D coordinates of detected landmarks along with labels of each class.

B. Landmark Extraction

For hand gesture recognition, it captures 21 key landmark points per hand and supports detection for both hands. In the case of Facial Expression Recognition, the MediaPipe extracts 468 facial landmarks to capture various expressions.

C. Data Preprocessing

To ensure the quality of data, various preprocessing steps such as normalization, handling missing landmark points and filtering out low-confidence detections are applied. This step increases the accuracy and robustness of machine learning models trained on the dataset.

D. Data Annotation and Labelling

Each captured frame or image is assigned a class label corresponding to the gesture or facial expression being performed. The labelling process is critical for supervised learning models and involves manual verification to reduce annotation mistakes or errors.

©IJRASET: All Rights are Reserved

TABLE I
LANDMARK KEY POINTS TABLE

Number of Landmarks	Details
21 per hand	Key points for fingers and palm tracking
468	Detects facial features for expression analysis

E. Implementation

Fig. 1. Illustrates a sample implementation of the technique discussed above, showcasing the process of extracting and storing landmark-based data

```python
import os
import csv
import cv2
import mediapipe as mp

# Initialize MediaPipe Hands
mp_hands = mp.solutions.hands
hands = mp_hands.Hands(static_image_mode=True, max_num_hands=2, min_detection_confidence=0.5)
mp_drawing = mp.solutions.drawing_utils

# Create/open a CSV file
csv_filename = "hand_landmarks.csv"
csv_file = open(csv_filename, 'w', newline='')
csv_writer = csv.writer(csv_file)
header = [f"hand{i+1}_x_{j}", f"hand{i+1}_y_{j}", f"hand{i+1}_z_{j}" for i in range(2) for j in range(21)]
header.append("label")
csv_writer.writerow(header)
```

Fig. 1 A sample Implementation in python

F. Data storage Format

Fig. 2. Shows a sample of the extracted landmark data stored in CSV file format. Each row represents a single frame, with columns corresponding to the 3D coordinates of the detected hand and face landmarks, together with their respective labels

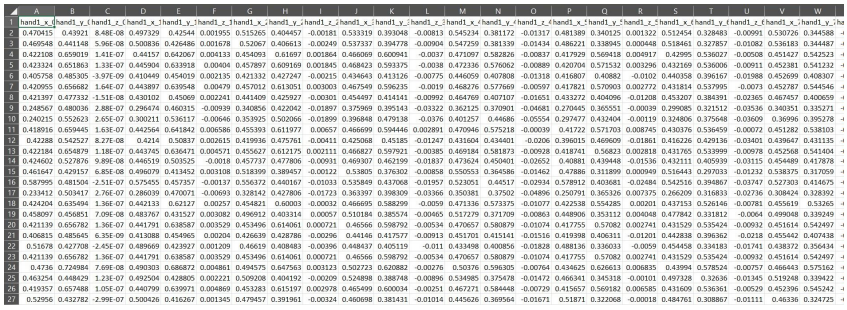


Fig. 2 Sample Landmark Data Stored in CSV Format

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

IV. APPLICATIONS OF LANDMARK-BASED DATASETS

Landmarks-based datasets have a wide range of applications, across various domains. They are extremely efficient for training models for sign language interpretation and touchless control systems. In facial expression analysis, they aid in emotion detection for healthcare and entertainment applications. Additionally, these datasets enhance privacy-protection by reducing reliance on raw images while maintaining recognition accuracy. Furthermore, they contribute to multimodal fusion techniques that combine facial and hand gestures for improved human-computer interaction models.

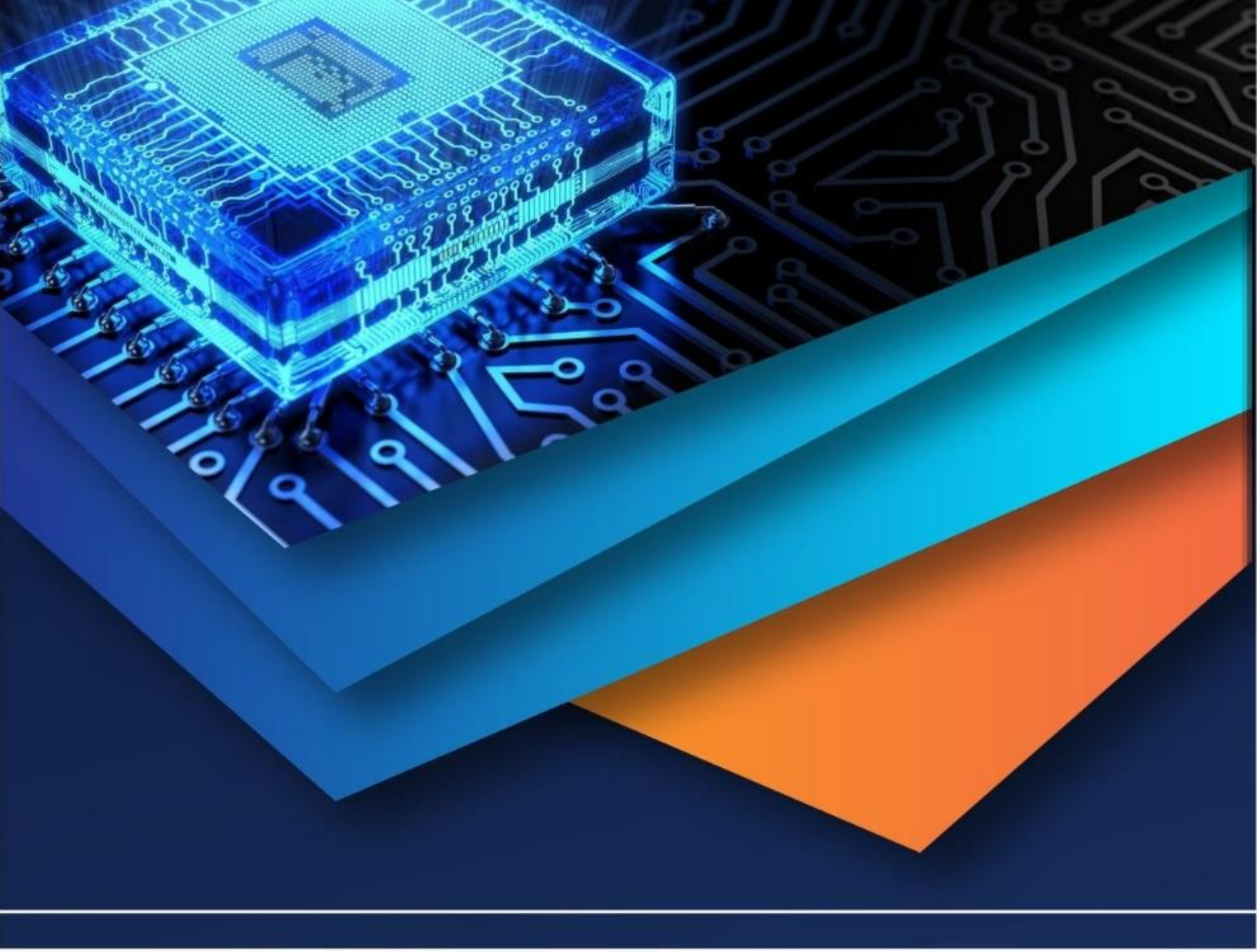
V. CONCLUSION

This paper introduces a structured, landmark-based dataset for hand gesture and facial expression recognition, giving a computationally efficient and privacy-focused alternative to traditional image-based datasets. By making use of the MediaPipe framework for three-dimensional landmark extraction, we enable the development of robust AI models for a wide range of applications like sign language detection, human-computer interaction and emotion detection. Future work includes expanding dataset diversity, incorporating motion dynamics, and benchmarking deep learning models on the dataset.

REFERENCES

[1] Google AI, “MediaPipe Solutions Setup for Python”, Google AI Developer Guide, Retrieved 10:49, February 10, 2025, from https://ai.google.dev/edge/mediapipe/solutions/setup_python
[2] Python Software Foundation, “csv- CSV File Reading and Writing,” *Python 3 Documentation*, Retrieved 10:49, January 12, 2025, fromhttps://docs.python.org/3/library/csv.html
[3] A. Weyand, M. A. Babenko, S. Cao, L. Shen, A. Kolesnikov, R. Philbin, and T. Weyand, "Google Landmarks Dataset v2 – A Large-Scale Benchmark for Instance-Level Recognition and Retrieval," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2020, pp. 1-10
[4] OpenCV Contributors, "OpenCV: OpenCV-Python Tutorials," OpenCV Documentation, Retrieved 10:49, January 15, 2025, from https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)