Face, Expression And Gesture Recognition & Compilation In Database

1Minal Tayde, 2Dr. Rupesh C. Jaiswal, 3Dr. Girish P. Potdar, 4Prof. Manish R. Khodaskar
1Student, 2Professor, 3Professor, 4Professor
1,2Department of E & Telecommunication Engineering,
3, Department of Computer Engineering,
4, Department of IT Engineering,
1,2,3,4SCTR’s Pune Institute of Computer Technology, Pune, India.

ABSTRACT

A database that encompasses all potential combinations of suspects' external variables is only available to the government's upper echelons, putting more pressure on researchers to collect their own data or compile groups of datasets. To fix this flaw, we suggest a system for acquiring facial and behavioural data in unregulated and real-life scenarios. For the government and private organizations to maintain security, this project would be beneficial. We intend to observe and analyse a fully computerize process of crime registration, investigation, prosecution etc. and to build a central database along with time and location stamps that will serve as a national search engine as well as crime analytics for police and other government agencies. The database that has been created will be utilised to identify suspects from CCTV footage of crimes that has been recorded by a number of systems that are situated near crime scenes and on major thoroughfares. We also make some attempts to electronically share data/information among Police Stations & Police Higher Offices.

1. INTRODUCTION

Using a distributed CCTV system, this technique captures the face, gesture, and expression of the intended individuals (criminals) and records that data in a database along with a time and location stamp. The process of recognizing and interpreting symbolic gestures of the hands[6], arms, face[2][4], and sometimes head is known as face/gesture[3][4] recognition. It is critical to create a user-friendly human-computer interface to automate this operation. The technology has been under investigation in recent years due to its potential for use in user interfaces. One of the key areas of study for engineers and scientists is gesture recognition [6-9]. Today, the industry is developing various applications to make interactions more normal, simple, and convenient without the use of any additional devices. Many prominent figures in the field of gesture recognition systems have promoted their solutions to solve different problems. We present what gesture recognition technology entails, as well as a possible solution to gesture recognition.

Also, role of ML and ESPs [10-76] are becoming important in recent applications, recognition and control.
2. Activity Diagram

The system shown in Figure 1 allows to record the targeted individuals' faces, expressions, and movements and save them in a database. Provide sample for matching and then extract features to recognize the below and to store data of criminal in CSV file.

1. Face detection and Recognition
2. Weapon detection
3. Posture detection and Recognition
4. Hand gesture recognition
5. Facial expression recognition
3. Algorithms

3.1. Face Detection & Recognition Module

MTCNN

A three-stage technique called MTCNN (Multi-task Cascaded Convolutional Neural Networks) recognises the bounding boxes and five-point face landmarks of faces in an image. By putting its inputs through a CNN, which produces candidate bounding boxes with their scores, each stage gradually improves the detection results.

The first step will be to give the software a picture to work with. We want to build a picture pyramid in this model in order to detect faces of various sizes. The image must be scaled into many copies, entered into the first neural network, P-Net, and its output must be accrued once the image has been entered. The weights and biases of P-Net have been tuned to provide a reasonably correct bounding box for each 12 x 12 kernel. In pictures, it's common for only a piece of a face to be visible peering in from the side of the frame. The network might then deliver a bounding box that is only partially visible in that situation. The new array for will contain many columns of 0s close to the right edge and pixel values distributed across the box's left side. The process of padding involves adding 0s to arrays. We can now feed these refined images into R-Net and collect the results. It provides the coordinates of the latest, more precise bounding boxes, as well as their trust levels. We must first pad any boxes from R-Net that are out-of-bounds before passing them to the O-Net. The bounding boxes can then be resized and moved into O-Net. Compared to P-Net and R-Net, O-Net results are significantly different. The three outputs provided by O-Net are the bounding box coordinates, the five facial landmark coordinates, and the confidence level of each box.

The data must next be compiled into a dictionary using the key terms "box, trust, and keypoints." This is the last phase. The terms "box" and "trust" refer to the bounding box coordinates, "keypoints" refers to the coordinates of each facial landmark (eyes, nose, and endpoints of the mouth).

3.2. Pose estimation module:

Steps:

- To extract function maps, the image is first passed through a baseline network.
- After that, the feature maps go through multiple stages of CNN processing to create: Firstly, a set of Part Confidence Maps, and secondly, a set of Part Affinity Fields (PAFs).
- The Confidence Map is a two-dimensional illustration of the idea that any given pixel can be used to find a certain body part.
- A collection of 2D confidence maps S for the locations of bodily parts. There is a map at each joint location.
- Part Affinity Fields (PAFs) are a collection of 2D vector fields L that express how closely related different parts are to one another.

![Figure 2: Overall Pipeline of the Open Pose architecture.](image)

IJCRT2212407 | International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org | d716
There are three steps that make up the parsing process:

Step 1: Using the confidence maps, locate every joint.

Step 2: Utilizing the part affinity fields and joints from step 1, determine which joints connect to create limbs (body parts).

Step 3: To gather the whole list of human positions, associate limbs that belong to the same person.

3.3. CNN-LSTM for

- Hand Gesture Recognition Module
- Weapon Detection Module
- Facial Expressions Recognition Module:

Long-short term memory networks (LSTMs) are a form of recurrent neural network that can account for sequential dependencies in a time series. Given the existence of associations between observations in a given time series (a phenomenon known as autocorrelation), a standard neural network will incorrectly treat all observations as independent, resulting in misleading results.

Convolutional Neural Network (CNN) layers are used in the CNN LSTM architecture to extract features from input data, and LSTMs are used to assist sequence prediction. For the purpose of producing textual descriptions from sequences of images and visual time series prediction issues, CNN LSTMs were created.

In order to generate textual descriptions from sequences of images, CNN LSTMs were created for visual time series prediction issues (e.g. videos). In particular, the issues with:

- **Activity Recognition**: Producing a textual explanation of a task that is illustrated by a series of visuals.
- **Image Description**: Producing a written description for a single image.
- **Video Description**: Creating a text description of an image series.

4. GUI

4.1. Implementing GUI:

Implementation of GUI involves designing and writing design code for each user screen. Implementation of GUI is done using HTML, CSS, and JavaScript with Electron framework and Flask framework is used to expose CSV files to open ended API.

HTML is used to design the GUI Screen of the desktop app which will be used by Admin. HTML and CSS are used along with each other. CSS describes how HTML elements are to be displayed on screen. Electron is a framework for creating native applications with web technologies like JavaScript, HTML, and CSS. An app built with Electron behaves...
like a web app, but it can read and write data in the computer’s file system. The Electron app usually reuses the business logic, design and general structure of a web app. All HTML files are stored under template folder and media under static folder. It’s a great way to save time and money on the business and development side.

Also, Flask framework is used to run this application on the web. Flask is a web framework. It provides tools, libraries and technologies that allow you to build a web application. Flask enables secure transmission of data from local machines, which other LAN connected machines can also use. In this system all results are stored in CSV File which is popular among machine learning models as a means of training. This CSV can be exposed to an API easily with the flask framework. A summary of collected data will be displayed.

GUI Screenshots:

1. Login Screen

2. Upload Video/ Use camera screen:

3. Face Detection, Weapon Detection, Pose Estimation, Hand Gesture Estimation, Facial Estimation Screen:
4. Final Result Screen:

CONCLUSION

Through a distributed CCTV system, our system will record the faces, gestures, and facial expressions of the targeted individuals (criminals or suspects) and store them in a database. The database created will be used to identify suspects using CCTV footage of crimes that was recorded by a number of systems situated along roads and near to the crime scene. This system will fully automate the suspect identification procedure and offer better ways to lower the rising crime rate.

ACKNOWLEDGEMENT

I would like to thank all the members who have helped me in completing this research work and paper. I express my heartiest gratitude to Dr. Rupesh Jaiswal Sir for his guidance and constant supervision as well as for providing necessary information to complete our research work. I would also like to express my gratitude towards our parents and friends for their kind co-operation and encouragement which helped me in completion of this research work.
REFERENCES


