International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST222938

Face, Expression and Gesture Recognition and Compilation in Database Using Machine Learning

Prof. Prashant Wakhare[1], Vaishnavi More[2], Rutuja Surdi[2], Kajal Patil[2], Vishwadip Ingale[2]

[1]Professor at Information Technology Department, AISSMS Institute of Information Technology, Pune, Maharashtra, India

[2]B.E Scholar, Information Technology Department, AISSMS Institute of Information Technology, Pune, Maharashtra, India

Article Info
Volume 8, Issue 3
Page Number : 348-352

Publication Issue
May-June-2021

Article History
Accepted : 18 May 2021
Published : 24 May 2021

ABSTRACT

In today's scenario, numbers of crimes have increased day by day. At many public places government has placed many CCTV cameras so police can get that CCTV footage to identify the suspects but sometimes it becomes difficult to recognize the criminals So here we have come up with a solution to make this process smooth, easier than the traditional one. The system which automates all the suspect recognition process and provides better solutions to reduce the increasing rate of crimes. We plan to design a system to capture face, expressions and gestures of the targeted people (Criminals) through distributed CCTV System and are maintaining it in a database along with time and location stamp. The compiled database will be used to identify suspects from video clips of crime related CCTV footage captured series of CCTV Systems located on routes and close to scene of crime. This research discusses the various types of methodologies that can be used to identify the suspects which are captured in CCTV footage and convert it into useful information for further analysis of particular crime cases.

Keywords : Artificial Intelligence, Machine Learning, OpenCV, CCTV, OpenFace, LSTM, YOLOV3

I. INTRODUCTION

Security is always a main concern in every domain, due to a rise in crime rate in a crowded event or suspicious lonely areas. Abnormal detection and monitoring have major applications of computer vision to tackle various problems. Due to growing demand in the protection of safety, security and personal properties, needs and deployment of video surveillance systems can recognize and interpret the scene and anomaly events play a vital role in intelligence monitoring. At many public places government has placed many CCTV cameras so Police can get that CCTV footage to identify the suspects but sometimes it becomes difficult to recognize the criminals So here we have come up with a solution to

make this process smooth and easier than the traditional one. We plan to design a system to capture face, expressions and gestures of the targeted people (Criminals) through distributed CCTV System and are maintaining it in a database along with time and location stamp. The database so compiled will be used to identify suspects from video clips of crime related CCTV footage captured series of CCTV Systems located on routes and close to scene of crime. Video surveillance systems using Closed Circuit Television (CCTV) cameras, is one of the fastest growing areas in the field of security technologies. However, the existing video surveillance systems are still not at a stage where they can be used for crime prevention. The systems rely heavily on human observers and are therefore limited by factors such as fatigue and monitoring capabilities over long periods of time. This work attempts to address these problems by proposing an automatic suspicious behaviour detection which utilises contextual information.

II. LITERATURE SURVEY

Various researchers have done excellent work on Expressions identification. Propositions of few authors are discussed in this section. M. Owayjan et al. have proposed a method with Viola-Jones object detection framework and ANN. With this method author identifies happiness, anger and neutral facial expression. For expression “happy”, “angry” and “neutral” 60%, 55% and 80%accuracy is achieved respectively [1].J. Jayalekshmi et al. have proposed a method in which Viola Jones algorithm is used for face detection and three methods Zernike moments, Local Binary Pattern and Discrete Cosine Transform used for feature point extraction which will take more time as compare to other methods. After extraction all the feature points are combined together by means of Normalized Mutual Information Selection method. The whole system is trained and classified using SVM, RF and KNN classifiers.[2]

Appearance-based techniques extract image features to model the visual appearance of the hand and compare these features with the extracted features from the video frames as our approach. They have real-time performance because of the easier 2-D image features that are used. A simple method, searching for skin coloured regions in the image, was used in [3].However, this method has some shortcomings; first, it is very sensitive to lighting conditions. Secondly, it is required that no other skin-like objects exist in the image. In [4], scale-space colour features are used to recognize hand gestures, which are based on feature detection and user independence. However, the system shows real-time performance only when no other skin-coloured objects exist in the image. The authors of [5] obtained a clear-cut and integrated hand contour to recognize hand gestures and then computed the curvature of each point on the contour. Due to noise and unstable illumination in the cluttered background, the segmentation of integrated hand contour had some difficulty. The eigenspace is another technique, which provides an efficient representation of a large set of high-dimensional points using a small set of basis vectors. However, eigenspace methods are not invariant to translation, scaling, and rotation.

Primarily, research on gun detection focuses on Concealed Weapon Detection (CWD) and knife detection. CWD is stand on some techniques of imaging like infrared imaging, millimeter wave imaging, in application of luggage control at airports.

Halima, N.B. et al.[6] demonstrated that BoWSS (Bag of Words Surveillance System) algorithm has a high potential to detect guns. They first extract features using SIFT, cluster the obtained functions using K-Means clustering and use SVM (Support Vector Machine)for the training. Sheen et al.[7] proposed a method of CWD, based on three dimensional millimeter (mm) wave imaging technique, to detect concealed weapon in the body at airports and other secure location. By using2-D millimeter wave imaging, they modelled a 3-D image for the target. 3-D image

from gathered data of 2-D image can be formed from three dimensional imaging systems or wide wand imaging.

Princeton Tracking Benchmark[8] introduced 2013 consists of 100 RGB-D tracking datasets with tracking software and online submission script. Datasets include real world footage of variety of actions performed by humans, pets and object presentations in form of RGB (8 bit PNG format) and 1100 accompanying depth (16 bit PNG format) images. Annotations are per-frame in a form of bounding box covering target object only. These datasets are more related to tracking than to action recognition but could serve as a starting point in segmentation of scenery. North western-UCLA Multi view Action 3D Dataset[9] contains RGB, depth and human skeleton data captured simultaneously by three Kinect cameras. This dataset include 10 action categories: pick up with one hand, pick up with two hands, drop trash, walk around, sit down, stand up, donning, doffing, throw, carry. Each action is performed by 10 actors in a library from a variety of viewpoints.

III. COMPARISON AND ANALYSIS

Literature [1] propose a Human face detection and facial Expression identification system and it is implemented using Zernike moment which is used to identify facial expressions. A support vector classification algorithm is used. The support vector classification algorithm provides high accuracy as compared to other classification techniques. Limitations for this survey is facial expression and emotion accuracy can be further explored to identify whether a person's emotions are in accordance with facial expressions or not. Similarly, In literature[4] Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support which is implemented using Hand detection and tracking using face subtraction using SIFT algorithm. And some of the limitations are three important factors affect the accuracy of the system, which is the quality of the webcam and testing stages, the number of training images, and choosing a number of clusters to build the cluster model.

Vector Machine Techniques In Literature [7] of Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications this paper uses SSD and Faster RCNN algorithms that are simulated for pre-labelled and self-created image dataset for weapon (gun) detection. In which this SSD provides real-time detection due to faster speed but Faster RCNN provided superior accuracy but similarly it has some limitations too that is it can be implemented for larger datasets by training using GPUs and high-end DSP and FPGA kits. In literature[8] A Survey on Human Activity Recognition from Videos this paper contains methodology as it uses Various Human Activity Recognition algorithms, various methodologies adapted for human object interaction in both still images and videos. In Human

Activity Recognition depicts major progress in various aspects. It has some limitation like it does not addressed the various challenges of activity recognition like incorporating the context of the scene, human-object interaction in videos.

TABLE 1: COMPARISON OF VARIOUS TECHNIQUES USED

SR.No.	Authors	Algorithms/Technologies Used
1	M. Owayjan et al.	Zernike polynomials, Support Vector Machine
2	Prem Jacob	SSD and Faster RCNN algorithms are simulated for pre-labeled and self-created image dataset for weapon (gun) detection.
3	Halima, N.B. et al.	Hand detection and tracking using face subtraction using SIFT algorithm.

4	Princeton Tracking Benchmark	combination of linear discriminant analysis and artificial neural networks.

Analysis:

According to survey, all the techniques which are used by previous authors have some limitations. To deploy this application the main challenge we were facing is to combine all this separate modules into one. So that we will get all models altogether and the efficiency will be increased accordingly. The second challenge is that we are using supervised machine learning algorithms so we need large amount of data for comparison and processing of each individual model. So we have found some algorithms which uses less amount of data but acquires high accuracy as compared to other techniques which we have seen briefly in literature survey. Basically proposed system is divided into five modules they are respectively:

1. Face expression detection
2. Weapon detection
3. Hand gesture recognition
4. Pose estimation.

All these modules are combined together to get the result in pie chart form for further analysis. And extension of this proposed system include Application Programming Interface which can be available publically for commercial and private users who need it so that can be used to deploy this kind of application that needs criminal's data for identification. The existing system does not have this feature. We tried to find out which systems are mostly preferred for analysing the details of the selected papers. However, we found that most of the papers containing general definitions and there were insufficient information on the technical implementation details. So we figured out following algorithms that can be used to reach higher accuracy and better results.

For Face and expression recognition we have decided to use OpenFace algorithm which is a Python and Torch implementation of face recognition with deep neural networks and is based on the CVPR 2015 paper FaceNet: A Unified Embedding for Face Recognition and Clustering by Florian Schroff, Dmitry Kalenichenko, and James Philbin at Google. Torch allows the network to be executed on a CPU or with CUDA.

For Weapon Detection we have planned to use YOLOv3 (You Only Look Once, Version 3) which is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. Versions 1-3 of YOLO were created by Joseph Redmon and Ali Farhadi. The first version of YOLO was created in 2016, and version 3, was made two years later in 2018. YOLO is implemented using the Keras or OpenCV deep learning libraries.

For hand gesture recognition we have planned to use OpenCV python libraries this can be implemented by following sequence of methods for recognizing hand gestures. Algorithm includes hand detection, fingers and palm segmentation, fingers recognition and finally whole hand gesture recognition.

For Pose estimation module in proposed system we have planned to use LSTM algorithms: Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning. It can be hard to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field. So this way all these models will be combined together to get all these functionalities in one application.

IV. CONCLUSION

As existing system contains all the separate modules which user needs to run manually. This will make confusion to perform all this tasks and store the results accordingly.It is very time consuming because of its manual approach. And in this fast moving world where CCTV is only capturing the scenes, the system aims at using these data like human face, expression, pose, gesture and what is human holding and converting it into useful information using different machine learning algorithms. This information will help the authorities to gain a better insight in the world of criminals/ suspects and get a readily available analysis. Thus it can act as both, crime detection as well as a crime prevention system. Which help India to be crime free nation.

V. FUTURE WORK

In future work, we will continue on the further implementation or changes in our system and we will try to research on its further performance. However, there are still some implementation that can be applied to our system. Basically our focus is on the development of more efficient and sophisticated system for face, expression and gesture recognition and compilation in database and its related variable datasets which can acquire higher accuracy as large as the dataset the result will be more accurate. And we will further explore some other algorithms which are better than the proposed system to reduce the complexity of the project.

VI. REFERENCES

[1]. M. Owayjan., R. Achkar. and M. Iskandar.: Face De-tection with Expression Recognition using Artificial NeuralNetworks. 3rd Middle East Conference on Bi-omedical En-gineering (MECBME), Beirut. IEEE MECBME, 115—119(2016).

[2]. J. Jayalekshmi and T. Mathew, Facial expression recog-nition and emotion classifi-cation system for sentiment analy-sis, 2017 International Conference on Networks & Advancesin Computational Technologies (NetACT), Thiruvanthapuram,India, 2017, pp. 1-8.

[3]. N.Paragios, (2018) Computer vision and understanding. Vol 116. pp 102-114.

[4]. Guojen Wen, Zhiwei Tong, et.al, (2009), Man machine interaction in machining center. International workshop on intelligent systems and applications. pp 1-4.

[5]. S.D. Bharkad, et.al. (2017). international conference on computing methodologies and communication, pp 1151-1155.

[6]. Nadhir Ben Halima and Osama Hosam. 2016. Bag of Words Based SurveillanceSystem Using Support Vector Machines. 10 (04 2016), 331–346.

[7]. D. M. Sheen, D. L. McMakin, and T. E. Hall. 2001. Three-dimensional millimeter-wave imaging for concealed weapon detection.IEEE Transactions on MicrowaveTheory and Techniques49, 9 (Sep 2001), 1581–1592. https://doi.org/10.1109/22.942570

[8]. S. Song and J. Xiao, “Tracking Revisited using RGBD Camera: Unified Benchmark and Baselines,”Princeton University Proceedings of 14th IEEE International Conference on Computer Vision (ICCV2013)

[9]. UCLA database http://users.eecs.northwestern.edu/~jwa368/

Cite this Article

Prof. Prashant Wakhare, Vaishnavi More, Rutuja Surdi, Kajal Patil, Vishwadip Ingale, "Face, Expression and Gesture Recognition and Compilation in Database Using Machine Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 3, pp. 348-352, May-June 2021. Available at doi : https://doi.org/10.32628/IJSRST222938
Journal URL : https://ijsrst.com/IJSRST222938