3D Face Recognition With Crowd Computing Using CNN

Nikhilesh R Vasist[1], Ravi Kiran K V[2], Sanjay P[3], Suresh Kumar C[4]
[1, 2, 3] Dept of Information Science and Engineering
[4]Asst. Professor, Dept of Information Science and Engineering
[1, 2, 3, 4] East West Institute Of Technology, Bengaluru, India

Abstract- *Face analytics has been gaining major traction in recent years along with advancement of computer science or algorithms and availability of cheap compute has been one of major reasons. Facial recognition technology is capable of detecting and identifying faces in an image. There are several methods available for this but mainly for facial recognition, each method extracts features from faces and compares them to features of given image from database. And what crowd computing does is it calculates the number of distinct faces in the photo and predicts number of people in the photo. This can be used in identifying number of people in an image as well as who they are.*

Keywords- Face recognition; Convolutional Neural Network; Softmax classifier; Deep learning

I. INTRODUCTION

Face recognition system has become one of the research topics in image processing and computer vision applied in various systems, starting from attendance systems, security systems such as CCTV, identity recognition, emotional recognition, granting system permissions and also applied to the robotics field as a way for the robot to recognize someone. Face detection itself is a pre-processing stage in facial recognition. In doing face recognition of a person, it will take a lot of face samples with various expressions, angle of retrieval and general differences attached to the face like glasses and moustache . To do face recognition of one face does not take a long time, but if we do it on the attendance system or security system on companies that have many faces to be recognized, it will take a long time. In addition, in some cases, there are deficiencies in some of the previous studies, such as face recognition not running in real time, using public datasets in testing, and computing processes that take a long time causing the computer cannot be used during the process. To solve those problems, several studies have using cloud computing technology. There are several methods available for this but mainly for facial recognition, each method extracts features from faces and compares them to features of given image from database. Cloud computing is a computing service that is done not on a local device, but on an internet connected to a data center infrastructure. The system of cloud computing also provides a scalability solution where cloud computing can increase the resources needed when doing larger data processing. Because computing is not done on a local device, then it is done between client communication with cloud server by utilizing web-service communication. In performing face recognition, it needs to perform several stages starting from face detection, feature extraction then doing face recognition. During such stages require adequate and available hardware resources at all times. If you want to do the local device, it will take a long time and besides, the device cannot be used for some time. Therefore, an approach is required to be able to perform face recognition without using the capabilities of local computing devices.

II. LITERATURE SURVEY

Face detection is a computer technology that determines the location and size of human face in arbitrary (digital) image. The facial features are detected and any other objects like trees, buildings and bodies etc are ignored from the digital image. It can be regarded as a specific case of object-class detection, where the task is finding the location and sizes of all objects in an image that belong to a given class.

Face detection, can be regarded as a more general case of face localization. In face localization, the task is to find the locations and sizes of a known number of faces . Basically there are two types of approaches to detect facial part in the given image i.e. Feature base and Image base approach. Feature base approach tries to extract features of the image and match it against the knowledge of the face features. While image base approach tries to get best match between training and testing images.

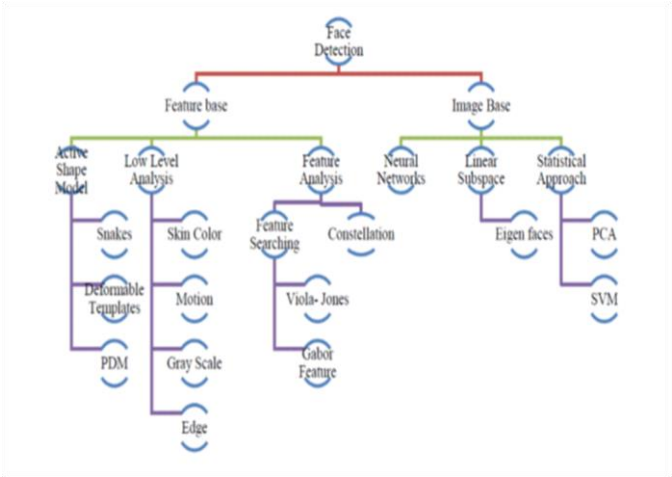

Fig 1 : Face Detection methods

A. FEATURE BASE APPROCH

Active shape models focus on complex non-rigid features like actual physical and higher level appearance of features means that Active Shape Models (ASMs) are aimed at automatically locating landmark points that define the shape of any statistically modelled object in an image. When of facial features such as the eyes, lips, nose, mouth and eyebrows. The training stage of an ASM involves the building of a statistical.

Facial model from a training set containing images with manually annotated landmarks. ASMs is classified into three groups i.e. PDM, Deformable templates

Internal energy is the part that depends on the intrinsic properties of the snake and defines its natural evolution. The typical natural evolution in snakes is shrinking or expanding. The external energy counteracts the internal energy and enables the contours to deviate from the natural evolution and eventually assume the shape of nearby features—the head boundary at a state of equilibria. Elastic energy is used commonly as internal energy. Internal energy is vary with the distance between control points on the snake, through which we get contour an elastic-band characteristic that causes it to shrink or expand. On other side external energy relay on image features. Energy minimization process is done by optimization techniques such as the steepest gradient descent. Huang and Chen and Lam and Yan both employ fast iteration methods by greedy algorithms. Snakes have some demerits like contour often becomes trapped onto false image features and another one is that snakes are not suitable in extracting non convex features.

B. LOW LEVEL ANALYSIS

Based on low level visual features like color, intensity, edges, motion etc. Skin Color Base Color is a vital feature of human faces. Using skin-color as a feature for tracking a face has several advantages. Color processing is much faster than processing other facial features. Under certain lighting conditions, color is orientation invariant. This property makes motion estimation much easier because only a translation model is needed for motion estimation. Tracking human faces using color as a feature has several problems like the color representation of a face obtained by a camera is influenced by many factors. Majorly three different face detection algorithms are available based on RGB, YCbCr, and HIS color space models.

In the implementation of the algorithms there are three main steps as follows

- Classify the skin region in the color space
- Apply threshold to mask the skin region and
- Draw bounding box to extract the face image

C. MOTION BASE

When use of video sequence is available, motion information can be used to locate moving objects. Moving silhouettes like face and body parts can be extracted by simply thresholding accumulated frame differences.

D. CONSTELLATION METHOD

All methods discussed so far are able to track faces but still some issue like locating faces of various poses in complex background is truly difficult. To reduce this difficulty investigator form a group of facial features in face-like constellations using more robust modelling approaches such as statistical analysis. Various types of face constellations have been proposed by Burl et al. They establish use of statistical shape theory on the features detected from a multiscale Gaussian derivative filter. Huang et al. also apply a Gaussian filter for pre-processing in a framework based on image feature analysis.

E. LINEAR SUB SPACE METHOD

An early example of employing eigen vectors in face recognition was done by Kohonen in which a simple neural network is demonstrated to perform face recognition for aligned and normalized face images. Kirby and Sirovich suggested that images of faces can be linearly encoded using a modest number of basis images.

The idea is arguably proposed first by Pearson in 1901 and then by HOTELLING in 1933 .Given a collection of n by m pixel training. Images represented as a vector of size m X n, basis vectors spanning an optimal subspace are determined such that the mean square error between the projection of the training images onto this subspace and the original images is minimized. They call the set of optimal basis vectors Eigen pictures since these are simply the eigen vectors of the covariance matrix computed from the vectorized face images in the training set. Experiments with a set of 100 images show that a face image of 91 X 50 pixels can be effectively encoded using only 50 Eigen pictures.

F. STATISTICAL APPROCH

SVMs were first introduced Osuna et al. for face detection. SVMs work as a new paradigm to train polynomial function, neural networks, or radial basis function (RBF) classifiers. SVMs works on induction principle, called structural risk minimization, which targets to minimize an upper bound on the expected generalization error. An SVM classifier is a linear classifier where the separating hyper plane is chosen to minimize the expected classification error of the unseen test patterns. In Osunaet al. developed an efficient method to train an SVM for large scale problems, and applied it to face detection. Based on two test sets of 10,000,000 test patterns of 19 X 19 pixels, their system has slightly lower error rates and runs approximately 30 times faster than the system by Sung and Poggio. SVMs have also been used to detect faces and pedestrians in the wavelet domain.

III. PROPOSED SYSTEM

The main objective of project is to estimate head pose and provide 3-D coordinates of same and using videos/images as inputs we prepare output containing 3-D coordinates of head pose (Roll, Yaw, Pitch) and 68 3-D key points.

The following data points will be captured :

- Face detection
- 68 3D facial key points.
- Pose data.

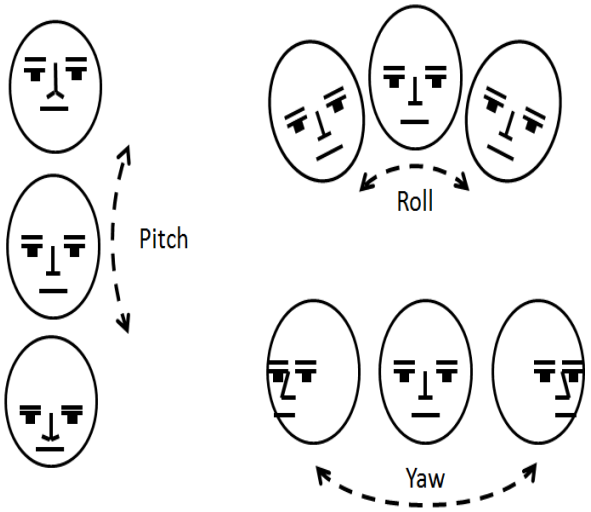

Fig 2 : 3D coordinates of the Head pose

PROCESS

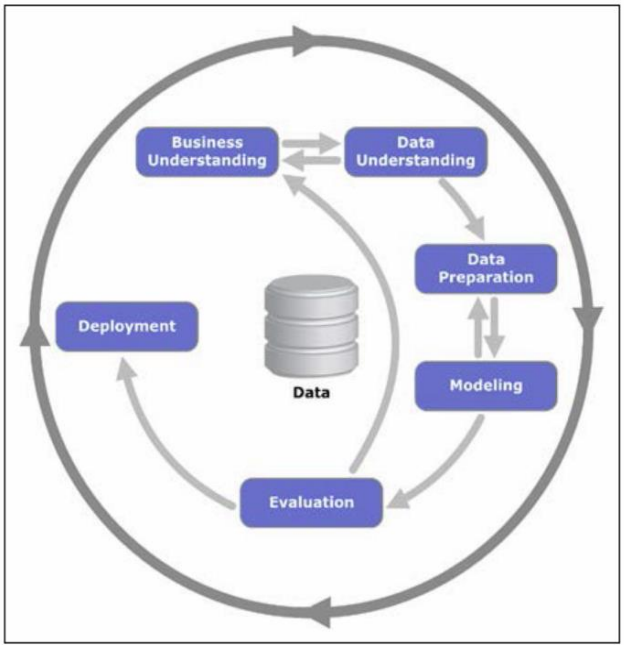

Fig 3 : Process of the Proposed System

i. Data collection, Data exploration, Data cleaning

- We collected video and picture from internet.
- Downloaded videos from YouTube and images using Bing Image search API.

ii. Analytical methods and Technology.

- We have used CUDA and GPU
- We have used TensorFlow
- Deep learning models and techniques

iii. Interpretation of Output/Visualization.

- We got combined set of images as output after combining our models.
- In the first image it is detecting facial landmarks of the face - that is 68 key points.
- Second gives the 3D points and 3D face detection.
- In the third part we get the bounding box around the face and pose that is the three axes, in Red, green, Blue, that is Roll, Pitch, Yaw.
- Reference image for this is given below

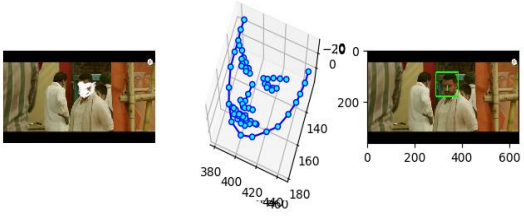

Fig 4 : Detection of 68 key points and Head pose

MODEL COMPARISION

Our main step was to get the three different models and see if they are giving us the result we are expecting. First model was to detect faces in a given image.We tried various models for face detection listed below so that we can get the 99% of the faces detected.

- ResNeXt
- Haar Cascades
- Mobilenet_v2
- MTCNN

MATLAB

MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar non interactive language such as C or Fortran.The name MATLAB stands for matrix laboratory. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects, which together represent the state-of-the-art in software for matrix computation.

MATLAB features a family of application-specific solutions called toolboxes. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others.

IV. EXPERIMENTAL DATA AND RESULTS

- We decided to go with MTCNN for face detection, as it gave us the desired result with 99% accuracy and almost all the faces detected given a frame, as we were looking for a model which detects all the faces present in the frame.
- The snapshots for the same are given below

Fig 5 : Sample image for Face Recognition

Fig 6 : Sample image with faces detected

- We can see from the above images that we were able to detect almost all the faces in a frame, having different angles. This is was not achieved while trying other Models.
- Second Model was to detect the 3D key points – Facial Landmarks for each face.
- There are models which does a lot of things with 2D, but as discussion with our mentor, we understood that 3D points give us more accuracy and good results.
- We used Face Alignment, 3D-Based, to get our 68 3D key points for a detected face.
- The snapshots are given below

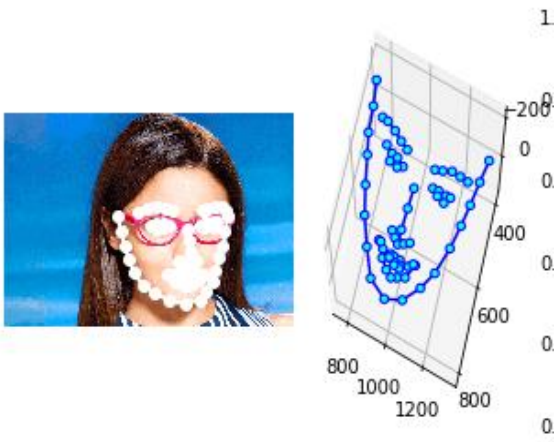

Fig 7 : 68 Key point face mapping

- The third model was to get the 3D poses that is, yaw, pitch, roll for the faces in the frame identified.
- We decided to go with the Hopenet - which is one of the good networks to identify the head poses.
- The snapshots are given below

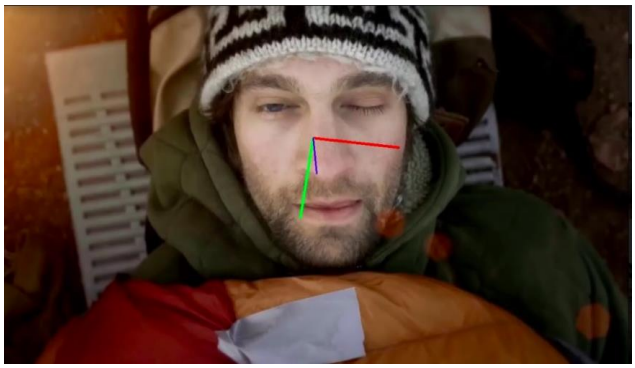

Fig 8 : User head pose detection

- Our next important goal was to combine the code or models in a way that we can get all our desired results in a frame, also a Json file in which we are writing all the 3D points and the poses (yaw, pitch, roll) for each face detected.
- So, we took hopenet as the base and combined rest of the code in the same to get the desired outcome.
- We generated huge datasets of the images from the videos first.
- Converted video in the frames like below, and then used the frames for further plots, just a snapshot of the images from a video.

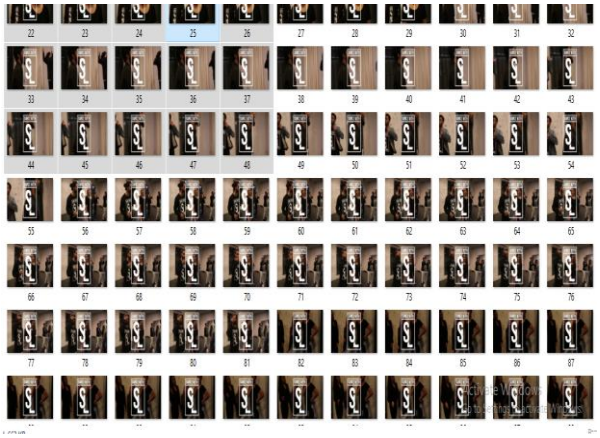

Fig 9 : Detection in videos to images

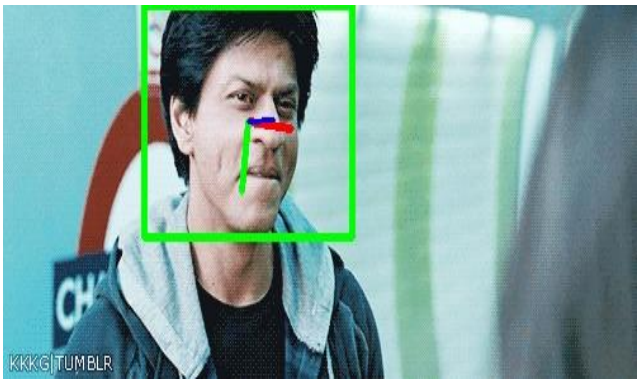

Fig 10 : Head pose detection in a frame

- We have also implemented it for the videos
- As seen in above image the video folder will have the video and respective text file which will have yaw, Roll, Pitch of every face detected in that particular frame.

iv. What are the additional insights you can provide to the organization based on the output from different models?

- The MTCNN model is the best, as it gives us the most accuracy.
- Also, it gives us the bounding boxes and five key points 2D
- We can further work on the architecture and modify the MTCNN network in a way that it can give us 3D points.

How will the project sponsor integrate the findings from the project back into his/her organization .

- The sponsors can now use the combined data code to work on their datasets.
- Also, they can use the dataset prepared by us to train their models, to get 3D points.

- This will allow them to modify the architecture if needed, very efficiently, as we are already getting the combined output.

V. CONCLUSION

The outcomes and results would help the business to understand and choose over the models wisely. Also, the combined outcome will help in easy data preparation and can be used to train the modified networks accordingly. Apart from this, for the expected outcomes like we did, for detection, pose, facial landmarks, it would be simpler for them to use or implement it directly as we have discussed and chosen the appropriate ones. Initially it will incur some cost Also, but for the long term the organization will create a different brand value in the market, increasing the efficiency of human resources and manpower. Facial analytics has lot of use cases in today's world. It depends on how and where the organization uses it. Also, once implemented for a good problem it will generate good profit.

We had some use cases where the organization can implement these facial analytics like airports for easy check ins, Retail outlets, Transport.

There are so many authentication techniques like password, OTP, Voice recognition, finger recognition, palm recognition etc. but still it has some drawbacks like at times password techniques are not feasible, password can be easily stolen by hacker or if user uses complex password, user may forget that password etc. So it is a better option to use face recognition system rather than traditional or other biometric authentication techniques.

REFERENCES

[1] W. Zhao, R. Chellapra, P.J. Phillips, A. Rosenfeld, "Face Recognition: A Literature Survey," ACM Computing Surveys, Vol. 35, No. 4, December 2003, pp. 399-458.

[2] M.A. Turk, A.P. Pentland. "Face Recognition Using Eigenfaces," IEEE Conference on Computer Vision and Pattern Recognition, pp.586--591, 1991.

[3] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. Fisherfaces Recognition using class specific linear projection," IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 711–720, May 1997.

[4] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, "Face Recognition by Independent Component Analysis", IEEE Trans. on Neural Networks, Vol. 13, No. 6, November 2002, pp. 1450-1464

[5] H. Ando, N. Fuchigami, M. Sasaki, A. Iwata, "A Prototype Software System for Multi-object Recognition and its FPGA Implementation," Proc. Third Hiroshima International Workshop on Nano-electronics for Terra-Bit Information Processing, 2004.

[6] Gottumukkal R., and Asari K.V., "System Level Design of Real Time Face Recognition Architecture Based on Composite PCA," Proc. GLSVLSI 2003, 2003, pp. 157-160.

[7] Hau T. Ngo, Rajkiran Gottumukkal, Vijayan K. Asari. "A Flexible and Efficient Hardware Architecture for Real-Time Face Recognition Based on Eigenface", isvlsi, pp. 280-281, Proc. IEEE Computer Society Annual Symposium on VLSI: New Frontiers in VLSI Design (ISVLSI'05), 2005.

[8] X. Li and S. Areibi, "A Hardware/Software Co-design Approach for Face Recognition," Proc. 16th International Conference on Microelectronics, Tunis, Tunisia, Dec 2004.

[9] Moritoshi Yasunaga, Taro Nakamura, and Ikuo Yoshihara, "A Fault-tolerant Evolvable Face Identification Chip," Proc. Int. Conf. on Neural Information Processing, pp.125-130, Perth, November 1999.

[10] In Ja Jeon, Boung Mo Choi, Phill Kyu Rhee. "Evolutionary Reconfigurable Architecture for Robust Face Recognition," ipdps, p. 192a, International Parallel and Distributed Processing Symposium (IPDPS'03)