International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 01 | Jan 2019 www.irjet.net p-ISSN: 2395-0072

Video based Face Recognition using Image Processing Techniques

Savitha G[1], Keerthi G S[2]

[1]*Associate Professor, Department of Computer Science and Engineering, BNMIT*
[2]*M.Tech Student, Department of Computer Science and Engineering, BNMIT, Bangalore, India*

Abstract - *Video-based facial recognition has received significant attention in recent years due to its widespread applications. The main challenges faced during face recognition has large variability of images due to pose variations, illumination conditions, facial expressions and image occlusion. Similarly surveillance and mobile cameras are of low cost devices which affects the video frames quality which results in severe motion blur, out-of-focus blur and large range pose variation. To achieve Face recognition from video image processing and machine learning techniques are used. The steps for processing involves Image acquisition, Image segmentation, feature extraction, classification and face detection. The features extracted are used in training the classifiers for images that are processed. Hence most recent algorithms developed give an idea of state of art of face recognition technology in video.*

Key Words: **Face Recognition, Pose variation, Occlusion, Still image, Convolution Neural Network (CNN).**

1. INTRODUCTION

Video based Face Recognition is a technique that is used to recognize the one or more human faces of the person present in the video, based upon their facial characteristics. Face recognition is most commonly used for two purposes that is verification and identification. Face recognition is ability of the computer to scan, store and recognize the face for identifying people. In video-based face recognition the faces are rich with appearance variations due to illumination, pose, lighting and expression. The human faces in video can be detected with various image processing and machine learning techniques. Face recognition in video has more applications when applied in setting security. Human beings possess a natural ability to recognize hundreds of faces by their visual system. It makes them recognize familiar faces even after a long period of time. Building an intelligent system which are similar to human perception system is still an active area of research. Human face is very challenging to recognize and detect due to its rigid anatomy that is all faces may have same structures, at same time environmental and personal factors affecting the facial appearance. The other challenges faces in VFR are face, expression and scene detection. The main difficult task is predicting whether or not the pair of facial images share same identity. There are many applications where face recognition from video techniques are been successfully used to perform specific tasks. Face recognition from video currently has been implemented for real-time applications, but yet it suffers severely from pose variation, occlusion, aging, lighting, motion blurriness and illumination. Recognizing the human faces is mainly used in companies for accessing and security purpose, which helps in detecting the theft in surveillance cameras and also in other fields. The face recognition in surveillance video is difficult due to its low image quality that caused due to pose variation, varying lighting conditions, noise and blurriness. Support Vector Machine (SVM) used to analyze the data, especially in the field of image processing, it works as a classifier model or as a regression model to give a meaningful information by analyzing the input data. Some of the datasets used in video-based face recognition are COX Face database, YouTube Face database, Honda/UCSD database, Labelled Faces in the Wild (LFW) and other databases. While detecting the human faces the main focus should be on facial expressions, pose, illumination conditions, hairstyle and use of cosmetics. Various techniques on video based face recognition are SVM, Convolutional Neural Network (CNN), Nearest Neighbor (NN) and other classifiers are used for detecting human faces. The methods for recognizing human face are image acquisition, preprocessing, segmentation, feature extraction and classification is used for experiment to achieve higher accuracy and performance.

2. RELATED WORK

L. Best-Rowden et al., [2] presented about video data captured by surveillance cameras and internet uploads that gathers facial information from video and finds the challenges faced such as illumination, expression and variations on pose. A Multi-Frame Fusion is used to track or detect human face. COTS (Commercial-off-the-shelf) technique is used to overcome these challenge and dataset used is YouTube Faces Database. Accuracy of the face track is based on the quality of the image frames and final matching results. Walter J. Scheirer et al., [3] use database such as PaSC (Point-and-Shoot Face Recognition Challenge) and VDMFP (Video Database of Moving Faces and People). The image processing technique used is Deep Convolutional Neural Network. For extraction of features Local Region Principal Components Analysis (LRPCA) for local facial patch matching. It provides high accuracy rate. Farshid Hajati et al., [4] faced challenge to detect the face expression and scene recognition. The datasets used are Honda/UCSD, CMU motion of body, YouTube and DynTex datasets. A non-binary texture coding technique are used to extract high order derivatives. For feature extraction Local Binary Patterns (LBP) are used. S. Mau et al., [6] discuss the challenge faced

in balancing computational efficiency with accuracy in recognition of human faces in video based face recognition. The dataset used is MOBIO (Mobile Biometrics) database. For still-to-still face recognition Probabilistic Multi-Region Histogram (MRH) method are been used. The steps used to detect face recognition are face detection, face selection, feature extraction, feature clustering and finally face matching. Feature extraction is done by using k-means algorithm and for matching face OpenCV is used. Z. Cui et al., [7] describes face recognition as matching of two imagesets that contains the facial images captured from each video of dataset. It uses the three datasets namely Honda, MoBo and YouTube Face databases. The main challenge is to achieve identifying a face from a single image that are under the controlled environment. The techniques used are LDA (Linear Discriminant Analysis) and also MSM (Mutual Space Method). The technique achieves the better consistent results on the three datasets. The techniques achieves better performance on the small scale videos but it need to work on large scale datasets. Xiaoou Tang et al., [8] utilises XM2VTS Face video database for detection of human faces. Multi-classifier technique is used to achieve accuracy rate of 98.6%. Along with capturing of human images in videos, it also makes use of audio sequences to match the frames. For extraction of feature the regular subspace analysis has been used. The final classification uses PCA (Probabilistic Correlation Analysis) to classify each frame using feature vector. Changbo Hu et al., [9] discuss challenges like variation in pose and occlusion. The main aim is to recognize the face from video based face patches. The dataset used is Yaleface database. The first step is to crop face patches from video frame by frame. To extract the face portion from video frames an alignment algorithm is used. The alignment algorithm align it and it also normalize the face image, but here the patches are not identified. A full face proper precision image stitching algorithm is developed and it reduces overlapping errors between the patches. The final task is to classify reconstructed face to candidate face in training dataset. Sparse representation technique is used for recognizing the reconstructed data. It achieves the accuracy rate of about 81% for video sequence database. The patch based method is flexible and the computational cost is also vulnerable, but the large changes in pose variation, illumination and expression cannot be detected and also noise need to be removed to provide the high quality image. XIE Liping et al., [11] and Cong Wang [12] uses Support Vector Machine (SVM) technique as a classifier. The SVM classifier increases the quality of image. It is widely used to analyze the data, especially in the field of image processing, it works as a classifier model or as a regression model to give a meaningful information by analyzing the input data. The challenge [11] is to extract dynamic features for video-based facial expression analysis. The database used is Cohn-kanade database. The Local Binary Patterns from three orthogonal planes (LBP-TOP) is used to convolve with multi-scale and multi orientation Gabor filters. Le An et al., [13] its challenge is to recognize faces in surveillance video as it is difficult due to its poor quality that is in terms of resolution, noise, blurring and also varying lighting conditions. The database used is ChokePoint database. Instead of using single cameras, multiple cameras can be used to get high image quality. The Nearest-neighbor (NN) classifier is used and for extracting the features LPB technique are used. Here it only focus on matching the images but further it need to focus on pose variation, occlusion and also on quality of image. Changxing Ding et al., [15] exhibit face video that suffers from severe image blur, pose variation and occlusion. It uses Trunk-Branch Ensemble Convolutional Neural Network (TBE-CNN) for detecting the human faces. It deals with training dataset, classification and feature extraction. Marko Arsenovic et al., [14] while training dataset CNN has two training data streams that is one with still face images and other is simulated video frames that creates artificial image blur. Then it classify the still image and its artificially blurred version into the same class. Finally includes efficient extraction of face image and facial components from layers of different CNN.

2.1 Limitations

Canonical Correlation Analysis (CCA) method [1] processing leads to loss of small number video frames. Probabilistic Linear Discriminant Analysis (PLDA) technique [5] lacks performance on large datasets. The Unified Face Image (UFI) technique [13] only focuses on matching images but it need to focus on pose, occlusion and also on quality of image. SVM classifier [11, 12] takes long training time for large datasets and it is difficult to understand and interpret the final models and variable weights.

2.2 Convolution Neural Network

CNN classifier is considered as one of the state-of-the-art machine learning approaches. CNN is a special kind of multi-layer perception, which has many specialized hidden layers used for classification. CNN improves the performance and it is cost effective. CNN classifier gives high accuracy rate when compared with other image processing techniques. N. Pattabhi Ramaiah et al., [18] discusses on varying illuminance and pose. CNN classifier is capable of learning local features from input data that is used to discriminate facial images. CNN as alternating sequence of convolution and sub-sampling layers for feature extraction and neural network is the last layer for classification. Xiaoguang Chen et al., [19] and Marko Arsenovic et al., [14] discusses on CNN network structure that include convolution layers, pooling layers and fully connection layers. Convolution layer is feature extraction layer. The convolution operation is performed on training convolution kernel and layer of feature maps. Pooling layer is used to reduce dimension on face image. The full connection layer is considered as convolution layer and it takes input of one-dimensional layer. Wen Yang et al., [20] provides CNN as regression model on spectrum images, which is a combined process time domain feature with its respective frequency feature.

CNN improves preprocessing method and it is more robust for time domain feature representation. By comparing the techniques used for detecting human faces, CNN gives high accuracy rate.

3. METHODOLOGY

In image processing there are four major steps that involves in detecting the human faces in video-based face recognition. It plays an important role in detection of human faces from videos. The steps involved are image acquisition, image pre-processing, segmentation and feature extraction.

3.1 Image Acquisition

The first step is image acquisition that is retrieving the facial images from video for identifying the human faces. Changbo Hu et al., [9] captures cropped face patches from videos frame by frame. It uses cropped face patches for detecting human faces. S. Mau et al., [6] uses the MOBIO dataset for facial detection. The dataset provides a high quality image so that human faces can be easily detected and high accuracy is achieved.

3.2 Image Pre-Processing

The image pre-processing is used to convert the original images into the grayscale images or into a pixels. During the pre-processing step the noise in the images are removed, so that a good quality image is obtained. Vignesh S et al., [17] discusses on processing the facial image obtained from the classifier and then cropping the facial region by removing the border. Finally cropped images are resized to 64 X 64 pixels. Z. Huang et al., [1] also uses processing for reduction of memory storing in the final database. The video recorded here as high resolution that is 1920 X 1080 which is of large memory requirement, hence there is a need to reduce the memory of database through cropping the facial images. The cropped images are saved into jpeg files, and the sizes of cropped face images vary from smallest size of 66 X 66 pixels to largest of 798 X 798 pixels.

3.3 Image Segmentation

Image segmentation is one of the process of partitioning the digital data into multiple segments. The image is segmented to extract the main region of interest that is the human face need to be extracted from whole image along with the facial characters. S. Mau et al., [6] utilises a face classifier for detecting region of interest face. A generic method is applied during processing step of a face or non-face binary classifier for all faces detected during processing. Then compare a face which are based on where landmarks are detected within the face (such as eyes and nose). Peng Jia et al., [10] discusses about image average technique for video based face recognition. Image Averaging Technique provides decision-making over testing frames. After morphing the image segmentation and tracking of images are done, it increase the efficiency of representing a face image.

3.4 Feature Extraction

The feature from images are extracted and these features helps in detecting the facial characters in human face that is captured from video is obtained. Vignesh S et al., [17] and Le An et al., [13] considers the feature extracted from each face image that forms the feature vector. Feature extraction is done using LBP (Local Binary Patterns) technique. LBP is used for both shape and texture information of the face. Cong Wang [12] defines the extraction of features which includes brightness of the facial image defined by means of the grayscale, illuminance of facial image is done through facial landmarks and facial landmarks calculates sharpness through algorithm FISH (Fast Image SHarpness).

4. CONCLUSION

In recent years detection of human faces from video plays an important role in many fields. In education and IT fields attendance are marked automatically by detecting the faces from the surveillance cameras, similarly one can predict whether the person is happy or sad based on the facial expression of the person by using video based face recognition. Therefore it is much needed to detect human face from mobile phones and surveillance cameras. Image processing are used to extract images from video. The techniques utilized are SVM, NN and also other image processing techniques. The performance in detecting human faces with high image quality is achieved using the CNN techniques compared with SVM and NN techniques. CNN technique gives high accuracy rate when compared with other state-of algorithms in detection of human faces.

REFERENCES

[1] Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, and X. Chen, "A benchmark and comparative study of video-based face recognition on cox face database," IEEE Transaction Image Processing, 2015.

[2] L. Best-Rowden, B. Klare, J. Klontz, and A. K. Jain, "Video-to-video face matching: Establishing a baseline for unconstrained face recognition," in Proceeding IEEE International Conference Biometrics, Theory, Applications and Systems, 2013.

[3] Walter J. Scheirer, Patrick J. Flynn, Changxing Ding, et al., "Report on the BTAS 2016 Video Person Recognition Evaluation", IEEE International Conference on Biometrics: Theory, Applications and Systems, 2016.

[4] Farshid Hajati, Mohammad Tavakolian, Soheila Gheisari, Yongsheng Gao and Ajmal S. Mian, "Dynamic Texture Comparison Using Derivative Sparse Representation:

Application to Video-Based Face Recognition", IEEE Transaction on Human-Machine Systems, 2017.

[5] Dong Chen, Xudong Cao, David Wipf, Fang Wen, and Jian Sun, "An Efficient Joint Formulation for Bayesian Face Verification", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.

[6] S. Mau, S. Chen, C. Sanderson, and B. C. Lovell, "Video face matching using subset selection and clustering of probabilistic multi-region histograms" arXiv preprint arXiv:1303.6361, 2013.

[7] Z. Cui, S. Shan, H. Zhang, S. Lao, and X. Chen, "Image sets alignment for video-based face recognition," in Proceeding IEEE Conference Computer Visual Pattern Recognition, 2012.

[8] Xiaoou Tang and Zhifeng Li, "Video Based Face Recognition Using Multiple Classifiers", IEEE International Conference on Automatic Face and Gesture Recognition, 2014.

[9] Changbo Hu, Josh Harguess and J. K. Aggarwal, "PATCH-BASED FACE RECOGNITION FROM VIDEO", 2009.

[10] Peng Jia and Dewen Hu, "Video-Based Face Recognition Using Image Averaging Technique", International Conference on Image and Signal Processing, 2012.

[11] XIE Liping, WEI Haikun, YANG Wankou and ZHANG Kanjian, "Video-based Facial Expression Recognition Using Histogram Sequence of Local Gabor Binary Patterns from Three Orthogonal Planes", IEEE Proceedings, 2014.

[12] Cong Wang, "A Learning-based Human Facial Image Quality Evaluation Method in Video-based Face Recognition Systems", IEEE International Conference on Computer and Communications, 2017.

[13] Le An, Bir Bhanu and Songfan Yang, "Face Recognition in Multi-Camera Surveillance Videos", International Conference on Pattern Recognition, 2012.

[14] Marko Arsenovic, Srdjan Sladojevic, Andras Anderla and Darko Stefanovic, "FaceTime – Deep Learning Based Face Recognition Attendance System", IEEE International Symposium on Intelligent Systems and Informatics, 2017.

[15] Changxing Ding and Dacheng Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-based Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[16] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proceeding IEEE Conference Computer Visual Pattern Recognition, 2015.

[17] Vignesh S, Manasa Priya K. V. S. N. L., Sumohana S. Channappayya, "Face Image Quality Assessment for Face Selection in Surveillance Video using Convolutional Neural Networks", IEEE Global Conference on Signal and Information Processing, 2015.

[18] N. Pattabhi Ramaiah, Earnest Paul Ijjina and C. Krishna Mohan, "Illumination Invariant Face Recognition Using Convolutional Neural Networks", IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES) , 2015.

[19] Xiaoguang Chen, Xuan Yang, Maosen Wang and Jiancheng Zou, "Convolution Neural Network for Automatic Facial Expression Recognition", IEEE International Conference on Applied System Innovation, 2017.

[20] Wen Yang, Xiaoqi Li and Bin Zhang, "Heart Rate Estimation from Facial Videos Based on Convolutional Neural Network", IEEE Proceedings, 2018.