IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Face Image Quality Assessment in Video Frame

T. Bhuvaneshwari[1], E. Kalpana[2], N. Ramadevi[3]

[1,2,3] Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad-500075, India.

Abstract: In order to improve the accuracy of recognition to reduce computing costs, quality-based frame selection is employed in video face recognition. In this article, the facial images for face recognition are chosen from a videos using a variety of criteria, such as face symmetry, sharpness, contrast, the proximity of the lips to the teeth, brightness, and eye openness. The Viola-Jones and KLT algorithms are used in our paper to continuously detect images in a video . We also give an overview of Deep Learning (DL) models that are applied to video frames, using a variety of techniques to assess image simplicity and selecting a limited selection of the best video frames. We recommend utilizing a lightweight convolutional neural network to determine the frame quality using deep learning methods.

Index Terms - Face quality, video frame, Face detection, Face recognition, Machine Learning, Deep Learning.

I. Introduction

Face detection is one of the research areas that is now most active. It is a fundamental stage for applications of enhanced processing. Even with face recognition, it appears that a large portion of the work is concentrated on real-time video processing to aid in surveillance or identification, among other things. Recent algorithmic advancements in this field operate with real-time movies of poor quality amazingly well and virtually quickly[1]. However, these algorithms grow and lose their effectiveness when dealing with high resolution, high quality footage used in movies or other sectors. Their processing time increases and their computing performance decreases [3]. From images in a database, a face recognition system automatically recognizes a human face. We were motivated by the work presented in that summarizes faces in surveillance video sequences and uses elements including head attitude, tilt, brightness, clarity, resolution, eye openness, eye direction, and mouth closeness to select high-quality frames. Frames with higher scores are used to represent the video sequence after normalized feature scores are integrated using a neural network. However, this approach cannot be tested in a face recognition trial and can only be used to provide a human operator with a summary of a sequence. [4]

Since it has to take into consideration every possible alteration in a face due to changes in facial characteristics, lighting, occlusions, etc., face recognition is a difficult subject. Systems for recognizing faces must process data very reliably and effectively. A person's visage is used at the recognition step to recognize them and verify their identification. Face recognition takes time due to the multiple comparisons that must be done. Most traditional algorithms can be broadly categorized into knowledge-based, feature-invariant, template-matching, or aesthetically pleasant groups. Although their average performance is still inferior to the second generation of algorithms due to the fact that they use feature-tracking from the KLT algorithm to delete non-facial portions in subsequent frames when an item is spotted in a prior frame, which consistently outperform others. These methods include Multi-view Face Detection, Linear Discriminant Analysis (LDA), Principle Component Analysis (PCA), and Viola-Jones. The most modern algorithms, such as Histogram of Oriented Gradients (HoG) and You Only Look Once (YOLO), in real-time video streams are more complex than conventional algorithms but provide results incredibly swiftly and precisely. [4]

II. Related work

Recently, other researchers have developed quality assessment techniques based on deep learning. The initial stage involves combining Face Quality Assessment (FQA) and Convolution Neural Networks [CNN] to extract keyframes from videos [5]. Given that digital forensic techniques for spoof audio and spoof visual deep fake detection include bidirectional recurrent structures, convolutional latent representations, and entropy-based cost functions [6]. Has suggests using Face in Video Recognition (FIVR) involves detecting faces using the MTCNN detector, which is capable of recognizing five landmarks (the eyes, the two corners of the mouth and the nose) [7]. Given KonIQ-10k is the largest ecologically valid IQA database to date, and we offered a fresh, methodical technique that might be scaled up[8]. A face recognition algorithm examines supervised machine learning to forecast the importance of images of faces across many state-of-the-art face identification systems[9].

A basic FIQA model is created. The gap between human perception and model performance is then filled using recently proposed strategies such as quality distribution alignment and contradicting samples selection.in future, we can improve accuracy [10]. A basic FIQA model is created. The gap between human perception and model performance is then filled using recently proposed strategies such as quality distribution alignment and contradicting sample selection.in future, we can improve accuracy[11].The method employs a liveness detection methodology to identify fraudulent biometric users. A threshold for the valid user is chosen depending on the IQA parameter and the nature of the image [12]. Image Quality Assessment (IQA) based features from two datasets, CASIA web face and VGGFace2, the author has suggested a novel method to identify fake faces [13]. The lightweight convolutional neural network (CNN) architecture is designed to recognise people's faces in hazardous conditions like mines, avalanches, underwater, or other places where their faces wouldn't be readily seen against the background [14].

III. Proposed System

The model utilizes facial detection successfully for high resolution films while maintaining the algorithm's integrity and keeping up with live, lower quality films. The suggested architecture consists of three sub-frameworks, each of which identifies and tracks faces in a stream of images while working in tandem to complete a specific task. (videos). To distinguish between various face sizes, the image is first scaled up and down many times. To complete the first detection, the P-network (Proposal) then looks over the images. Low detection thresholds result in numerous false positives even after NMS (Non-Maximum Suppression), even if this is done on purpose.[3]

1.To separate the frame work we will use Convolutional neural networks [CNN] algorithm.

2. Kanade-Lucas-Tomasi [KLT] Algorithm is used to extract the features.

3.Multitask Cascaded Convolutional Networks [MTCNN] is used to detect the face.

Paul Viola and Michael Jones developed the first algorithm of its attempted specifically for facial recognition. It is known as the Viola-Jones Algorithm. The application of this method occurs in two stages: first, it is trained using both facial and non-facial images, and then the actual detection is carried out using the Kanade Lucas Tomasi (KLT) algorithm. The initial work was finished by Kanade and Lucas, and Kanade and Tomasi later expanded it to track feature points. It continuously monitors qualities and employs a mechanism analogous to the Viola-Jones technique. The method monitors the amount of motion the feature points undergo between frames.[1] The proposed approach's flowchart is shown in Fig. 1.

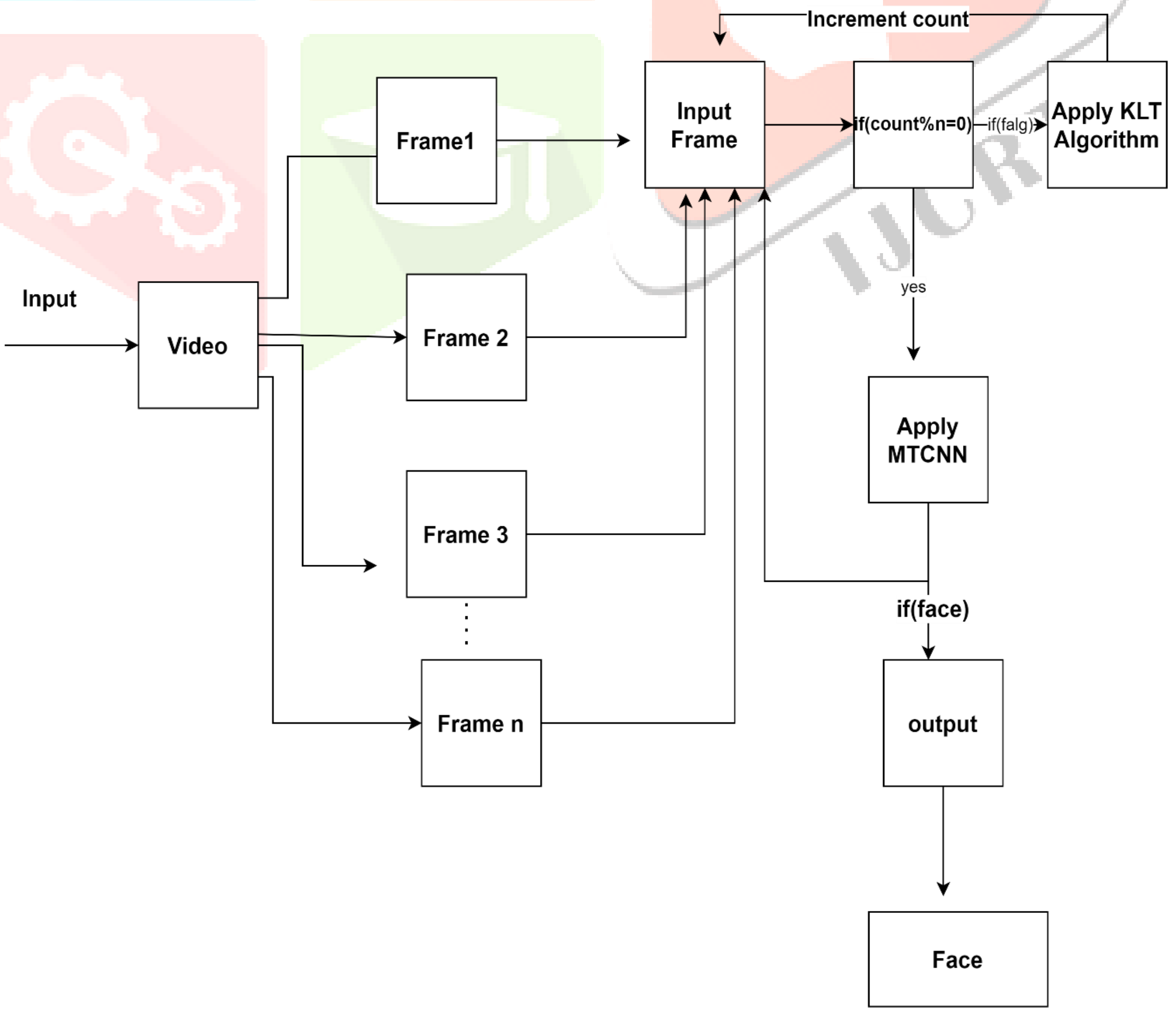

Fig 1: Flow chart of proposed Method

Any received stream of images must move more quickly than 24 frames per second in order for human eyes to recognise it as video. N should therefore perform at its best at a range larger than 8.300. The Videos in the Wild (300-VW) dataset is used to evaluate the performance of the proposed model, along with a number of other well-known and state-of-the-art techniques. The collection includes 114 films with a variety of frontal head poses, backdrops, and lighting. A frame rate of 25 to 30 was used to record the videos. (fps). The speed and accuracy tests use video samples with resolutions of around 1280 by 720 lines and 480 by 360 lines, respectively.[3] Figures 2 and 3 illustrate the identification of the eyes and nose in the facial frame, respectively. and fig 4 represents Facial Feature points Extracted.

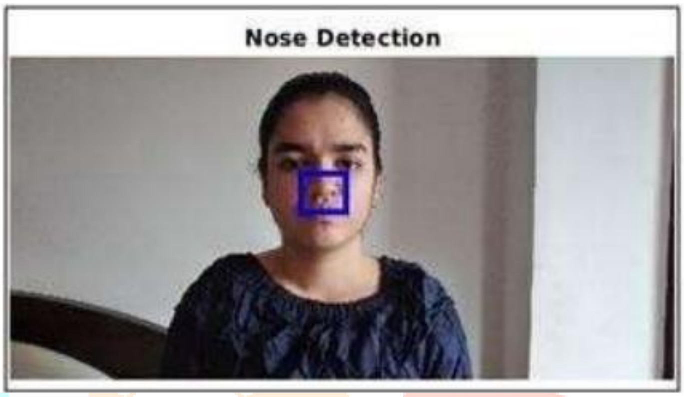

Fig 2: Nose detection in the facial frame detected

Fig 3: Eyes detection in facial frame detected

Fig 4: Facial Feature points Extracted

IV. Results And Discussion

Input:

The input we are giving is the video ,In a video they are number of images to detect the images in a video the algorithms used are MTCNN,KLT,CNN to detect the facial features in a video .fig 5 represents the detecting images in a video.

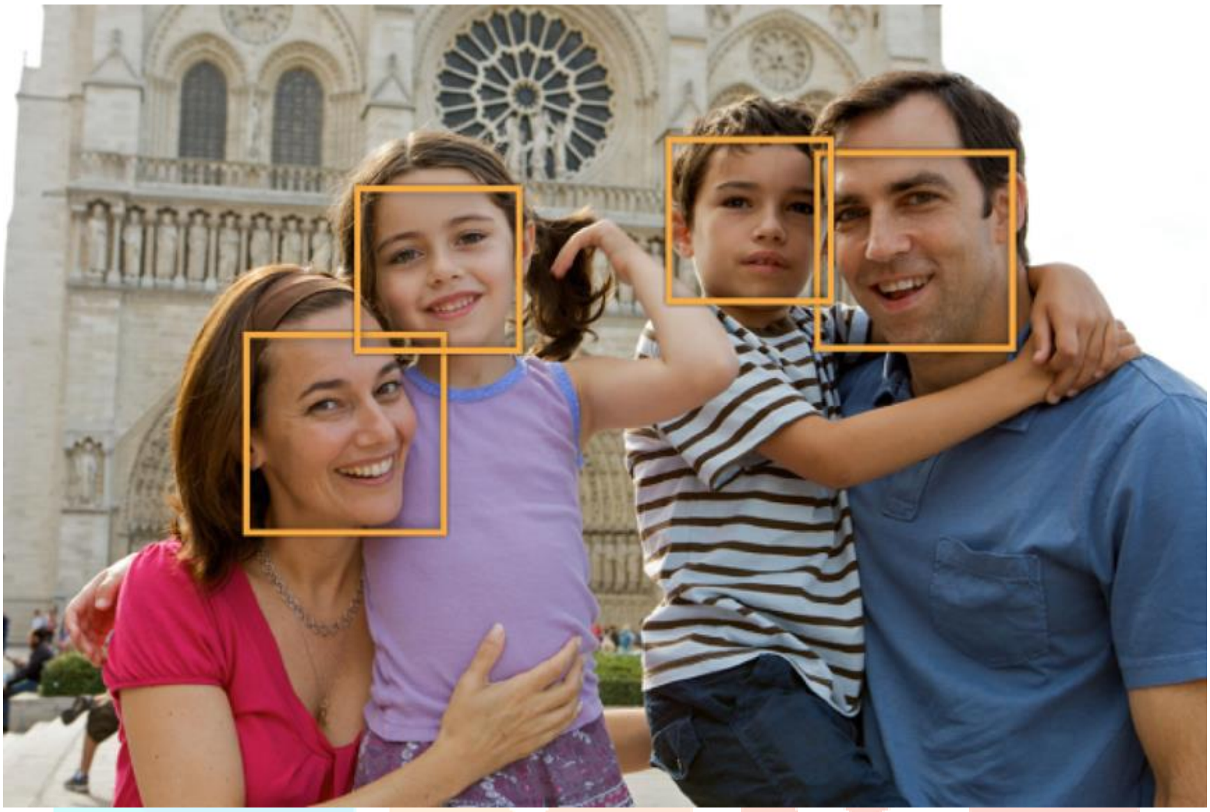

Fig 5: Detecting the images in a video

Output:

The initial sub-framework's input initiates the overall framework's activity. The MTCNN basic face identifying method is applied to the first frame of the incoming video, which is used as its input. The system separates faces in the frame based on five feature points. A rectangle box containing the recognised face and the five feature points identified on the frame are among the algorithm's outputs. The second sub-framework uses the first sub-framework's result as an input. The first sub-framework recognised and marked five feature points in order to speed up the entire framework without compromising the accuracy of the detected framework. Following the transmission of these points to the KLT feature point tracking algorithm, those points were monitored for an additional n frames, where n is the refresh rate. The detecting features of images are shown in fig 6[1].

Fig 6: Detecting the features of image.

V.Conclusion

The quality of a face image can be influenced by a number of factors. These explanations could be brought on by the utilisation of many image sensors, varied compression techniques, inadequate video or image capture settings, timing difficulties, etc. For all of these reasons, determining the quality of a facial photograph automatically is an extremely difficult problem. Numerous learning-based FIQA approaches have recently been reported that use the facial image quality score to properly forecast how well face recognition would work. Runtime effectiveness, face detection accuracy, and occlusion resolution between frames are the three performance measures that are discussed in this study. To recognize faces, the face net strategy will be employed, which focuses five key characteristics. These 5 features were identified using the KLT approach. When speed and precision are balanced, our model works at its best. This model still uses the original face detection technology that was used for it, therefore performance can be improved by choosing a quicker and more accurate base face identification method. Face tracking within a video frame will be possible very soon.[3]

References

[1]. K. Anantharajah, S. Denman, S. Sridharan, C. Fookes and D. Tjondronegoro, "Quality based frame selection for video face recognition," 2012 6th International Conference on Signal Processing and Communication Systems, Gold Coast, QLD, Australia, 2012, pp. 1-5, doi: 10.1109/ICSPCS.2012.6507950.

[2]. N. Zhang, J. Luo and W. Gao, "Research on Face Detection Technology Based on MTCNN," 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2020, pp. 154-158, doi: 10.1109/ICCNEA50255.2020.00040.

[3]. Mool, Akshay, J. Panda, and Kapil Sharma. "Optimizable face detection and tracking model with occlusion resolution for high quality videos." Multimedia Tools and Applications 81, no. 8 (2022): 10391-10406.

[4]. https://machinelearningmastery.com/introduction-to-deep-learning-for-face-recognition

[5]. Dean, Mark, Shih-Wei Wu, and Laurence T. Maloney. "Trading off speed and accuracy in rapid, goal-directed movements." *Journal of Vision* 7, no. 5 (2007): 10-10.

[6]. Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57, no. 2 (2004): 137-154.

[7]. Ismail, Nurulhuda, and Mas Idayu Md Sabri. "Review of existing algorithms for face detection and recognition." In 8th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics, pp. 30-39. 2009.

[8]. Bin Yang, J. Yan, Z. Lei and S. Z. Li, "Aggregate channel features for multi-view face detection," IEEE International Joint Conference on Biometrics, Clearwater, FL, USA, 2014, pp. 1-8, Doi: 10.1109/BTAS.2014.6996284.

[9]. Niu, Gang, and Ququ Chen. "Learning an video frame-based face detection system for security fields." Journal of Visual Communication and Image Representation 55 (2018): 457-463.

[10]. R. Ranjan *et al.*, "A Fast and Accurate System for Face Detection, Identification, and Verification," in IEEE Transactions on Biometrics, Behaviour, and Identity Science, vol. 1, no. 2, pp. 82-96, April 2019, Doi: 10.1109/TBIOM.2019.2908436.

[11]. W. Wu, Y. Yin, X. Wang and D. Xu, "Face Detection with Different Scales Based on Faster R-CNN," in IEEE Transactions on Cybernetics, vol. 49, no. 11, pp. 4017-4028, Nov. 2019, Doi: 10.1109/TCYB.2018.2859482.

[12]. R. Ranjan, V. M. Patel and R. Chellappa, "Hyper Face: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 1, pp. 121-135, 1 Jan. 2019, Doi: 10.1109/TPAMI.2017.2781233.

[13]. R. Abed, S. Bahroun and E. Zagrouba, "Face Retrieval in Videos using Face Quality Assessment and Convolution Neural Networks," 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2020, pp. 399-405, Doi: 10.1109/ICCP51029.2020.9266253.

[14]. A. Chintha et al., "Recurrent Convolutional Structures for Audio Spoof and Video Deepfake Detection," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 1024-1037, Aug. 2020, Doi: 10.1109/JSTSP.2020.2999185.