International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 672-676 www.ijaem.net ISSN: 2395-5252

Face Based Authentication from Video Streaming Data in Real Time Environments

S. Lavanya, M. Anupriya, N.Maheswari

Submitted: 25-05-2021 Revised: 01-06-2021 Accepted: 05-06-2021

ABSTRACT: For face recognition in surveillance scenarios, identifying a person captured on video is one of the key tasks. Human beings are recognized by their unique facial characteristics. Aim to create a attendance system for educational institutions to enhance and upgrade the current attendance system into more efficient. In the face recognition approach, a given face is compared with the faces stored in the database in order to identify the person. If the person may be unknown, the image of the unknown person from video clip will be send to the admin mail.

I. 1.INTRODUCTION:

Pictures are the most common and convenient means of conveying or transmitting information. A picture is worth a thousand words. Pictures concisely convey information about positions, sizes and inter-relationships between objects. They portray spatial information that we can recognize as objects. Human beings are good at deriving information from such images, because of our innate visual and mental abilities. Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them. It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too.

1.2 PURPOSE OF IMAGE PROCESSING

The purpose of image processing is divided into 5 groups. They are:

1. Visualization – Observe the objects that are not visible.
2. Image sharpening and restoration – To create a better image.
3. Image retrieval – Seek for the image of interest.
4. Measurement of pattern – Measures various objects in an image.
5. Image Recognition – Distinguish the objects in an image.

1.3 STEPS IN IMAGE PROCESSING

1.3.1. Image Acquisition

This is the first step or process of the fundamental steps of digital image processing. Image acquisition could be as simple as being given an image that is already in digital form. Generally, the image acquisition stage involves preprocessing, such as scaling etc.

1.3.2. Image Enhancement

Image enhancement is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interest in an image. Such as, changing brightness & contrast etc.

1.3.3. Image Restoration

Image restoration is an area that also deals with improving the appearance of an image. However, unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on mathematical or probabilistic models of image degradation.

1.3.4. Color Image Processing

Color image processing is an area that has been gaining its importance because of the significant increase in the use of digital images over the Internet. This may include color modeling and processing in a digital domain etc.

1.3.5. Wavelets and Multiresolution Processing

Wavelets are the foundation for representing images in various degrees of resolution. Images subdivision successively into smaller regions for data compression and for pyramidal representation.

1.3.6. Compression

Compression deals with techniques for reducing the storage required to save an image or

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 672-676 www.ijaem.net ISSN: 2395-5252

the bandwidth to transmit it. Particularly in the uses of internet it is very much necessary to compress data.

1.3.7. Morphological Processing

Morphological processing deals with tools for extracting image components that are useful in the representation and description of shape.

1.3.8. Segmentation

Segmentation procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually.

1.3.9. Representation and Description

Representation and description almost always follow the output of a segmentation stage, which usually is raw pixel data, constituting either the boundary of a region or all the points in the region itself. Choosing a representation is only part of the solution for transforming raw data into a form suitable for subsequent computer processing. Description deals with extracting attributes that result in some quantitative information of interest or are basic for differentiating one class of objects from another.

1.3.10. Object recognition

Recognition is the process that assigns a label, such as, "vehicle" to an object based on its descriptors.

1.3.11. Knowledge Base:

Knowledge may be as simple as detailing regions of an image where the information of interest is known to be located, thus limiting the search that has to be conducted in seeking that information. The knowledge base also can be quite complex, such as an interrelated list of all major possible defects in a materials inspection problem or an image database containing high-resolution satellite images of a region in connection with change-detection applications.

II. 2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

The term multi-view face recognition, in a strict sense, only refers to situations where multiple cameras acquire the subject (or scene) simultaneously and an algorithm collaboratively utilizes the acquired images/videos. But the term has frequently been used to recognize faces across pose variations. This ambiguity does not cause any problem for recognition with (still) images; a group of images simultaneously taken with multiple cameras and those taken with a single camera but at different view angles are equivalent as far as pose variations are concerned. However, in the case of video data, the two cases diverge. While a multi-camera system guarantees the acquisition of multi-view data at any moment, the chance of obtaining the equivalent data by using a single camera is unpredictable. Such differences become vital in non-cooperative recognition applications such as surveillance. For clarity, we shall call the multiple video sequences captured by synchronized cameras a multi-view video and the monocular video sequence captured when the subject changes pose, a single-view video. With the prevalence of camera networks, multi-view surveillance videos have become more and more common. Nonetheless, most existing multi-view video face recognition algorithms exploit single-view videos. Given a pair of face images to verify, they look up in the collection to "align" the face part's appearance in one image to the same pose and illumination of the other image. This method will also require the poses and illumination conditions to be estimated for both face images. This "generic reference set" idea has also been used to develop the holistic matching algorithm, where the ranking of look-up results forms the basis of matching measure.

2.2 PROPOSED SYSTEM

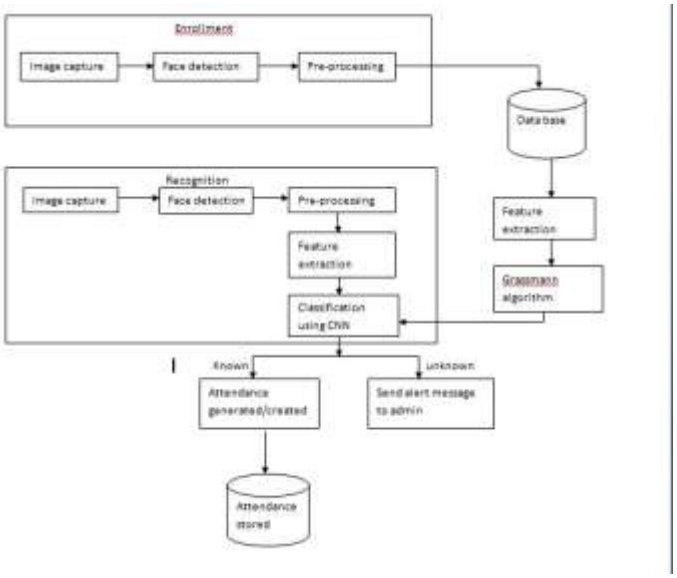
Face detection is the first stage of a face recognition system. A lot of research has been done in this area, most of which is efficient and effective for still images only & could not be applied to video sequences directly. Face recognition in videos is an active topic in the field of image processing, computer vision and biometrics over many years. Compared with still face recognition videos contain more abundant information than a single image so video contain spatio-temporal information. To improve the accuracy of face recognition in videos to get more robust and stable recognition can be achieved by fusing information of multi frames and temporal information and multi poses of faces in videos make it possible to explore shape information of face and combined into the framework of face recognition. The video-based recognition has more advantages over the image-based recognition. First, the temporal information of faces can be utilized to facilitate the recognition task. Secondly, more effective representations, such as face model or super-resolution images, can be obtained from the video sequence and used to improve recognition results. Finally, video- based recognition allows learning or updating the subject model over time to improve recognition results for future frames. So video based face recognition is also a very challenging problem, which suffers from following nuisance factors such as low quality

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 672-676 www.ijaem.net ISSN: 2395-5252

facial images, scale variations, illumination changes, pose variations, Motion blur, and occlusions and so on.

In the video scenes, human faces can have unlimited orientations and positions, so its detection is of a variety of challenges to researchers. In recent years, multi-camera networks have become increasingly common for biometric and surveillance systems. Multi view face recognition has become an active research area in recent years. In this paper, an approach for video-based face recognition in camera networks is proposed. Traditional approaches estimate the pose of the face explicitly. A robust feature for multi-view recognition that is insensitive to pose variations is proposed in this project. The proposed feature is developed using the spherical harmonic representation of the face, texture mapped onto a sphere. The texture map for the whole face is constructed by back-projecting the image intensity values from each of the views onto the surface of the spherical model. A particle filter is used to track the 3D location of the head using multi-view information. Videos provide an automatic and efficient way for feature extraction. In particular, self-occlusion of facial features, as the pose varies, raises fundamental challenges to designing robust face recognition algorithms. A promising approach to handle pose variations and its inherent challenges is the use of multi-view data. In video based face recognition, great success has been made by representing videos as linear subspaces, which typically lie in a special type of non-Euclidean space known as Grassmann manifold. To leverage the kernel-based methods developed for Euclidean space, several recent methods have been proposed to embed the Grassmann manifold into a high dimensional Hilbert space by exploiting the well-established Project Metric, which can approximate the Riemannian geometry of Grassmann manifold. Nevertheless, they inevitably introduce the drawbacks from traditional kernel-based methods such as implicit map and high computational cost to the Grassmann manifold.

III. SYSTEM ARCHITECTURE



IV. ALGORITHM USED:

GRASSMAN ALGORITHM:

For each frame in a video sequence, we first detect and crop the face regions. We then partition all the cropped face images into K different partitions. We partition the cropped faces by a Grassman algorithm type of algorithm that is inspired by video face matching algorithm. Sampling and characterizing a registration manifold is the key step in our proposed approach. The proposed algorithm presents a novel perspective towards frame selection by utilizing feature richness as the criteria. It is our assertion that quantifying the feature richness of an image helps in extracting the frames that have higher possibility of containing discriminatory features. In order to compute feature-richness, first the input (detected face) image I is preprocessed to a standard size and converted to grayscale. By performing face detection first and considering only the facial region, we ensure that other non-face content of the frame does not interfere with the proposed algorithm. Given a pair of face

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 672-676 www.ijaem.net ISSN: 2395-5252

coordinates, we determine a set of affine parameters for geometric normalization. The affine transformation maps the (x, y) coordinate from a source image to the (u,v) coordinate of a normalized image.

Input: A set of P points on manifold

$\{X_i\}_{i=1}^{P} \in G(d, D)$

Output: Karcher meanμK

1. Set an initial estimate of Karcher mean $\mu_K = X_i$ by randomly picking one point in $X_i\}_{i=1}^{P}$
2. Compute the average tangent vector

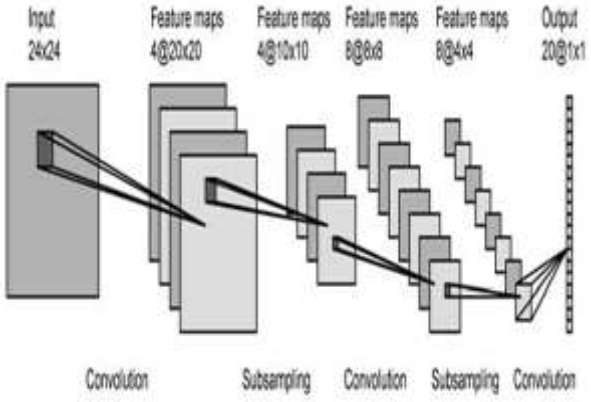
$A = \frac{1}{p}\sum_{i=1}^{P} \log_{\mu K}(X_i)$

3. If $\|A\| < \varepsilon$ then return μK stop, else go to Step 4
4. Move μK in average tangent direction $\mu K = \exp_{\mu K}(\alpha A)$, where$\alpha > 0$ is a parameter of step size. Go to Step 2, until μK meets the termination conditions (reaching the max iterations, or other convergence conditions

Thus, the video is transformed on a trajectory that links different points on Grassmann manifold. The projection on Grassmann manifold requires decomposition. The main advantages of this projection are being reversible and have no loss of information. The next step consists on similarity computing between human skeletal joint trajectories in order to identify the identity of a given skeleton sequence.

V. CONVOLUTIONAL NEURAL NETWORK ALGORITHM:

A convolutional neural network is a feed-forward network with the ability of extracting topological properties from the input image. It extracts features from the raw image and then a classifier classifies extracted features. CNNs are invariance to distortions and simple geometric transformations like translation, scaling, rotation and squeezing. Convolutional Neural Networks combine three architectural ideas to ensure some degree of shift, scale, and distortion invariance: local receptive fields, shared weights, and spatial or temporal sub-sampling. The network is usually trained like a standard neural network by back propagation. A convolutional layer is used to extract features from local receptive fields in the preceding layer. In order to extract different types of local features, a convolutional layer is organized in planes of neurons called feature maps which are responsible to detect a specific feature. In a network with a 5×5 convolution kernel each unit has 25 inputs connected to a 5×5 area in the previous layer, which is the local receptive field. A trainable weight is assigned to each connection, but all units of one feature map share the same weights. This feature which allows reducing the number of trainable parameters is called weight sharing technique and is applied in all CNN layers



With local receptive fields, elementary visual features including edges can be extracted by neurons. To extract the same visual feature, neurons at different locations can share the same connection structure with the same weights. The output of such a set of neurons is a feature map. This operation is the same as a convolution of the input image with a small size kernel. Multiple feature maps can be applied to extract multiple visual features across the image. Subsampling is used to reduce the resolution of the feature map, and hence reduce the sensitivity of the output to shifts and distortions.

A simple CNN is a sequence of layers, and every layer of a CNN transforms one volume of activations to another through a differentiable function. We have used three main types of layers to build CNN architectures: Convolution (CONV) Layer, Pooling Layer, and Fully-Connected Layer (exactly as seen in regular Neural Networks). The parameters in the CONV/FC layers have been trained with gradient descent so that the class

scores that the CNN computes are consistent with the labels in the training set for each image.

INPUT layer holds the raw pixel values of the image, in this case an image of width, height.

CONV layer computes the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume.

RELU layer applies an element wise activation function, such as the max(0, x) thresholding at zero. This leaves the size of the volume unchanged.

POOL layer performs a down-sampling operation along the spatial dimensions (width, height), resulting in volume.

FC (Fully-Connected) layer computes the class scores, resulting in volume of size [1 × 1 × 10], where each of the 10 numbers corresponds to a class score. As with ordinary Neural Networks and as the name implies, each neuron in this layer is connected to all the neurons in the previous volume.

And also provide CNN algorithm to classify faces with improved accuracy in alert system. Finally provide voice, SMS and Email based alert system with real time implementation.

VI. CONCLUSION:

In this project, we reviewed face recognition technique for still images and video sequences. Most of these existing approaches need well-aligned face images and only perform either still image face recognition or video-to video match. They are not suitable for face recognition under surveillance scenarios because of the following reasons: limitation in the number (around ten) of face images extracted from each video due to the large variation in pose and lighting change; no guarantee of the face image alignment resulted from the poor video quality, constraints in the resource for calculation influenced by the real time processing. So we can propose a local facial feature-based framework for still image and video-based face recognition under surveillance conditions. This framework is generic to be capable of vide to face matching in real-time. While the training process uses static images, the recognition task is performed over video sequences. Our results show that higher recognition rates are obtained when we use video sequences rather than statics based on Grassmann and Convolutional Neural network algorithm. Evaluation of this approach is done for still image and video based face recognition on real time image datasets with SMS alert system.

VII. FUTURE ENHANCEMENT

In future work, we can extend the framework to implement various algorithms to provide still to video face matching with improved accuracy rate. Videos provide an automatic and efficient way for feature extraction. And also implement in various applications with real time alert system.

REFERENCES

[1]. M.Ayazoglu, B. Li, C. Dicle, M. Sznaier, and O. Camps (2011). "Dynamic subspace-based coordinated multicamera tracking". IEEE International Conference on Computer Vision(ICCV), pages 2462–2469.

[2]. D. Baltieri, R. Vezzani, and R. Cucchiara (2013). "Learning articulated body models for people re-identification". In Proceedingsof the 21st ACM International Conference on Multimedia, MM '13, pages 557–560, New York, NY, USA, 2013. ACM.

[3]. D. Baltieri, R. Vezzani, and R. Cucchiara (2015). "Mapping appearance descriptors on 3d body models for people re-identification". International Journal of Computer Vision, 111(3):345–364.

[4]. I. B. Barbosa, M. Cristani, B. Caputo, A. Rognhaugen, and T. Theoharis(2017). Looking beyond appearances: "Synthetic training data for deep cnns in re-identification". arXiv preprintarXiv:1701.03153.

[5]. A. Bedagkar-Gala and S. Shah(2011). "Multiple person re-identification using part based spatio-temporal color appearance model". In Computer Vision Workshops (ICCVWorkshops), 2011 IEEE International Conference on, pages 1721–1728.

[6]. A. Bedagkar-Gala and S. K. Shah. "Part-based spatiotemporal model for multi-person re-identification". PatternRecognition Letters, 33(14):1908 – 1915, 2012. Novel Pattern Recognition-Based Methods for Re-identification in Biometric Context.

[7]. J. Berclaz, F. Fleuret, E. T¨uretken, and P. Fua(2011). "Multiple object tracking using k-shortest paths optimization". IEEETransactions on Pattern Analysis and Machine Intelligence.