ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

Person Re-Identification based on Facial Features in Real Time Video Streaming Environments

Ramyabharathi.B [1], Sudha.R [2]

P.G. Student, Department of Computer Engineering, Arasu Engineering College, Kumbakonam, India[1]

Associate Professor, Department of Computer Engineering, Arasu Engineering College, Kumbakonam, India[2]

ABSTRACT: For face recognition in surveillance scenarios, identifying a person captured on image or video is one of the key tasks. This implies matching faces on both still images and video sequences. Automatic face recognition for still images with high quality can achieve satisfactory performance, but for video-based face recognition it is hard to attain similar levels of performance. Compared to still images face recognition, there are several disadvantages of video sequences. First, images captured by CCTV cameras are generally of poor quality. The noise level is higher, and images may be blurred due to movement or the subject being out of focus. Second, image resolution is normally lower for video sequences. If the subject is very far from the camera, the actual face image resolution can be as low as 64 by 64 pixels. Last, face image variations, such as illumination, expression, pose, occlusion, and motion, are more serious in video sequences. The approach can address the unbalanced distributions between still images and videos in a robust way by generating multiple "bridges" to connect the still images and video frames. So in this project, we can implement still to video matching approach to match the images with videos using Grassmann manifold learning approach and Convolutional Neural network algorithm to know unknown matches. Using Grassmann learning algorithm to read the features vectors and matching feature vectors based on deep learning approaches. Finally provide voice alert at the time unknown matching in real time environments. And also provide SMS alert and Email alert at the time of unknown face detection.

KEYWORDS: Face recognition, Video Streaming, Convolutional neural network, Grassmann learning algorithm, Feature Vector, Deep Learning.

I.INTRODUCTION

1.1 IMAGE PROCESSING

An image is an array, or a matrix, of square pixels (picture elements) arranged in columns and rows. In a gray scale image each picture element has an assigned intensity that ranges from 0 to 255. A grey scale image is what people normally call a black and white image, but the name emphasizes that such an image will also include many shades of grey. There are two general groups of 'images': vector graphics (or line art) and bitmaps (pixel-based or 'images').

1.2 TYPES OF IMAGE PROCESSING:

The two types of Image Processing are

- Analog image processing
- Digital image processing

1.2.1 Analog image processing

Analog or visual techniques of image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. The image

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

processing is not just confined to area that has to be studied but on knowledge of analyst. Association is another important tool in image processing through visual techniques. So analysts apply a combination of personal knowledge and collateral data to image processing.

1.2.2 Digital image Processing

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of multidimensional systems.

Many of the techniques of digital image processing, or digital picture processing as it often was called, were developed in the 1960. The cost of processing was fairly high, however, with the computing equipment of that era. That changed in the 1970s, when digital image processing proliferated as cheaper computers and dedicated hardware became available. Images then could be processed in real time, for some dedicated problems such as television standards conversion. As general-purpose computers became faster, they started to take over the role of dedicated hardware for all but the most specialized and computer-intensive operations.

With the fast computers and signal processors available in the 2000s, digital image processing has become the most common form of image processing and generally, is used because it is not only the most versatile method, but also the cheapest.

II.RELATED WORK

"**DYNAMIC SUBSPACE-BASED COORDINATED MULTICAMERA TRACKING"by M. AYAZOGLU, B. LI, 2011** the paper contributes distributed surveillance systems use multiple cameras to cover wider areas and to provide different viewpoints of targets. Intuitively, the additional information provided by using multiple cameras with overlapping field of views can help a tracking system to overcome occlusion and clutter, specially when there are multiple similar targets in the scene. This paper considers the problem of sustained multicamera tracking in the presence of occlusion and changes in the target motion model. The key insight of the proposed method is the fact that, under mild conditions, the 2D trajectories of the target in the image planes of each of the cameras are constrained to evolve in the same subspace. This observation allows for identifying, at each time instant, a single (piecewise) linear model that explains all the available 2D measurements. In turn, this model can be used in the context of a modified particle filter to predict future target locations. In the case where the target is occluded to some of the cameras, the missing measurements can be estimated using the facts that they must lie both in the subspace spanned by previous measurements and satisfy epipolar constraints. Hence, by exploiting both dynamical and geometrical constraints the proposed method can robustly handle substantial occlusion, without the need for performing 3D reconstruction, calibrated cameras or constraints on sensor separation.

"LEARNING ARTICULATED BODY MODELS FOR PEOPLE RE-IDENTIFICATION by D. BALTIERI, 2013" the paper Contributes People re-identification is a challenging task extremely useful for video surveillance or forensics. It aims at discovering multiple instances of the same person captured from different points of view or after a significant temporal gap. However, differently from biometric techniques, re-identification trusts on appearance information only and thus it is mainly based on the color, texture and shape of people clothing. Features such as color and texture histograms have been deeply tested and applied rather than shape and geometrical properties, leading to an intrinsically low discriminability of the computed signatures. To solve this problem, 2D models and, recently, non-articulated 3D models of the human body have been introduced. Differently from motion capture or action analysis approaches, the models are not required to be extremely precise but fast. However, model based localization provides more coherent and representative descriptions and allows a correct comparison of corresponding body parts. Problems due to occlusions and segmentation errors can also be minimized. In this work we made a step forward by proposing a new original 3D approach to re-identification based on articulated body models. A 3D model is adopted to map appearance descriptors to skeleton bones. The color, depth and skeleton streams produced with the Microsoft Kinect sensor and the OpenNi libraries are exploited as input. The skeleton is further refined using a

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

learning approach in order to generate a "bone" set. The obtained signature is strictly related to the real body structure, thus it also allows an attribute based description, which could be useful is some applications. In addition, a learned metric is adopted which simultaneously acts as feature selection and body part weighting.

III.PROPOSED ALGORITHM

3.1 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 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. To overcome such limitations, we propose a novel method to learn the Projection Metric directly on Grassmann manifold rather than in Hilbert space. From the perspective of manifold learning, our method can be regarded as performing a geometry-aware dimensionality reduction from the original Grassmann manifold to a lower-dimensional, more discriminative Grassmann manifold where more favorable classification can be achieved. And also provide CNN algorithm to classify faces with improved accuracy in door control system. Finally provide voice, SMS and Email based alert system with real time implementation.

3.2 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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

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 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.

3.3 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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

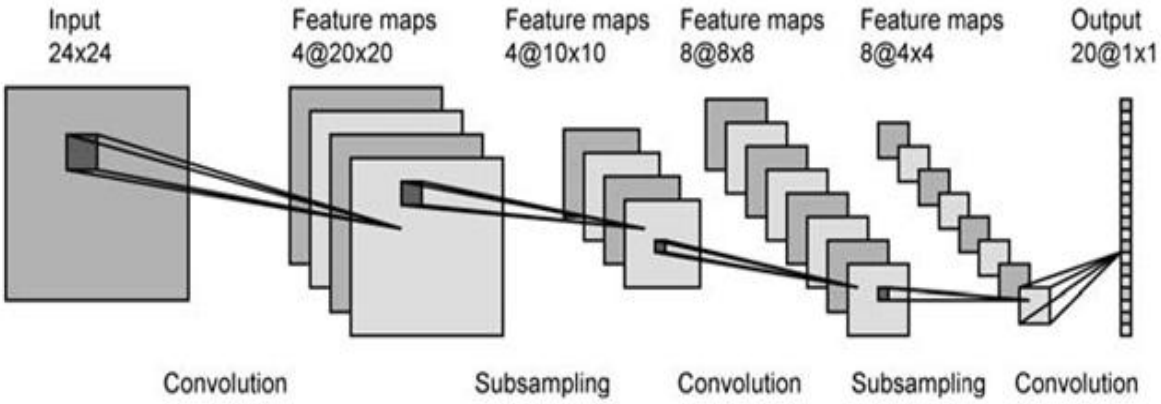

Fig 1. Convolutional Neural network approach

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.

IV .CONCLUSION AND FUTURE WORK

4.1 CONCLUSION

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.

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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 10, October 2019

REFERENCES

[1] M. Ayazoglu, B. Li, C. Dicle, M. Sznaier, and O. Camps. Dynamic subspace-based coordinated multicamera tracking. In 2011 IEEE International Conference on Computer Vision(ICCV), pages 2462–2469, Nov. 2011.
[2] D. Baltieri, R. Vezzani, and R. Cucchiara. 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. Mapping appearance descriptors on 3d body models for people reidentification. International Journal of Computer Vision, 111(3):345–364, 2015.
[4] I. B. Barbosa, M. Cristani, B. Caputo, A. Rognhaugen, and T. Theoharis. Looking beyond appearances: Synthetic training data for deep cnns in re-identification. arXiv preprintarXiv:1701.03153, 2017.
[5] A. Bedagkar-Gala and S. Shah. Multiple person reidentification using part based spatio-temporal color appearance model. In Computer Vision Workshops (ICCVWorkshops), 2011 IEEE International Conference on, pages 1721–1728, Nov 2011.
[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. Multiple object tracking using k-shortest paths optimization. IEEETransactions on Pattern Analysis and Machine Intelligence, 2011.
[8] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: the CLEAR MOT metrics. EURASIPJournal on Image and Video Processing, (246309):1–10, 2008.
[9] L. Beyer, S. Breuers, V. Kurin, and B. Leibe. Towards a principled integration of multi-camera re-identification and tracking through optimal bayes filters. CVPRWS, 2017.
[10] M. Bredereck, X. Jiang, M. Korner, and J. Denzler. Data association for multi-object Tracking-by-Detection in multicamera networks. In 2012 Sixth International Conference onDistributed Smart Cameras (ICDSC), pages 1–6, Oct. 2012.