International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878,Volume-8, Issue-1C, May 2019

Video Face Detection using Bayesian Technique

Seshaiah Merikapudi, Shrishail Math

Abstract*: Now a days, security based applications are developed widely and these systems are adopted in various real-time applications. Visual surveillance is considered as a most promising technique where certain objects can be detected, tracked and recognized using computer vision based approaches. In this field, face detection and recognition is considered as the important part of surveillance system. Several approaches have been developed for face recognition but existing approaches are applied on the face data. Recently, video face detection techniques are also introduced which provides more information to improve the security system. In this work, we emphasize on the detection of face, along with tracking and recognition using computer vision approach. In order to achieve this objective, first of all we utilized face detection and tracking approach using Kalman filtering. After face detection, we extract the combined features of the input image and stored the trained data. The learning process is developed using Bayesian learning approach. The proposed approach is implemented on benchmark datasets such as IARPA Janus Benchmark A (IJB-A), the YouTube Face repository and the Celebrity-1000 repository. A comparative performance evaluation is carried out which shows the robust performance of proposed approach.

***Index Terms*: Bayesian learning, computer vision, face detection, kalman filtering, visual surveillance**

I. INTRODUCTION

Recently, visual surveillance is an active research topic in industrial and academic field. According to the visual surveillance systems, certain objects detection, tracking and recognition tasks are performed to analyze the object behaviours. Generally, object detection and recognition is also performed using still images such as Hyperface [1] for face recognition and gender recognition, Openface [2] face detection for mobile applications and expression recognition [3]. Due to increased demand of visual surveillance application, video based visual surveillance systems has gained huge attraction from research community. In any visual surveillance system, object detection, tracking and recognition is considered as the main objectives of
the work where specific or suspicious objects can be detected, tracked and recognized to analyze the behavior of the detected objects. These systems are widely adopted in various real-time application such as action recognition [4], crowd behavior analysis [5], industrial applications [6] traffic monitoring [7] and intelligent video surveillance [8].

Revised Manuscript Received on May 21, 2019.
SeshaiahMerikapudi, Research Scholar, Dept. of CSE, SJCIT, Chickballapur, India.merikapudi@gmail.com.
Dr. Shrishail Math, Professor,Dept. of CSE, SKIT, Bangalore, India.

Several techniques have been introduced recently to improve the visual surveillance systems. These techniques are based on the video object detection using salient object detection [9], human detection [10] and object tracking [11]. The moving object detection is the main task in these types of surveillance systems. Now a days, human detection and tracking is considered as the important aspect of these applications. In object detection and tracking, there exists several challenges such as background complexity [12], illumination variation [13], object appearance [14], abrupt motion [15], occlusion, and object shadow [16] etc... On the other hand, the camera position as static or moving camera also causes complexity in the object detection and recognition. In these surveillance systems, object detection plays important role which is helpful to improve the segmentation of background and foreground objects. The appropriate segmentation also helps to improve the recognition and classification. The foreground and background object segmentation depends on the object detection. Recently, human detection and face recognition systems have gained attention in visual surveillance systems which are highly recommended for security systems. Face detection is widely studied using still images but video based face detection can provide several promising information related to the human and other information. Initial studies of video based face detection are based on the still-to-still approach where a good frame is selected and other processing mechanisms are performed on the identified frame to excerpt the information. According to these approaches, initially a frame which contains the frontal face, is selected as the key frame and face is detected. Later, upcoming frames are analyzed based on the size, pose, and illumination etc. If this criteria is satisfied then still-to-still matching process can be implemented to detect the faces. Conventional approaches of face detection initiated from the video frame and then statistical modeling, neural network approach, SVM based methods, color based approaches and Hidden Markov Model (HMM) based approaches are presented to detect the objects. However, these techniques ignore the temporal information of the video sequence which is considered as a majorshortcoming of these techniques. In these systems of face detection, incorporating face tracking is the second main objective of visual surveillance systems. According to the tracking systems, once the face is detected, is tracked throughout the all frames of the system. For complete video, the temporal relationship between frames is obtained which helps to detect the multiple faces in the considered video sequence.

Retrieval Number: A10160581C19/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Generally, these methods are divided into two stages as detection and prediction and then update the tracking process. Similarly, face tracking is also widely studied which uses model based approaches such as Active Appearance Model (AAM) [17], adaptive template tracking [18] and active contour model [19]. Similarly, color and shape based Approaches are also introduced to address the issue of face detection, tracking and recognition. Jairath et al. [20] presented the adaptive skin color model for face detection and tracking using skin color model. Dahal et al. [21] also developed skin color model for face detection and tracking purpose.

Now a days, Convolutional Neural network (CNN) has gained huge attraction from research community and it provides a promising solution for the object detection and recognition. The CNN models are further improved such as Fast and Faster RCNN [22]. Several CNN based systems are developed for wide range of applications such object detection and recognition [23], face detection [24], vehicle detection [25] etc. The existing approaches of R-CNN and its other variants are useful for extracting the deep convolution features with the help of region proposals and later these features areto the classification category. In this work, we focus on the video face detection and recognition using computer vision based approach learned and detected objects are classified according. According to the proposed approach, first of all we consider the input video sequence and extract their frames, later face detection scheme is applied using region proposal generation based RCNN, and later, we apply combined feature extraction model which is trained by using Bayesian Learning model. A general framework of proposed model is depicted in figure 1. Initially, the input video sequence is converted in to the frames and face detection model is applied for training process. This model provides the information about faces and non-faces. The identified faces are extracted and feature extraction mechanism is applied and stored in a trained database. During testing phase, an input video sequence frame is given as input where face detection and feature extraction processes are implemented. The extracted features of the input frames are later processed through feature matching process where testing is applied and this gives us the outcome of the face recognition model

The rest of the manuscript is structuredin the following segments: section II presents a literature review study about recent techniques in this field of visual surveillance systems. Section III presents proposed solution for face detection, tracking and recognition from the benchmark database. Section IV presents experimental study and comparative performance analysis using proposed approach. Finally, section V provides conclusion and future work direction of this approach.

II. LITERATURE SURVEY

In this section, we present a brief discussion on recent techniques used for object detection and recognition through computer vision approach. Moreover, we discuss about the recent techniques of video face detection and recognition using Convolution neural networks. Recently in 2019, Ranjan et al. [1] recommended a scheme i.e. named as HyperFace. In this method intermediate layers of a deep convolution neural networks is fused using another CNN subsequent by multitask learning algorithm based on fused features. It enhances theinteractions between the tasks which ultimately improves the performance of each layer tasks. Moreover, they presented two versions of HyperFace, first one is HyperFace-ResNet, whichuses ResNet-101 model and designed to improvise the detection efficiency and
second is Fast-HyperFace, which makes use of high recall quick face finder to generate ROI recommendations for faster operation speed.

Parkhie et al. [26] presented face recognition mechanism from individual photograph as well as face array tracking from video files. Lately, the progressions in this field of face recognition is because of two important aspects, first being complete learning scenario through CNN, and second, huge open source repositories for training models.

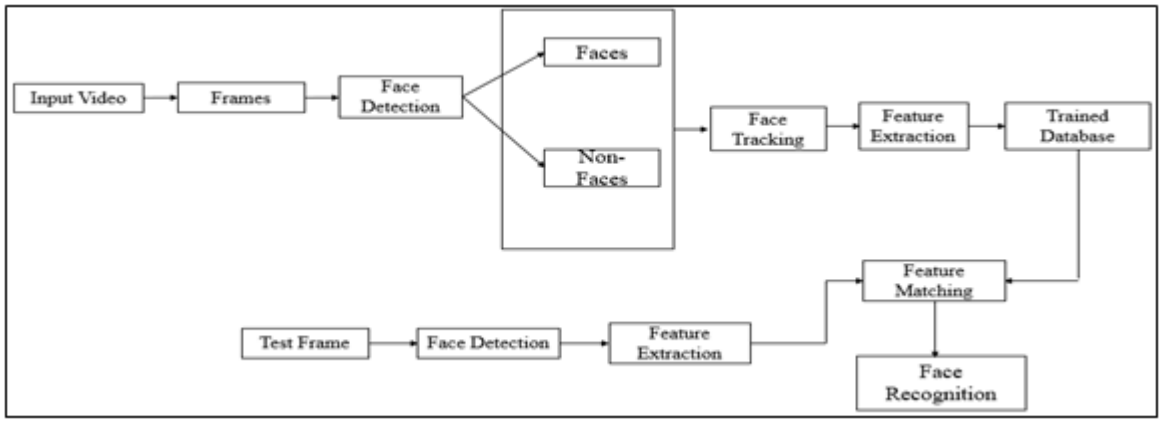

Fig 1: A general framework of proposed model

Yang et al. [27] in 2017, proposed a novel scheme for identification of faces in videos named as, Neural Aggregation Network (NAN).This scheme inputs an individual's face image array or face video with varied face positions and generates a condensed, uniform-dimension feature representation for identification. In this scheme, the entire network is made up of two elements i.e. embedding of feature and aggregation of features. The feature embedding element is basically a Deep CNN that associates every face image to a feature vector. Second element i.e. feature aggregator again involves two attention blocks that adjustably collect the feature vectors to build a single feature within the convex hull generated by them.

Ding et al. [28] developed an end-to-end architecture based on CNNs to counter the issues in VFR (Video-based Face Recognition). Here, the overall procedure can be split into three distinct phases, i.e. (i) learning of blur-robust face representations, where, clear still images are blurred while training to illustrate as real-time video training file. This way CNN is trained with both still images as well as artificially blurred images to learn blur-insensitive features automatically. Secondly, to increase the efficiency of CNN features performance to pose dissimilarities and obstructions, Trunk-Branch Ensemble CNN model (TBE-CNN) is presented, that retrieves corresponding information from whole face images and patches cropped near facial elements. TBE-CNN is a comprehensive model that retrieves features with high proficiently by distributing the low-level and middle-level convolutional layers among the trunk and branch networks. Third, an enhanced triplet loss method was proposed to additionally encourage the discriminative efficiency of the representations learnt by TBE-CNN.

Cao et al. [29] focused on face clustering in videos. Primarily, in contrast to the prevailing common video face clustering schemes that solely concentrates on restraints in the clustering step, their scheme fortify the pairwise restraints via complete video face clustering framework, both in sparse subspace illustration and spectral clustering.

Zheng et al. [30] consider challenging scenarios for unconstrained video-based face recognition from multiple-shot videos and surveillance videos with low-quality frames. In order to handle such type problem a robust and efficient system for unconstrained video-based face recognition, which is composed of face detection, face association, and face recognition. First, multi-scale single-shot face detectors to efficiently localize faces in videos is applied. The detected faces are then grouped respectively through carefully designed face association methods, especially for multi-shot videos. Finally, the faces are recognized by the proposed face matcher based on an unsupervised subspace learning approach and a subspace-to-subspace similarity metric.

Niu et al. [31] presented a scheme to differentiate face and the mask through Video Frame based face detection technique. Libface detection is utilized to spot and crop face from the images, which helped in enhancing speed and accurateness while detection, but then again, images with unclear faces (fractional blocking of the faces) still can be detected, so library is employed to align the faces and retrieve facial eigenvalue that does elimination of blocked faces and subsequently use GMM to classify the faces.

Chen et al. [32] presented a deep learning framework for unimpeded face detection and recognition that includes work on face region identification, linking, orientation and face recognition. Their work combines Deep CNN-based face region identification and facial mark identification components varies from its previous versions [32] in the following means:(i) it makes use of better features via two networks takes faces as input with diverse and (ii) it uses a better metric learning technique that employs inner-products based restraints among triplets to augment the embedding matrix in contrast to norm-based restraints employed in other schemes.

Yang et al. [33] focused on the efficacy of face characteristics based supervision for model training of a powerful face detector. They highlighted a remarkable fact that face component identifiers can be attained from a Convolution Neural Network that has made leaning on recognizing characteristics from uncropped face images, deprived of explicit part supervision. Subsequently, they presented the idea of 'faceness' grade value, which was cautiously computed while taking into consideration facial component responses and the related spatial provisions.

III. PROPOSED MODEL

Previous sections provides a brief discussion about recent techniques of face detection, challenges in face detection and their applications in surveillance systems. Now a days, the video based surveillance systems are widely adopted for real-time security systems and these surveillance videos suffer from low-contrast and poor quality of video frames hence it becomes a challenging task to detect and track the faces. However, various techniques are present for face detection and tracking in literature but achieving desired accurate performance is an open research challenge. Hence, in this work we focus on the accuracy improvement of face detection from videos using computer vision approach. According to the proposed model, first of all we perform the face detection and tracking process.

A. Face Detection & Tracking

Color space models have been adopted widely in various online and offline applications such as RBG color spaces used for display purpose, HSV color spaces are used for computer graphics and $YCbCr$ color spaces are used for video coding and storage application which can be helpful to extract the skin color related information. However, these techniques fails to provide the efficient performance for high complex and occlusion scenarios, hence, we present Kalman filter based approach for face detection and tracking. This scheme is applied for video frames. The Kalman filtering is a linear quadratic process which uses a series of measurements over the time and provides the

Retrieval Number: A10160581C19/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

precise measurement of the unknown variables.

Below given figure 2 shows the flow of Kalman algorithm for measurement and prediction of unknown variables. With the help of Kalman filter, the prediction can be expressed as:

$$X(k) = A(k-1)X(k-1) + W(k) \quad (1)$$

Similarly, the observation can be expressed as:

$$Z(k) = H(k)X(k) + V(k) \quad (2)$$

Where $X(k)$ denotes the state vector, $Z(k)$is the observation vector at the time $t(k)$, similarly, $A(k-1)$ is the state transition matrix and observation matrix is given as $H(k)$.
In this work, we focus on the detection and tracking of the face which is performed into three stages:

- First of all, Kalman filter is initiated with the initial state parameters. This stage is called as initialization of Kalman filter.
- In the next step, positions of the current detection window is predicted in the next upcoming frames, this process is known as state prediction process.
- Finally the states of filters are updated for each frame, this process is called as state updating process.

According to this process, the two consecutive frames are considered and the face can be approximated with the help of linear motion of two frame intervals.

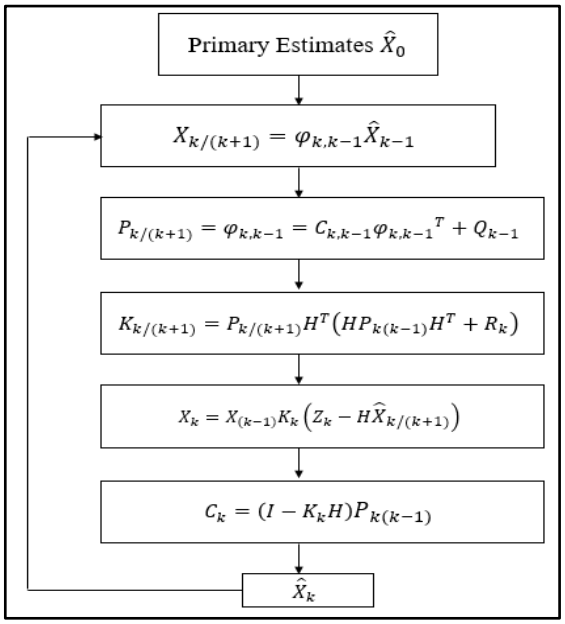

Fig 2: The flow of Kalman algorithm for measurement and prediction of unknown variables.

B. Face Recognition Model

In this section, we present the proposed approach for face recognition. In order to develop a robust model, input image part representation and pose variations need to examine efficiently which can be helpful to improve the accuracy of the system. Thus, this work introduces, path based approach to extract the intermediate information with the help of average pooling. This scheme provides intermediate representation of the detected face. Later, we present a principle component analysis model which reduces the dimension of intermediate patches. For the input video sequence or frames, spatial appearance model is computed to model the human faces. In this work, we use local binary pattern and Local phase quantization features to construct the descriptor.

(a) Local Binary Pattern

This feature extraction model is considered as a powerful process for texture feature extraction. According to this process, 256-bin histogram is computed over the image region to identify the texture descriptors [47]. Later, these features are extended in rotation invariant texture computation operator which identifies neighboring sets of P which are in the rangeof Rradius, represented as $LBP_{P,R}^{riu2}$ where $riu2$ represents the rotation invariant of uniform patterns.

To identify the neighboring radius or pixel values, bilinear interpolation is applied and similarly, rotation invariant parameters are achieved with the help of unique identifier assigned to the each pattern which is obtained using circular bit-wise right shifts. In this pattern, if two bitwise transitions from 0 to 1 are obtained then the pattern is called as uniform patterns, further, the occurrence of uniform pattern is achieved using image histogram. With the help of this LBP process, the histogram of labeled image can be expressed as:

$$\mathcal{H}_i = \sum_{x,y} I(f(x,y) = i), i = 0, \ldots n-1 \quad (3)$$

Where $f(x,y)$ denotes the input image, and (n) represents maximum number of labels and, $I(F)$ is the function which return 0 or 1 based on the occurrence of pattern.

(b) Local phase quantization

In this sub-section, we present the local quantization descriptor which provides the local phase information extracted using 2D- DFT. According to this process, an input image $f(x)$ is considered whose neighbourhood pixels are N_x in a rectangular form as $(H \times H)$ which are at the x position, on this pixel position, a short-term Fourier transform (STFT) is computed to generate the descriptor as:

$$F(u,x) = \sum_{y \in N_x} f(x-y)e^{-j2\pi u^T y} = W_u^T f_x \quad (4)$$

Where w_u represents the 2D DFT vector at u frequency and f_x denotes the total number of vectors which has total H^2 samples from the N_x. In this process of descriptor generation, the STFT is computed by performing 1D-Convlution for rows and columns successively. Further, Fourier coefficients are computed at the four different frequency points as $u_1 = [c,0]^T$, $u_2 = [0,c]^T$, $u_3 = [c,c]^T$ and $u_1 = [c,-c]^T$ where c is a smaller scalar coefficient.

Retrieval Number: A10160581C19/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Phase of these Fourier coefficients later is used for identifying the real and imaginary parts with the help of scalar quantizer, given as:

$$q_j = \begin{cases} 1 & if\ d_j \geq 0 \\ 0 & otherwise \end{cases} \quad (5)$$

Where d_j is the j^{th} function of real and imaginary vector given as $D_x = [Re\{F_x\}, Im\{F_x\}]$, with the help of this, the binary coefficients can be represented as:

$$f(x) = \sum_{j=1}^{B} q_j 2^{j-1} \quad (6)$$

Where B denotes the histogram bins.

With the help of these two descriptors, a combined feature descriptor model is created as $\{\mathcal{F}\} = \{[a_1, S_i]\}_{i=1}^{M}$ where a denotes the appearance model and S denotes the spatial feature part. During the training process, the patch based background model is created using Expectation minimization model. This training model is considered as Gaussian mixture model which contains K spherical Gaussian components, given as:

$$P([a\ S]|\Theta) = \sum_{k=1}^{K} w_k G([a\ S])\mu_k, \sigma_k^2 I) \quad (7)$$

Where, w_k is the mixture weight of Gaussian component, $G([a\ S])\mu_k, \sigma_k^2 I)$ represents thee spherical Gaussian with μ_k mean and $(\sigma_k^2 I)$ variance. From this generated descriptors, we compute the probability and highest probability patches are combined into a vector. Further, we present PCA based dimension reduction model which represents the image patches into the low-dimensional subspace.

(c) Bayesian Learning

In this section, we present the Bayesian learning framework for face recognition and verification using Bayesian model. This process is presented as $s = f(x_1, x_2)$ where x_1 and x_2 denotes the feature pairs as $(x_1, x_2) \in R^n$ and $s \in R$ represents the similarity. For training process, we use fully connected layers of Multi-Layer Perceptron (MLPs). In this process, two images as I_1 and I_2 which produces x_1 and x_2 features, the feature loss of this learning process can be given as:

$$\mathbb{L}_1(I_1, I_2, \Theta) = \frac{1}{2}\big(N(I_1, I_2\Theta) - J(x_1, x_2)\big)^2 \quad (8)$$

Similarly, the other loss is computed based on the similarity between original and reduced input image pairs. This is expressed as:

$$\mathbb{L}_2(I_1, I_2, y; \Theta) = (1-y) \times e^{\frac{1}{c}(N(I_1, I_2; \Theta) - t} + y \times e^{-\frac{1}{c}(N(I_1, I_2; \Theta) - t} \quad (9)$$

With the help of these two loss functions, the final loss can be measured by taking the weighted average of Eq. (6) and (7), as

$$\mathbb{L}(I_1, I_2, y; \Theta) = w_1 \mathbb{L}_1(I_1, I_2, \Theta) + w_2 \mathbb{L}_1(I_1, I_2, \Theta) \quad (10)$$

Based on this model, the detection verification can be measured using similarity value S, as:

$$P_i = \begin{cases} 1 & if\ S_i \geq t \\ 0 & else\ S_i \leq t \end{cases} \quad (11)$$

Where t denotes a threshold value for matching thesimilaity between image features.

IV. RESULTS AND DISCUSSION

In this section, we present a complete experimental analysis using proposed approach. In order to evaluate the performance of proposed approach, we have considered open source video face recognition database which are IARPA Janus Benchmark A (IJB-A) [34], the YouTube Face dataset [35], and the Celebrity-1000 dataset [36].

A. IJB-A Dataset

This subsection presents experimental analysis for IJB dataset [33]. In this dataset, total 500 subjects with 5397 image and 2040 videos with 20412 frames are present. The dataset contains various types of challenges such as pose variation, viewpoint and illumination variation. Moreover, still images are also incorporated which causes complexity during training process. To measure the performance, we consider two criteria as 1:1 verification where images belongs to the same category and another is 1: N Mixed search where data is mixed by using different images. The performance of proposed model is computed and measured in terms of true accept rates vs. false positive rates and true positive identification rate (TPIR) vs. false positive identification rate (FPIR). A comparative study for 1:1 verification is presented in table 1.

Table 1: A comparative study for 1:1 verification

Technique Used	FAR=0.01	FAR=0.1
LSFS [37]	0.733 ± 0.034	0.895 ± 0.013
DCNNmanual+ metric[38]	0.787 ± 0.043	0.947 ± 0.011
Triplet Similarity [39]	0.790 ± 0.030	0.945 ± 0.002
Deep Milti-Pose [40]	0.876	.954
DCNNfusion [41]	0.838 ± 0.042	0.967 ± 0.009
Triplet Embedding [39]	0.90 ± 0.01	0.964 ± 0.005
Proposed Model	0.92±0.01	0.97±0.002

Similarly, we evaluate the performance for 1: N scenario as depicted in table 2 where we compare the performance of proposed approach with the existing techniques.

Table 2: the performance for 1: N scenario

Technique Used	FPIR=0.01	FPR=0.1
LSFS [37]	0.383±0.063	0.613 ± 0.032
Triplet Similarity [39]	0.556±0.065	0.754±0.014
Deep Milti-Pose [40]	0.52	0.75
DCNNfusion [41]	0.577±0.094	0.790±0.033
Triplet Embedding [39]	0.753 ± 0.03	0.863±0.014
Proposed Model	0.78±0.01	0.894±0.0011

The above given comparative analysis in table 1 and 2 shows that the proposed approach achieves better performance when compared with the existing approaches such as LSFS [37], $DCNN_{manual\ +metric}$ [38] Triplet Similarity [39], Deep Multi-Pose [40], DCNNfusion [41], and Triplet Embedding [39].

B. YouTube Face Database

In this section, we present the face detection performance analysis of YouTube face database [35] which is developed for face detection in videos. This dataset contains total 3425 videos of 1595 people and the video length vary from 48 to 6,070 frames. The performance of these models is compared in terms of face detection accuracy and Area Under Curve (AUC). Table 3 shows a comparative performance for face detection for YouTube dataset.

Table 3: The performance comparison of face detection for YouTube dataset.

Technique Used	Accuracy	AUC
LM3L [42]	81.3 ± 1.2	89.3
DDML(combined)[43]	82.3±1.5	90.1
DeepFace Single[44]	91.4±1.1	96.3
DeepID2+ [45]	93.2±0.2	-
Wen *et al.* [46]	94.9	-
CNN+Max. L2	91.96±1.1	97.4
CNN+Min. L2	94.96±0.79	98.5
CNN+MaxPool	88.36±1.4	95.0
Proposed Model	95.22±1.1	98.22

According to the table 3, proposed approach achieves better performance in terms of face detection accuracy and AUC. The overall accuracy of proposed model is obtained as 95.22 which has improved by 6.06% when compared with the CNN+MaxPool model

C. The Celebrity-1000 dataset

The Celebrity-1000 dataset mainly focused on the video based face identification problem. This data contains total 159726 video sequences which includes total 1000 human subjects and total 2.4 M frames are available in this. This dataset provides two types of test protocols are open-set and close-set with the data [48]. The performance for close-set data is depicted in table 4 and performance of proposed approach is compared with the existing techniques. In order to evaluate the performance, we consider varied number of subjects and computed rank-1 frequency.

Table 4: The performance for close-set data

Technique Used	100	200	500	1000
MTJSR [48]	50.60	40.80	35.46	30.04
CNN+Mean L2	85.26	77.59	74.57	67.91
CNN+AvePool-VideoAggr	86.06	82.38	80.48	74.26
CNN+AvePool - SubjectAggr	84.46	78.93	77.68	73.41
Proposed Model	91.22	86.89	85.33	82.67

Similarly, we consider the open-set data base from Celebrity-1000 dataset and measured the performance. The performance of proposed model is compared with the existing techniques as depicted in table 5.

Table 5: The performance of proposed model is compared with the existing techniques.

Technique Used	100	200	400	800
MTJSR [48]	46.12	39.84	37.51	33.50
CNN+Mean L2	84.88	79.88	76.76	70.67
CNN+AvePool - SubjectAggr	84.11	79.09	78.40	75.12
Proposed Model	88.90	85.26	80.21	79.22

V. CONCLUSION

In this work, we have presented a novel approach for improving the visual surveillance system performance. According to the existing approaches, most of the techniques are based on the still image based face recognition however the increasing demand of security application require urge for the development of video based surveillance system where multiple faces can be detected, tracked and recognized. In this article, we use a Kalman filtering based approach for face detection and tracking, later combined feature extraction model is developed which is later user for generating the trained database. For recognition of detected face, Bayesian learning scheme is developed. A comparative study is presented for face detection which shows the robust performance of proposed approach. This work further can be extended for CNN based network model for faster recognition in occlusion and low illumination scenarios.

Retrieval Number: A10160581C19/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

REFERENCES

1. Ranjan, R., Patel, V. M., &Chellappa, R. (2019). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1), 121-135.
2. Amos, B., Ludwiczuk, B., &Satyanarayanan, M. (2016).Openface: A general-purpose face recognition library with mobile applications. CMU School of Computer Science, 6.
3. Gangopadhyay, I., Chatterjee, A., & Das, I. (2019). Face Detection and Expression Recognition Using Haar Cascade Classifier and Fisherface Algorithm. In Recent Trends in Signal and Image Processing (pp. 1-11). Springer, Singapore.
4. S. Wu, O. Oreifej, and M. Shah, Action recognition in videos acquired by a moving camera using motion decomposition of Lagrangian particle trajectories, IEEE International Conference on Computer Vision, (Nov. 2011), 1419-1426.
5. R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, Statistic and knowledge-based moving object detection in traffic scenes, IEEE Intelligent Transportation Systems, (2000), 27-32.
6. E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, and J. D. Legat, A survey on industrial vision systems, applications and tools, Image and vision computing, 21(2) (2003), 171-188.
7. W. Hu, T. Tan, L. Wang, and S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3), (2004), 334-352.
8. J. S. Kim, D. H. Yeom, and Y. H. Joo, Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems, IEEE Transactions on Consumer Electronics, 57(3), (2011), 1165-1170.
9. Wang, W., Shen, J., & Shao, L. (2018). Video salient object detection via fully convolutional networks. IEEE Transactions on Image Processing, 27(1), 38-49.
10. Gajjar, V., Gurnani, A., &Khandhediya, Y. (2017). Human detection and tracking for video surveillance: A cognitive science approach. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2805-2809).
11. Xu, R., Guan, Y., & Huang, Y. (2015). Multiple human detection and tracking based on head detection for real-time video surveillance. Multimedia Tools and Applications, 74(3), 729-742.
12. Javed, S., Mahmood, A., Bouwmans, T., & Jung, S. K. (2018). Spatiotemporal low-rank modeling for complex scene background initialization. IEEE Transactions on Circuits and Systems for Video Technology, 28(6), 1315-1329.
13. Kulchandani, J. S., &Dangarwala, K. J. (2015, January). Moving object detection: Review of recent research trends. In 2015 International Conference on Pervasive Computing (ICPC) (pp. 1-5). IEEE.
14. Kang, K., Ouyang, W., Li, H., & Wang, X. (2016). Object detection from video tubelets with convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 817-825).
15. Morimitsu, H., Bloch, I., & Cesar-Jr, R. M. (2017). Exploring structure for long-term tracking of multiple objects in sports videos. Computer Vision and Image Understanding, 159, 89-104.
16. Real, E., Shlens, J., Mazzocchi, S., Pan, X., &Vanhoucke, V. (2017). Youtube-boundingboxes: A large high-precision human-annotated data set for object detection in video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5296-5305).
17. Dornaika, F., &Ahlberg, J. (2004). Fast and reliable active appearance model search for 3-D face tracking. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34(4), 1838-1853.
18. J Saragih, R Goecke, Monocular and Stereo Methods for AAM Learning from Video CVPR '07. IEEE Conference on Computer Vision and Pattern Recognition, pp:1-8,2007
19. Le, T. H. N., &Savvides, M. (2016). A novel shape constrained feature-based active contour model for lips/mouth segmentation in the wild. Pattern Recognition, 54, 23-33.
20. Jairath, S., Bharadwaj, S., Vatsa, M., & Singh, R. (2016). Adaptive skin color model to improve video face detection. In Machine Intelligence and Signal Processing (pp. 131-142). Springer, New Delhi.
21. Dahal, B., Alsadoon, A., Prasad, P. C., &Elchouemi, A. (2016, March). Incorporating skin color for improved face detection and tracking system. In 2016 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI) (pp. 173-176). IEEE.
22. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).
23. Cai, Z., &Vasconcelos, N. (2018). Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6154-6162).
24. Jiang, H., & Learned-Miller, E. (2017, May). Face detection with the faster R-CNN. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 650-657). IEEE.
25. Fan, Q., Brown, L., & Smith, J. (2016, June). A closer look at Faster R-CNN for vehicle detection. In 2016 IEEE intelligent vehicles symposium (IV) (pp. 124-129). IEEE.
26. Parkhi, O. M., Vedaldi, A., &Zisserman, A. (2015, September). Deep face recognition. In bmvc (Vol. 1, No. 3, p. 6).
27. Yang, J., Ren, P., Zhang, D., Chen, D., Wen, F., Li, H., & Hua, G. (2017). Neural aggregation network for video face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4362-4371).
28. Ding, C., & Tao, D. (2018). Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE transactions on pattern analysis and machine intelligence, 40(4), 1002-1014.
29. Cao, X., Zhang, C., Zhou, C., Fu, H., &Foroosh, H. (2015). Constrained multi-view video face clustering. IEEE Transactions on Image Processing, 24(11), 4381-4393.
30. Zheng, J., Ranjan, R., Chen, C. H., Chen, J. C., Castillo, C. D., &Chellappa, R. (2018). An Automatic System for Unconstrained Video-Based Face Recognition. arXiv preprint arXiv:1812.04058.
31. Niu, G., & Chen, Q. (2018). Learning an video frame-based face detection system for security fields. Journal of Visual Communication and Image Representation, 55, 457-463.
32. Chen, J. C., Ranjan, R., Sankaranarayanan, S., Kumar, A., Chen, C. H., Patel, V. M., ...&Chellappa, R. (2018). Unconstrained still/video-based face verification with deep convolutional neural networks. International Journal of Computer Vision, 126(2-4), 272-291.
33. Yang, S., Luo, P., Loy, C. C., & Tang, X. (2018). Faceness-net: Face detection through deep facial part responses. IEEE transactions on pattern analysis and machine intelligence, 40(8), 1845-1859.
34. B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpajanus benchmark a. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1931–1939, 2015. 2, 4, 5, 6.
35. L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 529–534, 2011.
36. L. Liu, L. Zhang, H. Liu, and S. Yan. Toward large population face identification in unconstrained videos. IEEE Transactions on Circuits and Systems for Video

Retrieval Number: A10160581C19/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Technology, 24(11):1874–1884, 2014.
37. D. Wang, C. Otto, and A. K. Jain. Face search at scale: 80 million gallery. arXiv preprint arXiv:1507.07242, 2015
38. J.-C. Chen, R. Ranjan, A. Kumar, C.-H. Chen, V. Patel, and R. Chellappa. An end-to-end system for unconstrained face verification with deep convolutional neural networks. In IEEE International Conference on Computer Vision Work- shops, pages 118–126, 2015.
39. S. Sankaranarayanan, A. Alavi, C. Castillo, and R. Chellappa. Triplet probabilistic embedding for face verification and clustering. arXiv preprint arXiv:1604.05417, 2016.
40. W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, J. Choi, J. Lekust, J. Kim, P. Natarajan, et al. Face recognition using deep multi-pose representations. In IEEEWinter Conference on Applications of Computer Vision (WACV), 2016.
41. J.-C. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep cnn features. In IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.
42. J. Hu, J. Lu, J. Yuan, and Y.-P. Tan. Large margin multimetric learning for face and kinship verification in the wild. In Asian Conference on Computer Vision (ACCV), pages 252–267. 2014
43. J. Hu, J. Lu, and Y.-P. Tan. Discriminative deep metric learning for face verification in the wild. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1875–1882, 2014
44. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. DeepFace: Closing the gap to human-level performance in face verification. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1701–1708, 2014.
45. Y. Sun, X.Wang, and X. Tang. Deeply learned face representations are sparse, selective, and robust. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2892–2900, 2015
46. Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In European Conference on Computer Vision (ECCV), pages 499–515, 2016
47. S V N, Murthy, B K, Sujatha, “A Novel Graph-Based Technique to Enhance Video Compression Algorithm” in Emerging Research in Computing, Information, Communication and Applications, Springer, New Delhi , pp. 463-468.
48. L. Liu, L. Zhang, H. Liu, and S. Yan. Toward largepopulation face identification in unconstrained videos. IEEE Transactions on Circuits and Systems for Video Technology, 24(11):1874–1884, 2014.
49. X.-T. Yuan, X. Liu, and S. Yan. Visual classification with multitask joint sparse representation. IEEE TIP, 21(10):4349–4360, 2012.

AUTHORS PROFILE

SeshaiahMerikapudi,B.E,M.TECH,CSI Member, Research Scholar(VTU Belgavi)Dept. of CSE, SJCIT,Chickballapur-562101, India
merikapudi@gmail.com.

Dr.ShrishailMath,B.E,MTECH,Ph.D,MQCI ,MCSI,MISTE,FIE,FIETE,Professor, in theDept. of CSE, SKIT,Bangalore, India
shri_math@yahoo.com.

Retrieval Number: A10160581C19/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication