

Comparative Analysis of Various Face Detection and Tracking and Recognition Mechanisms using Machine and Deep Learning Methods

¹Prof. M. Seshaiyah, ²Prof. Shrishail Math,

¹Research Scholar in Dept of CS&E, VTU, Belgaum, India

²professor in Dept of CS&E, SKIT, Bangalore, India.

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract: Recently, the security of individuals has become the prime concern for the human community. Various real-time security management systems are developed widely. Visual surveillance is considered as one of the promising technique to improve the security which helps to detect and recognize the objects. Numerous techniques have been developed for real-time video surveillance. Face detection, tracking and recognition is one of the important part of visual surveillance systems. The existing face detection schemes suffer from various challenging issues such as pose variations, illumination conditions and occlusion and many more. To overcome these issues, we have developed new schemes which includes Bayesian learning, Region Based Convolutional Neural Networks(RCNN) and GoogleNet based CNN model for face detection, tracking and recognition. In this work, we compare the performance of proposed schemes with existing schemes for different datasets to conclude the robustness of proposed approach.

Keywords – Face detection, tracking and recognition, Bayesian learning, GoogleNet.

1. Introduction

The demand of surveillance applications have gained huge attraction and the surveillance systems are widely deployed in various real-time applications. Various techniques have been presented for security systems such as iris recognition, speech recognition, fingerprint verification and signature verification [1]. The input data obtained from irises, speech prints, fingerprints, signatures, individual faces, and so on are widely exploited by the technologies of biometric recognition systems for personal identification. Generally, these techniques of voice, fingerprint and signature recognition categorized as the passive methods that involve high quality camera to capture the biometric information of human [2]. Moreover, these cameras obtain the information from short range of region of interest. Thus, mostly these systems are used for authentication purpose and less suitable for visual surveillance systems. On the contrary, the face detection and recognition system has gained huge attraction for real-time visual surveillance systems [3].

Due to its value in various applications, such as video structuring, indexing, retrieval, and summarization, human face processing techniques for broadcast video, including face detection, tracking, and recognition, have attracted a lot of research interest. The main explanation for this is that rich knowledge is given by the human face to spot the presence of such individuals of interest [4].

Currently, several techniques are present for face detection in image and video sequences. Most of the existing methods are based on the concept of feature extraction and dictionary learning such as small-scale illumination invariant features [9], Gabor-features [27], 3D-DWT [28] and many more [29]. On the other hand, the dictionary based learning schemes are widely adopted for face recognition systems due to their significant performance for pose variation, illumination change and varied expressions. However, these techniques follow the unsupervised learning and bring the poor performance for various face sets. Moreover, these techniques require raw pixel which may contains noise data that can affect the dictionary learning performance. During last decade, several researches have been presented to overcome these challenging issues in this field of face recognition and verification. Similarly, video-based identification is one of the most difficult challenges in the area of real-time surveillance systems [9]. Video based recognition system records individual faces from different perspectives and offers various valuable details about the single face. However, the performance of video-based recognition system is affected for uncontrolled pose, and illumination scenarios which leads to poor classification performance. Now a days, deep learning schemes have gained huge attraction for computer vision based applications. The deep learning schemes are also adopted for video face detection and recognition due to their feature pooling and robust feature learning nature. In this context, we developed three techniques for video face recognition. First of all, we focused on the face detection, tracking and recognition using computer vision system. In this scheme, we apply Kalman filtering for face detection and tracking approach. Later, we extract the combined features of the input image and stored the trained

Comparative Analysis of Various Face Detection and Tracking and Recognition Mechanisms using Machine and Deep Learning Methods

data. The training is performed using Bayesian learning approach. In next phase, we developed CNN architecture for face detection and bounding box regression. Finally, a GoogleNet based architecture is developed for video face recognition and tracking.

2. Literature survey

Wang et al. [5] discussed that the huge amount of data is generated and uploaded on internet. However, face detection has been carried out efficiently but dealing with unconstrained face images still remains a challenging task for research community. In order to overcome this issue, authors developed a fast search process with state-of-the-art commercial off the shelf (COTS) matcher. This scheme uses cascade architecture to combine these modules. In this approach, a face image is given as input and deep features are extracted. These features are trained using convolutional neural network and top-k similar face images are identified. Further, these obtained k- similar faces are re-ranked based on similarities.

Chen et al. [6] introduced deep convolutional neural networks (DCNN) model for face detection and verification problem. This scheme uses a face pre-processing module where face detection and landmark detection are performed. In face detection phase, it constructs the pyramid modules for deep feature extraction, later, face association module is constructed where face tracker module is designed. Later, face alignment is applied by using global shape indexed features. Finally, DCNN model and joint Bayesian learning model is applied for face verification.

Sankaranarayanan et al. [7] also focused on unconstrained face recognition and presented a deep learning and triplet probability model for face verification. This triplet model uses low-dimensional discriminative embedding for learning.

AbdAlmageed et al. [8] developed a face recognition system by considering the several pose-specific deep convolutional neural network which helps to generate the multiple pose-specific features. In this work, a 3D rendering model is also used to generate the multiple face poses from the given input image.

Hu et al. [10] discussed that metric learning based schemes are widely adopted in face verification systems. Generally, the existing schemes learn one Mahalanobis distance metric from single image feature and fail to deal with multiple features. Due to increased complexity visual surveillance, it is necessary to extract multiple features. In order to deal with this issue, authors developed a new large margin multi-metric learning (LM3L) method for face verification. This scheme uses a distance threshold to obtain the distance correlation between different features.

In the same content, Hu et al. [11] again focused on metric learning scheme and developed deep learning approach for face verification. Similar to [10] in this work also, authors select the Mahalanobis distance metric to learn the features which is used to maximize the inter-class variation and minimize the intra-class variations.

Taigman et al. [12] reported that conventional deep learning based schemes are based on the four main stages which are detect, align, represent and classify. In this work, authors focused on redefining the alignment and representation step. By employing explicit 3D face modelling to apply a piecewise affine transformation, we revisit both the alignment step and the representation step, and derive a face representation from a nine-layer deep neural network. This deep network includes more than 120 million parameters, rather than the usual convolutional layers, utilizing multiple locally linked layers without weight sharing.

Sun et al. [13] developed a high-performance deep convolutional network (DeepID2+) for face recognition. This approach increases the dimensions of hidden representations and the convolution layers are supervised to improve the detection performance. However, the performance of deep learning approach depends on sparsity, selectiveness and robustness of the data. Authors concluded that activation functions are moderately sparse which increases the discriminative power of deep net and distance between images. Moreover, the DeepID2+ shows more robust performance for occlusion scenarios.

Wen et al. [14] discussed the importance of CNN in computer vision community. Most of the CNN architecture use softmax loss function to train the deep neural network. For further improvement in feature learning, a center loss function based supervision scheme is incorporated. Mainly, the center loss function learns the deep feature of each class and evaluates the distance between features and class. Moreover, this loss function is trainable and optimizes the working of CNN.

Yuan et al. [15] also discussed that face detection in unconstrained videos is a challenging task. Current researches for unconstrained videos use a very small size dataset and various datasets are captured in controlled laboratory environment. Thus, in this work, authors considered a large-scale video dataset in unconstrained environment. For face detection, this approach uses multitask joint sparse representation (MTJSR) that is a training free scheme and can be integrated with multiple frames of same tracking sequence. This makes it more suitable for video based identification. Moreover, a sparsity-induced scalable optimization method is also considered to solve the large-scale issues of MTJSR where these problems are solved by considering a smaller-scale sub problems.

Ortiz et al. [16] developed a video based face recognition system for huge datasets. The conventional l_1 -minimization technique performs frame by frame analysis which becomes computationally complex thus authors introduced Mean Sequence SRC (MSSRC) approach. This approach considers the entire information of the video data and face track information of individual.

Sivic et al. [17] discussed about the face detection and recognition of TV or movie characters by integrating the frontal and profile face detection in face detection pipeline. Moreover, this scheme uses a combined kernel strategy for recognition.

Parkhi et al. [18] presented a new study for face detection and recognition. The main contributions of this approach are as follows: it extracts the supervisory information from aligned faces, it is able to classify the background characters, it extracts the significant unique features using ConvNet and it labels the face tracks based on linear programming.

Yang et al. [19] introduced a Neural Aggregation Network (NAN) for video face recognition. The network takes as its input a face video or face picture set of a person with a variable number of face images, and produces for recognition a compact, fixed-dimension feature representation. Two modules compose the entire network. A deep Convolutional Neural Network (CNN) that maps each face image to a feature vector is the feature embedding module.

Crosswhite et al. [20] discussed the problem of template adaption in video face recognition. Template adaption is a process of transfer learning where target is defined by media data of subject in the considered template. The transfer learning approach uses source domain for feature encoding and target domain with limited available observations. Authors developed a simple method for template adaption using deep convolutional network and one-vs-rest linear SVMs.

In this context, we focused on the issues and challenges of face recognition and introduced novel schemes using Bayesian Learning [24], RCNN based technique [25] and GoogleNet based video face recognition [26].

3. Comparative analysis

In this section, we present the experimental analysis of proposed approach for face detection and recognition in still images, videos and real-time videos. The proposed approach is implemented on Python3.7 running on windows platform with NVIDIA GPU. 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) [21], the YouTube Face dataset [22], and the Celebrity-1000 dataset [23].

3.1. Comparative analysis using Bayesian Learning

3.1.1. IJB-A dataset

3.1.2.

This subsection presents experimental analysis for IJB dataset. 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.

Technique Used	FAR=0.01	FAR=0.1
LSFS [5]	0.733 ± 0.034	0.895 ± 0.013
DCNNmanual+metric[6]	0.787 ± 0.043	0.947 ± 0.011
Triplet Similarity [7]	0.790 ± 0.030	0.945 ± 0.002
Deep Milti-Pose [8]	0.876	.954
DCNNfusion [6]	0.838 ± 0.042	0.967 ± 0.009
Triplet Embedding [7]	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.

Comparative Analysis of Various Face Detection and Tracking and Recognition Mechanisms using Machine and Deep Learning Methods

Technique Used	FPIR=0.01	FPR=0.1
LSFS [5]	0.383 ± 0.063	0.613 ± 0.032
Triplet Similarity [7]	0.556±0.065	0.754±0.014
Deep Multi-Pose [8]	0.52	0.75
DCNNfusion [6]	0.577±0.094	0.790±0.033
Triplet Embedding [7]	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 , $DCNN_{manual+metric}$ Triplet Similarity, Deep Multi-Pose, DCNNfusion, and Triplet Embedding [39].

3.1.3. 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.

Technique Used	Accuracy	AUC
LM3L [10]	81.3 ± 1.2	89.3
DDML(combined)[11]	82.3±1.5	90.1
DeepFace Single[12]	91.4±1.1	96.3
DeepID2+ [13]	93.2±0.2	-
Wen <i>et al.</i> [14]	94.9	-
CNN+Max. L2 [14]	91.96±1.1	97.4
CNN+Min. L2 [14]	94.96±0.79	98.5
CNN+MaxPool [14]	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

3.1.4. 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. 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.

Technique Used	100	200	500	1000
MTJSRC [15]	50.60	40.80	35.46	30.04
CNN+Mean L2[15]	85.26	77.59	74.57	67.91
CNN+AvePool VideoAggr[15]	- 86.06	82.38	80.48	74.26
CNN+AvePool SubjectAggr[15]	- 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.

Technique Used	100	200	400	800
MTJSRC [15]	46.12	39.84	37.51	33.50
CNN+Mean L2 [15]	84.88	79.88	76.76	70.67
CNN+AvePool– SubjectAggr[15]	84.11	79.09	78.40	75.12
Proposed Model	88.90	85.26	80.21	79.22

3.2. Comparative analysis using Background removed Faster RCNN

In this section we present the experimental analysis using proposed approach for face detection and recognition from video datasets. The proposed method is evaluated for using publically available dataset which are known as YouTube Faces, YouTube Celebrities, Buffy. Below given figure shows some sample images of the YouTube celebrity dataset.



The performance of proposed approach is compared with the existing techniques. In this work we also compare the performance of proposed approach in terms of tracking. The tracking performance analysis is presented in below given next sub-section.

3.2.1. Video face tracking performance

We consider five movie trailer from the dataset which are 'Killer Inside', 'My Name is Khan', 'Beautiful', 'Eat, Pray, Love', and 'The Dry Land'. In order to measure the performance we use object tracking accuracy and object tracing precision. The obtained performance is presented in table 1.

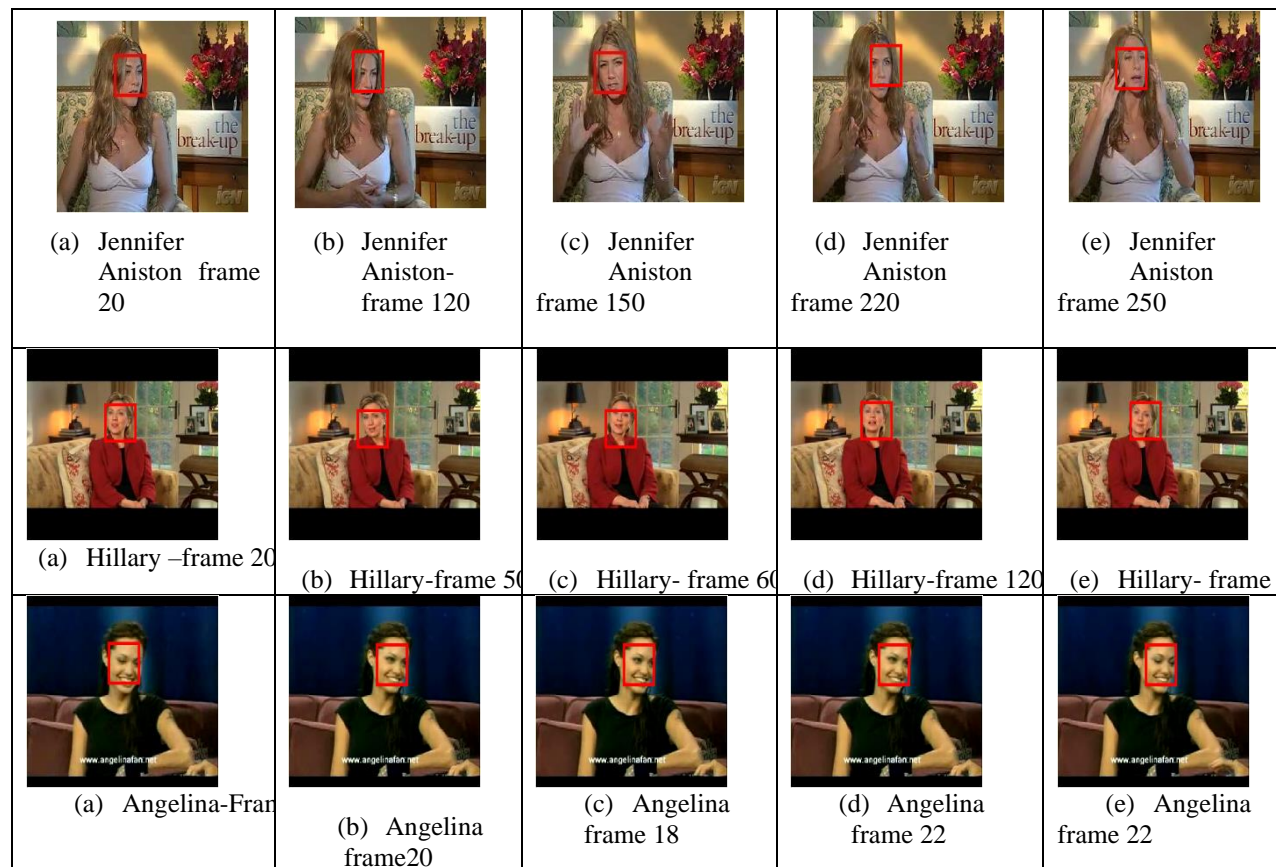
Table 1 Face tracking performance

Name of the Sequence		MSSRC [16]	BF-RCNN-VFR (Proposed Model)
'The Killer Inside'	Precision	69.35	86.33
	Accuracy	42.16	92.28
'My Name is Khan'	Precision	65.77	90.28
	Accuracy	48.24	91.24
'Biutiful'	Precision	61.34	88.56
	Accuracy	43.96	92.36
'Eat Pray Love'	Precision	56.77	85.21
	Accuracy	35.6	89.66

Comparative Analysis of Various Face Detection and Tracking and Recognition Mechanisms using Machine and Deep Learning Methods

‘The Dry Land’	Precision	62.7	93.52
	Accuracy	30.15	91.2

The obtained tracking results are depicted in below given figure. Each row of the figure shows the tracking results for different frames of 5 videos.



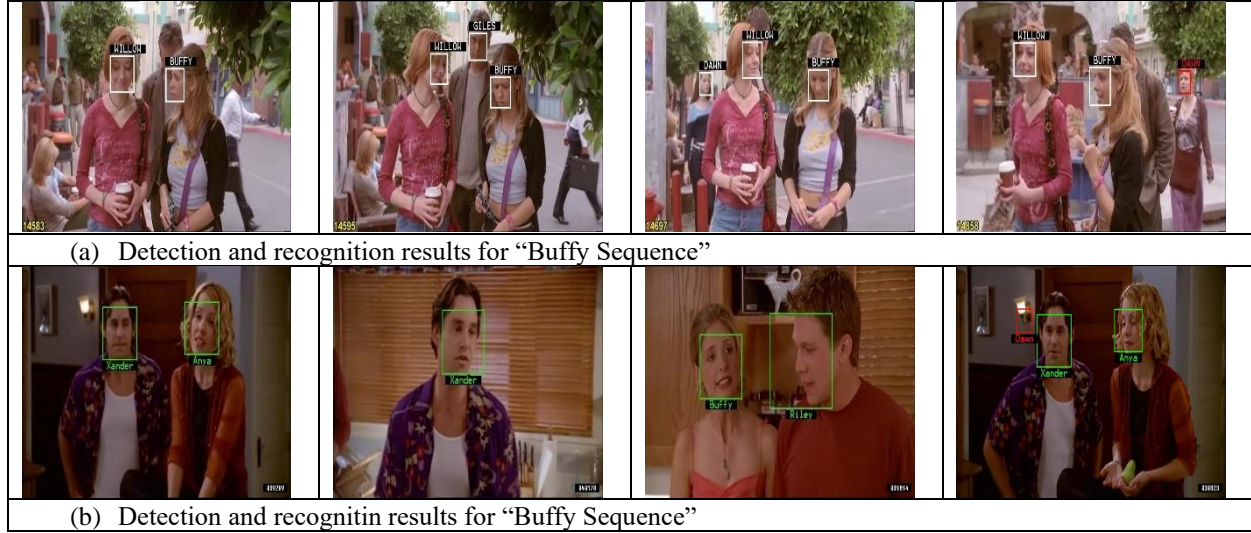
3.2.2. Video face recognition performance

In this section we present the face recognition performance using proposed approach. This experiment is carried out using YouTube Faces Dataset, YouTube Celebrities Dataset and Buffy Dataset

The YouTube face dataset is a huge dataset which contains total 3,425 videos which are acquired from 1,595 different people. These videos are obtained from the YouTube. The shortest clip duration is 48 frames, the longest clip is 6,070 frames, and the average length of a video clip is 181.3 frames. Similarly, YouTube celebrity data is obtained which contains total 1910 videos of 47 different people. The minimum frames are 8 and the maximum frame in a video are 400 in this dataset.

The buffy dataset contains total 639 face tracks which are extracted from the TV series “Buffy the Vampire Slayer”, this dataset is obtained from the episodes 9, 21 and 45.

The recognition results for buffy video sequence are presented in below given figure where the correctly detected faces are presented in white bounding box and incorrect recognition is depicted in red bounding box.



The performance of Buffy dataset is measured in terms of average precision and compared with the existing techniques. The obtained performance comparison is presented in table 2.

Table 2 Performance comparison for "Buffy Dataset"

Buffy Episode	[17]	[18]	[18] with LP	[18] Raw	Proposed Model
1	0.90	0.98	0.99	0.92	0.99
2	0.83	0.96	0.96	0.95	0.98
3	0.73	0.95	0.95	0.91	0.98
4	0.86	0.97	0.96	0.92	0.99
5	0.85	0.97	0.97	0.93	0.98

3.3. Comparative analysis using Googlenet

3.3.1. Results for IJB-A dataset

In this section, we present the face detection and recognition analysis for IJB-A dataset which contain videos and face images from different environment. This data has multiple variations and conditions thus it becomes a challenging task of face recognition. Below given table shows the comparative analysis in terms of the true accept rates (TAR) vs. false positive rates (FAR) where we compared the performance of proposed approach with existing techniques.

Method	1:1 Verification TAR		
	FAR =0.001	FAR =0.01	FAR -0.1
LSFS [5]	0.514±0.060	0.733±0.034	0.895±0.013
DCNN [6]	-	0.787±0.043	0.947±0.011
Triplet Similarity [31]	0.590±0.050	0.790±0.030	0.945±0.002
Triplet Embedding [31]	0.813±0.02	0.90±0.01	0.964±0.005
Template Adaption [35]	0.836±0.020	0.939±0.013	0.979±0.004
CNN + Max L2 [19]	0.202±0.029	0.345±0.025	0.601±0.024
CNN + Min L2 [19]	0.038±0.008	0.144±0.073	0.972±0.006
CNN + Mean L2 [19]	0.688±0.080	0.895±0.016	0.978±0.004
CNN + Soft Min L2 [19]	0.697±0.085	0.904±0.015	0.978±0.004
CNN + Max Pool [19]	0.202±0.029	0.345±0.025	0.601±0.024
CNN +Avg Pool [19]	0.771±0.064	0.913±0.014	0.978±0.004
NAN [19]	0.881±0.011	0.94±0.008	0.978±0.003
Proposed Model	0.921±0.012	0.97±0.011	0.981±0.001

The above given table shows the proposed approach achieves better performance when compared with existing techniques. We have adopted some comparative techniques from Yang et al. [19] where experimental study is extended by incorporating L2 distance measurements with CNN architecture such as CNN + Max L2, CNN + Min

Comparative Analysis of Various Face Detection and Tracking and Recognition Mechanisms using Machine and Deep Learning Methods

L2, CNN + Mean L2, and CNN + SoftMin L2 along with max and average pooling such as CNN + MaxPool, and CNN + AvePool. These techniques also achieve better performance as 0.978±0.004 but proposed aggregation module helps to reduce the noisy features resulting in improving the accuracy of the system.

3.4. YouTube Face dataset

In this section, we present the experimental analysis for YouTube face database. this data contains 3425 number of videos which are of 1595 different peoples. The number of frames in these vides varies from 48 to 6070 frames.

Method	Accuracy (%)	AUC
LM3L [10]	81.3±1.2	89.3
DDML [11]	82.3±1.5	90.1
Deep Face-single [12]	91.4±1.1	96.3
CNN + Min L2 [19]	94.96±0.79	98.5
CNN + Mean L2 [19]	95.30±0.74	98.7
CNN + Soft Min L2 [19]	95.30±0.77	98.7
CNN + Max Pool [19]	88.36±1.4	95
CNN +Avg Pool [19]	95.20±0.76	98.7
NAN [19]	95.72±0.64	98.8
Proposed Model	98.55±0.10	99.10

Prior to processing the video faces for recognition, we detect the faces, extract the features and align these features to generate the feature vector. Table 2 shows a comparative performance for video face recognition in terms of recognition accuracy and area under curve. We also consider base line methods such as CNN + Max L2, CNN + Min L2, CNN + Mean L2, CNN + Soft Min L2, CNN + Max Pool and CNN +Avg Pool. The comparative analysis shows that proposed approach achieves accuracy of 98.23% which shows a significant improvement when compared with existing techniques.

4. Conclusion

In this work, we have focused on the face detection, tracking and recognition and developed three schemes for video face recognition as novel schemes using Bayesian Learning, RCNN based technique and GoogleNet based video face recognition. The Bayesian learning follows conventional machine learning based method where trained database is classified using Bayesian classifier. The faster RCNN based model uses face detection along with the bounding box regression. Moreover, CNN based model for face recognition and bounding. Finally, we use GoogleNet based model for video based recognition.

References

1. Elrefaei, L.A., Alharthi, A., Alamoudi, H., Almutairi, S. And Al-Rammah, F., 2017, March. Real-Time Face Detection And Tracking On Mobile Phones For Criminal Detection. In *2017 2nd International Conference On Anti-Cyber Crimes (ICACC)* (Pp. 75-80). IEEE.
2. Đorić, D., Crnobrnja, S., & Punt, M. (2018, November). Implementation Of An Application For Real-Time Video Face Tracking. In *2018 26th Telecommunications Forum (TELFOR)* (Pp. 1-4). IEEE.
3. Kumar, A., Kaur, A. And Kumar, M., 2019. Face Detection Techniques: A Review. *Artificial Intelligence Review*, 52(2), Pp.927-948.
4. Rabhi, A., Sadiq, A. And Mouloudi, A., 2015, November. Face Tracking: State Of The Art. In *2015 Third World Conference On Complex Systems (WCCS)* (Pp. 1-8). IEEE.
5. D. Wang, C. Otto, And A. K. Jain. Face Search At Scale: 80 Million Gallery. Arxiv Preprint Arxiv:1507.07242, 2015
6. Chen, J.C., Ranjan, R., Kumar, A., Chen, C.H., Patel, V.M. And Chellappa, R., 2015. An End-To-End System For Unconstrained Face Verification With Deep Convolutional Neural Networks. In *Proceedings Of The IEEE International Conference On Computer Vision Workshops* (Pp. 118-126).
7. S. Sankaranarayanan, A. Alavi, C. Castillo, And R. Chellappa. Triplet Probabilistic Embedding For Face Verification And Clustering. Arxiv Preprint Arxiv:1604.05417, 2016.
8. 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.

9. Yu, Y. F., Dai, D. Q., Ren, C. X., & Huang, K. K. (2017). Discriminative Multi-Layer Illumination-Robust Feature Extraction For Face Recognition. *Pattern Recognition*, 67, 201-212.
10. 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
11. 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
12. 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.
13. 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
14. Wen, Y., Zhang, K., Li, Z. And Qiao, Y., 2016, October. A Discriminative Feature Learning Approach For Deep Face Recognition. In *European Conference On Computer Vision* (Pp. 499-515). Springer, Cham.
15. X.-T. Yuan, X. Liu, And S. Yan. Visual Classification With Multitask Joint Sparse Representation. *IEEE TIP*, 21(10):4349–4360, 2012.
16. Ortiz, E.G., Wright, A. And Shah, M., 2013. Face Recognition In Movie Trailers Via Mean Sequence Sparse Representation-Based Classification. In *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition* (Pp. 3531-3538).
17. J. Sivic, M. Everingham, And A. Zisserman. “Who Are You?” – Learning Person Specific Classifiers From Video. In *Proc. CVPR*, 2009.
18. Parkhi, O.M., Rahtu, E., Cao, Q. And Zisserman, A., 2018. Automated Video Face Labelling For Films And TV Material. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 42(4), Pp.780-792.
19. Yang, J., Ren, P., Zhang, D., Chen, D., Wen, F., Li, H. And 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).
20. Crosswhite, N., Byrne, J., Stauffer, C., Parkhi, O., Cao, Q. And Zisserman, A., 2018. Template Adaptation For Face Verification And Identification. *Image And Vision Computing*, 79, Pp.35-48.
21. 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.
22. 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.
23. 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 Technology*, 24(11):1874–1884, 2014.
24. Merikapudi, Seshaiyah And Math, Shrishail, Video Face Detection Using Bayesian Technique (May 17, 2019). Available At Ssrn: <https://Ssrn.Com/Abstract=3510956> Or [Http://Dx.Doi.Org/10.2139/Ssrn.3510956](http://Dx.Doi.Org/10.2139/Ssrn.3510956)
25. Seshaiyah M , Dr. Shrishail Math, 2019, Bsf-Rcnn-Vfr: Background Subtracted Faster Rcnm For Video Based Face Recognition, *International Journal Of Engineering Research & Technology (Ijert)* Volume 08, Issue 06 (June 2019)
26. Merikapudi. Seshaiyah, Shrishail Math, C. Nandini And Mahammed Rafi, 2021. A Googlenet Assisted Cnn Architecture Combined With Feature Attention Blocks And Gaussian Distribution For Video Face Recognition And Verification. *International Journal Of Electrical Engineering And Technology (Ijeet)*. Volume:12,Issue:1,Pages:30-42.
27. Li, L., Ge, H., Tong, Y., & Zhang, Y. (2018). Face Recognition Using Gabor-Based Feature Extraction And Feature Space Transformation Fusion Method For Single Image Per Person Problem. *Neural Processing Letters*, 47(3), 1197-1217.
28. Patil, C. M., & Ruikar, S. D. (2020). 3D-DWT And CNN Based Face Recognition With Feature Extraction Using Depth Information And Contour Map. In *Techno-Societal 2018* (Pp. 13-23). Springer, Cham.
29. Makhija, Y., & Sharma, R. S. (2019). Face Recognition: Novel Comparison Of Various Feature Extraction Techniques. In *Harmony Search And Nature Inspired Optimization Algorithms* (Pp. 1189-1198). Springer, Singapore.