Pattern Recognition Letters
journal homepage: www.elsevier.com

CrowdFaceDB: Database and Benchmarking for Face Verification in Crowd

Tejas I. Dhamecha*, Mahek Shah*, Priyanka Verma*, Mayank Vatsa**, Richa Singh

*IIIT-Delhi, New Delhi, 110020, India. * Equal contributions by student authors.*

ABSTRACT

Face recognition research has benefited from the availability of challenging face databases and benchmark results on popular databases show very high performance on single-person per image/video databases. However, in real world surveillance scenarios, the environment is unconstrained and the videos are likely to record multiple subjects within the field of view. In such crowd surveillance videos, both face detection and recognition are still considered as onerous tasks. One of the key factors for limited research in this direction is unavailability of benchmark databases. This paper presents CrowdFaceDB video face database that fills the gap in unconstrained face recognition for crowd surveillance. The two fold contributions are: (1) developing an unconstrained crowd video face database of over 250 subjects, and (2) creating a benchmark protocol and performing baseline experiments for both face detection and verification. The experimental results showcase the exigent nature of crowd surveillance and limitations of existing algorithms/systems.

1. Introduction

Research in face recognition started from cooperative individuals in constrained environment and is now attempting to address uncooperative individuals in unconstrained environment, such as surveillance. One of the key contributing factors in modern day face recognition algorithms is availability of challenging databases. Fig. 1 shows sample images from some of the most challenging video face databases currently being used for benchmarking. Most of these databases comprise videos where every frame contains one or two individuals performing certain actions in semi-controlled environments. However, the ultimate application of face recognition, i.e. surveillance in crowd, is significantly more challenging and requires the availability of databases where each frame contains multiple individuals with varying environments and actions. Current surveillance applications require face recognition algorithms to recognize face images in challenging crowd settings as shown in Fig. 2. Further, these applications require algorithms to handle variations due to low resolution, noise, pose, expressions, and illumination along with *multiple subjects in a frame*. The problem is exacerbated when both gallery and probe are captured in unconstrained conditions.

Table 1 lists the publicly available face databases used for benchmarking video based face recognition algorithms. A brief summary of these databases is presented below:

Fig. 1: Frames from some of the existing video face databases used for performance evaluation of face recognition algorithms.

1. Face-In-Action (FIA) by Goh et al. (2005): FIA database is specially created for border-security-passport-checking application and contains videos that requires user cooperation. It includes 6,470 videos covering total of 180 different subjects. However, in this dataset there is only one subject per video.
2. Honda UCSD by Lee et al. (2005): Honda UCSD dataset serves the dual purpose of face tracking as well as face recognition. The dataset has been created in constrained

**Corresponding author: Tel.: +91-9654653404; fax: +91-11-26907410;
e-mail: mayank@iiitd.ac.in (Mayank Vatsa)

Fig. 2: A law enforcement application scenario where subjects are matched using surveillance footage only. Top row of the figure shows four frames/images from the Boston bombing case. The suspects (the subject in black hat and the subject in white hat) can be seen walking along with other subjects. The bottom row show the face regions of the suspects.

Table 1: A summary of publicly available video face databases.

Database	No. of Subjects	No. of Videos
Honda UCSD	35	105
YouTube Faces	1,595	3,425
PaSC (handheld)	265	2,802
PaSC (control)	265	2,802
Face in Action	180	6,470
Chokepoint	54	48
SN-Flip	190	28
IJB-A	2,085	500
IJB-B	1,845	7,011
CrowdFaceDB	**257**	**385**

manner and with user acquaintance.

3. ChokePoint by Wong et al. (2011): ChokePoint Wong et al. (2011) database is designed to deal with person identification/verification under real-world surveillance conditions using prevailing technologies. It has 48 videos pertaining to 54 subjects.
4. YouTube Faces (YTF) by Wolf et al. (2011): The YTF dataset has been created with the primary purpose of studying face recognition in unconstrained environment. The YTF dataset contains 3,425 videos of 1,595 subjects. The dataset is composed of celebrity videos which are collected from YouTube with the constraint of only one subject present in a video.
5. Point and Shoot Challenge (PaSC) by Beveridge et al. (2013): The PaSC database contains videos captured using hand-held and high definition devices. PaSC dataset encompasses 2,802 videos of 265 subjects.
6. SN-Flip by Barr et al. (2014): SN-Flip database was created with the requirement of having multiple subjects in one video sequence. It includes 28 videos of 190 subjects.
7. IAPRA Janus Benchmark Datasets, IJB-A (Klare et al., 2015) and IJB-B (Whitelam et al., 2017) include face and non-face images and videos to facilitate face detection and recognition challenge. The IJB-B, which is superset of IJB-A, includes 7,011 videos of 1,845 subjects.

Fig. 3: Challenges for face detection and recognition in crowd videos.

The extent of challenges present in some of these databases such as Honda UCSD have been addressed and 100% accuracy has been achieved. On the other hand, challenging databases such as YouTube (Wolf et al., 2011) and Point and Shoot Challenge (Beveridge et al., 2013) are used to enhance the capabilities of modern algorithms. However, these databases do not help us understand the performance of current face recognition algorithms in unconstrained videos of crowd, i.e., two or more subjects in each video.

It is our assertion that there is a significant scope of improving the capabilities of face recognition performance in unconstrained environment, especially the crowd scenarios. Fig. 3 shows some of the challenges of crowd videos which make face detection and recognition tasks difficult. Specifically,

- in unconstrained environment, it is not easy to get acceptable level of performance in face detection due to variations in illumination, pose, and occlusion
- in low quality videos, it is difficult to differentiate between the face and the background; therefore, both face detection and recognition are challenging
- activities performed by different individuals in a video can lead to occluded face images. Such occlusions make detection and recognition tasks difficult for automated processing, and
- at times, it is possible that the subject is at a distance from the sensor and therefore, face area can be small. Such variations make detection as well as recognition very difficult.

To promote face detection and recognition (verification) research in challenging crowd scenarios, this paper presents CrowdFaceDB: an unconstrained video face database. The key contributions of this research are:

1. CrowdFaceDB dataset that includes total of 385 crowd videos pertaining to 257 subjects. The database includes manually annotated facial landmark points for every frame

Table 2: Details of the CrowdFaceDB.

Device (Resolution)	No. of Videos	No. of Frames	No. of Subjects	No. of Subjects/Video			No. of faces
				Min	Max	Avg	groundtruth
Device I (640×480)	157	21,683	176	1	14	2.5	27,747
Device II (2304×1296)	152	20,008	166	1	10	2.1	23,505
Device III (1920×1080)	76	8,461	106	1	8	2	10,557
Total	385	50,152	257	1	14	2.2	61,809

which has one or more face images in it. Along with the videos and landmark points, a set of protocols and end-to-end MATLAB software package are designed to evaluate the performance of face verification algorithms on this dataset.
2. Face detection baseline is provided by comparing the results of manual annotation and four publicly available codes: 1) Viola Jones (Viola and Jones, 2004) face detector (MATLAB open source), 2) HOG descriptor based C++ open source library dlib (King, 2009), 3) face detection aided by fiducial points (Everingham et al., 2009), and 4) Faster R-CNN (Ren et al., 2015; Ruiz and Rehg, 2017).
3. To establish the face verification baseline, results are reported with OpenBR (Klontz et al., 2013), VGG-Face (Parkhi et al., 2015), and a commercial-off-the-shelf system, FaceVacs.

2. CrowdFaceDB for Face Verification in Crowd

The proposed CrowdFaceDB dataset contains 385 videos (50,152 frames) of 257 subjects, captured at different locations and each video contains up to 14 subjects[1]. The videos are recorded using handheld devices without mounting on any tripod or similar structure. Consent for collecting these videos is taken from all the subjects. Fig. 4 shows samples from the database and dataset statistics are summarized in Table 2. Typically, subjects appear in groups in all the videos and therefore, almost all the video frames contain more than one subject. Since the grouping of volunteers is not restricted, the number of videos per person varies and this information is summarized in Table 3.

Further, Table. 1 shows the comparison of the proposed CrowdFaceDB with existing video face databases. This figure illustrates the unique property of CrowdFaceDB which is suitable for benchmarking face detection and verification algorithms for crowd scenario. We next present the details of the proposed CrowdFaceDB.

2.1. Device Details

The database is collected with the help three portable handheld devices with different resolutions:

1. Nikon Coolpix S570: resolution 640×480
2. Sony handycam DCR-DVD910E: resolution 2304×1296
3. iPhone (4s and 5c) resolution 1920×1080

[1] A significantly smaller version of this database was presented in IAPR ICB 2015 (Dhamecha et al. (2015)).

Table 3: In the CrowdFaceDB dataset, a subject can appear in multiple videos. This form of repetition allows to observe a subject under different crowd settings. For example there are 29 subjects appearing in exactly four videos.

No. of video (x)	1	2	3	4	≥ 5
No. of subjects (y)	77	52	38	29	61

y number of subjects appear in x number of videos.

These devices are referred to as Device I, Device II and Device III, respectively. The different devices lead to varying quality of captured videos. The selection of these devices also introduces cross-sensor and cross-resolution covariates in the database. Hence, one can use this dataset for cross resolution face matching problem.

2.2. Ground-truth Annotation: Detection and Identity

With unconstrained databases, it is very important to provide ground-truth annotations. Therefore, manual annotation has been performed for every visible face present in the frames of every video. Four POIs *(Points of Interest)* of frontal faces are annotated: Centers of eyes, nose tip, and center of lips, along with subject IDs of that face. This procedure is followed for each frame of each video. If any of the points are not visible due to occlusion or pose variation, the remaining points are annotated. The face bounding box is obtained by fitting the POIs on a canonical face frame. This procedure provides manually detected faces of 125×160 size from all frames of all videos. As mentioned in Table 2, a total of 61,809 face images are detected from 50,152 frames in 385 videos. The images are detected and normalized according to the three point registration. Along with the detected faces, the identity of the subjects is also annotated as

DeviceName_VideoID_FrameNo_SubjectID.jpg

Each registered output face image is named using the following convention and will be provided in the /Cropped/DeviceName/VideoID directory for easier access. A sample directory structure is shown in Fig. 5.

3. Evaluation Protocol and Package

In order to use the database and evaluate the performance of face detection and recognition (verification) algorithms, we have created independent evaluation packages for both of them. The package is designed to make the overall evaluation process convenient and user friendly. The database along with the

Fig. 4: Sample frames from the CrowdFaceDB: multiple subjects appear together in each video along with subjects appearing in unconstrained environment. The videos are captured using three different devices with different resolutions. The videos are captured while subjects are walking through a passage or passing through doors.

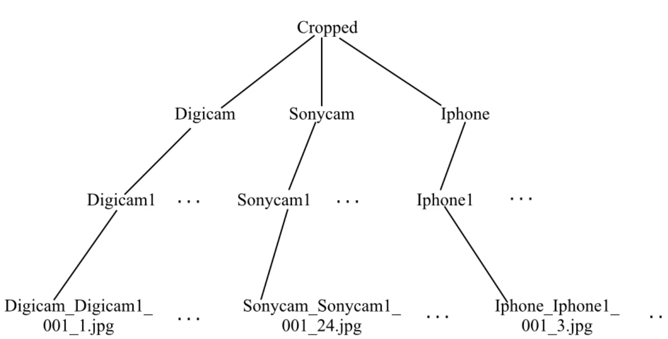

Fig. 5: Directory structure of the cropped face images provided as part of dataset package.

evaluation package will be made available to the research community at `http://iab-rubric.org/resources.html`.

3.1. *Face Detection Evaluation Package*

The evaluation package for face detection consists of end-to-end MATLAB code to evaluate the accuracy of face detection algorithms. To use this code, the user has to provide the file name and rectangular coordinates of the detected face in a file to the package. The evaluation code provides the accuracy for the corresponding face detection algorithm.

The performance of face detection algorithms is evaluated by computing the overlap of the automatically detected face bounding box with the one obtained using manual annotation. For each algorithm the percentage of intersection computed using the following steps.

- Obtain the coordinates of the rectangle for manual annotation and automatic detection.
- Find the intersection of the detected region with ground truth region, and union of both the regions, and calculate the intersection area percentage as follows:

$$Overlap = \frac{Ground\ truth\ rectangle\ \cap Detected\ rectangle}{Ground\ truth\ rectangle\ \cup Detected\ rectangle} \tag{1}$$

If the intersection is empty, then it is not considered as detected.

3.2. Face Verification Evaluation Package and Protocol

Similar to face detection, the accuracy of face verification algorithms can be evaluated using the package provided with the database. The experimental protocol for training and testing is also embedded in the evaluation package.

- 10 disjoint partitions of the training and testing sets have been created such that the training sets contain an average of 100 videos with around 115 subjects while the testing sets contain an average of 240 videos with around 135 subjects.
- Each testing set is further divided into gallery and probe sets without any overlapping subject IDs between the train and test sets.
- The details of the video IDs and subject IDs of each set are provided in the evaluation package.
- MATLAB code for end-to-end evaluation has been created and is provided in the package. This code requires score matrices for each testing set as input and performs all necessary calculations and provides a combined receiver operating characteristics (ROC) curve which is used for comparing the results.

CrowdFaceDB dataset focuses on unconstrained face recognition with multiple subjects in a video or frame. In this scenario, subjects from one video are matched with the subjects of another video. The gallery set is defined in terms of a *set of videos*. With respect to real world applications, there can be three types of evaluation settings: frame-to-frame matching, video-to-frame or frame-to-video matching, and video-to-video matching.

1. Frame-to-Frame matching: In this evaluation setting, every frame of gallery set is matched against every frame of probe set to get score matrix.
2. Video-to-Frame Matching: In this evaluation setting, face frame of probe set is compared against every video in the gallery set. A set of scores is obtained by comparing each probe face image with all the face frames in gallery video.
3. Video-to-Video Matching: In this evaluation technique, a set of probe videos is compared with the set of gallery videos. Each probe face image is compared with all the face images in a gallery video and for every match pair, scores are aggregated.

4. Face Detection and Recognition: Baseline Results

Along with the manually annotated ground-truth of both detection and recognition, baseline evaluation with automatic algorithms is also performed on the CrowdFaceDB dataset. The following existing face detection algorithms are used:

- Viola Jones Face Detector (Viola and Jones, 2004)
- Faster R-CNN face detector (Ren et al., 2015; Ruiz and Rehg, 2017)
- Face detection aided by fiducial points (Everingham et al., 2009)
- Face detection based on Histogram of Oriented Gradient (HOG) (King, 2009)

For establishing the baseline face verification performance, the results of three matchers are computed.

- Open Source Biometrics Recognition (OpenBR) (Klontz et al., 2013)
- VGG-Face (Parkhi et al., 2015)
- FaceVacs (Commercial System)

4.1. Face Detection Results

Each of the four automatic detectors provides a rectangle around the detected face region. Viola Jones (Viola and Jones, 2004), face detection aided by fiducial points (Everingham et al., 2009), HOG descriptor based detector from DLib (King, 2009) and Faster R-CNN, respectively, predicted that there are ~56K, ~47K, ~37K, and ~58K faces in the dataset. Note that, there are ~62K ground-truth faces, as mentioned in Table 2. Based on the detected face region and the ground-truth, *overlap* is computed using Eq. 1. The results are organized according to the percentage of overlap between the detected and manually annotated regions. Table 4 shows the true positive rates (fraction of actual face regions that overlap with detected faces) as a function of overlap along with the number of false positives (non-face regions detected as face). Ideally, for all overlap thresholds, TPR should remain 100%, i.e. all faces are detected with high overlap, and FP should be zero, i.e. no non-faces are detected as face. Among the four detectors, Faster R-CNN detector detects maximum number of faces. However, not all the faces have high overlap with ground-truth annotations. The performance of all the algorithms is very much dependent on the overlap threshold. With increasing threshold, the percentage of correctly detected faces decreases while the number of false positives increases. For instance, in case of Viola Jones detector, ~60% of ground-truth faces were detected with overlap greater than 25% but only 2.03% of them have more than 75% overlap with a false positive of over 54K. Further, for lower overlaps, all the four approaches have over 50% TPR but at threshold higher than 50% only Faster R-CNN detector exhibit relatively better detection performance with over 60% of TPR and lowest FPs. As mentioned by Jain and Learned-Miller (2010), receiver operating characteristics (ROC) curve can provide details of the strength of a face detection approach over the entire range. Fig. 6 shows the receiver operating curve for the best performing face detector (i.e. Faster R-CNN) on the proposed database at overlap ¿ 50

Fig. 7 shows sample detection results of all four algorithms on few frame. In our experiments, we have observed that for all four detectors, a large number of detected faces has less than 10% overlap with ground-truth and a few number of faces has more than 80% overlap. For instance, HOG feature based open source library detects approximately 10,000 faces, for which the percentage of overlap is 0-10%. Similarly, face detection

Table 4: Summarizing the detection results of the three automatic face detection algorithms on the CrowdFaceDB database. Reported are the True Positive Rates (TPR) and the number of False Positives (FP) at different thresholds of overlap.

	Overlap > 0%		Overlap > 25%		Overlap > 50%		Overlap > 75%	
Algorithm	TPR (%)	FP	TPR (%)	FP	TPR (%)	FP	TPR (%)	FP
Viola Jones	62.86	16,902	60.09	18,615	40.83	30,523	2.03	54,504
Faster R-CNN	79.81	8,854	77.82	9,943	70.47	14,458	61.78	19,753
Face detection aided with fiducial points	55.48	13,430	51.73	15,746	50.61	16,439	39.32	23,416
HOG-based	57.03	1,787	55.76	2,572	21.42	23,800	0.35	36,820

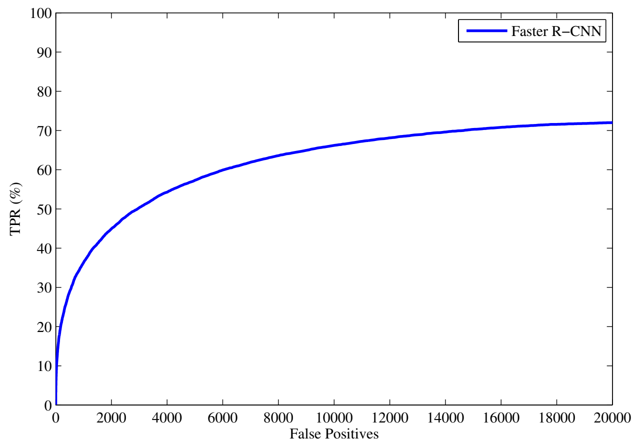

Fig. 6: ROC curve for the Faster R-CNN algorithm with overlap > 50% (this is similar to discrete score based ROC in FDDB protocol (Jain and Learned-Miller, 2010)).

with fiducial points and Viola Jones algorithms detect 11,000 and 27,500 faces respectively, with 0 to 10% overlap. Fig. 8 shows samples of inaccurate face detection due to the presence of pose and occlusion, and some samples of false detection as well. With these analysis, it can be concluded that in crowd or surveillance scenario, both false detection and false rejection are challenging. Considering detection as the first step in face recognition (verification) pipeline, it is important that the faces are detected with high overlap with ground-truth. Considering overlap as the measure, the performance of baseline detectors is very low and it is very important to improve state-of-the-art of face detection.

4.2. Baseline Results for Face Verification

Face verification evaluations on the CrowdFaceDB dataset have been performed using Open Source Biometrics Recognition (OpenBR) by Klontz et al. (2013), VGG-Face by Parkhi et al. (2015) and a commercial-off-the-shelf system, FaceVacs, with the pre-defined evaluation protocols explained in the previous section. The verification performance is reported in terms of receiver operating characteristic curve. The ROC curves obtained for the gallery-probe sets across cross validation trials are combined into one curve using vertical averaging. ROC curves for each of the system for manually annotated faces are also reported. The key results are summarized in Tables 5 and 6, Fig. 10 and Fig. 11. Verification is performance is reported in

Table 5: Face verification performance of FaceVacs, OpenBR, and VGG-Face in terms of GAR with manual face detection. Best GAR across four protocols settings is reported for each FAR.

FAR	**FaceVacs**	**OpenBR**	**VGG-Face**
0.001	0.023	0.000	0.007
0.01	0.059	0.010	0.068
0.1	0.192	0.110	0.366

terms of Genuine Accept Rate (GAR) as function of False Accept Rate (FAR). Note that, ROC curves are drawn with FAR on log-scale to emphasize on performance at low FAR.

- VGG-Face appears to outperform FaceVacs and OpenBR. At lower FAR, FaceVacs yields better GAR than OpenBR; however, at higher FARs, the performance difference is not significant. As shown in Table 6, at 0.01 FAR, the best verification rate of (only) 0.12 is achieved. This poor performance indicates the complexity of the problem (and dataset) as well as limitations of current systems.

- Score aggregation for video-to-video and video-to-frame matching is performed using two strategies: mean and max. Since both the systems provide similarity scores, the max strategy translates to selecting the scores corresponding to the best match. Both the systems suffer significantly in video-to-video matching using mean aggregation strategy and the best performance is observed with video-to-video matching with max aggregation strategy. This result underlines the importance of frame selection (Goswami et al., 2014, 2017).

- The face verification accuracies with manual face detection are lower than automatic face detection. This is due to the fact that the number of faces detected by automatic algorithms is different compared to manual, automatic algorithms able to detect significantly smaller number of faces. It is observed that automatic algorithms detect only frontal and near frontal faces whereas in manual annotation, all the faces are detected and many of which are unrecognized by the algorithms. Therefore, the results with manual detection are lower than automatic detection.

5. Usage for CrowdFaceDB Dataset

The proposed CrowdFaceDB dataset can be used for evaluating the performance of face detection, face recognition (veri-

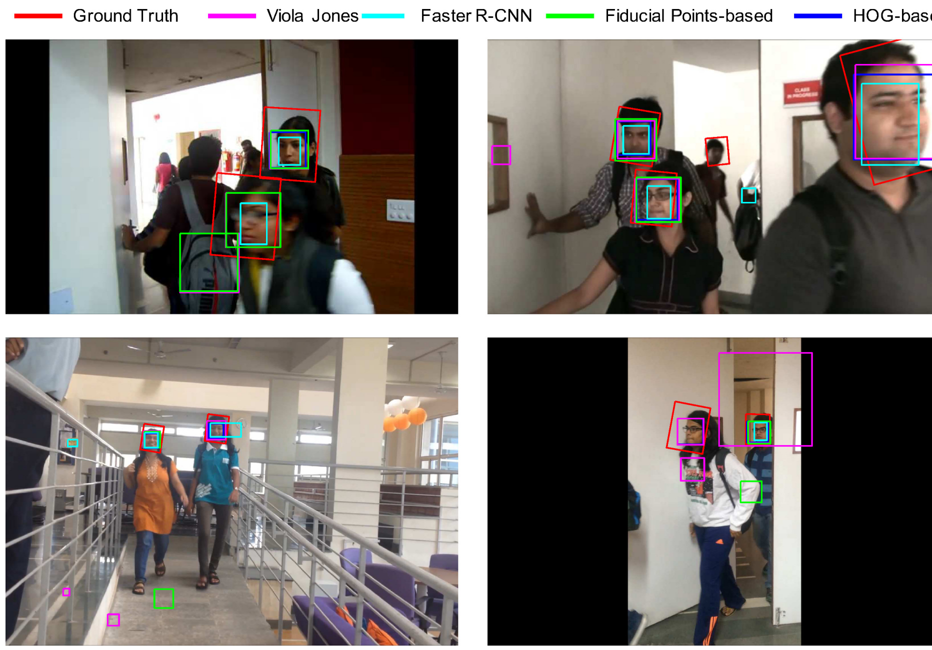

Fig. 7: Face detection bounding box of the ground-truth and automatic algorithms.

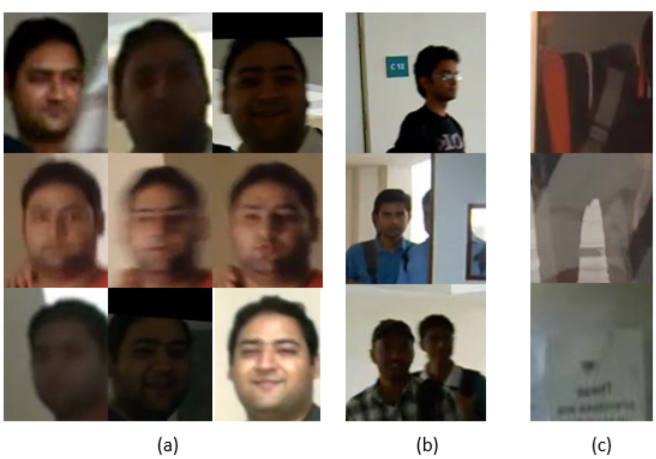

Fig. 8: (a) Examples of accurately detected faces corresponding to each of the three devices. (b) Examples of inaccurate face detection due to partial face and presence of extra non-face/background regions, and (c) sample images demonstrating falsely detected faces which are discarded based on POI annotations.

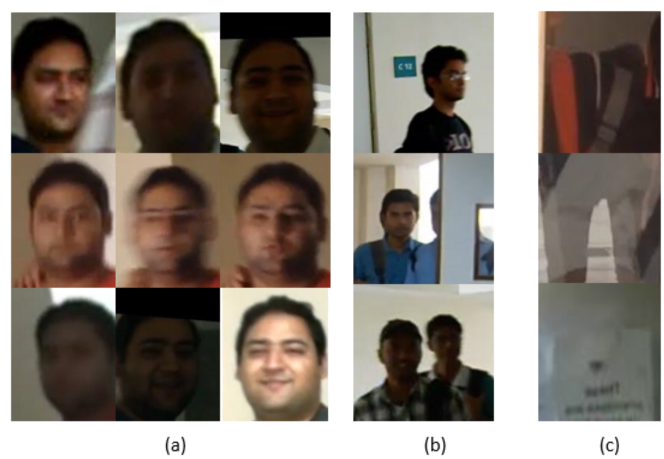

Fig. 9: (a) Examples of accurately detected faces corresponding to each of the three devices. (b) Examples of inaccurate face detection due to partial face and presence of extra non-face/background regions, and (c) sample images demonstrating falsely detected faces which are discarded based on POI annotations.

Table 6: Face verification performance of FaceVacs, OpenBR, and VGG-Face in terms of GAR with automatic (in-built) face detection. Best GAR across four protocols settings is reported for each FAR.

FAR	FaceVacs	OpenBR	VGG-Face
0.001	0.027	0.000	0.027
0.01	0.088	0.016	0.122
0.1	0.316	0.200	0.409

fication and identification), and re-identification algorithms.

- **Face Detection**: With the availability of manual annotations, the database can be used for evaluating the performance of face detection algorithms in crowd.
- **Face Verification and Identification**: With predefined training-testing splits and protocols, the dataset can be useful for evaluating the frame-to-video, video-to-video, and frame-to-frame comparison performances.
- **Person Re-identification**: Since the database is prepared in multiple sessions and using multiple cameras, it can also be potentially used for evaluating person re-identification algorithms.

Though we have provided the evaluation protocols for face detection and face verification, we solicit support from other researchers to build protocols and baseline evaluation for face identification and person re-identification.

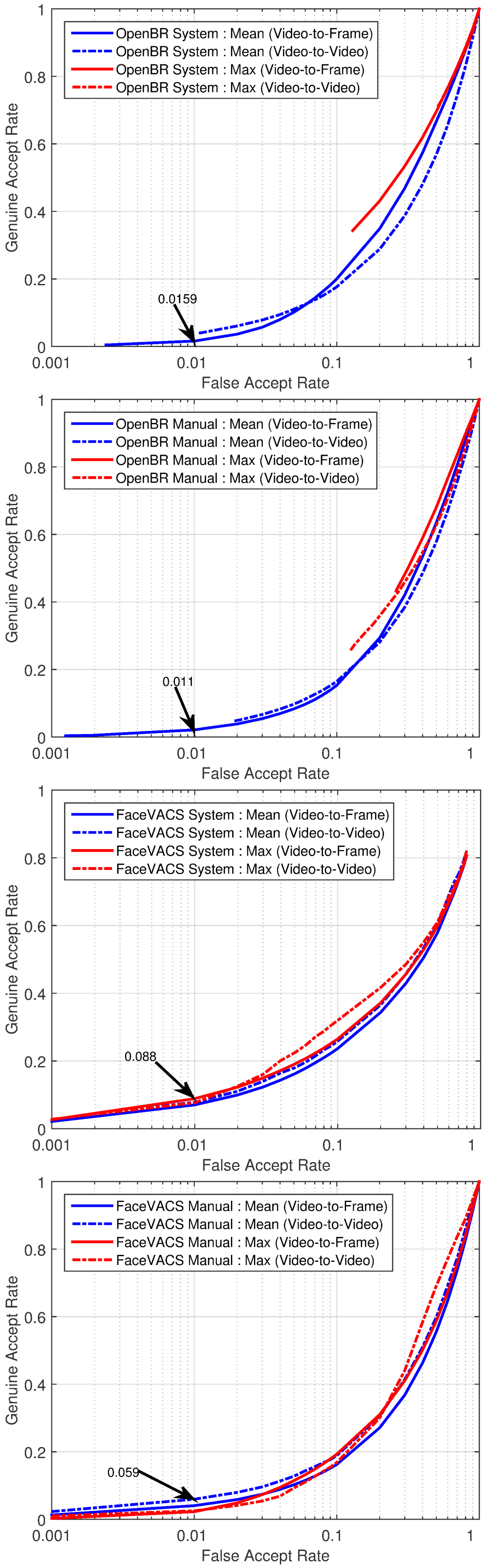

Fig. 10: Baseline results for face verification using OpenBR and FaceVacs: results are reported with in-built automatic face detection and manually annotated faces.

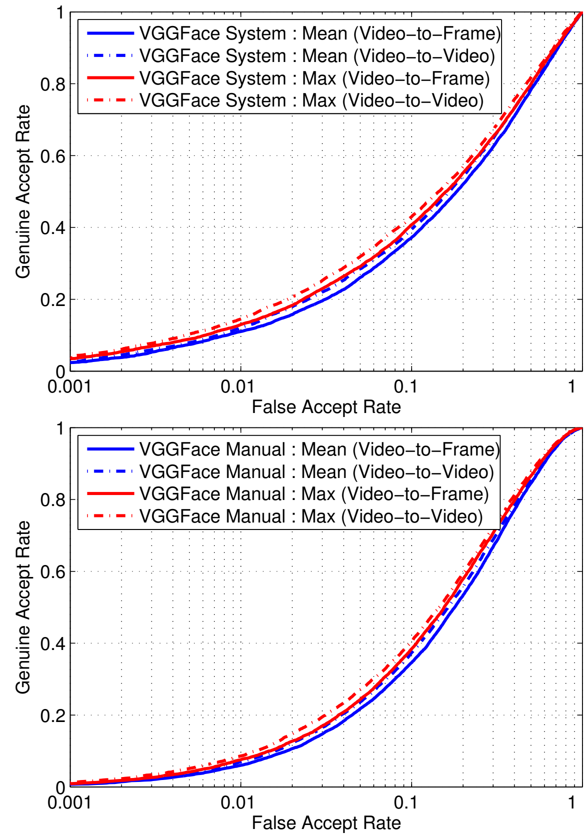

Fig. 11: Baseline results for face verification using VGGFace: results are reported with in-built automatic face detection and manually annotated faces.

6. Conclusion

Face recognition from video in unconstrained environment has attracted a lot of research interest due to its various applications. Multiple frames in a video provide temporal and intra-class variations that can be leveraged for efficient face recognition. While existing research has primarily focused on single-person-per-video, one of the key applications of face recognition is in crowd surveillance where multiple subjects appear in same frame/video. In order to instigate research in this arduous application scenario, in this paper, we propose a new dataset termed as CrowdFaceDB and present baseline experiments with existing systems. Baseline results for face detection and verification show that popular commercial and open source systems do not perform well on crowd videos in uncontrolled settings. We assert that the availability of the proposed database can help in advancing the state-of-art for both face detection and recognition in surveillance applications.

7. Acknowledgement

This research is supported through a grant from Ministry of Electronics and Information Technology, Government of India. T.I. Dhamecha was also partially supported through TCS PhD Fellowship. M. Vatsa and R. Singh are also partially supported by Infosys Center for Artificial Intelligence, IIIT-Delhi, India.

References

Barr, J.R., Cament, L.A., Bowyer, K.W., Flynn, P.J., 2014. Active clustering with ensembles for social structure extraction, in: IEEE Winter Conference on Applications of Computer Vision, pp. 969–976.

Beveridge, J.R., Phillips, P.J., Bolme, D.S., Draper, B.A., Given, G.H., Lui, Y.M., Teli, M.N., Zhang, H., Scruggs, W.T., Bowyer, K.W., et al., 2013. The challenge of face recognition from digital point-and-shoot cameras, in: IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1–8.

Dhamecha, T.I., Verma, P., Shah, M., Singh, R., Vatsa, M., 2015. Annotated crowd video face database, in: 2015 International Conference on Biometrics, pp. 106–112.

Everingham, M., Sivic, J., Zisserman, A., 2009. Taking the bite out of automated naming of characters in TV video. Journal of Image and Vision Computing 27, 545–559.

Goh, R., Liu, L., Liu, X., Chen, T., 2005. The CMU face in action (FIA) database, in: Analysis and Modelling of Faces and Gestures. Springer, pp. 255–263.

Goswami, G., Bhardwaj, R., Singh, R., Vatsa, M., 2014. MDLFace : Memorability augmented deep learning for video face recognition, in: IEEE/IAPR International Joint Conference on Biometrics, pp. 1–7.

Goswami, G., Vatsa, M., Singh, R., 2017. Face verification via learned representation on feature-rich video frames. IEEE Transactions on Information Forensics and Security 12, 1686–1698.

Jain, V., Learned-Miller, E., 2010. FDDB: A Benchmark for Face Detection in Unconstrained Settings. Technical Report UM-CS-2010-009. University of Massachusetts, Amherst.

King, D.E., 2009. Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research 10, 1755–1758.

Klare, B.F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., Grother, P., Mah, A., Jain, A.K., 2015. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1931–1939.

Klontz, J.C., Klare, B.F., Klum, S., Jain, A.K., Burge, M.J., 2013. Open source biometric recognition, in: IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1–8.

Lee, K.C., Ho, J., Yang, M.H., Kriegman, D., 2005. Visual tracking and recognition using probabilistic appearance manifolds. Journal of Computer Vision Image Understanding 99, 303–331.

Parkhi, O.M., Vedaldi, A., Zisserman, A., 2015. Deep face recognition, in: British Machine Vision Conference.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, in: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, pp. 91–99.

Ruiz, N., Rehg, J.M., 2017. Dockerface: an easy to install and use Faster R-CNN face detector in a Docker container. ArXiv .

Viola, P., Jones, M.J., 2004. Robust real-time face detection. International Journal of Computer Vision 57, 137–154.

Whitelam, C., Taborsky, E., Blanton, A., Maze, B., Adams, J., Miller, T., Kalka, N., Jain, A.K., Duncan, J.A., Allen, K., et al., 2017. IARPA Janus benchmark-b face dataset, in: IEEE CVPR Workshop on Biometrics.

Wolf, L., Hassner, T., Maoz, I., 2011. Face recognition in unconstrained videos with matched background similarity, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 529–534.

Wong, Y., Chen, S., Mau, S., Sanderson, C., Lovell, B.C., 2011. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 81–88.