Deep-Facial Feature-Based Person Re-identification for Authentication in Surveillance Applications

Pranjal Borse[1,], Aishwarya Gaikwad[2,**], Nachiket Dabhade[3,***], Harsha Saxena[4,****]*
[1]Department of Computer Engineering, Ramrao Adik Institute Technology, India
[2]Department of Computer Engineering, Ramrao Adik Institute Technology, India
[3]Department of Computer Engineering, Ramrao Adik Institute Technology, India
[4]Department of Computer Engineering, Ramrao Adik Institute Technology, India

Abstract. Nowadays, a large network of cameras is predominantly used in public places which provide enormous video data. These data are monitored manually and may be utilized only when the need arises to ascertain the facts. Automating the system can improve the quality of surveillance and be useful for high-level surveillance tasks like person identification, suspicious activity detection or undesirable event prediction for timely alerts. In this paper, we proposed a model that can Re-identify a person from a single camera tracking environment. This system will automatically extract face features of the person and generate the Unique Id for each person when it enters for the first time in the monitored area. Its face features are stored in the database which will help to Re-identify the person whenever the same person appears again. The challenges faced by the system are occlusion, pose, light conditions, and face orientation. The proposed system highlights, effect of different deep neural networks for Person Re-identification and compares based on the accuracy, GPU usage, Speed, Number of faces detected by overcoming the challenges like illumination and occlusion. The advantage of the system is it doesn't require the database of people in advance for recognition and it will be helpful for criminal identification for crime control and prevention.

Keywords: Person Re-identification, Face identification, CNN, HOG, MTCNN

1 INTRODUCION

Deep Facial Feature-Based re-identification is the technique of recognizing the person based on his Id which was assigned based on his previous appearance. Our human brain can recognize the unknown person which was seen previously. Similarly, we can develop a system which can recognize a person with his unique id based on his previous appearance in camera automatically. Humans can recognize strangers by face, height, body structure, skin color and hair color. Out of all people's faces, the most reliable feature for recognition. Earlier re-identification was not possible due to Real-Time challenges like variation in viewpoints, illumination conditions, Occlusion, Crowded environment, and different camera views. Few of the challenges are resolved due to immense growth in technology like capturing video in low illumination.

Person Re-identification came into the scene due to the incidents where technology couldn't help. The following three main incidents that lead to research on this topic are: First one is the London's Subway bomb blast on July 7, 2005. After this incident investigators had to parse the entire city's CCTV footage and yet there was no outcome. Second one was the Boston Marathon bombing on April 15th, 2013 and the Third one at Bangalore, a brutal attack in November of 2013 on an ATM. With the increase of crime rate, it is important to catch criminals and send them behind bars. Our security officers are struggling to catch all of them, but finding a few criminals in a huge crowd is not feasible. Security and crime control concerns are the motivating factors for the deployment of intelligent video surveillance systems.

The Proposed system has been intended to overcome the drawbacks of the past surveillance systems and to enhance security, adaptability, and efficiency. The main aim of this research is to automatically detect and Reidentify the person in public places. The model is designed for matching face structure with either existing one or set as new data. Initially there will be no data and then the system stores data as the new face is targeted. The system will have to work to detect human movements and recognize face efficiently, even in low lighting, all-natural environments. The system will reduce the cost of making the database and create a Real-Time database in public places automatically. It also helps to detect suspicious people and notify the Admin/Police as soon as the person is detected.

The paper's organizing is done as follows: Section 2 discusses the currently present techniques, analysis of previous research associated with our proposed methodology. The associated research is defined as a base for our approach. Section 3 presents models of computation and theoretical tools we adopted to Reidentify persons. Experimental results and also the comparative analysis of varied algorithms is shown in Section 4. Section 5 has Conclusions of the outcomes obtained by using our work.

*e-mail: pranjalborse158@gmail.com
**e-mail: aishwaryagaikwad21@gmail.com
***e-mail: nachiketsdabhade2@gmail.com
****e-mail: harsha.saxena@rait.ac.in

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

2 LITERATURE SURVEY

Many existing methods of Person Re-identification construct robust features that are, at the same time distinctive, and as robust as possible to describe a person's appearance in a variety of settings, some extensive survey and research s have been developed since then.we briefly introduce some research in the context of the proposed model done in person re-ID history.

To our knowlage the first work done in the feid of Multi-camera tracking with explicit "Re-identification" was done by W. Zajdel, Z. Zivkovic[1] in 2005,The method is designed as a unique, hidden, label - for-person dynamic Bayesian network defined to encode probabilistic relationships between symbols and signs(colors and spacetime) signals from tracklets. The user's person ID, which is determined by the reverse side of the signs, distributions are calculated using a different Bayesian deletion algorithm.

Many surveys[2][3] have been done where the features of a person are extracted either by using hand-crafted methods or the Deep neural networks methods in both image-based and video-based Re-Identification.

The handcrafted features technique[4] like Generalized Maximum Multi Clique optimization ,Local Maximal Occurrence Representation(LOMO) , Iterative Hankel Total Least Squares(IHTLS) provide less accuracy than the deep neural network proposed by D. Yi[5].This is the first work to apply deep learning in Re-id based on image classification.They use a Siamese neural network to determine whether a pair of login images belong to the same user identity. The reason for choosing the Siamese cat model is that the number of training samples for each of the personalities is limited (usually there are two). With the exception of some parameter variations, the main differences are that it adds that extra cost to network operation when a thinner partition is used at the same time. Experimental data that do not match each other, and the two methods do not directly compare. Even if the performance is not stable, even with small amounts of data, deep learning methods have become a very popular option for re-ID. Bai, Xiang[6] proposed part-based feature representation to fearn pedestrian feature.The pedestrian is described as a succession of body components from head to foot using Long Short-Term Memory (LSTM) in an end-to-end manner.They achieve a 90.84%mAP on Market-1501 which is a image based dataset.The image-based dataset didnt provide better accuracy when it is used in Real-Time environment.

Video-based Re-ID is suitable for the analysis of video in real world as it increase the accuracy of the system.Kasturi R, Ekambaram R.[7] focuses on the recent work done in the Re-identification research area using images and videos.They also perform Re-identification on videos and discuss the challenges faced by the system such as Camera calibration,non-overlapping views. E. Ristani and C. Tomasi[8] learnt features using CNN for Multi-Target Multi-Camera Tracking and Re-identification on Market-1501 and DukeMTMC-ReID dataset and achieved the training accuracy of 63.27% and 74.81% respectively.

The previous papers used the body features for Re-identification and acquired less accuracy. P. Li, M. L. Prieto[9] used face features for identification in real world by using DNN architectures and patch matching techniques. They achieve the Testing accuracy of 78.5% using Siameses network and overcome the challenge of low resolution. V. Mathew, T. Toby[10] also work on the face feature extracted from CNN model for Re-identification in Real time videos. They used Viola Jones algorithm and MTCNN model to detect face and train these models on ChokePoint dataset. The drawback of this paper is they tested this model on very few videos.

We have also done research on Generic Object Detection[11]and face detection techniques[12][13][14] using Deep learning which help to identify multiple person faces in a video-based data. Based on the research we concluded to compare the performance of three well known methods Histogram of Oriented gradients(HOG),Convolution Neural Network (CNN) and MTCNN with FaceNet for face identification and SVM model for recognition using faces embeddings

3 PROPOSED METHODOLOGY

The proposed methodology focuses on developing an efficient system to Reidentify the person in public places as shown in Figure 1. System takes video as the input and extracts frames to detect multiple people in the environment. By applying Various Deep learning Architecture like CNN, HOG and MTCNN, systems extract features from faces. Then, it compares facial features with known faces in the database. Initially, At the start of the system the database is empty. The face that appears first will be provided with Unique Id 0. After this the next face will be compared with the faces present in the database. If the person was recognized, which means it appears gain in the monitored area system will display its old id. If the person is not matched with the existing one, the system will store the new face and assign a new Unique Id to it. In the database file face features are stored along with its Unique ID as shown in Figure 1.

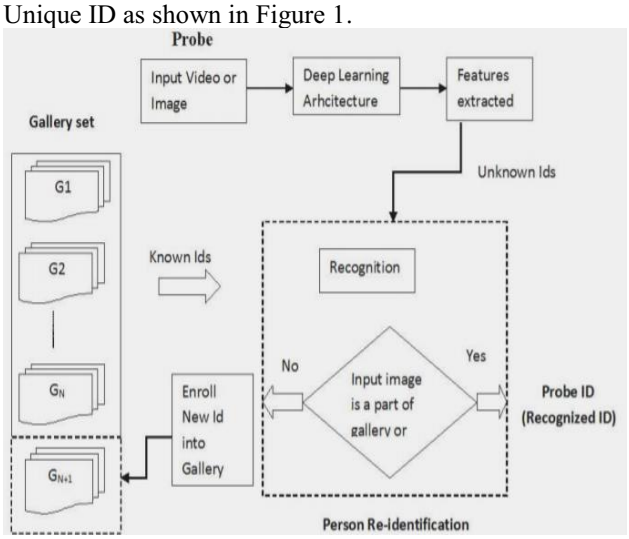


Figure 1: System Design

We have compared the working of three deep learning methods based on various parameters such as performance measure, accuracy, processing time, speed, GPU usage, side face detection and illumination conditions in Real-Time environment, which are as follows:

Method-1 Convolution Neural Network :

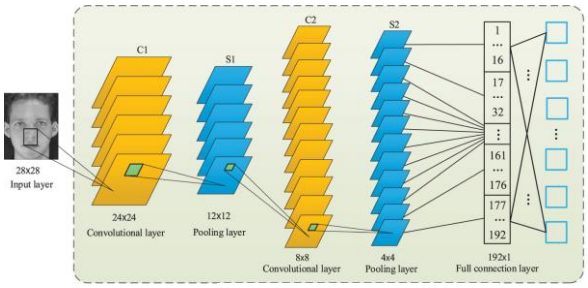


Figure 2 : Layers of CNN

CNNs work on multilayer perceptrons. Multilayer perceptrons are nothing but fully connected networks, in which each neuron in one layer is connected to any or all neurons within the next layer. CNN works in different types of layers. During a regular Neural Network there are various varieties of layers as shown in figure 2.

The model's structure is similar to that of the classic LeNet-5 model, however some parameters, such as the data file, network dimension, and full affiliation layer, are completely different. Two convolutional layers (C1 and C2), as well as two pooling layers, make up the developed CNN (S1 and S2). As seen in Figure 2, these layers are structured alternatively in the C1-S1-C2-S2 style. Within the input layer, there is only one feature map that is used to feed the normalised facial picture into the CNN model. C1 is the first convolutional layer, which has six feature maps and convolutes each somatic cell with a 5 to 5-size randomly produced convolution kernel. S1 is that the output of the previous layer was supported by the initial pooling layer, whose output is six calculated feature maps. Every component in the feature map is linked to the mean convolution kernel of the associated feature map in the C1 layer, ensuring that the weather's receptive fields do not overlap. The C2 and S2 units, respectively, are the second convolutional layer and the pooling layer, with twelve feature maps and similar calculation procedures as their predecessors. In addition, between the S2 layer and the output layer is a fully connected single-layer perceptron. The final word output might potentially be a 40-dimensional vector for face recognition of N people, as shown in Figure 2, when the sigmoid process is used for multi-label classification.

Method-2 Histogram of Oriented Gradient:

Histogram of Oriented Gradients (HOG) may be a feature definition that wants to process a picture, especially to detect an object. The fundamental premise of the Histogram of Oriented Gradients will then enter the way we calculate histograms and the way these feature vectors, found within the HOG dictionary, are employed by the editor as SVM to see the article involved as shown in figure 3.

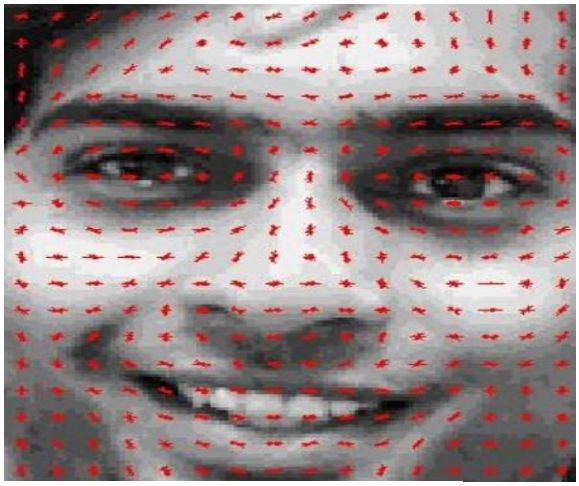


Figure.3 HOG Feature sample face

The HOG calculates image gradients per block.The Block is a pixel grid in which gradients are constituted from the value and route of alternate in the intensities of the pixel inside the block. As shown above, figure the sample image is fed to a feature descriptor which extracts useful information. In above example x and y are respectively the horizontal and vertical components of the change in the pixel intensity.The descriptors are calculated over blocks of pixels with 8 x 8 dimensions. These descriptor values are quantized into nine containers for each pixel over 8 × 8 blocks, with each bin representing a directional gradient attitude in that bin, which is the sum of the magnitudes of all pixels with the same perspective. The histogram is then normalised over a 16 × 16 block length, which implies four 8 × 8 blocks are normalised together.Similarly, the histogram is then normalized over a 16 x 16 block length, this means that four 8 × 8 blocks are normalised together to reduce the severity of mild problems. Because of the change in light, the system's accuracy suffers. For many faces, the SVM model is trained to use some HOG vectors.

Method-3 MTCNN with FaceNet:

Firstly we'd like to make sure whether the mtcnn library is installed correctly. After installation check the version. We used this library to form a face detector and extract the faces from the input images . This can be further integrated with Facenet face detector models. Initially we load a picture as a Numpy array. We create a numpy array using the PIL library and therefore the open() function. Further conversion of image to RGB is performed if the image is black and white or the alpha channel. We then create a MTCNN face detector class and use it to detect the faces from the loaded image. The output is the list of bounding boxes, each box states a lower-left-corner of bounding. It also defines the width and height of the box.We take absolutely the value of the coordinates if the library returns a negative pixel index. We will use these coordinates to extract the face. PIL library is employed to resize small frames to the desired size because the model expects the square input faces

with the form 160 x 160. The MTCNN architecture we've got used contains four convolution layers and two fully-connected layers with parameters Θ for feature learning. W= [Wa, Ws] from layer 6 to layer 7 are learned separately for every task respectively. Three stages of mtcnn are shown in figure 4.

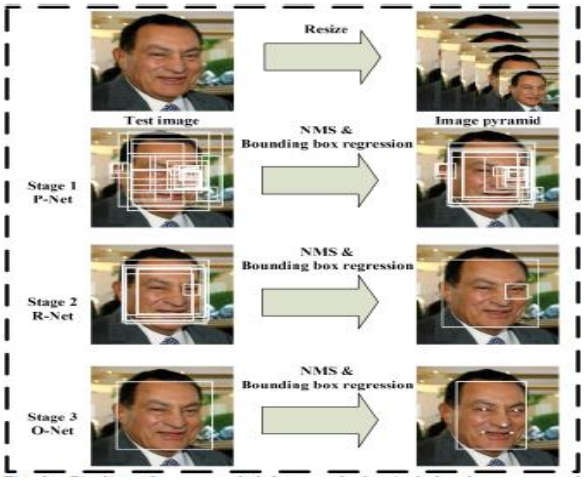


Figure 4 Three-stage multi-task deep convolution networks

We also produce a face embedding, which is a vector that represents the extracted features from a face and can be compared to other vectors of other faces. We create a classifier model that predicts identity using facial embedding as input. The embedding for a particular input frame will be generated by the Facenet model. Next we normalise the vector of every face embedding. It's important to normalise because we use them to check using distance metric. Vector normalisation means scaling the values until the magnitude is of unit length(1) of that vector. We set 'kernel' to 'linear' in a trial to suit linear SVM to schooling statistics. This can be performed using the SVC magnificence in scikit-analysis. We use face embedding as enter to form prediction with the fit version.

4 IMPLEMENTATION RESULTS AND ANALYSIS

The Table 1 highlights, effect of different deep neural networks for Person Re-identification and compares based on the accuracy, GPU usage, Speed, Number of faces detected by overcoming the challenges like illumination and occlusion. The comparative analysis shows that the MTCNN with facenet provides better accuracy in Real world and also handles issues like face orientation/angle variation and illumination as shown in Figure 5 and 6 respectively.

The testing results of the proposed models are shown below, When the first person appears, there is no data to match so system gives him a pickle file with ID '0' as shown in figure 7. When the second person appears, there is no data to match so system gives him a pickle file with ID '1' as shown in figure 8. When these person reappear in the system they will be Re-identified and assigned with the ID stored in the database. As shown in Figure 9, in input frame two faces are found and when the matching algorithm works this Reidentify people and matches up with ID '0' and ID '3' respectively.

Table 1 Comparison of different Algorithms

parameters	Method 1 (CNN)	Method 2 (HOG)	Method 3 (FaceNet)
1.Face Detection Algorithm	Pretrained models of Dlib or OpenCV	Pretrained models of CNN	Pretrained models of MTCNN
2.Face Recognition Algorithm	Face recognition module based on CNN	Face recognition module based on CNN	Linear Support Vector Machine (SVM)
3.Side Face Detection	Not Detected	Not Detected	Detected
4.Testing Accuracy	83-85%	80-82%	85-87%
5.GPU Usage	No	No	Yes
6.No of face Detected at a time	4-5	3-4	4-5
7.Result	Accurate result	Not accurate as CNN	More accurate result
8.No OF Frames Generated	Less or equal to 30 Frames/ Sec	Less or equal to 25 Frames/Sec	Less or equal to 10 Frames/ sec
9.Speed	Slower than method3	Faster than method1 and method 2	Slower than method3 but faster than method 1
10.Occlusion	It can handle Occlusions.	It can not handle Occlusions.	It can handle Occlusions

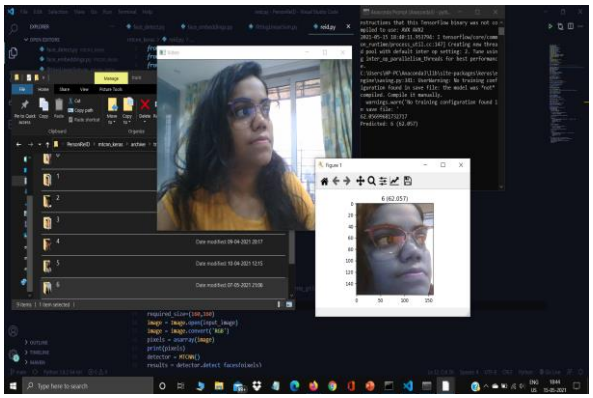
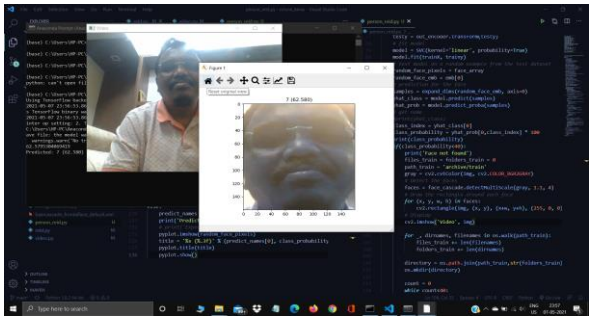


Figure 5 : Person Re-identification for Side view face



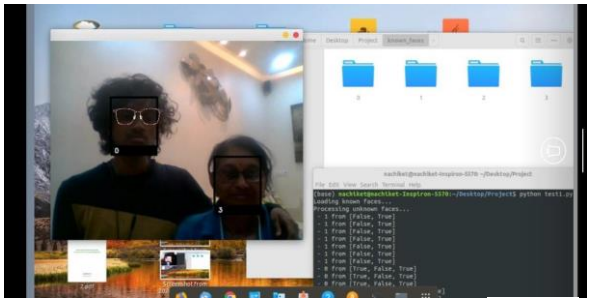


Figure 6 : Person Re-identification under different lighting condition

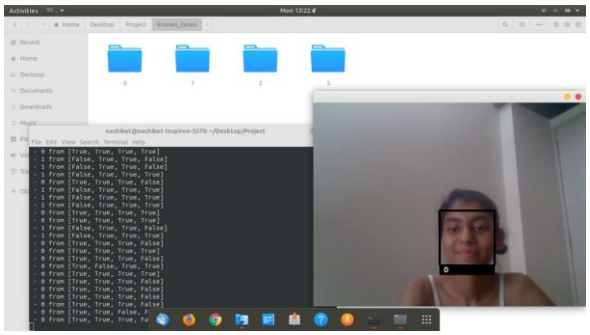


Figure 7 : Results of Unique ID '0' generation for First person enter in the system

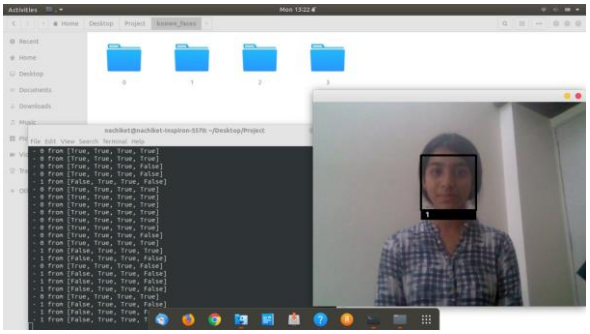


Figure 8: Results of Unique ID generation for next new person

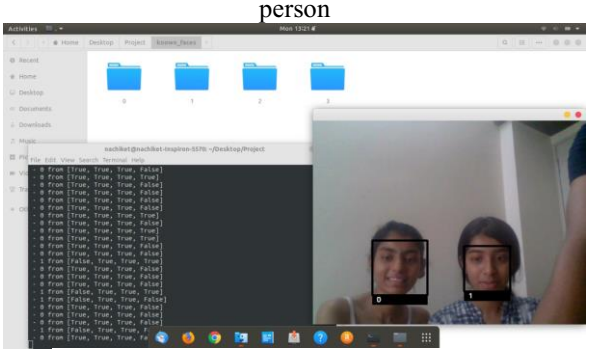


Figure 9 : Results of Person Re-identification

5 CONCLUSION

Person re-identification has gained a lot of attention in recent years. The Proposed model gives us accuracy around 87% and the HOG model gives accuracy 82% whereas the previous model gave accuracy under 85%. The system presented an approach to the challenges faced by the system are occlusion, pose, light condition, face orientation. Currently we are taking input as a single shot image from a single camera, in future we can extend this work by taking input across multiple cameras. The advantage of the system is it doesn't require the database of people in advance for recognition and it will be helpful for criminal identification for crime control and prevention. To increase performance of the system in future, we can use Hybrid models and effective feature learning techniques such as face matrices for face identification.

REFERENCES

[1] W. Zajdel, Z. Zivkovic, and B. Krose, "Keeping track of humans:Have i seen this person before?" in Proceedings of the 2005 IEEE International Conference on Robotics and Automation. IEEE, 2005,pp. 2081–2086.

[2] S. karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps and R. J. Radke, "A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets," in IEEE Transactions on Pattern Analysis and Machine Intelligence,
vol. 41, no. 3, pp. 523-536, 1 March 2019, doi: 10.1109/TPAMI.2018.2807450.

[3] Liang Zheng and Yi Yang and Alexander G. Hauptmann, Person Re-identification:Past, Present,

Future,CoRR,abs/1610.02984,2016,http://arxiv.org/abs/1610.02984
[4] Liu, Wenqian and Camps, Octavia and Sznaier, Mario. (2017). Multi-camera Multi-Object Tracking.
[5] D. Yi, Z. Lei, S. Liao and S. Z. Li, "Deep Metric Learning for Person Re-identification," 2014 22nd International Conference on Pattern Recognition, 2014, pp. 34-39, doi: 10.1109/ICPR.2014.16.
[6] Bai, Xiang & Yang, Mingkun & Huang, Tengteng & Dou, Zhiyong & Yu, Rui & Xu, Yongchao. (2017). Deep-Person: Learning Discriminative Deep Features for Person Re-Identification. Pattern Recognition. 98. 10.1016/j.patcog.2019.107036.
[7] Kasturi R., Ekambaram R. (2014) Person Reidentification and Recognition in Video. In: Bayro-Corrochano E., Hancock E. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2014. Lecture Notes in Computer Science, vol 8827. Springer, Cham. https://doi.org/10.1007/978-3-319-12568-8_35
[8] E. Ristani and C. Tomasi, "Features for Multi-target Multi-camera Tracking and Re-identification," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6036-6046, doi: 10.1109/CVPR.2018.00632.
[9] P. Li, M. L. Prieto, P. J. Flynn and D. Mery, "Learning face similarity for re-identification from real surveillance video: A deep metric solution," 2017 IEEE International Joint Conference on Biometrics (IJCB), 2017, pp. 243-252, doi: 10.1109/BTAS.2017.8272704.
[10] V. Mathew, T. Toby, A. Chacko and A. Udhayakumar, "Person re-identification through face detection from videos using Deep Learning," 2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 2019, pp. 1-5, doi: 10.1109/ANTS47819.2019.9117938.
[11] Liu, Li & Ouyang, Wanli & Wang, Xiaogang & Fieguth, Paul & Chen, Jie & Liu, Xinwang & Pietikäinen, Matti. (2018). Deep Learning for Generic Object Detection: A Survey.
[12] Murat Taskiran, Nihan Kahraman, Cigdem Eroglu Erdem,
Face recognition: Past, present and future (a review),Digital Signal Processing,Volume 106,2020,102809,ISSN 1051-2004,
[13] Florian Schorff, Dmitry Kalenichenko, James Phinbin,." FaceNet: A Unified Embeddings for Face Recognition and Clustering",2015,IEEE
[14] Kaipeng Zhang, Zhangpeng Zhang , Zhifeng Li , Joint Face Dectection and Alignment using Multi-task Cascade Convolution Networks, 2016,IEEE