Dileep J[1*]
Dr. Supriya Vedagiri[2]
Dr. Manjunath Ramachandra[3]

Detection of Principal Component Features for Person Identification in Video

Journal of Electrical Systems

Abstract
In the field of pattern recognition, computer vision and biometrics, various applications are found and brought into real-time application use. Person Detection system made possible to the modern world to achieve better and faster surveillance in crowded areas as it is impossible for human. Person Detection Model consists mainly of three stages namely: Pre-processing, Feature Extraction and Classification. This work explored with Subspace techniques, like Principal Component Analysis (PCA) and Fisher Linear Discriminant Analysis (FLDA) as feature extractors. Learning Vector Quantizer (LVQ) was used as multi-layer network classifier. 100% Recognition rate was obtained using PCA and FLDA through LVQ classifier for Visual Tracker Benchmark Database. 91.67% Recognition rate was obtained using PCA with LVQ classifier for Home Video database. Many challenges were addressed in this research such as variation in light, pose variations (>20 degree) and variation of facial expressions in video. Two minutes of minimal Execution time is required. 100% accuracy was obtained considering 1000 epochs.

Keywords: Principal Component Analysis Linear Discriminant Analysis, Feature Extraction, Learning Vector Quantizer, Multi-layer Classifier

1. INTRODUCTION

Computer vision basically understands and interprets two-dimensional images, as to identify objects individually and to understand the shape and size. An individual possesses different trait, which is distinct to oneself and differentiates from other individuals. Face recognition is one of the natural, nonintrusive and easy to use field, where in human's can be monitored without interruption. To identify a person, there exist multiple technologies like Iris based, Finger print based, retina-based person detection system. Biometrics gives a unique solution as to identify different individual based on traits obtained from the respective individual and can be identified for further implementation of various applications. It is recommended to identify a person using biometric traits rather than conventional passwords and PIN (Personal Identification Number)-based approaches. Biometric identification techniques have limitations. Research happened more in Face recognition with respect to images. Limited research is happening with respect to videos.

A Strong Video Database called Home Video database is taken which contains 6,000 frames. Frames are derived in different pose, varying illumination and facial expressions as to overcome the challenges which helps to identify a particular individual. A standard database called Visual Tracker Benchmark Database is taken in this paper to compare proposed methodology result with previous ones. Zhenguo Ding et al., proposed an architecture for multiple object tracking using block chain fusion architecture [1]. To extract pedestrians deep hash appearance attributes, it uses a HashNet deep neural network. Multiple Object Tracking 15 (MOT15) dataset was used. 72.8% of recognition accuracy was obtained. Tackhyun Jung et al., proposed a system which detects eye blinking pattern [2]. Based on the period, recurring number, and quantity of time that elapsed since the last eye blink, a Deep Vision neural network is employed as a metric to confirm an anomaly. 87.5% accuracy was obtained. Sudeep Thepade et al., proposed a paper which detects lieness of the face using fusion of features [3]. Proposed method showed better recognition result as 78.13% for Nanjing University of Aeronautics and Astronautics (NUAA) database. But the disadvantage is the execution time. Issam Hammad et al., proposed a method which includes two deep neural networks for identifying the people [4]. Highest classification performance of 94.86% was recorded utilizing UCI daily & sports activity raw data.

In paper [5] , author proposed 2D pose estimation-based person identification technique. This includes extracting several angles of a moving person. Convolutional Neural networks were used to get average recognition rate as 90%. 30 degrees pose variation limits were accepted. Object detection technology using deep neural networks [6], [7], [8], [9] were discussed by various authors to show the opportunities and challenges of the design. Oytun Ulutan et al., introduced a person identification approach on Order Preserving Bilinear Model [10]. Requirement was to find out the area of interest with an object moving around. Camera and seismic geophones are used to track the moving objects. Vinay et al., identified a method that classifies the person efficiently from a video database. Author had used Dominant Feature based two self-contained convolutional neural networks [11]. For every frame, key points are calculated and area density is found out.

[1*]Research Scholar, Department of Electronics & Communication Engineering, Sir. M V Institute of Technology, Visvesvaraya Technological University, Belagavi - 590018
[2]Professor, Department of Electronics & Communication Engineering, Sir. M V Institute of Technology, Visvesvaraya Technological University, Belagavi - 590018
[3]Professor, Department of Electronics & Communication Engineering, Atria Institute of Technology, Visvesvaraya Technological University, Belagavi – 590018

***Corresponding Author:** Dileep J
*Department of Electronics & Communication Engineering, Research Scholar, Sir M. Visvesvaraya Institute of Technology, Visvesvaraya Technological University, Belagavi – 590018, India, Email: dileep1721991@gmail.com

These data are used to extract overlapping windows from every frame. So that, classifier gives maximum efficiency of 91% for 100 epochs. Ruichi Yu et al., presented a paper on efficient relevant motion event detection [12].
Author addressed the complexity in identifying relevant movement caused by various objects of interest such as persons, animals or vehicles. This methodology takes some time to compute. To increase the speed, ReMotENet (Relevant Motion Event Detection Network), which is a three-dimensional convolutional neural network had been implemented. This methodology is efficient and has light-weight. Wang Zhiqiang et al., worked on different Object Detection Based methodologies [13]. This paper includes modern convolutional neural networks and its advantages over other neural networks. This paper initiates some public datasets and his perception of verification of data. Paper [14] was proposed for people counting and human detection in challenging situation. Goal is to guess accurate number of objects and group them in a low-resolution frame. 90% of result was obtained for Context Aware Vision using Image-based Active Recognition (CAVIAR) database. Zhen Zhou et al., introduced an approach called Joint Spatial and Temporal Recurrent Neural Networks person detection [15]. Merely feature-based learning or metric-based learning are indeed the primary emphasis of existing models. This network makes use of both the techniques to re-identify an individual. Yang et al., presented his work using Semi-supervised Learning [16]. New feature-based learning methodology on auto encoders has been implemented in this paper. 71% and 90% of accuracy were obtained for horse and pet detection. Machaca et al., describes about Violent Flow (ViF) descriptor to describe the person in voilent videoss[17]. Bank of Standardized Stimuli (BOSS) dataset was used to get maximum accuracy of 91% for 30 frames per second video. Paper [18] had used "Regularized Sparse Representation Classification (RSRC) algorithm" which includes two minimization methods. YouTube database was used to get 65.56% of average precision. 25 degrees was the pose angle variation limit. Various papers have been developed by many authors on convolutional network algorithms to recognize people in a video or from an image [19], [20], [21], [22], [23], [24]. Few papers gave information about introduction to various convolutional neural networks.
Paper [25] showed Long Short-Term Memory (LSTM) based person classification using temporal features. Institute of Automation, Chinese Academy of Sciences (CASIA) database had been used to get 90% classification accuracy. Paper [26] proposed convolutional neural network for identifying people. Blurring and face alignment issues were addressed in this paper to achieve 85% of average classification rate. Kim et al., proposed extended Kalman filter based 3D human tracking for identifying people in video [27]. Fall detection also considered in this paper to safeguard old people without background subtraction technique for which 87% precision is achieved. Paper [28] summarized 13 visual tracking techniques to detect number of people in the selected frame. Deep learning based person identification techniques had been proposed [29], [30], [31], [32], [33], [34]. Recognition rate was more than 90% from the papers. Complex structure and execution time was the limitation. Paper [35] used fuzzy logic with k-NN (k-nearest-neighbor) classifier. This system achieved 92.7% of maximum accuracy. For increased ROI scaling sizes, error class percentage is more. Further, limitations of these papers were tabulated in Results and Discussion section.

2. RESEARCH METHOD

Figure 1 shows the detailed architecture of Person Identification system. Pre-processing deals with many pre-determined steps which is helpful to produce clear frame for extracting features logically. The predominant objective behind feature extraction is to decrease the input frame's dimensions as small as possible. Many logics have been derived to obtain features from the frames obtained for further classification. Learning Vector Quantizer (LVQ) is used as multi-layer classifier in this work, which sorted recognized classes as a unique individuals. Training video is given to the network for converting video file into frames. Frames are in color in nature, which are converted to gray images for easy processing and takes less memory to store. Frame resize operation takes place to concentrate only on required facial parameters. Unwanted area is neglected to save execution time as well. All the three steps belong to pre-processing stage. Next comes feature extraction stage. It consists of three stages. The averaging technique is performed on frame data to compute mean vectors for an individual person. Difference vector is calculated at the end to bifurcate one individual characteristic with another person. Covariance matrix is computed for all individuals and given to the classifier. Preprocessing and feature extraction procedure is applied to testing video as well. Covariance vectors of both training and testing frames are compared at multi-layer classifier. Final output is generated for all given testing frames to validate the authenticity of a person. Recognition rate and execution time can be noted for performance measures.

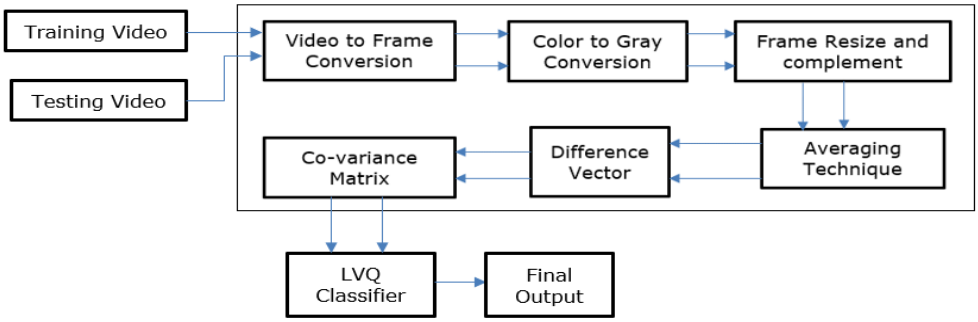

Figure 1. Proposed Method for Person Identification

2.1. Adopted Dataset

Each investigation was carried out employing HV (Home Video) Raw-Dataset and VTB (Visual Tracker Benchmark) Dataset. HV Database is considered from four different persons in Bangalore. Each video was taken for 50 seconds. For individual person, 1500 frames were extracted. Totally 6000 frames were used in this work. For every 30 frames, system was taking single frame for training or testing. The frames from videos are captured at critical junctures with various lighting, facial emotional reactions (closed or open eyes, smiling or sad), and posture that correspond to the change in time. All the videos were taken through RGB Color Camera with resolution 720P and 30f/s. Figure 2 shows sample frames of Person-1 from HV database taken at different time instances. A standard database namely Visual Tracker Benchmark Database is used. 4 person videos are considered. 200 frames are extracted from each video. Figure 3 shows sample frames from VTB database taken at different time instances. The database is randomly split into a training set and a testing set with no overlap between them in necessary to undertake out the person identification experiment. The database is segregated as follows: "n" is the cumulative number of frames in each experiment. Each person receives "k" randomly chosen frames for the test, while the leftover "n - k" frames serve as the training set and are used to calculate the projection matrix. The training and test sets' "n" frames were all projected into a dimension-reduced region. Each experiment decided to perform numerous runs.

Figure 2. Person-1 frames from HV database

Figure 3. Sample frames from VTB database

2.2. Feature Extraction

Since the raw data for face recognition has a very high degree of dimension, dimension reduction (Feature Extraction) is recommended before classification. Among the several feature extraction techniques, “Principal Component Analysis” (PCA) and “Fisher Linear Discriminant Analysis” (FLDA) are frequently employed. There are two primary sorts of dimension reduction strategies. The first is based on local characteristics, which typically extract a collection of face features from the frame, such as the eyes, nose, and so on. The face is categorized using these characteristics. The other is global or holistic techniques, which incorporates a broader feature extraction of face frames and takes a comprehensive view of the identification task.

2.2.1. Principal Component Analysis (PCA)

The high-dimensional data are mapped onto a lower-dimensional environment using a linear strategy. It pursues a weight estimate that best characterizes the data, which is called principal components. Let S(p, q) be a two-dimensional N × N array for a face frame. The training set frames are mapped onto a assembly of points in this gigantic region; these points are represented as subspace. These vectors are the “Eigen vectors” which is attained after the “covariance matrix” which defines the subspace of face frames.

$$L = \frac{1}{G}\sum_{n=1}^{G} \phi_n \, {\phi_n}^T = S\,S^T \qquad (1)$$

Matrix S= [ϕ_1, ϕ_2,....., ϕ_m]. The Matrix L nevertheless is N^2 x N^2.

2.2.2. Fisher Linear Discriminant Analysis (FLDA)

System can extract the facial features using the described fisher face or discriminant eigenfeature technique. This method keeps the principle of the Eigenface in projecting faces from a high-dimension image space to a noticeably lower-dimensional feature space while overcoming the shortcoming of the Eigenface method by incorporating "Fisher's Linear Discriminant" (FLD) criteria. Creating a concise internal representation of faces, also known as feature extraction, is the aim of face processing employing neural networks. In order to separate distinct classes of training data as much as feasible and accumulate the same classes of patterns as closely as possible, the PCA first develops a wide range of the most expressive features, and the FLD is subsequently employed to build a set of the most discriminant features. The projection is improved using "Fisher-LDA" by

$$J(H) = \frac{H^T S_B H}{H^T S_W H} \qquad (2)$$

S_B = “Between Class Scatter Matrix”

S_W = “Within Class Scatter Matrix”

$$S_B = \sum_C N_C \, (\mu_C - \bar{x})(\mu_C - \bar{x})^T \quad (3)$$

$$S_W = \sum_C \sum_{i \epsilon C} (x_i - \mu_C) \, (x_i - \mu_C)^T \quad (4)$$

$$where, \quad \mu_C = \frac{1}{N_C} \sum_{i \epsilon C} x_i \quad (5)$$

$$\bar{x} = \frac{1}{N} \sum_i x_i = \frac{1}{n} \sum_C N_C \mu_C \quad (6)$$

Even J could be defined using covariance matrix as its proportional to scatter matrix. The dimensions of scatter matrix are defined in equation (3) & (4). Where N_C denotes the quantity of cases in class C. The Fisherfaces approach, which circumvents the issue by projecting the image set to a lower-dimensional space so that the resulting within-class scatter matrix S_W is non-singular, was developed to address the challenge of a singular S_W. This is accomplished by initially deploying PCA to minimize the spatial size of the feature area to N − c and after adopting the typical FLDA to reduce the spatial size to c− 1. The feature matrix constructed from the image space is the within class scatter matrix.

2.3. Learning Vector Quantizer (LVQ)

"Learning Vector Quantizer" is a schematic based "supervised classification technique". Figure 4 shows LVQ Architectural diagram. It possesses a first layer of competition and a second layer of linearity. The linear layer converts the competitive layer's classes into the intended categorization as defined, while the competitive layer categorises input vectors. Each (sub or target) class contains one neuron in the competitive and linear stages. These gives an opportunity to learn up to E^1 subclasses and E^2 target classes (E^1 is always greater than E^2).

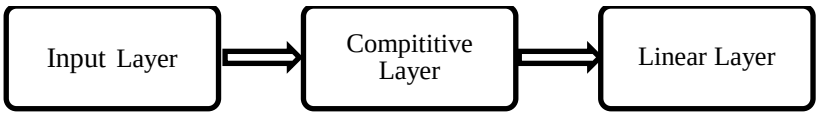

Figure 4. Learning Vector Quantizer Architecture

2.4. Algorithm for Proposed Approach

This Algorithm is designed for either PCA as feature extraction technique and LVQ as classifier. Figure 5 and Figure 6 demonstrate the Training and Testing phase algorithms respectively.

Training stage begins:
input: a collection of video training frames
Outcome: A Wisdom Database (WD) with Extracted Features
Technique:
1. Collect training aids.
2. Training samples + suggested model: generation of relevant features.
3. Conserve the features in the wisdom database.
Training stage ends

Figure 5. Algorithm for Training phase

Testing stage begins:
input: i. A Wisdom Database with Extracted Features
ii. Query frames, Q.
Output: Query frame label or identity.
Technique:
1. Extracted features for Query frames
2. Use classification approach to discover the uniqueness of query frames.
Testing stage ends

Figure 6. Algorithm for Testing phase

3. RESULTS AND DISCUSSION

In this work, subspace methods such as PCA and FLDA were used as feature extraction techniques. LVQ was used as classifier. The number of successfully identified frames to the total number of frames can be used to calculate the recognition rate or accuracy. This has to be multiplied by 100 to represent in terms of percentage.

$$\text{Recognition Rate} = \frac{\text{No. of classified frames}}{\text{Total No. of frames}} * 100 \quad (7)$$

Table 1 depicts the results of person detection using PCA and LVQ. For HV Database, system obtained maximum result of 91.67% for 35 Hidden Layers. If number of hidden layers were increased, then computation time was increasing slightly. For the standard VTB Database, proposed system produced 91.66%, considering equal number of training and testing frames for 15 Hidden layers. If hidden layer count is increased to 45, same network produced 100% recognition rate. Table 2 depicts the results of person detection using FLDA and LVQ. For HV Database, system obtained maximum result of 80% for 70 Hidden Layers. Computation time increased with increase in hidden layer count. For the standard VTB Database, proposed system produced 100% result, considering 32 number of training and 16 testing frames for 20 Hidden layers.

Table 1. PCA with LVQ

Trained Frames	Testing Frames	Hidden Layers	Classified Frames	Misclassified Frames	Recognition Rate in %	Database
8	4	15	16	0	100	
8	4	100	16	0	100	
6	6	15	22	2	91.67	VTB Database
6	6	25	23	1	95.83	
6	6	45	24	0	100	
44	6	100	22	2	91.67	
40	10	100	30	10	75	HV Database
44	6	35	22	2	91.67	
44	6	45	20	4	83.33	

Table 2. FLDA with LVQ

Trained Frames	Testing Frames	Hidden Layers	Classified Frames	Misclassified Frames	Recognition Rate in %	Database
8	4	20	16	0	100	
8	4	40	12	4	75	
8	4	60	16	0	100	VTB Database
8	4	140	13	3	81.25	
8	4	180	12	4	75	
40	10	70	32	8	80	
40	10	90	32	8	80	HV Database
45	5	20	14	6	70	

Figure 7 shows graph metric which illustrates the "comparison of False Positive Rate versus True Positive Rate". Figure 7(a), 7(b) denotes training ROC and Test ROC plot respectively. The above metric was obtained for PCA as feature extractor and LVQ as classifier by taking 176 total training frames for which 70.8% recognition accuracy was obtained. Training Region of Convergence (ROC) Graph metric should be towards True positive rate to get the accurate results. True positive rate can be illustrated as number of successfully categorized faces to the complete number of existed faces in a frame. In False positive rate, the face is predicted as "YES" but there will be no real face existed in the frame. Theoretically ratio of TPR to FPR should be infinite considering zero false predictions. With respect to diagram, the ratio exists and which should be greater than 1.

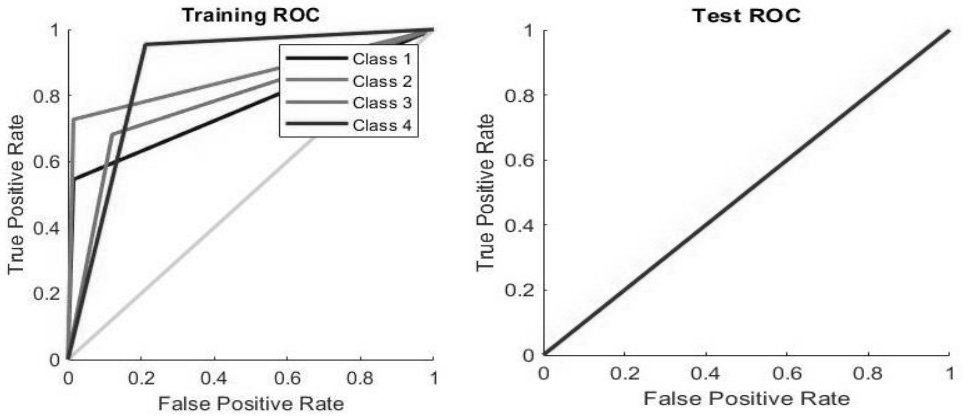

Figure 7. PCA LVQ ROC Metric (a)Training ROC (b) Test ROC

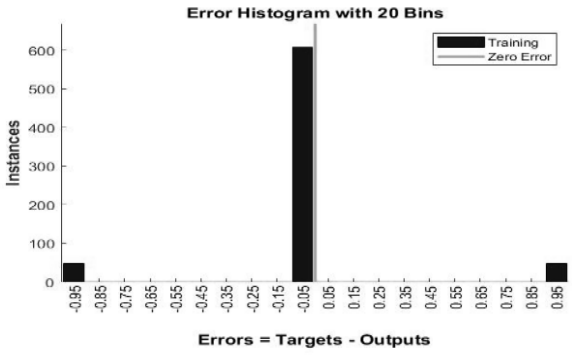

Figure 8. Error Histogram Metric

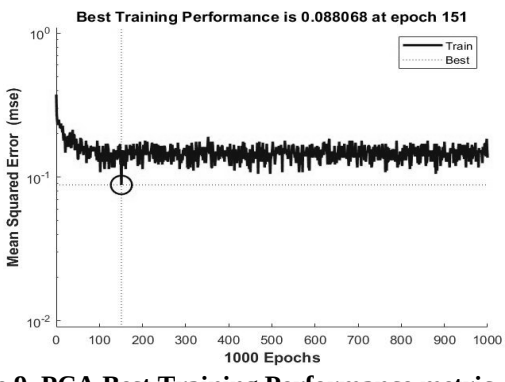

Figure 9. PCA Best Training Performance metric

Figure 8 shows the error histogram metrics for detection of a person using PCA and LVQ. For 176 training frames, the output of error histogram plot should be less than zero or negative to be accepted. Theoretically error histogram is zero, but practically it exists with small values. Figure 9 shows training performance metric using PCA and LVQ classifier. It gave the best training performance 0.088068 for 151st epoch, for which 176 frames were trained. This graph was plotted with respect to 1000 epochs and Mean square error (MSE). Ideal value for MSE was "Zero". Table 3 shows the performance measures comparison of various papers with respect to proposed method. It is evident that, proposed scheme performs well compare to existing methods.

Table 3. Performance Measures Comparison

Paper Number	Illumination Limit in %	Pose Variation Limit	Expression Variation Limit	Execution Time	Recognition Rate in %
[1]	Illumination: 40	15 deg	---	---	72.8
[5]	---	20 deg	---	---	90
[8]	---	30 deg	---	---	81
[9]	Background: 15, Occlusion: 30	---	---	---	96
[14]	Background: 15, illumination: 35	---	---	---	90
[18]	---	25 deg	---	---	65.56
[19]	Occlusion: 25	---	---	---	85
[20]	---	30 deg	---	---	92.6
[21]	---	25 deg	---	---	85
[23]	---	---	---	360 min	90
[24]	Occlusion: 30	---	Expression: 8	---	99
Proposed System	**Background: 20, illumination: 50**	**45 deg**	**Expression: 10**	**2-5 min**	**100**

4. CONCLUSION

By using LVQ classifier and PCA or FLDA feature extraction techniques on video database, it is possible to overcome the stated challenges such as Illumination, pose variation limit up to 45 degrees (One Eye partially / completely invisible) and different expressions (like Eye Open, Eye close, smile, sad). Recognition Accuracy is greater than 90% under robust conditions. PCA and FLDA along with LVQ provides best result even for a greater number of epochs (1000). Computation time is very less. For reading a video and converting it to number of frames, model took 2 to 5 minutes of average minutes. With a standard database called VTB database and strong "HV" database were taken to prove the results for all conditions. From Table III, it is evident that greater number of challenges had been solved to some extent by considering different expressions, varying illumination in the video. Background variations and pose angle problem had been addressed.

REFERENCES

[1] Z. Ding, S. Liu, M. Li, Z. Lian, and H. Xu, "A Blockchain-Enabled Multiple Object Tracking for Unmanned System With Deep Hash Appearance Feature," IEEE Access, vol. 9, pp. 1116–1123, 2021, doi: 10.1109/ ACCESS.2020.3046243.

[2] T. Jung, S. Kim, and K. Kim, "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," IEEE Access, vol. 8, pp. 83144–83154, 2020, doi: 10.1109/ACCESS.2020.2988660.

[3] S. Thepade, P. Jagdale, A. Bhingurde, and S. Erandole, "Novel Face Liveness Detection Using Fusion of Features and Machine Learning Classifiers," in 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar: IEEE, Feb. 2020, pp. 141–145. doi: 10.1109/ICIoT48696.2020.9089525.

[4] I. Hammad and K. El-Sankary, "Using Machine Learning for Person Identification through Physical Activities," in 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain: IEEE, Oct. 2020, pp. 1–5. doi: 10.1109/ISCAS45731.2020.9181231.

[5] G. Cao, Y. Pu, Y. Li, and Z. Zhao, "Human Motion Capture Using a Multi-2D Pose Estimation Model," in 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China: IEEE, Aug. 2019, pp. 64–67. doi: 10.1109/IHMSC.2019.00023.

[6] A. R. Pathak, M. Pandey, and S. Rautaray, "Application of Deep Learning for Object Detection," Procedia Comput. Sci., vol. 132, pp. 1706–1717, 2018, doi: 10.1016/j.procs.2018.05.144.

[7] S. A. Vahora and N. C. Chauhan, "Deep neural network model for group activity recognition using contextual relationship," Eng. Sci. Technol. Int. J., vol. 22, no. 1, pp. 47–54, Feb. 2019, doi: 10.1016/j.jestch.2018.08.010.

[8] F. Letsch, D. Jirak, and S. Wermter, "Localizing salient body motion in multi-person scenes using convolutional neural networks," Neurocomputing, vol. 330, pp. 449–464, Feb. 2019, doi: 10.1016/j.neucom.2018.11.048.

[9] N. Narayan, N. Sankaran, S. Setlur, and V. Govindaraju, "Re-identification for Online Person Tracking by Modeling Space-Time Continuum," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA: IEEE, Jun. 2018, pp. 1519–151909. doi: 10.1109/CVPRW.2018.00193.

[10] O. Ulutan, B. S. Riggan, N. M. Nasrabadi, and B. S. Manjunath, "An Order Preserving Bilinear Model for Person Detection in Multi-Modal Data," 2017, doi: 10.48550/ARXIV.1712.07721.

[11] A. Vinay, D. A. Mundroy, G. Kathiresan, U. Sridhar, K. N. B. Murthy, and S. Natarajan, "Dominant feature based convolutional neural network for faces in videos," in 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, Andhra Pradesh, India: IEEE, Mar. 2017, pp. 17–22. doi: 10.1109/ICBDACI.2017.8070802.

[12] R. Yu, H. Wang, and L. S. Davis, "ReMotENet: Efficient Relevant Motion Event Detection for Large-Scale Home Surveillance Videos," in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV: IEEE, Mar. 2018, pp. 1642–1651. doi: 10.1109/WACV.2018.00183.

[13] W. Zhiqiang and L. Jun, "A review of object detection based on convolutional neural network," in 2017 36th Chinese Control Conference (CCC), Dalian, China: IEEE, Jul. 2017, pp. 11104–11109. doi: 10.23919/ChiCC.2017.8029130.

[14] Y.-L. Hou and G. K. H. Pang, "People Counting and Human Detection in a Challenging Situation," IEEE Trans. Syst. Man Cybern. - Part Syst. Hum., vol. 41, no. 1, pp. 24–33, Jan. 2011, doi: 10.1109/TSMCA.2010.2064299.

[15] Z. Zhou, Y. Huang, W. Wang, L. Wang, and T. Tan, "See the Forest for the Trees: Joint Spatial and Temporal Recurrent Neural Networks for Video-Based Person Re-identification," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, Jul. 2017, pp. 6776–6785. doi: 10.1109/CVPR.2017.717.

[16] Y. Yang, G. Shu, and M. Shah, "Semi-supervised Learning of Feature Hierarchies for Object Detection in a Video," in 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA: IEEE, Jun. 2013, pp. 1650–1657. doi: 10.1109/CVPR.2013.216.

[17] V. E. Machaca Arceda, K. M. Fernández Fabián, P. C. Laguna Laura, J. J. Rivera Tito, and J. C. Gutiérrez Cáceres, "Fast Face Detection in Violent Video Scenes," Electron. Notes Theor. Comput. Sci., vol. 329, pp. 5–26, Dec. 2016, doi: 10.1016/j.entcs.2016.12.002.

[18] S. Nagendra, R. Baskaran, and S. Abirami, "Video-Based Face Recognition and Face-Tracking using Sparse Representation Based Categorization," Procedia Comput. Sci., vol. 54, pp. 746–755, 2015, doi: 10.1016/j.procs.2015.06.088.

[19] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, "Tracking People by Detection Using CNN Features," Procedia Comput. Sci., vol. 124, pp. 167–172, 2017, doi: 10.1016/j.procs.2017.12.143.

[20] Y. Li, R. Ge, Y. Ji, S. Gong, and C. Liu, "Trajectory-Pooled Spatial-Temporal Architecture of Deep Convolutional Neural Networks for Video Event Detection," IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 9, pp. 2683–2692, Sep. 2019, doi: 10.1109/TCSVT.2017.2759299.

[21] N. M. Ara, N. S. Simul, and Md. S. Islam, "Convolutional neural network approach for vision based student recognition system," in 2017 20th International Conference of Computer and Information Technology (ICCIT), Dhaka: IEEE, Dec. 2017, pp. 1–6. doi: 10.1109/ICCITECHN.2017.8281789.

[22] R. K. Kumar, G. A. R. Kumar, J. Garain, D. R. Kisku, and G. Sanyal, "Determine attention of faces through growing level of emotion using deep Convolution Neural Network," in 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur: IEEE, Jul. 2017, pp. 975–980. doi: 10.1109/ICICICT1.2017.8342699.

[23] K. C. P. Tanay, S. Khanna, V. Chandrasekaran, and P. K. Baruah, "Fast video super resolution using deep convolutional networks," in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore: IEEE, Mar. 2017, pp. 1–6. doi: 10.1109/ICIIECS.2017.8276067.

[24] K. Yan, S. Huang, Y. Song, W. Liu, and N. Fan, "Face recognition based on convolution neural network," in 2017 36th Chinese Control Conference (CCC), Dalian, China: IEEE, Jul. 2017, pp. 4077–4081. doi: 10.23919/ChiCC.2017.8027997.

[25] Z. Xu, S. Li, and W. Deng, "Learning temporal features using LSTM-CNN architecture for face anti-spoofing," in 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia: IEEE, Nov. 2015, pp. 141–145. doi: 10.1109/ACPR.2015.7486482.

[26] Y. Li, Y. Takashima, T. Takiguchi, and Y. Ariki, "Lip reading using a dynamic feature of lip images and convolutional neural networks," in 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan: IEEE, Jun. 2016, pp. 1–6. doi: 10.1109/ICIS.2016.7550888.

[27] S. Kim, M. Ko, K. Lee, M. Kim, and K. Kim, "3D fall detection for single camera surveillance systems on the street," in 2018 IEEE Sensors Applications Symposium (SAS), Seoul, Korea (South): IEEE, Mar. 2018, pp. 1–6. doi: 10.1109/SAS.2018.8336746.

[28] Jia Zhang, Lei Yang, and Xiaoyu Wu, "A survey on visual tracking via convolutional neural networks," in 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu: IEEE, Oct. 2016, pp. 474–479. doi: 10.1109/CompComm.2016.7924746.

[29] Z. Hu, W. Hou, and X. Liu, "Deep Batch Active Learning and Knowledge Distillation for Person Re-Identification," IEEE Sens. J., vol. 22, no. 14, pp. 14347–14355, Jul. 2022, doi: 10.1109/JSEN.2022.3181238.

[30] Bong-Nam Kang, Yonghyun Kim, and D. Kim, "Deep convolution neural network with stacks of multi-scale convolutional layer block using triplet of faces for face recognition in the wild," in 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary: IEEE, Oct. 2016, pp. 004460–004465. doi: 10.1109/SMC.2016.7844934.

[31] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Ouyang, and Y. Yang, "Exploit the Unknown Gradually: One-Shot Video-Based Person Re-identification by Stepwise Learning," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 5177–5186. doi: 10.1109/CVPR.2018.00543.
[32] Y. Wang, Y. Sun, Z. Lan, F. Sun, N. Zhang, and Y. Wang, "Occluded Person Re-Identification by Multi-Granularity Generation Adversarial Network," IEEE Access, vol. 11, pp. 59612–59620, 2023, doi: 10.1109 /ACCESS.2023.3285798.
[33] S. Raghavendra, Ramyashree, S. K. Abhilash, V. M. Nookala, and S. Kaliraj, "Efficient Deep Learning Approach to Recognize Person Attributes by Using Hybrid Transformers for Surveillance Scenarios," IEEE Access, vol. 11, pp. 10881–10893, 2023, doi: 10.1109/ACCESS.2023.3241334.
[34] M. O. Almasawa, L. A. Elrefaei, and K. Moria, "A Survey on Deep Learning-Based Person Re-Identification Systems," IEEE Access, vol. 7, pp. 175228–175247, 2019, doi: 10.1109/ACCESS.2019.2957336.
[35] M. Barry and E. Granger, "Face Recognition in Video Using a What-and-Where Fusion Neural Network," in 2007 International Joint Conference on Neural Networks, Orlando, FL, USA: IEEE, Aug. 2007, pp. 2256–2261. doi: 10.1109/IJCNN.2007.4371309.

BIOGRAPHY OF AUTHORS

Dileep J received M. Tech degree in 2015 from PESIT, Bengaluru, India. He is pursuing a Ph.D. degree at the Department of Electronics and Communication Engineering, Sir. M. Visvesvaraya Institute of Technology, Bengaluru, Affiliated to VTU, Belagavi. Currently, he works as an Assistant Professor in the Department of Electronics and Communication Engineering at Kammavari Sangham School of Engineering and Management, Bengaluru. His research areas include image processing, machine learning, deep learning, and computer vision. Mail: dileep1721991@gmail.com.

Supriya V G received Ph.D degree in 2016 from Jain University, Bengaluru, India. Currently, she works as a Professor and Head of the Department of ECE, Sir.M.VIT, Bengaluru. She has about 34 years of work experience in the field of communication and image processing- including medical image processing and wireless/mobile networking. She has 20 journal and conference publications. She worked as Principal of GTTC Women Polytechnic for 10 years and started Mechatronics Training Centre. Her areas of interest include signal & image processing, cryptography, and renewable energy. Mail: hod_ece@sirmvit.edu.

Manjunath Ramachandra received Ph.D degree in 2007 from Bangalore University. He has 24 years of work experience in the overlapping verticals of signal processing. He published 199 journal and conference publications, patent disclosures, and a book. He represented Philips in international standardization bodies such as the Wi-Fi Alliance, served as the editor for the regional profiles standard in the Digital living network alliance (DLNA), and as the industrial liaison officer for the CE-Linux Forum. He has chaired about 30 conferences. His areas of interest include signal & image processing database architecture.
Mail: drmanjunathramachandra@gmail.com.