Continuous hand gesture segmentation and acknowledgement of hand gesture path for innovative effort interfaces

Prashant Richhariya[1], Piyush Chauhan[1,2], Lalit Kane[1,3], Bhupesh Kumar Dewangan[4]
[1]Schools Computer Science and Engineering, University of Petroleum and Energy Studies, Dehradun, India
[2]Department of Computer Science and Engineering, JAIN (Deemed to be University), Bangalore, India
[3]Department of Computer Science and Engineering, MIT World Peace University, Pune, India
[4]Department of Computer Science and Engineering, OP Jindal University, Raigarh, India

Article Info

Article history:

Received Oct 12, 2022
Revised Sep 5, 2023
Accepted Sep 17, 2023

Keywords:

Accuracy
Fault-tolerance
Human computer interaction
Optimization
Performance
Resource cost
Service level agreement violation rate

ABSTRACT

Human-computer interaction (HCI) has revolutionized the way we interact with computers, making it more intuitive and user-friendly. It is a dynamic field that has found it is applications in various industries, including multimedia and gaming, where hand gestures are at the forefront. The advent of ubiquitous computing has further heightened the interest in using hand gestures as input. However, recognizing continuous hand gestures presents a set of challenges, primarily stemming from the variable duration of gestures and the lack of clear starting and ending points. Our main objective is to propose a solution: the framework for "continuous palm motion analysis and retrieval" based on "Spatial-temporal and path knowledge". Framework harnesses the power of cognitive deep learning networks (DLN), offering a significant advancement in the continuous hand gesture recognition domain. we conducted rigorous experiments using a diverse video dataset capturing hand gestures for boasting an impressive F-score of up to 0.99. The potential of our framework to significantly enhance the accuracy and reliability of hand gesture recognition in real-world applications.

Corresponding Author:

Bhupesh Kumar Dewangan
Department of Computer Science and Engineering, OP Jindal University
Punjipathra, Raigarh, Chhattisgarh 496109, India
Email: bhupesh.dewangan@gmail.com

1. INTRODUCTION

Effective communication involves not only spoken words but also gestures, they are essential for expressing and boosting communication's expressiveness. This applies to both the speaker and the audience. In the realm of human-computer interaction (HCI), gestures are instrumental in facilitating seamless interaction. Gestures serve as a bridge between the speaker's intent and the audience's understanding, forming the foundation of interaction [1]. When it comes to recognizing hand gestures, there are two primary approaches: non-vision-based and vision-based [2]. Among these, vision-based methods are particularly appealing due to their natural feel. Vision-based approaches can be further categorized as either active or passive. Active sensing techniques have emerged as a successful avenue for gesture recognition, notably through the utilization of devices such as Microsoft_kinect V2 [3], [4] and Leap_Motion cameras. These technologies offer a dynamic and responsive means of capturing gestures, making the recognition process more effective and accurate. In summary, effective communication relies not only on words but also on gestures, which are pivotal in both conveying and enhancing the overall message. In the context of HCI, gestures serve as a fundamental tool, bridging the gap between speakers and their audience. The methods used for recognizing hand gestures vary, with vision-based approaches, particularly those employing active sensing devices like Kinect V2 and leap motion cameras, proving to be highly successful in this endeavor.

Tools have been developed to aid linguists in analyzing gestures during interactions [4]. Different aspects of a gesture, such as stillness, cerebral infarction, planning, hold, and retraction, are involved in it. Classification is the first phase in applications that demand movements of the hands, and this provides a significant challenge for movement analysis [4], [5]. In the context of recognition and classification of frequent hand movements, two common approaches are considered: i) image division before recognition and ii) cooperative separation and identification.

The latter approach, synchronized segmentation and recognition, is often favored as it feels more natural and doesn’t require additional motion [5], [6]. The primary goal of this study is to create a framework that combines segmentation and recognition simultaneously. It is required for the system to perform categorization according to physical as well as ordered data when using inactive monitoring, which is often used for vision-based interactions between humans and computers applying devices like Microsoft. Whenever movement, the hand’s location in each frame can be determined using spatial categorization, and the gesture’s beginning and conclusion points can be determined using temporal fragmentation. Both spatial and temporal segmentation are important in a continuous video stream. It's crucial to keep within consciousness that when viewing such flows, the movements that are relevant are frequently enmeshed within a chaotic or dynamic background. Therefore, communicating knowledge of coordinates for position and path velocity is crucial to effective interactions. Variations in gesture velocity may also offer difficulties.

2. PROPOSED METHOD

The process of segmenting gestures into distinct phases introduces complexities in the analysis of these gestures. A number of obstacles must be solved in order to achieve the objective of building an architecture for ongoing palm motion interpretation and identification that utilizes spatial-temporal and path variables. We have identified three key problems in the framework of this research and have developed remedies for each. The multilayer perceptron (MLP) [7]-[18], which has a deep layer building and a suitable sampling method, is the deep learning system we ultimately use to achieve.

2.1. Challenge of vertical identification

In contactless sensor systems, the camera often captures not only intentional hand gestures but also unintended movements. In each frame of the input sequence, we assume the gesturing hand is reliable [19]. One of the primary challenges in this continuous stream of data is accurately determining the precise location of a gesture.

2.2. Temporal segmentation challenge

Certain gestures, such as composing numeric characters, pose difficulties in pinpointing their start and completion due to “trash movements” that occur between two consecutive images [20]. Many gesture recognition systems use a fixed-width sliding window approach to address these issue, which may not be the most effective strategy. Self supervised temporal domain adaptation (SSTDA) segmentation challenges help us to reduced the discrepancy by applying two main approach of binary and self supervised tasks (Figure 1). Figure 1 show the two self-supervised auxiliary tasks in SSTDA: i) binary domain prediction: discrimination single frame, and ii) sequential domain prediction: predict a sequence of domains for an untrimmed video. These two tasks contribute to local and global SSTDA, respectively.

2.3. Pathway-related problems

Individual variances in gesture path, including variances in velocity and position, can significantly impact recognition performance [21], [22]. To tackle the challenges, this work makes use of arm gesture-related spatial-temporal and trajectory data. The data originates from images routinely taken with the Xbox console’s camera device [23]. To capture diverse gesture behaviors, three different individuals record the video dataset during separate sessions.

This paper provides a framework that addresses categorization and classification at exactly the same time. It takes the motion participation and separates geographical and statistical gesture detection data. The segmentation procedure entails pulling apart, using an ongoing basis, individual image frames from the video and locating the positions of the conjunction, the arm, forehead, and vertebra within each frame. Furthermore, the essential coordinates are gathered from the acquired feature vectors, which include parameters like the acceleration and motion of the fingers and forearm. For organizing uninterrupted data from videos, further trajectory-related data is also extracted [24], including velocity fluctuations and spots. The fact that the proposed approach intends to perform both temporal as well as spatial differentiation, it can be applied to a variety of contexts and situations for the recognition of gestures. See Figure 2 for a depiction in visual form. The suggested deep learning connect accepts the retrieved features after being re-sampled with nearest neighbor based algorithms. The next section gives the planned network's features.

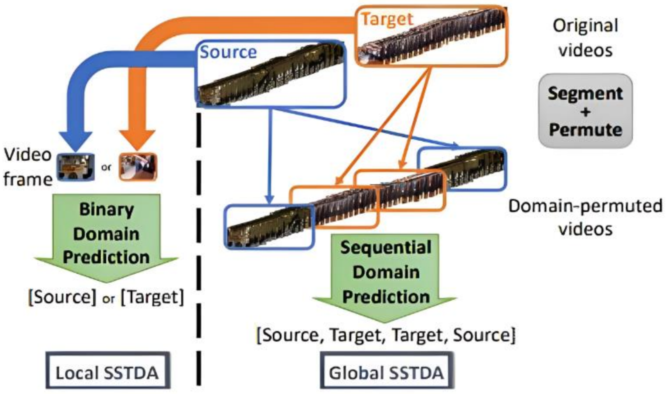

Figure 1. The two self-supervised auxiliary tasks in SSTDA

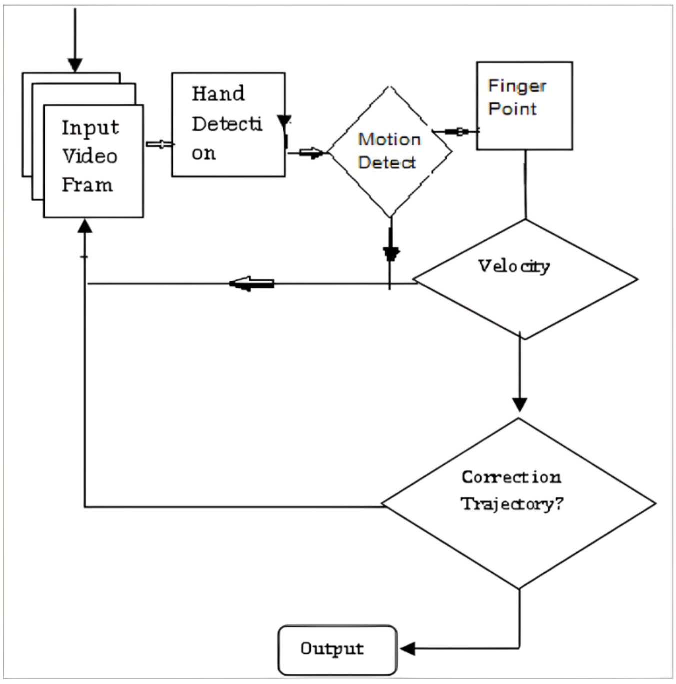

Figure 2. Framework for the designed layout

3. PROPOSED NETWORK OR ALGORITHM FOR REAL LEARNING

Neuronal networks, an important category of artificial intelligence (AI) techniques, govern the comprehensive computational intelligence field. Artificial neural networks (ANNs) are accountable for generating architectures that function to correspond with the creature's cerebral neuronal network. The adaptive method of learning helps the neural networks based architectures establish how to generate an order or a prediction. When the amount of data available grows, these systems perform classical procedures like support vector machine (SVM) and random forest. Concurrent neural networks [19], multi-layer [25], [26], and repetitive ANNs [27] are three distinct kinds of supervised neural networks. Whatever they learn, the way they learn it, and other factors define these networks. It has three different types of layers, an input layer, one or more hidden layers and an output layer.

In each node of the input layer, the input feature, weights and bias are taken as input and the output E is calculated:

$$E = WTxi \ + b \qquad (1)$$

whereas for numerous programs and deposits the formula will become:

$$E = \sum_{I=1}^{m} Whi * \ Gi \ + \ b \qquad (2)$$

inputs are represented by:

$$G = \ g1, g2 \ldots gm \qquad (3)$$

W^1: characterizes mass set at h=1, i.e., at first unseen layer.

$$W^1 = w^1, w^1 \ldots w^1 \qquad (4)$$

The weights and unfairness are values that are arbitrarily prepared original. The weights and unfairness are used to calculate the output E. These networks are tuned using optimal values of unfairness and masses to fit our data.

Finally, the system categorizes the input into output class ŷ based on an stimulation function a which is either rectified linear unit (ReLu) or sigmoid (σ) on z, such that.

$$\alpha \ = \ \sigma(z) \qquad (5)$$

$$\hat{y} = \ \{y1, y2, y3, y4, y5\} \qquad (6)$$

$$\hat{y} \ = \ \{D, P, S, H, R\} \qquad (7)$$

MLP: it is a profound, counterfeit neural system formed on more than one perceptron. The information layer gets the sign and settles on a choice or expectation. In any case, there are discretionary quantities of shrouded layers that goes about as an MLP’s apparent processing engine [28]. MLPs are often capable of carrying out instructional activities. Descent of gradients is the method of changing weights as well as biases in line with the cost model by means of back-propagation. There are many available loss functions. We will simply use (8).

$$Sum \ of \ Squares \ Error = \ \sum_{i=1}^{m}(y - \ \hat{y})2 \qquad (8)$$

The convolutional neural network (CNN) network has been recognized as one of the most important machine vision applications [29]. By analyzing low-level information, such as the movement of arms, and then setting up the representation, which is more abstract and specialized, through a series of levels of convolution, CNN model is able to execute categorization.

4. RESEARCH MODELS

Theoretical background, image and video processing, feature extraction, gesture representation and gesture recognition algorithms. In the next section, we will introduce some of the principles that we use in our research. In this we have processed the images through feature extraction techniques. After that we have applied the gestures recognition algorithm to classifying the images. Then after real-time processing have to be done.

4.1. Theoretical background

Hand gesture recognition is a field of computer vision and human-computer interaction that focuses on the development of algorithms and systems capable of interpreting and understanding gestures made by the human hand. These gestures can be used for various applications, including sign language recognition, virtual reality interactions, robotics control, and more. Here is a theoretical background of hand gesture recognition.

4.1.1. Image and video processing

Hand gesture recognition typically begins with the acquisition of image or video data. In most cases, this involves using cameras and sensors to capture the hand's movement and appearance [30]. Image and video processing techniques, such as image filtering, segmentation, and feature extraction, are often applied to isolate and enhance the hand region in the captured frames.

4.1.2. Feature extraction

Extracting relevant features from the hand image is a crucial step. Features can be geometric, appearance-based, or a combination of both. Geometric features may include hand shape, finger positions, and joint angles. Appearance-based features could involve color histograms, texture descriptors, or even deep learning-based representations [31].

4.1.3. Gesture representation

The extracted features are used to represent the gestures in a numerical or symbolic form. This representation allows for the comparison and recognition of different gestures. Common approaches include using vectors, templates, histograms, or neural network embeddings to represent gestures [32].

4.1.4. Gesture recognition algorithms

Various machine learning (ML) and computer vision algorithms can be employed for gesture recognition. Traditional ML techniques like SVMs, decision trees, and k-nearest neighbors (KNN) can be used. Deep learning approaches, particularly CNNs and recurrent neural networks (RNNs), have gained popularity due to their ability to learn complex representations from data [33].

4.1.5. Gesture classification

Gesture classification involves assigning a label or identity to a recognized gesture based on the extracted features and the trained model [34]. The hand landmark model bundle detects the keypoint localization of 20 hand-knuckle coordinates within the detected hand regions. The model was trained on approximately 30 K real-world images, as well as several rendered synthetic hand models imposed over various backgrounds. Figure 3 show the hand landmarker model bundle contains palm detection model and hand landmarks detection model. Palm detection model localizes the region of hands from the whole input image, and the hand landmarks detection model finds the landmarks on the cropped hand image defined by the palm detection model.

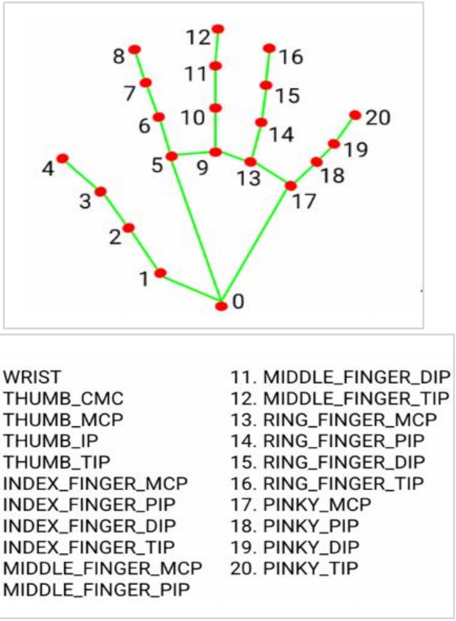

Figure 3. 20 Hand – knuckle coordinates

4.1.6. Real-time processing

For many practical applications, real-time processing is essential. This requires efficient algorithms and optimizations to ensure low latency in recognizing gestures. Figure 4 show frames demonstrate the gesture phases projected by Dewangan *et al*. [10], [11].

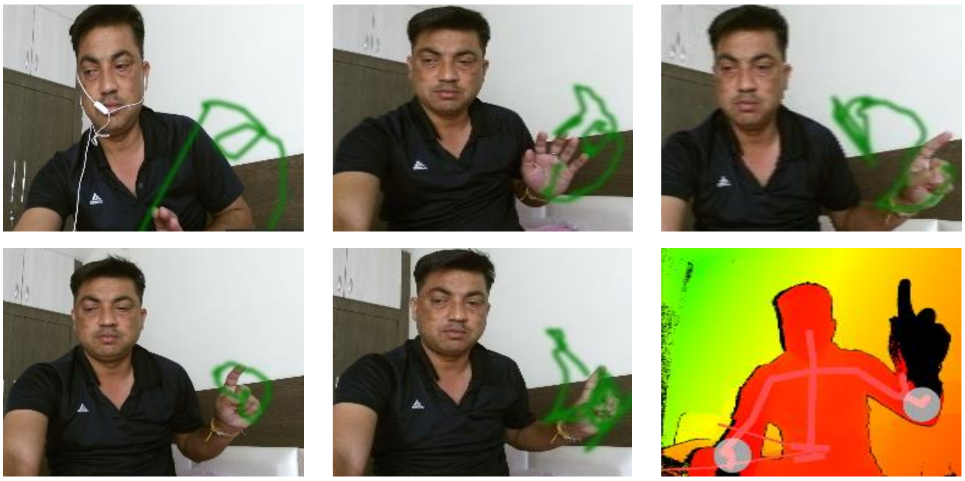

Figure 4. A gesture that represents the concept of “distortion” through selected images

4.2. Review of ML based techniques for continuous hand gesture segmentation and recognition

Krueger [20] applied the inductive MLP [21], [22], which is a supervised learning technique, for addressing the signal unit division issue. The back-spread strategy is executed utilizing angle plunge, and a versatile learning rate. Quan [23] demonstrated signal stage division as an issue of characterization, and utilized SVM to plan a model to get familiar with the motion designs of each phase. The work mainly addressed the limitations of the segmentation approach due to human behavior and conducted analysis by considering the standpoint of linguistics and psycholinguistics specialists. Cao *et al.* [24] demonstrated the issue as a characterization task, and applied SVM. The work exploited the transient parts of the issue and utilized a few kinds of information pre-preparing to consider time and recurrence area highlights. Sturman and Zeltzer [25] presents a survey about fleeting parts of hand motion examination, concentrating on applications identified with normal discussion and psycholinguistic investigation. Mitra and Acharya [26] constructed three separate identification models using different training techniques: First, a linguistic model using an empirical language model; second, a signal model using a Bayesian or the CART system selection tree; and third, a language model with a Bernoulli tree of choices. Then, in order to incorporate the outputs from these modules and provide finding results, the hidden Markov model also called the HMM, is used.

Few such inquiries about examinations are not straightforwardly practically identical. However, it gets helpful to analyses the exhibitions which were at that point achieved in such sort of issue. The outcomes are recorded in Table 1.

Table 1. Features and resulting eye vector

Location of pate	x	y	z

5. METHOD

The entire experiment’s steps are shown in Figure 5. A few analyses have been done to assess and improve the exhibition of the models worked with deep learning systems and utilizing the information portrayals depicted by including different parameter. Two arrangements of investigations have been concurred specifically for this investigation. In the first set, trials are conducted using a straightforward MLP decoder powered by AI. The suggested supervised deep learning network with the kernel neural network re-sampling approach is utilized in the final set of experiments. The recommended method utilises several parameters.

Key parameters:

- $IAccuracy$: these measures the accuracy of an organization depiction and can be presented as:

$$IAccuracy = 2(T\_P + T\_N)/2(T\_P + T\_N + F\_P - F\_N) \quad (9)$$

- $IPrecision$: over all favourable findings, it is the part of real results and is displayed as.

$$IPrecision = 2 * T\_P \;/\; 2(T\_P + F\_P) \tag{10}$$

- $IRecall$: it is a component of each correct result the model has returned.

$$IRecall = \; 2T\_P/2(T\_P + F\_N) \tag{11}$$

- IF_score: it is calculated as the weighted average based on accuracy and review. Its value ranges from 0 to 1, with 1 being the best result.

$$IF\_score = 2.0 * (IPrecision * IRecall) \;/\; (IPrecision + IRecall) \tag{12}$$

Tables 2 and 3 shows the outcomes of tests done using the deep learning networks (DLN) framework and neural networks (NN), respectively. Employing the thinking processes described in the present article, comparing the levels of accuracy, average precision, and recall in all situations. Here are the results shown in graphs for each context from the two experiment sets. Using the recommended framework, the highest achievable accuracy in U3, V3, and T3 is, respectively, 93, 86, and 84. The average accuracy gain across all situations was 18%, which is a substantial rise above previous works.

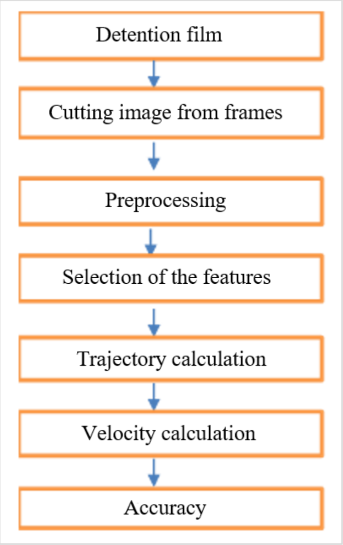

Figure 5. Steps of experiment

Table 2. Results experiment set-I

Contexts	Prediction accuracy (%)	Avg. precision	Avg. recall	F-score				
				D	S	R	P	H
A1	74	0.74	0.74	0.86	0.80	0.42	0.36	0.30
A2	64	0.64	0.64	0.79	0.69	0.26	0.46	0.33
A3	**84**	0.82	0.82	**0.93**	**0.85**	**0.68**	0.67	**0.80**
B1	80	0.79	0.79	0.83	0.78	0.62	**0.82**	0.85
B3	60	0.60	0.60	0.70	0.60	0.48	0.63	0.54
C1	62	0.61	0.61	0.64	0.62	0.59	0.5	0.64
C3	55	0.55	0.55	0.81	0.52	0.53	0.52	0.33

Table 3. Comparison of accuracy of NN and proposed DLN based framework

Context	D in NN	D in DLN
A1	0.86	0.92
A2	0.79	0.8
A3	0.93	0.9
B1	0.83	0.94
B3	0.7	0.95
C1	0.65	0.72
C3	0.81	0.83

6. RESULT AND DISCUSSION

Three people said: videos A, B, and C were recorded with Microsoft Xbox Kinect sensors to allow for different gestural behaviors. For A, three videos A1, A2 and A3 were recorded. Gesture behavior affects the classification performance generated for segmentation, so videos A1 and A2 were recorded in similar sessions, while A3 differs. Likewise, B1, B3, C1 and C3 were recorded in different sessions to obtain different gestural behaviors. All seven of these films are referenced here in context. The 3D coordinates (x; y; z) of hand movements were extracted from each image using software based on the Microsoft Kinect sensor [5]-[7]. The associated timestamps are also taken, as this is the functional way to maintain the entity tagging process. To obtain trajectory information, numerical velocity, and acceleration relative to manual activity were also calculated. The captured and exported properties are shown in Table 1. Table 4 shows the number of material types and their attributes before and after the extraction phase. The dataset is highly unnecessary and a dissimilar motions ratio is obtainable in Table 4.

Table 4. Class circulation in different contexts

Parameter	A	B	C	D	E	Borders
T1	94	56	91	63	29	173
T2	89	31	20	86	44	120
T3	58	35	79	28	10	180
U1	12	87	17	23	90	109
U3	69	20	21	10	10	140
V1	82	62	26	13	14	117
V3	89	58	38	25	14	144

7. CONCLUSION

Gesture segmentation and recognition has several inherent difficulties, as it does not indicate a clear starting point for the phase. Therefore, different segments of the same input video can be presented to different researchers. There is also difficulty establishing a resting position and maintaining posture. To better understand the classifier and its performance, the gesture behavior should be recorded in different sessions. We develop a framework that addresses three related questions. Experimentation and evaluation are performed by detecting, segmenting, and recognizing hand movements in videos. After resampling the image using a KNN-based method, a deep learning network was used to perform gesture recognition, achieving better accuracy than other base learning algorithms. It turns out that interesting motion embedded in a video stream can be easily learned and recognized by frame resampling. The performance of the framework is evaluated based on several metrics, including F-score and classification accuracy. We also compare the recital of this framework with recently proposed accepted works. We face the challenge of using deep learning algorithms based on spatiotemporal and path information in C_H__S_R. Finally, this work raises open questions for researchers about simultaneous segmentation and recognition at different stages or the definition of important gestures.

REFERENCES

[1] C. Yang, D. K. Han, and H. Ko, "Continuous hand gesture recognition based on trajectory shape information," *Pattern Recognition Letters*, vol. 99, pp. 39–47, Nov. 2017, doi: 10.1016/j.patrec.2017.05.016.

[2] V. Bhame, R. Sreemathy, and H. Dhumal, "Vision based hand gesture recognition using eccentric approach for human computer interaction," in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Sep. 2014, pp. 949–953, doi: 10.1109/ICACCI.2014.6968545.

[3] C. Keskin, F. Kıraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Consumer Depth Cameras for Computer Vision*, 2011, pp. 119–137, doi: 10.1007/978-1-4471-4640-7_7.

[4] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, "Robust hand gesture recognition with kinect sensor," in *MM'11-Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops*, 2011, pp. 759–760, doi: 10.1145/2072298.2072443.

[5] R. C. B. Madeo, S. M. Peres, and C. A. de M. Lima, "Gesture phase segmentation using support vector machines," *Expert Systems with Applications*, vol. 56, pp. 100–115, Sep. 2016, doi: 10.1016/j.eswa.2016.02.021.

[6] M.-C. Popescu, E. V. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Transactions on Circuits and Systems*, 2009, doi: 10.5555/1639537.1639542.

[7] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff, "A unified framework for gesture recognition and spatiotemporal gesture segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1685–1699, 2009, doi: 10.1109/TPAMI.2008.203.

[8] B. K. Dewangan, A. Jain, R. N. Shukla, and T. Choudhury, "An ensemble of bacterial foraging, genetic, ant colony and particle swarm approach EB-GAP: a load balancing approach in cloud computing," *Recent Advances in Computer Science and Communications*, vol. 15, no. 5, 2020, doi: 10.2174/2666255813666201218161955.

[9] B. K. Dewangan, A. Agarwal, M. Venkatadri, and A. Pasricha, "A self-optimization based virtual machine scheduling to workloads in cloud computing environment," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 4, pp. 91–96, 2019.

[10] B. K. Dewangan, A. Agarwal, M. Venkatadri, and A. Pasricha, “SLA-based autonomic cloud resource management framework by Antlion optimization algorithm,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 4, pp. 119–123, 2019.

[11] B. K. Dewangan and P. Shende, “The sliding window method: an environment to evaluate user behavior trust in cloud technology,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 2, pp. 1158-1162, 2013.

[12] S. K. Baghel and B. K. Dewangan, “Defense in depth for data storage in cloud computing,” *International Journal of Technology*, vol. 2, no. 2, pp. 58–61, 2012.

[13] B. K. Dewangan, M. Venkatadri, A. Agarwal, A. Pasricha, and T. Choudhury, “An automated self-healing cloud computing framework for resource scheduling,” *International Journal of Grid and High Performance Computing*, vol. 13, no. 1, pp. 47–64, 2021, doi: 10.4018/IJGHPC.2021010103.

[14] B. K. Dewangan, A. Agarwal, T. Choudhury, and A. Pasricha, “Workload aware autonomic resource management scheme using grey wolf optimization in cloud environment,” *IET Communications*, vol. 15, no. 14, pp. 1869–1882, 2021, doi: 10.1049/cmu2.12198.

[15] T. Choudhury, B. K. Dewangan, R. Tomar, B. K. Singh, T. T. Toe, and N. G. Nhu, “Autonomic computing in cloud resource management in industry 4.0,” *EAI/Springer Innovations in Communication and Computing*, 2021.

[16] L. Chen, M. P. Harper, Y. Liu, and E. Shriberg, “Multimodal model integration for sentence unit detection,” *ICMI’04-Sixth International Conference on Multimodal Interfaces*, 2004, pp. 121–128, doi: 10.1145/1027933.1027955.

[17] S. Haykin, *Neural networks and learning machines*, New Jersey: Pearson Prentice Hall, 2008.

[18] M. A. Nielsen, *Neural networks and deep learning*, Determination press San Francisco, CA, USA, 2015.

[19] P. K. Wagner, R. C. Madeo, S. M. Peres, and C. A. Lima, “Segmentation of gestural units with multilayer perceptrons (In Portuguese),” *Conference: X Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, 2013.

[20] M. W. Krueger, “Artificial reality II,” *Reading (Mass): Addison-Wesley*, p. 304, 1991.

[21] W. Fan *et al.*, “A method of hand gesture recognition based on multiple sensors,” *2010 4th International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2010*, 2010, doi: 10.1109/ICBBE.2010.5516722.

[22] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, “A framework for hand gesture recognition based on accelerometer and EMG sensors,” *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, 2011, doi: 10.1109/TSMCA.2011.2116004.

[23] Y. Quan, “Chinese sign language recognition based on video sequence appearance modeling,” *Proceedings of the 2010 5th IEEE Conference on Industrial Electronics and Applications, ICIEA 2010*, 2010, pp. 1537–1542, doi: 10.1109/ICIEA.2010.5514688.

[24] X. Y. Cao, H. F. Liu, and Y. Y. Zou, “Gesture segmentation based on monocular vision using skin color and motion cues,” *IASP 10-2010 International Conference on Image Analysis and Signal Processing*, 2010, pp. 358–362, doi: 10.1109/IASP.2010.5476096.

[25] D. J. Sturman and D. Zeltzer, “A survey of glove-based input,” *IEEE Computer graphics and Applications*, vol. 14, no. 1, pp. 30–39, 1994, doi: 10.1109/38.250916.

[26] S. Mitra and T. Acharya, “Gesture recognition: a survey,” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, 2007, doi: 10.1109/TSMCC.2007.893280.

[27] Y. Wu and T. S. Huang, “Vision-based gesture recognition: a review,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1739, pp. 103–115, 1999, doi: 10.1007/3-540-46616-9_10.

[28] Y. Hamada, N. Shimada, and Y. Shirai, “Hand shape estimation using sequence of multi-ocular images based on transition network,” *Proceedings of the International Conference on Vision Interface*, 2002, pp. 161–166.

[29] N. Tanibata, N. Shimada, and Y. Shirai, “Extraction of hand features for recognition of sign language words,” *The 15th International Conference on Vision Interface*, 2002, pp. 391–398.

[30] Y. Wu and T. S. Huang, “Nonstationary color tracking for vision-based human-computer interaction,” *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 948–960, 2002, doi: 10.1109/TNN.2002.1021895.

[31] C. Tomasi, S. Petrov, and A. Sastry, “3D tracking=classification+interpolation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2003, vol. 2, pp. 1441–1448, doi: 10.1109/iccv.2003.1238659.

[32] G. Ye, J. J. Corso and G. D. Hager, “Gesture recognition using 3D appearance and motion features,” *2004 Conference on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA, 2004, pp. 160-160, doi: 10.1109/CVPR.2004.356.

[33] J. Y. Lin, Y. Wu, and T. S. Huang, “3D Model-based hand tracking using stochastic direct search method,” in *Proceedings-Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 693–698, doi: 10.1109/AFGR.2004.1301615.

[34] R. Aggarwal, S. Swetha, A. M. Namboodiri, J. Sivaswamy, and C. V. Jawahar, “Online handwriting recognition using depth sensors,” *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2015, pp. 1061–1065, doi: 10.1109/ICDAR.2015.7333924.

BIOGRAPHIES OF AUTHORS

Prashant Richhariya received his bachelor degree in computer science and engineering from Chhatrapati Shivaji Institute of Technology, Durg, and his master of technology in computer science and engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal and Pursuing Ph.D. from University of Petroleum and Energy Studies, Dehradun. He is in IT industry full-time. His research lines are in machine learning and digital image processing. He can be contacted at email: prashant1579@gmail.com.

Piyush Chauhan graduated from IEET now Baddi University. He received the, M.Tech., computer science engineering and the Ph.D. degree from Jaypee University of Information Technology. He worked as an assistant professor in computer science and engineering at University of Petroleum and Energy Studies Dehradun. He has supervised over 30 master thesis and 6 Ph.D. dissertations. He received several prizes for the acknowledgment of his outstanding research and teaching performance. He can be contacted at email: pchauhan@ddn.upes.ac.in.

Lalit Kane graduated B.E., (computer science and engineering), Barkatullah University, M.Tech., information technology (artificial intelligence), Rajiv Gandhi Technological University, Ph.D. (computer science and engineering), (real-time static and dynamic hand gesture recognition using depth data), Indian Institute of Information Technology, Design, and Manufacturing Jabalpur. GATE in 2011 and 2004. Area of specialization computer vision, image processing, human-computer interaction member of IEEE. He has supervised over 35 master thesis and 10 Ph.D. dissertations. He can be contacted at email: lalit.kane@ddn.upes.ac.in.

Bhupesh Kumar Dewangan graduated his degree of B.E. in computer science and engineering from Pt. Ravi Shankar Shukla. His M.Tech. in computer science and engineering from Chhattisgarh Swami Vivekananda Technical University, Bhilai 2012. And his Ph.D. in computer science and engineering from University of Petroleum and Energy Studies, Dehradun 2021. Area of expertise cloud computing, high performance computing and resource management autonomic computing informatics. He is working as associate professor in OP Jindal University. He can be contacted at email: bhupesh.dewangan@gmail.com.