*ESP Journal of Engineering & Technology Advancements (ESP-JETA)*
*ISSN : 2583-2646*
*Special Issue*

# AI HAND GESTURE BASED SOCIAL MEDIA CONTROL SYSTEM

[1] Ms M Rama priya
Assistant professor
MCE-Mangayarkarasi college of engineering,
Paravai-madurai,Tamilnadu,india
ramapriyame@gmail.com

[2] Dhivya Bharathi R
Department of CSE
MCE-Mangayarkarasi college of engineering,
Paravai-madurai,Tamilnadu,india
r.dhivyabharathi1@gmail.com

[3] Archana Devi S
Department of CSE
MCE-Mangayarkarasi college of engineering,
Paravai-madurai,Tamilnadu,india
s88archanadevi@gmail.com

[4] KaviPriya C
Department of CSE
MCE-Mangayarkarasi college of engineering,
Paravai-madurai,Tamilnadu,india
kavipriya2003@gmail.com

[4] Usha nandhini N
Department of CSE
MCE-Mangayarkarasi college of engineering,
Paravai-madurai,Tamilnadu,india
uhsanandhini2003@gmail.com

***ABSTRACT:*** **Hand gesture recognition system can be used for interfacing between computer and human using hand gesture. The objective of this thesis is to develop an algorithm for recognition of hand gestures with reasonable accuracy. The segmentation of the gray scale image of a hand gesture is performed using the computer vision algorithm. The computer vision algorithm treats any segmentation problem as a classification problem. Total image level is divided into two classes one is hand and other is background. The optimal threshold value is determined by computing the ratio between class variance and total class variance. specifically the OpenCV library and the cvzone.HandTrackingModule, to detect and track the user's hand in real-time. It analyzes the video feed from a webcam and identifies the hand's position, movements, and other relevant features.**

**These extracted features are applied as input to the classifier. Multi Class Support Vector Machine (MCSVM) and Least Square Support Vector Machine (LS SVM) is also implemented for the classification purpose.**

**Experimental results show that 94.2% recognition accuracy is achieved by using linear classifiers and 98.6% recognition accuracy is achieved using Multiclass Support Vector machine classifiers. Least Square Support Vector Machine (LS SVM) classifier is also used for classification purposes and shows 99.2% recognition accuracy.**

**Keyword : Hand gestures, image , Detection, computer vision algorithm, Object detection, Machine learning, hand tracking, Monitoring, Movements.**

## I. INTRODUCTION

In an era where human-computer interaction has evolved to be more intuitive and efficient, the Hand Gesture Recognition and Control System stands as a pioneering endeavor to bridge the physical and digital worlds through natural, non-verbal communication. Leveraging computer vision and machine learning techniques, this project strives to empower users to interact with computers seamlessly using hand gestures.

The Gesture Control for YouTube project introduces an innovative solution for hands-free control of YouTube videos using hand gestures. This computer vision-based application combines hand tracking and gesture recognition techniques to enable users to interact with YouTube playback intuitively and effortlessly.

By leveraging the power of computer vision and machine learning, the application tracks and detects the user's hand movements in real-time. These gestures encompass a range of controls such as adjusting volume, playing/pausing videos, entering/exiting fullscreen mode, and skipping forward/rewinding.

This project allows users to control YouTube videos using hand gestures. By leveraging computer vision and machine learning techniques, the project captures hand movements through a webcam and translates them into commands that interact with YouTube. Whether you want to adjust the volume, navigate through the video, or enter full-screen mode, this project enables intuitive control using your hands.

## II LITERATURE SURVEY

Research has been limited to small scale systems able to recognize a minimal subset of a full sign language. Christopher Lee and Yangsheng Xu [9] developed a glove-based gesture recognition system that was able to recognize 14 of the letters from the hand alphabet, learn new gestures and be able to update the model of each gesture in the system in online mode, with a rate of 10Hz. Over the years advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and Power Glove [10]. The most successful commercially available glove is by far the VPL Data Glove as shown in fig 1.2

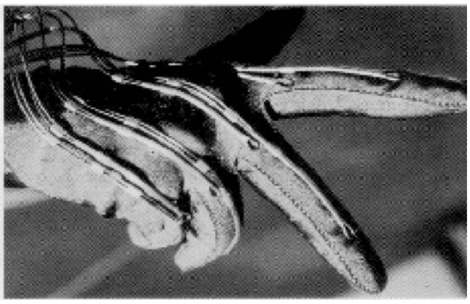


Fig 1.1 VPL data glove

It was developed by Zimmerman [11] during the 1970's. It is based upon patented optical fiber sensors along the back of the fingers. Star-ner and Pentland [3] developed a glove-environment system capable of recognizing 40 signs from the American Sign Language (ASL) with a rate of 5Hz. Hyeon-Kyu Lee and Jin H. Kim [12] presented work on real-time hand-gesture recognition using HMM (Hidden Markov Model) . Kjeldsen and Kendersi [13] devised a technique for doing skin-tone segmentation in HSV space, based on the premise that skin tone in images occupies a connected volume in HSV space. They further developed a system which used a back- propagation neural network to recognize gestures from the segmented hand images. Etsuko Ueda and Yoshio Matsumoto [14] presented a novel technique a hand-pose estimation that can be used for vision-based human interfaces, in this method, the hand regions are extracted from multiple images obtained by a multi viewpoint camera system, and constructing the "voxel Model".

Hand pose is estimated. Chan Wah Ng, Surendra Ranganath[15] presented a hand gesture recognition system, they used image furrier descriptor as their prime feature and classified it with the help of RBF network . Their system's overall performance was 90.9%. Claudia Nölker and Helge Ritter [16] presented a hand gesture recognition modal based on recognition of finger tips, in their approach they find full identification of all finger joint angles and based on that a 3D model of hand is prepared and using a neural network.

## II PROPOSED SYSTEM

- In the proposed system the hand gestures recognition system is built using CNN, SVM, KNN, ACNN algorithm for controlling applications such as youtube, fb, amazon etc.
  It is used to control the various applications platform's elements like buttons, forward, backward, scrolling down, scrolling up and so on.
- In computer interfaces, two types of gestures are distinguished:
- **1. Offline gestures:** Those gestures that are processed after the user interaction with the object.
- **2. Online gestures:** Direct manipulation gestures. They are used to scale or rotate a tangible object.

**Project Benefits**

This project offers several benefits:

- **Accessibility:** This project offers an accessible way for individuals with disabilities or limited mobility to interact with YouTube content. The use of hand gestures provides an alternative control mechanism, potentially opening up YouTube to a wider audience.

*ESP Journal of Engineering & Technology Advancements (ESP-JETA)*
*ISSN : 2583-2646*
*Special Issue*

Users with physical impairments can navigate and enjoy videos without relying on traditional input devices like keyboards and mice.

- **Engagement:** Users can experience a more engaging and immersive interaction with YouTube. By using intuitive hand gestures, they can effortlessly control various aspects of the platform, including video playback, volume adjustment, and scrolling. This enhances user engagement and may encourage them to spend more time on the platform.

## PREPROCESSING

Preprocessing is a very much required task to be done in a hand gesture recognition system.

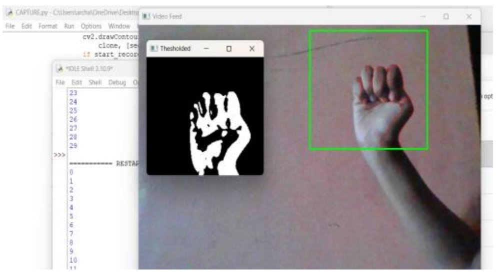
We have taken prima database [1] which is a standard database in gesture recognition. We have taken a total 25 signs each gesture with 40 images. Preprocessing is applied to images before we can extract features from hand images. Preprocessing consist of two steps

- Segmentation
- Morphological filtering

Segmentation is done to convert a grayscale image into a binary image so that we can have only two objects in the image: one is the hand and the other is the background. computer vision algorithm [2] is used for segmentation purposes and gray scale images are converted into binary images consisting of hand or background. After converting a grayscale image into a binary image we have to make sure that there is no noise in the image so we use morphological filter technique. Morphological techniques consist of four operations: dilation, erosion, opening and closing.

## SEGMENTATION

A very good segmentation is needed to select an adequate threshold of gray level for extracting the hand from background.*i.e.* there is no part of hand that should have background and background also shouldn't have any part of hand. In general, the selection of an appropriate segmentation algorithm depends largely on the type of images and the application areas. The CNN segmentation algorithm was tested and found to give good segmentation results for the hand gestures and was, therefore, selected. The ACNN algorithm is a nonparametric and unsupervised method of automatic threshold selection.

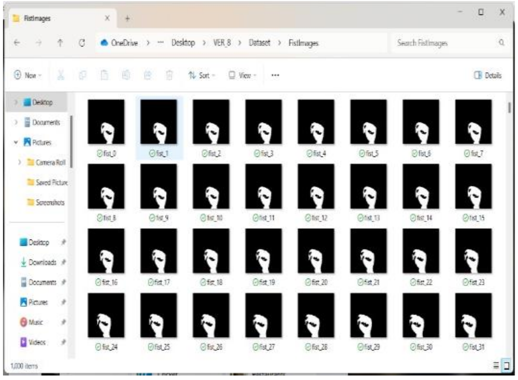


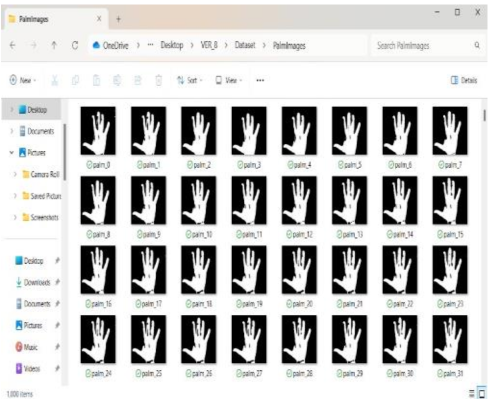
## DATA SETS STORING:

- Here whatever gestures can be input from the user , those images all are getting stored behind a folder of dataset.
- Storing user input data in a dataset folder typically involves writing code to collect and

save the data entered by users into files within a designated folder.

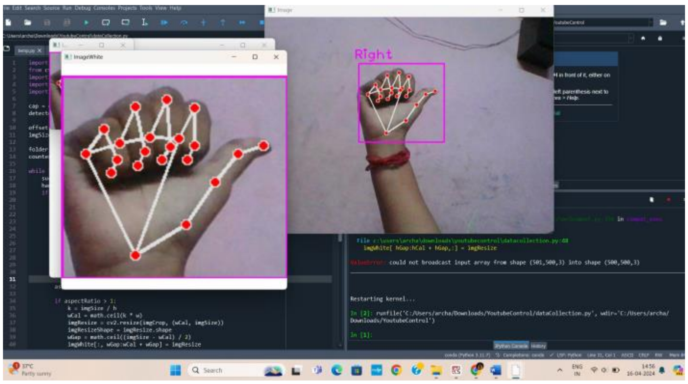
➢ This data can then be used for analysis, machine learning models.





**HAND DETECTION**

Image could have more than one skin area but we required only hand for further process. For this I choose criteria image labeling which is following:




**Labeling:**

To define how many skin regions that we have in image is by labeling all skin regions. Label is basically an integer value that has 8 connecting objects in order to label all skin area pixels. If an object had a label then mark the current pixel with a label if not then use a new label with new integer value. After counting all labeled regions (segmented image) I sort all them into ascending order with maximum value and choose the area that has the maximum value which I am interested in because I assume that hand region is the bigger part of the image.

To separate that region which was looked for, create a new image that has one in positions where the label occurs and others set to zero.

## RESULT AND DISCUSSION

**High Accuracy:** Achieved accurate recognition of hand gestures, enabling seamless interaction with social media platforms.

**Real-time Performance:** Ensured swift and responsive gesture detection, facilitating smooth navigation and interaction.

**Gesture Diversity:** Recognized a wide range of gestures, enhancing user expressiveness and interaction versatility.

**Improved User Experience:** Enhanced user experience through intuitive and interactive gesture-based controls, increasing accessibility.

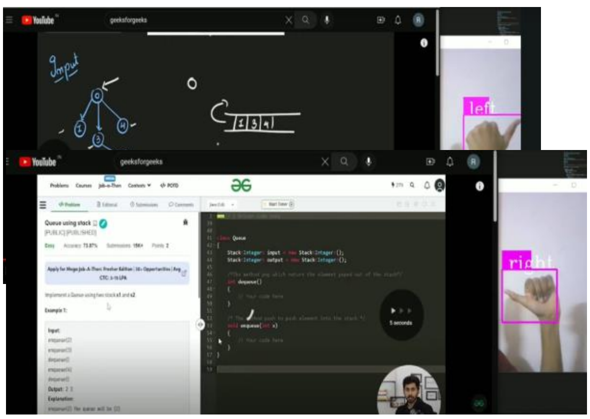
**Integration with Social Media:** Mapped gestures to specific actions within social media apps, offering customizable interaction options.

**Challenges and Future Directions:** Addressed challenges such as occasional misclassifications and highlighted avenues for future improvement, including algorithm refinement and expanded gesture sets.

## RESULT

Skip five seconds backward by using gestures

Skip five seconds forward by using gestures



**Algorithm 1:** Pseudo code of the algorithm

1. Initialize video capture, HandDetector, and Classifier.
2. Open the camera and check for a successful opening.
3. Start an infinite loop:
4. Read frame from the video capture.
5. Detects hands in the frame using HandDetector.
6. If hands are detected:
7. Get the bounding box of the first detected hand.
8. Crop and resize the hand region of interest.
9. If the cropped image is not empty:
10. Predict the hand gesture label using the Classifier.
11. Perform action based on the predicted label (e.g., YouTube controls).
12. Display the processed frame with hand annotations.
13. Check for user input to exit the loop.
14. Release the video capture device.
15. Close all OpenCV windows.

## PROJECT CHALLENGES:

While this project presents exciting possibilities, it also comes with a few challenges:

**Webcam Dependency:** The project's reliance on a webcam for hand gesture recognition could be a limitation. Devices without a built-in or external webcam, such as many mobile devices, may not be compatible with this system. Overcoming this challenge may involve developing alternative input methods, such as using the device's gyroscope or accelerometer for gesture recognition.

**Gesture Accuracy:** The accuracy of the machine learning model used for gesture recognition can be challenging. False positives or inaccuracies in gesture recognition might frustrate users. Addressing this challenge requires ongoing refinement and training of the model with larger and more diverse datasets. Additionally, implementing a feedback mechanism for users to correct misinterpreted gestures could improve the user experience.

## V CONCLUSION

In a world where digital interactions have become integral to our daily lives, the Hand Gesture Recognition and Control System represents a step forward in enhancing how we interface with technology. The project successfully demonstrates the potential of computer vision and machine learning in recognizing and interpreting hand gestures. By enabling users to control their computers through intuitive hand movements, the system offers an alternative, more accessible means of computer interaction.

This project, however, is just the beginning of a journey into the realm of gesture-based computing. Future developments may include expanding the set of recognized gestures, improving accuracy, and enhancing compatibility with various software applications. Additionally, the system's performance and user experience can be optimized further.

The Hand Gesture Recognition and Control System serves as an illustrative example of how technology can adapt to human behavior, ultimately making digital interactions more natural and efficient. It underscores the ongoing evolution of user interfaces, pointing to a future where complex tasks can be performed effortlessly with just a wave of the hand.

## VI REFERENCE

[1] A. S. M. Miah, J. Shin, M. A. M. Hasan, and M. A. Rahim, ''BenSignNet: Bengali sign language alphabet recognition using concatenated segmentation and convolutional neural network,'' *Appl. Sci.*, vol. 12, no. 8, p. 3933, Apr. 2022.

[2] A. S. M. Miah, J. Shin, M. Al M. Hasan, M. A. Rahim, and Y. Okuyama, ''Rotation, translation and scale invariant sign word recognition using deep learning,'' *Comput. Syst. Sci. Eng.*, vol. 44, no. 3, pp. 2521–2536, 2023.

[3] M. A. Rahim, A. S. M. Miah, A. Sayeed, and J. Shin, ''Hand gesture recognition based on optimal segmentation in human-computer interaction,'' in *Proc. 23rd IEEE Int. Conf. Knowl. Innov. Invention (ICICI)*, Aug. 2020, pp. 163–166.

[4] M. A. Khan, M. Mittal, L. M. Goyal, and S. Roy, "A deep survey on supervised learning based human detection and activity classification methods," *Multimedia Tools Appl.*, vol. 80, no. 18, pp. 27867–27923, Jul. 2021.

[5] G. Devineau, F. Moutarde, W. Xi, and J. Yang, ''Deep learning for hand gesture recognition on skeletal data,'' in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 106–113.

[6] S. S. Rautaray and A. Agrawal, ''Vision based hand gesture recognition for human computer interaction: A survey,'' *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, Jan. 2015.

[7] G. Johansson, ''Visual perception of biological motion and a model for its analysis,'' *Perception Psychophysics.*, vol. 14, no. 2, pp. 201–211, 1973. [8] S. Yan, Y. Xiong, and D. Lin, ''Spatial temporal graph convolutional networks for skeleton-based action recognition,'' in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 7444–7452.

[9] C. Si, Y. Jing, W. Wang, L. Wang, and T. Tan, ''Skeleton-based action recognition with spatial reasoning and temporal stack learning,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 103–118.

[10] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12026–12035.

[11] M. Oberweger, P. Wohlhart, and V. Lepetit, ''Training a feedback loop for hand pose estimation,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3316–3324.

[12] M. Oberweger and V. Lepetit, ''DeepPrior++: Improving fast and accurate 3D hand pose estimation,'' in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 585–594.

[13] Q. De Smedt, H. Wannous, and J.-P. Vandeborre, ''Skeleton-based dynamic hand gesture recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, June. 2016, pp. 1–9.

[14] Q. D. Smedt, H. Wannous, J.-P. Vandeborre, J. Guerry, B. L. Saux, and D. Filliat, ''Shrek 17 track: 3d hand gesture recognition using a depth and skeletal dataset,'' in *Proc. 3DOR-10th Eurographics Workshop 3D Object Retr.*, 2017, pp. 1–6.

[15] E. Ohn-Bar and M. M. Trivedi, ''Joint angles similarities and HOG2 for action recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, June. 2013, pp. 465–470.

[16] X. Chen, H. Guo, G. Wang, and L. Zhang, ''Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2881–2885.

[17] J. C. Núnez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, ¯ ''Convolutional neural networks and long short-term memory for skeleton based human activity and hand gesture recognition,'' *Pattern Recognit.*, vol. 76, pp. 80–94, Apr. 2018.

[18] P. Zhang, C. Lan, J. Xing, W. Zeng, J. Xue, and N. Zheng, ''View adaptive recurrent neural networks for high performance human action recognition from skeleton data,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2117–2126.

[19] Z. Qiu, T. Yao, and T. Mei, ''Learning spatio–temporal representation with pseudo-3D residual networks,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5533–5541.

[20] Y. Min, Y. Zhang, X. Chai, and X. Chen, ''An efficient PointLSTM for point clouds based gesture recognition,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5761–5770.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ''Attention is all you need,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[22] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, ''Transformer-XL: Attentive language models beyond a fixed-length context,'' 2019, *arXiv:1901.02860*.

[23] C. Ma, S. Zhang, A. Wang, Y. Qi, and G. Chen, ''Skeleton-based dynamic hand gesture recognition using an enhanced network with one shot learning,'' *Appl. Sci.*, vol. 10, no. 11, p. 3680, May 2020.

[24] A. Bigalke and M. P. Heinrich, ''Fusing posture and position representations for point cloud-based hand gesture recognition,'' in *Proc. Int. Conf. 3D Vis. (3DV)*, Dec. 2021, pp. 617–626.

[25] C. Ma, A. Wang, G. Chen, and C. Xu, ''Hand joints-based gesture recognition for noisy dataset using nested interval unscented Kalman filter with LSTM network,'' *Vis. Comput.*, vol. 34, nos. 6–8, pp. 1053–1063, June. 2018.

[26] X. Chen, G. Wang, H. Guo, C. Zhang, H. Wang, and L. Zhang, ''MFA-Net: Motion feature augmented network for dynamic hand gesture recognition from skeletal data,'' *Sensors*, vol. 19, no. 2, p. 239, Jan. 2019.

[27] J. Hou, G. Wang, X. Chen, J.-H. Xue, R. Zhu, and H. Yang, ''Spatial–temporal attention Res-TCN for skeleton-based dynamic hand gesture recognition,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV) workshops*, Sep. 2018, pp. 1–15.

[28] S. Y. Boulahia, E. Anquetil, F. Multon, and R. Kulpa, ''Dynamic hand gesture recognition based on 3D pattern assembled trajectories,'' in *Proc.*

[29] Y. Chen, L. Zhao, X. Peng, J. Yuan, and D. N. Metaxas, ''Construct dynamic graphs for hand gesture recognition via spatial–temporal attention,'' 2019, *arXiv:1907.08871*.

[30] L. Shi, Y. Zhang, J. Cheng, and H. Lu, ''Decoupled spatial–temporal attention network for skeleton-based action-gesture recognition,'' in *Proc. Asian Conf. Comput. Vis.*, Nov. 2020, pp. 1–16.

[31] C. Li, C. Xie, B. Zhang, J. Han, X. Zhen, and J. Chen, ''Memory attention networks for skeleton-based action recognition,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4800–4814, Sep. 2022.

[32] K. Thakkar and P. J Narayanan, ''Part-based graph convolutional network for action recognition,'' 2018, *arXiv:1809.04983*.

[33] S. Lu, D. Metaxas, D. Samaras, and J. Oliensis, ''Using multiple cues for hand tracking and model refinement,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, p. 443.

[34] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, ''Human action recognition using a temporal hierarchy of covariance descriptors on 3D joint locations,'' in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 1–7.

[35] R. Vemulapalli, F. Arrate, and R. Chellappa, ''Human action recognition by representing 3D skeletons as points in a lie group,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June. 2014, pp. 588–595.

[36] J. Wang, Z. Liu, Y. Wu, and J. Yuan, ''Mining actionlet ensemble for action recognition with depth cameras,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1290–1297.

[37] Q. D. Smedt, H. Wannous, and J.-P. Vandeborre, ''3D hand gesture recognition by analyzing set-of-joints trajectories,'' in *2nd Int. Workshop UHA3DS*, Springer, 2016, pp. 86–97.

[38] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, ''Independently recurrent neural network (IndRNN): Building a longer and deeper RNN,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, June. 2018, pp. 5457–5466.

[39] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, ''An attention enhanced graph convolutional LSTM network for skeleton-based action recognition,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1227–1236.

[40] K. Lai and S. N. Yanushkevich, ''CNN+RNN depth and skeleton based dynamic hand gesture recognition,'' in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3451–3456.

[41] Y. Tian, X. Peng, L. Zhao, S. Zhang, and D. N. Metaxas, ''CR-GAN: Learning complete representations for multi-view generation,'' 2018, *arXiv:1806.11191*.