A Proposed Technique For Recognizing And Analyzing Body Language Using Slantlet Transform And Deep Learning

Ban Amjed Saeed, Ali Mohammed Sahan
Technical College of Management, Middle Technical University
Baghdad, Iraq

DOI: 10.37648/ijps.v19i01.011

[1]*Received: 12/03/2025; Accepted: 05/05/2025; Published: 07/05/2025*

Abstract

Body language encompasses a variety of nonverbal cues, such as head movements, hand movements, body postures, or any other part of the body, from which intentions and feelings can be inferred. The expanding use of body language in various fields has made it necessary for recognition systems to achieve accuracy and efficiency in performance. This paper proposes a two-stage to body language recognition and analysis technique. The first stage is extracting frames from the video, after which the third level of Slantlet Transform is applied to reduce the dimensions of the frames while preserving the essential features and removing unnecessary features, thus reducing processing time. In the second stage, deep learning models are used to extract features and classify, followed by analyzing body language and human behavior in various legal contexts. This mechanism helps reduce processing time while maintaining high accuracy. The Slantlet Transform also makes the technique resistant to rotation and noise, as it removes unwanted features that typically contain noise. A variety of experiments were conducted to test the accuracy and efficiency of the proposed technique using the GEMEP dataset and The Real-Life Deception Detection dataset. The proposed technique achieved an accuracy of 99.56% and 98.22% in recognizing and analyzing body language, respectively, in addition to achieving a reduction in processing time of more than 75%.

Keywords: *Body Language; Body Language Recognition; Analyzing Body Language; Slantlet Transform; Deep Learning; GEMEP dataset; Real Life Deception Detection dataset.*

1. Introduction

Body language (nonverbal language) is interpreted through visible body signals such as eye movements, hand, shoulders, body posture, or any other part of the body to express feelings or impulses. Psychological studies have indicated that the largest part of understanding messages occurs using nonverbal language signals. These signals may be intentional to send a specific message to the other party or unintentional [1]. Most of these signals are innate, and there are several indicators that help in understanding the message of body language, including the nose. Touching the nose while speaking indicates confusion or lying, while looking directly at the other party indicates self-confidence. Dilation of the pupils while speaking indicates lying, while dilation of the eye openings indicates happiness. Therefore, it is necessary when interpreting body language to combine more than one indicator to obtain an accurate interpretation. This makes the process of recognizing emotions applicable to different human groups around the world.

[1] *How to cite the article:* Saeed B.A., Sahan A.M. (May 2025); A Proposed Technique For Recognizing And Analyzing Body Language Using Slantlet Transform And Deep Learning; *International Journal of Professional Studies*; Jan-Jun 2025, Vol 19, 136-150; DOI: http://doi.org/10.37648/ijps.v19i01.011

This universality has contributed to the process of recognizing body language taking on an essential role, applicable in various fields, even those that affect our daily lives [2].

The advancements in artificial intelligence have had a positive impact on body language recognition systems, making them more sophisticated in recognizing even subtle and rapid invisible signals that are difficult to detect with the naked eye, making the interpretation of nonverbal signals more accurate and efficient [3].

Body language recognition systems are diverse, ranging from simple systems that rely on individual features of body expressions to advanced systems that combine multiple features extracted from the body. This diversity is also influenced by the way body signals are collected and analyzed, and by the way datasets of human gestures are fed to the system to train the model, allowing it to interpret and classify them into multiple categories. Given their potential applications in a wide range of fields, such as healthcare, where it helps diagnose mental illnesses. It is also effective in analyzing nonverbal behavior during legal investigations and detecting abnormal behavior in security-critical areas such as government buildings, airports, and other areas. As well as improving the interaction between humans and smart devices [4]. These sensitive applications and fields make performance speed a critical factor, so reducing processing time is crucial to enhancing the efficiency of body language recognition systems [5-6].

The purpose of this research is to develop a technique for recognizing and analyzing body language by maintaining recognition accuracy and reducing processing time, while taking into account the effect of noise and some data loss on performance, with the aim of expanding the application of this technique in various fields.

2. Related Works

The development of technology has contributed to the introduction of many techniques and improvements in the field of body language recognition. We will review below some of the studies conducted in this field with the aim of presenting and understanding the progress achieved in previous methods.

Reference [7] presents a method that uses Open Pose technology to extract the positions of key parts such as the head, shoulder, and hands as the first stage of the recognition process. A one-nearest neighbor classifier is then used to recognition. In Reference [8 A two-stage integrated system is proposed. The first uses a proposed technique called Star RGB to convert each video into a single-color image. The second uses a RESTNET model for feature extraction and classification.

Reference [9] A proposed model is presented that integrates the extracted features into a single approach designed for human body recognition by using 3D-CDCN and 2D-MRCN models, each model extracting multiple features. The use of Long Short-Term Memory network with deformable convolution technique was combined in Reference [10] as an innovative approach to recognizing body expressions.

Reference [11] proposes a new framework for emotion recognition and body language analysis to identify psychological states by using Open Pose and K-Nearest Neighbors techniques for classification and then using LSTM technology for body language analysis.

A multi-stage approach has been developed for recognizing body expressions and emotions. In reference [12], presented a two-stage approach for recognizing emotional human gestures. In the first stage, several features are extracted to create a single image that combines the body positions from all the time sequences of the video. Then, in the second stage, a convolutional neural network (Conv Net) model is used for recognition.

Reference [13] The 3D-ResNet model was used to integrate several basic features, including facial, hand, and body movements, and presented as a system for nonverbal gesture recognition. Finally, Reference [14] presents a system for analyzing and interpreting emotions and bodily gestures by using a CNN architecture for classification.

3. Body Language Recognition And Analysis

Most body language recognition systems focus on the primary goal of enabling machines to detect emotions, intentions, and impulses by analyzing images and videos containing nonverbal body signals [15]. Despite this, the systems used to recognize body language vary, starting from simple systems that rely on individual features extracted from one part of the body, such as the hand or face, to systems that rely on integrating a variety of extracted characteristics and features [16]. Recently, artificial intelligence technologies have achieved development and progress in processing sensory or visual data containing movement signals and extracting a set of important information used in recognizing emotions and revealing human behavior, intentions, and motives, i.e. converting them into understandable information [17-18].

The operations group begins with the process of collecting data using sensors or digital cameras, the process of organizing the data so that it is ready for use in the advanced stages, and the pre-processing process represented by processing images or videos and extracting features to identify the important characteristics of each movement, then the classification process that uses deep learning models, so that the model is trained using a set of diverse and huge data on the important characteristics of each movement or behavior. All of the above processes are fundamental to body language recognition systems. The recognition process is followed by the inference stage, which relies on interpreting the connotations associated with movement. Which represents the final step in body language recognition and analysis techniques [19-20].

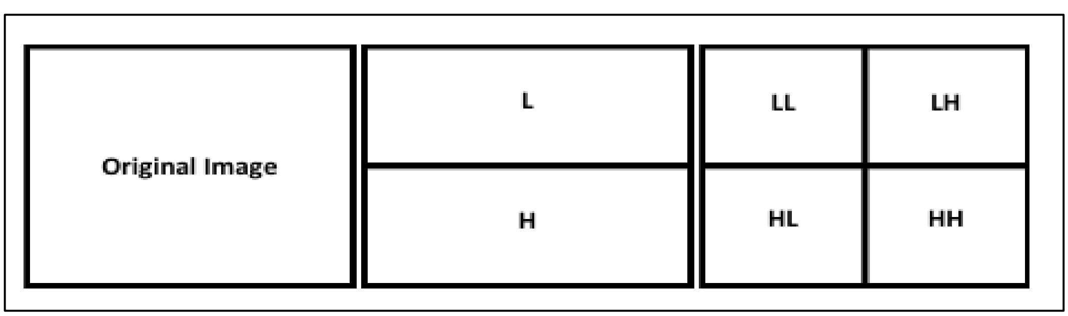
4. Slantlet Transform (SLT)

In 1999, the SLT was proposed by Andreas Volter after the need to improve the time resolution in signal analysis became apparent [21].

The primary goal of the SLT was to enhance signal localization in time better than the Discrete Wavelet Transform (DWT) by using different filters for each level instead of repeating the same filters across levels. The SLT filters are also shorter than their DWT counterparts, which reduces computational complexity and enhances the efficiency of the transformation process. [22].

A notable feature of SLT is its ability to meet the zero moments condition in its filters. This characteristic reduces interference between filter components, thereby improving the system's ability to eliminate unwanted noise. SLT is widely used in various applications that require enhanced time localization, such as in image processing, data compression, signal processing in noisy environments, and multi-scale data analysis. When applied to image analysis, SLT divides an image into four frequency components at each level: LL (Low-Low), which retains the image's fundamental information; LH (Low-High), which captures horizontal edge details; HL (High-Low), which highlights vertical edges; and HH (High-High), which provides details of diagonal edges and finer image features [23]. Figure 1 illustrates the application of 2D SLT.

Figure 1 illustrates the conventional approach for segmenting an image using the 2D SLT decomposition technique [24].



Through the formulas mentioned below [25,26,27], 1-D SLT is applied first to the rows and again to the columns to represent the data in two dimensions, obtaining 2D SLT [25].

$$low\ pass\ filter\ , h_i(n) = \begin{Bmatrix} b_{0,0} + b_{0,1}\, n, & for\ n = 0, \dots 2^i - 1 \\ b_{1,0} + b_{1,1}(n-2)^i & for\ \ n = 2^i, \dots\ 2^{i+1} - 1 \end{Bmatrix} \quad (1)$$

Where

$$b_{0,0} = \frac{u(v+1)}{2m}$$

$$b_{1,0} = u - \ b_{0,0}$$

$$b_{0,1} = \frac{u}{m}$$

$$b_{1,1} = -\, b_{0,1}$$

$$u = \frac{1}{m}$$

$$v = \sqrt{\frac{2m^2+1}{3}}$$

$$m = \ 2^i$$

$$Adjacent\ of\ low\ pass\ filter\ \ , f_i(n) = \begin{Bmatrix} c_{0,0} + c_{0,1}\, n, & for\ n = 0, \dots 2^i - 1 \\ c_{1,0} + c(n-2)^i & for\ \ n = 2^i, \dots\ 2^{i+1} - 1 \end{Bmatrix} \quad (2)$$

Where

$$c_{0,0} = \frac{c_{0,0}(v+1)}{2}$$

$$c_{1,0} = c_{1,1}\frac{u+1-2m}{2}$$

$$c_{0,1} = w(v+m)$$

$$c_{1,1} = -\, w(v+m)$$

$$w = \frac{\sqrt{\frac{3}{m(m^2-1)}}}{m}$$

$$v = \sqrt{\frac{2m^2+1}{3}}$$

$$m = \ 2^i$$

$$Remaining\ filter\ \ , g_i(n) = \begin{Bmatrix} a_{0,0} + a_{0,1}\, n, & for\ n = 0, \dots 2^i - 1 \\ a_{1,0} + a(n-2)^i & for\ \ n = 2^i, \dots\ 2^{i+1} - 1 \end{Bmatrix} \quad (3)$$

Where

$$a_{0,0} = \frac{s_0 + t_0}{2}$$

$$a_{1,0} = \frac{s_0 - t_0}{2}$$

$$a_{0,1} = \frac{s_1 + t_1}{2}$$

$$a_{1,1} = -s_1 \frac{m-1}{2}$$

$$w = \frac{\sqrt{\frac{3}{m(m^2-1)}}}{m}$$

$$v = \sqrt{\frac{2m^2+1}{3}}$$

$$s_0 = -s_1 \frac{m-1}{2}$$

$$s_1 = 6\sqrt{\frac{m}{(m^2-1)(4m^2-1)}}$$

$$t_0 = \frac{((m+1)\frac{s_1}{3} - m\,t_1)(m-1)}{2m}$$

$$t_1 = 2\sqrt{\frac{3}{m(m^2-1)}}$$

$$m = 2^i$$

5. Deep Learning

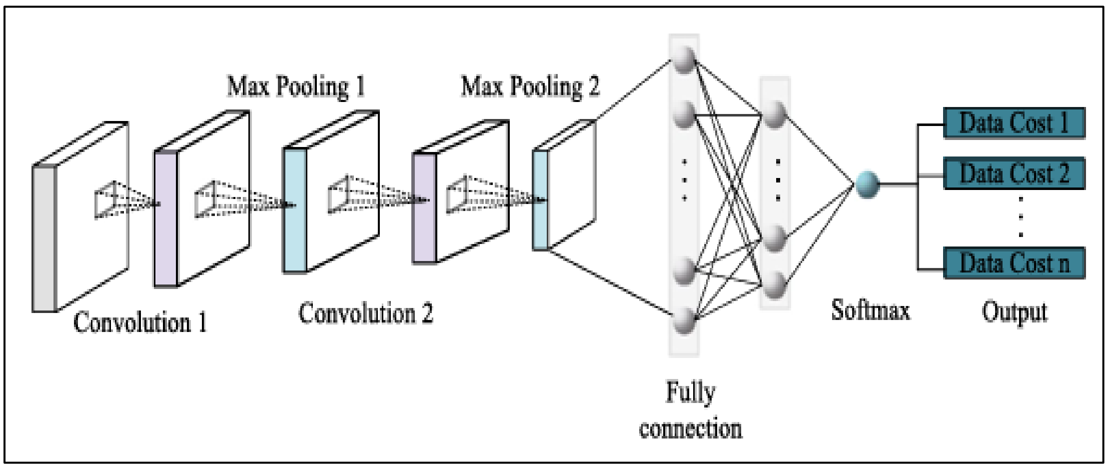
With the contribution of technological development, deep learning has witnessed significant development as it is one of the most advanced branches of machine learning, which has led to the diversity of architectural structures used to implement deep learning, including Deep Neural Networks (DNNs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs). Pre-trained networks and other hybrid methods combine two or more methods to achieve better performance. On the other hand, training and testing techniques contribute to enhancing the accuracy and efficiency of the model, which has made these models an essential part of modern applications such as human recognition, image classification, and face detection. [28]. What sets deep learning apart is its use of multiple sequential layers within neural networks, giving them greater complexity and data-processing power compared to traditional neural networks. The more layers a network has, the "deeper" it becomes—hence the name deep learning [29].

CNNs are the most popular and widely used deep learning networks. They are specifically designed for feature extraction and classification with high accuracy, and are efficient in processing image and video data. The network consists of multiple layers. The first layer's work to discover basic features, while the deeper layers work to discover more complex features and characteristics [30].

Multiple hidden layers in a CNN work in concert to achieve accuracy and efficiency in performance, starting with convolutional layers that use kernels or filters to apply the convolution process to form the feature map. Then the resulting feature map is fed into the deeper pooling layers in order to reduce dimensions using max pooling or mean pooling. After that, the data processed in the layers reaches the final output layer, where classification or prediction is made based on the previous analysis. The CNN architecture provides superior image processing performance, making it more applicable in the fields of face recognition, security surveillance systems, and object detection. [31].

Activation functions, such as ReLU, sigmoid, and tanh, are essential for improving network performance by controlling the outputs of each layer. Furthermore, regularization techniques such as dropout are used to reduce the risk of overfitting, helping to maintain model stability and generalization when dealing with new and previously unseen data. [32]. Figure 2 show the Basic components CNN.

Figure 2 Basic components of Convolutional Neural Networks [33].



6. Methodology

The proposed research framework consists of two main stages: the preprocessing stage and the feature extraction and classification stage. Figure 3 illustrates the general structure of the proposed technique.

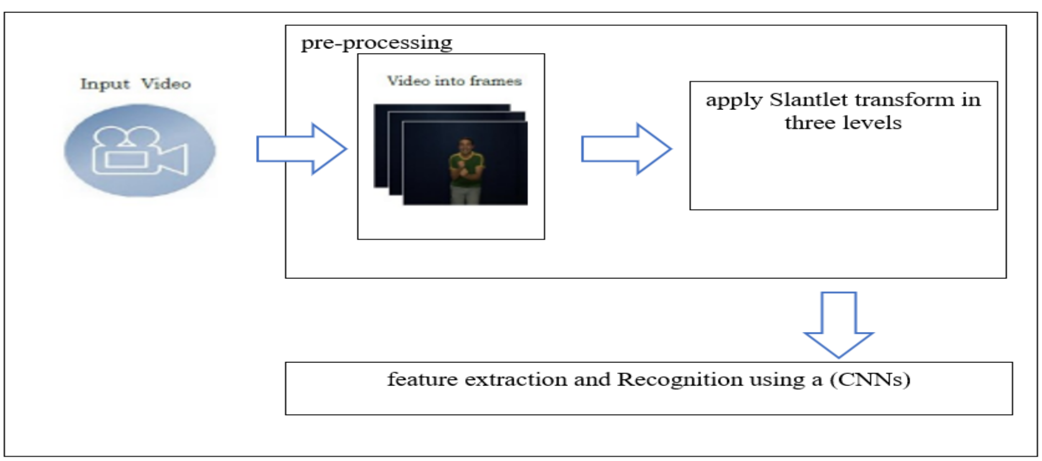


Figure 3 illustrates the general structure of the proposed technique.

A. Preprocessing

The preprocessing stage involves a series of essential steps aimed at preparing the video data in a way that enhances model performance and reduces computational complexity. This stage begins with standardizing the number of frames across the different video clips, as the number of frames varies between videos. Although it is possible to divide the videos into smaller segments, this approach often fails to adequately represent emotional states [10].

To address this issue, the Optical Flow technique was adopted to select the frames that best express motion and visible emotions. This ensures the sequential and systematic selection of a set of frames with high representative value. The data size was also reduced by converting each frame to a grayscale image, thus reducing the number of channels (from three color channels to one).

In the next step, the SLT was applied to the selected frames using a different filter bank at each level. This transformation reduces the frame dimensions by half by dividing the image into four equal quadrants representing different components, while preserving essential information in the LL (Low-Low) component. The same transformation is repeatedly applied to the LL component across three consecutive levels, which reduces computational complexity and enhances the model's robustness to noise. Figure (4) illustrates the application of SLT in the three levels.

Frames before applying SLT	**Frames after applying SLT**		
	Level 1	Level 2	Level 3
			

Figure 4 illustrates the application of SLT in the three levels.

B. Feature Extraction and Classification

The input shape to the model is defined as: (number of frames in the clip, frame height, frame width, number of channels), where the number of channels is set to 1 due to the use of grayscale images.

The model consists of three convolutional layers containing 16, 32, and 64 filters, respectively. Each layer uses 3×3×3 filters with ReLU activation, and L2 regularization with a coefficient of 0.001. A pooling layer with a pool size of 2×2×2 is included before each convolutional layer.

At the end of the model, a fully connected layer with 128 units is used, followed by an output layer with 12 units, corresponding to the number of target classes. A Softmax activation function is applied for prediction. The model was trained using the Adam optimizer with a learning rate of 0.001. A batch size of 32 was set, and the training lasted for 10 epochs. The dataset was divided into two subsets: 80% for training and 20% testing. Figure 5 illustrates the architecture for the model of CNN.

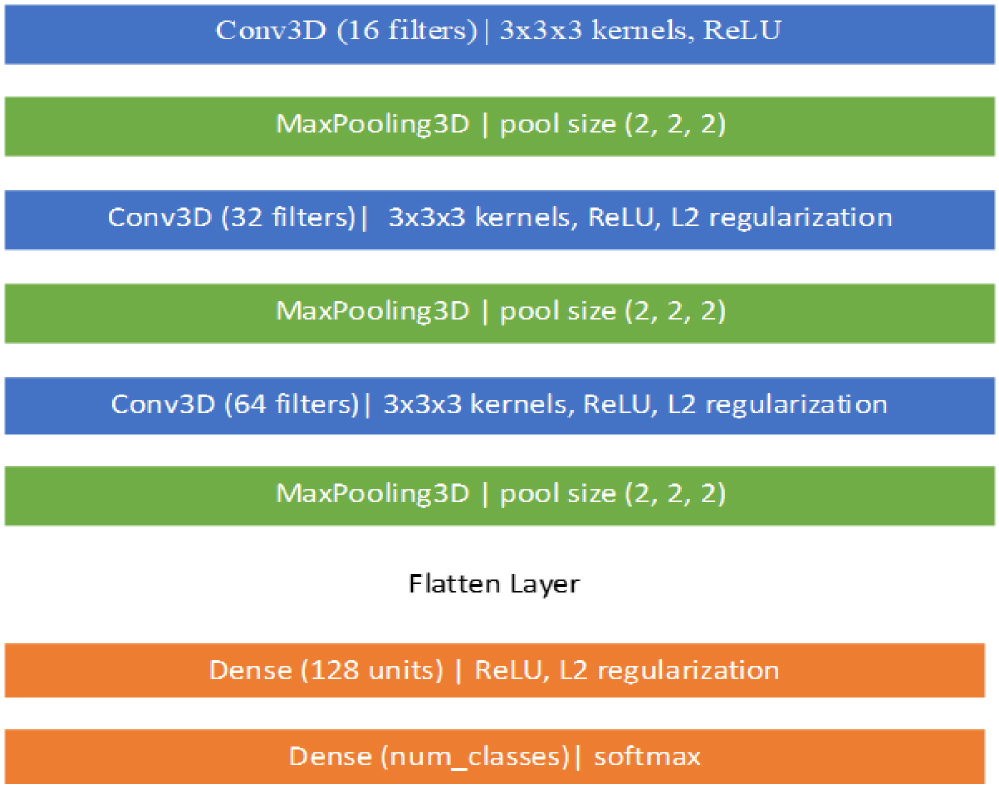


Figure 5 illustrates the architecture for the model of CNN.

7. Results And Discussion

A variety of experiments were conducted to evaluate the performance of the proposed technique in the field of body language recognition and analysis, as well as to compare the accuracy and efficiency of the proposed technique with previous studies to determine the extent of discrimination of the proposed approach and the extent of its integration with previous studies.

7.1 Dataset

a) The Geneva Multimodal Emotion Portrayals (GEMEP) It is an advanced database that is considered one of the leading datasets in the field of recognizing emotional expressions through facial expressions, body movements, and postures. The representation of emotions was performed by 10 professional actors (of both genders), each of whom used a unique expressive method to represent 17 subtle emotional states, some of which are common and well-known, and others are subtle and rare. [34-35].
Specific samples were selected from a dataset that included 12 emotional states (amusement, pride, sadness, anxiety, despair, curiosity, annoyance, fear, pleasure, joy, relief, and anger). To ensure that subtle changes in facial expressions and body movements were preserved, 25 frames were extracted from each video sequence, producing a data sample of 3,000 frames that represented the primary dataset used in the experiments within this study. Representative samples from this curated subset are illustrated in Figure 6.
Figure 6 Examples from the sub-dataset [36].

b) The real-life deception detection dataset includes 121 recordings, including 61 deceptive and 60 truthful videos, featuring defendants and witnesses from a range of court trials with varying outcomes, including convictions, acquittals, and dismissals. The clips were rated according to the credibility of the individuals' statements. Clips considered deceptive were typically those featuring defendants denying their involvement in the crimes, while truthful clips featured either witnesses or defendants providing independently verified information. These clips were carefully selected to capture subtle details of facial expressions and body movements. The participant sample included 21 females and 35 males, ranging in age from 16 to 60. [37].
To ensure the preservation of expressive dynamics over time, 25 frames were systematically extracted from each video clip. This resulted in a total of 3,025 frames, each resized to a uniform resolution of 300×300 pixels, forming a balanced dataset suitable for subsequent analysis. Figure 7 show some examples from the

real-life deception detection dataset.

Figure 7 some Examples from The real-life deception detection dataset [38]

7.2 Evaluate the use of the SLT

Based on a series of extensive experiments conducted to evaluate the impact of applying the SLT on the performance of the proposed technique, the GEMEP dataset—comprising 3,000 frames related to facial and body expressions—was utilized. The dataset was randomly divided into 80% for training and 20% for testing. As illustrated in Table 1, SLT demonstrated a notable ability to reduce frame dimensions while preserving key discriminative features.

At the first level (image size 150×150), processing time was reduced by 77.5%, with a slight improvement in accuracy by 0.79% Compared to the model without SLT. At the second level (75×75), processing time improved by 56.48%, while classification accuracy remained consistent with the first level. At the third level (38×38), a further 31.91% reduction in processing time was observed, again with stable accuracy. These results indicate that SLT effectively eliminates unnecessary features while preserving important features, improving processing efficiency without compromising classification performance.

At the fourth level, although it achieved a significant reduction in processing time, it lost high accuracy due to excessive feature reduction. Therefore, the third level of SLT was adopted for its application in the proposed technique.

Table 1 Result of The Test to Time and Accuracy Before and After Use of SLT

before using the SLT		After using SLT					
Time	Accuracy	Level 1		Level 2		Level 3	
		Time	Accuracy	Time	Accuracy	Time	Accuracy
480s	%98.60	108s	%99.56	47s	%99.56	32s	%99.56

7.3 Evaluation of the Proposed Technique in Body Language Analysis

A set of comprehensive experiments was conducted to evaluate the performance efficiency of the proposed technique using samples from the Real-Life Deception Detection database. The dataset was randomly divided into 80% for training and 20% for testing.

Experiment One:

In this experiment, 25 frames were selected from each video, so that the database sample size became 3025 frames. The outcomes were as follows:

Level 1 (150×150): Processing time was reduced by 76.56%, alongside a 0.82% Improve accuracy of the model before applying SLT.

Level 2 (75×75): A further 41.33% gain in speed was recorded, accompanied by a 0.22% rise in accuracy compared to the previous level.

- **Level 3 (38×38):** Efficiency improved by an additional 34.09%, with accuracy remaining consistent with that of Level 2.

Experiment Two:

In the second phase, data augmentation was employed to diversify the dataset. ءؤImages were randomly scaled within a 10–20% range and rotated at arbitrary angles (60°, 90°, 180°). As a result, the dataset expanded to 12,100 frames. The same 80/20 training-testing split was applied. The outcomes were as follows:

- **Level 1 (150×150):** Time efficiency improved by 71.83%, with a 0.85% boost Improve accuracy of the model before applying SLT
- **Level 2 (75×75):** Processing time decreased by 62.5%, and accuracy rose by 0.20% compared to Level 1.
- **Level 3 (38×38):** A 70% enhancement in speed was achieved, with a modest 0.14% increase in accuracy from the previous level.

As shown in Table 2, the results show a consistent trend toward reduced processing time across all SLT levels. While maintaining high accuracy, these improvements demonstrate the potential application of the proposed technique in various settings, such as forensic and security investigations.

Table 2 Result of the test to time and accuracy before and after use of SLT

Number of Frames	before using the SLT		After using SLT					
	Time	Accuracy	Level 1		Level 2		Level 3	
			Time	Accuracy	Time	Accuracy	Time	Accuracy
3025	320s	%97.20	75s	%98.00	44s	%98.22	29s	%98.22
12100	1420s	%95.05	400s	%95.86	150s	%96.05	45s	%96.18

7.4 Comparison with Previous Studies

The performance of the proposed technique was evaluated against several prior studies focused on body language recognition using facial expressions and body movements. All comparisons were conducted using the same standard dataset, **GEMEP**. As shown in Table 3, the proposed method achieved a notable accuracy of **99.56%** .The proposed approach showed good advantage in processing efficiency, reducing the processing time by more than 77 This improvement makes the proposed technique suitable for real-time applications where speed and accuracy are key factors

Table 3 comparison between the proposed method and previous approaches using the GEMEP dataset.

Methods	**Accuracy**
FDCNN [36]	95.4%
Deformable Convolutional LSTM [10]	98.8%
Dense Optical Flow (DOF) with (CNN) [28]	96.63%
Ours	99.56%

8. Conclusions

This paper presents a proposed technique for interpreting body language signals from video frames. The proposed technique operates in two stages: the first is dimensionality reduction while preserving important features and removing unnecessary ones using the Slantlet transform, and the second is the use of a dedicated model to classify the enhanced inputs.

A set of experiments was conducted using samples from the GEMEP and Real-Life Deception Detection databases to test the accuracy of the proposed technique in recognizing and analyzing body language. The results of the experiments showed that using the proposed technique achieved improvements in the accuracy of recognizing and analyzing body language. Also, applying the Slantlet transform at the third level reduced the processing time by more than 75% in all experiments while maintaining the performance and efficiency of the classifier.

Compared to previous studies, the proposed technique achieved high accuracy in recognizing body language, in addition to Reduce processing time, which makes it suitable for real-time applications. The proposed technique demonstrated the ability to withstand various conditions such as (increasing data volume, rotation, and decreasing data accuracy), which enhances its applicability and reliability. Confirming the efficiency of the proposed technique Thus, it has potential applications in multiple fields such as behavioral analysis and security monitoring.

9. Conflict of Interest

The authors declare that they have no conflict of interest.

10. Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

11. Reference

1. Abdulghafor, R., Turaev, S., & Ali, M. A. H. (2022). Body language analysis in healthcare: An overview. Healthcare, 10(7), 1251. https://doi.org/10.3390/healthcare10071251
2. Ahmad, Z., & Khan, N. (2022). A survey on physiological signal-based emotion recognition. Bioengineering, 9(11), 688. https://doi.org/10.3390/bioengineering9110688
3. Al-Tamimi, M. S. H. (2019). Combining convolutional neural networks and slantlet transform for an effective image retrieval scheme. International Journal of Power Electronics and Drive Systems, 9(5), 4382-4395. https://doi.org/10.11591/ijece.v9i5.pp4382-4395
4. Arif, S., Wang, J., Hassan, T. U., & Fei, Z. (2019). 3D-CNN-based fused feature maps with LSTM applied to action recognition. Future Internet, 11(2), 42. https://doi.org/10.3390/fi11020042

5. Artanto, H., & Arifin, F. (2023). Emotions and gesture recognition using affective computing assessment with deep learning. IAES International Journal of Artificial Intelligence, 12(3), 1419-1427. https://doi.org/10.11591/ijai.v12.i3.pp1419-1427
6. Banziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In Blueprint for affective computing: A sourcebook (pp. 217-294). Oxford University Press.
7. Darduh, M. T., Sahan, A. M., & Al-Itbi, A. S. A. (2024). Rotation invariant technique for sign language recognition. InfoTechSpectrum: Iraqi Journal of Data Science, 1(1), 16-27. https://doi.org/10.51173/ijds.v1i1.6
8. Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3-4), 169-200. https://doi.org/10.1080/02699939208411068
9. Elboushaki, A., Hannane, R., Afdel, K., & Koutti, L. (2020). MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences. Expert Systems with Applications, 139, 112829. https://doi.org/10.1016/j.eswa.2019.112829
10. Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: A survey. Foundations and Trends in Human-Computer Interaction, 1(3), 203-275. https://doi.org/10.1561/1100000005
11. Hanoon, A. (2009). Image zooming using inverse slantlet transform. Al-Khwarizmi Engineering Journal, 5(2), 54-65.
12. Khalifa, I., Ejbali, R., Schettini, R., & Zaied, M. (2021). Deep multi-stage approach for emotional body gesture recognition in job interview. The Computer Journal, 65(7), 1702-1716. https://doi.org/10.1093/comjnl/bxab011
13. Khazal, R. B., & Habeeb, N. J. (2022). Enhancement infrared-visible image fusion using the integration of stationary wavelet transform and fuzzy histogram equalization. Journal of Techniques, 4(4), 119-127. https://doi.org/10.51173/jt.v4i4.700
14. Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. IEEE Transactions on Affective Computing, 4(1), 15-33. https://doi.org/10.1109/T-AFFC.2012.16
15. Kolivand, H., Wee, T. C., Asadianfam, S., Rahim, M. S., & Sulong, G. (2023). High imperceptibility and robustness watermarking scheme for brain MRI using slantlet transform coupled with enhanced knight tour algorithm. Multimedia Tools and Applications, 83(8), 22221-22260. https://doi.org/10.1007/s11042-023-16459-7
16. Leong, S. C., Tang, Y. M., Lai, C. H., & Lee, C. K. M. (2023). Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing. Computer Science Review, 48, 100545. https://doi.org/10.1016/j.cosrev.2023.100545
17. Mohammed, R. T., & Khoo, B. E. (2012). Image watermarking using slantlet transform. 2012 IEEE Symposium on Industrial Electronics and Applications, 281-286. https://doi.org/10.1109/ISIEA.2012.6496644
18. Niewiadomski, R., Kolykhalova, K., Piana, S., Alborno, P., Volpe, G., & Camurri, A. (2019). Analysis of movement quality in full-body physical activities. ACM Transactions on Interactive Intelligent Systems, 9(1), 1-20. https://doi.org/10.1145/3132369
19. Noroozi, F., Corneanu, C. A., Kaminska, D., Sapinski, T., Escalera, S., & Anbarjafari, G. (2021). Survey on emotional body gesture recognition. IEEE Transactions on Affective Computing, 12(2), 505-523. https://doi.org/10.1109/TAFFC.2018.2874986
20. Ojeda-Castelo, J. J., Capobianco-Uriarte, M. L. M., Piedra-Fernandez, J. A., & Ayala, R. (2022). A survey on intelligent gesture recognition techniques. IEEE Access, 10, 87135-87156. https://doi.org/10.1109/ACCESS.2022.3199358
21. Perez-Rosas, V., Abouelenien, M., Mihalcea, R., & Burzo, M. (2015). Deception detection using real-life trial data. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 59-66.
22. Rehman, M. U., et al. (2022). Dynamic hand gesture recognition using 3D-CNN and LSTM networks. Computers, Materials & Continua, 70(3), 4675-4690. https://doi.org/10.32604/cmc.2022.019586

23. Santhoshkumar, R., & Geetha, M. K. (2019). Deep learning approach for emotion recognition from human body movements with feedforward deep convolution neural networks. Procedia Computer Science, 152, 158-165. https://doi.org/10.1016/j.procs.2019.05.038
24. Santos, C. C. D., Samatelo, J. L. A., & Vassallo, R. F. (2020). Dynamic gesture recognition by using CNNs and star RGB: A temporal information condensation. Neurocomputing, 400, 238-254. https://doi.org/10.1016/j.neucom.2020.03.038
25. Sarma, D., Kavyasree, V., & Bhuyan, M. K. (2022). Two-stream fusion model using 3D-CNN and 2D-CNN via video-frames and optical flow motion templates for hand gesture recognition. Innovations in Systems and Software Engineering. https://doi.org/10.1007/s11334-022-00477-z
26. Schneider, P., Memmesheimer, R., Kramer, I., & Paulus, D. (2019). Gesture recognition in RGB videos using human body keypoints and dynamic time warping. Lecture Notes in Computer Science, 11531, 281-293. https://doi.org/10.1007/978-3-030-35699-6_22
27. Sekar, V., & Jawaharlalnehru, A. (2022). Semantic-based visual emotion recognition in videos-a transfer learning approach. International Journal of Electrical and Computer Engineering, 12(4), 3674-3683. https://doi.org/10.11591/ijece.v12i4.pp3674-3683
28. Selesnick, I. W. (1998). The slantlet transform. Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, 53-56. https://doi.org/10.1109/TFSA.1998.721359
29. Sen, U. M., Perez-Rosas, V., Yanikoglu, B., Abouelenien, M., Burzo, M., & Mihalcea, R. (2022). Multimodal deception detection using real-life trial data. IEEE Transactions on Affective Computing, 13(1), 306-319.
30. Shafique, Z., Wang, H., & Tian, Y. (2023). Nonverbal communication cue recognition: A pathway to more accessible communication. 023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 5666-5674. https://doi.org/10.1109/CVPRW59228.2023.00600
31. Singh, H., Kumar, A., Balyan, L. K., & Singh, G. K. (2018). Slantlet filter-bank based satellite image enhancement using gamma corrected knee transformation. International Journal of Electronics, 105(10), 1695-1715. https://doi.org/10.1080/00207217.2018.1477199
32. Tahghighi, P., Koochari, A., & Jalali, M. (2021). Deformable convolutional LSTM for human body emotion recognition. Lecture Notes in Computer Science, 12663, 741-747. https://doi.org/10.1007/978-3-030-68796-0_54
33. Vaijayanthi, S., & Arunnehru, J. (2022). Human emotion recognition from body posture with machine learning techniques. Communications in Computer and Information Science, 1613, 231-242. https://doi.org/10.1007/978-3-031-12638-3_20
34. Yang, Z., Kay, A., Li, Y., Cross, W., & Luo, J. (2021). Pose-based body language recognition for emotion and psychiatric symptom interpretation. 2020 25th International Conference on Pattern Recognition, 294-301. https://doi.org/10.1109/ICPR48806.2021.9412591
35. Zhao, X., Wang, L., Zhang, Y., et al. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57, 99. https://doi.org/10.1007/s10462-024-10721-6

About Authors

Ban Amjed Saeed holds a Bachelor's degree in the Department of Informatics at Middle Technical University, Baghdad, Iraq. She is currently pursuing a Master's degree in the same field, specializing in Artificial Intelligence. Eager to apply her knowledge and skills in diverse work environments, she aims to contribute to developing technological solutions that enhance performance and support intelligent decision-making.

Prof. Dr. ALI Mohammed Sahan is a Professor at the Department of Informatics, Meddle Technical University, Iraq, Baghdad. His research interests include Artificial Intelligent., Pattern Recognition and Digital Image processing. He has published many articles in national and international scopus and clarivate journals of computer science, as well as he is reviewer and editor/chief in several national and international journals.