International Journal of Artificial Intelligience and Cognitive Psychology (IJAICP)
Volume 1, Issue 1, ISSN,ISBN-10:1-100-22300-2
DOI: 10.6084/m9.figshare.26000290

Application of Viola-Jones Control Model for Emotion Recognition Based on Facial Expression and Body Gestures

Fatima Isiaka
Department of Computer Science, Nasarawa State University, Keffi, Nigeria
Email: isiakafatima@nsuk.edu.ng
Zainab Adamu
Department of Computer , Ahmadu Bello University, Zaria, Nigeria.
Email: zadmu31@gmail.com

A Published Article of International Journal of Artificial Intelligience and Cognitive Psychology

Abstract

Body gestures are one of the crucial components of body language and can also be used for emotion recognition in people. This could include movement and depths and changes in the motion of the body in a confined environment. However, these aspects remain less explored to recognise emotion while facial expressions and examples such as speech-based approaches are widely investigated. This paper introduces a custom method of emotion recognition through facial expressions and body gestures in people based on Viola-Jones using Matlab images saved in the archive and a thousand images online collection of a thousand body gestures and facial expressions. A Viola Jones control model was used to save these images and to identify features that recognise different forms of emotion saved in the database. The surroundings of the person also help to map the right emotion and expression on the face of the person. Tracking points and error tracking were computed for each class of emotion detected; prediction accuracy is close to 78% for each predicted class of emotion.

Keywords: *Body gestures, facial expression, Voila Jones, Point tracking, emotion recognition, Speed-based emotion recognition*

Accepted: 14th September, 2023
Revised: 16th October, 2023
Published:11th November, 2023

Corresponding Author:
Fatima Isiaka
Correspondent Email:
fatima.isiaka@outlook.com

1 INTRODUCTION

A lot of research *(Kessous et al. (2010); Castellano et al. (2007); Zhao et al. (2013); De Gelder (2016); Kleinsmith and Bianchi-Berthouze (2012); Calbi et al. (2021); D'Mello and Graesser (2009); Sapinski et al. (2019); Isiaka and Adamu; Dael et al. (2012))* has viewed body gestures to be very crucial features of body language and less explored emotion recognition. This paper investigated a tentative experimental procedure that recognises emotion using a custom Voila Jones control model on facial expression and body gestures with a good number of Matlab images saved in archives and also online images of static body gestures and environment input including the contours and symmetry in clothes and atmospheric disturbance caught in static motion. Their neural control groups are from different persons or in a group, just like the simple model used as an example in this paper. A camera with high image capture was used to support the experiment. The novel approach here is to use the control model as a fuse skeleton and color feature both for background and foreground capture to support the experiment on the color image capture data. The results showed that the approach achieved extensive improvements in all categories and also in the overall dataset with stronger generalisation and capability of image processing. For the complete dataset that involves the image of a person, the surrounding environment (Figure 1) is captured which also includes certain body

postures and gestures as parameters used to identify the emotion on the face of the person. These are divided into training and test sets.

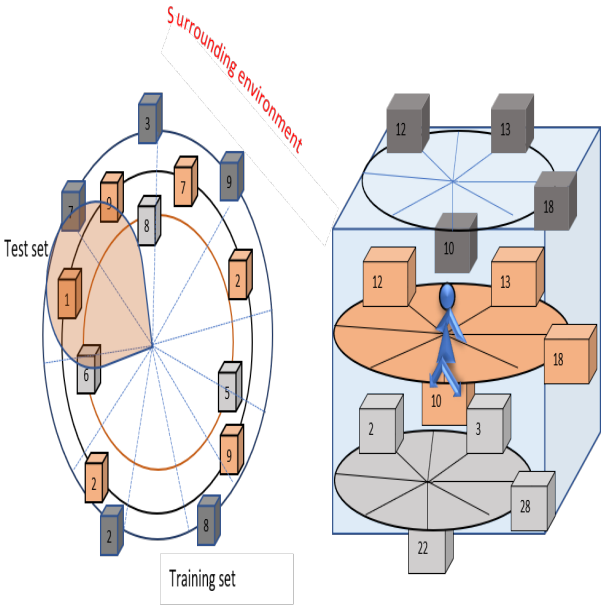


Figure 1: Emotion recognition based on multi-view body gesture perspective at different angles.

In different psychological investigations, it is revealed that bodily gestures convey very important information about emotion recognition as a valuable aspect but have not gained much recognition from the aspect of the science community when compared to modalities, such as facial expression and speech recognition related to emotion recognition. Also, there are certain limited problems in this area of research. Recent work has also exploited the hand-grafted features that propose an end-to-end deep learning approach *(Kim et al. (2023); Naik and Mehta (2018))* for gesture-based emotion recognition. In the initial stage, they adopted the hashing procedure that extracts the keyframes from some video files and in the second stage, a convolution LSTM network was used for exploiting the sequence information. The results exceed some of the had-grafted results and achieved the state-of-the-art results for an end-to-end deep learning-based technique for gesture-based emotion recognition on the FABO dataset they used and also indicate a promising future improvement *(Mohammed et al. (2019); Ly et al. (2018); Mahmoud et al. (2022); Ly et al. (2019); Goyal et al. (2020); Radoi et al. (2021); Wei et al. (2024); Cui et al. (2020))*.

Other r research has worked specifically on related changes in facial patterns that give more information about the emotional state of a person and contribute to regulating conversations with any person. These expressions help in understanding the overall mood of the person in a better way. Facial expressions play an important role in human interactions and non-verbal communication. The classification of facial expressions can be used as an effective tool in behavioural and cognitive studies as an analysis that deals with visually recognising and processing different facial motions and facial features' distinctness. Research *(Radoi et al. (2021); Wei et al. (2024); Cui et al. (2020); Ekman and Rosenberg (1997); Ekman and Friesen (1978); Donato et al. (1999); Correia-Caeiro et al. (2022); Bartlett et al. (2008); Gavrilescu and Vizireanu (2019); Tian et al. (2001); Essa and Pentland (1997))* was conducted on facial action using action coding systems to measure facial behaviour. The codes use different facial movements into action units that are based on the underlying muscular activity which produces momentary changes in the facial expression of the images. An expression can be recognised by correctly identifying the action unit or combination of the action units related to a particular expression.

1.1 Emotion recognition as a mental state

Using facial expressions for emotion recognition can be an intuitive reflection of a person's mental state, this contains rich enough emotional information and can serve as one of the most important forms of interpersonal communication. This interpretation can also be used in various fields such as psychology to understand a person's mood and psychological status. *(Salovey et al. (2000); Rauthmann et al. (2015); Russell (2003); Gross (1998); Lazarus (2006))* the book on wisdom in facial recognition and emotion recognition summarises eight different methods on how to recognise people's mood and identity by choosing the right attributes. One of the characteristics to identify a person is looking at the eyes and nose for evil and righteousness, the lips tend to curve up for truth and falsehood; the temperament for success and fame and also other attributes is based on these feature searches. Due to the complexity and variability of the human facial expression for emotional feature extraction, traditional facial expression technology has become more at a disadvantage for insufficient feature extraction and predisposition to external environmental influences. *(Pabba and Kumar (2022); Zhang et al. (2020); Kim (2007); Cowie et al. (2001); Prkachin and Hammal (2021); Jain et al. (2016); Fernandez et al. (2016); Koelstra et al. (2011); Isiaka et al. (2022))* proposes a new approach to feature fusion with a dual-channel expression recognition module that is based on machine learning theory and also some philosophical rationale on the changes in facial expression. The first module takes in a Gabor feature of the ROI area as input, to make full use of the detailed features of the active face. The first segment is set to an active state from the original face image and used in the Gabor transformation to extract the emotional features of the area on a face. An efficient channel attention network was proposed based on depth separable convolution to improve the linear bottleneck structural compatibility and reduction of network

complexity. Overfitting can also be prevented by designing an efficient attention module that combines the depth of the features mapped to spatial information *(ISI-AKA (2024); Ma et al. (2021); Fang et al. (2019))*. This part mostly focuses on extracting features, outperforming, and improving emotion recognition accuracy in the dataset used. The paper follows similar guidelines using image matching and prediction through sample and control image processing, the proceeding section briefly discusses this method.

2 METHOD

At the initial stage, a least of images were saved in Matlab archive, and online images were selected, all containing different images of people in groups or single appearances. The aim was to conduct an initial study on the Voila Jones control model to identify persons and their facial expressions from their surrounding environment, contours, and symmetry of the lines from the surroundings. The input images were divided into testing and training sets, where the model has to learn from the training set. Predictions are made from the images on each trained set to identify the facial expression and matching emotion recognised (Figure 3). The Viola Jones control model is embedded with both a dynamic control model and a linear model that is used to map out the facial features and surrounding environments. The index pages start with the input of multiple images to identify the facial expression. The image capture could be both static and still captured from real-time video surveillance. Though developments in image capture now allow the processing of practically any form of video content, additional modification in precision and efficiency is highly desirable, in particular via the development of real-time detection and feedback systems. The method here demonstrates the application of neural networks for process monitoring via visual observation of the workpiece during image processing.

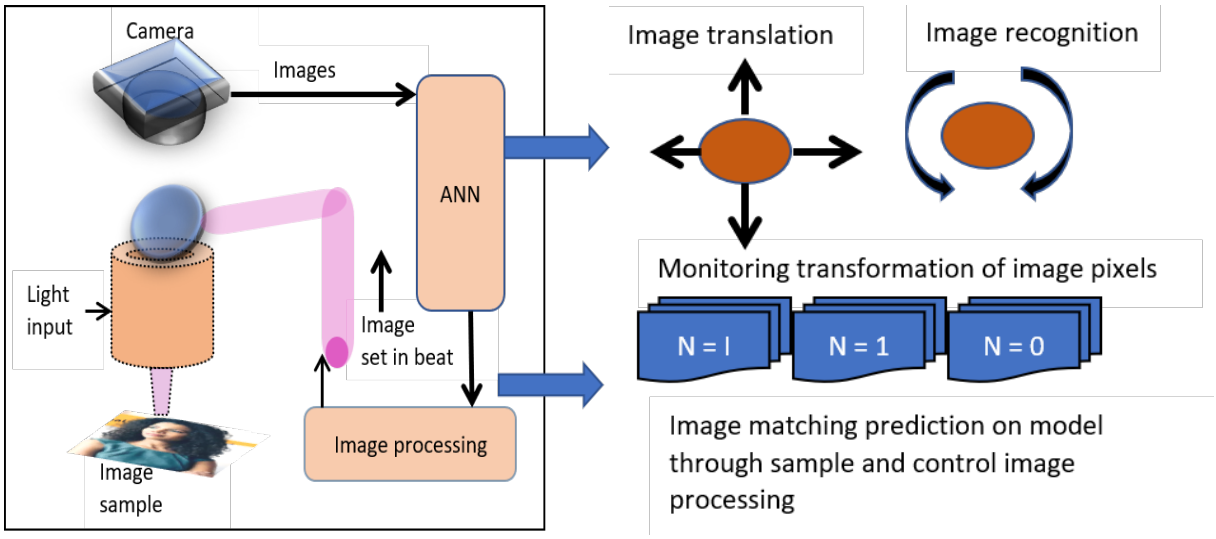
Specifically, the quantification of unintended image transfer modifications, namely emotion recognition and body gestures, along with real-time closed-loop feedback capable of halting image processing are illustrated immediately after matching set through sequential processing. This approach can detect translations in image movement and position that are smaller than the pixels of the camera used for observation. The method also utilises data augmentation that can be used to significantly reduce the quantity of experimental data needed for training an artificial neural network for the images. Inadvertently image translations and recognition are detected synchronously, hence representing the likelihood for simultaneous identification of many persons' image machining parameters. Neural networks are an ideal solution, as they require zero understanding of the physical properties of image and emotion matching, and instead are trained directly from input and experimental data from real-time image capture.

3 RESULT

The controllers in Viola Jones are embedded with dynamic and linear controls (Figure 2) to activate the camera and give choices for real-time image capture and input of images from the archive. These images can either contain facial images of a group of people or a single person, which are trained to recognise facial expressions and body gestures that correspond to the rounding environment. Tracking points are used to locate detection accuracy in real-time and an aggregate of the tracking points can also be computed. The tracking frequency for both background and foreground image detection can also be adjusted to a high speed for high performance in detection accuracy; the parameters for emotion recognition are set in Table 1, each emotion can be detected based on the class of each correlating emotional response detected on a particular face.

The database of emotions to be recognised is set to five classes, *"Smile", "Neutral", "Sad", "Angry"* and *"Happy"* at the pilot stage, each depends on the posture of the upper body or the lower body, the number of contours and symmetry also depends on a maximum value between eleven to fourteen. The default is set to zero if the value allocated to a given expression exceeds the set limit for each facial expression in the database. The given attributes of body parts depend on the intensity of the pose, for instance, if a person is happy or smiling the posture is a number that vibrates with different contours and lines surrounding them and the wrinkles are also part of the features. If a person is angry or sad the continuance is normally stiff with less symmetry around them.

Sometimes, the detected contours and symmetry can exist within the maximum set limit for an expression to be accurately detected, and this is one of the default effects of prediction accuracy. Figure 4 shows images in the foreground and background detected with facial expressions and the surrounding contours with purple cascaded makers; each facial expression is also assigned a different color for feature identification and accuracy. The happy and smiling facial expressions can also be assigned similar values and colors since they both have similar attributes with the addition of bare teeth for a wide and extremely happy face. The different colors in the foreground image detection are also part of the attributes and also affect the detection accuracy hence the detection accuracy mostly depends on the background image capture (Figure 4a). The background subtraction uses a technical concept, which allows an image and its facial expression to be extracted for further object recognition in the foreground image view. The control systems do not require information about everything in the evolution of body



(a) Image processing for emotion recognition based on closed loop feedback.

(b) Closed loop feedback.

Figure 2: : Image processing for emotion recognition based on closed loop feedback.

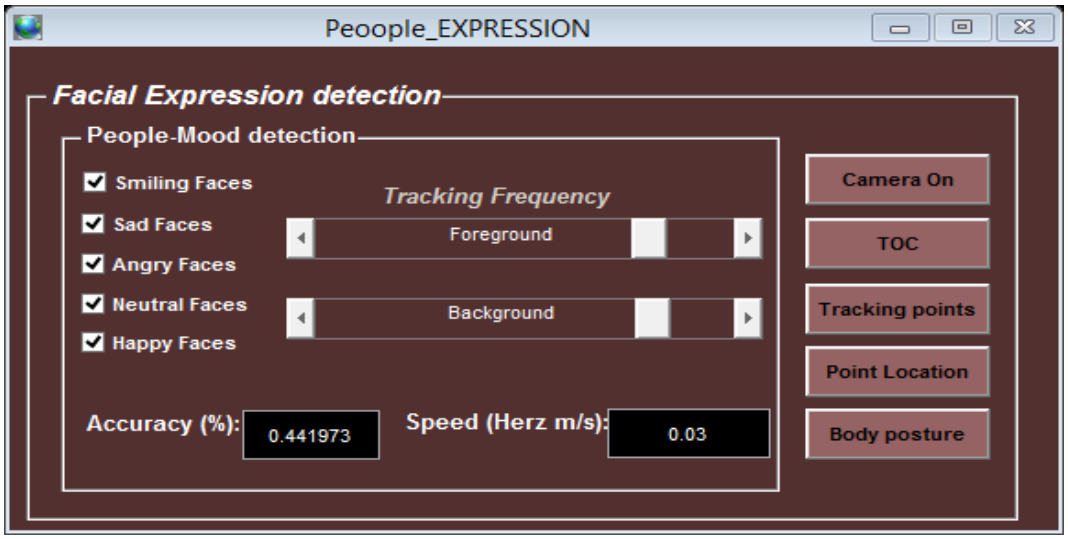


Figure 3: The index page f the Viola Jone control model for image capture and tracking.

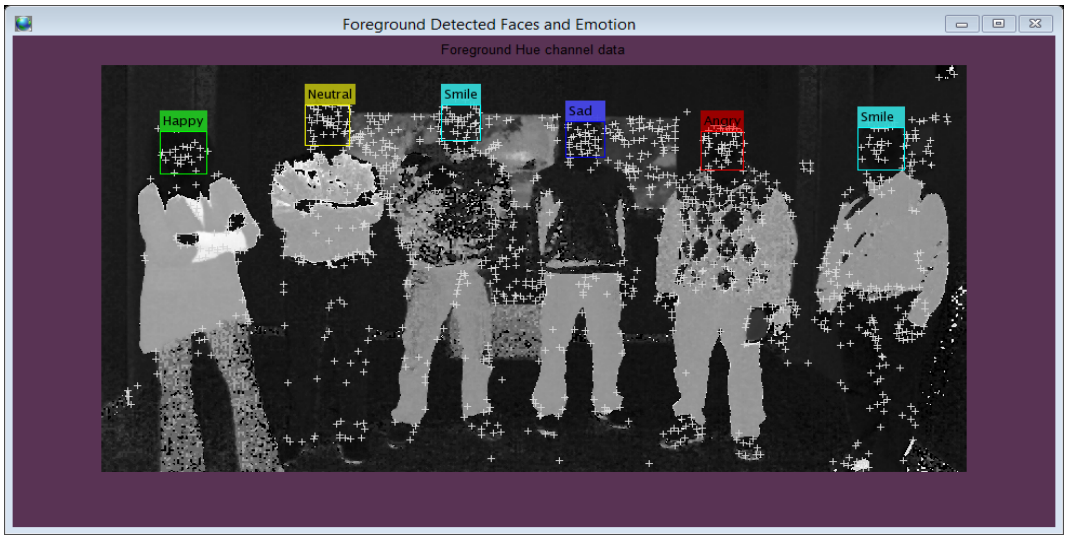
Table 1: Class of emotion and parameters to identify on a person.

Emotion Name	Body parts	Value	Contours	Frequency (Hz)	Color
Happy	Face and upper body	4	10	20	Green
Neutral	Face and upper body	5	12	12	Yellow
Sad	Face, Upper body and lower body	7	12	34	Red
Smile	Face and upper body	5	10	23	Light blue
Angry	Face and upper body	6	12	34	Red

movement in both the real-time and static images but require information about the scene changes, since the images' regions of interest are objects like the contours and symmetry of the body and facial surroundings.

Background subtraction (Figure 4b is used as an approach for detecting the moving surroundings in terms of the detecting features that relate to a facial expression based on body pose and contours in both real-time and

(a) Foreground image of body and facial expression detected with cascaded features on lines and symmetry on body postures and surrounding environment.



(b) Background images captured with detected facial expressions and contours of the surrounding environment

Figure 4: Foregroundand Background image of images and facial expression detected with cascaded features on lines and symmetry on body postures and surrounding environment.

static image capture. The rationale of is approach is that of detection of the moving surroundings for instance in a video sequence. It sets the difference between the current frame and a reference frame, often called the background image. The subtraction is mostly done when the image to be analysed is part of the features of the input images or video frame. This provides important for numerous applications designed for the sole purpose of image processing in surveillance tracking and people pose estimation. The background subtraction is also based on a static background hypothesis which is sometimes not often applicable to real environments. This is one of the novel approaches applied here, to add additional features that can be able to detect the surrounding environments based on the number of lines and contours surrounding the images from the input source.

Figure 5 shows the tracking accuracy in real-time on an aggregate of the input images, the tracking error for the contour and symmetry flow is reduced by the multi-view forward-backward tracking and the traced feature points are divided into the background and the moving or surrounding environment based on homograph of the Viola-Jones control model of both camera movement compensation and symmetry flow. The outlying points are filtered by the moving average filter and set the size as a classified emotion value for each facial image detected.

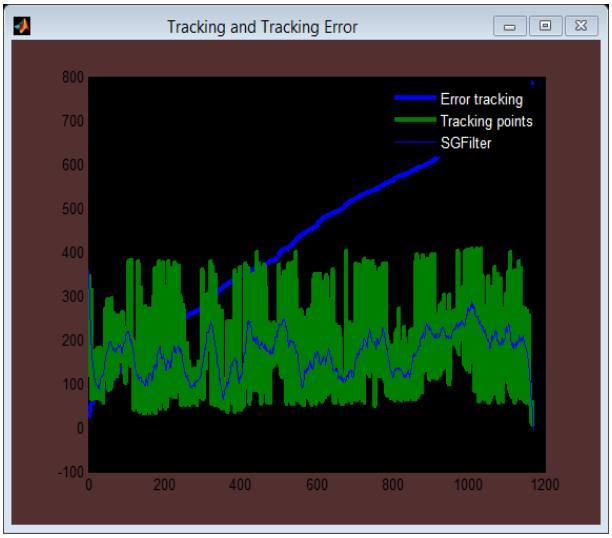


Figure 5: The tracking error and points of detected facial images with a moving average filter.

4 CONCLUSION

This paper set out to investigate Viola Jones's control model for emotion recognition based on facial expressions and body gestures in a real-time closed loop, the process uses Matlab images saved in an archive and a thousand images online with a collection of a thousand body gestures and facial expressions. A Viola Jones control model was used to save these images and to identify features that recognise different forms of emotion saved in the database. The surrounding environment of the person also helps to map the right emotion the expression on the face of the person. Tracking points and error tracking were computed for each class of emotion detected; prediction accuracy is close to 78% for each predicted class of emotion. The limitation of the process is the lack of authentic parameter settings for a background color for all attributes related to contours and symmetry of the surrounding environment, so basically, the colors, values, and body movement are used as the contributing parameters for object tracking and detection. Future work would be to compare the results obtained here to deep learning algorithms and set authentic parameters related to the environment for detecting emotion from facial expressions.

ACKNOWLEDGMENTS

The authors would like Nasarawa State University, Keffi, Nigeria, and Ahmadu Bellow University, Zaria, Nigeria for the sponsor of this paper.

References

Marian Bartlett, Gwen Littlewort, Esra Vural, Kang Lee, Mujdat Cetin, Aytul Ercil, and Javier Movellan. Data mining spontaneous facial behavior with automatic expression coding. In *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction: COST Action 2102 International Conference, Patras, Greece, October 29-31, 2007. Revised Papers*, pages 1–20. Springer, 2008.

M Calbi, N Langiulli, F Siri, MA Umilta, and V Gallese. Visual exploration of emotional body language: a behavioural and eye-tracking study. *Psychological Research*, 85:2326–2339, 2021.

Ginevra Castellano, Santiago D Villalba, and Antonio Camurri. Recognising human emotions from body movement and gesture dynamics. In *International conference on affective computing and intelligent interaction*, pages 71–82. Springer, 2007.

Catia Correia-Caeiro, Anne Burrows, Duncan Andrew Wilson, Abdelhady Abdelrahman, and Takako Miyabe-Nishiwaki. Callifacs: the common marmoset facial action coding system. *PloS one*, 17(5):e0266442, 2022.

Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

Heng Cui, Aiping Liu, Xu Zhang, Xiang Chen, Kongqiao Wang, and Xun Chen. Eeg-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowledge-Based Systems*, 205:106243, 2020.

Nele Dael, Marcello Mortillaro, and Klaus R Scherer. Emotion expression in body action and posture. *Emotion*, 12(5):1085, 2012.

Beatrice De Gelder. *Emotions and the Body*. Oxford University Press, 2016.

Sidney D'Mello and Art Graesser. Automatic detection of learner's affect from gross body language. *Applied Artificial Intelligence*, 23(2):123–150, 2009.

Gianluca Donato, Marian Stewart Bartlett, Joseph C. Hager, Paul Ekman, and Terrence J. Sejnowski. Classifying facial actions. *IEEE Transactions on pattern analysis and machine intelligence*, 21(10):974–989, 1999.

Paul Ekman and Wallace V Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.

Paul Ekman and Erika L Rosenberg. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

Irfan A. Essa and Alex Paul Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, 19(7):757–763, 1997.

Yuming Fang, Chi Zhang, Hanqin Huang, and Jianjun Lei. Visual attention prediction for stereoscopic video by multi-module fully convolutional network. *IEEE Transactions on Image Processing*, 28(11):5253–5265, 2019.

Alberto Fernandez, Ruben Usamentiaga, Juan Luis Carus, and Ruben Casado. Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11): 1805, 2016.

Mihai Gavrilescu and Nicolae Vizireanu. Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors*, 19(17):3693, 2019.

Samta Jain Goyal, Arving Kumar Upadhyay, and Rajesh Singh Jadon. A brief review of deep learning based approaches for facial expression and gesture recognition based on visual information. *Materials Today: Proceedings*, 29:462–469, 2020.

James J Gross. The emerging field of emotion regulation: An integrative review. *Review of general psychology*, 2 (3):271–299, 1998.

Fatima ISIAKA. Physiological metrics for adult and younger users based on a cognitive analytical modeling. *International Journal of Computer Science and Multidisciplinary Research*, 2(2):6, 2024.

Fatima Isiaka and Zainab Adamu. Ethnic classification using support vector machine for eye color recognition. *International Journal Computer Studies and Advance ment in Current Research*, 2(1).

Fatima Isiaka, Salihu Aish Abdulkarim, Kassim Mwitondi, and Zainab Adamu. Emotion detection on webpages using biosensors integrated to a window-based dynamic control system. *International Journal of Intelligent Computing and Cybernetics*, 15(2):277–301, 2022.

Anil K Jain, Karthik Nandakumar, and Arun Ross. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern recognition letters*, 79:80–105, 2016.

Loic Kessous, Ginevra Castellano, and George Caridakis. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 3: 33–48, 2010.

Donghyun Kim, Junmo Yang, and Dongwon Yun. Anthropomorphic robot hand using the principle of sweat and fingerprints of human hands. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10289–10295. IEEE, 2023.

Jonghwa Kim. Bimodal emotion recognition using speech and physiological changes. *Robust speech recognition and understanding*, 265:280, 2007.

Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4 (1):15–33, 2012.

Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

Richard S Lazarus. Emotions and interpersonal relationships: Toward a person-centered conceptualization of emotions and coping. *Journal of personality*, 74(1): 9–46, 2006.

Son Thai Ly, Guee-Sang Lee, Soo-Hyung Kim, and Hyung-Jeong Yang. Emotion recognition via body gesture: Deep learning model coupled with keyframe selection. In *Proceedings of the 2018 international conference on machine learning and machine intelligence*, pages 27–31, 2018.

Son Thai Ly, Guee-Sang Lee, Soo-Hyung Kim, and Hyung-Jeong Yang. Gesture-based emotion recognition by 3d-cnn and lstm with keyframes selection. *International Journal of Contents*, 15(4):59–64, 2019.

Xu Ma, Jingda Guo, Andrew Sansom, Mara McGuire, Andrew Kalaani, Qi Chen, Sihai Tang, Qing Yang, and Song Fu. Spatial pyramid attention for deep convolutional neural networks. *IEEE Transactions on Multimedia*, 23:3048–3058, 2021.

Rihem Mahmoud, Selma Belgacem, and Mohamed Nazih Omri. Towards an end-to-end isolated and continuous deep gesture recognition process. *Neural Computing and Applications*, 34(16):13713–13732, 2022.

Adam Ahmed Qaid Mohammed, Jiancheng Lv, and Md Sajjatul Islam. A deep learning-based end-to-end composite system for hand detection and gesture recognition. *Sensors*, 19(23):5282, 2019.

Niti Naik and Mayuri A Mehta. Hand-over-face gesture based facial emotion recognition using deep learning. In *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, pages 1–7. IEEE, 2018.

Chakradhar Pabba and Praveen Kumar. An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. *Expert Systems*, 39(1):e12839, 2022.

Kenneth M Prkachin and Zakia Hammal. Computer mediated automatic detection of pain-related behavior: prospect, progress, perils. *Frontiers in Pain Research*, 2:788606, 2021.

Anamaria Radoi, Andreea Birhala, Nicolae-Catalin Ristea, and Liviu-Cristian Dutu. An end-to-end emotion recognition framework based on temporal aggregation of multimodal information. *IEEE Access*, 9:135559–135570, 2021.

John F Rauthmann, Ryne A Sherman, and David C Funder. Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality*, 29(3):363–381, 2015.

James A Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003.

Peter Salovey, Alexander J Rothman, Jerusha B Detweiler, and Wayne T Steward. Emotional states and physical health. *American psychologist*, 55(1):110, 2000.

Tomasz Sapinski, Dorota Kaminska, Adam Pelikant, and Gholamreza Anbarjafari. Emotion recognition from skeletal movements. *Entropy*, 21(7):646, 2019.

Y-I Tian, Takeo Kanade, and Jeffrey F Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 23(2):97–115, 2001.

Jie Wei, Guanyu Hu, Xinyu Yang, Anh Tuan Luu, and Yizhuo Dong. Learning facial expression and body gesture visual information for video emotion recognition. *Expert Systems with Applications*, 237:121419, 2024.

Jianhua Zhang, Zhong Yin, Peng Chen, and Stefano Nichele. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, 59:103–126, 2020.

Yisu Zhao, Xin Wang, Miriam Goubran, Thomas Whalen, and Emil M Petriu. Human emotion and cognition recognition from body language of the head using soft computing techniques. *Journal of Ambient Intelligence and Humanized Computing*, 4:121–140, 2013.