Hand Alphabet Recognition for Dactylology Conversion to English Print Using Streaming Video Segmentation

Manuel B. Garcia*
College of Computer Studies, FEU
Institute of Technology

Teodoro F. Revano, Jr
College of Computer Studies, FEU
Institute of Technology

Armi Cunanan-Yabut
College of Engineering, FEU Institute
of Technology

ABSTRACT

Assistive technologies gained traction in the medical field over the last few decades. Novel approaches have been developed in order to support people with disability to communicate effectively. However, little research has been conducted on the other side of the coin, that is, assistive technologies to help people who do not have a disability to understand and comprehend the language of disabled. This study describes the early development of a hand alphabet recognition that intends to accomplish a functioning dactylology conversion from sign language to English print in a live streaming video. Through a video analysis, each frame is processed using a segmentation technique to partition it into different segments (e.g., pixels of hand gesture). The dactylology conversion algorithm was implemented in a mobile application where users can watch video containing an on-screen sign language interpreter and understand fingerspelling used as a communication by hearing- and speech-impaired people. Through the sample dataset of 13 videos of American Sign Language manually collected (N=10) and recorded (N=3), the application was tested for its accuracy in detecting the alphabet in a video (94.16%), and the correctness of conversion of the detected alphabet into English print (89.65%). This study contributes to the list of existing novel approaches that aims to promote social positive effects as well as improve the quality of life for both disabled and all the people they socialize with.

CCS CONCEPTS

• **Social and professional topics** → User characteristics; People with disabilities; • **Human-centered computing** → Interaction design; Interaction design process and methods; User centered design.

KEYWORDS

Dactylology, Assistive Technology, Sign Language, Video Segmentation, Fingerspelling, Hand Alphabet Recognition

*Corresponding Author. Email: mbgarcia@feutech.edu.ph

1 INTRODUCTION

In a world where millions of people are deaf-mute, Dactylology, or the science of communication using hands and fingers (e.g., one-handed alphabet, two-handed alphabet), is one of the, if not the only one, communication modalities that lets people with and without disability to express and send ideas and thoughts to each other. In fact, there are more than 120 distinctive sign lingos used in various nations such as American, French, German, Spanish, Filipino, Japanese, Indo-Pakistani, and more. The difficulty in establishing a universal communication modality between disabled people with different sign languages, and people with and without disability led to a stirring invitation of technology implementation. For instance, one study [1] developed a convolutional neural network model for American Sign Language alphabet recognition. The classification model was then combined with a multi-view augmentation strategy to exploit 3D information from depth images. On the other hand, artificial neural network was utilized in another study [2] to design and develop an American Sign Language recognition system with a sensory glove and a three-dimensional motion tracker for extracting gesture data features. These technology-based communication advancements are some of the most novel contributions to the vast reaches of healthcare information, intelligent application systems, and even communication technology.

In a deeper look, there has been a widespread of development of intelligent systems [3-6] and studies [7-9] that intend to establish and understand the usage of technology-based assistance to people with disabilities in performing basic communication tasks. Naves, Rocha, and Pino [10] developed an alternative communications system by uniting electromyographic (EMG) signals to the field of Human-Computer Interaction (HCI) to attend and serve patients severely disabled by amyotrophic lateral sclerosis. HCI and EMG were both incorporated to EDITH system – a computer software package consisting of communication features designed for a multimedia environment. In the robotics field, a mobile robotic arm was developed by Gushi, Higa, Uehara, and Soken for people with severe disabilities [11]. The robotic arm can perform several tasks by using eye movements, which are detected by an image processing technique. In another field, Garcia [12] developed a speech therapy game application to assist aphasic people in learning how to communicate again just like before their stroke occurred. Such technologies have been proven as an important tool and instrumental in promoting social positive effects as well as improving the

Figure 1: Image Segmentation in a Video Sequence

quality of life not only for disabled but also of their family and relatives. Santos et al. [13] confirmed the positive relationship between quality of life of people and assistive technology (e.g., VISIMP [14]) making it a sought-after invention of our time. These stigmatized and marginalized social groups have now a way to establish their position and promote inclusion within the society.

Granted, these novel approaches have been developed in order to aid people with disability to communicate effectively. Notwithstanding, little research has been conducted on the other side of the coin, that is, assistive technologies to assist people who do not have a disability to also understand and comprehend the language of disabled. In fact, most people do not clearly understand the sign language. Therefore, aside from the research gap, there is also a communication gap between the deaf communities and the public. This study describes the early development of a hand alphabet recognition that intends to achieve a functioning dactylology conversion from sign language to English print in a live streaming video. Through a video analysis, each frame is processed using a video segmentation technique to partition it into different segments (e.g., pixels of hand gesture). The dactylology conversion algorithm was implemented in a mobile application where users could watch video containing an on-screen sign language interpreter and understand fingerspelling used as a communication by hearing and speech-impaired people. Not only does the mobile application provide a new communication modality, it also sensitizes and offer awareness on how to communicate with deaf and mute people.

2 RELATED WORKS

The growth of multimedia information led to an extensive interest towards a video indexing of media information and video retrieval for accessing the acquired information stored in a database. For a better performance video indexing and retrieval system, a proper video segmentation algorithm must be applied [15]. According to Dhiman and Dhanda [16], video segmentation refers to the process of decaying a video data into meaningful segments that have strong correlation with the real world. In the computer science field, there are several schemes to use when performing video segmentation and every algorithm has different advantages and disadvantages. Figure 1 shows an example of how a single video frame is segmented to extract the signer.

Beevi and Natarajan [17] proposed a video segmentation algorithm for MPEG-4 encoding systems to build segmentation results with low computation load by using baseline, shadow cancellation, and adaptive threshold modes. Similarly, a frame-by-frame technique with computationally efficient results was proposed by Vora and Raman [18] through clustering of visually similar generic object segments in a video via extraction using top-k region proposal to generate preliminary masks of foreground object. Alternatively, Hassani et al. [19] managed to use a region merging process for spatial and motion information to implement their proposed time-consistent video segmentation algorithm for real-time application. Another technique in partitioning video information for further analysis is the graph-based hierarchical video segmentation [20]. This method used four main steps that starts with generating a graph for k-sized frame block followed by a calculation of hierarchical scales. Then, a calculation for the inference of video segmentations through a thresholding process is accomplished. Lastly, temporal coherence video segments are calculated by merging two consecutive segmented blocks. Li et al. [21], on the other hand, utilized an algorithm called suboptimal low-rank decomposition (SOLD) to decompose the representation coefficient matrix into sub-matrices of low ranks. The efficiency analysis revealed that this method is faster and more effective than *HGB* and *SHGB*. Further, intelligent approaches are likewise proposed for recognizing hand gestures in a natural manner. Chaudhary, Raheja, Das, and Raheja [22] grouped these approaches into fuzzy logic, genetic algorithm, and artificial neural networks. Verma and Dev [23] used fuzzy clustering based finite state machines to recognize hand gestures efficiently. Nolker [24], on the other hand, used neural network to distinguish fingertips transformable into finger joint angles of a hand model. This allowed a full reconstruction of a three-dimensional hand shape, with 16 segments and 20 joint angles. Hu, Yu, Li, and Ma [25] used an extraction of a human parametric 2D model to estimate human posture and recognize human activity. In their system, the genetic algorithm was applied to make a model with human silhouette.

Another area of related works that needs to be reviewed is the hand recognition system where the recognized gestures could be utilized as part of a more intelligent system [26], or for controlling a robot [27]. To produce gesture recognition systems, different approaches have been proposed from using additional hardware, such as gloves and color markers, to the use of skin-based segmentation for feature extraction [28-34]. The growing importance of gesture recognition in the society [35] led to a stimulating revolution among
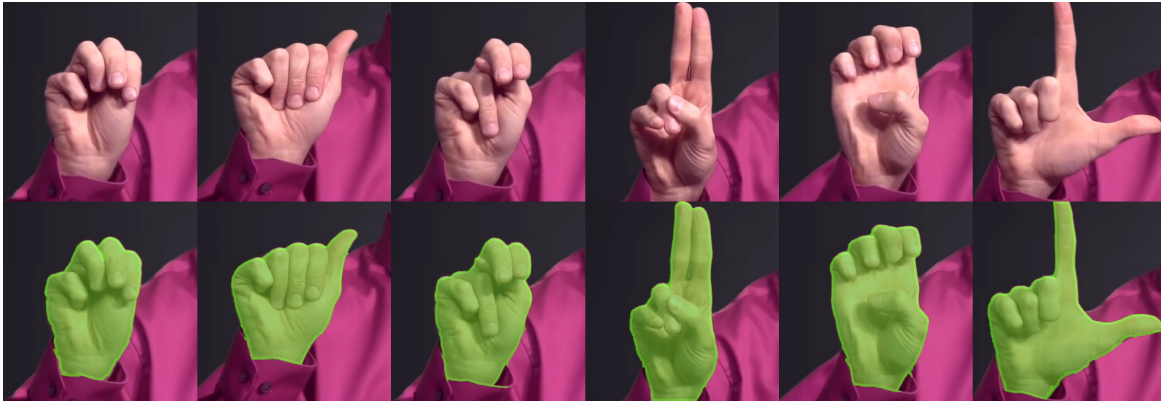
Figure 2: Frame-by-Frame Segmentation of Hand Gesture with Sign Language Alphabet Recognition

systems inventors and gave birth to essential applications in numerous areas like surveillance systems, robotics, HCI, healthcare, education, etc. In addition, sign language recognition has been likewise a beneficiary, and received a special attention from all the advancement of gesture recognition. Nevertheless, the development of gesture recognition systems has formed many lessons based on the drawbacks of existing and completed prototype. For instance, a neural network classifier is too time consuming (e.g., learning ten words in four days [36]) to make although progress in computer hardware created a faster result. An orientation histogram method becomes problematic when dealing with similar gestures but different histograms or different gestures with similar histograms [37].

3 METHODS

The main purpose of this study is to translate hand alphabet of the American Sign Language into English print in order to bridge a communication gap between people with and without disability when using dactylology as a communication modality. Towards the realization of this goal, various image and video processing techniques were utilized based on the experimental results of other existing studies. First, the strategy of Vora and Raman [18] in video object segmentation was the basis of the core processes of the hand gesture recognition. Moreover, a skin detection algorithm of Garcia et al. [38] was slightly mimicked particularly on image processing techniques used to enhance the accuracy result. For this application, each video frame undergoes several processes such us (1) color illumination restoration to recover details of a frame, (2) histogram equalization to redistribute color intensities, (3) lighting correction to re-adjust dark areas, and (4) noise reduction to remove unwanted pixels. For this to become possible, each frame must be extracted from the video media file. Afterwards, the extracted frames proceed to the core processes of the algorithm, such as (1) hand detection and extraction via an image segmentation with the mixture of background subtraction and three-frame-difference method, (2) object tracking based on a motion consistency algorithm [39], and finally, (3) hand alphabet recognition was performed using a convolutional neural network model for the classification process. Figure 2 shows a segmentation and tracking of hand gesture frame-by-frame to

automatically detect and recognize sign language alphabet. Meanwhile, the system block diagram that exhibits how the whole process works is illustrated on Figure 3.

3.1 Hand Detection and Extraction

After the preprocessing of video frames, the first step towards the recognition of hand features is segmentation. A common cue when segmenting body parts like a hand is the skin color [38] since it is invariant to scale and rotation changes [40]. However, the result of the segmentation is affected by illumination conditions. As such, a segmented skin colored-regions might not be a skin but of another region with similar color. Fortunately, a sign language interpreter customarily conveys the information using hand movements with a stationary body. Consequently, detecting a moving object in a video sequence will probably result to a moving hand. Weng, Huang, and Da [41] proposed a new interframe difference algorithm for detecting a moving object in a video through the combination of background subtraction and three frame-difference method. The first process is to subtract the current frame to the previous and next frames separately, and add together the results to generate a grayscale image. Afterwards, another grayscale image will be created by subtracting the current frame to the background image. The final output is a binary image made from the sum of two previously generated grayscale images, which helps in adding a bounding box on a region that has a constant moving object.

3.2 Hand Gesture Tracking

Once the target object is segmented and the feature is extracted, a hidden bounding box stays on that object for tracking purposes in order to perform the same process on the next frames until the end of a video sequence. He, Qiao, Wen, and Li proposed an object tracing based on a motion consistency (MCT) [39], which serves as the basis of the algorithm used in tracking the segmented hand. MCT states that the object that needs tracking must be known in the first frame. The segmented region during the hand detection and extraction processes determines the "known target object" prior to shifting to the state transition model, which is used to select the candidate samples in the current video frame. Subsequently, the target state prediction estimates the target position including
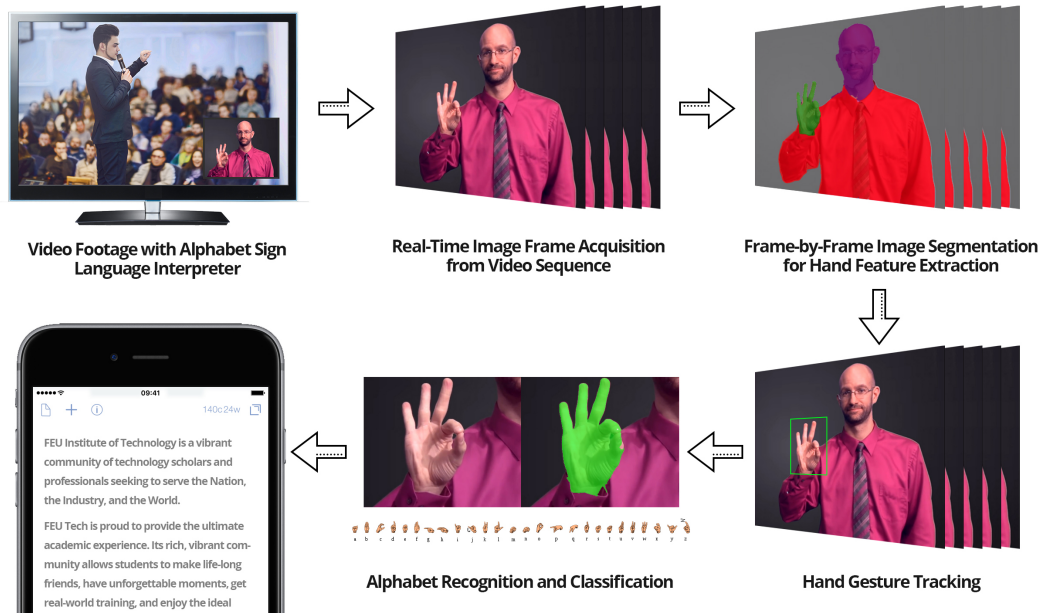
Figure 3: System Block Diagram of Hand Alphabet Sign Language Conversion to English Print

motion directions and distances based on a motion consistency. Hence, the tracking result is determined by the position factor with the holistic responses of each candidate.

3.3 Hand Alphabet Recognition

To classify the hand alphabet sign language, a convolutional neural network model for classification was tested, trained, and evaluated through the combination of various technologies such as Python programming language, openCV for real-time computer vision, and TensorFlow for machine learning with the inclusion of PIP packages such as *matplotlib*, *numpy*, *opencv*, and *tensorflow*. After classification, the alphabet sign language is converted into an English print and displayed on the mobile application. To recognize a word and/or to separate a string (converted sign language) into actual English words, a spellchecker and autocorrect algorithm was used through the comparison of translated sign language to an actual English word dictionary.

4 EXPERIMENTAL RESULTS

Through the sample dataset of 13 videos of American Sign Language manually collected (N=10) and recorded (N=3), the application was tested for its accuracy in detecting the alphabet in a video, and the correctness of conversion of the detected alphabet into English print. Upon the calculations, detection accuracy of 94.16% and conversion accuracy of 89.65% were obtained as shown on Table 1. Manual labeling was performed in all of the videos to determine the number of alphabets made by a sign language interpreter. A common error in the detection of alphabet is the recognition of fingerspelled letters that must be traced in the air such as "Z" and "J". In addition, "K" and "P" use a similar hand shape where the former is palm down while the latter is palm forward that confuses the

algorithm. Notwithstanding, the spellchecker and autocorrect algorithm was able to translate the sign language into correct English words especially for those words with problematic letters.

5 CONCLUSION AND RECOMMENDATION

In this study, a mobile application for hand alphabet recognition for dactylology conversion to English print was presented. Grounded from various algorithms and methodologies, the preliminary results of the experimental assessment specify a very encouraging outcome with a 94.16% detection accuracy and 89.65% conversion accuracy, at least based on the dataset supplied. One problem that needs to be addressed is the accuracy enhancement of the classification model particularly on letters with the same hand shape and fingerspelled alphabet that must be traced in the air. As of writing, the mobile application is still limited because it is an ongoing and the first phase of the project that focuses on conversion of sign languages. For future work and next phase of the project, a classifier will be modelled to detect words in a sign language to extend the hand alphabet. This extension of the algorithm is expected to be useful when communicating and enhancing interactions with people with disabilities. Conversion of sign language to English print then the text to speech audio is also feasible to remove the hassle of reading textual information but this is only a recommendation for future authors for now. Overall, this evaluation study presented a support to the list of existing novel approaches in promoting social positive effects as well as improving the quality of life for both disabled and all the people they socialize with.

ACKNOWLEDGMENTS

Table 1: Detection and Conversion Accuracy Results of Sample Sign Language Videos

Video Index	Number of Frames	Number of Alphabet	Detected Alphabet	Detection Accuracy	Correct Character	Conversion Accuracy
1	3744	145	144	99.31	121	84.03
2	3576	124	120	96.77	110	91.67
3	4176	169	151	89.35	140	92.72
4	3336	139	131	94.24	118	90.08
5	4104	171	169	98.83	149	88.17
6	5136	214	201	93.93	192	95.52
7	1824	76	72	94.74	66	91.67
8	2664	111	100	90.09	92	92.00
9	2352	98	92	93.88	89	96.74
10	3384	141	124	87.94	115	92.74
11	3096	129	120	93.02	103	85.83
12	1416	45	43	95.56	34	79.07
13	1512	56	54	96.43	46	85.19
Mean:	**3102**	**124**	**117**	**94.16**	**106**	**89.65**

REFERENCES

[1] W. Tao, M. C. Leu, and Z. Yin, "American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion," Engineering Applications of Artificial Intelligence, vol. 76, pp. 202-213, 2018/11/01/ 2018.

[2] C. Oz and M. C. Leu, "American Sign Language word recognition with a sensory glove using artificial neural networks," Engineering Applications of Artificial Intelligence, vol. 24, pp. 1204-1213, 2011/10/01/ 2011.

[3] B. L. Bayasut, G. P. Ananta, and A. K. Muda, "Intelligent biometric detection system for disabled people," in 2011 11th International Conference on Hybrid Intelligent Systems (HIS), 2011, pp. 346-350.

[4] D. Chiluisa-Castillo, F. Ortega-Barreto, V. Robles-Bykbaev, and F. Pesántez-Avilés, "An intelligent platform to design and develop low-cost assistive technologies and robotic assistants for children with disabilities," in 2018 IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), 2018, pp. 1-4.

[5] K. Kawamura, S. Bagchi, M. Iskarous, and M. Bishay, "Intelligent robotic systems in service of the disabled," IEEE Transactions on Rehabilitation Engineering, vol. 3, pp. 14-21, 1995.

[6] G. A. E. Khayat, T. F. Mabrouk, and A. S. Elmaghraby, "Intelligent serious games system for children with learning disabilities," in 2012 17th International Conference on Computer Games (CGAMES), 2012, pp. 30-34.

[7] A. Khetarpal, "Information and Communication Technology (ICT) and Disability," Review of Market Integration, vol. 6, pp. 96-113, 2014.

[8] M. Manzoor and V. Vimarlund, "Digital technologies for social inclusion of individuals with disabilities," Health and Technology, 2018.

[9] M. J. A. Zahid, M. M. Ashraf, B. T. Malik, and M. R. Hoque, "Information Communication Technology (ICT) for Disabled Persons in Bangladesh: Preliminary Study of Impact/Outcome," Berlin, Heidelberg, 2013.

[10] E. Naves, L. Rocha, and P. Pino, "Alternative communication system for people with severe motor disabilities using myoelectric signal control," in 2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), 2012.

[11] S. Gushi, H. Higa, H. Uehara, and T. Soken, "A mobile robotic arm for people with severe disabilities: Evaluation of scooping foods," in International Conference on Intelligent Informatics and Biomedical Sciences, 2017.

[12] M. B. Garcia, "A Speech Therapy Game Application for Aphasia Patient Neurorehabilitation – A Pilot Study of an mHealth App," International Journal of Simulation: Systems, Science & Technology, vol. 20, 2019.

[13] R. Santos, P. Sampaio, R. Sampaio, G. Gutierrez, and M. Almeida, "Assistive technology and its relationship to the quality of life of people with disabilities," Rev. Ter. Ocup. Univ. São Paulo, 2017.

[14] M. B. Garcia and N. U. Pilueta, "The VISIMP Portable Communications Device for Visually Impaired Individuals – Development and Feasibility Study of an Assistive Technology," Journal of Critical Reviews.

[15] D. M. Thounaojam, A. Trivedi, K. Manglem Singh, and S. Roy, "A Survey on Video Segmentation," New Delhi, 2014, pp. 903-912.

[16] P. Dhiman and M. Dhanda, "A Review on Various Techniques of Video Segmentation," International Journal for Innovative Research in Science & Technology, 2016.

[17] C. P. Y. Beevi and S. Natarajan, "A Novel Video Segmentation Algorithm with Shadow Cancellation and Adaptive Threshold Techniques," Berlin, Heidelberg, 2009, pp. 304-311.

[18] A. Vora and S. Raman, "Flow-free Video Object Segmentation," Computer Vision and Pattern Recognition, 2017.

[19] M. El Hassani, S. Jehan-Besson, L. Brun, M. Revenu, M. Duranton, Tschumperl, et al., "A Time-Consistent Video Segmentation Algorithm Designed for Real-Time Implementation," VLSI Design, vol. 2008, 2008.

[20] K. J. F. de Souza, A. D. A. Araújo, S. J. F. Guimarães, Z. K. G. do Patrocínio, and M. Cord, "Streaming Graph-Based Hierarchical Video Segmentation by a Simple Label Propagation," in 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, 2015.

[21] C. Li, L. Lin, W. Zuo, W. Wang, and J. Tang, "An Approach to Streaming Video Segmentation With Sub-Optimal Low-Rank Decomposition," IEEE Transactions on Image Processing, vol. 25, pp. 1947-1960, 2016.

[22] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, "Intelligent Approaches to interact with Machines using Hand Gesture Recognition in Natural way: A Survey," Human-Computer Interaction, 2013.

[23] R. Verma and A. Dev, "Vision based hand gesture recognition using finite state machines and fuzzy logic," in 2009 International Conference on Ultra Modern Telecommunications & Workshops, 2009, pp. 1-6.

[24] C. Nolker and H. Ritter, "Visual recognition of continuous hand postures," IEEE Transactions on Neural Networks, vol. 13, 2002.

[25] C. Hu, Q. Yu, Y. Li, and S. Ma, "Extraction of parametric human model for posture recognition using genetic algorithm," in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.

[26] R. Elakkiya, K. Selvamani, S. Kanimozhi, R. Velumadhava, and A. Kannan, "Intelligent System for Human Computer Interface Using Hand Gesture Recognition," Procedia Engineering, vol. 38, pp. 3180-3191, 2012.

[27] A. A. M. Faudzi, M. H. K. Ali, M. A. Azman, and Z. H. Ismail, "Real-time Hand Gestures System for Mobile Robots Control," Procedia Engineering, vol. 41, pp. 798-804, 2012/01/01/ 2012.

[28] N. Zengeler, T. Kopinski, and U. Handmann, "Hand Gesture Recognition in Automotive Human–Machine Interaction Using Depth Cameras," Sensors, vol. 19, p. 59, 2018.

[29] O. Sidek and M. Abdul Hadi, "Wireless gesture recognition system using MEMS accelerometer," in 2014 International Symposium on Technology Management and Emerging Technologies, 2014, pp. 444-447.

[30] L. Shi, Y. Wang, and J. Li, "A Real Time Vision-Based Hand Gestures Recognition System," Berlin, Heidelberg, 2010, pp. 349-358.

[31] S. S. Rautaray and A. Agrawal, "Real Time Gesture Recognition System for Interaction in Dynamic Environment," Procedia Technology, vol. 4, 2012.

[32] S. O. Oprea, A. Garcia-Garcia, S. Orts-Escolano, V. Villena-Martinez, and J. A. Castro-Vargas, "A long short-term memory based Schaeffer gesture recognition system," Expert Systems, vol. 35, p. e12247, 2018.

[33] J. I. Koh, J. Cherian, P. Taele, and T. Hammond, "Developing a Hand Gesture Recognition System for Mapping Symbolic Hand Gestures to Analogous Emojis in Computer-Mediated Communication," ACM Transactions on Interactive Intelligent Systems, vol. 9, pp. 1-35, 2019.

[34] G. Cicirelli, C. Attolico, C. Guaragnella, and T. D'Orazio, "A Kinect-Based Gesture Recognition Approach for a Natural Human Robot Interface," International Journal of Advanced Robotic Systems, vol. 12, p. 22, 2015.

[35] R. Khan and N. Ibraheem, "Survey on Gesture Recognition for Hand Image Postures," International Journal of Computer and Information Science, 2012.

[36] K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks," presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New Orleans, Louisiana, USA, 1991.

[37] W. T. Freeman and R. Michal, "Orientation Histograms for Hand Gesture Recognition," IEEE International Workshop on Automatic Face and Gesture Recognition, 1995.

[38] M. B. Garcia, T. F. Revano, B. M. Habal, J. Contreras, and J. Enriquez, "A Pornographic Image and Video Filtering Application Using Optimized Nudity Recognition and Detection Algorithm," in 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management, 2018, pp. 1-5.

[39] L. He, X. Qiao, S. Wen, and F. Li, "Robust Object Tracking Based on Motion Consistency," Sensors, vol. 18, p. 572, 2018.

[40] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," Engineering Applications of Artificial Intelligence, vol. 22, pp. 1141-1158, 2009/12/01/ 2009.

[41] M. Weng, G. Huang, and X. Da, "A new interframe difference algorithm for moving target detection," in 2010 3rd International Congress on Image and Signal Processing, 2010, pp. 285-289.