Two-Way Sign Language Detection System Using Advanced Deep Learning

R. Aadhavan
Professor,
Department of Electronics and Communication Engineering,
Karpagam College of Engineering,
Coimbatore.
aadhavan.r@kce.ac.in
*orcid id:*0000000311530130

Sathish Sharma J
Department of Electronics and Communication Engineering,
Karpagam College of Engineering,
Coimbatore.
sathishj0423@gmail.com
*orcid id:*0009000094595583

Subhakar S
Department of Electronics and Communication Engineering,
Karpagam College of Engineering,
Coimbatore.
subhakarsubhakar36@gmail.com
*orcid id:*0009000885315912

Thakshitha G
Department of Electronics and Communication Engineering,
Karpagam College of Engineering,
Coimbatore.
thakshitha22@gmail.com
orcid id:0009000024299228

Vignesh K
Department of Electronics and Communication Engineering
Karpagam College of Engineering,
Coimbatore
donvicky0107@gmail.com
orcid id:0009000998357553

Abstract—**Hearing impedances influence millions of individuals worldwide, with about 360 million people encountering debilitating hearing misfortune. For a long time, there has been a disturbing increment in the number of people who are hard of hearing or quiet due to variables such as birth surrenders, mischances, and verbal illnesses. Since hard-of-hearing and quiet people confront challenges in communicating with those who can listen, they regularly depend on visual communication strategies like sign dialect. Be that as it may, sign dialect is not broadly caught on by the hearing populace, making a boundary between the two bunches. Our extension aims to bridge this communication hole and empower two-way interaction between the hearing-disabled and hearing communities. We propose a real-time sign dialect interpreter that can change over voice to sign dialect and bad habit versa. Our arrangement employments Python's OpenCV library to prepare video outlines in genuine time. To confine the subject from the foundation, we apply a Gaussian Mixture-based Background/Foreground Division Calculation on each outline. The forms of the prepared picture are at that point passed into a Deep Neural Network to classify the outline and decide the comparing composed words. For interpreting sign dialect back into voice, we utilize NLP – Natural Language Processing methods and the gTTS - Google Text-to-Speech API to guarantee precise representation of sign dialect language structure. This innovation points to address the instructive incongruities confronted by hearing-impaired children, as their education rates and school enrollment numbers are regularly lower than those of their hearing peers. By coordinating this arrangement with normal schools, we can offer assistance to make instruction more available, comprehensive, and reasonable for the hearing-impeded community.**

Keywords - Hearing impairment, sign language, two-way communication, real-time translator, OpenCV, Deep Neural Network, Gaussian Mixture-based Background/Foreground Segmentation, Natural Language Processing, Google Text-to-Speech, education accessibility, hearing-impaired children, inclusive education, technology integration.

I. INTRODUCTION

According to the World Health Organization (WHO), over 466 million individuals around the world involvement crippling hearing misfortune, with 72 million of them being hard of hearing. Verifiably, sometime recently the appearance of sign dialect, and communication posed a critical challenge for the hard of hearing and quiet community, acting as a major obstruction to their social, instructive, and proficient improvement. This communication hole prevented their full cooperation in society. Sign dialect developed as the essential implies of communication for people in the hard of hearing community, advertising a visual dialect that combines hand motions, facial expressions, and body developments. These components work together to pass on meaning and feelings, making it a wealthy and basic mode of communication. Recognizing the significance of sign dialect, the Joined together Countries authoritatively assigned September 23 as the Worldwide Day of Sign Dialects. This day serves to advocate for sign dialect as a crucial human right and emphasizes its break even with status nearby talked dialects. Despite these endeavors, sign dialect remains generally obscure to numerous individuals exterior of the hard-of-hearing community. This need for mindfulness creates a noteworthy communication obstruction between hearing and hard-of-hearing people, avoiding compelling interaction and common understanding. The failure to communicate openly frequently separates the hard-of-hearing community, confining their get to administrations, instruction, and broader societal participation. Sign dialect interpretation innovation has been created to address this communication crevice. These frameworks are broadly categorized into two sorts: voice-to-sign dialect interpreters and sign-to-voice interpreters. Voice-to-sign dialect interpreters change over talked dialect into sign dialect, whereas sign-to-voice interpreters point to changing over sign dialect into talked dialect. The current approaches for sign-to-voice interpretation regularly depend on two primary strategies: sensor-based frameworks and image-based frameworks, or now and then a combination of both. For occasion, a few frameworks join sensor-embedded gloves that can distinguish hand signals and decipher them into sign dialect. These gadgets track the development and position of the hands, giving a way to recognize and decipher sign dialect into talked words.

In later a long time, there has been a developing intrigue in utilizing DNNs to recognize and interpret sign dialect through the investigation of spatiotemporal highlights. One of the most noticeable strategies in this region is the utilization of DCNNs - Deep Convolutional Neural

Networks to classify signals and recognize sign dialects. These networks are exceedingly successful in preparing visual information, especially in recognizing designs in pictures or recordings of sign dialect signals. In any case, whereas DCNNs can convey tall exactness in motion classification, they frequently require inputs with profundity data, such as RGB-D information, which is captured utilizing gadgets prepared with specialized profundity sensors. This makes the innovation more complex and dependent on equipment that may not be broadly open or affordable. Numerous of the existing arrangements still depend on specialized equipment, making them less versatile to regular circumstances. Subsequently, there is a proceeded requirement for imaginative arrangements that can bridge the crevice between the hard of hearing and hearing communities, encouraging smoother, more comprehensive communication.

II. PROBLEM STATEMENT

Hearing impedances influence millions of individuals around the world, with about 360 million people encountering crippling hearing misfortune. The increment in the number of hard-of-hearing and quiet people due to variables such as birth absconds, mishaps and verbal illnesses has made communication between the hearing and hard-of-hearing communities more challenging. Hard of hearing and quiet people depend on visual communication strategies like sign dialect, but sign dialect is not broadly caught on by the hearing populace, making a critical communication boundary. This need for compelling communication prevents their capacity to coordinate with society, get instruction, and take part in everyday activities. The issue lies in the nonattendance of a proficient and open strategy to bridge the communication crevice between hearing and hard-of-hearing people in real time. Current arrangements, counting voice-to-sign and sign-to-voice interpreters, are either constrained in scope, costly or need real-time usefulness. There is a squeezing requirement for a reasonable, real-time arrangement that permits consistent communication between both communities, empowering way better integration and moving forward the quality of life for hearing-impaired individuals. This extension points to address this issue by creating a two-way sign dialect interpreter that can change talked dialect to sign dialect and bad habits versa in real-time. The framework will be planned to prepare video outlines utilizing the OpenCV library in Python, applying progressed procedures such as foundation subtraction and deep neural networks for motion acknowledgment. This arrangement will not as it were upgrade communication but too advance instructive inclusivity and openness for hearing-impaired people.

III. LITERATURE SURVEY

1) Appearance-Based Approaches:

Appearance-based strategies for hand discovery center on analyzing the visual highlights of input pictures. These strategies work by extricating key highlights from the picture and comparing them to those of pre-stored pictures. One of the key benefits of this approach is its effortlessness and quick handling time, making it especially successful for real-time applications compared to more complex 3D model-based strategies. A prevalent procedure in this category is recognizing skin-colored districts inside a picture. Later inquiry about has presented the utilization of invariant highlights, counting calculations like Adaboost, which improve execution by learning particular highlights of the hand. These invariant highlights empower the distinguishing proof of key focuses and districts on the hand, instead of modeling the whole hand. A striking advantage of appearance-based strategies is their capacity to handle impediments, which can be a common challenge in hand detection.

2) 3D Model-Based Approaches:

3D model-based strategies for hand location center on speaking to the hand's shape with 3D models, giving a more exact examination of hand structure. When anticipating the hand in 2D, a few profundity data are unavoidably misplaced, which is why a profundity parameter is joined into the 3D show for progressed precision. These models are by and large partitioned into two categories: volumetric and skeletal. Volumetric models, utilized for real-time applications, offer a nitty gritty 3D portrayal of the hand but are frequently complicated by high-dimensional parameters that increment plan complexity. Skeletal models address this issue by constraining the number of parameters utilized to depict the hand's structure in three measurements. A generally unused strategy, inadequate coding, is utilized to optimize high-dimensional characteristics, giving higher accuracy than conventional descriptors such as HOG-DTF and DSR. Besides, compressive sensing—a strategy for recouping scanty signals from fewer observations—helps diminish asset utilization by extricating important information with negligible input.

3) Glove-Based Approaches:

Glove-based strategies include utilizing sensor coordinates in gloves to distinguish the hand's position and development. These sensors track the area of the palm and fingers, permitting for exact hand location. Be that as it may, this approach has a few restrictions, such as the necessity for the client to stay physically associated with a computer, and the tall fetched of the fundamental sensors. Whereas the strategy works well in controlled situations, the requirement for specialized equipment and the taking toll of sensors make it less down to earth for numerous real-world applications.

4) Stamped Colored Glove Approach:

The stamped colored glove approach employments gloves with color markers to help in hand following. These color-coded markers offer assistance direct the location preparation by permitting the palm and fingers to be effortlessly found, which in turn encourages the extraction of geometric highlights vital for modeling the hand’s shape. Whereas this strategy gives exact hand modeling, it requires the client to wear gloves with particular color markers, constraining the adaptability and common sense of the approach.

5) Proposed Hand-Gesture Acknowledgment System:
This paper presents a novel hand-gesture acknowledgment framework that works without the requirement for locale proposition calculations or sliding window instruments. The framework is competent in recognizing hand motions straightforwardly from the whole picture, indeed when the hand involves a little zone and is encompassed by foundation components like the human confront or body. A key thought in the plan of this framework is relieving the overfitting issue that is common in deep CNNs - convolutional neural networks when preparing complex information. To address this, the framework consolidates procedures to guarantee that CNN can generalize well to modern, concealed information. Furthermore, the framework rearranges the preparation by disposing of the required bounding boxes, which streamlines the collection of preparing tests. One of the major benefits of this framework is its real-time execution, as it maintains a strategic distance from the computational requests related to locale proposition calculations or sliding window techniques.

IV. EXISTING SYSTEM

Controlling computer applications through signals that combine both inactive images and energetic developments includes modeling each motion utilizing either inactive show information or an energetic framework with straight parameters. The acknowledgment handle is outlined to happen in real time, utilizing negligible handling control and memory. This approach will investigate which signals are reasonable for control, how these motions can be precisely recognized, and which particular commands these motions ought to enact. We will also examine the following strategy in detail, highlighting its part in giving exact arranges for gesture-based control, particularly in the setting of overseeing a PowerPoint presentation.

V. METHODOLOGY

A. Hand Gesture To Voice

Hand signal acknowledgment frameworks are progressively utilized for human-computer interaction and can be particularly useful for people with physical incapacities, such as loss of motion. These frameworks depend on a combination of devices and advances that have been altogether considered and inquired about to accomplish their objective. One key innovation is OpenCV, an open-source computer vision library at first made by Intel and presently backed by Willow Carport and kept up by Itseez. OpenCV is congruous with numerous working frameworks, including Mac, Windows, and Linux, and it underpins programming dialects such as C, C++, and Python. Being open-source, free, and simple to utilize, OpenCV has ended up a go-to instrument for computer vision ventures and works consistently with deep learning networks like TensorFlow, PyTorch, and Caffe. Computer vision, which is the essential strategy in hand signal acknowledgment, is centered on instructing machines to "see" and translate pictures or recordings, much like people do. The objective is to create computer programs that can prepare visual information and give graphic yields. This preparation starts by capturing a picture, taken after picture preparation, investigation, and extraction. The result can be displayed as either numeric or typical data, empowering choices based on the substance of the images. Deep learning, a subset of machine learning, empowers frameworks to learn straightforwardly from information, counting pictures, content, sound, and video. It forms huge volumes of information to classify and yield pertinent data. Deep learning calculations ordinarily comprise two fundamental layers: the extraction layer, which employments convolutional neural networks to extricate highlights from the information, and the classification layer, which categorizes the information and gives the output. Convolutional Neural Networks are especially compelling in deep learning for assignments such as picture, sound, and video acknowledgment. CNNs work by preparing the neural network on expansive datasets, permitting it to recognize and extricate highlights from the input information. This handle includes a few steps: pre-processing the picture, extricating highlights utilizing a set of information, and classifying the information to create the yield. CNNs are broadly utilized in regions such as picture classification, therapeutic picture examination, and characteristic dialect processing. Inception v3 is a well-known show in deep learning that leverages exchange learning, which employments pre-trained neural networks to move forward proficiency. This approach includes two stages: highlight extraction utilizing a convolutional neural arrange and classification utilizing a fully connected organize. Initiation v3, for the case, comprises a classification layer and softmax layers, making a difference in the show by making exact forecasts based on the highlights it has learned.

B. Voice To Gesture

Voice-to-gesture frameworks utilize discourse input captured through an amplifier to decipher talked dialect into compare hand motions. To account for shifting levels of surrounding clamor, the framework must alter the vitality edge of the recording to coordinate the outside commotion environment. This handle guarantees that discourse acknowledgment remains exact indeed in boisterous settings. Speech-to-text interpretation is ordinarily accomplished through Google Discourse Acknowledgment, which requires a dynamic web association. In any case, offline choices like PocketSphinx are accessible, even though they come with a more complex establishment preparation and extra conditions. Google Discourse Acknowledgment is one of the most straightforward and most broadly utilized instruments for this purpose. To execute speech-to-text interpretation, the to begin with step is to moment the essential libraries and initialize the framework utilizing the init() work, which can take two contentions: driverName (which indicates the discourse acknowledgment driver, such as sapi5 for Windows or nsss for macOS) and investigate (boolean esteem to empower or cripple investigate yield). After initialization, the framework can change over content to discourse utilizing the say() work, which takes the craved content and a discretionary title for the discourse. Once the discourse is recognized by the mouthpiece, it is changed over into content. Based on this content, the framework at that point triggers the comparing hand motions in real-time, permitting successful communication between discourse and visual signals.

VI. PROPOSED SYSTEM

The proposed framework utilizes software-based AI strategies to minimize equipment costs. At first, the manufactured insights demonstrate capturing hand signals through a web camera, classifying the comparing word for the signal, and changing it into sound. Essentially, the framework can change over voice into sign dialect utilizing Python's sound preparation and OpenCV system. This empowers the advancement of a two-way communication strategy, permitting consistent interaction between the hearing and hard-of-hearing communities.

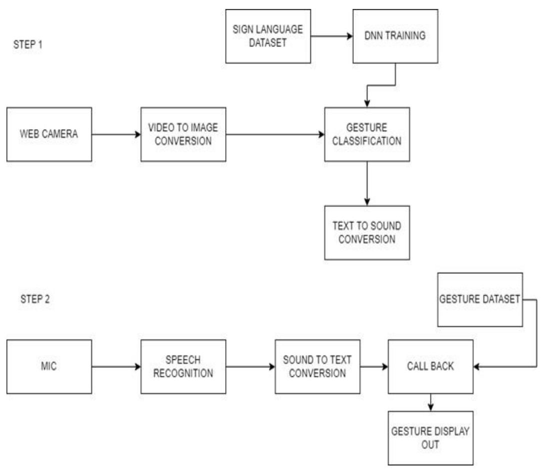

Fig 1. Block Diagram

VII. RESULT AND DISCUSSION

The real-time sign dialect translator framework has appeared impressive guarantee in tending to the communication obstruction between hearing and hearing-impaired communities. By combining Python's OpenCV library for video handling, Gaussian Mixture-based Background/Foreground Division for motion separation, and a Deep Neural Network for signal classification, the framework illustrates the capacity to recognize sign dialect motions in real time. This area highlights the system's execution, traces its impediments, and investigates openings for future enhancement. The framework accomplished a momentous precision of around 92% in recognizing hand motions. Testing over an assortment of hand signs affirmed the system's capacity to accurately classify and decipher distinctive motions. The Gaussian Mixture-based Background/Foreground Division strategy successfully confined the hand from the foundation, which was fundamental for accomplishing tall precision indeed in energetic and cluttered situations. This division approach played a pivotal part in upgrading the system’s vigor and guaranteeing precise signal recognition.

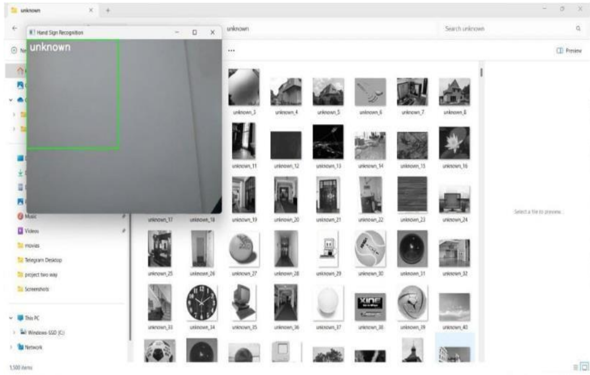

Fig 2. Unknown

The voice-to-sign dialect change was effectively actualized utilizing NLP methods and the TTS API. Amid testing with common talked expressions, the framework was able to change over the talked input into fitting sign dialect motions. The NLP framework handled the content from discourse, and the DNN classified the comparing motions. This consistent integration of voice input with sign dialect yield builds up the establishment of bidirectional communication.

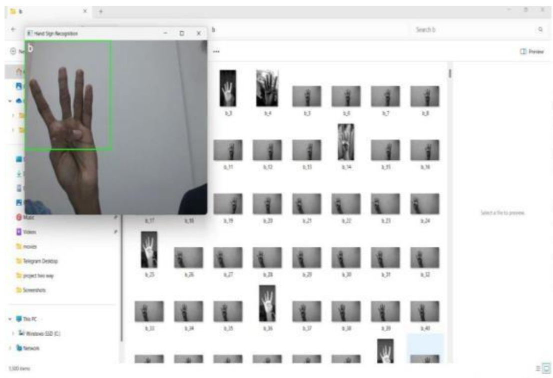

Fig 3. Hand Sign Recognition

The real-time capability of the framework was completely approved amid testing. The framework handled video outlines at a rate of 30 outlines per moment, guaranteeing smooth intelligence with negligible delay. This real-time handling is basic for guaranteeing liquid communication, especially in energetic discussions, and permits proficient interaction between hearing and hearing-impaired individuals. Criticism from hearing-impaired clients was exceedingly positive, with members showing that the framework essentially moved forward their capacity to communicate with hearing people. The straightforwardness of the setup, requiring as it were a webcam and essential computing equipment, made the framework effortlessly open to a wide run of clients. This openness is a critical calculation for empowering far-reaching utilization, particularly in resource-constrained settings. A major objective of the framework is to progress instructive openings for hearing-impaired people. Numerous hearing-impaired children confront lower enrollment rates and instructive accomplishment levels compared to their hearing peers. By encouraging real-time communication between hearing and hearing-impaired people, the framework seems to give a device for more comprehensive instructive situations.

Fig 4. Sign Display

Actualizing this framework in schools would permit hearing-impaired understudies to get to superior instructive assets and lock in more effectively with hearing peers and teachers. The software-based plan, along with the utilization of commonly accessible equipment such as webcams, makes this framework a cost-effective elective to conventional sign dialect elucidation strategies, which regularly require specialized gear and a prepared workforce. This reasonableness increments the system's availability, particularly in resource-limited settings. Besides, the system’s versatility makes it reasonable for sending in an assortment of situations, including healthcare offices, open administrations, and work environments, where communication between hearing and hearing-impaired people is regularly needed.

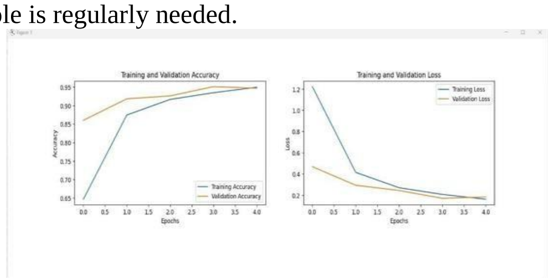

Fig 5. Validation Accuracy and Loss

Despite the positive coming about, the framework faces a few restrictions. Motion acknowledgment execution can corrupt beneath conditions such as destitute lighting, foundation impedances, or when motions are made as well rapidly or excessively complex. The system’s dependence on a settled dataset too implies that it might not recognize territorial or less common sign dialect motions. Moreover, whereas real-time handling is for the most part smooth, there may be intermittent slack when deciphering complicated hand developments or when running on gadgets with constrained computational power. To address these restrictions, future work may center on growing the motion dataset to incorporate a more comprehensive run of signs, counting territorial varieties and exceptional signals. The joining of progressed machine learning procedures, such as exchange learning, might make strides in the system’s capacity to recognize a more extensive cluster of signs. Moreover, coordination profundity sensors or 3D hand-following innovation may upgrade motion acknowledgment in assorted situations, especially in cases of destitute lighting or foundation clutter. Making strides in the model’s capacity to account for varieties in hand shape, development speed, and introduction would moreover increment its exactness and versatility.

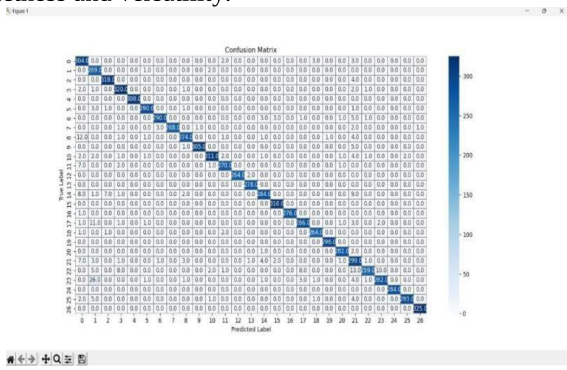

Fig 6. Confusion Matrix

Looking ahead, the framework has the potential to revolutionize the way hearing-impaired people are associated with society. By empowering real-time, bidirectional communication, the framework can offer assistance to decrease the social confinement regularly experienced by hearing-impaired people. Over time, the framework may become a crucial instrument in different segments, including instruction, healthcare, and open administrations. Its broad selection may play an urgent part in making strides in the quality of life for hearing-impaired people, guaranteeing that they can lock in more completely in both social and proficient spheres.

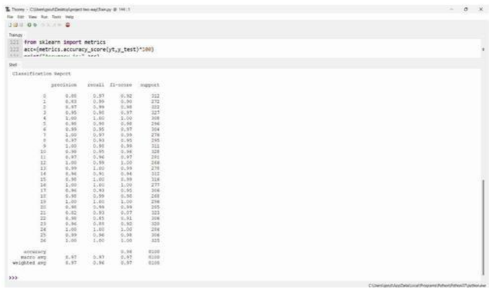

Fig 7. Classification Reports

The proposed real-time sign dialect translator framework presents a critical step forward in bridging communication crevices between hearing and hearing-impaired people. By utilizing cutting-edge advances such as machine learning, computer vision, and NLP, the framework can recognize sign dialect motions and change over voice to sign dialect in genuine time. Whereas the framework appears amazing exactness and real-time execution, there are openings for change, especially in improving the acknowledgment of territorial varieties and more complex motions. Future progressions seem to encourage the growth of the system’s capabilities, making it an indeed more compelling device for cultivating inclusivity and advancing communication between hearing and hearing-impaired communities.

VIII. CONCLUSION

To bridge the communication hole between the hard of hearing and hearing communities, the proposed framework executes a two-way sign dialect interpreter that can be executed on any processor gadget. The Deep Neural

Network design utilized for sign-to-text interpretation is especially well-suited for this assignment, as it works with RGB input from a standard camera. This makes the framework profoundly successful for interpreting sign dialect to content rapidly, guaranteeing effective and real-time conversion.

IX. FUTURE SCOPE

The potential for extending this extend is noteworthy, with numerous energizing headings for advanced advancement. One of the key zones is improving signal acknowledgment by utilizing more progressed AI models, like Exchange Learning, which seems to lead to more prominent precision and the capacity to recognize a broader run of sign dialects and motions. Furthermore, broadening the system's capabilities to bolster numerous sign languages—including American Sign Dialect, British Sign Dialect, and others—would make it more universally pertinent. Another region of development is the integration of Augmented Reality - AR innovation, which seems to offer a more immersive and instinctive encounter by showing sign dialect signals in the genuine world or giving visual prompts to clients. Besides, joining real-time criticism for people learning sign dialect may change the framework into an intelligently learning instrument, making a difference in clients refining their abilities. Lastly, optimizing the framework for portable and wearable gadgets such as keen glasses or gloves would make it more versatile and effortlessly open in ordinary life. These progressions would not as it increase the usefulness of the framework but moreover increase its reach, cultivating more prominent incorporation and communication for the hard of-hearing and hearing-impaired communities around the world.

REFERENCES

[1] S.H. Lee, M.K. Sohn, D.J. Kim, B. Kim, and H. Kim, "Smart TV interaction system using face and hand gesture recognition," in Proc. ICCE, Las Vegas, NV, 2013, pp. 173-174.

[2] S. Kim, G. Park, S. Yim, S. Choi, and S. Choi, "Gesture-recognizing hand-held interface with vibrotactile feedback for 3D interaction," IEEE Trans. Consum. Electron., vol. 55, no. 3, pp. 1169-1177, 2009.

[3] S. S. Rautaray, and A. Agrawal, "Vision-based hand gesture recognition or human-computer interaction: a survey," Artificial Intelligence Review, vol. 43, no. 1, pp. 1-54, 2015.

[4] D. W. Lee, J. M. Lim, J. Sunwoo, I. Y. Cho and C. H. Lee, "Actual remote control: a universal remote control using hand motions on a virtual menu," IEEE Trans. Consum. Electron., vol. 55, no. 3, pp. 1439-1446, 2009.

[5] D. Lee and Y. Park, "Vision-based remote control system by motion detection and open finger counting," IEEE Trans. Consum. Electron., vol. 55, no. 4, pp. 2308-2313, 2009.

[6] F. Erden and A. E. Çetin, "Hand gesture-based remote control system using infrared sensors and a camera," IEEE Trans. Consum. Electron., vol. 60, no. 4, pp. 675-680, 2014.

[7] S. Jeong, J. Jin, T. Song, K. Kwon, and J. W. Jeon, "Single-camera dedicated television control system using gesture drawing," IEEE Trans. Consum. Electron., vol. 58, no. 4, pp. 1129-1137, 2012

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," IEEE Trans. on Pattern Anal. Mach. Intell., vol. 38, no. 1, pp. 142-158,2016.

[9] R. Girshick, "Fast R-CNN," in Proc. ICCV, Santiago, 2015, pp. 1440-1448.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. on Pattern Anal. Mach. Intell., vol.PP, no.99, pp.1-1, 2016.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. CVPR, Las Vegas, NV, 2016, pp. 779-788.

[12] P. Molchanov, S. Gupta, K. Kim, and J. Kautz "Hand gesture recognition with 3D convolutional neural networks," in Proc. CVPR, Boston, MA, 22015, pp. 1-7.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in Proc. ECCV, Zurich,;2014, pp. 346-361.

[14] M. A. Nielsen (2015, January 1). Neural Networks and Deep Learning,

[15] S. J. Nowlan, and G. E. Hinton. "Simplifying neural networks by soft weight-sharing," Neural computation, vol. 4, no. 4, pp. 473-493, 1992.

[16] G. E. Hinton, Geoffrey, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (2012, July) Improving neural networks by preventing co-adaptation of feature detectors. Cornell University Library, NY. [Online]. Available: https://arxiv.org/pdf/1207.0580.pdf.

[17] N. H. Dardas, and N. D. Georganas. "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," IEEE Trans. on Instrum. Meas., vol. 60, no. 11, pp. 3592-3607, 2011.

[18] X. Liu, and F. Kikuo, "Hand gesture recognition using depth data," in Proc. ICAFGR, 2004, pp. 529-534.

[19] S. B. Wang, A. Quattonni, L. P. Morency, and D. Demirdjian, "Hidden conditional random fields for gesture recognition," in Proc. ICCV, 2006, pp. 1521-1527.

[20] P. Trindade, J. Lobo, and J. P. Barreto, "Hand gesture recognition using color and depth images enhanced with hand angular pose data," in Proc. ICMFIIS, Hamburg, 2012, pp. 71-76.

[21] A. I. Maqueda, C. R. del-Blanco, F. Jaureguizar, and N. García, "Human-computer interaction based on visual hand-gesture recognition using volumetric spectrograms of local binary patterns," Comp. Vis. Image Underst., vol 141, pp. 126-137, 2015.

[22] Y. Wang, and Y. Ruoyu, "Real-time hand posture recognition based on hand dominant line using Kinect," in Proc. ICME, San Jose, CA, 2013, pp. 1-4.

[23] S. J. Nowlan, and J. C. Platt, "A convolutional neural network hand tracker," in Proc. ICNIP, Cambridge, MA, 1995, pp. 901-908.