An Application for Hearing Impaired that Understands Signal Language from Hand Gestures

Muhammed Dogan, Berker Uysal and Pinar Kirci

Bursa Uludag University, Gorukle, Bursa, 16285, Turkey

Abstract

Today, computers, which are used in all areas of life and have become an inseparable part of our lives, contribute to human life by using advanced object recognition and image processing technologies. It is actively used in various fields such as medicine, military, security, traffic, industry, agriculture, astronomy, retail, environmental safety, geodesy and photogrammetry and provides great benefits. Based on all these, the recognition of sign language by computers and its translation into voice and writing will facilitate the work of individuals with disabilities and those who do not know sign language in daily life. In this study, it is aimed to recognize sign language letters by using image processing technologies and to convert them into sound and text at the same time.

Keywords

Sign language, object recognition, image processing, machine learning

1. Introduction

Object recognition is a technology that is actively used and utilized in many areas of daily life. With the development of technology from past to present, computers have taken over many jobs done by human beings. The usage areas of computers have expanded to many sectors from agriculture to textile, automotive and animal husbandry [1].

Image processing and object recognition technologies are among the technologies that help people the most. With the increasing workload in many areas, manpower has been insufficient and advanced technologies such as image processing and object recognition have been used instead. The usage areas of these technologies are increasing day by day [2].

Today, people with hearing and speech impairments communicate through sign language. Sign language is very important for people with disabilities to communicate [3]. They can easily communicate with sign language, express their needs and chat.

However, the work of people with disabilities, unfortunately, does not get easier when they know sign language. The biggest problem they face is that the people around them do not know sign language and therefore it becomes very difficult to communicate. Since sign language cannot be taught to all people, the best way to overcome this communication problem is to use an app. With this application, everyone will be able to understand each other easily without making any extra effort [4].

First of all, the most important part is to know the sign language well, understand its structure, pay attention to international standards and approach it as a whole. For the success of the application, it is necessary to know exactly what the sign means and to treat it as a constant. These operations can be done using object recognition technology. First of all, it is necessary to analyze the sign language letters and create a database for signs using the alphabet.

COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems, April 20–21, 2023, Kharkiv, Ukraine
EMAIL: 031890129@ogr.uludag.edu.tr (M. Dogan); 031890128@ogr.uludag.edu.tr (B.Uysal); pinarkirci@uludag.edu.tr (P.Kirci)
ORCID: 0000-0002-0442-0235 (P.Kirci)


However, there are certain points to be noted in this regard. One of the most important points to be decided will be the size of the database. The size of the database needs to be well adjusted. The fact that the number of data is neither too little nor too much will negatively affect the situations and functionality [1].

According to the data of the World Health Organization (WHO), there are 466 million hearing-impaired individuals worldwide. 34 million of these people are children under the age of 15. For this reason, different sign languages have been developed for hearing impaired individuals to communicate in every period of history [4].

In [5], a novel dynamic sign language recognition method was given. It utilizes trajectory and key hand type to extract features. It adopts a key frame weighted DTW (dynamic time warping) algorithm to implement hierarchical matching strategy. The method gradually matches sign language gestures from two levels of trajectory and key hand type.

The presented study in [6] improved a smart wearable American Sign Language (ASL) interpretation model. It used deep learning method. The presented model applied sensor fusion to integrate features from six inertial measurement units (IMUs). The presented smart wearable ASL interpretation model aimed to assist hearing-impaired person to communicate with society in best way.

The study in [7] focused on the use of the knowledge of phonology of Japanese sign Language (JSL) and dictionary to improve a real-time JSL sign recognition system. The system employs Kinect v2 sensor to collect sign features: hand shape, position, and motion. Depth sensor provides real-time processing and robustness against environmental changes.

Sign language words composed of three elements, these are, hand's motion, position, and shape, in terms of phonology. In [8], a recognition system was developed for Japanese sign language (JSL) with abstraction of manual signals based on these three elements. The abstraction of manual signals is performed with Japanese sign language words dictionary. Features like coordinates of hands and depth images are extracted from manual signals with the depth sensor and Kinect v2.

In [9], it presented an algorithm for segmenting videos of signs into sequences of still images and four techniques for Arabic sign language recognition: Modified Fourier Transform (MFT), Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), and combination of HOG and Histogram of Optical Flow (HOG-HOF). And, these techniques are evaluated using Hidden Markov Model (HMM).

A survey on dynamic SLR was presented in [10]. It includes two categories, typically mentioning HMM, some main datasets in variable languages and methods used for data preprocessing.

In [11], American sign language is used. The paper works on helping specially abled people to communicate with people who don't know sign language with utilizing the approaches of computer vision and deep learning. To solve this problem, the paper uses convolutional neural network. Firstly, the paper focuses on capturing variable hand expressions in the form of video by the person and translating them to text using a convolutional neural network. The other part targets on the reverse of it, showing GIF upon converting text. And, integrating these two parts helps in two-way communication.

The research paper [12] compared three optimization types of gradient descent method, that are stochastic gradient descent algorithm, adaptive gradient algorithm, root mean square propagation algorithm. According to the results, it is shown that root mean square propagation (RMSProp) is that the first-class optimizer to preserve the loss of model performs of Optimized Convolution Neural Network (OCNN) capacity in managing sign language recognition.

This paper focuses on state-of-the-art literature that identifies areas of interest in the non-visual inputs, image frames, and video frames to determine the features for a particular hand gesture. The literature survey also takes into account the approaches considered by researchers across different sign languages like American Sign Language, Taiwanese Sign Language, etc. which will help to develop a perspective for Indian Sign Language.

2. Sign language

Communication has been a very important factor for people to get along with each other throughout history. It can be defined as the transmission of information from one person to another, any kind of exchange of meanings such as feelings, thoughts and ideas. Verbal communication, which is based on hearing and speaking, is mostly used among people [13].

But people with hearing loss have difficulty communicating with people with hearing loss or other people like themselves. Sign languages have emerged to overcome this difficulty. Sign language is a silent and visual language that requires us to use gestures and facial expressions that enable us to communicate with individuals who do not have the ability to hear or speak [4].

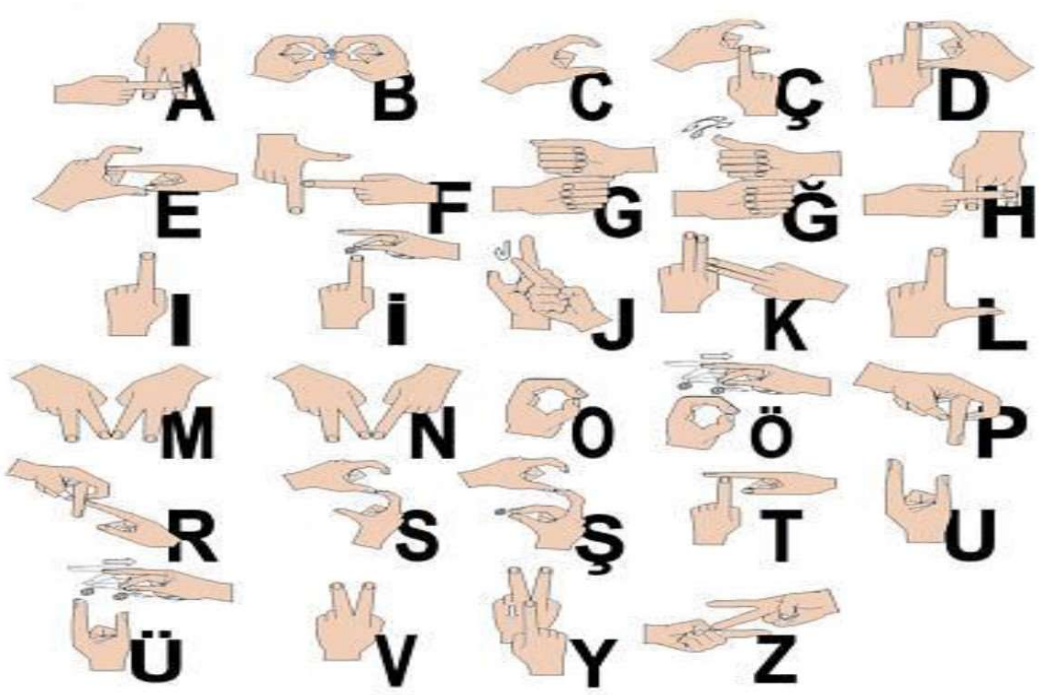

Figure 1: Turkish Sign Language Alphabet

The earliest written records of sign language dates back to the 5th century BC. Until the 19th century, most of the knowledge known about sign languages was limited to hand-made alphabets produced to facilitate the transfer of words from oral language to sign language rather than documents [14].

In the early 1770s in France, hand gestures used by hearing impaired individuals were accepted as a grammatical language and started to be taught to people in schools. Later, this method used in the language was carried to America by a French sign linguist. In 1817, Thomas Gallaudet established the first sign language school in the United States, which gave education only to hearing-impaired individuals [4].

The gesture is an activity. It can be characterized as movement, hand or face frequency that communicates thoughts, sentiments, or feelings, such as, raised eyebrows; shoulder movements are a portion of the activities people use in their lives. Gesture-based communication is a more formal and expressive type of correspondence. Here all words or letters are given a particular activity. Gesture-based communication is an all-around coded demonstration of code; every activity has an appointed significance. Communication with signing is the solitary method for correspondence for the hard of hearing [15].

Every country owns different Sign Languages. The gestures and movements of hands, fingers, and the meaning they represent differs. There are many research and implementations are done on SL: British Sign Language (BSL), Indian Sign Language (ISL), Chinese Sign Language (CSL), American

Sign Language (ASL). Implementations own the development of Automated Systems that convert or translate the evolved languages. Also, ISL was a recently developed language.

The paper [16] worked on the literature that identifies areas of interest in the non-visual inputs, image frames, and video frames to determine the features for a particular hand gesture. The literature survey considers the approaches worked by researchers across variable sign languages: American Sign Language, Taiwanese Sign Language, etc. that helps to develop a perspective for Indian Sign Language.

Human interpreters are not available for identifying ISL. Fort his reason, Sign Language Recognition System (SLRS) is focussed on for the identification of ISL. By the way designing an SLRS for ISL is very difficult when compared to other SL. ISL is very complex because it is composed of single- and double handed gesture and owns an extensive vocabulary with similar gestures.

The paper [17] presents a study of ISL, its syntax/vocabulary and different trending techniques for designing an ISL recognition system.

Sign language is a media for communicating with deaf and dumb people and it is not known by most of the normal people. Thus, it is a difficult task to form a communication between normal people and hearing impaired person. There are many tools are presented to help them, but unfortunately not produce accurate results. To be able to communicate with them, various fingers' gestures are used and then, a designed model converts those gestures into words or alphabets into a specific language [18].

It is not known how many sign languages exist in the world. Generally, each country has its own sign language. In some countries, this number is higher. Ethnologue counts 137 sign languages in its 2013 edition [14].

The Turkish Sign Language shown in Figure 1 is one of these sign languages. In the Ottoman Empire, there are records that sign language was used in palaces, baths and even courts since the 16th and 17th centuries. However, it is not known whether today's sign language is the sign language used in the Ottoman period [19].

3. Proposed system

In the study, it is aimed to add basic words in sign language to the system as well as letters. The system also aims to convey the result displayed on the screen to the user audibly after completing the necessary steps.

Working stages of the system;

- Saving alphabet letters and basic words in database,
- Taking new images with the camera,
- Comparison of the received image with the database,
- Output of the value obtained as a result of the comparison,
- It can be explained as the last output to be transmitted to the user audibly.

The default camera on the computer was used for the collection of images and recordings.
One of the limitations in the study is the environmental restriction. The environment has been a major constraint factor in the use of images, as the work is constantly intertwined with processes such as image data collection, processing and detection. The environment should be as uniform and stable as possible in order to properly collect and perceive images.

The operating performance of the system is largely dependent on the environmental factor. If there is a lot of noise during object detection in the environment, the detection process becomes very difficult. Since image processing operations are costly and demanding computer performance, they must be as fast as possible. For this case, GPU rather than CPU should be used. In this case, having a good GPU is an important criterion for determining system performance [20].

In the project, the subjects of image processing, object recognition and machine learning were studied intertwined. The main libraries, frameworks and plugins used; OpenCV, Tensorflow, Keras, Pyqt, NumPy, gTTS and Mediapipe.

PyCharm CE IDE (Integrated Development Environment) was used to develop the project. PyCharm is an integrated development environment that is highly advanced especially for the Python language

and offers many advantages to the developer. After the IDE was installed, the libraries and frameworks used in the project were added to the project.

Since hand movements must be recognized in the project, the hand must first be perceived by the program. In order for this process to take place, the camera is first activated and the camera is started with the program in Python with the imshow command. Then the hand image on the camera is customized for object detection.

Three different methods were used in this part.

In the first stage, hand tracking was performed using the Mediapipe library, that is, each joint point in the skeletal system of the hand was shown as connected to each other with lines and points. In this way, certain points of the hand can be controlled and interventions can be made where necessary [1].

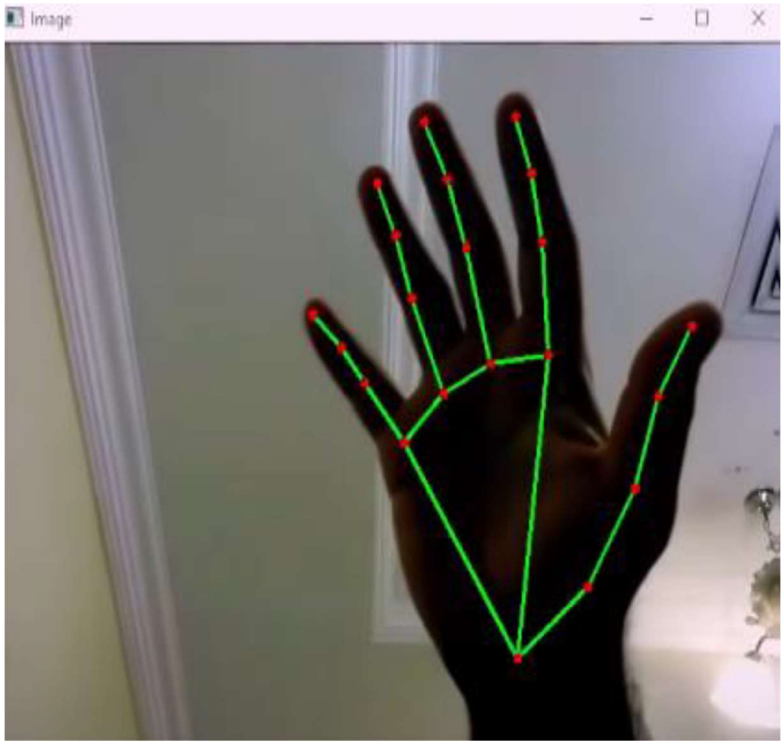

Figure 2: Hand Tracking Mediapipe

Object detection was performed by creating a certain area (frame) on the screen and masking the image by filtering in this area.

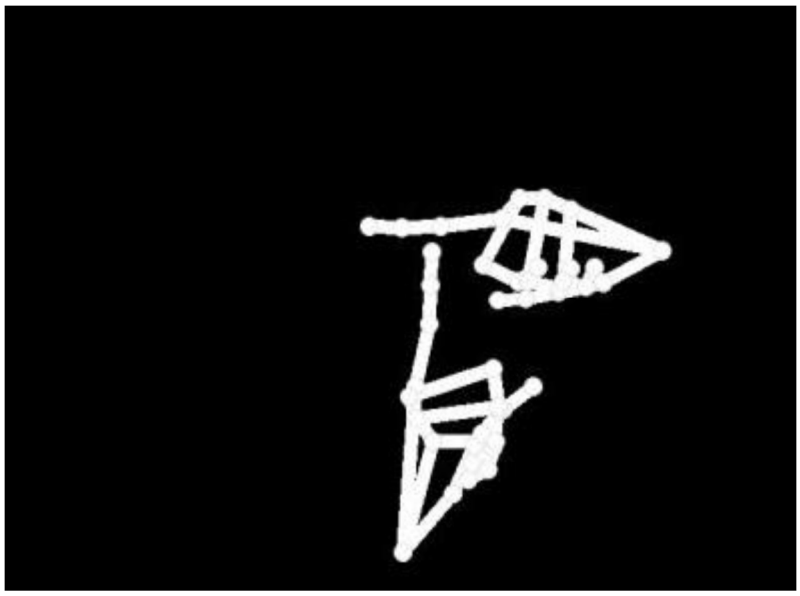

Figure 3: Filtering Method

The main reason for filtering and masking in the hand detection process is to highlight the object to be detected from other objects in the environment. If there is no image such as a green screen in the background, the main object to be detected should be highlighted from other objects. This is an important detail for both detection and tracking.

In the project, the frame creation method without filtering was used for the hand detection process. Adding an extra screen to detect the hand and evaluate it independently of environmental noise, it has been kept constant in the background. The added screen used only a certain part of the normal screen, so it was minimized to cover only the area that needs to be detected (an area large enough to show the hand).

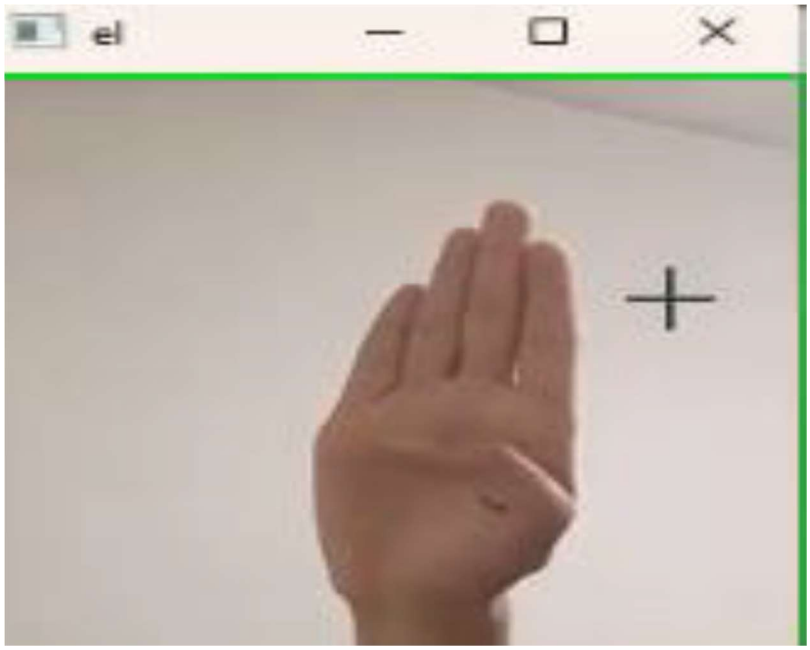

Figure 4: Unfiltered Frame Screen

Although using frames without a filter is effective in dataset creation and model training steps, it is a method whose performance may vary in different backgrounds. For this, it is important that the background is chosen by the user as a noise-free and flat background. However, in the researches, the hand detection process was carried out without filtering or hand tracking in a certain frame area. Therefore, this method was used in the study.

After performing the object detection and filtering processes in the project, a dataset was created. The dataset consists of sign language letters and words that should be briefly recognized. While creating the dataset, we created and collected the data. First of all, when the application is run, the letter that is wanted to be displayed on the frame screen is marked, and then the image on the frame screen is photographed. While creating the dataset, the most important point was to take the images from every angle and position.

4. The presented scenario

In the project, it is aimed to use 29 letters in the Turkish Sign Language alphabet and basic words in sign language. By adding all the letters in the alphabet and additionally basic hand gestures, it is aimed to put the letters side by side to form words.

The purpose of adding hand gestures that are separate from the alphabet is to switch to detecting the other letter in the word when the perception of a letter is completed. In this way, it is aimed to complete the word.

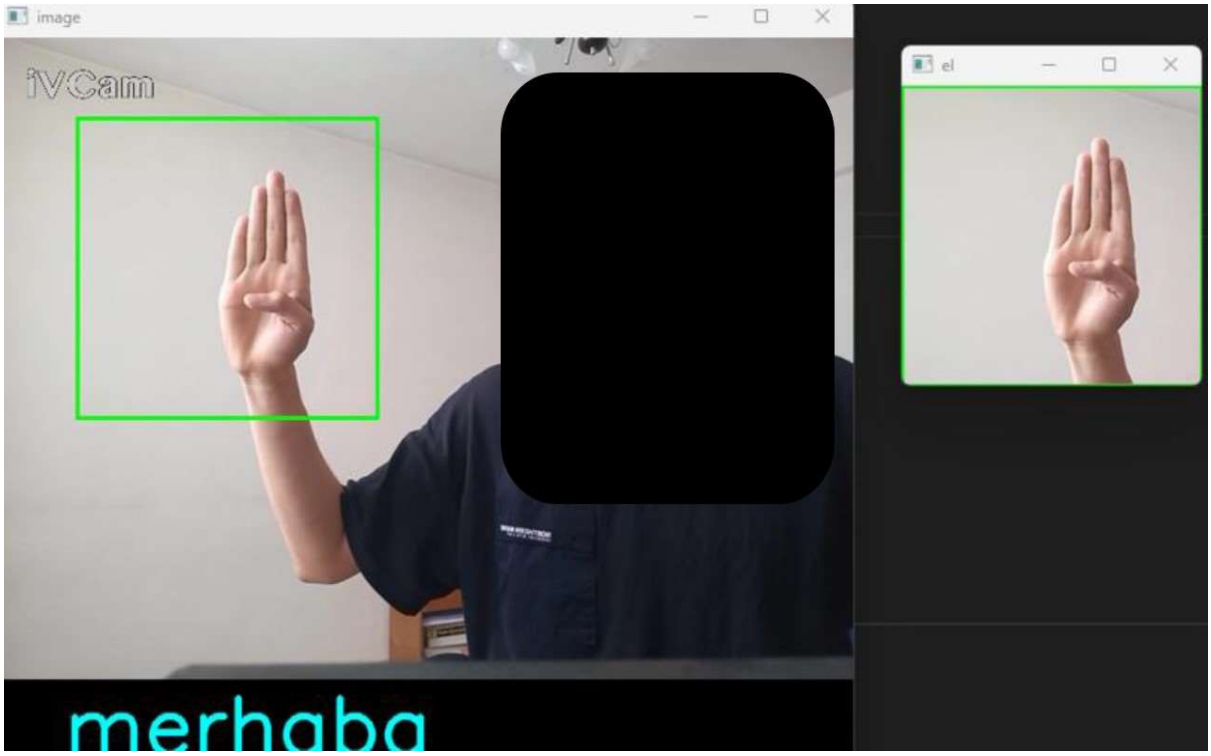

Figure 5: Representation of the Word “Hello” in Turkish Sign Language as “merhaba”

In the project, the Supervised Learning technique was used as a learning method. When the program is run, a small and a large (main) screen opens. The words are started to be shown by taking the hand into the green frame on the big screen.

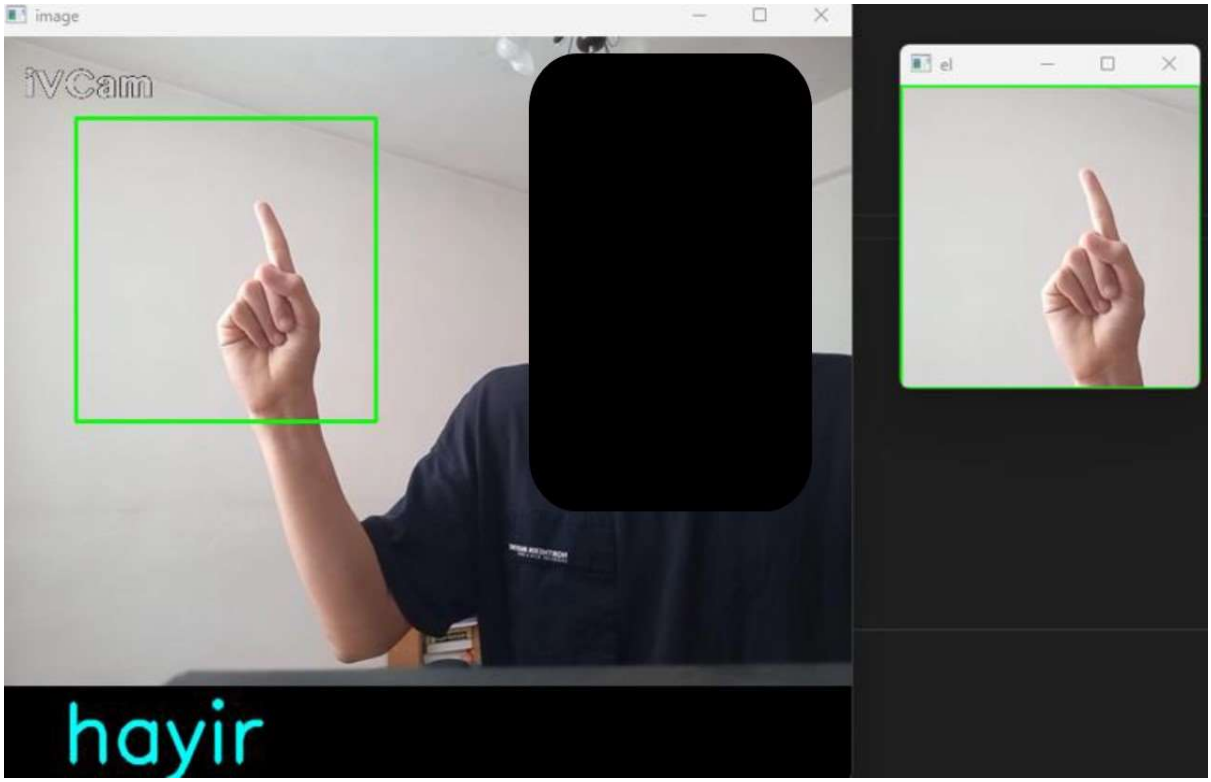

Figure 6: Representation of the Word “No” in Turkish Sign Language as “hayir”

If the program does not see any word in the frame field, then it just writes blank on the screen and waits for the word to be displayed. Afterwards, the words are displayed one by one and each displayed word is shown in writing on the screen and transmitted audibly as soon as it is shown. When the end sign is shown, the words displayed on the screen are reset and a new word is expected.

5. Conclusion

With today's advancing technology and advanced tools, image processing has become widespread and has been used frequently in every field. With the development and spread of image processing technology, it has been used in many sectors, making people's work much easier, accelerated, automated and considerably reduced the margin of error.

With the use of image processing techniques, even objects that are very difficult for the human eye to perceive and capture have become perceivable, processable and information can be obtained. One of the most important features of image processing is that it enables us to obtain information about the attributes of the desired object without the need to interact with the environment. Therefore, the present study has advanced in this direction.

In the study, a dataset was created to detect, transmit and recognize data very accurately. During the training, we considered and analyzed the successful studies in designing the model and determined our scenario accordingly. In the studies, the success of multi-layers in multi-class models was presented. Therefore, it was preferred to use multiple layers with more than one class in the study. The results showed that the study was successful in recognizing sign language letters by perceiving the hand.

6. References

[1] H. Yakut, Isaret Dili Harflerinin Goruntu Isleme Yontemleriyle Tanınması icin Bir Uygulama, Thesis, Fırat University, Engineering Science Institute, 2013

[2] B. Oktekin, N. Cavus, Isitme ve Konusma Engelli Bireyler için Isaret Tanıma Sistemi Gelistirme, Folklor/Edebiyat (2019) Vol:25, No:97-1, 2019/1

[3] A. Z. Oral, Turk Isaret Dili Cevirisi, Siyasal Kitabevi, 2016, 9786059221207

[4] Isaret dili öğreniyorum, 2022. URL: https://isaretdili.ego.gov.tr/isaret-dili-tarihcesi/

[5] S.W. Zhang, Z. S. Zhu, R. X. Zhu, Research on Dynamic Sign Language Recognition Based on Key FrameWeighted of DTW, W. Fu et al. (Eds.): ICMTEL 2021, LNICST 388, pp. 11–20, 2021.

[6] B. G. Lee, W. Y. Chung, Study of Sign Language Recognition Using Wearable Sensors, M. Singh et al. (Eds.): IHCI 2020, LNCS 12615, pp. 229–237, 2021.

[7] S. Sako, M. Hatano, T. Kitamura, Real-Time Japanese Sign Language Recognition Based on Three Phonological Elements of Sign, C. Stephanidis (Ed.): HCII 2016 Posters, Part II, CCIS 618, pp. 130–136, 2016.

[8] S. Awata, S. Sako, T. Kitamura, Japanese Sign Language Recognition Based on Three Elements of Sign Using Kinect v2 Sensor, C. Stephanidis (Ed.): HCII Posters 2017, Part I, CCIS 713, pp. 95–102, 2017.

[9] Alaaddin I. Sidig, Hamzah Luqman, Sabri A. Mahmoud, Arabic Sign Language Recognition Using Optical Flow-Based Features and HMM, F. Saeed et al. (eds.), Recent Trends in Information and Communication Technology, Lecture Notes on Data Engineering and Communications Technologies 5, 2018, DOI 10.1007/978-3-319-59427-9_32

[10] Z. Sun, A Survey on Dynamic Sign Language Recognition, S. K. Bhatia et al. (eds.), Advances in Computer, Communication and Computational Sciences, Advances in Intelligent Systems and Computing 1158, 2021, https://doi.org/10.1007/978-981-15-4409-5_89

[11] S. Rakesh, A. Bharadhwaj, E. S. Harsha, Sign Language Recognition Using Convolutional Neural Network, J. S. Raj et al. (eds.), Innovative Data Communication Technologies and Application, Lecture Notes on Data Engineering and Communications Technologies 59, 2021, https://doi.org/10.1007/978-981-15-9651-3_58

[12] S. K. Swarnkar, A. Ambhaikar, V. K. Swarnkar, U. Sinha, Optimized Convolution Neural Network (OCNN) for Voice-Based Sign Language Recognition: Optimization and Regularization, A. Joshi et al. (eds.), Information and Communication Technology for Competitive Strategies (ICTCS 2020), Lecture Notes in Networks and Systems 191, 2022, https://doi.org/10.1007/978-981-16-0739-4_60

[13] M. F. Karaca, S. Bayir, Turk Isaret Dili Incelemesi: Iletisim ve Dil Bilgisi, Ulusal Egitim Akademisi Dergisi (2018) Vol.2, No.2

[14] İsaret Dili, 2023. URL: https://tr.wikipedia.org/wiki/%C4%B0%C5%9Faret_dili

[15] R. K. Jain, S. K. Rathi, A Review Paper on Sign Language Recognition Using Machine Learning Techniques, R. Mathur et al. (eds.), Emerging Trends in Data Driven Computing and Communications, Studies in Autonomic, Data-driven and Industrial Computing, 2021

[16] A. Patil, A. Kulkarni, H. Yesane, M. Sadani, P. Satav, Literature Survey: Sign Language Recognition Using Gesture Recognition and Natural Language Processing, N. Sharma et al. (eds.), Data Management, Analytics and Innovation, Lecture Notes on Data Engineering and Communications Technologies 70, 2021, https://doi.org/10.1007/978-981-16-2934-1_13

[17] S. Das, S. Kr. Biswas, M. Chakraborty, B. Purkayastha, Intelligent Indian Sign Language Recognition Systems: A Critical Review, M. Tuba et al. (eds.), ICT Systems and Sustainability, Lecture Notes in Networks and Systems 321, https://doi.org/10.1007/978-981-16-5987-4_71

[18] V. Sannareddy, M. Barlapudi, V. K. R. Koppula, G. R. Vuduthuri, N. R. Seelam, Sign Language Recognition Using Convolution Neural Network, V. S. Reddy et al. (eds.), Soft Computing and Signal Processing, Advances in Intelligent Systems and Computing 1413, 2022, https://doi.org/10.1007/978-981-16-7088-6_59

[19] Türk İsaret dili, 2023. URL: https://tr.wikipedia.org/wiki/T%C3%BCrk_%C4%B0%C5%9Faret_Dili
[20] A. N. Erkan, C. Keskin, L. Akarun, Etkileşimli Ara Yuzler Icin Gercek Zamanlı El Izleme ve HMM Tabanlı Uc Boyutlu Hareket Tanima, in: Proceedings of the IEEE Conference on Signal Processing and Communications Applications, SIU 2003, Istanbul, Turkey, pp.192-195.