Hand Gesture Recognition Based on Image Processing for Supporting Deaf People

Phan Gia Luan, Tuong Phuoc Tho and Nguyen Truong Thinh

March 24, 2020

Hand gesture recognition based on Image Processing for Supporting Deaf People

[1,a]Phan Gia Luan, [1]Tuong Phuoc Tho, [1,c]Nguyen Truong Thinh
Department of Mechatronics, Ho Chi Minh City University of Technology and Education
Ho Chi Minh City, Vietnam
pgluan95@gmai.com, thotp@hcmute.edu.vn, thinhnt@hcmute.edu.vn

***Abstract*—One of the important problems our society is facing is that deaf and dumb people find difficulties in communicating with normal people who don't understand sign language. Even though sign language is the best way for deaf and dumb people to communicate with each other. Sign language is also used a little by normal people because it is an expressive and natural way for communicating. The motivation for our study is the improvement of accessibility to public information announcements for deaf and less hearing people. The whole idea is to build a service robot that enables communications between speech-hearing impaired individual and a normal person. In this paper, we will present two tasks, the first is how to translate sign language to speech, the second is focusing in the image processing algorithm to achieve hand gestures using depth sensor and then convert to speech. Several sign language visualization methods were evaluated, in order to perform this study a machine gesture translation system. It was concluded that system is suitable service-delivery platforms for sign language machine translation systems.**

Keywords— sign language, deaf, image processing, service robot,

I. NTRODUCTION

Today, the integration of deaf people has been improved in real life with supporting modern technology devices as well as medicines. The development and the unification of sign language has also created more advantages for deaf people in daily communication. However, the support of devices such as hearing aids is also limited, with a large number of hearing impaired people unable to receive signals to transmit through the 8th nerve. This shows that the sign or gesture language is still the main tool used in today's communication between people with hearing impairments together, as well as the communication between hearing impairment people and normal hearing people. It is this that interferes with daily communication because sign language is mostly used by people with normal hearing. Meanwhile, the image processing technology is supported by novel algorithms, so the analysis and processing of captured images is not too complicated and time-consuming as previously. From the analysis, the study of supporting hearing impaired people is described in this paper to get the purpose of researching new useful tool for communication between the hearing impaired and ordinary people. Hearing loss also known as hearing impairment, is a partial or total inability to hear. A deaf person has little to no hearing. Hearing loss may occur in one or both ears. In children, hearing problems can affect the ability to learn spoken language and in adults it can create difficulties with social interaction and at work [1]. Sign language or hand gesture language is a language that uses the expression of the hand, body posture and facial expressions in place of the sound of speech. Sign language was created to help deaf people communicate with each other in their community and acquire knowledge of society.

There are over one million deaf people in Viet Nam. Approximately 400,000 of these are in school age and use Vietnamese Sign Language as their first and only language. Contrary to common belief, sign languages are not visual-gestural representations of spoken languages. They are rich fully-fledged languages of their own. Additionally, different countries have sign languages of their own. Each of these sign languages also experience regional variations within the country. This is akin to dialect and idiolect variations in spoken languages. The objective of this study is contribution of active communication between deaf and hearing people and reduction of their communication barrier. As you known, deaf people need to communicate with hearing people on the very basic daily talks as a necessity of life.

Fig. 1. International alphabet of sign language

Besides, the deaf need to communicate with hearing people on situations include public services, buying bus and train tickets, seeking directions and purchasing the goods. The inability to access these services effectively because of communication constraints is a source of stress for the deaf. Like spoken language, sign language is very different from many countries or regions, that is because each country or region has different from history, culture, as well as customs, so the symbols to signify things and things are also different. In fact, sign language is life, because it originates from life. Whether or not we realize it, we have been using sign Language a lot in our daily lives. Science has proven that we transmit language 70% through non-verbal means, i.e gestures, gestures, facial expressions, etc. Thus, sign language is so pervasive in our lives that we may not be aware, but it still exists, develops and makes life more convenient and comfortable. In other words, it is the ordinary people who "invented" the sign language, the deaf people do one thing is to simulate and systematize it all into a language of their own. The communication between deaf and hearing people based on service robot required to have two distinct translation tasks: Translation from spoken language to gesture/sign language; Translation from gesture/sign language to spoken language. The deaf people cannot communicate with hearing people in directly, and require additional intermediary tool to be translation tool. Sign/Gesture-based tool carries out a gesture translation to deaf people message. The deaf person can stands opposite robot and camera attaching on robot's head. These captured images are then sent over the network to central processors of robot that recognize them. Such tool requires additional translation infrastructure in order for the deaf person to communicate with a hearing person. When using devices communicating between the different parties is indirect such as a text and speech based instant messenger by one party and a sign/gesture language application by the other party. The service robot's communication switches such as text to speech and vice versa, sign/gesture language to speech and vice versa. Gesture-based tool refers to specifically built to translate between gesture/sign and spoken language using translation tool.

In translating from gesture/sign to spoken language, gesture recognition algorithm is used to recognize gesture language from captured images. The output of this system is information of speech or message on screen. Conversely, in translation between hearing and deaf people, the input which is sentences of spoken language is translated into gesture information with animation avatar. The translation from spoken language to gesture/sign involves the following processes: capturing spoken speech input as audio; and converting it into text using speech recognition technology; translating the text into a gesture/sign transcription notation finally synthesizing like as clip on interactive screen. Capturing spoken speech is based on several microphones. Nowadays, many speech recognition technologies currently exist that can convert from spoken language audio to text, many of them open-source. On the other hand, translation from spoken language text to graphical user interface (GUI) that allows a hearing person to understand what the deaf person want to do from gesturing manual signs. The animation avatar is created using humanoid skeletons.

TABLE I. DIFFERENT TYPES OF DEAFNESS

Types of deafness	Description
Conductive hearing loss	Conductive hearing loss can occur when there is damage or a blockage in the outer and/or middle ear
Sensorineural hearing loss	Sensorineural hearing loss occurs when there is damage or malfunction of the hair cells in the cochlear
Mixed hearing loss	A mixed hearing loss occurs when both conductive and sensorineural hearing losses are present
Auditory neuropathy	Auditory Neuropathy occurs when there is a problem with the auditory nerve transmitting the signal from the cochlea to the brain.

The solution for barrier between deaf and hearing people was the use of highly skilled and trained interpreters proven inadequate and inefficient. And of course, the services are also very costly. We believe that most deaf people in Viet Nam cannot afford this service. Also, the use of interpreters is unsuitable in contexts in which confidentiality is vital such as when a deaf person seeks medical or psychological treatment. In such cases, the deaf person may not be keen to have a human interpreter present. The research carries out automated translation robot between sign/gesture language and Vietnamese/English and vice versa. The given gesture/sign as input, the system aims to produce corresponding English audio as well as display on screen of robot, and given Vietnamese/English speech as input, the robot produce and render the corresponding sign language clip. The system of robot produced gesture recognition systems that can track hand-motions, hand-shape recognitions as well as facial recognition. Besides, the robot's system produced an animation avatar displaying on robot's screen based on synthetic sign language phrases. The paper describes the use of mobile robot as service delivery devices in a gesture/sign language machine translation system.

One of the senses helps people communicate with others, help absorb human knowledge as well as share things in life that are hearing. If a person lacks hearing, or hearing is not good, it is a huge disadvantage for that person. Meanwhile, the part of people with hearing disabilities makes up a significant part of the human community [2]. The problem of the condition of people with hearing disabilities is more complicated because hearing loss has many different levels and manifestations (Table 1). Currently, according to the statistics of the World Health Organization (WHO), around the world there are about 278 million deaf and mute people, accounting for 4.5% of the world population. Most of them have average and below average living standards due to language barriers [1]. In Vietnam, according to statistics of the Ministry of Labor, Invalids and Social Affairs, Vietnam has about 2 million deaf and dumb people, accounting for about 13.5% of the population. In which there are about 400,000 people in school age [1]. In recent years, the integration of hearing impaired people has been improved thanks to medical assistance, technology equipment. The development and gradual integration of sign language has also created more advantages for deaf people in daily communication. However, those supports cannot solve the problems of hearing impaired people, but only partially solve them. Because the disability status has many very different levels, the methods used must also be different and the brand is specific to certain cases. For example, people with ear injuries can be treated with medical methods to help restore hearing. Or people with poor hearing can be supported by hearing aids. For those who have very poor or lost hearing that we often call deaf, the tool that helps them communicate is sign language. This has caused many obstacles in daily communication between deaf and ordinary people. Because, the sign language of most ordinary people who do not know or can use is very limited. Recently, there have been a number of studies to support people with this hearing loss. Study to transmit sound through skulls to patients with hearing loss. Researchers from Chalmers University of Technology in Sweden have successfully developed an implant that transmits sound into the inner ear by sending vibrations directly through the skull bone instead of the middle ear [3]. A research team from Ukraine studied a glove that converted speech-to-speech languages to complete Enable Talk [4]. Besides, the study recovers hearing by regenerating the auditory hair cells of specialists from the eye hospital and in Massachusetts and Harvard Medical School [5].

In order to make some moves from sign language to spoken language, and vice versa, robots that support deaf people have two main tasks: one is to recognize the sign language from the deaf and then switch to spoken language; the second is to recognize the voice of a normal person and then switch to sign language by output on the screen. In the first mission, the robot performs a series of following actions: identifying the positions and gestures of both hands and arms, capturing their movements and comparing them to the database, then export as audio. Thus, in this first task, the robot uses a camera to capture images in real time, using image processing algorithms to process retrieved signals, then compare them with the database for export audio signal. In the second mission, the robot will recognize the voice captured by a microphone. This audio data is processed by the cloud algorithm, then compared to the database and exported to the corresponding sign language screen. Figure 1 shows an overview of the activity of robots that support hearing impaired people. From the analysis and judgment above, along with the tendency to apply robots to support activities in people's daily lives such as entertainment robots, service robots, specialized robots,… are increasingly popular. A robot which can assist deaf people using image processing will be presented in this paper. For the purpose of studying another tool to support, create a visual and close-knit communication between the hearing impaired and ordinary people.

II. Hand Gesture Recognition Based on Image Processing

Figure 2 presents an algorithm to identify and capture hands. First, information about the depth of the object is captured by the data obtained from the depth camera, and creates moving images from images that accumulate differences between depth images [6]. Then reduce the noise by locating the spatial and the morphological operation. Motion clustering method is used to search for moving clusters. After that, the hand will be identified from moving clusters with wave motion (wave motion). Finally, after having identified the hand, the Kalman filter will be applied to catch the hand. Images taken from depth sensors often suffer from interference due to reflections. Sometimes, these noise is identified as moving information. Therefore, noise reduction is really necessary before identifying. Preprocessing part includes algorithm for grouping for hand identification in the original position.

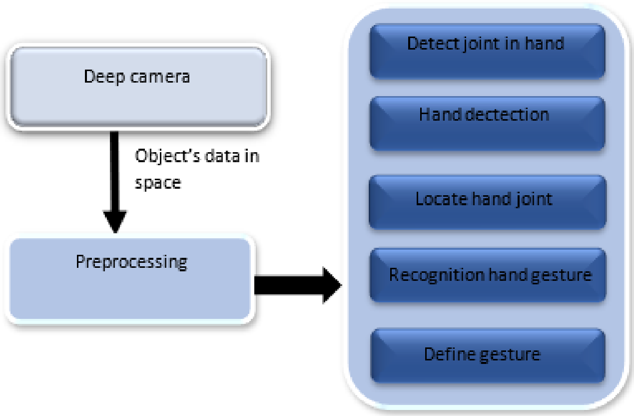


Fig. 2. Diagram hand gesture recognition

The identification process begins with the initialization of the variable to receive the data series from the depth sensor [7]. These settings are intended to identify the source of information that needs to be obtained, which is the coordinates of the joints in three-dimensional space. These data will be used to search for finding the joints of the hands. This is a very important step in the identification process. Because, the identification and tracking of the position and gesture of the hand are accurate or not, depending on the search for the basic hand, thereby calculating to find the joints of the fingers. After detecting the position of the hand joint, the next is to identify the hand area [7]. First, the search area is limited to an optimal size during hand identification. A reference gender of the hand will be defined, this reference limit is the distance from the hand joint to the fingertip and hand joint in the palm position. Meanwhile, the position of the fingertip is located on the middle finger, the longest finger. After finding the reference size, the next step is to identify the hand in space. After finding the tips of the fingers, the next step from these points finds the joints of the fingers [8]. To find the joints of the fingers, the algorithm applies the golden ratio to the calculation. Consider the straight lines consisting of two parts as shown in **Figure 3**, where the long paragraph is a, the short line is b, the sum of the two straight lines is a + b. When these numbers satisfy the condition (a + b) / a = a / b, then a / b ratio is called the golden ratio. By solving the above equation, find the ratio that is 1.61803398875; (almost equal to 1.62) people sign it as φ.

$$a = \frac{a+b}{a} = \frac{a}{b} = 1.62 \qquad (1)$$

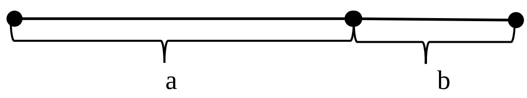


Fig. 3. Straight paragraph in the golden ratio.

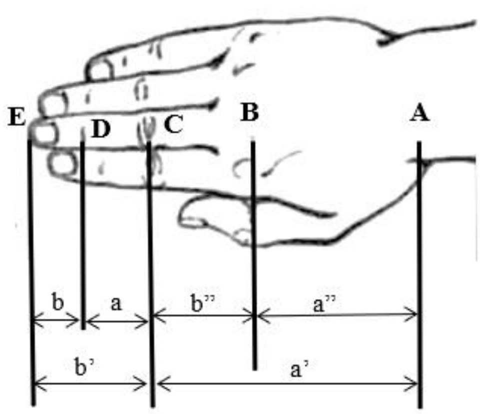


Fig. 4. The golden ratio on human hands.

Applying the golden ratio **(Figure 4)** in the calculation to interpolate the knuckles. Calling A, B, C, D and E are the points lying on the wrist joints, matching the middle finger, middle joints and middle finger. According to the golden ratio we have the equation:

$$\frac{a}{b} = \frac{a'}{b'} = \frac{a''}{b''} = \varphi \quad (2)$$

Meanwhile, the length from the wrist joint to the fingertip is the AE segment. So the equation for calculating joints will be:

Point C: $$a' = \frac{\varphi.AE}{(1+\varphi)} \quad (3)$$

Point B: $$a'' = \frac{\varphi.a'}{(1+\varphi)} \quad (4)$$

Point D: $$a' + a = \frac{\varphi.b'}{(1+\varphi)} + a = \frac{a'}{(1+\varphi)} + a \quad (5)$$

Same for the joints of the remaining fingers. The results obtained calculate the position of the joints of the fingers (Figure 5).

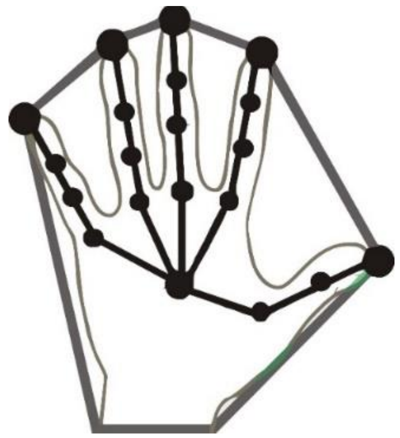


Fig. 5. The knuckles after identification

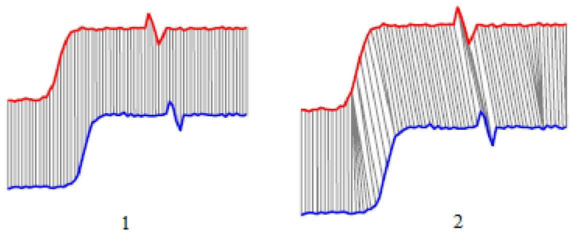


Fig. 6. Comparison of accuracy between Euclid method and DTW method. (1) Calculated by Euclid method; (2) Calculated by Dynamic Time Warping method

The next step is to define the gestures and compare them with the baseline data to identify what the gesture is. To accomplish this, the topic applied DTW dynamic time torsion algorithm to calculate. First of all, the base data used for comparison was created before to the base library [6]. When a gesture is identified and when compared to the base data will encounter a difficulty that is the difference in time. If you call the base data series T, then T will be the set of real numbers Q = $\{q_1, q_2, ..., q_n\}$, and call the gesture data sequence collected from the camera C then C = $\{c_1, c_2, ..., c_n\}$. To be able to compare these two strings, the algorithm applies Euclid truth:

$$D(Q,C) = \sqrt{\sum_{i=1}^{n}(q_i - c_i)^2} \quad (6)$$

The degree of distance Euclid has the advantage of being easy to calculate, easy to expand for calculation. However, it has a disadvantage that only gives accurate results when the two chains have similarities in time. Thus, with the comparison of the gesture obtained in real time with the database, the accuracy will be low if applied by Euclid. To overcome this, DTW algorithm is applied. Using the DTW algorithm will overcome the limitations when applying Euclid measurements **(Figure 6).** Using Dynamic Time Warping method (DTW) to compare the similarity between two strings Q and C in real time, DTW algorithm will build a square matrix T of size m in it:

$$T(i,j) = dist(i,j) + min[T(i,j-1), T(i-1,j), T(i-1,j-1)] \quad (7)$$

With dist (i, j) is calculated according to Euclid formula. Suppose that the two strings corresponding to the base data sample are Q = $\{q_1, q_2, ..., q_i\}$, and the data string retrieved from the camera is C = $\{c_1, c_2, ..., c_j\}$. The algorithm will perform finding the optimal path of series C according to Q series (i.e different positions between the two chains over time) so that the difference between the two chains is the smallest.

The path is determined by consecutive pairs of nodes $(i_{k-1}, j_{k-1}) \rightarrow (i_k, j_k)$. Use the symbol i_k to represent the index of the string Q at time k and j_k is the index of string C at time k. So the total distance between the two strings is:

$$D(i_k, j_k) = D(i_{k-1}, j_{k-1}) + d(i_k, j_k) \quad (8)$$

Finding the value min D (i, j) follows the following formula:

$$D^*(i_k, j_k) = \min\left[D(i_{k-1}, j_{k-1})\right] + d(i_k, j_k) \quad (9)$$

Thus, the gesture sequence received from the camera will be compared with the basic data series using DTW algorithm to standardize time. After comparing with baseline data, which pair has the smallest difference, it is the same pair.

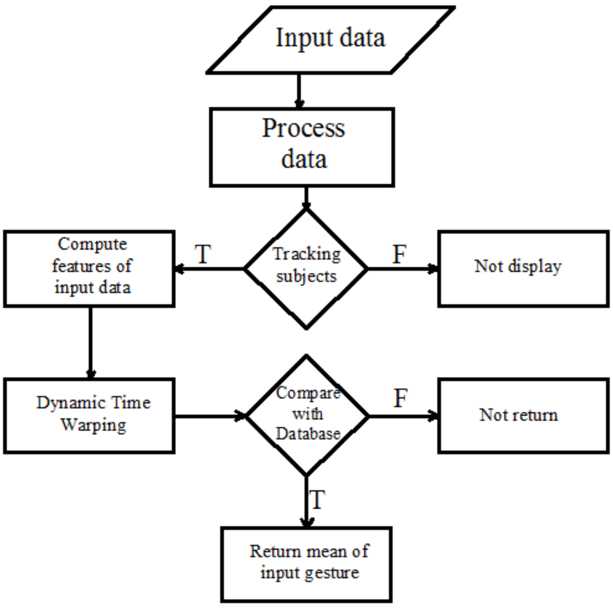


Fig. 7. Block diagram for translating gesture language.

Finally, translating sign language into language is the export of homologous strings in the form of sounds. In order to translate sign/gesture language, deaf people stand opposite robot, the following block diagram summarizes our approach is shown in **figure 7**.

III. Hand tracking based on depth information

Two factors to understand sign language are grasping the position of the arms and hands. This experiment presents an algorithm to identify the posture of arms and hands. To do this, the camera used is a kind of depth camera, combining color cameras. In addition to the information obtained by the x and y axes, the depth camera provides additional information of the z-axis. From these data we will build an algorithm to recognize gestures from two hand and hand scenes. The physical world around us is three-dimensional. Consequently, traditional cameras and image sensors only capture 2-dimensional images and omit depth information. This is a huge limit to being able to understand objects in the real world. In recent years, with the research, development and application, 3-dimensional image sensors have been fabricated and applied widely. One of the most popular 3-dimensional image sensors is Microsoft's Kinect. Methods of capturing objects based on depth information can be classified into two types: model or motion-based. The method of tracking the object based on the model using 3D skeleton model that fits the hand. The method of capturing objects based on motion uses the movement of hands in deep space. The model-based retrieval method uses 3D points in the x-y-z axis to estimate hand position and orientation. These 3D points are attached to the joints of the hand model, which can be used with some degrees of freedom in the joints of the fingers, or the model with all degrees of freedom. Dynamic model of the hand is shown in **Figure 5**. The human hand consists of 27 bones, including 8 bones in the wrist and 19 bones for the palms and fingers. These bones are connected by joints with one or more degrees of freedom.

Fig. 8. Tracking of hand based on Kalman filter.

Figure 5 shows the couplings and the corresponding degrees of freedom making up a total of 26 degrees of freedom. In particular, the wrist has 6 degrees of freedom with 3 degrees of freedom for translational motion in space and 3 degrees of freedom for movement around the axes. The parameters of the hand model are optimized to minimize the difference between the theoretical model and the actual observation of the hands. In the motion-based object capture method, the primitive motions are searched from accumulated images based on 3D data. It detects motion by using the depth value subtracted between the depth image before and after the depth image, accumulating those values. Images obtained from normal color cameras are heavily influenced by lighting conditions. Meanwhile, the image obtained from the depth camera is not affected by lighting conditions, however, the noise, especially in the subject's border areas, it is difficult to find differences in the depth images of subjects because the depth image depends only on the distance between the camera and the subject. Object tracking has a number of methods, but the Kalman filter has advantages for hand-tracking. Firstly, the Kalman filter only needs a small data store for the previous data to process, because we only need information about the previous state without all the previous frames. Second, the advantage of a Kalman filter is suitable for dealing with variable efficiency over time. Therefore, we apply the Kalman filter to capture the hand. Kalman filters are used to capture objects with many applications. Typically, they are used for the state of both x-y dimensions for images. However, we need to add in depth information. Thus, the state is designed with 3 dimensions x-y-z. We set depth information as the value of the z-axis. Therefore, we set the state to a 3-dimensional size:

IV. Experiments and discussions

To be able to assess the effectiveness of the proposed method in identifying sign language according to the sign language system in Ho Chi Minh City and to translate into spoken language, the topic was experimentally received. Small form of dialogue with the following statements: Hello. How are you. I'm fine. Where do you go? I go to school **(Figure 9).** The experiment was conducted to identify each word with the number of times it was performed 50 times in normal light conditions. **Table 2** shows the identifier and the exact proportion of the identification process with 50 times the number of attempts for each word. Based on the chart, it was found that the program with the recognition algorithm gave relatively stable results. The difference between the highest accuracy rate and the lowest accuracy rate of 15% (**Figure 10**).We repeated several times with difference about speed of gesture. The gesture was tracked and translated successfully with accuracy about than 80%. Besides that, we sent five clips sign language to center of deaf people to ensure that they can understand our database. The result is almost all deaf people can identify what we want to convey.

In image process, the distinctive images with human movements, objects, as well as noise are shown in these images then they are eliminated noise, moved groups, and identified hands. Besides, spatial filters and geometric processing to eliminate noise. When noise cancellation is applied, the movements actually become clearer. Median filter with open size 5x5 is used for space filter. The median filter changes the value of the pixels with the average of the small parts of the image. The median filter provides effective noise reduction without image blur. The noise parts of moving images are often mixed black and white particles, so the median filter is very effective when applied to noise reduction.

TABLE II. ACCURATE RECOGNITION RATE OF WORDS

Words	You	I	Hello	Go to	Where

Accuracy	80%	75%	87%	90%	83%

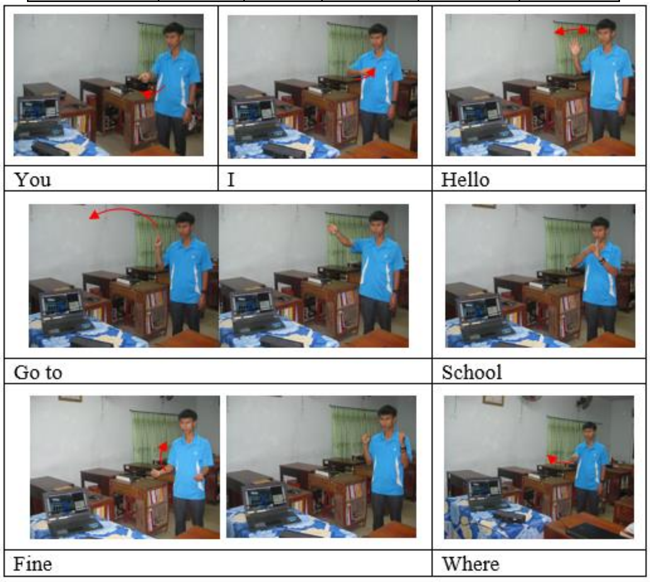


Fig. 9. Used words in experiment.

Fig. 10. The accuray of word recognition.

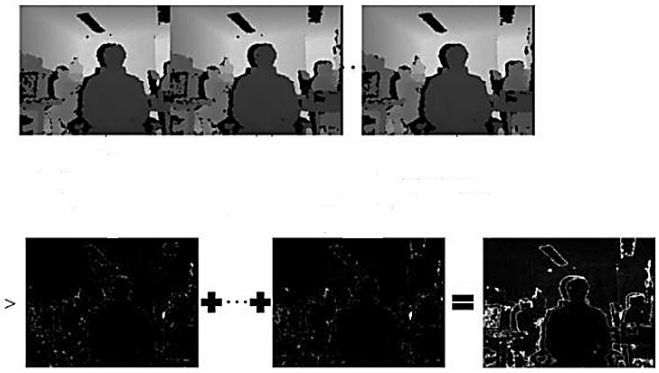


Fig. 11. Processing procedure for generating moving images.

We also use a geometric processor to eliminate noise. We reduce the outer shadow of the object by reducing and extending the outer area of the object. As a general rule, this method is to smooth out the outside, separate small sized areas and remove small parts. Therefore, this method eliminates random noise and minimizes the original image. Activities that reduce the object ignoring particles, eliminate unrelated pixels and small points of the image. The expansion is the opposite of abrasive activity. It attaches layers to the object and returns the original size of the object. These activities are very effective at eliminating noise in depth images. The elements of the motion picture are connected and gathered to re-generate digital data of hand. A polynomial regression method is used to determine dimensions of hands. With the hand's size data, the polynomial regression method is used to attach a curve to the data. We use the fifth polynomial model to describe.

In the preprocessing section, motion images are created by accumulated image data, noise reduction in motion images and find moving clusters. To detect the hand, the condition of the moving hand is set, which includes repeated movements from side to side. First, the direction of movement of the cluster is identified by using motion patterns. These motion patterns are an effective method for tracking movements in general, and are especially for detecting the position of the hand.

Fig. 12. Original image (a) and moving image after filting noise.

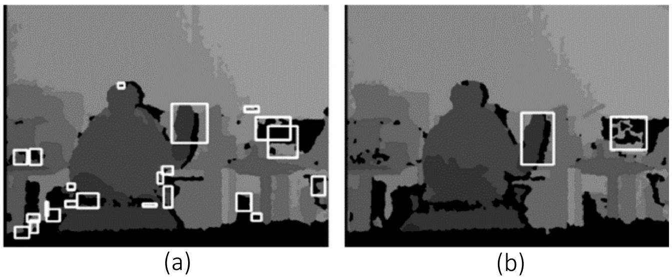


Fig. 13. Movement cluster of hand's size: (a) before applied threshold, (b) After applied threshold.

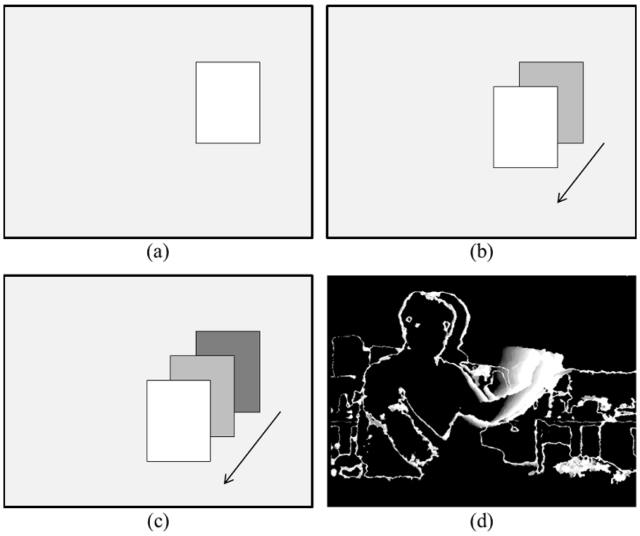


Fig. 14. Movement history and suggested pattern: (a) at time t; (b) at time t+1.

In **Figure 13**, we assume that we have a distinguished cluster that is a white rectangle. This image is referred to as the history motion picture. The white area of the image shows that all the pixels in this area are set to a floating point. As the rectangle moves, a new cluster is computed from the current moving image and superimposed on the history motion image. In **Figures 14b** and **c**, the white rectangle shows a new cluster and the old ones become darker. The darkest rectangle represents the most previous motion. The rectangle becomes brighter for subsequent clusters. All of this rectangular series represents the cluster's movement. **Figure 14d** shows the history motion picture in depth space. From history motion image, the direction of movement is determined by taking the gradient. The slope is also calculated by using the functions Sobel gradient and Scharr gradient. Some gradients calculated from the history motion picture are incorrect. The problem is that when the regions are not moving, the gradient value is 0 and the outer boundary of the clusters has a large gradient. Once we know the time between frames, we could calculate the sequence of gradients and remove the incorrect gradients. Finally, we can confirm the global gradient with the direction of movement. Figure 15 shows the direction of the clusters. The line in the circle indicates the direction the cluster is moving.

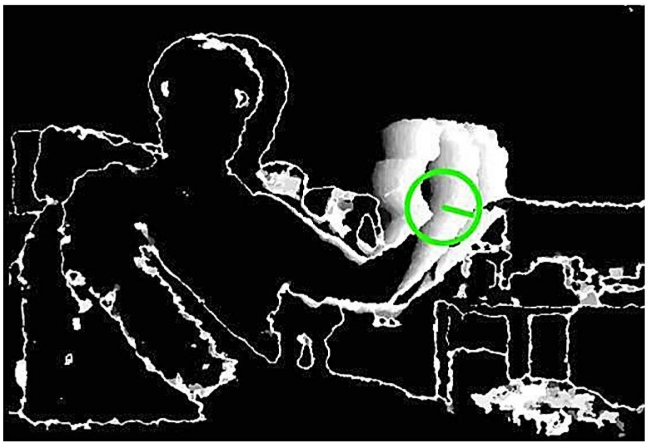


Fig. 15. Direction of movement of the cluster.

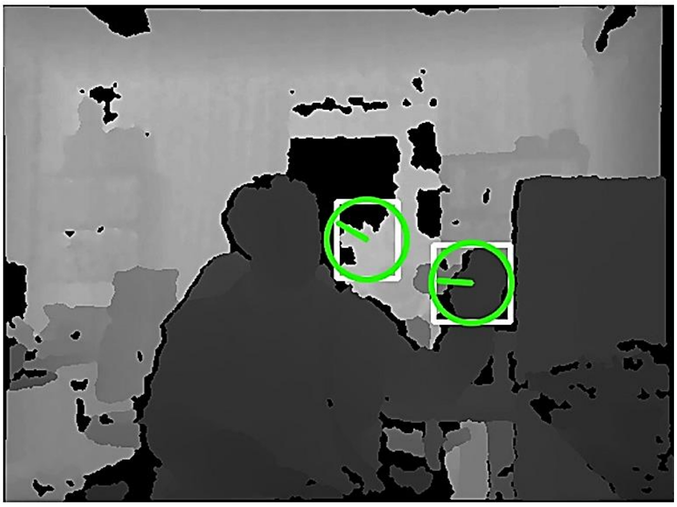
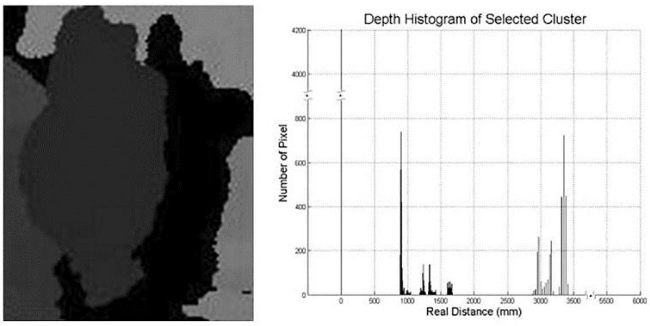


Fig. 16. Find the direction of motion

Next, the hand cluster is found by using the direction of hand movement. From these moving clusters, their directions of movement can be calculated. **Figure 16** shows the direction of movement of the clusters in the moving image. The method is used to determine the direction of reciprocal motion is to calculate the change in direction of the clusters. We set the condition that the number of times to and fro is 3 and calculate the number of times that the cluster moves from left to right. We also assume that the hand is closest to the camera. With this assumption, the hand is part of the entire hand cluster identified with the smallest depth value, dimensional histogram is used to find the hand in the selected cluster. **Figure 17** shows the back-light diagram of the cluster.



(a) (b)

Fig. 17. a) Selected moving cluster; (b) Depth light diagram of cluster.

Fig. 18. The result of hand recognition

In the depth light chart, we remove pixels below the 600mm level because the depth sensor cannot be captured at distances below 600mm. So we initialize the hand at the first vertex of the depth light chart at nearly 1000mm. **Figure 19** shows the result of the initial hand recognition. In this method of identification, it is stable even under strong lighting conditions. But noise at the object boundary and the effect of noise are sometimes referred to as moving clusters, and can satisfy both size and reciprocal conditions, this affects hand recognition. The tracking method will be used to eliminate these false identities.

An algorithm in this study is used like as independent device or embedded into service robot for supporting deaf people. Several experiments are conducted to evaluate the reality of the study. With dynamic moving ability and friendly display, robot can easily communicate with deaf people. We required to translate to Vietnamese Sign Language (VSL) with sentence: “Chào, tôi rất vui khi gặp bạn”. In grammar of VSL have not adjectives as “rất”, “thì”,”khi”…before adverbs and adjectives to add emphasis or exclamatory. Therefore, the translated sentence become: “Chào, tôi vui gặp bạn”. Gesture in real life is recorded from the 1st to the 32nd frames of 640x480 frames. The Dynamic Time Warping algorithm doesn’t care about how quickly the gesture is performed.

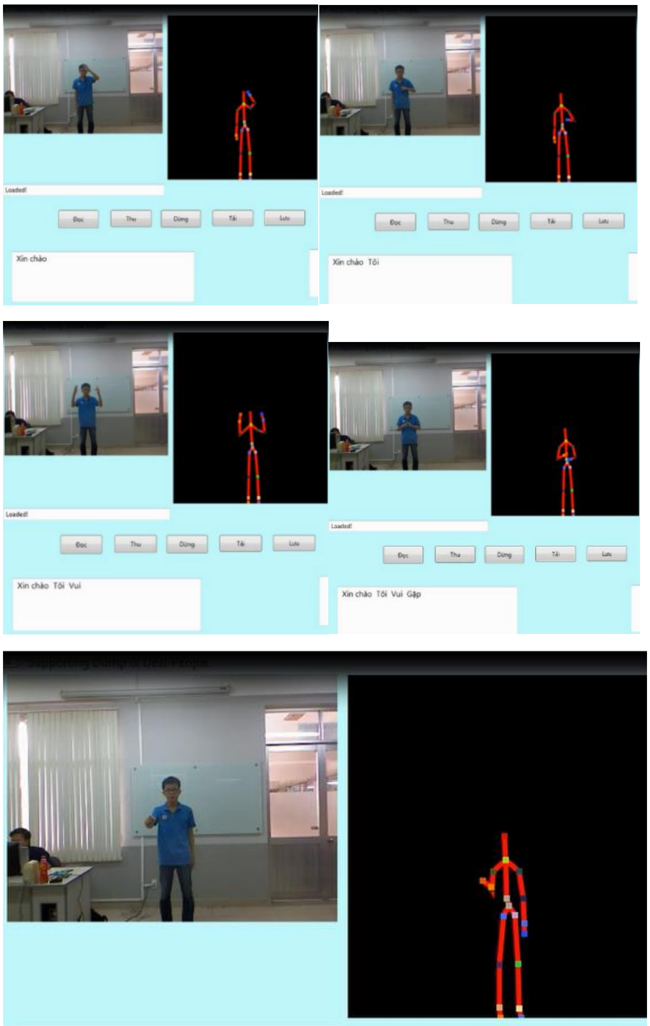


Fig. 19. The result of skeleton recognition

We repeated several times with difference about speed of gesture. The gesture was tracked and translated successfully with accuracy about than 80%. Besides that, we sent five clips sign language to center of deaf people to ensure that they can understand our database. The result is almost all deaf people can identify what we want to convey.

V. CONCLUSION

This paper presented a summary of a processing algorithm supporting for deaf and less hearing people. With dynamic moving ability and friendly display, robot can easily communicate with deaf people. The main result of this paper is the successful construction of the position recognition and gesture recognition hand. The main tasks are done in the paper: studying the structural and grammatical characteristics of the sign language used as a basis for identification; build identification algorithms, capture positions and gestures of hands; build algorithms to find the joints of fingers as data for identifying sign language; Sign language recognition is expressed in two hands, accessing the base data to translate into spoken language. The first step was to experiment with some words in sign language and translate into spoken language. Besides, the identification still exists many problems need to be solved next: overlapping gestures; the basic data library has not been built yet. In the coming time, the next research of the topic will be implemented: optimize the algorithm of identifying sign language and build larger database library.

ACKNOWLEDGMENT

We would like to say thank you Ho Chi Minh City University of Technology and Education for financially supporting us.

REFERENCES

[1] http://www.dphanoi.org.vn, *Cần nhân rộng mô hình dạy ngôn ngữ ký hiệu tại Trung tâm Giáo dục thường xuyên*.

[2] W.C. Stokoe. Sign language structure: An outline of the visual communication systems of the American deaf. Deaf Studies and Deaf Education, 10(1):3{37, 2005.

[3] A. Hameed. Information and communication technologies as a new learning tool for the deaf. Relation, 10(1.129):7856, 2009.

[4] Jinhua Xu, Sign language translation using Kinect and Dynamic time warping.

[5] Dương Tuấn Anh, Tổng quan về tìm kiếm tương tự trên dữ liệu chuỗi thời gian, trường đại học Bách khoa TpHCM, 2011.

[6] Le Thi Hang, General knowledge of hearing loss, Da Nang Pedagogical University, 2008, pages 3-6, 43 – 48.

[7] Higher Education Project for Vietnamese deaf people, Ho Chi Minh city sign language, Publisher of Culture Information, 2008, pages 5-8.

[8] Chalmers University of Technology, New bone conduction hearing aid passed long term endurance test, 2017

[9] Frederic Lardinois, Ukrainian Students Develop Gloves That Translate Sign Language Into Speech, 2012

[10] Mary Leach, Mass. Eye and Ear Researchers Regenerate Sensory Hair Cells, Restore Hearing to Noise-Damaged Ears, Massachusetts, 2013

[11] A. Lekova, D. Ryan. Fingers And Gesture Recognition With Kinect V2 Sensor. International Conference AUTOMATICS AND INFORMATICS'2016. 4-5 October 2016, Sofia, Bulgaria. P. 107-109.

[12] Yi Li. Hand Gesture Recognition Using Kinect. Electronic Teses and Dissertations. P. 823

[13] Andrew Weinfield. Gesture Recognition Using Microsoft's Kinect v2.