MINISTRY OF EDUCATION AND TRAINING
THE UNIVERSITY OF DANANG

VO DUC HOANG

VIETNAMESE SIGN LANGUAGE RECOGNITION

Major : COMPUTER SCIENCE
Code : 62 48 01 01

THESIS SUMMARY

Da Nang - 2018

The doctoral dissertation has been finished at:
THE UNIVERSITY OF DANANG

Advisors:

1. Prof. PhD. Jean Meunier

2. PhD. Huynh Huu Hung

Reviewer 1: Assoc. Prof. PhD. Do Nang Toan

Reviewer 2: Assoc. Prof. PhD. Tran Thi Thanh Hai

Reviewer 3: Assoc. Prof. PhD. Huynh Xuan Hiep

The dissertation is defended before The Assessment Committee at The University of Danang

Time: 08h30

Date: 07/12/2018

The dissertation is available at:

- National Library of Vietnam

- Learning & Information Resources Center, The University of DaNang

INTRODUCTION

1. Motivation

Vietnam is one of the countries with a relatively high numbers of people with disabilities in the Asia-Pacific region. Deaf people make up about 15% of people with disabilities. Deaf people use sign language as their hand gesture language with manual visual cues and facial expressions to convey meaning from words instead of using sound. The language used in the deaf community, however, is not popular in the community, so there is a huge barrier between deaf and normal people.

In recent years, many studies have developed an identification system with many different sign languages that is a big challenge for many areas of research, such as a method of using hand gestures, machine learning classification, human-machine interface, natural language processing ...

The requirement of the thesis is to develop identification methods converting sign language into text in order to create a convenient communication between deaf people and normal people. The study of improving methods of hand gesture recognition has important implications, helping deaf people better integrate into the community.

This motivates us to choose the problem **"Vietnamese sign language recognition"** for our doctoral dissertation.

2. Objectives of the study

The dissertation aims to address the Vietnamese sign language (VSL) and to overcome the technical difficulties of data acquisition, pretreatment and extraction of characteristics, helping deaf people

integrate into the community. Specifically, the thesis aims at the following objectives:

- Analyse basic characteristics of Vietnamese sign language.

- Develop pre-processing and extraction methods to improve recognition rate compared to previous studies.

- Apply machine learning models for testing, especially SVM-Support Vector Machine (SVM), for training and recognizing the gestures of the Vietnamese sign language.

- Construct sample data set of the Vietnamese sign language and study the video segmentation method to improve recognition rate and apply the system in real time.

3. Object and scope of the study

Research objects of the thesis:

- Algorithms and solutions for analysis and identification of sign language.
- The alphabet of static gestures in Vietnamese sign language.
- Words and phrases of the continuous gesture of Vietnamese sign language.

To determine the research objectives and objects as mentioned above, the research scope of the dissertation is as follows:

- Research on image processing techniques to support the system of general sign language identification, analysis, and evaluation towards the identification of sign language in Vietnamese.
- Research on the static sign language identification system, the sign language of the Vietnamese language, consisting of two main tasks: (1) to develop a data acquisition method, combining the basic characterization, (2) to search, select and improve the

recognition method to match the Vietnamese sign language recognition system.

- Research on building a system of continuous sign recognition including words, aiming at translating complete sentences of Vietnamese sign language.

4. Method of the study

The method used in the thesis is combining theory and experiment to test the effect:

- Analyze specific characteristics of the Vietnamese sign language, build a sample database for testing.
- Review relevant studies, compare and evaluate the strengths and weaknesses of different identification methods, and then propose ideas for Vietnamese sign language recognition. The evaluation is based on processing time criterion and successful recognition rate.
- Use appropriate mathematical tools to model the set of gestures for identification purposes.
- Design and implement experiments with a common database available to evaluate the effectiveness.

5. Dissertation outline

On the basis of research tasks mentioned above, in order to achieve its objectives and ensure the validity of the research problem, apart from the introduction, conclusion and development, the thesis structure consists of three chapters. The main contents of the chapters are as follows:

Chapter 1 introduces the overview of the current sign language in Vietnam and in the world. Related research on sign language

recognition under two classifications based on the process of gathering data and extract specific machine learning methods is then presented.

Chapter 2 presents two studies on static gesture recognition of sign language. The first one was proposed based on the basic image processing process. Data captured by the camera is a hand image, the pre-processing process uses a color filter to eliminate interference. Research applied geometric methods to determine the tops of the fingers, removing the arm. After extraction is characterized by vectors, the study uses a multi-layer vector-assisted learning model (SVMs) for training and identification. The other used in-depth cameras for data acquisition, character extraction based on the rank-order correlation matrix (ROCM). The research was tested on datasets of Vietnamese sign language symbols with single, dual and symbols associated with the mark.

Chapter 3 presents research on continuous gesture recognition of the Vietnamese sign language. The first study was tested based on real-time data acquisition of joints. Machine learning and identification are based on the Dynamic Time Wrapping (DTW). The second study used the depth camera to capture data and applied models to process 3D spatial data in real time. Support vector machine (SVM) learning models are used for classification and identification.

6. Contributions

Basic research on static and continuous sign language identification based on data collected from color cameras and features extracted using geometric model is presented. Experiments

are done using SVM machine learning methods; the performance is evaluated based on successful recognition rate.

Methods for obtaining data from in-depth cameras are suggested: (1) Characteristic extraction is based on matrix rating algorithm for an alphabet of Vietnamese sign language; (2) Dividing block by 3D model method is used for recognition of words, phrases and sentences of Vietnamese sign language.

Research on video segmentation method and application for recognition and combination of characters of Vietnamese sign language alphabet are presented.

For continuous gesture of Vietnamese sign language, study and experiment are implemented using two different data acquisition methods, joint coordinates and in-depth cameras, for analysis and evaluation.

CHAPTER 1: INTRODUCTION

The content of chapter 1 consists of two main sections: the first part is an overview of sign language in the world and in Vietnam (VSL - Vietnamese Sign Language); the second part is a synthesis of relevant studies on sign language identification, sign language to date.

1.1 Overview of sign language

Sign language is widely used by the deaf community. Sign language includes both common gestures and thousands of symbols that deaf people have developed over time.

Because each country, regional history, culture have different habits, signs to denote the phenomena are different. There is a

difference of system vocabulary and grammar of sign language between countries.

The alphabet of Vietnamese sign language corresponds to the alphabet of the written language, including 29 letters, compound letters, bookmarks bar, and digits.

Alphabet sign language is a kind of hand gestures. Vietnamese sign language is built similar to American Sign Language (ASL) which has been widely used in some countries. Alphabet includes 23 letters, compound words, circumflex and bars. The letters Ă, Â, Ê, Ô, Ơ, Ư, CH, GH, NGH is a combination of 2 or 3 consecutive hand gestures. The numbers from 0 to 5 are the number of fingers is often widely used in daily life, including ordinary people. Particularly the numbers from 6 to 9 differ from our imagination. The numbers from 10 onwards is a combination of two or more hand gestures.

For static gestures (hand image) we can show and graft~~ed~~ characters to form meaningful words and phrases, in the same way as the graft in written language. In addition sign language is also performed by continuous actions of the hand and arm.

1.2 Related Study

Based on the study of sign language recognition, it can be divided into two main groups based on the *method of data acquisition* and *machine learning classification, recognition*.

Classification based on data acquisition method

The first important step in sign language identification is to collect data. The data obtained are analyzed using different methods to extract features and included in the statistical model identification.

Data acquisition can also be classified into two separate groups: sensor data and computer vision.

EMG- Electromyography

Electromyography, a system of direct interaction between people and computers through the signals of the body or mind, has become an important component in the study of motion detection of the human body. The system helps the computer to understand human movements such as robot control, virtual games, and artificial limbs for people with disabilities. The computer will receive bioelectric signals from direct-attached sensors and classify. After synthesizing system data, the system usually uses artificial neural networks to classify and recognize action. The use of electromyographic signal is still continuing in research of many fields such as health, control by thoughts.

Data-Glove

Data gloves are special gloves used to monitor the shape change and movement of the hand. The device has sensors to be arranged on all the fingers and hands to detect movement and bending of the fingers, providing the location, direction, speed and direction of the hand in a fixed reference.

Camera

Data collection methods based on computer vision (camera) is widely deployed in the recognition of sign language. In this method, the gesture symbols are included with the camera fixed preset performers. Images of hand shape, the position of the fingers, palm, hand position compared with the body or facial expressions are focused. This method has the advantage of extracting both face

images and gestures of people, but often have image noises from the acquisition of images (camera resolution, light, color combination, background).

Microsoft Kinect

The first version of Kinect was announced on 2010, Kinect V2 was introduced in the summer of 2014 with many improved features: increased quality depth sensor, filming 1080p, improved Skeletal identification, enhanced infrared technology. Kinect is a device independent of the light environment, which can detect the movement of the human body in the dark.

Classification by machine learning techniques

There are many methods used to recognize sign language, these methods are based on the following parameters to extract selected features from the processing data after the acquisition by methods such as artificial neural network (ANN), hidden Markov model (HMM), support vector machine (SVM), dynamic time warping (DTW), Gaussian mixture model (GMM) ... Most of these methods are based on statistical models and self-learning, self-optimizing the parameters through the training process to improve the classification and identification based on the parameters offline.

Support Vector Machines - SVM

The SVM method was proposed by Vapnik in 1995. This is a methodology based on statistical theories that should have a rigorous mathematical foundation to ensure that the result is optimal. SVM is a method with high generality; it can be applied to many different identity problems.

Artificial Neural Network - ANN

An artificial neural network is a mathematical model or computational model built based on biological neural networks. The biggest advantage of an artificial neural network is its generality, it is capable of self-learning directly from data in predefined patterns, real-time response. There are many models of artificial neural networks trained in sign language recognition, but the most common model is the multilayered and simple recurrent network.

Hidden Markov Model - HMM

Hidden Markov model (HMM) is a statistical model in which the system is modeled supposedly Markov processes with parameters unknow before and the task is to identify the hidden parameters from the parameters observed based on this recognition. This statistical mathematical model has many applications in Bioinformatics.

Dynamic Time Warping-DTW

DTW was introduced in the 1960s, it is an algorithm to match the similarity between two sequences which may vary in time or speed. One of the characteristics of DTW that is very useful in the field of signature recognition is the ability to handle unequal signature lines (ie, the curve has a number of different x, y coordinates). This allows comparisons without resampling. This technique is now also being applied to the field of sign language recognition, gait recognition, the actions of humans, robots.

1.3 Conclusion

Chapter 1 arranges, classifies, analyses and evaluates of recent research on data acquisition, feature extraction and recognition

methods of the sign language recognition system. This will be the basis to propose and select feature extraction and recognition methods to achieve the ultimate goal of the thesis, building a Vietnamese sign language recognition system with its own characteristics.

CHAPTER 2: STATIC VIETNAMESE SIGN LANGUAGE RECOGNITION

This chapter focuses on the structure of the static sign language recognition system, the proposed object of treatment is the image of the alphabet language sign language in Vietnamese. Research is towards the application of alphabet teaching, static symbols of sign language for beginners with images from one or two hands in real time. Chapter 2 presents two directions for the study of the static hand gestures of the sign language.

2.1 Geometric modeling method

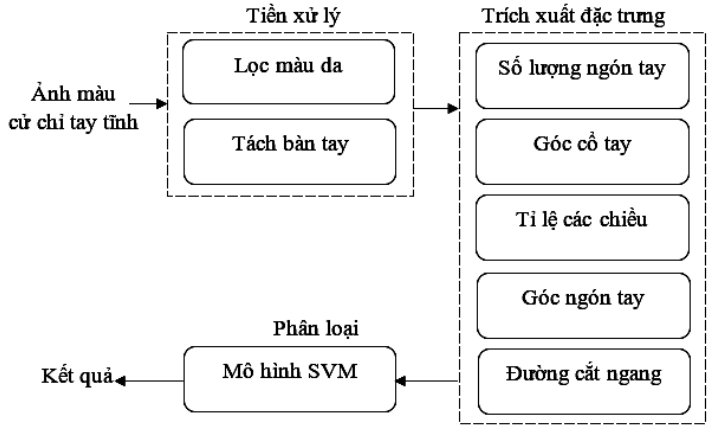

Figure 2.1: Recognition diagram based on geometry model

The first study on sign language recognition based on image processing model consists of three stages: preprocessing, extraction and classification as shown in Figure 2.1. The preprocessing phase involves two sub-phases: color filtering using color filters, sub-thresholding, and identification of color areas containing the hand image; separation will eliminate hand arm section containing little information of sign language. In the second stage features taken from the hand includes the width to height ratio, the hand angle, the finger number and the cross-section. The values obtained form feature vectors. Finally, support vector machine (SVM) with the "max-wins" strategy is used to classify hand gestures. The experiment is conducted on the color image data of the Ministry of Science and Higher Education of Poland. The research results are also tested directly on the data obtained from the Logitech QuickCam Sphere camera with a resolution of 640 * 480 performed in real time.

Preprocessing

Filter color: Using the HSV model with the available Polish data sets to achieve high generality and be able to test on the skin color of the Vietnamese. In the next step, the study uses a median filter to smooth the image and then use methods of background subtraction between the original image and the obtained image. The resulting image still has a slight interference hence we continue to use the median filter to completely remove noise.

Hand extraction: After determining the color of the object, the image obtained is often disturbed by small images, affecting hands extract information. In the next step morphological filters are used to remove noise and smooth object boundaries. In some cases, the

snapshot will be accompanied by the implementation of the human face. To remove the face, the AdaBoots algorithm is used to determine face and proceed to remove. The hand image is closest to the camera and is larger than the face. Based on this feature, the system determines a threshold value for the size of the image to proceed with the removal of the noise and get a hand image.

Feature extraction

Features proposed include general parameters: the width to height ratio, the hand angle, the finger number in addition to detailed parameters: the angle of the fingers, the number of points of intersection with the mesh cut. By combining 18 values as described above, the feature vector containing 18 elements are used in training and recognition.

Training and recognition

The multi-class SVM model is used in the study because the number of classes to classify is greater than two. Due to the large number of classes to be learned, the result obtained is determined by taking the largest value (MAX-WIN) closest to each, in which each basic classifier divides each model into a class. The model is assigned to the gesture corresponding to the class with the most value.

Experimental results

System testing is done using C ++ language. The hand gestures static data are collected from a project by the Ministry of Science and Higher Education of Poland, including 899 color photographs 27 gestures, corresponding to average 33 samples/gestures. To increase the number of samples, each image is rotated in two directions

(±20°), so there are 2697 samples for testing. This work is compatible with the fact that a gesture can be made in different directions. Dataset is divided into training and test sets with a ratio of 2: 1, corresponding to 1798 samples and 899 samples for training to the test.

In addition, the study was also tested with images obtained directly through the Logitech QuickCam Sphere camera at 30 fps and 640 * 480 resolution. Processing time per frame is about 20 milliseconds, aiming to build a recognition systems in real time, but the results do not achieve high accuracy and slow recognition response time.

2.2 *Rank-order correlation matrix (ROCM)*

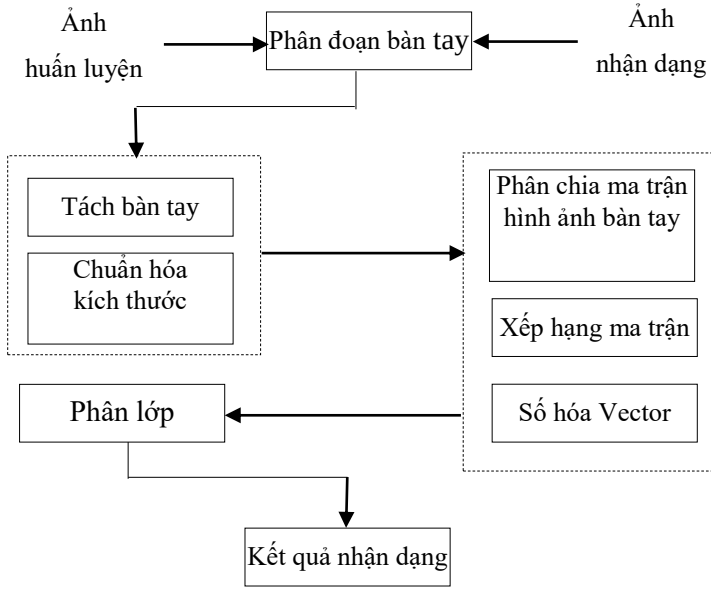

Figure 2.2: Single gesture recognition diagram

Hand Segmentation

With each image depth with the smallest value is found by scanning each line, the corresponding value dmin is the closest distance to the point of hand. Finally, we determine the area of interest the captured image contains.

Preprocessing

Cropping hand region: After selecting the appropriate image acquisition range, the image obtained by the light may be disturbed depending on the environment and sensors. Space morphology filter is used to remove interference and smooth the image, and algorithms is used to determine boundary and smooth object. Finally, we have the image of the hand based on its frame.

Size normalization: There are many methods to change the size of the hand images before the feature extraction stage. One disadvantage of the image obtained from the hand is that the size obtained is proportional to the vertical and horizontal ratios (the hand represents the vertical or the horizontal) and thus greatly affects the subsequent processing steps. So it is necessary to use a standard size for the hand images.

Feature extraction

Image subdivision: A grid is employed to divide the normalized hand image into equal blocks. The numbers of cells corresponding to two grid dimensions are equal, e.g. d (cells). Feature calculation is then performed on each cell. We aim to represent the image by a square matrix of order d.

Statistical information: In order to describe a cell, which corresponds to an image region, mean and standard deviation

parameters are used. After the calculation for all cells, the results obtained are two d-level square matrixs. The first matrix, M_at_m including d^2 average values, the second matrix M_at_{sd} d^2 including standard deviation values.

Rank matrix: Each level-2 square matrices is converted into a rank matrix with the same size based on ranking the element values. In details, each element value of M_at_m is arranged in an ascending order, in which the rank values are continuous numbers starting from 0. The M_at_{sd} calculation is similar.

Vector creation: To be appropriate for classification techniques, each rank matrix is represented as a vector, which is named a combined vector. Each element of the vector describes the relation between two neighbor image regions, corresponding to two consecutive elements of the rank matrix.

Classification and identification

Support Vector Machine (SVM) is a powerful model used for data analysis and pattern recognition, based on value-added features.

Experimental results

Table 2.1: Accuracy when testing 5 models

SVM \ ROCM	2 × 2	3 × 3	4 × 4	5 × 5	6 × 6
*Model*1	42.44 %	92.95 %	**94.22 %**	94.16 %	48.76 %
*Model*2	63.90 %	**98.27 %**	96.96 %	88.49 %	62.01 %
*Model*3	89.20 %	**99.67 %**	96.18 %	75.91 %	56.64 %
*Model*4	99.01 %	**100.0 %**	**100.0 %**	90.57 %	74.44 %
*Model*5	71.85 %	**99.01 %**	94.45 %	86.09 %	63.66 %

Similar to the recognition of digits from 0 to 9 data includes 2011 sample included 10 hand gestures.

Table 2.2: Accuracy when testing 10 gestures with 5 matrix divisions

Kích cỡ ma trận	2 × 2	3 × 3	4 × 4	5 × 5	6 × 6
Tỉ lệ	52.61 (%)	**97.61 (%)**	95.72 (%)	80.46 (%)	61.66 (%)

2.3 Automatic video segmentation in static gesture recognition

The basic concepts

A keyframe is an important information for the synthesis of the video content. A keyframe represents a video segment. After identifying, understanding the content and information of keyframes, we can accurately determine the video content. Keyframes are considered the most basic summary of the content. For processing and gathering information from the video, identifying keyframe helps to reduce a lot of time searching and handling. Based on previous studies comparison and feature extraction algorithms can be divided three groups: pixel comparison, block based comparison and color histogram comparison.

Treatment process

The treatment process can be summarized in the following description: A person performs symbols before the in-depth camera. A sequence of video frames is captured. The image segmentation process is applied to extract, capture the image of the hand area. In this data set, there are many redundant images, due to character or action transitions. The key frame identification algorithm is used to eliminate the same excess images, keeping only the mainframe. Finally features are extracted from the binary images of keyframes, applying SVM techniques for classification and identification.

Define keyframes

Symbol N is the number of images and M is a group of images with similarities, in which each movement represented in each frame is mi. Keyframes corresponding mi is determined k consecutive frames with mostly no or little change. To calculate the different degrees of successive frames, the following formula is used:

$$D_t(x,y) = |F_t(x,y) - F_{t-1}(x,y)| \qquad (2.15)$$

In that $F_t(x,y)$ is the pixel value at coordinates (x,y) $F_t(x,y), F_{t-1}(x,y)$ are two consecutive frames. When the object does not move, it means $D_k(x,y) \approx 0$ that $D_k(x,y) = |F_k(x,y) - F_{k-1}(x,y)| \approx 0$. To determine the difference of the frames $D_k(x,y)$ should be within a threshold that allows T_1 and T_2 ($T_1<T_2$). T_1 and T_2 are different average values of the pixels of two consecutive frames. In the trial for this study, the threshold values T_1=16 and T_2=48 were chosen.

2.3 Conclusion

In Chapter 2, the study proposes two methods to recognize static Vietnamese sign language based on color and depth. The first study tests the use of the available data in a color image. The data obtained from the color camera has brought comfort for implementers, not carrying devices such as gloves, sensors. For the second study, data are obtained from the depth camera of Kinect, which has fully overcome the dependence on the execution environment. The unique ROCM-based extraction pre-processing method completely eliminates the dependence on factors such as angles, geometric

edges, and finger numbers. Recognition results and training data from ROCM method provides quite satisfactory results.

CHAPTER 3: CONTINUOUS SIGN LANGUAGE RECOGNITION

The main content of this chapter is relevant studies on continuous gesture recognition of the Vietnamese sign language. Each study, the author has gradually improved and enhanced recognition results, the amount of vocabulary is built.

3.1 Recognition based on spherical coordinates

Data acquisition

Kinect v2 can identify 25 joint positions in the skeleton but after examining the Vietnamese sign language dictionary, we conclude that the movement of the hand is the most important factor, the other components of the face such as the mouth or eye movements are not used. Therefore, we only use 4 points related to arms including 2 points of the left and right hands, 2 points of the left and right elbows.

Feature extraction

In mathematics, a spherical coordinate system is a coordinate system for a 3D space in which a point position is determined by three parameters: the distance in the direction of the radius from the origin r, angle of elevation of that point from a fixed plane θ, and the angle of projection longitude perpendicular to that fixed plane φ.

After data normalization, the next thing we have to describe the normalized data. We will have a vector of 12 elements containing the data of 4 points at a time.

$$\vec{J} = \{r_{LE}, \theta_{LE}, \varphi_{LE}, r_{RE}, \theta_{RE}, \varphi_{RE}, r_{LH}, \theta_{LH}, \varphi_{LH}, r_{RH}, \theta_{RH}, \varphi_{RH}\}$$

Data will be an array of vectors $\vec{J}$ at each different time

$$Data = \{\overrightarrow{J_1}, \overrightarrow{J_2}, \overrightarrow{J_3}, \dots, \overrightarrow{J_n}\}$$

The training data will be saved to a file and will be labeled with each word of sign language.

Classification

An object is classified based on its k neighbors. K is a positive integer that is defined before executing the algorithm. Euclidean distance is used to calculate the distance between objects.

The method of applying the kNN algorithm to find the tag of gesture in the subject is to find the k-vector describing the gesture in each gesture class and the closest to the gesture of inclusion based on the DTW distance. The average distance of the k vector is calculated and they are considered the distance of the gesture. With 10 classes of gestures, we find the one that has the smallest possible distance and it is considered the class of gesture that needs identification.

The research proposes to build an improved KNN classification method combined with DTW algorithm (KNN-DTW) that is known as a cost function. When receiving a test data, the system will categorize and rank the input data with the data set with the nearest k. To ensure we continue using DTW to match and give recognition results. The input data includes 2 componets, elbow and hand data in the same array vector.

Results

The method is tested with 10 words in Vietnamese Sign Language dictionary. Each word is taken 30 samples including 20

samples for training and 10 samples for testing. Data are classified by DTW algorithm and Nearest Neighbor clustering method with the weights of 80% of arms, 20% of elbows. The operating system provides the results in real time:

Figure 3.1: Results of Vietnamese sign language recognition

WORD	RESULT
Morning	90%
Conference table	85%
Chung cake	95%
Overpass	90%
Traffic	95%
Warm	90%
Dress	80%
City	95%
Voting	100%
Volunteer	100%

3.2 Identification method divided blocks

Preprocessing

In the proposed approach, the depth image is captured one by one frame and an action is represented by a series of images.

The process of making a gesture from the action is continuous over time. The movement of hands, arms and head is concerned in the pre-processing, feature extraction. In this section, the study focuses on the movement of the hand of the subject's performance and the image acquired from a gesture depending on the time taken.

The person performing the gesture is at different distances, so the acquisition image of the object may be of different sizes, so in the first step, the preprocessor will determine the frame surrounding the object, after using the filter. The Otsu threshold will extract the image of the object.

Feature extraction

Feature extraction is an important step that greatly influences the classification and identification results. *d* is so-called the number of frames, h is the height, w is the width of the minimum bounding box. Combining the values we obtain a 3-D array of *h*w*d* elements. The objective of the next step in the pretreatment is to resize this matrix into a matrix of *n*n*n* elements without altering the properties of the matrix.

First, the array A is fixed over the time dimension *d* and the 2D space *h*w* is resized into *n*n*. As a results, the array B after the change has a size of *n*n*d*.

$$A(h,w,d) \rightarrow B(n,n,d)$$

In the next step column 1 to column n of the B array after processing, are considered as a 2D matrix of *n*d* elements, we proceed to convert the matrix *n*n* elements. The result is a 3D array of *n*n*n* elements.

$$B(n,n,d) \rightarrow C(n,n,n)$$

In this study, the Bicubic transformation is used to change the matrix size as it provides the best results and is commonly used in image processing, digital cameras, and printers. In this technique, the

new pixel value is calculated based on the average of the nearest 16 pixels closest to the average of the *4*4* matrix.

First, the elements in the array are adjusted based on the average value in the array. Let m be the mean of all elements in array C, the new value of each element in the array is determined by taking each element minus m. The sum of the elements in the new array equals zero to reduce the effect of the distance difference between the gesture and the Kinect in different gestures. Next, the C array is divided into blocks, each of which can be 4, 16 or 32 blocks for testing. Finally, each block is represented by a unique value corresponding to the average of the block elements. The result is a Z array of size *z^3*, significantly smaller than C. Figure 3.11 shows the 3D array Z obtained after dividing the 3D array C into *4^3* = 64 blocks.

Training and recognition

With the received characteristic vector Z, we continue to convert the data to fit different machine learning models, then proceed to identify and compare the results.

The dataset used in the experiment was constructed from 5 people with an average distance between the subject and the Kinect of 2.5 m. Each predefined gesture was done 30 and each symbol is performed 20 times, corresponding to 600 images for each character string. Each image depth is created at 30 frames / second with 512 * 424 resolution pixels. Figure 3.12 shows some words of the Vietnamese sign language in the collected data set.

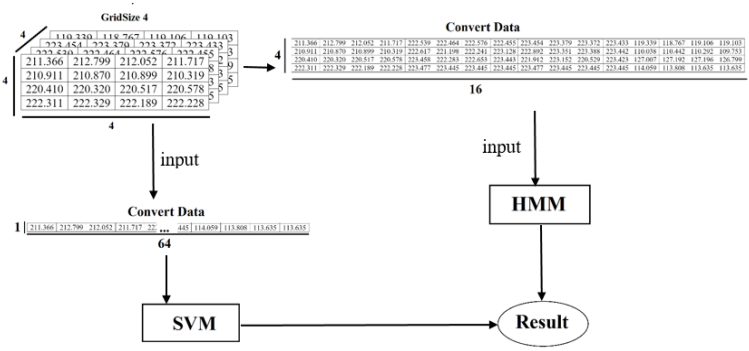

Figure 3.11: The array Z results and the vector value put into the test

Experimental results

In total 3000 models recorded by 5 people, training data is generated by the 1800 samples corresponding to three objects, and the rest is used for the testing phase. This test is done with different mesh sizes z. With each division and extract specific blocks we use two methods of classifying HMM and SVM for training and recognition.

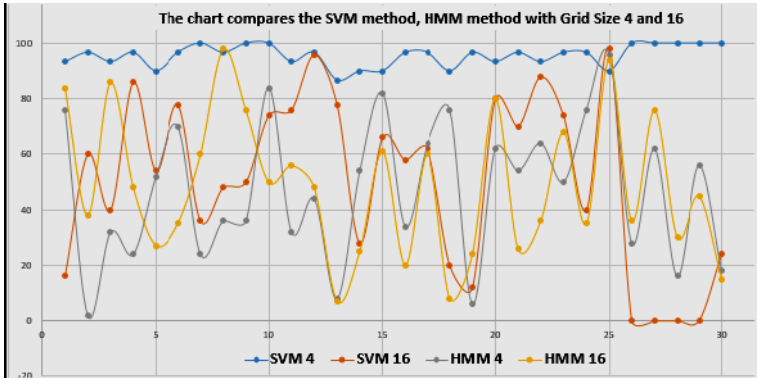

Figure 3.12: Results using SVM, HMM corresponding divided blocks 4 and 16

CONCLUSIONS AND FUTURE WORKS

The thesis has accomplished the objectives proposed for Vietnamese sign language. The following main tasks were implemented in the thesis:

- Apply geometric model method to identify static and dynamic gestures of sign language (1, 4, 5).
- Suggest method for receiving image data from the in-depth camera. Using the ROCM method for static gesture recognition studies (2).
- Study the segmentation method to extract keyframes, remove redundant frames that apply static signatures of the sign language in real time (3).
- Study continuous gestures of Vietnamese sign language based on information from joints and DTW identification model combined with kNN (6)
- Research on the continuous identification of sign language based on time division method into 3D space. Apply real-time deployment, eliminating dependency on performing gestures over time (7).

LIST OF PUBLICATIONS

1. Nguyen Trong-Nguyen, Duc-Hoang Vo, Huu-Hung Huynh, and Jean Meunier. "Geometry-based static hand gesture recognition using support vector machine." *In Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, pp. 769-774. IEEE, 2014.
2. Duc-Hoang Vo, Trong-Nguyen Nguyen, Huu-Hung Huynh, and Jean Meunier. "Recognizing vietnamese sign language based on rank matrix and alphabetic rules." *In Advanced Technologies for Communications (ATC), 2015 International Conference on*, pp. 279-284. IEEE, 2015.
3. Duc-Hoang Vo, Huu-Hung Huynh, Thanh-Nghia Nguyen, and Jean Meunier. "Automatic hand gesture segmentation for recognition of Vietnamese sign language." *In Proceedings of the Seventh Symposium on Information and Communication Technology*, pp. 368-373. ACM, 2016.
4. Duc-Hoang Vo, Huu-Hung Huynh, and Trong-Nguyen Nguyen. "Modeling dynamic hand gesture based on geometric features." *In Advanced Technologies for Communications (ATC), 2014 International Conference on*, pp. 471-476. IEEE, 2014.
5. Vo, Duc-Hoang, Huu-Hung Huynh, and J. Meunier. "Geometry-based dynamic hand gesture recognition." *Issue on Information and Communications Technology*, Vol 1 (2015): pp13-19.
6. Võ Đức Hoàng, Huỳnh Hữu Hưng, Nguyễn Hồng Sang, Jean Meunier, "Nhận dạng ngôn ngữ ký hiệu tiếng Việt với cử chỉ động dựa trên hệ tọa độ cầu." *Hội thảo quốc gia về điện tử, truyền thông và công nghệ thông tin*. Số: 1. Trang: 222-226, 2015.
7. Duc-Hoang Vo, Huu-Hung Huynh, Phuoc-Mien Doan and Jean Meunier, "Dynamic Gesture Classification for Vietnamese Sign Language Recognition" *International Journal of Advanced Computer Science and Applications(IJACSA)*, 8.3 (2017), pp. 415-420.