

# Application of Template Matching Algorithm for Dynamic Gesture Recognition of American Sign Language Finger Spelling and Hand Gesture

<sup>1</sup>Karl Caezar P. Carrera, <sup>1</sup>Alvin Patrick R. Erise, <sup>1</sup>Eliza Marie V. Abrena,  
<sup>1</sup>Sharmaine Joy S. Colot, <sup>1,2</sup>Roselito E. Tolentino

<sup>1</sup>Polytechnic University of the Philippines – Santa Rosa Campus and <sup>2</sup>De La Salle University - Dasmarias  
kenmetara@yahoo.com

Date Received: June 28, 2014; Date Published: August 15, 2014

**Abstract**—In this study the researchers developed a human computer interface system where the dynamic gestures on the American Sign Language can be recognized. This is another way of communicating by people who understands and do not understand American Sign Language. They proposed the application of template matching algorithm for the recognition of dynamic gestures where it is based on the number of templates per gesture, which must be taken by the user, to be trained and saved in the system. To be able to recognize the dynamic gestures three things must be considered. These are the number of templates required for the algorithm to be able to recognize the gestures, the factors in handling different hand orientation of other users, and the reliability of the system in terms of communication

**Keywords** - ASL, Dynamic gesture, Template Matching Algorithm

## I. INTRODUCTION

Sign language recognition is a research area involving pattern recognition, computer vision, and natural language processing. It is considered as a very important function in many practical communication applications, such as sign language understanding, entertainment, and human computer interaction (HCI). A comprehensive problem in Sign language recognition is the complexity of the visual analysis of hand gesture and the highly structured nature of sign language. There are two kinds of sign language these are Finger spelling and Hand Gesture Language. Finger spelling is a representation of letters in a writing system using only the hands while the Hand gesture language translate the words into sequence of hand gestures. Some cases that is being faced in the finger spelling recognition is the problem of identifying the dynamic gesture of other letters correctly because most of the letters create a static gesture which is easy to recognize[1]. On the other hand, problems occurring on the research about hand gesture language are the following: hand gestures are difficult to detect because of the movement they portrayed in translating words into gesture; hand gesture recognition have limited translated words which is why there is also a limitation of communication between human and computer interface. But the main problem occurred in most of the study about sign language

recognition is the detection of dynamic gestures.

The Sign language is very important for people who have hearing and speaking deficiency generally called Deaf and Mute [2]. American Sign Language is the language of choice for most deaf people in the United States. It is part of the “deaf culture” and includes its own system of puns, inside jokes, etc. ASL consists of approximately 6000 gestures of common words with finger spelling used to communicate obscure words or proper nouns. Finger spelling uses one hand and 26 gestures to communicate the 26 letters of the alphabet. The signs can be seen in Fig. 1 and Fig. 2 [3]. The use of hand gestures provides an attractive alternative to these cumbersome interface devices for human-computer interaction (HCI). In particular, visual interpretation of hand gestures can help in achieving the ease and naturalness desired for HCI. Recent researches in computer vision have established the importance of gesture recognition systems for the purpose of human computer interaction [4]. Gesture recognition is a phenomenon in engineering and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Gesture recognition enables humans to communicate with the machine (HMI) and interact naturally without any mechanical devices. This could

potentially make conventional input devices such as mouse, keyboards and even touch-screens redundant [5]. The template matching method is used as a simple method to track objects or patterns that we want to search for in the input image data [6].

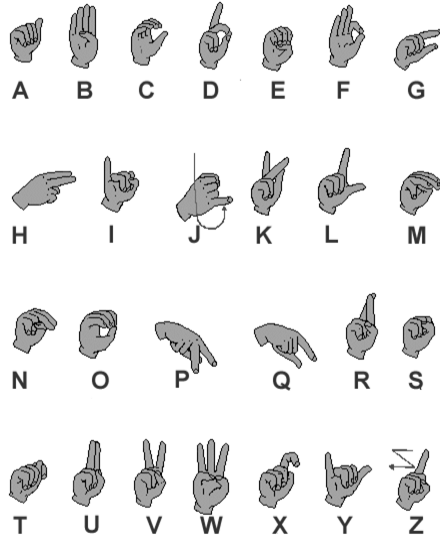


Fig. 1. American Sign Language Finger Spelling



Fig. 2. American Sign Language Hand Gesture

The problems of this research are the number of the templates required in recognizing dynamic gestures based on one user and the factor/s to be considered in recognizing different hand gestures of the other users. The main objective of the study is to be able to apply the template matching algorithm for the recognition of dynamic gestures. Other objectives are to be able to know how many templates must be applied to the proposed algorithm based on one user and to be able to know the factor/s to be considered that can affect the

recognition of different hand gestures of the users. The researchers proposed the applying of template matching algorithm for recognition of dynamic gestures for the deaf people. It is a new form of their communication using computer like chatting thru Skype and other messengers. This research will be able to help the deaf to have a conversation to others by just using their very own way of communicating which is through hand gesture.

## II. METHODOLOGY

### A. Determining the number of templates

The researchers used experimental method to know the required number of templates of a certain gesture to be taken that should be saved on the database for the matching process of the algorithm. The proponents will make experiments based on the data and research study of B.F. Johnson and J.K. Caird. From their study, they conclude that five (5) frames per seconds were sufficient for an ASL be able to recognize. Foulds similarly found out that six (6) fps can accurately identify ASL and finger spelling for an individual when smoothly interpolated to a thirty (30) frames per seconds' system. Referring from that idea, the experiment will have an initial three (3) templates per gesture. If the system will not be able to detect and recognize the gesture given with the initial templates, an additional template/s must be trained and stored in the database until the system accurately recognizes the gesture. Since performing these gestures will have different time results, the researchers used a timer in identifying the time of a certain gesture. After gathering the data, results will determine what will be the number of template per gesture is appropriate to use for the better recognition of dynamic gesture. In determining the average time, the proponents will sum up all the time in seconds under a certain number of templates and divide it by the number of gestures having the same number of templates per gesture. The researchers will use the linear regression equation to be able to know the exact number of templates to be used on a certain gesture based on the average time the gesture was performed. Below is the formula to get the linear regression of the time based on the time covered by the gestures being performed.

$$y' = a + bx \quad (1)$$

$$b = \frac{(n)(\sum xy) - (\sum x)(\sum y)}{(n)(\sum x^2) - (\sum x)^2} \quad (2)$$

$$a = \frac{(\sum y - (b)(\sum x))}{n} \quad (3)$$

**B. Identifying the factors to be consider in recognizing different hand gestures of other users**

The proponents used experimental method in order to know the factors to be considered in handling different hand gestures of the other users. There will be three (3) users that will utilize the system. The three (3) users will make trials to test the system if it can be able to identify their hand gestures without requiring them to train how they will perform a certain gesture. The main user is the one who will train the system and perform the gesture. During the training process, series of templates per gesture from the said user will be captured and saved in the database. The second user will test the system by performing the gestures done by the main user. Also the third user will do the same as what the second user perform.

The test will classify if the gesture portraying by the other user can be able to correctly recognize by the system or not. If not recognized, the proponents will capture the set of templates of the other user and compare it to the saved templates of the main user. From this data they will be able to categorize the factors regarding why the gesture of other user is not correctly recognized.

**III. RESULTS AND DISCUSSION**

**A. Determining the number of templates**

The number of templates to be used in the recognition of the dynamic gesture depends on the time it takes for a certain gesture to be performed. The number of templates for a certain time is shown in Table I.

Table 1. Number of templates per gesture based on how the gesture is performed

Word	Time of Performing the Gesture (sec)	No. of Templates	Word	Time of Performing the Gesture (sec)	No. of Templates
Always	1.49	4	Make	1.08	3
Class	1.42	4	Meet	1.30	3
Collect	1.25	3	Moment	1.50	4
Corner	1.19	3	Name	2.00	6
Doctor	1.55	4	Need	1.37	4
Family	1.75	5	Seems	1.64	5
Go	1.18	3	There	1.25	3
Group	1.45	4	Thing	1.20	3
Help	1.25	3	Turn	1.50	4
House	1.24	3	Until	1.52	4
Impossible	2.24	7	With	1.45	4
Last	1.05	3	You	1.15	3
Proud	1.30	3	Your	1.45	4

The average time being shown in Figure 3 is derived from the data gathered in Table II in calculating the time covered by each gesture portrayed by the user. Based on the collected information above, the number of templates is associated on how the system will be able to recognize a certain gesture with the appropriate number of template per gesture.

Table 2. Average Time of Each Gesture

	Number of templates per gesture				
	3	4	5	6	7
Time in seconds	1.05	1.37	1.64	2.00	2.24
	1.08	1.42	1.75		
	1.15	1.45			
	1.18	1.45			
	1.19	1.45			
	1.20	1.49			
	1.24	1.50			
	1.25	1.50			
	1.25	1.52			
	1.25	1.55			
	1.30				
	1.30				
	Average time	1.20	1.47	1.70	2.00

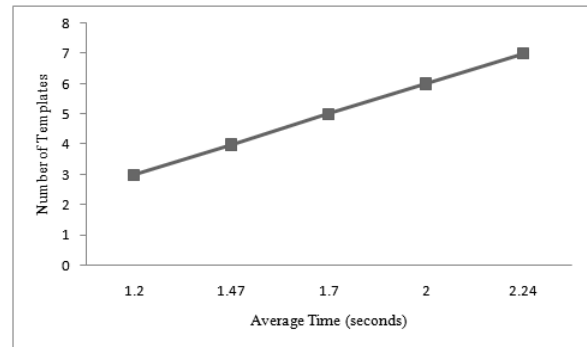


Figure 3 Number of templates with respect to time

Table 3. Values for Determining Linear Regression Equation

X	Y	xy	x <sup>2</sup>
1.2	3	3.6	1.44
1.47	4	5.88	2.16
1.7	5	8.5	2.89
2	6	12	4
2.24	7	15.68	5.02
$\sum x = 8.61$	$\sum y = 25$	$\sum xy = 45.66$	$\sum x^2 = 15.51$

Based on the values in Table 3 and using the formulas of linear regression the researchers calculated the values of the slope (b) and intercept (a) which is 0.261 and 0.417 respectively. The researchers came up to the line

regression equation given below. By this equation the researchers will be able to solve the appropriate number of templates per gesture with the respect to the time covered by a certain gesture being performed by the user.

$$y' = a + bx$$

$$y' = 3.818x - 1.575 \quad (4)$$

$$x = 0.261y' + 0.413 \quad (5)$$

**B. Factors to be considered in recognizing different hand gesture of the other users**

The following data shown on Table IV and Table V are the comparison of sample templates used in the experiment performed by the user who trained the system and the other users who did not trained the system and don't have their own template saved in the database.

Table 4. Number of Misclassified Templates for Finger Spelling

LETTERS	HAND SHAPE		HAND ORIENTATION	
	Number of misclassified sign	(%)	Number of misclassified sign	(%)
A	0	0	0	0
B	4	8.89	0	0
C	0	0	0	0
D	3	6.67	0	0
E	4	8.89	0	0
F	0	0	10	22.22
G	2	4.44	1	2.22
H	3	6.67	0	0
I	7	15.56	0	0
J	0	0	16	35.56
K	8	17.78	0	0
L	0	0	0	0
M	7	15.56	7	15.56
N	8	17.78	7	15.56
O	3	6.67	5	11.11
P	3	6.67	3	6.67
Q	0	0	0	0
R	2	4.44	0	0
S	0	0	3	6.67
T	5	11.11	1	2.22
U	7	15.56	0	0
V	4	8.89	0	0
W	4	8.89	0	0
X	3	6.67	4	8.89
Y	0	0	4	8.89
Z	0	0	13	28.89
<b>TOTAL</b>	<b>77</b>		<b>74</b>	
Average Percentage of Misclassified gesture		6.58 %		6.32 %

Each user conducted 15 trials for each gesture in the finger spelling and hand gesture. All misclassified gestures are then gathered to be able to know the error percentage of all the users in performing the gestures. It shows the percentage error of all the users based on hand shape and hand orientation. These tables illustrates that the most prevalent factor occur in recognizing finger spelling is the hand shape of the three users who performs the experiment. On the other hand the factor that affects the recognition of the hand gesture in the system is the hand orientation.

Table 5. Number of Misclassified Templates for Hand Gestures

WORDS	SHAPE		HAND ORIENTATION	
	Number of misclassified	(%)	Number of misclassified	(%)
Always	0	0	3	6.67
Class	0	0	5	11.11
Collect	4	8.89	0	0
Corner	0	0	0	0
Doctor	0	0	7	15.56
Family	0	0	11	24.44
Go	0	0	0	0
Group	0	0	3	6.67
Help	0	0	0	0
House	0	0	6	13.33
Impossible	0	0	14	31.11
Last	0	0	11	24.44
Make	0	0	5	11.11
Meet	0	0	0	0
Moments	0	0	13	28.89
Name	3	0	3	6.67
Need	2	4.44	3	6.67
Proud	0	0	9	20
Seems	0	0	4	8.89
There	0	0	0	0
Thing	5	11.11	0	0
Turn	2	4.44	3	6.67
Until	0	0	11	24.44
With	0	0	1	2.22
You	7	15.56	0	0
Your	2	4.44	0	0
<b>TOTAL</b>	<b>22</b>		<b>112</b>	
<b>AVERAGE PERCENTAGE OF MISCLASSIFIED GESTURE</b>		<b>1.88%</b>		<b>9.57%</b>

Table 6. Summary of Misclassified Gestures for Finger Spelling

USERS	SHAPE		HAND ORIENTATION	
	Number of misclassified	(%)	Number of misclassified	(%)
1 <sup>st</sup>	0	0	11	2.82
2 <sup>nd</sup>	44	11.28	26	6.67
3 <sup>rd</sup>	33	8.46	37	9.49
TOTAL	77		74	
AVERAGE	6.58 %		6.32 %	

Table 7. Summary of Misclassified Gestures for Hand Gestures

USERS	SHAPE		HAND ORIENTATION	
	Number of misclassified	(%)	Number of misclassified	(%)
1 <sup>st</sup>	0	0	4	1.03
2 <sup>nd</sup>	19	4.87	52	13.33
3 <sup>rd</sup>	3	0.77	56	14.36
TOTAL	22		112	
AVERAGE	1.88 %		9.57 %	

Table 6 and Table 7 summarize the results gathered from for determining the most common factor that causes the unrecognized gestures. Based on the results, the higher average error of misclassified gesture in finger spelling is caused by the hand shape while in the hand gesture the factor that affects the system not to recognize the gesture is the hand orientation of the different users.

#### IV. CONCLUSIONS

Based on the data gathered after conducting experiments:

1. The proponents consider the gesture being performed in considering the appropriate number of template per gesture. The number of templates depends on how long it takes for a person to perform a certain gesture. The time it takes to performed gesture is linear to the appropriate number of templates which must be used by the system in recognizing the hand gestures. With the initial time of 0.431 seconds, for every 0.261 seconds increase in time the gesture is performed, there is one template increase to be able to recognize the hand gesture.

2. From the results gathered in the experiments, the most common factor affecting the gesture not to be

recognized in finger spelling is caused by the hand shape. On the other hand, the hand orientation is the main cause why hand gestures of the users are being misclassified by the system. Another reason why gesture are unable to recognized by the system is due to the way a person delivers a certain gesture and also other user must first train and save their templates in the database before using the system for better recognition. There is also a major effect on why other user has lower accuracy rate than the main user, this is caused by the distinct hand shape of different person. If these factors were not considered there will be a greater possibility that the gesture will not be recognized.

#### V. RECOMMENDATIONS

Based on the results, there were errors/problems that the proponents encountered during the research study so the following things are highly suggested:

1. Video Segmentation Algorithm can be added to the system to prevent the misclassification of the templates having a similar template stored on a certain gesture.
2. An additional technique such as the EIGENSPACE for instances where the template may not provide a direct match because of different hand orientation from other users may be used.

#### REFERENCES

- [1] Vaishali.S. & Kulkarni et al. (2010). Appearance Based Recognition of American Sign Language Using Gesture Segmentation", *International Journal on Computer Science and Engineering* 2(3).
- [2] Sakshi G., Ishita S., & Shanu S. (2010). Sign Language Recognition System For Deaf And Dumb People, *International Journal of Engineering Research & Technology* (Vol. 2 Issue 4, April – 2013): 382 – 387.
- [3] Thad Eugene Starner (1991). Visual Recognition of American Sign Language Using Hidden Markov Models, *Massachusetts Institute of Technology*.
- [4] G. R. S. Murthy & R. S. Jadon "A Review of Vision Based Hand Gestures Recognition" *International Journal of Information Technology and Knowledge Management*, 2(2), 405 – 410.
- [5] Prof Kamal K Vyas, Amita Pareek and Dr Sandhya Vyas (2013). Gesture Recognition and Control Part 1 - Basics, Literature Review & Different Techniques, *International Journal on Recent and Innovation Trends in Computing and Communication*, 1(7), 575 – 581.
- [6] Jadhav, S.S. & Bamanikar, A. A. (2013). Recognition of Alphabets & Words using Hand Gestures by ASL, *International Journal of Emerging Trends & Technology in Computer Science*, 2(3), 56 – 59.