TELKOMNIKA, Vol.11, No.3, September 2013, pp. 583~590
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v11i3.1486

Isolated Sign Language Characters Recognition

Paulus Insap Santosa
Deparment of Electrical Engineering and Information Technology, UGM
Jalan Grafika No. 2, Yogyakarta 55281, 62 274 552305
e-mail: insap@jteti.gadjahmada.edu

Abstrak

Orang normal menggunakan bahasa percakapan untuk berkomunikasi dengan yang lainnya. Cara ini tidak bisa digunakan oleh mereka yang mempunyai cacat tuna rungu dan tuna wicara. Kedua kelompok ini akan mengalami kesulitan ketika harus berkomunikasi satu sama lain. Bahasa isyarat tidak mudah dipelajari karena banyak variasinya, dan tidak tersedia banyak tutor. Penelitian ini berfokus pada pengenalan karakter yang disebut dengan alfabet manual. Secara umum, karakter terbagi menjadi huruf dan angka. Kelompok huruf kemudian dikelompokkan lagi ke dalam beberapa kelompok menurut ciri-ciri gesturnya. Prosedur pengenalan dilakukan dengan membandingkan foto dari sebuah gestur karakter dengan kamus gestur yang telah dikembangkan sebelumnya. Kamus gestur dibuat dengan menggunakan metode jarak Euclidian, yang hasilnya kemudian dinormalisasi. Pengenalan gestur dilakukan dengan menggunakan metode tetangga terdekat dan jumlah galat absolut. Secara keseluruhan, tingkat akurasi dari metode yang digunakan adalah 96.36%.

Kata kunci: *tuna rungu, bahasa isyarat, pengenalan, gestur, metode tetangga terdekat*

Abstract

People with normal senses use spoken language to communicate with others. This method cannot be used by those with hearing and speech impaired. These two groups of people will have difficulty when they try to communicate to each other using their own language. Sign language is not easy to learn, as there are various sign languages, and not many tutors are available. This study focuses on the character recognition based on manual alphabet. In general, the characters are divided into letters and numbers. Letters were divided into several groups according to their gestures. Characters recognition was done by comparing the photograph of a character with a gesture dictionary that has been previously developed. The gesture dictionary was created using the normalized Euclidian distance. Character recognition was performed by using the nearest neighbor method and sum of absolute error. Overall, the level of accuracy of the proposed method was 96.36%.

Keywords*: hearing impared, sign language, recognition, gesture, nearest neighbor method*

1. Introduction

Until the 20th century, most people do not know exactly what sign language is, nor people with hearing impaired who often use it [1]. Most Americans think that sign language is a way of delivering English words by using special sign instead of pronunciation. In general, sign languages are not international, not all countries have unique sign languages [2]. Moreover, each sign languae has its own grammar and rule that learning one sign language does not automatically understand another one. One type of sign language is manual alphabet. It is also often referred to as finger spelling, finger alphabet, or alphabet. Manual alphabet is a letter representation in the letter writing system. The same system is also often used in a number system. By using manual alphabet, Latin alphabet (A, B, C and so on to Z) can be generated by one or a combination of two hands. Figure 1.1 shows the manual alphabet by using one hand (http://www.lifeprint.com/asl101/ fingerspelling/ images/abc1280x960.png).

Sign language is not easy to understand even by human because there are various sign languages. Moreover, because normal people have very little attention and intention on using sign language, there is a need to develop a system that can act like a tutor where normal human being can learn sign language from it. In sign language, a message is expressed as visual sign patterns that are sent its recipient. In a simple form, the visual sign consists of a combination of fingers that form certain sign, combined with wrist bending and/or hand and arm

Received May 8, 2013; Revised June 26, 2013; Accepted July 7, 2013

movement in certain direction. In a more complex sign situation, the visual sign may include arm and/or body movement. Facial expression is often embedded to increase the emotionality of the messenger. Hand gestures basically are divided into two groups, namely the sign language letter-by-letter and sign language words-by-word, or idiom.

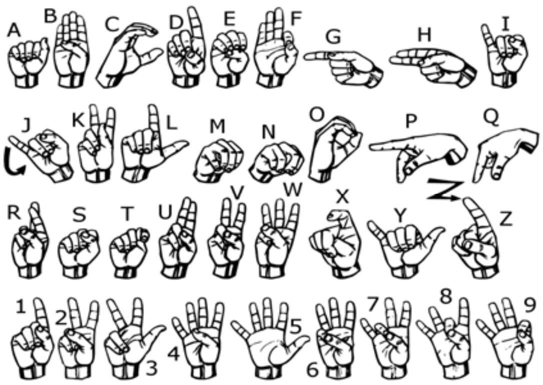


Figure 1. Manual alphabet.
(http://www.lifeprint.com/asl101/fingerspelling/images/abc1280x960.png)

Several studies have been done to make computer recognizes sign language. These studies used different input devices to represent the sign of each character and different methods in recognizing each character. Various gloves were often used as input device, including fabric glove [3][4] and various electronic and mechanical gloves [5][6]. Several method have been used in character recognition including principal component analysis (PCA) [7], a combination of statistical template matching and least mean square (LMS) [5], Gaussian model combined with thresholding [3], and Hidden Markov Model (HMM) [8][9]. In different setting, Yuniarti et al. (2012) employed k-nearest neighbor (kNN) method and binary support vector machine (SVM) [10], and Faridah et al. [11] employed object features and image texture to recognize certain image features.

In attempt to recognize characters in a sign language, Walker [7] observed that there are features in each letter. Each letter can be differentiated one to another based on its features. Walter divided features in a letter into 4 groups: holes, end point, line segments, and curves. Letter A is the letter A is said to have one hole, two end points, three line segments, and zero curve. Letter D is said to have one hole, zero end point, one line segment, and one curve. This combination can be represented as a four-item-tuple, i.e. (hole, end point, line, curve). With this tuple, letter A is represented as (1,2,3,0), and for letter D is represented as (1,0,1,1). In this study, images were manipulated images using an open source package named Octave. The training and feature extraction were done using principal component analysis (PCA). One drawback of this method was due to the fact that the direction of curves was not given. For example, the direction of curve in letter C and D could not be differentiated.

Alvi et al. [5] used a statistical template matching to recognize sign language based on the Pakistani characters. The system was built based on as many as 2500 samples from six respondents. Each respondent wore DataGlove as an input device. The status of each finger was represented using training data between 0-255. The sign language tested was based on English and Urdu. Least mean square (LMS) was used to reduce confusion in recognizing ambigious letters such as R and H. The results showed about 78.5% and 71% of accuracy rate for English alphabets with no ambiguities and the ones with ambiguities, respectively. For letters in Urdu, the system gave 85% and 69% accuracy rate for letters with no ambiguities and the ones with ambiguities, respectively. Almost similar with [5], Mohandes and Deriche [3] and Maraqa et al. [4] used colored gloves in their study to recognize an isolated Arabic Signs. In [3], single signer sat in front of video camera wore two different colored gloves for each hand. The hand tracking was done using Gaussian Model followed by adaptive thresholding. It was mentioned that the information provided by such model is sufficient to track the human face and hands in various positions and orientations. This method was then combined with region growing technique to enhance recognition accuracy. The camera was set to acquire 12 fps at a resolution of 352x576 pixels. The result showed a high accuracy of 98%.

In different settings, several methods of image identification had been used by [10] and [11]. Although their focuses were not sign language recognition, they provided different methods appropriate for sign language recognition. Specificaly, Yuniarti et al. [10] proposed a human identification system based on human dental structure. Tooth classification was done using binary support vector machine (SVM) method and k-nearest neighbor (kNN) method. These two methods gave different accuracies, i.e. 89.07% and 77.31% for SVM method and kNN method, respectively. Faridah et al. [11] used feature extraction to determine the coffee bean quality. Coffee bean features include size, shape, color, defect, and other materials. Since the coffee bean quality was determined based on its image previously taken, the calculation of its quality was combined with the image intrinsic characteristics or image texture including energy, entropy, contrast, homogeneity, and color parameter. The image texture was determined using ANOVA method with the confident level of 95%. The beans was grouped into grades I, II, III, IVA, IVB, V, and VI. The accuracy of the identification was 100%, 80%, 60%, 40%, 100%, 40%, and 100%, respectively.

This paper reports the result of a study to recognize isolated sign language characters recognition using Eucledian distance combined with k-nearest neighbor method. The research method in Section 2 explains how characters are groups according to certain criteria, followed by some discussion on how markers were placed on each finger. Section 3 presents the result and discussion, followed by conclusion on Section 4.

2. Research Method

2.1. Character Grouping

Manual alphabet in Figure 1 shows that there are two groups of characters, namely numbers and letters. In both groups there is some sort of regularities, as well as ambiguites, in term of finger-opening and closing that represent certain character. In group numbers, certain regularity is apparent from number 6 to number 9. It can be observed that from number 6 to 9 there are three fingers opening and two fingers closing in which one of them is the thumb. On the other hand, letters can be grouped into 5 groups. Table 1 shows the different groups of letters based on finger-opening and closing.

Figure 1 also shows that there are some sort of ambuguities between characters. These ambiguities need to be identified that the ambiguous characters can be treated more carefully. Table 2 presents those characters with possible ambiguities.

Tabel 1. Letters' grouping.

Group Name	Letter	Characteristic
Group 1	A, E, M, N, S, T	Letters with all fingers are closed
Group 2	B, D, F, I, K, L, R, U, V, W, Y	Letters with some fingers are opened and the rest are closed
Group 3	C, O, X	Letters with some sort of circle
Group 4	G, H, P, Q	Letters that require wrist bending
Group 5	J, Z	Letter that require movement in certain direction

Tabel 2. Characters with possible ambiguities.

No.	Possible ambiguous characters
1.	Letter B and number 4
2.	Letter F and number 9
3.	Letter D and number 1
4.	Letter W and number 6
5.	Letter K, R, U, V, and number 2
6.	Letter A, E, M, N, S, and T
7.	Letter C, O, and X

2.2. Marker Placement

In general, there two finger states, i.e. finger-opening and closing. The combination of finger-opening and closing in certain way represents certain character. For example, a combination of one finger-opening and four finger-closings forms a number 1. Another example shows that a combination of three finger-openings and two finger-closings-forms a number 3.

This arrangement needs only two different markers to differentiate betwen finger-opening and closing. Figure 2 shows the marker placement of the above examples, i.e. number 1 and number 3 as presented in Figure 2.a and Figure 2.b, respectively.

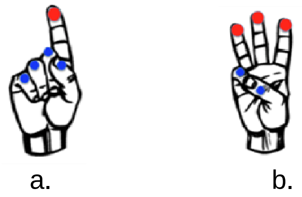


Figure 2. Marker placement for number 1 and number 3 gestures.

The marker placement as shown in Figure 2 is meant only to differentiate between finger-opening and closing. In one quick look, this arrangement seems appropriate for all characters. However, this arragement creates another problem when two different characters have similar gesture, e.g. number 1 and letter D (see Figure 1). In general, number 1 has similar gesture as letter D, i.e. there is only one finger-opening and four finger-closings. However, a closer look at these two gestures reveal that the position of the thumb in number 1 differs from the one in letter D. In number 1, the thumb covers almost all middle finger; in letter D, the thumb is only touches the tip of middle finger. This arrangement arises another problem when it comes to differentiate between number 6 to number 9. There are three finger-openings and two finger-closings in both number 6 and number 9. The only difference is the placement between the thumb and the other finger to represent different numbers. To overcome the difficulties arise from using only two markers, five different markers will be used. Figure 3 shows the placement of red, green, blue, light green, and orange markers for the thumb to the little finger, respectively.

Figure 4. Marker placement for all fingers.

2.3. The Gesture Dictionary

The gesture dictionary is prepared in two steps as can be seen in Figure 4. The first step is to determine the color feature of each marker by calculating its color components (RGB components). The second step is to create the dictionary using feature extraction by calculating the Euclidian distance between a pair of colors.

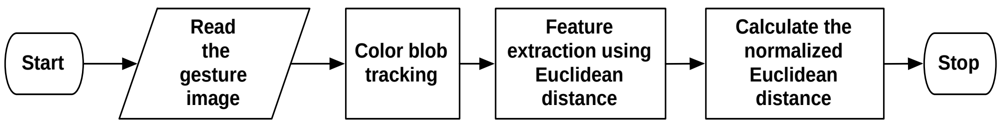


Figure 4. Steps in developing the gesture dictionary.

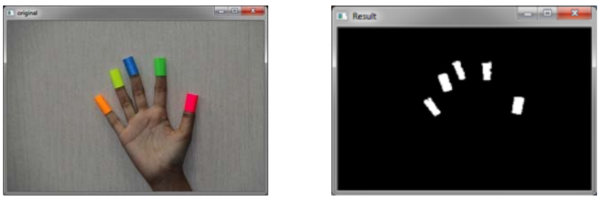
In the first step, the color components of each marker were calculated based on 30 still images from six different people, each people posed in five different fingers positions. The still images were taken using Canon 5000D. The threshold of each color component is calculated based on the average value of maximum and minimum values from the six different people for each image (see Table 3). For example, from image no. 1, the average value of the minimum and maximum value for R component of the red marker is 213 (decimal) and 255 (decimal), respectively. From image no. 1, the average value of the minimum and maximum value for R component of the green marker is 41 (decimal) and 73 (decimal), respectively. The final values that will be used for further analysis are the one located at the row entitled ‘Overall Average’. In short, the R, G, and B component of the red marker (the thumb) is in the range of 224-255, 13-63, 48-140, respectively; the R, G, and B component of the green marker (index finger) is in the range of 42-75, 168-200, 50-81, respectively; and so on. The values are ini decimal.

Table 3. Color component analysis.

Image		Thumb (red marker)			Index finger (green marker)			Middle finger (blue marker)			Ring finger (yellow marker)			Little finger (orange marker)		
		R	G	B	R	G	B	R	G	B	R	G	B	R	G	B
No. 1	Min	213	13	41	41	170	44	9	73	131	137	169	11	218	90	0
	Max	255	63	125	73	194	81	47	136	192	169	180	30	247	130	3
No. 2	Min	231	11	49	52	175	57	13	85	159	141	120	3	238	105	0
	Max	255	76	173	92	214	94	60	150	212	224	226	57	255	153	5
No. 3	Min	233	11	58	38	161	52	12	86	157	173	204	25	240	120	0
	Max	255	51	126	76	200	80	55	147	210	214	239	70	255	171	6
No. 4	Min	211	12	35	41	159	47	17	110	179	174	211	26	252	133	0
	Max	255	78	175	74	200	78	59	143	207	194	227	52	255	165	4
No. 5	Min	230	16	58	39	174	51	9	94	159	166	199	25	246	137	0
	Max	255	47	102	60	191	70	49	139	202	177	208	45	254	152	5
Overall Average	Min	224	13	48	42	168	50	12	90	157	158	181	18	239	117	0
	Max	255	63	140	75	200	81	54	143	205	196	216	51	253	154	5

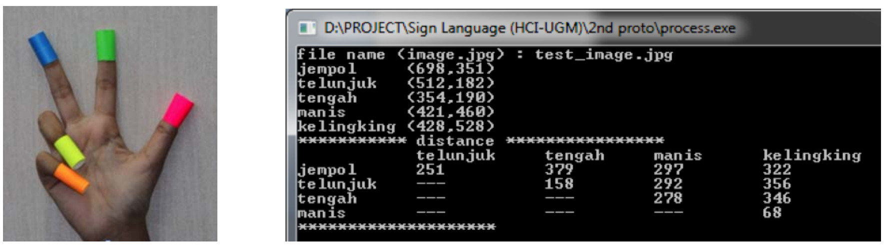
Once the threshold of each marker has been determined, the next step is to perform color thresholding using color blob tracking followed by calculating the centroid of each marker. The color blob tracking is used to determine the location of each marker. Figure 5 shows the result of the color blob tracking.

The second step starts with color feature extraction of the markers using Euclidian distance, i.e. by calculating the distance from any point in the center of one finger with another finger. Figure 6.b shows an example of Euclidian distance for the image in Fig 6.a. Figure 6.b shows that the location of the center of the marker of the thumb (*jempol*), index finger (*telunjuk*), middle finger (*tengah*), ring finger (*manis*), and little finger (*kelingking*) is at (698, 351), (512, 182), (354, 190), (421,469), and (428, 528), respectively. Figure 6.b also shows that the distance between the thumb and index finger is 251, the distance between ring finger and little finger is 68, and so on. The gesture dictionary is obtained from normalizing the Euclidian distance using linear scale normalization [12] as seen in Eq. 1.



a. b.

Figure 5. Result of color blob tracking.



a. b.

Figure 6. Example of Euclidian distance calculation

$$v_{in} = \frac{v_i - \min(v_1 \ldots v_n)}{\max(v_1 \ldots v_n) - \min(v_1 \ldots v_n)} \quad 3.1.$$

where: v_i is the value of one of several feautures in one image

v_{in} is the normalized v_i
n is the number of features

3. Results and Discussion

Table 4 shows several entries in the gesture dictionary (normalized Euclidian distance). The bigger value of certain entry means that the distance between two fingers is further than the one with smaller value. For example, the gesture for number 1 (see row 1), the distance between the thumb and the middle finger (0.077) is closer than the distance between the thumb and the index finger (0.644).

Tabel 4. Example of the gesture dictionary entries (normalized Euclidian distance).

Image			Data Dictionary			
			index finger	middle finger	ring finger	little finger
	uji_1.jpg	thumb	0.644	0.077	0.208	0.262
		index finger		0.819	0.956	1
		middle finger			0.040	0.128
		ring finger				0
	uji_2.jpg	thumb	0.834	0.757	0	0.060
		index finger		0.416	0.707	1
		middle finger			0,625	0.959
		ring finger				0.196
	uji_b.jpg	thumb	0.827	1	0.555	0.415
		index finger		0	0.285	0.375
		middle finger			0.255	0.393
		ring finger				0.375
	uji_u.jpg	thumb	0.674	0.787	0	0.061
		index finger		0.049	0.824	0.893
		middle finger			0.943	1
		ring finger				0.893
	uji_k.jpg	thumb	0	0.176	0.13	0.192
		index finger		0.166	0.92	1
		middle finger			0.943	0.889
		ring finger				1

As mentioned earlier, character recognition is conducted using nearest neighbor method. Character recognition is done by comparing the normalized distance of the test image with enties in the gesture dictionary. The absolute error (difference) of each feature is calculated, and aggregated to get a sum of absolute error (SAE) of all features of a single word. The minimum SAE shows the recognized character. Figure 7 shows the flowchart for recognizing isolated character.

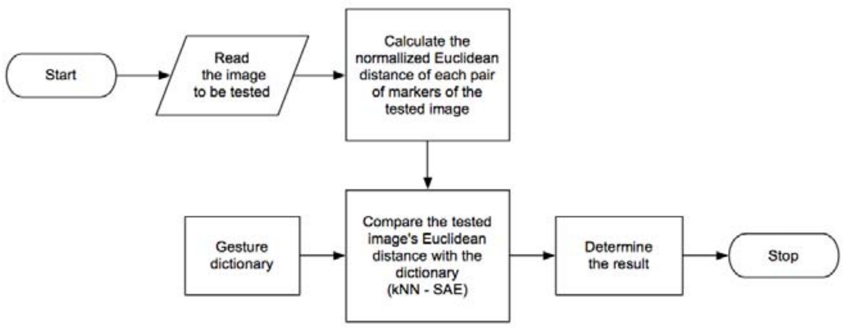


Figure 7. Flowchart of the isolated character recognition.

Figure 8.a and 8.b show the result of character recognition test for number 6 and letter U, respectively. Figure 8.a shows that the image of number 6 could be recognized directly without any ambiguity. On the other hand, letter U in Figure 8.b could not be recognized directly but through several testings until minimum SAE was found. As shown in the right picture of Figure 8.b, when letter U was recognized as number 3, the error was 3.2991 (see the line inside the top red oval). From the same picture, it can be observed that when the character in the test

image was recognized as letter U, the error was 0.5031 (see the line inside the bottom red oval). Since 0.5031 was the smallest error, the application determined that the character being tested was letter U.

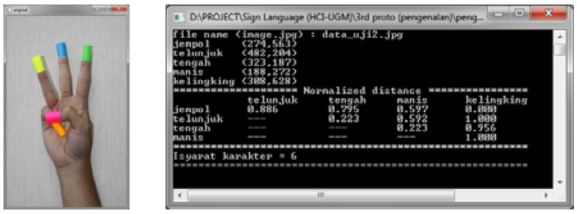


Figure 8.a Recognition of the number 2 gesture.

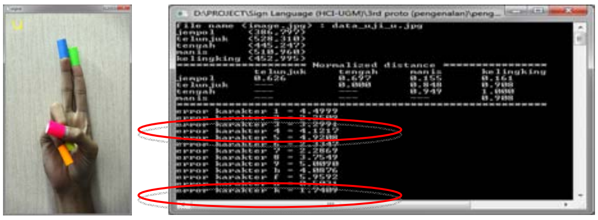


Figure 8.b Recognition of the letter U gesture.

A series of test was conducted to test the recognition accuracy. Five people were asked to display the letter and number gestures. The result is shown in Figure 9 where letters were grouped according to the first three groups as stated in Table 1, and named as Group 1, Group 2, and Group 3, respectively. Group 4 comprised gestures for number 1 to number 9. The average accuracy for each group and overall are shown in Table 5.

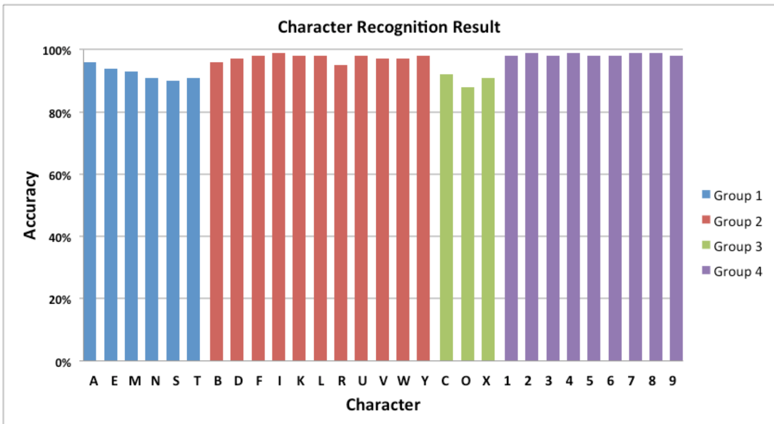


Figure 9. The recognition accuracy for each character.

Table 5. The accuracy of character recognition.

Group	Accuracy	Notes
Group 1	92.87%	Letters with all fingers are closed (A, E, M, N, S, T)
Group 2	97.76%	Letters with some fingers are opened and the rest are closed (B, D, F, I, K, L, R, U, V, W, Y)
Group 3	90,80%	Letters with some sort of circle (C, O, X)
Group 4	98.82%	Number 1 to number 9
Overall	96.36%	All of the above

Table 5 shows that the proposed procedure was able to recognize letters and numbers with different accuracy levels. In general, the proposed procedure gave the highest and the lowest accuracy in recognizing numbers and letters in Group 3, letters whose gestures have some sort of O-shape, respectively. Among all leters, the proposed procedure gave the highest accuracy for all letters in Group 2, i.e. letters with some finger-openings and closings. Overall, the accuracy of the proposed procedure is 96.36%.

Comparing the current result with the previous studies, especially [3] and [5], provides the following insight. The accuracy of the current study is lower than the one in [3] but it is higher compares to [5]. The current study was dealing with isolated character gestures performed with one hand; it utilized simple markers to differentiate one finger with the rest. In [3], the signer wore different colored glove in each hand, thus the gestures were presented using two hands. The accuracy is higher than the current study, i.e 98% compares to 96.36%. In [5], each signer wore DataGlove as an input device. The accuracy is lower than the current study, i.e. 78.5% and 71% for English alphabets with no ambiguities and the ones with ambiguities, respectively.

4. Conclusion

The intention of this study was to learn on how computer understand human gesture, especially the one related to sign language usually used by hearing impared people. This study proposed a simple procedure to recognized the gestures of isolated characters performed with one hand. The proposed procedure resulted in different accuracies for different groups of characters. Numbers were recognized with higher accuracy as compared to letters, i.e. 98.82% as compared to 92.87%, 97.76%, and 90.80%. Letters whose gestures have an O-shape were recognized with accuracy lower than the rest of the letters. Among letters, those with some finger-openings and closings were recognized higher as compared to the rest of letters, i.e. 97.76% as compared to 92.87% and 90.80%. In this study, two groups of letters were not included. The first group is the one comprises letters whose gestures need some sort of wrist bending, i.e. letter G, H, P, and Q. The second group is the one comprises letters whose gestures need some sort of movement, i.e. letter J and Z. The letter J and Z are certainly will not be able to be recognized with the proposed gesture dictionary that was based from still photograph. The future work shoud address the above limitation by finding certain representation of all letters, and other method to obtain higher recognition accuracy.

References

[1] Perlmutter DM. The Language of the Deaf. *New York Review of Books*, March 28 1991: pp. 65-72.
[2] Parton BS. Sign Language Recognition and Translation: A Multidisciplined Approach From the Field of Artificial Intelligence. *Journal of Deaf Studies and Deaf Education.* 2005; 11(1): 94-101.
[3] Mohandes M and Mohamed M. *Image Based Arabic Sign Language Recognition*. Proceedings of the Eighth International Symposium on Signal Processing and Its Applications. ISSPA. Sydney. 2005; 1: 86-89.
[4] Maraqa M, Al-Zboun F., Dhyabat M, and Zitar RA. Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks. *Journal of Intelligent Learning Systems and Applications*, 2012; 4: 41-52.
[5] Alvi AK, Azhar MY, Usman M, Mumtaz S, Rafiq S, Ur Rehman R, and Ahmed I. Pakistan Sign Language Recognition Using Statistical Template Matching. *World Academy of Science, Engineering and Technology*. 2005; 3: pp. 52-55.
[6] Hernandez-Rebollar J, Kyriakopoulos N, and Lindeman R. *A New Instrumented Approach For Translating American Sign Language Into Sound and Text*. Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea. 2004: 547 – 552.
[7] Walker E. Recognizing Images of American Sign Language Letters using Principal Component Analysis. *Integrated Research Project for Computer Vision performed in 2004 by Art Geigel*, III, http://cs.hiram.edu/~walkerel/ RASLUPCA.pdf. Accessed on 22th March 2011.
[8] Vassilia P and Konstantinos M. *Towards An Assistive Tool For Greek Sign Language Communication*. Proceedings of the Third IEEE International Conference on AdvancedLearning Technologies, Athens, Greece. 2003: 125-129.
[9] Vassilia P and Konstantinos M. *"Listening To Deaf": A Greek Sign Language Translator*. Proceedings of the Second International Conference on Information & Communication Technologies: From Theory to Applications. 2006; 1: 125-129.
[10] Yuniarti A, Nugroho AS, Amaliah B and Arifin AZ. Classification and Numbering of Dental Radiographs for an Automated Human Identification System. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(1): 137-146.
[11] Faridah, Parikesit GOF and Ferdiansjah. Coffee Bean Grade Determination Based on Image Parameter. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2011; 9(3): 547-554
[12] Pyle D. Data Preparation for Data Mining. Morgan Kaufmann Publishers, Inc. 1999.