Development of Learning Media for The Deaf Using a Webcam

Bradika Almandin Wisesa
Master of Information Technology
Unversitas Teknologi Digital
Indonesia
Yogyakarta, Indonesia
superman.wisesa@gmail.com

Widyastuti Andriyani
Master of Information Technology
Universitas Teknologi Digital
Indonesia
Yogyakarta, Indonesia
widya@utdi.ac.id

Totok Suprawoto
Information Systems
Universitas Teknologi Digital
Indonesia
totok@utdi.ac.id

Hamdani
Dept. of Informatics
Faculty of Engineering
Mulawarman University
Samarinda, Indonesia
hamdani@unmul.ac.id

2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 978-1-6654-5512-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISRITI56927.2022.10052934

***Abstract*—The lack of interactive sign language learning media makes it difficult for the deaf to find support for interactive sign language learning. Therefore, SIBI sign language interactive learning media is needed that can support interactive learning with computers as official sign language used in Indonesia. KNN method is used on this system because its suitable for pose recognition. The application can recognize all the alphabets except 'J' and 'Z because 'J' and 'Z' require gesture to be presented.**

Keywords—learning media, deaf, hand pose, recognition, SIBI, KNN.

I. Introduction

Deaf children are children who have lost part of their hearing or all of their hearing so that they experience difficulties in communicating which ultimately results in obstacles in their development, so that deaf children need special assistance or education. In general, a child is said to be deaf if his hearing does not function as normal children of the same age do. Sistem Isyarat Bahasa Indonesia (SIBI) is the sign language officially used in Indonesia approve by the government. SIBI was adopted from American Sign Language. SIBI has been inaugurated by the government, but is more often used in learning in special schools. SIBI is considered more difficult because it contains standard and complicated vocabulary, and has prefixes and suffixes. [1]

Lack of learning media that provide understanding hand pose and how we should practice the pose cause people to have difficulties in finding learning media to help them communicate with the deaf. Therefore learning media that provide learning about sign language is needed on this case. This learning media is used to make people easier learn and understanding sign language.

Learning media can clarify the presentation of messages so that they are not too verbal (in the form of written words or only spoken words). Learning media can overcome the limitations of space, time and sensory power, for example objects that are too large can be replaced with reality, images, film frames, films, or models. Small objects, assisted by micro projectors, film frames, films, or images. Movement that is too slow or too fast can be helped by timelapse or high-speed photography. Events or events that occurred in the past can be displayed again through film recordings, videos, film frames, or photos.

This process is based on machine learning and classification method used in this project is KNN (K-Nearest Neighbor). [2] All of the alphabet can be recognized on this application except letter "J" and "Z", because it require motion gesture.

II. Literature Review

A. Sign Language

Sign language is manual communication commonly used by people who are deaf. Sign language is not universal people who are deaf from different countries speak different sign languages. The gestures or symbols in sign language are organized in a linguistic way. Each individual gesture is called a sign. Each sign has three distinct parts: the handshape, the position of the hands, and the movement of the hands. Use of sign language in each country is very unique and may vary, for example, Britain and America use different sign languages even though they have the same written language, it is different with English and Spanish even though they have different written languages but the two countries use the same language. the same sign language, because until now there has been no successful international sign language to be implemented. In Indonesia, the sign language used is the Indonesian sign system (SIBI) *with the difference that lies in the shape of the letter t with American sign language (American Sign Language).* [3]

Fig. 1. American Sign Language Alphabet

B. SIBI

Indonesian Sign System (SIBI) is sign language communication used by the deaf. Consists of order systematic about a set of motion signals body, hand and lip movements represents Indonesian vocabulary. In its development Sign Language System Indonesia (SIBI) still uses the language past foreign signal added with Indonesian and artificial local signs. Indra who can help the deaf to communicate with other people is the sense of sight and must maximized, in order to help problems faced by persons with disabilities deaf. According to the physical potential possessed by deaf people are sight when it comes to socialization and understanding. Sign Language System Indonesian who follows Indonesian grammar easy to learn by teachers and parents alike hear, but it becomes a difficulty separately for deaf children who have not never know the use of Indonesian. Hoping to learn sign language, the government hopes that the deaf can communicate and obtain information like everyone else. System Dictionary appears Signs of Indonesian are also expected to be able to helping the deaf and deaf non-deaf to learn it and simplify the communication process. [4]

C. K-Nearest Neighbor

KNN algorithm is a classification algorithm that works by taking a number of K nearest data (neighbors) as a reference for determining the class of new data. This algorithm classifies data based on similarity or similarity or proximity to other data. In general, how the KNN algorithm works is as follows. [5]

- Determine the number of neighbors (K) that will be used for class determination considerations.
- Calculate the distance from the new data to each data point in the dataset.
- Take a number of K data with the shortest distance, then determine the class of the new data.

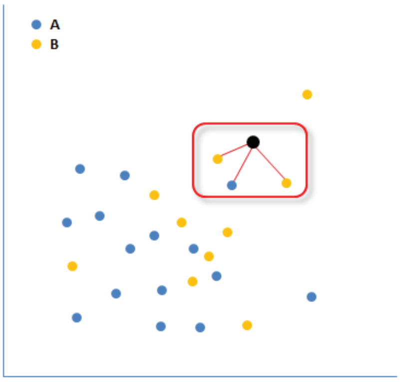


Fig. 2. Example of KNN with Blue Square and Red Triangle Neighbors

From the picture above, there are a number of data points which are divided into two classes, namely A (blue) and B (yellow). For example, there is new data (black) which we will predict the class using the KNN algorithm. From the example above, the value of K used is 3. After calculating the distance between the black points to each of the other data points, the 3 closest points are obtained consisting of 2 yellow points and one blue point as illustrated in the red box, then class for new data (black point) is B (yellow). [6]

III. ANALYSIS AND DISCUSSION

A. Overall System Stages

The following is a chart of the stages of the system in Initialize Train data and KNN Classes

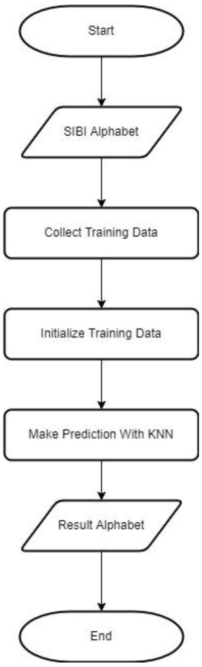


Fig. 3. Flowchart of the System

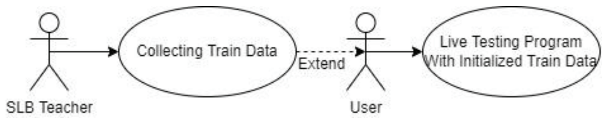


Fig. 4. Use Case of the System

B. Collecting Train Data

In order for the system to translate hand postures into letters, a gesture collection process is needed for training and testing. Each letter has 1 frame, system will take a hand posture reference 1 times. For example, below is taking the frame for the word ‘A’, ‘B’, and ‘C’. [7]

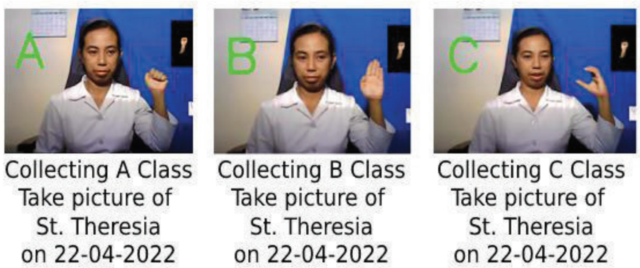


Fig. 5. Collecting Train Data

C. Train Data Initialization

Initialization of train data and classes KNN aims to initialize existing train data and classes (alphabet letters and train samples) into memory which will later be used for hand posture recognition. The image that becomes the train data must be in the form of a binary image without noise, so that the hand posture recognition process will be accurate. The letters of the alphabet that can be recognized are all the alphabets except ‘J’ and ‘Z’. Each letter of the alphabet has 50 train data (samples of hand posture) which stored in each different folders according to name of the letter of each alphabet. The steps in initializing KNN include :

- Classify KNN Class.
- Initialize all train data.
- Initialize all KNN Classes in memory.

Fig. 6. A Class

For example, a class A image consists of 2 points, 1 point consists of 3 pixels, namely :

Table I Picture Matrix Example

Piksel 1	Piksel 2	Piksel 3
3	4	5
3	5	2

The number of pixels determines how many dimensions the image has. Since Class A images consist of three pixels, three-dimensional coordinates are created. The above conditions can be illustrated as follows :

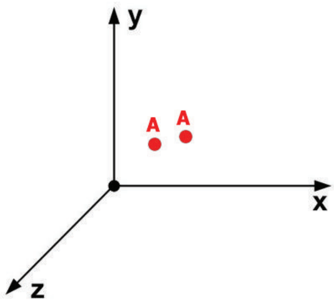


Fig. 7. Class A Coordinate

D. Preprocessing

Grayscale final result:

Fig. 8. Grayscale Result Image

Table II Grayscale Result Image Matrix

Gray 1	Gray 2	Gray 3
3	0	1
5	6	7
9	22	20
16	41	25
23	75	52
29	102	65
29	110	83
28	116	116

(0,11 x 0)+(0.58 x 8)+(0.3 x 26)

Fig. 9. Image Result of Binary Image Process

- To convert it into a binary image, we take the average of the maximum white value, which is 255/2=128. For pixels with a value < 128 will be replaced with a value of 0, otherwise for pixels with a value > 128 will be replaced with a value of 1. [8]

Table III Binary Result Image Matrix

R	G	B
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0

- The next step is to change the image to a binary inverted image because what we want is the black hand posture while the background of the hand is changed to white. Change the value 0 to 1 and the value 1 to 0. [9]

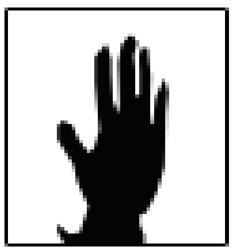


Fig. 10. Image Result of Binary Inverted Image Process

Table IV Grayscale Image Matrix

R	G	B
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1

E. KNN Classify

After the image is converted into a binary image, the next step is to match the image with the available KNN class classifications. The image that has become a binary image will be matched with the 24 knn classes stored in memory, if the input image does not have any resemblance at all to all existing classes, the system will still define it according to the closest class that has the most neighbors. This happens because the way knn works is to find the closest neighbor in image matching. [10]

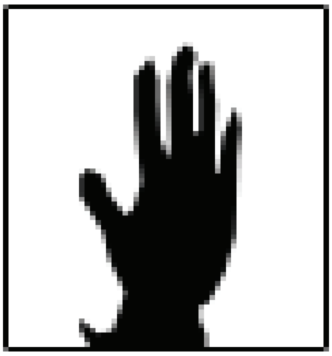


Fig. 11. Image Result of Binary Inverted Image Process

For example, the input image consists of 1 point, 1 point consists of 3 pixels, namely:

Table V Grayscale Image Matrix

Piksel 1	Piksel 2	Piksel 3
5	4	4

Fig. 12. Class A Image

For example, a class A image consists of 2 points, 1 point consists of 3 pixels.

Fig. 13. Class B Image

For example, a class B image consists of 2 points, 1 point consists of 3 pixels.

The number of pixels determines how many dimensions the image has. Since in this case the image consists of three pixels, three-dimensional coordinates are created. The condition of the three images above can be illustrated as follows:

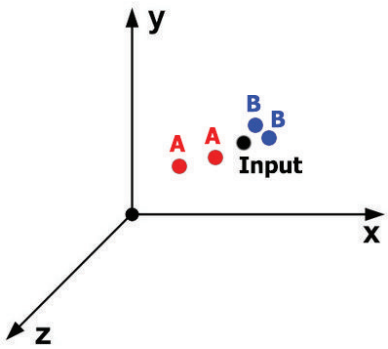


Fig 14. KNN Coordinate

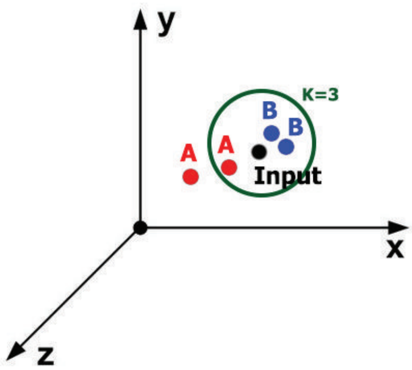


Fig. 15. KNN Classification

- With K=3 then 3 nearest neighbors are taken. Because the distance from the Input to Class A is smaller or closer than the distance from the Input to Class B, the hand detection will select Class B as the Output Class.

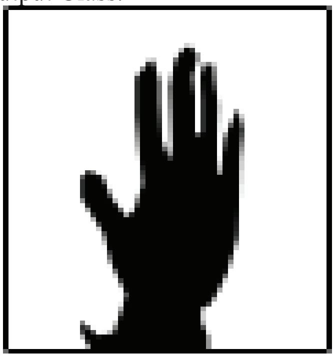


FIG. 16. IMAGE OF KNN CLASSIFICATION RESULTS

IV. RESULTS AND DISCUSSION

In the following discussion contains software testing. This test is intended to see whether the software that has been made is as planned or not. Testing is carried out by going through several stages continuously. In this test, the first thing to do is testing the routine parts that have been made, then testing the software as a whole.

A. Skin Detection

The following are the results of skin color detection testing on RGB and HSV models with different light sources:

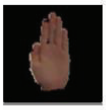
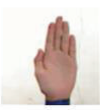
Original	Skin HSV	Skin RGB	Intensitas Cahaya
			Rendah
			Sedang
			Tinggi

Fig. 17. Skin Color Detection

The table above shows the results of cropping hand posture images taken from a webcam. The original column shows a picture of the original hand posture after the cropping process. The skin rgb column shows an image of hand posture that has been given a skin detection process (blackens pixels that are not included in human skin color) with an rgb color model. The skin hsv column shows an image of the hand posture that has been given a skin detection process (removing pixels that are not included in human skin color) with the hsv color model.

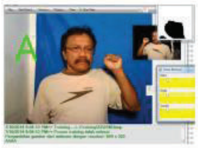
From the test results in table, it can be seen that the skin detection method in HSV colorspace has better consistency in environments with low, sufficient, or high light intensity. However, in order for the system to run smoothly, high light intensity is used, because at this light intensity, hand posture can be detected optimally.

B. Attempt

If the conditions as described in the previous chapter are met, the next step is to classify them. In the first test, 24 letter cues will be classified, namely the letters 'A' until 'Z' except 'J' and 'Z'. This test was carried out by 3 people and each of these people gave 24 hand signals that represented alphabet. This paper use 50 data per letter and the value of K variable used is 33. The following table shows the results of hand posture detected per letter of the alphabet by 3 different subjects with high light intensity, direction of light from the front, position hands straight, and within 110cm.

Table VII Results of Hand Posture Accuracy Performed by 3 Subjects

Testing A Class
Take picture of
Bradika
on 22-04-2022



Testing A Class
Take picture of
Nugroho
on 22-04-2022

Testing A Class
Take picture of
Awanta
on 22-04-2022

Fig. 17. Testing A Class

C. Confusion Matrix

After 5 attempts of hand gesture recognition for all words, i create confusion matrix to show all the accuracy results of each alphabet.

	A	B	C	D	E	F	G	H	I	K	L	M
A	4	0	0	0	0	0	0	0	0	0	0	0
B	0	4	0	0	0	0	0	0	0	0	0	0
C	0	0	4	0	0	0	0	0	0	0	0	0
D	0	0	0	2	0	0	0	0	0	0	0	0
E	0	0	0	0	1	0	0	0	0	0	0	2
F	0	0	0	0	0	4	0	0	0	0	0	0
G	0	0	0	0	0	0	4	0	1	0	0	0
H	0	0	0	0	0	0	0	4	0	0	0	0
I	0	0	0	0	0	0	0	0	4	0	0	0
K	0	0	0	0	0	0	0	0	0	4	0	0
L	0	0	0	0	0	0	0	0	0	0	4	0
M	0	0	0	0	0	0	0	0	0	0	0	2
N	0	0	0	0	1	0	0	0	0	0	0	2
O	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0

	N	O	P	Q	R	S	T	U	V	W	X	Y
A	0	0	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0
M	2	1	0	0	0	0	0	0	0	0	0	0
N	1	0	0	0	0	0	0	0	0	0	0	0
O	0	4	0	0	0	0	0	0	0	0	0	0
P	0	0	4	0	0	0	0	0	0	0	0	0
Q	0	0	0	4	0	0	0	0	0	0	0	0
R	0	0	0	0	4	0	0	0	0	0	0	0
S	0	0	0	0	0	4	0	0	0	0	0	0
T	0	0	0	0	0	0	4	0	0	0	0	0
U	0	0	0	0	0	0	0	4	0	0	0	0
V	0	0	0	0	0	0	0	0	4	0	0	0
W	0	0	0	0	0	0	0	0	0	4	0	0
X	0	0	0	0	0	0	0	0	0	0	4	0
Y	0	0	0	0	0	0	0	0	0	0	0	4

Fig. 19. Confusion Matrix

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \qquad Precision = \frac{TP}{TP + FP}$$

$$= \frac{83 + 0}{96 + 0 + 0 + 10} \qquad = \frac{83}{83 + 13}$$

$$= 0{,}86\% \qquad = 0{,}83$$

$$= 86\% \qquad = 83\%$$

$$Recall = \frac{TP}{TP + FN} \qquad F1 = \frac{2 \text{ x } Precision \text{ x } Recall}{Precision + Recall}$$

$$= \frac{86}{86 + 0} \qquad = \frac{2 \text{ x } 0{,}864 \text{ x } 0{,}864}{0{,}864 + 1}$$

$$= 1 \qquad = 0{,}927$$

$$= 100\% \qquad = 92\%$$

TP = True Positive is alphabet that have predicted value true and actual value true
('A', 'B', 'C', 'D', 'E', 'F', 'G', 'F', 'H', 'K', 'L', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W')

TN = True Negative is alphabet that have predicted value false and actual value false
('I', 'M', 'N')

FP = False Positive is words that have predicted value false and actual value true
(none)

FN = False Negative is words that have predicted value true and actual value false
(none)

V. CONCLUSION

The following conclusions can be drawn the letters of the alphabet that can be translated are the letters of the alphabet 'A' – 'Z' except 'J' and 'Z'. The letters of the alphabet j and z cannot be translated because they are in the form of hand poses.

REFERENCES

[1] D. DARMATASIA, "Pengenalan Sistem Isyarat Bahasa Indonesia (Sibi) Menggunakan Gradient-Convolutional Neural Network," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 6, no. 1, p. 56, 2021, doi: 10.24252/instek.v6i1.18637.

[2] O. Ocsirendi, "Sarung Tangan Penerjemah Huruf Sistem Isyarat Bahasa Indonesia (SIBI) Abjad (A~Z) Dengan Menggunakan Median Filter," *Semin. Nas. Tek. Elektro*, vol. 6, pp. 3–7, 2021, [Online]. Available: http://prosiding.pnj.ac.id/index.php/snte/article/view/3222

[3] M. Y. Andrian, D. Purwanto, and R. Mardiyanto, "Penerjemahan Bahasa Isyarat Indonesia Menggunakan Kamera pada Telepon Genggam Android," *J. Tek. ITS*, vol. 6, no. 1, pp. 221–224, 2017, doi: 10.12962/j23373539.v6i1.21981.

[4] L. K. Mahesta, "Pengembangan Media Kamus Digital Sistem Isyarat Bahasa Indonesia Berbasis Android di SMP-LB Karya Mulia Surabaya," *J. Mhs. Teknol. Pendidik.*, vol. 1, no. 1, pp. 1–7, 2018.

[5] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2888, no. August, pp. 986–996, 2003, doi: 10.1007/978-3-540-39964-3_62.

[6] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.

[7] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," *2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019*, no. January, pp. 1255–1260, 2019, doi: 10.1109/ICCS45141.2019.9065747.

[8] R. N. Yusra, O. S. Sitompul, and Sawaluddin, "Kombinasi K-Nearest Neighbor (KNN) dan Relief-F Untuk Meningkatkan Akurasi Pada Klasifikasi Data," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 1, pp. 0–5, 2021.

[9] M. B. S. Bakti and Y. M. Pranoto, "Pengenalan Angka Sistem Isyarat Bahasa Indonesia Dengan Menggunakan Metode Convolutional Neural Network," *Semin. Nas. Inov. Teknol.*, pp. 11–16, 2019.

[10] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," *Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017*, vol. 2018-Janua, pp. 294–298, 2018, doi: 10.1109/ICITISEE.2017.8285514.