

Hand Gesture Recognition Using Hidden Markov Model Algorithm

Roro Inda Melani

Abstract— Recognizing human hand gesture through the use of INS (Inertial navigation System) sensor, Hidden Markov Model (HMM) was used as a tool to recognize pattern statistically. Employing INS sensor to admit data input, it is assumed that hand gesture could be detected by analyzing the acceleration and fluctuation from data sensor and the difference of hand-position in 3-axis. The INS sensor that was being used was came with 6 channels to generate signals of a 3-axis gyroscope and a 3-axis accelerometer. The acceleration fluctuated in three perpendicular directions due to different hand gestures was detected by the accelerometer, while the change of hand-position in 3-axis was detected by gyroscope. Data from sensor was exported to computer via USB (Universal Serial Bus) port.

During the stage of data collection, a cut algorithm was developed to pick the most significant part of the sensor data. After finishing data comparison stage, DCT (Discrete Cosine Transform) was selected to transform the signal from time domain to frequency domain. Sequences of calculation were performed to analyze the best sampling frequency to select dominant frequency of every gesture to be picked as parameter value. The parameter value used in HMM as the approach to recognize and differs gestures.

Index Terms—3-axis gyroscope, accelerometer, gesture, gesture recognition, hand gesture, human gesture.

I. INTRODUCTION

COMPUTER distinguishes patterns in discrete digital measurement. In order to have a meaningful interaction, it is important to have such a control to make computer to feel the difference human movement. So the messages conveyed by human beings can be understood, recognized, utilized to. Therefore research in HCI (Human Computer interaction) is about to make computer understand and react in more humanize ways.

The objective of this project is to recognize human hand gesture through the use of INS sensor with the approach of statistical pattern recognition tool. It is assumed that hand gesture would be detected by the fluctuation of acceleration and the difference of hand-position in 3-axis. Therefore this project employed INS sensor to admit the information of human hand gesture. This sensor came with 6 sensor channels to generate

signals of a 3-axis gyroscope and a 3-axis accelerometer. The acceleration fluctuated in three perpendicular directions due to different hand gestures was detected by the accelerometer, while the change of hand-position in 3-axis was detected by gyroscope. Acquired data from sensor was exported to computer via USB port.

Human interacts each other by words and gestures. Usually human utters their words complements the words with facial expression, a bit of hand movements or body movements. People were convinced their words by facial expression, e.g., surprise, smirking, smiling face, and hand waving, pointing or shoulder movements. While others replying with similar ways. Those activities happened anytime and anywhere, and has been studied since human's childhood that human gestures is part of language. Among a variety of gestures, hand gesture is the most expressive and the most frequently used [1].

As a part of HCI (Human-Computer Interaction) topic, comprehending how human body moved is one of the major studies. Bridging interaction between human and computer became important since human had carried many kind of machine to daily life. The way human movement expressed in time-varying motion patterns, known as gesture.

Research on gesture recognition has many motivations. The center issue of those researches have strong relation on how human have a better interaction with computer. It is how to improve the understanding between them.

Human as computer creator appropriately have been understood how computer works and reacts according to its input. In the other side, computer need a special treatment to understand how human communicating through his natural way. Such as expressing words, voice intonation, facial expression, eyeball movement, hand waving, pointing, body movement, etc.

During the stage of data collection, a cut algorithm was developed to pick the most significant part of the sensor data. After finishing data comparison stage, DCT selected to transform the signal from time domain to frequency domain. Sequences of calculation were performed to analyze the best sampling frequency to select dominant frequency of every gesture to be picked as parameter value. The parameter value used in HMM as the approach to recognize and differs gestures.

This study is how to recognize human hand gestures through the signal produced from INS sensor using the

implementation of HMM as statistical pattern tool. In order to perform the recognition, therefore the works consist of : (1) Acquiring data input from INS sensor, (2) defining threshold, (3) signal processing (e.g., cutting the signal, restoring the most significant part, comparing the plot, state definition) , and (4) data training and hand gestures recognition.

Contribution of this is to help present the experimental results of hand gesture recognition with approaches of HMM. Further development can be directed to the need for closer relationship between human and computer through a human natural movement, especially hand gestures. This recognition can be applied to games, inputting commands trough hand movement, screening of hand movement for hand fracture patients, screening for anomalies of human body movement and other concerned purposes.

II. BACKGROUND

A. Definition of Gesture and Gesture Recognition

The definition of gesture articulated by Kurtenbach and Hulteen (1990): "A gesture is a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on its way to hitting a key is neither observed nor significant. All that matters is which key was pressed"[2].

B. INS Sensor

Until recently the weight and size of inertial sensors has prohibited their use in domains such as human motion capture. Recent improvements in the performance of small and lightweight micro-machined electromechanical systems (MEMS) inertial sensors have made the application of inertial techniques to such problems possible. This has resulted in an increased interest in the topic of inertial navigation [3].

C. Discrete Cosine Transform (DCT)

A discrete cosine transform (DCT) expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. DCTs are important to numerous applications in science and engineering, from lossy compression of audio (e.g. MP3) and images (e.g. JPEG) (where small high-frequency components can be discarded), to spectral methods for the numerical solution of partial differential equations. The use of cosine rather than sine functions is critical in these applications: for compression, it turns out that cosine functions are much more efficient (as described below, fewer are needed to approximate a typical signal), whereas for differential equations the cosines express a particular choice of boundary conditions.

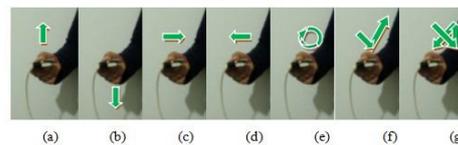
In particular, a DCT is a Fourier-related transform similar to the Discrete Fourier transform (DFT), but using only real numbers. DCTs are equivalent to DFTs of roughly twice the length, operating on real data with even symmetry (since the Fourier transform of a real and even function is real and even), where in some

variants the input and/or output data are shifted by half a sample. There are eight standard DCT variants, of which four are common.

The most common variant of discrete cosine transform is the type-II DCT, which is often called simply "the DCT"; its inverse, the type-III DCT, is correspondingly often called simply "the inverse DCT" or "the IDCT". Two related transforms are the discrete sine transform (DST), which is equivalent to a DFT of real and odd functions, and the modified discrete cosine transform (MDCT), which is based on a DCT of overlapping data

III. CONCEPT DESIGN

The gesture used in this project came from 7 hand gestures which were: up, down, right, left, circle, tick and cross. Generally all of them were representing the basic gesture how human moving their palm. But especially for tick and cross, they could seen as combination gesture. For the rest of the report those gestures would written in bold font format, such as : up, down, right, left, circle, tick and cross. Those gestures should be done by a human holding the INS sensor in a specific way as shown in Figure 1.



7 hand gestures: (a) up, (b) down, (c) right, (d) left (e) circle, (f) tick, (g) cross.

Fig. 1 Performing 7 hand gestures

The basic input-process-output was shown as in the Figure 2 below.

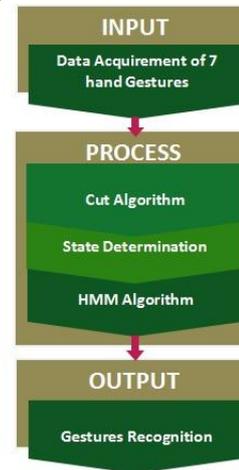


Fig. 2 Research Basic Input Process Output

IV. EXPERIMENTS

A. Data Acquisition

In the process of data acquisition every one type of hand gesture need to be performed 10 times. Every measurement took 10 seconds. A person who performed the hand gesture need to holds the INS sensor in a specific way because INS sensor works in axis system.

B. Sensor Holding Adjustment

To carry on hand gesture measurement, user need to perform specific way to hold the INS sensor. It became important thing how the user holds the sensor because INS sensor measured the gesture based on axis system, as shown in figure 3.



Fig. 3 Sensor Holding on Axis System

C. Data of 7 Hand Gestures from User

Every measurement of each hand gestures generated 70 files. Each of this file has a form in a 1000 x 7 matrix. All of one gesture measurement data gathered in that matrix. The explanation of a 1000 x 7 sized was that INS sensor employed in this project would generated 7 signals. Three signals came from accelerometer, three signals came from gyroscope and one signal of flag. One signal of flag always show same value which was=255.

Only 6 signals would be calculated and processed further in this project. The 1000 row of data represents 1000 points in the time where it was measured. There were 1000 points of measurement during 10 seconds of data acquisition

D. Gesture Cut Algorithm

The purpose of cut algorithm is to obtain the most significant part of the whole signal. During 10 seconds of measurement step, the whole signal did not representing the all significant part. Only a small part is significant. The reason of this is because when the gesture produced by the user it wasn't spend the whole 10 seconds. Therefore, it is very important to cut signal to obtain more accurate calculation.

The basic idea of cutting signal in the project was quite simple as to cut the signal where about the most significant part happen. It was assumed that its first change starts from the time=0 is the start point. Whereas the stop point was assumed as the time when the signal start to have its first change starts from opposite direction, time=100. By storing the start point and stop point, the cut signal generated by plot it only from start point until stop point.

E. Cut Threshold Value

Finding the best value of threshold that suited for every signal of gesture explained in this section. During the calculation of this step presumed first threshold value was 20000.

Next value provided were: 285, 320, 385 and 420 as the values that will be next to try. Every trial of a threshold value should be plotted and recorded for later comparison with results from other value. This process needs to be done by comparing the same gesture and also between different gestures. This process is constantly repeated and compared to obtain a graph that comes closest to natural vision of the human eyes which indicates that the only significant signal who just happens from the cutting process.

From all 4 values tested on those experiment, 285 showed its beneficial of perform the best cut in a way of how a user thought. But the threshold value= 285 did not suitable for other signals.

F. Methods to Averaging Data.

Two methods being used for the experiment to have the average from signal after it was transformed by DCT. Those two methods are absolute of average and average.

Absolute of average generates absolute value after averaging the data. While the methods of average left the data without absolute. The method 1 in this experiment refers to Absolute of average and method 2 is average without absolute.

TABLE 1. METHOD 1 APPLICATION OF GESTURE UP WITH REGION DIVISION 1-10 AND 11-20 FOR EVERY SIGNALS.

Gesture	Accelerometer signals						Gyroscope signals					
	Ax		Ay		Az		Gx		Gy		Gz	
	1-10	11-20	1-10	11-20	1-10	11-20	1-10	11-20	1-10	11-20	1-10	11-20
up1	1.3157	0.0797	0.4287	0.0944	4.7015	0.2016	50.4903	4.1128	27.8476	1.8764	14.2513	2.0079
up2	0.3396	0.0254	0.4459	0.1541	4.3285	0.1522	53.0870	0.4411	8.3082	0.6999	14.0617	0.3853
up3	0.8626	0.1998	0.3896	0.1014	4.7484	0.1635	37.8931	2.9896	29.9305	1.9597	17.1477	1.9874
up4	0.5008	0.0271	1.0588	0.0529	4.2229	0.0303	40.1126	1.7904	17.1995	0.9709	11.0684	0.9885
up5	0.8623	0.0257	1.0459	0.0336	4.3867	0.0309	30.9694	0.8086	4.8144	1.1253	1.9566	0.5710
up6	0.6597	0.0321	0.4722	0.0412	4.4199	0.0621	27.5580	1.9804	7.1687	0.4296	12.5976	0.9556
up7	1.0478	0.0746	0.8520	0.1376	4.6907	0.2102	52.1009	1.2205	79.0880	0.9163	14.8262	1.7486
up8	0.3171	0.0925	1.3380	0.0440	4.3692	0.0536	22.6983	3.9981	16.7060	2.7439	12.9680	5.1729
up9	0.8180	0.0607	0.7256	0.0142	4.3188	0.0133	11.1756	1.9028	7.6170	2.3388	4.2056	1.2539
up10	0.2294	0.0811	0.3759	0.0476	4.6446	0.1925	71.3775	3.0428	31.2030	1.1457	11.7770	0.8645

TABLE 2. METHOD 1 APPLICATION OF GESTURE UP WITH REGION DIVISION 1-5 AND 6-20 ONLY AT ACCELEROMETER SIGNALS.

Gesture	Accelerometer signals						Gyroscope signals					
	Ax		Ay		Az		Gx		Gy		Gz	
	1-5	6-20	1-5	6-20	1-5	6-20	1-10	11-20	1-10	11-20	1-10	11-20
up1	2.5558	0.0783	0.7224	0.1079	9.1742	0.2107	50.4903	4.1128	27.8476	1.8764	14.2513	2.0079
up2	0.6397	0.0301	0.8243	0.1252	8.5970	0.1215	53.0870	0.4411	8.3082	0.6999	14.0617	0.3853
up3	1.5109	0.2047	0.6773	0.1016	9.2541	0.1899	37.8931	2.9896	29.9305	1.9597	17.1477	1.9874
up4	0.9636	0.0307	2.0345	0.0629	8.4179	0.0295	40.1126	1.7904	17.1995	0.9709	11.0684	0.9885
up5	1.6888	0.0291	2.0222	0.0456	8.7339	0.0337	10.9694	0.8086	4.8144	1.1253	1.9566	0.5710
up6	1.2877	0.0320	0.8885	0.0461	8.7465	0.0725	27.5580	1.9804	7.1687	0.4296	12.5976	0.9556
up7	1.9207	0.1081	1.5816	0.1325	9.1315	0.2234	52.1009	1.2205	79.0880	0.9163	14.8262	1.7486
up8	0.4345	0.1282	2.5993	0.0516	8.6637	0.0606	22.6983	3.9981	16.7060	2.7439	12.9680	5.1729
up9	1.5223	0.0784	1.4242	0.0184	8.6147	0.0165	11.1756	1.9028	7.6170	2.3388	4.2056	1.2539
up10	0.3938	0.0757	0.6966	0.0501	9.1231	0.1837	71.3775	3.0428	31.2030	1.1457	11.7770	0.8645

Comparison between Table 1 and Table 2 gave the result that region division as shown in Table 2 promoted as better division. Therefore this division used in the experiment.

G. Combination of Two Methods

The purpose of region division is to obtain the most significant value to process to the next stage. So selecting first region from **method 1** and **method 2** was performed to obtain the purpose. The example was taken from gesture Up showed by table 1 and table 2 in fig. 4,5 and 6.

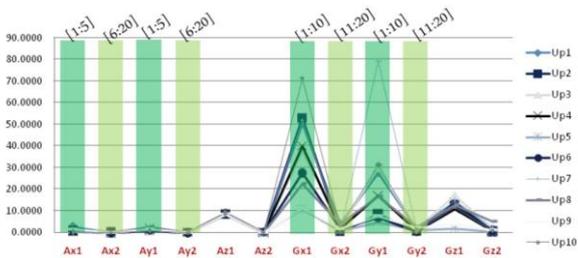


Fig. 4 Plot of method 1 gesture Up

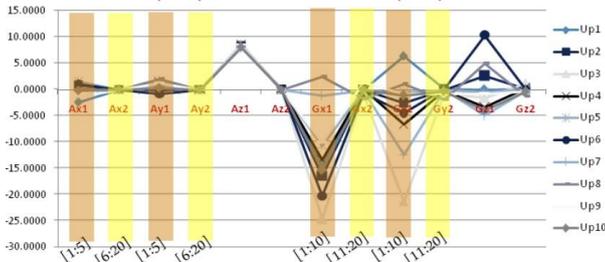


Fig. 5 Plot of method 2 gesture Up

The data picked only from first region. First region of method 1 in Figure 4 was characterized by the color of darker green, while the first region of method 2 in darker yellow as shown in Figure 5. So the combination of these two methods select only the first region of method 1 and the first region of method 2. The data combined in a new table and plotted as shown in the Figure 6.

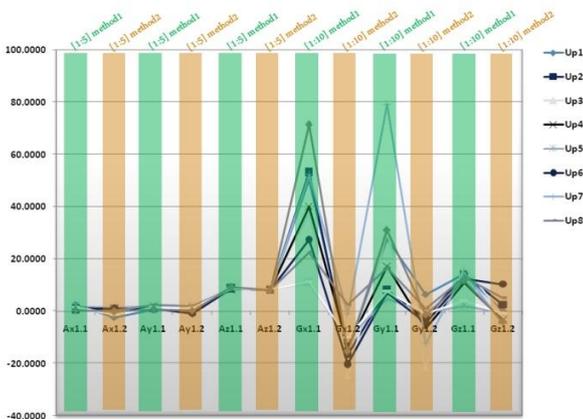


Fig. 6

Plot of method 1 and method 2 in gesture Up

V. RESULT

There were 10 times of measurement of every gesture. To find the likelihood of a gesture, 10 times of comparison need to be done. So the total is 10 as shown in the last column of Table 3.

Here come the explanation. The comparison was done by comparing one gesture to another. For an example 10 measurement of gesture Down compared to the maximum value of gesture Up.

TABLE 3.RESULT OF GESTURES RECOGNITION

Input	Up	Down	Left	Right	Circle	Tick	Cross	Total
Up	10	0	0	0	0	0	0	10
Down	1	6	2	0	1	0	0	10
Left	0	1	7	1	0	0	0	10
Right	0	0	3	7	0	0	0	10
Circle	0	0	2	0	7	0	1	10
Tick	1	0	0	0	0	9	0	10
Cross	0	1	2	1	3	1	2	10

The result in Table 3 indicates that among 10 gestures of Down : there was 1 gesture Down which were look alike with gesture Up, there were 6 gestures Down which were look alike with gesture Down itself, there were 2 gestures Down which were look alike with gesture Left, there was 1 gesture Down which were look alike with gesture Circle. Therefore the total number of comparison was always 10 times.

TABLE 4.CONCLUSION OF GESTURES RECOGNITION

Input	Right	Wrong
Up	10	0
Down	6	4
Left	7	3
Right	7	3
Circle	7	3
Tick	9	1
Cross	2	8

As be seen from table 4 there were confusion of recognizing between the gesture **Cross** from **Circle**. This could happen if the gesture was performed in a mixed way as example in Figure 7, when gesture **Cross** taken from right-up first with intersection with gesture **Circle**.

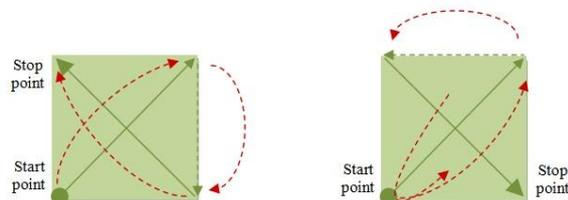


Fig. 7.

Mixture on Performing Gesture Cross

ACKNOWLEDGMENT

The authors wish to acknowledge LP2M for funding the research by the LP3S regulation. And also to the Department of Informatics Engineering for the facilities in laboratory of Multimedia and Mobile Computation.

REFERENCES

[1] Mohamed Alsheakhali, *Hand Gesture Recognition System*, Computer Engineering Department, The Islamic University of Gaza Gaza Strip , Palestine, 2011

[2] Kurtenbach, G. & Hulteen, E., *Gestures in Human-Computer Communications*. In B. Laurel (Ed.) *The Art of Human Computer Interface Design*. Addison-Wesley, 1990, 309-317.

[3] Oliver J. Woodman., *An introduction to inertial navigation, Technical Report*, Computer Laboratory, University of Cambridge, UK, 2007

Roro Inda Melani (1978) is a Lecturer in Informatics Engineering Department in UIN Maulana Malik Ibrahim Malang, East Java, Indonesia. She is a graduated students of with Double Degree program. Gained Master of Engineering from Graduate Study Program, Departement of Electrical Engineering, Faculty of Industrial Technology Institute of Sepuluh Nopember Surabaya, Indonesia (ITS), and Master of Science from Fachbereiit EIT, Hochschule Darmstadt (H-da), Germany.