

PERANCANGAN DAN IMPLEMENTASI SISTEM PENGENALAN BAHASA ISYARAT INDONESIA MENGGUNAKAN KOMBINASI SENSOR DEPTH IMAGE DAN HUMAN SKELETON KINECT DENGAN METODE HMM

DESIGN AND IMPLEMENTATION RECOGNITION SYSTEM OF INDONESIA SIGN LANGUAGE USING COMBINATION DEPTH IMAGE AND HUMAN SKELETON KINECT SENSOR WITH HMM METHOD

Yoga Satria Putra¹, Astri Novianty, S.T., M.T.², Nurfitri Anbarsanti, S.T., M.T.³

^{2,3}Fakultas Elektro dan Telekomunikasi Institut Teknologi Telkom, Bandung

¹yogasaputra93@gmail.com ²astrinov_akademiks1@yahoo.com

³anbarasanti@yahoo.com

ABSTRAK

Komunikasi adalah salah satu hal terpenting bagi kehidupan manusia. Terutama untuk tuna rungu, mereka memiliki keterbatasan dalam berkomunikasi dengan orang lain. Mereka dapat berkomunikasi dengan menggunakan bahasa isyarat. Sistem pengenalan bahasa isyarat dibutuhkan sebagai sarana komunikasi bagi tunarungu dan orang yang tidak mengerti bahasa isyarat.

Bentuk gerakan tangan dideteksi oleh sensor *depth image* dan *human skeleton* pada Kinect, dimodifikasi dan diterjemahkan menjadi sebuah teks. Hasil data rekaman mentah Kinect di segmentasi dan tracking hand dengan Haar Cascade dan diklasifikasi dengan metode Hiden Markov Model (HMM).

Dengan kombinasi sensor *depth image* dan *human sekeleton* pada Kinect serta klasifikasi metode HMM tingkat akurasi pengenalan bahasa isyarat dapat mencapai 82% dan rata-rata waktu komputasi sebesar 1.98 detik.

Kata kunci: bahasa isyarat indonesia, depth image, human skeleton, kinect.

ABSTRACT

Communication is one of the most important things for human life. Especially for the deaf, they have limitations in communicating with others. They can communicate using sign language. Sign language recognition system is needed as a means of communication for deaf and people who do not understand sign language.

Form of hand gestures detected by the sensor and the image depth human skeleton on Kinect, modified and translated into a text. Results of raw footage of data segmentation and tracking Kinect in hand with Haar Cascade and classified by the method Hiden Markov Model (HMM).

With the combination of image sensor and human depth sekeleton on Kinect and classification methods HMM sign language recognition accuracy rate can reach 82% and the average computation time of 1.98 seconds.

Keyword: indonesia sign language, depth image, human skeleton, kinect.

1. Pendahuluan

Bagi tuna rungu, bahasa isyarat merupakan sarana untuk berkomunikasi. Masalah utama adalah beberapa orang tidak pernah belajar sendiri untuk memahami bahasa isyarat. Masalah lain adalah banyak dari tuna rungu sendiri juga tidak dapat membaca atau menulis bahasa seperti pada umumnya^[7]. Efek yang paling jelas pada orang yang telah mengalami gangguan pendengaran sejak usia dini adalah lambatnya penguasaan bahasa, dan kemampuannya untuk berkomunikasi. Dalam rangka untuk mengatasi rintangan ini, pendengaran kebutuhan media dan materi pembelajaran yang tidak hanya terdiri dari komponen aural, tetapi juga dengan unsur-unsur visual yang lebih nyata dari gerakan bibir. Ini adalah alasan mengapa pengembangan bahasa isyarat diperlukan^[5]. Pada jaman sekarang banyak teknologi yang dapat digunakan untuk membantu sesuatu yang tidak dapat dilakukan oleh manusia, mungkin teknologi dapat berperan untuk bahasa isyarat, dengan mengembangkan aplikasi khusus untuk mengenali bahasa isyarat dan menerjemahkannya^[7].

Indonesia memiliki dua sistem bahasa isyarat yaitu BISINDO dan SIBI. Pada tahun 1994, Departemen Kebudayaan dan Pendidikan Indonesia merilis kamus SIBI. SIBI menjadi sistem bahasa isyarat resmi untuk bahasa Indonesia. Ini terdiri dari sistematis jari dan gerakan tangan serta tangan dan jari gerakan untuk mewakili kosa kata Indonesia^[5]. Sedangkan BISINDO merupakan bahasa isyarat alami budaya asli Indonesia yang dengan mudah dapat digunakan dalam berkomunikasi kaum tunarungu sehari-hari.

Pada penelitian tugas akhir ini, pengenalan bahasa isyarat menggunakan HMM sebagai klasifikasi, Haar Cascade untuk *tracking hand* segmentasi, dan depth image dikombinasikan dengan *human skeleton* pada Kinect. Kinect salah satu teknologi yang dapat mendeteksi gerakan tangan manusia. Dengan teknologi ini memungkinkan bahasa isyarat dapat dipahami dan dikenali. Bentuk gerakan tangan kata bahasa isyarat tersebut dimodifikasi menjadi sebuah tulisan.

2. Dasar Teori

2.1 Kinect

Teknologi yang digunakan dalam berkomunikasi oleh penderita tunarungu sangat sedikit, salah satunya ialah teknologi Kinect. Kinect merupakan perangkat yang dikembangkan oleh Microsoft yang memungkinkan untuk mengenali sendi-sendi utama pada tubuh manusia (joints) serta memperoleh posisi koordinat dari setiap sendi-sendi tersebut^[4]. Kinect (nama kode Project Natal selama pengembangan) adalah garis gerak penginderaan perangkat input oleh Microsoft untuk Xbox 360 dan Xbox One konsol video game dan PC Windows. Berdasarkan suatu kamera perifer, memungkinkan pengguna untuk mengontrol dan berinteraksi dengan konsol mereka / komputer tanpa perlu game controller, melalui antarmuka pengguna alam dengan menggunakan gerak tubuh dan perintah lisan.

Sensor Kinect dapat mendeteksi (detection) sendi tubuh manusia dan dilacak gerakannya (motion tracking). Terdapat delapan titik sendi yang digunakan yaitu: tangan kanan (hand right), tangan kiri (hand left), pergelangan tangan kanan (wrist right), pergelangan tangan kiri (wrist left), siku kanan (elbow right), siku kiri (elbow left), bahu kanan (shoulder right) dan bahu kiri (shoulder left). Fitur yang ada pada sensor Kinect meliputi kamera RGB, depth sensor, motorized tilt, dan multi-array microphone^[1].



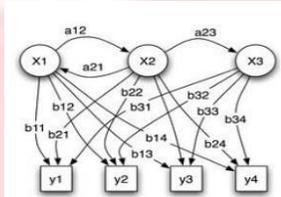
Gambar 1. Kinect^[2]

2.2 Hidden Markov Model

HMM adalah model Markov statistik di mana sistem yang dimodelkan diasumsikan oleh probabilitas transisi. HMM kuat di kemampuan belajar yang dicapai dengan menghadirkan data waktu sekuensial dan secara otomatis mengoptimalkan model dengan data. HMM terutama dikenal untuk aplikasi di pengenalan pola duniawi seperti pengenalan suara dan pengenalan tulisan tangan karena dapat menunjukkan proses waktu yang berbeda-beda dengan transit negara^[6].

Dalam Markov model biasa, setiap keadaan dapat terlihat langsung oleh pengamat. Oleh karena itu, kemungkinan dari transisi antar kondisi menjadi satu-satunya parameter teramat. Dalam HMM, keadaan tidak terlihat secara langsung. Tetapi output yang bergantung terhadap keadaan tersebut^[9].

Setiap kondisi memiliki distribusi kemungkinan disetiap output yang mungkin. Oleh karena itu, urutan langkah yang dibuat oleh HMM memberikan suatu informasi tentang urutan dari keadaan. Perlu dipahami, bahwa sifat hidden (tersembunyi) menunjuk kepada kondisi langkah yang dilewati model, bukan kepada parameter dari model tersebut. Walaupun parameter model diketahui, model tersebut tetap tersembunyi^[9].



Gambar 2 Representasi Parameter HMM^[9]

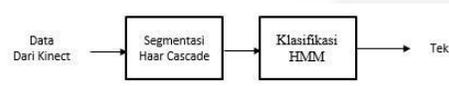
Penjelasan Gambar 2.2:

- x = kondisi
- y = observasi yang mungkin
- a = kemungkinan keadaan transisi
- b = kemungkinan output

HMM dapat digunakan untuk aplikasi dibidang *temporal pattern recognition* (pengenalan pola temporal) seperti pengenalan suara, tulisan, gestur, bioinformatika, kompresi kalimat, *computer vision*, ekonomi, finansial, dan pengenalan not balok^[9].

3. Perancangan Sistem

3.1 Gambaran Umum Sistem



Gambar 3 Gambaran Umum Sistem

Gambaran umum sistem pengenalan bahasa isyarat pada gambar di atas dapat dilihat sistem yang akan dibangun merupakan sistem pengenalan bahasa isyarat Indonesia menggunakan data Kinect sebagai alat media perekam dengan mengkombinasikan *depth sensor* atau *stream* dengan *human skeleton data tracking*. Data yang dimaksud adalah rekaman mentah gerakan bahasa isyarat dari Kinect. Pada sistem ini dirancang agar dapat membaca data inputan *gesture* yang sebelumnya sudah direkam dengan kamera pada Kinect dan disimpan kedalam *database*, data akan berupa kordinat yang sudah di tetapkan pada titik sendi pada *depth image* dan *human skeleton tracking* sebagai inputan yang sudah tersegmentasi pada bagian tangan dan *tracking hand* dengan teknik *haar cascade* untuk mendeteksi tangan, yang nantinya akan di uji performansi akurasi keluaran (kata) dari *gesture* bahasa isyarat dengan klasifikasi HMM.

Kemudian HMM melakukan klasifikasi memunculkan kata yang sesuai dengan *gesture* yang sudah di simpan dalam *database* sebelumnya dan melakukan uji performansi akurasi data dengan jumlah data yang benar dibagi dengan jumlah data yang diuji. Aplikasi ini berguna sebagai pembelajaran apakah metode yang diuji lebih baik dari metode yang lainnya.

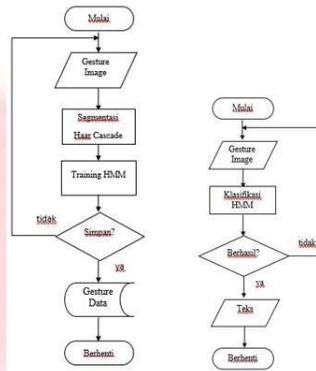
Pada Gambar 3 dapat dideskripsikan bahwa perancangan umum sistem adalah sebagai berikut:

1. Bahasa isyarat akan diperagakan dari kata yang sudah dipilih, *gesture* dari kata tersebut akan di rekam oleh kinect, kinect akan menjadi media perantara dari gerak *gesture* yang dapat diolah dengan mengkombinasikan *depth sensor* dan *human skeleton*.
2. Kinect akan mengolah data berupa *depth image* dan *human skeleton* sebagai inputan. Inputan tersebut berupa *gesture* yang diolah menjadi *depth image* dan *human skeleton data tracking*. Disegmentasi dan *tracking hand* oleh Haar Cascade pada bagian tangan.
3. Data inputan disimpan kedalam *database*, di *training* dan di klasifikasi oleh metode HMM.
4. *Gesture* yang telah direkam akan diterjemahkan menjadi berupa teks yang diuji dengan Kinect dan data sebelumnya yang sudah diolah yang disimpan pada *database* untuk mendapatkan teks atau kata dari *gesture* yang dipilih.

Gambar 3 menunjukkan alur kerja secara umum dari sistem pengenalan bahasa isyarat Indonesia ini. Dimulai dari pengambilan *gesture* lalu menyimpannya ke dalam *database*, mengolah hingga menampilkan keluaran kata dari *gesture* yang sudah direkam sesuai dengan kata yang akan di inginkan.

3.1.1 Diagram Alir Perancangan Umum Sistem

Diagram alir dari sistem pengenalan bahasa isyarat secara keseluruhan dapat dilihat pada diagram alir berikut.



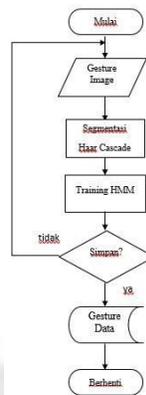
Gambar 4 Diagram Alir Perancangan Training Dan Testing Data

Pada Gambar 4 dapat didefinisikan alur dari sistem yang di bangun dalam aplikasi ini sebagai berikut:

- a. Kamera kinect merekam *gesture* dari kata bahasa isyarat.
- b. Hasil rekaman *gesture* atau data akan disimpan pada *database*.
- c. Melakukan proses segmentasi dan *tracking hand* dengan Haar Cascade.
- d. Data *gesture* atau inputan dari Kinect yang telah disimpan dalam *database* dimuat dan akan diklasifikasi oleh metode HMM.
- e. Setelah data diklasifikasi oleh metode HMM, maka akan mendapat kelauran dari *gesture* yang berupa kata atau teks.

3.2 Diagram Alir Proses Training Data

Dalam tahapan ini ditentukan proses untuk merekam dan menyimpan data. Pada Gambar 5 yang merupakan bagian proses merekam dan menyimpan data.

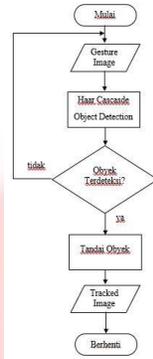


Gambar 5 Diagram Alir Proses Merekam dan Menyimpan Data.

Diagram Alir pada Gambar 5 dapat dilihat proses *training* data. Merekam gerakan yang sesuai dengan kata yang diinginkan, saat proses rekam sistem melakukan segmentasi dan training lalu akan disimpan kedalam *database*.

3.3 Perancangan Alir Proses Segmentasi Dan Tracking Hand

Kinect *runtime processes depth data* untuk mengidentifikasi hingga enam *player* dalam peta segmentasi. Peta segmentasi adalah bitmap dengan nilai-nilai pixel yang sesuai dengan indeks orang di jarak pandang yang paling dekat dengan kamera di posisi pixel. Segmentasi *player data* hanya tersedia pada *depth stream* ketika skeletal tracking diaktifkan.



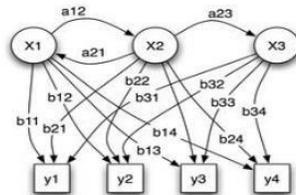
Gambar 6 Diagram Alir Hand Detection Haar Cascade

Pertama image terdeteksi yang berasal dari frame rekaman gerakan, lalu ditentukan area yang akan dideteksi apakah terdapat obyek atau tidak, obyek yang dimaksud adalah tangan. Proses berikutnya adalah melakukan pendeteksian obyek menggunakan Haar Cascade, apabila terdeteksi maka obyek tersebut ditampilkan atau ditandai dalam kotak persegi.

3.4 Perancangan Hidden Markov Model

Langkah-langkah yang dilakukan untuk mengklasifikasikan *gesture* menggunakan HMM adalah sebagai berikut.

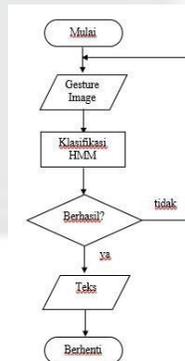
- a) Menentukan *gesture* bahasa isyarat dari BISINDO yang digunakan.
- b) Menentukan *state* yang digunakan seperti pada gambar di bawah ini.



Gambar 7 HMM state

- S1-SN adalah *hidden state* yang berisikan *database* kata-kata bahasa isyarat yang direkam.
- a_{ij} adalah probabilitas transisi dari *state* i ke *state* j.
- *gesture* yang direkam secara *real time* adalah observe state.
- c) Mengumpulkan data *training* yang telah diproses sedemikian rupa dari data *raw* dengan hanya mengambil informasi dari data *depth image* dan *skeleton kinect*.
- d) Melakukan proses *training data* terhadap HMM melalui data *training*. Pelatihan ini merupakan unsur yang sangat penting dalam menggunakan pendekatan HMM. Pada proses *training data* ini akan digunakan algoritma *Baum-Welch*.
- e) Melakukan proses *testing data* terhadap data yang telah di *training*. Sehingga dapat diklasifikasikan. Pada proses ini digunakan algoritma *Forward-Backward*.

3.5 Diagram Alir Proses Pengujian



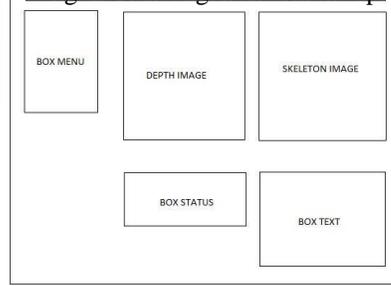
Gambar 8 Diagram Alir Pengujian

Diagram Alir pada Gambar 8 dapat dilihat proses pengujian akhir kata apakah inputan yang kita masukan akan sama seperti data yg telah direkam sebelumnya dalam database. Ketika memasukan input gesture, input gesture akan membuat kordinat data yang baru dan dicocokkan dengan data kordinat yang sudah disimpan sebelumnya dalam *database*. Kemudian metode HMM akan diterapkan apakah data sama atau ada kecocokan dengan data yang ada di *database*. Ketika terdapat kesamaan pola *feature* antara dataset dengan citra yang ditangkap maka status akan berhasil membaca kata, jika tidak maka kita kembali melakukan inputan kembali. Akurasi merupakan ukuran ketepatan sistem dalam mengenali masukan yang diberikan sehingga menghasilkan keluaran yang benar. Rumus akurasi yang dapat dituliskan sebagai berikut.

$$\text{Akurasi} = \frac{\text{Jumlah Data yang Benar}}{\text{Jumlah Data}} \times 100\%$$

3.6 Perancangan Antarmuka

Perancangan Antarmuka mendeskripsikan rancangan tampilan dari setiap halaman yang akan digunakan pada aplikasi yang dibuat. Perancangan antarmuka dibuat untuk merepresentasikan keadaan aplikasi yang akan dibangun. Rancangan antarmuka aplikasi yang dibangun seperti berikut.



Gambar 9 Tampilan Perancangan Sistem

Pada Gambar 9 *depth image* akan diisi gambar *depth*, *skeleton image* akan diisi gambar *human skeleton* data *tracking*.. *Text Box* akan berisisebagai keluaran kata yang diterjemahkan dari *gesture*. *Box Menu* berisi menu untuk merekam, menyimpan gesture, dan memanggil gesture. *Box status* berisi informasi kata yang dipilih, frame rate, dan status.

3.7 Skenario Pengujian

Skenario pengujian yaitu pengujian performansi yang berupa akurasi sistem dalam hal mengenal gesture perkata pada jarak (1m, 2m, dan 3.5m), intensitas cahaya (terang dan gelap), dan kecepatan waktu mengenali gerakan.

4. Perancangan Sistem

4.1 Implementasi Sistem

Pada bab ini akan dilakukan implementasi terhadap sistem yang telah dibuat dan dilakukan pengujian untuk menganalisis hasil dari sistem yang dibuat.

4.2 Implementasi Perangkat

Kebutuhan yang diperlukan untuk menunjang penelitian ini terdiri dari kebutuhan perangkat lunak dan perangkat keras. Berikut ini adalah perangkat-perangkat yang digunakan dalam penelitian ini, berupa perangkat keras (*hardware*), perangkat lunak (*software*).

4.2.1 Kebutuhan Perangkat Lunak

Spesifikasi perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut :

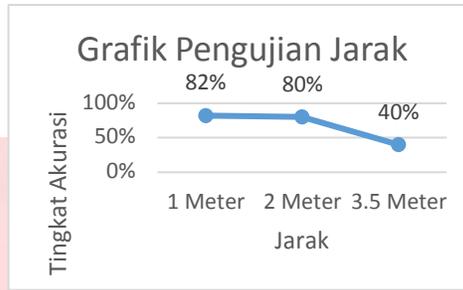
- a) Sistem operasi windows 10
- b) Kinect for windows SDK V1.8
- c) Microsoft Visual Studio 2015

4.2.2 Kebutuhan Perangkat Keras

Spesifikasi perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

- a) *amd phenom II x3 710 processor*
- b) *ram 8 gb*
- c) *vga geforce gtx 670*

4.3 Pengujian Pengujian Akurasi Sistem Berdasarkan Jarak

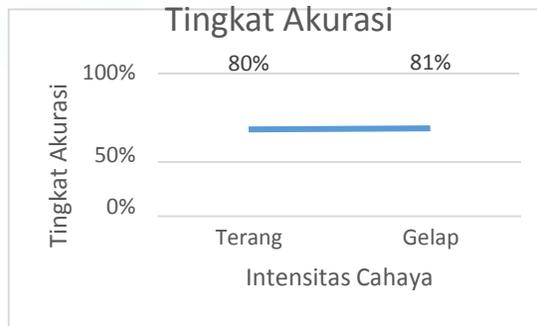


Gambar 10 Grafik Tingkat Akurasi Pada Jarak

Dari data hasil pengujian akurasi sistem berdasarkan jarak dapat dianalisis beberapa hal sebagai berikut.

- a) Akurasi sistem tertinggi terdapat pada jarak 1 meter dengan nilai rata-rata akurasi adalah 82%.
- b) Akurasi sistem terendah terdapat pada jarak 3,5 meter dengan nilai rata-rata akurasi adalah 40%.
- c) Jarak optimal sistem mengenali objek adalah 1-2 meter.

4.4 Pengujian Akurasi Sistem Berdasarkan Intensitas Cahaya



Gambar 11 Grafik Tingkat Akurasi Pada Intensitas Cahaya

Dari data hasil pengujian akurasi sistem berdasarkan jarak dapat dianalisis beberapa hal sebagai berikut.

- a) Akurasi sistem tertinggi terdapat pada intensitas cahaya gelap dengan nilai rata-rata akurasi adalah 81%.
- b) Akurasi sistem terendah terdapat pada intensitas cahaya terang dengan nilai rata-rata akurasi adalah 80%.

4.5 Pengujian Waktu Komputasi Sistem

No	Kata	Waktu (s)
1	Ampun	3.45
2	Apa	1.29
3	Berapa	1.19
4	Halo	0.97
5	Kanan	0.95
6	Kiri	0.94
7	Maaf	1.52
8	Makan	2.05
9	Minum	2.27
10	Permisi	1.04
11	Sampai	2.73
12	Selamat	3.97
13	Siapa	2.21
14	Terimakasih	3.98
15	Tolong	1.24
	Rata-Rata	1.98

Dari data hasil pengujian dapat disimpulkan bahwa rata-rata waktu komputasi sistem yaitu 1.98 detik (s).

5. Kesimpulan dan Saran

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari penelitian Tugas Akhir ini sebagai berikut.

1. Tingkat akurasi tertinggi pada jarak 1 meter dengan nilai 82%, pada jarak 2 meter yang menghasilkan nilai 80%. Jadi jarak optimal sistem mengenali objek dan gerakan adalah 1-2 meter.
2. Tingkat akurasi tertinggi pada intensitas cahaya, yaitu pada intensitas cahaya gelap dengan nilai 82%, hanya berbeda 1% pada kondisi terang yang menghasilkan nilai 80%.
3. Tingkat akurasi terendah dari semua pengujian yaitu pada jarak 3.5 meter yang hanya menghasilkan 40%.
4. Waktu komputasi yang dibutuhkan oleh sistem adalah 1.98 detik.

5.2 Saran

Saran untuk penelitian selanjutnya sebagai berikut.

1. Menggunakan inputan RGB pada Kinect, tidak mematikan salah satu karena dapat mempengaruhi tingkat akurasi sistem.
2. Menggunakan algoritma atau metode segmentasi yang lain dan tidak hanya pada tangan tetapi *tracking* pada jari juga, metode klasifikasi yang lain agar mendapat tingkat akurasi yang lebih tinggi.
3. Menggunakan Kinect dan Kinect SDK versi 2.0 atau yang terbaru, karena memiliki fitur yang lebih bagus dari versi sebelumnya.
4. Menggunakan komputer dengan spesifikasi yang lebih tinggi dari yang digunakan pada tugas akhir ini karena membuat sistem tidak stabil (*lag*) dan mempengaruhi nilai waktu komputasi.

DAFTAR PUSTAKA

- [1] Gunawan A A S , Salim A. 2013. *Pembelajaran Bahasa Isyarat Dengan Kinect Dan Metode Dynamic Time Warping*.
- [2] Le T L., Nguyen M Q, Nguyen T T M. 2011. *Human posture recognition using human skeleton provided by Kinect*. Grenobleinp Hanoi University of Science and Technology. IEEE-ComManTel.
- [3] Li K F, Lothrop K, Gill E, Lau S. 2011. *A Web-Based Sign Language Translator Using 3D Video Processing*. IEEE-NBIS.
- [4] Ekasari D S, Yuhana U L, Hariadi R R. 2013. *Rancang Bangun Modul Pengenalan Bahasa Isyarat Menggunakan Teknologi Kinect*.
- [5] Rakun E, Andriani, M, Wiprayoga I W, Danniswara K, Tjandra A. 2013. *Combining depth image and skeleton data from Kinect for recognizing words in the sign system for Indonesian language (SIBI [Sistem Isyarat Bahasa Indonesia])*. IEEE-ICACSIS.
- [6] Song Y, Gu Y, Wang P, Liu Y, Li A. 2013. *A Kinect Based Gesture Recognition Algorithm Using GMM and HMM*. IEEE-BMEI.
- [7] Tolba M F, Elons A S. 2013. *Recent Developments in Sign Language Recognition Systems*. IEEE-ICCES.
- [8] Catuhe D. 2012. *Programming With The Kinect For Windows Software Developer Toolkit*.
- [9] Prasetyo M E B. 2011. *Teori Dasar Hidden Markov Model*.
- [10] Mourad B, Tarik A, Karim A, Pascal E. 2014. *Real-Time System of Hand Detection And Gesture Recognition In Cyber Presence Interactive System For E-Learning*.