

**PERANCANGAN DAN IMPLEMENTASI SISTEM  
PENGENALAN BAHASA ISYARAT INDONESIA  
DENGANMENGKOMBINASIKAN RGB DAN  
SKELETON KINECT MENGGUNAKAN HIDDEN  
MARKOV MODEL**

*Design And Implementation Of Indonesia Sign Language  
Recognition System By Combining RGB And Skeleton Kinect  
Using Hidden Markov Model*

**Gunanda Hasdiansyah<sup>1</sup>, Astri Novianty, S.T., M.T.<sup>2</sup>, Nurfitri  
Anbarsanti, S.T., M.T.<sup>3</sup>**

<sup>2,3</sup>Fakultas Elektro dan Telekomunikasi Institut Teknologi Telkom, Bandung  
<sup>1</sup>[g.hasdiansyah@gmail.com](mailto:g.hasdiansyah@gmail.com) <sup>2</sup>[astrinov\\_akademiks1@yahoo.com](mailto:astrinov_akademiks1@yahoo.com)  
<sup>3</sup>[anbarsanti@yahoo.com](mailto:anbarsanti@yahoo.com)

---

**ABSTRAK**

Sistem pengenalan bahasa isyarat Indonesia sangat dibutuhkan sebagai sarana komunikasi antara tunarungu dengan orang yang tidak mengerti bahasa isyarat. Ada dua model bahasa isyarat Indonesia yang digunakan yaitu BISINDO dan SIBI. Pada tugas akhir ini digunakan model BISINDO karena lebih mudah dipahami oleh para tunarungu dibandingkan SIBI.

Pada tugas akhir ini dibangun sistem pengenalan bahasa isyarat Indonesia dengan memanfaatkan data RGB dan skeleton kinect. Metode yang digunakan adalah Hidden Markov Model (HMM) sebagai metode klasifikasi.

Dari hasil pengujian diperoleh akurasi sistem tertinggi sebesar 78% dengan jarak optimal 2 meter dan waktu komputasi 1.473311 detik.

**Kata kunci : bisindo, hmm, kinect.**

---

**ABSTRACT**

*Indonesian sign language recognition system is needed as a means of communication between deaf with people who do not understand sign language. There are two models of Indonesian sign language used is BISINDO and SIBI. In*

*this final project is used BISINDO because it is more easily understood by the deaf than SIBI.*

*In this final project constructed Indonesian sign language recognition system by utilizing RGB data and skeleton kinect. The method used is Hidden Markov Model (HMM) as a method of classification*

*From the test results obtained the highest system accuracy by 78% with the optimum distance of 2 meters and computation time 1.473311 seconds.*

**Keyword : bisindo, hmm, kinect.**

## 1. Pendahuluan

Dewasa ini, untuk mempermudah para tunarungu mendapatkan informasi maka diciptakanlah BISINDO (Bahasa Isyarat Indonesia) oleh para tunarungu. Tetapi walaupun telah diciptakan BISINDO, masih terdapat kesulitan komunikasi antara tunarungu dengan orang yang tidak mengalami disabilitas. Hal ini dikarenakan tidak semua orang mengetahui bahasa isyarat.

Perkembangan informasi dan teknologi di dunia saat ini semakin pesat, sehingga pada tugas akhir ini dilakukanlah perancangan dan implementasi gesture recognition untuk membuat suatu sistem pengenalan bahasa isyarat Indonesia. Untuk melakukan gesture recognition digunakanlah kinect sebagai alat perekamnya. Kemampuan kinect yang bisa memberikan data RGB dan skeleton menjadi alasan dipilihnya kinect.

Metode yang digunakan pada pengerjaan tugas akhir ini adalah histogram of oriented gradients (HOG) sebagai metode untuk segmentasi RGB. Segmentasinya akan difokuskan pada bagian tangan. Untuk klasifikasi kata akan digunakan metode hidden markov model (HMM).

Dengan teknologi ini memungkinkan bahasa isyarat dapat dipahami dan dikenali dengan baik oleh orang yang belum mengerti tentang bahasa isyarat.

## 2. Landasan Teori

### 2.1 Bahasa Isyarat

Bahasa isyarat adalah bahasa yang mengutamakan komunikasi gerak, tidak melalui lisan dan digunakan oleh tunarungu.[1] Tunarungu adalah sebutan untuk orang yang kehilangan daya pendengaran. Kehilangan daya pendengaran tersebut dapat terjadi sejak lahir maupun disebabkan oleh faktor lainnya seperti musibah, kecelakaan dan lanjut usia.[2]

Di Indonesia ada dua sistem bahasa isyarat yaitu SIBI (Sistem Isyarat Bahasa Indonesia) dan BISINDO (Bahasa Isyarat Indonesia). Pada penelitian tugas akhir ini digunakan BISINDO karena lebih mudah dipahami oleh para tunarungu dibandingkan SIBI. BISINDO adalah bahasa Ibu komunitas tunarungu di Indonesia. BISINDO berkembang secara alami di kalangan tuli Indonesia dan sudah digunakan turun temurun selama bertahun-tahun.[3]

### 2.2 Sensor Kinect<sup>[4][5]</sup>

Kinect adalah suatu kamera 3D yang mampu menangkap data RGB image serta data depth image. Selain itu pada kinect terdapat microphone array yang memungkinkan untuk merekam suara. Berikut ini adalah gambaran skematik dari struktur hardware kinect : Infrared (IR) Projector (1), the RGB camera (3), and the depth camera (2). An array of four microphones (6), the three-axis accelerometer (5), and the tilt motor (4).

#### 2.2.1 Kinect for Windows SDK<sup>[6]</sup>

Kinect for windows SDK adalah suatu library memungkinkan kita untuk mengembangkan berbagai aplikasi pada platform microsoft dengan menggunakan sensor kinect. Aplikasi yang dapat dibangun menggunakan microsoft SDK kinect adalah aplikasi WPF, winform, XNA dan bahkan berbasis browser yang berjalan sistem operasi windows.

#### 2.2.2 RGB<sup>[5]</sup>

RGB pada kinect hampir mirip dengan warna pada webcam secara umum. Tetapi bedanya RGB pada kinect mempunyai filter IR-cut yang mampu mencapai CMOS. RGB ini memiliki resolusi mencapai 1280x960 piksel dengan 12 fps. Dan bisa mencapai 30 fps pada resolusi 640x480 dengan 8 bit per channel menghasilkan output filter bayer dengan RGGBD. RGB ini juga mampu melakukan penghindaran warna flicker, warna operasi saturasi. Data ini digunakan untuk mendapatkan data dari objek secara detail.

#### 2.2.3 Skeleton<sup>[4]</sup>

Skeletal tracking pada kinect berfungsi untuk mengenali pengguna yang menghadap ke arah kinect. Sebanyak enam orang dapat dideteksi dan satu atau dua orang dapat dilacak pada waktu

bersamaan oleh sensor kinect.

Setiap orang yang berhasil dilacak, NUI akan menghasilkan satu set lengkap data titik posisi yang disebut skeleton. Skeleton berisi 20 posisi, satu untuk setiap joint pada tubuh manusia

### 2.3 Histogram of Oriented Gradients

Metode histogram of oriented gradients (HOG) digunakan sebagai detektor pada gambar yang dideteksi sehingga menghasilkan output berupa vektor.

Langkah pertama dalam metode histogram of oriented gradients adalah mengaplikasikan gambar yang dinormalisasi dengan pemerataan global yang opsional yang dirancang untuk mengurangi pengaruh efek pencahayaan. Yang digunakan adalah normalisasi menggunakan kompresi gamma, yaitu dengan cara menghitung akar kuadrat atau log dari setiap channel warna. Kekuatan tekstur gambar biasanya sebanding dengan pencahayaan permukaan lokal sehingga kompresi ini membantu mengurangi efek local shadow dan variasi iluminasi.

Langkah kedua adalah menghitung gradien dari gambar. Gradien dari suatu gambar dapat diperoleh dengan penyaringan dengan filter 2 dimensi yaitu filter vertikal dan horisontal. Yang pertama dilakukan adalah gambar dikonversi dalam bentuk grayscale untuk menghindari keharusan untuk mempertimbangkan kontribusi intensitas yang berbeda untuk setiap bidang warna (RGB). Metode yang biasa digunakan adalah 1-D centered, dengan matriks  $[-1 \ 0 \ 1]$ . Akan diperoleh nilai  $x$  dan  $y$  yang digunakan untuk menghitung gradient.

Langkah ketiga adalah membagi gambar menjadi beberapa daerah dengan ruang kecil yang disebut sel. Untuk setiap sel diakumulasikan histogram 1-D lokal atau orientasi tepi pada semua piksel dalam sel. Kombinasi cell-level histogram 1-D membentuk representasi histogram orientasi dasar.

Langkah keempat menghitung normalisasi, yang mengambil kelompok sel dan menormalisasi respon kontras secara keseluruhan. Hal ini dilakukan dengan mengakumulasi ukuran histogram dari grup sel yang disebut blok. Hasilnya digunakan untuk menormalisasi setiap sel dalam blok.

Langkah kelima yaitu mengumpulkan descriptor dari semua blok yang merupakan overlapping grid yang mencakup detection window ke dalam feature vector gabungan untuk digunakan dalam classifier.

### 2.4 Hidden Markov Model

Dalam HMM, bagian yang dapat diamati disebut observed state sedangkan bagian yang tersembunyi disebut hidden state. memungkinkan pemodelan sistem yang mengandung observed state dan hidden state yang saling terkait. Pada kasus POS tagging, observed state adalah urutan kata sedangkan hidden state adalah urutan tag. Selain itu, HMM dapat digunakan dalam pengenalan suara, parsing/chunking, ekstraksi informasi, dan peringkasan teks.

HMM terdiri atas lima komponen, yaitu:

- a) Himpunan observed state:  $O = o_1, o_2, \dots, o_N$ .
- b) Himpunan hidden state:  $Q = q_1, q_2, \dots, q_N$
- c) Probabilitas transisi:  $A = a_{01}, a_{02}, \dots, a_{n1} \dots a_{nm}$ ;  $a_{ij}$  adalah probabilitas untuk pindah dari state  $i$  ke state  $j$ .
- d) Probabilitas emisi atau observation likelihood:  $B = b_i(o_t)$ , merupakan probabilitas observasi  $o_t$  dibangkitkan oleh state  $i$ .
- e) State awal dan akhir:  $q_0, q_{end}$ , yang tidak terkait dengan observasi.

## 3. Perancangan Sistem

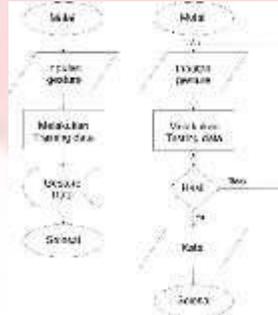
### 3.1 Gambaran Umum Sistem



Gambar 3.1 Gambaran Umum Sistem

Seperti yang terlihat pada gambar 3.1 sistem yang dibangun pada Tugas Akhir ini adalah sistem yang dapat melakukan gesture recognition serta menampilkan data gesture yang dikenal berupa kata jika gesture tersebut sesuai dengan database. Sistem ini berjalan dimulai dari perekaman video menggunakan kinect. Lalu dilakukan proses gesture recognition dengan melakukan proses training dan testing data menggunakan metode hidden markov model. Setelah itu akan menghasilkan prediksi kata sesuai dengan gesture yang direkam.

### 3.2 Diagram Alir Perancangan Umum Sistem



Gambar 3.2 Diagram Alir Sistem

Seperti yang terlihat pada gambar 3.2, sistem pengenalan bahasa isyarat Indonesia ini dibagi ke dalam dua proses untuk mendapatkan hasil yang diinginkan. Proses pertama Diawali dengan masukan rekaman video melalui kinect yang selanjutnya akan di training sebelum menghasilkan database. Sedangkan proses kedua diawali juga dengan masukan rekaman video melalui kinect dilanjutkan dengan testing data untuk dicocokkan data rekaman dengan database yang sudah ada sebelumnya. Jika datanya sesuai akan mendapatkan hasil berupa prediksi kata, jika tidak proses akan diulang dari awal.

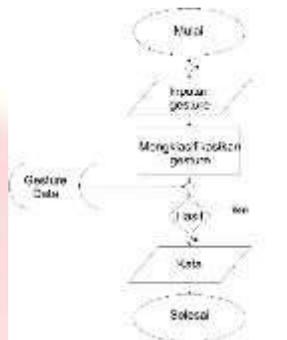
#### 3.2.1 Diagram Alir Training Data



Gambar 3.3 Diagram Alir Training Data

Pada tahap training data dilakukan beberapa tahap sebagai pemrosesan awal yang nantinya akan melakukan proses selanjutnya seperti yang terlihat pada gambar 3.2.2. Tahap-tahap tersebut adalah inputan rekaman video dari kinect yang kemudian diolah ke dalam data skeleton dan RGB dengan melakukan segmentasi bagian tangan menggunakan metode histogram of gradients. Selanjutnya kedua data tersebut digabungkan dan di training menggunakan algoritma baum welch. Sehingga menghasilkan training data model yang akan dijadikan sebagai database.

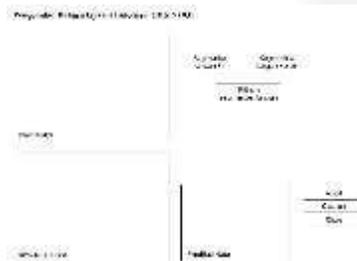
### 3.2.2 Diagram Alir Testing Data



Gambar 3.4 Diagram Alir *Testing Data*

Pada tahap testing data dilakukan beberapa tahap pemrosesan seperti yang terlihat pada gambar 3.2.3. Tahap tahap tersebut adalah inputan rekaman video dari kinect yang dilakukan secara real time yang. Kemudian diambil data gesture yang telah direkam untuk dilanjutkan ke proses klasifikasi dengan menggunakan metode hidden markov model dengan bantuan algoritma forward backward. Jika setelah dilakukan klasifikasi gesture tersebut cocok dengan database yang sudah ada sebelumnya maka akan dilanjutkan ke proses selanjutnya yaitu mendapatkan prediksi kata. Jika tidak, proses diulang dari awal.

### 3.3 Perancangan Antarmuka



Gambar 3.3 Tampilan Perancangan Enkripsi dan Dekripsi

Gambar 3.3 di atas merupakan perancangan antarmuka sistem pengenalan bahasa isyarat Indonesia. Pada RGB image akan diisi oleh tampilan video dalam bentuk RGB. Skeleton canvas diisi oleh video dalam bentuk skeleton. Prediksi kata akan berisi kata yang merupakan hasil dari proses sistem tersebut. Segmentasi tangan kiri dan segmentasi tangan kanan akan diisi oleh video yang telah disegmentasi pada bagian tangan. Pilihan segmentasi tangan merupakan combobox yang berisi pilihan untuk menampilkan segmentasi. Read adalah tombol yang berfungsi untuk membaca video ketika belum di capture. Capture adalah tombol yang berfungsi untuk merekam gesture. Store adalah tombol yang berfungsi untuk menyimpan gesture yang direkam

### 3.4 Skenario Pengujian Nilai Akurasi Sistem

Skenario pengujian nilai akurasi sistem ini dilakukan pada saat proses gesture recognition. Pengujian dilakukan berulang sebanyak 30 sampel per kata dengan jumlah kata adalah 15 kata. Skenario pengujian nilai akurasi ini akan dilakukan berdasarkan dua hal yaitu jarak dan intensitas cahaya. Jarak yang diuji adalah 1 meter, 2 meter dan 3,5 meter. Sedangkan untuk intensitas cahaya yang diuji adalah kondisi terang dengan menggunakan lampu phillip daya 8 watt dan kondisi gelap dengan lampu dimatikan

### 3.5 Skenario Pengujian Waktu Komputasi Sistem

Skenario pengujian waktu komputasi sistem ini dilakukan pada saat proses gesture recognition. Pengujian dilakukan berulang sebanyak 30 sampel per kata dengan jumlah kata adalah

15 kata. Dari pengujian ini akan diketahui berapa kecepatan sistem dan seberapa efisien sistem yang dibuat dilihat dari sisi waktu yang diperlukan dalam mengenali masukan yang diberikan sehingga akan menghasilkan output yang benar.

#### 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

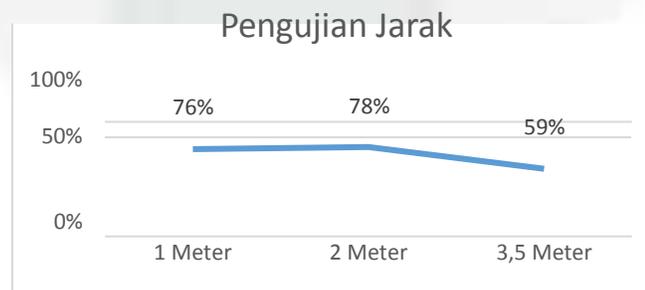
Spesifikasi perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

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

#### 4.3 Pengujian Pengujian Akurasi Sistem Berdasarkan Jarak

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

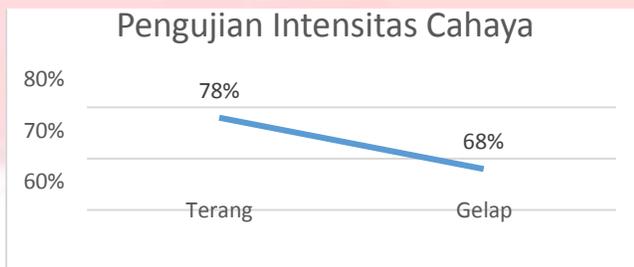
- a) Akurasi sistem tertinggi terdapat pada jarak 2 meter dengan nilai rata-rata akurasi adalah 78%.
- b) Akurasi sistem terendah terdapat pada jarak 3,5 meter dngan nilai rata-rata akurasi adalah 56%.
- c) Jarak optimal sistem mengenali objek adalah 2 meter, yaitu tidak terlalu dekat (1 meter) dengan sensor dan tidak terlalu jauh (3,5 Meter) dengan sensor.



Gambar. 4.1 Diagram Hasil Pengujian Jarak

#### 4.4 Pengujian Akurasi Sistem Berdasarkan Intensitas Cahaya

Dari data hasil pengujian tersebut dapat diketahui bahwa akurasi sistem lebih tinggi pada saat kondisi terang (menggunakan lampu phillip 8 watt) yaitu sebesar 78% dibandingkan kondisi gelap (lampu dimatikan) yang hanya 64%. Data ini dapat dilihat pada diagram di bawah ini.



Gambar. 4.2 Diagram Hasil Pengujian Intensitas Cahaya

#### 4.5 Pengujian Waktu Komputasi Sistem

Pengujian waktu komputasi sistem ini dilakukan sebanyak 30 sampel per kata, dengan jumlah kata sebanyak 15 kata. Dari data hasil pengujian waktu komputasi sistem dapat diketahui bahwa waktu komputasi sistem ini adalah 1.473311 detik.

Kata	Sampel	Waktu
Ampun	30	1.796667
Apa	30	0.96
Berapa	30	0.8
Halo	30	1.2
Kanan	30	1.728333
Kiri	30	1.663667
Maaf	30	1.119
Makan	30	1.892667
Minum	30	1.23
Permisi	30	2.436667
Sampai	30	2.303667
Selamat	30	1.283
Siapa	30	1.113333
Terimakasih	30	1.54
Tolong	30	1.032667
<b>Total</b>	<b>450</b>	<b>1.473311</b>

Tabel. 4.1 Data hasil pengujian waktu komputasi

### 5. Kesimpulan dan Saran

#### 5.1 Kesimpulan

Dari hasil pengujian dan analisis simulasi dalam Tugas Akhir ini, dihasilkan beberapa kesimpulan sebagai berikut:

- 1) Jarak optimal sistem mengenali objek adalah 2 meter, yaitu tidak terlalu dekat dengan sensor dan tidak terlalu jauh dengan sensor kinect dan intensitas yang menghasilkan akurasi sistem yang optimal adalah pada kondisi terang.
- 2) Nilai akurasi tertinggi didapatkan pada jarak objek 2 meter dari kinect dan dalam kondisi terang yaitu sebesar 78%.
- 3) Waktu komputasi yang dibutuhkan oleh sistem adalah 1.473311 detik.
- 4) Jika *gesture* hampir mirip satu sama lain atau saling bersinggungan, maka sering terjadi salah pengenalan terhadap *gesture* tersebut, seperti makan dengan tolong, minum dengan siapa, dan maaf dengan halo.

## 5.2 Saran

Saran untuk penelitian selanjutnya adalah :

- 1) Meningkatkan akurasi dari masukan RGB dan skeleton kinect menggunakan metode HMM untuk sistem pengenalan bahasa isyarat Indonesia.
- 2) Menggunakan masukan data kinect yang lain pada sistem pengenalan Bahasa isyarat Indonesia.
- 3) Menggunakan metode klasifikasi yang lain untuk sistem pengenalan Bahasa isyarat Indonesia.
- 4) Mencoba lebih banyak kata untuk dilakukan pengujian.
- 5) Menggunakan komputer dengan spesifikasi yang lebih bagus dari yang digunakan pada tugas akhir ini karena cukup mempengaruhi nilai waktu komputasi.
- 6) Menggunakan kinect v2 karena terdapat fitur-fitur baru yang tidak terdapat pada kinect v1.

## DAFTAR PUSTAKA

- [1]. Quan, Yang. 2010. Chinese Sign Language Recognition Based On Video Sequence Appearance Modeling. 2010 5th IEEE Conference on Industrial Electronics and Applications.
- [2].GERKATIN, 2014. Profil Organisasi. <http://gerkatin.com/profileprofil.NetLErp>. Diakses 05 Maret 2015, Pukul 22.54 WIB.
- [3].GERKATIN, 2014. Lebih Lanjut Tentang BISINDO. <http://gerkatin.com/detailberita-140-lebih-lanjut-tentang-bisindo.html>. Diakses 05 Maret 2015, Pukul 22.58 WIB.
- [4].Catuhe, David. Programming with Kinect for Windows Software Development Kit.
- [5].Giorio, Clemente. Fascinari, Massimo. Kinect in Motion – Audio and Visual Tracking by Example.
- [6].Jana, Abhijit. Kinect for Windows Programming Guide.
- [7].Prasetyo, Muhammad Eko Budi. 2010. Teori Dasar Hidden Markov Model. Program Studi Sistem dan Teknologi Informasi. Institut Teknologi Bandung.
- [8].Perdananti, Pramesti Festa. 2015. Implementasi Algoritma Ekstraksi Ciri Independent Component Analysis (Ica) Untuk Pengenalan Wajah Pada Video Pengawasan Ruang Terbuka. Fakultas Teknik Elektro. Telkom University.
- [9].Wibisono, Yudi. 2008. Penggunaan Hidden Markov Model untuk Kompresi Kalimat. Program Studi Informatika. Institut Teknologi Bandung.