

Integrasi Ekstraksi Fitur Statis dan Dinamis pada Gerakan Tangan Menggunakan Kinect 2.0 untuk Mengenali Bahasa Isyarat Indonesia

Ignatius Benedict¹, Wijayanti Nurul Khotimah², dan Nanik Suciati³

Departemen Informatika, Fakultas Teknologi Informasi dan Komunikasi, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: inyas.benedict@gmail.com¹, wijayanti@if.its.ac.id², nanik@if.its.ac.id³

Abstrak—Bahasa isyarat merupakan media komunikasi yang digunakan oleh penderita tunarungu untuk berkomunikasi dengan sekitar. Bahasa isyarat yang digunakan mengacu pada Sistem Bahasa Isyarat Indonesia (SIBI). Bahasa isyarat dibagi menjadi dua kategori, yaitu bahasa isyarat statis dan dinamis. Pada penelitian sebelumnya, pengenalan bahasa isyarat menggunakan teknologi Kinect sudah berhasil dibuat. Namun dalam prosesnya, pengaplikasiannya dilakukan secara terpisah antara statis dan dinamis.

Pada penelitian ini dilakukan integrasi pengenalan bahasa isyarat statis dan dinamis. Fitur-fitur yang digunakan adalah fitur statis dan fitur dinamis yang digabungkan untuk mengklasifikasikan bahasa isyarat statis dan dinamis dari gerakan yang dilakukan oleh pengguna. Hasil dari pengujian aplikasi ini memiliki persentase nilai akurasi 83,67 persen.

Kata Kunci—Bahsa Isyarat, Fitur Dinamis, Fitur Statis, Kinect 2, Sistem Isyarat Bahasa Indonesia.

I. PENDAHULUAN

Para penyandang tunarungu memiliki keterbatasan dalam berkomunikasi menggunakan bahasa sehari-hari. Bahasa isyarat merupakan bahasa yang digunakan oleh penyandang tunarungu. Bahasa ini menggunakan gerak visual tubuh untuk menyampaikan maksud kepada lawan bicaranya. Sistem yang digunakan di Indonesia mengacu pada Sistem Isyarat Bahasa Indonesia (SIBI).

Bahasa isyarat bagi penyandang tunarungu merupakan bahasa umum, tetapi bagi orang normal merupakan bahasa asing [1]. Hal ini dapat mengakibatkan kesenjangan komunikasi antara orang normal dengan penyandang tunarungu. Untuk mempermudah proses komunikasi, dibutuhkan penerjemah antara penyandang tunarungu dengan orang normal.

Sebelumnya sudah ada Tugas Akhir yang dibuat oleh Yohanes Aditya Sutanto dan Yahya Eka Nugyasa tentang pengenalan bahasa isyarat Indonesia menggunakan teknologi Kinect. Dalam Tugas Akhir yang dibangun oleh Yohanes menggunakan teknologi Kinect 1.0, sudah dapat mendeteksi bahasa isyarat statis [2]. Dalam Tugas Akhir yang dibangun oleh Yahya menggunakan teknologi Kinect 2.0, sudah dapat mendeteksi bahasa isyarat dinamis [3]. Oleh karena itu munculah ide untuk mengintegrasikan pendeteksi bahasa isyarat statis dan bahasa isyarat dinamis. Bahasa isyarat yang

digunakan mengacu pada SIBI.

II. TINJAUAN PUSTAKA

A. Bahasa Isyarat

Bahasa isyarat adalah bahasa yang digunakan untuk berkomunikasi oleh penyandang tunarungu. Bahasa isyarat yang sering digunakan di Indonesia berdasarkan pada SIBI.

Berdasarkan sifat geraknya, bahasa isyarat dibagi menjadi 2 kelompok yaitu gerak bahasa isyarat statis dan gerak bahasa isyarat dinamis.

B. Kinect 2.0

Generasi kedua dari Kinect yang dirilis oleh Microsoft pada tahun 2014 adalah versi terbaru Kinect dari yang pertama kali dikeluarkan pada tahun 2010. Perangkat Kinect 2.0 seperti yang terlihat pada Gambar 1, terdapat 3 lensa yaitu kamera RGB yang digunakan untuk menangkap spektrum warna, *infrared emitters* yang memproyeksikan spektrum inframerah dan sensor kedalaman yang menghasilkan gambar mendalam dari seseorang atau objek dengan menganalisis informasi inframerah. Dan sebuah *microphone array* yang dapat menentukan lokasi timbulnya suara. Alhasil, ada enam sumber data yang dihasilkan, termasuk warna, inframerah, kedalaman, indeks tubuh, tubuh, dan suara [4].



Gambar 1 Kinect 2.0

C. Decision Tree

Pohon Keputusan atau dikenal dengan Decision Tree adalah salah satu metode klasifikasi yang menggunakan representasi suatu struktur pohon yang berisi alternatif-alternatif untuk pemecahan suatu masalah. Pohon ini juga menunjukkan faktor-faktor yang mempengaruhi hasil alternatif dari

keputusan tersebut disertai dengan estimasi hasil akhir bila kita mengambil keputusan tersebut. Peranan pohon keputusan ini adalah sebagai *Decision Support Tool* untuk membantu manusia dalam mengambil suatu keputusan.

Decision Tree merupakan struktur hierarki untuk pembelajaran *supervised*. Proses dari Decision Tree dimulai dari *root node* hingga *leaf node* yang dilakukan secara rekursif. Dimana setiap percabangan menyatakan suatu kondisi yang harus dipenuhi dan pada setiap ujung pohon menyatakan kelas dari suatu data. Pada Decision Tree terdiri dari tiga bagian yaitu:

1. *Root Node*

Node ini merupakan *node* yang terletak paling atas dari suatu pohon.

2. *Internal Node*

Node ini merupakan *node* percabangan, hanya terdapat satu *input* serta mempunyai minimal dua *output*.

3. *Leaf Node*

Node ini merupakan *node* akhir, hanya memiliki satu *input* dan tidak memiliki *output*.

D. Ekstraksi Fitur

Fitur yang digunakan dalam Tugas Akhir ini terdiri dari dua kelompok, yaitu fitur statis dan fitur dinamis. Fitur statis akan digunakan untuk mengklasifikasikan gerak bahasa isyarat statis. Fitur dinamis akan digunakan untuk mengklasifikasikan gerak bahasa isyarat dinamis. *Skeleton joints* yang diambil oleh perangkat lunak ini sebanyak 11 *skeleton joints*. *Skeleton joints* yang dimaksud yaitu:

1. Leher (*N*)
2. Bahu tulang belakang (*SS*)
3. Tulang belakang tengah (*SM*)
4. Bahu tangan kanan (*SR*)
5. Siku tangan kanan (*ER*)
6. Pergelangan tangan kanan (*WR*)
7. Telapak tangan kanan (*HR*)
8. Bahu tangan kiri (*SL*)
9. Siku tangan kiri (*EL*)
10. Pergelangan tangan kiri (*WL*)
11. Telapak tangan kiri (*HL*)

Fitur statis adalah fitur yang bersifat statis yang didapatkan dari hasil kalkulasi beberapa *skeleton joints* yang dideteksi oleh Kinect 2.0. Fitur yang digunakan terdiri dari 3 bagian yaitu *vector2*, *angel*, *distance*. Dari hasil perhitungan *skeleton joints*, didapatkan 21 buah fitur yang akan digunakan [5]. Fitur-fitur yang digunakan untuk fitur statis dapat dilihat pada Tabel 1.

Fitur dinamis merupakan fitur gerak atau fitur yang bersifat dinamis yang didapatkan dari deteksi Kinect 2.0 terhadap tangan. Fitur dinamis yang digunakan pada metode ini adalah fitur untuk mengolah gerakan tangan (*hand gesture*).

Terdapat 40 *frame* yang akan diproses. Setiap *frame* akan menghasilkan koordinat gerak dinamis yang diproyeksikan ke dalam bidang XOY yang merupakan prinsip dalam bidang. Kemudian dari hasil setiap data tersebut, akan dicatat orientasi sudut mutlak $\alpha_t \in (0, 360^\circ)$ yang dapat dilihat pada persamaan (1) (2) dan (3) [4].

TABEL 1
FITUR YANG DIGUNAKAN PADA FITUR STATIS

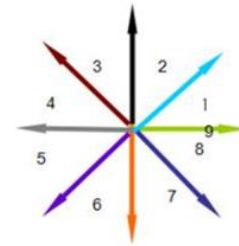
Vektor2 (x,y)	Sudut (float)	Jarak (float)
SR → ER	$\angle SS - SR - ER$	HR – HL
ER → WR	$\angle SR - ER - WR$	
WR → HR	$\angle ER - WR - HR$	
SL → EL	$\angle SS - SL - EL$	
EL → WL	$\angle SL - EL - WL$	
WL → HL	$\angle EL - WL - HL$	
HR → HL		

$$\Delta y = y_t - y_{t-1} \quad (1)$$

$$\Delta x = x_t - x_{t-1} \quad (2)$$

$$\alpha_t = \begin{cases} \arctan\left(\frac{\Delta y}{\Delta x}\right) * \left(\frac{180}{\pi}\right) + 180, & \Delta x < 0 \\ \arctan\left(\frac{\Delta y}{\Delta x}\right) * \left(\frac{180}{\pi}\right) + 360, & \Delta y < 0 \\ \arctan\left(\frac{\Delta y}{\Delta x}\right) * \left(\frac{180}{\pi}\right), & \Delta x > 0, \Delta y \geq 0 \end{cases} \quad (3)$$

Orientasi sudut mutlak tersebut kemudian diterjemahkan menjadi data kuantitasi hasil ekstraksi fitur dinamis seperti yang dapat dilihat pada Gambar 2.



Gambar 2 Pengelompokan Kuantitasi Data

Hasil dari α_t tersebut kemudian diterjemahkan menjadi data kuantitasi hasil ekstraksi fitur dinamis yang dapat dijelaskan pada Tabel 2.

III. ANALISIS DAN PERANCANGAN

TABEL 2
DATA KUANTITASI HASIL FITUR DINAMIS

Rule No.	α_t	Nilai Fitur
1	$\alpha_t = 0$	9
2	$\alpha_t > 314$	8
3	$\alpha_t > 269$	7
4	$\alpha_t > 224$	6
5	$\alpha_t > 179$	5
6	$\alpha_t > 134$	4
7	$\alpha_t > 89$	3
8	$\alpha_t > 44$	2
9	$\alpha_t > 0$	1

A. Deskripsi Umum Perangkat Lunak

Penelitian yang dikembangkan ini akan menghasilkan keluaran berupa perangkat lunak pengenalan bahasa isyarat statis dan dinamis menggunakan bantuan Kinect SDK dan dibangun dengan memasukkan melalui Kinect 2.0. Hal ini diperlukan untuk memeriksa secara langsung apakah pengenalan bahasa isyarat ini dapat mengeluarkan hasil sesuai yang diharapkan.

B. Rancangan Perangkat Lunak

Pada proses perancangan perangkat lunak, akan dijelaskan mengenai sistem yang ada di dalam perangkat lunak yang akan dibangun agar dapat memenuhi fungsional yang ada.

1) Proses Ekstraksi Fitur

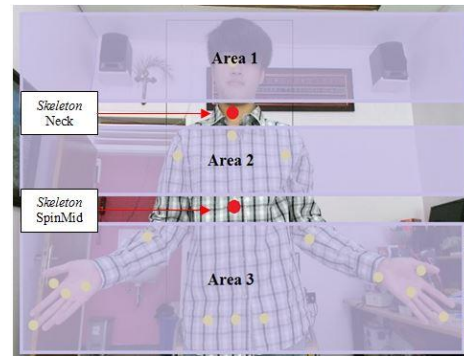
Proses ekstraksi fitur perangkat lunak akan mengambil total 40 *frame* untuk dikalkulasikan menjadi fitur-fitur yang dibutuhkan. Jumlah data tersebut ditentukan berdasarkan durasi gerakan bahasa isyarat dinamis yang paling lama dilakukan diantara seluruh *sample* gerakan yang sudah ditentukan sebelumnya. Apabila gerakan pengguna terlalu cepat atau terlalu lambat, maka pengambilan data harus diulang kembali guna mendapatkan hasil yang maksimal.

Saat perekaman *skeleton joints* oleh Kinect 2.0 yang kemudian diolah menjadi fitur data. Masing masing *skeleton joints* diinterpretasikan dalam bentuk *x* dan *y*. Data dari ke-11 *skeleton joints* yang sudah ditentukan dapat dihasilkan 59 fitur yang digunakan untuk menentukan gerakan isyarat yang dilakukan oleh pengguna. Fitur-fitur itu terdiri dari 21 fitur statis yang dapat dilihat pada Tabel 1 dan 38 fitur dinamis yang dapat dilihat pada Tabel 3.

TABEL 3
FITUR DINAMIS YANG DIGUNAKAN

Fitur	Indeks Fitur
Kuantitasi Tangan Kiri	Fitur ke-1 s.d. 18
Kuantitasi Tangan Kanan	Fitur ke-19 s.d. 36
Posisi Tangan Kiri	Fitur ke-37
Posisi Tangan Kanan	Fitur ke-38

Untuk mendapatkan hasil kuantisasi fitur dinamis, digunakan persamaan (1) dan (2) untuk mendapatkan koordinat perpindahan *skeleton joints* HR dan HL. Setelah data perpindahan didapatkan, digunakan persamaan (3) untuk mendapatkan sudut mutlak yang akan menjadi data kuantitasi dari rule yang telah ditetapkan pada Tabel 2 menjadi fitur kuantitasi tangan kanan dan kiri. Sedangkan data koordinat *skeleton joints* N dan SM digunakan untuk menentukan posisi gerakan tangan pengguna. Posisi gerakan tangan dibagi menjadi 3 (tiga) area seperti pada Gambar 3. Ke-3 area tersebut adalah Area 1 (kepala), Area 2 (dada), dan Area 3 (perut).



Gambar 3 Posisi Gerak Tangan pada Tubuh Mengacu Skeleton Joints Leher (N) dan Tulang Belakang Tengah (SM)

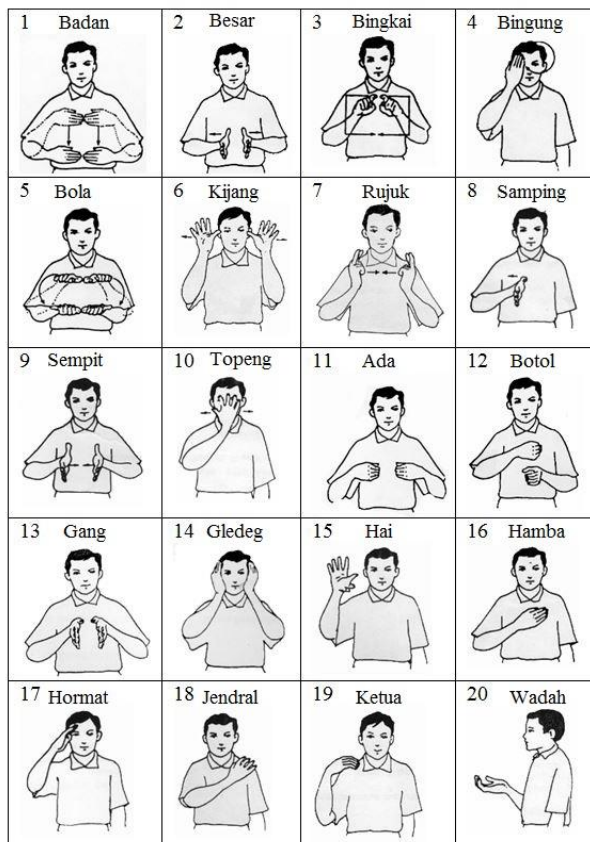
Dalam proses ekstraksi fitur dinamis, perangkat terlebih dahulu menentukan Δx dan Δy yang didapatkan dari hasil selisih masing-masing koordinat *x* dan *y* antara data ke-*n* dan *n+1 skeleton joints* HR dan HL. Nilai hasil dari kedua variabel tersebut didapatkan orientasi sudut mutlak (α_i) yang kemudian dijadikan data kuantitasi dengan *rule* yang sudah ditentukan sehingga menjadi 80 fitur yang dihasilkan oleh tangan kanan dan tangan kiri. Setelah fitur-fitur tersebut didapatkan, perangkat lunak kemudian menyeleksi 10% data awal sebagai tahap persiapan. Setelah itu data fitur yang disimpan hanyalah data pada frame genap, diharapkan agar mendapatkan hasil yang lebih akurat [4]. Posisi gerak tangan didapatkan dari posisi tangan pada saat frame ke-20, kemudian disimpan sebagai posisi gerak tangan kanan dan kiri.

IV. PENGUJIAN DAN EVALUASI

Pengujian dilakukan pada ke-20 bahasa isyarat yang sudah ditetapkan di awal. Ke-20 gerakan bahasa isyarat tersebut dapat dilihat pada Gambar 4. Skenario pengujian yang dilakukan menjadi dua bagian, yaitu menggunakan model *tree* A dan model *tree* B. Pengujian dengan model *tree* A dilakukan oleh penulis. Model *tree* A adalah model Decision Tree yang dibangun secara langsung menggunakan data statis dan dinamis yang telah ditentukan sebelumnya. Hasil dari uji coba tersebut dapat dilihat pada Tabel 4 bagian uji coba 1 (UC1). Kemudian pada pengujian berikutnya menggunakan model *tree* B yang dilakukan oleh penulis dan dua orang pengguna dengan karakteristik badan yang berbeda. Model *tree* B adalah model Decision Tree yang dibangun dengan cara mengelompokkan terlebih dahulu gerakan isyarat statis dan dinamis, lalu dibangun *tree* utama yang menggabungkan kedua *tree* tersebut. Hasil dari uji coba tersebut dapat dilihat pada Tabel 4 bagian uji coba 2 – 4 (UC2 – UC 4).

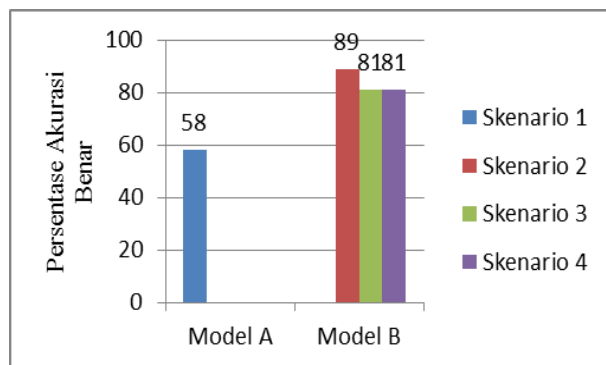
Hasil pengujian disusun dalam bentuk analisis secara keseluruhan sebagai berikut:

1. Klasifikasi fitur statis dan dinamis menggunakan model *tree* B (uji coba 2-4) di dalam perangkat lunak yang dibuat mempunyai akurasi rata-rata 83.67%. Grafik hasil akurasi dapat dilihat pada Gambar 5.



Gambar 4 Dua Puluh Bahasa Isyarat yang Digunakan

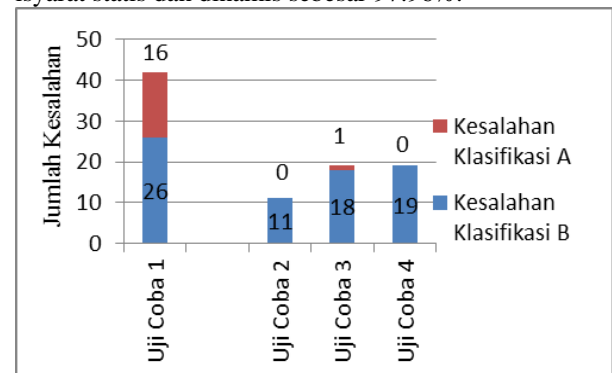
2. Pada percobaan skenario 1 dapat dilihat hasil dari klasifikasi kurang memuaskan. Dari hal ini dilakukan peningkatan dengan cara mengelompokkan data *training* terlebih dahulu. Dapat dilihat dari hasil skenario 2 bahwa hasil yang didapatkan lebih optimal.



Gambar 5 Akurasi Hasil Uji Coba

3. Pada hasil uji coba menggunakan model *tree* B dapat dilihat hasil klasifikasi penentuan antara gerak statis dan dinamis telah lebih baik daripada model *tree* A. Hasil dari percobaan ini dapat dilihat pada Gambar 6. Kesalahan Klasifikasi A adalah kesalahan klasifikasi bahasa isyarat statis menjadi bahasa isyarat dinamis atau sebaliknya. Kesalahan Klasifikasi B adalah kesalahan klasifikasi bahasa isyarat statis dengan bahasa isyarat statis lainnya atau bahasa isyarat dinamis dengan bahasa isyarat dinamis lainnya. Dari data berikut dapat

disimpulkan hasil klasifikasi model *tree* B (uji coba 2–4) memiliki tingkat akurasi penentuan antara gerakan bahasa isyarat statis dan dinamis sebesar 97.96%.



Gambar 6 Jumlah Kesalahan Pengelompokan Klasifikasi Statis dan Dinamis

4. Ada beberapa gerakan bahasa isyarat yang mempunyai kemiripan gerak dan posisi sehingga mengakibatkan salah dalam mengklasifikasikan data
5. Penggunaan klasifikasi decision tree yang mengakibatkan pengklasifikasian yang kurang akurat. Penggunaan decision tree dapat mempengaruhi hasil klasifikasi karena metode decision tree memungkinkan sebuah fitur dianggap lebih penting dibandingkan fitur lainnya. Hal ini dapat mempengaruhi tingkat akurasi dari perangkat lunak ini.

TABEL 4
AKURASI HASIL PENGUJIAN

Bahasa Isyarat	Model A		Model B	
	UC1 (%)	UC2 (%)	UC3 (%)	UC4 (%)
Badan	20	80	80	80
Besar	40	80	10	10
Bingkai	40	80	60	80
Bingung	60	60	60	40
Bola	100	10	80	40
Kijang	100	10	80	10
Rujuk	0	60	10	10
Samping	60	80	10	10
Sempit	100	10	10	80
Topeng	100	80	60	60
Ada	20	80	40	60
Botol	80	10	80	10
Gang	20	10	80	80
Gledag	80	10	10	10
Hai	60	10	10	10
Hamba	60	10	10	10
Hormat	60	10	80	80
Jendral	80	10	10	10
Ketua	40	10	60	10
Wadah	40	80	60	20
Rata-rata	58	89	81	81

V. KESIMPULAN DAN SARAN

Dalam proses penelitian selama perancangan, implementasi, dan proses pengujian aplikasi yang dilakukan, dapat diambil kesimpulan sebagai berikut:

1. Fitur-fitur yang digunakan berjumlah 59 fitur, terdiri dari 21 fitur statis dan 38 fitur dinamis. Fitur statis terdiri dari 7 fitur vektor, 6 fitur sudut, dan 1 fitur jarak. Fitur dinamis terdiri dari 18 fitur gerak tangan kiri, 18 fitur gerak tangan kanan, dan 2 fitur posisi tangan kanan dan kiri. Seluruh fitur tersebut didapatkan dari 11 *skeleton joints*.
2. Ketepatan posisi berdiri dan melakukan gerakan bahasa isyarat yang dilakukan ketika pengambilan data *training* dan *testing* sangat berpengaruh terhadap akurasi klasifikasi.
3. Perbedaan karakteristik tinggi badan berpengaruh terhadap penentuan koordinat *skeleton joints* dimana perbedaan tersebut mempengaruhi identifikasi posisi gerakan yang dilakukan oleh perangkat lunak.
4. Perangkat lunak yang dibangun pada Tugas Akhir ini dapat menerjemahkan bahasa isyarat pokok dengan akurasi rata-rata 83.67% dan ketepatan akurasi klasifikasi statis dan dinamis sebesar 97.96%.

Berikut adalah saran untuk penelitian kedepannya :

1. Memperbanyak data *training* guna meningkatkan akurasi dari klasifikasi perangkat lunak ini.
2. Identifikasi *skeleton joints* tidak hanya posisi tangan saja. Karena ada beberapa gerakan bahasa isyarat yang memiliki posisi gerak yang sama tetapi memiliki bentuk atau model tangan yang berbeda.
3. Melakukan normalisasi data koordinat masing-masing *skeleton joints* sebelum melakukan proses ekstraksi fitur guna mendapatkan hasil yang lebih akurat.
4. Menggunakan *classifier* yang lebih baik dan tepat dapat meningkatkan nilai akurasi dari perangkat lunak ini.
5. Fitur yang digunakan sebaiknya ditambahkan menjadi vektor 3D.

DAFTAR PUSTAKA

- [1] N. Sugianto dan F. Samopa, "Analisa Manfaat dan Penerimaan Terhadap Implementasi Bahasa Isyarat Indonesia pada Latar Belakang Komplek Menggunakan Kinect dan Jaringan Syaraf Tiruan (Studi Kasus SLB Karya Mulia 1)," JUISI, vol. 1, pp. 56-72, 2015.
- [2] Y. A. Sutanto, "Rancang Bangun Modul Pengenalan Bahasa Isyarat Indonesia Menggunakan Teknologi Kinect dan Metode Back Propagation Genetic Algorithm Neural Network," Departemen Teknik Informatika Institut Teknologi Sepuluh Nopember, Surabaya, 2016.
- [3] Y. E. Nugyasa, "Ekstaksi Fitur Dinamis Pada Gerakan Tangan Menggunakan Kinect 2.0 untuk Mengenali Bahasa Isyarat Indonesia," Institut Teknologi Sepuluh Nopember, Surabaya, 2017.
- [4] Y. Chen, B. Luo, Y.-L. Chen, G. Liang dan X. Wu, "A Real-time Dynamic Hand Gesture Recognition System Using Kinect Sensor," dalam *IEEE Conference on Robotics and Biomimetics*, Zhuhai, 2015.

- [5] C. Sun, T. Zhang, B.-K. Bao dan C. Xu, "Diskriminatif Eksemplar Koding untuk Bahasa Isyarat," dalam *IEEE TRANSACTIONS ON CYBERNETICS*, 2013.