

# Pengenalan Bahasa Isyarat Indonesia dengan Metode *Dynamic Time Warping* (DTW) Menggunakan Kinect 2.0

Tiara Anggita, Wijayanti Nurul Khotimah, dan Nanik Suciati

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

*e-mail*: nanik@if.its.ac.id

**Abstrak**—Bahasa isyarat merupakan media bagi penderita tuna rungu dan tuna wicara untuk saling berkomunikasi. Perkembangan teknologi untuk menerjemahkan bahasa isyarat Indonesia secara real-time bisa bermanfaat bagi komunitas tuna rungu. Pada penelitian sebelumnya, pengenalan bahasa isyarat menggunakan teknologi Kinect sudah berhasil dibuat. Namun dalam prosesnya, fitur bahasa isyarat yang digunakan hanya bahasa isyarat yang bersifat statis saja. Selain itu, terdapat pula penelitian sebelumnya yang menggunakan ekstraksi fitur untuk mengenali bahasa isyarat dinamis saja. Pada penelitian ini, dilakukan bahasa isyarat yang bersifat statis dan dinamis secara dengan metode *Dynamic Time Warping* (DTW), yang mampu mengenali secara *real-time* bahasa isyarat Indonesia dari 20 kata dan menerjemahkannya dalam gambar dan teks. Hasil pengujian menunjukkan bahwa persentase nilai akurasi rata-rata yang dihasilkan adalah 93%.

**Kata Kunci**—Bahasa Isyarat Indonesia, *Dynamic Time Warping*, Kinect, Tunarungu

## I. PENDAHULUAN

**K**EKURANGAN pada pendengaran sering berdampak pada kemampuan verbal pada orang dengan gangguan pendengaran, sehingga mereka menggunakan bahasa isyarat dan bahasa tubuh untuk berkomunikasi. Bahasa isyarat merupakan media bagi para penderita tuna rungu dan tuna wicara untuk berkomunikasi dengan sekitarnya. Penderita tuna rungu dan tuna wicara mengalami kesulitan dalam berkomunikasi dengan orang normal. Hal ini dikarenakan ketidakpahaman orang normal dengan system isyarat ini [1]. Dengan demikian, proses pertukaran informasi sulit terjadi. Untuk itu dibutuhkan sistem untuk menerjemahkan bahasa isyarat ke dalam Bahasa Indonesia agar dapat tercipta komunikasi yang lebih baik.

Bahasa isyarat adalah bahasa yang mengutamakan komunikasi manual, bahasa tubuh, dan gerak bibir untuk berkomunikasi. Penyandang tuna rungu adalah kelompok utama yang menggunakan bahasa ini, biasanya mengkombinasikan bentuk tangan, orientasi dan gerak tangan, lengan dan tubuh, serta ekspresi wajah untuk mengungkapkan pikiran mereka [2]. Bahasa Isyarat sangat dipengaruhi oleh latar belakang budaya dan kebiasaan dimana orang tersebut tinggal dan berasal. Di Indonesia, bahasa isyarat diterapkan dalam dua bentuk, yaitu Sistem Isyarat Bahasa Indonesia (SIBI) dan Bahasa Isyarat Indonesia (BISINDO). Bahasa Isyarat Indonesia adalah sistem komunikasi yang praktis dan efektif untuk

penyandang tuna rungu Indonesia yang telah dikembangkan oleh kaum tuna rungu, sedangkan Sistem Bahasa Isyarat Indonesia (SIBI) adalah sistem hasil rekayasa dan ciptaan dari orang normal untuk berkomunikasi dengan penyandang difabel tuna rungu dan bukan berasal dari penyandang difabel tuna rungu. Dalam kehidupan sehari-hari, penderita tuna rungu wicara berkomunikasi dengan bahasa isyarat yang mengacu pada SIBI [1].

Bahasa Isyarat tidak memiliki lingkup pengguna yang besar seperti bahasa lisan, sehingga tidak banyak orang yang dapat mengenali atau mengerti bahasa isyarat yang disampaikan lawan bicaranya. Hal ini menimbulkan kebutuhan alat bantu untuk mempelajari bahasa isyarat dengan mudah.

Pada tahun 2010, Microsoft meluncurkan teknologi baru berupa perangkat keras sensor Kinect. Pada awalnya, teknologi tersebut ditujukan sebagai konsol Xbox 360 sehingga pemain dapat menggunakan gerakan tubuhnya sebagai pengendali permainan. Teknologi sensor Kinect ini dapat dikembangkan lebih lanjut agar dapat digunakan untuk mengenali bahasa isyarat [3].

Penelitian mengenai pengenalan bahasa isyarat Indonesia masih terbatas dan masih membutuhkan pengembangan. Sebelumnya, sudah ada penelitian yang dibuat oleh Khotimah tentang pengenalan bahasa isyarat Indonesia menggunakan Kinect 1.0. Dalam penelitian tersebut, bahasa yang dapat dideteksi, yaitu Alquran, Bentuk, Gang, Hai, Hamba, Hormat, Ketua, dan Wadah. Bahasa yang dapat dikenali dalam penelitian ini hanya berupa gerakan statis dan kosa katanya pun masih terbatas, sehingga perlu ditelusuri lebih lanjut Hasil rata-rata akurasi yang didapat pada penelitian tersebut cukup tinggi, yaitu 96% [4].

Selain itu, penelitian sebelumnya mengenai pengenalan bahasa isyarat Indonesia juga pernah dilakukan oleh Yahya Eka. Penelitian tersebut membahas tentang bagaimana mengekstraksi fitur untuk mengenali bahasa isyarat dinamis. Sehingga, fitur bahasa isyarat yang dapat dikenali pada penelitian tersebut masih terbatas pada bahasa isyarat yang bersifat dinamis saja. Hasil rata-rata akurasi yang didapatkan pada penelitian tersebut, yaitu 82,9%.

Dalam pengenalan bahasa isyarat, terdapat beberapa metode yang digunakan untuk menangani gerakan dinamis, yaitu *Artificial Neural Networks*, *Hidden Markov Models*, dan *Dynamic Time Warping* (DTW). DTW diperkenalkan untuk komunitas *data mining* oleh Berndt and Clifford [5]. Metode ini digunakan untuk menghitung kesamaan antara dua data deret

waktu (*time series*) yang dapat berbeda dalam waktu dan kecepatan [3].

Dalam Tugas Akhir ini, akan dikembangkan suatu program aplikasi untuk pembelajaran bahasa isyarat dengan menggunakan sensor Kinect untuk mendeteksi gerakan tangan, kemudian membandingkan pola gerakan tangan tersebut dengan data-data referensi menggunakan metode *Dynamic Time Warping* (DTW).

## II. TINJAUAN PUSTAKA

### A. Bahasa Isyarat

Bahasa isyarat adalah sarana bagi penderita tuna rungu dan tuna wicara untuk berkomunikasi bagi sekitarnya. Bahasa isyarat dikembangkan dan memiliki karakteristik sendiri di berbagai negara. Di Indonesia, bahasa isyarat yang digunakan berdasarkan pada SIBI [6].

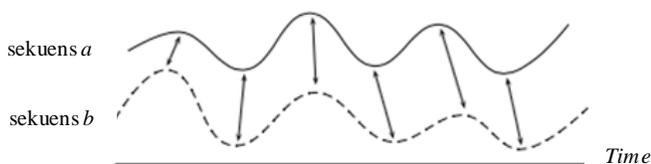
### B. Euclidean Distance

*Euclidean Distance* adalah perhitungan jarak dari dua buah titik dalam *Euclidean space*. Perhitungan ini biasanya diterapkan pada dua dimensi dan tiga dimensi. Pada sistem tiga dimensi, *Euclidean Distance* dihitung berdasarkan persamaan (1), dimana  $p$  dan  $q$  merupakan titik yang ingin dihitung dari masing-masing koordinat  $x$ ,  $y$ , dan  $z$ .

$$d = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2} \quad (1)$$

### C. Dynamic Time Warping (DTW)

*Dynamic Time Warping* (DTW) merupakan teknik yang terkenal untuk mencari penyesuaian yang optimal antara dua urutan yang diberikan (*time-dependent*) di bawah batasan tertentu. Penyesuaian antara dua urutan dapat dilihat pada Gambar 1 **Error! Reference source not found.** Algoritma ini juga disebut sebagai *non-linear sequence alignment*, sehingga algoritma ini lebih realistis untuk digunakan dalam mengukur kemiripan suatu pola (*pattern/template matching*) ketimbang hanya menggunakan algoritma pengukuran linier seperti *Euclidean Distance*, *Manhattan*, *Canberra*, *Mexican Hat*, dan lain-lain. Algoritma ini secara luas digunakan untuk aplikasi pengenalan suara (*speech recognition*), pengenalan tulisan tangan dan tandatangan, *data mining*, pengklusteran, pengolahan isyarat, musik, dan masih banyak lagi. Dalam bidang seperti pengumpulan data dan pengambilan informasi, DTW telah berhasil diterapkan untuk mengatasi deformasi waktu secara otomatis dan kecepatan yang berbeda yang terkait dengan data *time-dependent* [5].



Gambar 1 Penyesuaian Antara Dua Urutan Time Dependent [5]

Asumsikan kita mempunyai dua deret numerik sebagai berikut  $(a_1, a_2, \dots, a_n)$  dan  $(b_1, b_2, \dots, b_m)$ . Algoritma memulai dengan kalkulasi jarak lokal antara elemen dari dua deret dengan menggunakan tipe jarak yang berbeda. Metode yang paling sering digunakan untuk kalkulasi jarak adalah rumus

jarak Euclidean yang menghasilkan matriks dengan  $n$  baris dan  $m$  kolom. Kalkulasi jarak lokal dihitung dengan mencari nilai absolut dari selisih kedua deret data yang dirumuskan seperti pada persamaan (2).

$$c_{ij} = |a_i - b_j|, i = 1 \text{ s.d. } n, j = 1 \text{ s.d. } m \quad (2)$$

Memulai dengan matriks jarak lokal, langkah selanjutnya merupakan nilai-nilai yang harus kita lewati untuk menemukan *warping path*. Untuk menghitung elemen baris pertama pada matriks DTW dirumuskan seperti pada persamaan (3).

$$D_{(1,j)} = D_{(1,j-1)} + c_{(1,j)} \quad (3)$$

Selanjutnya, untuk menghitung kolom pertama pada matriks DTW dirumuskan pada persamaan (4).

$$D_{(i,1)} = D_{(i,j-1)} + c_{(i,1)} \quad (4)$$

Untuk menghitung elemen lainnya pada matriks DTW, dirumuskan seperti pada persamaan (5),

$$D_{ij} = c_{ij} + \min(D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}) \quad (5)$$

dimana  $D_{ij}$  adalah jarak minimal antara deret bagian  $(a_1, a_2, \dots, a_i)$  dan  $(b_1, b_2, \dots, b_j)$ . Sebuah jalan melengkung (*warping path*) adalah jalur lintasan atau *path* yang melalui matriks yang berisi jarak minimal dari elemen  $D_{ij}$  hingga elemen  $D_{nm}$  yang terdiri dari elemen-elemen  $D_{ij}$  itu sendiri.

## III. ANALISIS DAN PERANCANGAN

### A. Rancangan Proses Ekstraksi Fitur

Proses menentukan fitur dilakukan dengan menghitung jarak antara titik masing-masing tangan dengan 10 titik lainnya menggunakan jarak Euclidean. Dengan perhitungan ini akan didapatkan fitur data sebanyak 21 elemen, yang merepresentasikan 10 jarak minimum antara joint telapak tangan kanan dan joint lain, 10 jarak minimum antara joint telapak tangan kiri dengan joint lain, dan 1 jarak minimum antara joint telapak tangan kanan dengan joint telapak tangan kiri. Keseluruhan joint yang digunakan terdapat pada Gambar 2.



Gambar 2 Joint yang Digunakan

Untuk meminimalisir kesalahan pada saat pengenalan gerakan yang serupa, dilakukan proses ekstraksi fitur baru. Proses ini dilakukan dengan cara menghitung rata-rata, median, modus, varians, standar deviasi, dan nilai kedalaman minimum dari masing-masing titik tangan terhadap dua titik absolute joints, yaitu kepala dan pinggul. Perhitungan ini dilakukan

untuk mengurangi kesalahan pada saat pengenalan gerakan tangan. Dari hasil perhitungan ini akan didapatkan 24 elemen yang kemudian digabungkan dengan fitur data baru menjadi fitur data akhir dengan total 45 elemen.

**B. Rancangan Proses Implementasi DTW**

Dalam proses pengenalan gerakan tangan, perangkat lunak terlebih dahulu mencocokkan data uji dengan data *training* yang sudah ada dengan metode DTW. Pada penelitian ini, data uji dan data *training* masing-masing ada sebanyak 45 data. Pada penelitian ini, masukan gerakan yang dikenali sebagai gerakan pada referensi merupakan gerakan yang memiliki *warping path* keseluruhan yang paling kecil dari gerakan lainnya.

**C. Rancangan Proses Training Data**

Proses *training* data dilakukan dengan menyimpan fitur data akhir ke dalam sebuah berkas berekstensi .csv. Ada 20 gerakan yang dilakukan pada saat pembuatan data, dengan masing-masing kata dilakukan 15 kali gerakan. Sehingga, total gerakan yang disimpan pada saat *training* data ada sebanyak 300 gerakan.

**D. Rancangan Proses Testing Data**

Proses testing data dilakukan dengan mendapatkan fitur data uji yang dilakukan secara *real-time*. Kemudian, dilakukan pencocokkan antara hasil dari fitur data tersebut dengan fitur data yang didapatkan pada saat proses pembuatan data menggunakan metode DTW. Gerakan yang dikenali sebagai gerakan pada referensi merupakan gerakan yang memiliki nilai *warping path* paling kecil. Keluaran hasil klasifikasi yang dilakukan berupa prediksi gerakan bahasa isyarat yang dimaksud oleh pengguna dalam bentuk gambar dan juga tulisan bahasa isyarat.

**IV. PENGUJIAN DAN EVALUASI**

Pengujian dilakukan terhadap 20 bahasa isyarat yang dipilih oleh penulis, yaitu Halo, Salam, Pagi, Jumpa, Apa, Bagaimana, Keluarga, Ibu, Anak, Laki-laki, Saudara, Saya, Anda, Rumah, sekolah, Baik, Jahat, Masalah, Kerja, dan Marah. Pada awalnya penulis memilih bahasa isyarat yang mengacu pada *paper* referensi. Namun, karena ada beberapa kata yang tidak terdapat pada kamus Sistem Isyarat Bahasa Indonesia (SIBI), maka penulis mengubah kata yang dipilih sesuai dengan kategori kelasnya.

Skenario pengujian yang dilakukan dibagi menjadi tiga bagian, yaitu skenario A, skenario B, dan skenario C dengan penjelasan sebagai berikut:

1. Pada skenario A, pengujian bahasa isyarat dilakukan oleh penulis dengan tinggi badan 163 cm menggunakan data *training* yang dibuat oleh penulis.
2. Pada skenario B, pengujian dilakukan oleh pengguna lain dengan tinggi badan 173 cm menggunakan data *training* yang dibuat oleh penulis.
3. Pada skenario C, pengujian dilakukan oleh pengguna lain dengan tinggi badan 175 cm menggunakan data *training* yang dibuat oleh pengguna tersebut.

Pengujian pada skenario B dilakukan dengan tujuan untuk mengetahui apakah perbedaan tinggi badan yang signifikan ketika pembuatan dan *testing* data berpengaruh terhadap akurasi pada saat pengenalan gerakan. Pengujian bahasa isyarat

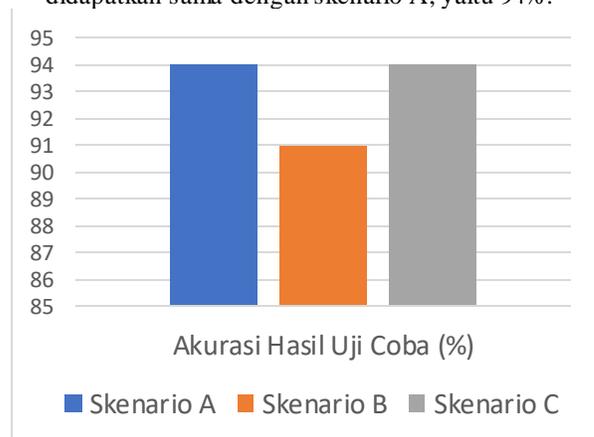
pada ketiga skenario dilakukan menggunakan 20 gerakan yang terdiri dari 3 bahasa isyarat statis dan 17 bahasa isyarat dinamis, serta dengan pergerakan yang menggunakan satu atau kedua tangan. Hasil pengujian seperti yang dapat dilihat pada Tabel 1.

Tabel 1  
Hasil Uji Coba

Kategori Makna	Kata	A(%)	B(%)	C(%)
Sapaan	Halo	100	100	100
	Salam	80	80	100
	Pagi	100	100	100
	Jumpa	100	100	100
Pertanyaan	Apa	100	80	100
	Bagaimana	80	60	60
Keluarga	Keluarga	80	60	100
	Ibu	80	80	100
	Anak	100	100	100
	Laki-laki	100	100	80
Kata Ganti	Saudara	100	100	100
	Saya	100	100	80
Orang	Anda	100	100	100
Bangunan	Rumah	80	100	100
	Sekolah	100	100	100
Lain-lain	Baik	100	100	100
	Jahat	100	100	100
	Masalah	100	80	60
	Kerja	80	100	100
	Marah	100	80	100
	Rata-rata Akurasi (%)		94	91

Hasil pengujian disusun dalam bentuk analisis secara keseluruhan sebagai berikut:

1. Implementasi DTW di dalam perangkat lunak yang dibuat mempunyai akurasi rata-rata yang cukup baik, yaitu 93%. Grafik hasil akurasi dari masing-masing percobaan dapat dilihat pada Gambar 3.
2. Adanya beberapa gerakan isyarat yang hampir sama antara satu dengan yang lain menyebabkan kesalahan pada proses klasifikasi.
3. Hasil yang didapatkan pada skenario A cukup baik, yaitu 94%. Sedangkan hasil akurasi yang didapatkan pada skenario B lebih rendah daripada skenario A, namun masih cukup baik, yaitu 91%. Hal ini membuktikan bahwa perbedaan tinggi pengguna pada *range* 163 cm – 173 cm memiliki pengaruh pada akurasi pengenalan gerakan. Untuk skenario C, hasil akurasi yang didapatkan sama dengan skenario A, yaitu 94%.



Gambar 3 Grafik Hasil Uji Coba

- Selain adanya beberapa gerakan bahasa isyarat dinamis yang hampir sama antara satu dengan yang lain, sensitifitas Kinect 2.0 dalam merekam koordinat masing-masing *skeleton joints* juga mempengaruhi tingkat akurasi pengenalan gerakan bahasa isyarat.

## V. KESIMPULAN DAN SARAN

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

- Titik representatif ditentukan dengan mengekstraksi fitur dari data rekaman asli. Fitur-fitur tersebut yaitu nilai minimum dari perhitungan jarak Euclidean antara joint telapak tangan dengan joint lain yang menghasilkan 21 elemen, dan perhitungan rata-rata, median, modus, varians, standar deviasi, dan nilai kedalaman minimum antara joint telapak tangan dengan absolute joints (kepala dan pinggul) yang menghasilkan 24 elemen. Jumlah keseluruhan titik representatif untuk masing-masing gerakan ada 45 elemen.
- Metode *Dynamic Time Warping* (DTW) yang digunakan untuk menerjemahkan bahasa isyarat berjalan dengan baik secara *real-time* dengan akurasi rata-rata 93%.
- Identifikasi posisi gerakan yang dilakukan ketika melakukan *training* dan *testing* data sangat berpengaruh terhadap akurasi pengenalan gerakan.
- Perbedaan karakteristik tinggi badan pengguna yaitu pada *range* antara 163 cm – 175 cm berefek terhadap penentuan koordinat *skeleton joints* dimana perbedaan tersebut mempengaruhi identifikasi posisi gerakan yang dilakukan oleh perangkat lunak, namun tidak mempengaruhi akurasi secara signifikan.
- Perbedaan posisi Kinect dengan pengguna, baik pada saat pengambilan data maupun uji coba data, mempengaruhi akurasi hasil pengenalan gerakan yang dilakukan oleh pengguna.

Kemudian saran untuk penelitian kedepannya adalah sebagai berikut:

- Memperbanyak data *training* dari berbagai macam pengguna yang mempunyai karakteristik tubuh yang berbeda-beda.
- Identifikasi *skeleton joints* tidak hanya dipergelangan tangan saja, tetapi juga pada jari tangan.

- Hasil akurasi yang didapatkan sudah cukup akurat namun masih lambat ketika mengenali gerakan. Untuk itu, menggabungkan metode DTW dengan teknik lain bisa menjadi solusi untuk meningkatkan kecepatan pada saat pengenalan gerakan.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Laboratorium Interaksi Grafika dan Seni (IGS) dan Departemen Informatika Institut Teknologi Sepuluh Nopember, Indonesia yang telah mendukung dan memfasilitasi penelitian ini.

## DAFTAR PUSTAKA

- A. Basuki, M. Zikky, J. Akhmad, N. Hasim, and N. I. Ramadhan, "Sensor Gerak Dengan Leap Motion Untuk Membantu Komunikasi Tuna Rungu / Wicara A-317 A-318," vol. 8, no. 1994, pp. 317–321, 2016. Retrieved from [https://www.researchgate.net/publication/320472422\\_Sensor\\_Gerak\\_Dengan\\_Leap\\_Motion\\_untuk\\_Membantu\\_Komunikasi\\_Tuna\\_RunguWicara](https://www.researchgate.net/publication/320472422_Sensor_Gerak_Dengan_Leap_Motion_untuk_Membantu_Komunikasi_Tuna_RunguWicara)
- I. Hidayat, A. Hasibuan, A. Mulyana, and A. B. O, "Menjadi Suara Berbasis Kinect Menggunakan Metode Dynamic Time Warping Design And Implementation Of Sign Language To Speech Application," pp. 1–7. Retrieved from [https://openlibrary.telkomuniversity.ac.id/pustaka/files/100124/jurnal\\_eproc/perancangan-dan-implementasi-aplikasi-penerjemah-bahasa-isyarat-menjadi-suara-berbasis-kinect-menggunakan-metode-dynamic-time-warping.pdf](https://openlibrary.telkomuniversity.ac.id/pustaka/files/100124/jurnal_eproc/perancangan-dan-implementasi-aplikasi-penerjemah-bahasa-isyarat-menjadi-suara-berbasis-kinect-menggunakan-metode-dynamic-time-warping.pdf)
- A. A. S. Gunawan, "Pembelajaran Bahasa Isyarat Dengan Kinect Dan Metode Dynamic Time Warping," *Univ. Binus*, no. 9, pp. 1–3, 2013. Retrieved from [http://library.binus.ac.id/eColls/eJournal/01\\_MatStat\\_Alexander%20Ashadi.pdf](http://library.binus.ac.id/eColls/eJournal/01_MatStat_Alexander%20Ashadi.pdf)
- Khotimah, W. N., Susanto, Y. A., & Suciati, N. (2017). "Combining Decision Tree And Back Propagation Genetic Algorithm Neural Network For Recognizing Word Gestures In Indonesian Sign Language Using Kinect", 95(2), 292–298. Retrieved from <http://www.jatit.org/volumes/Vol195No2/Vol195No2.pdf>
- G. García-bautista, F. Trujillo-romero, and S. O. Caballero-morales, "Mexican Sign Language Recognition Using Kinect and Data Time Warping Algorithm," 2010. Retrieved from <http://ieeexplore.ieee.org/document/7891832/?reload=true>
- M. Iqbal, "Pengenalan Bahasa Isyarat Indonesia Berbasis Sensor Flex dan Accelerometer Menggunakan Dynamic Time Warping," Institut Teknologi Sepuluh Nopember, Surabaya, 2011. Retrieved from <http://digilib.its.ac.id/TTS-Master-3100012045709/18061>