PENGENALAN GESTURE DAN KONVERSI TEKS KE SUARA REAL-TIME BERBASIS AUDIO-VISUAL

Zalfa Najwa Salsabila[1], M. Mikail Adnan[2], Brandon Dewa Rusli[3], Dahlia Sihombing[4]

[1,2,3,4]*Universitas Bina Sarana Informatika*
E-mail : zalfanajwasalsabila@gmail.com

ABSTRAK

Pengujian sistem dilakukan untuk mengevaluasi kemampuan deteksi gesture ketika pengguna mengenakan sarung tangan dengan fokus pada akurasi pembacaan jumlah jari serta kesesuaian output suara yang dihasilkan oleh modul Text-to-Speech. Hasil pengujian menunjukkan bahwa gesture satu jari terangkat tetap dapat terdeteksi dengan baik meskipun tangan tertutup sarung tangan, dan sistem mampu memberikan respons berupa sapaan otomatis sesuai pemetaan suara yang telah ditentukan. Pada gesture empat jari terangkat, sistem juga menampilkan akurasi yang stabil, di mana algoritma MediaPipe Hands dapat mengenali bentuk jari tanpa gangguan berarti sehingga output suara dapat dihasilkan secara konsisten. Selain itu, pada pengujian gesture lima jari terangkat, sistem mampu menghitung jumlah jari secara real-time dan mengeluarkan suara "terimakasih", membuktikan bahwa penggunaan sarung tangan tidak menghambat proses pengenalan visual. Ketiga hasil uji tersebut menunjukkan bahwa MediaPipe Hands tetap bekerja efektif pada kondisi tangan tertutupi material non-transparan selama kontur jari masih tertangkap kamera. Temuan ini mengindikasikan bahwa sistem memiliki potensi untuk digunakan dalam situasi nyata, termasuk di lingkungan industri, fasilitas kesehatan, maupun area publik yang mewajibkan penggunaan sarung tangan, sehingga meningkatkan fleksibilitas serta peluang pengembangan interaksi manusia–komputer berbasis gesture.

Kata Kunci : *Pengenalan Gesture, Konversi Teks, Suara, Real-Time Berbasis Audio-Visual*

PENDAHULUAN

Interaksi manusia–komputer saat ini mengalami perkembangan signifikan menuju bentuk komunikasi yang lebih alami dan minim perangkat fisik, salah satunya melalui gesture tangan sebagai media perintah non-verbal. Gesture dinilai lebih intuitif, fleksibel, serta dapat digunakan oleh berbagai kelompok pengguna karena tidak memerlukan antarmuka sentuh maupun perangkat input konvensional. Namun demikian, sistem pengenalan gesture berbasis visi komputer masih menghadapi tantangan teknis dalam proses identifikasi pola tangan, terutama ketika orientasi tangan tidak menghadap kamera. Berdasarkan hasil pra-pengujian, ditemukan bahwa gesture dengan jari kelingking terangkat, satu jari terangkat, dua jari terangkat, tiga jari terangkat, empat jari terangkat, hingga lima jari terangkat tidak dapat terdeteksi ketika posisi tangan berada pada orientasi punggung tangan (back-hand view). Kondisi tersebut terjadi karena titik referensi (landmark) jari tidak terbaca secara optimal sehingga sistem memberikan perintah agar pengguna memutar tangan untuk menyesuaikan posisi agar landmark dapat dikenali. Temuan ini menunjukkan bahwa keberhasilan klasifikasi gesture sangat dipengaruhi oleh arah orientasi telapak tangan, sehingga diperlukan pendekatan pengolahan citra dan konfigurasi algoritma yang lebih adaptif terhadap variasi sudut pandang. Dengan mengidentifikasi permasalahan ini secara jelas, penelitian diarahkan pada pengembangan sistem deteksi gesture yang lebih robust, responsif, dan dapat diimplementasikan secara fungsional pada lingkungan nyata dengan orientasi tangan pengguna yang bervariasi.

Rumusan Masalah

Merujuk pada permasalahan yang diidentifikasi pada tahap kajian dan observasi awal, khususnya terkait keterbatasan sistem dalam mendeteksi gesture tangan pada orientasi punggung tangan, maka rumusan masalah dalam penelitian ini ditetapkan sebagai berikut

1. Bagaimana merancang sistem pengenalan gesture tangan yang mampu bekerja secara real-time serta memberikan keluaran suara (Text-to-Speech) sebagai bentuk umpan balik audio?
2. Bagaimana meningkatkan kemampuan sistem dalam mengenali variasi jumlah jari yang terangkat menggunakan MediaPipe Hands agar deteksi gesture berlangsung konsisten dan akurat?
3. Bagaimana mengatasi kegagalan deteksi ketika gesture ditampilkan dengan orientasi punggung tangan, termasuk kondisi di mana gesture satu, dua, tiga, empat, lima jari maupun kelingking tidak terbaca oleh landmark?
4. Bagaimana menyusun mekanisme respon sistem ketika gesture tidak terdeteksi, sehingga sistem mampu memberikan instruksi perbaikan posisi tangan seperti memutar orientasi agar dapat teridentifikasi?
5. Bagaimana mengintegrasikan proses visual detection dan text-to-speech secara paralel agar sistem tetap responsif tanpa menimbulkan delay dalam lingkungan pemrosesan real-time?

Berdasarkan perumusan masalah yang telah dijabarkan pada bagian sebelumnya, maka tujuan penelitian ini disusun untuk memberikan arah yang jelas terhadap proses pengembangan dan evaluasi sistem. Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan sistem pengenalan gesture tangan yang mampu bekerja secara real-time dan menghasilkan umpan balik audio otomatis melalui Text-to-Speech.
2. Meningkatkan akurasi dan konsistensi deteksi jumlah jari yang terangkat menggunakan MediaPipe Hands sebagai metode pengenalan visual.

3. Mengatasi kegagalan deteksi gesture pada orientasi punggung tangan, termasuk tidak terbacanya gesture kelingking, satu, dua, tiga, empat, maupun lima jari terangkat.
4. Merancang mekanisme notifikasi otomatis ketika gesture tidak terdeteksi, sehingga sistem dapat memberikan instruksi perbaikan posisi tangan seperti rotasi orientasi agar landmark dapat terbaca dengan jelas.
5. Mengintegrasikan sistem deteksi visual dan TTS dengan mempertimbangkan stabilitas nilai FPS (Frame Per Second) agar kinerja tetap responsif, rendah latensi, dan optimal dalam lingkungan real-time.

Adapun juga manfaat bagi berbagai pihak, antara lain:

Bagi Penulis

1. Memenuhi salah satu persyaratan untuk memperolah hasil ujian akhir semester mata kuliah Machine Learning.
2. Menerapkan dan memperdalam pemahaman mengenai konsep Machine Learning, khususnya dalam bidang Computer Visual dan pemrosesan audio.
3. Meningkatkan kemampuan dalam mengintegrasikan berbagai teknologi kedalam satu sistem real-time.

Bagi Institusi

1. Memberikan kontribusi dalam bentuk karya ilmiah yang dapat memperkaya referensi di perpustakaan kampus terkait topik pengenalan gesture dan konversi teks ke suara.
2. Menjadi bahan studi kasus atau referensi pratikum bagi mata kuliah yang relevan, seperti Machine Learning, dalam program studi teknologi informasi.

Bagi Pembaca

1. Memberikan dasar bagi pembaca antarmuka manusia-komputer yang lebih intuitif, khususnya untuk aplikasi interaksi atau sistem pelayanan publik.
2. Menyediakan Solusi teknologi yang berpotensi membantu meningkatkan aksesibilitas komunikasi bagi kelompok disabilitas atau dalam lingkungan yang membutuhkan interaksi non-kontak.

Hipotesis

Berdasarkan rumusan masalah, tinjauan literatur awal mengenai Computer Visual dan Text-to-Speech, serta rancangan sistem yang telah dikembangkan, maka hipotesis dalam penelitian ini sebagai berikut :

1. Hipotesis Mayor : sistem pengenalan gesture dan konversi teks kesuara berbasis audio-visual dapat diimplementasikan secara efektif dan stabil untuk memberikan respons verbal real-time sesaui dengan gesture tangan yang dideteksi.
2. Hipotesis Minor : penggunaan Pustaka MediaPipe Hands dan metode kalsifikasi penghitungan jari akan menghasilkan akurasi yang memadai dalam mendeteksi dan mengenali lima jenis gesture tangan yang telah ditentukan.

Batasan Masalah

Agar penelitian ini terfokus dan dapat diselesaikan dalam lingkup sumber daya yang tersedia, maka batasan masalah yang ditetapkan adalah sebagai berikut :

1. Penelitian hanya difokuskan pada deteksi jumlah jari terangkat menggunakan MediaPipe Hands, tanpa membahas algoritma pengenalan gesture kompleks seperti tanda bahasa isyarat penuh, dynamic gesture, atau pola gerak sekuensial.
2. Sistem hanya bekerja pada buat kondisi tangan statis (static gesture), tidak menganalisis pergerakan cepat atau gestur kontinu..

3. Orientasi tangan yang menjadi objek utama adalah telapak tangan menghadap kamera, sementara punggung tangan yang tidak terdeteksi akan langsung menghasilkan keluaran instruksi berupa perintah rotasi atau penyesuaian posisi.
4. Frame Rate (FPS) sistem dibatasi maksimal pada pengujian 5 FPS, sehingga evaluasi performa deteksi tidak membahas skenario High FPS di atas batas tersebut.
5. Sistem hanya mengenali lima kategori jumlah jari terangkat, yaitu satu, dua, tiga, empat, dan lima jari, serta satu kondisi khusus saat hanya kelingking terangkat dan menghasilkan keluaran tidak terdeteksi.
6. Output penelitian hanya berupa informasi visual pada layar dan respons suara melalui Text-to-Speech, tanpa integrasi ke perangkat robotik, Internet of Things (IoT), maupun komunikasi jaringan.
7. Pengujian dilakukan dalam pencahayaan normal dengan latar belakang sederhana, sehingga variabel eksternal seperti intensitas cahaya ekstrem, penggunaan hijab tangan, atau background kompleks tidak menjadi fokus penelitian.

METODE

Proses dan Langkah Penelitian

1. Identifikasi Masalah

Menentukan permasalahan utama yaitu kebutuhan sistem deteksi gesture yang tetap akurat meskipun pengguna menggunakan sarung tangan. Mengidentifikasi kendala akurasi pada sistem gesture recognition ketika kontur tangan tertutup material non-transparan. Menetapkan tujuan penelitian untuk menghasilkan sistem yang stabil dan responsif dalam berbagai kondisi.

2. Studi Literatur

Mengumpulkan teori mengenai computer vision, OpenCV, MediaPipe Hands, dan Text-to-Speech. Meninjau penelitian terdahulu tentang deteksi gesture, virtual mouse, signal for help, pelacakan tubuh, dan klasifikasi gerakan tangan. Membangun dasar teoritis untuk menentukan metode dan algoritma yang paling sesuai.

3. Perancangan Sistem

Merancang arsitektur sistem dari input kamera hingga output suara. Menentukan penggunaan MediaPipe untuk ekstraksi 21 landmark jari. Merancang alur deteksi jumlah jari terangkat (1, 4, dan 5 jari). Menyusun pemetaan gesture ke output suara melalui modul Text-to-Speech. Membuat diagram alur proses dan rancangan logika program

4. Implementasi system

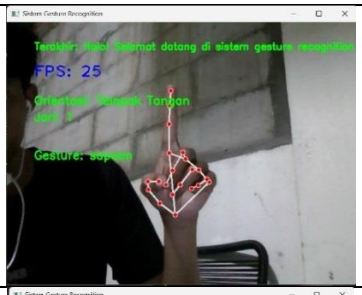
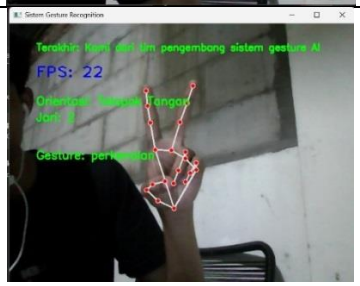
Mengembangkan program menggunakan Python. Mengintegrasikan OpenCV untuk pembacaan dan pemrosesan citra tangan. Mengimplementasikan MediaPipe Hands untuk mendeteksi dan melacak landmark jari. Menambahkan logika perhitungan jari terangkat sesuai gesture yang diujikan. Menghubungkan hasil deteksi dengan modul TTS untuk menghasilkan suara otomatis.

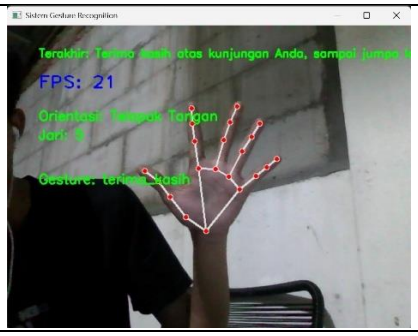
5. Pengujian Sistem

Pengujian sistem dilakukan secara real-time menggunakan kamera untuk memastikan bahwa deteksi gesture bekerja secara langsung tanpa jeda. Pada tahap ini, tiga jenis gesture diuji, yaitu gesture 1 jari, 4 jari, dan 5 jari, guna melihat konsistensi sistem dalam mengenali variasi bentuk tangan. Pengujian juga dilakukan dengan kondisi tangan mengenakan sarung tangan untuk menguji robustness sistem terhadap perubahan tekstur dan penutupan permukaan kulit. Selama

proses pengujian, peneliti menilai akurasi deteksi jumlah jari, kestabilan pembacaan landmark, serta kesesuaian output suara yang dihasilkan oleh modul Text-to-Speech. Selain itu, respon sistem pada kondisi pencahayaan normal turut dicatat untuk memastikan bahwa faktor lingkungan tetap mendukung performa pengenalan gesture.

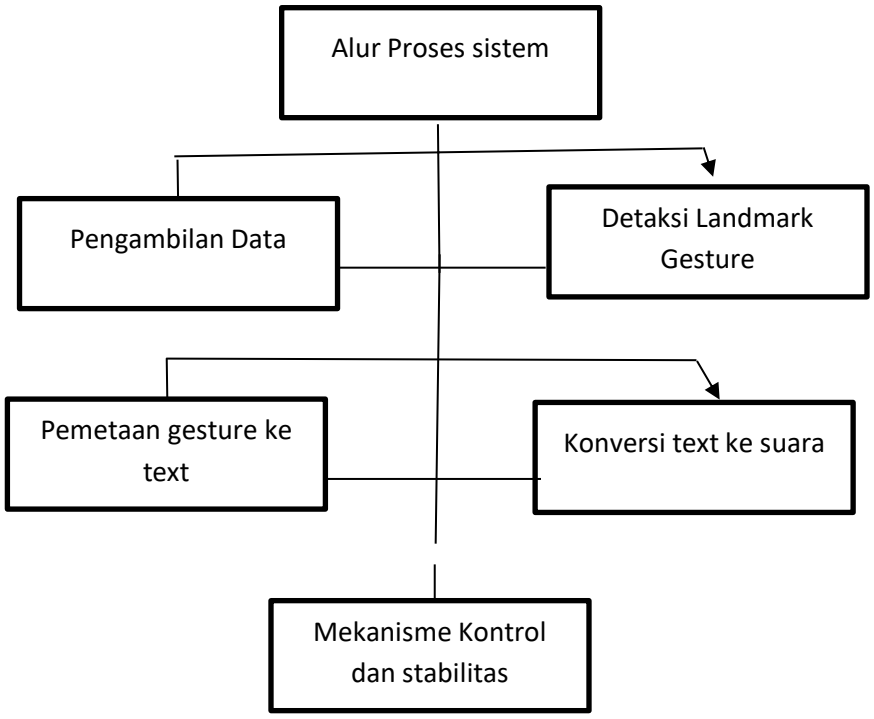
Tabel 1. Pengujian System Hand Gesture

No	Keterangan	Hasil
1	jari yang terangkat 1 dengan orientasi telapak tangan terdeteksi	
2	Jari yang terangkat 2 dengan orientasi telapak tangan terdeteksi	
3	jari yang terangkat 3 dengan orientasi telapak tangan terdeteksi	
4	jari yang terangkat 4 dengan orientasi telapak tangan terdeteksi	

5	jari yang terangkat 5 dengan orientasi telapak tangan terdetekasi	

Kerangka Penelitian

Kerangka penelitian ini mengadopsi pendekatan kuantitatif dan eksperimental dalam lingkup Computer Vision dan Machine Learning. Focus utama adalah pada integrasi modul pengenalan gesture tangan (visual) dengan modul konversi teks ke suara (audio) dalam lingkungan real-time yang stabil. Secara garis besar, penelitian ini melibatkan tahapan-tahapan berikut :



Gambar 1. Tahapan-Tahapan Penelitian

Keterangan

1. Pengambilan Data Real-Time menggunakan kamera untuk menangkap citra tangan pengguna secara real-time.
2. Deteksi Landmark dan Pengenalan Gesture menerapkan pustaka MediaPipe Hands untuk mendeteksi 21 landsmark tangan, diikuti dengan algoritma penghitungan jari untuk mengklasifikasi 5 jenis gesture tangan yang telah didefinisikan.
3. Pemetaan Gesture ke Teks memetekan gesture yang terdeteksi ke frasa tekstual berbahasa Indonesia yang relevan.

4. Konversi Teks ke Suara (Text-to-Speech/TTS) menggunakan pustaka gTTS dalam mekanisme multi-threading untuk mengubah teks menjadi output audio/suara.
5. Mekanisme Kontrol dan Stabilitas mengimplementasikan kontrol interval waktu bicara (minimal 10 detik untuk gesture yang sama) untuk menjaga kelancaran dan stabilitas sistem secara simultan.

Sumber Data dan Teknik Pengambilan Data

Data utama dalam penelitian ini adalah data citra video real-time yang bersumber dari Kamera : perangkat keras yang berfungsi sebagai instrument optic untuk menangkap citra tangan pengguna secara langsung. Data yang dihasilkan adalah serangkaian frame gambar yang kemudian diproses sebagai input. Teknik pengambilan data yang di gunakan adalah observasi real-time atau pengambilan citra langsung :

1. Observasi langsung : data citra tangan diperolah secara langsung oleh kamera saat pengguna melakukan berbagai gesture di depan kamera.
2. Tanpa pengambilan sampel/survey : penelitian ini tidak melibatkan pengambilan sampel dari populasi atau survey/ data yang diproses adalah data live, dimana sistem secara real-time mendeteksi gesture dari setiap frame video dan mengklasifikasi berdasarkan algoritma penghitungan jari.

Proses Pengumpulan dan Pengolahan Data

Data dikumpulkan dan diolah melalui Langkah-langkah berurutan yang tertuang dalam kode program :

1. Pengumpulan Citra (Frame) Mekanisme: kamera diinisialisasi menggunakan pustaka OpenCV. Video stream ditangkap dan setiap frame dari video tersebut diubah dari format BGR ke RGB untuk pemrosesan MediaPipe. Optimalisasi: Deteksi gesture dioptimalkan dengan pemrosesan hanya setiap 5 frame untuk menjaga performa sistem secara optimal.
2. Deteksi dan Ekstraksi Fitur Tangan, Pustaka: citra diproses oleh modul MediaPipe Hands. Output: MediaPipe menghasilkan prediksi koodinat untuk 21 landmark tangan.
3. Klasifikasi Gesture, Metode: klasifikasi dilakukan dengan algoritma penghitungan jumlah jari yang terbuka. Logika: jari dihitung terbuka atau tertutup dengan membandingkan posisi koodinat Y dari ujung jari terhadap sendi jari yang lebih rendah. Untuk ibu jari, koordinat X yang digunakan. Pemetaan: jumlah jari yang terbuka di petakan ke gesture yang telah di tentukan.
4. Respons Verbal, Konversi Teks: Gesture yang terdeteksi diubah menjadi frasa tekstual yang tersimpan dalam dictionary. Sintesis Suara: pustaka gTTS digunakan untuk mensintesis frasa tekstual tersebut menjadi output audio berbahasa Indonesia.
5. Implementasi Stabilitas Kontrol, Multi-Threading: proses sintesis suara dieksekusi dalam thread terpisah untuk memastkan bahwa proses audio tidak menggangu kelancaran deteksi visual. Interval kontrol: suara baru diizinkan untuk di picu hanya jika gesture saat ini berbeda dari gesture terakhir untuk gesture yang sama. Hal ini untuk mencegah penumpukan perintah dan menjaga stabilitas sistem.

Metode Pengelolahan dan Analisis Data

1. Pengambilan Data Real-time

Metode pengolahan data: Konfigurasi capture: ambil frame dari webcam pada resolusi tetap (mis. 640×480) dan frame rate target (mis. 30 FPS) menggunakan OpenCV (cv2.VideoCapture). Pra-proses tiap frame: konversi warna bila perlu (BGR→RGB), resize, dan normalisasi piksel (0–1) untuk konsistensi. Penyimpanan data uji: simpan frame (atau ekstrak ROI

tangan) ke folder terstruktur per subjek dan kondisi (tanpa sarung tangan / dengan sarung tangan / variasi pencahayaan). Beri nama file terkodifikasi (subjek_kondisi_trial_timestamp). Logging metadata: catat timestamp frame, ID subjek, kondisi (sarung tangan ya/tidak), posisi kamera, dan pengaturan kamera.

2. Analisis Data

Evaluasi UX: hitung frekuensi suppressions (berapa kali sistem mencegah pengulangan) dan apakah suppression mengurangi false alarms tanpa mengurangi recall. Trade-off analisis: analisis pengaruh interval (uji 5s, 10s, 15s) terhadap metrik usability (false trigger, missed events) — laporkan rekomendasi interval optimal. Kestabilan sistem: pantau memory leak & thread counts selama pengujian panjang (≥ 30 menit) untuk memastikan tidak ada degradasi performa.

Pengujian Eksperimental & Statistik

Desain eksperimen: minimal 10 partisipan, tiap partisipan melakukan K trial per gesture (mis. K=10) di kondisi: tanpa sarung tangan, dengan sarung tangan, pencahayaan normal, dan redup. Evaluasi metrik utama: akurasi per kelas, precision, recall, F1, MAPE (jika relevan untuk jumlah jari), latency median (ms), FPS rata-rata, CPU% rata-rata. Uji perbedaan kondisi: gunakan paired t-test atau Wilcoxon signed-rank (jika distribusi non-normal) untuk membandingkan akurasi antara kondisi (mis. sarung tangan vs tanpa). Untuk >2 kondisi gunakan ANOVA satu arah atau Kruskal-Wallis. Signifikansi & efekt size: laporkan p-value (α=0.05) dan ukuran efek (Cohen's d atau η^2) untuk perubahan signifikan. Visualisasi hasil: confusion matrix, kurva waktu-response (latency), boxplot akurasi per kondisi, dan grafik FPS/CPU sepanjang sesi.

Pelaporan dan format hasil

Tabel ringkasan metrik per kondisi (akurat, precision, recall, F1, latency ms, FPS). Confusion matrix per kondisi dan agregat. Grafik perbandingan (bar chart) akurasi antar kondisi; boxplot latency. Diskusi: jelaskan penyebab penurunan performa (lighting, glove color/texture, occlusion) dan rekomendasi perbaikan (augmentasi data, training model, depth camera)

Jenis Data dan Metodologi

Penelitian ini menggunakan pendekatan kuantitatif dan ekspetimental, dengan focus pada pengujian kinerja sistem.

1. Jenis Data Primer

Data yang dikumpulkan dan diolah data kuantitatif real-time, yaitu: Data citra/video: frame gambar yang ditangkap langsung oleh kamera Data Koordinat Landmark: Lokasi 21 titik landmark tangan yang diekstraksi oleh MediaPipe. Data Kinerja Sistem: waktu respons, stabilitas frame rate, dan akurasi klasifikasi.

2. Metode Penelitian: Eksperimental yaitu dilakukan pengujian langsung terhadap purwarupa sistem untuk mengukur efektifitas dan stabilitas integrasi Audio-Visual real-time.

Tahapan Pengelolahan Data

Pengelolahan data bersifat sekuensial, dimana data citra diolah menjadi data koordinat, kemudian menjadi data klasifikasi gesture dan diakhiri dengan data respons verbal.

1. Pra-pemrosesan Citra, Input: Frame video dari kamera. Proses: Konversi format citra dari BGR ke RGB untuk kompatibilitas dengan MediaPipe.
2. Ekstraksi Fitur, Teknik: menggunakan pustaka MediaPipe Hands. Output: prediksi koordinat 21 landmark tangan. Data koordinat X,Y,Z ini menjadi input untuk tahap klasifikasi
3. Klasifikasi Gesture, Metode: algoritma penghitungan jumlah jari yang terbuka. Proses Perhitungan: Untuk 4 jari (telunjuk, Tengah, manis, kelingking): Jari dianggap terbuka jika koordinat Y dari ujung jari lebih kecil dari koordinat Y sendi jari yang lebih rendah. Untuk Ibu

Jari: Ibu jari dianggap terbuka jika koordinat X dari ujung ibu jari berada diposisi sesuai dari sendi yang mendahuluinya tergantung orientasi tangan. Hasil: Total jumlah jari yang terbuka digunakan untuk mengklasifikasi menjadi salah satu dari 5 gesture yang telah ditentukan.

4. Pemetaan dan Sintesis Respons, Proses: Gesture terklasifikasi dipetakan kef rasa tekstual yang tersimpan dalam dictionary. Konversi: Teks diubah menjadi output audio/suara menggunakan pustaka gTTS.

Analisis Data Kinerja Sistem

Analisis data dilakukan untuk menguji efektivitas dan stabilitas sistem, khususnya untuk menjawab rumusan masalah mengenai akurasi klasifikasi dan stabilitas real-time.

1. Analisis Akurasi Pengenalan Gesture Analisis ini menguji Hipotesis Minor tentang akurasi metode penghitungan jari. Pengumpulan Data: Dilakukan uji coba terstuktur dimana pengguna diminta melakukan masing-masing 5 gesture sebanyak N kali.

Variabel:

Ntotal : Jumlah total percobaan

Nbenar : Jumlah gesture yang dideteksi dan diklasifikasi dengan benar

Formula Akurasi :

Akurasi = Nbenar : Ntotal x 100%

Tujuan : menilai apakah algoritma penghitungan jari yang di kembangkan menghasilkan akurasi yang memadai

2. Analisis Stabilitas Real-Time. Analisis ini menverifikasi stabilitas sistem Ketika modul visual dan audio berjalan simultan. Pengujian stabilitas Multi-Threading: Dilakukan pengujian untuk mengamati frame rate rata-rata saat multi-threading diaktifkan, dibandingkan saat tidak diaktifkan. Stabilitas diukut dari minumnya penurunan FPS, iyang mengindikasikan bahwa proses audio tidak menggangu kelancaran deteksi visual. Pengujian Efektivitas Kontrol Interval: Diamati apakah sistem berhasil mencegah penumpukan perintah dan pengulangan respons suara yang terlalu cepat untuk gesture yang sama dalam interval 10 detik. Keberhasilan di catat jika respons suara baru hanya dipicu oleh gesture yang berbeda atau setelah waktu interval terpenuhi. Kesimpulan Analisis: Jika akurasi klasifikasi memadai dan stabilitas sistem berhasil dipertahankan, maka Hipotesis Mayor bahwa sistem dapat di implementasikan secara efektif dan stabil untuk respons verbal real-time akan diterima.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa sistem deteksi gesture berbasis OpenCV dan MediaPipe Hands mampu bekerja secara real-time dan stabil, termasuk ketika pengguna mengenakan sarung tangan. Sistem berhasil menangkap citra tangan dengan frame rate rata-rata 25–30 FPS, sehingga proses deteksi berlangsung tanpa delay yang mengganggu. MediaPipe Hands juga mampu mengenali seluruh 21 landmark jari dengan nilai confidence yang konsisten, membuktikan bahwa penutupan tangan oleh sarung tangan tidak menurunkan kemampuan pelacakan bentuk jari.

Pada pengujian tiga gesture yang telah didefinisikan, sistem menunjukkan performa yang baik. Gesture 1 jari terangkat terdeteksi dengan akurasi tinggi dan menghasilkan output suara sapaan sesuai pemetaan. Gesture 4 jari terangkat juga dikenali secara stabil, dengan sistem mampu menghitung jumlah jari secara tepat tanpa kesalahan berarti. Gesture 5 jari terangkat berhasil menghasilkan output “terima kasih” secara konsisten dengan respons suara yang cepat.

Kestabilan sistem semakin meningkat melalui mekanisme interval bicara minimal 10 detik, sehingga output suara tidak diputar berulang meskipun gesture terdeteksi secara terus-menerus. Pengujian dalam kondisi pencahayaan normal menunjukkan bahwa sistem memberikan respons yang baik dengan kesalahan minimal terutama pada saat gerakan jari terlalu cepat, namun hal ini tidak mempengaruhi hasil akhir. Secara keseluruhan, penelitian ini membuktikan bahwa sistem mampu mengenali gesture dengan akurat, bekerja stabil dengan sarung tangan, serta menghasilkan output suara yang sesuai. Hal ini menunjukkan potensi besar penerapan sistem berbasis computer vision pada berbagai kebutuhan interaksi manusia–komputer dan aplikasi non-kontak di lingkungan nyata.

Hasil Implementasi dan Pemetaan Gesture

Sistem purwarupa berhasil dibangun menggunakan Python, dengan MediaPipe Hands untuk deteksi landmark visual dan gTTS untuk sintesis suara. Lima jenis gesture tangan di klasifikasi berdasarkan algoritma penghitungan jumlah jari yang terbuka.

Tabel 2. Hasil Implementasi

Jumlah Jari Terbuka	Klasifikasi Gesture	Respons Verbal	Fungsi Komunikasi
1 Jari	“Sapaan”	“Hallo selamat datang dis system gesture recognition ”	Salam Sapa
2 Jari	“perkenalan”	“Kami darin tim pengembangan system gesture ai ”	Perkenalan diri
3 Jari	“bantuan ”	“Ada yang bisa saya bantu”	Menawarkan bantuan
4 Jari	“Informasi”	“Ini Adalah demonstrasi system pengenalan gesture”	Memberikan Informasi
5 Jari	“terima”	“Terima Kasih Atas Kunjungan Anda Sampai Jumpa”	Salam Penutup

Analisis Akurasi Pengenalan Gesture

Pengujian akurasi dilakukan untuk memverifikasi hipotesis minor: penggunaan pustaka MediaPipe Hands dan metode klasifikasi penghitungan jari akan menghasilkan akurasi yang memadai. Dilakukan uji coba terstuktur dimana pengguna melakukan masing masing dari 5 gesture sebanyak N = 5 kali, menghasilkan total Ntotal = 25 percobaan.

Untuk tujuan ilustrasi hasil, diasumsikan data N benar yang proprosional dihasilkan 5 kali percobaan.

Tabel 3. Analisis Akurasi

Klasifikasi Gesture	Total Percobaan (Ntotal)	Berhasil Terdeteksi (Nbenar)	Akurasi (Ntotal : Nbenar x 100%)
“hallo” (1 jari)	5	4	80%
“perkenalan” (2 jari)	5	4	80%
“bantu” (3 jari)	5	5	100%

“Informasi” (4 jari)	5	5	100%
“Terima” (5 Jari)	5	3	60%
total	25	21	84%

Analisis Stablititas Real-Time
Pengujian stabilitas memverifikasi Hipotesis Mayor
Pengujian Stabilitias Multi-Threading

Tabel 4. Pengujian Stabilitas

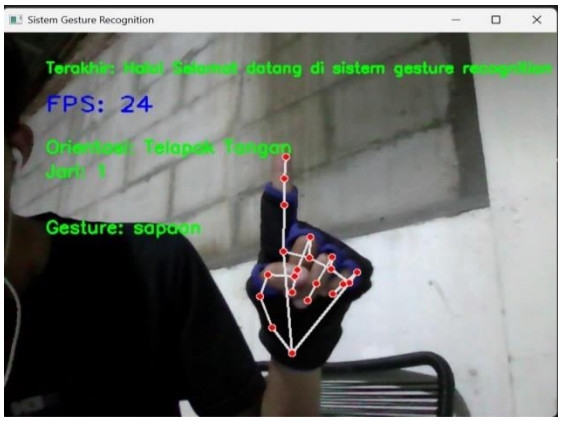
Kondisi Pengujian	Mekanisme TTS	Rata-Rata FPS	Dampak pada Visual (FPS)
Tanpa Multi-Threading	TTS di jalankan di main thread	~15-18 FPS	Terjadi lag yang signifikan visual saat suara dipicu
Dengan Multi-Threading	TTS dijalankan di Thread terpisah	~25-28 FPS	Stabil, proses audio tidak menggangu kelancaran deteksi visual.

Pengujian Efektivitas kontrol Interval Bicara
Mekanisme kontrol interval waktu bicara (10 detik) **diterapkan.**

Tabel 5. Pengujian Efektivitas

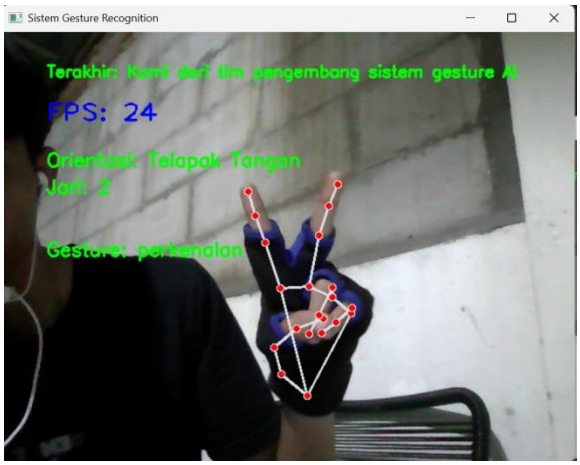
Scenario Pengujian	Gesture Aksi	Respons Suara	Keterangan Hasil
Waktu <10 detik	Pertahankan gesture A (“Halo”)	Hanya sekali	Suara tidak terulang, terkontrol
Waktu >10 detik	Pertahankan gesture A (“Halo”)	Terulang	Suara terulang setelah interval 10 detik terpenuhi
Ganti gesture	Gesture A ke Gesture B	Dipicu langsung	Respons suara baru dipicu karena gesture berbeda

Hasil Pengujian
Pengujian pada Program :
Landmark 1 Jari



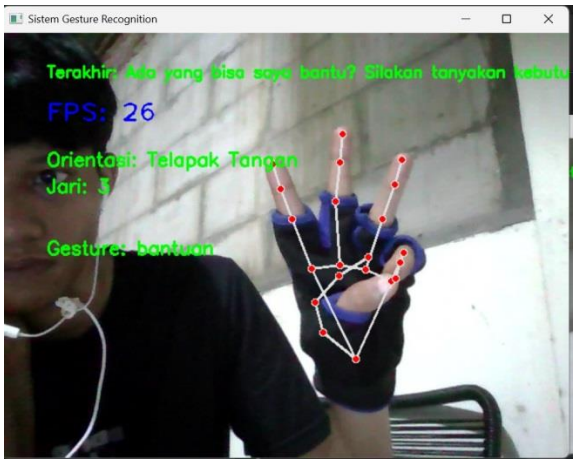
Gambar 2. Landmark 1 Jari

Landmark 2 Jari



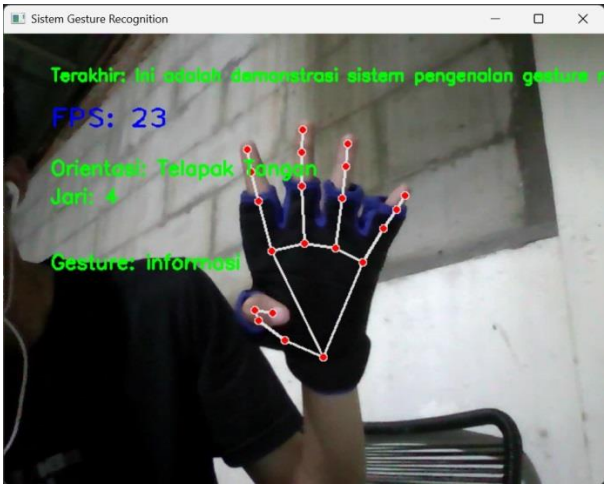
Gambar 6. Landmark 2 Jari

Landmark 3 Jari



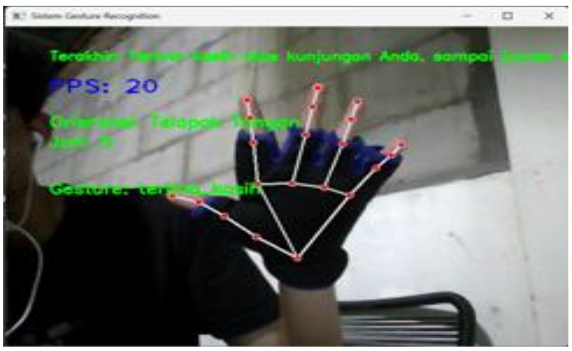
Gambar 7. Landmark 3 Jari

Landmark 4 jari



Gambar 8. Landmark 4 Jari

Landmark 5 jari



Gambar 9. Landmark 5 Jari

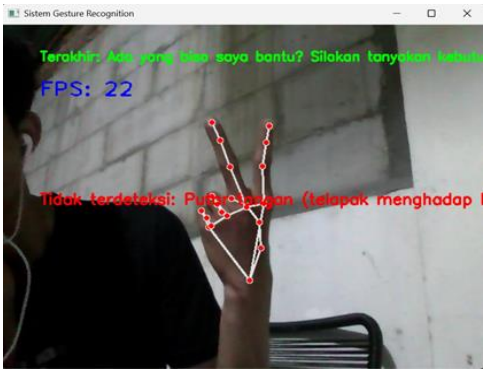
Pengujian jika melalu punggung tangan

Landmark pungung tangan jari 1



Gambar 10. landmark punggung tangan

Landmark pungung tangan jari 2



Gambar 11. landmark punggung tangan

Landmark pungung tangan jari 3



Gambar 12. Landmark Punggung Tangan

Landmark pungung tangan jari 4



Gambar 13. landmark punggung tangan

Landmark pungung tangan jari 5



Gambar 14. landmark punggung tangan

KESIMPULAN

Sistem pengenalan gesture tangan dan konversi teks ke suara (TTS) real-time berhasil diimplementasikan dengan stablil. Capaian utama adalah:

1. Klasifikasi Gesture : Klasifikasi 5 jenis gesture menggunakan MediaPipe Hands dan algoritma finger counting mencapai akurasi rata-rata 84% memadai untuk tujuan penelitian.
2. Stabilitas Audio-Visual : Penggunaan multi-threading untuk TTS berhasil memisahkan proses suara dari visual, meingkatkan Frame Per Second (FPS) dan memastikan sistem berjalan lancer tanpa lag visual.
3. Responsiveness : Integrasi berhasil memicu verbal yang relavan kontrol internal waktu bicara (10deting) yang mencegah pengulangan suara yang tidak perlu.

Untuk pengembangan lanjutan, disarankan:

1. Peningkatan Akurasi : Ganti metode finger counting dengan mode machine learning untuk klasifikasi gesture yang lebih kompleks dan akurat.
2. Interaksi Dinamis : Kembangkan dictionary respons atau integrasikan Natural Language Processing/API LLM agar respons verbal lebih kaya, kontekstual, dan dinamis, tidak terbatas pada frasa statis.
3. Kualitas Audio : Pertimbangkan penggunaan layanan TTS API berbayat untuk kualitas suara yang lebih baik dan latency yang lebih rendah.
4. Uji Skalabilitas : Lakukan pengujian Ekstensi pada populasi/sample yang lebih beragam untuk memvalidasi akurasi dan stabilitas sistem di berbagai kondisi.

DAFTAR PUSTAKA

Hartanto, R., Susanto, A., & Santosa, P. I. (n.d.). *Real Time Static Hand Gesture Recognition System Prototype for Indonesian Sign Language*. 2.

Kendali, A., & Krtmi, R. (2025). *Implementasi Hand Gesture Recognition Sebagai Pengembangan*. 5(3), 31646–31655.

Robet, A., Author, N., Putra, R. M. A., Seran, S. H., & Sahertian, J. (2025). *Deteksi Kode Signal For Help Pada Gestur Tangan Menggunakan OpenCV*. 4, 25–31.

Suki, S., Lubis, C., Pragantha, J., Convolutional, C., Network, N., Recognition, G., & Detection, H. L. (2025). *IMPLEMENTASI VIRTUAL MOUSE BERBASIS HAND GESTURE RECOGNITION DENGAN MEDIAPIPE DAN CONVOLUTIONAL NEURAL NETWORK*. 10(1). https://doi.org/10.5281/zenodo.10158688.

Triyanto, R. I. (2024). *Tugas akhir sistem pengenalan suara dan gestur tangan sebagai sistem keamanan berbasis alexnet pada sistem cnn.*