Subsistem Pengenalan Gestur dan Muka, Integrasi Sistem, serta Pengujian Perangkat Keras pada Sistem Identifikasi Tingkat Keamanan Tempat Tinggal Berdasarkan Analisis Kebiasaan

Dafa Faris Muhammad[1], Reza Darmakusuma[2], Aciek Ida Wuryandari[3]
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung
Jalan Ganesha No. 10 Bandung, 40132, Indonesia
[1]d.faris323@gmail.com, [2] reza.darmakusuma@gmail.com, [3]aciek@lskk.stei.itb.ac.id

***Abstrak*—Meningkatnya kemampuan komputasi berarti semakin banyak aspek kehidupan yang bisa tergantikan oleh komputer. Salah satu aspek penting yang berkembang pesat akibat peningkatan ini adalah sistem keamanan, sebab semakin banyak informasi yang bisa diperoleh tanpa perlu adanya campur tangan manusia secara terus menerus. Dengan memanfaatkan estimasi pose untuk mengambil kerangka tubuh manusia, suatu sistem didesain untuk mengenali gestur tubuh. Lapisan pengenalan lebih mendalam dilakukan dengan bantuan berbagai macam pemrosesan data lebih lanjut, sehingga suatu sistem bisa memberikan kesimpulan dengan level abstraksi yang lebih tinggi seperti aman atau tidaknya suatu ruangan. Penambahan sub-sistem pengenalan muka bermanfaat dalam mengatur wewenang yang dipegang setiap subjek dalam bertindak, sedangkan pengenalan objek untuk menambah variabel pengenalan. Pengujian dibagi atas dua, sistem itu sendiri serta spesifikasi perangkat keras minimum yang dibutuhkan untuk menjalankan sistem. Dalam skenario uji, sistem memiliki reliabilitas untuk tetap mencakup seminimalnya 88% dari area semula ketika salah satu kamera dari total empat mati, akurasi pengenalan muka sebesar 90,44%, serta akurasi penentuan aman tidaknya ruangan sebesar 93,75%. Sedangkan perangkat keras yang digunakan setidaknya harus mampu menjalankan sistem dengan kecepatan minimum 3 FPS, yang mana pengujian menunjukkan bisa dicapai dengan menggunakan GPU Nvidia dengan Compute Capability minimum 3,5 serta VRAM minimum 4 GB.**

Kata Kunci—gestur tubuh, muka, pemrosesan data, tingkat keamanan

I. PENDAHULUAN

Salah satu tren yang terjadi dalam masa perkembangan teknologi yang pesat ini adalah bergesernya ketersediaan pekerja, pekerjaan yang memerlukan pengulangan atau monoton semakin tidak diminati. Sehingga pekerja keamanan merupakan pilihan berat dalam segi ekonomi. Belum lagi meningkatnya kecenderungan untuk tempat tinggal ditinggal kosong karena pekerjaan, sekolah, dan lain-lain.

Itulah mengapa kamera keamanan merupakan salah satu opsi dalam menangani permasalahan keamanan tersebut. Sayangnya sistem kamera keamanan komersial umumnya tidak memiliki kemampuan untuk mengambil kesimpulan. Dalam kondisi tidak termonitor, sistem tersebut kurang lebih hanya bekerja sebagai perekam ruangan; bisa memberikan informasi berarti setelah suatu kejadian terjadi.

Dengan mendesain sistem yang dapat mengambil level keamanan, pemilik rumah dapat dibantu dalam berbagai hal. Mulai dari pemberian peringatan apabila ada suatu hal yang dianggap mencurigakan oleh sistem, mengeluarkan alarm yang bisa bekerja sebagai pencegah, dan lain-lain sesuai dengan keinginan pemilik; tentunya selama pemilik memahami karakteristik maupun kekurangan sistem.

II. FITUR PENDUKUNG

Ada berbagai fitur maupun fungsionalitas pendukung yang tidak berkaitan langsung dengan sistem utama demi meningkatkan performa sistem, penampilan sistem, memberikan lebih banyak kontrol, ataupun tujuan-tujuan lainnya. Fungsionalitas tersebut antara lain:

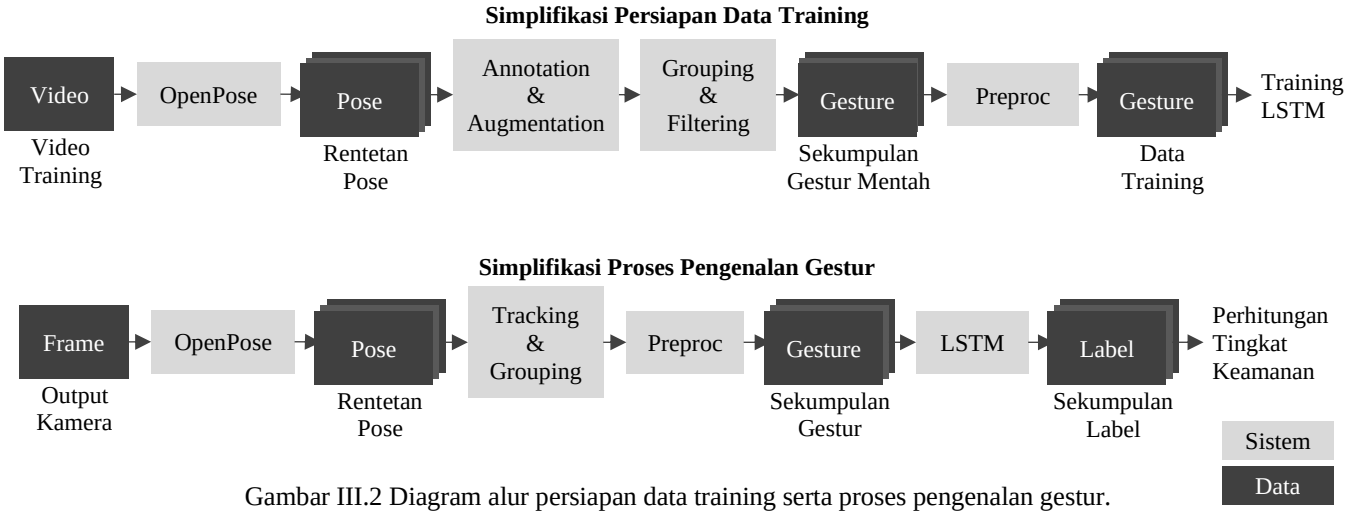
- Mengambil gambar dari empat kamera IP secara paralel.
- Visualisasi output sistem utama pada gambar.
- *Masking* untuk menutupi area-area tertentu, seperti lokasi yang tidak bisa dilewati namun memberikan banyak noise.
- Data dummy untuk mempermudah pengujian.
- Pembatas FPS untuk menjaga laju sistem.

(a)

(b)

Gambar II.1 Output visual sistem, dengan kotak atas menunjukkan level keamanan. Visualisasi hasil kamera muka pada gambar tengah. Masking aktif pada daerah meja

kerja, namun tidak ditampilkan. (a) Orang dikenal masuk ruangan. (b) Orang dikenal keluar ruangan.



Gambar III.2 Diagram alur persiapan data training serta proses pengenalan gestur.

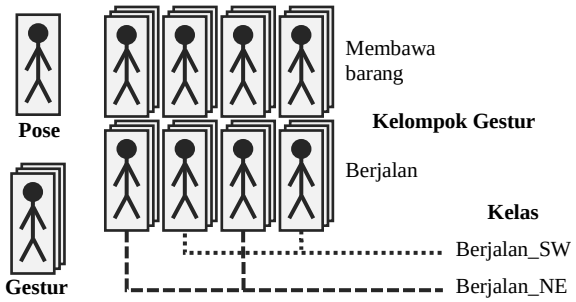
Sistem multi kamera diterapkan dengan menyatukan gambar dari masing-masing kamera ke dalam satu gambar besar. Hal ini berarti empat sisi dari kamera utama dimasukkan ke sistem layaknya satu buah gambar. Salah satu karakteristik dari dilakukannya implementasi ini antara lain sumber kamera dari suatu hasil recognition bisa ditentukan apabila koordinatnya melewati batas. Ada juga kekurangan berupa adanya kemungkinan subsistem salah mengenali adanya objek/kerangka yang muncul melewati batas.

Sedangkan satu buah kamera ditambahkan secara khusus untuk subsistem pengenalan muka sekaligus analisis warna pakaian. Gambar dari kamera ini tidak menghalang gambar kamera utama dalam pemrosesan, sehingga tidak mempengaruhi kerja subsistem lainnya secara fungsional. Inisial orang dikenal yang ada di dalam ruangan serta warna pakaiannya terletak pada pojok kiri bawah Gambar II.1.

III. Pemrosesan Data

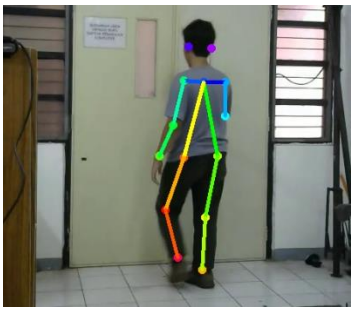
Terminologi yang digunakan dalam dokumen ini antara lain:

- Pose = kerangka tubuh manusia.
- Gestur = serentetan pose.
- Kelompok gestur = kumpulan gestur yang serupa.
- Kelas = sebagian dari satu kelompok gestur yang telah dipisah untuk keperluan sistem pengenalan gestur.



Gambar III.1 Simplifikasi data yang digunakan dan sebutannya.

Sebagaimana terlihat pada Gambar III.2, data pose atau kerangka tubuh diperoleh dari subsistem OpenPose. Pada sistem tersebut, gambar suatu tubuh manusia (ataupun sebagian) bisa dibuatkan representasi kerangkanya dalam bentuk vektor [1]. Gambar III.3 merupakan visualisasi contoh keluarannya.



Gambar III.2 Visualisasi pose dari keluaran OpenPose.

A. *Pengelompokan Kelas-Kelas Gestur*

Secara umum suatu gestur bisa dianalogikan sebagai kumpulan sinyal dan terdiri atas tiga komponen penyusun:

Gestur = Offset + Gerakan besar + Gerakan kecil

- Offset = lokasi terjadinya gestur
- Besar = translasi tubuh, gerakan full-body, dll
- Kecil = gait, gerakan anggota tubuh, dll

Dalam mengelompokkan suatu kelas, ada dua skema utama yang diuji dengan tujuan agar sistem pengenal gestur dapat mengambil fitur-fitur serta tingkatan generalisasi yang diharapkan:

1) Berdasarkan Kamera

Dengan empat kamera terletak pada sudut-sudut ruangan, akan ada variasi dalam bentuk kerangka yang terjadi pada masing-masing kamera untuk satu kelompok gestur yang sama. Baik disebabkan oleh lokasi ataupun sudut yang berbeda (relatif terhadap kamera terhadap subjek). Sehingga memisah setiap kelompok gestur berdasarkan sisi kamera akan mengurangi tercampurnya kelas-kelas dengan fitur yang berbeda pada sisi komponen offset maupun rasio ukuran dan kemiringan.

Di bawah ini merupakan kelas-kelas yang digunakan dalam skema data ini. “Jalan” untuk gestur berjalan, “Menyapu” untuk gestur menyapu, “Barang” untuk gestur berjalan dengan membawa barang dengan dua tangan, dan “Diam” untuk gestur yang tidak bertranslasi. Sedangkan NE,

SW, SE, dan SW berarti gestur diambil oleh kamera yang berlokasi di empat mata angin tersebut.

```
jalan_NE, jalan_NW, jalan_SE, jalan_SW,
menyapu_NE, menyapu_NW, menyapu_SE, menyapu_SW,
barang_NE, barang_NW, barang_SE, barang_SW,
diam_NE, diam_NW, diam_SE, diam_SW
```

2) Berdasarkan Gerakan Besar

Salah satu contoh gerakan besar adalah translasi tubuh. Dalam skema ini, setiap kelompok gestur dibagi ke dalam 4 kelas yaitu kanan-atas, kanan-bawah, kiri-atas, dan kiri-bawah (dikodekan sebagai UR, DR, UL, dan DL berturut-turut). Untuk menghilangkan faktor komponen offset dalam skema ini, digunakan preprocessing normalisasi. Alhasil setiap kelas dari skema ini akan memiliki komponen offset dan gerakan besar yang serupa. Kekurangan terbesar dalam skema ini terletak pada proses anotasi yang lebih mendetail.

Di bawah ini kelas-kelas yang digunakan dalam skema data ini. Penjelasan tiap kelas serupa dengan sebelumnya, kecuali beberapa hal. "Barang2" untuk gestur berjalan dengan membawa barang dengan dua tangan, "Barang1l" dan "Barang1r" berturut-turut untuk gestur berjalan dengan membawa/merangkul barang di tangan kiri dan kanan.

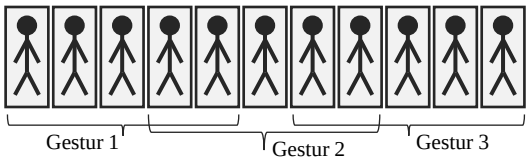
```
jalan_DR, jalan_UR, jalan_DL, jalan_UL,
barang2_DR, barang2_UR, barang2_DL, barang2_UL,
barang1l_DR, barang1l_UR, barang1l_DL, barang1l_UL,
barang1r_DR, barang1r_UR, barang1r_DL, barang1r_UL,
diam_ND
```

Sedangkan semua gestur pada dasarnya akan memberikan sentimen positif dalam level keamanan (gestur dikenal positif), kecuali gestur-gestur tertentu (gestur dikenal negatif) serta gestur yang tidak dikenal (confidence rendah). Dalam kasus ini, semua kelas gestur yang memiliki sebutan "Barang" dikelompokkan sebagai gestur dikenal negatif.

B. Augmentasi Data

Salah satu cara untuk memperbanyak variasi data dalam melatih sistem pengenalan gestur adalah dengan menambah dengan data yang sama namun telah ditransformasi. Transformasi yang dilakukan perlu memiliki tingkat validitas yang tinggi agar tidak malah menambah noise, sehingga perubahan yang diberikan tidak bisa terlalu besar. Tiga operasi transformasi dilakukan untuk keperluan ini yaitu pergeseran, perubahan ukuran, serta rotasi. Semua transformasi tersebut diterapkan pada satu kesatuan pose data training.

Selain itu, sebagian besar data gestur untuk latihan akan memiliki irisan (overlap) pose dengan gestur lainnya yang direkam pada saat yang sama. Sehingga bisa diperoleh lebih banyak potongan data gestur dari satu rentetan pengambilan data yang sama.



Gambar III.3 Irisan gestur-gestur dari satu perekaman data.

Sedangkan metode terakhir merupakan transformasi pada setiap titik dalam pose, dengan jenis pergeseran. Hal ini ditujukan untuk mengurangi kemungkinan sistem terlalu terfokus pada pergerakan satu titik ke titik lainnya yang terlalu ketat. Sehingga setiap gestur akan memiliki variasi lebih dengan nilai awal dan akhir titik gerak yang berbeda-beda pula.

C. Skema Preprocessing Gestur

Data pose ataupun gestur yang diperoleh dari estimasi pose dapat melewati berbagai skema pemrosesan lebih lanjut sebelum diberikan ke dalam sistem pengenalan gestur. Mula-mula untuk membedakan apakah pose yang baru diterima merupakan kelanjutan dari pose-pose sebelumnya dan dikelompokkan menjadi sebuah gestur, sistem perlu memiliki kemampuan untuk tracking. Fungsionalitas ini diimplementasikan dengan mengubah setiap pose menjadi representasi titik dengan dirata-rata, kemudian dilewatkan sistem klasifikasi k-Nearest Neighbor untuk dipasangkan dengan pose-pose sebelumnya. Gestur-gestur yang diperoleh akan dilewatkan pemrosesan lebih lanjut.

Di lapisan pertama preprocessing gestur, ada beberapa opsi yang dibuat (digunakan salah satu atau tidak sama sekali):

1) Amplify: Pose yang terletak pada sisi kamera yang berbeda digeser lebih jauh seraya ada jarak yang jauh antar gambar, memperkuat perbedaan (komponen offset) antara gestur yang terekam oleh masing-masing kamera.

2) Normalize: Setiap pose dinormalisasikan lokasinya ke pusat koordinat.

3) Normalize Once: Setiap gestur dinormalisasikan lokasinya ke pusat koordinat (pose pertama akan menentukan pergeseran pose-pose setelahnya dalam satu gestur).

4) Normalize Point: Setiap poin dalam gestur dinormalisasikan lokasinya ke pusat koordinat (poin pada pose pertama akan menentukan pergeseran poin pada pose-pose setelahnya dalam satu gestur).

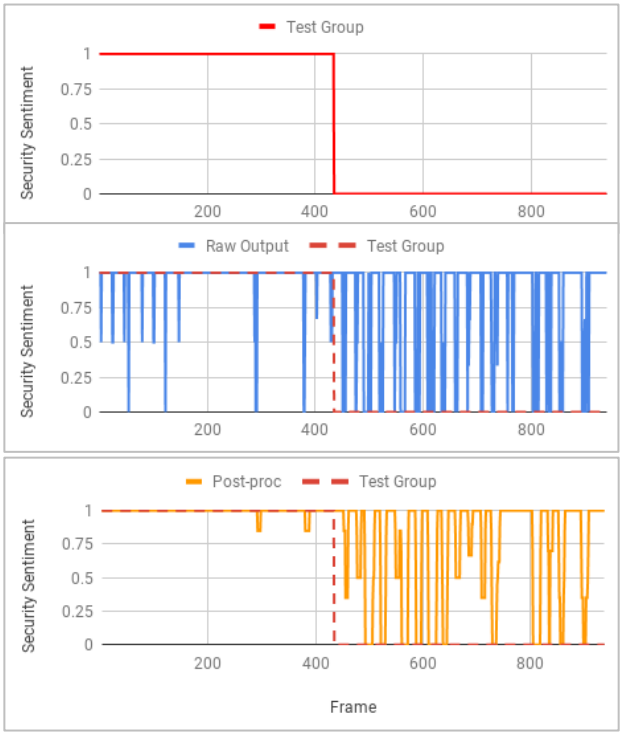
5) Reverse: Kebalikan dari skema Amplify, data dari empat kamera digeser seraya diambil dari hanya satu sisi kamera.

Pada lapisan kedua, ada opsi untuk menggunakan kompensasi dalam mendeteksi gestur diam. Gestur dengan komponen pose yang tidak bergerak melebihi ambang batas akan dipaksakan untuk menjadi totalitas diam.

D. Postprocessing Pengenalan Gestur

Hasil dari pengenalan gestur akan melewati pemrosesan lebih lanjut untuk dijadikan nilai level keamanan. Metode yang digunakan di sini hanyalah sekedar mengumpulkan hasil-hasil pengenalan, seperti seberapa banyak hasil yang memberikan sentimen negatif, positif, ataupun netral. Lapisan pengujian ini ada dua, yaitu pengujian tiap frame serta pengujian historis yang mana mengakumulasikan hasil-hasil sekumpulan frame sebelumnya.

Dalam menentukan metode spesifik atau rumusan yang paling tepat, pengujian terhadap hasil akhir pengenalan gestur akan dilakukan. Sekelompok kegiatan diuji dalam suatu ruangan dan diambil keluaran mentah sistem pengenalan gestur. Hasil yang sudah dilewatkan dengan suatu metode akhir kemudian dibandingkan dengan nilai keamanan yang seharusnya diberikan oleh sistem. Gambar III.4 memvisualisasikan hasil pada tahapan ini dalam grafik.



Gambar III.4 Grafik pengujian metode post-processing terhadap keluaran mentah pengenalan gestur dari kelompok kegiatan positif dan negatif (garis merah, Test Group).

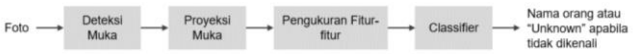
IV. Pengenalan Gestur dengan LSTM

Network yang digunakan didasarkan pada hasil modifikasi LSTM Guillaume Chevalier oleh Stuarteiffert. Sederhananya network tersebut berisikan dua hidden layer dengan masing-masing layer memiliki jumlah node sebanding dengan jumlah input, yaitu 36 dalam kasus ini (18 koordinat poin pose). Fitur lainnya antara lain menerapkan metode regularisasi L2, Decaying Learning Rate, dan Adam Optimizer [2]. Variasi lainnya dibuat dengan mengurangi hidden layer menjadi satu. Jumlah pose dalam satu gestur adalah 5, sebanding dengan gestur yang terjadi selama 1,2-1,6 detik untuk kecepatan sistem di rentang 3-4 FPS.

V. Pengenalan Muka

Dalam mengenali wajah, sistem ini memerlukan sebuah model pre-trained. Sedangkan untuk suatu muka yang diinginkan, minimal hanya satu buah foto dibutuhkan. Sederhananya sistem ini telah dilatih untuk membedakan muka, bukan mengenali suatu muka tertentu secara spesifik. Sehingga dalam menambahkan orang yang ingin dikenali, sistem tidak perlu melakukan training lagi. Hanya suatu foto sehingga parameter-parameter dari foto tersebut dapat diambil dan dibandingkan dengan gambar masukan.

Sistem pengenalan muka pada dasarnya memiliki diagram blok sebagai berikut, yang mana dirancang oleh Ageitgey [3].



Gambar V.1 Diagram blok sistem pengenalan muka.

Deteksi muka diterapkan menggunakan HOG atau *histogram of oriented gradients*, di mana suatu gambar bisa diubah menjadi representasi lain yang menangkap stuktur dasar suatu muka secara sederhana dan membandingkannya pada pola HOG yang diperoleh dari beragam muka hasil training [4]. Sedangkan proyeksi muka memanfaatkan estimasi landmark yang diimplementasikan dengan pendekatan oleh Vahid Kazemi dan Josephine Sullivan [5]. Dengan landmark, area-area penunjuk penting suatu muka bisa ditandai dan diproyeksikan. Sehingga diperoleh bentukan netral untuk semua muka yg dideteksi.

Pengukuran fitur-fitur muka dilakukan untuk memperoleh representasi numerik berdasarkan fitur-fitur muka. Nilai-nilai ini diperoleh melalui sistem OpenFace, di mana sistem tersebut dapat mengeluarkan nilai-nilai yang serupa antara foto-foto muka dari orang yang sama [6]. Dari nilai-nilai tersebut, sistem klasifikasi sederhana menentukan apakah ada orang pada foto input yang serupa dengan orang yang ada pada foto referensi. Keluarannya adalah nama orang tersebut jika iya, dan "Unknown" jika tidak dikenali.

Tahapan pemanfaatan informasi dari sistem pengenalan muka dalam perilaku sistem keseluruhan adalah sebagai berikut:

1. Orang dikenal masuk ruangan, mode “Authorized” aktif.
2. Mode tersebut aktif sampai batasan tertentu. Opsi:
   a. Suatu durasi konstan.
   b. Orang tersebut keluar ruangan. Diperiksa dengan:
      i. Memperoleh warna pakaian orang yang dikenal ketika masuk.
      ii. Menguji setiap warna pakaian orang yang keluar ruangan.
      iii. Jika ada yang serupa, orang tersebut dihapus dari daftar.
3. Kembali ke mode semula atau netral.

Di mana mode Authorized berarti sistem akan mengabaikan kegiatan-kegiatan yang seharusnya menurunkan level keamanan, sebab sistem mengetahui adanya orang yang berwewenang ada di dalam ruangan. Di lain sisi, memanfaatkan warna pakaian tentunya memiliki berbagai kekurangan seperti tidak dapat membedakan antara orang yang keluar dengan warna pakaian sama, orang yang dikenal keluar dengan pakaian berbeda ketika masuk, dan lain-lain. Namun implementasi ini dianggap sudah cukup menyelesaikan permasalahan sistem dalam membedakan kapan sistem mengaktifkan alarm dan kapan tidak.

VI. Hasil

Secara garis besar, pengujian sistem di sini akan dibagi atas dua tipe:

- Pengujian pengenalan gestur terhadap data training, namun dengan variasi augmentasi lain. Data dimodifikasi dengan parameter berupa nilai random yang berbeda dengan ketika data digunakan untuk training (data berbentuk representasi gestur & pose dalam angka).

- Pengujian sistem pengenalan gestur maupun pengukur level keamanan terhadap data nyata, dalam bentuk video yang dijadikan input sistem utama (tentunya dengan karakteristik dibuat serupa dengan input kamera).

Sedangkan pada pengujian jenis kedua, akan ada dua macam hasil utama yang diperoleh:

- Akurasi pengenalan, performa pengenalan gestur dalam mengenal setiap kelompok gestur dengan benar setiap framenya (misal Diam, Jalan, Barang2,

Barang1l, dan Barang1r dianggap berbeda; namun Jalan_DR, Jalan_UL, dan sejenisnya dianggap sama-sama Jalan).

- Akurasi level keamanan, performa keseluruhan sistem utama dalam memberikan trigger keamanan yang tepat dalam jangka waktu yang ditentukan setelah suatu kegiatan mencurigakan telah terjadi.

Melalui definisi dan standar pengujian di atas, berbagai macam pengujian dilakukan untuk segmen-segmen di dalam sistem pengenalan gestur dan penentu level keamanan.

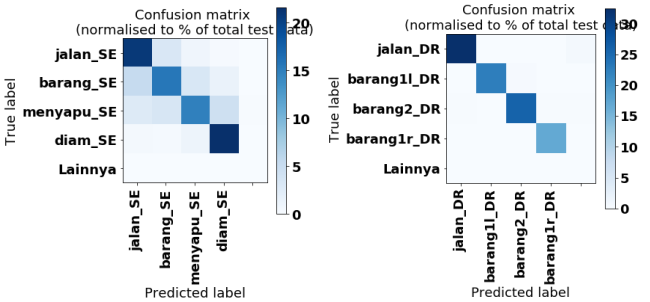
A. *Pengelompokan Kelas-Kelas Gestur*

Perbandingan antara performa skema pengelompokan berdasarkan kamera terhadap berdasarkan gerakan besar dapat dilihat di Tabel I. Bisa dilihat juga Confusion Matrix pada Gambar VI.1, dengan skema kamera memiliki kemampuan lebih rendah dalam membedakan gestur dengan kelompok kelas yang sama (NE, NW, SE, SW). Data di sini merupakan hasil pengujian tipe pertama, yaitu terhadap varian augmentasi lain.

Data tersebut memiliki arti bahwa skema kamera memiliki kemampuan yang jauh lebih rendah dalam meng-klasifikasikan suatu gestur secara spesifik jika dibandingkan skema gerakan, padahal variasi di sini baru sebatas perbedaan ukuran maupun pergeseran pose. Karakteristik ini sudah diduga melalui pemaparan komponen dari masing-masing skema. Meski komponen offset pada skema kamera sudah berkurang, masih ada offset yang beragam dalam satu kelasnya. Sedangkan skema gerakan memiliki komponen offset yang benar-benar hilang dalam satu kelasnya, ditambah lagi komponen gerakan besar yang serupa. Sehingga model pada skema ini hanya perlu fokus ke dalam generalisasi pada komponen gerakan kecil, sedikit ke gerakan besar, dan tidak sama sekali ke komponen offset.

TABEL I. PENGUJIAN PENGELOMPOKAN KELAS GESTUR

Skema Data	Akurasi vs. Varian Augmentasi Lain	
	Total	Kelompok Kelas Terburuk
Kamera	81,02%	72,84% (Kelompok kelas SW)
Gerakan	98,10%	87,40% (Kelompok kelas DR)



Gambar VI.1 Confusion Matrix untuk kelompok terburuk skema Kamera dan Gerakan berturut-turut dari kiri.

B. *Augmentasi*

Pengujian dilakukan terhadap data tanpa augmentasi, data dengan irisan pada gestur-gesturnya, data dengan irisan serta transformasi pose, dan semua metode beserta transformasi poin. Data merupakan hasil pengujian tipe kedua (input nyata ke dalam sistem utama). Ada dua hasil dalam pengujian tipe kedua sebagaimana dijelaskan sebelumnya, yaitu akurasi pengenalan kelompok gestur dan akurasi level keamanan (spesifiknya akan dibahas di subbab VI.D) yang bisa dilihat pada Tabel II.

TABEL II. PENGUJIAN DENGAN DAN TANPA AUGMENTASI

Jenis Augmentasi	Akurasi Pengenalan	Akurasi Level
Tanpa Augmentasi	52,24%	77,11%
Overlap	56,05%	91,78%
Tranformasi pose, & Overlap	58,07%	93,75%
Trans pose, Overlap, & Trans poin	59,42%	87,53%

Dari informasi yang diperoleh, data training dengan augmentasi transformasi pose dan overlap memiliki performa akhir yang terbaik. Namun dalam mengenal suatu kelompok gestur secara spesifik, data dengan tambahan transformasi point memiliki performa terbaik. Secara keseluruhan, tidak ada perbedaan yang signifikan dalam akurasi pengenalan, selama data melewati proses augmentasi.

Sedangkan tidak adanya korelasi antara akurasi pengenalan dan akurasi level atau akhir kembali kepada karakteristik dari setiap model beserta skema postprocessing pengenalan gestur. Karakteristik ini bisa dilihat pada Gambar III.4, sederhananya suatu model bisa salah dalam mengenal suatu gestur namun sentimen akhir akan bertitik berat mayoritas beberapa hasil sebelumnya. Model akan memiliki akurasi level yang buruk apabila salah dalam mengenali gestur secara berturut-turut dalam suatu kegiatan uji, sedangkan model yang selalu salah namun hanya sesekali saja dalam suatu kegiatan uji akan memiliki performa akhir yang lebih baik.

C. *Skema Preprocessing Gestur*

Dengan pengujian tipe kedua, serupa dengan subbab sebelumnya, diperoleh hasil performa skema preprocessing pada Tabel III.

TABEL III. PENGUJIAN SKEMA PREPROCESSING GESTUR

Skema Preprocessing	Akurasi Pengenalan	Akurasi Level
Reverse	32,29%	63,44%
Normalize	58,97%	92,62%
Normalize Once	58,07%	93,75%
Normalize Point	40,36%	57,92%

Sehingga pilihan terbaik adalah antara skema preprocessing Normalize beserta Normalize Once. Tren performa yang sama juga muncul dengan kombinasi konfigurasi lainnya.

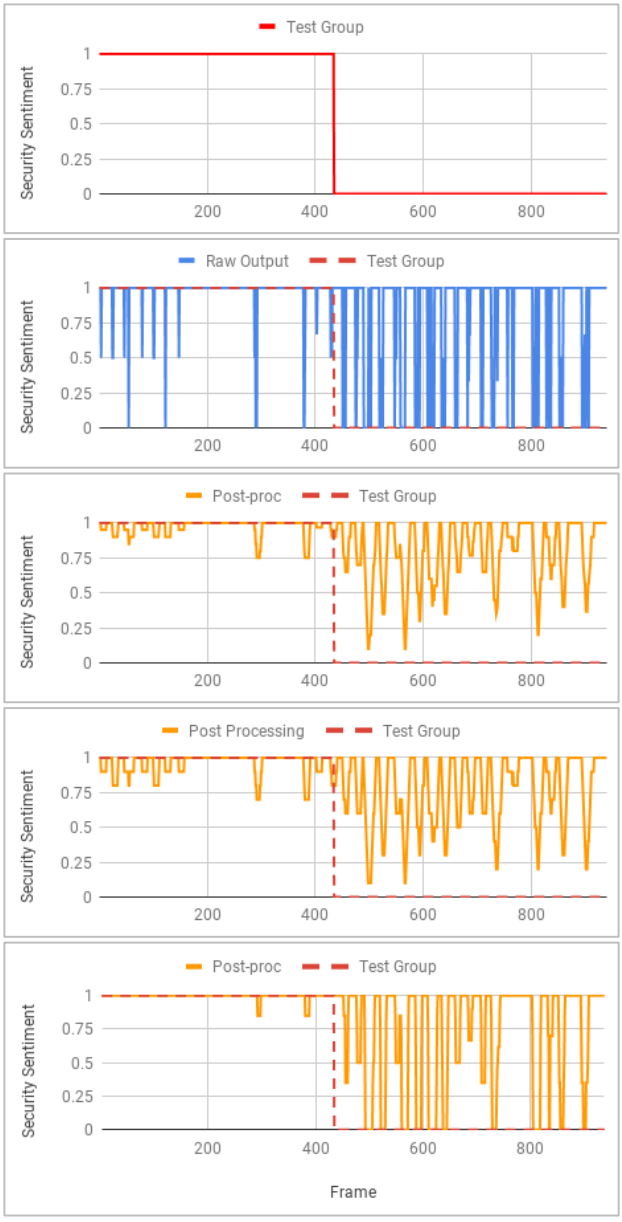
D. *Skema Postprocessing Pengenalan Gestur*

Dari analisis terhadap hasil keluaran pengujian tipe kedua, ada beberapa proses matematis yang dipilih untuk menentukan hasil akhir berupa level keamanan. Sebagaimana disebutkan sebelumnya, proses akhir ini memiliki dua lapisan. Untuk lapisan pertama, yaitu penilai hasil-hasil dalam satu frame input, dipilih proses sederhana berupa perhitungan persentase total hasil yang memberikan sentimen positif. Di sini nilai confidence dari suatu hasil pengenalan gestur juga diuji, yang mana akan memberikan sentimen negatif apabila rendah (tidak peduli gestur tersebut apa).

Pada lapisan kedua, yaitu penilai sekumpulan hasil dari lapisan pertama, diuji berbagai proses matematis dengan karakteristik yang beragam. Sebelum itu, lapisan ini diatur untuk menggunakan konfigurasi historis 10 frame. Nilai ini

sebanding dengan 2,4-3,2 detik untuk kecepatan sistem di rentang 3-4 FPS. Sedangkan operasi matematis yang diuji antara lain persentil, rata-rata, serta menghitung persentase nilai frame yang melewati ambang batas (Count if). Kecuali operasi rata-rata, operasi-operasi tersebut menggunakan nilai parameter yang bisa diubah-ubah untuk memberikan akurasi akhir terbaik. Nilai akhir yang keluar dari lapisan ini kemudian dibandingkan juga terhadap nilai ambang batas akhir, untuk memicu peringatan apabila nilai tersebut terlalu rendah.

Hasil dari masing-masing pengujian operasi tersebut bisa dilihat pada Tabel IV. Sedangkan representasi visual dari proses ini bisa dilihat pada Gambar VI.2.



Gambar VI.2 Representasi visual tiga postprocessing pengenalan gestur (tiga terbawah) terhadap data mentah. Berturut-turut operasi rata-rata, count if, serta persentil.

Dari visualisasi pada Gambar VI.2, bisa dilihat karakteristik dari masing-masing operasi yang diuji. Operasi rata-rata memiliki sifat yang linear dan memiliki karakteristik suatu filter Low-Pass, sehingga mampu menahan false recognition namun sekumpulan impuls atau sinyal kotak dari pengenalan gestur tidak bisa direpresentasikan dengan baik. Karakteristik serupa bisa dilihat juga pada operasi Count if, meskipun sebenarnya operasi tersebut non-linear. Sedangkan operasi persentil sanggup memberikan respons yang sangat representatif, namun tetap memiliki kemampuan untuk menahan false recognition. Hal ini dikarenakan operasi persentil juga mempertimbangkan populasi atau distribusi nilai-nilai masukannya.

TABEL IV. PENGUJIAN SKEMA POSTPROCESSING

Operasi Matematis	Akurasi Pengenalan	Akurasi Level
Rata-rata	58,07%	90,63%
Count if	58,07%	92,50%
Percentil	58,07%	93,75%

E. Kompleksitas Model

Tabel V merupakan hasil pengujian tipe pertama terhadap network LSTM yang semula dengan dua hidden layer terhadap yang hanya satu. Bisa dilihat bahwa model modifikasi bekerja sedikit lebih baik untuk segala aspek, dalam skenario ini. Tabel tersebut menunjukkan salah satu hasil dari sekian banyak pengujian (variasi terhadap skema data) yang memberikan hasil serupa. Belum lagi faktor bahwa model yang lebih sederhana memiliki beban pemrosesan yang lebih ringan.

TABEL V. PENGUJIAN TIPE SATU UNTUK MODEL

H. Layer	Akurasi	Loss	Waktu Training
1	98,10%	0,179	1028 detik
2	98,08%	0,187	1248 detik

Sedangkan Tabel VI menunjukkan hasil pengujian tipe kedua. Bisa dilihat bahwa model modifikasi bekerja jauh lebih baik. Hal ini memperkuat konklusi bahwa untuk skenario maupun konfigurasi sistem yang digunakan, model dengan kompleksitas yang lebih tinggi akan memiliki permasalahan overfitting yang tinggi. Sebab dalam pengujian tipe satu, perbedaan data tidaklah jauh dengan data training. Sedangkan pengujian tipe kedua memberikan sistem nilai nyata, yang benar-benar tidak ada dalam data training.

TABEL VI. PENGUJIAN TIPE DUA UNTUK MODEL

H. Layer	Akurasi Pengenalan	Akurasi Level
1	58,07%	93,75%
2	55,16%	83,44%

F. Model Terbaik

Melalui pengujian-pengujian tersebut, diperoleh model dengan performa terbaik. Konfigurasi lengkap dari model tersebut adalah:

- Model LSTM satu hidden layer.
- Pengelompokan kelas-kelas gestur berdasarkan gerakan besar.
- Augmentasi data: Overlap & transformasi pose.
- Preprocessing gestur skema Normalize Once.
- Postprocessing pengenalan gestur skema percentile (30 persentil dan threshold level keamanan 0,8).

Diperoleh performa dari model ini pada Tabel VII dengan skenario uji berisikan kegiatan-kegiatan yang seharusnya dikenal oleh sistem. Perlu diingat bahwa akurasi level keamanan merupakan hasil perbandingan nilai historis terhadap threshold, dengan pengujian nilai tersebut ada pada Tabel VIII.

TABEL VII. CONFUSION MATRIX & AKURASI

A\P	P	N
P	160	0
N	20	140

Sampel	320
Acc Pengenalan Spesifik	59.82%
Acc Pengenalan Sentimen	76.51%
Acc Level Keamanan	93.75%

P dan N berturut-turut Positif dan Negatif. Positif adalah kegiatan yang tidak mencurigakan, sebaliknya untuk negatif.
*Acc spesifik = Mengenali kelompok gestur dengan benar.
*Acc sentimen = Mengenali sentimen dengan benar.

TABEL VIII. PENGUJIAN NILAI THRESHOLD TERHADAP AKURASI

Threshold	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
Ak. Level	82	82	82	82	82	82	86	86	86	91	91	91	93	94	94	94	86	86

Sedangkan confusion matrix pengenalan spesifik bisa dilihat pada Tabel IX, yang mana menjadi dasar terhadap perhitungan akurasi pengenalan spesifik dan pengenalan sentimen pada Tabel VII. Dari situ, bisa dilihat bahwa sebenarnya pengenalan tiap gestur sendiri bisa memiliki akurasi yang cukup rendah, dengan 14% paling minimum pada "Barang1l". Namun kesalahan dalam pengenalan spesifik tersebut tidak seberapa signifkan efeknya, karena secara umum jatuh pada kelas dengan sentimen-sentimen yang sama (yaitu tetap sama-sama "barang"). Nilai ini hanya menunjukkan seberapa unikkah gestur-gestur spesifik yang diberikan ke dalam sistem.

TABEL IX. CONFUSION MATRIX PENGENALAN SPESIFIK

A \ P	Diam	Jalan	Barang2	Barang1r	Barang1l
Diam	0	0	0	0	0
Jalan	3	185	2	4	14
Barang2	0	28	49	11	35
Barang1r	0	42	5	14	18
Barang1l	0	11	0	1	11
Acc (%)	0.0	69.5	87.5	46.7	14.1

G. *Pengenalan Muka*

Pengujian secara sistematis dilakukan dengan cara menyiapkan video yang berisikan berbagai konfigurasi. Mulai dari masing-masing subjek menghadap kamera sendirian secara bergantian, beberapa subjek bersamaan dalam satu kamera, mengubah referensi foto subjek yang ingin dikenali (satu-satu atau kombinasi), serta memvariasikan sudut serta jarak muka subjek. Hasil yang diperoleh ada pada Tabel X.

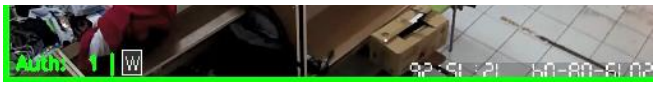
TABEL X. PENGUJIAN PENGENALAN MUKA

Konfigurasi		Benar	Salah Kenal	Salah Tidak Dikenal
Toleransi 0.6	Total	496	93	7
	Persen	83.22%	15.60%	1.17%
Toleransi 0.4	Total	539	0	57
	Persen	90.44%	0.00%	9.56%

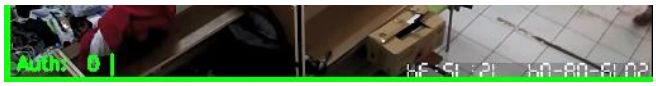
Visualisasi pemanfaatan dari implementasi pengenalan muka bisa dilihat pada Gambar II.1 yang sudah ditunjukkan sebelumnya. Dalam gambar tersebut, orang dikenal masuk ruangan sehingga mode Authorized aktif dan ditandai dengan bingkai hijau pada gambar akhir sistem (tidak berbingkai jika netral, dan merah apabila kegiatan negatif dideteksi tanpa ada orang dikenal dalam ruangan). Orang tersebut kemudian melakukan kegiatan negatif berupa membawa barang, sehingga menurunkan level keamanan. Namun mode Authorized berarti alarm tidak aktif hingga akhirnya orang tersebut keluar ruangan.

Gambar VI.3 Pengujian pengenalan muka.

Sedangkan catatan sistem terhadap orang-orang dikenal beserta warna pakaiannya bisa dilihat pada sisi pojok kiri bawah gambar akhir tersebut. Bagian ini bisa dilihat pada Gambar VI.4 dengan (a) dan (b) berturut-turut pembesaran Gambar II.1 (a) dan (b).



(a)



(b)

Gambar VI.4 Sebagian visualisasi output sistem yang menghitung dan menunjukkan inisial orang-orang yang dikenal serta warna pakaiannya. (a) Orang dikenal (William) masuk ruangan. (b) Orang dikenal keluar ruangan.

H. *Gestur Tidak Dikenal*

Salah satu limitasi utama dalam implementasi yang dilakukan adalah dalam mendeteksi gestur yang belum pernah dikenali sebelumnya. Metode ini membutuhkan sistem yang dapat mempelajari suatu data yang ada dan membuat region yang secukupnya, atau yang disebut dengan pengenalan Open-set. Sedangkan sistem pengklasifikasi tradisional tanpa metode tersebut akan membentuk region yang menuju tak terhingga selama tidak ada kelas berlawanan yang menghinggapinya, sehingga tidak ada definisi "tidak dikenal" yang pasti selain nilai confidence yang rendah. Kelemahan dalam menggunakan nilai confidence saja adalah nilai tersebut bukan nilai yang definitif dan non-linear sifatnya [7].

Pengujian sebelumnya hanyalah terhadap video yang berisikan peraga melakukan kegiatan yang seharusnya dikenal oleh sistem. Sedangkan untuk melengkapi kinerja sistem dalam mendeteksi perilaku mencurigakan, perlu diuji kemampuan sistem dalam mengenali gestur yang sebelumnya belum pernah dilatih. Sehingga data hasil sebelumnya disambung dengan data hasil sistem ketika diberi input video penuh berisikan kegiatan atau gestur yang benar-benar berbeda (contoh, merangkak, membungkuk kesakitan, terjatuh, terpapar, dan lain-lain). Sehingga confidence pada saat-saat tersebut harus di bawah ambang batas untuk dianggap pengenalan gestur yang benar, sebaliknya untuk gestur yang seharusnya dikenal.

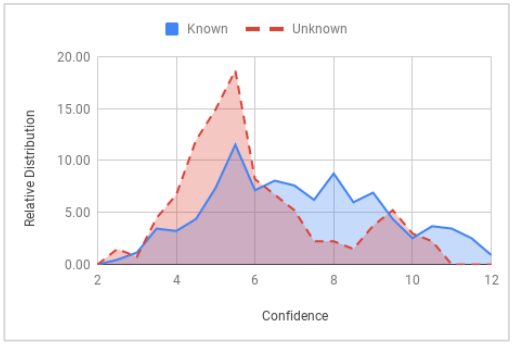
Hasil pengujian tersebut dapat dilihat pada Tabel XI. Bisa dilihat bahwa akurasi pengenalan gestur menurun. Melihat detail data lebih lanjut menunjukkan bahwa nilai level keamanan dalam kasus ini memang tidak memburuk, namun hal tersebut tidak lain karena gestur keluarannya memang memiliki sentimen yang sama-sama negatif.

Bahkan ada lebih banyak gestur-gestur yang seharusnya dikenali, menjadi tidak dikenali akibat adanya pengujian nilai confidence. Sehingga dalam kasus tersebut sistem akan memiliki performa lebih baik apabila tidak mengklasifikasi-kan suatu gestur input sebagai tidak dikenal. Hal ini menunjukkan kelemahan dalam sistem klasifikasi Multi-class.

TABEL XI. PENGUJIAN GESTUR TIDAK DIKENAL

Pengujian tipe kedua terhadap	Akurasi Pengenalan
Gestur yang seharusnya dikenal saja	59,4%
Gestur yang seharusnya dikenal dan tidak	45,6%

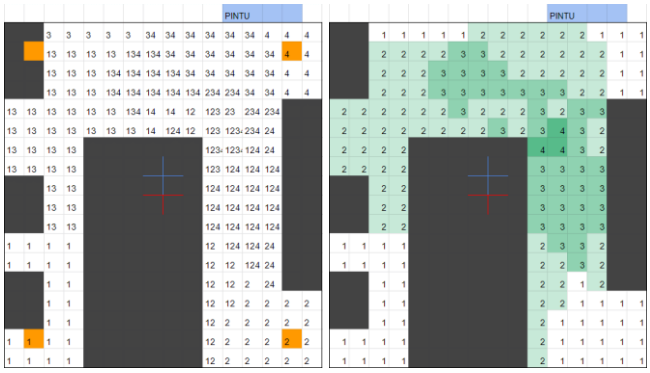
Hal ini semakin jelas ketika melihat perbandingan distribusi nilai confidence antara kegiatan yang seharusnya dikenal dan yang seharusnya tidak dikenal pada Gambar VI.5. Tidak ada pemisah yang jelas antara keduanya, meskipun memang kegiatan yang seharusnya tidak dikenal umumnya bernilai rendah.



Gambar VI.5 Distribusi nilai confidence hasil pengujian kegiatan yang seharusnya dikenal dan yang seharusnya tidak dikenal.

I. *Cakupan dan Reliabilitas Sistem Kamera*

Area uji sistem berukuran 4,8x5,4 m^2. Satu kotak menandakan area 30x30 cm^2. Gambar VI.6 menunjukkan area cakupan dari 4 kamera utama, bisa dilihat bahwa ada irisan antara kamera yang satu dengan yang lainnya. Irisan ini selain memperkaya informasi yang diperoleh sistem, juga dapat berlaku sebagai faktor redundansi.



Gambar VI.6 Hasil pengujian cakupan area dari tiap kamera. Kotak abu-abu adalah perabotan dan kotak oranye adalah keempat kamera (kamera 1,2,3, dan 4 urut dari kanan-atas, kiri-atas, kanan-bawah, dan kiri-bawah). (Kiri) Kamera mana saja yang bisa mencakup petak tersebut. (Kanan) Total kamera yang mencakup area tersebut.

Pengujian menunjukkan informasi luas area dari masing-masing kamera pada Tabel XII. Sedangkan tingkatan reliabilitas yang ditimbulkan karena adanya redundansi atau irisan area cakupan kamera bisa dilihat pada Tabel XIII. Dari tabel tersebut didapatkan info bahwa ketika salah satu kamera mati, sistem tetap bisa mencakup seburuk-buruknya 88% dari kondisi sebelumnya.

TABEL XII. PENGUJIAN LUAS AREA CAKUPAN KAMERA.

Cakupan dari	Luas (petak)	Luas (m^2)	Keterangan
Keseluruhan tanpa prabotan	176	15.84	61% Keseluruhan ruangan
Kamera 1 (kanan-atas)	112	10.08	64% Ruang tanpa prabotan
Kamera 2 (kiri-atas)	69	6.21	39% Ruang tanpa prabotan
Kamera 3 (kanan-bawah)	85	7.65	48% Ruang tanpa prabotan
Kamera 4 (kiri-bawah)	76	6.84	43% Ruang tanpa prabotan

TABEL XIII. PENGUJIAN LUAS AREA CAKUPAN KAMERA DALAM KASUS SALAH SATU KAMERA MATI.

Cakupan apabila	Luas (petak)	Luas (m^2)	Keterangan
Kamera 1 mati	154	13.86	88% Ruang tanpa prabotan
Kamera 2 mati	156	14.04	89% Ruang tanpa prabotan
Kamera 3 mati	171	15.39	97% Ruang tanpa prabotan
Kamera 4 mati	167	15.03	95% Ruang tanpa prabotan

J. *Komputer*

Melalui pengujian dengan simplifikasi sistem, yaitu dengan resolusi 432x368 saja dan hanya model OpenPose yang aktif (karena yang paling intensif pemrosesannya), bisa diperoleh beberapa informasi utama. Pengujian program hanya dengan CPU menunjukkan kecepatan di kisaran 1 FPS saja pada tiga perangkat dengan jumlah core empat. Sedangkan pengujian dengan memanfaatkan GPU Nvidia (Compute Capability di 3,5 ke atas) menunjukkan FPS rata-rata 6,4, 11, dan 18 dari tiga perangkat dengan berturut-turut jumlah core 384, 640, dan 1280.

Kecepatan pemrosesan pada saat implementasi total yang dibutuhkan seminimalnya adalah 3 FPS agar performa hasil sistem serupa dengan skenario uji. Dari pengujian di atas, bisa disimpulkan bahwa perangkat keras perlu memiliki GPU Nvidia untuk setidaknya bisa menjalankan sistem yang tersimplifikasi di atas batas tersebut. Alasan di balik ini adalah semakin banyak proses sederhana di dalam sistem yang bisa dijalankan secara paralel dengan memanfaatkan jumlah core pada suatu perangkat GPU yang jauh lebih banyak dibanding CPU pada umumnya.

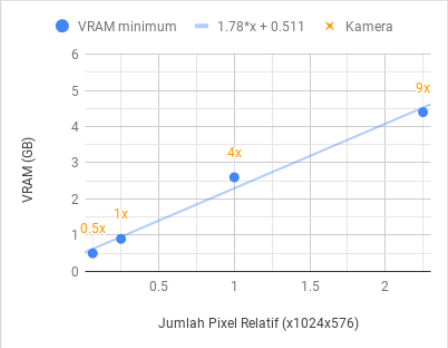
Sedangkan pengujian skalabilitas sistem bisa memberikan informasi VRAM (Video RAM) minimum yang dibutuhkan sistem untuk setiap jumlah kamera yang berbeda. Dalam skenario uji dengan konfigurasi 4 kamera dan total resolusi 512x288 (total 1024x576), VRAM yang dibutuhkan dari subsistemnya adalah sebagai berikut:

- Total = 3,7 GB
- Pose Estimation = 2,5 GB (~70% Total)
- Gesture Recognition = 0,3 GB
- Object Detection = 0,5 GB
- Face Recognition = 0,4 GB

Faktor skalabilitas dalam sistem ini pada dasarnya terletak pada jumlah kamera utama (kamera atas) yang digunakan. Dari paparan penggunaan VRAM di atas, mayoritas beban VRAM terletak pada OpenPose (Pose Estimation). Sedangkan sistem gesture recognition tidak bergantung pada resolusi gambar (inputnya dalam bentuk vektor), dan sistem face recognition bergantung pada kamera muka saja.

Sehingga penambahan satu buah kamera utama berarti menambah jumlah pixel sebesar 512x288 – pembesaran model OpenPose dengan nilai yang sama. Sebaliknya juga dilakukan untuk skenario dengan lebih sedikit kamera

digunakan. Hasil pengujian memberikan grafik pada Gambar VI.7.



Gambar VI.7 Pengujian jumlah kamera terhadap VRAM minimum OpenPose (Pose Estimation). Kamera 0,5x berarti 1 kamera dengan downscale 2x (resolusi 256x144).

Gambar di atas merupakan hasil pengujian kebutuhan VRAM minimum terhadap jumlah kamera utama yang berbeda. Dari grafik tersebut, diperoleh relasi bahwa nilai VRAM minimum OpenPose linear terhadap jumlah pixel gambar (linear terhadap jumlah kamera) dengan persamaan hasil regresi:

$$VRAM_{OpenPose}(GB) = 1{,}78 * (N_{kamera} * 512 * 288) + 0{,}51$$

Menambah dengan kebutuhan sistem lainnya, nilai ini menjadi sebagai berikut.

$$VRAM_{Sistem}(GB) = 1{,}78 * (N_{kamera} * 512 * 288) + 0{,}51 + 1{,}2$$

Perangkat keras dengan VRAM di bawah batas ini umumnya akan menolak untuk menjalankan sistem, atau kalaupun bisa akan memiliki performa jauh lebih buruk dari seharusnya.

VII. Kesimpulan

Sistem secara umum memiliki performa yang sangat baik dalam menentukan tingkatan keamanan dan cukup baik dalam mengenali kelompok gestur. Namun limitasi dalam jenis network atau skema training yang digunakan (Multi-class Classification dan bukan Open-set recognition) menimbulkan sistem tidak memiliki definisi yang kuat untuk mengenali gestur yang tidak dikenal. Dalam skenario uji, sistem memiliki reliabilitas untuk tetap mencakup seminimalnya 88% dari area semula ketika salah satu kamera dari total empat mati, akurasi pengenalan muka sebesar 90,44%, serta akurasi penentuan aman tidaknya ruangan sebesar 93,75%. Sedangkan perangkat keras yang digunakan setidaknya harus mampu menjalankan sistem dengan kecepatan minimum 3 FPS, yang mana pengujian menunjukkan bisa dicapai dengan menggunakan GPU Nvidia dengan Compute Capability minimum 3,5 serta VRAM minimum 4 GB untuk konfigurasi empat kamera.

Referensi

[1] CMU Perceptual Computing Lab, “OpenPose: Real-time multi-person keypoint detection library for body, face, hands, and foot estimation,” [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose. [Diakses 5 7 2019].

[2] Stuarteiffert, “Activity Recognition from 2D pose using an LSTM RNN,” GitHub, 2018. [Online]. Available: https://github.com/stuarteiffert/RNN-for-Human-Activity-Recognition-using-2D-Pose-Input.

[3] A. Geitgey, “Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning,” 2016. [Online]. Available: https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78.

[4] N. Dalal dan B. Triggs, “Histograms of Oriented Gradients for Human Detection,” [Online]. Available: http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf. [Diakses 17 8 2019].

[5] V. Kazemi dan J. Sullivan, “One Millisecond Face Alignment with an Ensemble of Regression Trees,” [Online]. Available: http://www.csc.kth.se/~vahidk/papers/KazemiCVPR14.pdf. [Diakses 17 8 2019].

[6] Carnegie Mellon University, “OpenFace,” GitHub, [Online]. Available: https://github.com/cmusatyalab/openface. [Diakses 17 08 2019].

[7] W. J. Scheirer, A. Rocha, A. Sapkota dan T. E. Boult, “Towards Open Set Recognition,” 2013.