

ANALISIS KINERJA *HAND TRACKING-BY-DETECTION* UNTUK HOLOGRAM INTERAKTIF MENGGUNAKAN MODEL *ADAPTIVE BOOSTING*

PERFORMANCE ANALYSIS OF *HAND TRACKING-BY-DETECTION* FOR INTERACTIVE HOLOGRAM TECHNOLOGY USING *ADAPTIVE BOOSTING*

Muhammad Yanuar Rizqi¹, Suryo Adhi Wibowo, Ph.D.², Desti Madya Saputri, S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik, Universitas Telkom

¹yanuarrizki@student.telkomuniversity.ac.id, ²suryoadhiwibowo@telkomuniversity.ac.id,

³destimadyasaputri@telkomuniversity.ac.id

Abstrak

Human Computer Interaction atau HCI adalah ilmu yang digunakan untuk menghubungkan manusia dengan komputer sehingga dapat berinteraksi secara efisien dengan meliputi perancangan, evaluasi, dan implementasi antarmuka dari komputer agar mudah digunakan. Pada Tugas Akhir ini dibuat sebuah sistem untuk menampilkan informasi tiga dimensi hologram menggunakan media proyeksi kaca mika yang diintegrasikan dengan metode pengenalan objek untuk mendeteksi gerakan tangan sebagai pengganti pointer, contohnya Haar-Cascade Classifier yang merupakan *weak learner* sehingga untuk meningkatkan performa diperlukan algoritma *boosting* (contohnya *Adaptive Boosting*). Penulis membagi permasalahan tersebut ke dalam 4 skenario, dimana skenario 1 yaitu tes akurasi, skenario 2 IoU, skenario 3 presisi dan skenario 4 waktu komputasi yang akan diujikan pada 2 model data uji, yaitu tangan terbuka dan tertutup. Lalu parameter yang akan dioptimalisasi, yaitu jumlah *dataset*, dan nilai *minimum neighbour*. Dari sistem yang dioptimalisasi, penulis mendapatkan hasil terbaik pada akurasi yaitu 99.72% untuk semua model, lalu pada IoU 0.58 untuk tangan terbuka dan 0.71 untuk tangan tertutup, sedangkan pada presisi 40.31 piksel untuk tangan terbuka dan 14.06 piksel untuk tangan tertutup, dan pada waktu komputasi 0.13 sekon untuk tangan terbuka dan 0.17 sekon untuk tangan tertutup.

Kata Kunci: Adaptive Boosting, Haar Cascade, hologram, HCI

Abstract

Human Computer Interaction is the science used for connecting humans with computers, so they can interact interactively efficient by covering the design, evaluation, and implementation of the interface from computer for easy use. In this Final Project, a system is created to display 3D information holographic use projection media into the mica glass who integrated with object recognition method for hand detection where the hand is use as pointer, one of them is the Haar-Cascade Classifier which is a weak learner, so boosting algorithms are needed to improve performance (e.g *Adaptive Boosting*). The writer divides the problem into 4 scenario, where scenario 1 is accuracy test, 2 is IoU, 3 is precision, and 4 is computation time which will be tested on 2 models data (i.e open hand and close hand). Then the parameter to be optimized, i.e. number of dataset and minimum neighbour value. The author get the best result on accuracy, which is 99.7% for all models, then on the IoU 0.58 for open hand and 0.71 for close hand, while precision get 40.31 pixel for open hand and 14.06 pixel for close hand, and on computation time which is 0.13 second for open hand and 0.17 second for close hand.

Keywords: Adaptive Boosting, Haar Cascade, hologram, HCI

1. Pendahuluan

Perkembangan media pertukaran informasi sudah semakin maju dan modern, akan tetapi masih sulit ditemukan media penyampaian informasi yang interaktif dan menarik. Ditunjang dengan berbagai alat seperti penggunaan sensor, layar, dan tiga dimensi hologram, media penyampaian informasi menjadi semakin canggih dan dapat dimanfaatkan untuk menyampaikan informasi ke banyak pengguna yang lebih menarik dan efisien sehingga pengguna dapat berinteraksi secara langsung dengan informasi tersebut yang disebut dengan *Human Computer Interaction* (HCI) [1]. Informasi tiga dimensi hologram dirasa dapat membantu dalam merekayasa informasi secara efisien dan interaktif dalam mengolah informasi yang akan direalisasi dan kembangkan. Selain merancang media untuk menampilkan gambar tiga dimensi, diperlukan juga metode yang dapat mengenali gerakan suatu objek yang dapat dimanfaatkan untuk interaksi antara pengguna dengan informasi yang ditampilkan dari suatu media. Objek

yang dimaksud adalah tangan yang dideteksi pergerakannya sebagai pengganti fungsi pointer pada komputer sehingga tidak ada kontak langsung terhadap sistem [2]. Menggunakan metode Haar-like Feature sebagai implementasi sistem ekstraksi ciri data dan dikombinasikan dengan Adaptive Boosting [3] sebagai metode classifier yang menghasilkan metode bernama Haarcascade Classifier. Sehingga nantinya pengguna dapat berinteraksi dengan cara merekayasa dan merancang informasi yang terdapat pada media seperti menyentuh dan menggeser objek secara langsung. Selain itu juga dilakukan analisis dengan berbagai parameter yang telah ditentukan sebagai tingkat kepercayaan pada sistem yang telah dibuat.

2. Desain dan Perancangan

Pada penelitian ini diperlukan adanya *dataset*. Dataset yang digunakan adalah citra yang memiliki kumpulan dari objek dan atributnya. Atribut merupakan sifat atau karakteristik dari suatu objek. Contohnya gerakan tangan, gerakan jari, dan lain-lain. Atribut juga dikenal sebagai variabel, field, dan karakteristik atau fitur. Kumpulan dari atribut menggambarkan sebuah objek bisa disebut record, titik, kasus, sample, dan entitas atau instance [4]. Citra merupakan matriks n dimensi dimana elemen dari matriks merupakan intensitas cahaya. Kumpulan dari intensitas cahaya tersebut dinyatakan sebagai suatu fungsi kontinu $f(x,y)$ dimana x dan y menyatakan indeks koordinat dari elemen matriks yang bersangkutan [5]. Citra yang digunakan memiliki 2 tipe, yaitu citra positif dan negatif. Citra positif adalah citra yang mengandung objek yang akan dideteksi sebagai dataset. Banyaknya dataset positif membuat fitur cascade yang digunakan juga semakin banyak dan akurat. Citra negatif adalah Citra apapun yang tidak mengandung objek yang digunakan sebagai latar belakang dari objek yang dideteksi. Dataset negatif boleh memiliki ukuran yang bebas yang disesuaikan dengan latar yang akan digunakan [6].

Setelah *dataset* didapat maka akan dilakukan proses ekstraksi fitur. Ekstraksi fitur atau sering disebut juga dengan ekstraksi ciri merupakan suatu teknik yang digunakan oleh komputer untuk mendapatkan fitur dari suatu citra. Fitur ini nantinya akan digunakan oleh komputer untuk membedakan suatu citra dengan citra yang lainnya [7]. Untuk mendapatkan nilai ciri dari citra, pada penelitian ini menggunakan metode Haar-Like Feture yang dikenalkan oleh Viola dan Jones yang merupakan fitur yang memiliki dasar dari fungsi matematika Haar-Wavelet. HaarWavelet adalah gelombang yang memiliki satu interval tinggi dan satu interval rendah (jika dalam 2D berarti terdapat sisi gelap dan cerah) yang dikombinasikan untuk pendeteksian objek visual [8]. Terdapat 3 fitur dasar yang ditampilkan pada gambar 2.1.

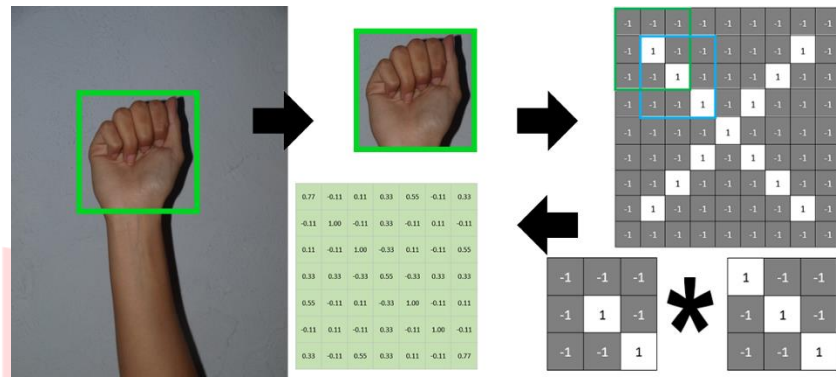


Gambar 2.1: Dasar fitur Haar-Like (kernel)

Fitur akan digunakan pada proses filtering dimana terjadi proses konvolusi antara nilai citra dengan nilai fitur. Jika diketahui matriks filter K , maka akan didapat notasi pada persamaan (1):

$$(I * K)_{r,s} = \sum_{u=-h_1}^{h_1} \sum_{v=-h_2}^{h_2} K_{u,v} I_{r+u,s+v} \tag{1}$$

Gambar 2.2 menjelaskan cara kerja konvolusi. Misalkan bagian dari gambar dipresentasikan dengan nilai-nilai pixel. Lalu nilai-nilai tersebut akan dikonvolusikan dengan matriks kernel dan akan dibuat matriks baru untuk hasil konvolusi tersebut. Proses ini akan dilakukan sampai semua elemen di dalam matriks gambar terkonvolusi. Window dari matriks yang dikonvolusikan akan bergeser setiap satu kolom dan satu baris. Hasil dari konvolusi ini adalah kumpulan matriks yang sudah terfilter [9].

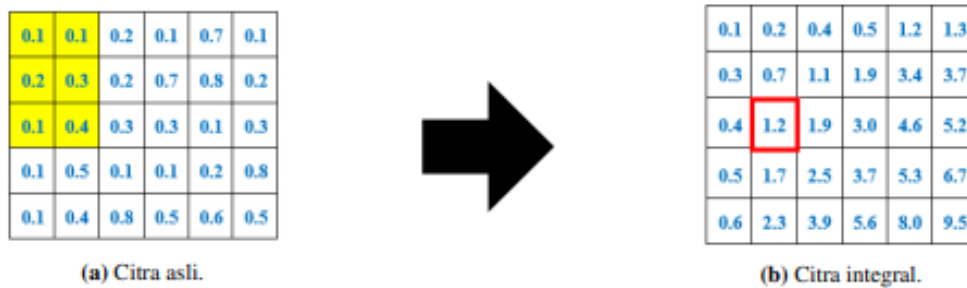


Gambar 2.2: Proses konvolusi

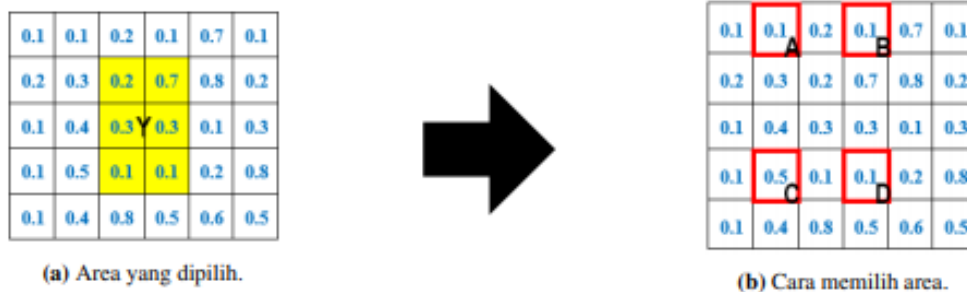
Kemudian nilai pada citra akan di integral untuk mengetahui perbedaan rata-rata nilai piksel (intensitas) pada citra, sehingga dapat ditentukan fitur Haar yang akan digunakan pada tiap piksel yang telah ditentukan [10]. Untuk menghitung integral pada suatu nilai citra dapat menggunakan persamaan (2) dan (3) yang ditunjukkan pada gambar 2.3 dan 2.4.

$$(x, y) = \sum_{x', y'}^{x, y} (x', y') \tag{2}$$

$$Y = A + D - (B + C) \tag{3}$$

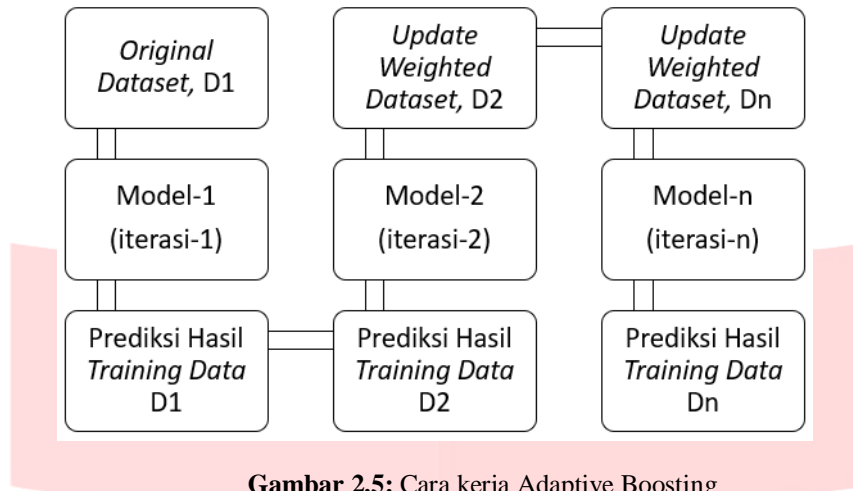


Gambar 2.3: Cara kerja integral image



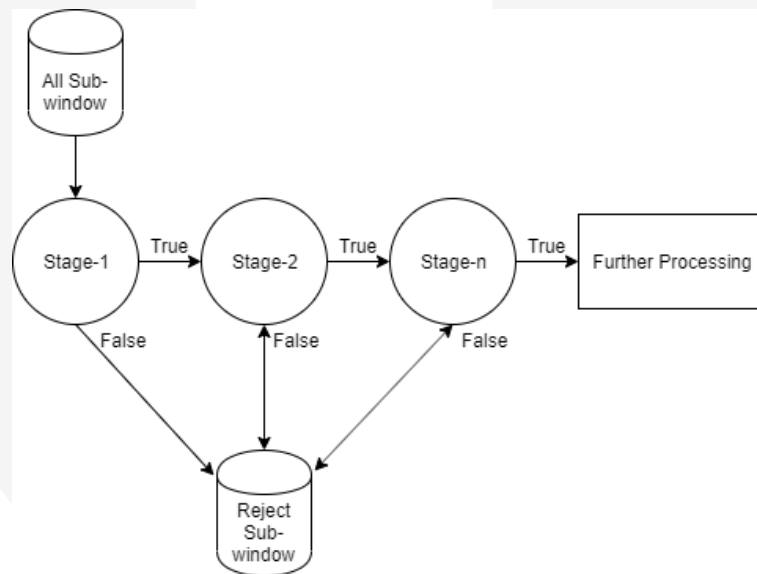
Gambar 2.4: Cara hitung area piksel

Hasil ekstraksi yang didapat akan diolah dalam *classifier* supaya dapat membedakan antara objek dengan latarnya. Metode yang digunakan pada penelitian ini adalah Adaptive Boosting yang merupakan metode meta-algorithm dari varian boosting. Boosting merupakan teknik ensemble learning yang dapat dikombinasikan dengan classifier algoritma lain untuk meningkatkan performa klasifikasi dan mengubah weak learner method menjadi strong learner method dengan tingkat akurasi maksimum [3].



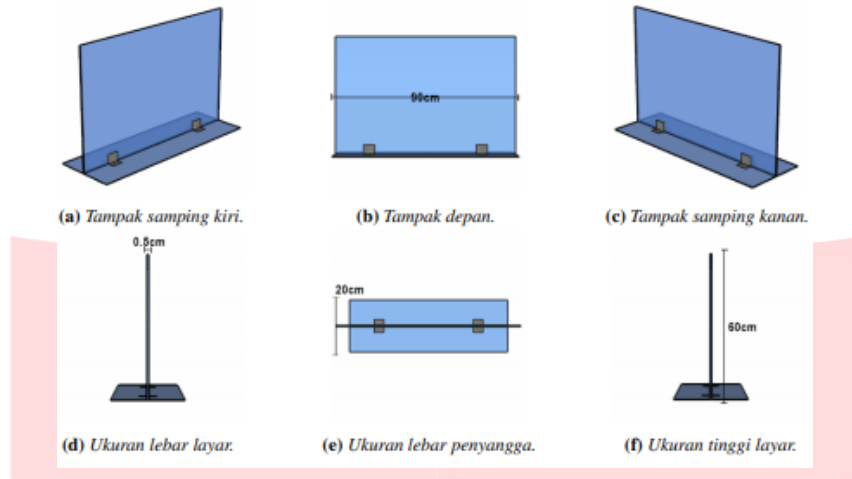
Gambar 2.5: Cara kerja Adaptive Boosting

Pada gambar 2.5 diperlihatkan cara kerja Adaboost dengan menetapkan bobot (weight) ke data latih. Jika terdapat kesalahan pada prediksi maka bobot akan ditambahkan pada pengamatan yang diprediksi salah dan mengurangi bobot dari pengamatan yang diprediksi benar. Lalu ulangi iterasi tersebut hingga jumlah iterasi yang ditentukan dan memperbarui nilai bobot hingga didapatkan akurasi yang maksimum [11]. Penerapan metode ini pada Haar-Like Feature terdapat pada Cascade Classifiernya yang digunakan untuk mempercepat proses deteksi, serta mengurangi tingkat false positif dan mempertahankan presisi. Cascade Classifier bersifat multistage yang berarti setiap hasil prediksi disimpan dan digunakan kembali pada stage berikutnya yang dijelaskan pada gambar 2.6. Setiap satu stage memiliki beberapa window dan fitur klasifikasi. Cascade Classifier akan mengeliminasi kandidat objek yang tidak sesuai sehingga tidak diloloskan pada stage berikutnya [6]. Cascade Classifier bekerja pada tiap citra dengan mengumpulkan informasi pada jumlah window setiap citra [8].



Gambar 2.6: Cara kerja Cascade Classifier

Desain perangkat yang digunakan sesuai pada gambar 2.7. Perangkat keras yang digunakan adalah kaca bening berbentuk persegi berukuran 90x0.5x60(cm) yang digunakan untuk memantulkan cahaya dari proyektor dengan video objek yang memiliki background hitam atau putih polos sehingga terlihat 3D. Pantulan di cermin bidang (kaca mika) menghasilkan gambar virtual untuk pengamat. Gambar ini nampak oleh penonton untuk diposisikan pada jarak yang sama di belakang permukaan cermin saat objek berada di depan [12]. Untuk menampilkan video, proyektor terhubung dengan personal computer (PC) sebagai pusat pemrosesan sistem.



Gambar 2.7: Desain perangkat keras

3. Analisis

3.1. Perancangan Sistem

Perancangan pada penelitian ini terbagi menjadi beberapa tahapan yaitu preprocessing, melatih sistem, dan menguji hasil prediksi dari sistem dari metode [8]. Preprocessing merupakan proses untuk membentuk dataset menjadi data yang terstruktur sesuai dengan kebutuhan. Sistem akan dilatih dengan dataset yang telah digunakan dan menggunakan metode [8] dan algoritma [6] [11]. Terdapat sebuah parameter yang harus diatur dalam proses ini, yaitu jumlah stage. Hasil dari proses ini adalah kumpulan dari bobot dan nilai threshold yang membentuk sebuah file prediksi berformat .xml sebagai prediktor yang nantinya akan diuji. Untuk melakukan prediksi maka hal yang harus dilakukan adalah memasukan dataset uji dan predictor hasil latih sistem atau dalam sistem pengenalan objek disebut dengan classifier [6].

3.2. Skenario Pengujian Sistem

Dalam tugas akhir ini terdapat 4 skenario pengujian, dimana pada setiap skenario akan dilakukan analisis untuk mencari konfigurasi terbaik agar hasil dari skenario tersebut maksimal.

3.2.1. Skenario 1: Pengujian terhadap akurasi

Akurasi merupakan suatu nilai yang mengukur ketepatan sistem dalam mengenali citra atau objek masukan [13]. Akurasi secara matematis dapat dituliskan pada persamaan (4).

$$Akurasi = \frac{Prediksi\ benar}{Total\ jumlah\ data\ uji} \times 100\% \tag{4}$$

Pada skenario ini, prediksi hasil keluaran dari sistem yang telah dilatih dengan dataset latih akan diuji untuk mengenali objek dalam citra yang tidak terdapat dalam dataset latih. Terdapat 700 citra dalam dataset uji ini yang terbagi kedalam 2 kelas, yaitu tangan terbuka dan mengepal. Skenario ini bertujuan untuk mengetahui seberapa baik sistem dalam melakukan pengenalan terhadap citra yang belum dipelajari oleh sistem.

3.2.2. Skenario 2: Pengujian terhadap Intersection over Union (IoU)

IoU adalah evaluasi untuk mengukur keakuratan detektor objek pada dataset tertentu, dan waktu komputasi (seberapa cepat sistem melakukan deteksi). Untuk mengetahui nilai IoU diperlukan nilai inter area, box area ground truth, dan box area uji dengan persamaan (5-8).

$$InterArea = (x_{max} - x_{min}) * (y_{max} - y_{min}) \tag{5}$$

$$BoxArea_{gt} = (gtx_{max} - gtx_{min}) * (gty_{max} - gty_{min}) \tag{6}$$

$$BoxArea_{uji} = (ujix_{max} - ujix_{min}) * (ujiy_{max} - ujiy_{min}) \tag{7}$$

$$IoU = \frac{InterArea}{BoxArea_{gt} - BoxArea_{uji} - InterArea} \tag{8}$$

Sistem akan dievaluasi perbedaan antara nilai ground truth yang ada pada prediktor dengan nilai ground truth pada hasil data uji. Tujuan skenario ini adalah mendapatkan hasil perhitungan prediksi yang cocok dan memiliki persimpangan yang sedekat mungkin antar ground truth yang dimiliki prediktor dengan data uji.

3.3.3. Skenario 3: Pengujian terhadap presisi

Presisi adalah tingkat keberagaman yang terletak pada beberapa nilai pengukuran dari faktor yang sama, selain itu presisi juga menunjukkan kedekatan dari masing-masing pengukuran. Presisi menunjukkan konsistensi dari

pengukuran dari tingkat kesalahan yang acak dan terjadi secara periodik tanpa pola yang dikenali. Semakin kecil nilai maka dianggap semakin presisi. Untuk mencari nilai presisi dapat menggunakan rumus (9-11).

$$CentroidX = \left(\frac{x_{max} - x_{min}}{2} \right) + x_{min} \tag{9}$$

$$CentroidY = \left(\frac{y_{max} - y_{min}}{2} \right) + y_{min} \tag{10}$$

$$Presisi = \sqrt{((CentroidX_{gt})^2 - (CentroidY_{gt})^2) + ((CentroidX_{uji})^2 - (CentroidY_{uji})^2)} \tag{11}$$

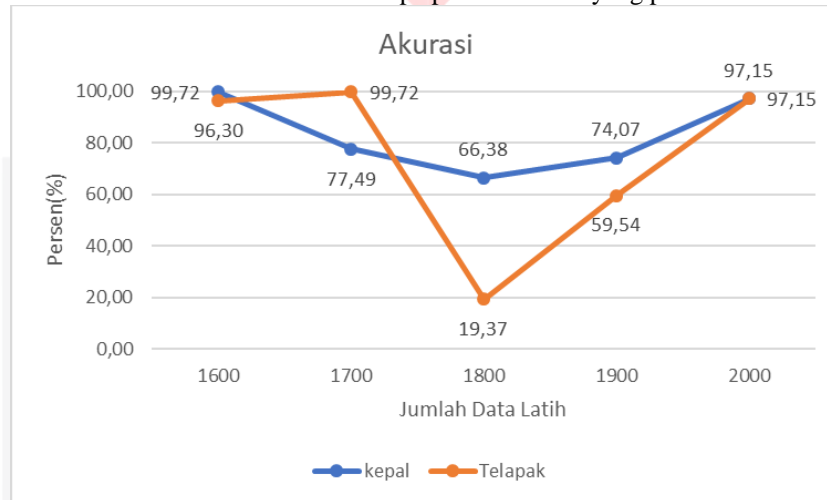
Pengujian ini digunakan untuk menunjukkan kedekatan dari masing-masing pengukuran sehingga hasil yang didapat tidak menyimpang jauh dan memiliki nilai antar hasil kedekatan yang rapat (minimal).

3.3.4. Pengujian terhadap waktu komputasi

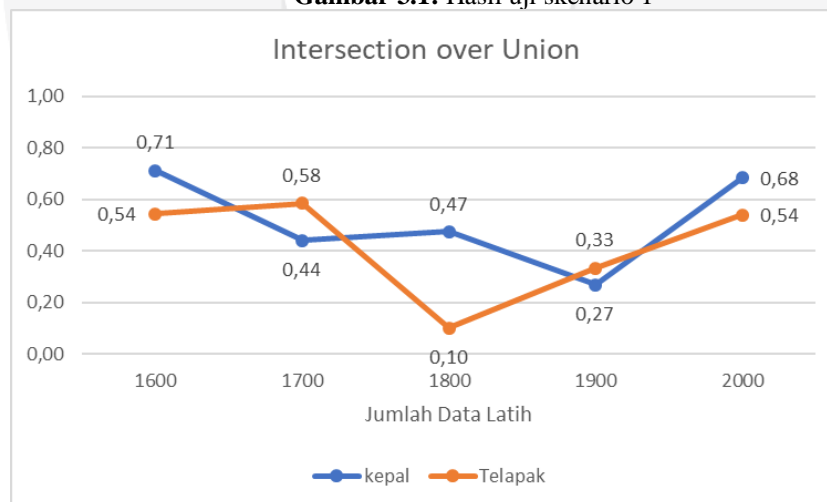
Pengujian ini bertujuan untuk menguji seberapa cepat sistem dapat mengenali suatu objek. Semakin sedikit waktu yang dibutuhkan sistem dalam mengenali suatu objek, maka sistem akan semakin mendekati realtime.

3.3. Data Hasil Pengujian Sistem

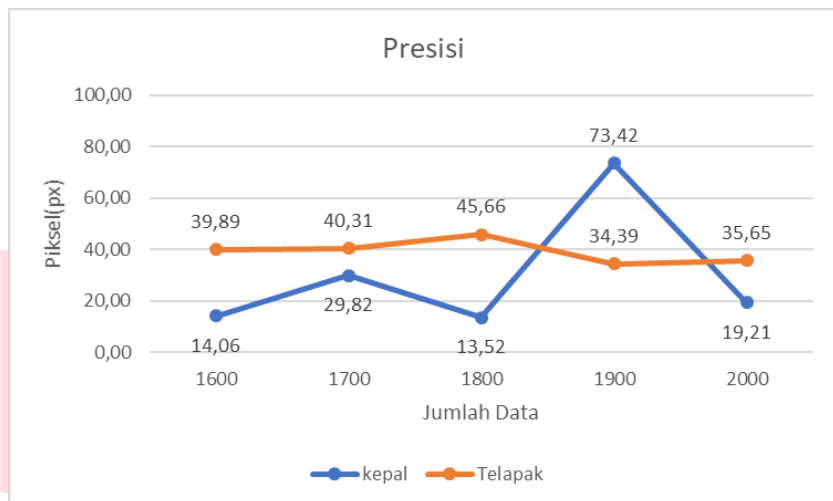
Pengujian dilakukan pada 350 data uji tangan terbuka dan 350 data uji tangan tertutup dengan parameter scale factor 1.1 pada semua model hasil latih dan beberapa parameter lain yang perlu dilakukan optimasi.



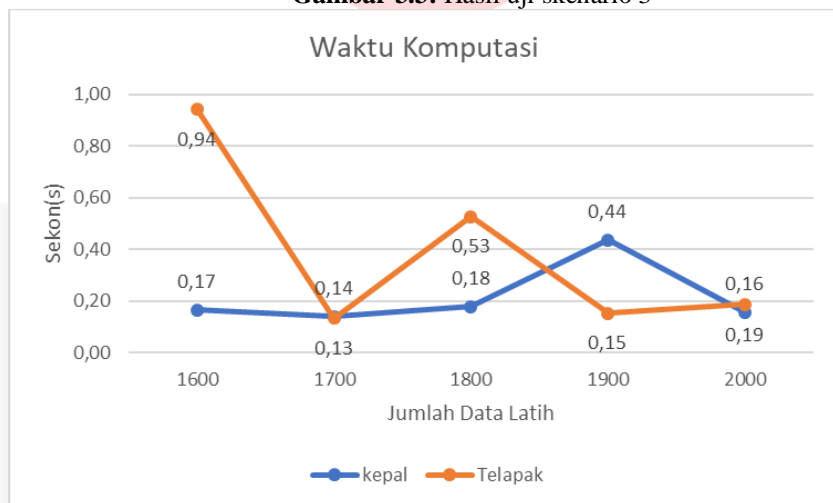
Gambar 3.1: Hasil uji skenario 1



Gambar 3.2: Hasil uji skenario 2



Gambar 3.3: Hasil uji skenario 3



Gambar 3.4: Hasil uji skenario 4

Pada gambar 3.1 merupakan grafik performansi dari parameter akurasi, dari grafik tersebut dapat dilihat bahwa saat menggunakan 1800 data latih akurasi mengalami penurunan dan merupakan hasil terendah dari jumlah data latih lainnya dari semua model dan mengalami kenaikan setelahnya. Sedangkan pada gambar 3.2 merupakan grafik performansi pada parameter IoU. Nilai IoU minimal objek terdeteksi adalah 0.5, sehingga pada 1700, 1800, dan 1900 data latih pada model tangan tertutup, 1800 dan 1900 tangan terbuka tidak memenuhi syarat dan dianggap banyak mendeteksi objek yang false positive. Gambar 3.3 menunjukkan grafik performansi dari presisi. Semakin kecil nilai presisi dianggap semakin baik sistem dalam mendeteksi suatu objek dan penggunaan pada 1800 dan 1900 data uji memiliki nilai presisi terbaik untuk tangan tertutup dan tangan terbuka. Gambar 3.4 menunjukkan grafik performansi dari waktu komputasi. Waktu komputasi tercepat didapat pada saat penggunaan 1700 data uji pada semua model. Konfigurasi konfigurasi tersebut merupakan konfigurasi terbaik dari setiap skenario.

4. Kesimpulan

Dari penelitian ini dapat disimpulkan dari 4 skenario yang diujikan penggunaan 1600 dataset pada model tangan tertutup dan penggunaan 1700 dataset pada tangan terbuka lebih baik dibandingkan dengan data uji lainnya dengan nilai hasil setiap skenarionya yaitu, pada akurasi yaitu 99.72% untuk semua model, lalu pada IoU 0.58 untuk tangan terbuka dan 0.71 untuk tangan tertutup, sedangkan pada presisi 40.31 piksel untuk tangan terbuka dan 14.06 piksel untuk tangan tertutup, dan pada waktu komputasi 0.13 sekon untuk tangan terbuka dan 0.17 sekon untuk tangan tertutup.

Daftar Pustaka

- [1] C. Q. a. J. Liang, "A simple and effective method for hand gesture recognition," in *International Conference on Network and Information Systems for Computers (ICNISC)*, 2016.
- [2] K. D. a. P. Katiyar, "Integrating Hand Gesture Modelling and Virtual," in *International Conference on Intelligent Circuits and Systems (ICICS)*, 2018.
- [3] E. L. a. M. A. Muslim, "Penerapan Adaboost untuk Klasifikasi Support," in *Seminar Nasional Teknologi dan Informatika*, 2017.
- [4] S. A. a. H. L. J. Tang, "Feature selection for classification: A review," in *CRC Press*, 2014.
- [5] P. F. a. T. B. O. Ronneberger, "'U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015.
- [6] S. D. a. R. A. F. Timbane, "The Experimental Comparison of Features for Hand Detection," in *International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, 2018.
- [7] D. L. X. W. a. X. T. Y. Sun, "Deepid3: Face recognition with very deep neural networks," in *arXiv preprint arXiv:1502.00873*, 2015.
- [8] M. K. a. V. Inmoonoy, "The Message Notification for Patients Care System using Hand Gestures Recognition," in *International Conference on Digital Arts, Media and Technology (ICDAMT)*, 2017.
- [9] D. Stutz, "Understanding Convolutional Neural Networks," in *Seminar Report, Fakultat fur Mathematik, Informatik und Naturwissenschaften Lehr-und Forschungsgebiet Informatik VIII Computer Vision*, 2014.
- [10] M. J. M. H. A. a. M. H. M. A. Rahaman, "Real-Time Computer Vision-based Bengali Sign Language Recognition," in *17th International Conference on Computer and Information Technology (ICCIT)*, 2014.
- [11] S. a. Shahad, "A Survey on Properties of Adaptive Boosting with Different Classifiers," in *International Journal of Science and Research (IJSR)*, 2017.
- [12] M. J. R. a. J. D. Wiltshire, "Important Optical Principles and their Occurrence in Nature," IEEE, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8068906>. [Accessed 25 11 2019].
- [13] S. A. W. a. R. P. A. Y. Nasirudin, "Performance Analysis on Fine-tuned Region-based CNN for Object Recognition," in *2nd Symposium of Future Telecommunication and Technologies (SOFTT)*, Bandung, 2018.