

Computer Vision Implementation for Detection and Counting the Number of Humans

Tresya Anjali Dompeipen¹⁾, Sherwin R.U.A Sompie²⁾, Meicsy E.I Najooan³⁾
Jurusan Teknik Elektro, Universitas Sam Ratulangi Manado, Jl. Kampus Bahu, 95115, Indonesia
E-mail: dompeipentesa08@gmail.com¹⁾, aldo@unsrat.ac.id²⁾, meicsynajoan@unsrat.ac.id³⁾
Diterima: 11 Desember 2020 ; Direvisi: 12 Januari 2021; Disetujui: 21 Maret 2021

Abstract — *Limiting the capacity of visitors to a place is very necessary in the current situation and conditions. This is because it is to prevent crowds of people in one place, because it is known together with the health protocols from the government that each person must conduct outreach. For this reason, it is necessary to know the number of people who enter and leave a building, so that it can be seen whether the visitor capacity has reached the maximum number or not. The job of counting the number of visitors to a place is still easy to do if it's on a small scale, but what if it's on a large scale. By utilizing technology in the field of computer vision, this study aims to perform and calculate the number of humans automatically. This research uses algorithms from deep learning, namely MobileNet-SSD and Centroid Tracking to detect and count objects. Based on the results of evaluations and evaluations that have been carried out on the five trial videos, the system built successfully produces data containing the number of humans detected with the highest level of detection and counting of objects from objects of 93.75%.*

Keywords — *Centroid Tracking, Computer Vision, Deep Learning, MobileNet-SSD.*

Abstrak — *Pembatasan kapasitas pengunjung pada suatu tempat sangat diperlukan pada situasi dan kondisi saat ini. Hal ini dikarenakan untuk mencegah adanya kerumunan orang disuatu tempat, karena diketahui bersama sesuai protokol kesehatan dari pemerintah tiap orang harus melakukan pembatasan sosial. Untuk itu perlu diketahui jumlah orang yang masuk dan keluar dari suatu bangunan, sehingga dapat diketahui apakah kapasitas pengunjung telah mencapai jumlah maksimal atau belum. Pekerjaan untuk menghitung jumlah pengunjung pada suatu tempat masihlah mudah dilakukan jika dalam skala kecil, namun bagaimana jadinya jika berada pada skala yang besar. Dengan memanfaatkan teknologi pada bidang computer vision yaitu deep learning, penelitian ini bertujuan untuk mendeteksi dan menghitung jumlah manusia secara otomatis. Pada penelitian ini menggunakan algoritma dari deep learning yaitu MobileNet-SSD dan Centroid Tracking untuk melakukan deteksi dan penghitungan objek. Berdasarkan hasil pengujian dan evaluasi yang telah*

dilakukan pada kelima video uji coba, sistem yang dibangun berhasil menghasilkan suatu data yang berisi jumlah manusia yang terdeteksi dengan tingkat akurasi tertinggi pada pendeteksian dan penghitungan objek manusia sebesar 93,75%.

Kata kunci — *Centroid Tracking, Computer Vision, Deep Learning, MobileNet-SSD.*

I. PENDAHULUAN

Perubahan yang diakibatkan oleh covid-19 memaksa munculnya kondisi yang baru. Secara global kehidupan sosial tercipta suatu kehidupan *new normal* (tatanan kehidupan baru). Namun meskipun telah diterapkan tatanan kehidupan baru, orang-orang masih merasa rasa curiga dan takut berlebihan (*paranoid*) jika ingin beraktifitas diluar rumah, apalagi jika berada disebuah tempat umum yang ramai. Maka dari itu banyak tempat-tempat umum telah memberikan batasan pengunjung, Untuk tempat berskala kecil masih mudah untuk mengetahui dan mengawasi jumlah pengunjung yang ada, namun bagaimana pada tempat berskala besar. Tidak efektif dan efisien jika hanya mengandalkan kemampuan manusia untuk menghitung jumlah setiap pengunjung yang masuk dan keluar pada suatu tempat. Maka dari itu dibutuhkan teknologi yang dapat memberikan informasi tentang jumlah kapasitas pengunjung yang ada disuatu tempat dengan cara menghitung tiap pengunjung yang masuk dan *keluar ditempat tersebut*.

Computer vision merupakan ilmu komputer yang bekerja dengan cara meniru kemampuan *visual* manusia. *Computer vision* mempunyai beberapa macam bidang salah satunya adalah *object detection*. *Object dection* bekerja dengan mengenali objek yang berada pada gambar dan letak objek tersebut. *Object detection* dapat dimanfaatkan untuk berbagai keperluan misalnya, untuk keamanan, informasi, monitoring, tingkat produktivitas dan lain-lain.

Berdasarkan permasalahan yang ada maka perlu untuk membuat suatu sistem yang dapat mendeteksi dan menghitung jumlah manusia guna mendapatkan informasi tentang jumlah objek manusia yang terdeteksi. Pendeteksian dan penghitungan manusia ini menggunakan *library* dari *OpenCv* yang memang difokuskan dalam hal pengolahan citra digital. Dalam pengimplementasian sistem perhitungan ini digunakan 2 metode yaitu *object detection* dan *object tracking*. Untuk pendeteksian,

algoritma yang digunakan adalah *MobileNet SSD (Single Shot Multibox Detectors)*, sedangkan untuk proses *tracking* digunakan algoritma *centroid tracking*. Alasan digunakan kedua metode ini dikarenakan *object detector* tidak memungkinkan untuk membaca setiap *frame* dari *input* video dengan mempertahankan kinerja waktu secara *real-time*. Maka dari itu diperlukan algoritma sebagai perantara untuk dapat menerima lokasi *bounding box* dari setiap objek, melacaknya dan kemudian secara otomatis memperbaharui sendiri saat objek bergerak didalam *frame* dari *input* video. Dengan menggabungkan kedua metode ini, yaitu *object detection* dan *tracking*, maka objek akan lebih cepat terdeteksi.

A. Penelitian Terkait

Dalam penelitian ini, penulis mengambil beberapa penelitian terkait sebelumnya untuk menjadi acuan sehingga dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan serta dapat menghindari adanya duplikasi. Hal ini bermanfaat agar penelitian yang dilakukan dapat memiliki arti dan kontribusi terhadap ilmu pengetahuan.

- 1) Penelitian oleh Khairul Umam dan Benny Sukma Negara pada tahun 2016 mengenai. Deteksi Obyek Manusia Pada Basis Data Video Menggunakan Metode Background Subtraction dan Operasi Morfologi: Penelitian menggabungkan dua metode untuk melakukan pendeteksian objek manusia. Metode Background Subtraction digunakan untuk mendeteksi subtraksi pada background, Operasi Morfologi digunakan untuk menghilangkan noise pada video [1].
- 2) Penelitian oleh Ferdinan Hasugian, Achmad Rizal, dan Rita Magdalena pada tahun 2008. Implementasi Perangkat Lunak Penghitung Jumlah Orang Berbasis Pengolahan Citra Menggunakan Webcam: Penelitian ini menggunakan metode Normalized Sum Differences (NSD) untuk menghitung jumlah objek manusia didalam ruangan. Nilai akurasi optimum yang didapat adalah 94% [2].
- 3) Penelitian oleh Afandi Nur Aziz Thohari dan Rifki Adhitama pada tahun 2019. Real-Time Object Detection For Wayang Punakawan Identification Using Deep Learning: Penelitian ini menggunakan metode Single Shot Multibox Detectors (SSD) dan digabungkan dengan MobileNet untuk mengidentifikasi objek wayang. Tingkat akurasi yang dihasilkan dari proses deteksi wayang ini adalah sebesar 98.86% [3].

B. Citra Digital

Citra digital adalah gambar dua dimensi yang dihasilkan dari gambar analog dua dimensi yang kontinu menjadi gambar diskrit melalui proses sampling. Di dalam citra digital terdapat sebuah larik (*array*) yang berisi nilai-nilai real maupun kompleks yang di representasikan dengan deretan bit tertentu. Sebuah citra digital dapat diwakili oleh sebuah matriks dua dimensi $f(x,y)$ yang terdiri dari M kolom dan N baris, dimana perpotongan antar kolom dan baris disebut piksel (pixel = picture element) atau elemen terkecil dari sebuah citra [4].

Proses perubahan citra analog menjadi citra digital dinamakan dengan digitasi. Digitasi adalah proses mengubah gambar, teks, atau suara dari benda yang dapat dilihat ke dalam data elektronik dan dapat disimpan serta diproses untuk keperluan lainnya. Citra digital pada komputer dipetakan menjadi bentuk grid atau elemen piksel berbentuk matriks 2 dimensi. Setiap piksel-piksel tersebut memiliki angka yang mempresentasikan channel warna. Angka pada setiap piksel disimpan secara berurutan oleh komputer dan sering dikurangi untuk keperluan kompresi maupun pengolahan tertentu [5].

C. Pengolahan Citra Digital

Pengolahan citra digital (*Digital Image Processing*) adalah sebuah disiplin ilmu yang mempelajari tentang teknik-teknik mengolah citra. Citra yang dimaksud disini adalah gambar diam (foto) maupun gambar bergerak (video). Sedangkan digital disini mempunyai maksud bahwa pengolahan citra/gambar dilakukan secara digital menggunakan komputer [4]. Secara matematis, citra merupakan fungsi kontinu (*continue*) dengan intensitas cahaya pada bidang dua dimensi. Agar dapat diolah dengan komputer digital, maka suatu citra harus dipresentasikan secara numerik dengan nilai diskrit. Representasi dari fungsi kontinu menjadi nilai-nilai diskrit disebut digitalisasi citra. Tujuan dari pengolahan citra ini adalah memperbaiki kualitas suatu citra sehingga dapat diinterpretasi dengan mudah oleh manusia atau sebuah mesin (komputer).

Proses perubahan citra menjadi citra digital dinamakan dengan digitasi. Digitasi merupakan proses sebuah gambar, teks, atau suara dari benda yang dapat dilihat ke dalam data elektronik dan dapat disimpan serta diproses untuk keperluan lainnya. tujuan dari pengolahan citra terbagi atas 3 bagian, yaitu [6]:

- 1) Memodifikasi sebuah citra untuk meningkatkan kualitas maupun menekankan pada sejumlah aspek informasi yang terkandung didalam citra.
- 2) Mengklasifikasikan, mencocokkan dan mengukur bagian-bagian tertentu dari sebuah citra.
- 3) Membagi bagian-bagian citra yang ingin dihilangkan atau digabungkan dengan bagian citra lain.

Pada penelitian ini terdapat beberapa teknik pengolahan citra digital yang digunakan antara lain:

1) Resizing

Resize merupakan proses pengubahan ukuran besarnya citra digital entah itu berubah menjadi lebih kecil atau lebih besar dari file aslinya.

2) Mean Subtraction

Mean subtraction digunakan untuk membantu mengatasi permasalahan iluminasi pada gambar *input* yang ada pada *dataset*. Untuk mengatasi variasi intensitas dan normalisasi, dilakukan dengan cara menghitung nilai piksel rata-rata pada *dataset* pelatihan dan mengurangnya dari setiap gambar selama pelatihan. Tahap ini dilakukan dengan inisialisasi 3 buah variabel yaitu : μ_R , μ_G , μ_B . Kemudian disaat gambar siap melalui jaringan yang ada, dilakukan prose pengurangan nilai rata-rata untuk setiap *input channel* pada citra *input*.

3) Scaling

Scaling terhadap standar deviasi oleh training set dibutuhkan pada gambar agar dapat diproses oleh jaringan *Deep Learning* berdasarkan citra masukan serta dapat mengkordinasikan skala *bounding box* yang diproses pada ukuran citra *input*, *display* maupun citra yang akan disimpan. *scaling factor*, σ , yang ditambahkan dalam sebuah normalisasi.

D. Computer Vision

Computer vision adalah proses pembelajaran dan menganalisa gambar ataupun video untuk memperoleh hasil sebagaimana yang bisa dilakukan manusia. Dengan bahasa lebih mudahnya, *computer vision* mencoba meniru cara kerja visualisasi manusia [1].

Computer vision mempunyai keterkaitan dengan beberapa bidang yaitu, *image processing* (pengolahan citra) dan *machine vision* (visi mesin). Ada kesamaan yang signifikan dalam berbagai teknik dan aplikasi yang mencakup tiga bidang ini. Hal ini menunjukkan teknik dasar yang digunakan dan dikembangkan kurang lebih sama. Secara luas *computer vision* berhubungan dan diterapkan dengan bidang lain seperti *artificial intelligence (AI)*, robotika, otomasi industry, pengolahan sinyal, optik fisik, *neurobiology*, dan lain-lain.

Computer vision sekarang ini telah sering digunakan untuk berbagai hal, contohnya saja mendeteksi wajah pada gambar (*face detection*), mengenali wajah (*facial expression recognition*) dan dalam prakteknya sering digunakan bersama dengan jaringan syaraf tiruan (*artificial neural network*).

E. Artificial Intelligence (AI)

AI bertujuan untuk mengetahui atau memodelkan proses berpikir manusia dan mendesain mesin sehingga bisa menirukan perilaku manusia. Terdapat dua hal yang menjadi bagian utama dalam pembuatan aplikasi kecerdasan buatan, yaitu [7]:

- 1) *Knowledge Base* (Basis Pengetahuan). Bagian ini berisi tentang fakta-fakta, teori, pemikiran dan hubungan antara satu dengan yang lainnya.
- 2) *Inference Engine* (Motor Inferensi) yaitu kemampuan menarik kesimpulan berdasarkan pengalaman.

Artificial Intelligence (AI) merupakan disiplin ilmu yang luas dan beberapa lingkup utama AI antara lain adalah Sistem Pakar (*Expert System*), Pengolahan Bahasa Alami (*Natural Language Processing*), Pengenalan Ucapan (*Speech Recognition*), *Computer Vision*, *Intelligence Computer-Aided Instruction*, dan lainnya.

Artificial Intelligence (AI) yang dibuat berdasarkan sistem yang memiliki keahlian seperti manusia pada domain tertentu disebut dengan *soft computing*. *Soft computing* merupakan dasar dalam membangun sistem cerdas yang mampu beradaptasi dan bekerja lebih baik jika terjadi perubahan lingkungan. *Soft computing* juga mengeksplorasi adanya toleransi terhadap ketidakpastian, dan kebenaran parsial sehingga dapat diselesaikan dan dikendalikan dengan mudah agar sesuai dengan realita. Metodologi yang sering digunakan dalam *soft computing* salah satunya yaitu Jaringan Syaraf Tiruan (*Artificial Neural Network / ANN*). Metodologi lain yang juga digunakan adalah Sistem *Fuzzy* (mengakomodasi

ketepatan), *Probabilistic Reasoning* (Mengakomodasi Ketidakpastian) dan *Evolutionary Computing* (Optimasi) .

F. Komponen Neural Network

Neural Network memiliki beberapa tipe yang berbeda-beda, akan tetapi hampir semua komponen yang dimiliki sama. Seperti halnya jaringan syaraf pada otak manusia, *neural network* juga terdiri dari beberapa *neuron* unit yang saling berhubungan. Masing-masing dari neuron tersebut akan melakukan transformasi informasi yang diterima melalui sambungan keduanya menuju neuron lain. Hubungan ini biasanya disebut dengan sebutan bobot (*weight*). Informasi tersebut disimpan pada suatu nilai tertentu pada bobot tertentu.

Cara kerja struktur neural network identik dengan struktur jaringan syaraf pada manusia. Informasi (*input*) akan dikirimkan dengan bobot kedatangan tertentu. Input tersebut kemudian diproses oleh suatu fungsi perambatan yang akan menjumlahkan nilai semua bobot yang datang. Hasil penjumlahan ini kemudian akan dibandingkan dengan suatu nilai ambang (*threshold*) tertentu melalui fungsi aktivasi setiap *neuron*. Jika *input* tersebut melewati suatu nilai ambang tertentu, maka *neuron* tersebut akan diaktifkan. Jika tidak, *neuron* tersebut tidak akan diaktifkan. Apabila *neuron* diaktifkan, selanjutnya *neuron* tersebut akan mengirimkan *output* melalui bobot-bobot *outputnya* ke semua neuron yang berhubungan dengannya, begitu seterusnya.

G. Machine Learning

Istilah *Machine learning* pertama kali didefinisikan oleh Arthur Samuel ditahun 1959 [8]. Menurut Arthur Samuel, *Machine Learning* adalah salah satu bidang ilmu komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas. Terdapat dua aplikasi utama di dalam *Machine Learning*, yaitu klasifikasi dan prediksi.

Pada dasarnya *Machine Learning* adalah proses komputer untuk belajar dari data. Tanpa adanya data komputer tidak akan bisa belajar apa-apa. Semua pengetahuan *Machine Learning* pasti akan melibatkan data. Data bisa saja sama, akan tetapi algoritma dan pendekatannya berbeda-beda untuk mendapatkan hasil yang optimal.

Didalam pembelajaran *Machine Learning* terdapat tiga kategori utama yaitu:

1) Supervised Learning

Supervised Learning merupakan pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label

2) Unsupervised Learning

Unsupervised Learning merupakan pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokkan data berdasarkan karakteristik-karakteristik yang ditemui

3) Reinforcement Learning

Reinforcement Learning merupakan fase pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan

sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajaran

H. Deep Learning

Deep Learning merupakan salah satu bidang dari *Machine Learning* yang memanfaatkan jaringan syaraf tiruan untuk implementasi permasalahan dengan dataset yang besar. Teknik *Deep Learning* memberikan arsitektur yang sangat kuat untuk *Supervised Learning*. Dengan menambahkan lebih banyak lapisan maka model pembelajaran tersebut bisa mewakili data citra berlabel dengan lebih baik. Pada *Machine Learning* terdapat teknik untuk menggunakan ekstraksi fitur dari data pelatihan dan algoritma pembelajaran khusus untuk mengklasifikasi citra maupun untuk mengenali suara. Namun, metode ini masih memiliki beberapa kekurangan baik dalam hal kecepatan dan akurasi.

Model *Deep Learning* dirancang untuk melalui berbagai layer. Pada situasi jaringan saraf tiruan, *Multi-Layer Perceptron (MLP)* dengan lebih dari dua *hidden layer* dapat dikenal sebagai *Deep Model*. Beberapa *layer* yang sering digunakan adalah *convolution layer*, *dropout layer*, *fully connected layer*, *pooling layer* and *Relu layer*. Setelah menerapkan beberapa *layer*, maka akan didapat lebih banyak data serta perlu menghindari *over-fitting*. Ada beberapa prosedur yang digunakan untuk menghindari *over-fitting*, yang paling umum adalah augmentasi data dan regularisasi [9].

Aplikasi konsep *Deep Learning* dapat diterapkan pada algoritma *Machine Learning* yang sudah ada sehingga komputer sekarang bisa belajar dengan kecepatan, akurasi, dan skala yang besar. Prinsip ini terus berkembang hingga *Deep Learning* semakin sering digunakan pada komunitas riset dan industri untuk membantu memecahkan banyak masalah data besar seperti *Computer vision*, *Speech recognition*, dan *Natural Language Processing*. *Feature Engineering* adalah salah satu fitur utama dari *Deep Learning* untuk mengekstrak pola yang berguna dari data yang akan memudahkan model untuk membedakan kelas.

I. Artificial Neural Network

Artificial Neural Network (ANN) merupakan suatu model komputasi paralel yang meniru fungsi dari sistem jaringan syaraf biologi otak manusia. Dalam otak manusia terdiri dari milyaran neuron yang saling berhubungan. Hubungan ini disebut dengan *Synapses*. Komponen *neuron* terdiri dari satu inti sel yang akan melakukan pemrosesan informasi, satu akson (*axon*) dan minimal satu *dendrit*. Informasi yang masuk akan diterima oleh *dendrit*. Selain itu, *dendrit* juga menyertasi akson sebagai keluaran dari suatu pemrosesan informasi.

Cara kerja dari sistem syaraf diatas adalah bermula pada sinyal masuk melalui *dendrit* menuju *cell body*. Kemudian sinyal akan di proses didalam *cell body* berdasarkan fungsi tertentu (*Summation Proses*). Jika sinyal hasil proses melebihi nilai ambang batas (*threshold*) tertentu maka sinyal tersebut akan membangkitkan *neuron* untuk meneruskan sinyal tersebut. Sedang jika dibawah nilai ambang batasnya maka sinyal tersebut akan dihalangi (*inhibited*). Kemudian sinyal yang

diteruskan akan menuju ke *axon* dan akhirnya menuju ke *neuron* lainnya melewati *synapse*.

ANN merupakan sistem adatif yang dapat mengubah strukturnya untuk memecahkan suatu masalah berdasarkan informasi internal maupun eksternal. penerapan ANN dapat mengidentifikasi beberapa aplikasi yaitu:

- 1) Estimasi/prediksi (aproksimasi fungsi)
- 2) Pengenalan Pola (klasifikasi, diagnosis, dan analisis diskriminan)
- 3) *Klustering* (pengelompokan tanpa adanya pengetahuan sebelumnya).

J. Object Detection

Pada saat manusia melihat sebuah objek dalam gambar maka otak manusia akan dapat langsung mengenali objek, letak, beserta kondisi interaksi yang terjadi. Dengan sistem visual manusia yang cepat dan akurat, memungkinkan manusia untuk dapat melakukan tugas-tugas kompleks. Namun bagaimana jika tugas-tugas kompleks tersebut dilakukan oleh sebuah komputer. Untuk itu dibutuhkan sebuah algoritma yang cepat dan akurat yang akan memungkinkan komputer dapat melakukan hal serupa hingga berpotensi untuk menyelesaikan tugas secara umum. Deteksi objek merupakan salah task dari computer vision. Deteksi objek adalah suatu proses yang digunakan untuk menentukan keberadaan objek tertentu didalam suatu citra digital.

Proses deteksi tersebut dapat dilakukan dengan berbagai macam metode yang umumnya melakukan pembacaan fitur-fitur dari seluruh objek pada citra input. Fitur dari objek pada citra input tersebut akan dibandingkan dengan fitur dari model yang digunakan atau *template*. Hasil perbandingan tersebut dapat digunakan untuk menentukan apakah suatu objek terdeteksi sebagai *template* yang dimaksud atau tidak. Sistem deteksi objek perlu melatih dan menguji *dataset* dengan *bounding box* dan diberi label untuk kelas per setiap objek untuk proses pengenalan. Demi mencapai tujuan ini, ada banyak *dataset* untuk menghasilkan model *Deep Learning* seperti *Pascal-VOC*, *MS COCO*, atau bahkan dapat menggunakan *dataset* yang dibuat sendiri [10].

K. MobileNet

MobileNet, merupakan salah satu arsitektur *Convolutional Neural Network (CNN)* yang dapat digunakan untuk mengatasi kebutuhan akan sumber komputasi berlebih. Para peneliti dari Google membangun *MobileNet* atas kebutuhan arsitektur CNN yang dapat digunakan untuk ponsel maupun sistem tertanam. Perbedaan mendasar antara arsitektur *MobileNet* dan arsitektur CNN pada umumnya adalah penggunaan lapisan atau layer konvolusi dengan ketebalan *filter* yang sesuai dengan ketebalan dari *input image*. *MobileNet* membagi konvolusi menjadi *depthwise convolution* dan *pointwise convolution*. Arsitektur *MobileNet* memanfaatkan *Batch Normalization (BN)* dan *Rectified-Linear unit (ReLU)* untuk *depthwise convolution* dan *pointwise convolution*. *MobileNet* dibangun di atas arsitektur jaringan yang efisien dengan menggunakan konvolusi yang dapat dipisahkan secara mendalam untuk menghasilkan *Deep Neural Network* yang ringan.

DSC menggantikan konvolusi standard dengan 2 tahap operasi: 1. *Depthwise Convolution* dimana setiap filter $DF \times DF$ hanya melakukan proses *filter* terhadap sebuah *feature map input* secara mendalam; 2. *Pointwise Convolution* yang merupakan 1×1 *convolution layer* yang digunakan untuk menggabungkan jalur informasi dari *depthwise layer*.

L. Single Shot Multi-Box Detectors (SSD)

SSD dirancang untuk pendeteksian objek secara *real-time*. Perbedaan utama antara pelatihan detector klasik seperti R-CNN dan pelatihan SSD adalah bahwa *detector* objek klasik menggunakan proposal wilayah dan metode SSD menggunakan *groundtruth* yang harus ditetapkan untuk hasil yang ditentukan dalam rangkaian hasil detektor yang jelas.

Dalam mendeteksi objek, SSD menandai area *bounding box* yang diprediksi kepada koleksi *default bounding box* melalui berbagai skala dan rasio untuk setiap lokasi *feature map*. SSD membandingkan objek dengan *default bounding box* dengan berbagai rasio selama masa *training*. Setiap *default box* dengan $IoU > 0.5$ dikategorikan cocok. Metode SSD juga menggunakan sejumlah *layer* dalam berbagai skala yang mampu memberikan hasil terbaik terhadap objek yang terdeteksi. Dalam penelitian ini, arsitektur *MobileNet* digunakan sebagai *feature extractor* pada metode SSD.

Ketika gambar diteruskan kepada arsitektur *MobileNet*, SSD menggunakan 6 *layer* konvolusi ekstra. Tiga dari *layer* ekstra tersebut dapat menghasilkan enam prediksi untuk setiap sel. Metode SSD secara total mampu menghasilkan 8732 prediksi dengan memanfaatkan 6 *layer* tersebut. *Layer* ekstra tersebut juga menghasilkan *feature maps* dalam berbagai ukuran untuk mendeteksi objek dalam berbagai ukuran sehingga mampu memberikan akurasi yang lebih baik terhadap objek-objek yang memiliki ukuran berbeda dalam suatu gambar [11]

M. Object Tracking

Dalam *computer vision*, *object tracking* adalah suatu proses untuk melacak satu objek atau lebih dari suatu citra. Ada tiga langkah penting dalam analisa video: deteksi objek yang bergerak, mendeteksi beberapa objek di setiap *frame*, dan analisa objek yang dilacak untuk mengenali pergerakan objek pada citra [12]. Sederhananya, *tracking* dapat didefinisikan sebagai suatu masalah untuk memperkirakan lintasan dari sebuah objek yang bergerak dalam gambar. Secara konsisten, pelacak memberikan label pada objek yang dilacak pada *frame-frame* yang berbeda dalam sebuah video. Berdasarkan dari pelacakan domain, sebuah pelacak juga dapat memberikan informasi suatu objek, seperti sebuah orientasi gerak, area, atau bentuk dari objek. Terdapat 3 proses dalam pelacakan objek, yaitu [13]:

- 1) Mengambil *set* awal deteksi objek seperti *set input* dari koordinat *bounding box*
- 2) Membuat ID unit untuk setiap objek deteksi
- 3) Lacak setiap pergerakan objek disekitar *frame* dalam video, dan tetap mempertahankan dalam pemberian ID unik.

N. Centroid Tracking

Centroid tracking merupakan sebuah algoritma yang bergantung pada jarak Euclidean antar centroid objek yang ada

dan centroid objek yang baru antara *frame* berikutnya dalam video.

O. Confusion Matrix

Confusion matrix sering disebut sebagai error matrix. Confusion matrix digunakan untuk mendapatkan informasi tentang perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. Matrix ini berbentuk tabel yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui

P. OpenCV

OpenCV adalah sebuah *library* yang berisi fungsi-fungsi pemrograman untuk teknologi *computer vision* secara *real time*. *OpenCV* awalnya dikembangkan oleh Intel, dan sekarang didukung oleh Willow Garage dan Itseez. *OpenCV* dapat berjalan di berbagai Bahasa pemrograman seperti, C, C++, Java, Python, dan juga *support* diberbagai platform seperti Windows, Linux, Mac OS, IOS dan Android.

II. METODE PENELITIAN

A. Waktu dan Tempat Penelitian

Penelitian ini dilakukan mulai dari bulan Februari – November. Penelitian dan pengambilan data diambil pada tempat yang ramai pengunjung, seperti pusat perbelanjaan atau taman.

B. Alat dan Bahan

Alat yang digunakan didalam penelitian ini adalah sebuah laptop Lenovo ideapad 110, Smartphone Samsung J5 untuk alat perekam, python v3.6 sebagai Bahasa pemrograman yang digunakan, command prompt sebagai aplikasi command line untuk menjalankan program. Bahan yang digunakan didalam penelitian ini adalah filr-file video rekaman.

C. Tahapan Penelitian

Penelitian ini terdiri dari delapan tahapan, yaitu yang pertama menentuka topik penelitian yang akan diangkat, kedua mengidentifikasi masalah yang apa yang harus dipecahkan, ketiga menentukan tujuan yang ingin dicapai dalam penelitian ini, keempat melakukan pengumpulan data dengan metode studi literatur pada penelitian sebelumnya yang terkait dengan penelitian ini, kelima melakukan perancangan sistem mulai dari algoritma yang akan digunakan serta data-data yang akan menjadi parameter dalam penelitian, keenam dilakukan pengkodean program yang kemudian dilanjutkan pada tahapan ke tujuh yaitu uji coba sistem dan menarik kesimpulan.

D. Pengumpulan Data

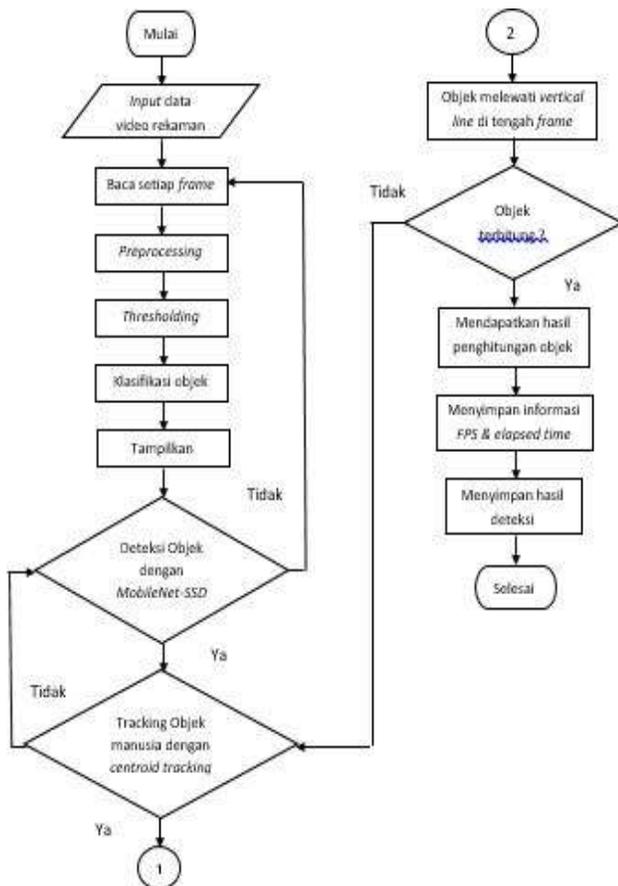
Data yang digunakan pada penelitian ini merupakan file-file video rekaman yang diambil menggunakan smartphone. File video rekaman diambil dengan kondisi yang berbeda-beda untuk digunakan sebagai bahan pengujian tingkat akurasi dari sistem yang akan dibangun. Informasi tentang file video rekaman dapat dilihat pada Tabel 1.

E. Perancangan Sistem

Proses perancangan sistem pada penelitian ini diawali dengan mengidentifikasi masalah serta menganalisa kebutuhan sistem yang akan dirancang. Di dalam penelitian ini ada beberapa hal penting untuk ada didalam sistem yang akan dibangun yaitu, sistem dapat mendeteksi objek manusia yang berada didalam file video rekaman, dan sistem dapat menghitung berapa objek manusia pada sebuah rekaman video.

Proses tahapan dari sistem yang akan dibangun dimulai dari perekaman objek yang bertempat dilokasi yang ramai pengunjung dengan kondisi tiap video rekaman yang berbeda-beda, selanjutnya file-file rekaman video tersebut dimasukan kedalam program sebagai input. File program yang dinamakan `people_counter.py` tersebut kemudian dijalankan didalam `command prompt`. Saat file program dijalankan, akan muncul sebuah tampilan deteksi dan penghitungan objek manusia.

Output dari sistem akan tersimpan pada sebuah file direktori yang telah ditentukan. Output tersebut kemudian akan dilakukan analisa menggunakan teknik `confusion matrix` untuk melihat tingkat akurasi pada masing-masing file video rekaman. Kemudian Hasil dari proses analisa akan dievaluasi oleh penulis untuk melihat hal-hal yang mempengaruhi dari tingkat akurasi sistem. Langkah selanjutnya adalah pembuatan bagan alir input dan output sistem yang dapat dilihat pada gambar



Gambar 1 Bagan Alir Sistem Pendeteksian dan Penghitung Jumlah Manusia

F. Metode Pengujian

Pengujian sistem pada penelitian ini menggunakan beberapa parameter untuk saling dibandingkan agar nantinya memperoleh informasi kondisi yang paling baik untuk menerapkan sistem pendeteksian dan penghitung jumlah manusia.

- 1) Pengujian sistem akan dilakukan dengan berbagai kondisi tingkat pencahayaan yang akan diambil pada waktu pagi, siang, sore ataupun malam hari.
- 2) Pengujian sistem akan dilakukan pada berbagai jarak objek terhadap kamera untuk mengetahui sejauh mana sistem dapat mendeteksi objek.
- 3) Pengujian sistem akan dilakukan dengan berbagai perspektif kamera dalam pengambilan gambar untuk mengetahui posisi atau sudut yang tepat dalam proses perekaman video.
- 4) Video rekaman akan mengambil kondisi ramai dan terdapat objek lain didalamnya.
- 5) Pengujian sistem akan dilakukan pada spesifikasi laptop yang berbeda-beda.

III. HASIL DAN PEMBAHASAN

A. Pengujian Sistem

Tahapan pengujian pada sistem dilakukan untuk mengetahui kemampuan dari sistem yang dibangun dalam melakukan pendeteksian dan penghitungan jumlah manusia pada setiap kondisi yang berbeda-beda dari file-file video rekaman. Selain itu, sistem juga akan diuji diberbagai spesifikasi laptop dan komputer yang berbeda untuk dapat melihat kemampuan dari sistem pada perangkat yang berbeda, serta dapat mengetahui spesifikasi perangkat yang direkomendasikan untuk menjalankan sistem. Pada Tabel II diperlihatkan kondisi dari tiap video yang akan dilakukan uji coba, tampilan sistem pendeteksian dan penghitungan jumlah manusia dapat dilihat pada gambar 2.

1) Pengujian Ke-1

Pengujian pertama dilakukan pada Video1 dan Video2 dikarenakan kondisi dari kedua video yang sama yaitu, intensitas cahaya keduanya adalah 550 lux, jarak objek terhadap kamera sekitar 5 meter dan sudut pengambilan video sekitar

TABEL I TABEL DATA VIDEO UJI COBA

Nama File	Durasi (Menit:Detik)	Total Frame	Jumlah Objek yang melintas
Video1.mp4	01:01	1830	9
Video2.mp4	00:25	750	6
Video3.mp4	00:46	1380	12
Video4.mp4	00:20	600	10
Video5.mp4	00:30	900	27
Video6.mp4	00:30	900	23
Video7.mp4	00:24	720	6
Video8.mp4	01:13	2190	37
Video9.mp4	01:09	2070	16
Video10.mp4	01:01	1830	13
Video11.mp4	00:37	110	7
Video12.mp4	00:15	450	5

45°. Kedua file video ini direkam di halaman rumah dengan kondisi objek manusia yang tidak terlalu ramai.

Dari hasil pengujian yang telah dilakukan, Video1 membutuhkan waktu 128.28 detik untuk proses deteksi dan penghitungan objek dari durasi video sebenarnya adalah 64 detik, sedangkan untuk Video2 membutuhkan waktu proses

59.68 detik dari durasi video sebenarnya adalah 25 detik. Untuk jumlah FPS yang berhasil terbaca oleh sistem pada Video1 adalah 14.99 *frame* per detik, untuk Video2 FPS yang berhasil terdeteksi adalah 12.92 *frame* per detik, dari jumlah *frame* video yang sebenarnya adalah 30 *frame* per detik. File Video1 dan Video2 memiliki 15 objek secara keseluruhan, namun terjadi kesalahan deteksi dan penghitungan objek oleh sistem diakibatkan oleh beberapa faktor, dalam kasus ini terjadi deteksi ganda dari beberapa objek, sehingga sistem mendeteksi dan menghitung objek manusia sebanyak 16 objek.

2) Pengujian Ke-2

Pada pengujian kedua dilakukan pada dua video yaitu Video3 dan Video4, Hal ini dilakukan karena kondisi dari kedua video yang sama, mulai dari intensitas cahaya yang berkisar 50-80 lux, jarak objek terhadap kamera sekitar 2 – 3 meter dan sudut pengambilan video sekitar 35° derajat. Kedua file ini direkam pada dua lokasi yang berbeda, yaitu pertama didepan halaman minimarket, dan kedua di dalam pusat perbelanjaan (Manado Town Square).

Dari pengujian yang telah dilakukan, mendapatkan waktu lamanya proses deteksi dan penghitungan objek pada Video3 yaitu 97.82 detik dari durasi video yang sebenarnya adalah 46 detik, Video4 mendapatkan waktu lamanya proses yaitu 35.32 detik dari durasi video sebenarnya adalah 20 detik. Jumlah *frame* pada kedua video adalah 30 *frame* per detik, namun Video1 hanya berhasil membaca sebanyak 14.21 *frame* per detik, Video2 mendeteksi 18.21 *frame* per detik. Jumlah total objek manusia yang ada di dalam Video3 dan Video4 sebanyak 22 objek, namun sistem hanya berhasil mendeteksi dan menghitung sebanyak 17 objek manusia.

3) Pengujian Ke-3

Pada tahap ini dilakukan pengujian pada empat video sekaligus dikarenakan memiliki kondisi yang sama. Video tersebut adalah Video5, Video6, Video7 dan Video8. Nilai intensitas cahaya dari keempat video ini adalah sekitar 200-300 lux, jarak objek terhadap kamera 2-3 meter dan sudut kamera dalam pengambilan video sekitar 35°. Keempat file video ini direkam pada dua lokasi yang berbeda, pertama di Supermarket Jumbo, kedua di pusat perbelanjaan Manado Town Square dengan kondisi objek manusia yang ramai.

Dari hasil pengujian keempat file video, Video5 mendapatkan 129.56 detik untuk lamanya proses deteksi dan penghitungan objek, dari durasi video yang sebenarnya adalah 30 detik. Video6 mendapatkan 105.86 detik untuk waktu proses, dari durasi video yang sebenarnya adalah 30 detik. Video7 mendapatkan 40.01 detik dari durasi video yang sebenarnya yaitu 24 detik. Video8 mendapatkan 126.51 detik untuk waktu proses deteksi dan penghitungan objek manusia, dari durasi video asli adalah 73 detik. Keempat video memiliki

jumlah pembacaan *frame* setiap 30 *frame* per detik, namun dari pengujian yang dilakukan, Video5 hanya mampu membaca sebanyak 7.05 *frame* per detik, Video6 8.66 *frame* per detik, Video7 sebanyak 19.40 *frame* per detik, dan yang terakhir Video8 hanya mampu membaca 16.41 *frame* per detik. Dari jumlah total 95 objek yang ada di dalam video, sistem hanya mampu mendeteksi 89 objek pada setiap *frame* yang terdeteksi. Artinya terdapat 6 objek yang tidak terhitung oleh sistem.

4) Pengujian Ke-4

Pada pengujian keempat ada dua video yang akan diuji secara bersama-sama, dikarenakan kondisi dari kedua video sama. Video tersebut adalah Video9 dan Video10. Intensitas cahaya dari kedua video berkisar antara 200-300 lux, jarak objek terhadap kamera 5-10 meter, dan sudut kamera dalam pengambilan video sekitar 60°. Kedua file video ini di rekam pada lokasi di salah satu pusat perbelanjaan di manado yaitu Mall manado, dengan kondisi objek manusia tidak terlalu ramai.

Dari hasil pengujian yang dilakukan, Video9 mendapatkan waktu proses deteksi dan penghitungan objek manusia yaitu 133.42 detik dari durasi video aslinya adalah 69 detik. Video10 mendapatkan 120.10 detik untuk melakukan proses deteksi dan penghitungan jumlah objek manusia dari durasi video asli adalah 61 detik. Dari 30 *frame* per detik sebagai default pembacaan *frame* dari kedua video, Video9 hanya mampu membaca 17.70 *frame* per detik, sedangkan untuk Video10 hanya mampu membaca 15.22 *frame* per detik. Total objek yang ada di dalam kedua video yaitu sebanyak 29 objek manusia, namun sistem hanya mampu menghitung 22 objek manusia, yang berarti ada 7 objek yang tidak terhitung ataupun terdeteksi oleh sistem.

5) Pengujian Ke-5

Pada pengujian kelima dilakukan uji coba pada dua video. Sama seperti pengujian yang sebelumnya, kedua video memiliki kondisi yang sama dengan intensitas cahaya berkisar antara 50-150 lux, jarak objek terhadap kamera 2-3 meter, dan sudut kamera dalam pengambilan objek sekitar 35°. Kedua file video ini direkam pada lokasi yang sama yaitu di pusat perbelanjaan Manado Town Square. Didalam kedua video ini terdapat objek lain selain manusia, yaitu Patung mannequin. Hal ini dilakukan untuk menguji apakah sistem dapat membedakan objek manusia dengan objek yang menyerupai manusia.

Waktu pemrosesan yang didapat dari hasil pengujian, Video11 mendapatkan 114.97 detik dari durasi video asli 37 detik dan Video12 mendapatkan 28.08 detik dari durasi video asli 15 detik. Untuk FPS, Video11 hanya dapat membaca sebanyak 9.24 *frame* per detik dari total 30 *frame* per detik, sedangkan untuk Video12 berhasil membacakan 17.35 *frame* per detik dari total 30 *frame* per detik. Total 12 objek yang berada di dalam video, tetapi sistem malah mendeteksi sebanyak 15 objek manusia. Hal ini dikarenakan oleh beberapa faktor, contohnya saja adanya pendeteksian ganda, dimana sistem mendeteksi objek secara berlebih.

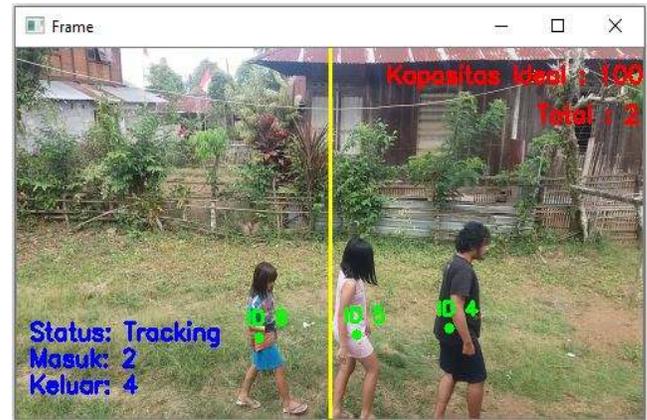
TABEL II TABEL PARAMETER YANG MEMPENGARUHI AKURASI SISTEM

No.	Nama Video	Parameter		
		Intensitas Cahaya	Jarak Objek Terhadap Kamera	Sudut Pengambilan Video
1	Video1.mp4	550 lux	5 meter	45°
2	Video2.mp4	550 lux	5 meter	45°
3	Video3.mp4	50-80 lux	2-3 meter	35°
4	Video4.mp4	50-80 lux	2-3 meter	35°
5	Video5.mp4	200-300 lux	2-3 meter	35°
6	Video6.mp4	200-300 lux	2-3 meter	35°
7	Video7.mp4	200-300 lux	2-3 meter	35°
8	Video8.mp4	200-300 lux	2-3 meter	35°
9	Video9.mp4	200-300 lux	5-10 meter	60°
10	Video10.mp4	200-300 lux	5-10 meter	60°
11	Video11.mp4	50-150 lux	2-3 meter	35°
12	Video12.mp4	50-150 lux	2-3 meter	35°

6) Pengujian Sistem Pada Spesifikasi Laptop dan Komputer yang Berbeda

Sistem yang dibangun juga akan diuji pada beberapa spesifikasi laptop atau komputer yang berbeda. Hal ini bertujuan untuk mengetahui kemampuan dari masing-masing spesifikasi laptop atau komputer dalam menjalankan sistem, dan sekaligus dapat mengetahui spesifikasi yang paling ideal diantaranya. Dalam hal ini, sistem akan diuji pada 4 spesifikasi laptop dan komputer yang berbeda. Hasil pengujian dapat dilihat pada tabel IX.

- 1) Laptop ideapad 110
 - Prosesor : AMD APU A9-9400
 - RAM : 4GB
 - Grafis Card : Radeon™ R5 Graphics
- 2) Laptop Asus A455L
 - Prosesor : Intel Core i3-5005U new 3th gen
 - RAM : 4GB
 - Grafis Card : nVIDIA Geforce 820m
- 3) Laptop Acer Aspire 3 A315-41-R66A
 - Prosesor : AMD Ryzen 3 2200U dual-core
 - RAM : 4GB
 - Grafis Card : AMD Radeon RX Vega 3 iGPU
- 4) Komputer
 - Prosesor : Proci i5 gen 9
 - RAM : 16GB
 - GPU : Nvidia Geforce Gtx 1070



Gambar 2 Tampilan sistem pendeteksian dan penghitungan jumlah manusia

B. Analisa Sistem

Pada tahapan ini akan dilakukan analisa sistem dengan tujuan untuk mengetahui seberapa banyak jumlah objek manusia yang terdeteksi benar dari keseluruhan objek yang terdeteksi oleh sistem. Di dalamnya juga akan dihitung tingkat akurasi dari sistem untuk melihat keefektifan model yang digunakan serta performa yang dihasilkan dengan menggunakan confusion matrix. Berikut merupakan parameter-parameter yang digunakan dalam pengukuran tingkat akurasi sistem:

- 1) *True Positive (TP)* adalah kondisi dimana objek manusia berhasil dideteksi oleh sistem.
- 2) *False Positive (FP)* adalah kondisi dimana sistem menganggap objek atau kondisi lain sebagai objek manusia.
- 3) *False Negative (FN)* adalah kondisi dimana ada objek manusia yang tidak terdeteksi oleh sistem.
- 4) *False In (FI)* merupakan kesalahan penghitungan objek pada sisi kanan atau sisi yang ditentukan sebagai arah masuk objek
- 5) *False Out (FO)* merupakan kesalahan penghitungan objek pada sisi kiri atau sisi yang ditentukan sebagai arah keluar objek
- 6) *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif dan salah negative. Rumus dari perhitungan recall dapat dilihat pada (1).

$$\text{Recal} = \frac{TP}{TP+FN} \quad (1)$$

- 7) *Precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Rumus dari perhitungan precision dapat dilihat pada (2).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

- 8) *F-Score* adalah nilai yang didapat untuk menunjukkan performa algoritma yang diterapkan. Rumus dari perhitungan f-score dapat dilihat pada (3).

$$\text{F-Score} = \frac{2x(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (3)$$

- 9) *Akurasi* adalah nilai rasio prediksi benar (positif) dengan keseluruhan data. Rumus dari perhitungan akurasi dapat dilihat pada (4).

$$\text{Akurasi} = \frac{TP}{TP+FN+FP+FI+FO} \times 100\% \quad (4)$$

1) *Analisa Pengujian ke-1*

Berikut merupakan perhitungan dari data analisa 1 yang ada pada tabel III.

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{15}{15+0} = \frac{15}{15} = 1$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{15}{15+1} = \frac{15}{16} = 0,94$$

$$\text{F-Score} = \frac{2x(\text{Precision} \times \text{Recall})}{\text{Precision}+\text{Recall}} = \frac{2x(0,94 \times 1)}{0,94+1} = \frac{1,88}{1,94} = 0,97$$

$$\text{Akurasi} = \frac{TP}{TP+FN+FP+FI+FO} \times 100\% = \frac{15}{16} \times 100\% = 93,75\%$$

Tingkat akurasi dari pendeteksian dan penghitungan objek manusia pada pengujian 1 mendapatkan nilai yang cukup tinggi, dengan nilai Recall 1, Precision mendapatkan nilai 0,94, F-Score dengan nilai 0,97 dan terakhir akurasinya mendapatkan nilai 93,75%. Hal ini dapat terjadi dikarenakan kondisi pencahayaan yang cukup baik, jarak objek terhadap kamera yang tidak terlalu jauh dan perspektif kamera yang tidak terlalu keatas ataupun kebawah. Namun dalam pengujian ini terjadi kesalahan deteksi sebanyak 1x yang diakibatkan adanya id objek yang ganda, yang membuat proses penghitungan objek mengalami kesalahan.

2) *Analisa Pengujian ke-2*

Berikut merupakan perhitungan dari data analisa 2 yang ada pada tabel IV.

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{17}{17+5} = \frac{17}{22} = 0,77$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{17}{17+0} = \frac{17}{17} = 1$$

$$\text{F-Score} = \frac{2x(\text{Precision} \times \text{Recall})}{\text{Precision}+\text{Recall}} = \frac{2x(1 \times 0,77)}{1+0,77} = \frac{1,54}{1,77} = 0,87$$

$$\text{Akurasi} = \frac{TP}{TP+FN+FP+FI+FO} \times 100\% = \frac{17}{22} \times 100\% = 77,27\%$$

Tingkat akurasi dari pendeteksian dan penghitungan objek manusia pada pengujian 2 mengalami penurunan jika dibandingkan dengan pengujian 1, dengan nilai Recall mendapatkan hasil sebesar 0,77, Precision mendapatkan nilai 1, F-Score dengan nilai 0,87 dan terakhir akurasinya mendapatkan nilai 77,27 %. Penurunan tingkat akurasi yang terjadi pada pengujian 2 ini disebabkan karena pencahayaan yang cukup rendah, objek yang bergerak cepat, terdapat beberapa objek yang memiliki jarak cukup jauh dari jangkauan kamera, serta adanya beberapa objek yang hanya diam ditempat sehingga menyebabkan kesalahan pemberian id objek oleh sistem yang membuat proses penghitungan objek manusia menjadi salah.

3) *Analisa Pengujian ke-3*

Berikut merupakan perhitungan dari data analisa 3 yang ada pada tabel V.

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{76}{76+15} = \frac{76}{91} = 0,83$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{76}{76+9} = \frac{76}{85} = 0,89$$

$$\text{F-Score} = \frac{2x(\text{Precision} \times \text{Recall})}{\text{Precision}+\text{Recall}} = \frac{2x(0,89 \times 0,83)}{0,89+0,83} = \frac{1,48}{1,72} = 0,86$$

$$\text{Akurasi} = \frac{TP}{TP+FN+FP+FI+FO} \times 100\% = \frac{76}{105} \times 100\% = 71,43\%$$

Tingkat akurasi dari pendeteksian dan penghitungan objek manusia pada pengujian 3 mendapatkan nilai Recall sebesar

0,83, Precision mendapatkan nilai 0,89, F-Score dengan nilai 0,86 dan terakhir akurasinya mendapatkan nilai 71,43 %. Hasil akurasi pada pengujian ini bisa dibilang cukup tinggi, hal ini dipengaruhi oleh pencahayaan yang cukup baik serta jarak objek yang tidak terlalu jauh dari jangkauan kamera. Namun tetap saja terdapat beberapa kesalahan deteksi dan penghitungan objek manusia yang disebabkan karena sudut kamera dalam pengambilan video yang dapat dikatakan rendah, sehingga ketika ada objek yang datang secara bergerombolan, sistem akan kesulitan dalam melakukan deteksi dan penghitungan objek secara benar.

4) *Analisa Pengujian ke-4*

Berikut merupakan perhitungan dari data analisa 4 yang ada pada tabel VI.

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{19}{19+8} = \frac{19}{27} = 0,70$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{19}{19+1} = \frac{19}{20} = 0,95$$

$$\text{F-Score} = \frac{2x(\text{Precision} \times \text{Recall})}{\text{Precision}+\text{Recall}} = \frac{2x(0,95 \times 0,70)}{0,95+0,70} = \frac{1,33}{1,65} = 0,81$$

$$\text{Akurasi} = \frac{TP}{TP+FN+FP+FI+FO} \times 100\% = \frac{19}{30} \times 100\% = 63,33\%$$

Tingkat akurasi dari pendeteksian dan penghitungan objek manusia pada pengujian 4 mengalami penurunan jika dibandingkan dengan pengujian yang lain. Nilai Recall mendapatkan hasil 0,70, Precision mendapatkan nilai 0,95, F-Score dengan nilai 0,81 dan terakhir akurasinya mendapatkan nilai 63,33 %. Penurunan tingkat akurasi yang terjadi pada pengujian 4 diakibatkan karena jarak objek yang terlalu jauh dari jangkauan kamera. Hal ini dapat membuat sistem akan mengalami kesulitan dalam mendeteksi objek manusia yang ada di dalam *frame* video, sehingga hasil penghitungan objek juga akan menjadi keliru.

5) *Analisa Pengujian ke-5*

Berikut merupakan perhitungan dari data analisa 5 yang ada pada tabel VII.

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{9}{9+2} = \frac{9}{11} = 0,81$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{9}{9+6} = \frac{9}{15} = 0,6$$

$$\text{F-Score} = \frac{2x(\text{Precision} \times \text{Recall})}{\text{Precision}+\text{Recall}} = \frac{2x(0,6 \times 0,81)}{0,6+0,81} = \frac{0,97}{1,41} = 0,67$$

$$\text{Akurasi} = \frac{TP}{TP+FN+FP+FI+FO} \times 100\% = \frac{9}{18} \times 100\% = 50\%$$

Tingkat akurasi dari pendeteksian dan penghitungan objek manusia pada pengujian 5 merupakan hasil akurasi terendah jika dibandingkan dengan hasil pada pengujian-pengujian sebelumnya, dengan nilai Recall mendapatkan hasil 0,81, Precision mendapatkan nilai 0,6, F-Score dengan nilai 0,67 dan terakhir akurasinya mendapatkan nilai 50%. Pada pengujian ini, penulis sengaja melakukan perekaman video pada tempat yang mempunyai objek lain yang menyerupai bentuk manusia, untuk mengetahui apakah sistem akan tetap mendeteksi objek tersebut atau mengabaikannya. Dari hasil pengujian yang dilakukan, ternyata objek tersebut masih terdeteksi oleh sistem. Karena cara sistem mendeteksi suatu objek dengan mengenali bentuk dari objek tersebut, sehingga jika objek tersebut masih

berbentuk persis dengan sasaran objek yang akan dideteksi, maka sistem masih akan mengenali objek tersebut sebagai sasaran objek deteksi.

C. Evaluasi Sistem

Nilai tingkat akurasi dari masing-masing pengujian dapat dilihat pada tabel VIII beserta dengan parameter-parameter yang mempengaruhinya. Pada tabel VIII, dapat dilihat bahwa tingkat akurasi dari tiap pengujian dipengaruhi oleh parameter-parameter yang ada didalam tiap video yaitu, intensitas cahaya, jarak objek terhadap kamera, dan sudut kamera dalam pengambilan video rekaman. Tetapi selain tiga parameter tersebut, terdapat juga faktor lain yang menyebabkan perbedaan dari tingkat akurasi yang dihasilkan, diantaranya kecepatan Bergeraknya objek manusia di dalam *frame* video, dan adanya objek lain yang menyerupai bentuk manusia seperti yang ada pada pengujian kelima.

Perbandingan akurasi dan parameter pada sistem juga divisualisasikan kedalam bentuk grafik yang dapat dilihat pada gambar 3. Seperti yang terlihat pada gambar 3, Grafik pada pengujian 1 mendapatkan nilai akurasi yang paling tinggi. Hal ini dipengaruhi oleh intensitas cahaya yang tinggi, serta jarak dan perspektif kamera yang stabil. Masuk pada grafik pengujian 2, nilai akurasi mengalami penurunan, hal ini diikuti dengan menurunnya nilai dari intensitas cahaya, jarak, dan perspektif kamera. Pada grafik pengujian 3 nilai akurasi semakin menurun, tetapi grafik pada nilai lux mengalami kenaikan walaupun tidak setinggi nilai nilai pada grafik pengujian 1, sedangkan untuk grafik 2 parameter lainnya yaitu jarak dan perspektif masih pada nilai yang sama. Nilai akurasi semakin menurun sampai pada grafik pengujian terakhir, yaitu pada pengujian 5. Tetapi nilai dari kedua parameter sempat naik pada grafik pengujian 4, yaitu jarak objek terhadap kamera dan perspektif kamera, namun hal ini tidak membuat meningkatnya nilai akurasi pada pengujian 4.

Dari grafik perbandingan pada gambar 3 dapat disimpulkan

TABEL III TABEL DATA ANALISA PENGUJIAN 1

No	Aktual	Sistem	TP	FP	FN	FI	FO
1.	1	1	1	-	-	-	-
2.	1	1	1	-	-	-	-
3.	1	1	1	-	-	-	-
4.	3	3	3	-	-	-	-
5.	3	3	3	-	-	-	-
6.	3	3	3	-	-	-	-
7.	3	4	3	1	-	-	-
Total			15	1	0	0	0

TABEL IV TABEL DATA ANALISA PENGUJIAN 2

No	Aktual	Sistem	TP	FP	FN	FI	FO
1.	1	1	1	-	-	-	-
2.	1	1	1	-	-	-	-
3.	1	1	1	-	-	-	-
4.	3	3	3	-	-	-	-
5.	3	2	2	-	1	-	-
6.	3	1	1	-	2	-	-
7.	3	2	2	-	1	-	-
8.	3	2	2	-	1	-	-
9.	4	4	4	-	-	-	-
Total			17	0	5	0	0

bahwa, untuk mendapatkan pendeteksian yang akurat serta nilai akurasi yang tinggi, haruslah ditunjang dengan kondisi pencahayaan yang terang. Hal ini dikarenakan, pencahayaan yang tinggi dapat memberikan kejelasan gambar dan perbedaan kromatik warna antar pixel yang cukup kontras, sehingga semakin memperlancar proses pemisahan antara objek dan latar. Selain itu, jarak objek terhadap kamera serta perspektif kamera juga dapat mempengaruhi keakuratan dalam proses deteksi. Dapat dilihat pada grafik pengujian 3, walaupun nilai lux terbilang cukup tinggi namun nilai akurasi masih terus menurun dikarenakan sudut kamera dalam pengambilan video tidaklah ideal. Sudut kamera masihlah terlalu rendah sehingga jika ada objek yang datang secara bergerombolan, maka sistem akan susah membedakan antara satu objek dengan objek yang lain. Lain halnya dengan grafik pada pengujian 4, nilai lux dan sudut kamera dapat dikatakan cukup tinggi, namun dikarenakan jarak objek terhadap kamera yang cukup sehingga membuat sistem tidak dapat mengenali objek secara jelas yang membuat proses deteksi tidak akurat. Untuk grafik pada pengujian 5, menurunnya tingkat akurasi disebabkan karena adanya objek lain yang berbentuk seperti objek manusia. Kemiripan bentuk objek tersebut dengan objek manusia membuat sistem keliru dalam mengenali target objek yang akan dideteksi, sehingga membuat sistem secara otomatis memberikan id kepada objek tersebut dan terjadi kesalahan pendeteksian.

TABEL V TABEL DATA ANALISA PENGUJIAN 3

No	Aktual	Sistem	TP	FP	FN	FI	FO
1.	27	31	26	4	-	-	1
2.	23	28	23	5	-	-	-
3.	3	2	2	-	1	-	-
4.	1	1	1	-	-	-	-
5.	2	1	1	-	1	-	-
6.	2	2	0	-	-	-	2
7.	2	2	2	-	-	-	-
8.	5	3	3	-	2	-	-
9.	2	1	1	-	1	-	-
10.	7	4	4	-	3	-	-
11.	7	4	4	-	3	-	-
12.	2	1	1	-	1	-	-
13.	1	1	1	-	-	-	-
14.	11	8	7	-	3	-	1
Total			76	9	15	0	5

TABEL VI TABEL DATA ANALISA PENGUJIAN 4

No	Aktual	Sistem	TP	FP	FN	FI	FO
1.	6	4	4	-	2	-	-
2.	4	1	1	-	3	-	-
3.	6	3	3	-	3	-	-
4.	1	1	1	-	-	-	-
5.	7	7	5	-	-	-	2
6.	5	6	5	1	-	-	-
Total			19	1	8	0	2

TABEL VII TABEL DATA ANALISA PENGUJIAN 5

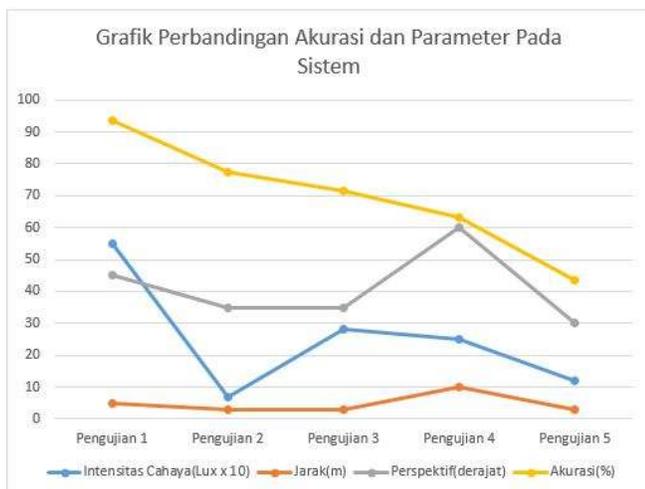
No	Aktual	Sistem	TP	FP	FN	FI	FO
1.	1	5	1	4	-	-	-
2.	2	4	2	2	-	-	-
3.	2	2	1	-	-	1	-
4.	2	1	1	-	1	-	-
5.	5	4	4	-	1	-	-
Total			9	6	2	1	0

TABEL VIII TABEL PERNDANDINGAN TINGKAT AKURASI DARI PENGUJIAN 1-5 BESERTA PARAMETER-PARAMETER YANG MEMPENGARUHI

	Pengujian 1	Pengujian 2	Pengujian 3	Pengujian 4	Pengujian 5
Lux	550	50-80	200-300	200-300	50-150
Jarak (m)	5	2-3	2-3	5-10	2-3
Perspektif	45°	35°	35°	60°	35°
Recall	1	0,77	0,83	0,70	0,81
Precision	0,94	1	0,89	0,95	0,6
F-Score	0,97	0,87	0,86	0,81	0,67
Akurasi	93,75%	77,27%	71,43%	63,33%	50%

TABEL IX TABEL PERBANDINGAN PENGUJIAN DARI 4 SPESIFIKASI LAPTOP DAN KOMPUTER YANG BERBEDA

No.	Nama File	Laptop Lenovo ideapad 110 (Processor:AMD APU A9-9400 up to 3.2 Ghz RAM: 4GB Graphic Card: Radeon™ R5 Graphics)	Laptop Asus A455L (Prosesor: Intel Core i3-5005U new 3th Gen RAM: 4GB Grafis nVIDIA GeForce 820m)	Laptop Acer Aspire 3 (Prosesor: AMD Ryzen 3 2200U dual-core up to 3,4GHz RAM: 4GB Grafis Card: AMD Radeon RX Vega 3 iGPU)	Komputer (Prosesor: Proci i5 gen 9 RAM: 16GB GPU: Nvidia Geforce Gtx 1070)				
		Waktu Proses	Frame terbaca per detik	Waktu Proses	Frame terbaca per detik	Waktu Proses	Frame terbaca per detik	Waktu Proses	Frame terbaca per detik
1.	Video1	128.28	14.99	125.29	15.35	80.56	23.87	47.15	40.79
2.	Video2	59.68	12.92	56.44	13.66	38.05	20.26	22.73	33.92
3.	Video3	97.82	14.21	86.47	16.07	58.25	23.86	36.74	37.84
4.	Video4	35.32	18.21	37.31	17.23	26.60	24.17	16.23	39.61
5.	Video5	129.56	7.05	135.94	6.72	100.07	9.13	52.78	17.32
6.	Video6	105.86	8.66	116.57	7.87	81.71	11.22	44.82	20.46
7.	Video7	40.01	19.40	36.54	21.23	26.06	29.77	15.89	48.84
8.	Video8	126.51	16.41	116.53	17.82	84.26	24.64	50.28	41.29
9.	Video9	133.42	15.70	109.44	19.14	70.80	29.59	43.59	48.06
10.	Video10	120.10	15.22	108.57	16.84	71.39	25.61	45.67	40.03
11.	Video11	114.97	9.24	111.63	9.51	82.28	12.91	46.66	22.76
12.	Video12	28.08	17.35	30.07	16.19	20.91	23.29	12.67	38.43



Gambar 3 Grafik Perbandingan Akurasi dan Parameter Penghitungan Objek

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan pada sistem yang dibangun, maka diambil kesimpulan bahwa sistem dapat mendeteksi objek manusia yang ada didalam frame video

dengan cukup baik. Pengujian sistem dilakukan 5 kali dengan jumlah video sebanyak 12 file video. Kondisi video dari tiap pengujian memiliki kondisi yang berbeda-beda, mulai dari intensitas cahaya, jarak objek terhadap kamera, dan sudut kamera dalam pengambilan video. Sistem mampu mendeteksi dan menghitung jumlah objek manusia dengan tingkat akurasi tertinggi mencapai 93,75%.

B. Saran

Beberapa saran penelitian untuk mengembangkan penelitian yang telah dilakukan, antara lain input didalam sistem yang dibangun hanya menggunakan file-file video rekaman, artinya sistem tidak mendeteksi objek manusia secara real-time. Untuk itu diharapkan dapat dikembangkan untuk mendeteksi objek manusia secara real-time. Pada penelitian ini sistem hanya dapat mendeteksi objek manusia, untuk itu pada penelitian selanjutnya diharapkan sistem dapat dikembangkan untuk melakukan proses pengenalan dan pengklasifikasian terhadap identitas objek manusia yang terdeteksi

I. KUTIPAN

- [1] B. Umam, Khairul Negara; Sukma, "Deteksi Obyek Manusia Pada Basis Data Video Menggunakan Metode Background Subtraction

- Dan Operasi Morfologi,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 2, pp. 31–40, 2016, doi: 10.24014/coreit.v2i2.2391.
- [2] R. Hasugian, Ferdinan; Rizal, Achmad; Magdalena, “IMPLEMENTASI PERANGKAT LUNAK PENGHITUNG JUMLAH ORANG BERBASIS PENGOLAHAN CITRA MENGGUNAKAN WEBCAM (SOFTWARE IMPLEMENTATION COUNTING PEOPLE USING WEBCAM BASED IMAGE PROCESSING),” *Telkom Univ.*, 2008.
- [3] A. N. A. T. R. Adhitama, “Real-Time Object Detection For Wayang Punakawan Identification Using Deep Learning,” *J. Infotel*, vol. 11, no. 4, pp. 127–132, 2019, doi: 10.20895/infotel.v11i4.455.
- [4] A. N. Kusumanto, RD; Tompunu, “PENGOLAHAN CITRA DIGITAL UNTUK MENDETEKSI OBYEK MENGGUNAKAN PENGOLAHAN WARNA MODEL NORMALISASI RGB,” *Semin. Nas. Teknol. Inf. Komun. Terap. 2011 (Semantik 2011)*, 2011, doi: 10.1016/S0166-1116(08)71924-1.
- [5] S. R. DEWI, “DEEP LEARNING OBJECT DETECTION PADA VIDEO MENGGUNAKAN TENSORFLOW DAN CONVOLUTIONAL NEURAL NETWORK,” 2018.
- [6] F. P. Tiara, “ANALISIS ALGORITMA EIGENFACE (PENGENALAN WAJAH) PADA APLIKASI KEHADIRAN PENGAJARAN DOSEN,” 2011.
- [7] J. McCarthy, “What Is Artificial Intelligence Anyway,” *Computer Science Department*, pp. 1–15, 2007.
- [8] A. L. Samuel, “Some studies in machine learning using the game of checkers. II-Recent progress,” *IBM*, vol. 6, no. PART 1, pp. 1–36, 1967, doi: 10.1016/0066-4138(69)90004-4.
- [9] M. A. Ponti, L. S. F. Ribeiro, and T. S. Nazare, “Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask,” *IEEE Comput. Soc.*, pp. 17–41, 2017, doi: 10.1109/SIBGRAP-T.2017.12.
- [10] F. Jalled, “Object Detection using Image Processing,” vol. 1, pp. 1–6, 2016, [Online]. Available: <http://arxiv.org/abs/1611.07791>.
- [11] A. C. Liu, Wei; Anguelov, Dragomir; Erhan, Dumitru; Szegedy, Christian; Reed, Scott; Fu, Cheng Yang; Berg, “SSD: Single Shot MultiBox Detector,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 1–17, 2016, doi: 10.1007/978-3-319-46448-0_2.
- [12] K. Kautsar, Havez Vazirani Al; Adi, “IMPLEMENTASI OBJECT TRACKING UNTUK MENDETEKSI DAN MENGHITUNG JUMLAH KENDARAAN SECARA OTOMATIS MENGGUNAKAN METODE KALMAN FILTER DAN GAUSSIAN MIXTURE MODEL,” *Youngster Phys. J.*, vol. 5, no. 1, pp. 13–20, 2016.
- [13] H. Howard, Andrew G; Zhu, Menglong; Chen, Bo; Kalenichenko, Dmitry; Wang, Weijun; Weyand, Tobias; Andreetto, Marco; Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *arXiv*, 2017.

TENTANG PENULIS



Penulis bernama lengkap Tresya Anjali Dompeipen lahir di Tondano pada tanggal 9 Agustus 1998. Penulis menempuh pendidikan pertama di TK GMIM ELIM Papakelan pada tahun 2003 – 2004, kemudian melanjutkan ke SD GMIM ELIM Papakelan 2004 – 2010, setelah itu melanjutkan studi di SMP Negeri 7

Tondano pada tahun 2010 – 2013, kemudian melanjutkan pendidikan di SMK Negeri 2 Tondano pada tahun 2013 – 2016. Pada tahun 2016, penulis melanjutkan studi di Program Studi Teknik Informatika, Fakultas Teknik, Jurusan Teknik Elektro, Universitas Sam Ratulangi Manado.