

Kombinasi Deteksi Objek, Pengenalan Wajah dan Perilaku Anomali menggunakan *State Machine* untuk Kamera Pengawas

MUHAMMAD FAUZI NURYASIN, CARMADI MACHBUB, LENNI YULIANTI

Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung, Indonesia
Email : mfauzinuryasin@gmail.com

Received 25 Oktober 2022 | *Revised* 24 November 2022 | *Accepted* 29 November 2022

ABSTRAK

Saat ini sistem kamera pengawas mengandalkan manusia dalam melakukan penerjemahan pada rekaman gambar yang terjadi. Perkembangan computer vision, machine learning, dan pengolahan citra dapat dimanfaatkan untuk membantu peran manusia dalam melakukan pengawasan. Penelitian ini merancang sistem kerja kamera yang terdiri dari tiga modul yaitu deteksi objek, pengenalan wajah, dan perilaku anomali. Deteksi objek memakai HOG-SVM, pengenalan wajah menggunakan CNN dengan arsitektur VGG-16 memanfaatkan transfer learning, dan perilaku anomali memakai spatiotemporal autoencoder berdasarkan threshold. Ketiga modul tersebut diuji menggunakan metrik akurasi, presisi, recall, dan f1-score. Ketiga modul diintegrasikan dengan state machine menjadi satu kesatuan sistem. Kinerja modul memiliki akurasi 88% untuk deteksi objek, 98% untuk pengenalan wajah, dan 78% untuk perilaku anomali. Hasil tampilan riil dapat diakses secara sederhana dan nirkabel melalui web.

Kata kunci: HOG-SVM, CNN, VGG-16, spatiotemporal autoencoder, state machine

ABSTRACT

Nowadays, the surveillance camera system relies on human to interpret the recorded images. Computer vision, machine learning, and image processing can be utilized to assist the human role in supervising. This study designed a camera work system consisting of three main modules, namely object detection, face recognition, and anomaly behavior. Object detection used the HOG-SVM combination. Facial recognition used CNN with the VGG-16 architecture that utilized transfer learning. Anomalous behavior used spatiotemporal autoencoder based on threshold. Modules are tested using the metrics of accuracy, precision, recall, and f1-score. The three modules are integrated using a state machine into one system. The performance of the module had 88% accuracy for object detection, 98% for facial recognition, and 78% for anomalous behavior. Real time video recording can be accessed wireless via web-based.

Keywords: HOG-SVM, CNN, VGG-16, spatiotemporal autoencoder, state machine

1. PENDAHULUAN

Perkembangan peradaban manusia selalu sejalan dengan perkembangan ilmu pengetahuan. Salah satu manfaat kemajuan ilmu pengetahuan adalah majunya teknologi yang mempermudah kehidupan manusia. Teknologi dapat menciptakan keamanan bagi manusia. Salah satu teknologi keamanan yang tengah pesat berkembang saat ini adalah sistem pengawasan kamera. Saat ini, Sistem pengawasan kamera telah diterapkan pada peradaban manusia selama bertahun-tahun. Penggunaan sistem pengawasan pertama kali menggunakan kamera yang merekam suatu kejadian lalu mengirimkan rekaman yang ditangkap ke layar tabung pada sekitaran tahun 1940 (**Shidik, dkk, 2019**). Sistem pengawasan kamera terdiri dari tindakan mengamati pergerakan dan perilaku dari pergerakan tersebut. Sistem pengawasan kamera biasanya digunakan untuk merekam kejadian yang bertujuan agar dapat mencegah pencurian, perlindungan fasilitas, keselamatan pekerja, tempat parkir, rekaman lalu lintas, dan keselamatan umum (**Fennelly & Perry, 2016**). Beberapa dekade terakhir, teknologi sistem pengawasan telah berkembang dari sistem analog menjadi digital atau hybrid. Sistem analog menggunakan kamera untuk mengirimkan sinyal analog kepada penyimpanan, biasanya memakai kabel *coax* (**Kalbo, dkk, 2020**). Seiring dengan perkembangan teknologi, keamanan pengiriman data membaik karena dikirim secara daring (**Puvvadi, dkk, 2015**). Pada saat ini, sistem pengawasan kamera pada umumnya melakukan perekaman setiap saat lalu menyimpannya pada media penyimpanan yang memerlukan banyak sekali kapasitas penyimpanan untuk dilakukan analisis (**Miao, dkk, 2016**). Sistem pengawasan pada umumnya masih amat bergantung pada pengamatan manusia dimana terdapat kemungkinan terjadinya distraksi serta kelelahan. Jika melihat dari segi ekonomi, penggunaan manusia dalam sistem pengamatan terbilang membebani biaya dengan efisiensi yang rendah. Manusia mudah mengalami kelelahan dibandingkan komputer atau mesin yang dapat bekerja 24 jam. Pada saat ini maka diperlukan inovasi pada sistem pengawasan untuk memperbaiki kekurangan dan memaksimalkan kinerja sistem pengawasan. Sebenarnya, perkembangan sistem pengawasan dapat dikembangkan dengan pesat akibat dari riset-riset yang berbasis aplikasi dimana mengintegrasikan *computer vision*, *machine learning*, dan pengolahan citra.

Pada umumnya, pengawasan masih mengandalkan manusia dalam melakukan pengamatan pada rekaman yang terjadi. Saat ini, seiring berkembangnya riset mengenai *computer vision*, *machine learning*, dan pengolahan citra dapat diharapkan membantu peran manusia dalam melakukan pengawasan. Teknologi *computer vision*, *machine learning*, dan pengolahan citra menjadi topik penelitian yang banyak diminati saat ini. Perkembangan teknologi *computer vision* telah banyak dikembangkan untuk keperluan identifikasi suatu perilaku atau keadaan (**Fang, dkk, 2020**). Perspektif riset *computer vision* memiliki beragam macam diantara lain mulai dari ekstraksi rekaman data mentah lalu digabungkan dengan teknik pengolahan citra, deteksi objek, *machine learning*, dan grafik komputer (**Wiley & Lucas, 2018**). Selain *computer vision*, perkembangan sistem pengawasan saat ini memanfaatkan *machine learning*. *Machine learning* merupakan studi yang menitik beratkan penggunaan algoritma komputer untuk mengubah data empiris menjadi suatu model yang dapat dimanfaatkan (**Edgar & Manz, 2017**). Terakhir, perkembangan riset yang menunjang kemajuan teknologi sistem pengawasan adalah pengolahan citra. Pengolahan citra merupakan suatu teknik algoritma untuk mengubah suatu penampilan visual atau sifat geometris dari gambar untuk ditransformasikan menjadi suatu himpunan data (**Aber, dkk, 2019**). Perkembangan *computer vision*, *machine learning*, dan pengolahan citra merupakan salah satu topik penelitian yang bisa terus berkembang belakangan ini.

Beberapa penelitian sebelumnya telah dilakukan di laboratorium sistem kendali dan komputer ITB yang berkaitan dengan penelitian ini antara lain penggunaan kombinasi viola-jones dan

algoritma *template matching* untuk deteksi wajah serta ditambahkan servo pada kamera sebagai kontrolnya (Yosafat, dkk, 2017), kombinasi fitur wajah dan postur pada manusia untuk pelacakan objek berupa manusia yang bergerak (Andriana, dkk, 2017), pelacakan objek manusia dengan kombinasi algoritma *Mobilenet*, pelacak KCF, dan fitur HOG (Kinasih, dkk, 2019), prediksi pergerakan manusia dengan algoritma *single shot multibox detector* yang ditambahkan *particle filter* (Maharani, dkk, 2020), pelacakan titik sentroid pada deteksi objek (Kinasih, dkk, 2020), dan peningkatan performa real time pada rekognisi wajah (Maharani, dkk, 2020). Pada saat ini, penelitian sebelumnya hanya menerapkan bagian-bagian tertentu seperti pengembangan algoritma, pelacakan gerakan, deteksi objek, dan deteksi wajah. Penelitian tersebut belum dimanfaatkan untuk dijadikan suatu kombinasi dan satu keutuhan suatu sistem dan diimplementasikan pada sistem pengawasan.

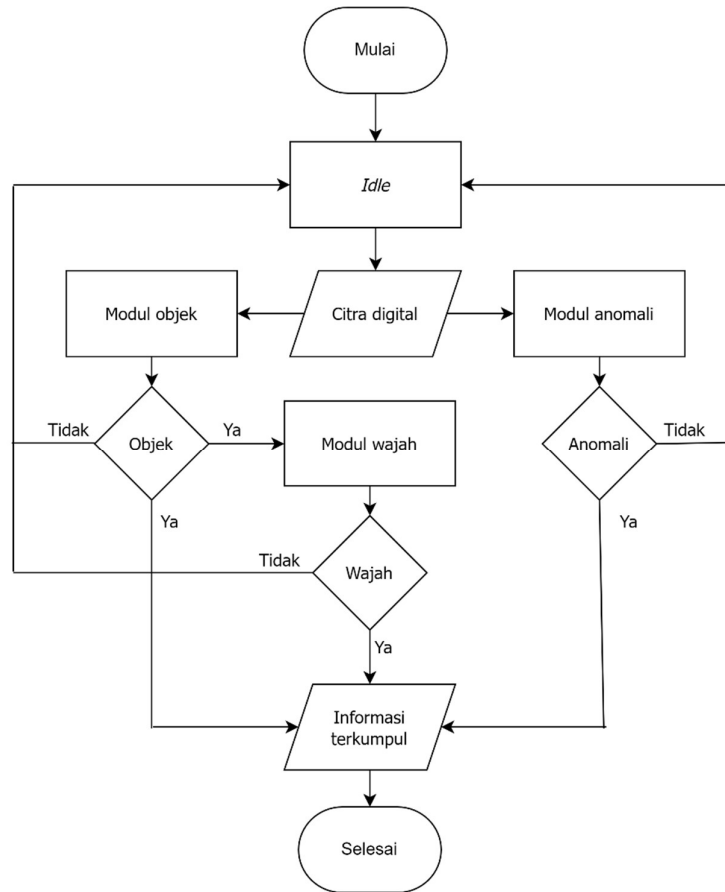
Berdasarkan latar belakang di atas, pada penelitian ini penulis merancang dan mempersatukan fitur-fitur yang telah dikembangkan oleh pekerjaan peneliti sebelumnya, dan juga menambahkan fitur lain untuk dijadikan suatu kombinasi algoritma untuk sistem pengawasan. Penelitian ini difokuskan pada rancang bangun *state machine* dan kebaruan penerapan kombinasi algoritma deteksi objek, pengenalan wajah, dan deteksi anomali melalui kamera. *State machine* yang dimaksud adalah pemanfaatan *state machine* untuk perpindahan *state* antar modul pada sistem yang terintegrasi. Penelitian ini diharapkan menciptakan sistem pengawasan yang dapat mengetahui anomali pada suatu keadaan di lingkungan tertentu contohnya digunakan untuk pengawasan rumah dimana yang berarti terjadi kejadian anomali berdasarkan batas nilai *threshold* perubahan citra yang terekam jika mengalami perubahan drastis antar setiap potongan citra. Selain itu, sistem dapat mendeteksi objek yang tertangkap kamera serta mengenali wajah seseorang jika tertangkap kamera.

2. METODE

2.1 Perancangan Sistem

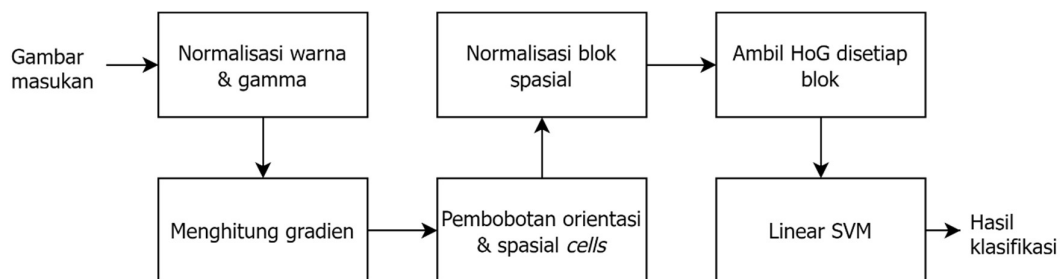
Pada penelitian ini, penulis merancang sistem yang terdiri dari perangkat keras dan perangkat lunak. Perangkat keras yang digunakan antara lain komputer *desktop*, Raspberry Pi 4, dan kamera Logitech C270. Pada perangkat lunak, penulis menggunakan bahasa pemrograman Python untuk membuat modul-modul dan integrasinya. Adapun *library opencv* dan *tensorflow* membantu untuk pengolahan citra, *computer vision*, dan *machine learning*. Sistem memiliki kerangka yang terdiri dari tiga elemen penting. Elemen penting tersebut diantara lain yakni modul objek, modul wajah, dan modul anomali. Modul objek berfungsi sebagai bagian dari sistem yang mendeteksi dan mengenali keberadaan objek yang tertangkap oleh sistem kamera. Modul wajah berguna sebagai bagian yang mengenali wajah seseorang ketika tertangkap oleh kamera. Terakhir, modul anomali berfungsi mengenali perilaku anomali terhadap lingkungan sistem yang diamati.

Hal yang sangat berpengaruh selain perangkat lunak terhadap performa sistem ini adalah perangkat keras yang digunakan, jika semakin canggih maka hasilnya akan semakin bagus. Namun hal ini perlu memikirkan antara biaya yang dikeluarkan dengan hasil yang didapatkan. Hal lain yang perlu diperhatikan adalah akses dan fleksibilitas sistem yang tentunya akan berbeda tiap perangkat keras yang dipilih. Ilustrasi diagram alir kerangka sistem ditunjukkan oleh Gambar 1. Ketika seluruh modul dan sistem berjalan dan berfungsi, implementasi sistem dapat diakses berbasis *web based*. Gambar 1 menunjukkan proses dari awal hingga selesai dari sistem yang dirancang.



Gambar 1. Diagram Alir Kerangka Sistem

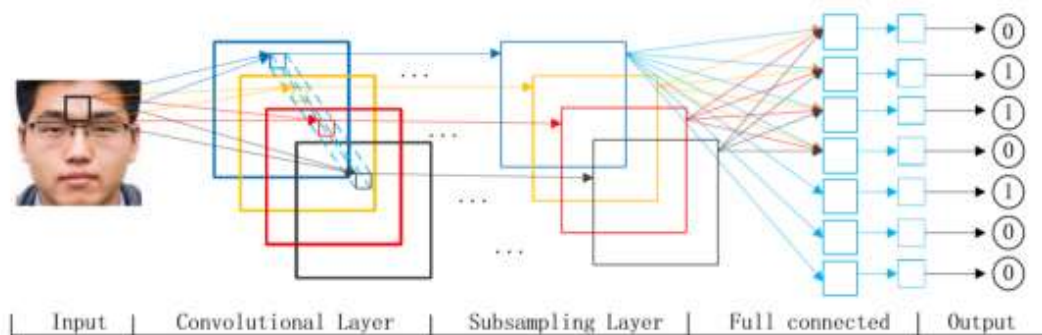
Pada modul deteksi objek memanfaatkan metode *Histogram of Oriented Gradients* (Dalal & Triggs, 2005) dan *Support Vector Machine* (Cortes & Vapnik, 1995) (HOG-SVM). HOG-SVM merupakan gabungan dari fitur deskripsi HOG dengan fitur klasifikasi SVM. Pada pemanfaatan HOG-SVM untuk deteksi objek, pertama-pertama fitur citra diekstraksi menggunakan deskriptor HOG. HOG bisa dideskripsikan sebagai suatu metode untuk mengekstraksi fitur suatu gambar dan mengelompokkan nilai-nilai piksel gradiennya berdasarkan orientasi arah disetiap bagian lokal dari gambar. Bentuk serta tampilan dari suatu gambar ternyata dapat dikarakterisasi dengan baik melalui distribusi gradien intensitas lokal atau disebut arah tepian meskipun tidak diketahui posisi gradien dan tepi tersebut. Blok diagram langkah-langkah fitur pada HOG-SVM ditunjukkan Gambar 2.



Gambar 2. Blok Diagram HOG-SVM

Berdasarkan Gambar 2, pertama citra masukan dilakukan normalisasi warna dan gamma. Citra dibagi ke beberapa potongan kecil yang saling terhubung yang disebut *cells*. Setiap bagian *cells* dilakukan komputasi histogram dari gradien arah atau tepian orientasinya untuk piksel yang ada di dalam *cell*. Selanjutnya, setiap piksel pada *cell* memiliki pengaruh terhadap pembobotan gradien. Sekumpulan *cells* yang berdekatan dapat dikategorikan menjadi wilayah spasial yang dinamakan blok. Pengelempokan dari *cells* menjadi blok ini merupakan dasar untuk pengelompokan dan normalisasi histogram. Terakhir, sejumlah histogram yang telah dinormalisasi merepresentasikan blok histogram. Sekumpulan blok histogram ini mempresentasikan sebagai deskriptor citra. Pada Gambar 2 terlihat bahwa metode HOG ini melakukan perhitungan gradien citra. Hasil akhir yang didapatkan menjadi satu vektor yang disebut vektor fitur HOG. Setelah diperoleh nilai vektor fitur HOG, maka dilakukan proses selanjutnya dengan linier SVM. SVM berkerja dengan cara memisahkan antara poin data satu sama lainnya dengan memperhatikan nilai margin maksimalnya.

Selanjutnya, pada modul pengenalan wajah menggunakan metode *Convolutional Neural Network* (CNN) dengan memanfaatkan *transfer learning* dan arsitektur VGG16. *Convolutional Neural Network* (CNN) merupakan bagian dari *deep learning* yang banyak digunakan berkaitan dalam proses citra. *Convolutional neural network* merupakan kelas dari *deep feed forward artificial neural network* yang telah dimanfaatkan untuk menghasilkan kinerja yang akurat dalam *computer vision*, seperti klasifikasi dan deteksi citra (Krizhevsky, dkk, 2012). Jaringan saraf pada CNN diatur secara volumetrik berupa tinggi, lebar, dan kedalaman. CNN terdiri dari beberapa lapisan atau *layer*. Pada Gambar 3, lapisan CNN terdiri dari *convolutional layer*, *pooling layer*, dan *fully connected layer*.

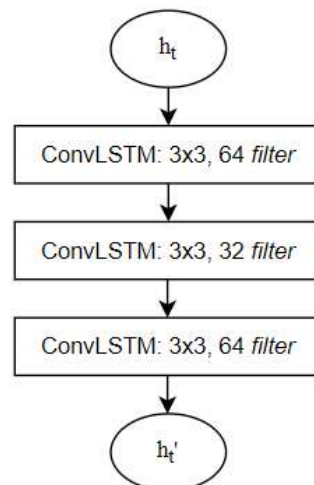


Gambar 3. Arsitektur CNN pada Rekognisi Wajah (Yuan, dkk, 2017)

Convolutional layer merupakan lapisan pertama yang menerima masukan citra misalnya melalui kamera. Pada lapisan tersebut terjadi operasi konvolusi, yaitu mengkombinasikan *filter* linier terhadap bagian lokal citra. Proses konvolusi memanfaatkan apa yang disebut dengan filter dengan ukuran panjang piksel, lebar piksel dan tebal piksel sesuai dengan data citra yang dimasukan. Nilai dari *filter* ini menjadi parameter yang akan diperbaharui dalam proses pembelajaran mesin. Ketiga *filter* ini akan bergeser keseluruhan bagian citra masukan. Pergeseran tersebut akan melakukan operasi dot matriks antara nilai masukan dan *filter* sehingga menghasilkan keluaran yang disebut *feature map*. *Pooling layer* menerima keluaran dari *convolution layer* yang telah ditambahkan dengan fungsi aktivasi. Pada lapisan ini ukuran data citra akan direduksi, prinsipnya ukuran matriks akan direduksi menggunakan operasi (Danukosumo, 2017). *Pooling layer* memiliki dua macam yakni *average pooling* dan *max pooling*. Dalam *average pooling*, nilai yang diambil adalah nilai rata-rata, sedangkan dalam *max pooling* nilai yang diambil adalah nilai maksimal. Lapisan terakhir adalah *fully connected*

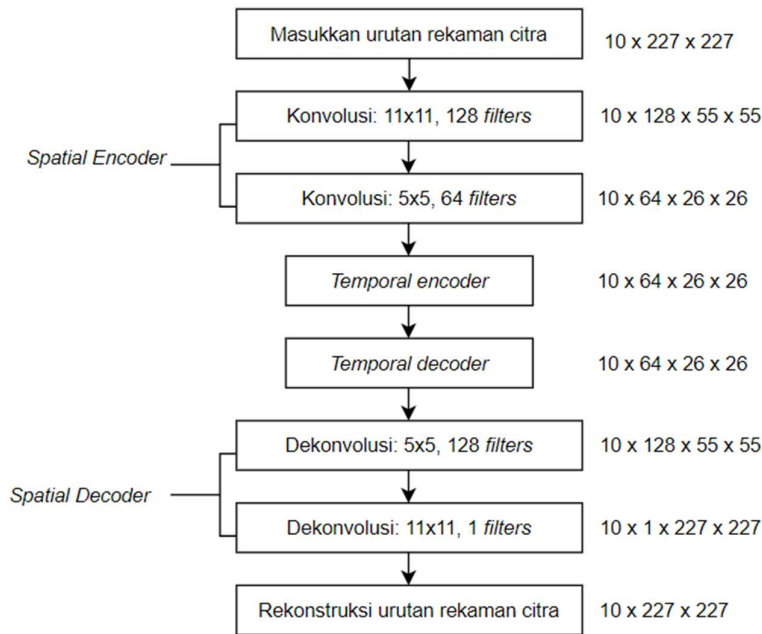
layer yang berfungsi untuk melakukan transformasi pada dimensi data agar data dapat diklasifikasikan secara linier. *Fully connected layer* dilakukan pada tahap terakhir dari CNN, dimana tahap ini data dari citra digital ditransformasikan menjadi data satu dimensi (**Putra, 2016**). Arsitektur VGG16 diadaptasi secara *transfer learning* pada penelitian ini. *Transfer learning* bekerja dengan mengadaptasi *pre-train* model lalu mengganti lapisan klasifikasi yang akan ditentukan. Model VGG-16 atau *Very Deep Convolutional Networks for Large-Scale Image Recognition* adalah jaringan 16-lapisan (konvolusi dan terhubung penuh) yang dibangun di atas basis data ImageNet, yang dibuat untuk tujuan pengenalan dan klasifikasi gambar (**Simonyan & Zisserman, 2015**).

Terakhir adalah modul anomali, modul anomali menggunakan metode *spatiotemporal autoencoder* (**Chong & Tay, 2017**). Terdapat beberapa tahap pada metode ini diantara lain proses awal, pelatihan ciri, dan nilai keteraturan. Proses awal yakni mengkonversi data mentah citra agar dapat diterima sebagai masukan ke model. Setiap citra diekstrak dan diubah ukurannya dari citra rekaman awal menjadi ukuran 227 x 227. Tahap selanjutnya adalah memastikan masukan citra dalam skala yang sama dengan cara skalalisasi antara 0 dan 1 serta dikurangi setiap potongan citra dari rata-rata citra global untuk normalisasi. Rataan citra dihitung dengan melakukan rata-rata nilai piksel setiap lokasi dari potongan citra dalam dataset pelatihan. Setelah itu, citra dikonversi ke abu-abu untuk mengurangi ukuran dimensinya. Citra yang telah diproses lalu dinormalisasi kembali agar memiliki nilai rataan dan varians nol. Masukan untuk model adalah berupa volume rekaman gambar dimana setiap volumenya terdiri dari potongan gambar yang berurutan. Pada tahap ini dibutuhkan data pelatihan yang besar. Pada pelatihan ciri bertujuan untuk mempelajari pola-pola biasa yang terjadi pada rekaman citra yang akan dilatih. Arsitektur pada tahap ini terdiri dari dua garis besar, pertama adalah *autoencoder* spasial untuk melatih spasial struktur dari setiap potongan rekaman citra, selanjutnya adalah *temporal encoder-decoder* untuk melatih pola *temporal* dari struktur spasial yang telah tersandi. Model *temporal encoder-decoder* terdiri dari tiga lapisan konvolusional *Long Short Term Memory* (LSTM) seperti ditunjukkan Gambar 4.



Gambar 4. Arsitektur Model *Temporal Encoder-Decoder*

Kombinasi Deteksi Objek, Pengenalan Wajah dan Perilaku Anomali menggunakan *State Machine* untuk Kamera Pengawas



Gambar 5. Arsitektur *Spatial Encoder-Decoder*

Encoder dan *decoder* spasial memiliki dua lapisan yaitu konvolusional dan dekonvolusional seperti ditunjukkan pada Gambar 5. Saat model berhasil melalui pelatihan, selanjutnya model dapat diketahui performanya dengan menguji menggunakan data tes yang telah disiapkan lalu dicek apakah dapat mendeteksi perilaku abnormal namun disini lain tetap menjaga nilai false alarm rendah. Rekonstruksi kesalahan dari keseluruhan nilai piksel I dalam potongan citra t dari urutan rekaman citra yang diambil sebagai jarak Euclidean antara masukan potongan citra dan potongan citra yang direkonstruksi. Persamaan yang digunakanw ditunjukkan persamaan 1 dimana fw merupakan bobot terlatih dengan model spatiotemporal. Selanjutnya melakukan komputasi skor keabnormalan $s_a(t)$ dengan skala antara 0 dan 1. Skor keabnormalan dapat diturunkan secara sederhana dengan pengurangan satu dengan skor keabnormalan.

$$e(t) = \|x(t) - fw(x(t))\|_2 \quad (1)$$

$$s_a(t) = \frac{e(t) - e(t)_{min}}{e(t)_{max}} \quad (2)$$

$$s_r(t) = 1 - s_a(t) \quad (3)$$

2.3 Alur Kerja Sistem

Alur kerja sistem dipertimbangkan dan didesain berdasarkan kemudahan dan kesederhanaan sistem yang ingin dicapai dan dapat memungkinkan untuk diimplementasikan. Setiap modul-modul yang dibangun diintegrasikan pada sistem dengan menggunakan metode *finite state machine* (FSM) sebagai pemilihan dan pergerakan *state*-nya seperti ditunjukkan Gambar 6. Sistem memiliki empat *state* antara lain *idle*, Anomali, Deteksi Objek, dan Pengenalan Wajah. *State* awal ditentukan adalah *idle*, lalu masukan *state* terdapat Citra, ModulObjek, ModulAnomali, ModulWajah dengan bentuk sinyal *pure*. Keluaran dari *state* memungkinkan tiga keadaan yaitu Objek, Wajah, dan Anomali dengan dua keadaan *present* atau *absent*. Algoritma tiap-tiap modul pada sistem dievaluasi dengan metrik seperti akurasi, presisi, *recall*, dan *f1-score* menggunakan Persamaan (4) hingga (7). Blok diagram umum sistem setiap modulnya ditunjukkan Gambar 7.

$$Akurasi = \frac{TP+}{TP+TN+FP+F} \tag{4}$$

$$Presisi = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

$$F1 - Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \tag{7}$$

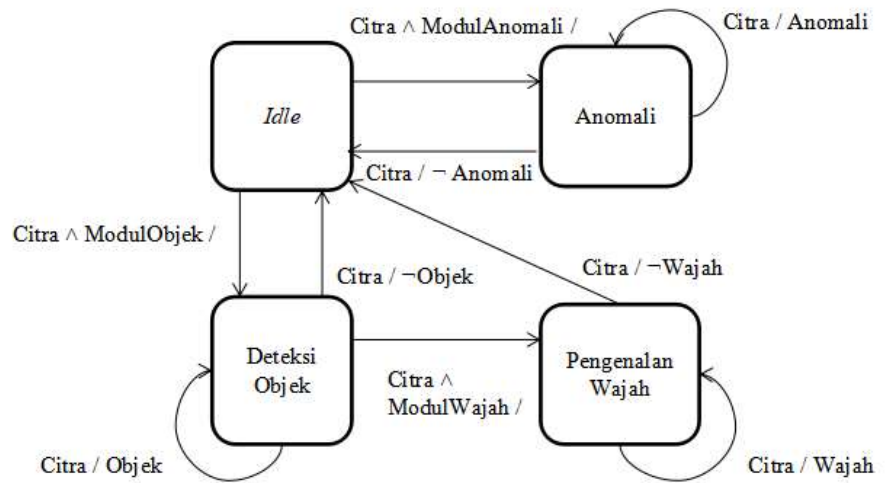
TP = True Positives

TN = True Negatives

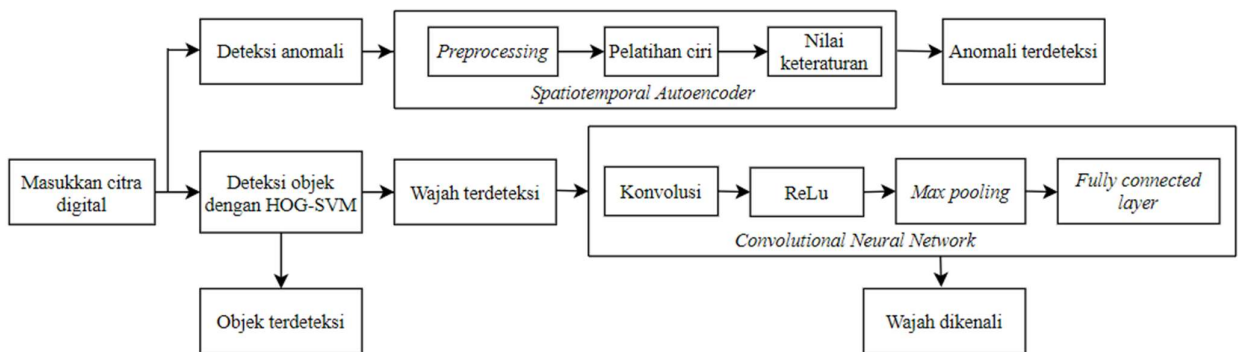
FP = False Positives

FN = False Negatives

States: Idle, Anomali, Deteksi Objek, Pengenalan Wajah
Input: Citra, ModulObjek, ModulAnomali, ModulWajah: pure
Output: Objek, Wajah, Anomali → {present, absent}
InitialStates: Idle



Gambar 6. Finite State Machine Sistem



Gambar 7. Blok Diagram Alur Kerja Modul Sistem

3. HASIL DAN PEMBAHASAN

Sistem yang dibangun serta dirancang pada tahap selanjutnya dilakukan pengujian berdasarkan algoritma arsitektur pembelajaran mesinnya serta implementasi secara riilnya. Pertama-pertama, setiap metode dan modul diuji kemampuan algoritmanya dengan melihat hasil metrik-metrik standarnya antara lain *loss*, akurasi, presisi, *recall*, dan *f1-score*. Metode yang diuji antara lain ketiga modul yang dibangun yakni deteksi objek, pengenalan wajah, dan deteksi anomali. Setelah melalui pengujian tiap modul, selanjutnya dilakukan kombinasi ketiga modul berdasarkan *state machine*. Pada pengujian kali ini dilakukan di dalam ruangan dengan kondisi pencahayaan terang. Skenario implementasi riil yang dilakukan adalah mencocokkan hasil rekaman langsung dengan hasil yang diolah oleh sistem lalu ditampilkan secara *web based* untuk implementasi secara nirkabel. *Dataset* yang digunakan adalah campuran dari pengambilan citra sendiri serta *dataset* yang tersedia daring. *Dataset* pada modul pengenalan wajah terdiri dari sebelas wajah berbeda dengan setiap wajahnya berjumlah 20 buah, lalu *dataset* pada modul deteksi objek terdiri dari 20 buah citra setiap objeknya, dan terakhir pada modul anomali menggunakan *dataset* sekunder dari rekaman *Avenue Dataset*.

Pengujian pertama dilakukan uji performa metrik dari modul pengenalan wajah, *dataset* yang digunakan adalah *data primer* terdiri dari dua bagian peruntukan, yaitu untuk pelatihan dan untuk validasi. Pemilihan banyak sedikitnya *epoch* didasari metode *best guess* dengan melihat performa pelatihan modulnya serta untuk *batch size* ditetapkan bernilai satu yaitu *stochastic gradient descent*. Berikut tabel hasil performa metrik modul pengenalan wajah ditunjukkan oleh Tabel 1. Tabel 1 menunjukkan performa yang cukup apik dari modul yang dibangun. Hal ini berkat memanfaatkan metode *transfer leaning* pada arsitektur pilihan VGG-16. Penulis menerapkan *transfer learning* VGG-16 yang pembobotannya dibuat tetap yang dilatih pada *database imagenet*. Selanjutnya yaitu penggantian *layer* pengklasifikasian *softmax* menggunakan kumpulan citra sebelas wajah berbeda. Proses pelatihan memakan waktu yang cukup lama setiap fasenya dimulai dari satu *epoch* hingga lima *epoch* dengan setiap fasenya memiliki perbaikan performa metrik.

Tabel 1. Hasil Pengujian Modul Pengenalan Wajah

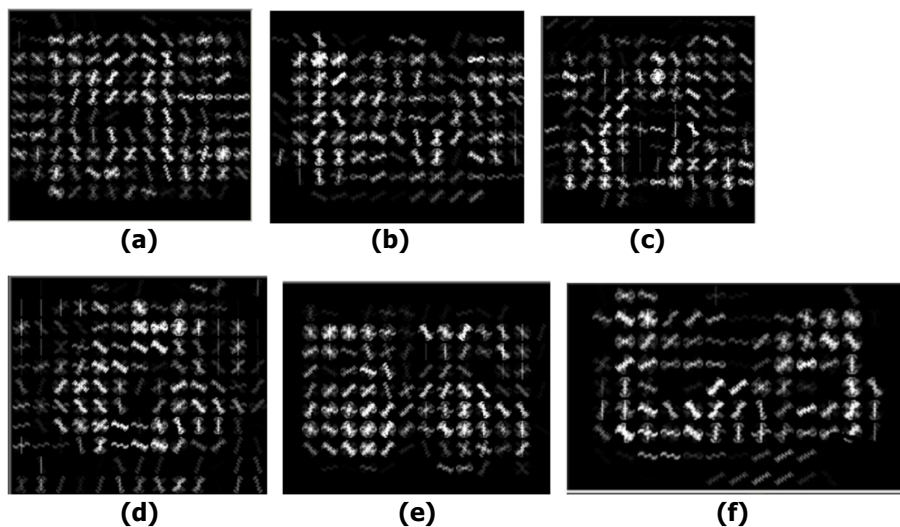
<i>Epoch</i>	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
1	0.9628	0.7981	0.7917	0.7949
2	0.979	0.8846	0.8846	0.8846
3	0.984	0.9132	0.9103	0.9117
4	0.9866	0.9263	0.9263	0.9263
5	0.9888	0.939	0.9375	0.9382

Selanjutnya performa modul kedua yakni modul anomali ditunjukkan Tabel 2. Pada Tabel 2 ditentukan *epoch* yang banyak hingga seribu kali guna mendapatkan performa paling baik. Setelah mencapai seribu, perubahan performa tidak terlalu signifikan. Pada modul anomali, *dataset* yang digunakan adalah *dataset* sekunder dari *Avenue Dataset* yang banyak digunakan untuk pengujian deteksi anomali. Pada modul anomali ini, performa metrik tidak sebaik modul pengenalan wajah. Prinsip kerja pada modul anomali yaitu membandingkan pergerakan potongan citra dari suatu rekaman citra apakah terdapat pergerakan yang perubahannya drastis, sehingga kunci dari modul anomali ini adalah penentuan nilai *threshold* untuk menentukan apakah suatu kejadian merupakan anomali atau tidak.

Tabel 2. Hasil Pengujian Modul Anomali

<i>Epoch</i>	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
1	0.5125	0.0779	0.0832	0.0779
10	0.6974	0.7125	0.6338	0.6703
20	0.7287	0.7917	0.6966	0.7408
40	0.7334	0.8105	0.7055	0.7541
60	0.7344	0.8081	0.7051	0.7527
80	0.7382	0.8003	0.7213	0.7584
100	0.746	0.7946	0.7351	0.7633
125	0.7638	0.8096	0.7778	0.7934
175	0.7734	0.8162	0.8072	0.8114
250	0.777	0.8181	0.8199	0.8189
375	0.7632	0.8134	0.8281	0.8235
500	0.7794	0.8211	0.8312	0.8293
750	0.7811	0.8264	0.8346	0.8305
1000	0.7828	0.8291	0.8363	0.8327

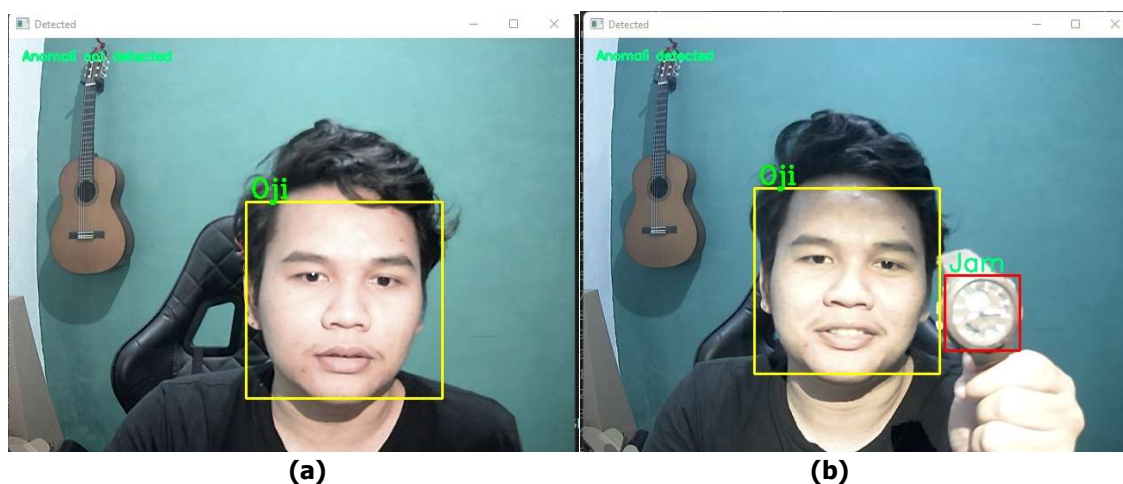
Terakhir, Gambar 8 menunjukkan fitur HOG yang diekstrak serta Tabel 3 menunjukkan hasil pengujian terbaik dari performa modul deteksi objek. Pada modul deteksi objek ini *dataset* yang digunakan adalah *data* sekunder terdiri dari enam kelas berbeda yang sering terlihat di kamera antara lain manusia, mobil, motor, sepeda, jam tangan, dan kucing. Akurasi pelatihan yang didapatkan relatif tinggi untuk mengklasifikasikan objek. *F1-score*, presisi, dan *recall* mencapai 89% keatas untuk setiap kelasnya. Hal ini menunjukkan bahwa sistem dapat menemukan kelas-kelas yang relevan pada dataset tersebut. Hasil Tabel 3 ini diintegrasikan dengan modul deteksi anomali dimana terdapat penentuan besaran *threshold* berdasarkan metode *best guess* untuk menentukan citra yang terekam adalah suatu anomali. Ketika nilai komputasi *spatiotemporal autoencoder* melebihi nilai *threshold* yang ditentukan, maka keadaan citra yang terekam merupakan anomali.



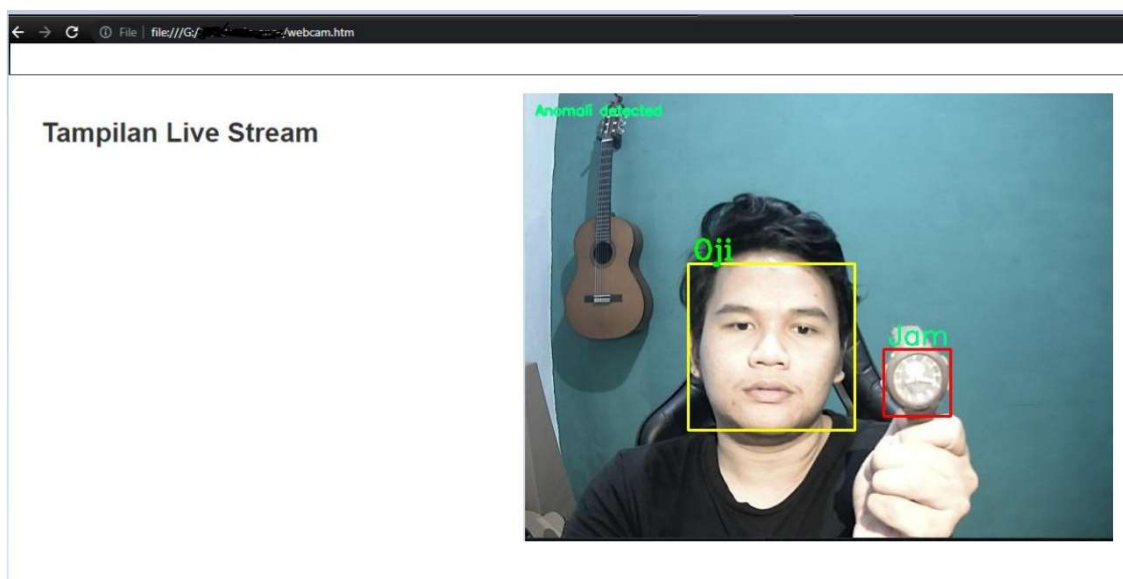
Gambar 8. Fitur HOG (a) Jam Tangan (b) Kucing (c) Manusia (d) Motor (e) Sepeda (f) Mobil

Tabel 3. Hasil Pengujian Modul Deteksi Objek

Kelas	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Manusia	0.9112	0.9223	0.9154	0.9188
Mobil	0.8834	0.9132	0.8944	0.9037
Motor	0.9012	0.9035	0.8985	0.901
Sepeda	0.8933	0.9013	0.9123	0.9068
Jam Tangan	0.9051	0.9111	0.9133	0.9122
Kucing	0.9122	0.9095	0.9185	0.914



Gambar 9. Integrasi Modul (a) Tidak Anomali (b) Anomali



Gambar 10. Siaran Langsung *Web-based*

Setelah dilakukan pengujian setiap modul secara terpisah, tahap selanjutnya adalah integrasi dari ketiga modul tersebut menjadi satu kesatuan sistem. Gambar 9 menunjukkan proses implementasi secara riil. Gambar 10 terdiri dari dua bagian yaitu bagian (a) dan bagian (b). Pada bagian A tidak terdeteksi anomali karena dibawah *threshold* yang ditentukan 0.6

sensitivitasnya. Penentuan sensitivitas *threshold* dapat menggunakan skenario *best guess* berdasarkan lingkungan implementasi riilnya. Pada bagian A pula hanya dicoba dengan wajah yang dikenali serta sistem tidak mengenali pergerakan anomali. Pada bagian B ditambahkan objek jam tangan yang terdeteksi serta modul anomali bekerja memberitahu hal tersebut adalah anomali serta dapat dilihat terdapat kenaikan nilai *loss* yang melebihi *threshold*. Terakhir, fleksibilitas akses oleh *user* dan pengiriman data sangat penting saat kamera mengambil rekaman gambar secara *real time*, maka dari itu rekaman yang terekam dan diproses bisa diakses melalui *web-based* seperti ditunjukkan oleh Gambar 10. Hasil proses ini dapat diakses secara kabel maupun nirkabel sehingga mempermudah aksesibilitas dari sistem yang dibangun.

4. KESIMPULAN

Dalam penelitian ini, penulis membangun sistem untuk menggabungkan dan menambahkan fitur-fitur kombinasi algoritma untuk diintegrasikan menjadi system pengawasan memanfaatkan *state machine*. Sistem terdiri dari tiga modul utama yaitu deteksi objek, pengenalan wajah, dan perilaku anomali. Sistem ini disatukan dengan memanfaatkan *finite state machine* dalam alur kerjanya. Sistem ditunjukkan untuk penggunaan kamera pengawas agar mereduksi dan memperbaiki pengawasan manusia secara konvensional agar terotomasi dan optimal. Hasil yang diperoleh dari modul-modul pada sistem ini berhasil mencapai performa metrik akurasi sekitar 98% untuk pengenalan wajah, 78% untuk deteksi anomali, serta diatas 88% untuk deteksi objek. Selain dari performa algoritma, hal lain yang mempengaruhi performa sistem Ketika implementasi riil adalah kemampuan penentuan perangkat keras yang digunakan. Sistem memerlukan perangkat keras yang berkinerja baik namun fleksibel dalam implementasinya. Penulis menyarankan untuk pengembangan selanjutnya berfokus pada optimasi performa agar ringan saat diimplementasikan secara riil.

DAFTAR RUJUKAN

- Aber, J., Ward Aber, S., Marzloff, I., & Ries, J. (2019). *Small-format aerial photography and UAS imagery* (2nd ed.). Elsevier.
- Andriana, D., Prihatmanto, A. S., Hidayat, E. M. I., & Machbub, C. (2017). Combination of face and posture features for tracking of moving human visual characteristics. *International Journal on Electrical Engineering and Informatics*, 9(3), 616–631.
- Chong, Y. S., & Tay, Y. H. (2017). Abnormal event detection in videos using spatiotemporal autoencoder. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10262 LNCS, 189–196.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3).
- Dalal, N., & Triggs, B. (2005). Histogram of Oriented Gradients for Human Detection. In *Computer Vision and Pattern Recognition* (Vol. 1). IEEE.
- Danukosumo, K. P. (2017). Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Candi Berbasis Gpu. *E-Journal Universitas Atma Jaya Yogyakarta*.

- Edgar, T., & Manz, D. (2017). *Research Methods for Cyber Security* (1st ed.). Elsevier.
- Fang, W., Ding, L., Love, P. E. D., Luo, H., Li, H., Peña-Mora, F., ... Zhou, C. (2020). Computer vision applications in construction safety assurance. *Automation in Construction*, *110*(November 2019), 103013.
- Fennelly, L., & Perry, M. (2016). *Physical Security 5th Ed*. Elsevier Science.
- Kalbo, N., Mirsky, Y., Shabtai, A., & Elovici, Y. (2020). The security of ip-based video surveillance systems. *Sensors (Switzerland)*, *20*(17), 1–27.
- Kinasih, Fabiola Maria Teresa Reetno, Saragih, C. F. D., Machbub, C., Rusmin, P. H., Yulianti, L., & Andriana, D. (2019). State machine implementation for human object tracking using combination of mobilenet, KCF tracker, and HOG features. *International Journal on Electrical Engineering and Informatics*, *11*(4), 697–712.
- Kinasih, Fabiola Maria Teresa Retno, Machbub, C., Yulianti, L., & Rohman, A. S. (2020). Centroid-Tracking-Aided Robust Object Detection for Hospital Objects. *6th International Conference on Interactive Digital Media, ICIDM 2020*, (Icidm).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks Alex. In *Advances in neural information processing systems*.
- Maharani, D. A., Machbub, C., Rusmin, P. H., & Yulianti, L. (2020). Improving the Capability of Real-Time Face Masked Recognition using Cosine Distance. *6th International Conference on Interactive Digital Media, ICIDM 2020*.
- Maharani, D. A., Machbub, C., Yulianti, L., & Rusmin, P. H. (2020). Particle filter based single shot multibox detector for human moving prediction. *2020 IEEE 10th International Conference on System Engineering and Technology*, (pp. 7–11).
- Miao, Z., Zou, S., Li, Y., Zhang, X., Wang, J., & He, M. (2016). Intelligent Video Surveillance System Based on Moving Object Detection and Tracking. *DEStech Transactions on Engineering and Technology Research*, (iect).
- Putra, W. (2016). KLASIFIKASI CITRA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA CALTECH 101. *E-Journal ITS*.
- Puvvadi, U. L. N., Di Benedetto, K., Patil, A., Kang, K. D., & Park, Y. (2015). Cost-effective security support in real-time video surveillance. *IEEE Transactions on Industrial Informatics*, *11*(6), 1457–1465.
- Shidik, G. F., Noersasongko, E., Nugraha, A., Andono, P. N., Jumanto, J., & Kusuma, E. J. (2019). A systematic review of intelligence video surveillance: Trends, techniques, frameworks, and datasets. *IEEE Access*, *7*, 170457–170473.

- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, (pp. 1–14).
- Wiley, V., & Lucas, T. (2018). Computer Vision and Image Processing: A Paper Review. *International Journal of Artificial Intelligence Research*, 2(1), 22.
- Yosafat, S. R., Machbub, C., & Hidayat, E. M. I. (2017). Design and implementation of Pan-Tilt control for face tracking. *2017 7th IEEE International Conference on System Engineering and Technology*, (pp. 217–222).