



ANALISIS DETEKSI PERILAKU ANOMALI MANUSIA  
BERDASARKAN *HUMAN POSE*

PROPOSAL SEMINAR BIDANG KAJIAN

Nurul Adhayanti

99216012

PROGRAM DOKTOR TEKNOLOGI INFORMASI

UNIVERSITAS GUNADARMA

SEPTEMBER, 2022

# Daftar Isi

Daftar Isi . . . . .	i
<b>1 Pendahuluan</b>	<b>2</b>
1.1 Latar Belakang . . . . .	3
1.2 Batasan Masalah . . . . .	5
1.3 Tujuan . . . . .	6
1.4 Kontribusi . . . . .	6
<b>2 Tinjauan Pustaka</b>	<b>10</b>
2.1 Citra Digital.....	10
2.1.1 Digitalisasi Citra .....	11
2.2.2 Citra Video .....	12
2.2 Pembelajaran Mesin.....	15
2.2.1 Deep Learning .....	16
2.2.2 Support Vector Machine .....	17
2.3 Generative Adversarial Networks (GANs).....	19
2.3.1 Deep Convolutional Generative Adversarial Networks (DCGANs).....	20
2.3.2 Fungsi Objektif GANs .....	22
2.3.3 Loss Function GANs.....	22
2.4 Analisis Gerak Manusia.....	23
2.5 Human Detection .....	26
2.5.1 Human Detection Methods For Moving The Human Body .....	26
2.5.2 Machine Learning-Based Human Detection.....	28
2.6 Analisis Struktur Manusia .....	29
2.7 Pengenalan tindakan manusia.....	31
2.7.1 Estimasi Pose .....	32
2.7.2 Tahapan Estimasi Pose .....	33
2.7.3 Anomali Manusia.....	35
<b>3 Metodologi</b>	<b>43</b>
3.1 Pendekatan Penelitian.....	43
3.2 Rencana Kerja .....	45

# Bab I

## Pendahuluan

### 1.1 Latar Belakang

*Human pose* atau Gerak manusia adalah suatu konsep dalam teknologi informasi yang mencakup beberapa komponen informasi objek di dalam gambar dan video: komponen mencakup seperti posisi dan pergerakan seluruh tubuh manusia, bagian tubuh, sampai kepada tindakan manusia. Pada dasarnya, analisis gerakan manusia terdiri dari serangkaian topik: deteksi manusia, pelacakan manusia, *estimasi pose* manusia, dan pengenalan tindakan manusia. Meskipun dalam praktiknya sistem yang berbeda menggunakan strategi pemrosesan yang berbeda, konsep gerak manusia dapat berasal empat topik ini, dimulai dengan analisa pergerakan tubuh manusia dan melakukan proses deteksi, kemudian pelacakan dan analisis struktur tubuh serta pengenalan tindakan manusia.

Deteksi manusia di *Computer Vision* menunjukkan kategori metode yang memperkirakan pergerakan tubuh manusia dalam gambar atau bingkai video. Objek yang akan ditempatkan dapat berupa seluruh atau sebagian (biasanya bagian atas) tubuh manusia. Metode deteksi manusia dapat dibagi menjadi dua kategori: metode berbasis pemrosesan sinyal dengan memiliki cara kerja mengklasifikasikan objek dengan mencocokkan informasi pergerakan tertentu, dan metode berbasis pembelajaran mesin, yang secara statistik mempelajari model dan mengklasifikasikan objek. Metode pelacakan manusia diterapkan untuk memperkirakan pergerakan tubuh manusia.

Pergerakan tubuh manusia dapat dicapai dengan menganalisis perbedaan antar frame pada video/gambar, atau mencari tahu hubungan antara tubuh manusia yang terletak dalam frame yang berurutan. Hal ini menjadi pertimbangan bahwa dalam urutan frame mungkin ada beberapa orang dengan benda diam, *Estimasi pose* atau gerak tubuh manusia mengklasifikasikan tubuh manusia ke dalam kategori pose. Secara umum, ini dapat dianggap sebagai masalah klasifikasi yang mengkategorikan sampel ke dalam beberapa model berbasis pemrosesan sinyal atau pembelajaran mesin.

Salah satu konsep dari *estimasi pose* atau gerak tubuh manusia adalah Melokalisasi posisi sendi tubuh manusia dalam gambar atau *video*. Dengan melokalisasi posisi dapat mengetahui gerak tubuh manusia dalam urutan gambar, sehingga menyimpulkan apa yang dilakukannya pada gerak manusia dengan dilihat dari tindakan atau perilaku. Dalam pengertian ini, tindakan dapat dilihat sebagai aplikasi alami manusia dalam *estimasi pose*.

*Estimasi pose* manusia dapat digunakan untuk pengenalan tindakan. Pengenalan perilaku atau tindakan manusia dilakukan dengan cara mengklasifikasikan gambar atau *video* input dengan tindakan perilaku yang terdeteksi di dalamnya. Untuk video, hubungan informasi tubuh manusia (yaitu bentuk, gerakan, dan lain-lain) dalam *frame* yang berbedadianalisis untuk menentukan kategori perilaku.

Bidang penelitian terkait analisis gerak manusia terutama mencakup pemrosesan gambar, video umumnya menggunakan pembelajaran mesin. Metode pemrosesan video umumnya mengubah frame video menjadi citra masukan menjadi fitur, sinyal, atau informasi yang sesuai. Informasi yang digunakan umumnya informasi gradien dan tepi penting dalam analisis gerakan manusia, dan diekstraksi dan dianalisis dengan metode pemrosesan video. Metode pembelajaran mesin digunakan untuk membangun model sesuai dengan data pelatihan, dan mengklasifikasikan atau memperkirakan data pengujian dengan menerapkan model ini. Algoritma seperti *Support Vector Machine (SVM)*, *Boosting* dan *Random Forest* adalah metode umum dan populer yang diterapkan.

Pembangunan analisa deteksi perilaku anomali manusia pada pendekatan tiga dimensi (3D), merupakan topik utama dalam analisis gerak. Isu penting dari analisis gerak 3D adalah cara memperoleh data 3D. Biasanya data 3D ditangkap oleh beberapa perangkat tertentu. Metode yang menggunakan beberapa kamera untuk mengambil gambar seseorang dengan menggunakan metode multi-warna, untuk merekonstruksi tampilan 3D orang tersebut dan memperkirakan posturnya. Sebuah kelompok riset di Universitas Carnegie Mellon membangun kumpulan data gerak manusia yang ditangkap secara 3D (*CMU multi Modal Activity Database, CMU MMAC*) yang menggunakan kumpulan perangkat penangkap gerak, yang mencakup 12 kamera inframerah, unit pengukuran internal berkabel dan nirkabel, dan perangkat yang dapat dikenakan. Perangkat seperti *BodyMedia* dan *eWatches*. Dataset berisi video tubuh manusia yang melakukan sejumlah tindakan beserta representasi 3D dari setiap bagian tubuh.

Contoh penelitian terbaru ini adalah *Microsoft Kinect* untuk *konsol game Xbox360*, yang memberikan kinerja *real-time* yang handal dibangun dalam perangkat keras yang terjangkau dan ringan. *Kinect* menggunakan sensor 3D yang terdiri dari proyektor laser infra merah bersama dengan sensor *CMOS (Complimentary Metal-Oxide Semiconductor)* untuk menangkap gambar kedalaman sesuai dengan intensitas laser infra merah yang dipantulkan. Gambar biasanya menunjukkan kedalaman tubuh manusia, karena perangkat ini dirancang untuk menangkap gerakan orang yang bermain *game*. Teknik kunci dalam sistem *Kinect* adalah regresi posisi sendi. Setelah mendapatkan kedalaman tubuh manusia, sistem mencoba untuk mendapatkan informasi strukturnya dan menguraikan tubuh menjadi lebih dari sepuluh sendi. Ketiga pendekatan di atas mencontohkan pengembangan 3D adalah cara pengambilan datanya. *Kinect* menggunakan sepasang perangkat inframerah: proyektor dan sensor, untuk mendapatkan gambar yang mendalam. Sementara itu, proyektor infra merah *Kinect* memiliki batasan jarak dan lingkungan, contoh pengawasan, permainan olahraga kelompok, dan lain-lain, meskipun teknologi 3D berkembang pesat, gambar dan video 2D masih mendominasi posisi saat ini.

## 1.2 Batasan Masalah

Penelitian ini merupakan pengembangan bidang analisis gerakan manusia dengan pendekatan 3D dan metode *Generative adversarial network (GAN)*, metode ini untuk meningkatkan kinerja deteksi manusia dan *estimasi pose* manusia, untuk merancang sistem yang membedakan perilaku manusia yang normal dan anomali, Adapun batasan masalah dari penelitian ini adalah sebagai berikut :

1. Gambar dan video diambil dengan kamera 3D RGB atau skala warna.
2. Resolusi gambar data harus lebih besar dari jendela deteksi minimum,  $64 \times 128$  untuk detektor HOG. Hal ini mempertimbangkan jika tubuh manusia jauh lebih kecil dari ukuran jendela minimum, akan terlewatkan dalam fase deteksi.
3. Data gambar dan video yang digunakan tidak terkompresi. Data terkompresi perlu didekompresi untuk diterapkan ke sistem. Oleh karena itu gambar terkompresi diubah menjadi data mentah dan video terkompresi diubah menjadi urutan bingkai. Kompresi *lossless* tidak berdampak pada sistem. Namun, kompresi *lossy* mempengaruhi gradien dan informasi tepi gambar, dan menyebabkan perbedaan hasil.
4. Gambar dan video yang digunakan seluruhnya merupakan sampel digital dengan kuantisasi 28 yang artinya setiap piksel memiliki rentang nilai  $[0, 255]$ .
5. Penelitian berfokus pada pemrosesan video yang berisi seluruh tubuh manusia dan membahas satu situasi tubuh dalam penelitian.
6. Jenis-jenis tindakan dalam suatu set perilaku yang digunakan untuk mengkategorikan tubuh manusia berbeda secara signifikan antara satu sama lain, dalam hal ini menggunakan *Generative adversarial network (GAN)* dalam kategori normal dan anomaly

## 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah mengembangkan analisis dasar/ sederhana manusia, seperti berjalan, berlari, melompat, untuk normal, dan perilaku *anomaly* seperti adanya tindakan yang merusak di tempat umum dengan demensi 3D umumnya menggunakan perangkat kamera pengawas dan subjek di tempat keramaian dan berfokus pada mendeteksi perbedaan antara perilaku normal dan anomaly berdasarkan pose manusia dalam perilakunya dan mempelajari gerakan serta tindakan berupa video.

## 1.4 Kontribusi Penelitian

Penelitian ini membuat kontribusi utama sebagai berikut :

1. Sebuah sistem pencatatan video secara *real time* yang dapat mengklasifikasi tindakan manusia dalam video kerumunan berdasarkan gerak tubuh manusia dalam bentuk 3D.
2. Mengembangkan metode klasifikasi tindakan atau perilaku manusia didalam video kerumunan, menggunakan *Generative adversarial network (GAN)* dalam kategori normal dan anomaly.
3. Membuat dataset gerak tubuh manusia dalam kategori normal dan anomaly khususnya pada masyarakat Indonesia.

## Bab II

# Tinjauan Pustaka

### 2.1 Citra Digital

Gambar atau citra merupakan informasi yang berbentuk visual. Menurut kamus *Webster* citra adalah suatu representasi, kemiripan atau imitasi dari suatu objek atau benda. Misalnya foto manusia mewakili entitas manusia tersebut di depan kamera, foto sinar X mewakili keadaan bagian dalam tubuh seseorang selanjutnya ada suatu file BMP yang mewakili apa yang digambarkan. Citra dari sudut pandang matematis merupakan fungsi menerus (*continue*) dari intensitas cahaya pada bidang dua dimensi, citra yang terlihat merupakan cahaya yang direfleksikan dari sebuah objek, sumber cahaya menerangi objek, objek memantulkan kembali sebagian dari berkas cahaya tersebut dan pemantulan cahaya ditangkap oleh alat-alat *optic*, misalnya mata manusia, kamera, scanner, sensor satelit dan sebagainya kemudian direkam.

Sebagai keluaran dari suatu system perekaman data citra dapat bersifat optik berupa foto, analog, berupa sinyal video seperti gambar pada monitor televisi dan digital yang dapat langsung disimpan pada media penyimpanan magnetik. Citra juga dapat dikelompokkan menjadi citra tampak dan citra tak tampak, contoh citra tampak misalnya foto, gambar, lukisan, apa yang tampak dilayar monitor/televisi, hologram sedangkan citra yang tidak tampak seperti data foto/gambar dalam *file* (citra digital), citra yang direpresentasikan dalam fungsi matematis, citra tak tampak ini harus dirubah menjadi citra tampak, misalnya dengan menampilkannya di layar monitor Citra digital yaitu citra yang disimpan dalam format digital (dalam bentuk file). Hanya citra digital yang dapat diolah dengan komputer, jenis citra lain jika akan diolah dengan komputer harus dirubah dulu menjadi citra digital.



Citra digital dapat didefinisikan sebagai fungsi dua variabel  $f(x,y)$ , dimana  $x$  dan  $y$  adalah koordinat spatial sedangkan  $f(x,y)$  intensitas citra pada koordinat tersebut. Menurut (Darma Putra, 2010) secara umum, pengolahan citra digital menunjuk pada pemrosesan gambar dua dimensi menggunakan komputer. Dalam konteks yang lebih luas, pengolahan citra digital mengacu pada pemrosesan setiap data dua dimensi. Citra digital merupakan sebuah larik (array) yang berisi nilai-nilai real maupun kompleks yang dipresentasikan dengan deretan bit tertentu. Komputer dapat mengolah isyarat-isyarat elektronik digital yang merupakan sinyal biner (bernilai dua: 0 dan 1). Untuk itu citra digital harus mempunyai format tertentu yang sesuai sehingga dapat mempresentasikan objek pencitraan dalam bentuk kombinasi data biner (Nugroho, 2012).

Pada prinsipnya citra pada video adalah sama dengan citra digital biasa. Citra video merupakan sekumpulan citra-citra yang digerakkan sepanjang durasi waktu tertentu. Pergerakan citra-citra yang terdapat pada video tersebut membentuk pergerakan dinamis sehingga mudah mengelabui mata manusia biasa (Fadlisyah dan Rizal, 2011).

### 2.1.1. Digitalisasi Citra

Suatu citra harus direpresentasikan secara numerik dengan nilai-nilai diskrit dengan tujuan agar dapat diolah dengan komputer digital, representasi citra, dari fungsi kontinu menjadi nilai-nilai diskrit disebut digitalisasi, citra yang dihasilkan inilah yang disebut citra digital. Pada umumnya citra digital berbentuk empat persegi panjang, dan dimensi ukurannya dinyatakan tinggi x lebar (atau lebar x panjang). Masing-masing elemen pada citra digital (elemen matrik) disebut *image element*, *picture element* atau *pixel (pixel)*. Citra digital yang berukuran  $N \times M$  umumnya dinyatakan dengan matrik yang berukuran  $N$  baris dan  $M$  kolom, jadi citra berukuran  $N \times M$  mempunyai  $NM$  buah pixel. Indek baris ( $y$ ) dan indeks kolom ( $x$ ) menyatakan suatu koordinat titik pada citra sedangkan  $f(x,y)$  merupakan intensitas (derajat keabuan) pada titik  $(x,y)$ . sebagai contoh

misalkan citra berukuran 256 x 256 dan direpresentasikan secara numeric dengan matrik yang terdiri dari 256 baris (indeks dari 0 sampai 255) dan 256 buah kolom (indeks dari 0 sampai 255).

$$f(x,y) = \begin{pmatrix} f(0,0) & f(0,1) \dots & f(0,M-1) \\ f(1,0) & f(1,1) \dots & f(1,M-1) \\ \vdots & \vdots & \vdots \\ f(N-1,0) & f(N-1,1) \dots & f(N-1,M-1) \end{pmatrix}$$

Gambar 2.1 Matriks Digital NxM  
Sumber (Bambang.dkk 2019)

### 2.1.2 Citra Video

Pada prinsipnya citra pada video adalah sama dengan citra digital biasa, citra video merupakan sekumpulan citra-citra yang digerakkan sepanjang durasi waktu tertentu, pergerakan citra-citra yang terdapat pada video tersebut membentuk pergerakan dinamis sehingga mudah mengelabui mata manusia. Pergerakan atau perubahan yang cepat tersebut yang menjadi tantangan kita untuk mengimplementasikan algoritma yang tepat dalam situasi yang cepat dan dinamis. Ada beberapa pendekatan yang telah dilakukan oleh beberapa peneliti didalam membangun sistem pendeteksi wajah dalam sebuah citra video. Secara umum metode yang diterapkan pada sistem pendeteksi wajah dapat diklasifikasikan menjadi (Fadlisyah dan Rizal, 2011)

1. *Knowledge-based method*. Metode ini kebanyakan digunakan untuk lokalisasi wajah.
2. *Feature invariant approach*. Metode ini kebanyakan digunakan untuk lokalisasi wajah.
3. *Template matching method*. Metode ini digunakan untuk lokalisasi wajah maupun deteksi wajah.
4. *Appearance-based method*. Metode ini kebanyakan digunakan untuk deteksi wajah.

**Knowledge-Based Method**, metode ini menggunakan dasar aturan-aturan yang biasanya digunakan oleh manusia untuk menentukan apa saja yang membentuk suatu wajah. Pada pendekatan ini, metode deteksi wajah dikembangkan berdasar pada aturan (*rule*) yang didapat dari

pengetahuan para peneliti tentang wajah manusia. Sebagai contoh, suatu wajah di dalam citra biasanya memiliki dua buah mata yang simetris, sebuah hidung, dan sebuah mulut. Relasi antara fitur-fitur tersebut dapat direpresentasikan sebagai jarak atau posisi. Pada tahap pertama fitur-fitur wajah diekstraksi lebih dulu, kemudian kandidat wajah ditentukan berdasarkan aturan yang dipakai. Masalah utama pada pendekatan ini adalah kesulitan dalam menterjemahkan pengetahuan manusia ke dalam aturan yang akan dipakai. Jika aturannya terlalu detail (*strict*), maka akan sering gagal mendeteksi wajah yang tidak memenuhi aturan tersebut. Jika aturannya terlalu umum (*general*), akan menghasilkan terlalu banyak *false positive*. Masalahnya akan bertambah sulit jika harus mendeteksi wajah dengan *pose* yang bervariasi karena aturan yang dipakai harus dapat menghadapi semua kemungkinan yang ada. Metode ini biasanya hanya dapat bekerja dengan baik pada wajah berhadapan dan tegak dengan latar belakang sederhana.

***Feature Invariant Approach***, algoritma pada metode ini bertujuan untuk menemukan fitur-fitur struktural dari wajah yang tetap eksis meskipun terdapat variasi *pose*, sudut pandang, dan kondisi cahaya. Pada pendekatan ini, para peneliti mencoba menemukan fitur-fitur yang tidak berubah (*invariant*) pada wajah. Asumsi ini didasarkan pada observasi bahwa manusia dapat dengan mudah mendeteksi wajah dengan berbagai *pose* dan kondisi cahaya, sehingga disimpulkan bahwa pasti ada sifat-sifat atau fitur-fitur yang bersifat *invariant*. Fitur wajah seperti alis, mata, hidung, mulut, biasanya diekstraksi dengan *edge detector*. Selanjutnya dibentuk suatu model statistik yang mendeskripsikan hubungan antara fitur-fitur tersebut untuk menentukan ada tidaknya wajah. Warna kulit manusia juga dapat digunakan untuk membantu memperkirakan area wajah. Namun biasanya deteksi warna kulit ini dikombinasikan dengan metode lainnya seperti *shape analysis dan motion information*.

**Template Matching**, pada metode ini akan disimpan beberapa pola wajah standar untuk mendeskripsikan wajah secara keseluruhan maupun bagian-bagiannya. Pada saat pendeteksian akan dihitung korelasi antara citra input dengan citra pola wajah yang tersimpan sebelumnya.

**Appearance-Based Method**, pada metode ini, model wajah dipelajari melalui proses training dengan menggunakan satu set data pelatihan yang berisi contoh-contoh wajah. Kemudian hasil training ini digunakan untuk mendeteksi wajah. Secara umum metode ini menggunakan teknik-teknik dari analisa statistik dan machine learning untuk menemukan karakteristik-karakteristik yang relevan dari wajah maupun non wajah. Yang termasuk dalam kelompok ini antara lain adalah metode *Eigenfaces* [Kirby, Sirovich, 1990], *distribution-based dan clustering* [Sung, Poggio, 1994], jaringan syaraf tiruan [Rowley, 1998], *support vector machines (SVM)* [Osuna, 1997], *Sparse Network of Winnows (SNoW)* [Yang, 2000], *Naive Bayes Classifier* [Schneiderman, 1998], *Hidden Markov Model (HMM)* [Nefian, 1998], *Kullback relative information* [Colmenarez, 1997], dan *decision trees* [Huang, 1996]. Tantangan yang dihadapi pada masalah deteksi wajah disebabkan oleh adanya faktor-faktor berikut (Nugroho, Setyo, 2004) Posisi wajah. Posisi wajah didalam citra dapat bervariasi karena posisinya bisa tegak, miring, menoleh, atau dilihat dari samping.

1. Komponen-komponen pada wajah yang bisa ada atau tidak ada, misalnya kumis, jenggot dan kaca mata.
2. Ekspresi wajah. Penampilan wajah sangat dipengaruhi oleh ekspresi wajah seorang, misalnya tersenyum, tertawa, sedih, berbicara dan sebagainya.
3. Terhalang objek lain. Citra wajah dapat terhalangi sebagian oleh objek atau wajah lain, misalnya pada citra berisi sekelompok orang.

4. Kondisi pengambilan citra. Citra yang diperoleh sangat dipengaruhi oleh faktor-faktor seperti intensitas cahaya, arah sumber cahaya, dan karakteristik sensor dan kualitas kamera.

## 2.2 Machine Learning

*Machine learning* dilakukan melalui 2 fase, yaitu *fase training* dan *fase application*. *Fase training* adalah proses pemodelan dari algoritma yang digunakan akan dipelajari oleh sistem melalui *training data*, sedangkan *fase application* adalah proses pemodelan yang telah dipelajari sistem melalui *fase training* akan digunakan untuk menghasilkan sebuah keputusan tertentu, dengan menggunakan *testing data*.

*Machine learning* dapat dilakukan dengan dua cara, yaitu *supervised learning* dan *unsupervised learning*. *Unsupervised learning* adalah pemrosesan *sample data* dilakukan tanpa mewajibkan hasil akhir memiliki bentuk yang sesuai dengan bentuk tertentu, dengan menggunakan beberapa *sample data* sekaligus. Penerapan *unsupervised learning* dapat ditemukan pada proses visualisasi, atau eksplorasi data. *Supervised learning* adalah pemrosesan *sample data* x akan diproses sedemikian rupa, sehingga menghasilkan output yang sesuai dengan hasil akhir y. *Supervised learning* dapat diterapkan pada proses klasifikasi (Hairani, 2018)

*Machine learning* merupakan serangkaian teknik yang dapat membantu dalam menangani dan memprediksi data yang sangat besar dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran. Istilah *machine learning* pertama kali didefinisikan oleh Arthur Samuel pada tahun 1959. Menurut Arthur Samuel, *machine learning* adalah suatu bidang ilmu komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas.

Menurut (Mohri et.al, 2012) *machine learning* dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Definisi pengalaman disini ialah informasi sebelumnya yang telah tersedia dan bisa

dijadikan data pembelajar. Dalam pembelajaran *machine learning*, terdapat skenario-skenario seperti :

### **1. *Supervised Learning***

Penggunaan skenario *supervised learning*, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.

### **2. *Unsupervised Learning***

Penggunaan skenario *unsupervised learning*, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokan data berdasarkan karakteristik-karakteristik yang ditemui.

### **3. *Reinforcement learning***

Pada skenario *reinforcement learning fase* pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajar.

## **2.2.1 *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.

Aplikasi konsep jaringan syaraf tiruan yang dalam (banyak lapisan) dapat ditangguhkan 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. *Feature Engineering* juga merupakan teknik yang paling penting untuk mencapai hasil yang baik pada tugas prediksi. Namun, sulit untuk dipelajari dan dikuasai karena kumpulan data dan jenis data yang berbeda memerlukan pendekatan teknik yang berbeda juga.

Dalam *deep learning*, metode *CNN (Convolutional Neural Network)* sangatlah bagus dalam menemukan fitur yang baik pada citra ke lapisan berikutnya untuk membentuk hipotesis nonlinier yang dapat meningkatkan kekompleksitasan sebuah model. Model yang kompleks tentunya akan membutuhkan waktu pelatihan yang lama sehingga di dunia *deep learning* penggunaan GPU sudah sangatlah umum (Danukusumo, 2017).

### **2.2.2 Support Vector Machine**

*Support Vector Machine (SVM)* dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali diperkenalkan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep dasar metode *SVM* sebenarnya merupakan gabungan atau kombinasi dari teori-teori komputasi yang telah ada pada tahun sebelumnya, seperti *marginhyperplane* (Dyda dan Hart, 1973; Cover, 1965; Vapnik, 1964), kernel diperkenalkan oleh Aronszajn tahun 1950, *Lagrange Multiplier* yang ditemukan oleh Joseph Louis Lagrange pada tahun 1766, dan demikian juga dengan konsep-konsep pendukung lain.

Menurut (Fachrurrazi, 2011) *SVM* merupakan suatu teknik untuk melakukan prediksi, baik prediksi dalam kasus regresi maupun klasifikasi. Teknik *SVM* digunakan untuk mendapatkan fungsi pemisah (*hyperplane*) yang optimal untuk memisahkan observasi yang memiliki nilai variabel target yang berbeda (William, 2011). *Hyperplane* ini dapat berupa *line* pada *two dimension* dan dapat berupa *flat plane* pada *multiple dimension*. Menurut (Nugroho, 2003), karakteristik *SVM* secara umum dirangkum sebagai berikut:

1. Secara prinsip *SVM* adalah *linear classifier*
2. *Pattern recognition* dilakukan dengan mentransformasikan data pada ruang input (*input space*) ke ruang yang berdimensi lebih tinggi (*feature space*), dan optimisasi dilakukan pada ruang vector yang baru tersebut. Hal ini membedakan *SVM* dari solusi *pattern recognition* pada umumnya, yang melakukan optimisasi parameter pada hasil transformasi yang berdimensi lebih rendah daripada dimensi input space.
3. Menerapkan strategi *Structural Risk Minimization (SRM)*.
4. Prinsip kerja *SVM* pada dasarnya hanya mampu menangani klasifikasi dua kelas, namun telah dikembangkan untuk klasifikasi lebih dari dua kelas dengan adanya *pattern recognition*.

**Metode Support Vector Machine** memiliki beberapa keuntungan yaitu:

### 1. *Generalisasi*

Generalisasi didefinisikan sebagai kemampuan suatu metode untuk mengklasifikasi suatu pattern atau pola, yang tidak termasuk data yang digunakan dalam fase pembelajaran metode itu.

### 2. *Curse of dimensionality*

Curse of dimensionality didefinisikan sebagai masalah yang dihadapi suatu metode pattern recognition dalam mengestimasi parameter dikarenakan jumlah sampel data yang relatif lebih sedikit dibandingkan dengan dimensional ruang vektor tersebut.

### 3. *Feasibility*

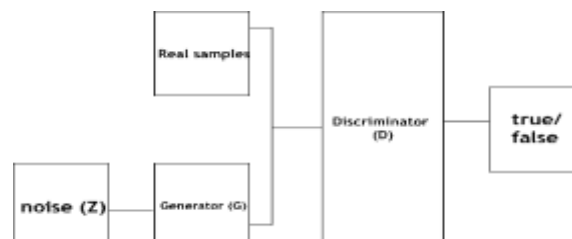


*SVM* dapat diimplementasikan relatif lebih mudah, karena proses penentuan *support vector* dapat dirumuskan dalam *Quadratic Programing (QP) problem* (Nugroho, 2003). Adapun kerugian dari metode *Support Vector Machine* adalah sebagai berikut:

1. Sulit dipakai pada problem berskala besar. Dalam hal ini dimaksudkan dengan jumlah sampel yang diolah
2. *SVM* secara teoritik dikembangkan untuk problem klasifikasi dengan dua kelas. Namun dewasa ini *SVM* telah dimodifikasi agar dapat menyelesaikan masalah dengan lebih dari dua kelas (Nugroho, 2003).

### 2.3 *Generative Adversarial Networks (GANs)*

*Generative Adversarial Networks (GAN's)* diperkenalkan pertama kali oleh Ian J. Goodfellow bersama tujuh orang lainnya pada tahun 2014 melalui Jurnal mereka yang berjudul *Generative Adversarial Nets*. Penggunaan *Generative Adversarial Networks (GAN's)* sangat berbeda dengan *Algoritma Deep Learning* lainnya, dimana Algoritma lainnya memanfaatkan penuh terhadap dataset yang ada dimana dataset tersebut dilatih untuk memaksimalkan hasil akurasi dari setiap pengujian yang dilakukan, semakin banyak dataset yang dimiliki maka semakin baik hasil yang didapatkan. *Generative Adversarial Networks (GAN's)* memiliki pendekatan yang berbeda dengan dengan algoritma lainnya, algoritma ini menggunakan dua jaringan syaraf buatan (*neural network*) untuk menyelesaikan permasalahan yang ada. Kedua jaringan syaraf buatan tersebut adalah Generator yang berfungsi untuk mengambil sampel data dari dataset, dan Diskriminator yang berfungsi untuk mengklasifikasikan bahwa sampel data itu bernilai asli atau palsu. Dengan adanya kedua neural network tersebut, maka akan dihasilkan suatu data yang baru yang sangat menyerupai data inputan-nya. (Eric dan Hendry, 2018)



Gambar 2.2 Struktur *Generative Adversarial Networks (GANs)*

Sumber : (Eric dan Hendry, 2018)

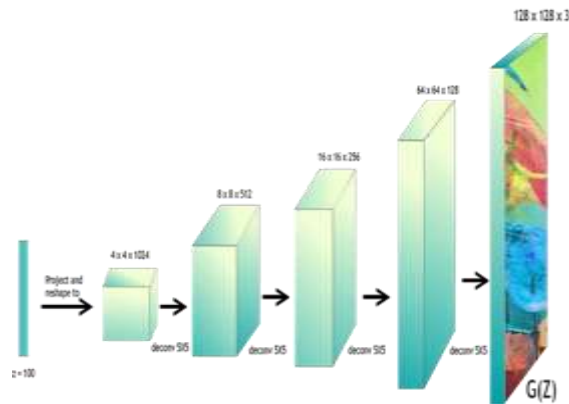
Struktur *Generative Adversarial Networks (GANs)* yang terdapat pada Gambar 2.2 menunjukkan bahwa fungsi dipisahkan menjadi Generator dan Diskriminator . Untuk mempermudah penjelasan disini Generator diinisialisasikan dengan fungsi G dan Diskriminator dengan fungsi D, dengan X sebagai data asli yang ada di dalam Dataset, terakhir terdapat Z sebagai variabel acak. Setiap hasil dari G akan diwakili  $G(z)$  sebagai sampel baru hasil dari Generator. Data X akan dimasukkan ke dalam Diskriminator D untuk memberikan label terhadap  $G(z)$ . jika Diskriminator D menghasilkan label 1(True) berarti  $G(z)$  dianggap mirip dengan X. Jika Discriminator D menghasilkan label 0(False) berarti  $G(z)$  dianggap tidak mirip dengan X. untuk hasil dari Generator G sangat baik, Diskriminator D harus mengklasifikasikan  $G(z)$  dengan label 0. tujuan G adalah untuk membuat kinerja data yang dihasilkan  $G(z)$  pada D ( $D(G(z))$ ) konsisten dengan kinerja Real Samples X pada D. (Eric dan Hendry, 2018)

### 2.3.1 Deep Convolutional Generative Adversarial Networks (DCGANs)

Untuk meningkatkan Performa dari model yang akan dibangun, disini ditambahkan *Deep Convolutional Generative Adversarial Networks (DCGANs)* untuk menggunakan *Convolutinal Layers* pada Diskriminator Nya dan *transpose convolution layers* pada Generator Nya. Sebelum membahas lebih lanjut, Diperlukan pengertian mengenai apa itu *Deep Convolutional Generative Adversarial Networks (DCGANs)*. *DCGAN* pertama kali diperkenalkan pada tahun 2016 melalui jurnal yang berjudul “*Unsupervised Representation Learning with Deep Convolutional Networks*” karya Alec Radford dan Luke Metz. Dalam jurnal tersebut menjelaskan bahwa DCGAN tercipta melalui proses perluasan *Generative Adversarial Networks (GANs)* dengan menambahkan *Convolution Neural Network (CNN)* kedalamnya . Untuk mengetahui perbedaan dari *GANs* dan *DCGANs* dapat dilihat dari arsitektur terbentuknya *DCGAN* yaitu :

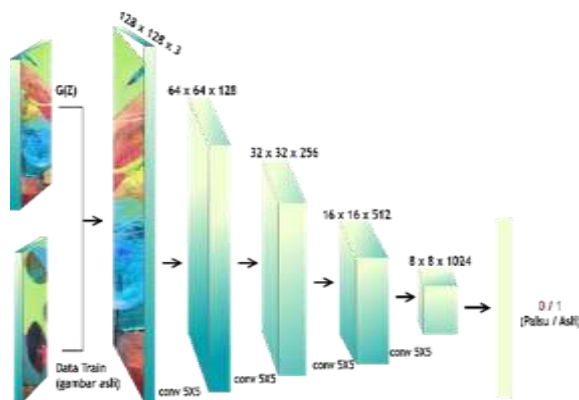
1. Mengganti *Pooling Layer* pada *Diskriminator* dengan *strided convolutions* dan di *fractional-strided convolutions* pada Generator nya.
2. Pada *Diskriminator* dan *Generator* menggunakan *Batch Normalization* yang berfungsi untuk meningkatkan kecepatan jaringan syaraf tiruan dan supaya lebih stabil.
3. *Fully connected hidden layers* dihapus untuk arsitektur yang lebih mendalam.
4. *Generator* menggunakan aktivasi ReLU untuk semua lapisan, tetapi terdapat pengecualian pada output yaitu menggunakan TanH.

5. Diskriminator menggunakan aktivasi LeakyReLU untuk semua lapisannya.



Gambar 2.3 Struktur Generator DCGANs  
Sumber : (Eric dan Hendry, 2018)

Pada Gambar 2.3 terdapat struktur dari Generator *DCGANs*[16]. Misalnya kita punya 100 random noise vektor sebagai input nya, kemudian kita proyeksikan dan bentuk kembali random *noise* vektor tersebut ke dalam bentuk  $4 \times 4 \times 1024$ . Pada Gambar tersebut, diinginkan untuk mengubah gambar  $4 \times 4 \times 1024$  menjadi bentuk  $128 \times 128 \times 3$ . Lakukanlah proses sebelumnya sampai memperoleh hasil yang diinginkan, kemudian setiap prosesnya harus mengikuti arsitektur dari pada *DCGANs*, yaitu untuk setiap Layer menggunakan *Batch Normalization* dan menggunakan aktivasi ReLU kecuali pada  $G(Z)$  menggunakan TanH. Seperti itulah cara untuk Upsample gambar yang dimiliki.



Gambar 2.4 Struktur Diskriminator DCGANs  
Sumber : (Eric dan Hendry, 2018)

Diskriminator berfungsi untuk mengetahui sebuah gambar bernilai asli atau palsu. Dengan gambar yang dihasilkan dari Generator, diskriminator akan bekerja untuk mengidentifikasi gambar tersebut dengan cara mencoba mengembalikan gambar tersebut ke bentuk awal. Seperti terlihat pada Gambar 2.4

bahwa Diskriminator merupakan kebalikan dari Generator, ditambah dengan gambar asli yang dimiliki di dalam Dataset, maka akan meningkatkan kemampuan Diskriminator untuk mengidentifikasi suatu gambar bernilai asli atau palsu. Untuk Diskriminator juga menggunakan kaidah arsitektur dari *DCGANs* yaitu menggunakan Batch Normalization dan menggunakan aktivasi LeakyReLU untuk semua layer. (Eric dan Hendry, 2018)

### 2.3.2 Fungsi Objektif GANs

Setiap model Deep Learning sebuah fungsi harus didefinisikan untuk memperoleh suatu hasil. Pada *Generative Adversarial Networks* memiliki suatu Fungsi Objektif yang digunakan bersamaan dengan model yang dibangun. Disini akan dibangun fungsi objektif dari struktur GANs.

$$\min_D \max_G \mathbb{E}_{x \sim p_x} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))] \quad (1)$$

Fungsi objektif (1) terdiri dari  $x$  sebagai real data (Sampel data) dan  $z$  adalah *noise* vektor,  $G$  adalah Generator dan  $D$  adalah Diskriminator,  $p_x$  adalah distribusi dari  $x$  (real data) dan  $p_z$  adalah masing-masing dari noise vektor.  $D(x)$  merupakan hasil dari training Diskriminator,  $G(z)$  adalah hasil dari Generator dan  $D(G(z))$  adalah Hasil dari Discriminator yang memberikan nilai/hasil dari  $G(z)$ . Diskriminator dilatih untuk memaksimalkan pemberian nilai terhadap hasil dari generator dengan memaksimalkan fungsi  $\log D(x) + \log(1 - D(G(z)))$ . Kemudian Generator dilatih untuk dapat menipu Diskriminator dengan meminimalkan fungsi  $\log(1 - D(G(z)))$ . Melalui pelatihan inilah Generator akan berusaha semaksimal mungkin untuk menghasilkan gambar tidak semirip mungkin dengan dataset(real data), dan Diskriminator juga akan berusaha untuk mengidentifikasi hasil dari Generator untuk mengetahui gambar tersebut asli atau palsu. (Eric dan Hendry, 2018)

### 2.3.3 Loss Function GANs

Pada dasarnya *loss function* digunakan untuk mengetahui performa suatu model yang sedang dikembangkan. Pada riset ini *loss function* digunakan untuk mengembangkan serta mengukur Generator dan

Diskriminator *DCGANs* untuk mengetahui sejauh apa performa dari kedua model ini. Seperti yang sudah dibahas sebelumnya bahwa Diskriminator bertujuan untuk memaksimalkan fungsi  $\log(1 - D(G(z)))$  sehingga Diskriminator akan meminimalkan fungsi  $D(G(z))$ . Sehingga loss function dari Diskriminator adalah:

$$L^D = \text{Max} [ \log D(x) + \text{Log}(1 - D(G(z))) ] \quad (2)$$

Dengan fungsi tersebutlah (2) Diskriminator akan memaksimalkan fungsi  $D(x)$  selagi meminimalkan fungsi  $D(G(z))$ . Untuk itu Generator harus berusaha semaksimal mungkin untuk menipu Diskriminator untuk hasil gambar yang dihasilkan Generator, dengan cara meminimalkan fungsi  $\log(1 - D(G(z)))$ . Sehingga loss function dari Generator (3) adalah:

$$L^G = \text{Min} [ \text{Log}(D(G(z))) ] \quad (3)$$

Generator tidak memiliki hak atas dataset (*real data*) sehingga generator tidak memiliki wewenang/kuasa atas  $D(x)$  sehingga Generator tidak dapat mengubahnya. Tetapi Generator dapat memaksimalkan fungsi  $G(z)$  untuk menghasilkan gambar yang dapat menipu Diskriminator. (Eric dan Hendry, 2018)

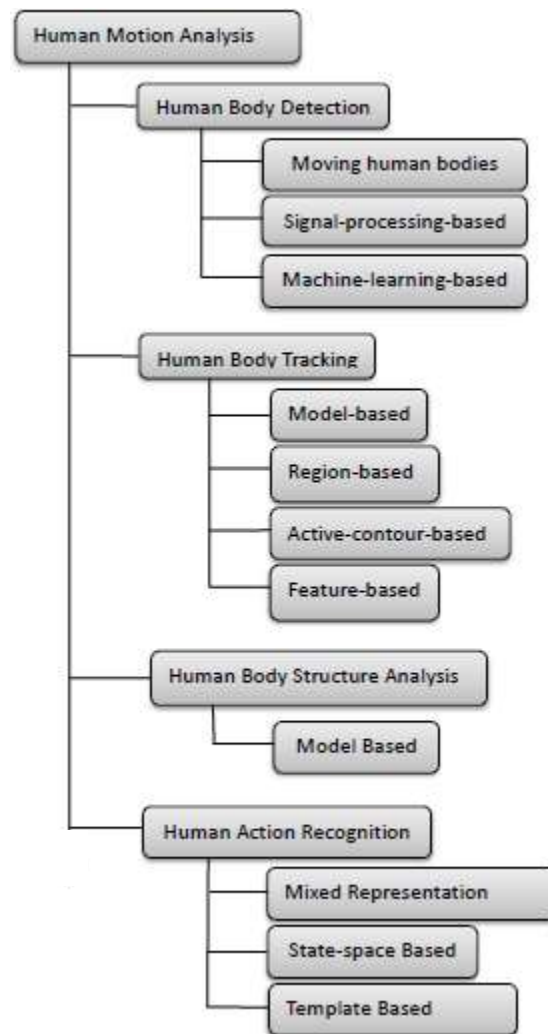
## 2.4 Analisis Gerak Manusia

Analisis gerak manusia adalah cabang dari analisis gerak yang besar dan penting dalam visi komputer dan pemrosesan gambar maupun video. Analisis Ini berkonsentrasi pada proses mengenali atau menggambarkan gerakan tubuh manusia dalam gambar atau video. Penelitian paling awal di bidang ini berawal dari tahun 1970-an. Johansson menggambarkan metode yang disebut *moving light display (MLD)* pada tahun 1975 untuk menganalisis pergerakan sendi pada tubuh manusia. Dalam penelitian mengembangkan suatu metode yang secara sederhana merepresentasikan tubuh manusia dengan bentuk tongkat yang merupakan garis-garis yang dihubungkan oleh persendian, yang gerakannya dilacak dan diperkirakan untuk mengenali aksi

tubuh manusia.

Analisis gerakan manusia dalam visi komputer umumnya memiliki empat topik: deteksi manusia, pelacakan manusia, pengenalan tindakan manusia (atau pengenalan aktivitas manusia) dan analisis struktur tubuh. Survei hanya menyebutkan 3 di antaranya, tanpa memasukkan deteksi manusia. Namun di sebagian besar proyek penelitian, deteksi adalah langkah dasar untuk proses lainnya. Penelitian yang memperkenalkan sistem analisis gerak keseluruhan biasanya menempatkan fase deteksi sebagai fase awal. Penelitian lain yang tidak membahas deteksi gerakan manusia juga mengambil hasil deteksi sebagai masukannya. Deteksi manusia diperlakukan sebagai bagian penting dari analisis gerakan manusia dan diperkenalkan secara rinci. Beberapa survei lain tidak memasukkan analisis struktur tubuh, yang saat ini populer dan diterapkan dalam banyak aplikasi praktis.

Deteksi manusia mendeteksi seluruh atau sebagian tubuh manusia dalam gambar masukan. Pelacakan manusia mempertimbangkan apakah dalam urutan bingkai tubuh manusia yang terdeteksi adalah untuk orang yang sama atau untuk orang yang berbeda. Dengan kata lain itu memperkirakan jejak gerak orang-orang dalam urutan *input frame*. Pengenalan gerakan manusia berfokus pada pengenalan perilaku orang yang terdeteksi. Analisis struktur tubuh mempertimbangkan lokasi, gerakan, dan hubungan bagian-bagian tubuh manusia. Setiap topik dalam analisis gerak manusia juga dapat dibagi menjadi beberapa *subtopic*, seperti pada Gambar 2.5 Menunjukkan keseluruhan area penelitian pada *human motion analysis*.



Gambar 2.5 Area penelitian pada *human motion analysis*.

(Sumber : Ziran Wu, 2012)

Secara umum, dua jenis data digunakan dalam analisis gerakan manusia: gambar diam dan urutan frame (atau disebut video). Dalam penelitian sebelumnya, kami menemukan bahwa sebagian besar metode pendeteksian manusia tidak mencapai kinerja yang sangat baik, dengan kata lain, metode tersebut tidak mencapai tingkat deteksi yang sangat tinggi dibandingkan dengan tingkat kesalahan yang sangat rendah. Jadi pada penelitian ini dilakukan percobaan di bagian pra-pemrosesan, untuk meningkatkan kinerja deteksi.

## 2.5 Human Detection

*Human Detection* atau Deteksi Manusia pada dasarnya mengklasifikasikan area yang mendeteksi tubuh manusia dan yang tidak terdeteksi. Menurut persyaratan sistem, target deteksi dapat berupa seluruh tubuh atau sebagian tubuh. Misalnya, gambar pengawasan yang memiliki beberapa jenis objek, seperti pejalan kaki, bangunan, kendaraan, hewan, latar belakang kosong, dan sebagainya. Tujuan dari deteksi manusia adalah menemukan tubuh manusia dan menolak objek lain. Tanpa membedakan secara akurat tubuh manusia dari objek lain, fase selanjutnya seperti pelacakan dan pengenalan tidak dapat dijalankan dengan benar, jika fase deteksi memberikan masukan ke fase selanjutnya.

Meskipun deteksi manusia adalah cabang dari klasifikasi objek, tiga faktor membuatnya lebih rumit dibandingkan dengan beberapa jenis klasifikasi objek lainnya: gaya tubuh, kondisi pakaian, dan *pose*. Penampilan objek target (tubuh manusia) sangat bervariasi karena ketiga faktor ini. Sangat mudah untuk membayangkan betapa banyak perbedaan antara seorang anak laki-laki gemuk, kecil, duduk mengenakan jaket dan seorang wanita tinggi kurus, menari mengenakan gaun.

Ada dua kategori pendekatan untuk mendeteksi tubuh manusia: pendekatan tubuh manusia bergerak, dan pendekatan berbasis pembelajaran mesin. Ada juga beberapa jenis target deteksi, seperti seluruh tubuh, sebagian tubuh, manusia yang mengendarai sepeda, atau bahkan kelompok manusia.

### 2.5.1 Human Detection Methods For Moving The Human Body

Metode semacam ini hanya dapat diterapkan pada masukann *frame video*, perbedaan antara *frame* yang disebabkan oleh gerakan digunakan sebagai informasi tentang objek yang bergerak. Gumpalan objek bergerak adalah fitur khas yang mewakili gerak. (Lipton dkk.) menggunakan dispersi dan wilayah gumpalan sebagai masukan untuk mengklasifikasikan objek bergerak ke dalam tubuh manusia, kendaraan dan kekacauan, dan mengoptimalkan hasilnya dengan batasan konsistensi temporal. (Collin dkk.) menggunakan informasi campuran waktu spasial, seperti



dispersi gumpalan, area gumpalan, rasio aspek nyata dari kotak pembatas gumpalan, dan *zoom* kamera, untuk mengklasifikasikan gumpalan di setiap bingkai untuk menemukan wilayah target yang mungkin dengan sudut pandang spesifik jaringan saraf. Kemudian mereka menyimpan hasilnya dalam histogram yang mencerminkan probabilitas dari kelas mana objek tersebut berada, dan memilih label kelas yang paling mungkin sebagai output. Dalam penelitian tersebut mengklasifikasikan objek menjadi manusia, kendaraan, kelompok manusia, dan kekacauan.

(Kuno dkk.) menggunakan siluet gambar masukan untuk membangun sistem deteksi manusia. Mereka mengekstrak daerah variasi kecerahan, daerah terpilih yang disebabkan oleh objek bergerak, menggabungkan daerah menjadi pola siluet menurut jarak. Hasil siluet sebagai kombinasi dari beberapa segmentasi tubuh sehingga fitur pola siluet dapat diekstraksi. Akhirnya mereka mengklasifikasikan siluet dengan membuat penilaian pola siluet dengan menggunakan parameter bentuk seperti mean dan standar deviasi histogram proyeksi siluet dan rasio aspek persegi panjang yang membatasi daerah yang bergerak.

Penelitian lain seperti Cutler dan Davis mempertimbangkan kesamaan gerak periodik manusia, dan menyajikan pendekatan yang mendeteksi gerak manusia periodik dengan memperkirakan kesamaan dengan prototipe gerak. Teknik ini mengimplementasikan deteksi, pelacakan, dan pengenalan saat melakukan analisis frekuensi waktu pada gerak periodik. Pendekatan serupa yang disajikan oleh Lipton menggunakan aliran optik. Dengan asumsi bahwa benda bergerak tidak kaku seperti tubuh manusia lebih sering menghasilkan aliran sisa daripada benda kaku, algoritme dapat membedakan gerakan manusia dari benda bergerak lainnya

seperti kendaraan. Pendekatan ini didasarkan pada asumsi bahwa gerak manusia dapat dianggap sebagai periodik, dan membangun model gerak agar sesuai dengan target.

Kelemahan dari jenis pendekatan ini jelas. Itu hanya dapat mendeteksi objek bergerak, sehingga pendekatan ini tidak dapat diterapkan pada gambar diam. Bahkan untuk video, jika targetnya statis, atau sebagian dari targetnya statis, pendekatan ini tidak akan berhasil. Oleh karena itu, aplikasi dari pendekatan ini terbatas.

### ***2.5.2 Machine Learning-Based Human Detection***

Deteksi manusia berbasis pembelajaran mesin sebagai metode yang mengklasifikasikan sampel objek yang tidak diketahui berdasarkan pengalaman sampel yang diketahui. Metode ini secara statistik merangkum informasi dari semua gambar pelatihan dan menggunakan metode pembelajaran mesin dan pengenalan pola. Tidak seperti metode berbasis pemrosesan sinyal, metode ini tidak memperhatikan sinyal di lokasi tertentu, tetapi memperlakukan informasi dalam deteksi secara keseluruhan, membangun model statistik yang menggunakan fitur yang diekstraksi dari data pelatihan, dan menggunakan model untuk membuat pengklasifikasi (detektor). Fitur adalah metode visi komputer yang dirancang untuk mewakili informasi penting dari gambar. *Haar like features* dan *histogram gradien* berorientasi adalah metode yang mencerminkan gradien dan informasi tepi, Komponen utama mewakili fitur statistik dari keseluruhan gambar. Metode klasifikasi dapat berupa *boosting/AdaBoost classifier*, *Bayes classifier*, *Support vector machine (SVM)*, jaringan saraf, lingkungan terdekat, dan lain-lain.

Konsep *Haar like features* dikembangkan dari teori *wavelet Haar*. Ini mengekstrak lokasi tertentu di jendela deteksi dan mewakili mereka dengan daerah yang berdekatan. Selisih yang digunakan sebagai fitur dalam klasifikasi, dihitung antara setiap pasangan daerah yang berdekatan dengan membandingkan intensitas pikselnya. Perbedaan ini kemudian digunakan untuk

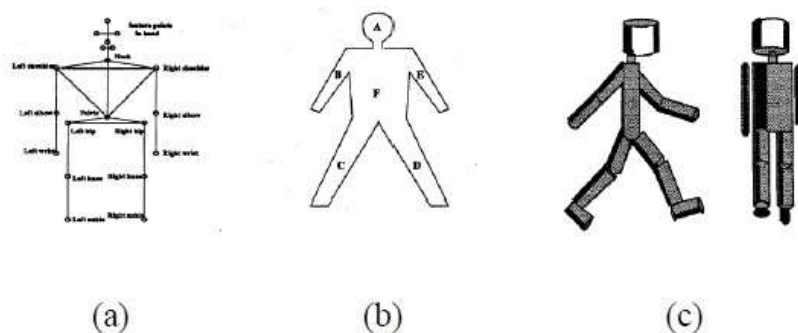
mengkategorikan subbagian dari suatu gambar. *Boosting/AdaBoost classifier* adalah jenis mekanisme yang menggabungkan *classifier* yang lemah menjadi *classifier* yang kuat. Pengklasifikasi ini dibangun dengan metode cascade. Ide dasarnya adalah: setiap pengklasifikasi memiliki tugas untuk mengklasifikasikan bagian dari fitur dan meneruskan hasil klasifikasi ke tingkat yang lebih tinggi. Mekanisme ini secara signifikan dapat mempercepat efisiensi klasifikasi. Jadi *boosting/AdaBoost* sering diterapkan untuk membangun sistem waktu nyata. Penelitian yang diterbitkan oleh Viola dan Jones menggabungkan garis dasar fitur seperti *Haar* dengan pendekatan *AdaBoost* (Kami menyebut kategori metode ini *Haar AdaBoost* untuk jangka pendek dalam tesis ini). Banyak peneliti mengembangkan pendekatan mereka berdasarkan ini. Fitur khas seperti *Haar* dan pendekatan *boosting* dijelaskan oleh Kruppa dan Schiele berdasarkan detektor wajah Lienhart dan ini adalah pendekatan standar di *OpenCV*. Penelitian menggambarkan fitur seperti *Haar* dan pendekatan peningkatan untuk deteksi wajah.

*Histogram of Orientated Gradients (HOG)* adalah topik populer di bidang deteksi manusia dalam beberapa tahun terakhir. Metode ini mengumpulkan dan menyimpan histogram dari gradien dan orientasi sebagai fitur. Pendekatan asli *HOG*, yang disajikan oleh Dalal dan Triggs menggabungkan *HOG* dengan *SVM*, sebagai pengklasifikasi. Baru-baru ini semakin banyak peneliti menggabungkan *HOG* dengan *classifier AdaBoost*, yang membuat komputasi jauh lebih efisien. Dalam penelitian *HOG* dari wilayah gambar kecil diekstraksi untuk membangun pengklasifikasi dengan *SVM*, dan pengklasifikasi kuat, yang mengeksekusi kategorisasi untuk wilayah besar, disusun oleh pengklasifikasi lemah. Sistem yang diusulkan dalam penelitian ini menerapkan metode *HOG* sebagai detektor manusia.

## 2.6 Analisis Struktur Manusia

Analisis struktur manusia berfungsi untuk menemukan lokasi, orientasi dan ukuran bagian tubuh manusia. Hal ini erat kaitannya dengan pendugaan postur karena dengan adanya informasi bagian-

bagian tubuh maka tidak akan sulit untuk mengidentifikasi pose-pose tersebut. Sebenarnya ada beberapa makalah yang membahas metode yang memperkirakan pose dari informasi bagian tubuh. Pembahasan penelitian ini dimasukkan ke dalam Bagian Untuk merepresentasikan bagian-bagian tubuh manusia, terutama ada tiga cara: kerangka model, model kontur dengan tandapada setiap bagian dan model volumetrik. Model kerangka atau *stickman* adalah representasi yang paling umum dan saat ini sebagian besar pendekatan untuk gambar dan video 2D didasarkan pada itu. Keuntungan dari model *stickman* adalah bahwa mereka dibangun oleh sendi yang memungkinkan bagian tubuh untuk bergerak dan berputar. Ini fleksibel dan sesuai dengan sifat gerak manusia. Itu model kontur biasanya tetap garis besar tubuh manusia. Jadi mereka hanya bisa memperkirakan perkiraan informasi untuk bagian tubuh. Apalagi bila jumlah bagian tubuh yang besar, menggunakan model kontur bukanlah solusi yang tepat. Ketika model *volumetrik* lebih mementingkan ukuran dan kedalaman bagian tubuh,

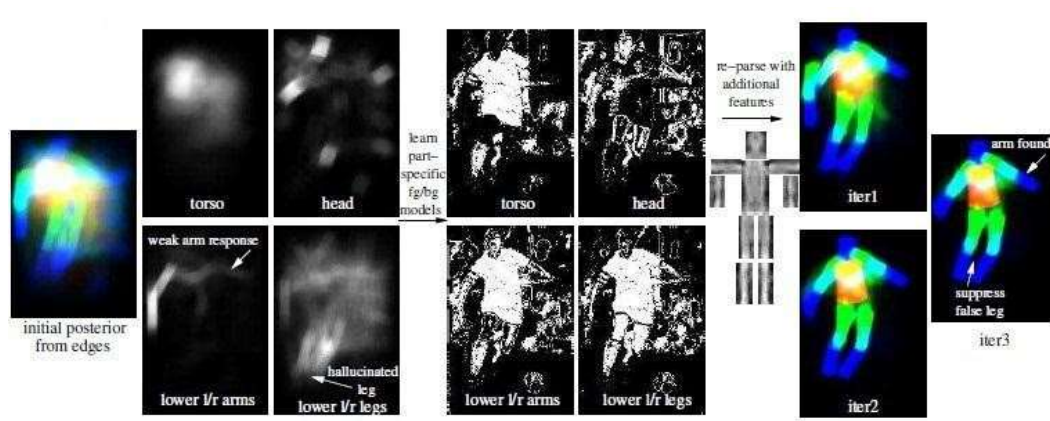


Gambar 2.6 Model struktur manusia  
Sumber : (San Juan, 1997)

(a) *model stickman 2D*, (b) *model kontur*, (c) *model volume 3D*.

Gambar 2.6 menunjukkan prosedur dalam penelitian Ramanan. Penelitian berawal, dari Gambar 2.6 (a), penelitian berulang dimulai dengan memperkirakan pelokalan bagian tubuh dengan mencocokkan dengan model pose yang dapat diubah bentuk tepinya. Hasil pencocokan merupakan *prosedur parse* awal, yang hanya memberikan perkiraan bentuk tubuh bagua kasar. Kemudian *inisial parse* diterapkan untuk membangun model histogram warna latar depan/latar belakang

untuk semua bagian tubuh yang merupakan daerah pada gambar pada tahap ini, seperti yang ditunjukkan Gambar 2.6(b). Di Gambar 2.6(c), model wilayah diterapkan pada citra masukan sebagai masking sehingga wilayah yang cocok akan berkurang. Hasilnya terlihat pada gambar 2.7



Gambar 2.7 *Prosedur metode parsing iteratif ramanan*  
Sumber : (Ramanan, 2006)

Ada beberapa masalah dengan penguraian berulang. Pertama-tama, itu tidak dapat mencapai kinerja yang baik untuk *figur* skala kecil dan beberapa *pose* umum. Kedua, karena ukuran bagian-bagian dalam model tumpang tindih antara bagian-bagian tubuh akan terjadi pada hasil keluaran. Ketiga, algoritma tidak tahu bagaimana menangani bagian tubuh yang tidak terlihat di beberapa gambar, dan ini adalah masalah utama yang dalam penelitian mengenai pose tubuh. Selain pencocokan *template*, penelitian melakukan pendekatan yang menggunakan lokasi bagian tubuh untuk membangun model klasifikasi. Namun kinerja pengklasifikasian sangat tergantung pada keakuratan lokasi bagian, dan *metode Ramanan* masih belum cukup akurat

## 2.7 Pengenalan tindakan manusia

Proses deteksi manusia dalam rekaman *video* sangatlah penting bagi banyak aplikasi termasuk interaksi manusia-komputer, pengakuan aktivitas manusia, dan pengawasan otomatis. Berbagai karakteristik manusia diusulkan sebagai fitur untuk klasifikasi. Klasifikasi tersebut seperti warna kulit, bentuk tubuh, dan pola berjalan. Meskipun setiap klasifikasi memberikan

informasi yang cukup untuk aplikasi tertentu, mereka saling melengkapi satu sama lain untuk proses deteksi manusia.

Pada penelitian (J. Russel, 2004), beliau mengusulkan sebuah sistem yang menggabungkan beberapa pengklasifikasi yaitu mengisolasi latar depan objek, pelacakan benda bergerak, dan klasifikasi objek sebagai manusia

Pengenalan tindakan manusia sangat berharga dalam aplikasi praktis. Metode bervariasi sesuai dengan tipe data. Untuk urutan frame, informasi dalam waktu domain adalah kuncinya. Pendekatan biasanya berbasis ruang negara. Untuk gambar diam, pengenalan tindakan dapat dianggap sebagai *estimasi pose*.

### **2.7.1 Estimasi Pose**

*Estimasi pose* dalam gambar diam biasanya terkait dengan deteksi manusia. *Estimasi pose* dapat dianggap sebagai masalah klasifikasi multi kernel, dan deteksi manusia dapat dianggap sebagai masalah klasifikasi *kernel* tunggal. Oleh karena itu metode yang digunakan untuk pendeteksian manusia, baik berbasis pemrosesan sinyal maupun berbasis pembelajaran mesin, dapat diterapkan untuk *estimasi pose*. Dalam sudut pandang lain, ia juga memiliki hubungan yang erat dengan analisis struktur manusia. Jika struktur tubuh, yang berarti posisi bagian-bagian tubuh, telah ditemukan, sangat mudah untuk mengidentifikasi *pose*.

(Wang dkk.2022) menemukan analisis *estimasi pose* dengan membandingkan gambar untuk diklasifikasikan. Dalam penelitian ini, algoritma pencocokan *template* yang dapat di ekstrak satu set titik sampel dari tepi dan menemukan penugasan antara titik sampel, digunakan. Setelah mekanisme pengelompokan *spektral*, gambar diam akan diklasifikasikan ke dalam kelompok aksi yang berbeda. Dari kedua pendekatan yang disebutkan yaitu mengekstraksi struktur tubuh terlebih dahulu, merepresentasikan struktur dengan beberapa jenis *fitur*, dan akhirnya mengklasifikasikan tubuh yang diekstraksi ke kelas yang telah ditentukan sebelumnya. Bagian dari penelitian

sebelumnya ini menggambarkan bahwa tindakan sederhana manusia dapat dikenali dengan memperkirakan pose, seperti pose yang bersifat normal dan bersifat *anomaly*

### 2.7.2 Tahapan *Estimasi Pose*

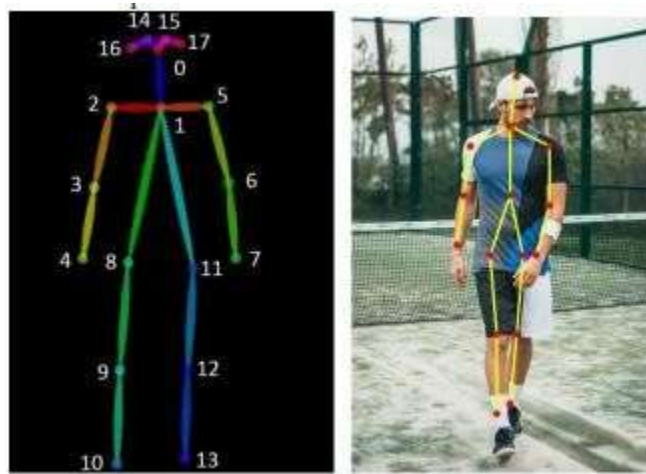
Proses utama *estimasi pose* manusia dibagi menjadi dua langkah-langkah dasar: I) melokalisasi sendi/titik kunci tubuh manusia dan; II) mengelompokkan sendi-sendi itu ke dalam konfigurasi *pose* manusia yang valid. Pada langkah pertama, fokus utama adalah menemukan lokasi masing-masing *keypoint* manusia seperti Kepala, bahu, lengan, tangan, lutut, pergelangan kaki.

Tabel 1. Contoh dataset *keypoint* Manusia

COCO output format	MPH output format
Nose - 0	Head - 0
Neck - 1	Neck - 1
Right Shoulder - 2	Right Shoulder - 2
Right Elbow - 3	Right Elbow - 3
Right Wrist - 4	Right Wrist - 4
Left Shoulder - 5	Left Shoulder - 5
Left Elbow - 6	Left Elbow - 6
Left Wrist - 7	Left Wrist - 7
Right Hip - 8	Right Hip - 8
Right Knee - 9	Right Knee - 9
Right Ankle - 10	Right Ankle - 10
Left Hip - 11	Left Hip - 11
Left Knee - 12	Left Knee - 12
Left Ankle - 13	Left Ankle - 13
Right Eye - 14	Chest - 14
Left Eye - 15	
Right Ear - 16	
Left Ear - 17	
Background - 18	Background - 15

Sumber Tabel. 1 (Tewodros, Yalew. dkk.2017)

Tabel 1. menunjukkan bagaimana dataset *keypoint* dari setiap tubuh manusia, diaman tahap ini mengumpulkan dan mengidentifikasi sendi-sendi ini dapat dilakukan melalui salah satu format kumpulan data populer yang berbeda, sedemikian rupa sehingga *keypoints* disimpan dalam dataset yang dipilih.

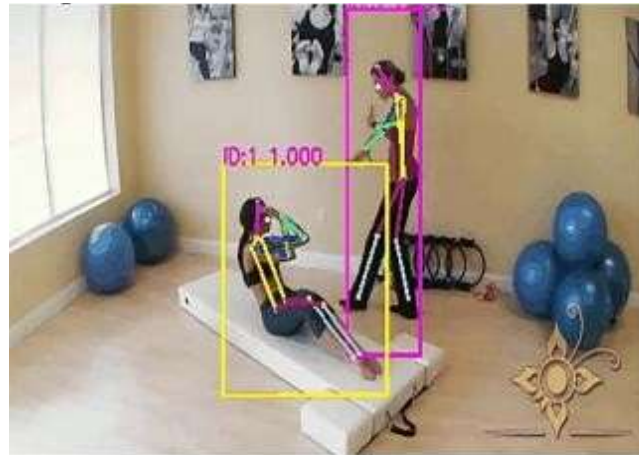


Gambar 2.8 Contoh *Estimasi Human Pose*  
Sumber : (Tewodros, Yalew. dkk.2017)

Langkah selanjutnya adalah mengelompokkan sendi-sendi itu menjadi manusia yang valid konfigurasi pose yang menentukan suku-suku berpasangan antara bagian tubuh seperti yang terlihat pada gambar 2.8

Dua pendekatan umum digunakan dalam memperkirakan pose individu dalam gambar tertentu, adalah 1) *Top-down Approach*, pemrosesan dilakukan dari rendah ke tinggi resolusi, ikuti deteksi contoh *individu* di gambar pertama menggunakan detektor objek kotak pembatas dan kemudian fokus menentukan *pose* mereka selanjutnya,





Gambar 2.9 *Top-Down Approach*  
Sumber : (Tewodros, Yalew. dkk.2017)

Pendekatan ini memiliki tingkat kegagalan tinggi, yang berarti jika deteksi individu gagal, tidak ada kemungkinan dapat terdeteksi. Juga, rentan ketika beberapa individu berada di dekatnya. Sehingga hasil komputasi tergantung pada jumlah orang di gambar, semakin banyak orang semakin rumit untuk komputasi

### 2.7.3 *Anomali Manusia*

Menurut (Jeffrey S,dkk.2002), Perilaku Anomali adalah kondisi emosional seperti kecemasan dan depresi yang tidak sesuai dengan situasinya. Perilaku *Anomali* terdiri dari dua kata yaitu Perilaku dan *Anomali*. Perilaku menurut kamus bahasa Indonesia adalah tingkah laku seorang manusia/sikap seorang manusia, sedangkan *Anomali* dapat didefinisikan sebagai hal yang jarang terjadi (seperti kidal) ataupun penyimpangan dari kondisi rata-rata (seperti tinggi badan yang ekstrem). *Anomaliitas* umumnya ditentukan berdasarkan munculnya beberapa karakteristik sekaligus dan definisi terbaik untuk ini menggunakan beberapa karakteristik kejarangan statistik, pelanggaran norma, distress pribadi, ketidakmampuan atau disfungsi, dan repons yang tidak diharapkan (*unexpectedness*).

Sumber lain mengatakan Perilaku *Anomali* adalah perilaku yang menyimpang dari norma sosial. Karena setiap masyarakat mempunyai patokan atau norma tertentu, untuk perilaku yang sesuai dengan norma maka dapat diterima, sedangkan perilaku yang menyimpang secara mencolok dari norma ini dianggap *Anomali*. sehingga perilaku yang dianggap normal oleh suatu masyarakat mungkin dianggap tidak normal oleh masyarakat lain, jadi gagasan tentang kenormalan atau ke *Anomalian* berbeda dari satu masyarakat lain dari waktu ke waktu dalam masyarakat yang sama. Perilaku *Anomali* yang terjadi pada kondisi emosional biasa terjadi kapan saja dalam kehidupan manusia, Mereka kadang-kadang bisa terjadi dan sudah terjadi dalam kehidupan orang lain. Sebuah masalah emosional dapat menyebabkan seseorang mengalami gangguan secara mental dan fisik. Para ahli kesehatan mental menggunakan berbagai kriteria dalam membuat keputusan tentang apakah suatu perilaku *Anomali* atau tidak. Kriteria umum yang digunakan adalah :

1. Perilaku yang tidak biasa. Perilaku yang tidak biasa sering dikatakan *Anomali*. Hanya sedikit dari kita yang menyatakan melihat ataupun mendengar sesuatu yang sebenarnya tidak ada.
2. Perilaku yang tidak dapat diterima secara sosial atau melanggar norma sosial. Setiap masyarakat memiliki norma-norma (Standar) yang menentukan jenis perilaku yang dapat diterima dalam beragam tertentu. Perilaku yang dianggap normal dalam satu budaya mungkin akan dipandang sebagai *Anomali* dalam budaya lainnya.
3. Persepsi atau interpretasi yang salah terhadap realitas. Sistem sensor dan kognitif yang memungkinkan seseorang untuk membentuk representasi mental yang akurat tentang lingkungan sekitar. Namun melihat sesuatu ataupun mendengar suara yang tidak ada objeknya akan disebut sebagai halusinasi, dimana dalam budaya sering dianggap sebagai tanda-tanda yang mendasari suatu gangguan.

4. Orang-orang tersebut berada dalam stress personal yang signifikan. Kondisi stress personal yang diakibatkan oleh gangguan emosi, seperti kecemasan, ketakutan, atau depresi, dapat dianggap *Anomali*. Namun kecemasan dan depresi terkadang merupakan respon yang sesuai dengan situasi tertentu. Gangguan emosi dapat mempengaruhi kejiwaan seseorang, sehingga seseorang yang menggunakan pikiran akan tetapi tidak bisa mengendalikan, maka bisa berakibat stress. Emosi ini menghalangi seseorang karena tindakan-tindakan yang dilakukannya tersebut pada umumnya merupakan tindakan fisik, dalam era sekarang tindakan fisik jarang untuk memecahkan suatu persoalan.
5. Perilaku *mal adaptif* atau *self-defeating*. Perilaku yang menghasilkan ketidakbahagiaan dan bukan pemenuhan diri dapat dianggap sebagai *Anomali*. Perilaku yang membatasi kemampuan kita untuk berfungsi dalam peran yang diharapkan atau untuk beradaptasi dengan lingkungan juga dapat disebut sebagai *Anomali*.
6. Perilaku berbahaya. Perilaku yang menimbulkan bahaya bagi orang itu sendiri ataupun orang lain dapat dikatakan *Anomali*.

## 2.8 Penelitian Sebelumnya

Tabel 2. Penelitian sebelumnya mengenai *estimasi pose* manusia adalah sebagai berikut :

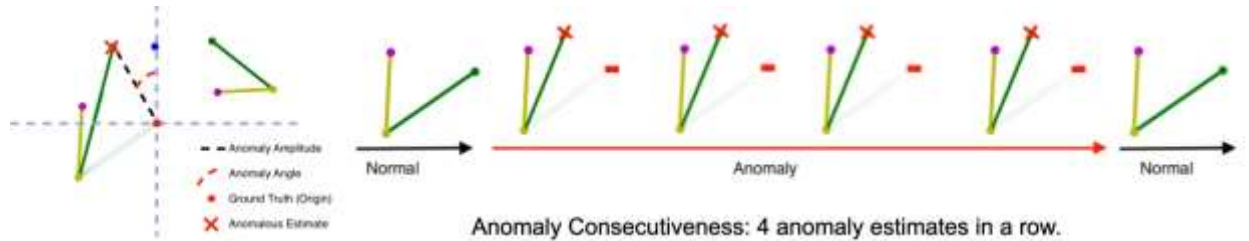
No	Penelitian	Metode	Hasil
1	M. Liu and J. Yuan, “Recognizing human actions as the evolution of pose estimation maps,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1159–1168.	Metode yang digunakan Body Pose Evolution Map.	Menggunakan Pose 2D dan 3D, Hasil pada NTU-RGB+D 120 dataset tingkat Accuracy(C-Subject) 64.6 % dan tingkat Accuracy (C-Setting) 66.9 %.
2	P. Zhang, C. Lan, W. Zeng, J. Xue, and N. Zheng, “Semantics-guided neural networks for efficient skeleton-based human action recognition,” arXiv preprint arXiv:1904.01189, 2019.	Metode yang digunakan Semantics - Guided Neural Network (SGN).	Perhitungan Pada posisi dan koordinat 3D dan kecepatan. Hasil pada NTU-RGB+D 120 dataset tingkat Accuracy(C-V) 93.4 % dan Accuracy(C-Subject) 86.6 %.
3	Mousavi, H., Mohammadi, S., Perina, A., Chellali, R., Mur, V.: Analyzing tracklets for the detection of abnormal crowd behavior. In: IEEE Winter Conference on Applications of Computer Vision Workshop (WACV), pp. 148–155 (2015)	Menggunakan Analyzing tracklets for the detection.	Penggunaan gambar, Performance comparison on violence-in-crowd dataset 78.3 %.
4	Tian, W., Snoussi, H.: Histograms of optical flow orientation for visual abnormal events detection. In: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 13–18 (2012)	Menggunakan Histogram of oriented optic flow Performance comparison on violence-in-crowd	Hasil dataset Accuracy sebesar 58.5%. Masih menggunakan gambar.

5	Sabokrou, M., Fathy, M., Hoseini, M., Klette, R.: <i>Real-time anomaly detection and localization in crowded scenes</i> . In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015, pp. 56–62 (2015)	Menggunakan metode <i>real-time anomaly detection and localization</i> pada keadaan ramai.	Performance comparison on UMN Dataset. EER 2.5% dan AUC 99.6%.  Melakukan estimasi 3D <i>Pose</i> tubuh manusia menggunakan gambar sebagai sumber.
6	Li, W., Mahadevan, V., Vasconcelos, N.: Anomaly detection and localization in crowded scenes. <i>IEEE Trans. Pattern Anal. Mach. Intell.</i> 36, 18–32 (2013)	Menggunakan model Normalcy And Anomaly.	Performance comparison on UMN Datas EER akurasi 3.7% dan AUC 99.5 %
7	Qiulei Han, Haofeng Wang <sup>2</sup> , Lin Yang, Min Wu, Jinqiao Kou, Qinsheng Du, Nianfeng Li. Real-time adversarial GAN-based abnormal crowd behavior detection. <i>Journal of Real-Time Image Processing</i> . 31 October 2020, in : <i>Journal of Real-Time Image Processing</i> .	Menggunakan <i>Generative Adversarial Network</i> (GAN).	Menganalisis <i>video</i> dari <i>youtube</i> sebanyak 246 video kekerasan nyata, Dataset dirancang memberikan dasar pengujian untuk klasifikasi kekerasan/Non Kekerasan dan standar kekerasan, kemudian deteksi kerumunan yang bersifat <i>anomaly</i> .  Performance comparison on UMN Dataset. EER 4.1 and AUC 99.8. Performance comparison on violence-in-crowd dataset Accuracy 81.8 %

## Bab 3

### Metodologi

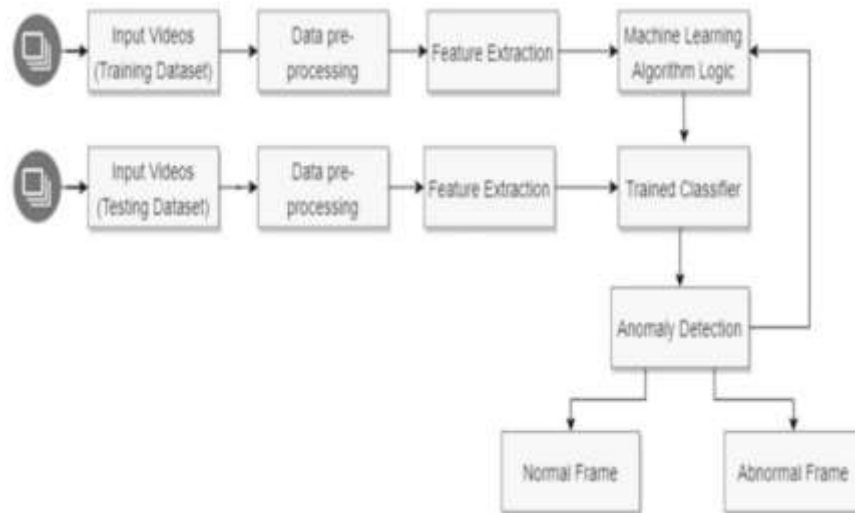
Deteksi manusia adalah tahap awal dari analisis gerak manusia, pada penelitian ini bertujuan membuat model anomali dari hasil video dan menghasilkan analisis dataset, dengan objek perilaku manusia di tempat keramaian Indonesia, motivasi dari penelitian menghasilkan deteksi pose anomali baru dari data pergerakan manusia di tempat keramaian, sehingga mendapat dataset dan sampel perilaku normal dan anomaly pada objek masyarakat di Indonesia. Penerapan sampel data anomaly memiliki pendekatan dengan memodelkan amplitudo, sudut, dan kontinuitas temporal atau "berturut-turut" dari yang terdeteksi anomali untuk setiap sendi. Yang dapat dikontrol tentang deteksi anomali pada sendi yang berbeda.



Gambar 3.1 Contoh visualisasi bentuk anomaly  
Sumber : (Haotian Zhang, 2019)

### 3.1 Pendekatan Penelitian

Untuk pendekatan penelitian model anomaly pada penelitian ini, memiliki langkah sebagai berikut :



Gambar 3.2 Alur Metode Penelitian  
(Sumber : Devashree dan Mayur, 2022)

Gambar 3.2 merupakan alur dari metode penelitian dengan tahapan sebagai berikut :

1. Input *video* adalah masukkan utama dari penelitian ini berupa kamera pengawas, atau *video* yang merekam aktivitas kerumunan
2. *Pre processing* adalah proses perbaikan citra untuk mengklasifikasi dari subjek *video*, dari tahap ini dapat terlihat pola *pose* dari pergerakan manusia normal atau pun *anomaly*, pada *epoch selection* menggunakan algoritma *HOG* dan *SVM*
3. *Feature Extraction* adalah Hasil data dari *preprocessing* yang disimpan dalam dataset pose dalam bentuk 3D dan dilakuka proses ekstrak data dengan memfokuskan pada objek manusia, dan mengabaikan objek yang lain
4. *Machine Learning Algorithm logic*, merupakan proses pelatihan hasil dari ekstrak yang dilakukan proses algoritma untuk melatih setiap pola individu yang terekam. Pada tahap ini *machine learning* menggunakan metode *Generative adversarial network*
5. Dalam menciptakan dataset pada *machine learning* dilakukan proses *trained classifier*, yaitu membuat klasifikasi dari pola yang didalam dari *Generative adversarial network* dan *machine learning* untuk membedakan satu pola dengan pola lainnya
6. Hasil dari *trained classifier* berupa pola dari deteksi anomaly pada individu, yang terdiri dari normal *frame* dan anomaly *frame*.





## Bibliografi

Kenny Chen, Paolo Gabriel, Abdulwahab Alasfour, Chenghao Gong, Werner K Doyle, Orrin Devinsky, Daniel Friedman, Patricia Dugan, Lucia Melloni, and Thomas Thesen. *Patient-specific pose estimation in clinical environments*. IEEE Journal of Translational Engineering in Health and Medicine, 6:1–11, 2018.

Diederik P Kingma and Max Welling. *Auto-encoding variational bayes*. arXiv preprint arXiv:1312.6114, 2013.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. *Isolation forest*. In 2008 Eighth IEEE International Conference on Data Mining, pages 413–422. IEEE, 2008.

Roberto Perdisci, Guofei Gu, and Wenke Lee. *Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems*. In ICDM, volume 6, pages 488–498. Citeseer, 2006.

Kevan Yuen and Mohan M Trivedi. *Looking at hands in autonomous vehicles: A convnet approach using part affinity fields*. IEEE Transactions on Intelligent Vehicles, 2018.

Zhijie Fang and Antonio M L´opez. *Is the pedestrian going to cross? answering by 2d pose estimation*. In 2018 IEEE Intelligent Vehicles Symposium (IV), pages 1271–1276. IEEE, 2018.

Yao-Jen Chang, Shu-Fang Chen, and Jun-Da Huang. *A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities*. Research in developmental disabilities, 32(6):2566–2570, 2011.

Zhengyou Zhang. *Microsoft kinect sensor and its effect*. IEEE multimedia, 19(2):4–10, 2012.

Aouaidjia Kamel, Bin Sheng, Po Yang, Ping Li, Ruimin Shen, and David Dagan Feng. *Deep convolutional neural networks for human action recognition using depth maps and postures*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, (99):1–14, 2018.

Zhenghua Chen, Qingchang Zhu, Yeng Chai Soh, and Le Zhang. *Robust human activity recognition using smartphone sensors via ct-pca and online svm*. IEEE Transactions on Industrial Informatics, 13(6):3070–3080, 2017.

Yaqiang Yao, Yan Liu, Zhenyu Liu, and Huanhuan Chen. *Human activity recognition with posture tendency descriptors on action snippets*. IEEE Transactions on Big Data, 4(4):530–541, 2018.

Gunnar Farnebäck. *Two-frame motion estimation based on polynomial expansion*. In Proceedings of the 13th Scandinavian Conference on Image Analysis (SCIA), 2003.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. *Microsoft coco: Common objects in context*. In European conference on Computer Vision, pages 740–755. Springer, 2014.

Leonid Sigal, Alexandru O Balan, and Michael J Black. *HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion*. International journal of Computer Vision, 87(1-2):4, 2010.

Laurens van der Maaten and Geoffrey Hinton. *Visualizing data using t-sne*. Journal of machine learning research, 9(Nov):2579–2605, 2008.

Yunchen Pu, Zhe Gan, Ricardo Henao, Xin Yuan, Chunyuan Li, Andrew Stevens, and Lawrence Carin. *Variational autoencoder for deep learning of images, labels and captions*. In Advances in neural information processing systems, pages 2352–2360, 2016.

Daniel Im Jiwoong Im, Sungjin Ahn, Roland Memisevic, and Yoshua Bengio. *Denoising criterion for variational auto-encoding framework*. In Thirty-First AAAI Conference on Artificial Intelligence, 2017.

Jeffrey S. Nevid, Spencer A. Rathus, Beverly Greene. *Psikologi Abnormal*. Edisi 5 Jilid 1. Jakarta: Erlangga. hlm. 5, 2002.

Meng Xua , Youchen Wangb , Bin Xuc , Jun Zhangc , Jian Renb , Stefan Poslada , Pengfei Xuc. *A Critical Analysis of Image-based Camera Pose Estimation Techniques*. Preprint submitted to Journal of Elsevier, 2022.

Atmoko Nugroho, *Pengembangan Model Pembelajaran Jarak Jauh Berbasis Web*. P-ISSN : 1693-3656, E-ISSN : 2460-6731 Vol 9 No 2, 2012.

Fadlisyah dan Rizal, *Pemrograman computer vision pada video menggunakan Delphi+ Vision Lab VCL 4.0.1*, 2011.

Putra, Darma. *Pengolahan Citra Digital (Edisi 1)* .Yogyakarta:Andi. 2010

Bambang Yuwono, Yusri Ilza Sania, Mangaras Yanu F., Panji Dwi Ashrianto, *Ecognition Face & Facial Expressions Recognition*, Buku Saku Pengolahan Citra, 2019.

Eric Secada Purba, Hendry, *Implementation Of Generative Adversarial Networks For Creating Digital Artwork In The Form Of Abstract Images*, 2018.

P. Zhang, C. Lan, W. Zeng, J. Xue, and N. Zheng, “Semantics guided neural networks for efficient skeleton-based human action recognition,” arXiv preprint arXiv:1904.01189, 2019.