

## **BAB 2**

### **TINJAUAN PUSTAKA**

#### **2.1 Gesture Recognition**

*Gesture* atau bahasa tubuh adalah gerakan tubuh secara spontan yang biasanya menyertai komunikasi verbal. Bagian tubuh yang umum digunakan adalah tangan, jari tangan, kepala, wajah, mata, alis dan badan. cara yang paling alami yang dapat menghubungkan citra visual dengan bahan yang disajikan adalah melalui *gesture*. Dengan kata lain *gesture* merupakan komunikasi *non-verbal* yang dilakukan manusia dengan menggunakan anggota tubuhnya.

*Gesture Recognition* adalah salah satu alternatif cara yang bisa digunakan untuk berinteraksi dengan komputer. *Gesture* dapat diterapkan di berbagai macam aplikasi, robotik, *game* bahasa isyarat dan sebagainya [2]. Teknologi ini merupakan salah satu jawaban para peneliti guna memberikan pengalaman yang baru dalam hal interaksi manusia dengan komputer. *Gesture Recognition* memerlukan suatu teknologi khusus yang mampu membaca serta menerjemahkan gerakan yang dibuat oleh manusia menjadi suatu perintah yang dapat dibaca dan diproses oleh komputer. *Gesture recognition* memungkinkan interaksi dengan komputer tanpa perlunya *Graphical User Interface (GUI)*. Saat ini kamera biasa dapat untuk melakukan *gesture recognition* melalui algoritma tertentu.

##### **2.1.1 Pose Estimation**

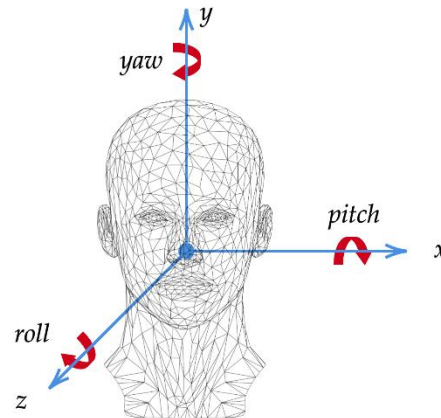
*Pose Estimation* adalah metode menggunakan model *Machine Learning* untuk memperkirakan pose seseorang dari gambar atau video dengan memperkirakan lokasi spasial sendi atau *keypoints*.

*Pose Estimation* mengacu pada teknik pengolahan citra yang mendeteksi sosok manusia dalam gambar dan video, sehingga dapat ditentukan dimana posisi siku seseorang muncul dalam sebuah gambar. Penting disadari fakta bahwa estimasi pose hanya memperkirakan dimana sendi tubuh utama berada dan tidak mengenali siapa yang ada dalam video atau gambar [14].

### 2.1.2 Head Pose Estimation

*Head pose estimation* adalah salah satu bagian dari *computer vision* yang ada. Pada bagian difungsikan sebagai bagaimana cara mengetahui pose objek dari transisi dan rotasinya. *Head pose estimation* memiliki banyak fungsi, misalnya dapat digunakan sebagai mendeteksi pengemudi memperhatikan jalan atau tidak, contoh kedua adalah pose estimation dapat memeriksa bahwa seorang siswa ketika belajar merasa terganggu atau tidak, dan masih banyak lagi fungsi dari *head pose estimation*.

Secara formal, *head pose estimation* berfungsi sebagai deteksi gerakan arah kepala dari hasil perhitungan deteksi wajah berupa nilai *yaw*, *pitch*, and *roll* dengan menggunakan salah satu metode yaitu *triangle geometry* yang kemudian diklasifikasikan kedalam *range* nilai arah untuk menghasilkan keluaran berupa kalimat yang menyatakan arah pergerakan kepala seperti kanan, kiri, atas dan bawah[3]. Pada gambar 2.3 berikut gambaran dari *axis head pose estimation*



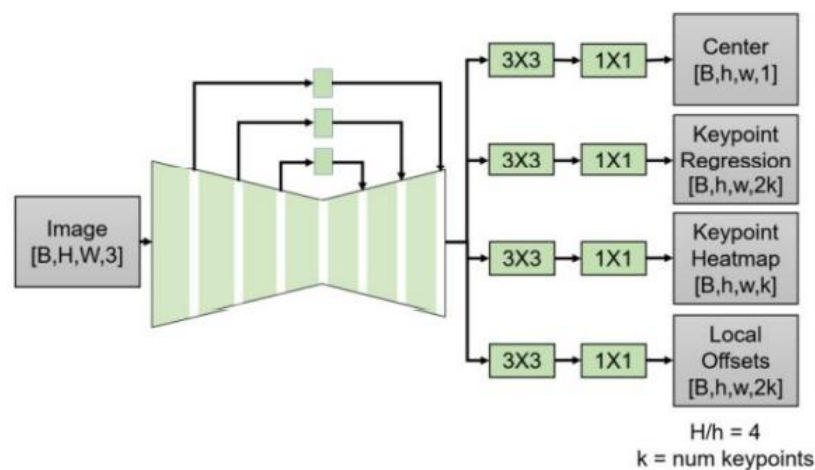
Gambar 2. 1 Main Axis Dari Head Pose Estimation

### 2.2 Movenet

Movenet adalah model inferensi berbasis Google yang dikembangkan oleh IncludeHealth, sebuah perusahaan kesehatan digital. IncludeHealth meluncurkan model tersebut pada tahun 2021, dan meminta bantuan dari Google untuk mendukung perawatan pasien jarak jauh. Mirip dengan *PoseNet*, versi web

MoveNet menggunakan TensorFlow.js, dan versi seluler menggunakan TensorFlow Lite. Ada dua versi MoveNet: *Lightning* yang berorientasi pada kinerja, dan *Thunder* yang berorientasi pada akurasi. Kedua model berbeda dalam ukuran input dan pengganda kedalaman. Dalam hal input, *Lightning* menerima video atau gambar dengan ukuran tetap (192x192) dan tiga saluran, dan menggunakan pengganda kedalaman 1,0. Sebaliknya, *Thunder* menerima input berukuran 256x256 dan tiga saluran, dan menggunakan pengali kedalaman 1,75. Pengganda kedalaman mengubah jumlah saluran video atau gambar input, yang umumnya mengadopsi format merah-hijau-biru (RGB). Namun peta fitur juga dapat dianggap sebagai satu saluran di setiap lapisan. Sementara itu, *Thunder* memiliki lapisan 1,75 kali lebih banyak untuk pembelajaran mendalam daripada *Lightning*.

MoveNet adalah model bottom-up yang mengandalkan API deteksi objek TensorFlow dan MobileNet V2 sebagai ekstraktor fitur. Dalam API deteksi objek TensorFlow, ada beberapa model deteksi yang mendukung TensorFlow 1 dan TensorFlow 2. MoveNet mengikuti CenterNet, sebuah API deteksi. Berbeda dari model deteksi berbasis *anchor (bounding box)* standar, CenterNet mengambil titik tengah sebagai satu-satunya jangkar, dan mencari dan mengklasifikasikan objek dengan memproses proposal regional, daripada menyimpulkan objek dari nilai IoU.



Gambar 2. 2 Alur Dari Library Movenet

Seperti yang ditunjukkan pada Gambar 2.1, MoveNet menghitung keempat proses secara bersamaan. Setelah peta panas pusat orang disiapkan untuk mengidentifikasi setiap orang, lokasi dengan skor tertinggi dipilih. Kemudian, satu set poin-poin kunci untuk orang tersebut diinisialisasi berdasarkan poin-poin kunci yang diperoleh melalui regresi. Seseorang diidentifikasi, ketika regresi cocok dengan pengaturan setiap *keypoints* yang disiapkan.

Selanjutnya, setiap piksel dikalikan dengan bobot, yang berbanding terbalik dengan jarak dari titik kunci yang diregresi. Dengan cara ini, poin-poin kunci dari orang-orang di latar belakang dikeluarkan dari perhitungan. Pada akhirnya, kumpulan titik kunci diselesaikan sesuai dengan nilai peta panas maksimum di setiap saluran titik kunci [4].

### **2.3 Robot Ikon Unikom**

Robot Nakula merupakan robot ikon milik Universitas Komputer Indonesia, yang dikembangkan oleh Divisi Robotika Unikom. Robot tersebut dibuat dan diperkenalkan pertama kali pada bulan september tahun 2017 pada acara penyambutan mahasiswa baru Unikom di gedung Sasana Budaya Ganesha (Sabuga). Robot ini merupakan robot semi humanoid yang berukuran sekitar 1.2 meter berwarna putih polet biru, robot ini memiliki 3 *DOF*, dan beberapa sensor seperti sensor kamera dan *microphone*. Robot Nakula mempunyai kemampuan untuk bernavigasi, bersalaman, berbicara dengan manusia, dan berekspresi melalui mimik wajahnya. Dibekali dengan beberapa Mikrokontroler yang difungsikan untuk *bernavigasi* dan berinteraksi.

Fungsi utama dari robot ini yaitu digunakan sebagai robot pentas yang biasa ditampilkan pada acara khusus Universitas Komputer Indonesia, seperti acara penyambutan mahasiswa baru dan acara wisuda. Robot ini juga pernah diundang untuk acara besar pada saat Seminar Digital & Risk Management in Insurance (DRiM) pada tahun 2019 di Nusa Dua Bali. Berikut merupakan photo dari robot Ikon Nakula, ditunjukan oleh gambar 2.2.



Gambar 2. 3 Photo Robot Ikon Nakula

## 2.4 Tensorflow

Tensorflow adalah perpustakaan perangkat lunak yang dikembangkan oleh tim Google Brain dalam organisasi penelitian Mesin Cerdas Google untuk tujuan melakukan pembelajaran mesin dan penelitian jaringan syaraf dalam. Tensorflow kemudian menggabungkan aljabar komputasi teknik pengoptimalan kompilasi, mempermudah penghitungan banyak ekspresi matematis dimana masalahnya adalah waktu yang dibutuhkan untuk melakukan perhitungan. Fitur utamanya meliputi:

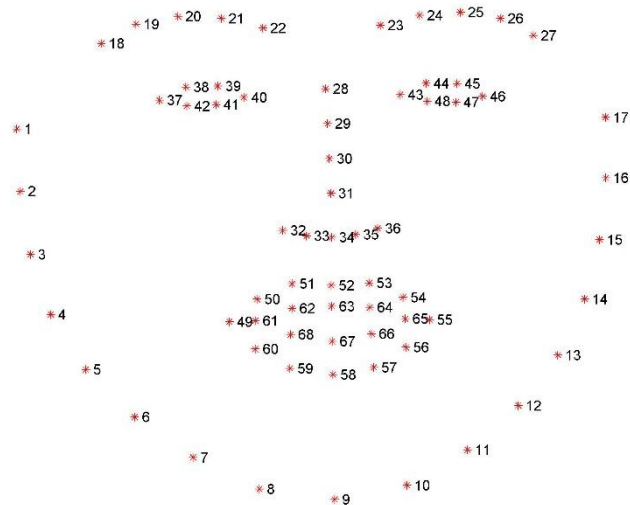
1. Mendefinisikan, mengoptimalkan, dan menghitung secara efisien ekspresi matematis yang melibatkan array multi dimensi (tensors).
2. Pemrograman pendukung jaringan saraf dalam dan teknik pembelajaran mesin
3. Stabilitas komputasi yang tinggi di seluruh mesin dan kumpulan data yang besar

4. Penggunaan GPU yang transparan, mengotomatisasi manajemen dan optimalisasi memori yang sama dan data yang digunakan. Tensorflow bisa menulis kode yang sama dan menjalankannya baik di CPU atau GPU. Lebih khusus lagi, tensorflow akan mengetahui bagian perhitungan mana yang harus dipindahkan ke GPU.

Tensorflow adalah sebuah perpustakaan *software* sumber terbuka yang digunakan untuk mesin belajar dalam berbagai macam tugas pemahaman persepsi dan bahasa. Tensorflow adalah API generasi kedua, menggantikan *DistBelief* (API generasi pertama) yang digunakan oleh 50 tim yang berbeda untuk penelitian dan pengembangan puluhan produk komersial milik Google, seperti *Speech Recognition*, *Gmail Image*, dan *Google Search*. Tensorflow pada awalnya dikembangkan oleh Brain Google Team untuk tujuan penelitian dan produksi Google, kemudian dirilis dengan sumber terbuka dibawah lisensi Apache 2.0 pada 9 november 2015. Tensorflow dapat berjalan pada *crossplatform* seperti Windows, Linux dan Mac OS [5].

## **2.5 DLIB**

Dlib merupakan *library* yang berfungsi menganalisis bagian wajah dengan mengesktrasi nilai gambar. Dengan mengekstrasi nilai pada wajah manusia DLIB akan menghasilkan 68 koordinat. DLIB akan digunakan untuk membantu mengolah wajah pada metode *facial landmark*[6]. Indeks dari 68 koordinat ditunjukan oleh gambar 2.3.



Gambar 2. 4 Dlib Facial Landmark

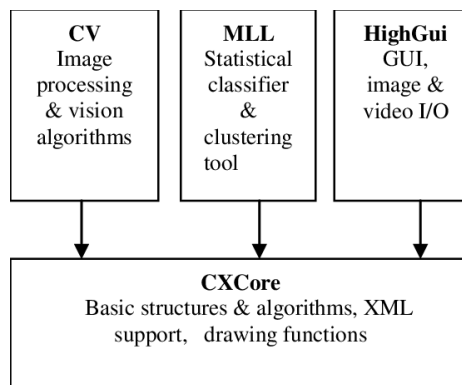
Dlib merupakan *library* yang berfungsi menganalisis bagian wajah dengan mengekstraksi nilai gambar. Dengan mengekstraksi nilai pada wajah manusia Dlib akan digunakan untuk membantu mengolah wajah pada metode *facial landmark*.

## 2.6 Computer Vision

*Computer Vision* merupakan salah satu dari banyak bidang dalam ilmu komputer. *Computer Vision* adalah penggunaan komputer untuk mengenali dan mengklasifikasikan isyarat visual, dengan tujuan untuk mendapatkan informasi dari gambar, *computer vision* juga bisa digunakan pada video dikarenakan video pada dasarnya hanyalah serangkaian gambar. Ada beberapa cara *computer vision* dapat dicapai, dari menggunakan model matematika hingga menggunakan teknik yang lebih modern, seperti *machine learning*. Pada zaman sekarang yang serba canggih ini, *computer vision* digunakan untuk berbagai macam hal, mulai dari pengenalan dan klasifikasi [7].

Salah satu *library* yang seringkali dipakai, yaitu adalah OpenCV (Open Computer Vision). OpenCV (Open Computer Vision) adalah sebuah *library* yang sudah sangat familiar pada bidang pengolahan citra yang bersifat open source. *Library* OpenCV ditulis menggunakan bahasa pemrograman C dan C++ dan dapat dijalankan pada sistem operasi Linux, Windows dan MAC OS X. *Library* OpenCV memiliki lebih dari 2500 algoritma yang sangat berkaitan dengan *computer vision* dan *machine learning*. Algoritma tersebut digunakan untuk mendeteksi, mengenali

dan mengklasifikasikan objek [8]. OpenCV juga bisa digunakan untuk meningkatkan kualitas dari gambar. Beberapa pengimplementasian dari OpenCV diantaranya adalah *Face Recognition*, *Face Detection*, *Object Detection*, *Pose Estimation* dan masih banyak yang lainnya [9]. Struktur dari OpenCV terdiri dari *basic image processing* dan algoritma *computer vision* tingkat tinggi. MLL merupakan *Machine Learning Library* yang terdiri dari *Statistical Classifiers* dan *Clustering Tools*. HighGUI terdiri dari *Input/Output* dan fungsi untuk menyimpan dan memanggil video dan gambar. CXCore terdiri dari struktur basis dan konten. Struktur dari OpenCV dapat dilihat pada Gambar 2.4.



Gambar 2. 5 Struktur Open CV

## 2.7 Perangkat Keras

Perangkat Keras adalah perangkat yang dapat di sentuh secara langsung dan juga dapat dilihat oleh manusia, sedangkan perangkat keras komputer adalah perangkat fisik dari sebuah komputer yang bisa beroperasi dengan baik sehingga dapat mempermudah kegiatan atau pekerjaan dari manusia dan berfungsi mendukung suatu proses komputasi. Berikut beberapa perangkat keras yang ada untuk membangun sistem ini.

### 2.7.1 Laptop

Laptop atau komputer jinjing adalah sebuah komputer pribadi yang berukuran relatif kecil dan ringan. Beratnya sekitar 1-3 kg, tergantung ukuran, bahan, dan spesifikasi laptop tersebut. Dipilihnya laptop pada sistem ini dikarenakan laptop sudah mewakili kebutuhan minum sistem untuk berjalan, selain itu laptop memiliki mobilitas yang tinggi sehingga ketika digunakan [10]. Performa laptop jika



dibandingkan dengan komputer pada umumnya tidak kalah hebat, lalu laptop menggunakan sumber daya dari baterai yang akan memudahkan jika dalam suatu acara tidak terdapat sumber listrik yang tidak memadai . Gambar Laptop dapat dilihat pada gambar 2.5.



Gambar 2. 6 Laptop

### 2.7.2 Kamera

Kamera merupakan sebuah alat yang dipasang dengan lensa untuk mengambil sebuah citra digital yang diperoleh secara *realtime*, pada sistem digunakan untuk diproses menangkap gambar dari operator pengendali [11]. Citra yang diambil yaitu bentuk tubuh dan gerakan manusia. Gambar Kamera dapat dilihat pada Gambar 2.7



Gambar 2. 7 Kamera

### 2.7.3 Wireless Router

*Wireless Router* merupakan suatu perangkat keras pada jaringan komputer yang berfungsi untuk memancarkan gelombang radio sehingga komputer yang telah memiliki *wireless card* bisa terhubung ke jaringan *wireless*. *Wireless router* ini dikenal juga dengan nama *wireless broadband* atau *access point* [12]. Setiap *router* memiliki fasilitas *DHCP* (*Dynamic Host Configuration Protocol*) yang dapat diatur dengan sedemikian rupa sehingga dapat membagi *IP address*. *Router* juga terdapat *NAT* (*Network Address Translator*) yaitu fasilitas yang memungkinkan suatu alamat IP atau koneksi internet dapat dibagi dengan alamat IP lain. *Router* dapat dilihat pada gambar 2.7.



Gambar 2. 8 Wireless and Router