



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## BAB II

### LANDASAN TEORI

#### 2.1 Computer Vision Berbasis Pengenalan Postur dan Isyarat Tangan

Menurut Harshith (2010), *computer vision-based hand posture and gesture recognition* merupakan teknologi yang mengenali dan menginterpretasikan posisi serta gerak tangan yang direkam oleh kamera video. Video yang ditangkap ini dibagi berdasarkan fitur dan diproses per *frame*. Gambar tangan akan dipisahkan dari bagian tubuh lain serta latar belakang, dan menjadi *point of interest* dalam pemrosesan. Gambar tangan yang telah dipisahkan akan dikenali postur serta gerakannya, dibandingkan dengan *data set* berisi data postur gerak tangan yang telah dikenali, kemudian digunakan sebagai kontrol untuk aplikasi. Konsep ini yang menjadi dasar metode pembuatan logika dan algoritma dalam pembangunan aplikasi ini.

Caputo (2012) juga menjelaskan pada *hand gesture recognition* berbasis sensor, posisi, postur, serta gerak tangan dapat dikenali secara 3D (memasukkan posisi tangan dalam bidang  $x, y, z$ ) dengan menggunakan kombinasi kamera 2D untuk menangkap gambar RGB dan *Time-of-flight camera*, yakni sistem kamera yang menggunakan sinyal cahaya untuk mengukur jarak antar kamera ke objek, lalu menghasilkan keluaran *image stream* yang mengandung info kedalaman (*depth*) dari objek

Dua dari beberapa kategori klasifikasi gerak tangan menjadi perhatian utama dalam *computer vision-based interface*, yaitu *controlling gestures*, bagaimana menggunakan tangan sebagai alat kontrol dan manipulasi komputer; dan

*manipulating gestures*, yang menyediakan interaksi natural dengan objek virtual (Wu dan Thomas, 2013). *Manipulating gesture* umumnya terkait dengan translasi, rotasi, dan skalasi objek virtual.

### 2.1.1 Contoh Metode Hand Recognition Berdasarkan Fingertips dan Palm

Penelitian-penelitian berikut mengembangkan teknik *hand recognition* berdasarkan *fingertips* dan *palm*, atau jari dan telapak tangan.

#### 1. *Binary Representation of Hand Recognition*

Pradhan (2012) pada penelitiannya menangkap gambar tangan dan mengubahnya dari model warna RGB (*Red, Green, Blue*) ke HSV (*hue-saturation-value*). Dari gambar tersebut diberlakukan *threshold* kepada nilai *Hue* dan *Saturation* dengan nilai yang sesuai dengan warna kulit manusia. Lalu gambar dikonversi ke *image* biner, tiap pixelnya berisi 1 jika nilai *Hue* dan *Saturation*-nya sesuai dengan *threshold*, dan diberikan nilai 0 pada tiap pixel jika tidak memenuhi *threshold*. Kemudian digunakan operasi morfologi, erosi terhadap *noise* di sekitar gambar untuk memperjelas hasil gambar.

Teknik *skeletonization* digunakan untuk mendapatkan posisi *fingertips*. Dimana dilakukan penipisan *pixel* pada gambar jari. Suatu *pixel* ditentukan sebagai sebuah *fingertips* jika hanya memiliki satu *pixel* tetangga dalam area  $3 \times 3$  *pixel*. Posisi *palm* didapat dari posisi rata-rata dari tiap *pixel* yang berada pada area tangan.

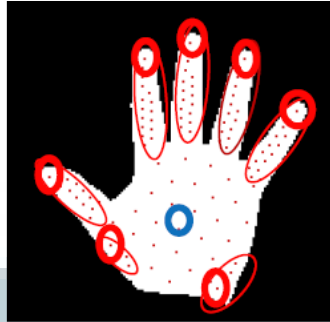


Gambar 2.1 *Skeletonization* pada Representasi Biner (sumber: Daren, 2013)

## 2. *Circle Template Matching*

Pada pengembangan algoritma pendeteksi *fingertips* dan *palm* menggunakan metode *circle template matching*, Mokhtar M. Hasan (2012) menggunakan teknik yang hampir sama dengan metode *Binary Representation of Hand Recognition*. Mengubah gambar RGB pada tangan menjadi HSV pada segmen tertentu yang memiliki warna kulit manusia. Pada penelitiannya, Hasan menggunakan  $0 \leq H \leq 0,1228$  untuk *threshold Hue* dan  $0,2666 \leq S \leq 0,8777$  untuk *threshold Saturation*. Lalu diberlakukan operasi morfologi untuk mengurangi *noise* pada gambar.

Setelah didapatkan *image* tangan bebas *noise*, diberlakukan metode *circle template matching* untuk mendapatkan posisi *fingertips* dan *palm*. Pada metode ini sudah disediakan dua *template* lingkaran, yaitu *finger circle template* (FCT) dengan radius berapapun yang dapat masuk ke area jari, dan *palm circle template* (PCT) dengan radius berapapun yang dapat masuk ke area tangan tanpa menyentuh area jari. Bagian tangan pada *image* yang didapat kemudian di cocokkan per *pixel* dengan FCT dan PCT untuk mendapatkan titik-titik kandidat *fingertips* dan *palm*. Titik-titik kandidat merupakan titik pusat FCT dan PCT yang dapat masuk ke dalam *image* tangan. Proses dilakukan dengan pencocokan PCT terlebih dahulu, sehingga saat pencocokan dengan FCT, area tangan yang dapat di cocokkan sudah berkurang karena sudah digunakan oleh PCT. Dengan proses ini pencocokan dapat di percepat. Ilustrasi metode *Circle Template Matching* dapat dilihat pada gambar 2.2.



Gambar 2.2 Ilustrasi Metode *Circle Template Matching*  
(sumber: Dareen, 2013)

## 2.2 Perceptual Computing dan RealSense

*Perceptual Computing* dapat mengubah cara interaksi terhadap *device* yang digunakan. Dengan gerakan tangan, deteksi jari, deteksi suara, pengenalan wajah, dan lain-lain. *Perceptual Computing* membuat *user* lebih natural, intuitif, dan imersif dalam melakukan interaksi (Doss, 2013). Untuk membantu dalam pengembangan, penelitian ini menggunakan Intel RealSense SDK 2013, sebuah *library* untuk mengimplementasikan *pattern detection* dan *recognition algorithm*.

Penelitian ini mengimplementasikan *perceptual computing* dengan bantuan alat dan *software* berupa Intel RealSense 3D Camera, sebuah kamera dari Creative beserta RealSense SDK dari Intel. Kamera tersebut mampu mendeteksi suara atau *speech recognition*, *facial tracking*, *finger tracking and gesture recognition*, dan *2D/3D object tracking*. Pada penelitian ini pembuatan aplikasi menggunakan *finger tracking and recognition* untuk menginterpretasi atau menerjemahkan bahasa isyarat.

### 2.2.1 Intel RealSense Camera F200

Intel RealSense Camera F200 adalah sebuah perangkat *input motion sensor* yang dibangun oleh perusahaan Singapura, Creative, untuk komputer yang menggunakan prosesor Intel. Perangkat dapat mengenali gerakan tangan dan jari, pose yang dilakukan oleh tangan dan jari, *face recognition* dan *emotion recognition*, serta *voice recognition*. Perangkat ini adalah kamera generasi kedua dari Creative Interactive Gesture Camera yang dibuat untuk *perceptual computing technology* dari Intel. Kamera ini dapat menangkap kedalaman suatu gambar dengan *depth sensor* yang menggunakan *infrared*, dimana gambar yang di dapat di ilustrasikan sebagai objek tiga dimensi dengan sumbu pengukuran x, y, dan z.



Gambar 2.3 Intel RealSense Camera F200  
(sumber: intel.com)

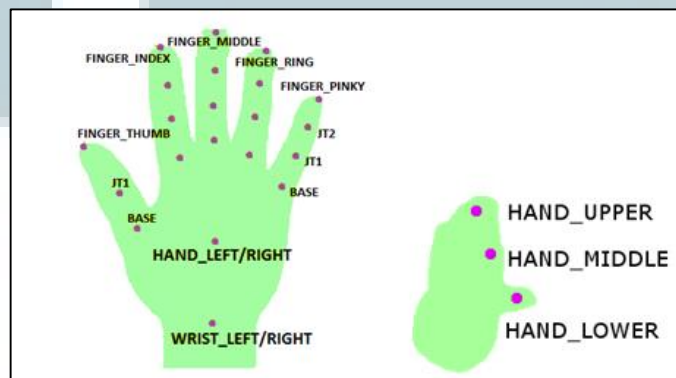
Berikut ini merupakan detail spesifikasi dari Intel RealSense Camera F200

1. *VGA depth resolution*
2. *1080p RGB Camera*
3. *0.2 – 1.2 meter jarak optimal pemakaian*
4. *USB 3.0 Interface*
5. *Dual-array microphone*

Berikut merupakan syarat perangkat keras yang diperlukan sebuah komputer untuk mengimplementasikan Intel RealSense Camera F200.

1. 4th generation Intel® Core™ processors, code name Haswell
2. 4GB hard disk
3. 4GB RAM
4. USB 3.0 port
5. Sistem Operasi Microsoft Windows 8.1 64-bit

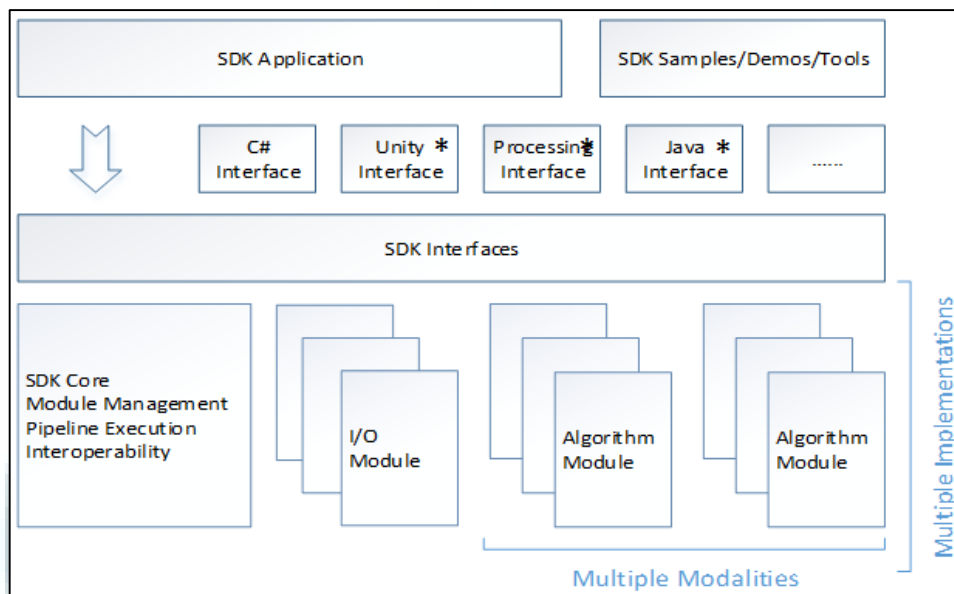
Pengembangan menggunakan kamera tersebut memerlukan sebuah *Software Development Kit* (SDK) yang dinamakan Intel RealSense SDK. SDK tersebut mengubah sebuah masukan *gesture* tangan menjadi *Geometric Node*. *Geometric Node* adalah sendi-sendi tulang pada tubuh manusia, dalam hal ini adalah tangan dan jari.



Gambar 2.4 Hand Labels  
(sumber: intel.com)

Pada arsitektur Intel RealSense SDK, *SDK Core*, *I/O Module*, dan *Algorithm Modules* merupakan pondasi dari susunan SDK. *SDK Core* membantu dalam menangani eksekusi *pipeline* dan *I/O Modules* (kamera). *Algorithm Modules* merupakan *middleware* untuk *hand tracking*, *gesture recognition*, *face detection*,

voice detection dan modul-modul lainnya. Gambar 2.5 menjelaskan struktur arsitektur dari Intel RealSense SDK.



Gambar 2.5 Arsitektur Intel RealSense SDK  
(sumber: intel.com)

### 2.2.2 Contoh Aplikasi

Berikut ini merupakan beberapa contoh aplikasi yang menggunakan teknologi RealSense .

1. Kagura ([www.kagura.cc](http://www.kagura.cc))

Kagura merupakan aplikasi instrumen musik yang menggunakan teknologi Intel RealSense. Aplikasi ini memungkinkan pengguna untuk memainkan, membuat, merekam, dan memanipulasi musik hanya dengan gerakan tangan. Kagura memanfaatkan kamera *motion sensor* sebagai masukan pengguna untuk beroperasi dalam aplikasi. Kagura merupakan *Grand Prize Winner* dari Intel Perceptual Computing Challenge 2013. Gambar 2.6 menggambarkan tampilan antarmuka dari aplikasi Kagura.

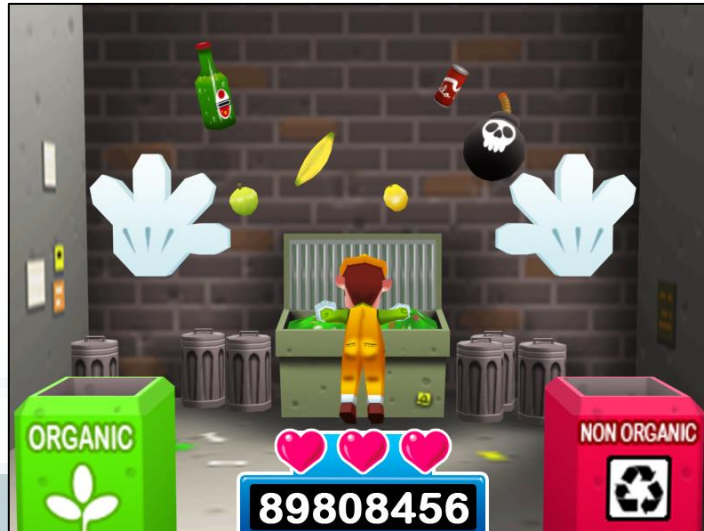




Gambar 2.6 Tampilan antarmuka aplikasi Kagura

## 2. The Catcher

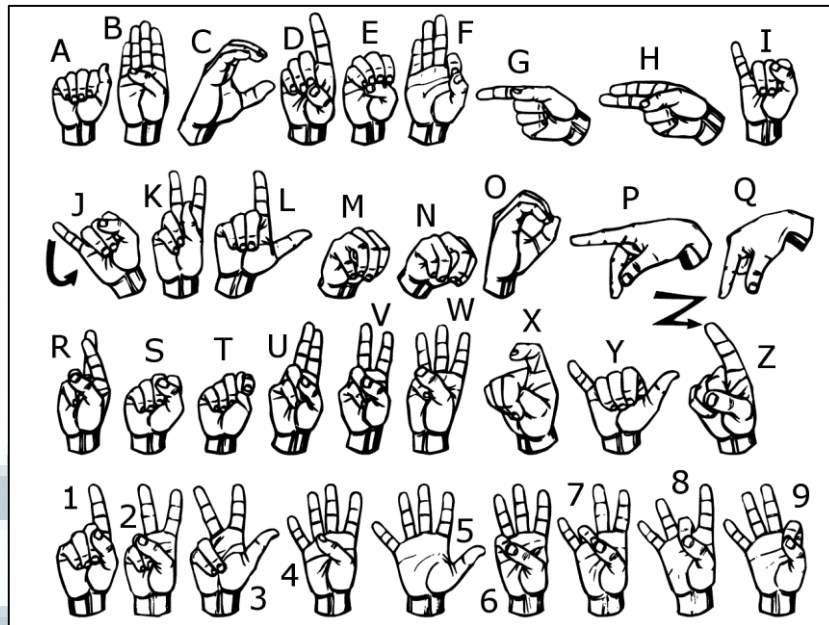
The Catcher merupakan aplikasi permainan (*game*) yang menggunakan teknologi Intel *Perceptual Computing*. Aplikasi ini adalah sebuah permainan dimana pengguna harus menangkap benda yang bergerak dan memasukkan benda tersebut ke dalam beberapa kategori. Aplikasi ini menggunakan *motion sensor camera* untuk menangkap gerakan tangan sebagai masukan dalam permainan. The Catcher adalah salah satu finalis Intel Perceptual Computing Challenge 2013 buatan Aditya Anwar yang berasal dari Indonesia. Tampilan antarmuka aplikasi The Catcher dapat dilihat pada gambar 2.7.



Gambar 2.7 Tampilan antarmuka aplikasi The Catcher

### 2.3 Bahasa Isyarat (Sign Language)

Menurut Oxford University Press, bahasa isyarat atau *sign language* merupakan suatu sistem komunikasi yang menggunakan gerakan visual dan isyarat yang biasanya digunakan oleh orang yang tuna rungu. Ada banyak macam-macam bahasa isyarat, antara lain ASL (*American Sign Language*), BSL (*British Sign Language*), dan BISINDO (Berkenalan dengan Sistem Isyarat Indonesia). Dalam hal ini, peneliti memilih menggunakan ASL karena menurut peneliti bahasa tersebut lebih mudah untuk diinterpretasikan. Latar belakang dari *American Sign Language* adalah bahasa alami dengan struktur bahasa yang berbeda dengan bahasa Inggris, digunakan sebagai sarana komunikasi utama untuk kurang lebih satu setengah juta orang di Amerika Serikat (Mitchell, 2006).



Gambar 2.8 Alfabet dan numerik dalam *American Sign Language* (ASL)  
(Sumber: lifeprint.com)



Gambar 2.9 Alfabet dalam BISINDO  
(Sumber: gerkatinsolo.or.id)

Dapat dilihat pada gambar 2.8 dan 2.9, perbandingan bahasa isyarat dalam hal huruf atau alfabet antara ASL dengan BISINDO. ASL hanya menggunakan satu tangan, sedangkan BISINDO cenderung menggunakan dua tangan. Karena hal ini peneliti memilih mengembangkan aplikasi dengan ASL.