



### **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 I

## PENDAHULUAN

### 1.1 Latar Belakang

Komputer menjadi merasuk dalam kehidupan kita sehari-hari, orang-orang memiliki kebutuhan besar terhadap interaksi manusia-komputer yang alami dan efisien (Shan, Tan & Wei, 2007). Saat ini cara yang paling populer dari interaksi manusia-komputer yakni dengan menggunakan *keyboard*, *mouse* dan layar sentuh, yang tidak wajar bagi manusia untuk berinteraksi dengan komputer dan lambatnya kecepatan dalam berinteraksi. Semakin komputer dapat mengerti cara komunikasi alami manusia, akan memberikan pengalaman berinteraksi yang lebih ramah, mudah, dan efisien. Hal ini membuat adanya tekanan besar terhadap penelitian di bidang interaksi manusia-komputer dalam membuat antar muka yang mudah digunakan dengan secara langsung menggunakan komunikasi alami dan ketrampilan manipulasi manusia (Jalab & Omer, 2015).

Di antara tubuh manusia yang berbeda bagian, tangan adalah bagian yang paling efektif untuk tujuan umum alat interaksi, karena ketangkasannya. (Jalab & Omer, 2015). Gerakan tangan adalah gerakan yang memiliki makna atau disengaja dari tangan manusia dan lengan (Shan, Tan & Wei, 2007). Sebagai bahasa tubuh manusia yang universal, gerakan tangan adalah salah satu yang paling alami dan memiliki arti yang efektif untuk manusia berkomunikasi secara non verbal kepada orang lain (Shan, Tan & Wei, 2007). Dengan pertumbuhan teknologi yang begitu pesat khususnya kemampuan komputer yang telah membantu manusia dalam

mejalankan aktivitas sehari-hari, diperlukan komputer yang mengerti bahasa alami manusia untuk meningkatkan produktivitas manusia sebagai pengguna. Komunikasi antara manusia dengan komputer menggunakan bahasa alami manusia akan meningkatkan fungsi komputer agar semakin ekspresif terhadap manusia. Kemampuan mengenali gerak tangan dapat digunakan untuk meningkatkan kemampuan sebuah mesin dalam interaksi manusia-komputer (Maung, 2009). Saat ini mayoritas orang menggunakan *mouse*, *keyboard* dan layer sentuh sebagai alat untuk berinteraksi dengan manusia dibandingkan dengan menggunakan cara alami manusia berkomunikasi menjadi masalah interaksi antara manusia dengan komputer (Jalab & Omer, 2015). Mengadopsi kemampuan mengenali gerak tangan pada sebuah komputer dirasa dapat menciptakan interaksi manusia-komputer yang lebih efektif dan dapat membangun jembatan baru antara mesin dan manusia yang akan menggantikan teks dan bahkan penggunaan *GUI* sebagai alat untuk berinteraksi.

Tujuan keseluruhan adalah untuk membuat komputer mengerti gerak tubuh manusia, untuk mengontrol jarak jauh melalui postur tangan dan dapat digunakan di berbagai perangkat (Hasan & Abdul-Kareem, 2014). Untuk itu, diperlukan otomatisasi pengenalan gerak tangan berdasarkan penglihatan yang dapat mendeteksi dan melacak objek baik secara statis (gambar) ataupun secara kontinu (video). Meskipun untuk melakukan pengenalan gerak tangan cukup sulit, sudah banyak hasil penelitian yang menunjukkan performa yang sangat baik dalam mengenali objek dan klasifikasi objek pada gambar, sebagai contoh pada penelitian Semanet et al. (2013), menunjukkan bahwa pada perbandingan banyak model tidak

ada yang menunjukkan nilai *error rate* melebihi 16%, model tersebut digunakan untuk mendeteksi objek pada dataset ILSVRC 2012 dan 2013. Penelitian (Krizhevsky, Sutskever & Hinton 2012) berhasil menaklukkan *ImageNet* 2012 *classification benchmark*, model yang dibuat berhasil mencapai *error rate* yang cukup rendah 16.4% dibandingkan dengan posisi kedua 26.1% (Zeiler & Fergus., 2013). Namun, masih banyak ruang untuk mengembangkannya. Sebagian besar dari penelitian terdahulu hanya fokus terhadap memperkirakan sikap tangan yang tunduk terhadap beberapa asumsi atau mengandalkan detektor lemah untuk mendeteksi tangan manusia (Chen, Wu, Hsieh & Fu, 2016).

Agar dapat membangun sistem yang mampu melacak tangan pada sebuah video, diperlukan sebuah model yang mampu mengenali corak atau ragam dari postur tangan yang ingin dilacak. Model pelacak memerlukan model deteksi sebagai alat bantu untuk memberi informasi berupa keberadaan tangan dan posisi tangan pada sebuah video untuk masing-masing *frame*. Maka dari itu, diperlukan sebuah model yang memiliki akurasi yang cukup tinggi untuk mendeteksi tangan manusia. Kemudian setelah model deteksi mampu mendeteksi tangan, model tersebut dapat digunakan sebagai salah satu modul penting dalam membuat model pelacak tangan. Pembuatan model pelacak dapat dilakukan dengan menseleksi pola pada gambar yang kemudian pola tersebut akan menjadi masukkan dan sebagai informasi yang dibutuhkan oleh model deteksi dan nantinya model deteksi akan menghasilkan hasil dari klasifikasi objek.

Untuk bagian deteksi, kita perlu mencari tahu posisi tangan dari lingkungan yang kompleks (Chen, Wu, Hsieh & Fu, 2016). Pertama, tangan manusia mungkin

memiliki beberapa bentuk dan skala karena sudut pandang yang berbeda, jarak, dan subjek yang menghasilkan variasi intra-kelas yang tinggi. Kedua, hasil deteksi dapat digunakan untuk melacak tangan sebagai objek dalam gambar baik pada gambar statis (gambar) maupun gambar kontinu (video).

Beberapa metode penelitian telah bermunculan untuk mengenali sikap pada penampilan 2D. Hal ini terbukti dari hasil survei dalam penelitian Chaudhary, Raheja, Das & Raheja (2011) dimana ada beberapa metode yang berhasil digunakan untuk mengenali sikap tangan sebagai objek dalam gambar. *Artificial neural network*, *genetic algorithm*, dan *fuzzy logic based approach* merupakan salah satu metode yang digunakan untuk mengenali sikap. Penelitian terkait dengan deteksi dan pengenalan sikap tangan yang dilakukan Nguyen, Hyunh & Meunier (2013), Chen, Geroganas & Petriu (2007), dan Savaris & Wangenheim (2010) berhasil untuk mengenali tangan sebagai objek dalam gambar. Namun, dari metode yang dipakai dari ketiga penelitian tersebut, masih memiliki keterbatasan dalam mengenali objek. Metode yang digunakan belum dapat mempelajari sudut, gradien, intensitas piksel, dan bentuk secara bersamaan. Penelitian-penelitian tersebut masih memerlukan sistem terpisah untuk membuat komputer mengerti warna, bentuk, sudut, dan gradien suatu objek.

Untuk dapat melakukan pengenalan objek dan klasifikasi objek *convolutional neural network* menjadi alat yang kuat dalam belajar mengenali objek terlebih jika dikombinasikan dengan *max-pooling* (Nagi et al., 2011). Penelitian terhadap deteksi dan pengenalan objek sampai pengenalan sikap menggunakan *convolutional neural network* yang dilakukan oleh Tompson, Stein,

Lecun & Perlin (2014), Chen, Wu, Hsieh & Fu (2016), Nagi et al. (2011), dan Nowlan & Platt (1995) berhasil mendeteksi dan mengenali objek menggunakan teknik CNN. Penelitian-penelitian tersebut memberikan solusi terhadap sulitnya menggunakan skema atau algoritma *machine learning* dalam mengenali seluruh aspek pada sebuah gambar seperti warna, sudut, bentuk, dan gradien. Akan tetapi penelitian-penelitian tersebut juga masih memiliki kekurangan diantaranya adalah keterbatasan melacak tangan ketika tangan berinteraksi dengan objek lain, pelacakan tangan memerlukan komputasi yang cukup banyak dan memerlukan pelatihan sebelum dapat melacak objek, model yang digunakan belum dapat mendeteksi objek berdasarkan warna alami dan memerlukan bantuan alat seperti sarung tangan dalam pembuatan dataset, deteksi objek memiliki variasi bentuk yang minim dan sulitnya mendeteksi objek ketika mempunyai latar yang kompleks atau memiliki corak yang mirip dengan objek sekelilingnya.

Banyak model *convolutional network* telah menunjukkan klasifikasi yang mengesankan belakangan ini (Zeiler & Fergus, 2013). Namun tidak ada pemahaman jelas mengapa mereka dapat bekerja lebih baik dalam pengolahan citra dibandingkan dengan model *network* lainnya seperti halnya *artificial neural network*. *Convolutional neural network* menunjukkan pemahaman yang lebih baik terhadap sebuah gambar. Tidak heran jika sekarang teknik CNN digunakan pada banyak penelitian terhadap *computer vision*.

Pelacakan adalah salah satu bidang dalam *computer vision* yang mendasar dan cukup penting. Menurut penelitian yang dilakukan oleh Fan, Xu, Wu & Gong (2010), pelacakan dilakukan dengan cara tradisional yaitu berdasarkan *featured*

*based method* yakni warna ataupun gerakan gumpalan pada gambar. Penelitian (Beleznai, Frühstück & Bischof, 2006) menggunakan cara tradisional yang disebutkan sebelumnya dengan *background subtraction*. Penelitian tersebut menggunakan algoritma *mean-shift* untuk menghitung perpindahan objek pada sebuah video. Hasil dari penelitian tersebut menunjukkan performa yang lebih baik dari *blob-based method* atau gerakan gumpalan dimana tingkat error penyimpangan antara posisi sebenarnya dengan estimasi pada sistem pelacak 31% untuk *blob-based method* dan 14% untuk *background subtraction* atau pendekatan yang diusulkan. Kekurangan dari pendekatan yang diusulkan ialah deteksi yang dilakukan hanya berdasarkan bentuk dasar tubuh manusia pada kejauhan sehingga ketika objek memiliki bentuk yang tidak umum hasil deteksi mungkin tidak dapat mengenali objek.

Penelitian (Fan, Xu, Wu & Gong, 2010) mengusulkan metode yang lebih modern yaitu menggunakan *convolutional neural network* untuk melacak manusia. Pada penelitian tersebut CNN digunakan untuk mendeteksi manusia dan untuk menghasilkan *probability map* yang menunjukkan posisi dari target pada *frame* tertentu dalam video. Hasil penelitian tersebut berhasil membuktikan CNN dapat digunakan untuk melacak objek dan dengan *shift-variant*, CNN dapat mengestimasi posisi objek pada video. Namun, kelemahan pada hasil penelitian ini adalah model memerlukan data yang cukup banyak untuk pelatihan model dan kompleksitas data pelatihan cukup tinggi dan model yang dihasilkan masih belum dapat menangani gangguan penuh pada oklusi yang lama.

Skripsi ini adalah penelitian yang bermaksud membuat sebuah model yang mampu melacak tangan manusia dengan membangun sebuah arsitektur jaringan yang mampu mendeteksi tangan pada gambar dengan kondisi tangan tidak menggunakan alat bantu untuk menseleksi warna, variasi sikap tangan yang akan dideteksi memiliki bentuk yang bervariasi, deteksi dapat dilakukan pada latar yang kompleks. Teknik deteksi pada penelitian ini menggunakan *convolutional neural network*, dimana metode ini pada umumnya sangat efektif untuk mengenali objek atau pengolahan citra. *Convolutional neural network* terbukti telah dipakai selama bertahun-tahun untuk mengenali sebuah kategori dominan pada objek dalam gambar dengan berbagai macam pola dan corak pada objek (Semanet et al., 2013). Setelah deteksi tangan dapat dilakukan, tahap kedua yaitu mengembangkan model tersebut agar dapat melacak tangan dengan menggunakan metode *image pyramid* dan *mean-shift* untuk melakukan pelacakan. Penggunaan *image pyramid* digunakan untuk mencari posisi objek yang terdeteksi dan penggunaan *mean shift* digunakan sebagai pelacak objek saat berpindah posisi pada gambar.

Penelitian ini akan dibagi menjadi dua tahap yaitu (1) penelitian dan pembuatan model deteksi tangan (2) penelitian dan pembuatan model pelacakan tangan. Tahap pertama akan dikerjakan selama skripsi 1 dan tahap kedua akan dikerjakan selama skripsi 2. Penelitian dan pembuatan model deteksi tangan akan ditargetkan selesai diakhir skripsi 1 dimana model deteksi berhasil dilatih dan mampu mendeteksi tangan dan kemudian akan dilakukan analisa terhadap performa pelatihan, validasi, dan pengujian model deteksi tangan, sedangkan penelitian dan pembuatan model pelacakan tangan akan ditargetkan selesai diakhir skripsi 2

dimana model pelacak tangan mampu memproses data dan melacak tangan pada data tersebut dan kemudian model pelacak tangan akan diuji dan dianalisa performanya.

## 1.2 Rumusan Masalah

Untuk melakukan pelacakan tangan dibutuhkan sistem yang mampu mendeteksi tangan dan kemudian sebuah sistem yang mampu melacak posisi tangan yang bergerak pada gambar. Oleh karena itu, masalah yang muncul adalah sebagai berikut:

1. Bagaimana cara implementasi *convolutional neural network* untuk mendeteksi tangan manusia dan implementasi *image pyramid* dan *mean shift* untuk melacak posisi tangan pada gambar?
2. Bagaimana akurasi yang dihasilkan pada model arsitektur *convolutional neural network* yang dibangun dalam mendeteksi tangan manusia dan *image pyramid* serta *mean shift* dalam melacak tangan?

## 1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah: (1) Membuat model pengolahan citra yang dapat melacak tangan manusia. (2) Mengetahui tingkat akurasi pada model yang dibangun untuk melacak tangan manusia.

#### **1.4 Hasil Keluaran**

Adapun hasil keluaran dari penelitian yang diharapkan adalah sebuah model komputasi pengolahan citra yang mampu melacak tangan manusia.

