

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Banyak *website* yang memuat CAPTCHA sebagai alternatif untuk meningkatkan keamanan *webnya* dari *spam*. Tes yang dimunculkan, seperti menebak huruf atau angka yang muncul, mencocokkan dua buah gambar, sampai menemukan apa tema dari gambar yang dimunculkan. Namun dalam kenyataannya, masih ada beberapa jenis CAPTCHA yang mampu dipecahkan oleh *spam*. Ini dapat berdampak sangat buruk bagi pihak pengguna maupun pemilik *website*. Contoh *website* yang membutuhkan keamanan yang lebih tinggi dibandingkan *web* lainnya adalah *web* judi *online*. Di dalamnya terjadi proses registrasi yang akan memuat data diri *user* hingga nomor rekening para *user*. Atas dasar itu, perlu dibuat sebuah pembaca otomatis yang bisa mengautentifikasi apakah CAPTCHA yang ditampilkan pada *web* tersebut sudah memiliki tingkat keamanan yang tinggi atau perlu ditingkatkan.

Penelitian sebelumnya tentang segmentasi CAPTCHA:

- Shih-Yu Huang, Yeuan-Kuen Lee, Graeme Bell and Zhan-he Ou dengan judul makalah “*A Projection-based Segmentation Algorithm for Breaking MSN and Yahoo CAPTCHAs*”.

Diawali dengan proses erosi dan kemudian akan dilebarkan beberapa kali. Erosi akan menghapus batas karakter satu *pixel* per waktu, sedangkan pelebaran akan memperbaiki perbatasan satu piksel per waktu. Setelah *clutter* tipis berhasil dihapus oleh proses erosi dilanjutkan dengan proses pelebaran yang akan mampu menghapus beberapa *item* dari *clutter*. Algoritma segmentasi yang digunakan adalah *CHELLAPILLA's Algorithm* yang bekerja ketika *clutter* lebih tipis daripada karakter. Namun, algoritma ini memiliki kelemahan karena tidak dapat mengenali perbedaan antara karakter dan *noise* dengan lebar yang sama. Tingkat akurasi segmentasi yang dicapai sistem ini adalah tidak lebih dari 70%.

Berdasarkan masalah tersebut pada tugas akhir ini dibuat simulasi pembaca otomatis untuk CAPTCHA yang memiliki tantangan segmentasi yang cukup tinggi, yaitu CAPTCHA *multiline background*. Dimana segmentasi bertujuan

untuk memisahkan obyek yang satu dengan yang lain dalam satu gambar. Selain itu pada penelitian sebelumnya sistem hanya dibuat hingga proses segmentasi, sehingga penulis mencoba untuk membuat sebuah sistem hingga proses pengenalan karakter.

Pada penelitian kali ini, CAPTCHA yang dianalisis adalah CAPTCHA *multiline background* yang mana memiliki karakteristik yang unik sehingga tantangan dalam menghilangkan *noise* serta mensegmentasi merupakan salah satu alasan penulis memilih CAPTCHA dengan jenis ini. Dengan adanya sistem pembaca otomatis untuk CAPTCHA ini, diharapkan perkembangan CAPTCHA akan semakin meningkat karena *programmer* akan berusaha membuat CAPTCHA dengan kesulitan yang lebih tinggi demi meningkatkan keamanan situs *web*.

## 1.2 Rumusan Masalah

Rumusan masalah yang terkait dalam tugas akhir ini yaitu yang berkaitan dengan proses adalah sebagai berikut :

1. Bagaimana cara menghilangkan *noise* pengganggu yang terdapat pada CAPTCHA *multiline background*?
2. Bagaimana proses kerja segmentasi CAPTCHA dengan *multiline background*?
3. Bagaimana merancang algoritma yang sesuai untuk membuat simulasi segmentasi CAPTCHA dengan *multiline background* pada MATLAB R2009a.
4. Bagaimana menguji kualitas hasil segmentasi dan akurasi sistem, serta waktu komputasi sistem dalam menjawab karakter yang dimunculkan pada CAPTCHA dengan *multiline background*.
5. Bagaimana membandingkan tingkat akurasi sistem pada Tugas Akhir ini dengan OCR?

## 1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penulisan tugas akhir ini antara lain:

1. Mengetahui cara menghilangkan *noise* pengganggu yang terdapat pada CAPTCHA *multiline background*.
2. Mengetahui proses kerja segmentasi CAPTCHA *multiline background*.

3. Merancang simulasi segmentasi CAPTCHA *multiline background* dengan menyusun algoritma yang sesuai pada MATLAB 2009a
4. Mengetahui kualitas hasil segmentasi dan keakuratan sistem serta waktu komputasi sistem dalam menjawab karakter yang dimunculkan pada CAPTCHA *multiline background*.
5. Mengetahui perbandingan sistem pembaca pada Tugas Akhir ini dengan OCR dari sisi tingkat akurasinya.

#### 1.4 Batasan Masalah

Untuk menghindari meluasnya pembahasan materi pada tugas akhir ini, maka permasalahan akan dibatasi pada hal-hal sebagai berikut:

1. Pada tugas akhir ini proses segmentasi hanya dilakukan berdasarkan CAPTCHA dengan karakteristik memuat latar belakang garis yang berjumlah lebih dari satu dengan ketebalan yang berbeda-beda, terdapat *noise* pengganggu berupa *noise salt and pepper*, jumlah karakter yang dimunculkan adalah enam angka, serta memiliki dua kemungkinan warna, yaitu latar belakang hitam dengan latar depan putih atau latar belakang putih dengan latar depan hitam.
2. Citra *input* menggunakan *file* format \*.bmp yang diambil dari *website*, [www.sbobet.com](http://www.sbobet.com) dengan ukuran 150x50 piksel.
3. *Input* sistem bersifat tidak *real time*.
4. Dalam penelitian ini hanya menggunakan alat simulasi berupa perangkat lunak MATLAB 2009a.
5. Algoritma ekstraksi ciri yang dibandingkan adalah algoritma *sum* dan *square sum*.
6. Parameter pengujian performansi sistem yang digunakan adalah tingkat akurasi segmentasi, akurasi karakter, akurasi *file*, serta waktu komputasi.

#### 1.5 Metode Penelitian

Metodologi yang digunakan dalam penulisan tugas akhir ini adalah:

1. Studi Literatur

Mempelajari konsep-konsep tentang pengolahan citra *digital* serta mempelajari tentang *preprocessing*, segmentasi, normalisasi karakter, ekstraksi ciri serta pengenalan karakter tulisan pada gambar.

2. Pengumpulan data

Bertujuan untuk mendapatkan data citra CAPTCHA *multiline background* yang akan digunakan sebagai masukan.

3. Perancangan sistem

Berdasarkan studi literatur dan pustaka, dibuat perancangan sistem sesuai kondisi yang diinginkan dengan menggunakan *software* MATLAB.

4. Studi analisis dan penarikan kesimpulan

Bertujuan untuk menguji kualitas *output* sistem serta menarik kesimpulan.

## 1.6 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika pembahasan sebagai berikut:

**BAB I            PENDAHULUAN**

Bab ini menguraikan latar belakang penulisan, rumusan masalah, batasan masalah, tujuan penulisan, metodologi penulisan, dan sistematika penulisan.

**BAB II           LANDASAN TEORI**

Pada bab ini berisi teori-teori yang mendukung dan mendasari penulisan tugas akhir ini, yaitu pengolahan citra *digital*, *preprocessing*, normalisasi karakter, metode segmentasi, serta pengenalan karakter tulisan pada gambar.

**BAB III          PERANCANGAN SISTEM DAN SIMULASI**

Bab ini menjelaskan tentang proses perancangan tahap pemrosesan segmentasi awal dan segmentasi akhir hingga teks yang ada pada CAPTCHA *multiline background* dapat terjawab.

**BAB IV          ANALISA DAN HASIL SIMULASI**

Pada bab ini berisi analisa terhadap hasil yang diperoleh dari tahap perancangan sistem dan simulasi.

**BAB V           KESIMPULAN DAN SARAN**

Bab ini berisikan kesimpulan dari analisa yang telah dilakukan dan saran untuk pengembangan ke depannya.