

ANALISIS DAN SIMULASI SISTEM PEMBACA OTOMATIS BERBASIS PENGOLAHAN CITRA DIGITAL UNTUK CAPTCHA MULTILINE BACKGROUND

I Dewa Ayu Putri Kartika Wulantari¹, Koredianto Usman², Suryo Adhi Wibowo³

¹Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

Abstrak

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) merupakan sebuah aplikasi yang mampu mengetahui bahwa yang melakukan autentikasi atau menginputkan data bukan spam atau aplikasi yang bekerja secara otomatis. Pada umumnya CAPTCHA berupa image yang memuat huruf atau angka yang dirusak dengan berbagai jenis noise. Tes ini melibatkan user untuk mengetikkan hasil tebakan huruf atau angka yang ada pada image. Tetapi terkadang masih banyak web yang tidak mengetahui tingkat keamanan dari CAPTCHA yang ditampilkannya, sehingga spam masih bisa menyamar sebagai manusia untuk mengakses dan menyerang web.

Pada tugas akhir ini, telah dibuat sebuah simulasi sistem pembaca otomatis untuk CAPTCHA multilinebackground, dimana memuat teks berupa angka yang disertai dengan gangguan berupa latar belakang garis dengan ketebalan yang berbeda-beda, memuat noise salt and pepper, serta memunculkan dua kemungkinan latar belakang warna, yaitu hitam dan putih. Untuk memecahkan teks yang ada pada CAPTCHA, dilakukan beberapa proses tahapan yaitu menghilangkan semua garis serta gangguan tambahan berupa noise salt and pepper, menghapus background yang bisa menyulitkan proses pengenalan karakter tulisan, serta melakukan pengenalan terhadap karakter angka itu sendiri, serta membaca tulisan apa yang dimuat pada CAPTCHA tersebut.

Dari hasil analisis dan pengujian, sistem mampu mencapai tingkat akurasi 25% untuk CAPTCHA berhasil keseluruhan, 78.3% untuk akurasi karakter, 90% untuk akurasi segmentasi dan waktu komputasi yang mencapai 0.2 detik.

Kata Kunci : CAPTCHA, multilinebackground, noise salt and pepper.

Abstract

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humamns Apart) is an application to authentication user, it is a human or a spam. Generally, CAPTCHA is an image that contain letter or number which is broken by noise. This test involve user to type the result of letter or number displayed. But almost all website owner didn't know how secure their CAPTCHA, so that spam can attack their website in act like human.

In this final project, already made a simulation for automated reader system for CAPTCHA multilinebackground, which is contain text with number, noise salt and pepper, multilinebackground with different thickness, and have two possibility background colour (white and black). To solve text displayed, there are many stage to remove all background which can make the detection is more difficult, and then find the feature to do the final stage recognition.

According to the analysis and testing, system have an accuration until 25% for whole accuration, 78.3% for character accuration, 90% for segmentation accuration, and the duration is about 0.2 seconds.

Keywords : CAPTCHA, multilinebackground, noise salt and pepper.

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 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.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis terhadap pengujian sistem pembaca otomatis CAPTCHA *multiline background*, maka dapat ditarik kesimpulan sebagai berikut:

1. Akurasi segmentasi mencapai angka 90%, dimana dari 120 CAPTCHA.
2. Tingkat akurasi berhasil keseluruhan mencapai angka 25% yang mana terjadi pada skema *resizing* 48x48 dengan algoritma *Square sum*, sedangkan pada skema dengan algoritma *Sum*, tingkat akurasi tertinggi mencapai angka 8.3333% dengan *resizing* 48x48.
3. Tingkat akurasi karakter berhasil mencapai angka 78.3333% yang mana terjadi pada skema *resizing* 48x48 dengan algoritma *Square sum*, sedangkan pada skema dengan algoritma *Sum*, tingkat akurasi tertinggi mencapai angka 62.6389% dengan *resizing* 24x24.
4. Untuk tingkat akurasi berhasil keseluruhan, pengukuran ulang pada CAPTCHA menghasilkan angka yang lebih baik ketika dilakukan *resizing* mendekati ukuran asli citra *input*.
5. Algoritma Ekstraksi Ciri dengan menggunakan *Square Sum* lebih baik dibandingkan dengan *Sum*.
6. Sistem mampu mencapai tingkat akurasi yang lebih tinggi dibandingkan dengan OCR, baik akurasi karakter, maupun akurasi *file*.
7. Waktu komputasi yang tercepat mencapai angka 0.2013 detik yang terjadi pada skema *Sum* dengan *resizing* 24x24, dimana diantara keseluruhan proses, *preprocessing* membutuhkan waktu yang lebih lama.

5.2 Saran

1. Penambahan jumlah ciri sebaiknya dilakukan untuk menambah acuan karakter pada saat pengenalan, disesuaikan dengan berbagai macam kemungkinan yang muncul pada data uji.
2. Upaya peningkatan akurasi dapat dilakukan dengan menambahkan skema pembobotan setelah algoritma *Square sum*, sehingga meminimalisir kesalahan pengenalan pada angka.
3. Dilakukan perbandingan antara sistem dengan OCR dengan database khusus angka, sehingga lebih akurat dalam membandingkan kehandalan sistem.
4. Sistem dibuat dalam skema *real time*.
5. Dilakukan sebuah verifikasi dalam menyatakan bahwa CAPTCHA yang dianalisis terdeteksi benar atau tidak.

DAFTAR PUSTAKA

- [1]Adipranata, Rudy.Kombinasi Metode Morphological Gradient dan Transformasi Watershed pada Proses Segmentasi Citra *Digital*. Universitas Kristen Petra. Surabaya
- [2]Adipranata,Rudy.Andreas Handojo. Prayogo, Yuliana Ivan, Oviliani Yenty. Perancangan dan Pembuatan Aplikasi Segmentasi Gambar dengan Menggunakan Morpological Watershed.Universitas Kristen Petra. Surabaya
- [3]Anifah, Lilik. Pengenalan Plat Mobil Indonesia menggunakan Learning Vector Quantization. Institut Teknologi Sepuluh November.
- [4]Budhi, Gregorius s. Metode Jaringan Syaraf Tiruan Backpropagation untuk Pengenalan Huruf Cetak pada Citra *Digital*. Universitas Kristen Petra.
- [5]Fitria, Whielda. (2010). *Desain dan Implementasi Sistem Penerjemah Huruf Jepang Kana (Hiragana-Katakana) Ke Huruf Latin Berbasis Pengolahan Citra Digital dan Jaringan Syaraf Tiruan Self-Organizing Map (SOM) Menggunakan Webcam*. Institut Teknologi Telkom.
- [6]Gonzales, Rafael C., Woods, Richard E., dan Eddins, Steven L.(2004) *Digital Image Processing using MATLAB*, India.
- [7]Gonzales, Rafael C. dan Richard E Woods.(2002)*Digital Image Processing Second Edition*.USA.
- [8]Goswami,Gaurav dkk. Face Recognition CAPTCHA.IIIT-Delhi.India.
- [9]Huang, Shih-Yu; Lee, Yeuan-Kuen; Bell ,Graeme; and Zhan-he Ou (2008). A Projection-based Segmentation Algorithm for Breaking MSN and Yahoo CAPTCHAs. London
- [10]Iqbal, Muhammad(2009). Dasar Pengolahan Citra Menggunakan MATLAB. Institut Pertanian Bogor.
- [11]Liliana; Budhi, Gregorius Setia; Hendra. Segmentasi Plat Nomor Kendaraan Dengan Menggunakan Metode Run Length Smearing Algorithm(RLSA). Universitas Kristen Petra.
- [12]Munir, Rinaldi. Pengolahan Citra *Digital*. Penerbit Informatika.
- [13]Taufik, Yudil. (2009). *Identifikasi Plat Nomor Mobil Menggunakan Support Vector Machine (SVM)*. Institut Teknologi Telkom. Bandung.

- [14]Wei-Bin Lee.Che-Wei Fan. Kevin Ho.dan Chyi-Ren Dow. A CAPTCHA with Tips Related to Alphabets Upper or Lower Case.Feng Chia Univesity.Taiwan.
- [15]Wijaya, Marvin Ch dan Agus Priyono. (2007). *Pengolahan Citra Digital Menggunakan Matlab Image Processing Toolbox*. Informatika. Bandung
- [16]NN. Segmentasi Otak pada Citra Magnetik Resonance Imaging (MRI) dengan Menggunakan Metode Multi Direction Gradient Vector Flow (MDGVF). Institut Teknologi Telkom
- [17]www.captcha.net diakses terakhir pada tanggal 30 April 2013
- [18]www.itelkom.ac.id dengan judul Modul 6 : Analisis Tekstur dan Ekstraksi Ciri diakses terakhir pada tanggal 6 Mei 2013
- [19]<http://napsters91.blogspot.com/2011/11/captcha-ternyata-tidak-sepenuhnya-aman.html> terakhir diakses pada tanggal 15 Januari 2014.



Telkom
University