

Analisis Dan Perancangan Sistem Identifikasi Kata Menggunakan *Shape Analysis Chain Code Dan Thinning* Safriatullah

Program Studi Teknik Informatika Universitas Islam Kebangsaan Indonesia
safriatullah@gmail.com

ABSTRAK

Penelitian ini dilakukan untuk mengidentifikasi kata yang terdapat pada citra digital. Proses identifikasi kata pada citra digital dilakukan dengan mengidentifikasi karakter – karakter yang menyusun kata pada citra digital. Algoritma yang diterapkan menggunakan kombinasi dua metode yang sederhana namun efektif dalam pengenalan karakter yaitu Chain Code dan thinning. Thinning digunakan untuk melakukan normalisasi karakter pada citra, menggunakan thinning maka diperoleh kerangka dari karakter sehingga lebih memudahkan dalam melakukan ekstraksi fitur bentuk dari karakter. Chain Code analysis digunakan untuk mengekstraksi fitur karakter yang terdiri dari jumlah titik ujung, titik sendi dan titik terpisah. Fitur – fitur yang telah diekstraksi kemudian dijadikan referensi dalam mengidentifikasi kata atau kalimat yang terdapat pada citra digital dengan melakukan analisis tiap karakter penyusun atau kata atau kalimat. Kombinasi kedua metode tersebut memberikan hasil yang cukup baik pada kasus identifikasi kata atau kalimat pada citra digital walaupun terdapat beberapa kelemahan seperti perlunya metode tambahan dalam hal segmentasi untuk memberikan hasil yang lebih akurat.

Kata Kunci: *Pengenalan Kata, Pengenalan Karakter, Analisis Bentuk, Thinning, Chain Code, Citra Digital.*

PENDAHULUAN

Penelitian identifikasi objek pada citra digital telah berkembang dari tahun ke tahun. Identifikasi objek merupakan kegiatan dalam mengekstrak informasi pada citra digital sehingga diperoleh informasi objek yang terdapat pada citra. Identifikasi objek terbagi lagi menjadi beberapa bidang mulai dari identifikasi objek secara umum, identifikasi manusia, identifikasi wajah, identifikasi sidik jari, identifikasi karakter atau huruf dan sebagainya.

Pengenalan atau identifikasi kata yang merupakan pengembangan dari pengenalan karakter atau *character recognition* adalah kemampuan komputer menerima dan menafsirkan input rangkaian karakter yang membentuk kata yang dapat dimengerti dari sumber citra digital. Pengenalan kata termasuk salah satu masalah yang menjadi perhatian banyak peneliti saat ini sebagai pengembangan dari pengenalan karakter. Alasan penelitian ini adalah kemampuan dalam mengenali secara efektif rangkaian karakter yang membentuk kata secara otomatis.

Penelitian mengenai identifikasi karakter pada citra digital merupakan penelitian yang telah lama dilakukan dan telah berkembang dari waktu ke waktu. Beberapa penelitian menerapkan jaringan syaraf tiruan dan teknik-teknik ekstraksi fitur seperti morfologi dan teknik-teknik lainnya. Pemilihan teknik yang digunakan pada proses identifikasi karakter pada citra digital sangat mempengaruhi kinerja dan akurasi dari hasil identifikasi. Akurasi yang baik diperoleh dari kombinasi metode dan teknik yang kompleks namun memiliki efek samping terhadap kinerja dan waktu yang dibutuhkan. Beberapa penelitian mencoba memanfaatkan fitur sederhana dari objek untuk meningkatkan kinerja dalam proses identifikasi karakter.

Berdasarkan hal tersebut penulis melakukan penelitian identifikasi rangkaian karakter yang membentuk kata menggunakan teknik *Chain Code* dan *Thinning*. Kedua metode di atas merupakan metode dasar yang banyak diterapkan pada penelitian morfologi pada citra digital. Untuk itu penulis ingin melakukan penelitian dan analisa terhadap kombinasi kedua metode tersebut pada kasus identifikasi kata pada citra digital.

METODE PENELITIAN

Masalah yang diangkat dari skripsi ini adalah pembuatan sistem *optical recognition* yang mampu mengidentifikasi kata pada kalimat. *Optical recognition* pada umumnya membutuhkan berbagai kombinasi metode yang rumit dan sangat kompleks, mulai dari penggabungan metode pengolahan citra, statistik sampai jaringan saraf tiruan. Metode chain code dan thinning merupakan metode sederhana yang pada umumnya digunakan untuk mengekstraksi fitur bentuk atau morfologi dari objek yang terdapat pada citra digital. Penerapan kedua metode tersebut pada *optical recognition* memberikan solusi terhadap kompleksitas yang dibutuhkan dalam mengidentifikasi karakter pada citra digital.

Pada penelitian ini digunakan UML sebagai bahasa pemodelan untuk mendesain dan merancang sistem *Identifikasi kata menggunakan chain code shape analysis dan thinning*. Model UML yang digunakan adalah *use case* dan *activity diagram*. *Activity Diagram* proses pengenalan melakukan ekstraksi fitur chain code dari karakter dan kemudian menyimpan fitur ke dalam database.

Proses identifikasi diawali dengan pengguna melakukan input citra yang akan di-identifikasi yang mana terdapat kata yang akan di-identifikasi. Karakter yang terdapat pada citra, kemudian di-normalisasi menggunakan proses *thinning* yang kemudian dilanjutkan dengan ekstraksi fitur menggunakan *chain code*. Hasil ekstraksi fitur dari tiap karakter kemudian dicocokkan dengan fitur yang terdapat pada database untuk memperoleh identifikasi karakter penyusun kata pada citra yang di-identifikasi.

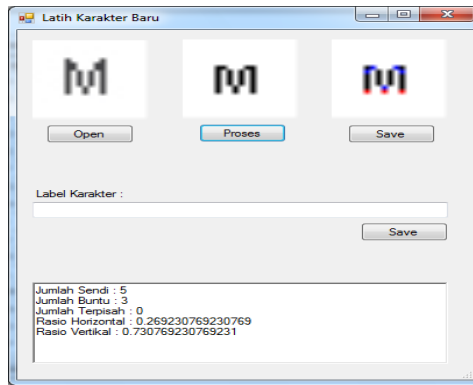
HASIL DAN PEMBAHASAN

Topik utama pada tugas akhir ini adalah merancang sistem yang mampu mengidentifikasi kata menggunakan metode *chain code shape analysis* dan *thinning*. Kata yang akan diidentifikasi pada sistem yang dibangun terdiri dari kumpulan karakter yang akan dikenali terlebih dahulu oleh sistem melalui fitur pelatihan.

Karakter-karakter yang telah dikenali disimpan ke dalam basis data yang dapat disebut dengan sumber pengetahuan bagi sistem sebagai referensi fitur dalam mengidentifikasi karakter-karakter yang terdapat pada kata sehingga kata dapat diidentifikasi. Secara garis besar sistem memiliki dua fungsi utama, yaitu fungsi pembelajaran dan fungsi pengenalan. Masing-masing fungsi dijalankan pada halaman yang terpisah sehingga memudahkan dalam penggunaan dari sistem yang dikembangkan. Berikut penjabaran dari fungsional sistem secara umum.

Fungsi pembelajaran merupakan fungsi yang dimiliki oleh sistem untuk mengenali fitur-fitur yang dimiliki oleh karakter yang diperoleh dari proses *chain code shape analysis*. Fitur-fitur karakter yang diperoleh kemudian disimpan ke dalam basis data untuk digunakan pada fungsi pengenalan. Fungsi pengenalan merupakan fungsi yang dimiliki oleh sistem yang dikembangkan yang digunakan untuk mengidentifikasi kata yang terdapat pada citra digital. Proses identifikasi kata dilakukan dengan mengidentifikasi karakter-karakter yang menyusun kata tersebut sehingga kata dapat diidentifikasi.

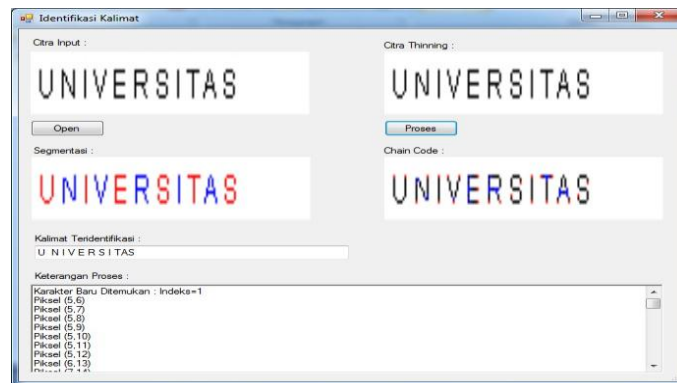
Fungsi pembelajaran merupakan komponen sistem yang melakukan pendataan terhadap fitur-fitur karakter penyusun kata. Fitur-fitur dari karakter di ekstraksi oleh proses *thinning* dan *chain code analysis* dan kemudian menyimpan fitur dari karakter tersebut ke dalam basis data disertai dengan label dari karakter.



Gambar 8. Pembelajaran

Fungsi pengenalan merupakan komponen sistem yang melakukan identifikasi terhadap kata yang terdapat pada citra digital. Proses pengenalan atau identifikasi dilakukan dengan mengekstraksi fitur dari tiap karakter penyusun kata dan membandingkan fitur dari karakter-karakter tersebut dengan

fitur yang ada pada basis data sehingga dapat di-identifikasi kata yang disusun dari karakter-karakter tersebut.



Gambar 9. Identifikasi

Proses identifikasi dilakukan dengan melakukan operasi *thinning* pada setiap karakter yang terdapat pada citra. Operasi dilanjutkan dengan melakukan ekstraksi fitur dari setiap karakter menggunakan *chain code shape analysis*.

Pembahasan

Pada sub bab ini akan dilakukan pengujian pada sistem yang dibangun. Pengujian dilakukan dengan melakukan pengenalan pada beberapa kata yang berbeda-beda. Pengujian dilakukan dengan tujuan untuk memperoleh tingkat akurasi dan kinerja dari sistem identifikasi kata menggunakan metode *chain code shape analysis* dan *thinning*. Jumlah pengujian yang akan dilakukan pada penelitian tugas akhir ini adalah sebanyak lima pengujian dengan menggunakan lima buah citra uji seperti yang dapat dilihat pada tabel 1 berikut.

Tabel 1. Pengujian

NO	CITRA UJI	KATA YANG TERKANDUNG
1	SAYAPERGI	SAYA PERI
2	SAYAMAKAN	SAYA MAKAN
3	TESSAFRI	TES SAFRI
4	SAYAWA	SAYAWA
5	SAFRISAYA	SAFRI SAYA



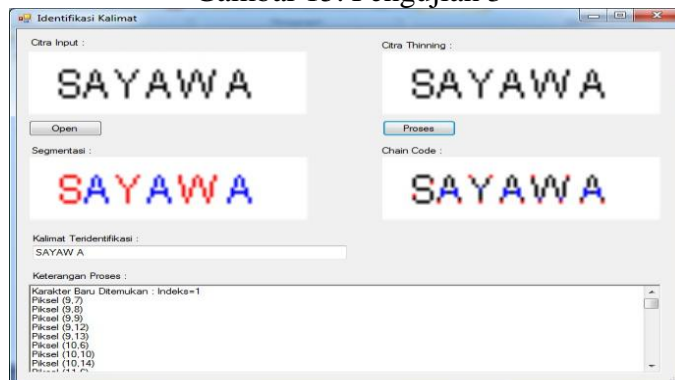
Gambar 12. Pengujian 1



Gambar 13. Pengujian 2



Gambar 15. Pengujian 3



Gambar 15. Pengujian 4



Gambar 15. Pengujian 5

Pengujian yang telah dilakukan pada sub-bab sebelumnya menunjukkan akurasi yang cukup baik pada hasil yang diberikan oleh sistem terhadap citra uji. Selain citra uji 1 semua kata yang terdapat pada citra uji yang diberikan dapat di-identifikasi dengan baik. Rincian hasil pengujian dapat dilihat pada tabel 2 berikut.

Tabel 2. Hasil Pengujian

No	Citra Uji	Kata yang terkandung	Kata Teridentifikasi	Akurasi
1	SAYAPERGI	SAYA PERGI	SPERGI	66.675
2	SAYAMAKAN	SAYA MAKAN	SAYA NAKAN	1005
3	TESSAFRI	TES SAFRI	TES SAFRI	100%
4	SAYAWA	SAYAWA	SAYAWA	100%
5	SAFRISAYA	SAFRI SAYA	SAFRI SAYA	1005

PENUTUP

Simpulan

Berdasarkan hasil analisis, implementasi dan pengujian pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut;

1. *Chain code shape analysis* dan *thinning* dapat diterapkan pada kasus identifikasi kata pada citra digital. Karakter penyusun kata diproses sehingga diperoleh fitur jumlah titik ujung, jumlah titik sendi, jumlah titik terpisah, rasio horizontal dan rasio vertical. Menggunakan fitur tersebut karakter-karakter penyusun kata dapat di-identifikasi sehingga kata tersebut dengan sendirinya teridentifikasi pula.
2. Pengujian yang dilakukan sebanyak lima kali dengan menggunakan lima citra yang berbeda menghasilkan kinerja akurasi sebagai berikut:
3. Pengujian pertama memberikan akurasi sebesar 66.67% di mana kata yang dikenali adalah “SPERGI” dari kata “SAYA PERGI”.
4. Pengujian kedua memberikan akurasi sebesar 100% di mana kata yang dikenali adalah “SAYA MAKAN” dari kata “SAYA MAKAN”.
5. Pengujian ketiga memberikan akurasi sebesar 100% di mana kata yang dikenali adalah “TES SAFRI” dari kata “TES SAFRI”.
6. Pengujian ke-empat memberikan akurasi sebesar 100% di mana kata yang dikenali adalah “SAYAWA” dari kata “SAYAWA”.
7. Pengujian kelima memberikan akurasi sebesar 100% di mana kata yang dikenali adalah “SAFRI SAYA” dari kata “SAFRI SAYA”.
8. Kinerja rata-rata dari sistem yang dikembangkan diukur dari akurasi sistem selama pengujian. Akurasi sistem rata-rata selama pengujian diperoleh sebesar 93.33%.

Saran

Berikut ini adalah saran-saran untuk pengembangan dan perbaikan penelitian ini;

1. Fitur-fitur dari karakter yang di-ekstraksi akan lebih baik jika ditambahkan jenisnya sehingga dapat meningkatkan akurasi sistem dalam mengidentifikasi karakter penyusun kata pada citra digital.
2. Operasi segmentasi yang dilakukan pada penelitian ini masih bersifat segmentasi biasa berdasarkan ruang antar karakter sehingga terkadang menyebabkan kesalahan identifikasi. Untuk penelitian yang akan datang penulis menyarankan adanya metoda tambahan dalam proses segmentasi karakter pada saat identifikasi.
3. Dalam identifikasi kata pada penelitian ini tidak mendukung multi-line sehingga untuk kedepannya penelitian disarankan mendukung keadaan multi-line sehingga dapat mengidentifikasi kalimat sampai paragraf.

DAFTAR PUSTAKA

- Sutoyo, T, dkk. 2009. *Teori Pengolahan Citra Digital*. Yogyakarta: Andi Offset, Page 9.
- Putra, Darma. 2010. *Pengolahan Citra Digital*. Yogyakarta: Andi Offset, Page 228.
- Chalechale, A., Naghdy, G., & Pramaratne, P. 2013. Image Analysis Using Line Segments Extraction by Chain Code Differentiation. *Wollongong: School of Electrical, Computer & Telecommunications Engineering*, Page 347.
- Cheriet, M. 2007. *Character Recognition System*. Canada: John Willey & Son, Page 70.
- Plamondon, R. 2000. On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Page 63.
- Siddiqi, I., & Vincent, N. 2009. A Set of Chain Code Based Features for Writer Recognition. *International Conference on Document Analysis and Recognition*, Page 981.
- Singh, G. 2012. Recognition of Handwritten Hindi Characters using Backpropagation Neural Network. *International Journal of Computer Science and Information Technologies*, Page 4892.
- Tarabek, P. 2007. Morphology Image Pre-Processing for Thinning Algorithms. *Journal of Information, Control and Management Systems*, Page 131.