

SEGMENTASI KARAKTER MENGGUNAKAN PROFIL PROYEKSI

Anindita Septiarini

Program Studi Ilmu Komputer FMIPA, Universitas Mulawarman

Email: aninditaseptiarini@yahoo.com

Abstrak

Karakter merupakan bagian dari suatu teks, dimana keberadaan teks dapat kita jumpai diberbagai tempat. Karakter dapat kita jumpai pada suatu dokumen, media cetak maupun pada gambar. Untuk pengambilan teks atau karakter pada suatu gambar diperlukan adanya proses segmentasi. Proses segmentasi karakter bertujuan untuk mendapatkan area karakter tunggal yang nantinya dapat digunakan pada untuk sistem pengenalan karakter. Hasil dari keseluruhan proses segmentasi ini bertujuan untuk mendapatkan kumpulan citra karakter tunggal.

Pada penelitian ini proses segmentasi karakter dilakukan dengan menggunakan metode binerisasi dan profil proyeksi. Metode binerisasi yang digunakan berupa binerisasi manual, dimana user menentukan sendiri nilai ambangnya, sesuai dengan warna *background* maupun karakter yang ada pada suatu citra.

Untuk tahap pengujian pada penelitian ini dilakukan dengan menggunakan beberapa file citra dengan format *.jpeg*, dimana beberapa citra memiliki warna *background* dan karakter yang berbeda. Perbedaan tersebut bertujuan untuk mengetahui ketahanan dari metode yang diusulkan. Dari hasil pengujian menunjukkan bahwa metode yang diusulkan untuk proses segmentasi karakter dapat menghasilkan citra karakter tunggal yang cukup baik dengan akurasi hingga 90%.

Kata Kunci : Segmentasi, Binerisasi, Profil Proyeksi

PENDAHULUAN

Pada saat ini pembacaan suatu karakter yang terdapat pada citra yang mengandung teks dapat dilakukan oleh komputer. Untuk membuat suatu sistem pembacaan karakter pada citra yang mengandung teks, maka terlebih dahulu harus melakukan proses segmentasi area karakter. Hal tersebut dilakukan karena terkadang karakter telah bercampur dengan objek-objek lainnya, sehingga objek-objek lain tersebut dapat dianggap sebagai *background* dari citra tersebut.

Pada saat area karakter secara keseluruhan dapat diketahui keberadaanya maka selanjutnya dilakukan proses segmentasi terhadap area karakter tersebut sehingga menghasilkan citra karakter tunggal. Hasil dari proses segmentasi karakter menjadi citra karakter tunggal tersebut nantinya dapat digunakan pada suatu sistem pengenalan karakter.

METODE PENELITIAN

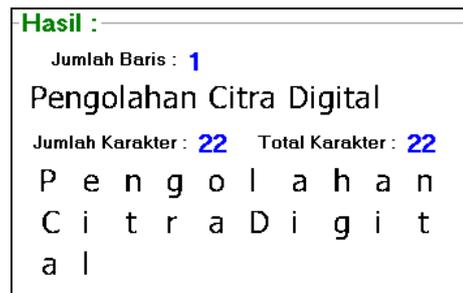
Proses segmentasi pada sistem ini dibedakan menjadi dua tahap. Tahap pertama yaitu proses segmentasi baris yang digunakan untuk mendapatkan area karakter setiap barisnya. Proses segmentasi baris dibutuhkan karena terkadang pada suatu citra karakter tersusun pada lebih dari satu baris. Tahap kedua adalah proses segmentasi

karakter, dimana pada proses ini bertujuan untuk mendapatkan area karakter tunggal.

Pada kedua tahap proses segmentasi tersebut dilakukan dengan menggunakan metode binerisasi dan profil proyeksi. Sistem ini dilakukan pada citra teks, adapun contoh dari citra teks dapat dilihat pada gambar 1 dan hasil yang diharapkan dari sistem ini adalah berupa kumpulan citra karakter tunggal yang dapat dilihat pada gambar 2.



Gambar 1. Data Masukan Untuk Proses Segmentasi



Gambar 2. Hasil Keluaran dari Proses Segmentasi

Segmentasi

Segmentasi merupakan suatu proses pemisahan objek yang satu dengan objek yang lain dalam suatu citra, berdasarkan sifat-sifat tertentu dari citra yang dapat dijadikan sebagai pembeda [2]. Pada penelitian ini objek yang dimaksud adalah objek karakter pada citra teks. Proses-proses yang dilakukan pada penelitian sistem segmentasi karakter ini secara keseluruhan terdiri dari lima proses.

Proses (1) *Load Citra* menerima masukkan berupa file citra yang dipilih oleh user. Citra teks tersebut akan menjadi masukkan untuk proses (2) *Binerisasi*, proses ini akan menghasilkan citra biner. Hasil dari proses (2) *Binerisasi* ini akan menjadi masukkan untuk proses selanjutnya, yaitu proses (3) *Segmentasi Baris*. Proses (3) *Segmentasi Baris* bertujuan untuk mengetahui jumlah baris teks dan menghasilkan citra per barisnya.

Setelah area citra baris dihasilkan maka sistem akan melakukan proses (4) *Segmentasi Karakter*, dimana pada proses ini akan dihasilkan area citra tiap karakternya dan diketahui jumlah karakter tiap baris dan jumlah total karakter pada citra teks tersebut.

Proses (5) *Tampil Data* merupakan proses untuk menampilkan data-data yang dapat dilihat oleh user, yaitu data citra teks, citra biner, citra baris, jumlah baris, jumlah karakter tiap baris dan jumlah total karakter.

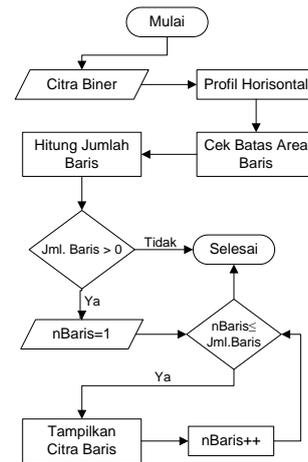
Proses Binerisasi

Binerisasi merupakan proses untuk mengubah suatu citra menjadi citra yang hanya terdiri dari warna hitam dan putih [1]. Pada proses ini membutuhkan masukkan dari user berupa nilai ambang untuk membedakan dan memisahkan area objek dan *background*. Pada penelitian ini suatu piksel akan menjadi area objek jika nilai intensitas suatu piksel kurang dari nilai ambang yang telah ditentukan oleh user. Jika piksel tertentu memiliki nilai intensitas lebih dari nilai ambang, maka piksel tersebut akan menjadi bagian dari area *background*.

Proses Segmentasi Baris

Segmentasi baris merupakan proses segmentasi yang dilakukan dengan tujuan untuk menghitung jumlah baris karakter pada citra dan mendapatkan area baris-baris karakter [4]. Hal ini dimaksudkan agar bagian-bagian yang tidak diperlukan untuk proses berikutnya tidak diikuti sertakan pada proses selanjutnya.

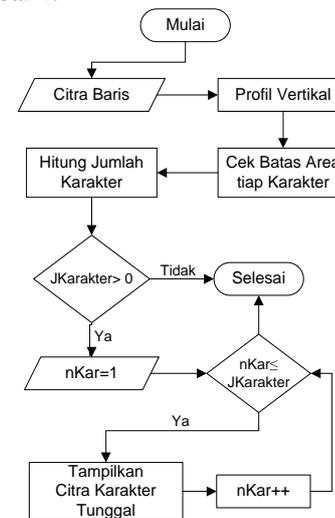
Segmentasi baris diperlukan karena proses segmentasi karakter dilakukan perbaris. Adapun tahapan yang dilakukan untuk menghitung dan mendapatkan area baris karakter dapat dilihat pada gambar 3.



Gambar 3. Diagram Alir Tahapan Proses Segmentasi Baris

Proses Segmentasi Karakter

Segmentasi karakter merupakan proses segmentasi yang tujuan untuk memisahkan setiap karakter dalam sebuah baris [4]. Hasil dari proses segmentasi karakter berupa kumpulan citra karakter tunggal, jumlah karakter tiap baris dan jumlah total keseluruhan karakter. Adapun tahapan untuk menghasilkan citra karakter tunggal dapat dilihat pada gambar 4.



Gambar 4. Diagram Alir Tahapan Proses Segmentasi Karakter

Profil Proyeksi

Apabila terdapat sebuah citra teks biner $S(N,M)$, dimana N menyatakan banyaknya baris citra dan M adalah banyaknya kolom citra seperti ditunjukkan pada gambar 4, maka dapat ditentukan profil vertikal dan profil horisontal dari citra teks tersebut. Profil vertikal adalah banyaknya piksel hitam yang tegak lurus dengan sumbu y , sedangkan profil horisontal adalah banyaknya piksel hitam yang tegak lurus sumbu x [4].



Gambar 5. Profil Proyeksi Horizontal dan Vertikal

Profil vertikal direpresentasikan dengan suatu vektor P_v berukuran N . Profil vertikal pada baris ke- i , yaitu $P_v[i]$, didefinisikan sebagai berikut :

$$P_v[i] = \sum_{j=1}^M S[i, j]$$

Sedangkan profil horizontal direpresentasikan dengan suatu vektor P_h berukuran M . Profil horizontal pada kolom ke- j , yaitu $P_h[j]$, didefinisikan sebagai berikut :

$$P_h[j] = \sum_{i=1}^N S[i, j]$$

HASIL DAN PEMBAHASAN

Pada penelitian ini ada dua data masukan yang dibutuhkan yaitu file citra teks dan nilai ambang. File citra teks merupakan data citra yang akan diolah pada sistem ini, sedangkan nilai ambang digunakan untuk proses binerisasi untuk membedakan antara area objek dan area *background*.

Pada penelitian ini data masukan yang digunakan sebagai pengujian berupa data citra teks dengan warna teks dan warna *background* yang berbeda. Adapun data citra yang digunakan sebagai data uji dan hasil pengujian dapat dilihat pada tabel 2 berikut ini.

Tabel 2. Contoh Data Citra Teks

Data Citra	Ambang	Kebenaran (%)
Pengolahan Citra Digital	30	100%
Pengolahan Citra Digital Digital Image Processing	30	100%
Segmentasi karakter dengan Metode Profil Proyeksi	30	100%

Pengolahan Citra	90	100%
Segmentasi Citra Digital	200	100%
Metode Profil Proyeksi	150	100%
Data Uji Coba	130	100%
Pengolahan Citra Digital	50	100%

Pada proses binerisasi ketepatan pemilihan nilai ambang sangat berpengaruh terhadap citra biner yang dihasilkan serta berpengaruh pada proses selanjutnya, yaitu proses segmentasi baris dan segmentasi karakter. Jika pemilihan nilai ambang kurang atau tidak tepat kemungkinan yang dapat terjadi yaitu :

1. Jumlah baris yang terdeteksi tidak sesuai dengan yang ada pada citra teks misal pada citra teks terdapat dua baris teks, tetapi yang terdeteksi oleh sistem hanya ada satu baris atau lebih dari dua baris.
2. Pada proses segmentasi karakter, karakter yang seharusnya terhitung ada dua karakter hanya terhitung satu karakter, hal tersebut dikarenakan karakter yang satu dengan yang lain tersambung.
3. Pada proses segmentasi karakter, karakter yang seharusnya terhitung ada satu karakter terhitung menjadi dua atau lebih karakter, hal tersebut dikarenakan ada bagian dari karakter tersebut yang terputus.

Dari keseluruhan hasil uji coba pada tabel 1 dapat dilihat bahwa dari sepuluh data citra yang diujikan menghasilkan persentase kebenaran mencapai 90%. Hasil tersebut dipengaruhi oleh ketepatan pemilihan nilai ambang. Jika pemilihan nilai ambang tidak tepat maka akan terjadi kesalahan hasil segmentasi seperti yang telah dijelaskan pada penjelasan sebelumnya.

KESIMPULAN

Dari hasil penelitian dan pembahasan yang telah dilakukan menunjukkan bahwa proses

binerisasi baik digunakan untuk segmentasi, dimana keberhasilannya mencapai hingga 90%. keberhasilan tersebut dipengaruhi oleh faktor ketepatan pemilihan nilai ambang, karena jika citra teks mengalami perubahan warna dan *background* maka nilai ambang yang tepat juga kemungkinan akan berubah.

Untuk penelitian lebih lanjut dapat digunakan metode segmentasi yang lain, atau bisa juga menggunakan metode binerisasi tetapi proses binerisasinya dibuat secara otomatis.

DAFTAR PUSTAKA

- [1] Ahmad. U, 2005, *Pengolahan Citra Digital dan Teknik Pemrogramannya*, Graha Ilmu.
- [2] Gonzalez. R. C. dan Woods. R. E., 2001, *Digital Image Processing*, Addison-Wesley Publishing.
- [3] Munir. R., 2004, *Pengolahan Citra Digital dengan Pendekatan Algoritmik*, Informatika Bandung.
- [4] Widiarti. A. R., 2006, *Pengenalan Citra Dokumen Sastra Jawa Konsep dan Implementasinya*, Tesis-S2 Pascasarjana Universitas Gadjah Mada, Yogyakarta