

Handwritten Image Segmentation Carakan Madura Based Projection And Connected Component Labeling

Segmentasi Citra Tulisan Tangan Carakan Madura Berbasis Proyeksi Dan Connected Component Labeling

<i>Miftah Farid</i>	Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya
<i>Joan Santoso</i>	Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya
<i>Endang Setyati</i>	Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya

At the time the Carakan Madura is currently found on sign of street names in the Madura region. It is also found in the historical sites and museum. However, many people do not understand about Carakan Madura because it is rarely used in daily life. There are many ways to keep Carakan Madura sustainable and not extinct, including preservation and maintenance of historical objects and also from education. The maintenance of these historical objects is usually carried out in museums and places where historical objects, including the Sumenep palace museum and Asta Tinggi. In the Sumenep palace museum, documents with Carakan Madura are still stored. In the Asta tinggi there are also stone carvings with the words Carakan Madura. There are other important things in the preservation of Carakan Madura, namely that Carakan Madura is included in the local content education curriculum for Madura language so that students can learn about Carakan Madura properly and well so students are not only knowing, but also understanding. For this reason, tools are needed so that students can more understand the Carakan Madura. The purpose of this research is to segment of the Carakan Madura handwritten image based on projection and connected component labeling. The results of this research can be concluded that the segmentation process in Carakan Madura handwriting has been successfully carried out in the form of the composing characters of Carakan Madura.

Pendahuluan

Carakan Madura merupakan identitas penting bagi masyarakat Madura. Hal ini harus dilestarikan agar tidak punah. Carakan Madura terdiri dari aksara ghajang, pangangguy sowara, pangangguy panyeghek, pangangguy panamba, pasangan dhampeng dan ghantongan, pacerrekk dan ngalellet, rengkessan, aksara raja, pasangan aksara raja, aksara rekaan¹. Bukan hal yang mudah dalam mempelajari Carakan Madura. Untuk itu di era digital saat ini diperlukan sarana agar orang dapat belajar Carakan Madura dengan mudah. Untuk itu pada bidang citra digital dibutuhkan handwriting recognition untuk mempermudah peserta didik dalam kegiatan belajar Carakan Madura. dengan cara melakukan pengenalan terhadap tulisan tangan Carakan Madura. Dalam proses pengenalan tulisan tersebut di butuhkan proses segmentasi atau pemisahan karakter pada citra tersebut. Berdasarkan permasalahan tersebut perlu dilakukan penelitian tentang segmentasi pada citra tulisan tangan Carakan Madura.

Carakan Madura dan aksara daerah lainnya hampir sama juga terdiri dari aksara dasar (aksara ghajang) atau konsonan dan aksara pasangan. Penelitian yang berkaitan dengan segmentasi tulisan tangan aksara daerah juga pernah dilakukan sebelumnya. Penelitian sebelumnya meneliti tentang segmentasi pada aksara dasar saja yang hanya berada pada satu baris yang sejajar dan berjarak. Penulis mengembangkan pada Carakan Madura namun bukan pada aksara dasar saja tapi juga aksara-aksara pasangan lainnya. Aksara-aksara pasangan ini bisa berada diatas aksara dasar,

dibawah aksara dasar dan disamping aksara dasar. Aksara dasar dan aksara pasangan yang saling berkaitan ini harus dapat dijadikan dalam satu segmen karena aksara dasar dan aksara pasangan ini jika digabungkan akan menimbulkan makna yang berbeda.

Tujuan dari penelitian ini untuk menyegmentasi Carakan Madura yang terdiri dari aksara dasar dan aksara pasangan. Segmentasi aksara dasar dan aksara pasangan lainnya tersebut menggunakan dua metode yaitu metode proyeksi dan *connected component labeling*. Metode proyeksi digunakan untuk menyegmentasi setiap aksara satu persatu. Metode proyeksi terdiri dari proyeksi vertikal dan proyeksi horisontal. Proyeksi vertikal digunakan untuk memisahkan karakter di sampingnya. Sedangkan proyeksi horisontal digunakan untuk memisahkan bagian atas dan bawah. Metode *connected component labeling* digunakan untuk menghubungkan karakter yang terdekat baik yang ada di atas, di bawah atau disamping aksara dasar sehingga makna dari Carakan Madura menjadi tepat.

Metode Penelitian

Data Citra

Penelitian ini menggunakan dataset tulisan siswa SMKN 1 Sumenep sebanyak 6 siswa. Setiap siswa menuliskan 25 kata Carakan Madura dengan total 150 kata. Setiap siswa menuliskan Carakan Madura pada selembar kertas putih bersih dengan menggunakan ballpoint warna hitam. Dalam pengambilan citra dilakukan dengan pemotretan pada selembar kertas tersebut. Kemudian dari hasil citra tersebut dipotong perkata menggunakan aplikasi pengolah gambar sehingga didapatkan potongan perkata Carakan Madura. Gambar potongan perkata tersebut yang digunakan sebagai input gambar pada penelitian ini. Adapun hasil tulisan tangan siswa pada penelitian ini bermacam-macam baik dari kemiringan dan jarak antar karakter.

Figure 1. *CitraTulisan Tangan Carakan Madura*

Rancangan Sistem

Figure 2. *RancanganSistem*

Input Citra Tulisan Tangan

Data yang diinputkan berupa citra tulisan tangan Carakan Madura yang terdiri dari aksara ghajang, pangangguy sowara, pangangguy panyegheh, pangangguy panamba, pasangan dhampeng dan ghangtongan yang di tulis oleh siswa. Jenis aksara di Carakan Madura tersebut sangat bervariasi posisi antar hurufnya.

Figure 3. *Contoh Citra Tulisan Tangan yang Di-input-kan*

Gray Scale

Citra input diubah menjadi *grayscale* untuk menghasilkan efek warna abu-abu. Pada tahap ini warna dinyatakan dengan intensitas berkisar antara 0 sampai dengan 255. Nilai 0 menyatakan hitam dan nilai 255 menyatakan warna putih².

Contrast Stretching

Kontras dari suatu citra dapat dipengaruhi oleh pencahayaan, bidang dinamika dari citra. *Contrast stretching* dapat diterapkan untuk mendapatkan citra baru dengan kontras yang lebih baik dari asalnya, yang artinya proses ini hanya tergantung nilai intensitas (*gray level*) satu piksel dan tidak tergantung pada piksel lain di sekitarnya. Proses dari *contrast stretching* dilakukan dengan meningkatkan bidang dinamika *gray level* pada citra yang akan diproses. Dengan kontras yang baik pada citra, dapat meningkatkan ketajaman warna obyek pada citra. Obyek yang terlihat jelas pada citra dapat membantu proses segmentasi citra³.

Citra Biner

Citra biner hanya mempunyai dua nilai derajat keabuan yaitu hitam yang bernilai 0 dan putih yang bernilai 1. Citra biner banyak dimanfaatkan untuk keperluan pengenalan angka, huruf maupun tanda tangan. Piksel-piksel objek bernilai 1 dan piksel-piksel latar belakang bernilai 0.

Resize

Resize ini terdiri dari citra dan faktor penskalaan. Kemudian citra diskalakan. Faktor penskalaan yang dipakai adalah 0.5, maka setiap dua piksel pada gambar asli dipetakan ke satu nilai piksel pada gambar keluaran untuk kedua dimensinya.

Imcomplemen

Imcomplement atau citra negatif merupakan citra yang berkebalikan dengan citra asli. Transformasi citra negatif ini digunakan ketika terdapat bagian tertentu yang perlu di-*enhance* menjadi putih atau detail abu-abu yang menempel pada warna hitam, khususnya ketika daerah gelap menjadi ukuran yang sangat dominan. Selanjutnya dilakukan penyusunan kembali elemen citra dengan menaikkan kontras untuk mendapatkan citra dengan tingkat kecerahan lebih menajamkan objek dan mereduksi *background*⁴.

Thinning

Karakter tebal yang berupa piksel gelap diubah menjadi karakter tipis dengan menggunakan piksel terdekat untuk pada kisaran tertentu⁵. *Thinning* merupakan operasi morfologi.

Skew Correction

Pada tulisan tangan siswa yang menggunakan kertas putih polos terkadang menyebabkan kemiringan pada tulisan. Untuk itu *skew correction* perlu dilakukan untuk mengurangi kemiringan pada citra tersebut sehingga berpengaruh pada saat proses segmentasi. Deteksi kemiringan adalah langkah yang menentukan prosedur pengenalan⁶.

Segmentasi

Segmentasi teks tulisan tangan merupakan fase penting dalam analisis dokumen, pengenalan karakter dan banyak bidang lainnya⁷. Segmentasi merupakan cara pemisahan karakter sebuah kata dan inilah bagian yang paling penting. Pada proses segmentasi dilakukan 2 tahap yaitu proyeksi dan *connected component labeling*.

a. Proyeksi

Proyeksi digunakan untuk menghasilkan batas-batas karakter. Proyeksi citra merupakan bentuk satu dimensi citra yang dihitung berdasarkan sumbu koordinat.

Persamaan:

$$P_{hor}(b) = \sum_{j=1}^N | (b_j, k) |$$

$$P_{ver}(k) = \sum_{j=1}^M | (b_k, k) |$$

Keterangan:

Dalam hal ini M adalah tinggi citra dan N adalah lebar citra. Definisi diatas menyatakan Proyeksi horizontal $P_{ver}(b)$ adalah jumlah nilai piksel pada baris b citra. $P_{ver}(K)$ adalah jumlah nilai piksel pada kolom k citra.

Pseudocode:

Masukan: F = Citra

Keluaran: ProyHor = Proyeksi horizontal

ProyVer = Proyeksi vertikal

[m,n] = size (F)

ProyHor = zeros (m,1);

ProyVer = zeros (n,1);

for y=1 : m

for x=1 : n

ProyHor (y) = ProyHor (y)+F(y,x);

ProyVer (x) = ProyVer (x)+F(y,x)

End

end

b. Connected Component labeling

Connected Component Labeling (CCL) adalah metode untuk memberikan label unik ke setiap objek dalam sebuah gambar. Dengan memberi label pada setiap objek, citra diubah menjadi gambar simbolis dimana semua piksel termasuk setiap komponen yang terhubung diberi label unik.

Persamaan:

B(x,y)=0 Piksel latar belakang1 Piksel latar depan2,3 Label objek

Keterangan:

Metode pelabelan dengan menggunakan *Connected Component Labeling* dapat dilakukan melalui langkah-langkah cek nilai pertama pada citra yang di-input-kan pada baris 1 kolom 1. Apabila nilai tersebut 0 (background), lewati dan pindah ke kolom selanjutnya pada baris yang sama. Apabila piksel bernilai 1, cek nilai piksel tetangganya. Apabila tetangganya belum terlabel, beri label pada piksel yang ditinjau dengan label 1. Jika memiliki satu tetangga terdekat yang sudah terlabel, beri label yang sama dengan dengan tetangga. Jika memiliki lebih dari satu tetangga terdekat dengan

label berbeda, beri label menggunakan label terkecil dari tetangga-tetangga tersebut. Setiap selesai melakukan pelabelan dan bertemu kembali dengan *background*. Perbarui indeks label. Lakukan proses ini sampai keseluruhan baris dan kolom di cek.

Pseudocode:

Masukan:

F (M,N): Citra Masukan berukuran M baris dan N kolom

Keluaran: G (M.N): Hasil citra yang telah diberi label

g <- f

Label <- 2

FOR baris <- 1 TO M

 FOR kolom <- 1 TO N

 IF g(naris,kolom) = 1

 banjiri(g, baris, kolom, label)

 label <- +1

 END-IF

END-FOR

END-FOR

RETURN g

Hasil dan Pembahasan

Dataset yang digunakan sebanyak 6 siswa. Setiap siswa menulis sebanyak 25 kata Carakan Madura jadi total ada 150 kata.

Figure 4. *Citra Hasil Proses Dari Grayscale Sampai Segmentasi*

Figure 5. *Hasil Segmentasi*

Jadi pada penelitian ini 1 kata berisi beberapa segmen (kotak merah). ada 1 segmen yg berisi beberapa karakter karena karaktertersebut berpasangan berada diatas atau dibawah karakter lainnya. Contoh pada Gambar 5 terdiri dari 3 segmen (kotak merah). segmen pertama berisi 1

karakter, segmen kedua berisi 1 karakter, segmen ketiga berisi dua karakter.

Figure 6. Segmen-segmen dalam 1 kata

Siswa	Kesesuaian Segmentasi per-kata	
	Sesuai	Tidak Sesuai
Ke-1	19	6
Ke-2	22	3
Ke-3	11	14
Ke-4	23	2
Ke-5	24	1
Ke-6	20	5
Total	119	31

Table 1. Hasil Kesesuaian Segmentasi

Total jumlah kata yang ditulis oleh siswa ada 150 kata. Yang sesuai segmentasinya ada 119 kata. Persentase hasil kesesuaian segmentasi:

$$(\text{hasil yg sesuai/jumlah kata}) \times 100 = \text{Hasil}$$

$$(119/150) \times 100 = 79\%$$

Dari hasil kesesuaian segmentasi yang ditulis oleh siswa dapat diketahui bahwa hasil 79%.

Kesimpulan

Pada paper ini telah dibahas penerapan metode proyeksi dan connected component labeling pada segmentasi citra tulisan tangan Carakan Madura. Dari Hasil pengujian dapat diterapkan metode proyeksi dan connected component labeling untuk menyegmentasi tulisan tangan Carakan Madura menggunakan matlab. Dapat dilihat penelitian ini dapat menyegmentasi antara aksara dasar dengan aksara dasar dan aksara dasar dapat bergabung dalam satu segmen dengan aksara pasangannya. Hal tersebut dapat dilihat pada enam tulisan tangan siswa. lima siswa diantaranya yaitu siswa ke-1, ke-2, ke-4, ke-5 ke-6 memiliki kesesuaian segmentasi yang lebih tinggi. Cara penulisan Carakan Madura yang baik juga berpengaruh pada hasil segmentasi karena jarak per karakter aksara dasar dengan aksara dasar dan aksara dasar dengan aksara pasangannya juga sangat menentukan hasil segmentasi.

References

1. Bastari Y. Ika, Fiandarti 2009.

2. Kadir A, Susanto A, Teori dan Aplikasi Pengolahan Citra. ANDI: Yogyakarta; 2013.
3. Dwi Putra Asana I Made, Oka Widyantara I Made, Wirastuti N.M.A.E.D, Adnyana Ida Bagus Putu, Metode Contrast Stretching untuk Perbaikan Kualitas Citra pada Proses Segmentasi Video. *Majalah Ilmiah Teknologi Elektro*. 2017; 16(2):1-1.
4. Suleman S, Identifikasi Jumlah Bibit Bandeng Menggunakan Metode K-Means Berbasis HSV Color dan Morfologi. *J. Teknol. Inf. Indones*. 2019; 2(1):27-32.
5. Vellingiriraj E K, Balamurugan M, Balasubramanie P, Information extraction and text mining of Ancient Vattezhuthu characters in historical documents using image zoning. *Proc. 2016 Int. Conf. Asian Lang. Process. IALP 2016*. 2017;37-40.
6. Francis S, *Neural Network And Hidden Markov Model*. 2017;595-598.
7. Dave N, Segmentation Methods for Hand Written Character Recognition. *Int. J. Signal Process. Image Process. Pattern Recognit*. 2015; 8(4):155-164.
8. Firdaus F I, Khumaini A, Utaminingrum F, Arabic letter segmentation using modified connected component labeling. *Proc. - 2017 Int. Conf. Sustain. Inf. Eng. Technol. SIET 2017*. 2018; 2018:392-397.