PENGENALAN KARAKTER PADA CITRA DIGITAL

MENGGUNAKAN MODEL JARINGAN SYARAF TIRUAN

TjutAwaliyah1, Andi Chairunnas1

1Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pakuan, Bogor

tjut.awaliyah@gmail.com

**ABSTRAKS.** Berbagai sumber informasi tulisan masih banyak yang belum tersedia dalam bentuk elektronik, diperlukan proses pemasukan data dari *hardcopy* ke *softcopy* untuk kebutuhan pengolahan lebih lanjut ke komputer. Untuk itu akan dikembangkan suatu sistem pengenalan karakter secara otomatis yang dilakukan pada citra digital berupa image dari hasil *scan*.Secara umum sistem pengenalan karakter ini akan dibagi atas 3 modul meliputi : modul image preprocessing, modul pelatihan (*training*) dan pengenalan (*recognize*). Pada tahap *image processing* meliputi *image preprocessing*, proses *grayscale*, *thresholding*, segmentasi, normalisasi karakter dan ekstrasi ciri. Pada modul pelatihan dan pengenalan menggunakan algoritma jaringan saraf tiruan *Hopfield*. Pengujian yang akan dilakukan meliputi : huruf A-Z, a-z, angka 0-9 dan tanda baca.Aplikasi ini telah melalui tahap proses validasi dan verifikasi. Uji coba pelatihan yang dilakukan denganmenggunakan *image* yang berurutan dengan99 karakter. Adapun persentase akurasi hasil pengenalan tiap karakter untuk ukuran yang sama dengan yang dilatihkan sampai 88,90% dan rata-rata akurasi hasil pengenalan untuk 4 ukuran karakter dan dengan *font* yang berbeda dengan yang dilatihkan

adalah 76,01%.

1. **PENDAHULUAN**

Bentuk elektronik pada saat ini merupakan sumber informasi yang belum banyak tersedia. Ini menyebabkan kendala dalam pengaksesan, pengelolahan dan pendayagunaan data tersebut. Sehingga dilakukan *entry manual* yang memerlukan waktu yang tidak sedikit, oleh karena itu akan dikembangkan suatu sistem pengenalan karakter secara otomatis pada citra digital. Proses *entry*  data dari *hardcopy* kedalam *softcopy*  menggunakan input berupa file *image*  dari hasil scan dan menghasilkan file dalam bentuk teks.

Pengenalan karakter merupakan salah satu bidang dalam ilmu komputer yang dapat membantu proses pengolahan data. Salah satu teknik pengenalan karakter adalah metode jaringan syaraf tiruan[1]. Metode ini menggunakan prinsip kerja otak manusia yang terdiri dari neuron sebagai pemrosesan input untuk menghasilkan output berdasarkan bobot yang ada.

Penelitian ini akan mengembangkan pengenalan karakter dengan inputan banyak karakter dalam satu *image* dengan menggunakan *software* MATLAB.Keunggulan dari penelitian ini adalah kemampuan pendeteksi serangkaian karakter meliputi huruf A-Z, a-z, angka 0-9 dan tanda baca.

1. **Perancangan Sistem**

Rangkaian proses yang dilakukan sistem. Rangkaian proses penelitian ini adalah baca gambar, *grayscale, thresholding,* Segmentasi karakter, normalisasi, ekstraksi ciridan *neural network.* Proses *neural network* sendiri membutuhkan proses pelatihan *(training)* agar *output* yang dihasilkan benar dan akurat. Jadi secara garis besar dapat dikatakan bahwa alur aplikasi ini dimulai dari diinputkannya sebuah gambar dari sebuah gambar teks*,* dimana gambar tersebut akan melalui serangkaian manipulasi citra dalam proses *image preprocessing* diantaranya proses *grayscaling, thresholding,* segmentasi karakter, normalisasi dan ekstrasi ciri yang akan mengubah gambar tersebut menjadi serangkaian nilai-nilai yang berguna untuk proses utama berikutnya yaitu : proses pelatihan *(training process)* ataupun proses pengenalan *(recognition process)*[1]*.*

**2.1. Modul *Image Processing***

Modul ini merupakan proses yang pertama kali dilakukan sebelum pelatihan dan proses pengenalan. Pertama, gambar yang dibuka akan dibaca kemudian setiap karakter akan dicari posisi dan ukurannya (segmentasi). Sebelum gambar disegmentasi terdapat proses *grayscale dan thresholding,* hingga menjadi gambar hitam-putih untuk memudahkan dalam pengelolahan data selanjutnya.

Proses selanjutnya adalah mencari baris demi baris yang ada pada gambar dan mensegmentasikan tiap kata dan tiap karakter yang ada pada gambar *input*. Proses pengambilan tiap-tiap karakter yang ada pada gambar ini sangatlah penting karena tanpa akurasi *input* yang baik, maka proses pengenalan (*recognition)* tidak akan berfungsi dengan baik. Proses ini ditunjukan oleh *flowchart* pada gambar 2.



**Gambar 2.** Flowchart Modul *Image Processing*

**2.2. *Grayscale dan Thresholding***

 Citra *Grayscale* merupakan citra digital yang hanya memiliki satu nilai pada kanal pada setiap pixelnya, dengan kata lain nila RED=GREEN=BLUE[4]. Nilai tersebut digunakan untuk menunjukan tingkat intensitas. Warna yang dimiliki adalah warna hitam, keabuan dan putih. Proses ini akan menghasilkan gambar ke abu-abuan.

**2.3. Segmentasi**

Segmentasi dilakukan dengan memetakan jumlah titik hitam setiap baris pada gambar pada sumbu Y dan setiap karakter hasil pemetaan tersebut dipetakan lagi ke sumbu X. *Flowchart* untuk segmentasi baris dan kolom pada gambar 3.



**Gambar 3.***Flowchart* Segmentasi Baris dan Kolom [2]

**2.4. Normalisasi**

Pemisahan huruf gabung yang telah dipisahkan maka kita akan melakukan proses normalisasi atau biasa disebut dengan resizing. Proses ini dilakukan agar semua huruf maupun angka akan mempunyai ukuran pixel yang sama, yaitu 10x10.

**2.5. Ekstraksi Ciri**

Ekstraksi ciri berfungsi mengubah suatu pola menjadi bit-bit digital sehingga dapat dimengerti oleh komputer.

Citra karakter dibagi menjadi matriks berukuran m x n (m baris dan n kolom). Pada contoh ini yang digunakan adalah matriks 10 x 10. Pada setiap kotak (elemen), dilakukan penelusuran. Jika ditemukan nilai 0 (pixel hitam) dalam jumlah tertentu, maka diberikan nilai 0 (warna hitam) pada kotak tersebut. Namun jika demikian, makan nilai 1 (warna putih) untuk kotak tersebut.

**2.6. Pelatihan Jaringan Syaraf Tiruan**

Jaringan yang dilatih berdasarkan vektor karakter yang akan diinputkan yang telah mengalami proses dari *image processing* sampai ekstraksi ciri. Output yang dihasilkan adalah file data informasi jaringan syaraf tiruan yang berisi bobot koneksi antar *neuron* dari pola vektor yang terbentuk dan vektor pola yang dipanggil secara stabil (konvergen)[2]. *Flowchart* pelatihan jaringan syaraf tiruan dilihat pada gambar 4. Proses pelatihan pada huruf 99 karakter tersebut dilakukan dengan cara berurutan, dari A-Z, a-z, 0-9 dan simbol.



**Gambar 4.** *Flowchart*Pelatihan Jaringan Syaraf Tiruan *Hopfield*[2]

**2.7. Pengenalan Jaringan Syaraf Tiruan**

Input dari pengenalan jaringan syaraf tiruan sama dengan input pelatihan yaitu berupa file berkas digital yang sudah di *encode input* berupa matrik 1x100. Selain input berupa vektor pola gambar, pengenalan juga memerlukan data tambahan yaitu berupa nilai-nilai bobot koneksi jaringan dan vektor pola konvergen dari tiap karakter yangdilatih sebelumnya melalui pelatihan. Pada dasarnya pengenalan ini hanya membandingkan dan memanggil *output* suatu pola yang sudah dilatihkan sebelumnya. *Flowchart* pengenalan jaringan syaraf tiruan *Hopfield* [2]dilihat pada gambar 4.

Sifat JST *Hopfield* yaitu *auto-associative* ini berarti JST *Hopfield* bisa berprilaku seperti memori yaitu bisa menyimpan informasi yang pernah diberikan dan menampilkan kembali informasi tersebut. Pengenalan karakter ini tergantung pada proses segmentasi yang baik, apabila terdapat proses segmentasi yang tidak baik maka pola yang diberikan tidak sama seperti salah satu pola dalam pembelajaran maka JST *Hopfield*[2]mampu memilih salah satu pola dalam memorinya yang paling mirip.

1. **Hasil dan Pembahasan**

Pada modul pelatihan terdapat 2 proses utama yaitu proses input pola gambar (pengambilan gambar) dan proses pelatihan. Pada modul proses pengambilan gambar diinputkan file dalam bentuk \*.jpg.

**Tabel1.**HasilPersentasePengujiandengan Citra danUkuran yang Sama

|  |  |
| --- | --- |
| Ukuran Huruf | Keterangan |
| Jumlah Karakter | Jumlah yang Dikenali | huruf yang tidak dikenali (%) | Akurasi(%) |
| 10 pt | 99 | 88 | 10,1% | 88,9% |
| 12 pt | 99 | 88 | 10,1% | 88,9% |
| 14pt | 99 | 88 | 10,1% | 88,9% |
| 16 pt | 99 | 88 | 10,1% | 88,9% |

Hasil persentase pada tabel 1 adalah hasil pengujian yang dilakukan dengan menghitung semua karakter yang muncul dan dapat dikenali. Pada pengujian ini jumlah yang dikenali rata-rata 88,90 % dikarenakan huruf yang tidak dikenali (*error*) adalah merupakan suatu simbol dan itu tidak terdapat dalam MATLAB [3]. Huruf yang tidak dikenali tersebut dapat dilihat pada tabel 1.

**Tabel2.**HasilPersentasePengujiandengan Citra berbedadanUkuran yang Sama

|  |  |
| --- | --- |
| Ukuran Huruf | Keterangan |
| Jumlah Karakter | Jumlah yang Dikenali | huruf yang tidak dikenali (%) | Akurasi(%) |
| 10 pt | 99 | 79 | 24,24% | 75,76% |
| 12 pt | 99 | 69 | 30,30% | 69,70% |
| 14pt | 99 | 88 | 11,10% | 88,90% |
| 16 pt | 99 | 69 | 30,30% | 69,70% |

Hasil persentase pada tabel 2 adalah hasil pengujian yang dilakukan dengan menghitung semua karakter yang muncul dan dapat dikenali. Pada pengujian ini jumlah yang dikenali rata-rata 76,01 % dikarenakan huruf yang tidak dikenali (*error*) adalah merupakan karakter memiliki kemiripan bentuk.

**5.Simpulan**

 Pada aplikasi pengenalan karakter ini meliputi 3 tahap utama yaitu : Proses Image preprocessing, pelatihan dan pengenalan. Aplikasi program ini diimplementasikan menggunakan bahasa pemrograman Matlab. Pada proses Image preprocessing dilakukan uji segmentasi, uji pelatihan dilakuakn sampai kondisi stabil hingga siap dimasukan ke database. Sedangkan pada uji pengenalan dilakukan hal yang sama seperti tahap pelatihan skondisi stabil setelah stabil dikonversi kedalam file txt.Uji coba pelatihan yang dilakukan denga 99 karakter menghasilkan pengenalan tiap karakter untuk ukuran yang sama dengan yang dilatihkan sampai 88,90% dan rata-rata akurasi hasil pengenalan untuk empat ukuran karakter dan dengan *font* yang berbeda dengan yang dilatihkan adalah 76,01%.

**Daftar Pustaka**

1. DiyahPuspitaningrum. 2006. *PengantarJaringanSyarafTiruan*. Yogyakarta :Andi.
2. Sri Kusumadewi. 2003. *Artificial Intelligence (Teknik&Aplikasinya).* Yogyakarta: GrahaIlmu.
3. Sianipar, Herry, Wirjayati, 2013. *MatlabUntukPemrosesan Citra Digital*. Bandung: Informatika.