

DAMPAK REDUKSI SAMPEL MENGGUNAKAN *PRINCIPAL COMPONENT ANALYSIS (PCA)* PADA PELATIHAN JARINGAN SARAF TIRUAN TERAWASI (STUDI KASUS : PENGENALAN ANGKA TULISAN TANGAN)

Diyah Puspitaningrum¹, Dyan Kemala Sari², Boko Susilo³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Bengkulu.
Jl. WR. Supratman Kandang Limun Bengkulu 38371A INDONESIA
(tel: 0736-341022; fax: 0736-341022)

¹diyahpuspitaningrum@gmail.com

²kemalasari.dyan@yahoo.com

³masboko@gmail.com

Abstrak: Makalah ini membahas tentang bagaimana sebuah komputer mengenali sebuah pola citra digital berupa pengenalan angka tulisan tangan yang menggunakan metode *Principal Component Analysis (PCA)* pada pelatihan jaringan saraf tiruan terawasi model *Backpropagation*. Dalam penelitian ini terdapat dua pelatihan untuk mengetahui dampak reduksi sampel menggunakan PCA atau tanpa menggunakan PCA. Pada penelitian ini terdapat 1060 sampel citra angka tulisan tangan yang berformat *.jpg, sebanyak 660 sampel citra sebagai citra latih (*training*) dan sebanyak 400 sampel citra sebagai citra uji (*testing*). Sampel tersebut diambil dari angka tulisan tangan 106 responden yang berbeda. Hasil analisis menunjukkan bahwa sistem memiliki tingkat pengenalan bergantung pada dimensi dan jumlah zona. Pada pelatihan pertama, jika dimensi yang digunakan adalah 40 piksel x 30 piksel maka tingkat pengenalan sebesar 86.75% dengan waktu rata-rata 0.4946 detik. Jika dimensi yang digunakan adalah 50 piksel x 40 piksel maka pengenalan sebesar 82.75% dengan waktu rata-rata untuk memproses setiap masukan adalah selama 1.063 detik. Untuk pelatihan kedua, dengan jumlah zona 5 : 5 tingkat pengenalan adalah sebesar 86% dengan waktu rata-rata adalah 0.6582 detik. Jika jumlah zona yang digunakan adalah 5 : 6 maka tingkat pengenalan sebesar 86.25% dengan waktu rata-rata selama 0.7068 detik. Kesimpulannya metode *Principal Component Analysis (PCA)* dapat dijadikan sebagai alternatif untuk tingkat kecepatan dan keakuratan dari pelatihan jaringan saraf tiruan.

Kata Kunci: *Principal Component Analysis (PCA)*, *Backpropagation*, *Neural Network*, Pengenalan Angka Tulisan Tangan.

Abstract: This paper discuss about how computer recognize hands writing number recognition using Principal Component Analysis.(PCA) In supervised neural network backpropagation model training. In this method, there were two training method with and without PCA to know what is the effect of sample reduction. In this research there were 1060 pictures in JPG format. 660 pictures used as training data and 400 as testing data. Analysis output show that the system depend on dimension and zone. In first training 40x20 dimension was used in the experiment and have recognition level 86.75% and average time 0.4946 second. If 50x60 dimension used in experiment then recognition level is 82.75% and 1.603 second time. For the second kind of training method with 5:5 zone have recognition level 86% and 0.7608 second average time. The summary is PCA method could be used as speed and accuracy enhancer in neural network training.

Key Word :Principal Component Analysis (PCA), Back Propagation, Neural Network, Hands writing recognition.

I. PENDAHULUAN

Klasifikasi pola merupakan permasalahan yang banyak digunakan dalam berbagai aplikasi. Dalam pengolahan citra maupun pengolahan sinyal seringkali pengelompokan maupun pengklasifikasian dilakukan hanya didahului dengan transformasi data saja tanpa melakukan ekstraksi fitur. Selain itu, sampel sebagai data pembelajaran yang digunakan seringkali sangat terbatas. Analisis komponen utama atau *Principal Component Analysis (PCA)* adalah teknik yang

digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan variansi maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan [1].

Sebuah citra merupakan kumpulan piksel yang membentuk sebuah matriks, dimana pada sebuah piksel tersebut mengandung suatu informasi. PCA merupakan sebuah metode yang dapat mengolah informasi yang terdapat pada sebuah citra dengan hanya mengambil informasi yang signifikan dari citra tersebut sehingga informasi yang didapatkan menjadi lebih sederhana [2].

Makalah ini membahas tentang pengenalan pola karakteristik tulisan tangan khususnya angka yang berbasis citra dengan menggunakan metode PCA. *Principal Component Analysis* (PCA) merupakan salah satu metode paling sering digunakan dalam pengenalan citra dan kompresi data.

II. LANDASAN TEORI

A. *Principal Component Analysis* (PCA)

Analisis komponen utama atau *Principal Component Analysis* (PCA) [1] adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan variansi maksimum. *Principal Component Analysis* (PCA) dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Metode ini mengubah dari sebagian besar variabel asli yang saling berkorelasi menjadi satu himpunan variabel baru yang lebih kecil dan saling bebas (tidak berkorelasi lagi). Langkah-langkah metode *Principal Component Analysis* (PCA):

1. *Pelatihan (Training)*

Berikut adalah langkah-langkah dalam melakukan proses pelatihan (*training*) yang terdiri dari 7 (tujuh) langkah sebagai berikut :

1. Pembuatan T matriks citra

Pembuatan *database* yaitu menyusun suatu matriks

$$[T] = (M*N) * \text{Jumlah_Sampel} \dots (2.1)$$

2. Menghitung matriks *mean*

$$\text{Mean}[T] = \text{Sum}(T_Training) / N \dots (2.2)$$

3. Menghitung matriks normalisasi

$$[A] = [T] - \text{mean} \dots (2.3)$$

4. Menghitung matriks kovarian

$$[L] = [A^T] * [A] \dots (2.4)$$

5. Perhitungan *eigen vector* dan *eigen value* dari matriks L

Misalkan sebuah matriks bujur sangkar dengan order $n \times n$ yaitu L, dan sebuah *vector* kolom X dalam ruang *Euclidian* R^n yang dihubungkan dengan persamaan

$$LX = \lambda X \dots (2.5)$$

Dengan λ adalah suatu skalar dan X adalah *vector* $\neq 0$ maka λ dinamakan dengan *eigen value* dari matrix L. *Eigen value* merupakan nilai karakteristik dari matriks bujur sangkar dan matriks X dinamakan dengan *eigen vector* yang berkorespondensi dengan *eigen value* λ . Untuk menghitung *eigen vector* dan *eigen value* dapat menggunakan *class EigenvalueDecomposition* pada *package Accord.Math.Decompositions* dengan *properties Eigenvectors* sebagai *eigen vector* dan *DiagonalMatrix* sebagai *eigen value*.

6. Urutkan *eigen vector* berdasarkan *eigen value*-nya dari nilai terbesar sampai terkecil

$$\text{Eigen value} = X \dots (2.6)$$

$$\text{Eigen vector} = Y$$

7. Perhitungan (ruang wajah) matriks *eigen face*

$$[A] * Eigen\ vector \quad \dots (2.7)$$

Setelah melakukan proses pelatihan (*training*), maka langkah selanjutnya yaitu melakukan proses pengujian (*testing*).

2. *Pengujian (Testing)*

Proses pengujian (*testing*) yaitu melakukan pengujian terhadap masukan (*input*) matriks yang akan dihitung jarak *euclidean* terhadap matriks yang ada dalam matriks masukan (*training*). Terdapat 5 (lima) langkah untuk melakukan pengujian sebagai berikut :

8. Menghitung *projectedImages(n)*

$$eigenface^T * A(n) \quad \dots (2.8)$$

9. Perhitungan delta

$$T_Testing - mean \quad \dots (2.9)$$

10. Menghitung *projectedTestImages*

$$eigenface^T * delta \quad \dots (2.10)$$

11. Menghitung jarak *euclidean(n)*

$$projectedTestImages, projectedImages(n) \dots (2.11)$$

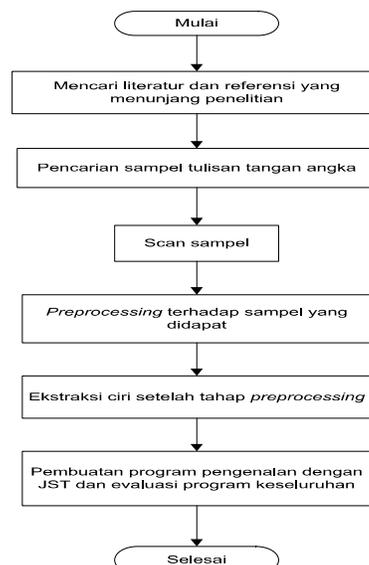
12. Jarak terkecil yaitu pada $n = 2$, sehingga untuk *T_Testing* mirip dengan matriks *T_Training* nomor 2

B. Jaringan Saraf Tiruan (*Artifical Neural Network*)

Jaringan saraf tiruan (*artifical neural network*) [3] adalah sistem komputasi yang arsitektur dan operasinya diilhami dari pengetahuan tentang sel saraf biologis di dalam otak. Jaringan saraf tiruan merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba menstimulasi proses pembelajaran pada otak manusia tersebut. Jaringan saraf tiruan dapat digambarkan sebagai model matematis dan komputasi untuk fungsi aproksimasi non-linear, klasifikasi data *cluster* dan regresi non-parametrik atau sebuah simulasi dari koleksi model jaringan saraf biologi.

C. Pengenalan Angka Tulisan Tangan

Pengenalan pola adalah mengelompokkan data numerik dan simbolik (termasuk citra) secara otomatis oleh komputer. Tujuan pengelompokkan pola ini adalah untuk mengenali suatu objek di dalam citra [4]. Klasifikasi pola merupakan permasalahan yang banyak digunakan dalam berbagai aplikasi. Dalam pengolahan citra maupun pengolahan sinyal seringkali pengelompokan maupun pengklasifikasian dilakukan hanya didahului dengan transformasi data saja tanpa melakukan ekstraksi fitur. Selain itu, sampel sebagai data pembelajaran yang digunakan seringkali sangat terbatas [1]. Pengenalan tulisan tangan dapat dilakukan secara *offline* maupun *online*. Pengenalan angka tulisan tangan secara *offline* adalah sebuah teknik dimana *input* data yang berupa lembaran kertas hasil *scan* menggunakan gambar pada komputer yang dikenali sebagai titik-titik (*bitmap*), *bitmap* inilah yang kemudian diproses lebih lanjut dengan menggunakan algoritma tertentu menjadi karakter sehingga dapat dikenali dan diolah menjadi informasi [4]. Tahapan Pengenalan Pola Tulisan Tangan Angka



Gambar 1. Alur proses penelitian [5]

D. Teknik *Zoning*

Metode ekstraksi fitur berbasis zona memberikan hasil yang baik bahkan ketika langkah sebelum proses tertentu dimulai seperti *filtering*, *smoothing*, dan menghapus zona yang tidak dianggap. Pada bagian ini, akan dijelaskan konsep metode ekstraksi, ciri yang digunakan untuk mengekstraksi fitur untuk klasifikasi yang efisien dan pengenalan.

Tahapan yang harus dilakukan untuk ekstraksi fitur ini, yaitu :

1. Hitung *centroid* utama dari citra.
2. Bagi ke dalam n buah zona yang sama besar proporsinya.
3. Hitung jarak antara titik *centroid* dengan koordinat *pixel* yang memiliki nilai.
4. Ulangi langkah 3 untuk *pixel* yang ada di semua zona.
5. Hitung rata-rata dari jarak yang telah didapat pada langkah 3. Jarak akan dihitung menggunakan metode *Euclidean* dua dimensi.

$$d(P, C) = \sqrt{(x_p - x_c)^2 + (y_p - y_c)^2}$$

Keterangan :

d = jarak antara dua titik

P = koordinat titik berat

C = koordinat *pixel*

x_p = koordinat x titik berat

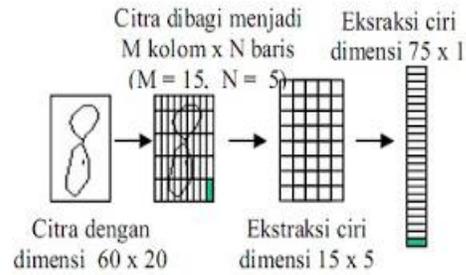
y_p = koordinat y titik berat

x_c = koordinat x *pixel*

y_c = koordinat y *pixel*

6. Ulangi langkah 5 hingga didapat masing-masing rata-rata jarak dari setiap zona.

7. Akhirnya jarak (*distance*) inilah yang merupakan hasil dari ekstraksi fitur *Zoning*.



Gambar 2. Visualisasi ekstraksi ciri [5]

E. *Backpropagation*

Backpropagation adalah jaringan saraf tiruan *feed-forward* yang terdiri dari sejumlah neuron yang dihubungkan oleh bobot-bobot penghubung. Neuron-neuron tersebut disusun dalam lapisan-lapisan yang terdiri dari satu lapisan *input* (*input layer*), satu atau lebih lapisan tersembunyi (*hidden layer*), dan satu lapisan *output* (*output layer*). Lapisan input menerima sinyal dari luar, kemudian melewatkannya ke lapisan tersembunyi pertama, yang akan diteruskan sehingga akhirnya mencapai lapisan *output*.

F. *Matrix Laboratory* (MATLAB)

Matrix Laboratory (MATLAB) [6] adalah sebuah bahasa pemrograman dengan kemampuan tinggi untuk komputasi teknis. MATLAB menggabungkan komputasi, visualisasi, dan pemrograman dalam satu kesatuan yang mudah digunakan dimana masalah dan penyelesaiannya dijelaskan dalam suatu notasi, pemakaian MATLAB.

III.METODOLOGI

A. Metode Pengumpulan Data

Objek penelitian dari paper ini adalah sampel angka tulisan tangan dari 106 orang responden yang dipilih yang masing-masing menulis semua sampel angka. Sampel angka terdiri dari 10 angka

untuk masing-masing responden. Rincian dari objek penelitian ini adalah:

Angka : $106 \times 10 = 1060$ sampel.

B. Metode Pengembangan Sistem

Pengembangan aplikasi dampak reduksi sampel untuk pengenalan angka tulisan tangan dalam Tugas Akhir ini menggunakan model *waterfall*. Adapun penjelasan langkah-langkah yang dilakukan dalam pengembangan sistem ini secara garis besar adalah sebagai berikut.

1. Analisis Kebutuhan

- a. Kebutuhan data masukan
- b. Kebutuhan data keluaran
- c. Kebutuhan *interface*

2. Perancangan Aplikasi

3. Melakukan operasi pemotongan (*cropping*) pertama citra untuk meng-hilangkan margin tahap pertama.
4. Melakukan operasi pe-ngubahan ukuran (*resize*) pertama citra untuk melakukan penyesuaian dimensi tahap pertama.
5. Melakukan operasi penipisan (*thinning*).
6. Melakukan operasi pemotongan (*cropping*) citra kedua
7. Melakukan operasi pe-ngubahan ukuran (*resize*) citra kedua untuk melakukan penyesuaian dimensi tahap kedua
8. Melakukan proses *Principal Component Analysis* (PCA)
9. Melakukan proses JST (Pengenalan citra angka tulisan tangan)

IV. HASIL DAN PEMBAHASAN

Proses pelatihan dibagi menjadi 2 (dua) tahap proses yaitu ekstraksi fitur dan pelatihan jaringan yaitu sebagai berikut :

1. Metode pertama

- a. Ekstraksi fitur menggunakan metode *Principal Component Analysis* (PCA)
- b. Pelatihan Jaringan Saraf Tiruan dari ekstraksi fitur menggunakan metode *Principal Component Analysis* (PCA)

2. Metode kedua

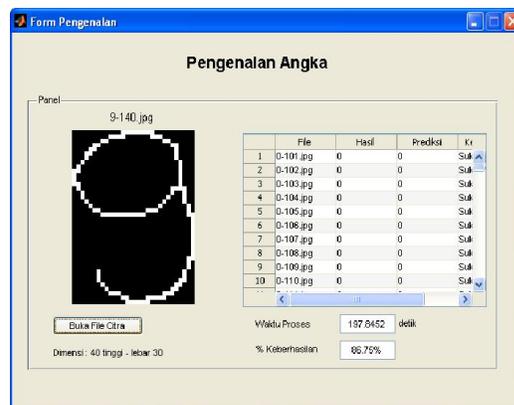
- a. Ekstraksi fitur menggunakan teknik *Zoning*
- b. Pelatihan Jaringan Saraf Tiruan dari ekstraksi fitur menggunakan teknik *Zoning*

4.1. Ekstraksi Fitur Menggunakan Metode

Principal Component Analysis (PCA)

Langkah-langkah ekstraksi fitur dengan menggunakan PCA ini adalah sebagai berikut :

1. Membaca lokasi *file* citra
2. Mengubah citra ke dalam format hitam dan putih (*black and white*)



Gambar 3 Pengujian citra angka 0 sampai 9 menggunakan PCA

Gambar 3 merupakan gambar pengujian citra secara keseluruhan yaitu angka 0 sampai dengan 9. Jadi, sistem membaca citra angka tulisan tangan yang ada *folder* citra_test yang berjumlah sebanyak 400 citra angka tulisan tangan. Dari hasil pengujian citra angka secara keseluruhan dapat dilihat bahwa waktu proses untuk mengenali citra sebanyak 400 citra angka tulisan tangan dengan dimensi tinggi 40 pixel dan lebar 30 pixel adalah selama 197.8425 detik dengan persentase keberhasilan adalah sebesar 86.75%.

4.2. Ekstraksi Fitur Menggunakan Teknik *Zoning*

Langkah-langkah ekstraksi fitur dengan menggunakan teknik *Zoning* adalah sebagai berikut:

1. Membaca citra *training *.jpg*
2. Melakukan operasi *black and white* (BW)
3. Melakukan operasi pemotongan (*cropping*) pertama citra untuk menghilangkan margin tahap pertama.
4. Melakukan operasi pe-ngubahan ukuran (*resize*) pertama citra untuk melakukan penyesuaian dimensi tahap pertama.
5. Melakukan operasi pe-motongan (*cropping*) citra tahap kedua
6. Melakukan operasi penipisan (*thinning*) atau proses *skeletonisasi*.
7. Melakukan operasi *remove noise*
8. Melakukan ekstraksi fitur dengan teknik *Zoning*
9. Melakukan proses JST (Pengenalan citra angka tulisan tangan)



Gambar 4 Pengujian citra angka 0 sampai 9 menggunakan teknik *Zoning*

Gambar 4 merupakan gambar pe-ngujian citra secara keseluruhan yaitu angka 0 sampai dengan 9. Jadi, sistem membaca citra angka tulisan tangan yang ada *folder* citra_test yang berjumlah sebanyak 400 citra angka tulisan tangan. Dari hasil pengujian citra angka secara keseluruhan dapat dilihat bahwa waktu proses untuk mengenali citra sebanyak 400

citra angka tulisan tangan dengan zona 4 : 3 adalah selama 223.6367 detik dengan persentase keberhasilan adalah sebesar 81%.

V. KESIMPULAN

Kesimpulan yang dapat diambil dari proses penelitian yang dilakukan adalah sebagai berikut :

1. Metode *Principal Component Analysis* (PCA) layak dipertimbangkan sebagai metode alternatif untuk pengujian jaringan saraf tiruan selain dengan teknik ekstraksi fitur (khususnya teknik *Zoning*).
2. Kelebihan lain dari Metode *Principal Component Analysis* (PCA) adalah pada kompresi data.
3. Untuk pelatihan pertama, besarnya tingkat keakuratan suatu pengujian dan lamanya waktu proses dipengaruhi oleh besarnya dimensi yang dimasukan serta nilai-nilai parameter dalam jaringan saraf tiruan.
4. Untuk pelatihan kedua, besarnya tingkat keakuratan suatu pengujian dan lamanya waktu proses dipengaruhi oleh jumlah zona yang dimasukan serta nilai-nilai parameter dalam jaringan saraf tiruan.
5. Waktu proses dipengaruhi oleh dua faktor yakni : ukuran angka dan jumlah segmen yang dimiliki oleh angka yang akan dikenali.

VI. SARAN

Berikut saran yang penulis ajukan guna pengembangan sistem pengenalan angka tulisan tangan menggunakan metode *Principal Component Analysis* (PCA) pada pelatihan jaringan saraf tiruan terawasi dan ekstraksi ciri pada pelatihan jaringan saraf tiruan terawasi :

1. Untuk pengembangan aplikasi selanjutnya diharapkan agar sudah ada aturan lain dalam mengenali suatu angka tulisan tangan tanpa

harus melakukan *trial and error* terlebih dahulu.

2. Untuk pengembangan sistem selanjutnya diharapkan agar dapat mengenali angka tulisan tangan dengan menggabungkan ketiga metode dalam pelaksanaannya yaitu menggunakan ekstraksi ciri - Metode *Principal Component Analysis* (PCA) - Jaringan Saraf Tiruan.
3. Kelemahan dari teknik *Zoning* ini untuk ekstraksi fitur adalah komputasi yang didapatkan tinggi sehingga memerlukan waktu yang cukup lama untuk proses pengekseskiannya. Metode ekstraksi fitur lain yang dapat digunakan untuk proses pengenalan pola tulisan tangan adalah *Fuzzy Feature Extraction* yang merupakan metode yang cukup baik untuk mengenali pola tulisan tangan

dengan waktu pengekseskuan yang tidak terlalu lama.

REFERENSI

- [1] Susetyoko, R., & Purwantini, E. (2010). Teknik Reduksi Dimensi Menggunakan Komponen Utama Data Partisi Pada Pengklasifikasian Data Berdimensi Tinggi dengan Ukuran Sampel Kecil. *Jurnal Politeknik Elektronika Negeri Surabaya*.
- [2] Faridh, M. M. (2013). Pengenalan Karakter Huruf Tulisan Tangan Dengan Menggunakan Metode *Principal Components Analysis*. *Jurnal Program Studi Teknik Informatika S-1*. Semarang : Universitas Dian Nuswantoro
- [3] Hermawan. (2006). Pengertian Jaringan Saraf Tiruan. *Jaringan Saraf Tiruan*.
- [4] Santosh. (2009). Pengenalan Pola Tulisan Tangan. *Jurnal Pengenalan Tulisan Tangan*.
- [5] Setiawan, W., & Asri, S. A. (2005). Aplikasi Jaringan Saraf Tiruan Perambatan Balik Pada Pengenalan Angka Tulisan Tangan. *Jurnal Politeknik Negeri Bali*. Bali : Kampus Bukit Jumbaran Bali.
- [6] Bisonerich. (2009). Pengantar Matlab. *Matlab*.