

EKSTRAKSI FITUR MENGGUNAKAN PRINCIPAL COMPONENT ANALISYS (PCA)

Puteri Noraisya Primandari *¹, Bagus Hardiansyah ²

Universitas 17 Agustus 1945 Surabaya,

puterinoraisya@untag-sby.ac.id *¹

Universitas 17 Agustus 1945 Surabaya

bagushardiansyah@untag-sby.ac.id *²

Abstrak

Penggunaan sistem teknologi biometrika dengan karakteristik ekspresi wajah memungkinkan untuk mengenali emosi seseorang. Beberapa ekspresi wajah yang dikelompokkan adalah ekspresi netral, gembira, terkejut, mata kedip, mengantuk dan sedih. Walaupun bentuk atau tampilan ekspresi tersebut berbeda-beda antara setiap individu, manusia tetap dapat mengenalinya. Pada penelitian ini telah dibuat suatu sistem yang dapat digunakan untuk mengenali citra ekspresi wajah manusia menggunakan metode PCA untuk ekstraksi fitur. Hasil ekstraksi fitur dengan PCA merupakan inisialisasi untuk dapat di kenali sebagai eigen value dari tiap-tiap ekspresi wajah

Keywords: Jaringan Syaraf Tiruan, Ekspresi wajah. Principal Component Analysis (PCA)

Abstract

The use of biometric technology systems with the characteristics of facial expressions makes it possible to recognize one's emotions. Some grouped facial expressions are expressions of neutral, joyful, surprised, blinking, sleepy and sad eyes. Although the form or appearance of the expression varies between individuals, humans can still recognize it. In this study a system has been created that can be used to recognize human facial expression images using the PCA method for feature extraction. The result of feature extraction with PCA is an initialization to be recognized as the eigen value of each facial expression

Keywords: Artificial Neural Networks, Facial Expressions. Principal Component Analysis (PCA)

Pendahuluan

Teknologi biometrika mulai banyak terapannya diberbagai bidang. Teknologi ini diterapkan dengan menggunakan karakteristik pembeda (*distinguishing traits*). Penggunaan sistem teknologi biometrika dengan karakteristik ekspresi wajah memungkinkan untuk mengenali emosi seseorang, sehingga analisis ekspresi wajah berhubungan dengan pengenalan secara visual gerakan wajah dan perubahan fitur wajah.

Komponen dasar sistem analisis ekspresi wajah adalah deteksi wajah, ekstraksi data wajah, dan pengenalan ekspresi wajah (Abidin, 2011).

Jaringan Syaraf Tiruan adalah salah satu metode komputasi yang berusaha meniru cara kerja sistem syaraf otak manusia. Keunggulan metode ini dibanding metode lain adalah kemampuan proses pelatihan dan memecahkan hubungan yang kompleks dan rumit, sehingga sulit untuk dideskripsikan antara data masukan dan data keluaran. Hal ini dimungkinkan karena pengetahuan yang ada pada jaringan syaraf tiruan tidak di program, namun dilatih berdasarkan informasi atau masukan yang diterimanya (Kusumto, 2010).

Metode PCA merupakan metode dalam statistika yang digunakan untuk ekstraksi fitur input dengan kehilangan informasi yang minimum. Beberapa peneliti telah menggabungkan PCA dengan neural network. (Rama, 2010) menggunakan model PCA dan *backpropagation* dengan tingkat akurasi 97.25 membandingkan dengan metode PCA dan *radial basic function* dengan tingkat akurasi 97.46. (Te-Hsiu Sun dan Fang-Chih Tien) menggunakan 2D dan 3D dengan menggunakan metode *backpropagation* dan PCA + LAC berhasil menentukan perbedaan ekspresi.

Metode Penelitian

Principal Component Analisis (PCA)

PCA pertama kali ditemukan oleh (Turk, 1991) yang digunakan untuk pengenalan citra wajah. Dalam metode PCA informasi yang paling baik mendeskripsikan wajah diturunkan dari citra wajah secara keseluruhan. kegunaan dari PCA adalah dapat mengkompres data dengan cara ekstraksi fitur data tanpa menghilangkan informasi-informasi dalam data tersebut (Anandita, 2007).

PCA lebih banyak digunakan untuk keperluan ekstraksi fitur citra, dimana jumlah piksel dari citra jauh lebih besar dibandingkan dengan jumlah data sampel yang digunakan. Untuk melakukan proyeksi sampel matriks dari citra pelatihan, setiap citra pelatihan disusun dalam bentuk vektor baris (soesanto, 2010).

Proyeksi pada PCA merupakan representasi himpunan data X ke dalam bentuk vektor eigen dari matrik varian-kovarian dari X . Vektor eigen dengan nilai eigen yang besar mempunyai peranan penting dalam proses perubahan pada nilai matrik piksel. dengan cara membuang nilai eigen yang mendekati nol tidak akan membuat kita kehilangan informasi data atau hanya kehilangan informasi yang minimum (Farida, 2014).

PCA memerlukan masukan data yang mempunyai sifat *zero-mean* pada setiap fiturnya. Sifat *zero-mean* pada masing-masing fitur data bisa didapatkan dengan mengurangkan semua nilai dengan rata-ratanya. Set data X dengan dimensi $M \times N$, dimana M adalah baris dan N kolom, akan tampak seperti berikut:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ x_{M1} & x_{M2} & \dots & x_{MN} \end{bmatrix} \quad (0.1)$$

Untuk fitur ke- j , semua nilai pada kolom tersebut dikurangi dengan rata-ratanya, diformulasikan dengan

$$x'_{ij} = x_{ij} - \bar{x}_j \quad (0.2)$$

dengan $X' = [x'_{ij}]$

$i = 1, 2, \dots, M$

$j = 1, 2, \dots, N$

\bar{x}_j = nilai rata-rata kolom ke - j

Selanjutnya dilakukan proses untuk mendapatkan matriks kovarian dari matriks X' , yaitu C , dengan menggunakan formula berikut:

$$C = \frac{1}{M} X' \cdot X'^T \quad (0.3)$$

X'^T adalah matriks transpose dari X' .

Pada matriks C , elemen ke- ij adalah hasil kali elemen baris matriks X' dengan kolom matriks X'^T . Sifat-sifat yang dimiliki oleh matriks C adalah sebagai berikut:

1. C adalah matriks simetris bujur sangkar berukuran $M \times M$.
2. Bagian diagonal utama (dari kiri atas ke kanan bawah) adalah nilai varian masing-masing fitur sesuai dengan indeks kolomnya.
3. Bagian selain diagonal utama adalah kovarian di antara pasangan dua fitur yang berkesesuaian.

Jadi, matriks C merepresentasikan kovarian di antara semua pasangan yang mungkin dari fitur data set matriks X' . Nilai kovarian merefleksikan noise pada fitur.

Nilai eigen dan vektor eigen dari matriks kovarian dihitung dengan menggunakan persamaan karakteristik berikut ini:

$$C - \lambda I = 0 \quad (0.4)$$

$$(C - \lambda I)v = 0 \quad (0.5)$$

Dengan C adalah matriks kovarian, I adalah matriks Identitas, λ adalah nilai eigen dan v adalah vektor eigen.

Nilai eigen yang terbesar yang berkorespondensi terhadap nilai vektor eigen yang terbesar dipilih menjadi *Principal Component*. Vektor eigen yang disusun dari yang terbesar ke yang terkecil dipilih menjadi vektor fitur.

$$\text{Vektor Fitur} = (eig_1, eig_2, eig_3 \dots eig_n) \quad (0.6)$$

Untuk mencari *Principal Component* dengan X' sebagai rata-rata dihitung dengan rumus.

$$PC = X' \times v \quad (0.7)$$

Langkah berikutnya melakukan transformasi data untuk menghasilkan data PCA dengan X sebagai data awal.

$$PCA \text{ data} = PC^T \times X'^T \quad (0.8)$$

Hasil dan Diskusi

Ekstraksi Fitur dengan PCA

Ekstraksi fitur merupakan proses yang bertujuan untuk menentukan ciri dari suatu citra ekspresi wajah. Metode PCA merupakan salah satu metode yang digunakan untuk ekstraksi fitur.

Berikut ini adalah tahap dalam algoritma PCA

1. Hitung rata-rata seluruh sampel data diperoleh dengan menggunakan persamaan:

$$\bar{x}_j = \frac{\sum_{ij=1}^n x_{ij}}{n} \quad (3.1)$$

2. Adjusted data (data yang telah disesuaikan) adalah hasil pengurangan dari setiap data dengan rata-rata setiap data yang diperoleh dengan rumusan berikut ini:

$$\text{Adjusted data} = x_{ij} - \bar{x}_j \quad (3.2)$$

dengan $X' = \text{Adjusted Data}$

3. Hitung matrik kovarian (c) dihitung dengan menggunakan persamaan berikut:

$$c = \frac{1}{M} X' X'^T \quad (3.3)$$

Dimana X' adalah *Adjusted Data* dan X'^T adalah transpose dari matrik X'

4. Hitung nilai eigen dan vektor eigen dari matrik kovarian dihitung dengan menggunakan persamaan karakteristik berikut ini:

$$c - \lambda I = 0 \quad (3.4)$$

$$(c - \lambda I)v = 0 \quad (3.5)$$

Dimana c adalah matrik kovarian, I adalah matrik Identitas, λ adalah nilai eigen dan v adalah vektor eigen.

5. Hitung nilai eigen yang terbesar yang berkorespondensi terhadap nilai vektor eigen yang terbesar dipilih menjadi *Principal Component*. Vektor eigen yang disusun dari yang terbesar ke yang terkecil dipilih menjadi vektor fitur.

$$v = (eig_1, eig_2, eig_3 \dots eig_n) \quad (3.6)$$

6. Untuk mencari *Principal Component* dengan X' sebagai rata-rata.

$$PC = X' \times v \quad (3.7)$$










7. Langkah berikutnya untuk melakukan transformasi data untuk menghasilkan data PCA.










$$PCA \text{ data} = PC^T \times X'^T \quad (3.8)$$

3.2 Contoh Ekspresi Wajah dari Database Yale Face

Tabel 0.1 Contoh Ekspresi Wajah dari Database Yale Face

<http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html>

EKSPRESI	WAJAH		
NETRAL			
GEMBIRA			
TERKEJUT			

SEDIH			
MENGANTUK			
MATA MENGEDIP			

Kesimpulan

Berdasarkan hasil dan pembahasan dapat diambil kesimpulan sebagai berikut:

1. Dengan ditemukannya PCA, telah membawa perubahan yang sangat besar pada ekstraksi fitur yang berbasis *"appearance"*
2. PCA merupakan teknik linier reduksi menggunakan teori-teori sederhana dari statistik seperti varian, standart deviasi, zero mean, kovarian dan persamaan karakteristik
3. PCA lebih banyak digunakan untuk keperluan ekstaksi fitur citra, dimana jumlah dimensi dari citra jauh lebih besar dibandingkan dengan jumlah data sampel yang digunakan.
4. Untuk melakukan proyeksi sampel vektor dari citra pelatihan, setiap citra pelatihan disusun dalam bentuk vektor baris. Jika jumlah data pelatihan adalah sebanyak m , maka dimensi vektornya adalah $m \times n$. Apabila vektor citra pelatihan mempunyai dimensi $m \times n$ tersebut diortogonlisasi dengan menggunakan *eigenvector* dan *eigenvalue*.

Saran

Persoalannya adalah dimensi yang tidak sama antara data sampel ($m \times n$), sedangkan dimensi dari rata-rata seluruh data sampel ($1 \times n$). Oleh karena itu perlu disamakan dimensinya dengan menggunakan rata-rata seluruh wajah sebanyak m , sehingga rata-rata seluruh data sampel mempunyai dimensi ($m \times n$).

Notation

X	Vektor input
C'	Konstan
X	ukuran citra dengan dimensi $M \times N$
c_x	Perhitungan matriks kovarian dari matriks X
C	adalah matriks kovarian
I	matriks indeks
λ	nilai eigen
v	vektor eigen.

Referensi

- Hardiansyah, Bagus. Pengenalan Ekspresi Wajah Menggunakan Jaringan Syaraf Tiruan Kohonen Self Organizing Maps (K-SOM). (2015). Seminar Nasional Matematika dan Pendidikan Matematika Unesa Surabaya. Surabaya. Unesa University Press. 51 – 57.
- Majumder, A., Behera, L. dan Subramanian, V. K. (2014). *Emotion Recognition From Geometrical Facial Features Using Self-Organizing Maps*. Pattern Recognition (47) 1282-1293.
- Sun T. dan Tien F. (2008). *Using backpropagation neural network for face recognition with 2D + 3D hybrid information*. Expert System with Application (35) 361-372.
- Leksmi P. dan Sasikumar M. (2009). *Analysis of Facial Expression using Gabor and SVM*. International Journal of Recent Trends in Engineering 2 (1).
<http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html>
- Reddy K. R. L., Babu G.R., dan Kishore L. (2010). *Face Recognition on Eigen Feature of Multy Scale Face Components and an Artificial Neural Network*. Procedia Computer Science (2) 62-74.
- Boediono, S. (2009). *Pemanfaatan Jaringan Syaraf Tiruan Kohonen Self Organizing Maps Untuk Pengenalan Wajah*. Tugas Akhir Jurusan Matematika ITS Surabaya

- Amin, M. (2012). *Pengelompokan Potensi Daerah di Bidang Komunikasi dan Informasi Menggunakan Principal Component Analysis dan Self Organizing Map*. Jurnal Penelitian Komunikasi, Informasi dan Media Massa- PEKOMMAS, 3 (15).
- Fausett, L. (1994). *Fundamentals of Neural Networks. Architectures, Algorithms and Applications*. Prentice-Hall, New Jersey: USA.
- Irawan, M. I. (2008). *Exploratory Data Analysis dengan JST- Kohonen SOM: Struktur Tingkat Kesejahteraan Daerah Tk II se Jawa Timur*. Institut Teknologi Sepuluh Nopember.
- Gonzales, R. C. dan Wood, R. E. (2001), *Digital Image Processing*. Edition, Prentice-Hall, inc.
- Li, J., Hao, W., dan Zhang, X. (2015). *Learning Kernel Subspace for Face Recognition*. Neurocomputing (151), 1187-1197.
- Kurdthongmee, W. (2008). *Color Classification of Rubberwood Boards for Fingerjoint Manufacturing Using a SOM Neural Network and Image Processing*. Computer and Electronics in Agriculture 64, 85-92.
- Yusob, B., Shamsuddin, S. M., dan Hamed, H. N. A. (2013). *Spiking Self-Organizing Maps for Classification Problem*. Procedia Technologi 11, 57-64.
- Zaenal, A. (2012). *Rancang Bangun Sistem Pengenalan Ekspresi Wajah Menggunakan Fisherface dan Jaringan Syaraf Tiruan Backpropagation*. Jurnal FMIPA UNNES 35 (2), 194-203
- Wasista, S., Bayu, B. S., dan Putra, S. A. (2011). *Sistem Pengenalan Wajah Pada Mesin Absensi Mahasiswa Menggunakan Metode PCA dan DTW*. Industrial Electronics Seminar, 224-229.
- Sutarno. (2010). *Identifikasi Ekspresi Wajah Menggunakan Alihragam Gelombang Singkat (Wavelet) dan Jaringan Syaraf Tiruan Learning Vector Quantization (LVQ)*. Seminar nasional Informatika UPN Veteran Yogyakarta, 87-94
- Siang, J. J. (2005). *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab*. ANDI Yogyakarta.
- Santosa, B. (2007). *Data Mining, Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Graha Ilmu, Edisi Pertama.
- Prasetyo, E. (2012). *Data Mining, Konsep dan Aplikasi menggunakan Matlab*. ANDI Yogyakarta, Edisi pertama.
- Purnomo, M. H. dan Muntasa A. (2010). *Konsep Pengolahan Citra Digital dan Ekstraksi Fitur*. Graha Ilmu, Edisi Pertama.

Santoso, (2015). Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas.