SISTEM PENGENALAN WAJAH MENGGUNAKAN METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DENGAN ALGORITMA FUZZY C-MEANS (FCM)

Andri Suryadi

Abstrak

Berkembangnya ilmu pengetahuan dan teknologi dalam computer vision menjadikan keamanan komputer menjadi sangat penting. Salah satu contoh keamanan komputer adalah dengan cara pengenalan wajah. Skripsi ini membahas algoritma tentang pengenalan suatu wajah agar dapat dikenali oleh sistem komputer berdasarkan data training yang telah ada dalam database. Fitur-fitur yang terdapat dalam wajah akan dicari menggunakan Principal Component Analysis (PCA), sedangkan untuk tahap identifikasi menggunakan algoritma Fuzzy C-means (FCM). Principal Component Analysis akan digunakan untuk mereduksi citra wajah yang menghasilkan output berupa feature yang akan dijadikan inputan ke dalam algoritma fuzzy C-means. FCM mengelompokan data menjadi beberapa cluster yang masingmasing cluster diwakili pusat cluster. Pusat cluster inilah yang akan dijadikan dasar untuk mengenali data baru. Hasil pengujian yang telah dilakukan menggunakan PCA dan FCM dengan menggunakan 150 data latih yaitu sebesar 84%, 300 data latih yaitu 76% dan 450 data latih yaitu 76% sedangkan nilai akurasi rata-rata normal adalah 74%. Kemudian pengujian dengan tambahan noise yaitu menggunakan 150 data latih yaitu sebesar 26%, 300 data latih yaitu 14% dan 450 data latih yaitu 8% sedangkan nilai akurasi rata-rata noise adalah 16%. Dapat disimpulkan dengan menggunakan PCA dan FCM sistem pengenalan wajah ini menghasilkan akurasi cukup baik namun, sebaliknya dengan tambahan noise sistem pengenalan wajah tidak dapat berjalan dengan baik.

Kata Kunci: Pengenalan wajah, Principal Component Analysis (PCA), Fuzzy C-means (FCM).

1. Pendahuluan

Sistem Pengenalan wajah merupakan pengenalan pendekatan pola keperluan identifikasi seseorang disamping pendekatan pola lainnya. Pengenalan citra berhubungan dengan obyek yang tidak pernah sama, karena adanya bagian-bagian yang dapat berubah.perubahan ini dapat disebabkan oleh ekspresi wajah, intensitas cahaya dan sudut pengambilan gambar, atapun perubahan akasesoris pada wajah. Dalam kaitan ini, obyek yang sama dengan beberapa perbedaan tersebut harus dikenali sebagai satu obyek yang sama (Joko Hartono, 2009).

Terdapat dua hal utama yang sangat penting dalam proses pengenalan wajah yaitu: proses ekstraksi fitur dari sampel wajah yang ada dan juga teknik pengelompokan yang digunakan untuk mengenali suatu wajah agar wajah tersebut sesuai yang diharapkan (Fitri Damayanti).

Secara garis besar tahapan yang dilakukan dalam sistem pengenalan wajah adalah sebagai berikut (Bambang Dwi Liestiawan, 2009):

1. PraProses

Tahapan yang digunakan untuk menyeleksi gambar sehingga citra gambar tersebut sesuai yang diinginkan.

2. Feature Extraction

ISSN 2086-4280 58

Tahapan untuk mengambil data informasi penting yang terdapat dalam citra gambar.

3. Pengelompokan

Tahapan ini berfungsi untuk mengelompokan citra-citra gambar sehingga proses pengenalan wajah tersebut dapat dicari.

Principal Component Analysis (PCA) teknik yang digunakan adalah sebuah untuk menyederhanakan suatu data dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan varians maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Dalam metoda ini citra wajah akan diproyeksi sebuah ruang fitur yang menonjolkan variasi yang signifikan diantara citra yang diketahui. Fitur signifikan inilah sering disebut "Eigenface" karena fitur-fitur tersebut adalah komponen utama dari set citra wajah untuk training. Tujuan PCA adalah menangkap variasi total pada citra latih dan merepresentasikan variasi tersebut dalam variabel-variabel yang jumlahnya lebih sedikit. PCA dikenal juga dengan sebutan transformasi Karhunen-Loeve dan transformasi Hotteling (Kartika Gunadi).

Terdapat beberapa metode yang digunakan dalam pengelompokan *feature vector* diantaranya: pendekatan statistik, *neural network*, algoritma genetic, *fuzzy clustering* (Krasteva,2002:1)

Dalam pembuatan sistem pengenalan wajah ini digunakan fuzzy c-mean sebagai pengelompokan metode data. **FCM** pertama dikenalkan kali oleh JIM tahun BEZDEK pada 1981 yang merupakan suatuu teknik peng*clusteran* data dimana tiap-tiap data dalam suatu cluster ditentukan oleh derajat keanggotaannya (Cavendis, 2010). Penggunaan FCM sendiri telah lebih dahulu dilakukan oleh Ahmad Ridwan (2010)dengan skripsi beriudul "Pengenalan Tulisan Tangan Aksara Sunda Ngalagena dengan Algoritma Fuzzy C-Mean" dengan menghasilkan nilai akurasi relatif 100%. Sistem pengenalan wajah ini menggunakan algoritma yang sama dalam pengklasifikasian. Dengan metode tersebut diharapkan nilai akurasi tidak jauh berbeda dengan apa yang telah dilakukan dalam penelitian sebelumnya.

2. Sistem Pengenalan Wajah

2.1 Pra Proses

Pra proses dalam pengenalan wajah ini bertujuan untuk membuat data mentah menjadi data yang dapat diolah oleh sistem. Dalam proses pengenalan wajah digunakan image grayscale sebagai data yang akan diolah. Hal ini dikarenakan image dalam bentuk grayscale lebih mudah untuk dianalisa apabila dibandingkan dengan image berwarna. *Image* berwarna adalah kombinasi dari tiga warna utama yaitu: merah, hijau dan biru yang disebut dengan sistem warna RGB. Sistem warna RGB ini terdiri dari 24 bit, masing-masing bit untuk merah, hijau dan biru. Agar lebih mudah untuk dianalisa maka image tersebut terlebih dahulu menjadi image grayscale dengan cara memberi nilai yang sama untuk masingmasing 8 bit itu (Robert Hoo, 2003:12).

Dalam sistem pengenalan wajah yang akan dibuat ini terdapat dua Pra proses yang akan dilakukan yaitu : menangkap objek wajah dari sebuah *webcam* dalam hal ini menggunakan sebuah libray yaitu aplikasi *Open CV*, kemudian membuat sebuah *image* warna menjadi sebuah *grayscale*.

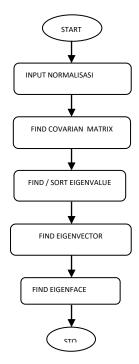
2.2 Feature Extraction PCA

Prinsip dasar dari *Principal Componen Analisys* (PCA) adalah dengan memproyeksi *image* ke dalam ruang *eigen*nya / ruang wajah. Cara mendapatkannya adalah dengan mencari eigen *vector* yang dimiliki setiap *image* dan memproyeksikannya ke dalam ruang wajah. Sasaran dari *Principal Componen*

Analisys (PCA) adalah untuk menangkap variasi total didalam kumpulan wajah yang dilatihnya (Kartika Gunadi,2001 :2). Dibawah ini adalah langkah-langkah yang dilakukan *Principal Componen Analisys* (PCA) dalam proses pengenalan wajah (Robert Hoo, 2003 :13).

PCA termasuk dalam bidang multivariate analysis pada ilmu statistik. multivariate analysis secara sederhana dijelaskan sebagai metode yang berhubungan dengan variable dalam jumlah besar pada satu atau banyak percobaan. Beberapa bidang lain pada multivariate analysis adalah common factor analysis, multiple regression, discriminant multiple analysis, multivariate analysis of variance dan covariance, conjoint analysis, canonical correlation, cluster analysis, multidimensional scaling, correspondence analysis, linear probability model, dan / structurak simultaneous equation modeling (Prentice-Hall International, 1993:13).

Adapun *flowchart* dalam pencarian sebuah nilai PCA adalah sebagai berikut (Robert Hoo, 2003: 15):



Gambar 2.1 flow chart PCA

a. Memasukan *image* kedalam sebuah matriks

Langkah awal dalam proses pengenalan wajah adalah dengan mengolah image mentah. Misal ada sebuah image dengan ukuran 100 x 100 piksel = 10.000.Kemudian matriks tersebut diubah menjadi 10.000 x m.

$$\mathbf{u} = \begin{bmatrix} u_{1,1} & u_{1,2} & \dots & u_{1,m} \\ u_{2,1} & u_{2,2} & \dots & u_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{10000,3} & u_{10000,2} & \dots & u_{10000,m} \end{bmatrix}$$

b. Mencari rata-rata *image*

Langkah selanjutnya setelah pengolahan image awal adalah mencari image *rata-rata* adalah rata-rata dari seluruh *image training* yang ada dalam database. Rumus untuk mencari rata-rata adalah sebagai berikut:

$$\tilde{\mathbf{u}} = \frac{1}{M} \sum_{n=1}^{M} u_{i,k}$$

Image rata-rata merupakan jumlah dari u_{i,k} dibagi dengan banyaknya *image* yang ada dalam database. Selanjutnya dari data diatas didapat sebuah matriks dengan ukuran 10.000 x 1. Berikut adalah *image* rata-rata yang didapat:



Gambar 2.2 Rata-rata image database

c. Mencari kovarian matriks PCA

Matriks kovarian adalah semua variasi yang memungkinkan yang diperoleh dari pasangan vector kolom (Jonathon Shlens, 2005:7). Untuk mencari kovarian matriks yaitu dengan cara mengalikan u dengan transposenya. Rumusnya adalah sebagai berikut:

$$C = u \times u'$$

Selanjutnya dilakukan pencarian nilai *eigen* sehingga berlaku rumus sebagai berikut:

$$C x v = \lambda x v$$

 λ merupakan matriks *eigen value* sedangkan ν adalah matriks *eigen vector*. λ dan ν adalah matriks berdimensi n x n yang mana n adalah jumlah piksel image.

Eigen value yang didapat kemudian diurutkan dari yang terbesar ke yang terkecil. Hasil dari matriks ini adalah matriks V berukuran 10.000 x 10.000.

d. Mencari feature PCA

Feature PCA merupakan informasi paling penting yang ada dalam sebuah image yang sebelumnya dilakukan pencarian nilai eigen vector. Feature PCA dapat dicari dengan cara mentransformasikan image asal ke dalam ruang wajah dengan persamaan berikut menggunakan (Robert Hoo, 2003:16):

$$f = \sum_{i=1}^{m} (I_i - \tilde{\mathbf{u}})^T \times \mathbf{V}$$

I merupakan data tiap piksel dari setiap training I, m adalah jumlah image training, V adalah matriks eigen vector sedangkan f adalah matriks berukuran m x 10.000.

Setelah mendapatkan *feature* setiap *image* maka selanjutnya adalah melakukan proses pengelompokan *clustering* menggunakan *Fuzzy C-Means* (FCM).

3. Fuzzy C-Means

Fuzzy clustering adalah salah satu teknik untuk menentukan cluster optimal

dalam suatu ruang vector yang didasarkan pada bentuk normal *Euclidian* untuk jarak antar *vector* (Kusumadewi dan Purnomo, 2004:83). Fuzzy c-means merupakan salah satu metode dari *fuzzy clustering*. *Fuzzy clustering* memperbolehkan satu bagian data dimiliki oleh dua atau lebih *cluster*.

Ada dua metode dasar dalam fuzzy clustering. Metode pertama disebut dengan fuzzy c-means. Metode ini dinamakan demikian karena dengan clustering ini akan dibentuk sebanyak *c-cluster* yang sudah ditentukan sebelumnya. Metode kedua adalah metode yang banyaknya cluster tidak ditentukan sebelumnya. Metode ini dinamakan (Kusumadewi, subtractive clustering 2004) atau fuzzy equivalence relation (Klir, 1995).

Fuzzy C-means (FCM) dikenalkan pertama kali oleh JIM BEZDEK pada tahun 1981 yang merupakan salah satu metode clustering menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau cluster dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Adapun konsep dari Fuzzy C-means (FCM) adalah sebagai berikut:

- 1. Penentuan pusat cluster yang menandai lokasi rata-rata untuk setiap *cluster*, dengan kondisi awal tidak akurat.
- 2. Tiap data memiliki derajat keanggotaan untuk masing-masing *cluster*.
- 3. Tiap perulangan yang didasarkan pada minimasi fungsi obyektif, pusat cluster dan nilai keanggotaan diperbaiki sehingga lokasi *cluster* bisa berada pada posisi yang benar.

Algoritma FMC secara lengkap diberikan sebagai berikut (Zimmerman,1991);(Yan,1994);(Ross,200 5):

1. Tentukan:

a. Matriks X berukuran n x m, dengan n=jumlah data yang akan di*cluster*, dan m adalah jumlah variable criteria.

- b. Jumlah *cluster* yang akan dibentuk $(C \ge 2)$.
- c. Pangkat (pembobot w > 1)
- d. Maksimum iterasi
- e. Criteria penghentian (ε = nilai positif yang sangat kecil)
- 2. Bentuk matriks partisi awal U (derajat keanggotaan dalam *cluster*), matriks awal biasanya diberikan secara acak

$$U = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1_n}(x_n) \\ \mu_{11}(x_1) & \mu_{21}(x_2) & \cdots & \mu_{2_n}(x_n) \\ \vdots & & & \vdots \\ \mu_{2}(x_1) & \mu_{c1}(x_2) & \cdots & \mu_{c_n}(x_n) \end{bmatrix}$$

3. Hitung pusat Cluster V untuk setiap cluster

$$V_{kj} = \frac{\sum_{j=1}^{n} ((\mu_{ik})^{w} * x_{ij})}{\sum_{i=1}^{n} (\mu_{ik})^{w}}$$

4. Perbaiki derajat keanggotaan setiap data pada setiap *cluster* (perbaiki matriks partisi)

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(x_{ij} - V_{kj})^{2}\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(x_{ij} - V_{kj})^{2}\right]^{\frac{-1}{w-1}}}$$

 Tentukan criteria penghentian iterasi, yaitu perubahan matriks partisi pada iterasi sekarang dan iterasi sebelumnya.

$$\Delta = \left\|U^{\tau} - U^{\tau-1}\right\|$$

Apabila $\Delta < \epsilon$ maka iterasi dihentikan.

Pengenalan Wajah

Cara pengenalan wajah dapat diilustrasikan sebagai berikut:

- 1. Citra database akan diekstrasi dan dihitung nilai PCA nya.
- 2. Hitung derajat keanggotaan dengan rumus sebagai berikut:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m} (x_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c} \left[\sum_{j=1}^{m} (x_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}}$$

- 3. Cari derajat keanggotaan tertinggi untuk menentukan kecenderungan terkuat suatu wajah masuk kedalam suatu *cluster*.
- 4. Didapatkan *cluster* pusat dan wakil dari masing-masing *cluster*.

Untuk pengenalan wajah data *testing* dapat dilakukan sebagai berikut:

- 1. Data *Testing* akan dihitung nilai PCAnya.
- 2. Hitung derajat keanggotaannya dengan pusat *cluster* dengan rumus sebagai berikut:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{1} (x - V_{kj})^2\right]^{\frac{-1}{2-1}}}{\sum_{i=1}^{3} \left[\sum_{j=1}^{1} (x - V_{kj})^2\right]^{\frac{-1}{2-1}}}$$

- 3. Tentukan kecenderungan data masuk kedalam suatu *cluster*.
- 4. Tampilkan hasil kecocokan wajah. Dari pengenalan wajah dapat nilai akurasi dengan perhitungan sebagai berikut:

$$akurasi\ relatif$$

$$= \frac{jumlah\ benar}{jumlah\ seluruh\ data}\ x\ 100\%$$

Pengujian

Dalam penelitian ini terdapat beberapa skenario yang akan dilakukan dalam pengujian. Pengujian ini dimaksudkan untuk mengetahui nilai akurasi sistem pengenalan wajah menggunakan PCA dan FCM. Berikut langkah-langkah yang akan dilakukan dalam pengujian:

a. Pembuatan data latih menggunakan parameter yang telah dilakukan sebelumnya oleh Ahmad Ridwan dengan judul skripsi "Pengenalan Tulisan Tangan Aksara Sunda Ngalagena Dengan Algoritma Fuzzy C-Means" dengan ketentuan sebagai berikut:

1. Jumlah *cluster* : 30

2. Epsilon : 0,01

3. Iterasi : 100 4. Fuzzy : 2

- b. Pengujian dilakukan dengan menggunakan dataset yang telah tersedia baik dataset dari hasil *capture webcam* ataupun dataset hasil *download* dari *internet* (http://cswww.essex.ac.uk/mv/allfaces/faces94.zip).
- c. Jumlah dataset yang digunakan adalah 150, 300, 450 dataset.
- d. Pengujian dilakukan sebanyak dua jenis yaitu, pengujian normal dan penggujian menggunakan tambahan noise.
- e. Pengujian dilakukan dengan mencoba memasukkan citra wajah yang berbeda-beda sebanyak 10 jenis wajah, setiap jenis wajah terdapat 5 wajah yang berbeda-beda yang masing-masing wajah akan dicoba dimasukan ke dalam sistem sebanyak 5 kali.
- Pengujian menggunakan tambahan noise dilakukan dengan mencoba memasukkan citra wajah yang berbeda-beda sebanyak 10 wajah, setiap jenis wajah terdapat 5 berbeda-beda wajah yang masing-masing wajah akan dicoba dimasukan ke dalam sistem sebanyak 5 kali.

Hasil Pengujian

Hasil pengujian dapat dilihat dalam tabel dibawah ini:

Tabel 6.1 Hasil Pengujian

No	Penguj ian	Jumlah Data		Pengenalan		Akur
140		Traini	Uji	Ben	Sala	asi
		ng	Cji	ar	h	
1.	1	150	50	42	8	84%
2.	2	300	50	38	12	76%
3	3	450	50	37	13	74%
			Rata-rata		78%	

Dari hasil percobaan pertama diatas dengan data *training* sebanyak 150 data didapat 84% kemudian pada percoaan kedua dengan data *training* sebanyak 300

data didapat 76%, sedangkan pada percobaan terakhir dengan data *training* 450 data didapat 74%. Dari semua percobaan diatas maka dapat rata-rata persentase sebesar 78%. Dari hasil penelitian diatas dapat terlihat bahwa dengan data *training* semakin banyak tingkat akurasi semakin menurun.

Tabel 6.1 Hasil Pengujian

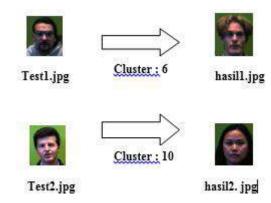
No	Penguj ian	Jumlah Data		Pengenalan		Akur
140		Traini	Uji	Ben	Sala	asi
		ng	Oji	ar	h	
1.	1	150	50	13	37	26%
2.	2	300	50	7	43	14%
3	3	450	50	4	46	8%
				Rata	16%	

Dari hasil percobaan diatas pertama dengan bantuan *noise* dan data *training* sebanyak 150 data didapat 26% kemudian dengan data *training* 300 data didapat 14%, sedangkan dengan data *training* sebanyak 450 data didapat 8%. Dari semua percobaan dengan bantuan *noise* didapat persentase sebesar 16%, hal ini menunjukan bahwa penggunaan *noise* sangat mempengaruhi tingkat akurasi.

Analisis Hasil Penelitian

Dari hasil penelitian langsung dengan menggunakan citra wajah berbeda, didapat nilai akurasi 78% dan untuk tambahan noise didapatkan nilai akurasi 16%. Penggunaan noise sangat mempengaruhi pada nilai akurasi sistem pengenalan wajah.

Beberapa kesalahan yang didapat dari penelitian diatas, misalnya:



Dari data contoh kesalahan yang diambil misalnya image test.jpg menempati cluster Hasil yang didapatkan tidak sesuai dengan yang diinginkan, hal ini disebabkan cluster 6 diwakili oleh image hasil.jpg. Begitupun dengan contoh kesalahan yang kedua yaitu image test2.jpg menempati cluster yang salah yaitu cluster 10 yang diwakili oleh image hasil2.jpg. Beberapa faktor yang dapat menyebabkan kesalahan dalam menempati cluster diantaranya:

- a. Jumlah data *training* yang dapat menyebabkan banyaknya variasi wajah.
- b. Adanya gangguan pada citra wajah *input*-an yang dapat menyebabkan nilai wajah berubah.

Kesimpulan

Beberapa kesimpulan yang dapat diambil dari penelitian yang telah dilakukan diatas diantaranya :

- 1. Principal Compnent Analysis (PCA) dan algoritma Fuzzy C-Means telah mampu mengenali citra wajah dengan akurasi relatif 86,16%. Nilai akurasi sangat dipengaruhi oleh fitur ekstraksi PCA dan parameter FCM.
- 2. Dari hasil penelitian diatas didapat rata-rata persentasi kebenaran dengan pengujian normal yaitu sebesar 86,16%, dengan bantuan noise sebesar 40,50%, dengan

- menggunankan *webcam* yaitu sebesar 80,66%, dan menggunakan *webcam+noise* sebesar 8,66%.
- 3. Keakuratan sistem pengenalan wajah ini sangat dipengaruhi oleh citra *image* yang masuk hal ini terlihat dengan penggunaan *noise*.

Saran

Beberapa saran yang dianjurkan untuk penelitian lebih lanjut dalam sistem pengenalan wajah ini adalah sebagai berikut:

- 1. Untuk menghasilkan nilai akurasi yang lebih tinggi disarankan menambah atau mengganti fitur ekstrasi dengan fitur ekstrasi yang lebih akurat.
- Dapat menggunakan metode lain sebagai bahan perbandingan dalam penelitian ini misalnya : KNN, ANN dan lain sebagainya.

Daftar Pustaka

Annoymous. *A Tutorial Clustering Algorithm*. Online.Tersedia di :http://home.dei.polimi.it/matteucc/C lustering/tutorial_html/cmeans.html. [5 September 2010]

Fatta, Al, Hanif. 2009. Rekayasa Sistem
Pengenalan Wajah. Yogyakarta:
Andi Luthfi, Emha
Taufiq. 2007. Fuzzy C-Means Untuk
Clustering Data. Yogyakarta:
STMIK AMIKOM

Gunadi, Kartika. Posngsitanan, Sonny Reinard.2001. Pembangunan Perangkat Lunak Pengenalan Wajah Menggunakan Principal Component Analysis. Jurnal Informatika Vol. 2 No. 2. Surabaya: Universitas Kristen Petra.

- M.Turk, A Pentland.1991. *Eigenface for Recognition*. Journal of Cognitive Neuroscience. 71-86
- Ridwan, Achmad. 2010. Pengenalan Tulisan Tangan Aksara Sunda Ngalagen Dengan Algoritma Fuzzy C-Means. Skripsi tidak terpublikasi. Bandung: Universitas Pendidikan Indonesia.
- Shalens, Jonathan. 2005. A Tutorial
 Principal Componen
 Analysis. Sandiego La Jolla:
 University of California.
- Smith, Lindsay I.2002. *A Tutorial Principal Componen Analysis*. Online. Tersedia di : www.cs.otago.ac.nz/cosc453/...tutori als/principal_components.pdf . [19 Juni 2010]
- Trivedi, Shubhendu.2009. Face
 Recognition using Eigenfaces and
 Distance Classifiers:
 A Tutorial.Online.Tersedia di:
 http://onionesquereality.wordpress.c
 om/.[20 Juni 2010]
- Yudha.2009. *Biometric*.Online.Tersedia di : http://nuxer1.multiply.com/ [20 Juni 2010].