

Pengenalan Wajah Wanita Berkerudung Menggunakan Metode 2DPCA dan *K-Nearest Neighbor*

Endra Budi Utomo

*Program Studi Teknik Informatika - S1, Fakultas Ilmu
Komputer, Universitas Dian Nuswantoro Semarang*

URL : <http://dinus.ac.id/>

Email : 111201106232@mhs.dinus.ac.id

Abstract—Research on face recognition have been carried out by various methods, but research is still interesting to do with the development of technologies and methods used in image processing. The Indonesian state many female citizens who wear the veil. This is certainly going to be a problem in the process of face recognition, because the veil is used mostly have the same models. Area faces become narrower because around his face covered by a veil. Forms of expression and lighting conditions are different in the image of course become increasingly complex problems in the process of face recognition. In this study, researchers used a method of feature extraction Two-Dimensional Principal Component Analysis (2DPCA). The data used is the face of a woman with a veil that some models measuring 180 x 200 pixels. Where previously will do the conversion process to form grayscale. After testing the feature extraction 2DPCA facial image by using K-Nearest Neighbor (KNN) based methods Euclidian Distance. From the results of a series of tests that have been carried out it was found that her face veiled women with 2 models veil have a more difficult level compared with women who wear the face veil 1 models. This is because they are due veil different models will produce different characteristics, so that the recognition process will be more difficult.

Index Terms—Face Recognition, Veiled Women, 2DPCA, KNN

I. LATAR BELAKANG

Pada zaman modern sekarang ini perkembangan teknologi komputer sekarang sangat pesat, hal ini ditandai dengan hampir semua pengolahan data dan informasi telah dilakukan dengan menggunakan komputer. Dengan demikian maka akan semakin banyak timbul permasalahan informasi yang harus ditangani, salah satu adalah dalam hal pengenalan pola.

Salah satu dari pengenalan pola yang penting dalam dunia usaha adalah pengenalan wajah. Sekarang ini semakin banyak pengaplikasian teknologi pengenalan wajah, diantaranya adalah pada sistem absensi, verifikasi daftar penjahat di kepolisian, pemrosesan Visa, dan verifikasi Paspor.

Pengenalan wajah digunakan untuk membandingkan satu citra wajah masukan dengan suatu database wajah dan menghasilkan wajah yang paling cocok dengan citra tersebut jika ada. Pada kasus tertentu seperti pemotretan untuk pembuatan KTP, SIM, dan kartu kredit, citra yang

dihasilkan biasanya hanya berisi satu wajah dan memiliki latar belakang seragam dan kondisi pencahayaan yang telah diatur sebelumnya sehingga deteksi wajah dapat dilakukan dengan lebih mudah. Namun pada kasus lain sering didapatkan citra yang berisi lebih dari satu wajah, memiliki latar belakang yang bervariasi, kondisi pencahayaan yang tidak tentu, dan ukuran wajah yang bervariasi di dalam citra. Sebagai contoh citra yang diperoleh di stasiun, tempat parkir, pintu masuk ruangan, dan pusat perbelanjaan. Selain itu juga pada citra yang didapatkan dari foto di media massa atau hasil rekaman video. Pada kasus tersebut pada umumnya wajah yang ada di dalam citra memiliki bentuk latar belakang yang sangat bervariasi.

Di negara Indonesia yang mayoritas penduduknya beragama Islam banyak sekali warga berjenis kelamin perempuan yang mengenakan kerudung. Ini sangatlah menjadi masalah dalam proses pengenalan wajah, hal ini dikarenakan kerudung yang digunakan kebanyakan mempunyai model sama. Area wajah yang menjadi lebih sempit karena sekeliling wajah tertutup oleh kerudung. Bentuk ekspresi yang berbeda-beda pada saat proses pengambilan foto. Pada kasus ini kondisi pencahayaan citra yang digunakan juga beragam. Dengan demikian

permasalahan pada proses pengenalan wajah akan menjadi semakin kompleks.

II. TINJAUAN PUSTAKA

Dalam melakukan penelitian ini penulis menggunakan jurnal nasional dan internasional sebagai referensi. Berdasarkan tabel *state of the art* yang telah dibuat penelitian pengenalan wajah perempuan berkerudung yang akan dilakukan nantinya akan digunakan metode *Two-Dimensional Principal Component Analysis* (2DPCA) yang digabungkan dengan algoritma klasifikasi *K-Nearest Neighbor* (KNN). Metode ini dipilih karena memiliki tingkat akurasi yang cukup tinggi dan waktu proses yang lebih cepat [2][5][6][7][8][9].

A. Pengolahan Citra Digital

Pengolahan citra adalah sebuah disiplin ilmu yang mempelajari hal-hal yang berkaitan dengan perbaikan kualitas gambar (peningkatan kontras, transformasi warna, restorasi), transformasi gambar (rotasi, translasi, skala, transformasi geometrik), melakukan pemilihan ciri citra (*feature extraction*) yang optimal untuk bertujuan analisis, melakukan proses penarikan informasi atau deskripsi objek atau pengenalan objek yang terkandung pada citra, melakukan kompresi atau reduksi data untuk tujuan penyimpanan data, transmisi data dan waktu proses data. *Input* dari pengolahan citra adalah citra, sedangkan *output*-nya adalah citra hasil pengolahan [10].

B. Definisi Citra Digital

Citra yang ditangkap oleh kamera dan telah dikuantisasi dalam bentuk nilai diskrit disebut sebagai citra digital (*digital image*) [1]. Sedangkan foto hasil cetak dari printer tidak dapat disebut sebagai citra digital, namun foto yang terimpan pada file gambar (bmp, jpg, png atau format lainnya) pada komputer dapat disebut sebagai citra digital.

C. Format File Citra

Format file citra standar yang digunakan saat ini terdiri dari beberapa jenis. Format-format ini digunakan dalam penyimpanan citra dalam sebuah file. Setiap format memiliki karakteristik masing-masing [11].

D. Face Recognition (Pengenalan Wajah)

Pengenalan seseorang dengan menggunakan citra wajah merupakan karakteristik biometrika yang paling umum digunakan oleh manusia. Pendekatan yang paling umum untuk pengenalan wajah didasarkan pada bentuk dan penempatan atribut wajah, seperti mata, alis, hidung, bibir dan dagu serta hubungan antara atribut tersebut atau analisis wajah secara keseluruhan yang menghadirkan suatu wajah sebagai suatu kombinasi dari sejumlah wajah kumpulan orang [12].

E. Two-Dimensional Principal Component Analysis (2DPCA)

Pada proses pengenalan wajah representasi gambar dan ekstraksi fitur merupakan teknik yang umum digunakan. Pada PCA matriks dari suatu gambar diubah menjadi

matriks vektor berdimensi tinggi, yang bermanfaat untuk menghitung matriks kovarians ruang vektor dimensi tinggi. Namun kendala utama yang terjadi adalah bahwa matriks kovarians menjadi ukuran yang besar, dimana akan mengakibatkan sejumlah besar pelatihan dengan sampel kecil akan sulit dalam melakukan evaluasi secara akurat. Selain itu dibutuhkan banyak waktu untuk menghitung vektor eigen berikutnya. Untuk mengatasi ini kesulitan *Two-Dimensional Principal Component Analysis* (2DPCA) memberikan cara untuk menangani keterbatasan ini [8].

Metode 2DPCA dapat didefinisikan sebagai berikut [2]:

Jika A merupakan matriks citra asli berukuran $m \times n$ dan X merupakan matriks dengan kolom *orthonormal*, maka 2DPCA merupakan proyeksi A pada X dengan persamaan:

$$Y = AX \quad (2.1)$$

Cara menentukan matriks X optimal untuk proyeksi citra adalah sebagai berikut. Terdapat persamaan :

$$J(X) = \text{trace}(Sx) \quad (2.2)$$

Sx merupakan matriks kovariansi dari proyeksi vektor fitur dari data training, $\text{trace}(Sx)$ merupakan $\text{trace} Sx$. Secara matematis Sx didefinisikan sebagai:

$$SX = E[(YE(Y))(Y-E(Y))^T] \quad (2.3)$$

Maka dari persamaan (2) dan (3) dapat dibuat:

$$J(X) = \text{trace} \{E[(Y-E(Y))(Y-E(Y))^T]\} \quad (2.4)$$

$$J(X) = \text{trace} \{E[(AX - E(AX))(AX - E(AX))^T]\}$$

$$J(X) = X^T E[(A - E(A))^T (A - E(A))] X$$

Berikutnya dimisalkan G yaitu matriks kovariansi citra definit non negatif berukuran $n \times n$ sebagai berikut:

$$G = E[(A - E(A))^T (A - E(A))] \quad (2.5)$$

Maka persamaan (4) dapat diubah menjadi:

$$J(X) = X^T G X \quad (2.6)$$

Jika terdapat sebanyak M citra data latih, dengan merupakan rata-rata semua citra latih sebagai berikut:

$$\bar{A} = \frac{1}{M} \sum_{k=1}^M A_k \quad (2.7)$$

Maka G dapat dihitung sebagai:

$$G = \frac{1}{M} \sum_{k=1}^M (A_k - \bar{A})^T (A_k - \bar{A}) \quad (2.8)$$

Pada percobaan, tiap-tiap citra dapat di-shift terlebih dahulu dengan dikurangi mean citra itu sendiri, sehingga nilai $\bar{A} =$

$0_{(m \times n)}$. Dimisalkan $X_{optimal}$ yaitu merupakan himpunan matriks proyeksi optimal. Berdasarkan persamaan (6):

$$\begin{aligned} X_{optimal} &= \arg \max J(X) \\ X_{optimal} &= \arg \max (X^T G X) \end{aligned} \quad (2.9)$$

Secara intuitif didapatkan $X_{optimal}$ merupakan himpunan sejumlah d eigenvektor dari G berkorespondensi dengan sejumlah d eigenvalue tertinggi.

$$X_{optimal} = \{X1, X2, \dots, Xd\} \quad (2.10)$$

Maka hasil ekstraksi fitur citra A menggunakan 2DPCA adalah $Y_{optimal}$.

$$\begin{aligned} Y_{optimal} &= A X_{optimal} \\ Y_{optimal} &= \{Y1, Y2, \dots, Yd\} \end{aligned} \quad (2.11)$$

F. Klasifikasi

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [13]. Terdapat berbagai macam algoritma klasifikasi yang banyak digunakan secara luas, diantaranya adalah *Decision/classification trees*, *Bayesian classifiers/Naïve Bayes classifiers*, *Neural networks*, *Analisa Statistik*, *Algoritma Genetika*, *Rough sets*, *k-nearest neighbor*, *Metode Rule Based*, *Memory based reasoning*, dan *Support vector machines (SVM)*.

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi. Proses klasifikasi didasarkan pada empat [14]:

1. Kelas

Variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa.

2. Predictor

Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.

3. Training dataset

Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

4. Testing dataset

Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi.

G. K-Nearest Neighbor (KNN)

K-Nearest Neighbor adalah contoh algoritma berbasis pembelajaran, di mana data set pelatihan (*training*)

disimpan, sehingga klasifikasi untuk record baru yang tidak diklasifikasi didapatkan dengan membandingkannya dengan record yang paling mirip dengan *training set* [15].

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya *euclidean distance* dan *manhattan distance (city block distance)*, yang paling sering digunakan adalah *euclidean distance* [13]. Selain daripada itu *euclidean distance* mempunyai hasil akurasi paling baik dibandingkan dengan metode berdasarkan jarak pada klasifikasi *K-Nearest Neighbor* [16]. Berikut merupakan persamaan yang digunakan dalam mencari jarak terdekat dengan *euclidean distance*.

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.12)$$

Dimana matriks d adalah jarak skalar dari kedua vektor x dan y dari matriks dengan ukuran n dimensi.

Pada fase *training*, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi data *training sample*. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk testing data (yang klasifikasinya tidak diketahui). Jarak dari vektor baru yang ini terhadap seluruh vektor training sample dihitung dan sejumlah k buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut.

III. IMPLEMENTASI

A. Jenis dan Sumber Data

Dalam pengumpulan data penulis menjumpai dan mengambil gambar (foto) secara langsung wajah wanita berkerudung. Pada penelitian ini nantinya akan dilakukan dengan menggunakan beberapa varian kerudung. Berikut merupakan dataset wajah wanita berkerudung yang akan digunakan dalam penelitian:

1. Wajah wanita berkerudung sebanyak 40 orang, dimana setiap orang diambil 10 foto dengan ekspresi wajah yang berbeda-beda.



Gambar 3.1 Wajah wanita berkerudung 1 model

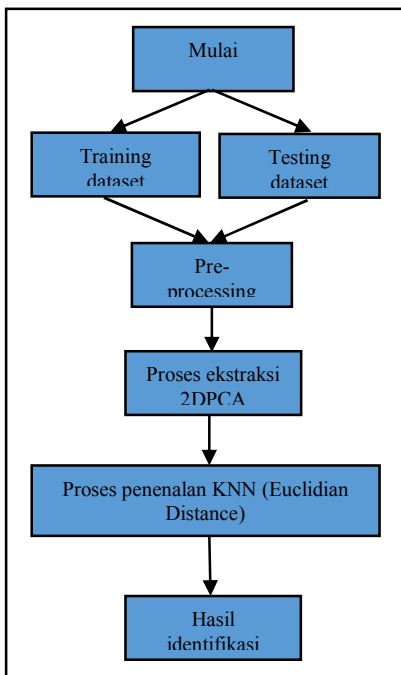
2. Wajah wanita berkerudung yang sebelumnya diambil langsung sebanyak 7 orang, dimana setiap orang diambil 10 foto dengan 2 model kerudung pada setiap orangnya.



Gambar 3.2 Wajah wanita berkerudung 2 model

B. Proses Identifikasi Wajah

Metode yang digunakan penulis dalam melakukan proses pengenalan wajah ini adalah dengan menggunakan penggabungan dari dua metode yang sudah ada yaitu metode *Two-Dimensional Principal Component Analysis* (2DPCA) dan *K-Nearest Neighbor* (KNN). Kedua metode ini digunakan untuk dapat mengidentifikasi wajah wanita berkerudung melalui inputan data *testing* yang kemudian diekstrak sehingga menghasilkan pola atau ciri tertentu yang nantinya akan dicari dan dicocokkan dalam data *training*. Berikut adalah gambar alur proses pengenalan wajah.



Gambar 3.3 Diagram alir proses identifikasi wajah

Tahap proses pengenalan wajah adalah sebagai berikut:

1. Pre-processing

Pre-processing merupakan proses konversi citra dari RGB ke *grayscale*. Konversi citra kedalam bentuk *grayscale* pada Matlab menggunakan fungsi *rgb2gray* seperti yang digunakan pada penelitian ini.



Gambar 4.4 Contoh citra wajah digital yang sudah mengalami pre-processing

2. Proses ekstraksi 2DPCA

Citra wajah digambarkan dalam matrik $A_{m \times n}$ kemudian ditransformasikan ke dalam bentuk $Y = AX$ sehingga diperoleh vektor dengan dimensi m yang merupakan proyeksi vektor ciri dari citra wajah A . Kemudian didefinisikan matrik G_t , yaitu:

$$G_t = E[(A-EA)^T(A-EA)] \quad (3.1)$$

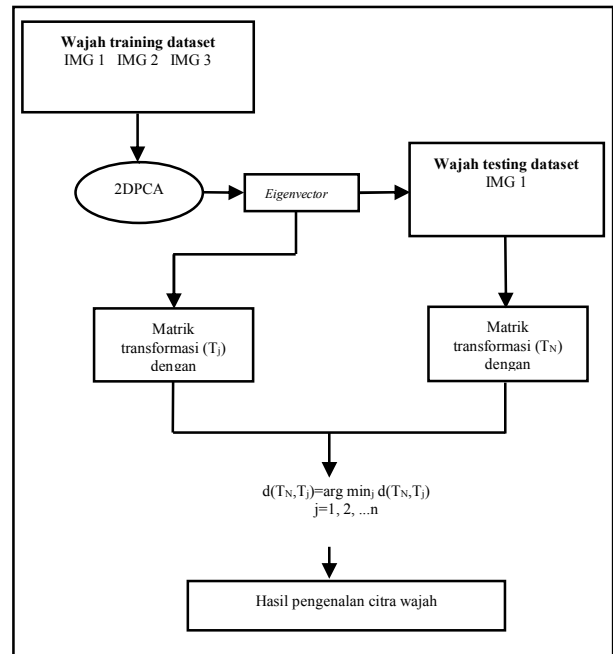
Matrik G_t disebut sebagai matrik kovarian citra wajah dan matrik G_t ini dievaluasi dengan citra wajah dari training dataset keseluruhan. Pada citra wajah ke- j dinotasikan sebagai matrik $A_{j(m \times n)}$ dimana $j=1, 2, \dots, M$, serta rata-rata dari seluruh training dataset dinotasikan dengan \bar{A} . Untuk G_t dievaluasi dengan rumus:

$$G_t = \frac{1}{M} \sum_{j=1}^M (A_j - \bar{A})^T (A_j - \bar{A}) \quad (3.2)$$

Diperoleh sejumlah proyeksi vektor X_1, \dots, X_d yang disebut komponen utama dalam bentuk vektor dari citra wajah A . Setiap komponen utama 2DPCA merupakan vektor, dan vektor komponen utama yang digunakan untuk membuat matrik $T_{m \times d} = [Y_1, \dots, Y_d]$ yang disebut sebagai matrik ciri dari citra wajah A .

3. Proses pengenalan KNN (*Euclidian Distance*)

Pada proses klasifikasi *K-Nearest Neighbor* (KNN) ini penulis menggunakan metode *euclidian distance*. Matrik ciri citra wajah yang diperoleh dari transformasi dengan menggunakan 2DPCA, dimana *training dataset* yang diekstraksi akan menghasilkan *eigenvector* yang digunakan untuk mendapatkan *eigenface*.



Gambar 5.5 Metode KNN dengan *euclidian distance* untuk pengujian

Didefinisikan jarak antar matrik $T_i = [Y_1^i, Y_2^i, \dots, Y_d^i]$ dan $T_j = [Y_1^j, Y_2^j, \dots, Y_d^j]$ sebagai berikut:

$$d(T_i, T_j) = \sum_{k=1}^d |Y_k^{(i)} - Y_k^{(j)}|_2 \quad (3.3)$$

Dimana $|Y_k^{(i)} - Y_k^{(j)}|_2$

$$\sqrt{(y^{1i} - y^{1j})^2 + \dots + (y^{1i} - y^{1j})^2} \quad (3.4)$$

Berdasarkan data minimum jarak *euclidian* pada *training dataset* T_1, T_2, \dots, T_M (M = jumlah *training dataset*), kemudian setiap data diberi identitas klasifikasi ω_k untuk *training dataset*. Apabila suatu testing dataset dengan matrik transformasi T_N , maka $d(T_N, T_j) = \arg \min_j d(T_N, T_j)$ dengan $j = 1, 2, \dots, n$, sehingga hasilnya $T_N \in \omega_k$.

C. Pengukuran Akurasi dan Validasi

Untuk mengevaluasi metode-metode yang telah diterapkan digunakan suatu metode tersendiri yaitu metode *K-fold cross validation*. Metode ini dipilih karena tidak membutuhkan waktu yang banyak dalam membuat variasi data uji dalam pengujian dan lebih mudah daripada metode *leave-one-out cross validation* [17]

K-fold cross validation digunakan dalam metode pembelajaran pada data *training* untuk menentukan keakuratan suatu algoritma sehingga dapat menguji data *testing*. Dengan metode ini data dikelompokkan menjadi k kelompok.

Berikut merupakan langkah-langkah dalam pengelompokan data:

1. Membagi data yang ada menjadi k kelompok.
2. Untuk setiap k , buat T himpunan data yang memuat semua data *training* kecuali yang berada di kelompok ke- k .
3. Kerjakan algoritma yang dimiliki dengan sejumlah T data *training*.
4. Uji algoritma ini dengan menggunakan data pada kelompok " k " sebagai data *testing*.
5. Lakukan pencatatan hasil algoritma.

Sebagai contoh terdapat 10 data, dimana akan digunakan 5 fold. Maka pembagian kelompok data ini menjadi:

- 1) Fold 1 : Citra 1 dan 2 sebagai data uji dan citra 3,4,5,6,7,8,9,10 sebagai data latih.
- 2) Fold 2 : Citra 3 dan 4 sebagai data uji dan citra 1,2,5,6,7,8,9,10 sebagai data latih.
- 3) Fold 3 : Citra 5 dan 6 sebagai data uji dan citra 1,2,3,4,7,8,9,10 sebagai data latih.
- 4) Fold 4 : Citra 7 dan 8 sebagai data uji dan citra 1,2,3,4,5,6,9,10 sebagai data latih.
- 5) Fold 5 : Citra 9 dan 10 sebagai data uji dan citra 1,2,3,4,5,6,7,8 sebagai data latih

Pada setiap fold akan di uji dan dihitung hasil akurasinya, selanjutnya hitung rata-rata seluruh akurasi untuk mendapatkan hasil akurasi secara keseluruhan. Tingkat akurasi dapat dihitung dengan rumus:

$$\text{Tingkat akurasi} = \frac{\sum \text{data uji benar diklasifikasi}}{\sum \text{total data uji}} \times 100\% \quad (3.5)$$

IV. HASIL & PEMBAHASAN

A. Hasil Pengujian

Berikut ini hasil dari beberapa pengujian yang telah dilakukan:

1. Pengujian pertama, menggunakan data *training* wajah wanita 1 model berkerudung dengan jumlah 40 sampai 360 dengan menggunakan semua *eigenvector* hasil ekstraksi.

Tabel 4.1 Hasil pengujian wajah wanita 1 model kerudung dengan semua *eigenvector*

No	Jumlah training	Jumlah testing	Bena r	Sala h	Akurasi	Waktu (s)
1	40	360	268	92	74.4444 %	3.4232
2	80	320	249	71	77.8125 %	3.6727
3	120	280	227	53	81.0714 %	4.5093
4	160	240	202	38	84.1667 %	5.406
5	200	200	171	29	85.5 %	6.6386
6	240	160	144	16	90 %	7.9726
7	280	120	111	9	92.5 %	9.49
8	320	80	74	6	92.5 %	11.4273
9	360	40	38	2	95 %	13.2484

Semakin banyak data training yang digunakan maka hasil akurasi yang didapatkan akan semakin tinggi akan tetapi waktu pemrosesan akan lebih lama. Dimana pada pengujian pada 360 data *training* dihasilkan akurasi 95 % dengan waktu proses 13.2484 detik.

2. Pengujian kedua, menggunakan data *training* wajah wanita 1 model berkerudung dengan jumlah 40 sampai 360 dengan menggunakan beberapa *eigenvector* hasil ekstraksi.

Selain penggunaan semakin banyak data *training* yang akan meningkatkan hasil akurasi, penggunaan jumlah nilai eigen dalam pengenalan juga mempengaruhi tingkat akurasi. Meskipun hasil akurasi yang dihasilkan tidak jauh signifikan, namun semakin banyak nilai eigen yang digunakan akan menambah waktu pemrosesan. Dimana pada pengujian pada 360 data *training* dihasilkan akurasi paling tinggi 97.5 % dengan penggunaan 2 nilai eigen didapat waktu proses 13.4942 detik lebih cepat dibandingkan dengan

penggunaan nilai eigen 4 dan 5 dengan waktu proses 14.123 detik dan 14.2723 detik.

- Pengujian ketiga, menggunakan data *training* wajah wanita 2 model berkerudung dengan jumlah 42 sampai 63 dengan menggunakan semua *eigenvector* hasil ekstraksi.

Tabel 4.2 Hasil pengujian wajah wanita 2 model berkerudung dengan semua *eigenvector*

No	Jumlah <i>training</i>	Jumlah <i>testing</i>	Benar	Salah	Akurasi	Waktu (s)
1	42	28	4	24	14.2857 %	1.1939
2	49	21	3	18	14.2857 %	1.2207
3	56	14	2	12	14.2857 %	1.2462
4	63	7	1	6	14.2857 %	1.3369

Didapatkan hasil akurasi yang sama pada setiap pengujian dengan beberapa jumlah data *training*. Hanya waktu pemrosesan yang semakin bertambah lama dengan bertambahnya jumlah data *training* yang digunakan.

- Pengujian keempat, menggunakan data *training* wajah wanita 2 model berkerudung dengan jumlah 42 sampai 63 dengan menggunakan beberapa *eigenvector* hasil ekstraksi.

Hasil akurasi yang dihasilkan yang dihasilkan tidak mengalami perubahan yang signifikan pada setiap jumlah data *training* yang digunakan. Akan tetapi penggunaan jumlah nilai eigen sangat berpengaruh pada waktu pemrosesan. Dimana pada pengujian pada 63 data *training* dihasilkan akurasi paling tinggi 14.2857 % dengan penggunaan 3 nilai eigen didapat waktu proses 1.3064 detik lebih cepat dibandingkan dengan penggunaan nilai eigen 4 dan 5 dengan waktu proses 1.3515 detik dan 1.3551 detik.

B. Analisis Hasil Pengujian

Dengan memaksimalkan jumlah data training pada setiap dataset, berikut merupakan hasil akurasi tertinggi dan waktu proses terkecil yang dihasilkan oleh masing-masing dataset:

Tabel 1.3 Hasil akurasi tertinggi dengan waktu terkecil dari masing-masing dataset

Dataset	Jumlah <i>training</i>	Jumlah <i>testing</i>	Jumlah <i>eigenvector</i>	Akurasi	Waktu (s)
Berkerudung 1 model	360	40	2	97.5 %	13.4942
Berkerudung 2 model	63	7	3	14.2857 %	1.3064

Semakin banyak jumlah data *training* yang digunakan maka tingkat akurasi yang dihasilkan akan semakin tinggi. Semakin berbeda model kerudung yang digunakan akan semakin sulit dikenali dan membutuhkan *eigenvector* yang lebih banyak dalam proses pengenalan. Tetapi semakin banyak jumlah *eigenvector* yang digunakan maka waktu pemrosesan maka akan menjadi semakin lama.

Berdasarkan hasil serangkaian pengujian yang telah dilakukan bahwa dengan menggunakan beberapa nilai eigen saja sudah dapat dihasilkan tingkat akurasi yang tinggi daripada menggunakan seluruh nilai eigen. Berikut merupakan validasi akurasi dataset dengan menggunakan *K-fold cross validation*.

Tabel 2.4 Hasil pengujian wajah wanita berkerudung dengan *k-fold cross validation*

Jumlah <i>eigenvector</i>	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Rata-rata
All	86.25 %	87.5 %	86.25 %	91.25 %	92.5 %	88.75 %
1	73.75 %	83.75 %	78.75 %	83.75 %	86.25 %	81.25 %
2	78.75 %	93.75 %	92.5 %	91.25 %	87.5 %	88.75 %
3	81.25 %	91.25 %	90 %	91.25 %	92.5 %	89.25 %
4	82.5 %	90 %	90 %	91.25 %	93.75 %	89.5 %
5	82.5 %	88.75 %	88.75 %	92.5 %	93.75 %	89.25 %

Akurasi rata – rata tertinggi yang dihasilkan adalah 89.25 % dengan penggunaan nilai eigen 3 dan 5 lebih tinggi dibandingkan dengan penggunaan semua nilai eigendengan akurasi 88.75 %..

V. PENUTUP

A. Kesimpulan

Dari hasil penelitian yang dilakukan dengan menerapkan metode *Two-Dimensional Principal Component Analysis* (2DPCA) dengan *K-Nearest Neighbor* (KNN) pada proses identifikasi wajah wanita berkerudung dengan varian model kerudung dapat diambil kesimpulan bahwa wajah wanita berkerudung dengan 2 model kerudung lebih sulit dikenali dibandingkan dengan wajah wanita berkerudung dengan 1 model kerudung. Hal ini dapat disebabkan karena selain ruang lingkup wajah yang semakin sempit karena tertutup oleh kerudung, bentuk model kerudung yang berbeda-beda juga mempersulit dalam identifikasi wajah, hal ini dibuktikan dengan hasil akurasi yang rendah pada pengenalan wanita berkerudung dengan 2 model kerudung.

B. Saran

Berdasarkan pada hasil analisis dan pembahasan dari hasil penelitian, terdapat beberapa saran untuk pengembangan penelitian wajah khususnya wajah wanita berkerudung:

1. Pengembangan software pengenalan wajah wanita berkerudung dengan menggunakan tools lainnya untuk hasil yang lebih akurat dalam mengidentifikasi dan meminimalisir kesalahan yang terjadi.
2. Pengembangan software dengan menggunakan gabungan dari beberapa metode selain *Two-Dimensional Principal Component Analysis* (2DPCA) dan *K-Nearest Neighbor* (KNN), dengan harapan didapat metode yang lebih baik dalam identifikasi wajah wanita berkerudung.

REFERENCES

- [1] M. H. Purnomo and A. Muntasa, *Konsep Pengolahan Citra Digital dan Ekstraksi Fitur*, Yogyakarta: Graha Ilmu, 2010.
- [2] B. Purnama dan D. Salsabila, "Identifikasi Telapak Tangan Menggunakan 2DPCA plus PCA dan K-nearest Neighbor," dalam *Seminar Nasional Aplikasi Informasi (SNATI)*, Yogyakarta, 2013.
- [3] Y. Zeng, D. Feng and L. Xiong, "An Algorithm of Face Recognition Based on the Variation of 2DPCA," *Journal of Computational Information Systems*, pp. 303-310, 2011.
- [4] P. Dhoke and M. Parsai, "A MATLAB based Face Recognition using PCA with Back Propagation Neural network," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 8, 2014.
- [5] Y. Elglaly and F. Quek, "Isolated Handwritten Arabic Characters Recognition using," *Computer Science Department, Virginia Polytechnic Institute and State University, USA*, pp. 1-6, 2012.
- [6] S. Fernandes and J. Bala, "Performance Analysis of PCA-based and LDA-based Algorithms for Face Recognition," *International Journal of Signal Processing Systems*, vol. 1, no. 1, pp. 1-6, 2013.
- [7] M. A. Kashem, M. Akhter, S. Ahmed and M. M. Alam, "Face Recognition System Based on Principal Component Analysis (PCA) with Back Propagation Neural Networks (BPNN)," *International Journal of Scientific & Engineering Research*, vol. 2, no. 6, pp. 1-10, 2011.
- [8] J. Shah, M. Sharif, M. Raza and A. Azeem, "A Survey: Linear and Nonlinear PCA Based Face," *Department of Computer Sciences, COMSATS Institute of Information Technology*, 2011.
- [9] C. Chen and K. Xie, "Face Recognition Based on Two-dimensional Principal Component Analysis and Kernel Principal Component Analysis," *Information Technology Journal*, vol. 11, no. 12, pp. 1781-1785, 2013.
- [10] T. Sutoyo, E. Mulyanto, D. V. Suhartono, O. D. Nurhayati and Wijanarto, *Teori Pengolahan Citra Digital*, Yogyakarta: Andi, 2009.
- [11] D. Putra, *Pengolaha Citra Digital*, Yogyakarta: Andi, 2010.
- [12] D. Putra, *Sistem Biometrika: Konsep Dasar, Teknik Analisis Citra dan Tahapan Membangun Aplikasi Sistem Biometrika*, Yogyakarta: Andi, 2009.
- [13] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, vol. 1, no. 1, pp. 65-76, 2013.
- [14] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, Verlag Berlin Heidelberg: Springer, 2011.
- [15] L. Hermawanti and S. G. Rabiha, "Penggabungan Algoritma Backward Elimination dan K-Nearest Neighbor Untuk Mendiagnosis Penyakit Jantung," in *Seminar Nasional Sains dan Teknologi ke-5*, Semarang, 2014.
- [16] M. Mustafa, M. N. Taib, Z. H. Murat and N. Sulaiman, "Comparison between KNN and ANN Classification in Brain Balancing Application via Spectrogram Image," *Journal of Computer Science & Computational Mathematics*, vol. 2, no. 4, pp. 17-22, 2012.
- [17] Pratiwi, "Metode Ekstraksi Ciri 2DPCA Pada Pengenalan Citra Wajah Dengan Matlab," *Jurnal Teknologi*, vol. 7, no. 1, pp. 1-5, 2014.