

Pengenalan Jenis Kelamin Berbasis *Kernel Principal Component Analysis*

Achmad Rizal

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo
Email : rizaliverpudlian@gmail.com

ABSTRAK

Gender Recognition adalah salah satu penelitian di bidang biometrik dan *computer vision* yang cukup populer. *Gender Recognition* adalah pengembangan dari *Face Recognition*, *Gender Recognition* dapat mengklasifikasikan citra menjadi 2 kelas yaitu perempuan dan laki-laki. Penelitian ini menggunakan 400 citra, 200 citra perempuan dan 200 citra laki-laki dan memakai *database JAVE* yang telah teruji sebelumnya. Ada 2 tahapan penting dalam penelitian ini, tahap pertama ekstraksi fitur menggunakan metode *Kernel Principal Component Analysis*. Metode kernel dapat membuat representasi data pada ruang kernel dengan menggunakan fungsi non-linear dan kernel yang digunakan adalah linier kernel. Tahap kedua adalah pengukuran jarak kemiripan citra *testing* terhadap citra *training* menggunakan metode *2D Correlation Coefficient* yang bekerja dengan cara mengalikan data *training* dan data *testing* kemudian membagi dengan hasil akar kuadratnya. Metode kernel berjalan cukup baik karena memperoleh akurasi lebih tinggi dari metode dari penelitian sebelumnya yang telah digunakan. Akurasi tertinggi yang dihasilkan pada penelitian ini mencapai 92%.

Kata Kunci : *Face Recognition, Gender Recognition, Kernel Principal Component Analysis, Correlation Coefficient.*

ABSTRACT

Gender Recognition is one of the research areas in the field of biometrics and *computer vision* that are quite popular. *Gender Recognition* is the development of *Face Recognition*. *Gender Recognition* can classify the image into two classes, namely women and men. This thesis used 400 images, 200 images and 200 images of women men and used the *JAVE* database that had been tested previously. There are two important stages in this research. The first phase is a feature extraction by using *Kernel Principal Component Analysis* method. Kernel methods can make a representation of data in kernel space using a non-linear function and the kernel used is a linear kernel. The second stage is the measurement of the distance to the testing image similarity of training images by using *2D Correlation Coefficient* method that works by multiplying the training data and testing data and then dividing them by the results of the square root. Kernel methods work quite well in which the accuracy is higher than previous methods of the research that have been used. Produced the highest accuracy in this study reached 92%.

Keywords: *Face Recognition, Gender Recognition, Kernel Principal Component Analysis, Correlation Coefficient.*

PENDAHULUAN

Jenis Kelamin atau *gender* pada manusia itu terbagi menjadi dua yaitu laki-laki dan perempuan. Manusia dengan mudah dapat mengklasifikasikan jenis kelamin dengan banyak cara. Contohnya dengan membedakan dari bentuk tubuh, cara berpakaian, dan juga dari bentuk wajah. Dari beberapa cara pengklasifikasian yang dapat dilakukan oleh manusia tersebut, cara mengenali dari bentuk wajah ini yang digunakan pada Penelitian

ini yang disini berbentuk citra. Wajah dapat dengan mudah dikenali dan merupakan fitur karakteristik dari manusia yang berupa identitas dan emosi [1]. Dan antara wajah laki-laki dan perempuan secara umum mempunyai ciri yang mencolok dan berbeda sehingga dapat dibedakan dengan mudah.

Dengan semakin pesatnya komputerisasi khususnya di bidang *Computer Vision* dan *Image Processing*, suatu sistem dapat mengenali jenis kelamin dari seseorang melalui citra wajah. *Gender Recognition* atau Pengenalan Jenis Kelamin berbasis citra wajah manusia adalah salah satu sistem yang mengimplementasikan cara kerja dari indera manusia. Jadi sistem mengambil obyek yang disini adalah citra wajah, kemudian citra tersebut diproses dan dilakukan ekstraksi fiturnya. *Gender Recognition* adalah pengembangan dari *Face Recognition* yang sudah lebih dulu ada. Perbedaan *Face Recognition* dan *Gender Recognition* terletak pada sistem klasifikasinya. *Gender Recognition* hanya mengklasifikasikan pada 2 kelas yaitu laki-laki dan perempuan.

Secara umum dalam teknik *Face Recognition* dan *Gender Recognition* mempunyai tiga macam pendekatan, yaitu: pendekatan holistik yang berdasarkan ciri secara keseluruhan citra, pendekatan *feature-based* yang berdasarkan ciri lokal seperti mata dan hidung, dan pendekatan *hybrid* yang berdasarkan gabungan dari holistik dan *feature-based*. Pada penelitian tentang pengenalan jenis kelamin berdasarkan citra wajah ini menggunakan pendekatan holistik. Pada pendekatan holistik seluruh bagian atau ciri global wajah digunakan sebagai data masukan untuk pengenalan [2].

Secara garis besar tujuan dari penelitian ini untuk membuat sistem dari masukan berupa citra wajah yang akan di cari tingkat kemiripannya di dalam *database*. Penelitian ini mempunyai dua tahapan yaitu pertama adalah ekstraksi fitur, pada tahap ini dilakukan untuk mendapatkan ekstraksi fitur atau reduksi dimensi agar nantinya mendapatkan fitur-fitur penting dari citra sehingga dari fitur itu dapat dibedakan antara satu citra dengan citra yang lain, istilahnya fitur yang penting dipertahankan dan fitur yang tidak penting dibuang atau dihapus. Tahap selanjutnya adalah menghitung kemiripan antara citra pelatihan dan citra *testing*.

Pada Penelitian ini metode yang digunakan adalah *Kernel Principal Component Analysis*. *Kernel Principal Component Analysis (KPCA)* merupakan metode pengembangan dari *Principal Component Analysis (PCA)*. Perbedaan antara *PCA* dan *KPCA* adalah pada *KPCA* dapat mereduksi dimensi data yang non-linier dan *non-separable*. Untuk pengukuran jarak kemiripannya digunakan *2D Correlation Coefficient*. *2D Correlation Coefficient* mengurangi nilai matrik *training* dengan rata-rata dari matrik *training* kemudian dikalikan dengan mengurangi nilai matrik *testing* dengan rata-rata dari matrik *testing*. Metode *kernel* dapat bekerja lebih optimal dari pada *PCA* untuk ekstraksi fitur. *PCA* tidak dapat melakukan klasifikasi data secara non-linier. *Kernel PCA*

merupakan *PCA* yang diaplikasikan pada *input* data yang telah ditransformasikan pada ruang fitur [3].

METODE

Kernel Principal Component Analysis

Kernel Principal Component Analysis merupakan metode pengembangan dari *Principal Component Analysis* yang bertujuan untuk memperbaiki proses ekstraksi fitur dari metode-metode sebelumnya yang sudah ada khususnya memproses data non-linier dan *non-separable*.

Perbedaan *Kernel Principal Component Analysis* dari *Principal Component Analysis* adalah *PCA* hanya dapat mereduksi dimensi data yang linier saja sedangkan *Kernel PCA* dapat mereduksi dimensi data non-linier dengan menggunakan fungsi *kernel* yang dilakukan pada awal komputasi. Fungsi *kernel* adalah fungsi yang memetakan dari *input space* menjadi *feature space*. Seperti ditampilkan pada Persamaan 1.

$$\begin{aligned}\phi &: R^n \rightarrow F \\ k(x,y) &= (\phi(x), \phi(y))\end{aligned}\quad (1)$$

Ada beberapa macam metode fungsi *kernel* yang telah diketahui sampai saat ini yaitu *kernel linier*, *kernel polinomial*, dan *kernel Gaussian* [4]. Pada penelitian ini menggunakan fungsi *kernel* linier yang dapat dilakukan dengan mengalikan matrik citra awal dengan *transpose* dari matrik citra awal tersebut. Berikut ini adalah langkah-langkah algoritma dari *Kernel Principal Component Analysis*.

1. *Input image*

Citra yang dimaksud adalah citra yang sudah tersimpan di *database*, jumlah citra yang diproses tergantung skenario yang dilakukan. Citra *database* mempunyai dimensi 200x200 kemudian citra dijadikan menjadi matrik baris sehingga dimensi citra menjadi 1x40.000. Jika ada 200 citra yang diproses maka dimensi menjadi 200x40.000. Ciri yang digunakan untuk pengenalan adalah seluruh bagian ciri global dari citra masukan dan tidak berdasarkan ciri lokal seperti hidung, rambut, mata, maupun lainnya.

2. Membentuk Matrik *Kernel*

Selanjutnya membentuk matrik awal citra menjadi matrik *kernel* menggunakan *kernel* linier dengan cara mengalikan matrik awal citra dengan *transposenya* seperti ditunjukkan pada Persamaan 2.

$$k(x,y) = X * Y \quad (2)$$

dengan:

$k(x,y)$ = matrik linear *kernel*

X = matrik awal dari citra

$Y =$ matrik *transpose* dari X

3. Menghitung *Kernel Mean*

Kernel Mean adalah nilai matrik rata-rata dari matrik *kernel* dan dapat dicari dengan menjumlahkan semua data dan dibagi sejumlah data. Selanjutnya matrik *kernel* dilakukan proses pengurangan dengan matrik *Kernel Mean* dan menghasilkan nilai matrik baru yang disebut *Kernel Zeromean*. seperti ditunjukkan oleh Persamaan 3.

$$\begin{aligned} \mu_i &= \frac{(x_{1,i} + x_{2,i} + x_{3,i} + \dots + x_{m,i})}{m} \\ &= [\mu_1, \mu_2, \dots, \mu_n] \end{aligned} \tag{3}$$

dengan:

$\mu_i =$ *mean* dari tiap piksel pada matrik baris dengan indeks i

$x =$ nilai piksel pada matrik citra dengan index m, i pada citra matrik baris

$m =$ banyaknya nilai piksel pada matrik citra dalam matrik baris

4. Menghitung *Kernel ZeroMean*

Menghitung *ZeroMean* dapat dilakukan dengan cara nilai dari setiap matrik *kernel* dikurangi *Kernel Mean* seperti ditunjukkan pada Persamaan 4.

$$\phi_{j,i} = x_{j,i} - \mu_i \tag{4}$$

Atau dapat dimodelkan menggunakan matrik seperti ditunjukkan pada persamaan 5.

$$= \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \dots & \dots & \dots & \dots \\ x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix} - \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \dots & \mu_{1,n} \\ \mu_{2,1} & \mu_{2,2} & \dots & \mu_{2,n} \\ \dots & \dots & \dots & \dots \\ \mu_{m,1} & \mu_{m,2} & \dots & \mu_{m,n} \end{bmatrix} \tag{5}$$

dengan:

$\phi =$ nilai *ZeroMean* dari tiap piksel matrik citra dengan indeks j, i

$x =$ nilai piksel pada matrik citra dengan index m, i pada citra matrik baris.

$\mu =$ *mean* dari tiap piksel pada matrik baris dengan indeks i

5. Menghitung Matrik Kovarian.

Matrik Kovarian terbentuk dari *Kernel Zeromean* dikurangi *Kernel Mean*. Selanjutnya Matrik Kovarian dapat dihitung dengan *Kernel Zeromean* yang dikalikan dengan *Tranpose* dari *Kernel Zeromean*, seperti ditunjukkan pada Persamaan 6.

$$C = \frac{1}{m} \sum_{j=0}^m \phi(x_j) \phi^T(x_j) \tag{6}$$

6. Menghitung *Eigen Value* dan *Eigen Vector*.

Proses selanjutnya adalah menghitung nilai *eigenvector* dan *eigenvalue* dari matrik Kovarian. Untuk mendapatkan ciri suatu data sampel yang direpresentasikan dalam bentuk matrik, seperti ditunjukkan pada Persamaan 7.

$$C v = \lambda v \tag{7}$$

Skalar λ disebut sebagai *eigenvalue* dari C dan disebut sebagai *eigenvector* dari C yang berpadanan terhadap λ . Untuk mendapatkan *eigenvector* dan *eigenvalue* seperti ditunjukkan Persamaan 8.

$$\begin{aligned} C \Lambda &= \lambda \Lambda \\ (\lambda I - C) \Lambda &= 0 \\ \text{Det} = (\lambda I - C) &= 0 \end{aligned} \quad (8)$$

Hasil dari Persamaan 8 adalah berupa vektor, yaitu *Eigenvalue* (λ) diurutkan secara menurun dari nilai paling besar menuju nilai yang paling kecil. Sedangkan *Eigenvector* menyesuaikan dengan nilai terbesar dari *Eigenvalue* yang mempunyai ciri paling dominan. Nilai *Eigenvector* yang bersesuaian dengan *Eigenvalue* terkecil mempunyai ciri paling tidak dominan [2].

7. Proyeksi Citra

$$\text{FinalData} = \text{Eigen Vector}' \times X \quad (9)$$

EigenVector' adalah matrik dengan kolom *Eigenvector* yang di *transpose*, dengan urutan *Eigenvector* yang signifikan adalah posisi paling atas. X adalah nilai awal matrik citra *training*.

8. Matrik Bobot

$$\text{Matrik Bobot} = X \times \text{FinalData}' \quad (10)$$

X adalah nilai awal dari citra *training*, sedangkan *FinalData'* adalah *transpose* dari final data atau matrik proyeksi.

2D Correlation Coefficient

Hasil dari ekstraksi fitur citra wajah dari proses *Kernel Principal Component Analysis* atau *Kernel PCA* digunakan sebagai acuan untuk mengukur kemiripan terhadap hasil ekstraksi fitur pada data ini. Pada Penelitian ini untuk mengukur jarak kemiripan antara citra *testing* dan citra *training* menggunakan metode *2-D Correlation Coefficient* [5], seperti ditunjukkan pada Persamaan 11.

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (11)$$

dengan :

r = nilai *correlation coefficient*

A = data *training*

B = data *testing*

A dan B merupakan matrik dengan ukuran yang sama

\bar{A} = rata – rata dari A

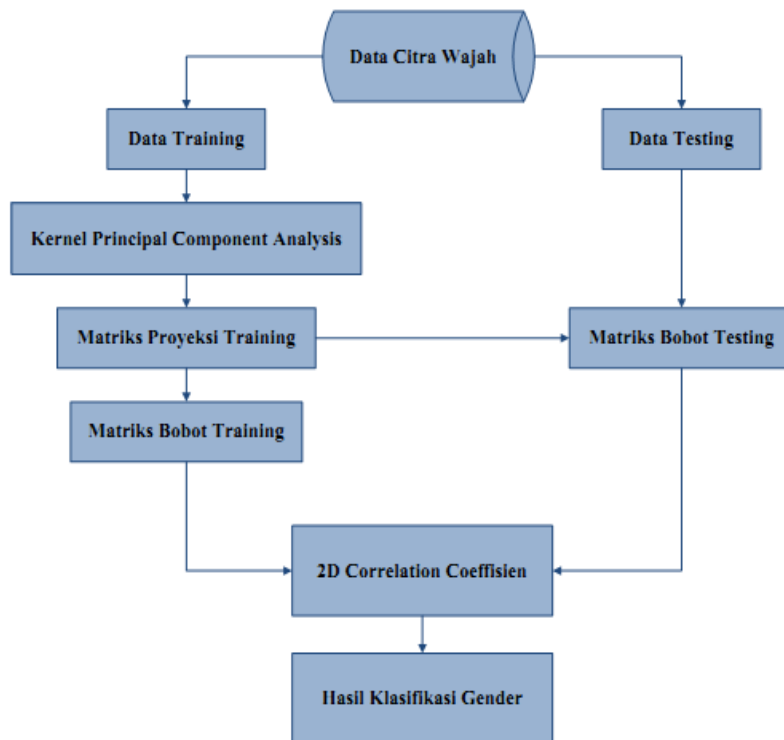
\bar{B} = rata – rata dari B

Σ_{mn} =jumlah vektor matrik sebanyak $m \times n$

Inti dari metode *2D Correlation Coefficient* adalah tiap nilai matrik *training* dikurangi nilai rata-rata matrik *training* selanjutnya dikalikan tiap nilai matrik *testing*, setelah itu dikurangi nilai rata-rata matrik *testing*. Kemudian dibagi akar dari tiap nilai matrik *training* dan dikuadratkan dikali tiap nilai matrik *testing* dikurangi nilai rata-rata matrik *testing* dan dikuadratkan juga, dan nilai yang paling maksimal yang diambil sebagai nilai paling mirip.

Flowchart Sistem

Pada Gambar 1 menunjukkan sistem yang dibuat. Citra masukan dibagi menjadi citra *training* dan citra *testing*. Selanjutnya citra *training* melalui proses ekstraksi fitur terlebih dulu menggunakan metode *Kernel Principal Component Analysis* untuk mendapatkan fitur-fitur yang dibutuhkan dan sudah mewakili dari nilai citra asal. Setelah itu citra *testing* langsung dikalikan dengan proyeksi *training* untuk mendapatkan bobot *testing*. Untuk pengukuran jarak kemiripan citra *testing* terhadap citra *training* menggunakan metode *2D Correlation Coefficient*. Setelah itu diperoleh hasil paling maksimal dari proses pengukuran jarak kemiripan.



Gambar 1. Rancangan umum sistem *Gender Recognition*

HASIL DAN PEMBAHASAN

Penerapan Skenario pada Data Uji Coba

Skenario uji coba dilakukan sebanyak empat kali. Jumlah data uji coba yang digunakan sebesar 400 citra, yaitu 200 citra perempuan dan 200 citra laki-laki. Tiap citra mempunyai ukuran dimensi yang sama yaitu 200x200. Data uji coba menggunakan *database JAVE* (<http://www.advancedsourcecode.com/>). Pada Gambar 2 merupakan contoh data uji coba yang dipakai.



Gambar 2. Contoh gambar citra *database* yang dipakai

Semua data citra uji coba tanpa perlu melakukan proses *preprocessing* saat *input* citra karena data uji coba sudah berbentuk *grayscale* dan mempunyai dimensi yang sama yaitu 200x200. Setiap skenario mempunyai jumlah citra *training* dan citra *testing* yang berbeda. Semua detail skenario uji coba ditunjukkan pada Tabel 1.

Tabel 1. Simulasi Skenario data uji coba pada sistem

Skenario		Training		Testing	
Skenario 1	200 citra	100 citra perempuan 100 citra laki-laki	200 citra	100 citra perempuan 100 citra laki-laki	
Skenario 2	250 citra	125 citra perempuan 125 citra laki-laki	150 citra	75 citra perempuan 75 citra laki-laki	
Skenario 3	300 citra	150 citra perempuan 150 citra laki-laki	100 citra	50 citra perempuan 50 citra laki-laki	
Skenario 4	350 citra	175 citra perempuan 175 citra laki-laki	50 citra	25 citra perempuan 25 citra laki-laki	

Hasil Uji Coba

Penelitian sebelumnya melakukan 2 macam ekstraksi fitur yaitu *Principal Component Analysis* yang dikombinasikan dengan *Local Binary Pattern (LBP)* dan ekstraksi fitur hanya menggunakan *PCA*. Hasil uji coba menyebutkan bahwa saat

dikombinasikan dengan *LBP*, *PCA* kurang bagus digunakan pada *Gender Recognition* [5]. Saat pengenalan menggunakan *Principal Component Analysis (PCA)* saja akurasi mempunyai nilai yang lebih tinggi. Hal ini pada data yang digunakan terdapat variasi pencahayaan yang tidak terlalu mencolok sedangkan *LBP* membuat variasi pencahayaan semakin cerah, dan hal itu membuat akurasi semakin rendah seperti ditunjukkan pada Tabel 2.

Tabel 2. Hasil penelitian sebelumnya [6].

Skenario		Akurasi					
		LBP-PCA			PCA		
		Benar	Salah	%	Benar	Salah	%
Skenario 1	100 Citra	80	20	80%	85	15	85%
Skenario 2	120 Citra	92	28	76,6667%	102	18	85%
Skenario 3	140 Citra	112	28	80%	120	20	85,7143%
Skenario 4	180 Citra	128	32	80%	139	21	86,8750%

Perbedaan *PCA* dan *KPCA* terletak pada matrik awal citra. Misalkan data yang diproses berjumlah 200 dan semua citra berdimensi 200x200, maka langkah pertama adalah merubah matrik citra menjadi matrik baris sehingga dimensi citra menjadi 200x40.000. Untuk *KPCA* proses awalnya sama merubah matrik citra menjadi matrik baris sehingga berdimensi 200x40.000. Selanjutnya membentuk matrik *kernel*, *kernel* linier dibentuk dengan mengalikan citra awal dengan *transposenya*. Matrik citra (200x40.000) * (40.000x200) sehingga matrik *kernel* berdimensi 200x200. Jika proses inialisasi disimbolkan dengan Ω , jumlah citra adalah m dan dimensi adalah n maka pembentukan dimensi awal *PCA* dan *KPCA* bisa dianalogikan seperti ditunjukkan pada Persamaan 13 dan Persamaan 14.

$$\Omega = m \times n^2 \quad (13)$$

$$\Omega = m \times n \quad (14)$$

Persamaan 13 adalah pembentukan awal citra *PCA* dan Persamaan 14 adalah pembentukan matrik *Kernel PCA (KPCA)* yang digunakan pada penelitian ini. *Kernel PCA* berhasil mereduksi dimensi matrik awal. *KPCA* juga berhasil meningkatkan tingkat akurasi jika dibandingkan dengan *PCA* [6].

Berdasarkan hasil uji coba, menunjukkan bahwa semakin banyak jumlah data citra *training* maka tingkat akurasinya juga meningkat. Pada Gambar 3 dan 4 merupakan contoh hasil pengenalan yang benar dan salah pada citra wajah perempuan dan laki-laki.



Gambar 3. Contoh gambar perempuan dan laki-laki hasil pengenalan yang benar

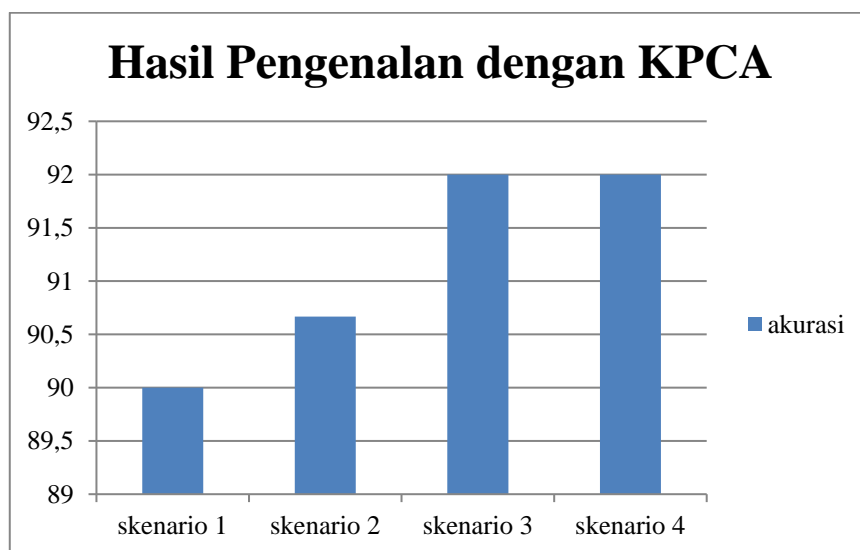


Gambar 4. Contoh gambar perempuan dan laki-laki hasil pengenalan yang salah

Pada Tabel 2 dan Gambar 5 menunjukkan uji coba sistem menggunakan 400 data citra berukuran 200x200. Skenario 1 menghasilkan akurasi 90%, sedangkan skenario 2 mempunyai akurasi 90.666%, skenario 3 mempunyai akurasi 92%, dan skenario 4 menghasilkan akurasi 92%.

Tabel 2. Hasil uji coba

Skenario	Citra <i>Training</i>	1. Citra <i>Testing</i>	KPCA		
			Benar	Salah	Akurasi
Skenario 1	200	200	180	20	90%
Skenario 2	250	150	136	14	90.666%
Skenario 3	300	100	92	8	92%
Skenario 4	350	50	46	4	92%



Gambar 5. Hasil tingkat akurasi menggunakan metode KPCA

SIMPULAN DAN SARAN

Pengenalan Jenis Kelamin menggunakan metode *Kernel Principal Component Analysis* dan *2D Correlation Coefficient* memperoleh hasil akurasi tertinggi pada skenario 4 yaitu sebesar 92%. Semakin banyak citra *training* maka semakin tinggi pula akurasi yang diperoleh. *KPCA* berhasil mereduksi dimensi matrik awal dan juga berhasil meningkatkan tingkat akurasi jika dibandingkan dengan *PCA*.

DAFTAR PUSTAKA

- [1] Purnomo, M.H., dan Muntasa, A., *Konsep Pengolahan Citra Digital dan Ekstraksi Fitur*. Yogyakarta: Graha Ilmu. 2010.
- [2] Riko, A.S., *Pengenalan Wajah Menggunakan Metode Fisherface*. Tugas Akhir. Universitas Kristen Maranatha, Bandung. 2007.
- [3] Quan, W., *Kernel Principal Component Analysis and its Application in Face Recognition and Active Shape Models*. Thesis. Rensselaer Polytechnic. USA. 2012.
- [4] Kekre, H.B., Thepade, S.D., Chopra, T., Face and Gender Recognition Using Principal Component Analysis. (*IJCSE*) *International Journal on Computer Science and Engineering*, 2: 1-6. 2010.
- [5] Eugene, K., Yen., dan Roger G.J., *The In Effectiveness of The Correlation Coefficient for Image Comparisons*. Thesis. Los Alamos National Laboratory. Los Alamos, New Mexico.
- [6] Ferdiansah, A. *Gender Recognition Menggunakan Metode Principal Component Analysis Dikombinasikan Dengan Local Binary Pattern*. Bangkalan: Universitas Trunojoyo Madura. 2013.