

Implementasi Principal Component Analysis Untuk Sistem Temu Balik Citra Digital

Okfan Rizal Ferdiansyah*¹, Ema Utami², Armadyah Amborowati³

¹STMIK AMIKOM Yogyakarta

^{2,3}Magister Teknik Informatika STMIK AMIKOM Yogyakarta

E-mail: *okfanrizal.f@gmail.com, emma@nrar.net, armadyah.a@amikom.ac.id

Abstrak

Tingginya kebutuhan citra digital dalam berbagai bidang, menuntut suatu mekanisme untuk manajemen data citra tersebut. Pada saat ini, teknik penotasian banyak digunakan untuk manajemen data citra digital. Namun, teknik ini menimbulkan perbedaan persepsi antara pengguna dengan penotasinya. Untuk itu perlu dikembangkan sebuah sistem temu balik citra digital yang mampu mengatasi permasalahan tersebut. Besarnya dimensi citra menjadi sebuah masalah tersendiri bagi bidang komputasi. Principal Component Analysis (PCA) mampu menjawab tantangan tersebut dengan melakukan proyeksi dari dimensi tinggi ke dimensi yang rendah. Pada penelitian sebelumnya, PCA sudah sering digunakan untuk sistem temu balik citra digital tetapi selalu memakai 100% komponen PCA. Tulisan ini memaparkan hasil analisa performa PCA untuk mengetahui pengaruh jumlah pemakaian komponen PCA terhadap akurasi sistem dengan dataset University Washington, Visual Geometri Group dan dataset wajah. Dari berbagai ujicoba, diperoleh hasil bahwa 10% s.d 30% komponen PCA yang dipakai menghasilkan akurasi sistem yang paling tinggi untuk masing-masing dataset, yaitu 91,4% untuk dataset University Washington, 92,0% untuk dataset Visual Geometri Group, dan 75,3% untuk dataset wajah.

Kata Kunci — Sistem temu balik, citra digital, Principal Component Analysis

Abstract

Increasing needs of digital image in many fields, requires a mechanism for managing the image data. At this time, notation techniques widely used to managing of digital image data. However these techniques cause differences of perception between the users and programmer. So it is necessary to develop a digital image retrieval system which is able to overcome these problems. Image dimension becomes a problem for the field of computing. Principal Component Analysis (PCA) answer the challenge by doing a high-dimensional projection to the lower dimension. In previous studies the PCA has been frequently used for digital image retrieval system but always use 100% PCA components. This article presents the results of the analysis of the PCA performance to know the influence usage amount PCA components against the accuracy of the system with dataset of University Washington (UW), Visual Geometri Group (VGG) and face. Of the various tests carried out showed that 10% to 30% of components used PCA system produces the highest accuracy for each dataset, 91.4% to 92.0% UW dataset to VGG dataset and 75.3 % for face dataset.

Keywords — Retrieval system, digital image, Principal Component Analysis

1. PENDAHULUAN

Kebutuhan citra digital meningkat dengan jumlah yang sangat besar untuk keperluan berbagai bidang seperti kedokteran, arsitektur, kesehatan, olahraga dan bidang-bidang lainnya [1]. Besarnya jumlah citra digital membuat proses pencarian memerlukan waktu yang lama, sampai saat ini masih banyak yang menggunakan teknik notasi untuk melakukan proses pencarian citra digital[2], cara seperti ini sering terjadi kesalahan dalam proses pencarian sebuah citra karena proses penotasian sangat bergantung pada persepsi user.

Beberapa metode pernah diusulkan oleh para peneliti sebelumnya untuk proses sistem temu balik citra digital seperti *Sub Block Base Image Retrieval Modified Region Matching*[3], *Shape, Color and Relevance Feedback* [4]. *Sub Block Base* merupakan metode yang mampu melakukan reduksi dimensi dengan membagi sebuah citra digital menjadi beberapa blok region dan mengabaikan region-region yang tidak begitu memiliki informasi[3]. Akan tetapi metode ini memerlukan penelitian lebih lanjut untuk menentukan jumlah block region serta ukuran untuk mendapatkan hasil yang optimal. Metode *Shape, Color and Relevance Feedback* menggabungkan beberapa fitur untuk menghasilkan tingkat akurasi yang tinggi, tapi dengan jumlah data citra digital yang besar dapat mengakibatkan proses berjalan lambat, *Relevance feedback* juga sangat bergantung pada asumsi dari user [5].

Principal Component Analysis (PCA) adalah suatu analisis yang menjelaskan struktur varian-kovarian dari suatu himpunan variabel yang melalui beberapa kombinasi linear dari variabel-variabel tersebut [6]. PCA mampu menjawab tantangan input sebuah citra digital dengan melakukan proses reduksi dari dimensi tinggi ke dimensi yang lebih rendah. Penelitian sebelumnya tentang PCA untuk CBIR adalah Implementasi Pengenalan Wajah Menggunakan PCA [7], Implementation of CBIR System for CAD Jewellery Images Using PCA [8], dan CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences [9]. Semua penelitian sebelumnya selalu menggunakan 100% komponen dari PCA, padahal terdapat potensi untuk lebih meningkatkan efisiensi waktu komputasi dengan menganalisa pengaruh prosentase komponen PCA yang digunakan. Sehingga pada penelitian ini peneliti fokus untuk menganalisa pengaruh reduksi komponen PCA terhadap akurasi sistem.

Penelitian sebelumnya tentang PCA dan CBIR pernah dilakukan pada tahun 2013 yang dilakukan oleh Kaur dan Jyoti yang berjudul Implementation of CBIR System for CAD Jewellery Images Using PCA[8], pada penelitian tersebut peneliti fokus pada presisi dan recall sistem. Dari hasil ujicoba didapatkan kesimpulan bahwa sistem mempunyai presisi sebesar 54%, recall 30% dan kecepatan waktu komputasi sistem bergantung pada query yang diinputkan oleh user.

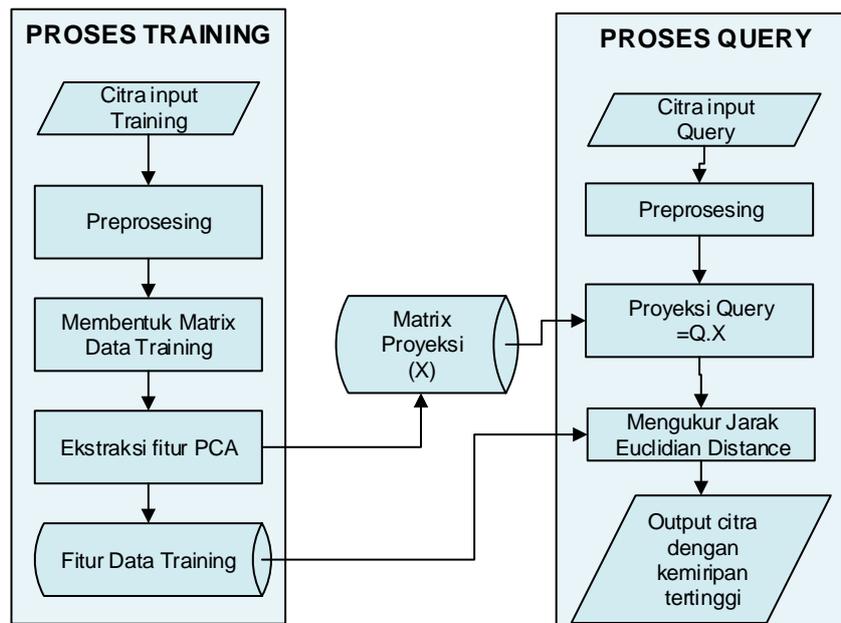
PCA dan CBIR pernah diteliti oleh H.B. Kekre pada tahun 2010 berjudul CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences [9]. Pada penelitian tersebut peneliti fokus pada analisa kombinasi matrix covarian yang digunakan terhadap performa dari PCA, didapatkan kesimpulan bahwa menggunakan matrix vector lebih baik dari pada menggunakan variasi vector didapatkan pula hasil dengan menggunakan citra digital true color mempunyai hasil yang lebih baik dibandingkan dengan citra grayscale.

Selain itu PCA juga pernah diteliti oleh Dian Esti Pratiwi dan Agus Harjoko pada tahun 2013 dengan judul Implementasi Pengenalan Wajah Menggunakan PCA mempunyai akurasi sebesar 82,81%[7]. Pada penelitian tersebut penulis ingin mengembangkan sebuah sistem identifikasi berdasarkan wajah, dari hasil ujicoba didapatkan kesimpulan bahwa banyak faktor yang mempengaruhi akurasi sistem diantaranya adalah faktor cahaya, jarak dengan webcam, dan banyaknya gambar yang tersimpan di dalam database.

2. METODE PENELITIAN

2.1. Perancangan Sistem

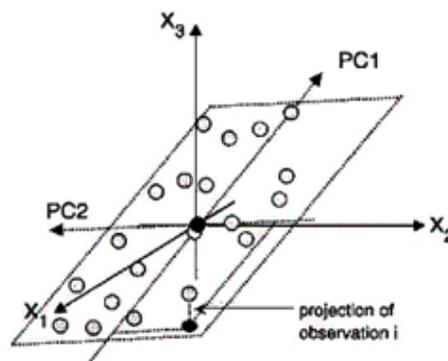
Sistem yang dibangun adalah suatu prototipe yang digunakan untuk membantu peneliti dalam menganalisa CBIR, sistem memiliki beberapa attribute yang bisa diubah-ubah sesuai dengan kebutuhan peneliti untuk mendapatkan hasil yang diinginkan. Sistem yang dibangun mempunyai fungsionalitas yang terbatas yaitu untuk proses testing dan ujicoba, gambar 1 menunjukkan flowchart sistem yang dibangun.



Gambar 1. Flowchart Sistem

2.2. Principal Component Analyst

Dalam melakukan proses sistem temu balik dihadapkan dengan banyak variabel dengan data yang berdimensi tinggi. Operasi-operasi yang dilakukan terhadap citra query maupun citra yang ada di dalam dataset membutuhkan biaya komputasi yang sangat besar. Oleh karena itu, dibutuhkan PCA yang dapat mengurangi besarnya dimensi dari data yang di observasi yang signifikan dalam menggambarkan keseluruhan data seperti digambarkan pada gambar 2[10].



Gambar 2. Diagram proyeksi dimensi tinggi ke dimensi yang lebih rendah di PCA

2.2.1. Standar deviasi

Untuk dapat memahami standar deviasi, kita memerlukan sebuah himpunan data. Seorang ahli statistik biasanya hanya mengambil sampel dari sebuah populasi. Keuntungan dari statistik adalah hanya dengan melakukan perhitungan pada data sampel dari populasi, sudah cukup baik menggambarkan perhitungan data sebenarnya [11].

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \dots\dots\dots (1)$$

Dimana S merupakan standar deviasi, X adalah sebuah himpunan data dan \bar{X} adalah rata-rata dari X. Dengan menggunakan rumus (1), maka himpunan data [0 6 15 40] akan memiliki nilai standar deviasi 17.6139 sedangkan himpunan data [4 6 8 9] akan memiliki nilai standar deviasi 2.2173. Nilai standar deviasi yang besar menunjukkan bahwa nilai data-data yang ada di dalam sebuah himpunan tersebar jauh dari rata-rata nilai himpunan tersebut, sedangkan nilai standar deviasi yang kecil menunjukkan bahwa nilai data-data yang ada di dalam sebuah himpunan berkumpul di sekitar rata-rata nilai himpunan tersebut [10].

2.2.2. Varian

Varian merupakan cara lain untuk mengukur seberapa tersebar nilai data-data yang ada di dalam sebuah himpunan. Faktanya adalah varian hampir sama dengan standar deviasi. Rumusan dari varian adalah sebagai berikut,

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \dots\dots\dots (2)$$

Dapat kita lihat pada rumus (2) bahwa varian hanya merupakan kuadrat dari standar deviasi, S^2 merupakan varian dari sebuah himpunan data [10].

2.2.3. Kovarian

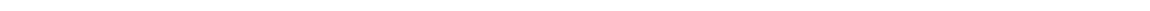
Standar deviasi dan varian hanya beroperasi pada data dengan satu dimensi. Padahal banyak himpunan data lebih dari satu dimensi, dan tujuan dari analisis statistik adalah untuk melihat apakah ada hubungan antar dimensi. Kovarian dapat melakukan perhitungan tersebut. Kovarian adalah ukuran untuk mengetahui seberapa kuat hubungan antar dimensi satu sama lain pada himpunan data.

Kovarian selalu diukur antar dua dimensi. Jika kita memiliki himpunan data yang berdimensi tiga (x, y, z), maka kita dapat menghitung kovarian antara dimensi x dan y, antara dimensi x dan z, dan antara dimensi y dan z. Melakukan perhitungan kovarian antara x dan x, y dan y, dan z dan z masing-masing akan memberikan nilai varian dari x, y, dan z.

Jika nilai kovarian positif, maka hal ini menunjukkan bahwa kedua dimensi berbanding lurus. Sedangkan jika nilai kovarian negatif, maka hal ini menunjukkan bahwa kedua dimensi berbanding terbalik. Dan jika nilai kovarian nol, maka hal ini menunjukkan bahwa kedua dimensi tidak berhubungan sama sekali.

Jika X dan Y adalah dua dimensi yang tidak independen, a dan b adalah suatu konstanta, maka kovarian di antara dua dimensi ini memiliki beberapa properti sebagai berikut [12]:

1. $Cov(X, X) = Var(X)$
2. $Cov(X, Y) = Cov(Y, X)$
3. $Cov(aX, bY) = abCov(X, Y)$
4. $Cov(X+a, Y+b) = Cov(X, Y)$



2.2.4. Matrix Kovarian

Kovarian selalu mengukur antara dua dimensi, jika kita memiliki himpunan data yang lebih dari dua dimensi, maka akan lebih dari satu nilai kovarian yang dapat dihitung. Sebagai contoh, dari himpunan data dengan tiga dimensi (dimensi x, y, dan z) kita dapat menghitung $cov(x,y)$, $cov(x,z)$, dan $cov(y,z)$. Dan untuk himpunan data dengan n dimensi, maka akan ada $\frac{n!}{(n-2)! \times 2}$ nilai kovarian yang berbeda

Cara yang paling mudah untuk mendapatkan semua nilai kovarian yang mungkin dari suatu himpunan data adalah dengan menghitung semua nilai kovarian yang mungkin dan meletakkannya pada sebuah matriks, yang sering disebut dengan matriks kovarian.

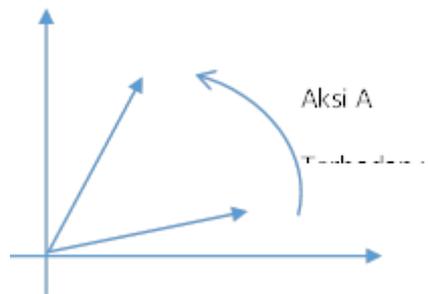
Jika kita memiliki himpunan data dengan tiga dimensi, yang terdiri dari dimensi x,y, dan z, maka matriks kovarian akan memiliki tiga baris dan tiga kolom. Matriks kovarian merupakan matriks yang simetrik pada diagonal utama. Hal ini sesuai dengan properti yang dimiliki oleh kovarian [13]

2.2.5. Vector Eigen dan Nilai Eigen

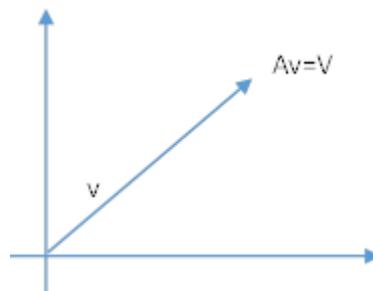
Jika A adalah suatu matriks persegi berdimensi $n \times n$ dalam ruang C^n , serta x dan b adalah suatu vektor berdimensi $n \times 1$, dan terdapat persamaan linear

$$Ax = b \dots\dots\dots (3)$$

Maka maksudnya adalah transformasi dilakukan oleh matriks A terhadap vektor x ke suatu vektor baru[6]. Ilustrasi dari transformasi yang dilakukan oleh matriks A terhadap vektor x dapat dilihat pada gambar 3 dan gambar 4



Gambar 3. Transformasi A terhadap Vector



Gambar 4. Transformasi Matriks A terhadap Vector Eigen v

2.2.6. Pincipal Component

Vektor eigen dengan nilai eigen yang besar memiliki peranan yang paling penting dalam proses transformasi [11]. Oleh karena itu, mereduksi dimensi dengan cara membuang vektor-vektor eigen dengan nilai eigen yang sangat kecil tidak akan membuat kita kehilangan data yang penting. Vektor-vektor eigen dengan nilai eigen yang terbesar disebut sebagai principal component dari sebuah matriks.

Cara untuk mendapatkan principal component dari sebuah matriks dilakukan dalam beberapa langkah. Langkah pertama yang dilakukan adalah mengurangi setiap nilai dari matriks dengan rata-rata nilai dari setiap masing-masing dimensi. Langkah berikutnya adalah menghitung matriks kovariannya dan dilanjutkan dengan mencari vektor eigen dan nilai eigen dari matriks kovarian tersebut. Kemudian vektor eigen yang ada diurutkan berdasarkan nilai eigen-nya dari yang paling besar ke yang paling kecil. Dengan memilih vektor-vektor eigen yang memiliki nilai-nilai eigen terbesar maka kita telah mendapatkan principal component dari matriks awal dan kita dapat membentuk feature vector. Feature vector berguna untuk mendapatkan data-data dari matriks awal yang bersesuaian dengan vektor eigen yang telah kita pilih dengan cara mengalikan transpose dari feature vector dengan transpose.

3. HASIL DAN PEMBAHASAN

3.1. Skenario Ujicoba

Beberapa skenario ujicoba dilakukan untuk mengetahui tingkat akurasi sistem, pada table 1 berikut ini adalah skenario-skenario yang digunakan dalam sistem temu balik citra digital.

Tabel 1. Tabel Skenario Ujicoba

No	Skenario	Training	Testing	Keterangan
1	Skenario I	50	75	Dataset UW
2	Skenario II	70	55	Dataset UW
3	Skenario III	90	35	Dataset UW
4	Skenario IV	40	60	Dataset VGG
5	Skenario V	60	40	Dataset VGG
6	Skenario VI	80	20	Dataset VGG
7	Skenario VII	30	120	Dataset Wajah
8	Skenario VIII	60	90	Dataset Wajah
9	Skenario IX	90	60	Dataset Wajah

3.2. Implementasi Ujicoba

Proses pertama yang dilakukan dalam implementasi sistem adalah memasukkan data *training* dan data *testing*, setelah itu sistem secara otomatis melakukan *preprocessing*, pembentukan data latih, ekstraksi fitur, proyeksi matrik query dan yang akan ditampilkan oleh sistem adalah hasil pengukuran jarak yang terdekat antara query dengan data training.

Selain itu setiap skenario akan dilakukan perubahan jumlah komponen PCA yang digunakan untuk mengetahui persentase reduksi dimensi yang paling baik dalam hal akurasi sistem seperti pada gambar 4.

```
>> myFinalThesis(data, '3-25.jpg', 8)

ans =

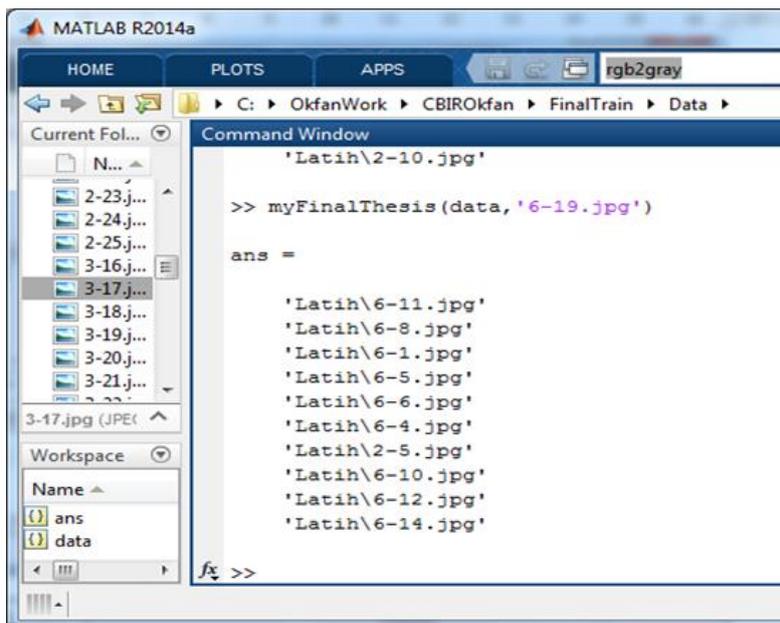
    'Latih\3-3.jpg'
    'Latih\3-20.jpg'
    'Latih\3-9.jpg'
    'Latih\3-7.jpg'
    'Latih\3-6.jpg'
    'Latih\3-11.jpg'
    'Latih\2-12.jpg'
    'Latih\2-4.jpg'
    'Latih\2-14.jpg'
    'Latih\3-18.jpg'
```

Gambar 4. Hasil running program

Tabel 2. Hasil ujicoba dataset VGG

Query	Hasil		
			
3-25.jpg	3-3.jpg	3-20.jpg	3-9.jpg
			
	3-7.jpg	3-6.jpg	3-11.jpg
			
	2-12.jpg	2-4.jpg	2-14.jpg
			
	3-18.jpg		

Pada gambar 4 dan tabel 2 ditunjukkan bahwa sistem temu balik citra digital dijalankan dengan input sebuah query berupa file gambar “3-25.jpg”, program tersebut memiliki 3 parameter, parameter pertama adalah list data training, parameter yang kedua adalah nama file query dan parameter yang terakhir adalah jumlah komponen yang digunakan, pada gambar 4 juga terlihat parameter pertama adalah “data”, parameter kedua merupakan query gambar yaitu “3-25.jpg” dan parameter yang ketiga adalah angka 8 menunjukkan bahwa komponen PCA yang digunakan dalam proses tersebut berjumlah 8. Setelah program dijalankan muncul hasil “3-3.jpg”, “3-10.jpg”, “3-9.jpg”, “3-6.jpg”, “3-7.jpg”, “3-11.jpg”, “2-12.jpg”, “2-14.jpg”, “2-14.jpg”, “3-18.jpg”, Sistem hanya akan menampilkan 10 gambar yang mempunyai jarak terdekat dengan gambar query yang diukur menggunakan algoritma *Euclidian Distance*. Sehingga bisa dikatakan gambar “3-25.jpg” mempunyai akurasi 70%.



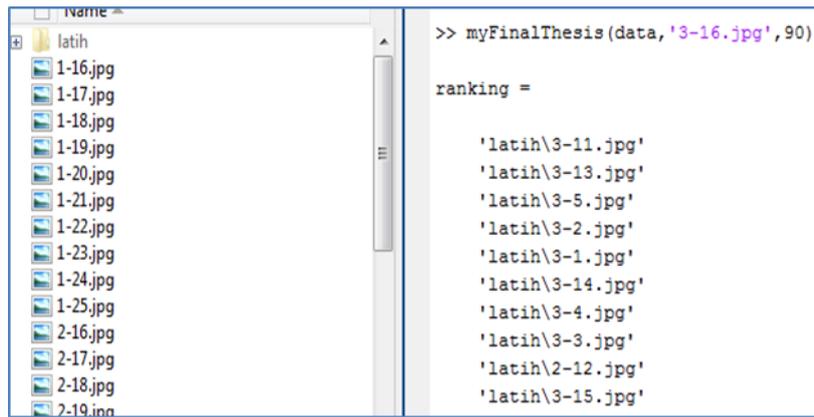
Gambar 5. Implementasi Uji Coba Gambar Dataset UW

Tabel 3. Hasil ujicoba dataset University Washington

Query	Hasil		
			
6-19.jpg	6-11.jpg	6-8.jpg	6-1.jpg
			
	6-5.jpg	6-6.jpg	6-4.jpg
			
	2-5.jpg	6-10.jpg	6-12.jpg
			
	6-14.jpg		

Dari hasil ujicoba seperti pada gambar 5 dan tabel 3, ketika sebuah sistem diinputkan gambar dari kelas 6 (gambar kuda), gambar yang ditampilkan oleh sistem merupakan data-data gambar yang didominasi dari kelas 6, yaitu “6-11.jpg”, “6-8.jpg”, “6-1.jpg”, “6-5.jpg”, “6-6.jpg”, “6-4.jpg”, “2-5.jpg”, “6-10.jpg”, “6-12.jpg”, “6-14.jpg”, dari 10 data gambar yang ditampilkan oleh sistem 9 diantara berasal dari data kelas 6 dan hanya ada 1 yang berasal dari kelas 2. Ini menunjukkan bahwa ujicoba menggunakan data bunga sangat relevan dalam sistem balik dengan nilai presisi mencapai 90%.

Untuk memudahkan dalam hal pemrosesan semua gambar diberi nama yang sesuai dengan kontennya, misalnya jika sebuah gambar mempunyai nama file “2-9.jpg” mempunyai arti bahwa gambar tersebut adalah kelas 2 dengan pose ke-19. Proses perhitungan akurasi dihitung menggunakan pendekatan presisi, dan akurasi sistem didapatkan dari rata-rata presisi setiap gambar yang dilakukan uji coba.



Gambar 6. Implementasi uji coba dataset wajah

Gambar 6 dan tabel 4 menunjukkan hasil ujicoba dataset wajah, terlihat bahwa sistem mampu mengenali hingga 90%.

Tabel 4. Ujicoba dataset wajah

Query	Hasil				
3-16.jpg	3-11.jpg	3-13.jpg	3-5.jpg	3-2.jpg	3-1.jpg
	3-14.jpg	3-4.jpg	3-3.jpg	2-12.jpg	3-15.jpg

3.3. Analisa Hasil Uji Coba

Setelah dilakukan beberapa kali percobaan didapatkan hasil seperti pada tabel 4 berikut.

Tabel 4. Hasil ujicoba training vs testing

No	Dataset	Training	Testing	Gambar	Akurasi
1	UW	50	75	Bunga	75,3%
				Bus	46,6%
				Dino	97,3%
				Gajah	82,6%
				Kuda	64,6%
				Rata-rata	73,3%
2	UW	70	55	Bunga	94,5%
				Bus	37,2%
				Dino	100%
				Gajah	86,3%
				Kuda	62,7%
				Rata-rata	76,1%

Tabel 4. (lanjutan)

No	Dataset	Training	Testing	Gambar	Akurasi
3	UW	90	35	Bunga	100%
				Bus	55,7%
				Dino	100%
				Gajah	98,5%
				Kuda	61,4%
				Rata-rata	83,1%
4	VGG	40	60	Pesawat	100%
				Daun	72%
				Motor	76%
				Gitar	96%
				Rata-rata	86%
5	VGG	60	40	Pesawat	100%
				Daun	91%
				Motor	73%
				Gitar	95%
				Rata-rata	89,7%
6	VGG	80	20	Pesawat	100%
				Daun	88%
				Motor	68%
				Gitar	96%
				Rata-rata	88%
7	Wajah	30	120	Orang ke-1	24,5%
				Orang ke-2	47,0%
				Orang ke-3	49,0%
				Orang ke-4	46,0%
				Orang ke-5	45,5%
				Orang ke-6	45,0%
				Rata-rata	42,8%
8	Wajah	60	90	Orang ke-1	36,0%
				Orang ke-2	51,3%
				Orang ke-3	81,3%
				Orang ke-4	56,6%
				Orang ke-5	71,3%
				Orang ke-6	63,8%
				Rata-rata	63,8%
9	Wajah	90	60	Orang ke-1	46,0%
				Orang ke-2	53,0%
				Orang ke-3	92,0%
				Orang ke-4	99,0%
				Orang ke-5	71,0%
				Orang ke-6	79,0%
				Rata-rata	73,3%

Tabel 5 berikut ini adalah hasil ujicoba dengan berbagai skenario dan variasi jumlah data training dan data testing.

Tabel 5. Hasil ujicoba berbagai nilai komponen PCA

No	Dataset	Training	Testing	Komponen PCA	Akurasi
1	UW	50	75	100%	73,3%
2				90%	71,7%
3				80%	71,8%
4				70%	71,4%
5				60%	71,4%
6				50%	70,9%
7				40%	71,6%
8				30%	71,6%
9				20%	76,6%
10				10%	79,3%
11	UW	70	55	100%	76,1%
12				90%	75,8%
13				80%	75,6%
14				70%	74,5%
15				60%	73,8%
16				50%	73,4%
17				40%	74,7%
18				30%	75,0%
19				20%	77,6%
20				10%	78,7%
21	UW	90	35	100%	83,1%
22				90%	80,5%
23				80%	80,5%
24				70%	80,8%
25				60%	82,2%
26				50%	83,1%
27				40%	84,8%
28				30%	86,8%
29				20%	88,0%
30				10%	91,4%
31	VGG	40	60	100%	86,1%
32				90%	85,6%
33				80%	85,3%
34				70%	85,1%
35				60%	85,1%
36				50%	84,3%
37				40%	83,5%
38				30%	86,1%
39				20%	86,3%
40				10%	92,1%

Tabel 5. (lanjutan)

No	Dataset	Training	Testing	Komponen PCA	Akurasi
41	VGG	60	40	100%	89,7%
42				90%	89,5%
43				80%	89,7%
44				70%	90,0%
45				60%	89,4%
46				50%	88,5%
47				40%	89,5%
48				30%	89,2%
49				20%	90,7%
50				10%	92,0%
51	VGG	80	20	100%	88,0%
52				90%	88,0%
53				80%	88,0%
54				70%	86,0%
55				60%	86,5%
56				50%	85,0%
57				40%	86,5%
58				30%	88,0%
59				20%	90,0%
60				10%	91,5%
61	Wajah	30	120	100%	42,8%
62				90%	42,7%
63				80%	42,8%
64				70%	42,8%
65				60%	42,8%
66				50%	42,5%
67				40%	42,5%
68				30%	42,8%
69				20%	43,1%
70				10%	42,6%
71	Wajah	60	90	100%	63,8%
72				90%	64,0%
73				80%	63,8%
74				70%	64,0%
75				60%	63,4%
76				50%	63,4%
77				40%	63,8%
78				30%	65,0%
79				20%	64,2%
80				10%	64,4%
81	Wajah	90	60	100%	73,3%
82				90%	73,1%
83				80%	72,8%
84				70%	73,3%
85				60%	73,1%
86				50%	73,6%
87				40%	73,6%
88				30%	75,3%
89				20%	74,5%
90				10%	73,6%

4. KESIMPULAN

1. Akurasi rata-rata sistem yang paling baik untuk algoritma PCA dalam sistem temu balik citra digital dengan dataset *University Washington*, *Visual Geometri Grup*, dan dataset Wajah adalah sebagai berikut
 - a. Akurasi maksimal rata-rata sistem untuk dataset *University Washington* adalah 91,4% terjadi saat menggunakan 10% komponen dari PCA dengan 90 data *training* dan 35 data *testing*.
 - b. Akurasi maksimal rata-rata sistem untuk dataset *Visual Geometri Group* adalah 92.0% terjadi saat menggunakan 10% komponen dari PCA dengan 60 data *training* dan 40 data *testing*.
 - c. Akurasi maksimal rata-rata sistem untuk dataset *Visual Geometri Group* adalah 75.3% terjadi saat menggunakan 30% komponen dari PCA dengan 90 data *training* dan 60 data *testing*.
2. Reduksi dimensi mempengaruhi tingkat akurasi sistem
 - a. Penggunaan antara 30% s.d. 10% dari komponen PCA menghasilkan akurasi yang paling baik untuk semua jenis gambar dataset yang digunakan.
 - b. Pemilih 30% s.d. 10% komponen mampu memangkas waktu komputasi hingga 90%.
 - c. Tingkat akurasi dengan reduksi dimensi tidak berbanding lurus maupun tidak berbanding terbalik, sehingga diperlukan penelitian lebih lanjut untuk menemukan pola reduksi dimensi dengan sebuah dataset.
 - d. Jumlah data training yang digunakan mempengaruhi tingkat akurasi sistem, semakin banyak data yang digunakan untuk proses training akurasi sistem cenderung semakin tinggi.

5. SARAN

1. Diharapkan pada penelitian selanjutnya menggunakan kombinasi variasi dataset yang lebih variatif sehingga dapat diketahui hubungan antara dataset, proses reduksi dimensi dan akurasi.
2. Diharapkan pada penelitian selanjutnya mengkombinasikan algoritma PCA dengan algoritma ekstraksi fitur lainnya.
3. Diharapkan pada penelitian selanjutnya digunakan jumlah obyek yang berbeda-beda pada dataset yang sama, untuk mengetahui hubungan obyek gambar tertentu dengan akurasi sistem.

DAFTAR PUSTAKA

- [1] Shortliffe, E. H., Cimino, J. J., 2006, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, Springer-Verlag London, United States of America.
 - [2] Kulkarni, R. K., Amoda, N., 2013, Efficient Image Retrieval using Region Based Image Retrieval, *IJAIS Proceedings on International Conference and workshop on Advanced Computing 2013*, No. 2, New York, Juni 2013.
 - [3] Vimina E. R., Jacop K. P., 2013, A Sub-bloc Based Image Retrieval Using Modified Integrated Region Matching, *International Journal of Computer Science Issue*, Vol 10, No 2, Hal 686-692.
 - [4] Yasmin, M., Mohsin, S., Irum, I., Sharif, M., 2013, Content Based Image Retrieval by Shape Color and Relevance Feedback, *Life Science Journal*, Vol 10, No 4s, Hal 593 - 598.
-

-
- [5] Singh, A., Shinde, A. G., Gulhane, S., M., 2013, Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patters, *Journal of Engineering, computer & Applied Sciencec(JEC&AS)*, Vol 2, No 3, Hal 50-54.
- [6] Johnson, R. A., Wichern, D. W., 2007, *Applied Multivariate Statistical Analysis*, 6thed, Prentice Hall, New Jersey.
- [7] Pratiwi, D. E., Harjoko, A. 2013, Implementasi Pengenalan Wajah Menggunakan PCA (Principal Component Analysis), *Indonesian Journal of Electronics and Instrumentation Systems (IJEIS)*, Vol. 3, No. 2, Hal 175-184.
- [8] Kaur, P., Jyoti, K., 2013, Implementation of CBIR System for CAD Jewellery Images Using PCA, *International Journal of Scientific & Engineering Research*, Vol 4, No 2.
- [9] Kekre, H. B., Thepade S. D., Maloo, A., 2010, CBIR Feature Vector Dimension Reduction with Eigenvector of Covariance Matrix using Row, Column and Diagonal Mean Sequences, *International Jurnal of Computer Applications*, Vol 3, No 12, Hal 39-46.
- [10] Cahyadi, D., 2007. Ekstraksi dan Kemiripan Mata pada Sistem Identifikasi Buron. <http://lontar.ui.ac.id/file?file=digital/123280-SK-691>, diakses tanggal 10 Mei 2014.
- [11] Smith, L. I., 2002, A Tutorial on Principal Component Analysis, http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_component.pdf, diakses tanggal 17 Mei 2014.
- [12] Weisstein, E. W., 2007, Distance, <http://mathworld.wolfram.com/distance.html>, diakses tanggal 17 Mei 2004
- [13] Manning, C. D., Raghavan, P., Schütze, H., 2009, *Introduction to Information Retrieval*, Cambridge, Cambridge University Press.
-