

3_ANALISIS KOMPONEN UTAMA PADA HASIL EKSTRAKSI FITUR CITRA DIGITAL

by Veronica Lusiana

Submission date: 18-May-2023 10:58AM (UTC+0700)

Submission ID: 2095955768

File name: ISIS_KOMPONEN_UTAMA_PADA_HASIL_EKSTRAKSI_FITUR_CITRA_DIGITAL.pdf (617.68K)

Word count: 3679

Character count: 22579

ANALISIS KOMPONEN UTAMA PADA HASIL EKSTRAKSI FITUR CITRA DIGITAL

Veronica Lusiana¹, Imam Husni Al Amin², Budi Hartono³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank

Email: ¹verolusiana@yahoo.com, ²imam_husni@yahoo.co.id, ³pakbudi@yahoo.com

Abstrak

Penelitian ini bertujuan untuk memperoleh jumlah dimensi data yang optimal untuk proses pencarian citra. Proses ekstraksi fitur citra menggunakan matriks GLCM (*Grey-Level Co-occurrence Matrix*), dengan memilih fitur tekstur citra. Fitur yang digunakan adalah mean, entropi, energi, kontras, dan homogenitas. Matriks ini menyimpan data *co-occurrence* tingkat keabuan piksel untuk ketetanggaan pada arah 0°, 45°, 90°, dan 135°. Proses reduksi dimensi data menggunakan analisis komponen utama (*Principal Component Analysis*, PCA).

Melalui hasil percobaan diperoleh, fitur yang dominan (signifikan) adalah mean dan entropi untuk ketetanggaan pada sudut 0 derajat. Pada ketetanggaan sudut 45 derajat maka fitur yang dominan adalah entropi dan kontras untuk. Fitur dominan disini adalah fitur utama yang dapat mewakili beberapa fitur-fitur lain yang digunakan untuk pencarian isi citra.

Kata Kunci: GLCM (*Grey-Level Co-occurrence Matrix*), PCA (*Principal Component Analysis*), fitur tekstur citra.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Penelitian ini akan berkaitan dengan proses mereduksi atau mengurangi dimensi data hasil ekstraksi fitur citra. Terdapat beberapa buah fitur citra yang dapat digunakan untuk memperoleh ciri citra. Dengan memperoleh ciri citra yang baik yang dapat merepresentasikan citra secara utuh, maka ciri citra ini dapat digunakan untuk aplikasi pencarian citra. Tiga buah fitur citra yang umum adalah: warna (*color*), bentuk obyek (*shape*), dan tekstur (*texture*). Penjelasan ekstraksi fitur secara lebih lanjut dijelaskan oleh Nixon dan Aguado (2012).

Disini data yang dihasilkan dari ekstraksi fitur citra masih memiliki dimensi yang cukup besar yaitu 17 buah sehingga dibutuhkan sebuah proses untuk lebih menyederhanakan dimensi ini. Proses untuk menyederhanakan jumlah dimensi data akan menggunakan metode analisis komponen utama atau *Principal Component Analysis* (PCA). Metode PCA dapat digunakan untuk mengurangi atau mereduksi dimensi data dengan tetap mempertahankan data yang memiliki pengaruh yang besar (signifikan). Setelah dilakukan reduksi maka diharapkan akan diperoleh pengurangan jumlah dimensi data ciri citra dan dapat diketahui data yang memiliki pengaruh signifikan.

Proses ekstraksi fitur citra menggunakan matriks *co-occurrence* tingkat keabuan piksel atau dikenal dengan istilah *Grey-Level Co-occurrence Matrix* (GLCM), dengan memilih fitur tekstur citra. Disini, matriks GLCM berisi data *co-occurrence* tingkat keabuan piksel untuk ketetanggaan pada arah 0°, 45°, 90°, dan 135°. Pada penelitian ini terdapat tujuh belas fitur yaitu:

- a. Fitur mean
- b. Fitur entropi arah 0°
- c. Fitur energi arah 0°
- d. Fitur kontras arah 0°
- e. Fitur homogenitas arah 0°
- f. Fitur entropi arah 45°
- g. Fitur energi arah 45°
- h. Fitur kontras arah 45°
- i. Fitur homogenitas arah 45°

- j. Fitur entropi arah 90°
- k. Fitur energi arah 90°
- l. Fitur kontras arah 90°
- m. Fitur homogenitas arah 90°
- n. Fitur entropi arah 135°
- o. Fitur energi arah 135°
- p. Fitur kontras arah 135°
- q. Fitur homogenitas arah 135°

Berikut ini adalah beberapa buah penelitian sejenis yang pernah dilakukan. Penelitian oleh Hardika dan kawan-kawan (Hardika, dkk., 2013), menggambarkan keragaman variabel dan mereduksinya sehingga mendapatkan faktor dominan yang mempengaruhi prestasi belajar siswa di SMAN 1 Medan. Proses analisis faktor mencoba menemukan hubungan (*relationship*) antara sejumlah variabel yang saling independen satu dengan yang lain, sehingga bisa dibuat satu atau beberapa kumpulan variabel yang lebih sedikit dari jumlah variabel awal.

Berdasarkan hasil penelitian yang dilakukan di SMA Negeri 1 Medan menunjukkan bahwa terdapat 3 faktor yang mempengaruhi prestasi belajar siswa diantaranya faktor dalam diri siswa dan lingkungan dengan nilai eigen paling tinggi yaitu 3,372 dan varians sebesar 33,721%, faktor pendukung dengan varians 14,176% dan faktor tambahan dengan varians 11,248%. Total varians yang dapat dijelaskan dari ketiga faktor tersebut sebesar 59,145%.

Dari kesepuluh variabel yang diteliti, setelah dilakukan analisis faktor dengan metode analisis komponen utama, diperoleh 3 faktor yaitu: faktor pertama (F1) merupakan faktor yang paling dominan mempunyai nilai eigen sebesar 3,372 dan mampu menjelaskan keragaman total sebesar 33,721%. Faktor ini terdiri atas variabel X1=bakat, X2=minat, X3=motivasi siswa dan X10=pergaulan siswa yang dinamakan faktor dalam diri siswa dan lingkungan. Faktor kedua (F2) terdiri atas variabel X4=motivasi orang tua, X6=kualitas pengajaran guru, X7=fasilitas sekolah dengan nilai eigen sebesar 1,148 dan mampu menjelaskan keragaman total sebesar 14,176%. Faktor ini dinamakan faktor pendukung. Faktor ketiga (F3) yaitu faktor tambahan, terdiri atas variabel X5=fasilitas belajar di rumah, X8=ekstra kurikuler dan X9=les tambahan dengan nilai eigen sebesar 1,125 dan mampu menjelaskan keragaman total sebesar 11,248%.

Penelitian oleh Ferdiansyah dan kawan-kawan (Ferdiansyah, 2015) mengimplementasikan *Principal Component Analysis* (PCA) untuk sistem temu balik citra digital. Besarnya dimensi citra menjadi sebuah masalah tersendiri bagi bidang komputasi. PCA mampu menjawab tantangan tersebut dengan melakukan proyeksi dari dimensi tinggi ke dimensi yang rendah. Hasil analisa performa PCA untuk mengetahui pengaruh jumlah pemakaian komponen PCA terhadap akurasi sistem temu balik citra memanfaatkan data University Washington, Visual Geometri Group, dan dataset wajah. Dari ujicoba, diperoleh hasil bahwa 10% sampai dengan 30% komponen PCA yang dipakai menghasilkan akurasi sistem tertinggi untuk setiap data, yaitu 91,4% untuk data University Washington, 92,0% untuk data Visual Geometri Group, dan 75,3% untuk data wajah.

Reduksi dimensi mempengaruhi tingkat akurasi sistem. Penggunaan antara 30% sampai dengan 10% dari komponen PCA menghasilkan akurasi yang paling baik untuk semua jenis gambar data yang digunakan. Pemilihan 30% sampai dengan 10% komponen mampu memangkas waktu komputasi hingga 90%. Jumlah data uji yang digunakan mempengaruhi tingkat akurasi sistem. Semakin banyak data yang digunakan untuk proses pelatihan (*training*) maka akurasi sistem cenderung semakin tinggi.

Penelitian oleh Kustian (Kustian, 2017), meneliti tentang analisis komponen utama menggunakan metode Eigenface terhadap pengenalan citra wajah. Input berupa citra wajah dalam bentuk citra skala abu-abu dengan ukuran dan resolusi yang sama. Output aplikasi ini berupa *class* terdekat dari wajah yang ingin dikenali. Salah satu metode pendekatan yang digunakan adalah Eigenface, sebuah metode yang dikemukakan oleh Turk dan Pentland. Dalam metode ini citra wajah akan diproyeksikan dalam sebuah ruang fitur yang menonjolkan variasi yang signifikan di antara citra wajah yang diketahui. Fitur signifikan inilah yang disebut dengan Eigenface karena fitur-fitur tersebut adalah komponen utama dari suatu set citra wajah untuk pelatihan. Eigenface hanya akan menangkap *point-*

point pada citra yang menyebabkan variasi yang signifikan antara wajah-wajah dalam database yang membuat mereka dapat dibedakan.

Terdapat beberapa faktor yang mempengaruhi hasil pengenalan yaitu posisi citra masukan, tingkat kecerahan, dan kemiripan nilai Eigenface. Sistem menampilkan beberapa citra wajah yang mempunyai kemiripan paling tinggi dengan citra masukan sehingga lebih mudah untuk mencarinya. Hasil pencarian dipengaruhi oleh faktor cahaya, sudut, dan lainnya. Aplikasi yang dibuat mendukung pengukuran dan analisis ciri biologi seseorang, sehingga memberikan informasi yang saling terkait merujuk karakteristik atau bagian tertentu pada seseorang sebagai proses pengenalan.

Penelitian oleh Rahayu dan Mustakim (Rahayu dan Mustakim, 2017), bertujuan untuk mengetahui kota-kota mana saja yang akan digunakan untuk klasterisasi wilayah di Indonesia berdasarkan tingkat pengajar yang sudah profesional yaitu melalui sertifikasi. Data yang digunakan terdiri dari 3 jenjang pendidikan yaitu SD, SMP dan SMA yang berada di sembilan puluh delapan kota diseluruh Indonesia. Proses pemetaan dapat dilakukan menggunakan salah satu teknik clustering yaitu K-Means. *Principal Component Analysis* (PCA) merupakan pendekatan fitur *selection* untuk pengurangan dimensi tanpa pengawasan teknik. Metode PCA dapat digunakan untuk mereduksi dimensi data tanpa harus kehilangan banyak informasi.

Dari hasil penelitian, reduksi dimensi data menggunakan metode PCA terbukti dapat meningkatkan kualitas kluster yang dihasilkan dengan algoritma K-Means. Dari pembagian 3 buah kluster ($K=3$), diperoleh hasil kluster 0 sebanyak 13 kota, kluster 1 sebanyak 44 kota, dan kluster 2 sebanyak 41 kota. Hasil pengukuran validitas kluster dengan Davies-Bouldin Index menunjukkan PCA K-Means menghasilkan nilai DB Index lebih kecil dibandingkan dengan K-Means. Penelitian ini menghasilkan sebuah metode yang lebih efektif menggunakan PCA yang dimodifikasi dengan K-Means.

1.2 Perumusan Masalah

Perumusan masalah pada penelitian ini adalah untuk memperoleh jumlah dimensi data yang optimal untuk proses pencarian citra. Dimensi data awal dapat dilihat pada Tabel 5.1. data ekstraksi fitur citra.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah,

- Data citra berukuran 128X128 piksel.
- Proses komputasi menggunakan perangkat lunak Matlab 7.
- Bahan penelitian adalah citra yang berasal dari pengambilan langsung di lapangan dan citra yang dibuat untuk mengkondisikan atau meminimalkan variasi warna dan tekstur.

2. TUJUAN DAN MANFAAT PENELITIAN

2.1 Tujuan dan Manfaat Penelitian

Tujuan penelitian ini adalah melakukan analisa proses ekstraksi fitur citra menggunakan matriks GLCM (*Grey-Level Co-occurrence Matrix*) dan proses reduksi dimensi data menggunakan analisis komponen utama (*Principal Component Analysis*, PCA). Manfaat penelitian ini adalah menghasilkan model ekstraksi fitur citra yang dapat digunakan untuk proses pencarian isi citra, serta menerapkan metode PCA untuk aplikasi pencarian isi citra.

3. TINJAUAN PUSTAKA

3.1 Fitur tekstur

Pengertian fitur tekstur disini adalah keteraturan pola tertentu yang terbentuk dari susunan piksel-piksel pada citra. Informasi tekstur yaitu apabila memiliki pola atau karakteristik di suatu daerah pada citra yang muncul secara berulang dengan interval jarak dan arah tertentu. Beberapa contoh fitur tekstur yang umum digunakan yaitu: entropi, energi, kontras, homogenitas, nilai rata-rata intensitas, dan deviasi standar. Fitur tersebut telah diusulkan oleh Haralick dan rekan-rekannya pada tahun 1973. Fitur tekstur lain yang cukup populer adalah fitur yang diusulkan oleh Tamura dan rekan-rekannya,

yaitu: coarseness, contrast, directionality, linelikeness, regularity, dan roughness (Davis, dan kawan-kawan, 1979). Sampai saat ini belum ada sebuah formulasi yang dapat digunakan secara umum untuk mendapatkan informasi tekstur dari citra.

Pola pada citra yang memiliki informasi tekstur dapat dikenali secara visual oleh mata kita meskipun pola tersebut tersusun oleh beberapa piksel dengan susunan yang tidak mudah untuk didefinisikan atau diformulasikan. Sistem visual manusia tidak menerima informasi citra secara independen untuk setiap piksel tetapi menerimanya dalam sekumpulan piksel. Tekstur adalah ciri yang dimiliki oleh citra dimana tidak mudah untuk dijelaskan dengan kalimat namun kita dapat mengenalinya pada saat melihat (Ahmad, 2005).

3.2 Matriks intensitas co-occurrence untuk ekstraksi fitur

Matriks intensitas co-occurrence adalah suatu matriks yang menggambarkan frekuensi munculnya pasangan dua piksel dengan intensitas tertentu dalam jarak dan arah tertentu di dalam citra. Pada Gambar 3.1 dapat dilihat besarnya sudut dan arah vektor dengan jarak pasangan piksel sebesar 1 piksel. Matriks intensitas co-occurrence $P(i_1, i_2)$, didefinisikan dengan dua langkah:

- Menentukan jarak antara dua titik dalam arah vertikal dan horizontal (vektor $d=(dx, dy)$), dimana besaran dx dan dy dinyatakan dalam piksel sebagai unit terkecil dalam citra digital.
- Menghitung pasangan piksel-piksel yang mempunyai nilai intensitas i_1 dan i_2 dan berjarak d piksel dalam citra. Hasil perhitungan setiap pasangan nilai intensitas diletakkan pada matriks sesuai dengan koordinatnya, dimana absis untuk nilai intensitas i_1 dan ordinat untuk nilai intensitas i_2 .

Matriks co-occurrence digunakan untuk menghitung fitur tekstur seperti: entropi, energi, kontras, dan homogenitas. Pada fitur tekstur yang lain, seperti: nilai rata-rata (*mean*) intensitas dan deviasi standar dapat diperoleh melalui perhitungan statistik. Berikut ini adalah rumus-rumus untuk fitur tekstur tersebut.

Nilai mean (x') didefinisikan sebagai berikut:

$$Mean = \frac{\left(\sum_{i_1=1}^m \sum_{i_2=1}^n x(i_1, i_2) \right)}{m \cdot n} \quad (1)$$

Nilai deviasi standar (σ) didefinisikan sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i_1=1}^m \sum_{i_2=1}^n (x(i_1, i_2) - x')^2}{m \cdot n}} \quad (2)$$

Nilai entropi digunakan untuk mengukur keacakan dari distribusi intensitas yang didefinisikan sebagai berikut:

$$Entropi = - \sum_{i_1=1}^m \sum_{i_2=1}^n p(i_1, i_2) \log p(i_1, i_2) \quad (3)$$

Nilai entropi akan maksimum apabila semua elemen $p(i_1, i_2)$ sama, yaitu tidak terdapat susunan tertentu dalam pasangan intensitas dengan jarak vektor d tertentu.

Nilai energi digunakan untuk mengukur konsentrasi pasangan intensitas pada matriks co-occurrence yang didefinisikan sebagai berikut:

$$Energi = \sum_{i_1=1}^m \sum_{i_2=1}^n p^2(i_1, i_2) \quad (4)$$

Nilai energi membesar apabila pasangan piksel yang memenuhi syarat matriks intensitas co-occurrence terkonsentrasi pada beberapa koordinat, dan sebaliknya akan mengecil jika letaknya menyebar.

Nilai kontras digunakan untuk mengukur kekuatan perbedaan intensitas dalam citra yang didefinisikan sebagai berikut:

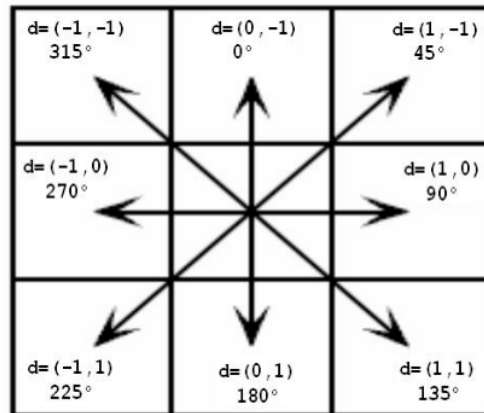
$$Kontras = \sum_{i_1=1}^m \sum_{i_2=1}^n (i_1 - i_2)^2 P(i_1, i_2) \tag{5}$$

Nilai kontras membesar jika variasi intensitas dalam citra tinggi, dan sebaliknya akan mengecil jika variasinya rendah

Nilai homogenitas adalah kebalikan dari nilai kontras yaitu untuk mengukur kehomogenan variasi intensitas dalam citra yang didefinisikan sebagai berikut:

$$Homogenitas = \sum_{i_1=1}^m \sum_{i_2=1}^n \frac{P(i_1, i_2)}{1 + |i_1 - i_2|} \tag{6}$$

Nilai homogenitas membesar jika variasi intensitas dalam citra rendah, dan sebaliknya akan mengecil jika variasinya tinggi.



Gambar 1. Sudut dan arah vektor untuk pasangan piksel berjarak 1 piksel

3.3 Pengurangan Dimensi Data

Metode reduksi atau pengurangan dimensi (*dimensionality reduction*) data bekerja dengan cara menangkap karakteristik data dengan memetakan set data dari dimensi awal ke dimensi baru yang relatif lebih rendah. Pemetaan ini menghasilkan komponen utama (*Principal Component*) yang dapat diambil komponen atau fitur dari dimensi baru yang mempunyai pengaruh besar pada set data dan mengabaikan komponen yang tidak berpengaruh besar atau bahkan tidak berpengaruh sama sekali. Apabila reduksi dimensi data dilakukan secara langsung dengan membuang fitur tertentu, maka dapat terjadi ada karakteristik data yang hilang dan bisa jadi fitur yang dibuang tersebut ternyata adalah fitur yang berpengaruh (Prasetyo, 2012).

3.4 Analisis Komponen Utama

Analisis Komponen Utama (*Principal Component Analysis*) melakukan pemetaan atau transformasi set data dari dimensi lama ke dimensi baru yang relatif berdimensi lebih rendah dengan memanfaatkan teknik dalam aljabar linier. PCA merupakan teknik linier untuk memproyeksikan data vektor yang berdimensi tinggi ke vektor yang mempunyai dimensi lebih rendah. PCA banyak digunakan untuk keperluan ekstraksi fitur citra dimana dapat terjadi jumlah dimensi citra

jauh lebih besar dibandingkan dengan jumlah data sampel yang digunakan (Pumomo dan Muntasa, 2010).

PCA adalah sebuah teknik untuk membangun variabel-variabel baru yang berjumlah tidak lebih dari variabel lama dan merupakan kombinasi linier dari variabel-variabel lama. Variabel baru disebut sebagai *principle component* dan nilai-nilai bentuk dari variabel ini disebut sebagai *principle component score*. Variabel baru pertama berhubungan dengan *variance maximum* dari data. Variabel baru kedua menunjukkan *variance maximum* yang belum terhitung pada variabel pertama. Variabel baru ketiga menunjukkan *variance maximum* yang belum terhitung pada kedua variabel pertama. Variabel baru ke-n adalah *variance maximum* yang belum terhitung pada n-1 variabel terdahulu. Variabel baru yang dihasilkan tidak saling berkorelasi satu sama lain.

Jika tujuan dari analisa data adalah untuk mengurangi jumlah variabel (*principle component*) yang merupakan kombinasi linear dari variabel-variabel asli, maka jumlah *principle component* harus kurang dari jumlah variabel asli. Dalam kasus ini, PCA dilakukan bila data dapat direpresentasikan dalam beberapa buah *principle component* tanpa menghilangkan substansi informasi.

Nilai *principal component scores* dari dua variabel baru (*principal component*) utama, dapat digunakan sebagai masukan variabel untuk data *clustering*. Secara umum skor dari *principal component* juga dapat digunakan sebagai masukan variabel untuk analisa regresi dan analisa diskriminant (*discriminant analysis*). Keuntungan menggunakan *principal component score* adalah variabel-variabel baru tidak saling berkorelasi sehingga masalah *multicolinearity* dapat dihindari

4. METODOLOGI PENELITIAN

4.1 Metode Penelitian

Metode penelitian pada penelitian ini adalah sebagai berikut,

1. Mencari literatur yang menunjang proses penelitian. Pada penelitian ini, yang utama adalah referensi mengenai penyusunan matriks *co-occurrence* (GLCM) dan penyusunan komponen utama (PCA).
2. Mencari dan atau membuat data citra.
3. Praproses terhadap data citra.
4. Penyusunan matriks *co-occurrence* (GLCM).
5. Penyusunan komponen utama (PCA).
6. Menganalisa hasil yang diperoleh.

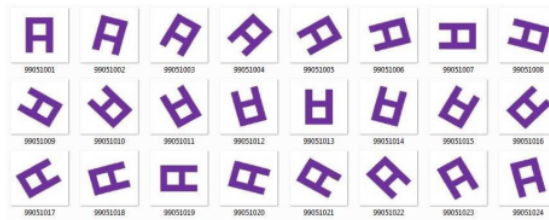
5. HASIL DAN ANALISA

5.1 Penyusunan data sub-Citra

Data citra yang digunakan untuk penelitian ini berukuran 512X512 dan 256X256 piksel. Proses *resize* bertujuan untuk menyesuaikan ukuran data citra sehingga ukuran data citra adalah sama dengan ukuran citra query, pada penelitian ini adalah 128X128 piksel. Data citra yang telah berukuran 128X128 piksel disebut juga dengan istilah data sub-citra. Data sub-citra dapat dilihat pada gambar 5.1, Gambar 5.2., Gambar 5.3, Gambar 5.4, dan Gambar 5.5.



Gambar 2. Penyusunan data sub-citra (i)



Gambar 3. Penyusunan data sub-citra (ii)



Gambar 4. Penyusunan data sub-citra (iii)



Gambar 5. Penyusunan data sub-citra (iv)



Gambar 6. Penyusunan data sub-citra (v)

5.2 Penyusunan matriks co-occurrence

Pada citra yang memiliki G tingkat keabuan, maka matriks *co-occurrence* P akan berukuran G x G. Apabila G cukup besar, maka jumlah pasangan piksel yang berkontribusi ke setiap elemen pij di dalam P akan rendah. Sedangkan jika tingkat keabuan G rendah yang disebabkan oleh proses kuantisasi citra, maka informasi tekstur semakin berkurang.

Dengan G=8 untuk ukuran citra 32x32 piksel, merupakan pilihan yang baik untuk membentuk P. Sedangkan kombinasi pasangan piksel tetangga terdekat dapat menggunakan sudut 0°, 45°, 90°, dan 135°. Nilai G sebaiknya dibatasi sampai dengan 16 dan jarak piksel dengan piksel tetangganya adalah 1 atau 2 piksel, mengingat lamanya waktu komputasi yang dibutuhkan (Ahmad, 2005; Gonzalez dan Woods, 2008). Pada Gambar 5.6 dapat dilihat proses penyusunan matriks co-occurrence dan pada Tabel 5.1 adalah matriks yang digunakan untuk menyimpan nilai ekstraksi fitur tekstur citra.

```

file_ori =
    3    0    0    2    2    2    2    1
    3    3    1    2    1    2    1    2
    0    2    2    2    1    3    3    3
    0    1    3    1    3    2    2    2
    0    2    3    3    2    3    3    0
    1    3    2    3    2    0    0    0
    1    2    3    0    0    0    1    2
    1    2    3    3    3    2    1    1

===== jarak 1 piksel arah 45 derajat (d=(1,-1)) =====
Matriks intensitas co-occurrence:  Matriks intensitas co-occurrence ternormalisasi:
    3    1    0    5          0.0612  0.0204          0  0.1020
    1    2    3    2          0.0204  0.0408  0.0612  0.0408
    2    7    5    5          0.0408  0.1429  0.1020  0.1020
    2    1    6    4          0.0408  0.0204  0.1224  0.0816

f_mean =      f_stdeviasi =      entropi =      energi =      kontras =      homogenitas =
1.7031      68.5201          2.5409          0.0887          2.1633          0.5901

===== jarak 1 piksel arah 90 derajat (d=(1,0)) =====
Matriks intensitas co-occurrence:  Matriks intensitas co-occurrence ternormalisasi:
    5    0    1    3          0.0893          0  0.0179  0.0536
    2    1    5    2          0.0357  0.0179  0.0893  0.0357
    3    6    7    5          0.0536  0.1071  0.1250  0.0893
    0    4    5    7          0  0.0714  0.0893  0.1250

f_mean =      f_stdeviasi =      entropi =      energi =      kontras =      homogenitas =
1.7031      68.5201          2.5059          0.0886          1.6071          0.6354
    
```

Gambar 7. Penyusunan matriks co-occurrence

Tabel 1. Data nilai ekstraksi fitur tekstur citra

nomor	nama citra	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		0 derajat					45 derajat				90 derajat				135 derajat			
		mean	entropi	energi	kontras	homogenitas	entropi	energi	kontras	homogenitas	entropi	energi	kontras	homogenitas	entropi	energi	kontras	homogenitas
1	13001	3.3148	3.01	0.0709	0.7241	0.8138	3.198	0.0561	0.9849	0.7565	2.8059	0.0833	0.4277	0.8596	3.2117	0.0552	1.0119	0.7523
2	13002	3.2737	2.8629	0.0814	0.5728	0.8521	3.0862	0.0628	0.7965	0.7899	2.7411	0.0879	0.3778	0.8747	3.0977	0.0624	0.8371	0.7875
3	13003	3.3043	2.9707	0.0742	0.6991	0.8211	3.2506	0.052	1.0854	0.736	2.9341	0.0735	0.5655	0.8256	3.2374	0.0533	1.0629	0.7413
4	13004	3.2507	2.779	0.0893	0.4943	0.8673	2.9589	0.0747	0.6978	0.8219	2.5868	0.1024	0.283	0.9062	2.9228	0.0773	0.6334	0.8318
5	13005	3.2084	3.1112	0.0626	0.8826	0.78	3.2378	0.0543	1.1319	0.7436	2.8148	0.0842	0.4818	0.8593	3.2746	0.0521	1.2325	0.7325
6	13006	3.2815	3.0672	0.071	0.9329	0.7949	3.268	0.0556	1.2929	0.7316	2.8916	0.0807	0.5782	0.8319	3.2701	0.0553	1.2845	0.7298
7	23001	3.3253	3.1252	0.0642	0.9035	0.7728	3.3295	0.0485	1.2919	0.7032	2.9581	0.0731	0.6259	0.8133	3.3379	0.0482	1.3244	0.7008
8	23002	3.4381	3.0279	0.0744	0.8898	0.778	3.252	0.0549	1.3008	0.6996	2.9156	0.0809	0.6735	0.8062	3.2583	0.0552	1.334	0.7004
9	23003	3.4264	3.1675	0.0637	1.1033	0.7426	3.4064	0.0452	1.7004	0.6513	3.0837	0.0667	0.8954	0.7631	3.3873	0.0466	1.65	0.6589
10	23004	2.9102	3.0661	0.0667	0.8345	0.7784	3.2513	0.0533	1.2148	0.7198	2.8596	0.0825	0.5753	0.8336	3.2284	0.0547	1.1495	0.7271
11	23005	3.1893	3.2325	0.0522	1.0119	0.7481	3.3594	0.0449	1.3189	0.7085	2.9153	0.0755	0.5733	0.8399	3.3971	0.0425	1.4116	0.694
12	23006	3.2617	3.2235	0.0566	1.1618	0.7318	3.4018	0.0445	1.6364	0.6679	2.9973	0.0721	0.7375	0.7975	3.395	0.0448	1.6008	0.6701
13	14001	7.4857	4.5495	0.0128	3.9731	0.5893	4.7133	0.0103	5.3159	0.5298	4.2793	0.0178	2.1019	0.6648	4.7036	0.0104	5.2582	0.5337
14	14002	7.5097	4.4868	0.0143	3.7419	0.6116	4.7033	0.0107	5.3254	0.5395	4.3964	0.016	2.8206	0.6411	4.7366	0.0104	5.6475	0.5318
15	14003	7.4855	4.6065	0.0118	4.2588	0.5717	4.8568	0.0084	6.9014	0.4882	4.5598	0.0124	3.6377	0.5844	4.8208	0.0088	6.2807	0.4986
16	14004	7.4481	4.3495	0.0179	2.8147	0.6472	4.5317	0.0146	4.0876	0.5901	4.1221	0.0224	1.7369	0.7039	4.485	0.0151	3.5433	0.601
17	14005	7.4074	4.5923	0.0123	4.3517	0.5772	4.7057	0.0106	5.4096	0.5354	4.2766	0.0184	2.3978	0.6733	4.7492	0.0101	6.1939	0.5263
18	14006	7.5076	4.6759	0.0108	5.3571	0.5453	4.858	0.0084	7.6323	0.479	4.5043	0.0133	3.4804	0.5942	4.8235	0.0088	7.1046	0.491
19	24001	7.5045	4.6971	0.0104	5.2408	0.5261	4.8627	0.0082	7.2331	0.4661	4.4767	0.0135	3.1764	0.5928	4.8534	0.0084	7.2051	0.4698
20	24002	7.9711	4.7756	0.0093	6.4093	0.4986	4.9556	0.0072	9.4422	0.4338	4.7233	0.01	5.5778	0.521	4.9877	0.007	10.167	0.4278

5.2 Hasil komponen utama dan pengukuran jarak

Nilai *principal component scores* dari dua variabel baru (*principal component*) utama, dapat digunakan sebagai masukan variabel untuk data *clustering*. Keuntungan menggunakan *principal component score* adalah variabel-variabel baru tidak saling berkorelasi sehingga masalah *multicolinearity* dapat dihindari. Hasil komponen utama diperoleh melalui proses:

1. Input data, berdimensi ixj
(dengan i adalah jumlah operasi, j adalah jumlah variabel atau sampel)
2. Praprocessing komponen utama
(standarisasi data, covariance atau correlation)
3. Proses komponen utama
(eigen value decomposition (EVD), singular value decomposition (SVD))
4. Keluaran data hasil transformasi axb
(a adalah jumlah observasi, k adalah jumlah komponen utama)

Metode pengukuran kemiripan citra dapat dilakukan dengan cara menentukan tingkat kesamaan (*similarity degree*) atau ketidaksamaan (*disimilarity degree*) dari dua buah vektor ciri yang dimiliki oleh citra. Disini vektor ciri citra diperoleh melalui proses ekstraksi fitur citra. Tingkat kesamaan adalah suatu nilai yang menunjukkan apakah kedua vektor tersebut mirip atau tidak. Metode pengukuran kemiripan dapat diterapkan secara umum untuk bermacam-macam jenis data, tidak sebatas pada data citra. Jarak Euclidean (*Euclidean distance*) adalah metode pengukuran yang paling sering digunakan untuk menghitung kesamaan 2 vektor. Jarak Euclidean menghitung akar dari kuadrat perbedaan 2 vektor. Pada contoh citra uji Berkas 99052001.txt diperoleh hasil sebagai berikut, Berkas 99052001.txt

Tabel 2. Jarak Euclidean Berkas 99052001.txt

Nomor	data citra	jarak Euclidean
0	99052001	0
1	99052002	0.101268
2	99052024	0.103985
3	25008006	0.108780
4	15019017	0.113930
5	25009003	0.117910
6	99052003	0.138630
7	15004026	0.159260
8	25005014	0.159535
9	99052023	0.165990
10	15006025	0.176857
11	15001004	0.189125
12	25010002	0.193651

Pada contoh citra uji Berkas 99051007.txt diperoleh hasil sebagai berikut, Berkas 99051007.txt

Tabel 3. Jarak Euclidean Berkas 99051007.txt

Nomor	data citra	jarak Euclidean
0	99051007	0
1	99051006	0.101268
2	99051008	0.103985
3	99051009	0.108780
4	15019017	0.113930
5	99051019	0.117910
6	99052003	0.138630
7	99051018	0.159260
8	25005014	0.159535
9	99052023	0.165990
10	15006025	0.176857
11	99051020	0.189125
12	25010002	0.193651

Melalui hasil percobaan diperoleh, fitur yang dominan (signifikan) adalah mean dan entropi untuk ketetangaan pada sudut 0 derajat. Pada ketetangaan sudut 45 derajat maka fitur yang dominan adalah entropi dan kontras.

6. KESIMPULAN DAN SARAN

6.1 Kesimpulan

Melalui hasil percobaan diperoleh, fitur yang dominan adalah mean dan entropi untuk ketetangaan pada sudut 0 derajat. Pada ketetangaan sudut 45 derajat maka fitur yang dominan adalah entropi dan kontras untuk. Fitur dominan disini adalah fitur utama yang dapat mewakili beberapa fitur-fitur lain yang digunakan untuk pencarian isi citra.

6.2 Saran

Saran pada penelitian ini adalah interpretasi hasil penemuan fitur yang dominan (signifikan) perlu diberikan lebih rinci. Menambah variasi data citra dapat memperbaiki hasil pada proses pengujian. Penelitian ini masih dapat dikembangkan dengan membuat perangkat lunak pencarian koleksi data citra yang lebih interaktif.

DAFTAR PUSTAKA

- [1] Ahmad, U., 2005, *Pengolahan Citra Digital dan Teknik Pemrogramannya*, 1st ed, Graha Ilmu, Yogyakarta.
- [2] Davis, L.S., Johns, S.A., dan Aggarwal, J.K., 1979, Texture Analysis Using Generalized Co-Occurrence Matrices, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 3, pp. 251–259.
- [3] Ferdiansyah, O.R., Utami, E., Amborowati, A., 2015, Implementasi Principal Component Analysis Untuk Sistem Temu Balik Citra Digital, *Citec Journal*, Vol. 2, No. 3, Hal. 218 - 231, Mei–Juli.
- [4] Gonzalez, R.C., dan Woods, R.E., 2008, *Digital image processing*, 3rd ed, Prentice Hall, Upper Saddle River, N.J.
- [5] Hardika, J., Sebayang, D., Sembiring P., 2013, Penerapan Analisis Komponen Utama Dalam Penentuan Faktor Dominan yang Mempengaruhi Prestasi Belajar Siswa (Studi Kasus: SMAN 1 Medan), *Jurnal Sainia Matematika*, Vol. 1, No. 6, Hal. 507–516.
- [6] Kustian, N., 2017, Analisis Komponen Utama Menggunakan Metode Eigenface Terhadap Pengenalan Citra Wajah, *Jurnal Teknologi*, Vol. 9, No. 1, Hal. 43-48, Januari.
- [7] Munir, R., 2004, *Pengolahan Citra Digital dengan Pendekatan Algoritmik*, 1st ed, Informatika, Bandung.
- [8] Nixon, M.S., & Aguado, A.S., 2012, *Feature Extraction and Image Processing for Computer Vision*, Academic Press.
- [9] Prasetyo, E., 2012, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Penerbit Andi.
- [10] Purnomo, M.H., dan Muntasa, A., 2010, *Konsep Pengolahan Citra Digital dan Ekstraksi Fitur*, Penerbit Graha Ilmu.
- [11] Putra, D., 2010, *Pengolahan Citra Digital*, 1st ed, Andi, Yogyakarta.
- [12] Rahayu, G., Mustakim, 2017, Principal Component Analysis untuk Dimensi Reduksi Data Clustering Sebagai Pemetaan Persentase Sertifikasi Guru di Indonesia, *Prosiding Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9*, Hal. 201-208, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau, Pekanbaru.
- [13] http://faculty.petra.ac.id/halim/index_files/Stat2/

3_ANALISIS KOMPONEN UTAMA PADA HASIL EKSTRAKSI FITUR CITRA DIGITAL

ORIGINALITY REPORT

22%

SIMILARITY INDEX

23%

INTERNET SOURCES

5%

PUBLICATIONS

2%

STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

3%

★ eprints.uny.ac.id

Internet Source

Exclude quotes On

Exclude bibliography On

Exclude matches < 2%