

PENGENALAN POLA BENTUK BUNGA MENGUNAKAN PRINCIPLE COMPONENT ANALYSIS DAN K-NN

Herfina ¹⁾

¹⁾ Program Studi Ilmu Komputer, FMIPA Universitas Pakuan
Jl. Pakuan PO BOX 452, Ciheuleut Bogora
email : herfinario@yahoo.com¹⁾

Abstrak

Pengenalan pola atau *pattern recognition* merupakan teknik yang bertujuan untuk mengklasifikasikan citra yang telah diolah sebelumnya berdasarkan kesamaan atau kemiripan ciri yang dimilikinya. Dengan banyaknya jenis dan bentuk bunga maka dapat diklasifikasikan berdasarkan kesamaan cirinya melalui proses pengenalan pola. Tujuan penelitian ini adalah melakukan pengenalan pola bentuk citra bunga. Metode yang digunakan pada penelitian ini adalah *Principle Component Analysis (PCA)* dan *k-nearest neighbor (K-NN)*. *PCA* digunakan untuk proses ekstraksi fitur yaitu untuk mendapat penciri dari citra bunga tersebut. Sedangkan *K-NN* digunakan untuk proses klasifikasi citra bunga. Hasil evaluasi pengenalan pola pada bentuk citra bunga menggunakan *K-NN* berdasarkan ekstraksi ciri dengan *PCA* menghasilkan akurasi sebesar 97,70%.

Kata kunci :

pengenalan pola, bunga, *PCA*, *K-NN*

1. Pendahuluan

Bunga merupakan bagian dari tumbuhan yang memiliki jenis dan bentuk yang beragam. Untuk mengetahui pola bentuk dari satu jenis bunga tahap pertama adalah mengetahui ciri dari citra bunga tersebut melalui proses ekstraksi ciri. Kemudian tahap selanjutnya adalah mengklasifikasikannya berdasarkan kesamaan atau kemiripan ciri yang dimilikinya. Proses ini dikenal dengan pengenalan pola atau *pattern recognition*.

Pada pengenalan pola bunga salah satu metode ekstraksi ciri yang dapat digunakan adalah metode *Principle Component Analysis (PCA)*. Sedangkan untuk klasifikasinya dapat digunakan *K-Nearest Neighborhood (K-NN)*. Melalui *PCA* setiap titik sampel data citra bunga ditransformasikan berdasarkan *principal component*-nya sehingga variasi penyebaran data hasil transformasi yang dimensinya lebih kecil dapat mewakili variasi penyebaran data asli yang dimensinya jauh lebih besar [7].

KNN digunakan pada pengenalan pola untuk melakukan klasifikasi terhadap citra bunga berdasarkan

jarak terdekat antara data citra bunga yang akan dievaluasi dengan *K* tetangga (*neighbor*) terdekatnya dalam data training. *Nearest Neighborhood* didasarkan pada suatu asumsi bahwa sekumpulan sesuatu yang mirip (mendekati) mestinya merupakan satu kelas yang sama [1].

2. Tinjauan Pustaka

Principle Component Analysis (PCA)

Principle Component Analysis (PCA) adalah teknik statistik untuk menyederhanakan kumpulan data banyak-dimensi menjadi dimensi yang lebih rendah (*extration feature*) [6]. *PCA* merupakan transformasi linier ortogonal yang mentransformasi data kesistem koordinat baru, sehingga keragaman terbesar dengan suatu proyeksi berada pada koordinat pertama (disebut *principal komponen pertama*), keragaman terbesar kedua berada pada koordinat kedua dan seterusnya. Konsep penggunaan *PCA* meliputi perhitungan nilai-nilai simpangan baku, matriks kovarian, nilai karakteristik (*eigen value*) dan vector karakteristik (*eigen vector*). *PCA* dapat menggunakan metode kovariansi atau korelasi [6]. Jika diperlukan, data distandardisasi terlebih dahulu sehingga menghampiri sebaran normal baku. Dalam hal ini digunakan metoda kovariansi dengan algoritma berikut:

- Mengumpulkan data dalam bentuk matrix tingkat-keabuan *X* berukuran *M x N*. Misalkan adalah vector *N x 1* :

- Menghitung rata-rata:

$$\bar{x} = \frac{1}{M} \sum_{i=1}^M x_i \dots\dots\dots(1)$$

- Menghitung selisih rata-rata:

$$\Phi_i = x_i - \bar{x} \dots\dots\dots(2)$$

- Menentukan matriks kovarian
Dari matriks $X = [\begin{matrix} 1 & 2 & \dots & M \end{matrix}]$ (matriks $N \times M$),
Hitung kovarian:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = XX^T \dots\dots\dots(3)$$

- Menentukan nilai karakteristik dan vector karakteristik dari matrik kovarian

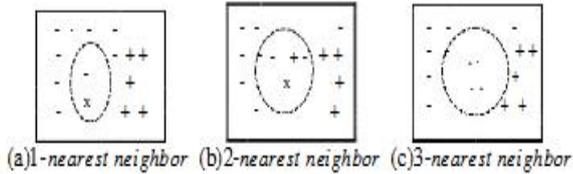
$$C: \lambda_1 > \lambda_2 > \dots > \lambda_N \text{ dan } C: u_1, u_2, \dots, u_n \dots \dots \dots (4)$$

- Mengurutkan vector karakteristik u dan nilai karakteristik dalam matriks diagonal dalam urutan menurun sesuai dengan nilai peluang kumulatif terbesar untuk tiap vector karakteristik sehingga diperoleh nilai-nilai karakteristik yang dominan.

K-Nearest Neighbour (KNN)

Algoritme k-nearest neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut. Pengelompokan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (neighbor) terdekat. Dalam hal ini jumlah data/tetangga terdekat ditentukan oleh user yang dinyatakan dengan k. Misalkan ditentukan k=5, maka setiap data testing dihitung jaraknya terhadap data training dan dipilih 5 data training yang jaraknya paling dekat kedata testing.

Setelah diperiksa output atau labelnya masing-masing, kemudian tentukan output mana yang frekuensinya paling banyak.. Representasi KNN dengan 1, 2 dan 3 tetangga data terhadap data baru x ditunjukkan pada Gambar 1 [4].



Gambar 1 Ilustrasi nearest neighbour terhadap data baru (x).

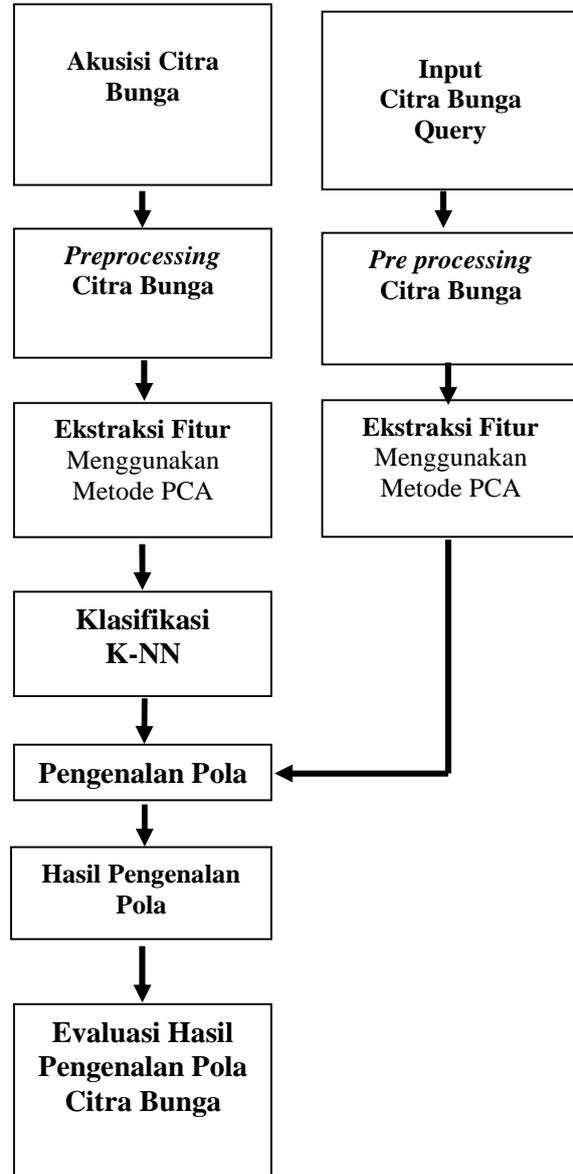
Untuk mendefinisikan jarak antara dua titik yaitu titik pada data training (x) dan titik pada data testing (y) maka digunakan rumus Euclidean.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (5)$$

Dengan d adalah jarak antara titik pada data training x dan titik data testing y yang akan diklasifikasi, dimana $x = x_1, x_2, \dots, x_i$ dan $y = y_1, y_2, \dots, y_i$ dan I merepresentasikan nilai atribut serta merupakan dimensi atribut [2].

3. Metode Penelitian

Tahapan penelitian yang dilaksanakan ditunjukkan pada Gambar 2.



Gambar 2 Tahapan Penelitian

Akuisisi Citra Bunga

Akuisisi citra bunga dilakukan menggunakan sensor larik (sensor array) yaitu kamera digital. Data yang digunakan pada penelitian sebanyak 300 data, terdiri atas 10 jenis bunga yaitu: bunga dahlia, bunga sepatu, bunga mawar, bunga anggrek bulan, bunga kamboja, bunga matahari, bunga teratai, bunga melati, bunga aster dan bunga anyelir. Masing masing bunga terdiri 30 sampel. Pada penelitian ini data dibagi menjadi

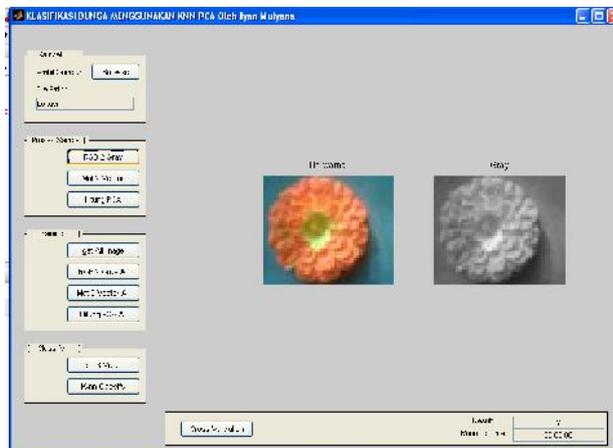
data latih dan data uji masing-masing 67% dan 33 % (20 data latih dan 10 data uji).

Pre processing Citra Tumbuhan Obat

Preprocessing yang dilakukan pada penelitian adalah menseragamkan ukurannya menjadi 40 x 50 piksel dan mengubah bentuk citra ke dalam format *grayscale* 8 bit. ($2^8 = 256$ derajat keabuan). Untuk mengubah warna RGB ke *grayscale* digunakan persamaan 6 [3].

$$\text{Gray} = 0,2989 \times R + 0,5870 \times G + 0,1140 \times B \dots\dots\dots(6)$$

Pada penelitian ini tahap *Preprocessing* telah dilaksanakan secara otomatis menggunakan bahasa pemrograman matlab versi 7.7.0 (R2008b). Seperti di tunjukan pada Gambar 3.



Gambar 3 Tampilan citra bunga dalam warna asli dan *grayscale*

Ekstraksi Fitur Citra Bunga dengan PCA

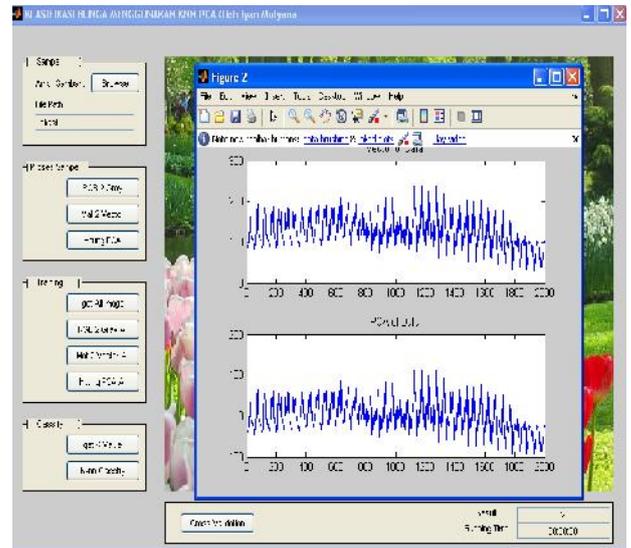
Fitur merupakan karakteristik unik dari suatu objek [5]. Ekstraksi fitur adalah proses mendapatkan penciri atau fitur dari suatu citra. Ada beberapa bagian citra yang dapat dijadikan fitur citra, antara lain bentuk dan tekstur.

Secara umum tahapan tahapan ekstraksi fitur yang dilakukan adalah sebagai berikut :

- Data Citra bunga direpresentasikan ke dalam Matriks X berdimensi 40 x 50.
- Matriks X direduksi menjadi matriks Y menggunakan algoritma PCA.
- Mengambil beberapa nilai ciri yang merepresentasikan satu citra bunga

Seperti pada tahap *preprocessing*, tahap ekstraksi fitur citra bunga menggunakan metode PCA diimplementasikan menggunakan bahasa pemrograman matlab versi 7.7.0 (R2008b). Hasil dari Ekstraksi Fitur

Citra Bunga menggunakan Metode PCA ditunjukan pada Gambar 4.

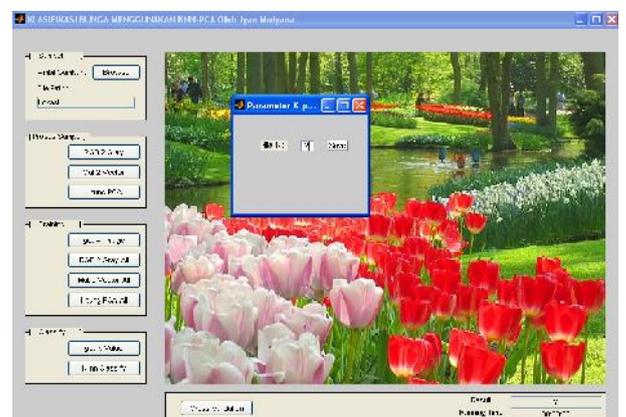


Gambar 4 Tampilan hasil PCA dari citra bunga

Klasifikasi Citra Bunga Menggunakan K-NN

Pada Penelitian ini Algoritma *k-nearest neighbor* (*k*-NN atau KNN) digunakan untuk melakukan klasifikasi terhadap data training dari citra bunga. Pada fase training, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data training citra bunga. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data testing (yang klasifikasinya tidak diketahui).

Jarak dari vektor yang baru ini terhadap seluruh vektor data training dihitung, dan sejumlah *k* buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut. Tampilan aplikasi untuk tahap Klasifikasi citra bunga dengan K-NN ditunjukan Pada Gambar 5.

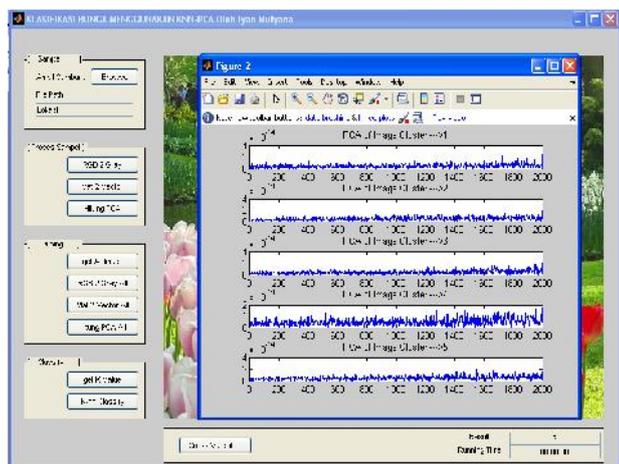


Gambar 5 Tampilan aplikasi klasifikasi citra bunga dengan K-NN

4. Hasil dan Pembahasan

Hasil Ekstraksi Fitur Menggunakan Metode PCA

Ekstraksi fitur menggunakan metode PCA menghasilkan vektor ciri untuk masing masing citra bunga seperti ditunjukkan pada Gambar 6. Citra bunga yang memiliki pola bentuk seragam memiliki vektor ciri yang hampir sama.



Gambar 6 Vektor ciri hasil ekstraksi ciri menggunakan metode PCA

Hasil Klasifikasi Citra Bunga dengan Metode K-NN untuk data training

Hasil klasifikasi untuk 200 data training citra bunga menggunakan metode K-NN diperoleh tingkat akurasi 97,70 %. *Confussion matrix* hasil klasifikasi ditunjukkan pada Tabel 1.

Tabel 1 *Confussion matrix* hasil klasifikasi citra bunga dengan K-NN

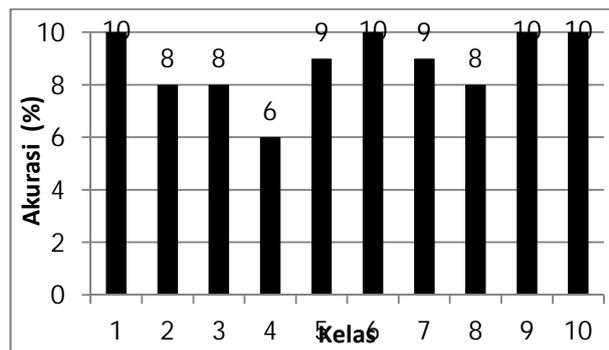
	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	Jumlah
c1	20										20
c2		17								3	20
c3			15				5				20
c4			2	14				4			20
c5					19			1			20
c6						20					20
c7						1	19				20
c8				5				15			20
c9									20		20
c10										20	20
Jumlah	20	17	17	19	19	21	24	20	20	23	200

Pada *Confussion matrix* (Tabel 1) terdapat empat kelas dengan hasil klasifikasi 100% yaitu kelas 1 (bunga dahlia), kelas 6 (bunga matahari), kelas 9 (bunga aster) dan kelas 10 (bunga anyelir).. Sedangkan satu kelas dengan hasil klasifikasi terendah yaitu kelas 4 (bunga anggrek bulan) dengan akurasi 70 %. Kelas dengan akurasi 100 % pada umumnya memiliki pola

bentuk dan vektor ciri yang hampir seragam. Sedangkan kelas dengan klasifikasi paling rendah yaitu kelas 4 (bunga anggrek bulan) memiliki pola bentuk vektor ciri yang hampir tidak seragam.

Hasil Pengenalan Pola Bentuk Bunga dengan metode K-NN untuk data testing

Setelah dilaksanakan pengenalan pola bentuk terhadap 100 data uji citra bunga diperoleh akurasi untuk setiap kelas seperti ditunjukkan pada Gambar 7.



Gambar 7 Grafik akurasi pengenalan pola bentuk setiap kelas citra bunga

Pada Gambar 7 dapat dilihat bahwa citra bunga yang memiliki tingkat akurasi tinggi seperti ditunjukkan pada tabel 1 memiliki tingkat pengenalan pola yang tinggi demikian pula sebaliknya. Hal ini disebabkan citra bunga dengan bentuk seragam memiliki pola vektor ciri yang hampir sama sehingga lebih mudah dikenali pada saat proses pengenalan pola.

5. Kesimpulan dan Saran

Kesimpulan

Berdasarkan penelitian yang dilakukan dapat diperoleh kesimpulan sebagai berikut:

1. Metode PCA dan klasifikasi dengan K-NN dapat diimplementasikan untuk mengembangkan sistem pengenalan pola bentuk citra bunga
2. Hasil ekstraksi ciri dengan PCA dipengaruhi oleh pola bentuk citra bunga. Makin seragam pola bentuk citra bunga maka memiliki vektor ciri yang hampir sama sehingga lebih mudah dikenali pada saat proses pengenalan pola
3. Hasil evaluasi pengenalan pola pada citra bunga menggunakan klasifikasi K-NN berdasarkan ekstraksi ciri dengan PCA menghasilkan akurasi sebesar 97,70%.

Saran

Dalam penelitian ini tahapan *preprocessing* citra yang dilakukan hanya menseragamkan ukuran citra 40 x 50 piksel dan mengubah bentuk citra ke dalam format

grayscale. Untuk mendapatkan kualitas citra yang lebih baik sebelum di ekstraksi perlu dilakukan beberapa tahapan *preprocessing* yang lainnya seperti segmentasi citra, perbaikan citra (*enhancement*) dan normalisasi arah citra dengan perputaran rotasi.

Daftar Pustaka

- [1] Barber, David., 2005, *Learning from Data Nearest neighborhood Classification*. <http://www.enc.ac.uk/~amos/ldf>.
- [2] Han j, Kamber M (2001). *Data Mining : Concepts and Techniques* , Morgan Kaufmann Publishers
- [3] Jianxin Xu, Wang Hua, Fang Hui. 2011. Characterization of Periodic, Quasiperiodic and Chaotic States in Nonpremixed Biodiesel/Air Jet Flames. *Hindawi Publishing Corporation* 2011 (861436) <http://www.hindawi.com/journals/mpe/2011/861436.htm> [13 Feb 2012]
- [4] Pramudiono, 2004, *Pengantar Data Mining*. Universitas Katolik Parahiyangan Bandung.
- [5] Putra D. 2010. *Pengolahan Citra Digital*. Yogyakarta : Penerbit ANDI.
- [6] Scrofano, Roland., Klassen Myungsook., 2001, “*Pattern Recognition with Feedforward, Artificial Neural Networks*”, California Lutheran University.
- [7] Smith, Lindsay., 2002, *A Tutorial on Principal Component Analysis*, Montana University.

Biodata Penulis

Herfina, memperoleh gelar Sarjana Komputer (S.Kom), Program Studi Manajemen Informatika Universitas Gunadarma, lulus tahun 1997. Tahun 2010 memperoleh gelar Magister Pendidikan (M.Pd) dari Program Pascasarjana Manajemen Pendidikan Universitas Pakuan. Saat ini sebagai Staf Pengajar Program Studi Ilmu Komputer, Fakultas MIPA Universitas Pakuan.

