

KLASIFIKASI GAMBAR MENGGUNAKAN METODE K-NEAREST NEIGHBOR

IMAGE CLASSIFICATION USING K-NEAREST NEIGHBOR METHOD

Yessi Yunitasari¹
Universitas PGRI Madiun¹
e-mail: yessi@unipma.ac.id

Abstract: *Classification is an important process to recognize and distinguish one thing from another, this can be humans, animals, or plants. This identification is done by recognizing the characteristics that something has. Indonesia has a variety of citrus varieties. These many varieties make it difficult to distinguish one orange variety from another. One way to classify citrus fruits can be done using the k-nearest neighbor method based on images. Image processing focuses on efforts to transform an image or image into another image by using certain techniques. Image processing stage is the first, namely image acquisition. Image acquisition is the initial stage for obtaining digital images. Preprocessing aims to improve image quality. Important things that are done in this process include size transformation, change to gray level, quality improvement (histogram equalization). After the preprocessing process, it is continued with the feature extraction process. Then proceed with the classification process using the k-nearest neighbor method.*

Keywords: *classification, k-nearest neighbor, size transformation, gray level, histogram equalization.*

Abstrak: Klasifikasi merupakan proses yang penting untuk mengenali serta membedakan sesuatu hal dengan hal lainnya, hal ini dapat berupa manusia, hewan, atau tumbuhan. Identifikasi ini dilakukan dengan mengenali ciri khas yang dimiliki sesuatu hal tersebut. Negara Indonesia memiliki beragam varietas jeruk. Varietas yang banyak ini mengakibatkan susahnyanya membedakan varietas jeruk satu dan lainnya. Salah satu cara untuk mengklasifikasikan buah jeruk bisa dilakukan menggunakan metode k-nearest neighbor dengan berbasis citra. Pengolahan citra berfokus pada usaha untuk melakukan transformasi suatu citra atau gambar menjadi citra lain dengan menggunakan teknik tertentu. Tahapan pengolahan citra adalah yang pertama yaitu akuisisi citra. Akuisisi citra adalah tahap awal untuk mendapatkan citra digital. Preprocessing bertujuan untuk meningkatkan kualitas citra. Hal-hal penting yang dilakukan pada proses ini diantaranya adalah transformasi ukuran, perubahan ke gray level, peningkatan Kualitas (equalisasi histogram). Setelah proses preprocessing dilanjutkan dengan proses ekstraksi fitur. Kemudian dilanjutkan dengan proses klasifikasi dengan metode k-nearest neighbor.

Kata kunci: klasifikasi, k-nearest neighbor, transformasi ukuran, gray level, equalisasi histogram.

PENDAHULUAN

Pemrosesan gambar merupakan salah satu fokus ilmu dari pengenalan pola. Pemrosesan gambar secara spesifik dalam sistem ini meliputi tahap preprocessing, ekstraksi fitur dan kemudian klasifikasi tanpa melalui segmentasi. Pada tahap preprocessing dilakukan manipulasi data gambar agar data gambar mempunyai dimensi ukuran yang sama. Termasuk di dalam tahap preprocessing ada tahap grayscale agar data gambar hanya mempunyai intensitas untuk memperkecil kompleksitas pada komputasi. Kemudian tahap selanjutnya dilakukan thresholding terhadap warna dengan intensitas lebih dari nilai tertentu agar data intensitas background tidak disertakan dalam komputasi. Tahap preprocessing ini dilakukan oleh library dari python yang bernama scikit-image, baik untuk grayscale dan normalisasi.

Fitur yang dihasilkan dari tahap sebelumnya adalah intensitas pada masing-masing piksel. Setiap data gambar akan menghasilkan matriks berukuran 100x100 yang merepresentasikan intensitas (di dalam library direpresentasikan antara 0-1) untuk setiap piksel pada koordinat tertentu. Proses ekstraksi fitur intensitas dilakukan untuk semua

gambar, baik itu data uji maupun data training untuk kemudian dibandingkan jarak kemiripannya. Hasil dari perhitungan matriks gambar yang berukuran 100x100 tersebut yang akan dijadikan sumber dari proses klasifikasi. Algoritma yang digunakan adalah k-Nearest Neighbor dengan nilai k dilakukan eksperimen untuk sistem ini. Algoritma ini bekerja dengan mencari sejumlah k tetangga terdekat dari data uji. Tetangga mayoritas dengan kelas tertentu akan dijadikan kelas dari data uji tersebut. Ketetanggaan dalam algoritma ini didasarkan dari jarak tertentu yang dihitung dengan jumlah selisih absolut dari 2 data atau biasa disebut L1 distance.

KAJIAN TEORI

A. Citra

Suatu citra dapat dikenali dengan cara mengekstrak cirinya yaitu dengan mengambil beberapa bagian dari citra tersebut seperti warna, pola citra, diameter fisik, bentuk dan lain-lain. Teknik yang digunakan adalah dengan mengekstrak citra RGB (Red-Green-Blue) (Manik & Saragih, 2017).

Pada model HSV (Hue, Saturation dan Value) Hue menyatakan warna sebenarnya, Saturation menyatakan tingkat kemurnian suatu warna, yaitu warna putih diidentifikasi seberapa banyak diberikan pada warna, Value adalah atribut yang menyatakan banyaknya cahaya yang diterima oleh sebuah warna (Syahid et al., 2016).

B. Histogram Equalization

Histogram equalization bertujuan untuk meratakan distribusi nilai piksel sehingga kontras dari sebuah citra dapat diperbaiki (Yoon et al., 2009). Proses histogram equalization ditunjukkan dalam persamaan (1). Proses ekualisasi ini memanfaatkan Cumulative Distribution Function (CDF) dari nilai piksel pada sebuah citra yang dapat dilihat pada persamaan (2). Histogram equalization ini memperbaiki kontras sehingga berpotensi membantu proses klasifikasi.

$$s_k = T(r_k) = \text{round} \left(\frac{C(r_k) - 1}{n - 1} \times L \right) \quad (1)$$

$$C(r_k) = \sum_{i=0}^k \frac{n_i}{n} \quad (2)$$

s_k = tingkat keabuan sebuah citra setelah ekualisasi

r_k = tingkat keabuan sebuah citra

$T(r_k)$ = transformasi dari ekualisasi

$C(r_k)$ = CDF dari r_k

n = jumlah piksel dari sebuah citra

L = rentang nilai tingkat keabuan

$k = 0, 1, 2, \dots, L-1$

C. K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing – masing dimensi merepresentasikan fitur dari data. Algoritma KNN termasuk metode yang menggunakan algoritma supervised. Perbedaan antara supervised learning dengan unsupervised learning adalah pada supervised learning bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada unsupervised learning, data belum memiliki pola apapun, dan tujuan unsupervised learning untuk menemukan pola dalam sebuah data (Huchon et al., 1981).

Tujuan dari algoritma KNN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training samples. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari contoh data uji yang baru. Jarak yang digunakan adalah jarak Euclidean Distance. Jarak Euclidean adalah jarak yang paling umum digunakan pada data numeric. Algoritma KNN merupakan algoritma yang menentukan nilai jarak pada pengujian data testing dengan data training berdasarkan nilai terkecil dari nilai ketetanggaan terdekat (Ismanto & Wardoyo, 2016).

METODE

1. Preprocessing Dan Ekstraksi Fitur

1.1 Grayscale

Citra Grayscale, suatu citra dikatakan sebagai citra grayscale apabila sebuah citra tidak memiliki warna RGB atau dapat dikatakan sebuah citra yang memiliki nilai dari putih yang memiliki intensitas paling besar sampai hitam yang memiliki intensitas paling rendah. Citra grayscale terdiri dari x dan y dalam koordinat spasial dan memiliki nilai intensitas antara 0 (hitam) hingga 255 (putih) dalam citra 8 bit (Putra, 2010). Khusus dalam library scikit-image direpresentasikan dengan bilangan float antara 0 sampai 1. Grayscale lebih mudah diproses karena mengandung nilai lebih sedikit, yaitu 8 bit warna dibandingkan RGB yang mengandung 24 bit warna. Untuk mengubah citra berwarna yang mempunyai nilai matrik masing-masing R, G, dan B menjadi citra grayscale dengan nilai s , maka konversi dapat dilakukan dengan mengambil rata-rata dari nilai r, g , dan b .

1.2 Histogram equalization

Tahap Histogram equalization adalah sebuah proses yang mengubah distribusi nilai derajat keabuan pada sebuah citra sehingga menjadi seragam (uniform). Tujuan dari histogram equalization adalah untuk memperoleh penyebaran histogram yang merata sehingga setiap derajat keabuan memiliki jumlah piksel yang relatif sama. Perataan histogram diperoleh dengan cara mengubah derajat keabuan sebuah piksel (r) dengan derajat keabuan yang baru (s) dengan sebuah fungsi transformasi T (Rafael C Gonzalez, 2009). Nilai equalisasi histogram merupakan rentang antara 0-1. 0 hitam dan 1 semakin berwarna putih. Hasil equalisasi adalah matriks ukuran 120×120 . Hasil matriks ini sudah menjadi fitur ekstraksinya yang nantinya akan dihitung jaraknya menggunakan euclidean distance dalam klasifikasi k-nearest neighbor.

2. Klasifikasi K-NN

Pada tahap klasifikasi, data testing akan dibentuk matriks grayscale atau intensitas kemudian dibandingkan dengan seluruh data training untuk memeriksa data manakah yang lebih dekat dengan data yang akan di test tersebut. Prinsip dari K-NN adalah memeriksa apakah data testing tersebut dekat dengan sejumlah k tetangga terdekat. Jumlah kelas yang mempunyai lebih banyak kedekatan dengan data tersebut dianggap merupakan kelas yang tepat.

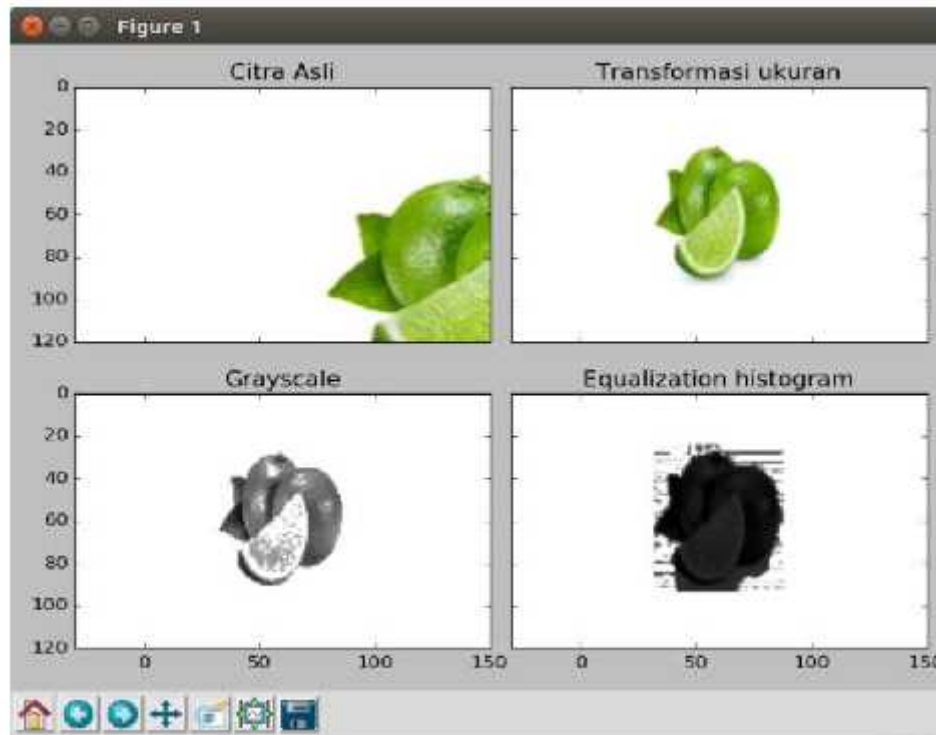
Penentuan kedekatan menggunakan L1 distance. Langkah awal dari tahap ini adalah dengan membentuk matriks Intensitas yang merupakan matriks 100×100 . Langkah selanjutnya adalah menggunakan L1 distance untuk perbedaan piksel. Tahap terakhir yaitu dengan melakukan sorting atau pengurutan terhadap data-data jarak yang telah dihitung berdasarkan jarak paling pendek. Berdasarkan data yang telah diurutkan tersebut dapat dilihat jumlah k tetangga paling dekat dari data tersebut untuk menentukan kelas.

HASIL

1. Contoh Pengambilan fitur data uji

Gambar yang akan diklasifikasikan merupakan gambar jeruk nipis dan jeruk purut. Masing-masing kelas terdiridari 8 gambar sehingga total keseluruhan adalah 16 gambar. Masing-masing gambar akan dilakukan transformasi ukuran menjadi 120×120 piksel. Setelah proses transformasi ukuran maka dilakukan proses gray

level untuk membuat gambar menjadi grayscale. Hasil dari grayscale akan dilakukan equalisasi histogram. Sehingga didapatkan hasil preprocessing dan ekstraksi fitur seperti Gambar 1 dibawah ini :



Gambar 1 Hasil preprocessing dan ekstraksi fitur

2. Metode Pengujian K-Fold dan Hasil Pengujian

Pengujian dilakukan dengan K-fold. Metode K-Fold membagi data menjadi 2 bagian yaitu data uji dan data latih (training). Selanjutnya dilakukan silang antara data uji dan data latihnya. Dalam system ini terdapat 4 K-Fold. Seperti table dibawah ini :

Tabel 1. Pembagian K-Fold

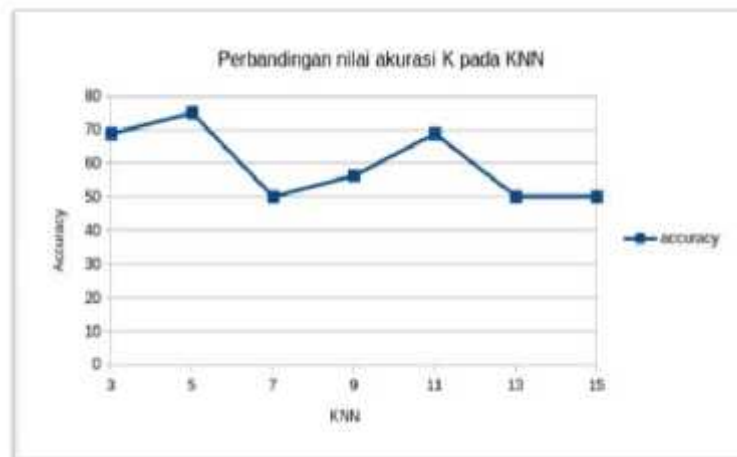
| Kelas Nipis | Kelas Purut | K-Fold |
|-------------|-------------|--------|
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 3 | 3 | 2 |
| 4 | 4 | 2 |
| 5 | 5 | 3 |
| 6 | 6 | 3 |
| 7 | 7 | 4 |
| 8 | 8 | 4 |

Hasil dari K-Fold nantinya akan didapatkan nilai akurasinya. Hasil K-Fold didapatkan dengan :

$$\text{Akurasi} : \frac{\text{Hasil K1} + \text{Hasil K2} + \text{Hasil K3} + \text{Hasil K4}}{4}$$

PEMBAHASAN

Sebelum menentukan nilai K untuk K-Nearest Neighbor maka sebelumnya dilakukan eksperimen untuk nilai k yang menghasilkan nilai akurasi tertinggi. Berikut ini adalah hasil eksperimen nilai K terhadap nilai akurasinya dapat dilihat pada Gambar 2 di bawah ini :



Gambar 2 Perbandingan nilai akurasi K pada KNN

Berdasarkan gambar diatas nilai K yang dipakai adalah 5 dengan hasil akurasinya sebesar 75 %. Rincian untuk setiap K-Fold akan dipaparkan seperti di bawah ini :

a. Hasil K-Fold 1

Hasil K-Fold 1 untuk data uji ke 0 sampai data uji ke 3 terklasifikasi dengan benar. Sehingga nilai akurasi dari K-Fold 1 adalah 100%. Hasil dari K-Fold 1 dapat dilihat pada Gambar 3.

```
Data Uji Ke - 3
label : purut
Daftar distance :
1355.67968681,1262.11929804,
837.245742215,791.6851044,
1043.58106079,1462.88428171,
1405.96454392,1159.69318832,
1170.04825105,1140.33729511,
1629.59914843,1073.00121297,

5 k terpilih : purut,nipis,nipis,purut,purut,
Diklasifikasikan dalam kelas : purut
klasifikasi benar

ACCURACY 100.0%
```

Gambar 3 Hasil dari K-Fold 1

b. Hasil K-Fold 2

Hasil K-Fold 2 ada 1 data uji yang terklasifikasi salah, yaitu data uji ke 1. Sehingga nilai akurasi dari K-Fold 2 adalah 75%. Hasil dari K-Fold 2 dapat dilihat pada Gambar 4.

```
Data Uji Ke - 3
label : purut
Daftar distance :
1501.73877809,1387.16864545,
489.797324612,791.6851044,
1159.71964021,1546.94139611,
767.756950021,508.344079447,
1782.49402091,993.409357449,
1122.58920143,725.259806596,

5 k terpilih : nipis,purut,purut,nipis,purut,
Diklasifikasikan dalam kelas : purut
klasifikasi benar

ACCURACY 75.0%
```

Gambar 4 Hasil dari K-Fold 2

c. Hasil K-Fold 3

Hasil K-Fold 3 ada 2 data uji yang terklasifikasi salah, yaitu data uji ke 0 dan ke 3. Sehingga nilai akurasi dari K-Fold 3 adalah 50%. Hasil dari K-Fold 3 dapat dilihat pada Gambar 5.

```
Data Uji Ke - 3
label : purut
Daftar distance :
1578.72675642,1461.34233267,
437.452625969,1159.69318832,
409.28729951,369.968062724,
1305.97786282,508.344079447,
1913.64051434,601.459703257,
471.404154348,610.149326547,

5 k terpilih : purut,nipis,nipis,nipis,purut,
Diklasifikasikan dalam kelas : nipis
klasifikasi salah

ACCURACY 50.0%
```

Gambar 5 Hasil dari K-Fold 3

d. Hasil K-Fold 4

Hasil K-Fold 4 ada 1 data uji yang terklasifikasi salah, yaitu data uji ke 2. Sehingga nilai akurasi dari K-Fold 4 adalah 75%. Hasil dari K-Fold 4 dapat dilihat pada Gambar 6.

```
Data Uji Ke - 3
label : purut
Daftar distance :
1680.73124094,1244.41586836,
411.513007995,1073.00121297,
778.88320896,564.534384542,
1314.99062011,725.259806596,
1065.05841651,667.043901404,
294.624667095,610.149326547,

5 k terpilih : nipis,nipis,purut,purut,purut,
Diklasifikasikan dalam kelas : purut
klasifikasi benar

ACCURACY 75.0%
*****
TOTAL ACCURACY = 75.0
```

Gambar 6 Hasil dari K-Fold 4

e. Hasil Akurasi Total

Berdasarkan hasil K-Fold dapat diperoleh nilai akurasi total yaitu :

$$\text{Akurasi} = \frac{\text{Hasil K1} + \text{Hasil K2} + \text{Hasil K3} + \text{Hasil K4}}{4}$$

$$\text{Akurasi} = \frac{100\% + 75\% + 50\% + 75\%}{4}$$

$$\text{Akurasi} = 75\%$$

KESIMPULAN

Percobaan yang telah dilakukan untuk mengklasifikasikan gambar jeruk dengan menggunakan metode K-Nearest Neighbor (KNN) memperoleh akurasi akhir dengan nilai ketetapan 5 adalah $100+75+50+75/4 = 75\%$.

DAFTAR PUSTAKA

- Huchon, D., Crozet, N., Cantenot, N., & Ozon, R. (1981). Germinal vesicle breakdown in the *Xenopus laevis* oocyte: Description of a transient microtubular structure. *Reproduction Nutrition Developpement*, 21(1), 135–148. <https://doi.org/10.1051/rnd:19810112>
- Ismanto, H., & Wardoyo, R. (2016). Analysis of C4 . 5 and K-Nearest Neighbor (KNN) Method on Algorithm of Clustering For Deciding Mainstay Area. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 18(2), 86–92. <https://doi.org/10.9790/0661-1802048692>
- Manik, F. Y., & Saragih, K. S. (2017). Klasifikasi Belimbing Menggunakan Naïve Bayes Berdasarkan Fitur Warna RGB. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(1), 99. <https://doi.org/10.22146/ijccs.17838>
- Putra, D. (2010). *Pengolahan Citra Digital*. Andi Publisher.
- Rafael C Gonzalez. (2009). *Digital Image Processing*. Pearson Education.
- Syahid, D., Jumadi, J., & Nursantika, D. (2016). Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighbor (KNN) Berdasarkan Nilai Hue, Saturation, Value (HSV). *Jurnal Online Informatika*, 1(1), 20. <https://doi.org/10.15575/join.v1i1.6>
- Yoon, H., Han, Y., & Hahn, H. (2009). Image contrast enhancement based sub-histogram equalization technique without over-equalization noise. *World Academy of Science, Engineering and Technology*, 38, 176–182. <https://doi.org/10.5281/zenodo.1084411>