

# Identifikasi Jenis Buah Apel Menggunakan Algoritma *K-Nearest Neighbor* (KNN) dengan Ekstraksi Fitur *Histogram*

Ion Ataka Halela<sup>1</sup>, Bowo Nurhadiyono, S.Si, M.Kom<sup>2</sup>, Farah Zakiyah Rahmanti<sup>3</sup>

<sup>1</sup>Fakultas Ilmu Komputer<sup>2</sup>Universitas Dian Nuswantoro

Jl. Nakula 1 No. 5-11, Jawa Tengah 50131 Telp: (024) 3517261

E-mail :

[ionataka@gmail.com](mailto:ionataka@gmail.com)<sup>1</sup>, [bowo.nurhadiono@dsn.dinus.ac.id](mailto:bowo.nurhadiono@dsn.dinus.ac.id)<sup>2</sup>, [farah\\_zakiyah@dsn.dinus.ac.id](mailto:farah_zakiyah@dsn.dinus.ac.id)<sup>3</sup>

---

## Abstrak

*Pemilihan jenis buah apel berdasarkan kriteria sangatlah mudah dilakukan manusia, tetapi tidak mudah dilakukan oleh komputer. Persepsi manusia biasanya cenderung subyektif terhadap suatu objek, hal ini dikarenakan adanya faktor komposisi warna yang dimiliki oleh objek tersebut. Oleh karena itu, diperlukan alat untuk melakukan pemilihan buah apel berdasarkan jenis – jenisnya yang dilakukan secara otomatis. Penelitian ini bertujuan untuk mengidentifikasi jenis buah apel menggunakan algoritma *K – Nearest Neighbor* (KNN) dan ekstraksi fitur histogram. Pada penelitian yang dilakukan memiliki tiga modul utama yaitu prapengolahan, ekstraksi fitur, dan klasifikasi. Data yang digunakan dalam penelitian ini ada dua yaitu data training yang berjumlah 90 dan data testing yang berjumlah 10. Hasil dari penelitian ini adalah sistem yang dapat membedakan jenis – jenis buah apel dengan tingkat akurasi sebesar 90%.*

Kata Kunci : KNN, Ekstraksi Fitur Histogram, Apel.

## Abstract

*The apples selection based on criteria are very easy to do by human but hard to do by computer. The human perception is usually subjective to the object, it because of colour composition of the object. Therefore, tools is needed to do the apples selection automatically. This research purpose is identify kind of apples using *K – Nearest Neighbor* (KNN) algorithm and histogram extraction. It has three main moduls, pre – processing, histogram feature extraction and classification. There are two data, which are use in this research, they are 90 training datas and 10 testing datas. Result from this research are the system which can distinguish type of apples with 90% accuracy level.*

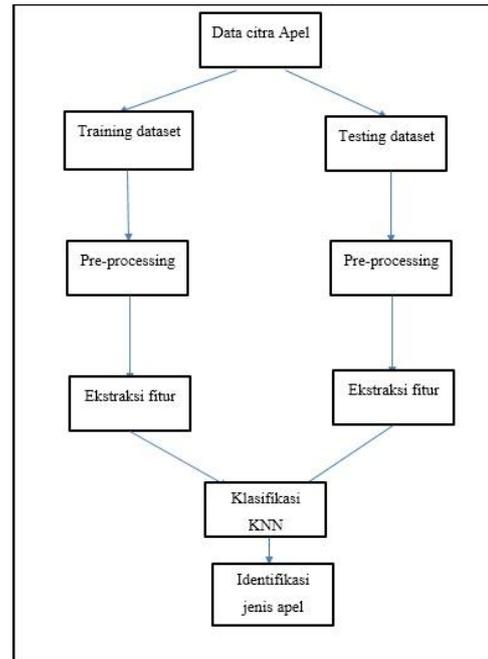
Keywords : KNN, Feature Extraction Histogram, Apple.

## 1. PENDAHULUAN

Buah apel merupakan salah satu jenis buah apel yang ada di Indonesia dan sangat di gemari oleh masyarakat umum baik muda sampai tua suka mengonsumsi apel. Hal ini menunjukkan bahwa buah apel sudah sangat dikonsumsi masyarakat secara luas dan memiliki daya saing juga[1]. Memilih jenis buah apel berdasarkan ciri – ciri sangatlah mudah dilakukan oleh manusia, tetapi tidak mudah dilakukan oleh komputer. Persepsi manusia biasanya cenderung subyektif terhadap suatu objek, hal ini dikarenakan adanya faktor komposisi warna yang dimiliki oleh objek tersebut. Oleh karena itu, diperlukan alat untuk melakukan pemilihan buah apel berdasarkan jenis – jenisnya yang dilakukan secara otomatis[2]. Penelitian ini bertujuan untuk mengidentifikasi jenis buah apel menggunakan algoritma *K – Nearest Neighbor* (KNN) dan ekstraksi fitur histogram. Algoritma *K – Nearest Neighbor* (KNN) adalah algoritma yang menggunakan algoritma *supervised* dimana hasil *query instance* diklasifikasikan berdasarkan mayoritas dari kategori pada KNN[3]. Sedangkan ekstraksi fitur histogram merupakan suatu metode untuk mendapatkan tekstur dengan dasar pada *histogram*. *Histogram* citra merupakan grafik yang menggambarkan perbedaan pada nilai – nilai intensitas piksel suatu citra atau bagian yang tertentu pada dalam suatu citra[4].

## 2. METODE PENELITIAN

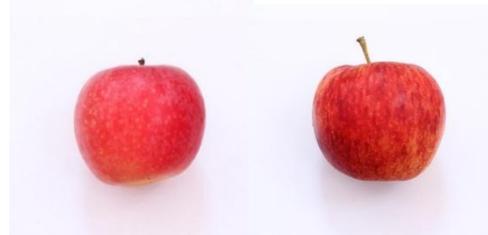
Metode penelitian identifikasi jenis buah apel menggunakan KNN dan ekstraksi fitur histogram adalah sebagai berikut:



Gambar 1 Metode yang diusulkan

## 2.1 Pengumpulan Data

Pada pengumpulan data ini dijelaskan bagaimana dan darimana data ini di ambil. Pengumpulan data premier buah apel diambil dari pedagang buah. Untuk data sekunder pengambilan data diambil dari jurnal/*paper* yang menyangkut penelitian pada buah apel dan untuk eksperimen data sebagai berikut:



Gambar 2 contoh citra apel envy dengan apel pacific rose

## 2.2 Metode Pengumpulan Data

Persiapan untuk pengambilan data adalah sebagai berikut:

- Diambil menggunakan kamera DSLR Canon 600D menggunakan lensa 18-55mm dengan menggunakan *tripod*.

- b. Ketinggian pada *tripod* setinggi 50 cm dengan kemiringan *head tripod* sekitar 30 derajat kebawah.
- c. *Setting* kamera meliputi *focal legth* lensa 35mm, ISO 400, dan diafragma  $f/6,3$ .
- d. Data di ambil di luar ruangan pada pukul 15.00 – 17.00 WIB.
- e. Buah apel diletakan diatas kertas putih HVS.
- f. Data meliputi 100 foto buah apel yang didapatkan secara acak atau *random*.

Dalam Penelitian ini menggunakan perangkat lunak sebagai berikut:

- a. Windows 7 Ultimate 64-bit digunakan sebagai sistem operasi.
- b. Matlab R2014a sebagai aplikasi untuk mengolah *dataset*.
- c. Microsoft Word sebagai penulisan dalam penelitian.

### 2.2.1 Bahan Penelitian

Adapun kebutuhan bahan yang digunakan untuk penelitian identifikasi jenis buah apel adalah menggunakan jenis buah apel envy dan apel pasific rose.

### 2.2.2 Peralatan Penelitian

Penelitian yang dilakukan untuk mengerjakan penelitian ini diantaranya:

1. Perangkat Keras
  - a. Kamera DSLR
  - b. *Processor* yang digunakan adalah Intel(R) Core(TM) i5-337U CPU @ 1.80GHz (4CPUs), ~1.8GHz.
  - c. Ram dengan ukuran 4GB
  - d. Harddisk berkapasitas 500GB
  - e. Monitor 14,1”
2. Perangkat Lunak

## 2.3 Teori Dasar

### 2.3.1 $K$ – Nearest Neighbor

Algoritma  $K$ -Nearest Neighbor (KNN) merupakan metode yang menggunakan algoritma *supervised* dimana hasil *query instance* diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Tujuan daripada algoritma KNN yaitu mengklasifikasikan objek - objek yang baru menurut atribut dan *training sample*[3].

### 2.3.2 Ekstraksi Fitur Histogram

Histogram merupakan suatu metode untuk mendapatkan tekstur dengan dasar pada *histogram*. *Histogram* citra merupakan grafik yang menggambarkan penyebaran pada nilai – nilai intensitas piksel suatu citra atau bagian yang tertentu pada dalam suatu citra. Dari segi

*histogram* dapat diketahui bahwa frekuensi kemunculan nisbi (*relative*) dari suatu intensitas pada citra tersebut. Metode *histogram* adalah metode statis orde satu untuk mendapatkan fitur – fitur tekstur[4].

Fitur – fitur yang bisa dikenali dari metode *histogram* antara lain yaitu: rerata intensitas, deviasi standar, *skewness*, energi, entropi dan kehalusan (*smoothness*). Persamaan pada perhitungan ke-6 ciri tersebut berbasis *histogram* dapat dilihat pada dibawah ini[4]:

Fitur pertama yaitu rerata intensitas rumusnya sebagai berikut:

$$m = \sum_{i=0}^{L-1} i \cdot p(i) \quad (1)$$

Dalam hal ini,  $i$  merupakan aras keabuan pada citra  $f$  dan  $p(i)$  merupakan probabilitas kemunculan  $i$  dan  $L$  merupakan nilai aras keabuan tertinggi. Rumus yang diatas akan  $energi = \sum_{i=0}^{L-1} [p(i)]^2$  (4)

menghasilkan rerata kecerahan pada citra[4].

Fitur yang kedua yaitu deviasi standar. Rumus nya sebagai berikut:

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (i - m)^2 p(i)} \quad (2)$$

Pada rumus deviasi standar,  $\sigma^2$  dinamakan varians, atau momen orde dua ternormalisasi dikarenakan  $p(i)$  adalah fungsi peluang. Fitur deviasi akan memberikan ukuran kekontrasan.

Pada fitur *skewness* adalah ukuran ketidaksimetrisan terhadap rerata intensitas. Untuk rumus fitur *skewness* adalah sebagai berikut:

$$Skewness = \sum_{i=1}^{L-1} (i - m)^3 p(i) \quad (3)$$

*Skewness* juga sering disebut dengan momen orde tiga ternormalisasi dimana nilai negatif menyatakan bahwa distribusi pada kecerahan lebih condong ke kanan terhadap rerata. Pada praktiknya, nilai *skewness* dibagi menjadi  $(L - 1)^2$  supaya lebih ternormalisasi.

Pada penjelasan untuk fitur energi adalah ukuran yang menyatakan distribusi intensitas piksel pada jangkauan aras keabuan. Rumus fitur energi adalah sebagai berikut:

Citra yang sama dengan satu nilai aras keabuan akan mempunyai nilai pada energi yang maksimum, yaitu sebesar 1. Secara umum, citra dengan sedikit aras keabuan akan mempunyai

energi yang lebih tinggi daripada yang mempunyai banyak nilai aras keabuan. Energi sering disebut juga sebagai keseragaman.

Fitur entropi mengindikasikan kompleksitas citra. Rumus daripada entropi adalah sebagai berikut:

$$entropi = \sum_{i=0}^{L-1} p(i) \log_2(p(i)) \quad (5)$$

Semakin tinggi nilai entropinya maka akan semakin kompleks citra tersebut. Seperti diketahui, entropi serta energi berkecendungan berkebalikan. Entropi juga merepresentasikan jumlah informasi yang terkandung pada sebaran data.

Pada rumusan ekstraksi fitur histogram yang terakhir yaitu fitur kehalusan (*smoothness*) merupakan fitur untuk mengukur tingkat kehalusan atau kekasaran pada intensitas citra. Penjabarannya adalah sebagai berikut:

Pada rumusan diatas,  $\sigma$  adalah deviasi standar. Berdasarkan rumus yang diatas, nilai pada R yang rendah menunjukkan bahwa citra memiliki intensitas yang kasar. Seperti yang diketahui, di dalam menghitung *smoothness*, varians perlu dinormalisasi sehingga nilainya berada di

dalam jangkauan [0-1] dengan cara membagi dengan  $(L - 1)^2$

### 2.3.3 Jarak Minkowski

Jarak *Minkowski* merupakan salah satu perhitungan jarak pada metode K-Nearest Neighbor. Prinsip perhitungan *minkowski* adalah mengukur jarak  $v_1$  dan  $v_2$  yang merupakan dua vektor yang jaraknya dihitung dan N yaitu menyatakan panjang dari vektor. Sebagai contoh, dengan dua vektor yang sama dengan di depan ( $v_1 = [4, 3, 6]$  dan  $v_2 = [2, 3, 7]$ ), jarak *Minkowski* kedua vektor tersebut untuk p berupa:

$$j_{v_1, v_2} = \sqrt[p]{\sum_{k=1}^n |v_1(k) - v_2(k)|^p}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Citra

Data citra yang digunakan berupa citra yang berwarna dengan ukuran

$$R = 1 - \frac{1}{1 + \sigma^2} \quad (6)$$

625x417. Citra tersebut berjumlah 100 citra buah apel yang di dapat dari toko – toko buah dengan rinciannya 50 Jenis buah apel envy dan 50 jenis buah apel pasific rose. Dari jumlah 100 citra 90 data citra apel digunakan untuk data *training* (data latih) serta 10 data cita apel digunakan untuk data *testing* (data uji). Langkah

selanjutnya yaitu pengelompokan dari 100 dataset yang terbagi menjadi 2, yaitu 10 data test dan 90 data training. Dari data 100 citra tersebut, 10 data citra digunakan sebagai data test sedangkan 90 data citra digunakan sebagai data training.

Data citra apel yang digunakan yaitu data *train* adalah sebagai berikut:

Tabel 3.1 data training

No	Citra apel	Kelas
1	1.jpg	Envy
2	2.jpg	Envy
3	48.jpg	Pasific rose
4	49.jpg	Pasific rose

Sedangkan data *testing* (data uji) adalah sebagai berikut:

Tabel 3.2 data testing

No	Citra Apel	Kelas
1	1.jpg	Envy
2	2.jpg	Envy
3	7.jpg	Pasific rose
4	8.jpg	Pasific rose

### 3.2 Ekstraksi Fitur Histogram

Pada ekstraksi fitur menggunakan metode *histogram* ini terdapat 6 fitur yang diantaranya adalah: rerata intensitas, deviasi, *skewness*, energi, entropi dan kehalusan (*smoothness*). Pada tabel dibawah ini adalah 100

data citra yang sudah di ekstraksi fitur *histogram* pada citra apel dari 2 jenis citra apel. Citra tersebut kemudian diproses menggunakan ekstraksi fitur histogram, dari hasil pengolahan tersebut.

Tabel 3.3 ekstraksi fitur histogram

No	Kode jenis apel	m(intensitas)	deviasi	skewness	energi	smoothness
1	1.jpg	223.6202999	57.43430595	-7.442057986	0.042946167	0.048280458
2	2.jpg	228.8955311	52.68079407	-6.177258418	0.055131106	0.040932967
3	3.jpg	229.1726312	51.21496373	-5.772865184	0.053675179	0.038773853
4	4.jpg	230.0007827	48.67951739	-4.929525448	0.052691333	0.035161453
5	5.jpg	229.4541173	48.70738011	-4.87348868	0.049657338	0.035200298
6	51.jpg	230.608958	46.56886525	-4.604856144	0.072294022	0.03227476
7	52.jpg	233.9313471	40.82728272	-3.265443166	0.074910945	0.024993558
8	53.jpg	231.1548905	41.26323216	-3.178948152	0.056493498	0.025516472
9	54.jpg	229.1815341	51.98512328	-6.036646332	0.072901133	0.039901885
10	55.jpg	226.6954756	56.57729005	-7.277637139	0.069910651	0.046917454

Dari data diatas merupakan beberapa data citra apel yang sudah di ekstraksi fitur menggunakan ekstraksi fitur histogram yang menghasilkan angka – angka piksel tersebut yang nantinya hasil dari angka – angka piksel tersebut akan di klasifikasi ke dalam algoritma K-Nearest Neighbor.

### 3.3 Algoritma K – Nearest Neighbor

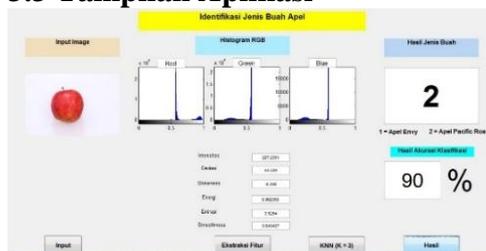
Dari hasil proses klasifikasi menggunakan algoritma KNN dengan data uji dan data latih serta menggunakan K=3 adalah sebagai berikut:

Tabel 3.4 klasifikasi percobaan pertama menggunakan metode K-NN

No	Nama	Hasil klasifikasi
1	Data uji1	Apel envy
2	Data uji2	Apel envy
3	Data uji3	Apel envy
4	<b>Data uji4</b>	<b>Apel pasific Rose</b>
5	Data uji5	Apel envy
6	Data uji6	Apel pasific rose
7	Data uji7	Apel pasific rose
8	Data uji8	Apel pasific rose
9	Data uji9	Apel pasific rose
10	Data uji10	Apel pasific rose

Pada tabel diatas, kata yang di *marking* merah merupakan *error* pada tahap klasifikasi. *Error* disini yang dimaksudkan adalah pada pengujian satu dan dua terdapat masing – masing *error* satu. Ini terjadi karena data tersebut memiliki kemiripan pada kelasnya meskipun secara fakta jenis apel tersebut bukan dari kelasnya.

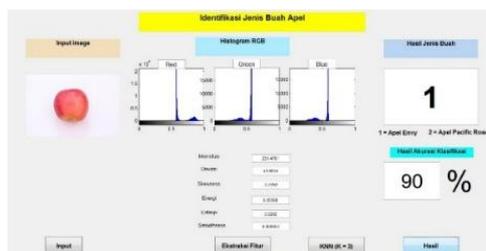
### 3.3 Tampilan Aplikasi



**Gambar 4. 1** menampilkan hasil output

Pada gambar diatas citra kelas buah apel 1 yaitu buah apel envy setelah di proses melalui ekstraksi fitur *histogram* dan menggunakan algoritma *K – Nearest Neighbor* hasilnya adalah benar citra tersebut masuk ke dalam kelas 1 yaitu kelas buah apel envy.

**Gambar 4. 2** menampilkan hasil output



Sama seperti pada gambar sebelumnya bedanya input citra buah apel dari kelas 2 yaitu apel *Pacific Rose*. Dan hasil output berupa angka 2 yang mana angka 2 yaitu dari kelas 2 yang bernama apel *Pacific Rose*.

## 4. KESIMPULAN

Berdasarkan hasil dari penelitian tentang Identifikasi Jenis Buah Apel Menggunakan Algoritma *K – Nearest Neighbor* (KNN) dengan Ekstraksi Fitur Histogram, dapat disimpulkan beberapa hal berikut yaitu:

1. Pada identifikasi pada citra apel menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan 100 data citra apel dengan terdiri dari 50 jenis apel envy dan 50 jenis apel pasific rose. Data uji sebanyak 10 citra yang terdiri dari 5 jenis citra apel envy dan 5 jenis citra apel pasific rose memperoleh akurasi sebesar 90%.
2. Sehingga hasil diatas dapat disimpulkan bahwa dengan menggunakan algoritma *K-Nearest Neighbor* serta menggunakan ekstraksi fitur *histogram* dapat diterapkan dengan baik pada identifikasi jenis apel.

## 5. SARAN

Beberapa saran untuk penelitian tentang citra apel dalam penelitian selanjutnya yang dapat memberikan manfaat dalam pelaksanaannya diantaranya sebagai berikut:

1. Penelitian selanjutnya dapat diharapkan dapat identifikasi jenis buah apel yang lebih banyak dari penelitian sebelumnya.
2. Pengujian dengan data citra latih lebih banyak agar dapat meningkatkan akurasi dalam proses identifikasi data.
3. Citra pun bisa dideteksi pada cahaya, kemiringan, jarak pengambilan data yang berbeda.

4. Untuk mencoba dengan menggunakan metode lain yang lebih optimal.

#### **DAFTAR PUSTAKA**

- [1] S. Jatmika dan D. Purnamasari, "Rancang Bangun Alat Pendeteksi Kematangan Buah Apel dengan Menggunakan Metode Image Processing Berdasarkan Komposisi Warna," vol. 8, 2014.
- [2] R. T. Sari, "Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation dan Alogirtma Pencocokan dalam Mengidentifikasi Kematangan Tomat Buah Berdasarkan Ciri Warna RGB," 2013.
- [3] M. I. Sikki, "Pengenalan Wajah Menggunakan K – Nearest Neighbor dengan Praproses Transformasi Wavelet," Paradigma, vol. 10, p. 2, 2009.
- [4] S. Q. Nugroho dan R. A. Pramuendar, "Pengelompokan Kayu Kelapa Menggunakan Algoritma K – Means Berdasarkan Tekstur Citra Kayu Kelapa Dua Dimensi (2D)," Dokumen Karya Ilmiah, 2015