

KLASIFIKASI KEMATANGAN SEMANGKA BERDASARKAN EKSTRAKSI CIRI STATISTIK ORDER PERTAMA DENGAN EKUALISASI HISTOGRAM

Classification of Watermelon Maturity Based on the Extraction Characteristics of First Order Statistics with Equalization of Histograms

Febri Liantoni¹, Hendro Nugroho²

¹Pendidikan Teknik Informatika dan Komputer, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sebelas Maret,
Jl. Ir. Sutami 36 Kentingan, Jebres, Surakarta - Indonesia

² Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya - Indonesia
Telp. (0271) 646994, Fax. (0271) 636268
E-mail: febri.liantoni@gmail.com, dosh3ndro@gmail.com

(Makalah diterima, 30 Agustus 2018 – Disetujui, 03 Juni 2019)

ABSTRAK

Semangka (*Citrullus Lanatus* Tunb) merupakan tanaman buah yang banyak disukai di Indonesia. Kematangan semangka bisa dibedakan berdasarkan tekstur kulit semangka. Adanya kemiripan tekstur kulit semangka mengakibatkan orang kesulitan dalam mengidentifikasi tingkat kematangan semangka yang matang dan belum matang. Berdasarkan permasalahan tersebut, maka penelitian ini dilakukan untuk proses klasifikasi berdasarkan tekstur kulit buah semangka.

Ekstraksi ciri statistik orde pertama digunakan sebagai metode untuk mengenali kematangan semangka dari segi tekstur kulit buah dalam proses klasifikasi. Parameter yang digunakan yaitu *mean*, *varians*, *skewness* dan *kurtosis*. Berdasarkan nilai ekstraksi ciri statistik tersebut, kemudian dijadikan acuan proses klasifikasi menggunakan metode *k-nearest neighbor* dan *naïve bayes*. Data dalam penelitian ini sejumlah 100 citra, dimana 70 untuk data training dan 30 untuk data testing. Dari hasil pengujian klasifikasi kematangan semangka menggunakan *K-Nearest Neighbor* diperoleh 26 data sesuai dengan tingkat akurasi 86.66%, sedangkan menggunakan *Naïve Bayes* diperoleh 27 data sesuai dengan tingkat akurasi 90%. Hal ini membuktikan bahwa kematangan buah semangka dapat diketahui berdasarkan warna kulit buah semangka.

Kata kunci: semangka, ciri statistik orde pertama, *K-Nearest Neighbor*, *Naïve Bayes*

ABSTRACT

Watermelon (*Citrullus Lanatus* Tunb) is a fruit plant that is widely preferred in Indonesia. The maturity of watermelon can be distinguished based on watermelon skin texture. The similarity in watermelon skin texture causes people difficulty in identifying the level of maturity of mature and immature watermelon. Based on these problems, this research was conducted for the classification process based on watermelon skin texture.

First-order statistical feature extraction is used as a method to recognize watermelon maturity in terms of fruit skin texture in the classification process. The parameters used are mean, variance, skewness, and kurtosis. Based on the statistical feature extraction value, then used as a reference classification process using the method of *K-Nearest Neighbor* and *Naïve Bayes Classifier*. The data in this study are 100 images, of which 70 are for training data and 30 for testing data. From the results of testing the classification of watermelon maturity using *K-Nearest Neighbor* obtained 26 data in accordance with 86.66% accuracy rate, while using *Naïve Bayes* obtained 27 data in accordance with 90% accuracy rate. These results prove the maturity of the watermelon can be determined based on the color of the skin of the watermelon.

Key words: watermelon, the first-order statistical feature, *K-Nearest Neighbor*, *Naïve Bayes*

PENDAHULUAN

Semangka mempunyai manfaat yang menjadikan semangka menjadi buah yang dicari oleh masyarakat. Adanya kemiripan tekstur kulit semangka yang matang dengan yang belum matang mengakibatkan orang kesulitan dalam mengidentifikasi buah semangka. Karena manusia bersifat subyektif dalam menentukan kematangan buah semangka, menyebabkan penilaian kematangan buah semangka berbeda dari satu orang dengan orang yang lainnya. Proses klasifikasi kematangan buah dapat dilakukan dengan cara mengidentifikasi citra dari tekstur kulit buah tersebut (Rocha *et al.*, 2010). Dengan cara tersebut maka dapat dilakukan langkah-langkah pengenalan pola tekstur kulit dengan mengenali karakteristik struktural kulit seperti bentuk dan warna semangka menggunakan ekualisasi histogram. Perkembangan teknologi untuk teknik pengolahan citra juga berkembang pesat. Berbagai teknik untuk mempermudah pekerjaan manusia, seperti pengolahan citra digital, analisis citra maupun penggunaan citra untuk berbagai tujuan (Rocha *et al.*, 2010).

Teknik pengolahan citra digital ini dilakukan pada tahapan praproses citra sampai didapatkan bentuk dan ciri struktural dari masing-masing kulit semangka. Ekualisasi histogram digunakan pada tahap praproses sebelum dilakukan ekstraksi fitur. Metode yang digunakan pada ekstraksi fitur ini berdasarkan nilai ciri statistik orde pertama. Setelah itu dilakukan ekstraksi fitur dari citra kulit semangka kemudian digunakan sebagai data klasifikasi kematangan semangka menggunakan *k-nearest neighbor* dan *naïve bayes classifier*.

Berdasarkan penelitian sebelumnya, Novita ditiadakan 2017 menggunakan ekstraksi ciri orde pertama yang diimplementasikan pada klasifikasi batik dengan algoritma backpropagasi. Pada penelitian tersebut digunakan nilai parameter *mean*, *skewness*, *kurtosis* dan *entropy* sebagai dasar pengolahan proses klasifikasi (Ningrum *et al.*, 2017). Guan ditiadakan 2017 melakukan eksplorasi nilai difusi pada *apparent diffusion coefficient* (ADC) ada ciri statistik orde pertama dalam pengenalan klinis kanker serviks (Guan *et al.*, 2017).

Ekualisasi Histogram merupakan metode untuk memperbaiki kualitas citra dengan cara mengubah sebaran tingkat keabuan citra (Kong dan Ibrahim, 2011). Ekualisasi histogram juga digunakan untuk meningkatkan kontras warna citra. Beberapa penelitian tentang peningkatan kontras menggunakan ekualisasi histogram dilakukan oleh Celik, Chen, dan Wang di penelitian sebelumnya. Celik melakukan perbaikan citra dengan ekualisasi histogram dua dimensi yang efektif meningkatkan kualitas dari tipe citra yang berbeda (Celik,

2012). Chen mengembangkan model ekualisasi histogram untuk peningkatan kontras gambar. Pendekatannya yang dilakukan dengan model variasional yang mengandung energi fungsional untuk menyesuaikan nilai-nilai pixel dari suatu gambar masukan secara langsung sehingga histogram yang dihasilkan dapat didistribusikan kembali menjadi lebih seragam (Chen dan Ng, 2016). Wang juga melakukan teknik peningkatan kontrak berdasarkan algoritma ekualisasi histogram lokal. Usulan dilakukan dengan melakukan segmentasi citra menjadi beberapa sub-blok menggunakan nilai gradien, algoritma ini berhasil meningkatkan kontras lokal tanpa menambahkan derau pada citra (Wang dan Pan, 2017).

Penelitian tentang klasifikasi semangka berbiji dan tanpa biji menggunakan *k-nearest neighbor* dilakukan oleh (Mohd Ali *et al.*, 2017). Pada penelitian sebelumnya yang telah dilakukan klasifikasi tanaman mangga menggunakan metode *k-nearest neighbor* dan *support vector machine* berdasarkan struktur tulang daun (Liantoni dan Hermanto, 2017). Pada tahun 2015, peneliti melakukan klasifikasi daun herbal menggunakan metode *k-nearest neighbor* dan *naïve bayes classifier* (Liantoni dan Nugroho, 2015).

Pada penelitian ini dilakukan klasifikasi kematangan buah semangka berdasarkan ekstraksi ciri statistik orde pertama dari tekstur kulit buah menggunakan *k-nearest neighbor* dan *naïve bayes classifier*. Penelitian ini diharapkan mampu membantu seseorang untuk mengetahui kematangan semangka melalui hasil foto dari buah semangka.

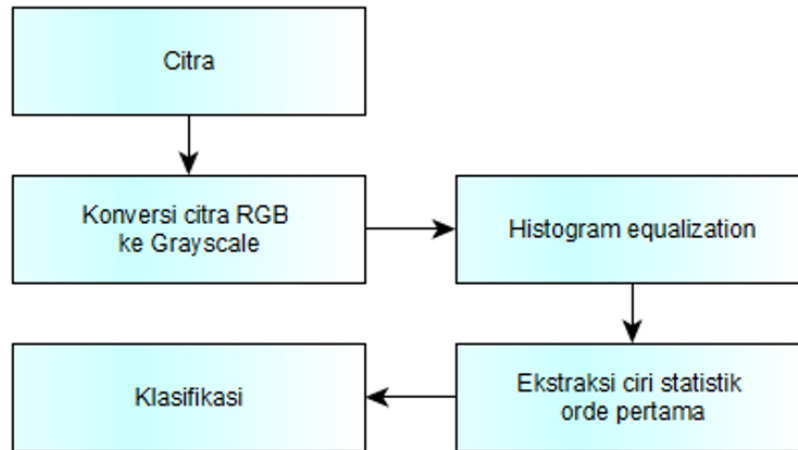
MATERI DAN METODE

Salah satu cara untuk mengetahui kematangan buah dapat dilakukan dengan melihat struktur kulit dari buah tersebut. Biasanya warna kulit buah akan berubah jika buah sudah mulai matang, sehingga dapat dijadikan acuan untuk identifikasi kematangan buah tersebut.

Secara umum, pengkategorisasian citra bergantung pada kombinasi pendekatan ciri statistik, struktural dan spektral. Pendekatan ciri statistik deskripsikan objek menggunakan deskriptor global dan lokal seperti mean, varians, dan entropi. Pendekatan struktural dapat mewakili objek tersebut. Pendekatan spektral dapat menggambarkan objek menggunakan beberapa representasi ruang spektral seperti Fourier spektrum (Gonzales dan Wood, 1992).

Pada penelitian ini dilakukan beberapa tahapan meliputi praproses, ekstraksi ciri orde pertama, klasifikasi dengan *k-nearest neighbor* dan *naïve bayes classifier*.

Tahapan penelitian yang dilakukan seperti ditunjukkan pada Gambar 1.



Gambar 1. Tahapan penelitian

Praproses

Pada penelitian ini foto buah semangka yang tampak berupa tekstur kulit digunakan sebagai data penelitian. Proses awal yang dilakukan dengan mengubah citra RGB menjadi citra keabuan. Hasil ini bertujuan untuk menyederhanakan citra sehingga lebih mudah dalam pengolahan pada tahap berikutnya.

Ekstraksi ciri statistik orde pertama

Ekstraksi ciri ini merupakan metode pengambilan ciri yang didasarkan pada karakteristik histogram citra dengan mengabaikan hubungan antar piksel tetangga. Ekstraksi ciri statistik orde pertama dengan menghitung fitur warna dari citra tekstur kulit semangka. Fitur warna dapat diperoleh melalui perhitungan statistis seperti *mean*, *varians*, *skewness*, dan *kurtosis*.

Mean merupakan ukuran dispersi citra mengenai distribusi warna dari citra. Penghitungan nilai mean ditunjukkan pada Persamaan 1.

$$\mu = \frac{1}{MN} \sum_{i,j=1}^{M,N} P_{ij} \quad (1)$$

Varians menyatakan luas sebaran distribusi. Akar kuadrat varians dinamakan sebagai standar deviasi. Varians skan variasi elemen pada suatu citra, ditunjukkan pada Persamaan 2.

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i,j=1}^{M,N} (P_{ij} - \mu)^2} \quad (2)$$

Skewness atau kecondongan menyatakan ukuran mengenai ketidaksimetrisan. Distribusi dikatakan condong ke kiri apabila memiliki nilai *skewness* berupa

bilangan negatif. Sebaliknya, distribusi dikatakan condong ke kanan apabila memiliki nilai *skewness* berupa bilangan positif. Penghitungan *skewness* ditunjukkan Persamaan 3.

$$\theta = \frac{\sum_{i,j=1}^{M,N} (P_{ij} - \mu)^3}{MN\sigma^3} \quad (3)$$

Kurtosis merupakan ukuran yang menunjukkan sebaran data bersifat meruncing atau menumpul. Penghitungan *kurtosis* seperti pada Persamaan 4.

$$\gamma = \frac{\sum_{i,j=1}^{M,N} (P_{ij} - \mu)^4}{MN\sigma^4} - 3 \quad (4)$$

K-Nearest Neighbor

Metode *K-Nearest Neighbour* (K-NN) menjadi salah satu metode berbasis *nearest neighbor* (NN) yang bisa digunakan untuk membandingkan nilai dari histogram (Mohamad *et al.*, 2011). Nilai K yang digunakan disini menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas pada data uji. Prinsip kerja metode *K-Nearest Neighbor* (KNN) dengan menghitung nilai jarak terdekat antara data yang akan dievaluasi dengan K ketetanggaan terdekatnya pada data yang diolah dalam pelatihan.

Naïve Bayes Classifier

Pada *machine learning*, *naïve bayes classifier* termasuk bagian dari *probabilistic classifiers* yang didasarkan pada penerapan teorema bayes dengan menggunakan asumsi independensi yang kuat. *Naïve Bayes* telah dipelajari secara ekstensif sejak tahun 1950-an. Ini diperkenalkan di bawah nama yang berbeda ke dalam komunitas text retrieval pada awal 1960-an (Russell dan Norvig,

2009). Metode *naïve bayes* merupakan metode yang populer untuk pengkategorian teks, dokumen, deteksi spam berdasarkan frekuensi kata sebagai fitur.

HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan sejumlah 100 citra, dimana 70 untuk data training dan 30 untuk data testing. Pengujian ini bertujuan untuk mengetahui kematangan semangka berdasarkan warna kulit dari buah semangka. Contoh citra kulit semangka ditunjukkan pada Gambar 1.

Tahap awal pada penelitian ini yaitu mengubah RGB citra menjadi *grayscale*. Berdasarkan citra *grayscale* tersebut selanjutnya dilakukan proses ekualisasi histogram. Hasil citra ekualisasi histogram seperti ditunjukkan pada Gambar 2.

Ekstraksi ciri statistik orde pertama meliputi nilai mean, varians, skewness, dan kurtosis. Beberapa contoh

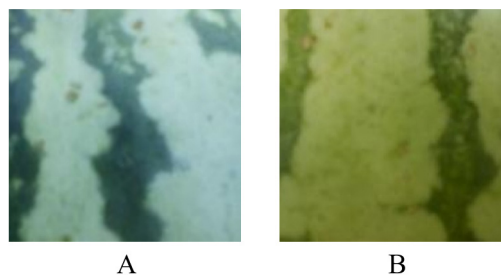
hasil penghitungan ciri statistik ditunjukkan pada Tabel 1.

Proses klasifikasi kemudian dilakukan berdasarkan nilai ciri statistik tersebut. Dari hasil proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* menggunakan $K=3$ dan *Naïve Bayes Classifier* seperti ditunjukkan pada Tabel 2.

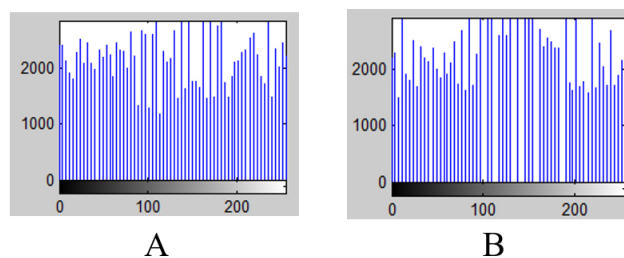
Pada pengujian dengan *k-nearest neighbor* diperoleh 26 data sesuai dengan tingkat akurasi $26/30 = 86.66\%$ sedangkan pengujian dengan *naïve bayes classifier* diperoleh 27 data sesuai dengan tingkat akurasi $27/30 = 90\%$. Hal ini membuktikan bahwa kematangan semangka dapat dilihat berdasarkan warna kulit buah semangka.

KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan untuk proses klasifikasi kematangan semangka berdasarkan ciri statistik orde pertama menggunakan *K-Nearest*



Gambar 2. Contoh citra kulit semangka a) semangka mentah, b) semangka matang



Gambar 3. Hasil ekualisasi histogram a) semangka mentah, b) semangka matang

Tabel 1. Contoh ekstraksi ciri statistik orde pertama

Data	Mean	Variance	Skewness	Kurtosis
Mentah	4700	252019112	-0.56743	-2.92653
Mentah	3284	177451453	-0.49763	-2.90486
Mentah	2355	216254661	-0.15793	-2.91873
Mentah	1353	134535344	-0.23446	-2.96454
Matang	1307	158984	-0.70742	-2.99325
Matang	1174	138453	-0.87463	-2.99856
Matang	1261	59763	-0.80735	-2.99893
Matang	1112	67488	-0.84533	-2.99877

Tabel 2. Hasil klasifikasi

Data	Class			Result	
	Ground truth	KNN	Bayes	KNN	Bayes
1	Matang	Matang	Matang	TRUE	TRUE
2	Matang	Mentah	Matang	FALSE	TRUE
3	Matang	Matang	Matang	TRUE	TRUE
4	Matang	Matang	Matang	TRUE	TRUE
5	Matang	Matang	Matang	TRUE	TRUE
6	Matang	Matang	Matang	TRUE	TRUE
7	Matang	Matang	Matang	TRUE	TRUE
8	Matang	Matang	Matang	TRUE	TRUE
9	Matang	Mentah	Mentah	FALSE	FALSE
10	Matang	Matang	Matang	TRUE	TRUE
11	Matang	Matang	Matang	TRUE	TRUE
12	Matang	Matang	Matang	TRUE	TRUE
13	Matang	Matang	Matang	TRUE	TRUE
14	Matang	Matang	Matang	TRUE	TRUE
15	Matang	Mentah	Matang	FALSE	TRUE
16	Mentah	Mentah	Mentah	TRUE	TRUE
17	Mentah	Mentah	Mentah	TRUE	TRUE
18	Mentah	Mentah	Mentah	TRUE	TRUE
19	Mentah	Mentah	Mentah	TRUE	TRUE
20	Mentah	Mentah	Mentah	TRUE	TRUE
21	Mentah	Matang	Matang	FALSE	FALSE
22	Mentah	Mentah	Mentah	TRUE	TRUE
23	Mentah	Mentah	Mentah	TRUE	TRUE
24	Mentah	Mentah	Mentah	TRUE	TRUE
25	Mentah	Mentah	Mentah	TRUE	TRUE
26	Mentah	Mentah	Mentah	TRUE	TRUE
27	Mentah	Mentah	Mentah	TRUE	TRUE
28	Mentah	Mentah	Matang	TRUE	FALSE
29	Mentah	Mentah	Mentah	TRUE	TRUE
30	Mentah	Mentah	Mentah	TRUE	TRUE

Neighbor didapatkan akurasi sistem sebesar 86.66%, sedangkan dengan menggunakan *Naïve Bayes classifier* didapatkan akurasi sistem sebesar 90%. Hasil pengujian ini menunjukkan jika warna kulit semangka dapat dijadikan sebagai acuan dalam menentukan kematangan buah semangka.

DAFTAR PUSTAKA

Celik, T. 2012. Two-dimensional histogram equalization and contrast enhancement. *Pattern Recognition* 45 (10) : 3810–3824.

Chen, C., and Ng, M. K. 2016. An image pixel based variational model for histogram equalization. *Journal of Visual Communication and Image Representation* 34 : 118–134.

Gonzales, R., and Wood, R. 2007. *Digital Image Processing*. 3rd edition. Prentice-Hall, Inc., United State, America.

Guan, Y., Li, W., Jiang, Z., Zhang, B., Chen, Y., Huang, X., Zhang, J. 2017. Value of whole-lesion apparent diffusion coefficient (ADC) first-order statistics and texture features in clinical staging of cervical cancers. *Clinical Radiology* 72 (11) : 951–958.

- Kong, N. S. P. and Ibrahim, H. 2011. Multiple layers block overlapped histogram equalization for local content emphasis. *Computers and Electrical Engineering, Special Issue on Image Processing* 37 (5) : 631–643.
- Liantoni, F., and Nugroho, H. 2015. Klasifikasi daun herbal menggunakan metode naïve bayes classifier dan k-nearest neighbor. *Jurnal Simantec, Universitas Trunojoyo Madura* 5 (1) : 9–16.
- Liantoni, F. and Agus Hermanto, L. 2017. Adaptive ant colony optimization on mango classification using k-nearest neighbor and support vector machine. *Journal of Information Systems Engineering and Business Intelligence* 3: 75.
- Mohamad, F. S., Manaf, A. A., and Chuprat, S. 2011. Nearest Neighbor For Histogram-based Feature Extraction. *Procedia Computer Science, Proceedings of the International Conference on Computational Science, ICCS* 4 : 1296–1305.
- Mohd Ali, M., Hashim, N., Bejo, S. K., and Shamsudin, R. 2017. Laser-induced backscattering imaging for classification of seeded and seedless watermelons. *Computers and Electronics in Agriculture* 140 : 311–316.
- Ningrum, N., Defri, K., and Novi, H. 2017. Penerapan ekstraksi ciri orde satu untuk klasifikasi tekstur motif batik pesisir dengan algoritma backpropagasi. *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer* 8 (2):
- Rocha, A., Hauagge, D. C., Wainer, J., and Goldenstein, S. 2010. Automatic fruit and vegetable classification from images. *Computers and Electronics in Agriculture* 70 (1) : 96–104.
- Russell, S., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach* (3 edition.). Upper Saddle River: Pearson.
- Wang, Y., and Pan, Z. 2017. Image contrast enhancement using adjacent-blocks-based modification for local histogram equalization. *Infrared Physics and Technology* 86 : 59–65.