

SIMULASI DAN ANALISIS SISTEM KLASIFIKASI BATUBARA MENGGUNAKAN FUZZY COLOR HISTOGRAM, DISCRETE COSINE TRANSFORM, DAN K-NEAREST NEIGHBOR PADA CITRA DIGITAL

Simulation and Analysis of Image Processing in Coal Classification System Using Fuzzy Color Histogram, Discrete Cosine Transform, and K-Nearest Neighbor

Galuh Laksmi Ranggi¹, Dr.Ir.Bambang Hidayat, DEA², Suci Aulia S.T.,MT.³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom
Jln. Telekomunikasi No.1 Terusan Buah Batu Bandung 40257 Indonesia

¹galuhlr@students.telkomuniversity.com, ²bhidayat@telkomuniversity.ac.id, ³suciaulia@telkomuniversity.ac.id

ABSTRAK

Tugas Akhir ini membahas tentang simulasi dan analisis sistem klasifikasi batubara dengan analisis ciri warna pada citra batubara. Penelitian Tugas Akhir ini dapat membantu masyarakat menjadi lebih jeli dalam memilih jenis batubara yang akan dibeli. Sehingga, angka kejahatan terkait kualitas batubara dapat menurun. Penelitian ini menggunakan 3 jenis batubara untuk dianalisis, yaitu: batubara dengan nilai kalori tinggi, sedang, dan rendah. Secara kasat mata, ketiga jenis batubara tersebut dapat dibedakan berdasarkan warnanya. Adapun tahapan yang dilakukan dalam penelitian ini ialah: akuisisi citra, *preprocessing*, ekstraksi ciri, dan klasifikasi. Metode utama yang digunakan untuk keperluan analisis adalah *Discrete Cosine Transform* (DCT), *Fuzzy Color Histogram* (FCH), dan *K-Nearest Neighbor* (K-NN). Penelitian ini membandingkan sistem ekstraksi ciri DCT dengan sistem ekstraksi ciri FCH. Parameter yang diukur adalah waktu komputasi dan tingkat akurasi. Untuk mengetahui nilai dari masing-masing parameter tersebut, digunakan 10 data latih dan 20 data uji untuk tiap kelasnya. Sehingga, total data yang digunakan dalam penelitian ini ada 90 buah data. Akurasi terbaik pada metode DCT terjadi ketika ukuran bloknya sebesar 8x8, yaitu mencapai 78,33%. Sedangkan, metode FCH hanya memperoleh akurasi sebesar 61,67%. Apabila dilihat dari waktu komputasi, model FCH membutuhkan waktu lebih sedikit dari DCT.

Kata kunci : klasifikasi batubara, DCT, FCH, K-NN

ABSTRACT

This final project discusses about the simulation and analysis of image processing in coal classification using color features. This research is expected to help people to be more intelligently chooses the type of coal that will be purchased. So, the amount of criminal in coal cases can be decrease. It used 3 types of coal which are analyzed below: coal with a high calorie value, medium, and low. Visually, these three types of coal can be distinguished based on its color. As for the step that is carried out in this study are: image acquisition, preprocessing, features extraction, and classification. The main method that is used for the purposes of the analysis is the Discrete Cosine Transform (DCT), Fuzzy Color Histogram (FCH), and K-Nearest Neighbor (K-NN). This research compares between DCT and FCH as a feature extraction method. Parameters measured is the accuracy level and computation time. This research uses 90 data. It contains 10 trial data and 20 test data for each class. The best accuracy in DCT method when the block size is 8x8, was 78,33%. Meanwhile, FCH method got 61,67%. Based on computation time, both method need less than 20 secons to get the output. Overall, FCH need less computational time than DCT.

Keywords: Coal Classification, DCT, FCH, K-NN

1. Pendahuluan

1.1 Latar Belakang Masalah

Dalam penelitian sebelumnya, penentuan kualitas batubara dilakukan dengan melihat unsur-unsur kimia pembentuk batubara. Sedangkan, penelitian di lain tempat menggunakan nilai HGI dengan standar *American Society for Testing and Material (ASTM)* sebagai faktor penentu kualitas. HGI adalah suatu tolok ukur secara laboratorium dari mudah atau sulitnya batubara digerus atau di *pulverizing*. Kedua penelitian tersebut memakan waktu lama dalam prosesnya. Oleh karena itu, dibutuhkan suatu sistem yang dapat mengklasifikasikan batubara dengan waktu yang relatif singkat dan akurasi yang tinggi.

Sistem ini dapat mengklasifikasikan kualitas batubara cukup dari citra batubara. Penelitian ini menitik beratkan menggunakan metode *Discrete Cosine Transform* (DCT) dan *Fuzzy Color Histogram* (FCH) sebagai ekstraksi ciri yang digunakan. Sedangkan, jenis classifier yang digunakan adalah *K-Nearest Neighbor* (K-NN). Performansi ssstem dilihat berdasarkan tingkat akurasi dan waktu komputasi ssstem.

Penelitian yang dilakukan hanya berdasarkan ciri warna pada citra batubara. Batubara yang diteliti berasal dari daerah Samarinda, Kalimantan Timur. Jenis data uji dan data latih yang digunakan adalah batubara kelas rendah (4400 kalori), batubara kelas sedang (5100 kalori), dan batubara kelas tinggi (7000 kalori). Total jumlah data uji dan data latih yang digunakan sebanyak 90 data. Jenis kamera yang digunakan adalah kamera DSLR Canon EOS 600D resolusi 18 MP. Jarak kamera ke objek saat pengambilan citra ± 20 cm dengan penerangan lampu *indoor* dan *angle* pengambilan citra, tepat diatas objek (90°). Input dari sistem adalah citra batubara format .jpg yang di-*cropping* secara manual dengan ukuran 512×512 pixel.

2. Dasar Teori

2.1 Batubara

Batubara terbentuk dari sisa tumbuhan yang ada zaman prasejarah. Selanjutnya, sisa tumbuhan tersebut mengalami proses fisika dan kimia sehingga menjadi rawa dan lahan gambut. Kemudian, rawa dan lahan gambut tersebut lambat laun akan berubah menjadi batubara [1].

Jika dilihat dari kalori yang terkandung di dalamnya, klasifikasi batubara dapat dilihat pada tabel berikut.

Tabel 1 Klasifikasi batubara berdasarkan jumlah kalori (Sumber: [5])

No	Jenis batubara	Jumlah kalori yang terkandung (Kal/gr, adb)
1	Batubara kalori rendah	<5100
2	Batubara kalori sedang	5100-6100
3	Batubara kalori tinggi	6100-7100
4	Batubara kalori sangat tinggi	>7100

Selain diklasifikasikan berdasarkan proses terjadinya dan kalori yang terkandung di dalamnya, batubara juga biasanya diklasifikasikan berdasarkan warnanya. Warna dari batubara yang memiliki nilai karbon yang tinggi akan terlihat lebih gelap mengilap jika dibandingkan dengan batubara yang memiliki nilai karbon rendah [2].

2.2 Citra Digital

Menurut Syarif *et al* (2012), citra digital adalah citra atau gambar yang disimpan dalam format digital (dalam bentuk file). Sedangkan, Putra (2010) mendefinisikan citra digital sebagai fungsi berukuran M baris dan N kolom, dengan x dan y adalah koordinat spasial [9] dan [11].

$$f(x, y) = \begin{matrix} f(1,1) & f(1,2) & \dots & f(1,M) \\ f(2,1) & f(2,2) & \dots & f(2,M) \\ \vdots & \vdots & \dots & \vdots \\ f(M,1) & f(M,2) & \dots & f(M,M) \end{matrix} \quad (1)$$

2.3 Ekstraksi Ciri

2.3.1 Discrete Cosine Transform

Discrete Cosine Transform atau DCT sangat berhubungan dengan DFT (*Discrete Fourier Transform*), yang membedakannya hanyalah komponen yang terkandung didalamnya. DFT adalah jenis transformasi yang terdiri dari komponen riil dan imajiner, yaitu kosinus dan sinus. Sedangkan, DCT hanya memiliki komponen kosinus untuk merubah suatu citra dari domain spasial ke domain frekuensi [11] [4].

Definisi dari DCT dua dimensi untuk citra input A dan output gambar B adalah sebagai berikut [10].

$$B_{(u,v)} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} A_{(x,y)} \cos \frac{\pi(2x+1)u}{2M} \cos \frac{\pi(2y+1)v}{2N} \quad 0 \leq u \leq M-1 \quad 0 \leq v \leq N-1 \quad (2)$$

dimana,

$$A_{(x,y)} = \begin{matrix} \frac{1}{\sqrt{M}} & & & & 0 \\ & \frac{1}{\sqrt{2}} & & & \\ & & \frac{1}{\sqrt{2}} & & \\ & & & \ddots & \\ & & & & \frac{1}{\sqrt{2}} \\ & & & & & \frac{1}{\sqrt{M}} \end{matrix}$$



dan

$$C_{u,v} = \begin{cases} \frac{1}{\sqrt{2}} & u=0 \\ \frac{1}{2} & 1 \leq u \leq M-1 \end{cases}$$

Nilai dari $C_{u,v}$ disebut koefisien DCT dari input citra A. Fungsi ini disebut fungsi basis dari DCT. Koefisien DCT $C_{u,v}$ dapat dianggap bobot untuk setiap basis fungsi [4].

Frekuensi horizontal meningkat dari kiri ke kanan, frekuensi vertikal meningkat dari atas ke bawah. Fungsi basis nilai konstan pada kiri atas dan disebut fungsi basis DC dan berhubungan dengan koefisien DCT $C_{0,0}$ dan disebut koefisien DC [4].

2.3.2 Fuzzy Color Histogram

Fuzzy Color Histogram (FCH) merupakan salah satu metode untuk merepresentasikan informasi warna dalam citra digital ke dalam bentuk histogram. Inti dari FCH adalah tiap warna direpresentasikan dengan himpunan fuzzy (*fuzzy set*) dan hubungan antar warna dimodelkan dengan fungsi keanggotaan dari himpunan fuzzy [14].

2.4 Klasifikasi K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (*K-NN*) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data uji yang jaraknya paling dekat dengan objek tersebut. Pada data latih biasanya diambil lebih dari satu tetangga terdekat dengan data uji kemudian algoritma ini digunakan untuk menentukan kelasnya. Pada *K-NN* terdapat beberapa aturan jarak yang dapat digunakan, yaitu: *euclidean*, *cityblock*, *cosine*, dan *correlation* [12].

3. Pembahasan

3.1 Perancangan Model Sistem

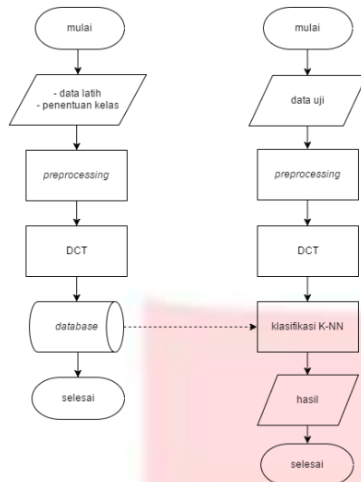
Dalam kasus ini, pengolahan citra dilakukan untuk mengelola informasi pada citra batubara agar dapat mendapat ciri yang ada pada citra tersebut. Setelah mendapatkan ciri dari objek penelitian, citra tersebut klasifikasi berdasarkan kelas yang telah ditentukan.



Gambar 1 Diagram blok sistem

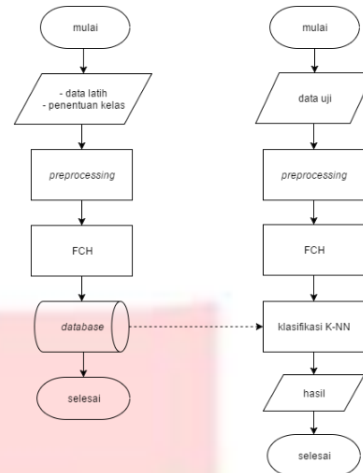
Diagram blok pada Gambar 3.1 akan dijelaskan secara rinci pada sub-bahasan berikutnya. Sedangkan, cara kerja penelitian secara umum atau menyeluruh dibagi kedalam dua skenario. Skenario pertama, metode yang digunakan untuk ekstraksi cirinya adalah DCT. Sedangkan, pada skenario kedua, metode yang digunakan ialah FCH.

Skenario 1:



Gambar 3 Flowchart model skenario 1

Skenario 2:



Gambar 2 Flowchart model skenario 2

3.1.1 Akuisisi citra

Akuisisi citra adalah tahapan untuk mendapatkan citra digital yang sesuai dengan kebutuhan. Citra digital yang digunakan sebagai data uji dan data latih pada penelitian ini didapat dari hasil *capture* kamera DSLR Canon EOS 600D dengan resolusi 18 MP. Citra digital yang belum diolah tersebut kemudian di-*crop* dengan ukuran 512×512 pixel tanpa ada latar belakang (*background*) warna. Citra yang berukuran 512×512 pixel tersebut yang selanjutnya digunakan sebagai citra *input*.

3.1.2 Preprocessing

Tahapan *preprocessing* adalah proses awal pada pengolahan citra. Citra hasil *preprocessing* nantinya digunakan sebagai input sistem. *Preprocessing* yang terjadi pada model sistem DCT dan FCH berbeda. Pada DCT, *preprocessing* yang terjadi adalah pengambilan layer *Red* sebagai komponen yang akan diolah sistem. Sedangkan pada FCH, *preprocessing*-nya berupa pengambilan nilai rata-rata dari masing-masing layer *Red*, layer *Green*, dan layer *Blue* sebagai input sistem.

3.1.3 Ekstraksi Ciri

3.1.3.1 Discrete Cosine Transform

Pada skenario penelitian 1, hasil *preprocessing* menjadi masukan pada proses ekstraksi. Dalam tahapan ini, layer *Red* suatu citra ditransformasikan ke domain frekuensi menggunakan metode DCT. Kemudian, hasil transformasi tersebut dibagi menjadi beberapa blok dengan ukuran 8×8, 16×16, 32×32, ..., 128×128. Selanjutnya diambil 1 blok dengan frekuensi terendah (DC) untuk mendapatkan ciri yang merepresentasikan citra. Alasan diambilnya frekuensi DC karena informasi terdapat didalamnya.

3.1.3.2 Fuzzy Color Histogram

Pada skenario penelitian 2, hasil *preprocessing* yang berupa rata-rata dari layer R, G, dan B digunakan sebagai input pada proses ekstraksi ciri. Selanjutnya, masuk ke 3 tahap utama sistem fuzzy, yaitu fuzifikasi, *inference*, dan defuzifikasi.

Pada fuzifikasi, masukan-masukan yang nilai kebenarannya bersifat pasti diubah ke dalam bentuk fuzzy input. Pada tahap *inference*, dibuatlah aturan-aturan dari kombinasi antara variabel dan nilai linguistiknya. Selanjutnya, tahap defuzifikasi berfungsi untuk mengubah fuzzy output menjadi *crisp value* berdasarkan fungsi keanggotaan yang telah ditentukan. Output dari defuzifikasi tersebut yang digunakan sebagai ciri.

3.1.4 Klasifikasi

Jenis klasifikasi yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor* atau K-NN. Klasifikasi dalam penelitian ini dibedakan menjadi 3 kelas yaitu, batubara kualitas tinggi, sedang, dan rendah. Ekstraksi ciri yang dimiliki data uji dicocokkan dengan kriteria pada tiap kelas yang tersimpan dalam database menggunakan metode ketetanggaan terdekat. Ketetanggaan terdekat ini dikelompokkan berdasarkan nilai k. Nilai k yang

digunakan pada klasifikasi K-NN sistem ini adalah $k=1, 3, \text{ dan } 5$. Sedangkan, jenis jarak yang digunakan dalam sistem ini yaitu *euclidean, cityblock, cosine, dan correlation*.

3.2 Performansi Sistem

Parameter yang diamati untuk mengetahui performansi sistem adalah akurasi dan waktu komputasi.

4. Analisis

4.1 Analisis Hasil Pengujian Sistem

Pengujian sistem dilakukan berdasarkan beberapa parameter yang diperoleh selama proses pemrograman sistem. Masing-masing pengujian memiliki hasil yang direpresentasikan dalam bentuk tabel dan kesimpulan dalam bentuk grafik. Skenario pengujian pada sistem ini adalah sebagai berikut:

1. Skenario 1 pengujian sistem dengan menggunakan metode ekstraksi ciri DCT dengan klasifikasi K-NN, yaitu dengan melihat perubahan parameter akurasi dan waktu komputasi terhadap ukuran blok, jenis jarak, dan nilai k yang digunakan pada K-NN.
2. Skenario 2 pengujian sistem dengan menggunakan metode ekstraksi ciri FCH dengan klasifikasi K-NN yaitu dengan melihat perubahan nilai akurasi dan waktu komputasi terhadap perubahan nilai k dan jenis jarak yang digunakan pada K-NN.

4.1.1 Analisis Resolusi Citra

Penelitian ini menggunakan 2 jenis resolusi citra untuk diuji, yaitu citra dengan resolusi 256×256 dan citra dengan resolusi 512×512 . Akurasi yang diperoleh kedua jenis resolusi citra tersebut ditunjukkan pada berikut.

Tabel 2 Akurasi DCT saat resolusi citra 256×256 dan 512×512

Resolusi citra	Akurasi (%)
256×256	58,33
512×512	78,33

Tabel 3 Akurasi FCH saat resolusi citra 256×256 dan 512×512

Resolusi citra	Akurasi (%)
256×256	33,33
512×512	61,67

Sedangkan untuk waktu komputasi, citra dengan resolusi 256×256 memiliki waktu komputasi yang lebih sedikit dibanding resolusi 512×512 (lihat Tabel 4 dan 5).

Tabel 4 Waktu komputasi DCT saat resolusi citra 256×256 dan 512×512

Resolusi citra	Waktu komputasi (s)
256×256	9,9294
512×512	19,4060

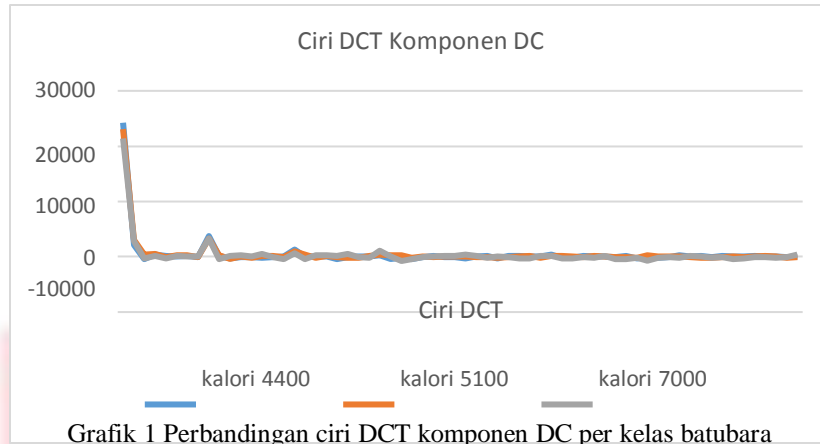
Tabel 5 Waktu komputasi FCH saat resolusi citra 256×256 dan 512×512

Resolusi citra	Waktu komputasi (s)
256×256	5,1562
512×512	9,7317

Dalam penelitian ini, terbukti bahwa makin besar resolusi citra maka waktu komputasi yang dibutuhkan oleh sistem untuk mengolah data lebih lama. Namun, berbeda halnya dengan akurasi. Akurasi dari ukuran citra 512×512 lebih tinggi karena ketika ukuran citra di-*resize* menjadi 256×256 , secara otomatis citra asli akan terkompres sehingga informasi pada citra tersebut sudah tidak dapat terbaca oleh sistem.

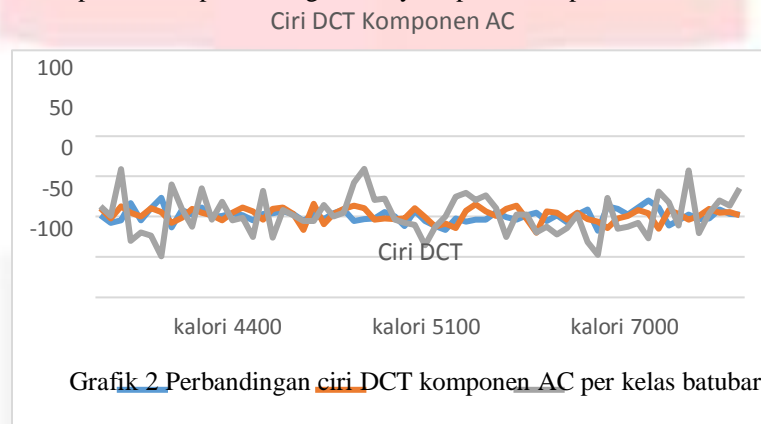
4.1.2 Analisis Hasil DCT

Dalam pengujian sistem DCT, komponen yang diujikan untuk memperoleh ciri adalah komponen DC dan komponen AC. Grafik 4.1 menunjukkan perbandingan ciri DCT komponen DC per kelas batubara.



Grafik 1 Perbandingan ciri DCT komponen DC per kelas batubara

Sedangkan untuk komponen AC, perbandingan cirinya dapat dilihat pada Grafik 2 berikut.



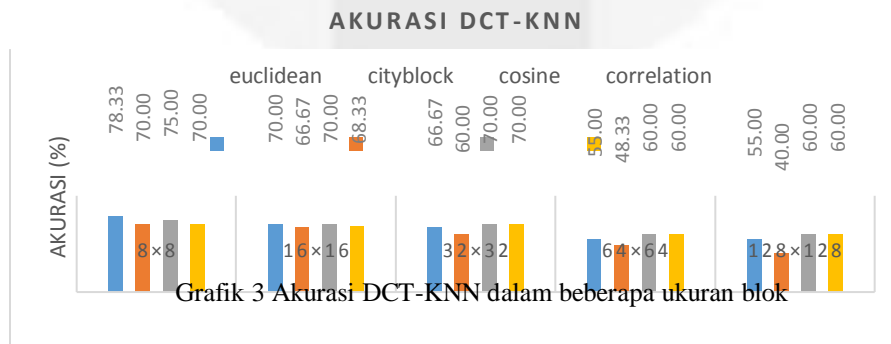
Grafik 2 Perbandingan ciri DCT komponen AC per kelas batubara

Akurasi tertinggi terjadi saat ciri yang digunakan adalah ciri pada komponen DC (lihat Tabel 6). Hal ini disebabkan ciri pada komponen DC lebih merepresentasikan karakteristik objek citra batubara.

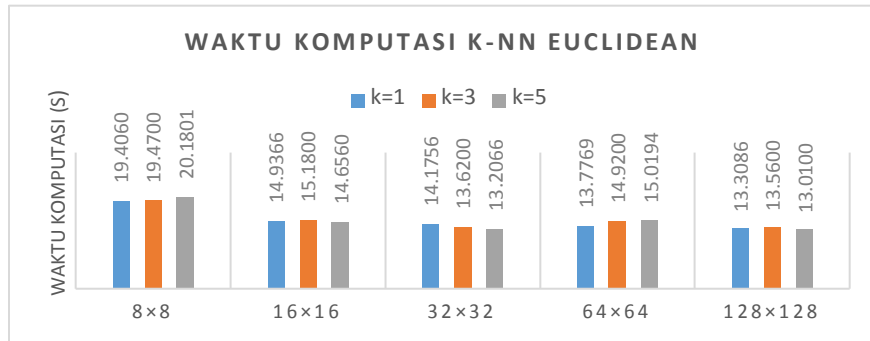
Tabel 6 Akurasi komponen DCT

Jenis komponen DCT	Akurasi (%)
DC	78,33
AC	38,33

Grafik 3 menyatakan bahwa akurasi DCT-KNN tertinggi terjadi saat ukuran bloknya 8x8 dan jenis jarak yang digunakan pada K-NN berupa *euclidean distance*. Jarak *euclidean*, *cosine*, dan *correlation* mulai mengalami saturasi ketika ukuran bloknya 64x64. Pada grafik tersebut terlihat bahwa makin besar ukuran bloknya, maka makin kecil akurasi yang dihasilkan.



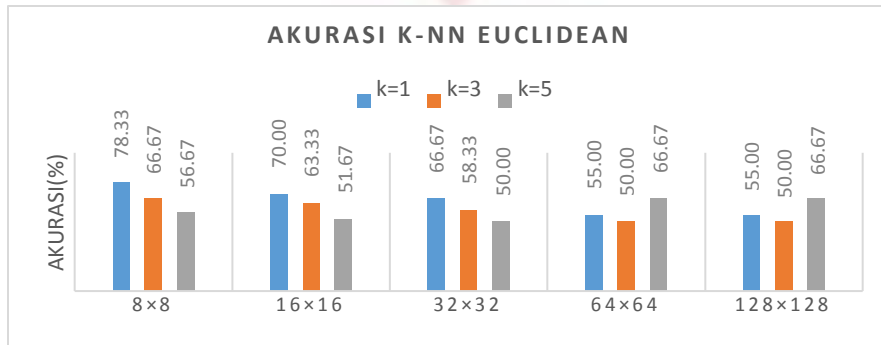
Waktu komputasi terlama yang dibutuhkan sistem untuk mengolah data terjadi ketika ukuran bloknnya sebesar 8×8 dan nilai k pada K-NN adalah 5 (lihat Grafik 4).



Grafik 4 Waktu komputasi K-NN dengan tipe jarak Euclidean saat k=1, 3, 5

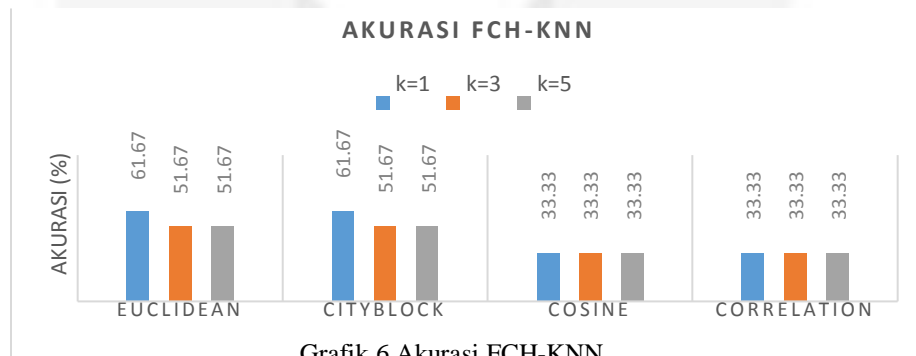
Adapun pengaruh nilai k pada K-NN terhadap performansi sistem ditunjukkan dapa Grafik 5. Akurasi tertinggi diperoleh saat k=1 pada blok 8×8, yaitu sebesar 78,33%. Dalam sistem ini k yang digunakan adalah 1, 3, dan 5.

Grafik 5 Akurasi K-NN dengan tipe jarak Euclidean saat k=1, 3, 5



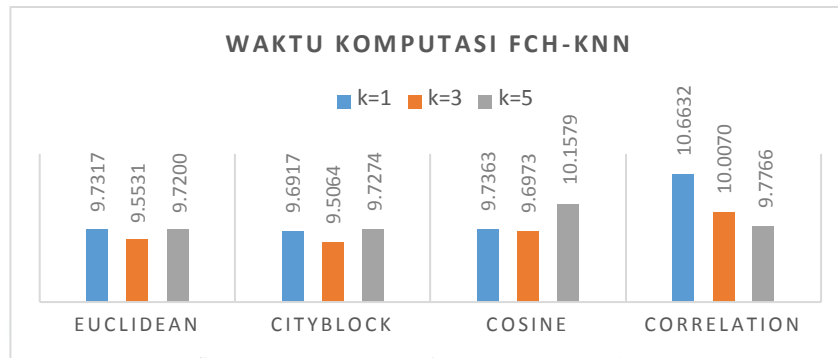
4.1.3 Analisis Hasil FCH

Jenis jarak *euclidean* dan *cityblock* memiliki akurasi yang sama baik untuk k=1, 3, ataupun 5. Sama halnya dengan jenis jarak *cosine* dan *correlation* (lihat Grafik 6). Akurasi tertinggi diperoleh saat jenis jaraknya adalah *euclidean* dan *cityblock* ketika nilai k=1. Akurasi yang diperoleh mencapai 61,67%.



Grafik 6 Akurasi FCH-KNN

Dilihat dari jenis jaraknya, waktu komputasi yang dibutuhkan memiliki selisih 1-2 sekon (lihat Grafik 7).



Grafik 7 Waktu komputasi FCH-KNN saat k=1, 3, 5

5. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian ini, antara lain:

1. Sistem ini dapat mengklasifikasikan kualitas batubara. Model DCT-KNN dengan komponen ciri DC dan ukuran blok 8×8 mampu mencapai akurasi tertinggi yaitu 78,33%. Sedangkan, model FCH-KNN mencapai akurasi sebesar 61,67%.
2. Baik model DCT-KNN ataupun model FCH-KNN, nilai k untuk mendapatkan hasil optimal adalah ketika k=1 dan jenis jaraknya *euclidean*. Namun, pada sistem FCH, jarak *euclidean* dan *cityblock* memiliki akurasi yang sama. Begitu juga dengan *cosine* dan *correlation*.
3. Performansi sistem ditinjau dari segi waktu komputasi, sistem DCT membutuhkan waktu untuk mengolah input lebih lama jika dibandingkan FCH. Faktor yang mempengaruhi lamanya waktu komputasi adalah resolusi citra yang digunakan. Makin besar resolusi citra maka waktu komputasi yang dibutuhkan. Sedangkan dilihat dari akurasinya, sistem DCT, makin besar ukuran bloknnya, akurasi yang dihasilkan justru menurun. Untuk akurasi FCH, karena objek penelitian memiliki warna homogen, pada tahap fuzifikasi terjadi banyak irisan yang menyebabkan hasil kurang presisi.

Daftar Pustaka:

- [1] _, 2009, *Sumberdaya Batubara (Tijauan Lengkap Mengenai Batubara)*, World Coal Assosiation.
- [2] Arif, Irwandy, 2014, *Batubara Indonesia*, Jakarta: Gramedia.
- [3] http://www.kirupa.com/imagesrgb_image.png diakses pada tanggal 1 Novem-ber 2015.
- [4] <http://www.mathworks.com/help/images/ref/dct2.html> diakses pada 7 Juni 2016.
- [5] http://www.tekmira.esdm.go.id/data/files/Batubara_Indonesia.pdf diakses pa-da tanggal 30 September 2015.
- [6] http://www.worldcoal.org/coal_facts_2014 diakses pada tanggal 30 September 2015.
- [7] Kadir, Abdul, Susanto, A., 2013, *Pengolahan Citra, Teori dan Aplikasi*, Yogyakarta: Andi.
- [8] Krisandi, Nobertus, et al., 2013, *Buletin Ilmiah: Algoritma K-Nearest Neighbor dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. Minamas Kecamatan Parindu*, Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster): Vol.02, No.1, hal. 33-38.
- [9] Putra, Darma, 2010, *Pengolahan Citra Digital*, Yogyakarta: Andi.
- [10] S. Aksoy. 2008. "Non Bayesian Classifier, K-Nearest Neighbor Classifier and Distance Functions". *Ankara: Bilkent University.*, vol. I, pp. 5-6.
- [11] Sutoyo, T., et al., 2009, *Teori Pengolahan Citra Digital*, Yogyakarta: Andi; Semarang: Universitas Dian Nuswantoro.
- [12] Suyanto, 2008, *Soft Computing Membangun Mesin ber-IQ Tinggi*, Bandung: Informatika.
- [13] Syarif, Syafruddin, et al., 2012, *Prosiding: Sistem Cerdas Deteksi Citra dengan Metode Discrete Cosine Transform*, Makassar: Universitas Hasanuddin.
- [14] Zhang, Roufei, dan Z. Zhongfei, 2004, *A Robust Color Object Analysis Approach to Efficient Image Retrieval*, *EURASIP Journal on Applied Signal Processing*, hal. 871–885.