

Pre-Processing Citra Untuk Biji Kopi Hijau Arabika Menggunakan Segmentasi K-Means

M. Naufal Syawali. A
Fakultas Informatika
PCD 40-01
Telkom University Bandung
kasevsyawal@student.telkomuniversity.ac.id

Muhammad Zaky Ramadhan
Fakultas Informatika
PCD 40-01
Telkom University Bandung
muridmaupintar@student.telkomuniversity.ac.id

line 1: 4th Given Name Surname
line 2: dept. name of organization
(of Affiliation)
line 3: name of organization
(of Affiliation)
line 4: City, Country
line 5: email address or ORCID

line 1: 5th Given Name Surname
line 2: dept. name of organization
(of Affiliation)
line 3: name of organization
(of Affiliation)
line 4: City, Country
line 5: email address or ORCID

Abstract—Pre-Processing menggunakan segmentasi K Means merupakan salah satu dari banyak metode pre-processing yang dapat digunakan pada biji kopi hijau, kopi menjadi salah satu minuman terpopuler pertama saat ini, salah satu alasannya adalah dapat dinikmati oleh berbagai kalangan, tetapi untuk mendapatkan rasa kopi yang sempurna diperlukan kualitas biji yang baik, pendekatan menentukan kualitas biji kopi yang dapat dilihat dari warna serta tekstur. Maka dari itu penulis melakukan percobaan Pre-Processing citra untuk biji kopi hijau menggunakan segmentasi adapun beberapa metode pendukung dalam percobaan ini antara lain *DCT Compression*, *Basic Operator* (yang digunakan adalah *Zoom* dan *Crop*) dan *Sharpening* yang dibangun menggunakan bahasa pemrograman python dengan GUI Tkinter.

Keywords—Biji Kopi Hijau, Pre-processing, Segmentasi K Meanse, DCT Compression, GUI Tkinter, Basic Operator, dan Sharpening.

I. PENDAHULUAN

Kopi menjadi salah satu minuman populer dunia saat ini. Berdasarkan jumlah produksi kopi menurut International Coffee Organization (ICO) tahun 2019 Indonesia menempati urutan kedua di Asia [4]. Dalam dunia kopi salah satu hal penting yaitu kualitas biji yang baik.

Pada Tahun 2017 dilakukan penelitian klasifikasi biji kopi hijau melalui media citra berdasarkan jenis kecacatannya secara satu per satu biji saling berjauhan dengan jarak tertentu dengan menggunakan metode Morphological Feature Extraction dan Convolutional Neural Network (CNN) dan berhasil mendapatkan jenis - jenis kecacatan pada biji kopi hijau salah satunya adalah kehitaman dan tidak layak untuk dipakai, dan akurasi yang didapat dari metode Convolutional Neural Network (CNN) untuk klasifikasi kategori jenis kecacatan sebesar 90% dari 6.500 dataset biji kopi dan 13.000 gambar [5].

Biji kopi hijau yang dicari adalah biji kopi hijau yang rusak, patah atau tidak berwarna hijau dari populasinya hal tersebut dapat dilihat dari perbedaan warna, dengan begitu penulis akan mencoba untuk mengaplikasikan preprocessing menggunakan beberapa metode seperti *segmentation*,

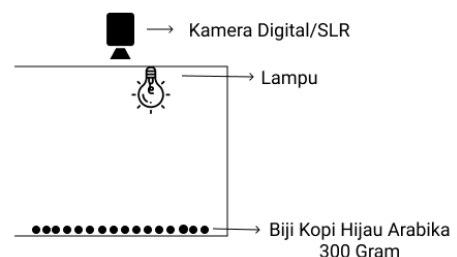
sharpening, dan *compressing*, yang bertujuan untuk mengenali biji kopi selain warna hijau.

II. METODE PENDEKATAN

A. Datasets

Teknik Pengambilan dataset yang dilakukan oleh penulis adalah dengan cara memakai 20 biji kopi hijau secara random dan disimpan diatas kain putih secara berjarak yang nantinya akan diambil citra oleh kamera digital, penulis akan mengambil sebanyak 20 citra yaitu 10 akan digunakan untuk dataset Training, dan 10 dataset citra akan digunakan untuk Testing, biji kopi hijau yang digunakan oleh penulis adalah biji kopi hijau Puntang Arabika.

Penulis menggunakan mini foto studio untuk membantu pencahayaan, agar pengambilan citra yang stabil dan menggunakan kamera digital dengan ISO 300 dan 800. Gambar 2.1, Gambar 2.2 ,dan Gambar 2.3 merupakan pengambilan dataset citra dan hasil citra yang diambil.



Gambar 2.1 Ilustrasi Pengambilan datasets



Gambar 2.2 Alat dan Biji kopi hijau diatas kain



Gambar 2.3 Contoh dataset yang digunakan

B. DCT Compression

Discrete Cosine Transform (DCT) merupakan teknik kompresi citra yang populer, DCT menyeleksi standar citra salah satunya JPEG, selain bidang DCT juga mampu melakukan kompresi terhadap video. DCT Menggunakan transformasi untuk melakukan kompresi terhadap data, dengan transformasi orthogonal yang memiliki set dasar yang tetap. DCT memiliki keunggulan seperti dapat membuat frekuensi yang lebih rendah untuk data citra [7].

C. Sharpening

Sharpening adalah salah satu tahapan dalam preprocessing citra yang digunakan untuk perbaikan citra dengan cara mempertajam kualitas citra yang bertujuan meningkatkan kontras citra sehingga objek lebih tajam dari sekitarnya. Jenis filter yang digunakan untuk melakukan sharpening adalah *High Pass Filter*. *High Pass Filter* merupakan filter yang mengambil frekuensi tinggi dan membuang frekuensi rendah. Cara kerja *High Pass Filter* dapat digunakan formula 1 berikut [1]:

$$Si, j = xi, j + \lambda F(xi, j) \quad (1)$$

1. Si, j = Nilai piksel pada koordinat i dan j yang sudah dilakukan *Sharpening*.
2. Xi, j = Nilai Piksel pada koordinat i dan j semula
3. $F()$ = High Pass Filter
4. λ = Parameter ketajaman, semakin besar semakin tajam

D. Segmentasi K Means

Segmentasi citra bertujuan untuk membagi memisahkan suatu wilayah berdasarkan atribut yang diambil dari citra tersebut dengan membagi citra ke dalam daerah intensitas masing-masing seperti tekstur atau warna yang dapat membedakan antara objek dengan *background*-nya [2].

Segmentasi citra memiliki tujuan menemukan karakteristik yang dimiliki citra. Semakin baik kualitas segmentasi maka akan semakin baik juga pengenalan pola yang akan didapatkan[3].

Algoritma K Means terdiri dari dua fase, fase pertama adalah menentukan setiap titik yang paling mendekati dengan nilai pusat dari titik masing – masing.

Ada tiga metode untuk menentukan jarak nilai terdekat dengan pusat dan yang paling sering digunakan adalah metode jarak Euclidean. Jarak Euclidean akan dikalkulasikan antara data pusat dengan titik, lalu titik dimasukkan kedalam kelompok yang memiliki nilai minimal dari jarak Euclidean.

Setiap kelompok didalam partisi mendefinisikan anggota objek dari pusatnya. Dengan begitu dapat disimpulkan K-Means adalah algoritma yang berulang dimana meminimalisasi jumlah dari jarak setiap objek ke pusat dari semua kelompok [6]. Algoritma untuk pengelompokan K Means memiliki langkah – langkah sebagai berikut:

1. Inisialisasi angka dari nilai k dan pusat.
2. Untuk setiap piksel dari citra, kalkulasikan terlebih dahulu jarak Euclidean, antara pusat dengan piksel citra dari persamaan berikut

$$d = ||p(x, y) - c_k|| \quad (2)$$

3. Masukkan semua piksel ke pusat terdekat dengan jarak dari d
4. Setelah semua dimasukan, kalkulasikan kembali dengan posisi terbaru dari pusat menggunakan persamaan berikut

$$c_k = \frac{1}{k} \sum_{y \in c_k} \sum_{x \in c_k} p(x, y) \quad (3)$$

5. Ulang langkah – langkah diatas untuk mendapatkan hasil terbaik atau nilai error.
6. Bentuk kembali kelompok piksel kedalam gambar.

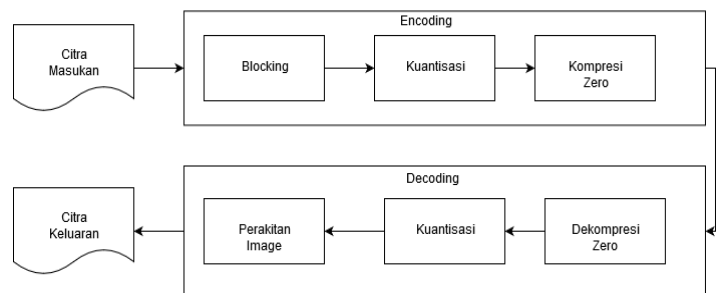
Hasil akhir pengelompokan tergantung dengan pemilihan inisialisasi pusat di awal, jika inisialisasi pusat dipilih secara acak, maka akan memberikan hasil yang berbeda dengan inisialisasi pusat. Maka dari itu inisialisasi pusat perlu dipilih dengan baik untuk mendapatkan segmentasi yang sesuai dengan harapan.

E. Segmentasi K Means

Basic operation yang akan digunakan adalah *crop* dan *resize*. *Crop* digunakan untuk membuang bagian dari citra, khususnya bagian tepi yang tidak terdapat konten atau hanya berupa *background* sehingga tidak memberikan beban komputasi terhadap sistem. *Resize* digunakan untuk memperbesar dimensi citra, karena fitur yang akan di ekstraksi oleh segmentasi k means akan menghasilkan pemetaan segmentasi yang tidak terlalu terlihat, sehingga dengan memperbesar dimensi citra keberhasilan segmentasi dapat terlihat lebih jelas.

III. PERANCANGAN SISTEM

A. DCT Compression

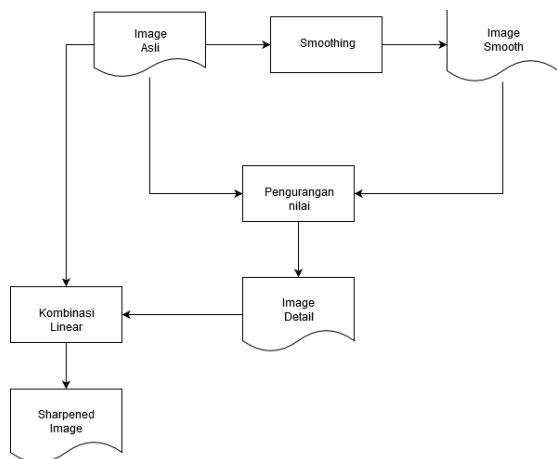


Gambar 3.1 Skema DCT Compression

Skema DCT compression dapat dilihat pada Gambar 3.1. Citra masukan akan dilakukan encoding. Pertama-tama citra akan dilakukan proses encoding. Proses encoding didalamnya terdapat proses-proses blocking, kuantisasi dan kompresi zero.

Blocking adalah memecah citra menjadi blok-blok bagian. Setelah itu dilakukan kuantisasi informasi frekuensi pada blok tersebut sehingga akan menghilangkan beberapa informasi pada citra. Hasil dari kuantisasi akan menghasilkan banyak nilai nol, sehingga nilai nol tersebut akan dikompresi. Setelah proses encoding maka akan dilakukan proses decoding, yaitu kebalikannya proses encoding yang akan melakukan perakitan blok-blok tadi menjadi sebuah citra baru [7].

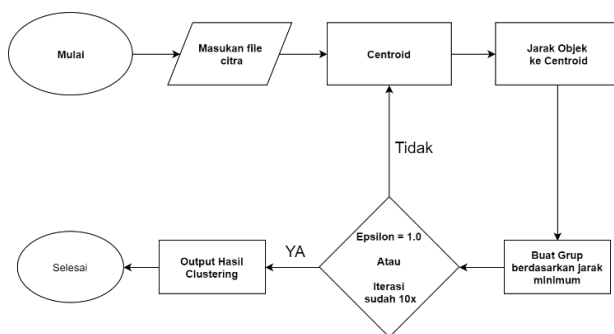
B. Sharpening



Gambar 3.2 Skema color sharpening

Skema color sharpening dapat dilihat pada Gambar 3.2. Image asli akan dilakukan smoothing atau diperhalus terlebih dahulu sehingga mengeluarkan output image yang lembut, terlihat lebih buram dan noise yang lebih sedikit. Image asli akan dilakukan pengurangan nilai dengan nilai image smooth sehingga menjadi citra yang detail, yang akan dilakukan kombinasi linear dengan nilai alpha sebagai parameter ketajaman sehingga menghasilkan sharpened image yaitu citra dengan tingkat ketajaman yang lebih tinggi. berikut merupakan contoh hasil proses sharpening citra yang telah mengalami kompresi dan menghasilkan citra dengan ketajaman warna yang lebih mencolok.

C. Segmentasi K Means



Gambar 3.3 Skema segmentasi K Means Clustering

Untuk skema atau ilustrasi penggunaan segmentasi K Means dapat dilihat pada gambar 3.3 Segmentasi dimulai dengan memasukkan file citra ke sistem, lalu sistem akan menentukan letak centroid dan jarak objek ke centroid, nantinya akan membuat grup berdasarkan hasil jarak minimum jika epsilon 1.0 atau sudah melalui iterasi sebanyak n kali maka sistem akan mengeluarkan hasil clustering, jika tidak maka akan membangkitkan centroid baru. Hasil segmentasi menggunakan K Means clustering citra yang telah dilakukan sharpening menghasilkan citra seperti yang telah tersegmentasi.

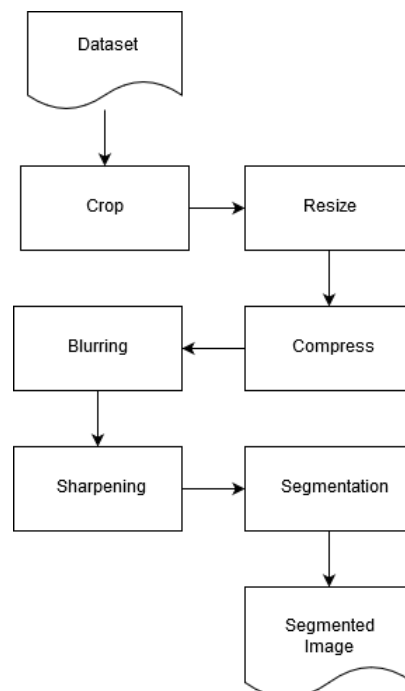
IV. HASIL PROGRAM

Sistem yang dibangun dengan bahasa pemrograman Python dan GUI Tkinter ini memiliki tampilan antarmuka seperti Gambar 4.1 berikut:



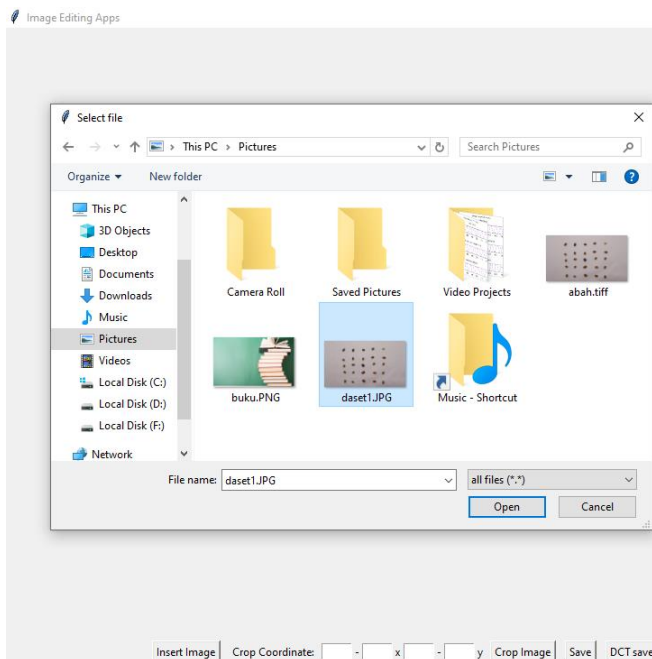
Gambar 4.1 Tampilan antarmuka sistem

Skema menjalankan program akan mengikuti skema pada Gambar 4.2 berikut:



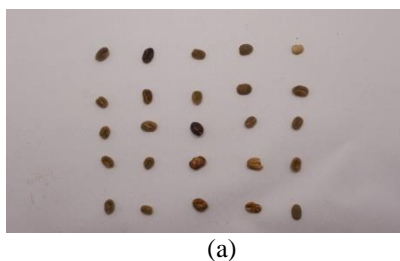
Gambar 4.2 Skema perancangan sistem

Pertama yang harus dilakukan adalah menginputkan citra yang akan di proses. Pemilihan citra yang akan diproses dapat dilihat pada Gambar 4.3.



Gambar 4.3 Tampilan *insert* citra sistem

Dataset akan dilakukan *cropping* dan *resizing* terlebih dahulu sebagai *basic operation* yang akan menghasilkan citra seperti pada Gambar 4.4, yang dimensinya telah diperbesar sebesar 20%.



(a)



(b)

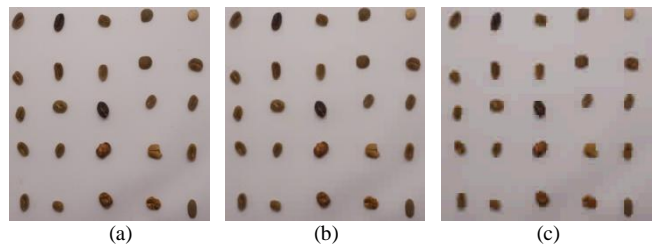
Gambar 4.4 Hasil dari *basic operation* *cropping* dan *resizing*:

(a) citra asli dan (b) citra hasil *basic operation*

Tahap selanjutnya adalah melakukan *compression*. Percobaan proses pengujian dilakukan dengan citra hasil *basic operation*, hasil proses DCT *compression* yang dilakukan dapat menurunkan ukuran citra yang semula 251 KB dengan menggunakan parameter *quality scale* sebesar 1,

5 dan, 10. Hasil proses DCT *compression* dapat dilihat pada Gambar 4.5.

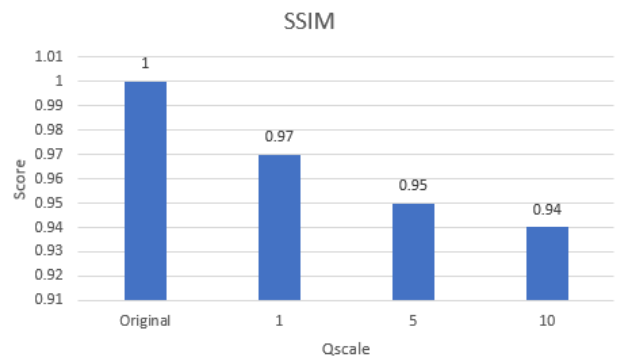
Pada Gambar 4.5 terlihat terjadi pengurangan ukuran dan detil citra, dengan citra hasil kompresi *quality scale* bernilai 1, 5 dan, 10 berturut-turut memiliki ukuran 82KB, 22KB dan, 17.3 KB.



Gambar 4.5 Hasil dari DCT *compression* dengan *qscale*:

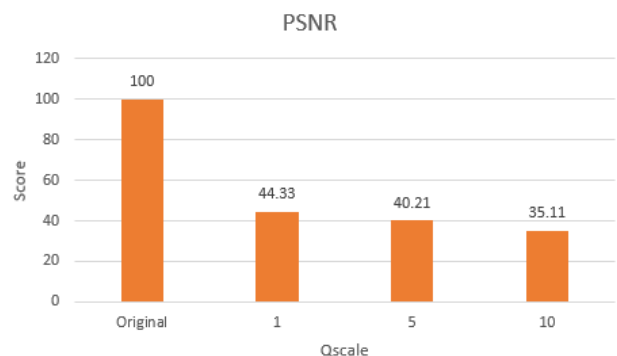
(a) *qscale* = 1 (b) *qscale* = 5 (c) *qscale* = 10

Kompresi DCT adalah kompresi *lossy* sehingga terdapat detail yang hilang, untuk menghitung seberapa besar detail citra yang hilang secara objektif maka digunakan SSIM dan PSNR yang dapat dilihat pada Gambar 4.6 dan Gambar 4.7.



Gambar 4.6 Grafik SSIM

Perbandingan kualitas citra menggunakan SSIM membandingkan kualitas sebuah citra, atau mengukur penurunan kualitas sebuah gambar (*degradasi*) yang bisa diakibatkan oleh kompresi data atau kehilangan data pada saat transmisi. Berdasarkan Gambar 4.6 hasil kompresi yang didapatkan tidak mengakibatkan skor SSIM yang jauh dari citra semula.



Gambar 4.7 Grafik PSNR

Perbandingan kualitas citra menggunakan metode PSNR adalah perbandingan antara nilai maksimum dari sinyal yang diukur dengan besarnya derau yang berpengaruh pada sinyal

tersebut. Berdasarkan Gambar 4.7 nilai PSNR citra mengalami penurunan yang signifikan sebesar 60 poin lebih.

Perbandingan penurunan kualitas dan ukuran file citra dapat dilihat pada TABLE I. berikut:

TABLE I. PENURUNAN KUALITAS DAN UKURAN

Qscale	Ukuran (KB)	PSNR	SSIM
Original	251	100	1
1	82	0.97	44.33
5	22	0.95	40.21
10	17.3	0.94	35.11

Setelah dilakukan DCT *compression* adalah melakukan *enhancement* citra dengan tujuan menghasilkan citra yang lebih tajam. Dengan menggunakan *high pass filter* maka citra terlebih dahulu dilakukan *blurring* dengan formula 4. Citra yang akan digunakan seterusnya adalah citra dengan *qscale* = 1, karena citra tersebut memiliki ukuran signifikan lebih kecil namun nilai SSIM yang tinggi.

$$Bi,j = Xi,j + \gamma G(xi,j) \quad (4)$$

1. Bi,j = Nilai piksel pada koordinat i dan j yang sudah dilakukan *blurring*.
2. Xi,j = Nilai piksel pada koordinat i dan j semula
3. $G()$ = Gaussian filter
4. γ = Parameter *smoothing*, semakin besar semakin *blur*.

Hasil blurring dapat dilihat pada Gambar 4.8 sebagai berikut:



Gambar 4.8 Citra *blur*

Selanjutnya citra hasil *blurring* akan digunakan untuk membentuk citra *detail* dengan formula 5 berikut:

$$DTi,j = Xi,j - Bi,j \quad (5)$$

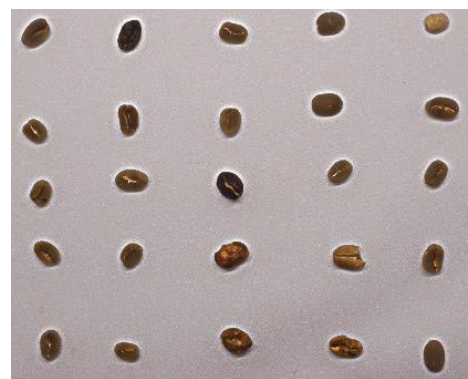
1. DTi,j = Nilai piksel pada koordinat i dan j pada citra *detail*.
2. Xi,j = Nilai piksel pada koordinat i dan j semula.
3. Bi,j = Nilai piksel pada koordinat I dan j citra *blur*.

Hasil dari operasi formula 5 akan menghasilkan citra seperti Gambar 4.9.



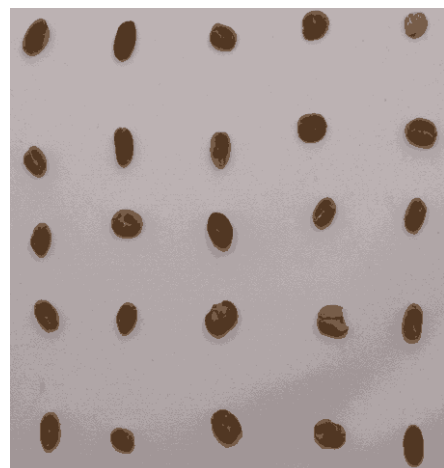
Gambar 4.9 Citra *detail*

Dari citra *detail* yang didapatkan maka akan dilakukan perhitungan menggunakan formula 1 dengan nilai alpha 5 untuk mendapatkan citra yang lebih tajam.



Gambar 4.10 Citra hasil *sharpening*

Hasil dari operasi formula 1 adalah seperti Gambar 4.10 yang merupakan citra yang telah dilakukan *sharpening*. Tahap terakhir adalah melakukan segmentasi K means. Nilai K pada segmentasi yang dilakukan adalah 5 dengan iterasi sebanyak 10 kali. Setiap segmen pada citra yang didapatkan akan diganti dengan nilai *centroid* segmen tersebut. Hasil dari segmentasi k means yang didapatkan dapat dilihat pada Gambar 4.11.



Gambar 4.11 Citra hasil segmentasi

Pada Gambar 4.11 terdapat bagian pada biji kopi yang berwarna keputihan, warna putih tersebut menandakan kondisi biji yang tidak sempurna, yaitu memiliki warna tidak hijau atau patah.

V. KESIMPULAN

Dari percobaan Pre-processing yang telah dilakukan oleh penulis, dapat disimpulkan mengalami keberhasilan, dan dapat digunakan sebagai bahan penelitian mengenai biji kopi.

Pada tahap awal penggunaan program menggunakan beberapa basic operasi seperti zoom dan crop, setelah itu dataset akan dikompresi oleh metode DCT Compression agar ukuran dataset tidak terlalu besar, dataset selanjutnya akan dimasukkan ke Tkinter, Tkinter merupakan GUI yang digunakan oleh penulis untuk memulai beberapa tahap – tahap selanjutnya.

Setelah dimasukan metode selanjutnya adalah *Sharpening* yang berguna untuk mempertajam citra, dari mulai blur, hingga dipertajam menggunakan *high pass*, alasan penulis menggunakan high pass karena ketika menggunakan kernel akan menampilkan banyak guratan pada hasil citra, setelah melalui metode *Sharpening* langkah selanjutnya adalah metode segmentasi.

Segmentasi K Means digunakan karena tujuan utama penelitian ini adalah mencari biji kopi hijau yang memiliki warna selain hijau atau tekstur yang patah, pada hasil akhir penelitian dapat terlihat tekstur biji kopi hijau yang patah dan warna biji kopi hijau yang memiliki kehitaman (dapat terlihat dari warna biji yang tidak memiliki guratan berwarna putih).

Hasil kompresi yang didapatkan menurut perhitungan SSIM terhitung baik, sedangkan menggunakan perhitungan

PSNR terdapat penurunan kualitas yang tinggi, namun sistem tetap berhasil melakukan segmentasi untuk mendeteksi kecacatan pada biji kopi.

REFERENSI

- [1] NPTEL, "Digital Image Processing," [Online]. Available: https://nptel.ac.in/courses/117104069/chapter_8/8_32.html. [Accessed 25 October 2019].
- [2] Sutoyo,T. Mulyanto, edy.Suhartono, Vincent.Dwi Mulyati, Oky. Wijanarto. " Teori Pengolahan Citra Digital ", Yogyakarta, 2009
- [3] S. Meilinda. " Implementasi Segmentasi Citra Menggunakan Metode Graph yang Efisien ", Medan , 2009.
- [4] ICO, "World Coffee Production - International Coffee Organization", [Online] Available: <http://www.ico.org/prices/po-production.pdf>. [Accessed 25 October 2019]
- [5] Pinto, C., Furukawa, J., Fukai, H., Tamura, S.: Classification of green coffee bean images based on defect types using convolutional neural network (CNN). In: 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), pp. 1–5. IEEE (2017)
- [6] Nameirakpam Dhanachandra, Yabem Jina Chanu and Kumathem Manglem Singh, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," Procedia Computer Science. 2015; 54; 764-771.
- [7] A.M.Raid, W.M.Khder, M.A.El-dosuky, and Wesam Ahmed, —Jpeg Image Compression Using Discrete Cosine Transform -A Survey || , International Journal of Computer Science & Engineering Survey (IJCSES), Vol. 5, No. 2, April 2014.