

HAAR WAVELET UNTUK EKSTRAKSI FITUR ENERGI, STANDAR DEVIASI, DAN HISTOGRAM DALAM SISTEM TEMU KEMBALI CITRA

(Haar wavelet for Extracting Energy, Standard Deviation and Histogram Features in Content Based Image Retrieval)

Adi Sugita Pandey*, I Gede Pasek Suta Wijaya, Fitri Bimantoro
Program Studi Teknik Informatika, Fakultas Teknik, Universitas Mataram
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA
Email: devil.sugita@gmail.com, gpsutawijaya@unram.ac.id, bimo@unram.ac.id

Abstract

Image retrieval initially uses a query in the form of text to search for images in the database. Image search using text query has a weakness because of the limited description of information stored or given by humans to the metadata on an inconsistent image that greatly affects the duration of searching an image in a database. Content based image retrieval (CBIR) is an image processing application to find the image sought in a large image database based on a query or user request. CBIR technique utilizes features that exist in images, namely color, texture, and shape. These features will be used as a basis for searching images in an image database. In this study the authors used the Haar wavelet method to extract color features (global color histogram and local color histogram) and texture features (energy and standar deviation) . Then the features found are matched with features stored in the database using the Euclidian distance method. In this study the authors used the Corel dataset as research material. The dataset used is classified into 4 categories: bus, dinosaurus, flower, and horse. Each category consists of 100 images where 70% are training images and 30% are test images. The retrieval experiments show that the combination of features has the best result with precision 95,83% and recall 80,93%

Keywords: Content Based Image Retrieval, Haar wavelet, Global Color Histogram, Local Color Histogram, Energy, Standar Deviation

*Penulis Korespondensi

1. PENDAHULUAN

Perkembangan teknologi informasi dewasa ini begitu pesat terutama di dalam bidang media penyimpanan. Hal-hal yang bersifat fisik kini dapat disimpan dalam bentuk digital di media penyimpanan, salah satunya adalah citra. Penyimpanan citra digital dalam jumlah banyak tentu akan menyebabkan proses pencarian citra yang diinginkan menjadi lebih sulit. Salah satu teknik pencarian citra untuk mengatasi masalah ini adalah metode *Image Retrieval*. *Image Retrieval* bekerja dengan mencocokkan data citra yang diinginkan dengan citra yang disimpan pada database.

Pada awalnya pencarian citra dilakukan secara tekstual dengan memasukan kata-kata (*keyword*) atau disebut dengan *Text Base Image Retrieval* (TBIR) dimana setiap citra diberikan *meta-data* sebelum disimpan ke dalam database. Namun teknik tersebut memiliki kekurangan dimana informasi yang dicari tidak sesuai dengan citra yang ditemukan oleh sistem.

Karena kekurangan tersebut kemudian dikembangkan teknik pencarian citra menggunakan konten visual atau disebut dengan *Content Base Image Retrieval* (CBIR). Dengan CBIR, pengguna tidak lagi menggunakan kata-kata atau *keyword* dalam mencari sebuah citra pada database melainkan menggunakan citra yang memiliki kemiripan yang akan ditemukan kembali. CBIR mulai berkembang dan banyak digunakan sebagai bahan penelitian sejak tahun 1990-an, salah satu metodenya adalah menggunakan fitur ciri pada citra seperti warna, tekstur, dan bentuk untuk pencarian citra. Hingga saat ini masih banyak penelitian tentang CBIR yang menantang peneliti dari berbagai disiplin ilmu untuk meneliti teknik tersebut.

Teknik CBIR memiliki beberapa metode yang digunakan dalam ekstraksi fitur citra. Fitur yang diekstraksi berupa ciri warna, tekstur, bentuk atau gabungan dari ciri warna, tekstur dan bentuk. Salah satu contoh metode yang digunakan penelitian CBIR

dengan ekstraksi fitur warna adalah penelitian dengan metode transformasi *Haar wavelet* dan histogram warna. Hasil penelitian menunjukkan kemiripan hasil pencarian citra query pada database rata-rata diatas 65% [1]. Selain digunakan untuk ekstraksi fitur warna, metode *Haar wavelet* juga dapat digunakan untuk mengekstraksi fitur tekstur pada citra. Hal ini dibuktikan pada penelitian pada 1000 citra yang diekstraksi ciri teksturnya menggunakan *Haar wavelet* menghasilkan akurasi yang baik dan kecepatan komputasi 1000 citra pada database yang hanya berlangsung selama 2 menit 30 detik [2].

Penelitian diatas merupakan hal yang mendasari penulis melakukan penelitian tugas akhir tentang sistem temu kembali citra dengan metode *Haar wavelet* dan histogram warna. Yang membedakan penelitian sebelumnya dan yang akan dilakukan penulis adalah jumlah fitur yang akan digunakan dalam pencarian kemiripan citra. Untuk fitur tekstur terdapat 24 fitur dan pada fitur warna terdapat 5 fitur. Pada penelitian yang dilakukan digunakan *dataset* corel untuk menguji metode *Haar wavelet* dan histogram warna pada sistem yang akan dibuat. *Dataset* corel merupakan dataset standar yang digunakan oleh peneliti karena terdiri dari citra dengan tekstur, warna, dan bentuk yang beragam.

2. TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Content Based Image Retrieval (CBIR) mulai berkembang sejak tahun 1990-an. Mengingat perkembangan penyimpanan data pada media penyimpanan yang sangat masif, penelitian tentang CBIR masih banyak dilakukan. Penggunaan CBIR saat ini tidak hanya terbatas pada pencarian objek pada media penyimpanan tapi juga pada bidang keamanan suatu sistem. Proses umum CBIR adalah melakukan ekstraksi fitur pada citra query dan citra latih. Parameter fitur citra yang dapat digunakan dalam sistem temu kembali dapat berupa histogram, susunan warna, tekstur, bentuk, tipe spesifik objek, event tertentu, nama individu, lokasi, emosi.

Penelitian CBIR menggunakan metode *Haar wavelet* dilakukan pada dataset corel 200 citra dengan 4 kategori yaitu bus, dinosaurus, bunga mawar, dan kuda. Masing-masing kategori terdiri dari 50 citra. Pada penelitian dilakukan dekomposisi *Haar wavelet* pada citra dengan model warna RGB hingga pada dekomposisi level-4. Lalu fitur citra diekstraksi menggunakan F-norm teori. Hasil dari penelitian menunjukkan nilai *recall* rata-rata adalah 71,25. Dengan

hasil *recall* yang ditunjukan itu berarti metode *Haar wavelet* dapat digunakan dalam sebuah sistem temu kembali citra [3].

Penelitian tentang CBIR menggunakan transformasi *Haar Wavelet* dan histogram warna dilakukan dengan data latih berukuran 256x256 piksel sejumlah 47 citra, dan citra query dengan ukuran yang sama sejumlah 13 citra dengan warna utama yang berbeda. Perhitungan jarak yang digunakan untuk menghitung kemiripan pada sistem temu kembali ini adalah Euclidian distance. Pengujian dilakukan oleh 3 orang responden yang berbeda, sehingga didapatkan rata-rata kemiripan sebesar 66,35% [1].

Penelitian CBIR menggunakan transformasi *Haar wavelet* dengan dekomposisi hingga level-3 menghasilkan akurasi temu kembali yang lebih baik. Fitur yang digunakan dalam penelitian adalah fitur histogram warna dan energi yang didapat dari dekomposisi *Haar wavelet* pada tiap level dekomposisi. Penghitungan jarak yang dilakukan menggunakan Swain & Ballard distance dan Euclidian distance. Hasil yang diperoleh dari penelitian menunjukkan kecepatan temu kembali citra meningkat yaitu 2 menit 3 detik untuk mencari citra pada 1000 data citra dataset corel pada database [2].

Penelitian CBIR menggunakan metode Local Color Histogram, Global Color Histogram dan gabungan keduanya dilakukan menggunakan 30 citra uji berwarna RGB yang dikelompokkan menjadi 3 kategori yaitu binatang 10 citra, bunga mawar 10 citra, dan mobil bus 10 citra. Penghitungan jarak dilakukan dengan menggunakan Euclidian distance. Pada metode Local Color Histogram (LCHs) ditemukan 21 citra, metode Global Color Histogram (GCHs) ditemukan 18 citra, dan metode gabungan (LCHs+GCHs) ditemukan 15 citra. Kemiripan citra hasil query lebih banyak dihasilkan oleh metode local color histogram (LCHs) dibandingkan dengan global color histogram atau keduanya. Hal ini dikarenakan informasi yang dihasilkan LCHs lebih detail berdasarkan blok-blok piksel [4]. Kemudian penelitian menggunakan histogram warna pada sistem kembali citra dilakukan dengan membagi citra menjadi $n \times n$ blok. Semakin besar blok maka nilai akurasi akan semakin besar, namun kecepatan komputasi lebih lambat. Sebaliknya semakin kecil blok akan membuat proses komputasi semakin cepat [5].

Berdasarkan tinjauan pustaka yang dijadikan acuan penelitian tugas akhir ini, penulis akan membuat sistem temu kembali berbasis konten/isi dengan menggunakan gabungan metode *Haar wavelet* dan histogram warna. Yang membedakan penelitian ini

dengan penelitian sebelumnya yaitu fitur yang digunakan sejumlah 29 fitur yang terdiri dari 24 fitur tekstur dan 5 fitur warna. Fitur tekstur didapat dari hasil dekomposisi citra hingga level-3 dekomposisi berupa energi dan standar deviasi. Dari tiap level dekomposisi Haar wavelet masing-masing level dekomposisi menghasilkan 4 nilai energi dan 4 nilai standar deviasi. Fitur warna didapat dari nilai aproksimasi dekomposisi Haar wavelet level-1. Fitur warna yang digunakan dalam 1 histogram warna secara global dan 4 histogram blok-blok warna lokal sebuah citra. Serta dataset yang digunakan yaitu dataset corel dengan jumlah 400 citra yang terdiri dari 4 kategori yaitu bus, dinosaurus, mawar dan kuda dengan masing-masing kategori terdiri dari 30% citra uji dan 70% citra latih.

2.2 Dasar Teori

2.2.1 Citra

Secara matematis citra (*image*) merupakan fungsi menerus (*continue*) dari intensitas cahaya pada bidang dua dimensi. Ada dua jenis citra yaitu citra diam (*still image*) yaitu citra tunggal yang tidak bergerak, dan citra bergerak (*moving image*) yaitu rangkaian citra diam disebut *frame* yang ditampilkan secara beruntun (sekuensial) sehingga memberi kesan bergerak pada indera manusia. Jenis-jenis citra berdasarkan nilai pikselnya adalah *biner*, *greyscale*, dan citra warna [6].

Pengolahan citra merupakan proses manipulasi citra menggunakan komputer menjadi citra yang kualitasnya lebih baik. Pengolahan citra bertujuan untuk memperbaiki kualitas citra agar sudah diinterpretasi oleh indera mata manusia atau mesin (komputer). Operasi – operasi pengolahan citra biasanya dilakukan pada citra apabila:

1. Manipulasi citra dilakukan untuk meningkatkan kualitas citra atau untuk menonjolkan beberapa aspek informasi pada citra tersebut.
2. Elemen pada citra perlu dicocokkan, dikelompokkan atau diukur.
3. Sebagian citra perlu digabungkan dengan citra yang lain.

2.2.2 Warna

Selain RGB warna juga dapat dimodelkan berdasarkan atribut warnanya. Setiap warna memiliki tiga atribut, yaitu *hue*(H), *saturation*(S) dan *value*(V) yang didasari oleh pendekatan yang digunakan pelukis dalam menggunakan *tint*, *shade* dan *tone*.

a. Hue

Menyatakan warna sebenarnya pada sebuah objek, misalnya merah, hijau, biru. *Hue* digunakan untuk membedakan warna satu sama lain seperti kemerahan (*redness*), kehijauan (*greenness*), kebiruan (*blueness*) dari cahaya. *Hue* didefinisikan dari 0 hingga 255 dimana 0 menyatakan merah lalu memutar nilai nilai *spectrum* kembali ke 0 untuk menyatakan merah lagi, ini dapat dipandang sebagai sudut 0° sampai 360°.

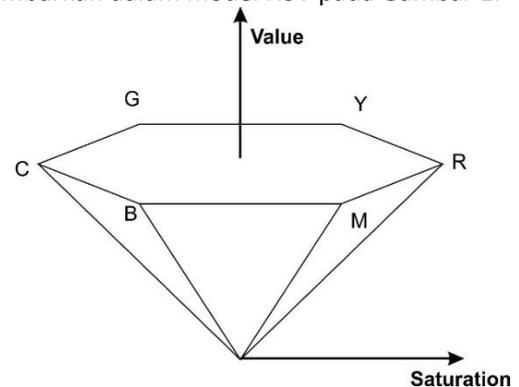
b. Saturation

Menyatakan tingkat kemurnian warna cahaya, yaitu mendefinisikan banyaknya warna putih yang diberikan pada warna. Contoh warna merah adalah 100% warna jenuh (*saturated color*), sedangkan warna *pink* (merah muda) adalah warna merah dengan tingkat kejenuhan yang rendah (karena ada warna putih di dalamnya).

c. Value

Merupakan atribut yang menyatakan banyaknya cahaya yang diterima oleh mata tanpa mepedulikan warna. Kisaran nilainya adalah antara 0 = gelap (hitam) dan 255 = terang (putih).

Ketiga atribut warna *hue*, *saturation* dan *value* digambarkan dalam model HSV pada Gambar 1.



Gambar 1 HSV hexagon

2.2.3 Kuantisasi Warna

Kuantisasi warna adalah proses untuk mereduksi warna yang ada dalam citra tanpa menghilangkan informasi penting dalam citra tersebut. Untuk mengkuantisasi warna dapat dilakukan dengan cara mengelompokkan warna ke dalam partisi, jumlah partisi yang ditentukan dalam kuantisasi warna disebut bins. Warna dimasukkan ke dalam bins sesuai rentang warna yang digunakan dalam pengolahan citra.

Kuantisasi warna pada citra dengan model warna HSV dapat dilakukan dengan mengelompokkan masing-masing kanal warna HSV ke dalam bins yang berbeda. Misalnya kanal hue dikuantisasi menjadi 8 bins, kanal saturation dikuantisasi menjadi 3 bins dan kanal value menjadi 3 bins. Kemudian nilai kuantisasi

pada tiap kanal digabung menjadi satu sehingga didapatkan $8 \times 3 \times 3 = 72$ total warna. Untuk menghitung kuantisasi dapat dilakukan dengan menggunakan Persamaan (1) dan Persamaan (2)

$$QH = \left[H \times \frac{\text{bins } H}{\max(H)} \right], QS = \left[S \times \frac{\text{bins } S}{\max(S)} \right], \text{ dan } QV = \left[V \times \frac{\text{bins } V}{\max(V)} \right] \quad (1)$$

$$QT = (\text{bins } S \times \text{bins } V \times QH) + (\text{bins } V \times QS) + QV \quad (2)$$

Dimana H,S,V merupakan nilai piksel dari masing-masing kanal HSV. Nilai *bins* merupakan nilai integer positif dari nilai kuantisasi yang ditentukan.

2.2.4 Histogram Citra

Informasi pada citra digital dapat diketahui dengan membuat histogram citra tersebut. Histogram citra merupakan sebuah grafik yang menggambarkan penyebaran nilai-nilai intensitas piksel dari suatu atau bagian dari sebuah citra. Dari sebuah histogram dapat diketahui frekuensi kemunculan nisbi (*relative*) dari intensitas pada citra tersebut. Histogram juga dapat menunjukkan tentang kecerahan (*brightness*) dan kontras (*contrast*) dari sebuah citra. Karena ini histogram merupakan alat bantu yang berharga dalam pekerjaan pengolahan citra baik secara kualitatif maupun kuantitatif.

Histogram warna terdiri dari 2 tipe yaitu *Global color histogram* (GCH) dan *Local color histogram* (LCH).

1. *Global color histogram*

Global color histogram mengambil distribusi warna pada sebuah citra secara global untuk membandingkan citra tersebut dengan citra lain.

2. *local color histogram*

Local color histogram membagi citra menjadi beberapa bagian kemudian menghitung histogram warna tiap bagian tersebut. LCH memiliki lebih banyak informasi daripada GCH namun metode ini membutuhkan lebih banyak proses komputasi.

2.2.5 Sistem Temu Kembali Citra Citra

Sistem temu kembali dapat diartikan sebagai upaya untuk mengingat, menemukan informasi yang relevan dengan kebutuhan pemakai untuk digunakan kembali. Sistem temu kembali citra adalah aplikasi pengolahan citra yang digunakan untuk mengambil atau mencari dengan cepat suatu citra pada database berdasarkan query atau permintaan pengguna.

Teknik yang digunakan untuk menentukan kemiripan fitur pada suatu citra adalah sebagai berikut:

1. Warna

Penghitungan jarak pada suatu citra dengan fitur warna digunakan dengan menghitung histogram warna yang ada pada tiap piksel citra. Penghitungan kemiripan citra dengan fitur warna banyak digunakan dalam sistem temu kembali citra karena warna tidak bergantung pada ukuran dan orientasi citra.

2. Tekstur

Penghitungan jarak menggunakan fitur tekstur dilakukan dengan mencari pola visual dari citra dan bagaimana pola tersebut didefinisikan secara spasial. Tekstur representasikan oleh *texture element* yang ditempatkan dalam sejumlah himpunan.

3. Bentuk

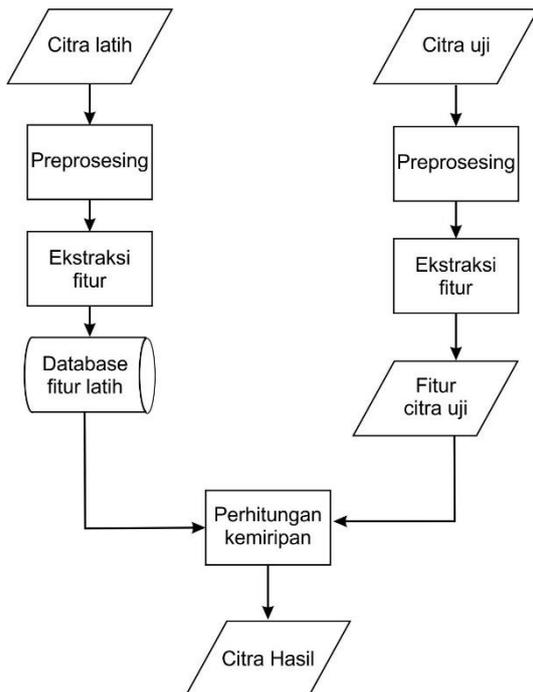
Fitur bentuk tidak merujuk pada bentuk citra, tapi bentuk wilayah tertentu pada citra. Fitur bentuk pada umumnya dilakukan dengan segmentasi atau deteksi piksel tepi pada citra. Pada beberapa kasus, deteksi yang akurat memerlukan campur tangan manusia karena metode seperti segmentasi sangat sulit dilakukan secara otomatis.

Selanjutnya fitur-fitur tersebut disimpan dan digunakan untuk mencari jarak antara citra latih dan citra uji. Penghitungan jarak artinya membandingkan kesamaan antara citra uji dan citra latih dalam berbagai aspek seperti warna, tekstur, bentuk dan sebagainya. Sebagai contoh apabila jarak antara dua citra adalah 0, maka keduanya sama persis. Jarak yang lebih besar dari 0 menunjukkan tingkat kemiripan kedua citra tersebut. Kemudian hasil penghitungan jarak tersebut diurutkan mulai dari jarak terendah, lalu ditampilkan kepada pengguna sesuai urutan. Alur sistem temu kembali citra ditunjukkan oleh Gambar 2.

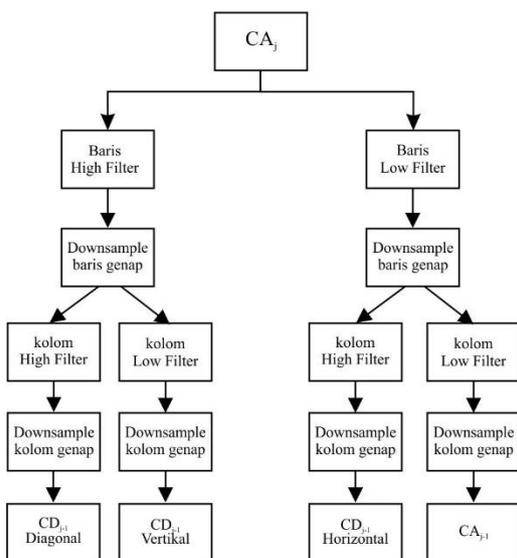
2.2.6 Haar wavelet

Transformasi wavelet merupakan suatu proses perubahan data dalam bentuk lain agar lebih mudah dianalisis. Transformasi wavelet menghasilkan energi citra yang terkonsentrasi pada sebagian kecil koefisien transformasi dan kelompok lain yang mengandung sedikit energi [2]. Proses transformasi wavelet dapat dilakukan dengan konvolusi atau dengan proses pererataan dan pengurangan secara berulang. Proses ini banyak digunakan pada proses dekomposisi, deteksi, pengenalan (*recognition*), temu kembali citra (*image retrieval*), dan lainnya yang masih dalam penelitian.

Transformasi wavelet merupakan dekomposisi citra dengan cara melewatkan sinyal frekuensi tinggi (*highpass filter*) dan frekuensi rendah (*lowpass filter*). Adapun gambaran tentang algoritma ini ditunjukkan pada Gambar 3.



Gambar 2. Diagram alir CBIR



Gambar 3 Ilustrasi Haar wavelet

Citra awal (CA_j) setelah dilakukan transformasi Haar wavelet (HWT) akan memiliki 4 output atau dalam penelitian ini kita sebut *subband* yaitu:

1. Koefisien Aproksimasi (CA_{j+1}) atau disebut *subband* LL
2. Koefisien Detil Horizontal (CD(h)_{j+1}) atau disebut *subband* LH
3. Koefisien Detil Vertikal (CD(v)_{j+1}) atau disebut juga *subband* HL
4. Koefisien Detil Diagonal (CD(d)_{j+1}) atau disebut juga *subband* HH

Fitur yang digunakan dalam penelitian ini adalah fitur energi dan standar deviasi pada wavelet. Energi menggambarkan keseragaman tekstur pada citra. Nilai energi sendiri dapat diambil dari 4 nilai-nilai subband LL, LH, HL dan HH yang nilainya tergantung pada nilai wavelet-nya[8]. Energi dibagi menjadi 4 ciri yaitu:

1. Energi yang berhubungan dengan nilai pendekatan (aproksimasi)/E_a, E_a dihitung berdasarkan persentase jumlah kuadrat dari nilai koefisien aproksimasi (c_a) dibagi dengan seluruh koefisien c.
2. Energi yang berhubungan dengan nilai detail arah horizontal/E_h, E_h dihitung berdasarkan persentase jumlah kuadrat dari nilai koefisien arah horizontal (c_h) dibagi dengan seluruh koefisien c.
3. Energi yang berhubungan dengan nilai detail arah vertikal/E_v, E_v dihitung berdasarkan persentase jumlah kuadrat dari nilai koefisien arah horizontal (c_v) dibagi dengan seluruh koefisien c.
4. Energi yang berhubungan dengan nilai detail arah diagonal/E_d, E_d dihitung berdasarkan persentase jumlah kuadrat dari nilai koefisien arah horizontal (c_d) dibagi dengan seluruh koefisien c.

Energi dan standar deviasi dari subband wavelet dirumuskan dengan Persamaan (3) dan Persamaan (4)

$$Energi = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N |X_{ij}| \tag{3}$$

$$Standar Deviasi = \left[\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (X_{ij} - \mu_{ij})^2 \right]^{\frac{1}{2}} \tag{4}$$

Dimana M×N adalah ukuran dari *subband* wavelet, X_{ij} adalah koefisien wavelet, dan μ_{ij} adalah nilai rata-rata dari koefisien wavelet. Nilai dari setiap energi dan standar deviasi inilah yang akan disimpan ke dalam database sebagai representasi fitur pada setiap citra.

2.2.7 Pengukuran Jarak

Untuk membandingkan sebuah citra dengan yang lain, perlu dilakukan perhitungan jarak. Citra yang memiliki jarak paling kecil merupakan citra yang paling mirip dengan citra yang dicari. Cara termudah untuk mengukur jarak yaitu dengan menggunakan Persamaan (5)

$$d(A, B) = \sum_{j=1}^n |H_j^A - H_j^B| \tag{5}$$

dimana : d(A,B) = jarak A dengan B
 n = jumlah piksel
 H_j^A = nilai nilai ke j pada histogram A
 H_j^B = nilai nilai ke j pada histogram B

Cara lain untuk menghitung jarak antara dua citra yaitu dengan menggunakan rumus *Euclidian distance* pada Persamaan (6).

$$d(A, B) = \sqrt{\sum_{j=1}^n (|H_j^A - H_j^B|)^2} \quad (6)$$

2.2.8 Precision dan Recall

Precision dan *Recall* merupakan perhitungan untuk mengukur efektivitas dalam sistem temu kembali informasi maupun citra. *precision* menunjukkan jumlah data yang relevan ditemukan kembali, sedangkan *recall* munjukan jumlah data relevan pada database citra.

Dalam perhitungan *precision* dan *recall*, terdapat himpunan citra mirip yang berhasil ditemukan kembali (*relevant images retrieved*) dan himpunan citra yang ditemukan (*Total images retrieved*). Perhitungan *precision* dapat dilakukan dengan Persamaan(7) [9].

$$Precision = \frac{\text{Citra relevan yang ditemukan}}{\text{Jumlah citra yang ditemukan}} \quad (7)$$

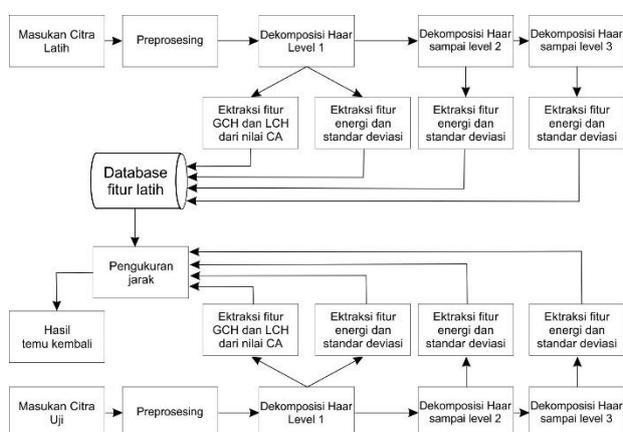
Precision mengambil semua citra untuk ditampilkan, namun hanya menampilkan sejumlah citra sesuai urutan dengan jumlah yang ditentukan. Untuk menghitung *recall* dapat menggunakan Persamaan (8).

$$Recall = \frac{\text{Jumlah citra relevan yang ditemukan}}{\text{Jumlah citra relevan pada database}} \quad (8)$$

3. METODE PENELITIAN

3.1 Prosedur Penelitian

Langkah-langkah pada pengembangan sistem ini dapat dilihat pada Gambar 4



Gambar 4 Blok diagram sistem

Gambar 4 menunjukkan proses pengolahan data latih dan uji pada sistem. Proses ekstraksi fitur, lalu diukur kemiripannya pada citra yang telah dilatih. Langkah-langkah sistem usulan adalah:

Langkah 1 : Preprocessing proses menyesuaikan citra yang digunakan dalam sistem untuk mempermudah cara kerja sistem. Karena dalam usulan digunakan dataset dengan ukuran citra yang sama yaitu 120x80 piksel (*landscape*) dan 80x120 piksel (*portrait*) pada preprocessing hanya dilakukan rotasi untuk citra dengan orientasi *portrait* sehingga ukuran citra yang diolah sistem sama yaitu 120x80 piksel.

Langkah 2 : Dekomposisi Haar level 1, setelah citra melalui proses preprocessing, kemudian dilakukan proses dekomposisi Haar wavelet level 1 pada citra sehingga didapatkan nilai koefisien aprosimasi(Ca), koefisien detil horizontal (Ch), koefisien detil vertical (Cv) dan koefisien detil diagonal (Cd).

Langkah 3 : Ekstraksi fitur GCH dan LCH, koefisien aprosimasi (Ca) dari hasil dekomposisi Haar level 1 membawa informasi warna pada citra. Untuk mendapatkan fitur GCH dan LCH dilakukan konversi ruang warna Ca dari RGB ke ruang warna HSV. Kemudian dilakukan perhitungan histogram untuk mendapatkan fitur GCH. Untuk mendapatkan fitur LCH dilakukan segmentasi pada hasil kuantisasi Ca menjadi empat blok. Kemudian masing masing blok dilakukan perhitungan histogram sehingga didapat 4 fitur *local color histogram*.

Langkah 4 : Ekstraksi fitur energi dan standar deviasi, pada tiap dekomposisi wavelet didapatkan nilai Ca, Ch, Cv dan Cd untuk mendapatkan fitur energi dan standar deviasi dilakukan perhitungan menggunakan Persamaan (3) dan Persamaan (4). Dilakukan proses dekomposisi Haar wavelet hingga level 3 sehingga pada sistem usulan didapatkan 12 fitur energi dan 12 fitur standar deviasi.

Langkah 5 : Penyimpanan fitur dan pengukuran jarak, selanjutnya dilakukan penyimpanan hasil ekstrasi fitur citra latih pada database citra.

Langkah 6 : Pengukuran jarak, setelah dilakukan ekstraksi pada citra uji, dilakukan pengukuran jarak fitur-fitur yang telah diekstraksi sebelumnya, adapun beberapa scenario pengukuran jarak adalah sebagai berikut:

1. Fitur GCH
2. Fitur LCH
3. Fitur energi
4. Fitur standar deviasi

Serta kombinasi dari persentasi hasil temu kembali terbaik dari fitur fitur tersebut.

Langkah 7 : Menghitung nilai *precision* dan *recall*, selanjutnya dilakukan perhitungan nilai *precision* dan *recall* pada tiap fitur yang diuji untuk mengukur seberapa efektif sistem yang diusulkan.

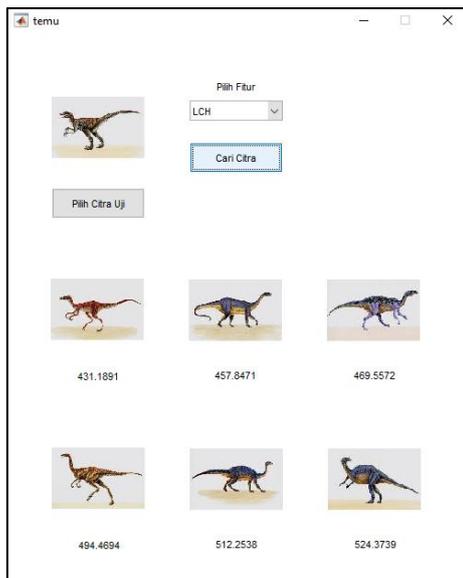
4. HASIL DAN PEMBAHASAN

4.1. Data Pengujian

Data yang digunakan dalam pengujian ini adalah dataset corel dari Corel Photo Gallery yang diklasifikasikan menjadi 4 kategori yaitu bus, dinosaurus, bunga dan kuda. Citra pada masing-masing kategori terdiri dari 100 citra dengan resolusi 120x80 piksel (*landscape*) dan 80x120(*portrait*) dengan kedalaman warna 24 bit. Jumlah data yang digunakan dalam penelitian adalah 70% untuk data latih dan 30% untuk data uji, jadi terdapat 280 citra latih dan 120 citra uji.

4.2. Tampilan Sistem

Untuk mempermudah interaksi pengguna dengan sistem dibuat tampilan sederhana menggunakan *Guide* pada matlab. Adapun tampilan saat sistem dijalankan ditunjukkan oleh Gambar 5.



Gambar 5 Hasil temu kembali

Pada saat sistem temu kembali dijalankan, user diminta untuk memilih citra dan fitur yang akan diuji pada sistem. Kemudian hasil pencarian akan ditampilkan pada *axes* diikuti dengan hasil perhitungan jarak pada bagian bawah citra yang ditunjukkan pada Gambar 5.

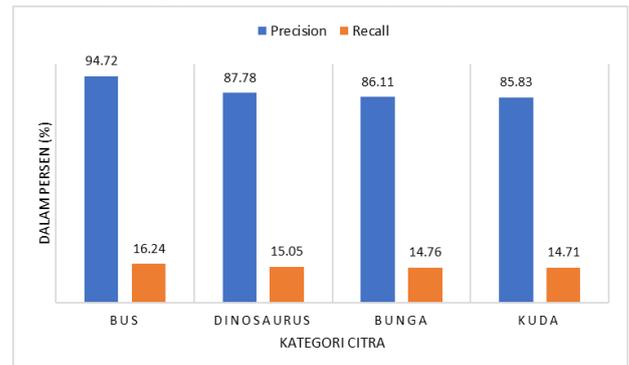
4.3. Hasil Pengujian

Pengujian dilakukan dengan menggunakan 400 citra dengan 4 kategori yaitu bus, dinosaurus, bunga dan kuda. Masing-masing kategori terdiri dari 100 citra

dimana 30% merupakan citra uji dan 70% merupakan citra latih. Dilakukan pengujian fitur terhadap semua citra uji dengan menampilkan hasil 12 citra dengan jarak terdekat lalu dihitung nilai *precision* dan *recall* dari masing-masing kategori. Kemudian dilakukan pengujian dengan mengkombinasikan berbagai fitur. Hasil dari pengujian adalah sebagai berikut:

1. Fitur GCH

Hasil pengujian fitur GCH disajikan pada Gambar 6

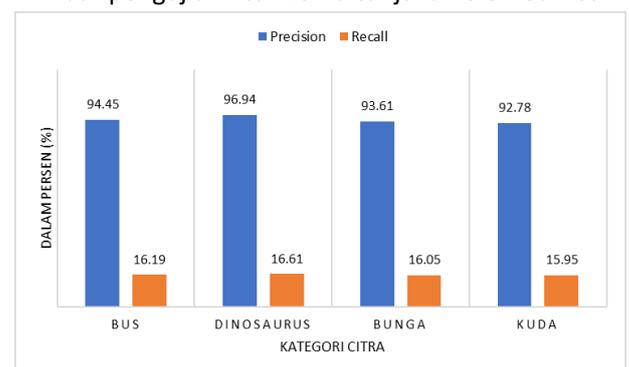


Gambar 6 Hasil pengujian fitur GCH

Hasil pengujian fitur GCH menunjukkan nilai *precision* dan *recall* dari hasil uji 120 buah citra uji pada citra dengan kategori bus memiliki nilai *precision* paling tinggi yaitu sekitar 94,72%. Rata-rata nilai *precision* masing-masing kategori citra adalah 88,61% dan nilai rata-rata *recall*nya adalah 15,19%, hal ini berarti performa sistem menggunakan fitur *global color histogram* sangat baik.

2. Fitur LCH

Hasil pengujian fitur LCH ditunjukkan oleh Gambar 7



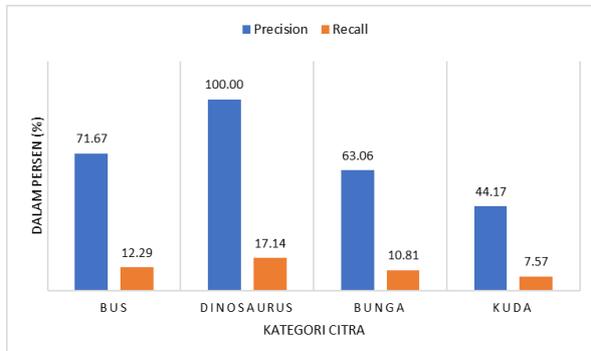
Gambar 7 Hasil pengujian fitur LCH

Pada fitur LCH nilai *precision* terbesar ditunjukkan oleh citra dengan kategori dinosaurus dengan nilai 96,94% diikuti dengan citra bus 94,45%, bunga 93,61% dan kuda 92,78%. Nilai rata-rata *precision* fitur LCH adalah 94,45% dan nilai rata-rata *recall*nya adalah

16,20% maka performa sistem menggunakan fitur LCH dapat dikatakan sangat baik.

3. Fitur energi

Hasil pengujian sistem menggunakan fitur energi ditunjukkan oleh Gambar 8

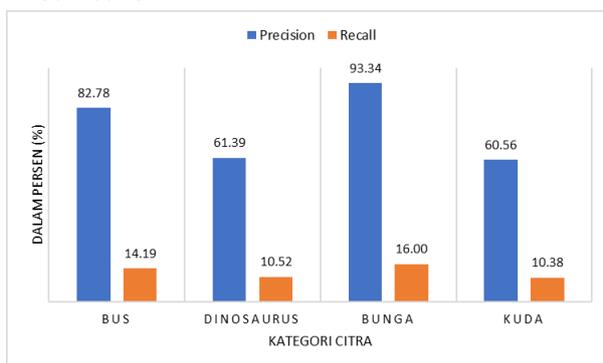


Gambar 8 Hasil pengujian fitur energi

Pada pengujian dengan fitur energi citra dengan kategori dinosaurus menghasilkan nilai *precision* yang sempurna yaitu 100% sedangkan citra dengan kategori kuda menghasilkan nilai *precision* yang cukup rendah yaitu 44,16%. Ini berarti performa sistem dengan menggunakan fitur energi sangat dipengaruhi dengan citra yang digunakan untuk pelatihan data dan pengujian data. Nilai rata-rata *precision* fitur energi adalah 69,73% dan nilai rata-rata *recall*nya 11,95%. Hal ini berarti penggunaan fitur energi tidak menghasilkan hasil sebaik fitur GCH dan LCH.

4. Fitur standar deviasi

Hasil pengujian fitur standar deviasi disajikan pada Gambar 9

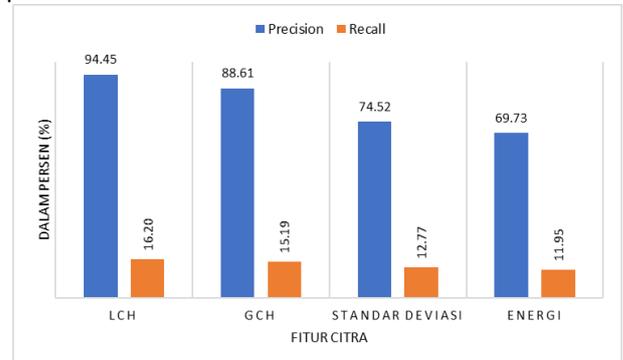


Gambar 9 Hasil pengujian fitur standar deviasi

Pada pengujian fitur standar deviasi citra bus menghasilkan nilai *precision* 82,78% dinosaurus dengan nilai 61,38%, bunga 93,34% dan kuda 60,55%. Fitur standar deviasi memiliki nilai rata-rata *precision*

74,52% dan recall 12,77%. Ini berarti fitur standar deviasi tidak cukup baik.

Dari hasil pengujian, nilai *precision* dan *recall* rata-rata fitur terbaik dalam sistem sementara ditunjukkan pada Gambar 10.



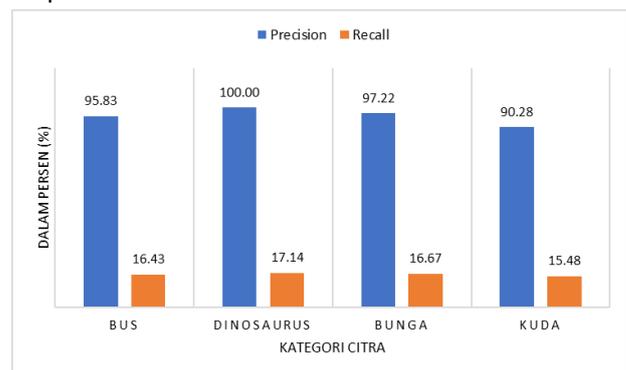
Gambar 10 Hasil pengujian fitur temu kembali

Gambar 10 menunjukkan nilai *precision* rata-rata fitur terbaik hingga fitur yang paling tidak baik. Dimana pada grafik ditunjukkan fitur LCH merupakan fitur terbaik untuk temu kembali dengan nilai *precision* rata-rata sebesar 94,45% dan *recall* 16,20% sedangkan fitur energi merupakan fitur yang paling tidak baik dengan nilai *precision* rata-rata 69,73% dan *recall* 11,95%.

Selanjutnya untuk menguji fitur terbaik yang dapat digunakan pada sistem, dilakukan pengujian dengan mengkombinasikan semua fitur.

5. Fitur gabungan energi, standar deviasi, GCH dan LCH.

Hasil pengujian gabungan semua fitur disajikan pada Gambar 11



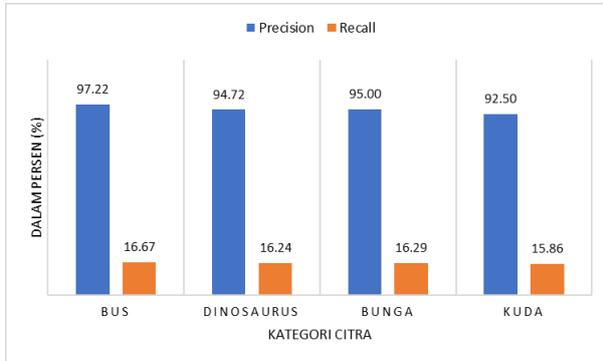
Gambar 11 Hasil pengujian gabungan semua fitur

Pengujian dengan menggabungkan semua fitur energi, standar deviasi, GCH dan LCH menghasilkan nilai *precision* dan *recall* citra dengan kategori dinosaurus memiliki nilai *precision* tertinggi yaitu 100% diikuti dengan bunga 97,22%, bus 95,83% dan citra

kuda 90,28%. Nilai rata-rata *precision* dan *recall* dari fitur gabungan ini adalah 95,83% dan 16,43%.

6. Fitur GCH, LCH dan standar deviasi

Hasil pengujian fitur GCH, LCH dan standar deviasi disajikan pada Gambar 12.

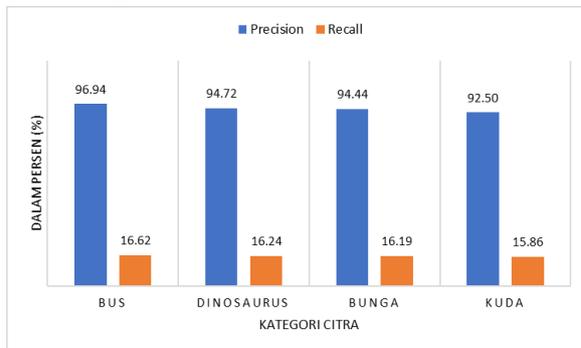


Gambar 12 Hasil pengujian fitur standar deviasi, GCH dan LCH

Pada Gambar 12 ditunjukkan nilai *precision* tertinggi pada citra bus sebesar 97,22%, citra bunga 95%, citra dinosaurus 94,72% dan citra kuda 92,50%.

7. Fitur GCH dan LCH

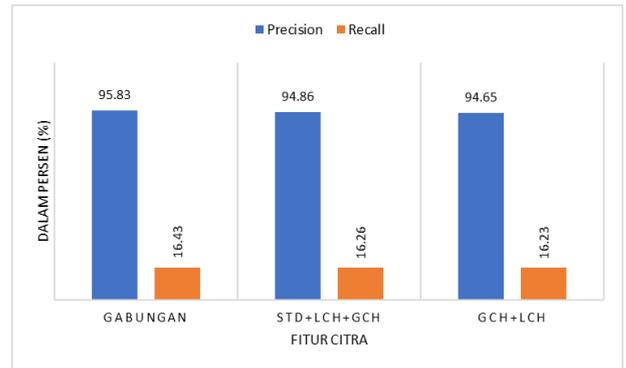
Hasil pengujian fitur GCH dan LCH pada Gambar 13



Gambar 13 Hasil pengujian fitur GCH dan LCH

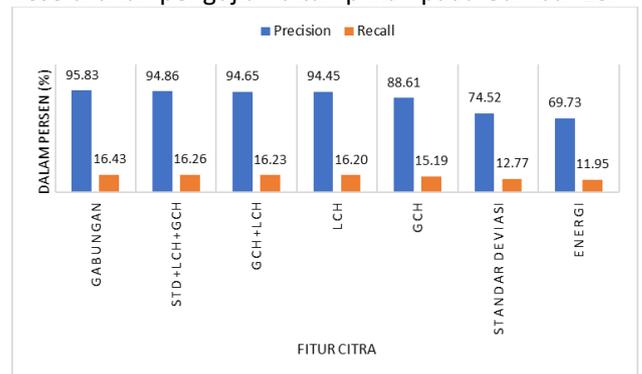
Pada Gambar 12 ditunjukkan nilai *precision* dari gabungan dari fitur GCH dan LCH. Nilai *precision* paling tinggi didapatkan pada pengujian citra dengan kategori bus dengan nilai 96,94%. Nilai rata-rata *precision* gabungan fitur GCH dan LCH ini adalah 94,65% dan nilai *recall*-nya 16,23%.

Hasil pengujian kombinasi fitur-fitur pada sistem disajikan pada Gambar 14.



Gambar 14 Hasil pengujian kombinasi fitur

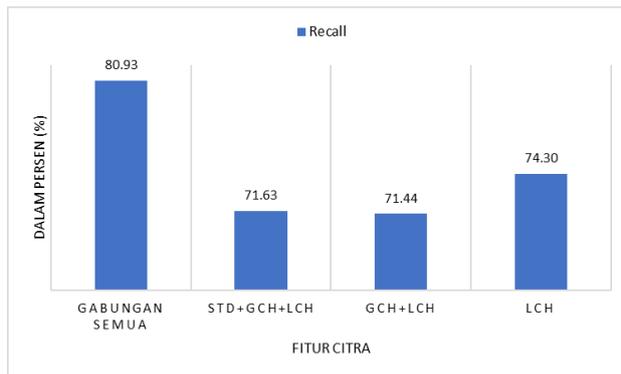
Pada pengujian kombinasi fitur didapatkan nilai rata-rata *precision* dan *recall* pada gabungan semua fitur adalah 95,83% dan 16,43%, gabungan fitur standar deviasi GCH dan LCH adalah 94,86% dengan nilai *recall* 16,26% dan kombinasi fitur GCH dan LCH sebesar 94,65% dengan nilai *recall* 16,23%. Hasil keseluruhan pengujian ditampilkan pada Gambar 15.



Gambar 15 Nilai precision dan recall hasil keseluruhan pengujian fitur.

Pada Gambar 15 dapat dilihat fitur terbaik dalam pengujian sistem ini adalah gabungan semua fitur dengan nilai *precision* 95,83% dan *recall* 16,43% sedangkan fitur terburuk adalah fitur energi dengan nilai *precision* 69,73% dan *recall* 11,95%.

Karena citra yang ditemukan (retrieved) hanya 12 nilai recall maksimal yang dihasilkan pada sistem hanya 17,14%, karena itu dilakukan pengujian terhadap fitur citra dengan nilai precision lebih dari 90% (Gabungan semua fitur, gabungan standar deviasi gch dan lch, gabungan GCH dan LCH, dan fitur LCH) dengan meningkatkan jumlah citra yang ditampilkan (retrieved) hingga 70 citra, sehingga didapatkan nilai recall yang ditunjukkan pada Gambar 16.



Gambar 16 Nilai recall rata-rata 4 fitur terbaik pada sistem

Pada Gambar 16 diketahui nilai recall gabungan semua fitur adalah 80,93%, nilai recall kombinasi fitur standar deviasi, GCH dan LCH sebesar 71,63%, nilai recall kombinasi GCH dan LCH 71,44% dan nilai recall LCH sebesar 74,30%. Nilai recall sangat dipengaruhi oleh jumlah citra yang ditemukan(ditampilkan). Semakin banyak citra relevan pada citra yang ditemukan, semakin besar recallnya[11].

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian yang sudah didapatkan, dapat disimpulkan bahwa:

- 1) Fitur terbaik dalam pengujian sistem ini adalah gabungan semua fitur dengan nilai *precision* 95,83% dan *recall* 16,43% sedangkan fitur terburuk adalah fitur energi dengan nilai *precision* 69,73% dan *recall* 11,95%.
- 2) Pengujian menggunakan fitur energi sangat dipengaruhi oleh jenis citra yang digunakan oleh sistem, karena pada pengujian fitur energi didapatkan nilai *precision* 100% pada citra dengan kategori dinosaurus namun hanya mendapat nilai *precision* 44,17% pada citra dengan kategori kuda.
- 3) Walaupun performa fitur energi dan standar deviasi menghasilkan nilai *precision* yang tidak cukup baik yaitu 69,73% dan 74,52% namun dapat dikombinasikan dengan fitur LCH dan GCH. Hal ini ditunjukkan pada gabungan semua fitur nilai rata-rata *precision* yang dihasilkan sebesar 95,83%
- 4) Ekstraksi fitur ciri warna (GCH dan LCH) dapat digabungkan dengan fitur dengan ciri tekstur (energi dan standar deviasi) dengan performa yang baik karena menghasilkan nilai *precision* lebih dari 90%.

5.2. Saran

Dari hasil penelitian yang dilakukan didapatkan catatan untuk dapat memperbaiki performa sistem agar menjadi lebih baik pada penelitian selanjutnya, antara lain:

- 1) Pemilihan citra sebagai bahan penelitian sangat berpengaruh terhadap hasil penelitian, untuk itu sebaiknya pemilihan citra disesuaikan dengan klasifikasi fitur yang akan diuji.
- 2) Untuk mendapatkan hasil yang lebih baik sebaiknya kemiripan citra menggunakan berbagai metode pengukuran jarak yang berbeda-beda.

DAFTAR PUSTAKA

- [1] A. S. Mukti, "Sistem Temu Kembali Citra Berbasis Warna Menggunakan Transformasi Wavelet Haar dan Histogram Warna," J. Informatics Technol., vol. 2, no. 3, 2013.
- [2] Y. T. Mshari and H. A. Younis, "Content Based Image Retrieval using Haar Wavelet to Extracted Color Histogram and Texture Features," vol. 4, no. 8, pp. 322–329, 2015.
- [3] S. Iqbal, I. H. Sarker, and M. I. Khan, "Content-Based Image Retrieval using Haar Wavelet Transform," no. November 2013, 2016.
- [4] A. Sumarna, "CBIR Berdasarkan Ekstraksi Fitur Warna Menggunakan Java," J. Exp. Psychol. Gen., vol. 136, no. 1, pp. 23–42, 2007.
- [5] J. Yue, Z. Li, L. Liu, and Z. Fu, "Content-based image retrieval using color and texture fused features," Math. Comput. Model., vol. 54, no. 3–4, pp. 1121–1127, 2011.
- [6] "The COREL Database for Content based Image Retrieval." [Online]. Available: <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>. [Accessed: 05-May-2019].
- [7] R. Munir, "Bab 1 Pengantar Pengolahan Citra," pp. 1–10.
- [8] M. Mustikasari, S. Madenda, E. Prasetyo, D. Kerami, and S. Harmanto, "Content Based Image Retrieval Using Local Color Histogram," no. August, 2014.
- [9] L. A. Sunjoyo, R. G. Santosa, and K. A. Nugraha, "Implementasi transformasi," vol. 12, no. 2, pp. 165–173, 2016.
- [10] N. S. S, P. R. S, and S. Singh, "Efficient CBIR Using Color Histogram Processing," vol. 2, no. 1, pp. 94–112, 2011.
- [11] L. R. Purwandi, "Sistem Temu Kembali Citra Berbasis Isi Menggunakan Metode Micro-Structure Descriptor (MSD) dan Grey Level Co-Occurrence Matrix (GLCM)," 2017.