

Implementasi Algoritma Genetika pada Temu Kembali Citra

Yeni Herdiyeni, Agus Buono, Gibtha Fitri Laxmi

Departemen Ilmu Komputer, Fakultas Matematika dan IPA, Institut Pertanian Bogor
e-mail: yeniherdiyeni2006@yahoo.com

Abstract

This research proposed to implement content based image retrieval using genetic algorithm to increase information retrieval efficiency. Genetic algorithm is used to get the best of color at vector space models. The vector space models obtained from color extraction by fuzzy color histogram. Under genetic algorithm, a chromosome represents each vector of color. These chromosome feed into genetic operator process such as selection, crossover, and mutation until maximum generation and find optimize chromosome for image retrieval. The optimal chromosome showed that chromosome with high similarity to query are more relevant to the query.

Retrieval evaluation used average precision value for every recall value. Based on the experience the average precision using genetic algorithm is 66.65%. In addition, this result shows that genetic algorithm can be used in image retrieval.

Keywords: content based image retrieval, genetic algorithm, fuzzy color histogram, cosine coefficient.

1. Pendahuluan

Perkembangan teknologi saat ini menyebabkan jumlah citra yang tersedia di internet menjadi sangat banyak. Hal tersebut menyebabkan pengguna mesin pencari (*search engine*) mengalami kesulitan memukul balik citra untuk mendapatkan informasi citra dengan cepat dan relevan. Hal ini disebabkan oleh beberapa hal, salah satunya adalah metode pencarian yang dipakai oleh pengguna pada umumnya masih dilakukan secara tekstual. Metode tekstual menjadi kurang relevan digunakan saat deskripsi yang diberikan terhadap citra tidak sesuai dengan isi dari citra tersebut. Oleh karena itu, perlu dikembangkan metode pencarian dengan menggunakan ciri citra yang disebut *content based image retrieval* (CBIR).

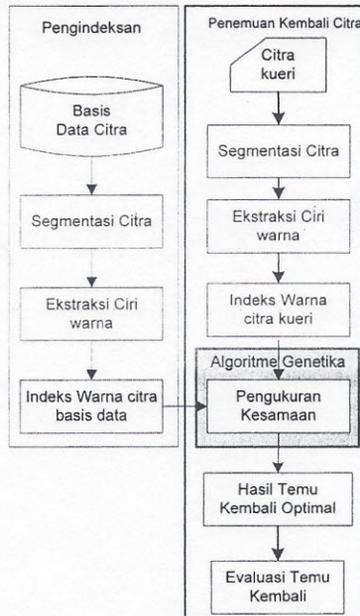
CBIR merupakan suatu pendekatan dalam temu kembali citra yang didasarkan pada informasi yang terkandung di dalam citra seperti warna, bentuk, dan tekstur dari citra (Han & Ma 2002). Warna lebih dominan dalam merepresentasikan karakteristik visual (*visual-content*) dari citra dibandingkan ciri citra lainnya, sehingga ciri ini paling banyak digunakan dalam temu kembali citra saat ini [14].

Teknik pengukuran kemiripan pada temu kembali citra menggunakan jarak *Euclidean* antara kueri dan basis data citra [2] [5] [11] menghasilkan hasil yang kurang optimal. Kurang optimalnya temu kembali tersebut dapat diatasi dengan menggunakan metode optimasi. Penelitian yang dilakukan Klabbankoh dan Pinngern (1999) menggunakan metode optimasi algoritma genetika untuk meningkatkan atau mengoptimalkan hasil temu kembali pada teks. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan temu kembali citra menggunakan algoritma genetika agar pemilihan

warna pada citra menjadi optimal dan temu kembali citra pun lebih relevan.

2. Metode Penelitian

Penelitian ini dilakukan dalam beberapa tahap, seperti yang terlihat pada Gambar 1.



Gambar 1. Metode penelitian.

Segmentasi Citra

Tahap *preprocessing* dalam penelitian ini adalah tahap segmentasi. Proses ini dilakukan secara otomatis berdasarkan kuantisasi yang berbeda.

Tahapan segmentasi yang dilakukan sebagai berikut [2]:

- 1 Mengkuantisasi warna piksel. Kuantisasi berguna untuk mengurangi jumlah warna pada citra. Variasi warna pada objek dapat diklasifikasikan menurut warna yang sama secara kenampakannya. Nilai kuantisasi segmentasi citra yang digunakan ialah warna hitam, merah, hijau, dan biru yang dapat dilihat pada Tabel 1.
- 2 Menghitung jumlah piksel tiap warna kuantisasi pada pusat citra ukuran 20×20 . Dua warna kuantisasi dengan jumlah piksel di atas 70 diambil, kecuali warna hitam. Seluruh piksel dengan warna kuantisasi tersebut menjadi objek citra yang menjadi perhatian utama.
- 3 Menjadikan citra biner. Daerah bunga pada citra diberi nilai 1, sedangkan daerah yang tidak diperhitungkan diberi nilai 0.

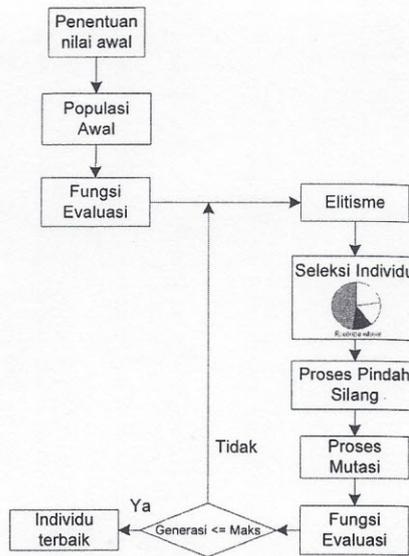
Tabel 1. Nilai warna kuantisasi pada proses segmentasi

Segmentasi			
	0	0	0
	255	0	0
	0	255	0
	0	0	255

Nilai warna kuantisasi pada segmentasi yang digunakan pada penelitian kali ini berdasarkan pengujian basis data citra.

Ekstraksi Ciri Warna

Ekstraksi warna dengan FCH tidak terlalu beragam hasilnya pada ruang warna seperti RGB, HSV dan *Lab* [14]. Penelitian ini menggunakan RGB untuk mempermudah pengolahan citra. Hasil segmentasi dari semua citra, warna yang sering muncul diambil yang terbaik tanpa ada warna yang sama. Pengambilan warna awal yang terlalu banyak memerlukan waktu komputasi yang besar untuk ekstraksi ciri citra, oleh karena itu *Fuzzy C-Means* digunakan untuk mengelompokkan warna-warna tersebut ke dalam sejumlah pusat *cluster* yang ukurannya lebih kecil.



Gambar 2. Proses algoritma genetika.

Algoritma Genetika

Algoritma genetika (GA) merupakan suatu metode pencarian yang didasarkan pada mekanisme dari seleksi dan genetika natural [7]. Proses tersebut dapat dilihat pada Gambar 2.

GA digunakan untuk menemukan solusi dalam masalah yang kompleks melalui kumpulan-kumpulan metode atau teknik seperti fungsi evaluasi (*fitness function*), pindah silang, mutasi, dan seleksi alam [12].

GA dikarakterisasi dengan 5 komponen dasar yaitu:

1. Representasikan kromosom untuk memudahkan penemuan solusi dalam masalah pengoptimasian.
2. Inisialisasi populasi.
3. *Fitness function* yang mengevaluasi setiap solusi.
4. Proses genetik yang menghasilkan sebuah populasi baru dari populasi yang ada.
5. Parameter seperti ukuran populasi, probabilitas proses genetik, banyaknya generasi, dan lain-lain.

Temu Kembali Citra

Hasil populasi akhir dari GA adalah populasi yang setiap kromosomnya merepresentasikan sebuah citra yang memiliki tingkat kemiripan yang tinggi dengan kueri. Hasil citra tersebut ditemukembalikan dengan cara pengukuran tingkat kemiripan untuk mengetahui citra yang relevan dengan basis data. Pengukuran tingkat kemiripan ini menggunakan *cosine coefficient* dengan *threshold* sebesar 0.75.

Evaluasi Temu Kembali Citra

Tahap evaluasi temu kembali dilakukan untuk menilai tingkat keberhasilan dalam proses temu kembali citra terhadap sejumlah koleksi pengujian. Evaluasi tersebut dilakukan dengan menghitung nilai *recall* dan *precision* dari proses temu kembali citra berdasarkan penilaian relevansinya (gugus jawaban). *Recall* adalah perbandingan jumlah citra relevan yang terambil terhadap jumlah citra dalam basis data. *Precision* adalah perbandingan jumlah citra yang relevan terhadap jumlah citra yang ditemukan kembali.

3. Hasil Percobaan

Penelitian ini menggunakan algoritma genetika (GA) dalam mencari kemiripan antara kueri dan basis data FCH yang ada. Kemiripan tersebut merepresentasikan nilai kedekatan ciri citra kueri dan ciri citra basis data. Kedekatan jarak ini merepresentasikan relevansi citra terhadap kueri yang diberikan.

Segmentasi Citra

Tahap segmentasi menggunakan kuantisasi warna hitam, merah, hijau, dan biru. Parameter kuantisasi yang berbeda menghasilkan wilayah objek utama yang berbeda juga. Tahap segmentasi pada Tabel 2 ini bertujuan untuk mendapatkan objek utama citra yang paling baik dengan cara yang sederhana [2].

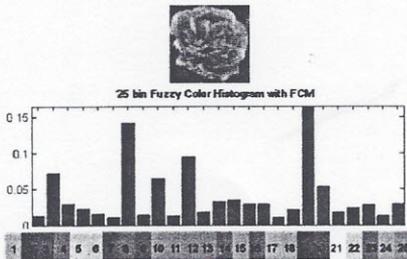
Ekstraksi Ciri Warna

Ekstraksi ciri warna pada citra di-lakukan dengan cara merepresentasikan peluang atau frekuensi piksel setiap citra ke dalam nilai warna (*bin*). *Bin* tersebut diperoleh dengan menggunakan FCH dengan FCM. Nilai warna yang digunakan 25 *bin* [2] yang dilakukan secara manual. Contoh Citra dan nilai warna citra dapat dilihat pada Gambar 3.

Tabel 2. Tahapan segmentasi

	Citra segmentasi	Citra hitam putih	Objek Utama
Seg			

Citra

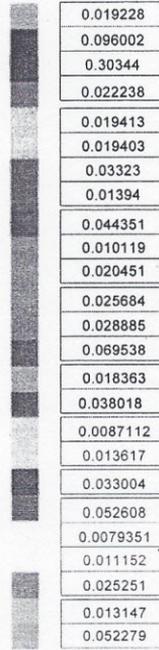


Gambar 3 Citra dan nilai warna. Algoritma Genetika

Tahapan algoritma genetika dalam penelitian ini adalah sebagai berikut:

- Representasi Kromosom dan Populasi Awal

Representasi kromosom merupakan cara untuk mengkodekan suatu alternatif solusi itu menjadi kromosom yang akan diproses menggunakan GA. Representasi kromosom dalam penelitian ini digunakan dalam berbentuk ciri citra yang berupa histogram. Panjang gen dalam kromosom yang berisi nilai warna sebanyak 25 *bin* merupakan nilai warna yang akan memberikan solusi pemecahan dalam masalah temu kembali citra berdasarkan nilai warna [2]. Ilustrasi sebuah kromosom dalam penelitian ini dapat dilihat pada Gambar 4. Ukuran kromosom pada populasi yang digunakan berdasarkan *threshold* yang telah ditetapkan sebelumnya yaitu sebesar 0.4. *Threshold* tersebut menjelaskan nilai jarak antara citra kueri dan basis data yang dimiliki lebih besar dari 0.4.



Gambar 4. Ilustrasi sebuah kromosom dalam algoritma genetika.

- Fungsi Evaluasi

Pada penelitian ini, permasalahan optimasi yang ingin dicapai adalah mendapatkan hasil temu kembali yang optimal yaitu *bin-bin* warna yang ditemukan kembali mirip dengan kueri. Oleh karena itu fungsi evaluasi yang digunakan pada penelitian kali ini adalah sebagai berikut:

$$Eval(V_k) = \frac{\sum_{i=1}^k x_i y_i}{\sqrt{\sum_{i=1}^k x_i^2 \sum_{i=1}^k y_i^2}} \quad (1)$$

Dengan V_k adalah nilai evaluasi, x_i adalah vektor *bin-bin* warna yang dimiliki kueri, dan y_i adalah vektor *bin-bin* warna yang dimiliki kromosom.

Hasil dari fungsi evaluasi tersebut berada dalam interval 0 dan 1. Nilai evaluasi mendekati 1 mengartikan bahwa dokumen semakin relevan dengan kueri. Sebaliknya, nilai evaluasi mendekati 0 mengartikan bahwa semakin tidak relevan antara kueri dan dokumen [9].

Elitisme

Nilai evaluasi yang dimiliki tiap individu atau kromosom akan diurutkan untuk mengetahui kromosom yang memiliki nilai terbaik. Kromosom dengan nilai terbaik tersebut akan disalin atau disimpan agar tidak rusak akibat proses genetik [13]. Jumlah kromosom yang digunakan pada penelitian ini sebanyak 2 nilai terbaik.

• Seleksi Individu

Seleksi adalah proses memilih individu pada populasi yang memiliki nilai evaluasi baik untuk dilanjutkan ke proses pindah silang dan mutasi [3]. Proses seleksi yang digunakan adalah *roulette wheel*. Seleksi dilakukan dengan cara mengambil nilai acak antara nilai minimum dan maksimum evaluasi tiap generasi. Jika nilai tersebut lebih kecil dari nilai probabilitas kumulatif maka kromosom yang ditunjuk akan dipilih sebagai kromosom induk.

Tahap awal dari *roulette wheel* adalah dengan menghitung probabilitas seleksi dengan rumus:

$$P_k = \frac{Eval(V_k)}{PopSize}, \quad k = 1, 2, \dots, PopSize$$

$$F = \sum_{i=1}^k P_i \quad (2)$$

Probabilitas seleksi kumulatif (q_k) untuk setiap kromosom V_k :

$$q_k = \sum_{i=1}^k P_i \quad (3)$$

Algoritma seleksi sebagai berikut:

Langkah 1: Bangkitkan bilangan acak r antara $[0,1]$

Langkah 2: Jika $r \leq q_1$, pilih kromosom V_1 , kalau tidak pilih kromosom k dengan ketentuan:

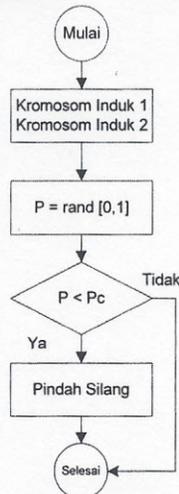
$$V_k \quad (2 \leq k \leq Pop.Size) \text{ dan } q_{k-1} \leq r \leq q_k$$

Kromosom yang memiliki nilai evaluasi yang besar atau mendekati 1 akan memiliki kemungkinan terpilih yang lebih besar sebagai populasi baru untuk proses

genetik selanjutnya. Hal tersebut menyebabkan kromosom akan terpilih lebih dari satu kali.

• Pindah Silang

Pindah silang merupakan proses paling penting dalam GA pada proses genetik. Proses ini melakukan pindah silang antar kromosom induk yang telah dipilih sebelumnya. Pemilihan antarkromosom induk yang akan dipindahsilang ialah dengan cara mengambil nilai acak yang bernilai lebih kecil dari peluang pindah silang P_c . Peluang pindah silang yang digunakan sebesar 0.85. Proses pindah silang dapat dilihat pada Gambar 5.



Gambar 5. Proses pindah silang.

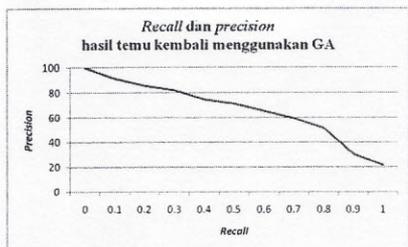
Kromosom yang akan dipindah-silangkan menggunakan *one point crossover*. Gen-gen yang dimiliki kromosom induk setelah titik potong akan dipindahsilangkan antar-induk sehingga tercipta dua buah kromosom baru.

• Mutasi

Proses mutasi yang digunakan adalah sistem gen dengan nilai gen mutasi yaitu bilangan bulat. Gen akan mengalami perubahan yang berguna untuk mengembalikan kerusakan gen akibat genetik lainnya [1]. Proses mutasi merupakan proses pengubahan nilai gen pada kromosom yang telah dipilih sebelumnya. Pemilihan gen ialah dengan cara mengambil nilai acak yang menyatakan posisi gen-gen yang akan dimutasi. Setelah diketahuinya posisi gen yang akan dimutasi maka nilai kuantisasi warna tersebut akan diubah dengan ketentuan sebagai berikut:

```

if (allele_kueri > allele_kromosom)
  allele_kromosom += 0.01;
else
  allele_kromosom -= 0.01;
end
  
```

Gambar 8. Grafik rata-rata *precision* hasil temu kembali menggunakan GA.

Berdasarkan Tabel 5 dapat dijelaskan bahwa nilai *precision* algoritma genetika memiliki rata-rata yang cukup baik untuk seluruh tingkat *recall*. Grafik pada Gambar 8 memperlihatkan rata-rata *precision* yang cukup stabil, yaitu dengan rata-rata sebesar 66.65%. Nilai *precision* cukup baik pada tingkat *recall* antara 0.1–0.8 karena pada *recall* 0.9 ke atas nilai rata-rata *precision* turun secara signifikan.

4. Kesimpulan

Hasil penelitian menunjukkan bahwa temu kembali menggunakan algoritma genetika memberikan hasil temu kembali citra yang baik terutama pada nilai *recall* 0.1–0.8, dengan rata-rata *precision* sebesar 72.64% untuk tingkat tersebut. Nilai rata-rata *precision* untuk temu kembali citra menggunakan algoritma genetika secara keseluruhan yaitu sebesar 66.65%.

Daftar Pustaka:

- [1] Aly AA. 2007. *Applying Genetic Algorithm in Query Improvement Problem*. www.foibg.com/ijitk/ijitk-vol01/ijitk01-4-p02.pdf [3 Juli 2008].
- [2] Balqis DP. 2006. *Fuzzy Color Histogram untuk Temu Kembali Citra Bunga*. [skripsi]. Bogor: Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- [3] Cox E. 2005. *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration*. USA: Morgan Kaufman Publishers.
- [4] Gen M, Cheng R. 1997. *Genetics Algorithms and Engineering Design*. John Wiley & Sons, Inc. Canada.
- [5] Hadi S. 2004. *Pengembangan Model Generatif Pengenalan Wajah pada Latar Belakang, Pose dan Iluminasi yang Bervariasi* [disertasi]. Bandung: Program Pascasarjana, Institut Teknologi Bandung.
- [6] Han J, Ma K. 2006. *Data Mining: Concepts and Techniques*. USA: Morgan Kaufman Publishers.
- [7] Haupt RL, Haupt SE. 2004. *Practical Genetic Algorithms*. New Jersey: John Wiley.
- [8] Hermawanto D. 2003. *Algoritma Genetika dan Contoh Aplikasinya*. <http://dennyhermawanto.webhop.org> [17 Maret 2008].

- [9] Klabbankoh B, Pinnern O. 1999. *Applied Genetic Algorithms in Information Retrieval*. Bangkok : Faculty of Information Technology King Mongkut, Institute of Technology Ladkrabang. <http://www.journal.au.edu/ijcim/sep99/02-drouen.pdf> [17 Juli 2007].
- [10] Mitchell M. 1998. *An Introduction to Genetic Algorithms*. London : Massachusetts Institute of Technology.
- [11] Noorniawati VY. 2007. *Metode Support Vector Machine untuk Klasifikasi pada Sistem Temu Kembali Citra*. [Skripsi]. Bogor: Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- [12] Owais SSJ, Kromer P, Snasel V. 2005. *Query Optimization by Genetic Algorithm*. ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-129/paper16.pdf [3 Juli 2008].
- [13] Suyanto. 2005. *Algoritma Genetika dalam MATLAB*. Penerbit ANDI. Yogyakarta
- [14] Vertan C dan Boujemaa N. 2000. *Using Fuzzy Color Histogram and Distance for Color Image Retrieval*. Prentice Hall. United Kingdom. http://www.rocq_inria.fr/media/Articles/cir2000.pdf [28 Oktober 2006].