

Sistem Temu-Balik Citra Menggunakan Jarak Histogram Dalam Model Warna YIQ

Arif Rahman

Program Studi Sistem Informasi, Universitas Ahmad Dahlan
Jl. Prof. Soepomo, Yogyakarta 55164, Telp. (0274) 381523, 379418
E-mail: arif@uad.ac.id

ABSTRAKS

Komposisi warna merupakan salah satu ciri dari citra yang dapat digunakan sebagai pembeda dalam sistem temu-balik citra. Komposisi warna piksel dalam suatu citra dapat direpresentasikan dalam histogram warna. Tingkat kemiripan warna antar citra dapat ditentukan berdasarkan nilai jarak antar histogramnya. Semakin kecil nilai jarak antar histogram, semakin tinggi tingkat kemiripan suatu citra. Jarak antar histogram pada citra berwarna dihitung untuk setiap komponen warna. Penelitian ini menggunakan model warna YIQ. Tiap komponen warna YIQ nilainya dikuantisasi menjadi 128, 64 dan 64 level. Komponen warna Y memiliki jumlah level lebih banyak karena memiliki pengaruh lebih dominan pada persepsi mata manusia terhadap warna dibanding komponen yang lain. Jumlah piksel dalam citra berbeda-beda sesuai dengan ukuran citra, sehingga histogram perlu dinormalisasi agar invarian terhadap ukuran citra. Normalisasi dilakukan dengan membagi jumlah piksel tiap level dengan jumlah total piksel citra, sehingga didapatkan jangkauan nilai $[0,1]$ untuk tiap level. Total jarak antar histogram merupakan hasil jumlahan dari jarak tiap levelnya yang dihitung menggunakan rumus jarak Manhattan. Hasil temu-balik diranking berdasarkan nilai jarak antar histogramnya.

Kata Kunci: Temu-balik, citra, histogram, YIQ

1. PENDAHULUAN

Seiring dengan perkembangan teknologi informasi dan semakin luasnya pemanfaatan teknologi komputer di berbagai bidang kehidupan, kebutuhan akan efisiensi pengelolaan data yang meliputi data teks, suara, citra dan video dewasa ini semakin meningkat. Salah satu aspek dalam hal pengelolaan data adalah penemuan kembali informasi yang diinginkan pengguna atau yang disebut dengan istilah temu-balik informasi atau *Information Retrieval* (IR). Tujuan utama dari IR adalah menemukan kembali dokumen yang berisi informasi yang relevan dengan query yang diberikan oleh user.

Dalam penelitian ini pembahasan retrieval difokuskan pada salah satu jenis dokumen yaitu dokumen berbentuk citra (*image*). Relevansi isi (*content*) dokumen citra relatif lebih kompleks dibandingkan dengan dokumen teks karena dipengaruhi faktor persepsi seseorang terhadap suatu citra. Untuk menyederhanakan hal tersebut, dalam sistem temu-balik dokumen citra berdasarkan isi atau *Content-Based Image Retrieval* (CBIR) pada umumnya digunakan deskripsi citra yang bersifat obyektif yaitu berupa *low-level feature* meliputi komposisi warna, tekstur atau bentuk (*shape*) yang dimiliki citra (Rui, 1999).

Dokumen citra digital tersusun dari elemen-elemen yang disebut piksel. Ukuran citra pada umumnya dinyatakan dalam jumlah piksel, misalnya citra berukuran 640 x 480 piksel, berarti citra ini

memiliki ukuran panjang 640 piksel dan lebar 480 piksel. Penentuan fitur dari suatu citra diperlukan untuk efisiensi proses pencarian kemiripan citra, karena dokumen citra terdiri piksel-piksel yang berjumlah ratusan, ribuan bahkan jutaan.

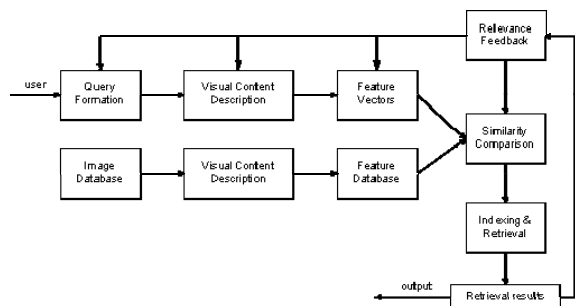
Persepsi manusia terhadap suatu citra cenderung didominasi oleh faktor komposisi warna yang dimiliki citra. Kita sering membedakan suatu citra dengan citra yang lain berdasarkan komposisi warna yang dimilikinya. Asumsi tersebut menjadi ide dasar pengembangan sistem Image Retrieval yang menggunakan komposisi warna sebagai fitur yang mewakili suatu citra. Histogram warna dari piksel-piksel dalam citra dapat digunakan sebagai representasi komposisi warna yang dimiliki citra.

2. LANDASAN TEORI

2.1 Sistem Temu-Balik Citra

Sistem Temu-Balik Citra (*Image Retrieval*) pada awal pengembangannya yaitu sekitar akhir 1970-an, masih menggunakan teks untuk menandai atau memberi keterangan (*annotation*) pada citra. Pertama-tama citra diberi keterangan berbentuk teks kemudian untuk melakukan proses temu-balik digunakan DBMS (*Database Management System*) berbasis teks. Pemberian keterangan tersebut memiliki kelemahan yaitu: jika koleksi citra memiliki jumlah yang sangat besar, maka menjadi tidak efisien karena proses dilakukan secara manual dan keterangan yang diberikan pada citra bersifat subyektif, sangat tergantung pada persepsi pemberi keterangan. Untuk mengatasi persoalan tersebut,

pada awal 1990-an mulai dikembangkan CBIR (*Content-Based Image Retrieval*) yang melakukan proses temu-balik berdasarkan muatan visual berupa komposisi warna yang dimiliki citra (Long, 2002), (Remco, 2002). Secara umum sistem CBIR dapat dinyatakan dalam diagram pada Gambar 1.



Gambar 1. Diagram *Content-Based Image Retrieval*

Muatan visual citra dalam basis data diekstrak, kemudian dideskripsikan sebagai vektor ciri (*feature vector*) dan disimpan dalam basis data ciri. Untuk mendapatkan kembali suatu citra, pengguna memberi masukan kepada sistem berupa contoh citra yang akan dicari, proses ini dinamakan QBE (*Query By Example*). Sistem kemudian mengubah contoh citra tersebut ke dalam bentuk vektor ciri dan membandingkan tingkat kemiripannya (*similarity comparison*) dengan vektor-vektor ciri dalam basis data ciri. Dalam proses perbandingan kemiripan digunakan indeks agar pengaksesan vektor ciri dalam basis data lebih efisien. Selanjutnya dilakukan proses temu-balik dan pengurutan citra berdasarkan nilai yang dihasilkan pada proses perbandingan tingkat kemiripan. Sistem temu-balik dewasa ini juga telah melibatkan umpan-balik dari user apakah suatu citra hasil retrieval relevan atau tidak (*relevance feedback*) yang digunakan sebagai acuan untuk memodifikasi proses temu-balik agar mendapatkan hasil yang lebih akurat (Long, 2002)

2.2 Histogram

Komposisi warna merupakan salah satu fitur yang dapat digunakan dalam sistem temu-balik citra. Komposisi warna dapat direpresentasikan dalam bentuk histogram. Histogram warna merepresentasikan distribusi jumlah piksel untuk tiap intensitas warna dalam citra. Untuk mendefinisikan histogram, warna di kuantisasi menjadi beberapa level diskrit, kemudian untuk tiap level tersebut dihitung jumlah piksel yang nilainya sesuai (Acharya, 2005) seperti ditunjukkan Gambar 2.

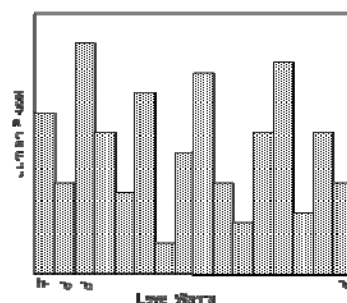
2.3 Jarak Histogram

Dengan histogram dapat dicari citra yang memiliki kemiripan komposisi warna. Pengukuran tingkat kemiripan dilakukan dengan menghitung jarak antar histogram. Jika $G = \{g_1, g_2, \dots, g_N\}$

dan $H = \{h_1, h_2, \dots, h_N\}$ adalah histogram warna dari dua buah citra, dimana g_i dan h_i adalah jumlah piksel pada level ke i dari kedua histogram dan N adalah jumlah level untuk tiap histogram, maka jarak (d) antara dua histogram dapat dinyatakan dalam jarak Manhattan seperti pada Persamaan (1).

$$d = \sum_{i=1}^N |g_i - h_i| \quad (1)$$

Citra dengan nilai jarak yang lebih kecil dianggap memiliki tingkat kemiripan komposisi warna yang lebih tinggi atau lebih mirip dibandingkan dengan citra yang memiliki nilai jarak yang lebih besar.



Gambar 2. Histogram warna

2.4 Model Warna YIQ

Model YIQ merupakan salah model warna yang berfokus pada persepsi mata manusia terhadap warna. Model ini merupakan standar warna pada penyiaran TV. YIQ merepresentasikan warna dalam tiga komponen, yaitu komponen Y mewakili pencahayaan (*luminance*), komponen I mewakili corak warna (*hue*) dan komponen Q mewakili intensitas atau kedalaman warna (*saturation*) (Gonzalez, 2002). Perangkat keras pengolah citra pada umumnya menerapkan model warna RGB dengan pertimbangan kemudahan pada teknis penampilan warna. Konversi warna diperlukan untuk menjembatani perbedaan kedua model warna tersebut agar dapat diproses dan ditampilkan dengan benar. Rumus konversi warna dari RGB ke YIQ ditunjukkan oleh Persamaan (2)

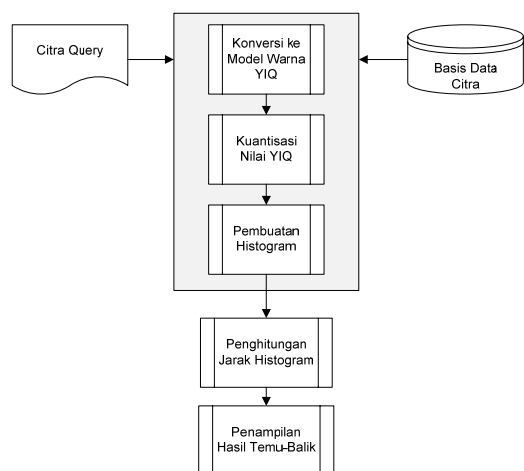
$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.144 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

3. PEMBAHASAN

3.1 Deskripsi Sistem

Dalam penelitian ini dibangun sistem CBIR untuk melakukan pencarian dokumen citra berwarna. Sistem ini menggunakan model QBE, pengguna memberi input berupa citra, kemudian sistem akan mencari citra-citra lain yang mirip dengan citra input. Kemiripan ditentukan berdasarkan nilai jarak antar histogram dalam model

warna YIQ. Hasil pencarian akan ditampilkan terurut berdasarkan nilai jarak histogramnya. Proses temu-balik citra digambarkan pada Gambar 3



Gambar 3. Proses Temu-Balik Citra

Citra *query* dan citra dalam basis data dikonversi dari model warna RGB ke model warna YIQ menggunakan persamaan (2). Masing-masing komponen YIQ memiliki jangkauan nilai 0-255. Penyederhanakan histogram dilakukan agar proses penghitungan histogram lebih cepat yaitu dengan kuantisasi nilai masing-masing komponen. Kemudian masing-masing komponen dibuat histogramnya. Histogram yang dibentuk adalah histogram ternormalisasi. Penghitungan jarak antara histogram citra *query* dengan citra dalam basis data dilakukan menggunakan persamaan (1). Nilai jarak yang didapat menentukan tingkat kemiripan antara citra *query* dengan citra dalam basis data. Hasil pencarian diurutkan berdasarkan nilai jarak histogram dari yang terkecil sampai yang terbesar.

3.2 Kuantisasi Nilai YIQ

Dalam penelitian ini tiap komponen warna YIQ nilainya dikuantisasi menjadi 128, 64 dan 64 level. Komponen warna Y memiliki jumlah level lebih banyak karena memiliki pengaruh lebih dominan pada persepsi mata manusia terhadap warna dibanding komponen yang lain (Gonzalez, 2002). Nilai hasil kuantisasi tiap komponen warna (q) dihitung menggunakan Persamaan (3)

$$q = \left\lceil \frac{v}{v_{\max} - v_{\min}} (q_{\max} - q_{\min}) \right\rceil \quad (3)$$

v adalah nilai yang akan dikuantisasi, v_{\max} , v_{\min} nilai maksimum dan minimum awal, q_{\max} , q_{\min} nilai maksimum dan minimum target. Penelitian ini menggunakan target jangkauan nilai [0,127] untuk Y, [0,63] untuk I dan Q. Jangkauan nilai awal adalah [0,255].

3.3 Normalisasi Histogram

Jumlah piksel citra sangat beragam, maka histogram perlu dinormalisasi agar nilai selisih jarak histogram invarian terhadap ukuran citra. Normalisasi dilakukan dengan cara membagi jumlah piksel untuk masing-masing level dengan jumlah total piksel dalam citra (N) sehingga didapatkan nilai minimum 0 dan maksimum 1 untuk tiap level warna. Normalisasi tiap level histogram (h_i) dinyatakan dalam Persamaan (4)

$$\bar{h}_i = \frac{h_i}{N} \quad (4)$$

3.4 Pengukuran Jarak Histogram

Jarak antara histogram citra *query* dengan citra dalam koleksi dihitung untuk setiap komponen warna YIQ menggunakan persamaan (1). Jadi jarak kedua histogram merupakan total jumlah selisih antar level dalam komponen warna yang dapat dirumuskan pada persamaan (5)

$$D = d_y + d_i + d_q \quad (5)$$

D memiliki jangkauan nilai [0,3]. Karena untuk masing-masing komponen warna nilai jarak histogram ter-normalisasi memiliki jangkauan [0,1].

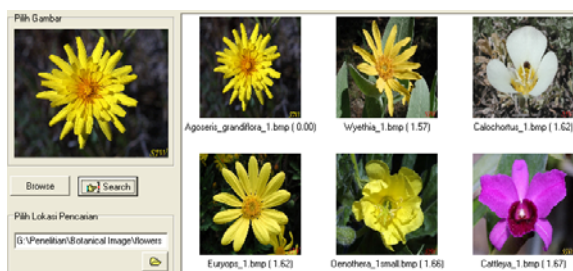
4. PERCOBAAN DAN HASIL

4.1 Percobaan

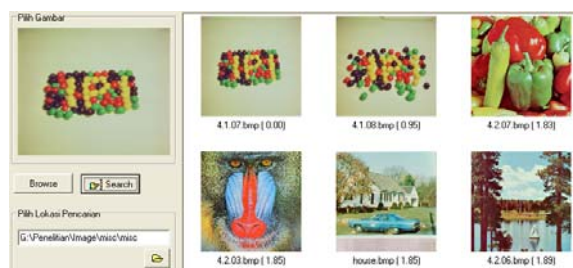
Percobaan dilakukan dengan membuat program komputer menggunakan bahasa pemrograman Delphi untuk melakukan proses temu-balik. Citra *query* dapat dipilih oleh pengguna, kemudian pengguna juga menentukan lokasi pencarian dalam komputer untuk dokumen citra yang akan dicari kemiripannya. Hasil pencarian akan ditampilkan dengan keterangan berupa nama file dan nilai jarak histogramnya dengan citra *query*. Untuk menguji program, dilakukan percobaan pencarian terhadap beberapa citra *query* dalam koleksi citra flowers dan misc.

4.2 Hasil

Hasil percobaan pencarian kemiripan pada koleksi citra menunjukkan bahwa citra yang memiliki kemiripan distribusi warna dengan citra *query* memiliki nilai jarak yang kecil dan akan muncul di awal atau menempati ranking tinggi. Semakin besar nilai jaraknya maka semakin tidak mirip distribusi warnanya. Citra yang sama persis akan menghasilkan nilai jarak 0. Penilaian kesesuaian antara nilai jarak dengan kemiripan citra bersifat subyektif, artinya tergantung pada persepsi seseorang terhadap citra. Hasil percobaan ini mencari kemiripan distribusi warna, bukan pada bentuk obyek atau tekstur. Contoh hasil pencarian dapat dilihat pada Gambar 4 dan Gambar 5



Gambar 4. Hasil temu-balik koleksi citra flowers



Gambar 5. Hasil temu-balik koleksi citra misc

5. KESIMPULAN DAN SARAN

Sistem temu-balik yang telah dibuat dapat menemukan kemiripan antar citra berdasarkan nilai jarak antar histogramnya. Ini dibuktikan dengan hasil percobaan pada program yang telah dibuat menunjukkan bahwa citra yang memiliki kemiripan distribusi warna masuk dalam ranking atas, dan citra yang sama persis, masuk diranking satu dan memiliki selisih jarak sama dengan nol.

Program yang telah dibuat belum menggunakan skema indeks dan struktur data yang mendukung, untuk pengembangan dan keberlanjutan penelitian, penulis menyarankan untuk membuat skema indeks dan struktu data yang sesuai sehingga mempercepat proses pencarian.

PUSTAKA

- Acharya T., Ray A.K. (2005), *Image Processing Principles and Applications*, by John Wiley & Sons,
- Gonzalez, Rafael C. (2002), *Digital Image Processing*, Addison-Wesley.
- Long F., Zhang H., Feng D. (2002), *Fundamentals of Content-based Image retrieval*, in *Multimedia Information Retrieval and Management - Technological Fundamentals and Applications*, D. Feng, W.C. Siu, and H.J.Zhang. (ed.), Springer.
- Remco C. V, Mirela T. (2002), *Content-Based Image Retrieval Systems: A Survey*, Department of Computing Science, Utrecht University.
- Rui Y., Huang T.S, Chang S. (1999), *Image Retrieval: Past, Present and Future*, *Journal of Visual Communication and Image Representation*, Vol. 10, pp. 1 - 23.