

PERBANDINGAN PENCARIAN CITRA BERDASARKAN TEKS DAN WARNA

¹⁾Devi Natalia, ²⁾Mumun Nurmilawati, dan ³⁾Daniel Swanjaya

^{1,3)}Jurusan Teknik Informatika Universitas Nusantara PGRI Kediri

²⁾FKIP Universitas Nusantara PGRI Kediri

¹⁾oknataliadevi@gmail.com, ²⁾mumunnurmila86@gmail.com, ³⁾swanjayadaniel@gmail.com

Abstrak. Metode pencarian citra dapat dilakukan dengan berbagai cara yaitu pencarian citra berdasarkan teks dan pencarian citra berdasarkan konten (warna, bentuk, tekstur). Akan tetapi untuk saat ini sistem pencarian citra berdasarkan teks lebih sering digunakan oleh user dalam proses pencarian citra daripada sistem pencarian citra berdasarkan warna. Penelitian ini digunakan untuk membandingkan pencarian citra berdasarkan teks dan berdasarkan konten warna. Metode pencarian citra berdasarkan teks yang digunakan adalah Latent Semantic Indexing. Dalam pencarian citra berdasarkan warna ini metode yang digunakan adalah histogram warna citra digital dengan menggunakan pengukuran jarak eucliden. Untuk mengukur performa sistem digunakan precision dan recall sebagai evaluasi performa.

Kata Kunci: histogram, jarak eucliden, Latent Semantic Indexing, precision, recall.

Perkembangan teknologi informasi telah banyak melahirkan sistem dan aplikasi yang sangat bermanfaat dan membantu pengguna dalam menyelesaikan pekerjaan. Salah satu contohnya, gambar atau citra telah memegang peranan penting dalam berbagai segi kehidupan manusia baik dalam bidang kedokteran, jurnalis, periklanan, disain, pendidikan, perdagangan, maupun hiburan.

Dulu gambar masih diabadikan dalam bentuk karya fisik seperti foto, lukisan, dll. Akan tetapi di era digital saat ini gambar juga telah berubah menjadi karya digital yang disebut sebagai citra digital. Citra digital ini juga bertambah banyak setiap waktu, sehingga sering kali sulit untuk mencari suatu gambar dari koleksi atau kumpulan gambar yang jumlahnya sangat besar sehingga para programmer membuat sebuah aplikasi pencarian citra untuk mempermudah pencarian tersebut. Teknik pencarian citra saat ini dapat dilakukan dengan berbagai cara. Seperti pencarian citra digital dengan menggunakan berbasis teks (*Text Based*) maupun basis konten (*Content Based*).

Dalam pencarian citra basis teks, citra akan diindeks dan dicari berdasarkan deskripsi-deskripsi seperti ukuran, tipe, tanggal pembuatan, identitas pemilik, kata kunci atau deskripsi lainnya yang bersangkutan dengan citra tersebut. Deskripsi pencarian citra basis teks biasanya diinputkan secara manual oleh operator atau admin untuk setiap masing-masing citra karena pembangkitan secara

otomatis sulit dilakukan tanpa informasi visual dan ekstraksi fitur.

Pencarian citra berdasarkan konten dapat dilakukan dengan berbagai cara, salah satunya berdasarkan konten warna atau yang sering disebut *Color Based Images Rertieval*. Dalam pencarian citra berdasarkan warna ini metode yang digunakan adalah histogram warna citra digital. Pencarian citra dengan menggunakan metode ini yang dilihat bukan kemiripan dari citra melainkan kemiripan dari sebaran warna pada citra yang akan dibandingkan.

Untuk saat ini sistem pencarian citra berdasarkan teks lebih sering digunakan oleh user dalam proses pencarian citra daripada sistem pencarian citra berdasarkan warna. Dan pencarian citra berdasarkan warna kurang dikenal oleh masyarakat. Padahal banyak sekali jurnal yang mengangkat tentang pencarian citra berdasarkan warna.

Penelitian ini bertujuan untuk : (1) mengetahui perbedaan antara sistem pencarian citra berdasarkan teks (*text based*) dan sistem pencarian citra berdasarkan warna (*content based*). (2) menentukan mana yang lebih baik dalam menampilkan hasil pencarian citra, antara pencarian citra berdasarkan teks dan pencarian citra berdasarkan warna. (3) mengetahui kelebihan dan kekurangan sistem pencarian berdasarkan teks, kelebihan dan kekurangan sistem pencarian berdasarkan warna, serta perbandingan dari kedua metode tersebut.

Berdasarkan uraian diatas maka peneliti melakukan suatu penelitian untuk membandingkan dari kedua metode tersebut manakah yang dapat menampilkan hasil pencarian citra yang lebih baik atau yang lebih dikenal dengan nama penelitian komparatif. Penelitian ini akan membahas mengenai analisa pencarian citra berdasarkan teks (*Latent Semantic Indexing*) dengan pencarian citra berdasarkan konten warna dengan menggunakan histogram warna pada citra digital. Untuk mengukur performa sistem digunakan *precision* dan *recall* sebagai evaluasi performa.

I. Metodologi

Image Retrieval (pengambilan citra) adalah suatu sistem komputer untuk menelusuri (*browsing*), mencari (*searching*), dan pengambilan citra (*image retrieval*) dari suatu database yang besar pada citra-citra digital (*digital images*). Sebagian besar cara lama dan umum metode pada *image retrieval* memanfaatkan metode penambahan metadata seperti *captioning*, *keyword*, atau deskripsi pada citra itu sendiri sehingga dapat disajikan dengan penjelasan kata-kata.

Pencarian Citra Berdasarkan Teks

Dalam pencarian citra basis teks, citra akan diindeks dan dicari berdasarkan deskripsi-deskripsi seperti ukuran, tipe, tanggal pembuatan, identitas pemilik, kata kunci atau deskripsi lainnya mengenai citra tersebut. Deskripsi pencarian citra basis teks biasanya diinputkan secara manual untuk masing-masing citra oleh operator karena pembangkitan secara otomatis sulit dilakukan tanpa informasi visual dan ekstrasi fitur.

Latent Semantic Indexing

Latent Semantic Indexing (LSI) adalah metode pengindeksan dan pencarian yang menggunakan teknik matematika yang disebut *Dekomposisi Nilai Singular* (SVD) untuk mengidentifikasi pola hubungan antara istilah dan konsep-konsep yang terkandung dalam sebuah koleksi teks yang tidak terstruktur [1]. LSI dalam *Context Based Information Retrieval* dapat berdasarkan penjelasan-penjelasan yang dimana ditambahkan untuk menyingkapkan citra (kata kunci (*keyword*), keterangan), atau pada teks tambahan yang sengaja tersedia dengan sebuah citra (judul halaman, judul

tambahan, dekat teks). Dari teks tersebut, pengindeksan dapat diciptakan menggunakan teknik standart pengambilan teks. Kesamaan antara teks dimana pada gilirannya sering didasarkan pada kesamaan penggunaan kata. Masalah penting dengan pendekatan *Context Based Information Retrieval* ini adalah perbedaan pada penggunaan kata diantara dokumen-dokumen, yaitu:

- Dokumen dapat membahas topik yang sama dengan menggunakan kata-kata berbeda (sinonim)
- Menggunakan kata-kata yang sama menggambarkan konsep yang berbeda.

Singular Value Decomposition

Singular Value Decomposition (SVD) adalah teknik reduksi dimensi dengan mengantikan matrik tern-dokumen A dengan matrik baru A_k . Tujuan SVD adalah untuk menghitung persamaan dekomposisi : [3]

$$A = USV^T$$

Dimana :

S : dengan nilai singular.

U : vector singular kiri.

V^T atau V^1 : vector singular kanan.

Untuk mengukur kemiripan teks query dengan teks database digunakan *Cosine Similarity*.

$$\text{sim}(Q,D) = \frac{Q \cdot D}{\|Q\| \|D\|}$$

Q = Query teks

D = teks dalam database

Pencarian Citra Berdasarkan Warna

Dalam pencarian citra berdasarkan warna ini metode yang digunakan adalah histogram warna citra digital. Pencarian citra dengan menggunakan metode ini yang dilihat bukan kemiripan dari citra melainkan kemiripan dari sabaran warna pada citra yang akan dibandingkan.

Content Based Image Retrieval

CBIR (*Content Based Image Retrieval*), atau disebut juga *Query By Image Content* (QBIC) merupakan suatu teknik pencarian kembali gambar yang mempunyai kemiripan karakteristik atau content dari sekumpulan gambar. Proses umum dari CBIR adalah gambar yang menjadi *query* dilakukan proses ekstrasi fitur, begitu halnya dengan gambar

yang ada pada sekumpulan gambar juga dilakukan proses seperti pada gambar *query* [4]. *Content based* berarti pencarian akan menganalisis isi sebenarnya dari citra. Menganalisa dalam konteks ini dapat berarti berupa warna, bentuk, tekstur, atau informasi lain yang terdapat dalam citra tersebut. Tanpa kemampuan dalam memeriksa konten citra, pencarian harus tergantung pada metadata seperti judul dan kata kunci yang susah payah atau mahal untuk diproduksi.

Fitur Warna

Model warna (*color model*) adalah sebuah cara untuk mempresentasikan warna yang diindera manusia dalam komputasi. Berikut penjelasan singkat tentang berbagai macam model atau format warna :

- Format warna RGB
- Format warna HSV atau HSI atau HSL
- Format warna CIE
- Format warna YcrCb
- Format warna CMYK

Histogram

Histogram adalah grafik yang menunjukkan frekuensi kemunculan setiap nilai gradiasi warna. Histogram warna mempresentasikan distribusi jumlah piksel untuk tiap intensitas warna dalam citra [5].

Histogram yang normal memiliki rentan dari 0...1 untuk setiap frekuensi derajat intensitas, sehingga secara matematis dapat dihitung sebagai berikut :

$$h_i = \frac{n_i}{n} \quad i = 0,1,\dots,L-1$$

n_i = jumlah piksel yang memiliki derajat keabuan i

n = jumlah seluruh piksel di dalam citra

Color Quantization

Dalam pembuatan histogram, nilai RGB yang punya range dari 0 – 255 akan mempunyai kemungkinan kombinasi warna 16777261 (didapat dari : 255 x 255 x 255). Pada proses komputasi, tentu saja ini proses yang menghabiskan banyak waktu (*time consuming*).

Masalah tersebut dapat diatasi dengan menggunakan *Color quantization* (kuantisasi warna), yaitu suatu prosedur untuk mengurangi kemungkinan jumlah warna. Dengan cara ini, jumlah warna yang besar bisa dikurangi,

sehingga proses yang dibutuhkan dalam pencarian citra akan semakin mudah [6].

Pengukuran Jarak Histogram

Fitur warna merupakan fitur yang paling banyak digunakan pada sistem pencarian citra berbasis konten. Banyak diantaranya menggunakan *image color histogram*. Untuk mengetahui kesamaan antara dua citra perlu dihitung jarak antara histogram warnanya. Citra yang memiliki jarak terkecil, merupakan solusinya.

Pengukuran tingkat kemiripan dilakukan dengan menghitung jarak antar histogram menggunakan rumus jarak *Euclidean* [6]:

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

Dimana H^A , H^B adalah histogram citra A dan citra B.

Precision Dan Recall

Terdapat dua ukuran evaluasi yang digunakan untuk mengevaluasi efektifitas dari sistem mesin pencarian gambar. Pengukuran pertama adalah *recall* dan yang kedua adalah *precision*. *Recall* merupakan ukuran dari kemampuan sebuah sistem untuk menampilkan seluruh gambar yang relevan.

Precision merupakan ukuran dari kemampuan sebuah sistem untuk menampilkan hasil retrieve image yang relevan dari keseluruhan image yang ter-retrieve.

Hasil temu kembali mempresentasikan hasil yang relevan jika hasil temu kembali tersebut termasuk dalam kategori gambar yang sama dengan gambar query [2].

$$\text{Recall} = \frac{\sum \text{gambar relevan yang didapat}}{\sum \text{gambar relevan pada kumpulan gambar}}$$

$$\text{Precision} = \frac{\sum \text{gambar relevan yang didapat}}{\sum \text{gambar yang didapat}}$$

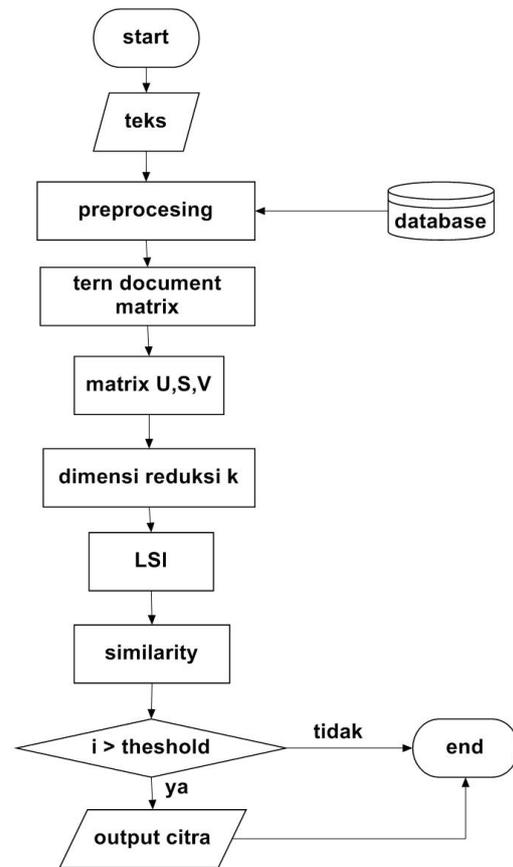
Contoh Basis Data Gambar

Contoh basis data gambar yang digunakan dalam penelitian ini adalah sebagai berikut :

Tabel 1. Contoh basis data gambar

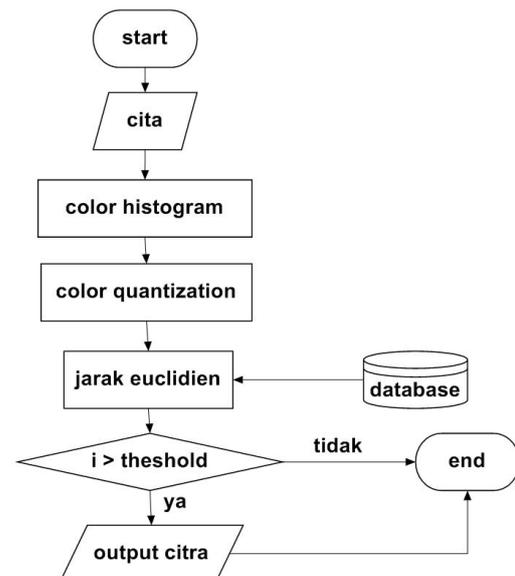
Beach	 father and son playing sand beach	 people sunbathing on the beach	 people playing on the beach
Flower	 Pink rose in full bloom	 Red roses with a few flower buds	 Three pink rosebud
Horse	 White horse running in grassland	 Mother and son horse grazing in a field	 Brown horse ran with his son
Mountain	 Flower garden snowy mountain area	 someone playing sky in the snowy mountains	 View of the mountain with forests

Alur Pencarian Citra Berdasarkan Teks



Gambar 1. Flowchart Program Pencarian Citra Berdasarkan Teks

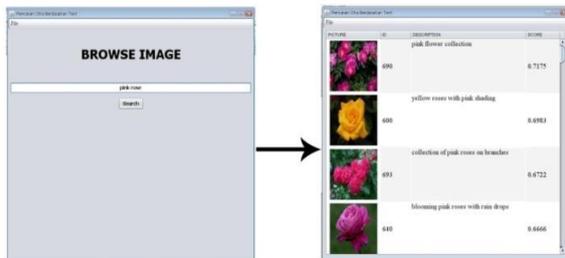
Alur Pencarian Citra Berdasarkan Warna



Gambar 2. Flowchart Program Pencarian Citra Berdasarkan Warna

II. Hasil dan Pembahasan

Berikut ini merupakan gambar dari hasil input dan output sistem pencarian citra berdasarkan warna maupun sistem pencarian citra berdasarkan warna.



Gambar 3. Hasil Input Dan Output Pencarian Citra Berdasarkan Teks



Gambar 4. Hasil Input Dan Output Pencarian Citra Berdasarkan Warna

Evaluasi performa

Untuk menguji ketepatan dan keakuratan aplikasi ini dilakukan percobaan dengan menghitung nilai *precision* dan nilai *recall* yang akan dilakukan terhadap 5 query acak setiap kategori citra (*beach*, *flower*, *horse*, dan *mountain*) yang terdiri dari masing-masing kategori sebanyak 50 gambar.

Untuk mengetahui performa sistem pencarian citra berdasarkan warna maupun teks, nilai *precision* dan *recall* dari setiap kategori dihitung nilai rata-ratanya. Sistem yang mempunyai nilai *precision* dan *recall* lebih tinggi berarti sistem tersebut mempunyai performa lebih baik daripada sistem yang lainnya. Dibawah ini merupakan perhitungan *precision* dan *recall* dari sistem pencarian citra berdasarkan teks maupun sistem pencarian citra berdasarkan warna. Hasil evaluasi dapat dilihat pada tabel 2 dan 3.

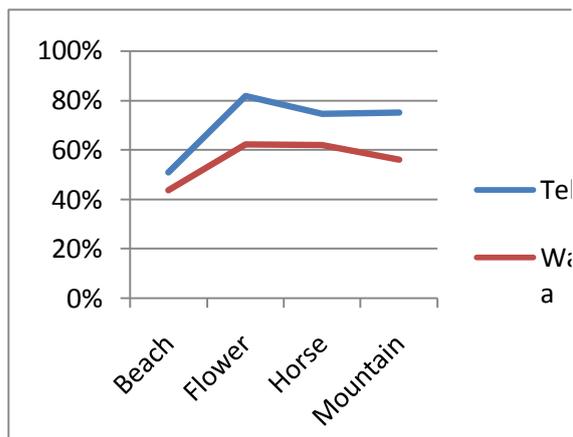
Tabel 2. Precision Dan Recall Sistem Pencarian Citra Berdasarkan Teks

Kategori	Precision	Recall
Beach	50,90%	43,60%
Flower	81,80%	74,80%
Horse	74,69%	74,40%
Mountain	75,20%	41,20%
Rata-rata	70,65%	58,50%

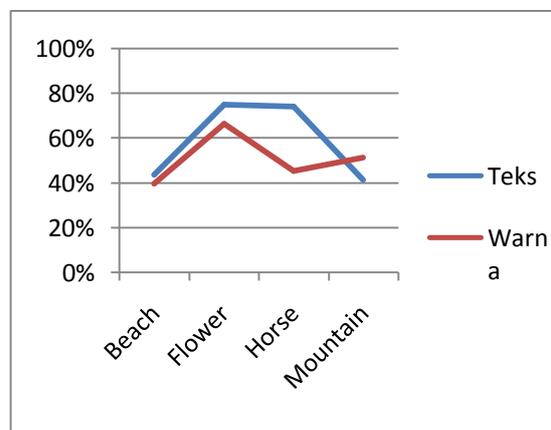
Tabel 3. Precision Dan Recall Sistem Pencarian Citra Berdasarkan Warna

Kategori	Precision	Recall
Beach	43,69%	39,60%
Flower	62,20%	66,40%
Horse	61,86%	45,20%
Mountain	56,09%	51,20%
Rata-rata	55,96%	50,70%

Tabel 2 dan 3 merupakan hasil *precision* dan *recall* dari sistem pencarian citra berdasarkan teks maupun warna setelah dilakukan rata-rata dari hasil pengujian terhadap 5 query acak setiap kategori. Dari hasil rata-rata semua kategori (rata-rata sistem) yang ditampilkan dalam tabel 2 dan 3, menunjukkan bahwa sistem pencarian citra berdasarkan teks lebih baik dibandingkan sistem pencarian citra berdasarkan warna karena mempunyai nilai rata-rata *precision* dan *recall* lebih tinggi. Selisih nilai *precision* dan *recall* antara sistem pencarian citra berdasarkan teks dan sistem pencarian citra berdasarkan warna adalah 14,69% untuk nilai *precision* dan 7,80% untuk nilai *recall*.



Gambar 5. Grafik Precision Pencarian Citra Berdasarkan Teks Dan Warna



Gambar 6. Grafik Recall Pencarian Citra Berdasarkan Teks Dan Warna

Pada gambar 5 grafik *precision* sistem pencarian citra berdasarkan teks lebih baik dari pada sistem pencarian citra berdasarkan warna dilihat dari semua kategori. Sedangkan untuk gambar 6 grafik *recall* sistem pencarian citra berdasarkan warna dengan kategori *mountain* mempunyai nilai *recall* lebih tinggi dari pada sistem pencarian citra berdasarkan teks, akan tetapi untuk kategori *beach*, *flower*, dan *horse* sistem pencarian citra berdasarkan teks mempunyai nilai *recall* lebih tinggi dari pada sistem pencarian citra berdasarkan warna.

III. Simpulan

- Ada perbedaan hasil dari sistem pencarian citra berdasarkan teks maupun pencarian citra berdasarkan warna. Hasil dari proses pencarian sistem pencarian citra berdasarkan teks ada beberapa data yang ditampilkan tetapi tidak ditampilkan

oleh sistem pencarian citra berdasarkan warna ataupun sebaliknya.

- Kemampuan sistem untuk menampilkan citra yang relevan dari seluruh citra yang berhasil ditemukembali oleh sistem pencarian citra berdasarkan teks dengan metode LSI lebih bagus 14,69% daripada sistem pencarian citra berdasarkan warna menggunakan metode color histogram dengan pengukuran jarak eucliden.
- Kelebihan sistem pencarian berdasarkan teks dapat menampilkan hasil pencarian yang lebih bervariasi. Sedangkan kekurangannya sistem pencarian citra berdasarkan teks dalam pendeskripsian citra masih diinputkan secara manual oleh user.
- Kelebihan sistem pencarian berdasarkan warna pengindeksan citra dilakukan oleh komputer. Sedangkan kelemahan pencarian citra berdasarkan warna sistem ini kurang cocok untuk pencarian citra secara global seperti pencarian citra berdasarkan teks dan lebih cocok untuk pengidentifikasian suatu citra.

IV. Daftar Pustaka

- [1] Anonym. *Latent semantic Indexing*, http://en.wikipedia.org/wiki/Latent_semantic_indexing, diakses 19 April 2014.
- [2] Anonym. *Precision And Recall*, http://en.wikipedia.org/wiki/Precision_and_recall, diakses 15 September 2014.
- [3] Ferdian, Edward, dkk. *Penerapan Metode Latent Semantic Indexing Pada Search Engine*. Depatemen Teknik Informatika. Institut Teknologi Bandung. Bandung.
- [4] Karmilasari, and Agus, Sumarna. 2011. *Temu Kenali Citra Basis Konten Warna*. Jurusan Teknik Informatika. Fakultas Teknologi Industri. Universitas Gunadarma. Yogyakarta.
- [5] Munir, Rinaldi. 2002. *Diktat Kuliah Pengolahan Citra, Edisi Kedua*. Depatermaen Teknik Informatika ITB.
- [6] Widodo, Yanu. 2007. *Penggunaan Color Histogram Dalam Image Retrieval*. Surabaya : Ilmu Komputer.