

IMPLEMENTASI DAN ANALISIS PENCARIAN CITRA VISUAL BERDASARKAN LOW LEVEL FEATURE CITRA DENGAN CONTENT-BASED IMAGE RETRIEVAL

Wahyu Pratomo¹, Tjokorda Agung Budi Wirayuda², Warih Maharani³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Content-Based Image Retrieval (CBIR) adalah suatu teknik untuk melakukan pencarian citra dengan mencocokkan ciri visual antara citra query dengan setiap citra pada dataset. Perbandingan dilakukan dengan menghitung jarak antar vektor visual citra yang didapat dari hasil ekstraksi ciri. Ciri pada citra dapat dibagi menjadi high level feature dan low level feature. Pada tugas akhir ini dilakukan analisis dan implementasi sistem CBIR dengan memanfaatkan low level feature. Metode ini mengekstrak fitur warna dalam model warna HSV dan mengekstrak tekstur pada model warna yCbCr. Hasilnya berupa matriks visual vektor dari citra yang disimpan dalam file indeks. Perhitungan jarak kemiripan dilakukan dengan metode Euclidean distance, Cosine distance, dan Histintersection distance. Tingkat keberhasilan sistem Content Based Image Retrieval diukur dengan precision, recall, dan F-measure. Kombinasi terbaik yang didapat adalah dengan menggunakan histintersection distance untuk fitur warna dan euclidean distance untuk fitur tekstur. Kombinasi ini menggunakan bobot yang sama antar masing-masing fitur (masing-masing 0,5).

Kata Kunci : CBIR, ekstraksi fitur, low level feature, visual vector, index file

Abstract

Content-Based Image Retrieval (CBIR) is a technique to perform image retrieval by comparing the visual content between query image with every image in the dataset. The comparison was done by calculating the distance between visual vectors of images obtained from the feature extraction. Characteristic of the image can be divided into high-level features and low-level features. In this final analysis and implementation of CBIR systems by using low-level features. This method is done by extracting color features in the HSV colorspace and extract the texture feature on the YCbCr colorspace. The result is a visual matrix of image vectors that are stored in an index file. Similarity distance calculation method with Euclidean distance, Cosine distance, and Histintersection distance. The success rate for Content-Based Image Retrieval system measured by precision, recall, and F-measure. The best combination is obtained by using histintersection distance for color features and euclidean distance for texture features. This combination uses the same weight between each feature (each 0.5).

Keywords : CBIR, Feature Extraction, low level feature, visual vector, index file

BAB I. PENDAHULUAN

1.1 Latar Belakang

Perkembangan yang sangat pesat di bidang komputer dan penyimpanan data elektronik mengakibatkan penambahan koleksi data gambar yang sangat besar setiap harinya. Seiring banyaknya data gambar yang ada, maka pencarian terhadap gambar semakin mengalami kesulitan sehingga perlu dibuat suatu sistem untuk melakukan *retriving* gambar-gambar tersebut secara otomatis.

Pencarian citra pada umumnya dilakukan menggunakan atribut yang melekat pada citra seperti tanggal pembuatan, lokasi penyimpanan, ukuran atau dengan atribut-atribut yang telah didefinisikan sebelumnya[9]. Namun hasil pencarian dengan cara ini tidak efisien dan sangat tergantung pada kepiawaian *user* dalam mendeskripsikan citra. Selain itu pencarian tidak bisa didasarkan pada isi visual pada citra[9]. Pada referensi [7] , sudah dibuat sistem pencarian yang menggunakan *low level feature* untuk *content based image retrieval*, namun belum mengimplementasikan kombinasi setingan lamda, pemilihan *distance measure* yang tepat untuk fitur warna, pemilihan *distance measure* yang tepat untuk fitur tekstur, dan normalisasi jarak antar *distance* yang memiliki skala yang berbeda.

Maka dari itu, pada tugas akhir ini yang berjudul ” Implementasi dan Analisis pencarian citra visual berdasarkan *low level feature* citra dengan *Content-Based Image Retrieval*” , akan dibuat sebuah aplikasi yang dapat melakukan pencarian citra berdasarkan informasi visual dari citra dengan menggunakan kombinasi fitur warna dan tekstur citra, dimana proses perbandingan antara citra *query* dengan citra dalam *database* citra akan dilakukan dengan beberapa kombinasi penghitungan jarak antara lain *euclidean distance*, *cosine distance*, dan *hist intersection distance*, dengan variasi nilai lamda antara (0,3) , (0,5) , dan (0,7). Sehingga nantinya dengan aplikasi ini, *user* tidak perlu lagi mendeskripsikan setiap citra secara manual karena sistem sudah mampu mendeskripsikan masing masing citra secara otomatis menurut *low level feature* citra tersebut.

1.2 Tujuan Dan Manfaat

Tujuan dari penulisan Tugas Akhir ini adalah :

1. Mengimplementasikan serta menganalisis teori *content-based image retrieval* ke dalam perangkat lunak yang dapat digunakan untuk mencari citra berdasarkan ciri visual citra *query*.
2. Mencari faktor-faktor yang berpengaruh terhadap hasil pencarian sehingga didapat konfigurasi yang terbaik untuk melakukan pencarian.

Manfaat penyusunan tugas akhir ini adalah :

1. Dapat membantu pengembangan aplikasi pencarian citra dari sudut pandang IR .

2. Dapat membantu mempermudah pengelolaan citra yang berjumlah banyak dan bersifat umum tanpa perlu mendefinisikan *metadata* secara manual.

1.3 Rumusan Masalah

Permasalahan yang dijadikan obyek penelitian dan pengembangan tugas akhir ini adalah sebagai berikut :

1. Bagaimana membuat aplikasi pencarian citra berdasarkan isi visual dari citra?
2. Bagaimana memproses query yang berupa citra agar dapat dimanfaatkan untuk *me-retrieve* data citra pada database citra (query by sample)?

1.4 Batasan Masalah

Dalam penulisan tugas akhir ini, penulis membatasi permasalahan mencakup hal-hal berikut :

- Tidak memperhatikan noise pada citra.
- Citra yang akan diamati berupa citra diam.
- Sistem bersifat *offline* dimana gambar didapat dari *dataset* didalam komputer.
- Mode warna yang digunakan pada ekstraksi ciri adalah RGB,HSV,dan YCbCr.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian tugas akhir ini adalah sebagai berikut :

1. Studi Literatur

Studi literatur dilakukan dengan mengumpulkan sumber-sumber yang berhubungan dengan pengolahan citra dan *information retrieval*. Adapun referensi utamanya yaitu buku textbook *Introduction To Information Retrieval* karangan *Christopher D. Manning , Prabhakar Raghavan, and Hinrich Schutze* [11]. Sedangkan untuk teori grafika & citra dipercayakan pada textbook *Image Processing : Principles And Applications* karangan *Tinku Acharya and Ajoy K. Ray* [15]. Selain itu, untuk teori grafika & citra digunakan juga paper tentang *Color and Texture Descriptors* karangan *B. S. Manjunath, Jens-Rainer Ohm, Vinod V. Vasudevan, and Akio Yamada* [2]. Dasar-dasar dari CBIR itu sendiri bersumber dari paper yang berjudul *An Overview of Content-Based Image Retrieval Techniques* karangan *Deb, S. and Zhang, Y* [1].

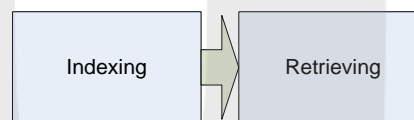
Bahan dan materi yang dipelajari meliputi teori dasar tentang *feature extraction*, *indexing*, dan *retrieving*. *Feature extraction* meliputi ekstraksi warna dan tekstur. *Indexing* mencakup cara-cara untuk menyimpan hasil *feature extraction* dalam bentuk *index file*. *Retrieving* mencakup pemilihan, pembobotan, dan penghitungan *distance* antar *visual vector*, serta pencarian *distance threshold* yang dapat memberikan hasil yang baik untuk setiap kelas.

2. Pengumpulan Data

Pengumpulan data bertujuan untuk mendapatkan dataset yang tepat guna mendukung analisis. Untuk tugas akhir ini, data yang dibutuhkan adalah dataset gambar untuk menguji sistem temu kembali citra. Dataset ini di download dari <http://wang.ist.psu.edu/%7Ejwang/test1.zip>

3. Perancangan Sistem

Berdasarkan hasil studi literatur maka gambaran sistem pencarian citra visual dengan *content-based image retrieval* secara umum dibagi menjadi dua bagian proses besar yaitu *indexing* dan *retrieving*.



Gambar 1-1-Skema Prancangan Utama

Proses pertama yaitu *indexing*. Dalam sistem pencarian citra visual dengan *content-based image retrieval*, maka proses *indexing* adalah proses melakukan ekstraksi fitur dari citra yang ada pada dataset dan menyimpannya kedalam sebuah file indeks. Setelah fitur didapat kemudian dilakukan proses *retrieving*, dimana dalam proses ini akan dilakukan proses temu kembali citra yang sudah terindeks pada proses *indexing*. Proses *retrieving* ini dilakukan dengan memasukkan citra yang dijadikan sebagai *query example* dan dilakukan ekstraksi fitur terhadap citra *query* tersebut. Kemudian dilakukan penghitungan kemiripan dan *thresholding*, lalu citra yang tergolong mirip akan ditetapkan sebagai hasil pencarian dan ditampilkan ke *user*. Langkah selanjutnya adalah menentukan konfigurasi *distance measure* yang akan digunakan untuk sistem CBIR ini. Dari *distance measure* yang telah ditetapkan, kemudian dilakukan variasi skenario pengujian sistem untuk mencari konfigurasi yang terbaik dalam melakukan proses pencarian.

4. Implementasi Sistem

Bertujuan untuk melakukan implementasi metode pada sistem sesuai dengan analisa perancangan yang telah dilakukan dengan simulasi menggunakan software Matlab 7.4

5. Analisa Performansi

Akan dilakukan uji coba pada sistem yang telah dibuat dan akan di analisa data hasil implementasi yang telah dilakukan.

6. Pembuatan Laporan Penelitian

Setelah data dari hasil implementasi dianalisa, maka langkah terakhir adalah mengambil kesimpulan dari hasil analisa untuk kemudian dibuat buku laporan hasil penelitian (TA) ini.

1.6 Sistematika Penulisan

Tugas Akhir ini disusun dengan sistematika penulisan yang terdiri atas lima bab yang disusun sebagai berikut :

BAB I Pendahuluan

Pada bab ini akan dibahas mengenai latar belakang masalah, rumusan masalah, tujuan&manfaat penelitian, batasan masalah, metodologi penelitian, hipotesis, dan sistematika penulisan tugas akhir.

BAB II Dasar Teori

Pada bab ini akan dipaparkan berbagai dasar teori yang mendukung dan menjelaskan istilah-istilah yang ada pada proposal tugas akhir ini.

BAB III Perancangan dan Simulasi Sistem

Pada bab ini akan dijelaskan mengenai model sistem dan diagram blok sistem aliran pengerjaan penelitian.

BAB IV Pengujian Sistem Dan Analisa Hasil Simulasi

Pada bab ini akan dilakukan implementasi dan pengujian sistem serta analisa terhadap output yang dihasilkan.

BAB V Kesimpulan dan Saran

Bab ini merumuskan kesimpulan dari perancangan dan analisa yang telah dilakukan dan saran untuk pengembangan lebih lanjut.

BAB V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis terhadap pengujian yang dilakukan dalam sistem pencarian citra visual berdasarkan low level feature citra dengan content based image retrieval, dapat disimpulkan beberapa hal sebagai berikut:

1. Implementasi perangkat lunak untuk melakukan pencarian citra visual dengan low level feature citra untuk content based image retrieval terbukti mampu bekerja dengan baik.
2. Keberhasilan sistem dalam menemukan citra yang dicari dipengaruhi low level feature dari citra baik itu citra pada dataset maupun citra query.
3. Hasil pengujian menunjukkan bahwa hasil terbaik diperoleh dengan settingan *distance measure* warna digunakan *histogram intersection distance* (HD) dan untuk tekstur digunakan *euclidean distance* (ED).
4. Hasil pengujian juga menunjukkan bahwa hasil pengujian terbaik untuk varian *distance measure* HD-ED diperoleh dengan settingan lamda sama dengan 0,5 (penggunaan fitur warna dan tekstur digunakan secara seimbang).
5. Low level feature citra kadang tidak sesuai dengan makna semantik dari citra tersebut. Sebagai contoh, ciri pada low level feature tidak akan mampu membedakan matahari terbit dengan matahari terbenam, laut & langit biru dengan gunung salju, dan lain sebagainya. Selama content nya sesuai secara low level feature (mirip secara warna dan tekstur) maka akan ditetapkan sebagai hasil pencarian.

5.2 Saran

Penelitian lebih lanjut diharapkan dapat memperbaiki kekurangan yang ada dan diharapkan dapat mengembangkan apa yang telah dilakukan pada penelitian ini. Untuk itu disarankan hal-hal berikut.

1. Dalam pengerjaan Tugas Akhir selanjutnya, sebaiknya dibuat suatu dataset yang memiliki *relevance judgement* untuk gambar query yang tidak terdapat dalam database atau gambar query yang merupakan potongan dari gambar pada database.
2. Pada penelitian selanjutnya diharapkan dataset dapat berupa database yang online dari website di seluruh dunia.
3. Untuk penelitian selanjutnya, fitur dapat ditingkatkan menjadi high level feature sehingga program mampu mengenali objek objek yang terdapat pada gambar.
4. Pengerjaan Tugas Akhir menggunakan metode *image preprocessing* yang lain agar dapat dibandingkan dengan penelitian sebelumnya.

DAFTAR PUSTAKA

- [1] Deb, S. and Zhang, Y. 2004. An Overview of Content-Based Image Retrieval Techniques. IEEE.
- [2] B. S. Manjunath, Member, IEEE, Jens-Rainer Ohm, Member, IEEE, Vinod V. Vasudevan, Member, IEEE, and Akio Yamada. Color and Texture Descriptors . IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 6, JUNE 2001
- [3] John R. Smith , 1997, Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression, Columbia University.
- [4] Jia Li, and James Z. Wang, 2003 , Automatic Linguistic Indexing of Pictures
by a Statistical Modeling Approach
- [5] James Z. Wang, Jia Li, and Gio Wiederhold, 2001 , SIMPLiCity: Semantics-Sensitive Integrated Matching for Picture Libraries
- [6] Muhammad Erwin Ashari Hariyono¹ dan Wahyudi: 2005 , CUSTOMER INFORMATION GATHERING MENGGUNAKAN METODE TEMU KEMBALI INFORMASI DENGAN MODEL RUANG VEKTOR, Laboratorium Pemrograman dan Informatika Teori, Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.
- [7] Michele Saad , May 09, 2008 , Low-Level Color and Texture Feature Extraction for Content-Based Image Retrieval, Final Project Report, EE 381K: Multi-Dimensional Digital Signal Processing
- [8] Jorma Laaksonen , Erkki Oja , Markus Koskela , and Sami Brandt , 2000 , Analyzing Low-Level Visual Features Using Content-Based Image Retrieval , Published in the Proceedings of ICONIP 2000, Taejon, Korea, November 2000
available at <URL:<http://www.cis.hut.fi/picsom/publications.html>>
- [9] PENCARIAN CITRA VISUAL BERBASIS ISI CITRA MENGGUNAKAN FITUR WARNA CITRA Kusrini¹, Agus Harjoko²; ¹Dosen STMIK AMIKOM Yogyakarta : ²Dosen FMIPA Universitas Gadjah Mada Yogyakarta.

- [10] Yutaka Sasaki , 2007 , The truth of the F-measure
- [11] Christopher D. manning , Prabhakar Raghavan , and Hinrich Schutze . 2008 . INTRODUCTION TO INFORMATION RETRIEVAL , Cambridge University.
- [12] Michael J. Swain and Dana H. Ballard . 2005 . Color Indexing
- [13] Achmad Basuki , Nana Ramadijanti . 2006 . CONTENT BASED IMAGE RETRIEVAL BERDASARKAN CIRI TEKSTUR MENGGUNAKAN WAVELET
- [14] Alvy Ray Smith, "Color Gamut Transform Pairs", SIGGRAPH '78 , New York Institute of Technology, Old Westbury, NY 11568
- [15] Tinku Acharya and Ajoy K. Ray .2005 . Image Processing : Principles And Applications
- [16] William K.Pratt. 2001. Digital Image Processing Third Edition.
- [17] Rafael C.Gonzalez and Richard E.woods .2002 . Digital Image Processing Second Edition . Prentice-Hall, Upper Saddle River, New jersey.
- [18] Anil K.Jain .1989 . Fundamentals Of Digital Image Processing, Prentice-Hall International.
- [19] Mark Nixon and Alberto Aguado .2008 .Feature Extraction & image Processing, second edition, elseiver press.
- [20] Donald Hearn and M.Pauline Baker.1997.Computer Graphics C Version, second edition, Prentice Hall,upper sadle river NJ.
- [21]Slide Presentasi Dengsheng Zhang dan Guojun Lu, Generic Fourier Descriptor for Shape-based Image Retrieval, presented in Gippsland School of Comp. & Info Tech, Monash University Churchill, Australia.
- [22]Slide Presentasi Yong Rui, Thomas Huang and Shih-Fu Chang, Image Retrieval: Current Techniques, Promising Directions, and Open Issues Published in the Journal of Visual Communication and Image Representation. Presented by: Deepak Bote .
- [23]Slide Presentasi Michael saad, Content Based Image Retrieval, EE-381K-14: Multi-Dimensional Digital Signal Processing, March 06, 2008.