

QUERY IMAGE DENGAN METODA SIMILARITY-BASED IMAGE QUERIES USING SIMILARITY-BASED METHOD

Andra Andarini¹, Dhinta Darmantoro², Eddy Muntina Dharma³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Salah satu kesulitan user dalam pencarian image adalah ketika mencari kata kunci yang tepat untuk mendeskripsikan image yang diinginkan. Terlebih lagi hasil query yang seringkali tidak sesuai dengan keinginan user. Oleh sebab itu, ditawarkan satu sistem yang dapat mencari image dengan menggunakan image. Adapun pendekatan yang digunakan adalah metoda similarity-based, yaitu teknik untuk mengembalikan image berdasarkan jenis dan tingkat kemiripan tertentu. Sehingga yang menjadi tujuan dari tugas akhir ini adalah membuat database image, membuat query berupa image, dan menguji performansi query image yang menggunakan metoda similarity-based.

Langkah-langkah yang ditangani sistem saat memproses query, dimulai setelah user menginputkan query. Kemudian sistem mengekstraksi fitur warna dan teksturnya dengan menggunakan histogram warna dan co-occurrence matrix. Setelah itu, dihitung tingkat similarity antara image query dengan image database. Hasil query kemudian diurutkan menurun sesuai dengan similarity-level. Dalam sistem, disediakan database berisi image berukuran 384 X 256 pixel dengan karakteristik bervariasi, yaitu image non-texture, texture, landscape, background monoton warna, wajah full frame, hitam putih, contrast tinggi/rendah, dan image terang/gelap.

Hasil dari tugas akhir ini adalah perangkat lunak yang dapat melakukan pencarian image dalam database dengan jenis dan tingkat kemiripan tertentu. Setelah dilakukan pengujian terhadap perangkat lunak akan diketahui metoda similarity-based yang sesuai dengan karakteristik image.

Kata Kunci : query image, database multimedia, similarity-based, content image

Telkom
University

Abstract

One of difficulties in image searching is looking the correct keywords to describe the desired image. Moreover the query result often doesn't match with the intended. To minimize that kind of error a solution is given by providing an image searching system using image. The approach that is used is similarity-based method, it is a technique that take an image as input and return images as output, based on similarity kind and level. So that this final assignment objectives are, first, to build image database; second, to make image query to retrieve data from the database, and third, to test image query performance with similarity-based method.

The steps handled by system in query processing started after user inserting image query. Then system extracting its color and texture features based on color histogram and co-occurrence matrix. After that, system will count similarity-level between query image and database image. The query result then sorted by descending order based on similarity-level.

This system provide database contains images with 384 X 256 pixel with various characteristics such as non-texture image, texture image, landscape, monotonicity background color, picture of full-frame face, black and white image, image with low /high contrast, and bright/dark image.

The result of this final assignment is a software that can search images in the database with similarity type and similarity level based on what user want. After software testing, the appropriate similarity-based method that compatible with image characteristic will be known

Keywords : query image, database multimedia, similarity-based, content image

1. Pendahuluan

1.1 Latar belakang

Dalam model data relasional biasanya user dihadapkan pada model query tekstual, yang mengharuskan user memasukkan teks untuk mengambil/menampilkan data dari database. Namun jika database yang digunakan adalah database multimedia, bisa jadi query yang digunakan akan berbeda bentuknya. Database jenis ini akan menyimpan data-data multimedia, seperti *image*, suara, dan video. Untuk melakukan query pada jenis data tersebut seringkali tidak tepat jika menggunakan query tekstual.

Pada database multimedia dengan data *image* misalnya, untuk banyak kasus akan lebih praktis jika digunakan query yang melibatkan *image* sebagai bagian di dalamnya. Meskipun tidak menutup kemungkinan untuk menggunakan query tekstual jika database tersebut menyimpan deskripsi yang berhubungan dengan *image* yang disimpan. Namun penggunaan query dengan teks pada database *image* memiliki beberapa kelemahan. Pertama, dengan menggunakan teks akan sukar untuk mendeskripsikan ciri-ciri persepsi pada fitur visual. Kedua, penggunaan teks tidak sesuai untuk memodelkan kesamaan persepsi. Dan terakhir karena deskripsi teks berasal dari orang yang membuat database, terkadang isinya berbeda dengan persepsi user yang menggunakan sistem. Oleh karena itu digunakan metoda *similarity-based* untuk mengaplikasikan query *image*. Metoda *similarity-based* merupakan teknik pencarian *image* yang serupa dengan melakukan penghitungan kesamaan antar fitur *low-level image* (misalnya, warna, tekstur, bentuk dan sebagainya).

Untuk menggunakan query berupa *image* harus ada aplikasi yang memproses fitur *low-level* dari *image* query maupun *image* dalam database, *image*, setelah itu baru bisa dilakukan penghitungan/pengukuran kesamaan *image*. Namun sebenarnya tingkat kemiripan *image* satu dengan *image* lainnya tidak ada alat ukurnya yang pasti, oleh sebab itu harus didefinisikan terlebih dahulu jenis dan tingkat kemiripannya.

1.2 Perumusan masalah

Masalah-masalah yang terkait dengan tugas akhir ini adalah:

1. Bagaimana uji performansi *image* query dilakukan pada database multimedia dengan data *image*.
2. Apakah *image* dalam database bisa di-query dengan menggunakan *image*.
3. Seperti apakah content-based *image* retrieval digunakan pada query.
4. Bagaimana metoda *similarity-based* digunakan pada database multimedia.
5. Bagaimana peran mekanisme peringkat/ranking dalam query result, apakah sesuai dengan tingkat kemiripan mata manusia.
6. Apa saja fitur-fitur yang digunakan untuk mengukur kesamaan *image* query dengan *image* dalam database.

1.3 Tujuan

Tujuan pembahasan tugas akhir ini adalah:

1. Membuat database untuk objek *image*.
2. Membuat query *image* untuk mencari *image* dari database.
3. Menguji performansi query *image* berdasarkan mirip tidaknya data yang di-*retrieve*, dengan menggunakan metoda *similarity-based*.

1.4 Metodologi penyelesaian masalah

Untuk menyelesaikan permasalahan yang telah dirumuskan sebelumnya, ada beberapa tahapan yang harus dilaksanakan. Tahapan-tahapan ini didefinisikan sebagai metodologi pemecahan masalah, yang terdiri dari:

1. Studi Literatur
Sebagai langkah awal untuk mengerjakan tugas akhir ini, sangatlah diperlukan untuk mengerti dan memahami konsep database. Selain itu juga memanfaatkan fitur-fitur yang dimiliki oleh DBMS Microsoft SQL Server 2000 dalam menyimpan *image* dan mengerti bagaimana representasinya. Skema database yang digunakan dalam tugas akhir ini akan berisi *image*, untuk itu perlu memahami bagaimana *image* disimpan dalam database, info-info apa saja yang perlu disimpan untuk dalam database, dan bagaimana cara mengolah *image*. Hal-hal tersebut bisa dipelajari dalam pengolahan citra dan basis data multimedia Adapun yang paling penting untuk dipahami dalam tugas akhir ini adalah bagaimana metoda *retrieve* data, yang dalam hal ini menggunakan *similarity-based*.
2. Analisis
Tahapan setelah melakukan studi literatur adalah membuat skema yang merepresentasikan database *image*. Adapun *image* yang disediakan dalam database berasal dari database *image* Corel, *image* hasil *download* dari web, dan koleksi pribadi, dengan ukuran *image* maksimal 384 X 384 pixel. Dalam melakukan analisis ini yang menjadi fokus adalah bagaimana metoda *similarity-based* memproses *image* sebagai query untuk mencari *image* yang "sama". Ukuran kesamaan *image* tersebut biasanya berada antara 0 sampai 1. Jika kesamaan bernilai 1 artinya *image* sama persis (100%) dan jika bernilai di bawah 1 artinya *image* tidak sama persis (misal, query *image* dengan *recognition degree* 0.2 berarti query hanya mengambil *image* lebih dari 20% bagian *image*-nya mirip).
3. Perancangan
Dalam tahap perancangan, akan didefinisikan rancangan basis data dan rancangan antarmuka dengan database *image*. Aplikasi ini akan memiliki beberapa fitur pilihan dalam mendefinisikan kesamaan *image*, yaitu berdasarkan warna, tekstur, maupun keduanya. Sistem akan melakukan pencarian *image* dengan menggunakan *query-by-example*.
4. Implementasi
Tahapan ini merupakan proses pengkodean atau pembuatan program dari desain yang telah didefinisikan sebelumnya.
5. Testing
Dalam tahapan ini aplikasi akan diuji dengan menggunakan beberapa skenario query kemudian dilakukan analisis terhadap hasil pengujian.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Setelah melakukan seluruh tahapan pengembangan sistem, yang diawali dengan analisis sampai dengan pengujian sistem, maka dapat diambil beberapa kesimpulan sebagai berikut:

- 1) Sistem yang dibuat dapat melakukan query *image* menggunakan *image*.
- 2) Pencarian *image* dengan menggunakan *color histogram* yang dikuantisasi menjadi 4-bit menghasilkan jenis warna yang masih terlalu banyak, sehingga query dengan *similarity-based on color* diproses dengan waktu cukup lama.
- 3) Berdasarkan pengujian, masing-masing karakteristik gambar memiliki jenis *similarity-based* sendiri untuk mendapatkan hasil yang "sama". Karakteristik gambar dan metoda yang dianjurkan adalah sebagai berikut:
 - a. *Image* non tekstur, gunakan *similarity-based on color and texture*.
 - b. *Image* bertekstur, gunakan *similarity-based on texture*.
 - c. *Image* dengan *contrast* tinggi/rendah, gunakan *similarity-based on texture* atau *similarity based on color and texture*.
 - d. *Image* yang terang/gelap, gunakan *similarity-based on color and tekstur*.
 - e. *Image* hitam putih, gunakan *similarity-based on color and texture*.
 - f. *Image* berupa *landscape*, gunakan *similarity based on texture*.
 - g. *Image* berupa wajah *full frame*, gunakan *similarity based on color and texture*.
 - h. *Image* dengan warna *background* monoton, gunakan *similarity based on color and texture*.
- 4) Berdasarkan pengujian, dengan menggunakan *image* yang dimodifikasi, maka dapat disimpulkan bahwa penggunaan operasi zoom tidak memberi pengaruh yang besar terhadap hasil query, operasi rotasi akan berpengaruh jika *image* yang dirotasi ada bagiannya yang berwarna putih, operasi crop akan memberi hasil berbeda dengan *image* asli jika *image* query memotong hingga 50% bagian *image*, dan operasi blur akan menghasilkan query yang semakin sedikit jika *image* semakin banyak di-blur.
- 5) Masing-masing metoda pencarian *similarity* memiliki kelebihan dan kekurangan sesuai dengan *image* query yang diinputkan.

5.2 Saran

Untuk pengembangan lebih lanjut dari sistem yang telah dihasilkan, adapun saran yang dapat diberikan adalah:

- a. Sistem dapat ditambahkan dengan metoda ekstraksi fitur lainnya, misalnya bentuk dan lain sebagainya
- b. Metoda ekstraksi fitur warna maupun tekstur dapat dikembangkan dengan metoda selain *color histogram* dan *co-occurrence matrix*, misalnya untuk

- ekstraksi fitur warna dapat menggunakan *corelogram*, dan teksturnya menggunakan *wavelet*.
- c. Sistem basis data yang digunakan bisa diganti dengan basis data yang mendukung tipe data objek, seperti Oracle.
 - d. Pengembangan perangkat lunak bisa dibuat untuk jaringan, sehingga bisa user dan database *image* tidak perlu berada dalam satu PC.



Daftar Pustaka

- [1] Aksoy, Haralic, 2003, "Using Texture in *Image* Similarity and Retrieval", Univ. of Washington, didownload pada bulan November 2006.
- [2] B. Ilaria, 2002, "Efficient and Effective Similarity Search in *Image* Databases", Thesis, didownload pada bulan Januari 2007.
- [3] Carson, Belongie, Greenspan, Malik, 1997, "Region-Based *Image* Querying", Published by IEEE. (CVPR '97 Workshop on Content-Based Access of *Image* and Video Libraries).
- [4] Copeland, Ravichandran, Trivedi, 2001, "Texture synthesis using gray-level co-occurrence models: algorithms, experimental analysis, and psychophysical support", University of California, San Diego Computer Vision and Robotics Research (CVRR) Laboratory Electrical and Computer Engineering Department.
- [5] Dharma, Eddy Muntina, 2006, "Slide Mata Kuliah Grafika dan Citra", STT Telkom, Bandung.
- [6] J. Sangoh, 2001, "Histogram-Based Color *Image* Retrieval", Project Report, didownload pada bulan Januari 2007.
- [7] L. Adrian, 1991, "Introductory Computer Vision and *Image* Processing", Singapore: McGraw-Hill Co.
- [8] Mocofan, Vasiu, 2002, "Algorithm Based on *Image* Similarity for Content Search in an *Image* Database", Univ. of Timisoara, didownload pada bulan September 2006.
- [9] Pitas, Ioannis, 1993, "Digital *Image* Processing Algorithms", UK: Prentice-Hall
- [10] Schalkoff Robert J, 1989, "Digital *Image* Processing and Computer Vision", Singapore: John Willey & Sons Inc.
- [11] Subrahmanian V. S., 1998, "Principles of Multimedia Database Systems", San Francisco: Morgan Kaufmann Publisher Inc.
- [12] Ortega Michael, 2002, "Content-Based Retrieval Database Management System with Support for Similarity Searching and Query Refinement", (Technical Reports).
- [13] Yu, Niu, Agrawal, Abbadi, Singh, 2004, "Efficient Similarity Search on Vector Sets", Computer Science Dept. Univ. of California.
- [14] www.corel.com
- [15] www.textureking.com
- [16] www.bigfoto.com
- [17] www.luminous-landscape.com