

**IMPLEMENTASI *CONTENT BASED IMAGE RETRIEVAL* (CBIR)
UNTUK PENCARIAN GAMBAR BUNGA MENGGUNAKAN QUERY
GAMBAR DENGAN METODE *CHI-SQUARE DISTANCE***

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer



Oleh :

REINALDY RAHMADIAN

1405061

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA**

2022

**IMPLEMENTASI *CONTENT BASED IMAGE RETRIEVAL* (CBIR)
UNTUK PENCARIAN GAMBAR BUNGA MENGGUNAKAN QUERY
GAMBAR DENGAN METODE *CHI-SQUARE DISTANCE***

Oleh
Reinaldy Rahmadian
NIM 1405061

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer pada Fakultas Pendidikan Matematika Dan Ilmu
Pengetahuan Alam

© Reinaldy Rahmadian
Universitas Pendidikan Indonesia
Januari 2022

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis.

**IMPLEMENTASI *CONTENT BASED IMAGE RETRIEVAL* (CBIR)
UNTUK PENCARIAN GAMBAR BUNGA MENGGUNAKAN QUERY
GAMBAR DENGAN METODE *CHI-SQUARE DISTANCE***

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



Lala Septem Riza, M.T., Ph.D.

NIP 197809262008121001

Pembimbing II,



Herbert Siregar, M.T.

NIP 197005022008121001

Mengetahui,

Ketua Program Studi Ilmu Komputer

Dr. Rani Megasari, M.T.

NIP 198705242014042002

**IMPLEMENTASI *CONTENT BASED IMAGE RETRIEVAL* (CBIR)
UNTUK PENCARIAN GAMBAR BUNGA MENGGUNAKAN QUERY
GAMBAR DENGAN METODE CHI-SQUARE DISTANCE**

ABSTRAK

Salah satu proses yang penting dalam pengelolaan data yaitu adanya proses penemuan kembali informasi yang dibutuhkan oleh pengguna. Proses ini disebut dengan istilah temu balik informasi atau *Information Retrieval* (IR). Tujuan dari IR adalah memberikan dokumen yang berisi informasi yang relevan dengan query yang dimasukkan oleh pengguna. Pada mesin pencari, proses pencarian sebagian besar dilakukan dengan query berbentuk teks. Teknik pencarian ini dapat digunakan jika pengguna hanya mengingat nama file atau keterangan pada citra, namun tidak dapat digunakan jika pengguna hanya mengingat gambar dari citra. Untuk itu, penelitian ini membuat sebuah model dan implementasi sistem untuk melakukan pencarian dengan menggunakan query berupa citra, yaitu disebut sebagai *content based image retrieval* (CBIR). Sistem ini dibuat untuk mencari citra bunga dengan dataset yang digunakan diambil dari beberapa web seperti flickr, google images, dan yandex images. Sedangkan data testing didapatkan dari hasil foto bunga secara manual. Model sistem ini memiliki beberapa tahapan yaitu (1) *feature extraction* (2) *Color Histogram* (3) Normalisasi *Histogram* (4) *Chi-Square Distance* (5) *Rangking*. Penelitian ini menghasilkan sebuah sistem yang dapat melakukan pencarian dengan menggunakan citra sebagai *query* nya. Eksperimen yang dilakukan pada penelitian ini dibagi menjadi 2 eksperimen, yaitu *fitting validation* atau menggunakan dataset sebagai *query* nya, dan eksperimen menggunakan data testing sebagai *query* nya. Keseluruhan eksperimen masing – masing mendapatkan persentase presisi tingkat keberhasilan 0 – 100%, dengan rata – rata presisi tingkat keberhasilan 38,10%, 34,29%, dan 28,57%

Kata Kunci: *Content Based Image Retrieval*, *Chi-Square Distance*, ekstraksi fitur, histogram warna

**IMPLEMENTASI *CONTENT BASED IMAGE RETRIEVAL* (CBIR)
UNTUK PENCARIAN GAMBAR BUNGA MENGGUNAKAN QUERY
GAMBAR DENGAN METODE CHI-SQUARE DISTANCE**

ABSRTACT

One of the important processes in data management is the process of rediscovering the information needed by the user. This process is known as information retrieval or Information Retrieval (IR). The purpose of IR is to provide a document containing information relevant to the query entered by the user. In search engines, the search process is mostly done with text queries. This search technique can be used if the user only remembers the file name or description of the image, but cannot be used if the user only remembers the image from the image. For this reason, this study creates a model and implementation of a system to perform searches using a query in the form of images, which is known as content based image retrieval (CBIR). This system was created to search for flower images with the dataset used taken from several websites such as flickr, google images, and yandex images. While the testing data obtained from the results of photos of flowers manually. This system model has several stages, namely (1) feature extraction (2) Color Histogram (3) Histogram normalization (4) Chi-Square Distance (5) Ranking. This research resulted in a system that can perform a search by using the image as the query. The experiments carried out in this study were divided into 2 experiments, namely fitting validation or using a dataset as the query, and experiments using testing data as the query. The entire experiment each got a precision percentage of 0-100% success rate, with an average precision success rate of 38.10%, 34.29%, and 28.57%

Keywords ; Content Based Image Retrieval, Chi-Square Distance, feature extraction, color histogram

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL	v
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah	5
1.6 Sistematika penulisan	5
BAB II KAJIAN PUSTAKA	7
2.1. Search Engine	7
2.2. Image Processing	8
2.3. Content-Based Image Retrieval.....	11
2.4. Ekstraksi Fitur / Feature Extraction	12
2.5. Histogram	16
2.6. Chi-Square	17
BAB III METODOLOGI PENELITIAN	20
3.1 Desain Penelitian	20
3.2 Metode Penelitian	22
3.2.1 Metode Pengumpulan Data	22
3.2.2 Metode Pengembangan Perangkat Lunak	23
3.3 Alat dan Bahan Penelitian	24
3.3.1 Alat Penelitian	24
3.2.2 Bahan Penelitian.....	24
BAB IV HASIL DAN PEMBAHASAN	25
4.1. Data Penelitian.....	25
4.2. Perancangan Model Komputasi	30
4.3. Pembangunan Perangkat Lunak	50

4.4.	Skenario Eksperimen.....	56
4.4.1.	Fitting Validation	58
4.4.2.	Data Testing	59
4.5.	Hasil Eksperimen dan Pembahasan	60
BAB V	KESIMPULAN DAN SARAN	137
5.2.	Kesimpulan.....	137
5.2.	Saran	138
DAFTAR PUSTAKA		139

DAFTAR GAMBAR

Gambar 2.1 Alur Proses Pengolahan Citra.....	10
Gambar 2.2 Proses CBIR Secara Umum	12
Gambar 3.1 Desain Penelitian.....	20
Gambar 3.2 Metode Waterfall (Sukamto & Shalahuddin, 2011).....	23
Gambar 4.1 Contoh Hasil Nilai Histogram Dataset.....	29
Gambar 4.2 Perancangan Model Aplikasi.....	31
Gambar 4.3 Flickr	32
Gambar 4.4 Google Images.....	32
Gambar 4.5 Yandex Images	32
Gambar 4.6 Hasil Konversi Ruang Warna RGB ke HSV	33
Gambar 4.7 Implementasi Kode <i>Feature Extraction</i>	53
Gambar 4.8 Implementasi Kode <i>Color Histogram</i>	54
Gambar 4.9 Implementasi Kode untuk <i>Chi-Square Distance</i>	54
Gambar 4.10 Implementasi Kode untuk Rangkings Nilai Distance	55
Gambar 4.11 Proses Bisnis Sistem CBIR ReImg	57
Gambar 4.12 Skenario Fitting Validation	58
Gambar 4.13 Data Testing Eksperimen	59

DAFTAR TABEL

Tabel 2. 1 Warna HSV	16
Tabel 4. 1 Contoh Dataset Setiap Sumber.....	26
Tabel 4. 2 Contoh Hasil Masking	34
Tabel 4. 3 Hasil Perhitungan Histogram.....	36
Tabel 4. 4 Hasil Normalisasi Nilai Histogram.....	40
Tabel 4. 5 Hasil Normalisasi Setelah Ditampung Kedalam Bentuk Vector	46
Tabel 4. 6 Perancangan Fitur Perangkat Lunak	52
Tabel 4. 7 Pengujian Setiap Fungsi dan Error Handling pada Program	56
Tabel 4. 8 Hasil Eksperimen Pencarian Top 3 dengan Query Dataset	61
Tabel 4. 9 Hasil Eksperimen Pencarian Top 5 dengan Query Dataset	65
Tabel 4. 10 Hasil Eksperimen Pencarian Top 10 dengan Query Dataset	71
Tabel 4. 11 Hasil Eksperimen Pencarian Top 3 dengan Query Data Testing.....	81
Tabel 4. 12 Hasil Eksperimen Pencarian Top 5 dengan Query Data Testing.....	85
Tabel 4. 13 Hasil Eksperimen Pencarian Top 10 dengan Query Data Testing.....	91
Tabel 4. 14 Skor Distance Hasil Eksperimen Top 3 Query Dataset	102
Tabel 4. 15 Skor Distance Hasil Eksperimen Top 5 Query Dataset	103
Tabel 4. 16 Skor Distance Hasil Eksperimen Top 10 Query Dataset	105
Tabel 4.17 Skor Distance Hasil Eksperimen Top 3 Query Data Testing.....	109
Tabel 4.18 Skor Distance Hasil Eksperimen Top 5 Query Data Testing.....	112
Tabel 4.19 Skor Distance Hasil Eksperimen Top 10 Query Data Testing.....	115
Tabel 4.20 Rekapitulasi Hasil Eksperimen Top 3 Query Dataset	123
Tabel 4.21 Rekapitulasi Hasil Eksperimen Top 5 Query Dataset.....	123
Tabel 4.22 Rekapitulasi Hasil Eksperimen Top 10 Query Dataset.....	124
Tabel 4.23 Rekapitulasi Hasil Eksperimen Top 3 Query Data Testing	125
Tabel 4.24 Rekapitulasi Hasil Eksperimen Top 5 Query Data Testing	128
Tabel 4.25 Rekapitulasi Hasil Eksperimen Top 10 Query Data Testing	131
Tabel 4.26 Kinerja Chi-Square Distance Semua Hasil Eksperimen Top 3.....	134
Tabel 4.27 Kinerja Chi-Square Distance Semua Hasil Eksperimen Top 5.....	135
Tabel 4.28 Kinerja Chi-Square Distance Semua Hasil Eksperimen Top 5.....	136

DAFTAR PUSTAKA

- Agustina, M. (2009). Search Engine Optimization,. Andi.
- Ajie, S. (2010). Analisis Deteksi Tepi Untuk Mengidentifikasi Pola Bunga. Teknik Elektro Universitas Diponegoro.
- Arli Fauzi, M. (2020). Implementasi Metode On-Page Search Engine Optimization (SEO) Pada Website Bukuwarung. Technical Report.
- Bashkov, E., & Shozda, N. (2012). Content-Based Image Retrieval Using Color Histogram Correlation. Graphicon proceedings.
- Ferdiansyah, d. (2012). Sistem Temu Kembali Citra Berbasis Histogram Warna Fuzzy Untuk Pencarian Citra Berwarna. Journal of Informatics and Technology, No1: 130136.
- Gonzales, R. C., Woods, R. E., & Eddins, S. L. (2004). Digital Image Processing Using Matlab. *Pearson Prentice Hall*.
- Gonzalez, R. C., & Woods, R. E. (2002). Digital Image Processing Second Edition. *Prentice Hall*.
- Han, J., & Ma, K. (2002). Fuzzy Color Histogram and Its Use in Color Image Retrieval. IEEE Transaction on Image Processing, Vol 11 No 8.
- Hardini, A. (2020). Penerapan Metode Chi Square Terhadap Hubungan Tingkat Pendidikan Dengan Lapangan Kerja Di Provinsi Sumatera Utara. Journal Universitas Sumatera Utara.
- Hastuti, I., Hariadi, M., & Purnama, I. (2009). Content Based Image Retrieval berdasarkan Fitur Bentuk Menggunakan Metode Gradient Vector Flow Snake. Seminar Nasional Informatika 2009, 140-145.
- Hendrawan, R., & Puspitasari, D. (2016). Sistem Temu Kembali Citra Dengan Histogram Fuzzy dan Chi-Square Distance. Jurnal Informatika Polinema, Malang.
- Hussain, C., Rao, D., & Praveen, T. (2013). *Color Histogram Based Image. International Journal of Advanced Engineering Technology*, 63–66.

- Kurniawan, T. (2015). Pemanfaatan Content Based Image Retrieval Berbasis Color Histogram Dengan *Histogram Equalization* (HE) Dan *Euclidean Distance* Untuk Menentukan Kematangan Cabai Merah Besar. Universitas Dian Nuswantoro.
- Maridana, A. E. (2014). Aplikasi Content Based Image Retrieval Motif Batik Menggunakan. *Skripsi Malang: Universitas Islam Negeri Maulana Malik Ibrahim*.
- Meivi, K., & Chaulina, A. O. (2014). Implementasi Content Based Image Retrieval Untuk Menganalisa Kemiripan Bakteri Yoghurt Menggunakan Metode Latent Semantic Indexing. *IC-ITECHS*.
- Pramiswari, Aisyah, N., Purwananto, Yudhi, S., & Rully. (2010). Implementasi Denoising Citra RGB Menggunakan Metode *Wavelet* Berbasis Logika *Fuzzy*. Fakultas Teknologi Informasi ITS.
- Purnamasari, F., Ramadijanti, N., & Setiawardhana. (2010). System Online CBIR Menggunakan Identifikasi Dominan Warna Pada Foreground Objek. *Politeknik Elektronika Negeri Surabaya*.
- Rakhmawati, & Puji R. 2013. Sistem Deteksi Bunga Menggunakan Nilai HSV dari Citra Mahkota Bunga. Skripsi. Universitas STIKBANK.
- Smith, J., & Chang, S. (1995). Automated Image Retrieval using Color and Texture. Technical Report Columbia University.
- Suhandy, Diding, & U, A. (2003). Pengembangan Image Processing untuk Menduga kemasakan Buah Manggis Segar. Buletin Keteknikan Pertanian Vol. 17, No. 02.
- Suhendri, Susanti, D., & Mubarak, K. (2020). Identifikasi Kematangan Buah Mangga Gedong Gincu Berdasarkan Warna Menggunakan Metode Content Based Image Retrieval (CBIR). Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika.
- Sukanto, R. A., & Shalahuddin, M. (2011). Modul Pembelajaran Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek). Bandung: Modula.

- Susanto, A. K. (2013). Teori dan Aplikasi Pengolahan Citra. *CV.ANDI OFFSET*.
- Tantri Hidayati Sinaga, E. (2019). Implementasi Teknik Search engine Optimization Dalam Meningkatkan Trafik Website Bima Utomo Waterpark. *Jurnal Sistem Informasi*.
- Wicaksono, T. (2005). Text Mining Untuk Pencarian Dokumen Bahasa Inggris Menggunakan Suffix Tree Clustering. Eepis Final Project.
- Yang, W., Xu, L., Chen, X., Zheng, F., & Liu, Y. (2015). *Chi-Squared Distance Metric Learning for Histogram Data. Mathematical Problems in Engineering Volume 2015*.
- Zitouni, H., Sevil, S., Ozkan, D., & Duygulu, P. (2008). *Re-ranking of Web Image Search Results using a Graph Algorithm. Proceedings of 19th International Conference on Pattern Recognition (ICPR'08) Tampa, FL, USA, 8-11*.