

RANCANG BANGUN MESIN PENCARI CITRA DENGAN PENDEKATAN TEMU BALIK BERBASIS KONTEN DAN MODEL RELEVANCE FEEDBACK

Phie Chyan¹⁾, Sean Coonery Sumarta²⁾

^{1,2}Fakultas Teknologi Informasi, Universitas Atma Jaya Makassar
E-mail: phie_chyan@lecture.uajm.ac.id¹⁾, sean_chiko@yahoo.com²⁾

ABSTRACT

This research aimed at improving the reliability of the system by developing a machine learning framework as relevance feedback mechanism that aims to harness human perception in the training system to learn to map between the vector feature extraction results to the specifications given query. In the developed relevance feedback mechanism, users are involved in providing feedback to the system via the query by example (QBE) interface model. Basically, this module works with the user to decide how to involve the relevant results or irrelevant that emerged as a result of the system output user-defined queries. Then the machine-learning mechanisms will reformulate return query results based on user ratings and displays the new results. This process can take place iteratively until the user is satisfied on the relevance of the results obtained. Outcomes of this study will produce an image search system that can be integrated with information systems of third party services through the modules developed.

Keywords: Content Based Image Retrieval, Histogram, relevance feedback

1. PENDAHULUAN

Perkembangan teknologi mutakhir dalam bidang penyimpanan data dan pemrosesan citra digital memungkinkan proses penciptaan, transmisi, manipulasi dan penyimpanan koleksi dokumen dan citra digital dapat dilakukan dengan cepat dan mudah. Pertumbuhan pesat dari internet juga ikut memberi sumbangsih terhadap membludaknya koleksi citra digital yang tidak terorganisasi dengan baik, Studi menunjukkan pada tahun 2004 saja telah terdapat lebih dari 180 juta berkas citra digital yang berasal dari berbagai web yang diindeks di seluruh dunia dan jumlah ini terus bertambah jutaan berkas setiap harinya [1]. Untuk dapat menangani jumlah data yang sangat besar ini dibutuhkan pengembangan sebuah sistem informasi yang dapat mengelola berkas citra secara efisien. Pencarian citra (*image searching*) merupakan suatu layanan sangat penting yang harus disediakan oleh sistem tersebut. Secara umum terdapat 2 metode berbeda yang dapat diimplementasikan dalam proses pencarian citra pada koleksi berkas citra yang besar: pertama, berbasis informasi tekstual / metadata dari citra dan yang kedua,

menggunakan informasi konten yang terkandung dalam berkas citra.

Metode *retrieval* pada berkas citra dapat dilakukan dengan pendekatan ekstraksi informasi berbasis tekstual / metadata, tapi metode ini sangat tidak efektif karena ukuran basis data citra yang besar dan penilaian subjektif dalam mengasosiasikan hubungan antara teks dan citra. Penamaan atau kata kunci yang terbatas untuk masing-masing referensi citra kurang dapat menjelaskan isi citra secara akurat. Lebih jauh lagi beberapa citra yang menjadi hasil dari *query*, mungkin saja akan sangat berbeda dari citra yang diinginkan pencari berkas [2]. Metode *retrieval* berbasis konten citra merupakan pendekatan lain yang dapat digunakan dalam melakukan temu balik citra. Dalam pendekatan menggunakan metode ini, informasi yang digunakan merupakan informasi visual dari citra yaitu fitur – fitur *low level* dari citra seperti distribusi warna, tekstur dan bentuk (*shape*) yang terkandung dalam suatu citra. Saat ini terdapat banyak aplikasi *search engine* yang digunakan untuk membantu melakukan temu balik data baik yang berupa *desktop searching* untuk melakukan pencarian berkas di komputer

lokal maupun internet searching yaitu *search engine* internet seperti *Google search*, *Yahoo Search*, *Bing* dan banyak lagi untuk mencari informasi dan berkas di internet. Berbagai *Search engine* yang telah disebutkan sebelumnya hanya bekerja melakukan temu balik dengan menggunakan informasi tekstual / metadata sehingga tidak efektif untuk digunakan dalam mencari berkas multimedia termasuk berkas citra.

Mesin Pencari (*Search engine*) yang mampu melakukan temu balik data berkas citra sangat dibutuhkan dalam berbagai bidang yang memanfaatkan basis data citra yang besar seperti galeri seni, desain teknik dan arsitektur, desain interior, penginderaan jarak jauh, manajemen sumber daya bumi, GIS (Geographical Information System), pencitraan medis, prakiraan cuaca, sistem komunikasi hingga ke bidang penegakan hukum dan investigasi kriminal. Oleh karena itu penelitian ini hendak menggali potensi tersebut dengan merancang bangun suatu model mesin pencari yang diperkuat dengan algoritma sistem temu balik citra berbasis konten untuk mendukung proses temu balik berkas citra dengan tingkat akurasi yang lebih baik dibandingkan mesin pencari konvensional yang berbasis informasi tekstual.

Berdasarkan pemaparan sebelumnya, maka perumusan masalah dirumuskan sebagai berikut:

1. Bagaimana meningkatkan akurasi sistem pencari citra dengan melibatkan pengguna dalam menentukan citra yang relevan dan tidak relevan terhadap kriteria citra yang dicari.
2. Bagaimana mengintegrasikan model umpan balik sebagai modul sistem pencari citra untuk dapat digunakan dalam model interface *query by example* (QBE).

2. TINJAUAN PUSTAKA

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 subjektif, sangat tergantung pada persepsi pemberi keterangan. Untuk mengatasi persoalan tersebut maka, 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 [3].

Muatan visual citra dalam basis data diekstrak, kemudian dideskripsikan sebagai vektor ciri (*feature vector*) dan disimpan dalam basis data ciri. Sistem kemudian mengubah contoh citra tersebut ke dalam bentuk vektor ciri dan membandingkan tingkat kemiripannya (*similarity comparison*) dengan vektor-vektor ciri dalam basis data [4]. 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 [5].

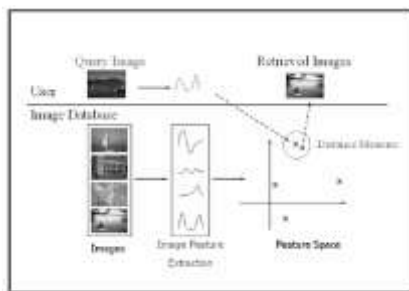
2.1 Arsitektur Search Engine Dengan Pendekatan Temu Balik Berbasis Konten

Gambar 1 menunjukkan arsitektur umum dari sebuah *search engine* dengan pendekatan temu balik berbasis konten [6]. Terdapat 2 fungsionalitas utama yang didukung: *data insertion* dan *query processing*

Subsistem *data insertion* berfungsi untuk mengekstrak fitur-fitur tertentu dari data citra (*image feature Extraction*) dan menyimpannya pada basis data citra. Proses ini umumnya dilakukan secara *off-line*.

Query Processing, diorganisasikan sebagai berikut : antarmuka memungkinkan pengguna untuk merincikan *query* dengan beberapa metode misalnya memvisualisasi citra yang sama baik menggunakan citra yang

lain atau sketsa (*Query By Example*), metode lain bisa dengan *user-defined attribute query* dan *feature combination query*, misalnya seorang pengguna bisa melakukan *query* ke sistem dengan statemen “Tampilkan Citra berwarna biru dengan tekstur bergaris dimana kedua attribut sama bobotnya”. Modul *query processing* mengekstrak vektor fitur dari pola *query* dan mengaplikasikan suatu formula *distance measure* (misalnya Jarak Euclidean) untuk mengevaluasi kemiripan diantara citra *query* dan citra pada koleksi basis data. Kemudian, citra dari koleksi basis data diranking secara berurutan berdasarkan kemiripannya dengan citra *query*. Citra yang memiliki ranking teratas adalah citra yang memiliki jarak yang terkecil dan menjadi citra yang *retrieve* oleh sistem.



Gambar 1 Arsitektur Umum Model Temu Balik Berbasis Konten

2.2 Interaksi Pengguna Dalam Search Engine Dengan Temu Balik Berbasis Konten

Dari perspektif pengguna, *search engine* dengan temu balik berbasis konten menawarkan fleksibilitas yang lebih baik dibandingkan yang berbasis informasi tekstual. Tetapi, di sisi lain, hal ini menghadirkan banyak tantangan. Pertama, bagaimana mendukung pengguna dalam proses mendeskripsikan / merincikan *query*. Masalah lain yaitu *information overload* – bagaimana sistem dapat memberikan hasil yang berguna. Isu ketiga adalah bagaimana memberikan pengguna suatu tools atau kakas bantu untuk berinteraksi dengan sistem dengan tujuan untuk membantu pengguna dalam mendeskripsikan *query* nya sebagai masukan bagi sistem.

2.3 Spesifikasi Query

Beberapa mekanisme *query* telah diciptakan untuk mendukung pengguna mendefinisikan kebutuhan informasi mereka. Ada beberapa strategi pendefinisian *query* yang memungkinkan dalam *search engine* dengan temu balik konten seperti yang dikemukakan oleh [7] seperti *simple visual feature query*, *feature combination query*, *localized feature query*, *query by example*, *object relationship query* dan *concept query*.

2.4 Visualisasi Hasil

Teknik presentasi hasil yang paling umum adalah menggunakan *thumbnail* citra berbasis 2D. *Thumbnail* merupakan bentuk miniatur dari citra yang diorganisasi dalam bentuk petak / *grid*. Selain metode tersebut ada beberapa metode lain, diantaranya yang dihasilkan oleh [8] yang menjabarkan sistem eksplorasi terhadap basis data citra dengan kakas untuk visualisasi basis data pada tingkat detail yang berbeda berdasarkan teknik *scaling* multidimensional dan juga oleh [9] yang mempresentasikan 2 teknik visualisasi berbasis spiral dan *concentric rings* untuk mengeksplorasi hasil. Struktur visual ini dipusatkan untuk menjaga fokus pengguna pada citra *query* dan hasil citra temu balik yang paling mirip.

2.5 Relevance Feedback

Relevance Feedback (RF) merupakan metode atau model yang digunakan untuk meningkatkan efektifitas dari temu balik secara interaktif berdasarkan umpan balik yang diberikan user terhadap sistem. Pada dasarnya terdiri dari 3 tahap : (a) pencarian awal dibuat oleh sistem untuk pola *query* yang dimasukkan pengguna, menghasilkan sejumlah kecil citra kembalian; (b) pengguna kemudian berperan untuk menunjuk citra hasil temu balik yang berguna atau relevan; (c) selanjutnya yang terakhir dalam model ini, sistem secara otomatis menformulasi ulang *query* asli berdasarkan atas penilaian relevansi yang dilakukan oleh pengguna. Proses ini dapat berlangsung secara iteratif sesuai keinginan pengguna hingga pengguna puas terhadap hasil yang diberikan. Strategi *Relevance feedback* ini berguna untuk dapat mereduksi masalah gap semantik karena memungkinkan sistem temu balik untuk

belajar mengenai persepsi pengguna terhadap suatu citra. Strategi RF biasanya berhubungan dengan data latih asimetri dan kebutuhan waktu nyata [10]. Beberapa isu penting lainnya berkaitan dengan desain implementasi dari mekanisme pembelajaran. Strategi yang paling umum diimplementasikan adalah menggunakan *weight-based learning approaches*, Algoritma Genetika dan *Support vector machines*.

3. METODOLOGI PENELITIAN

3.1 Studi Literatur

Studi literatur dilakukan untuk mendapatkan dasar teoritis dan juga metode analisis terkini agar diperoleh hasil yang sesuai dengan perkembangan dalam bidang ilmu teknologi informasi dan multimedia. Kegiatan yang dilakukan pada tahapan ini adalah mengumpulkan dan mempelajari artikel, jurnal dan referensi lain nya yang mutakhir berkaitan dengan image processing dan metode - metode didalam teknik Content Based Image Retrieval (CBIR).

3.2 Pengumpulan Data

Mengumpulkan data yang berkenaan dengan data pendukung dalam proses implementasi aplikasi image retrieval yang akan dijalankan. Data yang dikumpulkan mencakup data gambar/citra random yang dikumpulkan dari website yang memiliki koleksi citra dengan bantuan search engine internet dan juga citra yang dikumpulkan langsung melalui hasil capture dari kamera digital. Data citra yang terkumpul akan disimpan dalam basis data untuk digunakan dalam simulasi proses temu balik citra

3.3 Tahapan Penelitian

Dalam mencapai tujuan penelitian ini, proses tahapan penelitian akan difokuskan pada rancang bangun *search engine* memanfaatkan algoritma temu balik berbasis konten dan merancang mekanisme *query processing* dari sistem. Kemudian PADA tahap berikutnya penelitian difokuskan untuk merancang bangun suatu modul *relevance feedback* yang akan diintegrasikan pada sistem yang telah dihasilkan dan melakukan implementasi sistem secara menyeluruh kemudian melakukan analisis dan evaluasi

terhadap unjuk kerja sistem. Rincian kegiatan pada masing – masing tahap akan dibahas berikut ini

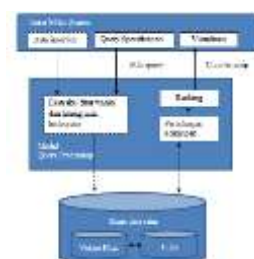
3.4 Perancangan Sistem

Tujuan yang ingin dicapai pada kegiatan penelitian ini adalah untuk menghasilkan sistem *search engine* dengan pendekatan temu balik berbasis konten dengan menerapkan model *relevance feedback* bagi pengguna untuk memberi umpan balik terhadap sistem. Untuk mewujudkan hal tersebut maka perancangan sistem dititikberatkan pada rancang bangun 2 fungsionalitas utama sistem yaitu *data insertion* dan *query processing*

Subsistem *data insertion* berfungsi untuk mengekstrak fitur- fitur tertentu dari data citra pada basis data. Pendekatan yang digunakan untuk mengekstrak fitur dari data citra adalah menggunakan algoritma ekstraksi fitur warna dan perhitungan nilai histogram yang telah diteliti pada studi pendahuluan untuk penelitian ini.

Subsistem *Query Processing* untuk penelitian ini akan dirancang menggunakan model antar muka *Query By Example*. Antar muka ini memungkinkan pengguna untuk merincikan *query* dengan memvisualisasi citra yang sama dengan menggunakan berkas citra tertentu sebagai kunci atau menggambar suatu sketsa dan menginputkan ke sistem.

Modul *query processing* kemudian akan digunakan untuk mengekstrak vektor fitur warna dari pola *query* dan menghitung jarak histogram untuk mengevaluasi kemiripan distribusi warna diantara citra *query* dan citra pada koleksi basis data menggunakan formula jarak Euclidean dan jarak Laplace. Kemudian selanjutnya citra dari koleksi basis data diranking secara berurutan berdasarkan kemiripannya dengan citra *query* dari pengguna dan divisualisasikan kepada pengguna. Gambar 2 memperlihatkan rancangan arsitektur sistem *image search engine* yang akan dihasilkan.



Gambar 2 Rancangan Arsitekur sistem *Image Search Engine*

4. HASIL DAN PEMBAHASAN

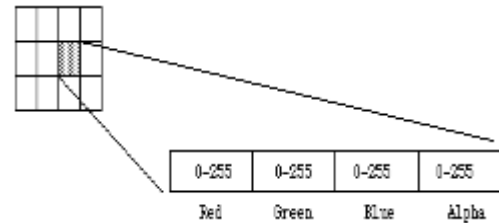
4.1 Analisa Sistem

Sistem dirancang untuk mampu mencari citra yang sama atau memiliki kemiripan berdasarkan citra yang diberikan sebagai masukan. Proses temu balik citra didasarkan pada jarak histogram dengan fitur warna YIQ. Dalam proses temu balik citra semua koleksi citra yang tersimpan dalam database akan diekstrak ciri-ciri low level nya kedalam basis data, untuk proses retrieve citra masukan, yaitu citra yang dicari berfungsi sebagai query yang kemudian akan dibandingkan dengan koleksi citra yang ada dalam basis data. Dalam prosesnya citra masukan akan diekstrak informasi low-level nya untuk didapatkan feature vector yang berguna dalam proses perbandingan kemiripan (*similarity comparison*) terhadap koleksi citra dalam basis data. 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. Prosedur dalam melakukan temu balik dilakukan dengan menghitung nilai komponen Red, Green, Blue (RGB) dari piksel citra selanjutnya dilakukan konversi dari RGB ke model warna YIQ, dari hasil kuantisasi nilai YIQ dibentuk histogram citra yang kemudian digunakan dalam perhitungan jarak histogram untuk perbandingan kemiripan antara citra. Selanjutnya sistem akan menampilkan image dalam koleksi basis data yang memiliki kemiripan (*similarity*) yang paling tinggi dengan image masukan yang digunakan sebagai query, informasi yang ditampilkan adalah jumlah piksel yang serupa beserta dengan persentase kemiripannya.

4.2 Pemrosesan Citra

Dalam citra tidak terkompresi, piksel citra umumnya disimpan dengan kedalaman warna tertentu misalnya 1, 4, 8, 16, 24, 32, 48, atau 64 bits per piksel. Piksel yang terdiri dari 8 bit atau lebih kecil dapat direpresentasikan dengan *greyscale* (skala

abu-abu) atau warna terindeks. *Alpha channel* (transparansi) yang disimpan pada bitmap terpisah atau sebagai *channel* keempat selain tiga *channel* warna dasar yaitu Red, Green, Blue (RGB) menyebabkan konversi citra 24 bit menjadi 32 bit. Setiap elemen data image bitmap 24 bit panjangnya 3 byte untuk menyatakan elemen RGB ditambah 8 bit sebagai alpha channel yang merepresentasikan informasi transparansi.

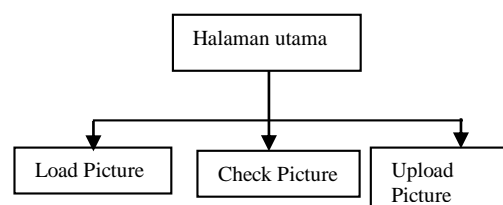


Gambar 3. Format Citra Bitmap

Pada citra bitmap 24 bit, tiap piksel direpresentasikan dengan masing-masing 8 bit warna dasar (RGB), dengan nilai intensitas antara 0 hingga 255 desimal untuk tiap komponen warna tersebut seperti terlihat pada gambar 3.

4.3 Struktur Implementasi Sistem

Sistem dirancang menggunakan antar muka bagi pengguna. Perancangan struktur menu program ini membantu dalam merancang bagian-bagian dari sistem yang sebenarnya dan untuk mengetahui bagian mana yang terlebih dahulu yang akan diakses setelah program tersebut selesai. Struktur menu padaprogram ini ditunjukkan pada gambar 4 berikut.



Gambar 4 Struktur Menu Aplikasi

4.4 Pengujian Sistem

Pada pengujian ini akan dilakukan *image retrieval* terhadap gambar *plane1.bmp*. Untuk proses *image retrieval* gambar yang ingin di cari kemiripannya dengan yang ada pada database pada menu klik Load Picture pada halaman utama dan pilih image

plane1.bmp maka kemudian gambar tersebut akan tampil di *picture box*, selanjutnya diklik tombol check picture untuk memulai proses image retrieval pada gambar tersebut, maka sistem akan mulai mencari gambar yang mempunyai kemiripan paling tinggi dalam koleksi gambar di basis data. Hasil yang diperoleh menampilkan gambar plane2.bmp sebagai gambar yang memiliki kemiripan paling tinggi dengan gambar plane1.bmp. Informasi mengenai hasil image retrieval dapat dilihat pada panel di sebelah kanan. Seperti yang terlihat bahwa kedua gambar memiliki kemiripan sebesar 73% sebagaimana terlihat pada gambar 5.



Gambar 5. Hasil pencarian citra

4.5 Pembahasan

Berdasarkan pengujian sistem yang dilakukan, dilakukan pembahasan terhadap hasil-hasil tersebut. Adapun hasil pembahasan difokuskan pada kelebihan dan kelemahan dari sistem yang dihasilkan

a) Kelebihan Sistem

Adapun kelebihan dari sistem yang dihasilkan ini adalah sebagai berikut :

- a. Dengan menyimpan gambar dalam bentuk nilai biner, memungkinkan proses *image retrieval* menjadi lebih cepat karena gambar tidak perlu di-load ulang pada saat akan diambil nilai histogramnya.
- b. Dengan mengukur jarak histogram antara dua buah gambar, dapat dilihat persentase kemiripan kedua gambar tersebut. Hasil pengukuran dengan menggunakan jarak histogram ini dapat menampilkan jumlah pixel yang sama dan berbeda dari kedua gambar yang dibandingkan.

b) Kekurangan Sistem

Sistem yang dikembangkan tidak dapat memproses citra yang mempunyai ukuran lebih besar dari 250 X 250 *pixel*. Karena dengan ukuran citra yang besar membutuhkan waktu yang lebih lama dalam proses komputasi nilai histogram dan perhitungan kemiripannya dengan koleksi citra pada basisdata

5. KESIMPULAN

Berdasarkan hasil penelitian sistem temu balik citra berbasis isi citra menggunakan fitur warna dan jarak histogram, dapat disimpulkan beberapa hal sebagai berikut:

1. fitur low level citra dapat digunakan dalam proses temu balik citra berbasis konten dengan menggunakan warna sebagai vektor yang diekstraksi.
2. Sistem image search engine menggunakan umpan balik dari pengguna sebagai model relevance feedback untuk menentukan citra yang paling relevan diantara citra yang mempunyai presentasi kemiripan paling tinggi berdasarkan hasil temu balik sistem.

6. DAFTAR PUSTAKA

- [1] Ozbay, O.C., 2004, An Image Retrieval System Based On Region Classification, Thesis, Graduate School Of Natural And Applied Science, Middle East Technical University.
- [2] [2]Sardar, M., and Basuli, K., 2008, Content-Based Image Retrieval System, Thesis, Department of Computer Science and Engineering, University of Calcutta, India.
- [3] Rahman, A., 2009, Sistem Temu Balik Citra Menggunakan Jarak Histogram. Prosiding Seminar Nasional Aplikasi Teknologi Informasi (SNATI) 2009, hal : I 58-61.
- [4] Kusriani, Harjoko, A., 2009, Pencarian Citra Visual Berbasis Isi Citra Menggunakan Fitur Warna

Citra, Jurnal Ilmiah DASI Vol 9 No. 2 2008, hal: 90 – 106.

- [5] Long, L., Thoma, G., Antani, S., 2004, A Prototype Content-Based Image Retrieval System For Spine X-Rays, CBMS'03 Proceeding Of The 16th IEEE Conference On Computer-Based Medical Systems page:156-162
- [6] Torres, R., and Falcao, A., 2006, Content-Based Image Retrieval : Theory and Applications, Revista de Informática Teórica e Aplicada Journal Vol.13 No.2 2006 page 161-185.
- [7] Y. A. Aslandogan and Yu, C., 2000. Techniques and Systems for Image and Video Retrieval. IEEE Transactions on Knowledge and Data Engineering, 11(1) Page:56–63.
- [8] Stan, D., and I. K. Sethi, 2003, A System for Exploration of Image Databases, Information Processing and Management, 39(3):335–365.
- [9] Koradiya, J.T.,and Swadas, P., 2013, Content Based Image Retrieval, International Journal of Advanced and Innovative Research (IJAIR) Vol.2 No.4 Page: 1324-1329.
- [10] Zhou, X.S and Huang, S.T, 2003, Relevance Feedback in Image Retrieval: A