

類似領域特徴量を用いた類似画像検索手法 A New CBIR Approach Using Similar Region Feature

呉宇星[†]
Yuxing Wu

六井淳[†]
Jun Rokui

1. まえがき

画像検索の研究では, CBIR(Content-based Image Retrieval) 内容に基づく画像検索の研究が 30 年ほど前から行われている [9].

TBIR(Text-based Image Retrieval) は文書に基づく画像検索で, 画像データの文書情報(画像名, タグなど)を検索に用いる. TBIR に使われる画像には人手によるタグ付けが必要であり, 飛躍的に増える画像数に対応できなくなっている. この問題を回避する方法の一つが CBIR である.

CBIR は画像の視覚情報を検索に用いることで, 検索の自動処理を可能にした. CBIR の研究では画像の色, 形状, 模様などの視覚特徴量を検索に用いる. 早期では 1995 年の QBIC[1] や, 1996 年の VisualSEEK[3], 日本では 1999 年の ExSight[4] などがある. 現在では画像から得られる視覚的情報を階層的に分析し, 低階層特徴量からより高階層の特徴量を求める手法が提案されている [7], [10]. また, 検索手法の研究でも進展があり, 従来の特徴量をそのまま用いて検索する手法以外に, 内容から文書を推定する手法, フィードバック手法などの研究が行われている [6], [8].

現在実用化されている CBIR エンジンの一つとして Google 画像検索[‡]が挙げられる. 検索画像からタグを推定し, タグと視覚情報両方を用いて検索する手法である. オブジェクト単一, 背景一色な簡単画像のタグ推定および検索では精度は若干高いが, 多オブジェクトを含む複雑画像のタグ推定および検索精度では極めて悪い. 数少ないフィードバック式 CBIR エンジンとして pixolu[§]がある. これはタグ使用が前提で, 画像を直接入力して検索することはできない. また, 文書不用の CBIR エンジンとして, TinEye[¶]がある. このエンジンは検索画像と極めて類似する画像しか出力しないため, 用途に限られる. そのほかにも実用化された検索エンジンはいくつもあるが, ほとんどが内容と文書両方を用いた手法である. 文書不用の画像検索は, 「用途に限られる」, 「精度が悪い」, 「計算が遅い」などの理由で実用化できていないのが現状である. 検索精度が低い原因として, 画像にユーザーにとっての不要情報が入っていることが考えられる. この問題に対し, ユーザーが画像内注目部分を多角形を用いて切り出す手法が提案されている [2]. この手法は自由な切り出しではないため, 不要情報は存在する. 自由な切り出しはユーザーの負担を増加させるため, 望ましい方法とは言えない.

精度が低い原因としてもう一つ考えられるのが, 有効な特徴量が分からないことである. 一枚の検索画像による類似画像検索では, 計算機にユーザーの意図を理解させることはむずかしい. 本論文では二枚の検索画像から類似領域特徴量を推定して類似画像検索を行う手法を提案する.

2. 提案法

本論文では, 類似オブジェクトを含んだ画像を類似画像とし, 提案する類似画像検索手法は類似オブジェクトを含む画像を検索することを目的とする. 類似オブジェクトは同一カテゴリに属するオブジェクトとする.

2.1. アルゴリズム

現在の類似画像検索手法の研究ではほとんどが 1 枚だけの画像で検索することを前提としている. 複数枚を用いるのに比べて, 1 枚だけを用いると処理時間が短く, かつ操作がシンプルで実用的観点からみても妥当である. しかしそれだけで計算機にユーザーの必要情報を判断させるのはむずかしい. 色の類似, 形の類似, 用途の類似, 画像内分布の類似など, 人にとっての類似は多種多様に存在する. 一方, 機械学習では大量の学習データと処理を必要とし, 大量画像に対して高速処理を必要とする検索とは相性が悪い. そこで提案手法は二枚の検索画像を用いる事で問題の解決を試みる.

提案法の処理の流れを図 1 に示す.

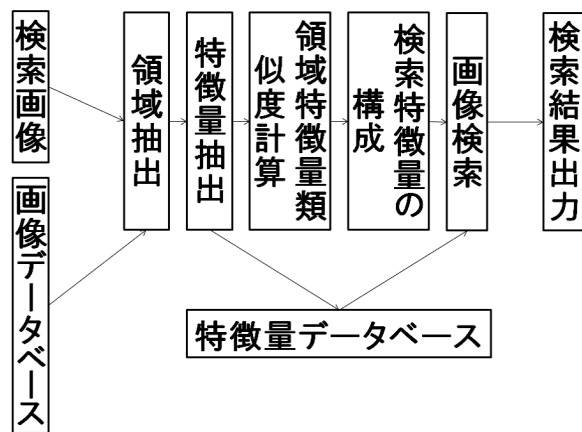


図 1: 提案法の処理の流れ

以下に提案法の手順を説明する.

STEP1 ユーザーは検索エンジンに二枚の検索画像を入力する. そのうち一枚を主検索画像とする.

STEP2 検索画像に対し領域分割を行う.

[†] 島根大学大学院総合理工学研究科数理・情報システム学科, Department of Mathematics and Computer Science, Interdisciplinary Graduate School of Science and Engineering, Shimane University

[‡] <http://www.google.com/>

[§] <http://www.pixolu.de/>

[¶] <http://www.tineye.com/>

STEP3 領域から特徴量を抽出する。

STEP4 検索画像間の領域および特徴量の類似度を計算し、類似度ベクトルを作成。

STEP5 類似度ベクトルから検索用特徴量マトリックスを構成する。

STEP6 検索用特徴量マトリックスを用いて画像データベースから類似画像を検索する。

STEP2 では検索画像に対して領域分割を行う。検索画像に対して領域分割を行い、問題をユーザーにとっての類似部分から画像の類似領域に置き換える。

STEP3 では抽出された領域から特徴量を抽出する。あらゆる可能な特徴量を抽出することが理想であるが扱う特徴量が増えれば計算時間も増える。本論文で扱う特徴量は色情報と形状情報である。

STEP4 では検索画像間の領域特徴量の類似度を計算。主検索画像の個々の領域に対して副検索画像内すべての領域と類似度計算を行い、類似度ベクトルを得る。主検索画像内領域の類似度ベクトルの構造を表 1 に示す。

ABS	ABSa	ABSb	...
ACS	ACSa	ACsb	...
...	...	...	...

表 1: 主検索画像内領域の類似度ベクトル

A は主検索画像内の領域を表す。B,C,... は副検索画像内の領域を表す。S は類似度を表す。a,b,... は特徴量を表す。例えば、ABS は主検索画像領域 A と副検索画像領域 B の類似度、ACSa は主検索画像領域 A と副検索画像領域 C の特徴量 a に関する類似度。

主検索画像領域 A と副検索画像領域 B の類似度 ABS の計算式を式 (1) に示す。

$$ABS = \sum_{a \in P} Ra * ABSa \quad (1)$$

Ra は特徴量 a の重みであり、あらかじめ与えておく。すべての重み和を 1 とする。ABSa は領域 A と B の特徴量 a に関する類似度。P は全特徴量の集合。得られたベクトル群から主検索画像領域の最大類似度を求める。

主検索画像領域 A と副検索画像の最大類似度の計算式を式 (2) に示す。

$$AS = \max_{B \in Q} (ABS) \quad (2)$$

B は副検索画像領域の一つ。Q は副検索画像領域の集合。

STEP5 では主検索画像領域の類似度ベクトルおよび特徴量を用いて検索用特徴量マトリックスを構成する。検索用特徴量マトリックスは主検索画像複数領域の類似ベクトルと特徴量の合成である。時間削減のため、類

似ベクトルと特徴量は主検索画像領域類似度 (AS 等) の高い順に最大 100 個までを用いた。検索用特徴量マトリックスの構造を表 2 に示す。

AS	ASa	ASb	...	Aa1	...	Ab1	...
BS	BSa	BSb	...	Ba1	...	Bb1	...
...	...	...	...	...	...	...	...

表 2: 検索用特徴量マトリックス

A,B は主検索画像内の領域を表す。S は類似度を表す。a,b は特徴量を表す。例えば Ab1 は主検索画像領域 A の特徴ベクトル b の一番目の要素、AS は主検索画像 A の類似度を表している。ASa は STEP4 で得られた主検索画像領域 A と副検索画像の特徴量 i の類似度。例えば STEP4 で AS を求めたとき主検索画像領域 A は副検索画像領域 K との類似度が副検索画像領域内で最大とすると、AS は AKS の値で、ASa は AKSa、ASb は AKSb の値となる。

STEP6 では検索用特徴量マトリックスを用いてデータベース内の画像を検索する。

検索用特徴量ベクトル A と被検索画像との類似度 AS' を式 (3) で求める。

$$AS' = \max_{B \in Q} \frac{\sum_{a \in P} Ra * ASa * ABSa}{\sum_{a \in P} ASa} \quad (3)$$

Ra は特徴量 a の重み、式 1 と同じである。Q は被検索画像領域の集合。ASa は STEP4 で得られた主検索画像領域 A と副検索画像の特徴量 a の類似度。ABSa は主検索画像領域 A と被検索画像領域 B の特徴量 a の類似度。P は全特徴量の集合。

主検索画像と被検索画像の類似度を式 (4) で求める。

$$S = \sum_{A \in Q} \frac{AS * AS'}{\sum_{A \in Q} AS} \quad (4)$$

Q は主検索画像領域の集合。最終的に得られた類似度 S を大きい順から対応する画像を出力する。

ここまでは提案法のアルゴリズムである。続いて提案法の実装について具体的に説明する。

2.2. 代表色を用いた領域分割

提案法の STEP1 では画像に対して領域分割を行う。本論文では代表色を用いた領域分割を行った。用いた色の表現法は rgb である。

実装では、代表色を求める前にまず 255 階調の rgb データを均等量子化を用い 8 階調まで減色処理を行った。代表色の求め方の手順を以下に示す。

- 1 画像の rgb 情報からカラーヒストグラム C1[8][8][8] を作成。
- 2 要素が 0 のカラーヒストグラム C2[8][8][8] を作成。

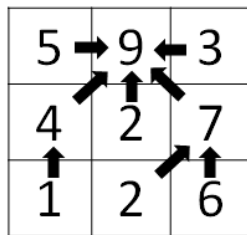
3 カラーヒストグラム $C1[8][8][8]$ 内の値が 0 でない個々のブロック $C1[R][G][B]$ に対し、 $C1[R][G][B]$ および $C1[R][G][B]$ と 26 隣接するブロック内で最も高い値を持つ $C1[R][G][B]$ を探す。

4 カラーヒストグラム $C2[8][8][8]$ に対し、 $C2[R][G][B] = C2[R][G][B] + C1[R][G][B]$ と計算する。同時に画像内の色データに対し、 $C1[R][G][B]$ が対応する色 RGB を探し出して $C1[R][G][B]$ が対応する色 $R'G'B'$ に書き換える。

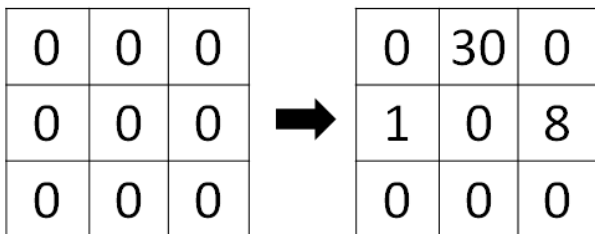
5 $C1[8][8][8]$ すべてのブロックに手順の 3,4 が行われた後、 $C1[8][8][8]$ と $C2[8][8][8]$ が同じなら終了。 $C1[8][8][8]$ と $C2[8][8][8]$ が同じでないなら、 $C2[8][8][8]$ を $C1[8][8][8]$ とし、手順 2 に戻る。

本論文の rgb データは 3 チャンネル 8 階調であるため、手順 1 と手順 2 で作成されたカラーヒストグラムのブロック数は $8 \times 8 \times 8 = 512$ 個である。

全手順を終え、カラーヒストグラム $C2[8][8][8]$ 内で値が 0 でないブロックが存在する。これらのブロックが対応する色が代表色である。



$C1[8][8][8]$ 内



$C2[8][8][8]$ 内の計算

図 2: 2 チャンネル内の計算

図 2 は 2 チャンネルを見たときの計算例である。図 2 上部分は $C1[8][8][8]$ 内一部的の様子で、矢印はそれぞれのブロックから周囲 3×3 範囲内で最も高い値を持つブロックを指している。中央ブロック ($C1[R][G][B]$ に相当) の値は 2 で、中央ブロックを中心とする周囲 3×3 ブロック内の最大値は上ブロック ($C1[R][G][B]$ に相当) の 9 であるので、カラーヒストグラム $C2$ 内の上ブロック ($C2[R][G][B]$ に相当) に 2 が加算される。全部のブロックに手順 3,4 が行われた時、図 2 左下部分の全要素が 0 の $C2[8][8][8]$ は図 2 右下部分のようになる。

本論文の代表色を求める手法で新しいカラーヒストグラムが作成された時、画像データも代表色で塗り替え

られる。この代表色データを用いて画像データの領域分割を行う。領域分割の手順を以下に示す。

I 画像データの個々のセルに対して唯一のラベルを付ける。

II 画像内全部のセルを探索し、4 隣接のセルで同じ色があればラベルを統一する。

III ラベルの統一が起こらなくなるまで、手順 II を繰り返す。

領域分割の処理が終われば、同じ領域のマスに同じラベルが付けられる。

2.3. 領域特徴量抽出と類似度計算

本論文の実験で使用された領域特徴量について説明する。

• 色特徴量 (rgb)

色特徴量は、領域内の元 RGB データの平均値を用いた、要素数 3。

• 形状特徴量 1 (sharp1)

重心点から上下左右 45 度斜め計 8 方向の領域輪郭までの距離、要素数 8。

• 形状特徴量 2 (sharp2)

重心点の値と縦軸、横軸の最大値、最小値の差、要素数 4。

重心点は領域内マス縦軸、横軸の平均値を取る。

本論文で用いた類似度計算式はコサイン類似度である。コサイン類似度の計算式を式 (5) に示す。

$$s = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} * \sqrt{\sum_{i=1}^n B_i^2}} \quad (5)$$

s はコサイン類似度、 A_i, B_i は特徴ベクトル A, B の i 番目の値、 n は特徴ベクトルの長さである。

rgb 特徴量に対してそのままコサイン類似度を計算する。sharp1 と sharp2 特徴量は対称、回転を考慮して類似度を求め、最も大きい値を出力する。

3. 評価実験

3.1. 実験条件

実験環境は表 3 に示す通りである。

CPU	Intel Core i5-2400 3.10GHz
OS	Windows 7 Professional(SP1)(64bit)
メモリ	8GB
プログラミング言語	C
コンパイラ	cygwin gcc
データベース	Caltech256 の前 10 カテゴリ

表 3: 実験環境

カテゴリ番号	カテゴリ名	枚数	割合 (%)
001	ak47	97	7.8
002	american-flag	96	7.8
003	backpack	151	12.2
004	baseball-bat	127	10.3
005	baseball-glove	148	12.0
006	basketball-hoop	85	6.9
007	bat	105	8.5
008	bathtub	231	18.7
009	bear	102	8.3
010	beer-mug	94	7.6

表 4: データベース詳細

データベース Caltech256^{||}の前 10 ジャンル,001 から 010 までの計 1236 枚の画像をデータベースとして用いた。データベースの詳細を表 4 に示す。

割合は個々のカテゴリ枚数を全カテゴリ枚数である 1236 で除算したものである。

3.2. 実験方法

実験では複数回の検索を行い、出力結果を統計し、正解率を用いて提案法を評価する。提案法と特徴量の有効性を見るため、表 5 に示された 4 セットで実験を行った。なお、rs1s2 を用いたときの各特徴量の重みは均等とした。

セット名	用いた特徴量
rs1s2	rgb+sharp1+sharp2
r	rgb のみ
s1	sharp1 のみ
s2	sharp2 のみ

表 5: 実験に用いた特徴量セット

検索手順は次の通りである。

- 1 カテゴリーから二枚の画像を取り出し、一枚を主検索画像とし、もう一枚を副検索画像とする。
- 2 主検索画像と副検索画像で領域特徴量の類似度判定を行い、検索用特徴量マトリックスを作成。
- 3 検索用特徴量マトリックスを用いてデータベース内で検索を行い、類似度の大きい順に上位 51 枚の結果を出力する。

検索画像は各カテゴリ画像番号 0001 から 0005 までの 5 枚*10 カテゴリーを用いた。5 枚から 1 枚を選んで主検索画像とし、もう 1 枚を選んで副検索画像とする。よって 1 セットに対して実験した検索回数は $5*4*10$ の計 200 回。評価に用いた検索結果の出力画像枚数は $200*50 = 10000$ 枚である。1 回の検索で 1 位から 51 位までの結果が出力されるが、1 位は常に主検索画像で

あるため実験結果の統計には入れず、検索結果の 2 位から 51 位までの計 50 枚を用いて統計し正解率で評価する。実験結果で述べる上位 n 枚は検索結果の 2 位から数えたものとなる。

本実験での正解率を式 (6) に示す。

$$\text{正解率} = \frac{\text{検索結果中正解枚数}}{\text{検索結果枚数}} \quad (6)$$

3.3. 実験結果

検索結果をセット別に上位 1 枚, 上位 10 枚, 上位 50 枚を統計した時の正解率と 1 回の検索で得た正解数の最大値を表 6 に示す。

セット名	T1(%)	M1	T10(%)	M10	T50(%)	M50
rs1s2	22.0	1	19.5	6	15.9	20
r	18.0	1	17.3	5	13.4	21
s1	15.5	1	13.6	5	12.6	13
s2	18.0	1	11.8	5	11.3	15

表 6: セット別上位 1 枚, 上位 10 枚, 上位 50 枚の正解率と最大正解数

T1,T10,T50, はそれぞれ上位 1 枚, 上位 10 枚, 上位 50 枚を統計した時の正解率。M1,M10,M50 はそれぞれ上位 1 枚, 上位 10 枚, 上位 50 枚で 1 回の検索で得た正解枚数の最大値。例えば、rs1s2 の T1 は 22.0%で、特徴量に rgb+sharp1+sharp2 を用いた 200 回の検索で出力結果の 1 枚目だけを統計した時の正解率が 22.0%であることを表し、r の M50 は特徴量 rgb だけを用いた 200 回の検索で上位 50 枚中最大正解枚数が 21 枚であることを表している。

セット rs1s2 の正解率は r,s1,s2 を単独で用いた時より高い値を出している。これは提案法における類似領域特徴量を用いた領域特徴量の重み推定が有効であることを示している。また、r の M50 は 21 枚で、rs1s2 の 20 枚を上回る結果となった。提案法では検索画像領域間の特徴量類似度を重みとして用い不要特徴量による影響の軽減を図っている。しかし不要特徴量は検索時の計算で用いられ、rs1s2 の結果を悪化させたと考えられる。

セット、カテゴリ別の上位 50 枚の正解率と一回の検索での最大正解数を表 7 に示す。

rs1s2,r,s1,s2 はそれぞれの特徴量セットを用いたときの上位 50 枚の正解率。M は検索結果の上位 50 枚で得た正解枚数の最大値。特徴量が有効であることを示すには正解率が少なくとも各カテゴリが占める割合の値を上回る必要がある。表 7 で正解率の前に*が付いているものは各カテゴリが占める割合の値を下回るものである。

カテゴリ baseball-bat と baseball-glove のオブジェクト形状はほぼ単一であり、形状特徴量の有効性を否定する結果となっている。原因の一つとして考えられるのは領域分割手法の問題である。本論文の領域分割手法では色情報を用いるため撮影環境に影響されやすく、単一オブジェクトに複数の色領域が含まれると複数の

^{||}http://www.vision.caltech.edu/Image_Datasets/Caltech256/

番号	rs1s2(%)	M	r(%)	M	s1(%)	M	s2(%)	M
001	11.8	12	*5.2	4	9.4	11	*6.3	7
002	25.3	18	24.1	21	12.0	8	*6.4	5
003	12.6	15	13.8	11	13.7	9	12.8	11
004	11.5	10	14.4	9	*9.5	7	10.8	9
005	19.8	20	14.0	13	12.4	8	*11.3	8
006	8.6	6	8.1	7	10.4	8	10.2	6
007	21.5	17	10.0	7	15.6	13	11.2	11
008	*16.2	12	*13.6	14	*18.6	13	19.8	15
009	20.7	17	20.8	17	15.5	12	12.7	11
010	11.3	9	10.0	9	9.3	9	11.2	6

表 7: セット, カテゴリ別の上位 50 枚の正解率と最大正解数

領域として分離してしまう現象が多く確認された。また、画像内オブジェクトどうしの色が類似すると一つの領域として抽出する問題も多数確認された。更に本論文で用いた形状特徴量の問題である。sharp1 と sharp2 は rgb に劣る結果となったことから見て、形状特徴量に問題があると考えられる。

4. まとめ

本論文では二枚の画像を用いた類似画像検索を試みた。二枚の検索画像を領域分割してから特徴量を抽出して類似度を求め、得られた類似度を検索時の領域特徴量の重みとして使用し、ユーザーが注目する類似領域特徴量の推定および検索を行った。実験では、複数の特徴量セットを用い、検索結果を統計して提案法を評価した。結果、複数特徴量を用いた時の正解率は全体的に個々の特徴量を単独で用いるときの正解率より高く、提案法の趣旨であるユーザーが注目する類似領域特徴量の推定は有効であることが示された。一部の検索結果で複数特徴量を用いた時の正解率が単独特徴量を用いた時の正解率を下回る結果となり、不要情報の影響問題が判明した。カテゴリ別に正解率を出し、各カテゴリに対する特徴量の性能を検証した。結果、カテゴリ別の特徴量有効性が一部判明し、領域分割手法の問題、特徴量の問題が判明した。

今後の課題の一つは不要情報の切り捨てである。検索画像間のそれぞれの特徴量類似度に対して閾値を設定し、閾値を下回る類似度を出した特徴量は不要情報として切り捨て、検索には用いないよう改良する必要がある。もう一つの課題として領域分割手法の見直しがある。本論文の代表色を用いた領域分割手法では画像データから代表色を求めて、得られた代表色で画像を書き換えて領域を抽出するが、色情報は撮影環境、光源などに影響されやすく、領域分割の結果で過分割や分割不足が多く見られた。この問題に対して、模様特徴量を用いた領域分割手法で解決を図ることができると考えられる。三つ目の課題は特徴量の追加である。本論文で用いた特徴量はどれも要素数が少なく、低階層のものである。ユーザーが注目する類似がわからないゆえ、より多く、より高階層の特徴量を取り入れる必要がある。検索時間短縮のために、提案法に用いる多次元インデックスの構築も今後の課題の一つとなる。

提案法の二つのデータから類似部分を推定する考え

は画像データに特化したものではなく、時系列の問題を解決すれば動画や音声などのデータにも用いることができる。将来的には複雑データ内の類似特徴推定と少量データを用いた特徴量評価に用いることができると思われる。

参考文献

- [1] Myron Flickner, etc, Query by Image and Video Content: The QBIC System, IEEE Computer, vol.28, no.9, pp.23-32, 1995.
- [2] D.Kishore Kumar, E.Usha sree, K.Suneera, P.V.Chaitanya Kumar, CONTENT BASED IMAGE RETRIEVAL - EXTRACTION BY OBJECTS OF USER INTEREST, IJCSE, vol.3, no.3, pp.1068-1074, Mar 2011.
- [3] John R. Smith, Shih-Fu Chang, VisualSEEK: a fully automated content-based image query system, In Proceedings of the fourth ACM international conference on Multimedia (MULTIMEDIA '96), ACM, New York, USA, no.87-98, 1997.
- [4] 串間和彦, 赤間浩樹, 紺谷精一, 木本晴夫, 山室 雅司, "オブジェクトに基づく高速画像検索システム: ExSight, " 情報処理学会論文誌, vol.40, no.2, pp.732-741, Feb 1999.
- [5] Wei-Ying Ma, B.S. Manjunath, EdgeFlow: A Technique for Boundary Detection and Image Segmentation, IEEE TRANSACTIONS ON IMAGE PROCESSING, vol.9, no.8, pp.1375-1388, Aug 2000.
- [6] Neetesh Gupta, R.K.Singh, P.K.Dey, A New Approach for CBIR Feedback based Image Classifier, International Journal of Computer Applications, vol.14, no.4, pp.14-18, Jan 2011.
- [7] 北本朝展, 高木幹雄, "類似画像検索システム構築のフレームワークとしての階層モデル, " 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎 97(206), pp.25-32, 1997-07-24.
- [8] Yu Sun, Bir Bhanu, Image Retrieval with Feature Selection and Relevance Feedback, IEEE 17th International Conference on Image Processing, pp.3209-3212, Sep 2010.
- [9] Yamamoto.A, Takagi.M, Extraction of Object Features from Image and its Application to Image Retrieval, Pattern Recognition, 9th International Conference on, pp.14-17, 1988.
- [10] Hui Hui Wang, Dzulkifli Mohamad, N.A.Ismail, Semantic Gap in CBIR: Automatic Objects Spatial relationships Semantic Extraction and Representation, International Journal of Image Processing(IJIP), vol.4, no.3, pp.192-286, July 2010.