

# Color $N \times M$ -grams과 CCV를 이용한 내용기반 영상 검색 시스템

이 은 주\*, 이 상 미\*\*, 정 성 환\*

\* 창원대학교 전자계산학과 멀티미디어 연구실(MIPS)

\*\* 한국전자통신연구원 정보유통연구실

## Content-based Image Retrieval System Using Color $N \times M$ -grams & CCV

Eun-Joo Lee\*, Sang-Mi Lee\*\*, Sung-Hwan Jung\*

\* MIPS Lab., Dept. of Computer Science, Changwon National University

\*\* Electronics and Telecommunications Research Institute

ejlee@cosmos.changwon.ac.kr, sjung@sarim.changwon.ac.kr

### 요 약

칼라 히스토그램의 단점을 보완할 수 있는 CCV(Color Coherence Vector) 방법이 소개되었다. CCV는 구현이 쉽고, 칼라 히스토그램과 달리 같은 색상 분포를 가지는 다른 영상을 구별하는 것이 가능하다. 그러나, CCV는 계산량이 많아 많은 처리 시간이 요구된다. 본 논문에서는 효율적인 계산을 위하여  $N \times M$ -grams과 계층적인 검색 방법을 이용하여 처리 시간을 줄이는 검색 방법을 제시한다. 먼저, 영상의 구조적 특징을 잘 반영하는  $N \times M$ -grams를 사용하여 주어진 질의 영상과 같은 부류(category)에 속하는 모든 영상들을 찾는다. 그리고, 찾은 영상들만을 대상으로 CCV를 계산하여 검색한다. 200개의 영상을 가지고 실험한 결과, 검색율은 약 79%이고, CCV만을 사용한 방법보다 시간이 약 37%감소하였다.

### 1. 서 론

최근 컴퓨터 기술의 발달로 단일미디어 뿐만 아니라 정지영상, 동영상, 비디오, 오디오 등의 다양한 형태의 멀티미디어 정보를 응용하여 사용하고 있다. 원격진단 시스템, VOD(Video On Demand), 전자도서관, 홈쇼핑 등은 멀티미디어 정보를 응용한 서비스들이다. 이러한 응용을 위해서는 멀티미디어 정보를 보다 효과적으로 저장하고 관리하는 기술이 필요할 뿐만 아니라, 멀티미디어 정보를 효율적으로 질의하고 검색하는 기술도 요구된다.

멀티미디어 정보를 검색하는 방법은 문자기반 접근 방법(text-based approach)과 내용기반 접근 방법(content-based approach)으로 나눌 수 있다[1,2].

문자 중심의 접근은 검색 대상이 되는 모든 주제어를 사람이 수작업으로 생성하고, 검색을 하기 위한 생성된 주제어의 색인도 숙달된 작업자에게만 익숙한 일이 되어 시간과 비용에 있어서 비효율적이다. 내용 중심의 접근은 멀티미디어 정보의 내용 표현 요소들로 얻어진 특징 값을 자동으로 추출하여 검색에 사용하기 때문에 시간과 비용에 있어서 효율적이다. 따라서 본 논문에서는 내용기반 접근을 기반으로 멀티미디어 정보 중 영상 정보에 대한 검색을하고자 한다.

영상의 내용 표현 요소로는 질감(texture), 모양(shape), 칼라(color) 등이 있다. 이중 칼라는 대상 식별을 아주 단순화 시켜주고 대략적인 구분정보를 제공해 주는 특성이 있다. 본 논문에서는 검색을

위하여 칼라 정보를 사용하고, 영상의 패턴과 칼라의 분포정도도 고려하였다.

CCV(Color Coherence Vector)[3]는 칼라 히스토그램 기법과는 달리 칼라 정보의 지역적 특성을 고려하여 같은 색상 분포를 가지는 다른 영상을 구별하는 것이 가능하며 구현이 간단한 장점이 있다. 그러나, 계산량이 많아 많은 처리 시간이 요구된다. 또한, 대용량 영상 데이터베이스 검색 시, 데이터베이스내의 영상 모두를 비교 대상으로 한다면 상당한 처리 시간이 요구된다. 즉, 전혀 유사하지 않은 집단의 영상까지도 비교하는 redundancy가 발생한다.

본 논문에서는 CCV의 처리 시간을 줄이고 영상 검색에서 redundancy를 없애기 위해, 먼저 영상 데이터베이스에서 질의 영상과 유사한 영상들을  $N \times M$ -grams[4]를 사용하여 1차 분류한 후, 분류된 영상들을 대상으로 최종적으로 CCV를 사용하여 영상을 검색한다[5].

본 논문의 구성은 다음과 같다. 2장에서는 Color  $N \times M$ -grams과 CCV를 이용한 내용기반 영상 검색 시스템을 제시하고, 3장에서는 Color  $N \times M$ -grams을 사용한 1차 영상 분류에 대해 설명한다. 그리고 4장에서는 CCV를 설명하고, 5장에서는 실험 및 결과를 서술한다. 마지막으로 6장에서는 결론을 맺는다.

## 2. 제안한 영상 검색 시스템의 구조

본 논문에서 제안한 영상 검색 시스템의 구조는 그림 1과 같다. 질의 영상이 입력되면 Color  $N \times M$ -grams을 사용하여 질의영상과 같은 부류(category)에 있는 영상들을 모두 찾는다. 다음, 분류된 영상들에 대하여 CCV를 구하여 비교한 후 최종적으로 후보 영상을 검색한다.

## 3. Color $N \times M$ -grams를 사용한 1차 영상 분류

$N \times M$ -grams를 사용한 영상 분류[4]는 text documents의 유사성을 결정하는데 사용되는  $N$ -gram 기술을 기본으로 한다. 이것은 영상을 분류하기 위하여  $N$ -gram 정의를 이차원으로 확장한 것이다.  $N \times M$ -gram은 영상의 작은 패치 또는 패

턴의 형태를 나타낸다.

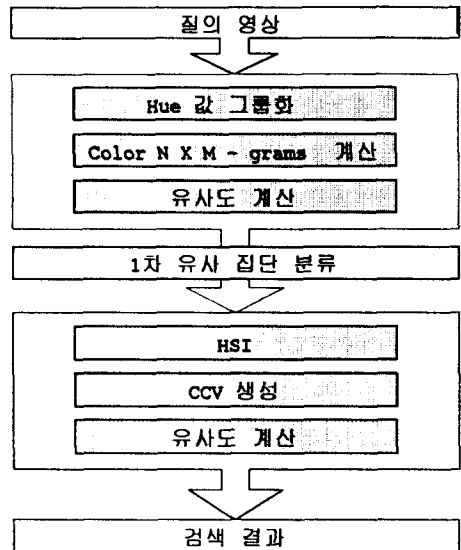


그림 1. Color  $N \times M$ -grams과 CCV를 이용한 내용기반 영상 검색 시스템

### 3.1 $N \times M$ -grams

$N \times M$ -grams은  $N \times M$  화소 값의 시퀀스이다.  $N \times M$ -grams은 row-wise 방법으로  $N \times M$  크기의 window 내에 있는 화소들 scan하면서 vector 값을 산출하는 방법이다. 예를 들면, 그림 2에 보여지는 화소 값을 가진 영상은 "001011110", "011110100", "011110100", "110100000"인 4개의  $3 \times 3$ -grams으로 구성된다.

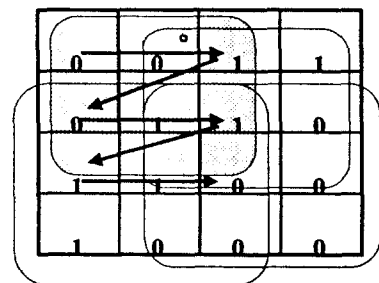


그림 2.  $N \times M$ -grams 생성의 예

영상의  $N \times M$ -gram 주파수 측면도(frequency profile)는 각 가능한  $N \times M$ -gram의 빈도 수의 계산으로 구성한다. 예를 들면, 이진 영상의 경우  $3 \times 3$

window를 사용하면  $2^{3 \times 3} = 512$ 개의 가능한  $3 \times 3$ -grams가 된다.

먼저,  $N \times M$ -grams를 얻기 위해서 일반적으로 입력 영상에 global threshold operation을 사용해서 영상을 이진 표시로 변형한다. 다음, 전체 영상에 크기  $N \times M$ 의 window를 움직여서 이진 영상에 대해  $N \times M$ -grams를 구하고, 각  $N \times M$ -gram의 발생 빈도 수를 계산한다. 만일, 두 영상의  $N \times M$ -gram vectors가 유사하다면 두 개의 영상이 같은 부류일 것이라는 가정을 전제로 한다.

### 3.2 Color $N \times M$ -grams

$N \times M$ -grams을 칼라 영상에 사용하기 위해서  $N \times M$ -grams을 확장할 필요가 있다. 본 논문에서는 영상을 단순히 이진화 하지 않고, 본 논문에서 제안한 Color  $N \times M$ -grams를 나타내기 위하여, Hue 값을 이용해서 그림 2와 같이 Red, Green, Blue 각각을 0, 1, 2로 그룹화 하였다. 따라서,  $3 \times 3$  window에 대하여 칼라 영상은  $3^{3 \times 3} = 19683$  개(예를 들면, 000000000, 000000001, 000000010, ..., 222222222)의 가능한  $3 \times 3$ -grams으로 나타내어진다.

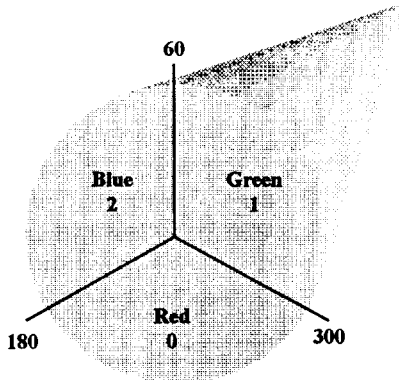


그림 3. Color  $N \times M$ -grams 생성을 위한 Hue 값의 그룹화

### 3.3 $N \times M$ -gram vectors의 유사도 계산

계산된  $N \times M$ -gram vector를 비교하기 위하여  $N \times M$ -gram vector intersection인  $S_{int}(A, B)$ 를 사용한다.  $S_{int}(A, B)$ 를 사용한 두 개의 영상 A와 B 사이의 유사도는 식 (1)에 의해 계산된다[4].

$\min(t(j, A), t(j, B))$ 는 영상 A, B의  $N \times$

$M$ -grams의 vector j의 빈도 중 작은 값이다.  $S_{int}(A, B)$ 의 범위는 0에서 1사이에 있다. A와 B가 같은 영상이면 1이 되고, 전혀 다른 영상이라면 0이 된다.

$$S_{int}(A, B) = \frac{2 * \sum_{j=1}^T \min(t(j, A), t(j, B))}{\sum_{j=1}^T t(j, A) + \sum_{j=1}^T t(j, B)} \quad (1)$$

## 4. HSI 모델과 CCV

일반적으로 영상의 칼라 정보를 이용한 방법으로는 Swain에 의해 제안된 칼라 히스토그램이 있다 [6]. 칼라 히스토그램은 간단하게 칼라 분포로 영상을 나타낼 수 있고, 회전과 변형에 둔감하다는 장점을 가지고 있어서 영상 비교에 자주 사용된다. 그러나, 칼라 히스토그램은 다른 공간적 칼라분포를 가지는 다른 영상도 같은 히스토그램일 수 있어서 다른 두 개의 영상을 유사하다고 판단하는 오류를 범할 수 있다. 이런 단점을 보완하여 칼라의 공간적 분포 정도를 고려한 CCV(Color Coherence Vector)[3]는 각 화소의 칼라 값을 연속적인가 비연속적인가로 구분하여 표현하는 방법으로 구현이 쉬운 장점을 가지고 있다.

### 4.1 HSI 계산

칼라를 일반적으로 RGB(Red, Green, Blue), HSI(색조(Hue), 채도(Saturation), 명도(Intensity)), YIQ(밝기(Y), 색상의 색차성분(I, Q)) 등의 여러 가지 모델로 표현된다.

영상 데이터베이스 검색에서는 칼라 정보를 사용할 때 사람의 눈이 인식하는 색과 유사한 HSI모델을 일반적으로 많이 사용한다[6,7]. HSI모델은 색의 순수성을 나타내는 H 성분, 색의 농도를 나타내는 S 성분, 색의 밝기를 나타내는 I 성분으로 표현된다. 본 논문에서는 RGB 칼라 특징 값보다 검색에서 더 좋은 결과를 산출하는 HSI 칼라 값을 사용하였다 [7]. 그림 3은 HSI 칼라 모델을 삼각형의 형태로 표현한 것이다.

입력 영상의 Red, Green, Blue의 세 가지 칼라 특징 값을 식 (2), 식 (3), 식 (4)에 의하여 HSI 칼라 공간으로 변환한다.

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2} [(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\} \quad (2)$$

$$S = 1 - \frac{3[\min(R, G, B)]}{(R + G + B)} \quad (3)$$

$$I = \frac{1}{3}(R + G + B) \quad (4)$$

HSI로 변환된 화소 값들은 CCV 계산에 사용되어 진다.

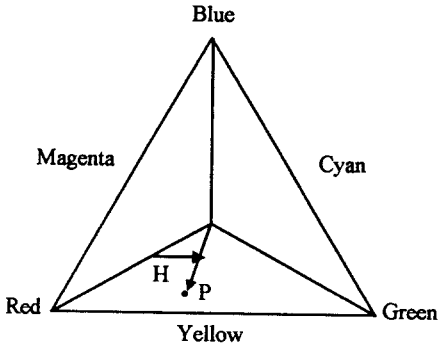


그림 4. HSI color triangle

4.2 CCV의 예

크기가 6×6인 영상을 예로 간단한 CCV 계산 과정을 보이코자 한다. 그림 5는 간단한 영상의 예이다.

CCV는 화소들을 n개의 칼라 범위로 분류하여 각 범위의 칼라 분포 정도(연속성 여부)를 비교한다. 화소들의 칼라 범위 값의 연속성 여부를 결정하기 위해서는 기준이 되는 값(R)을 정한다. 예에서는 R=5로 한다. 연속되는 화소들이 5개 이상이 되면 연속이라고 하고(Coherence vector), 5개 미만이면 비연속이라고 할 수 있다(Incoherence vector).

|    |    |    |    |    |    |
|----|----|----|----|----|----|
| 20 | 19 | 35 | 31 | 22 | 26 |
| 13 | 22 | 12 | 33 | 22 | 27 |
| 17 | 32 | 14 | 2  | 1  | 29 |
| 19 | 8  | 5  | 36 | 37 | 2  |
| 3  | 10 | 19 | 32 | 8  | 6  |
| 25 | 13 | 16 | 19 | 10 | 5  |

그림 5. 6×6 영상의 예

본 예제에서는 영상의 화소의 값 1에서 10까지를 bucket 1, 11에서 20을 bucket 2, 21에서 30을 bucket 3, 31에서 40까지를 bucket 4로 하여 칼라 공간을 4개로 구분하였다. 그림 6은 각 화소들을

칼라 공간으로 표시한 것이다.

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 2 | 2 | 4 | 4 | 3 | 3 |
| 2 | 3 | 2 | 4 | 3 | 3 |
| 2 | 4 | 2 | 1 | 1 | 3 |
| 2 | 1 | 1 | 4 | 4 | 1 |
| 1 | 1 | 2 | 4 | 1 | 1 |
| 3 | 2 | 2 | 2 | 1 | 1 |

그림 6. 그림 5의 칼라 공간 분류

4개의 칼라 공간으로 구분한 후 문자를 사용하여 label을 표시하면 그림 7과 같다. label 표시는 8-connected components를 사용하였다. 같은 칼라 공간에 속하지만 연속되지 않으면, 다른 label로 표시한다.

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| a | a | b | b | c | c |
| a | d | a | b | c | c |
| a | e | a | f | f | c |
| a | f | f | g | g | f |
| f | f | i | g | f | f |
| h | i | i | i | f | f |

그림 7. 그림 6을 label로 표시

표 1은 그림 7의 각 label의 개수를 계산하고 bucket과 연결한 결과이다. 개수가 5보다 크면 연속성으로 분류하고, 작으면 비연속성으로 분류하여 표 2와 같이 CCV를 생성한다.

표 1. 각 label의 연속 화소 수

| Label    | a | b | c | d | e | f  | g | h | i |
|----------|---|---|---|---|---|----|---|---|---|
| bucket   | 2 | 4 | 3 | 3 | 4 | 1  | 4 | 3 | 2 |
| 연속 화소 개수 | 7 | 3 | 5 | 1 | 1 | 11 | 3 | 1 | 4 |

여기서, 화소들이 연속적인 경우는  $\alpha_j$  vector, 비연속적인 경우는  $\beta_j$  vector이다. 표 2는 vector쌍  $(\alpha_j, \beta_j)$ 을 나열하여  $\langle (\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n) \rangle$ 로 표기하기도 한다.

표 2에서 bucket 1과 2는 연속적인 화소( $\alpha$ )만

포함하고, bucket 3은 화소가 연속적인 것( $\alpha$ )과 비연속적인 것( $\beta$ )이 섞였고, bucket 4는 비연속적인 화소( $\beta$ )만 포함한다. 예제 영상에 대하여 CCV는  $\langle(11,0), (11,0), (5,2), (0,7)\rangle$ 로 표현된다.

표 2. 그림 5의 CCV

|          |        |    |    |   |   |
|----------|--------|----|----|---|---|
|          | bucket | 1  | 2  | 3 | 4 |
| vector   |        |    |    |   |   |
| $\alpha$ |        | 11 | 11 | 5 | 0 |
| $\beta$  |        | 0  | 0  | 2 | 7 |

### 5. 실험 및 결과

실험에 사용된 영상 데이터베이스는 WWW의 Corel Draw Photo Album에 분류되어 있는 영상들로 구성되어 있다. 표 3과 같이 총 200개의 데이터 각각은 4개의 class로 분류된다. 질의 영상으로는 각 class에서 10개의 영상을 무작위로 선택하여 총 40개를 사용하였다. 실험에 사용된 영상들은 128×128 크기로 정규화 되어있고, 화소들은 Red, Green, Blue의 3차원 값들로 구성된다. 실험은 IBM 호환 기종인 586 PC로 하였으며, 시스템은 GNU C 언어를 사용하여 구현하였다.

표 3. 실험에 사용된 영상의 종류

|         |     |
|---------|-----|
| 영상의 종류  | 개수  |
| 경주용 자동차 | 56  |
| 구름, 하늘  | 54  |
| 꽃       | 41  |
| 대리석     | 49  |
| 총 계     | 200 |

질의 영상과 같은 부류(category)의 영상을 모두 찾기 위한 Color N×M-gram은 128×128 영상을 32×32로 축소하여 계산하였다. 각 영상의 Hue를 계산하여 Hue의 범위에 따라 0, 1, 2 값으로 변형한 후, 3×3-grams를 계산하면 각각의 영상에 대해 900개의 vector가 생성된다. vector 각각의 빈도수를 계산하여 식 (1)에 적용하면 질의 영상에 대한 대상 영상들의 유사도가 산출된다. 본 논문에서는 유

사도가 0.5 이상인 영상을 질의 영상과 같은 부류(category)로 1차 분류하였다

최종 검색과정인 CCV 계산을 하기 위해서 색조(Hue), 채도(Saturation), 명도(Intensity)에 따라 26개의 bucket으로 영상의 칼라 범위를 구분하였다. 칼라의 연속성 여부를 결정하는 R 값은 영상 전체 화소(128×128 = 16384)의 약 1%인 150으로 하였다[3]. 따라서, vector의 개수는 coherent vector 26개, incoherent vector 26개가 된다.

본 연구의 검색 결과에 대한 효율성을 분석하는 방법으로 Recall과 Precision을 사용했다. Recall은 영상 데이터베이스에서 질의와 관련된 영상 중 검색된 영상의 비율이고 Precision은 검색된 영상 중에서 질의와 관련된 영상의 비율이다[8-10].

$$Recall = \frac{R_r}{T}, \quad Precision = \frac{R_r}{T_r} \quad (5)$$

식 5에서  $R_r$ 은 검색 항목 중에서 질의와 관련된 항목의 수를 나타내고,  $T$ 는 검색 대상 중에서 질의와 관련된 항목의 총수,  $T_r$ 은 검색된 항목의 총수를 나타낸다. 표 5는 CCV만을 사용해서 검색한 결과와 계층적으로 검색한 결과를 비교한 것이다. 표에서 CCV만을 사용한 검색보다 계층적 검색은 Recall과 Precision이 조금 더 높게 나왔다.

그림 8은 본 논문에서 제시한 Color N×M-grams와 CCV를 이용한 검색 결과를 보여준다. 질의 영상과 유사한 순서로 검출한 결과 칼라의 영역별 분포 정도와 패턴이 잘 반영되었다는 것을 보여주고 있다. 표 6은 검색시간을 측정하여 비교한 것이다.

표 5. 계층적 검색 결과와 CCV만으로 검색한 결과 (질의 영상 : 40개, 대상 영상 : 200개)

| 방법      | 계층적 검색 |           | CCV    |           |
|---------|--------|-----------|--------|-----------|
|         | Recall | Precision | Recall | Precision |
| Class 1 | 0.91   | 0.50      | 0.94   | 0.48      |
| Class 2 | 0.79   | 0.76      | 0.73   | 0.68      |
| Class 3 | 0.81   | 0.79      | 0.80   | 0.77      |
| Class 4 | 0.66   | 0.59      | 0.66   | 0.57      |
| 평균      | 0.79   | 0.66      | 0.78   | 0.62      |

참고문헌

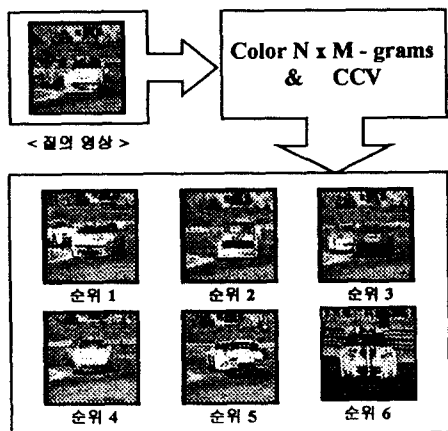


그림 8. Color N×M-grams와 CCV를 이용한 영상 검색 결과의 예

표 6. 검색 시간 비교

| 방법      | 계층적 검색 | CCV      |
|---------|--------|----------|
| 검색시간(초) | 5.8(1) | 9.3(1.6) |

6. 결론

본 논문은 Color N×M-grams와 칼라의 공간적 분포 정도를 고려한 CCV를 사용하여 계층적으로 영상 데이터베이스를 검색하는 방법을 제시하였다. 1차 검색으로는 Color N×M-grams를 사용하여 영상의 작은 패턴들을 비교하였고, 2차 검색으로는 HSI 칼라모델로 칼라 범위를 분류한 CCV를 사용하여 최종 검색하였다.

총 200개의 영상 중 데이터베이스의 각 class에서 질의 영상을 10개씩 선택하여 실험한 결과, 계층적 검색은 CCV만을 사용한 검색보다 Recall과 Precision은 조금 향상되었으며, 검색 시간은 약 37% 감소했다. 따라서, 대용량의 데이터베이스 검색시 계층적 검색이 더 효과적임을 알 수 있다.

앞으로의 연구는 효율성과 검색 시간을 향상시키는 방법과 본 시스템을 확장하여 WWW상에서 구현하는 연구를 하고자 한다.

- [1] Rajiv Mehrotra and James E. Gray, "Feature-based Retrieval of Similar Shapes," 9th International Conference on Data Engineering, pp.108-115, 1993.
- [2] Greg Pass and Ramin Zabih, "Histogram Refinement for Content-based Image Retrieval," Workshop on the Application of Computer Vision, 1996.
- [3] Greg Pass, Ramin Zabih and Justin Miller, "Computer Images Using Color Coherence Vectors," In ACM Multimedia 96, pp.65-73, 1996.
- [4] A.Soffer, "Image categorization using N×M-grams," In Proceeding of the SPIE, Storage and Retrieval of Still Image and Video Database V, vol. 3022, pp.121-132, San Jose, CA, February 1997.
- [5] 이은주, 정성환, "계층적 칼라 정보를 이용한 영상 검색 시스템," 대한전자공학회 한국통신학회 부산경남지부 추계 합동 학술논문발표회 논문집, 제 6권 제 2호, pp.130-134, 1997.
- [6] M. J. Swain and D. H. Ballard, "Color Indexing," International Journal of Computer Vision, Vol. 7, No. 1, pp.11-32, 1991.
- [7] 김진숙, 김호성, "색과 공간 정보를 이용한 영상 검색 기법," 한국정보과학회 가을학술 발표논문집, 제 24권 2호, pp.483-486, 1997.
- [8] 배회정, 정성환, "칼라와 질감을 이용한 칼라 영상 데이터베이스 검색시스템," 한국정보처리학회 '96 추계 학술발표논문집, 제 3 권 2호, pp.326-331, 1996.
- [9] 김진아, 정성환, "Wavelet 기반의 질감 특징을 이용한 영상 검색," 한국정보과학회 '97 가을 학술발표논문집, Vol. 24, No. 2, p.379-382, 1997.
- [10] Gerard Salton and Chris Buckley, "Improving Retrieval Performance by Relevance Feedback," Department of Computer Science Cornell University Ithaca, NY14853-7501, 1988.