

# 가중치 칼라 히스토그램을 통한 이미지 검색의 성능평가

이용환\*, 이유경\*,이준환\*\*, 이상범\*, 김영섭\*\*\*

\*단국대학교 전자컴퓨터공학과

\*\*극동대학교 컴퓨터정보표준학부

\*\*\*단국대학교 공학대학 전자공학과

## An Evaluation of Image Retrieval used Weighted Color Histogram

Yong-Hwan Lee\*, Yu-Kyong Lee\*, June-Hwan Lee\*\*,

Sang-Burm Rhee\*, YoungSeop Kim\*\*\*

\*Dept. of Electronics and Computer Engineering, Dankook University,

\*\*School of Computer Science, Information and Standard, Far East University,

\*\*\*Dept. of Computer Science & Electronics, Dankook University

E-mail : \*hwany1458@empal.com, \*angelyk@hanmail.net, \*\*rainbow@infomail.kdu.ac.kr

\*sang107@dku.edu, \*\*\*wangcho@dankook.ac.kr

### Abstract

본 논문에서는 이미지 검색을 위해 가장 기본적인 요소인 이미지 색상에 따라 칼라 분포정보를 이용하고 다양한 요소에 따라 가중치를 부여한 칼라기반의 검색 기술자를 제안하였고 실험적 평가를 통하여 제안 기술자의 성능을 평가하였다. 칼라 히스토그램을 통한 이미지 검색 기술자를 설계하는데 있어 칼라모델은 HSV, 웨이블릿 변환 필터는 D9/7, 웨이블릿 분해는 2 레벨을 적용하였을 때 가장 좋은 검색효율성을 보였다.

### I. 서론

병원, 학교, 웹사이트 및 전자상거래 등 다양한 분야에서 대용량의 이미지들이 제작되어 사용되고 있으며, 이러한 이미지의 저장은 파일 압축과 전송, 검색분야에 새로운 연구분야로 부각되고 있다[1]. 야후, 구글, 네이버와 같은 전통적인 웹 기반의 이미지 검색은 사용자 관점에서 해당 이미지에 대한 적절한 주제어 또는 키워드를 지정하고 텍스트 기반으로 검색을 수행한다[2]. 그러나 최근에 이미지 자체에 대한 색상, 질감, 형태 등과

같은 시각적인 특징을 추출하고 유사도를 측정하여 질의 이미지와 유사한 이미지를 검색하는 내용기반의 이미지 검색이 활발히 연구되고 있다[3,4].

본 논문에서는 콘텐츠 기반의 이미지 검색에서 가장 많이 사용되고 있는 이미지 색상(Color)정보에 대한 분포를 기반으로 이미지의 특징 추출과 유사도 검색을 수행하였다. 이미지 칼라분포에 대한 정보는 이미지 히스토그램이 가장 일반적으로 활용되고 있으며, 칼라 이미지의 히스토그램을 추출하는데 있어 다양한 요소와 가중치를 부여한 검색 기술자(Weighted Color Histogram Descriptor)를 제안하고 실험적으로 성능을 평가하였다.

### II. 본론

칼라(Color)는 콘텐츠 기반의 이미지 검색에서 가장 기본적이면서 가장 많이 사용되는 특징중의 하나이며, 특징 추출방법이 간단하고 이미지 사이즈나 방향에 독립적인 장점을 가진다. 본 논문에서는 이미지 특징을 추출하기 위한 요소로 칼라 이미지의 히스토그램을 사용하고 칼라 모델에 따라 가중치를 부여한 칼라 히스토그램 기술자를 제안한다.

본 연구는 2005년도 산학협동재단 연구비지원사업의 지원에 의해 수행되었습니다.

검색 기술자의 추출과정을 요약하면 다음과 같다.

- 1) 입력이미지에 대한 칼라 모델 변환
- 2) 칼라에 따른 채널 분리
- 3) 각 채널에 대한 히스토그램 추출
- 4) 웨이블릿 변환
- 5) 선형적 양자화
- 6) 가중치를 적용한 특징 벡터 병합
- 7) 인덱스 생성 및 저장

칼라모델 변환에서는 RGB, HSV 와 YCbCr 의 3 가지 모델에 대한 변환을 실험적으로 수행하였으며, 웨이블릿 변환에서는 Haar, D4, D5/3 와 9/7 필터를 사용하였고 웨이블릿 분해는 5 레벨까지 단계적으로 적용하였다. 특징 벡터 병합에 적용된 가중치는 기본으로 1:1:1 을 부여하였으며, 다양한 가중치를 실험적으로 적용하였다.

본 논문에서는 질의 이미지와 이미지 레퍼지토리에 저장된 검색 대상의 이미지들과의 유사도를 측정하는 방법으로 두 벡터간의 절대크기 차이의 합(L1 Norm) 을 이용하였다. 이를 구하는 수식은 다음과 같다.

$$D(I_q, I_t) = \sum_{i=1}^b |H_q(i) - H_t(i)| \quad (1)$$

### III. 구현 및 실험결과

전체 이미지는 500 개를 포함하고 있으며, Corel 이미지 200 개, 워싱턴대학과 스웨덴대학 웹사이트를 통해 다운받은 샘플이미지 130 개와 직접 제작한 일반 이미지(Natural Image) 170 개로 구성되었다[5]. 테스트 이미지는 384x256 에서 756x504 까지의 서로 다른 해상도와 파일 사이즈를 가지며 파일 포맷은 JPEG 을 사용하였다.

본 논문에서는 검색 효율성을 평가하기 위한 방법으로 순위 검색 결과에 따른 재현도(Recall)과 정확도(Precision)을 계산하고 R-P 그래프를 통하여 성능을 비교하였다. 그림 1 과 2 는 검색결과에 대한 랭킹을 통하여 검색 효율성을 계산한 그래프이다. 랭킹 검색결과에 대해 각각 상의 5%, 10%, 20%와 50%에 해당하는 이미지를 추출하고 이에 대한 검색효율성을 계산하였다. 유사한 이미지의 객관적 평가 기준은 질의 이미지와 동일 카테고리에 포함되는지의 여부를 통하여 측정하였다. 실험적 결과, 칼라 히스토그램을 활용하여 이미지 검색 기술자를 설계할 경우, 칼라모델은 HSV 가 가장 좋은 검색 결과를 얻었으며, 웨이블릿 변환 필터는 D9/7,

D5/3, D4, Haar 순으로 나타났다. 웨이블릿 분해레벨은 각각 2,3,1,4,5 레벨순으로 검색 효율성을 보였으며 가중치는 HSV 모델에서 채널별 2:1:1 을 적용하였을 때 가장 성능이 좋았다.

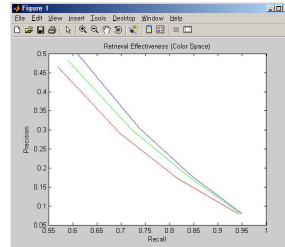


그림 1) 칼라모델변경에 따른 성능비교  
(Red=YCbCr, Green=RGB, Blue=HSV)

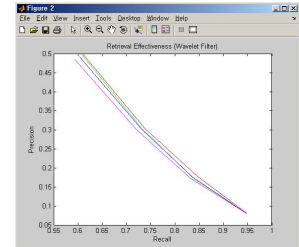


그림 2) 웨이블릿 필터변경에 따른 성능비교  
(Red=D9/7, Green=D5/3, Blue=Db4, Magenta=Haar)

### IV. 결론 및 향후 연구 방향

본 논문에서는 이미지검색에 가장 많이 사용되는 색상정보를 이용하여 칼라분포에 따른 특징을 추출하고 유사도검색을 수행하여 유사 이미지의 순위를 계산하였다. 실험결과 특징벡터 추출 평균시간은 이미지당 0.24 초, 유사도검색시간은 0.45 초가 소요되었다. 향후 과제로서 Color Layout 과 Correlogram 등과 같은 이미지의 공간적 정보를 포함하여 이미지 특징을 추출할 수 있도록 검색기술자의 확장과 더불어 실시간 처리를 위한 지속적인 연구와 성능개선이 진행되어야 할 것이다.

### 참고문헌

- [1] Remco C.Veltkamp, Mirela Tanase, "Content-based Image Retrieval System: A Survey", Technical Report UU-CS-2000-34, Oct.,2000.
- [2] Scott Houchin, ISO/IEC JTC1/SC29/WG1N3506, "Image Search System Components and Standardization Scope Recommendations", Jan., 2005.
- [3] Mun-Kew Leong, Wo Chang, ISO/IEC JTC1/SC29/WG1 N3684, "Framework and System Components" July, 2005.
- [4] ISO/IEC JTC1/SC29/WG11/N4063 "MPEG-7 Visual Part of eXperimentation Model version 10.0" March, 2001.
- [5] Website: <http://muvis.cs.tut.fi>, <http://wang1.ist.psu.edu>, <ftp://ftp.sunet.se>, <http://www.cs.washington.edu>