

# 모양과 색상 정보를 결합한 효율적인 내용 기반 이미지 검색

김현중, 박영배  
명지대학교 컴퓨터공학  
e-mail : [yjkfc@mail.yeojoo.ac.kr](mailto:yjkfc@mail.yeojoo.ac.kr)

## An Efficient Content-based Image Retrieval merged Color and Shape Information

Hyun-Jong Kim \*, Young-Bae Park\*  
\*Dept. of Computer Engineering, Myong-ji University

### 요 약

영상의 객체가 갖고 있는 중요도를 계산하는 것은 매우 주관적이기 때문에, 객체의 의미 정도를 판단하는 것은 매우 어려운 문제이다. 따라서 본 논문에서는 효율적인 내용기반 이미지 검색 시스템을 구축하기 위하여 원 영상에서 Wavelet 변환을 이용하여 RGB 이미지를 각각 분리 및 병합하는 방법을 제안한다. 또한, 각각의 RGB 이미지들 중 검색에 필요한 특징들을 추출하여 원 영상과 가장 근접한 이미지들을 찾아내는 방법을 제안한다. 제안한 세가지 특징 정보를 이용한 내용 기반 이미지 정보 검색은 기존 방법에 비해 높은 검색 성능을 보였으며, 사람 얼굴이나 캐릭터 이미지인 경우에는 더욱 효율적이라 예상된다

### 1. 서론

영상분할은 오랫동안 연구된 분야지만 여전히 어려운 문제로서 다양한 시도들이 계속되고 있다. 효율적인 영상분할 방법[1]이 되기 위해서는 우선 영역이 과잉 분할되는 것을 방지해야 한다. 또한 고속처리 및 다양한 입력영상에 대한 적절한 임계값 설정이 가능해야 된다. Qian[2]에서는 색상이나 패턴의 변화가 거의 없는 평탄한 배경을 가진 영상에서 피사체를 분리하고 있다. 이들이 제안한 방법은 복잡한 배경을 가지는 영상에서는 배경과 피사체를 분리하기 힘들다. 이러한 문제점들을 해결하기 위해서 본 논문에서는 총 다섯 가지 단계로 처리할 수 있는 내용 기반 검색 시스템을 구축하였다. 첫 번째 단계인 전처리 과정은 원 영상의 윤곽선 정보를 명확하게 구분할 수 있다. 또한 원 영상의 Noise 를 제거하여 특징을 추출하는 방법을 제공한다. 두 번째 단계인 Wavelet 변환에서는 원 영상에서 RGB 각각을 Wavelet 에 대입하여 부밴드(subband)별 저대역 대표값 만을 추출한다. 또한,

Wavelet 변환에서의 저대역 대표값 만을 사용함으로써, 저장 공간상의 문제점을 해결할 수 있으며, 찾아내고자 하는 영역을 정확하게 찾을 수 있다. 세 번째 단계로 RGB 로 분리된 각각의 이미지들을 Watershed 를 이용하여 이미지의 조각들을 더욱 세분화 시키며, 반대로 세분화된 영역들은 Watermerge 를 이용하여 결합한다. 이로 인해서 두 번째 단계에서 찾아내기 어려웠던 영역들까지 찾아짐을 실험결과를 통하여 나타내고자 한다. 네 번째 단계에서는 RGB 채널 선택을 통하여 세가지 특징을 추출한다. 세가지 특징 중에서 첫 번째 특징 정보는 배경이 가장 큰 것과 배경이 가장 작은 이미지를 찾아내며, 각각의 이미지는 Zernike Moment 처리를 거쳐서 이들간에 차이를 구한다. 두 번째와 세 번째 특징 정보는 배경이 가장 큰 이미지와 원 영상을 비교 연산하여 피사체의 색상과 배경 색상을 추출한다. 마지막 단계로 RGB 채널 선택에서 추출된 세가지 특징 정보를 이용하여, 원 영상과 가장 근접한 이미지를 검색하게 된다. 이를 위해 특징 정보들로 질의 영상과 데이터베이스 영상 간의 가중치 기반의 유사도 척도를 정의하여, 유사도

를 측정 비교함으로써 영상을 검색할 수 있도록 하였다.

서론에 이어, 2 장에서는 논문과 관련 있는 관련 연구들을 소개하며, 3 장은 본 시스템에서 제안하는 모양과 색상 정보를 결합한 검색 시스템 방법과 4 장에서는 제안된 방법을 이용하여 시행한 실험 및 결과에 대한 분석과 고찰을 서술하고, 마지막으로 5 장에서 결론을 맺는다.

## 2. 기존 내용 기반 시스템

1992 년 Nagasaka 등은 영상의 지역 정보를 사용한 Color-pair 를 이용했고, 1995 년 Chua 는 이를 개선하여 영상의 두드러진 특징만을 고려하는 효율적인 방법을 제시했다[3]. 색상 정보를 이용하는 이러한 연구들은 서로 연관된 화소 간의 위치정보를 잘 나타내지 못하고 대략적인 구분정보만을 가지게 된다. 1990 년 Grosky, 1993 년 James 등은 모양 정보를 이용하는 검색 방법을 제시한바 있으나, 이 방법은 객체를 이루는 윤곽선이 모양의 변환이나 방향에 따라 민감하게 반응하고, 응용 분야에 이용되는 모양 데이터가 미리 등록되어야 한다는 조건이 따른다. 그리고 1993 년 Wang 등은 이러한 통계적 질감 정보를 이용하여 이진 문서 영상 분석에 사용하였다[4]. 하지만 질감 정보의 통계 값을 구하는 알고리즘이 복잡하고 유사도를 측정하기 위한 구성도 복잡하다. 그러나 질감은 영상의 표면특성을 나타내는 중요한 특징이며, 서로 다른 영상들을 구별하기 위한 중요한 요소이다.

Jian[5]은 모양의 경계선에 있는 점들을 2 차원 평면상에 두고 이 점들을 2 차원 신호로 만들어 푸리에 변환을 통해 얻어지는 계수를 벡터로 표현하기도 하였다. 하지만 이와 같은 윤곽선을 이용하는 방법은 윤곽선의 작은 변형에도 민감하게 반응하는 단점을 가지며 푸리에 서술자나 모멘트는 영역의 지역적인 정보를 제공하지 못하는 단점을 가진다.

## 3. 제안된 시스템 구성도

영상의 객체가 갖고 있는 중요도를 계산하는 것은 매우 주관적인 문제이기 때문에, 객체의 의미 있는 정도를 판단하는 것은 매우 어려운 문제라고 볼 수 있다. 사람은 영상에 포함된 객체들의 중요도를 판단하여 관심이 가는 의미 있는 객체를 자신이 선택할 수 있지만, 컴퓨터를 통하여 영상을 처리하는 경우에 있어서는 쉽지 않은 문제이다. 그러한 이유로 본 논문에서는 영상 내 존재하는 두드러진 피시체를 추출하기 위해서 다음과 같은 시스템을 구성하였다.

전체적인 시스템의 구성은 그림 3.1 과 같다. 본 시스템에서 제시하고 있는 방법에서 처음 처리되는 것은 전처리 과정이다. 전처리 과정에서는 원 영상에

세가지 방법을 사용하여 수정하게 된다. 전처리 과정에서 첫 번째는 원 영상을 그대로 Wavelet 으로 처리한다. 두 번째는 Contrast 와 Sharpen 필터를 거쳐서 수행하는 경우이다. 이는 원 영상에 색상 대조를 강하게 하며, 또한 Edge 의 처리를 진하게 함으로써, Edge 는 명확하게 구분 지을 수 있다. 또한 마지막으로 Blur 필터는 원 영상을 흐리게 하는 것으로, Noise 를 제거하는 특징을 가지고 있다.

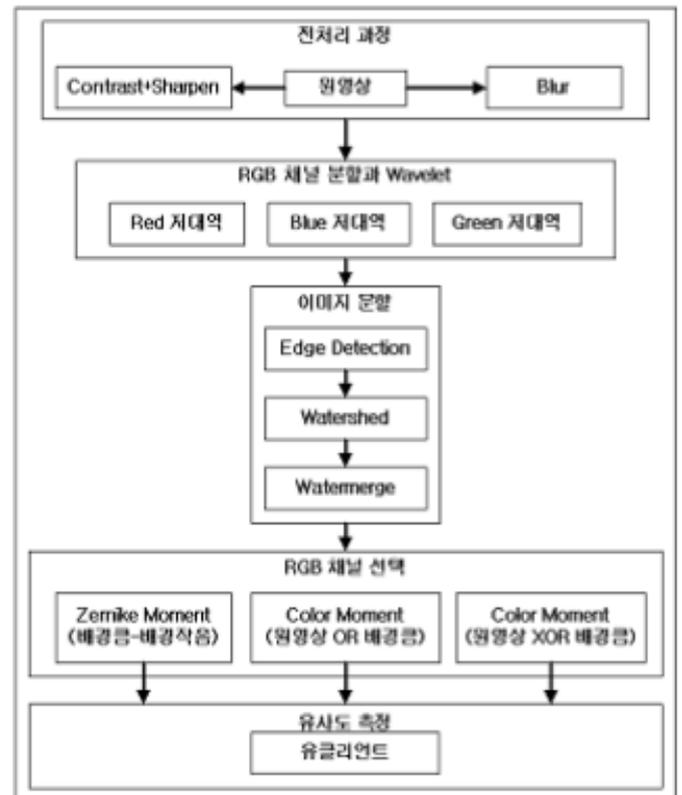


그림 3.1. 모양과 색상 정보를 결합한 검색 시스템.

원 영상, Contrast + Sharpen, 그리고 Blur 과정을 거친 세 가지 영상 중에서 최적의 영상을 찾기 위해서, PSNR(Peak Signal Noise Ratio)을 이용하여 선택 명도가 가장 높은 것을 선정한다. 또한 Wavelet 에서는 각 영역마다 한 개의 저대역과 세 개의 고대역 서브밴드를 만들게 되며, 그 중에서 Edge 에 대한 정보를 포함하고 있는 고대역 부밴드를 사용하지 않고, 시각적으로 잘 표현할 수 있는 질감 정보를 가지고 있는 저대역 부밴드만을 사용한다.

그리고 Wavelet 에서 변환된 저대역 정보를 이용한 이후, 배경을 제외한 가장 중요한 영역만을 추출하기 위해서 Edge Detection, Watershed, 그리고 Watermerge 를 수행하게 된다.

Edge Detection 은 Edge 정보만 표현하며, 그 밖에 정보는 소멸된다. 그 다음으로 Watershed 과정을 거쳐서 Edge 정보에서 더욱 세분화된 정보로 분리하게 된다. 이렇게 분리된 정보를 이용하여

Watermerge 에서는 통합을 하게 되며, 실제 Edge Detection 에서 찾지 못한 부분까지도 Watermerge 에서는 찾을 수 있다.

Watermerge 에서는 임계값을 가지고 있으며, 실험 치로 30, 40, 그리고 50 으로 정의하였다. 그 중에서 영역이 가장 큰 모양을 찾아내서 선정을 하게 된다. 만약 임계값 50 에서 영역이 가장 큰 부분이 발견 되었다면, 임계값 50 에서 생성된 세 가지 RGB 영상 정보를 선정하게 된다. 임계값의 값이 크면 클수록, 통합되는 그룹의 크기가 커지며, 그룹의 수는 작아진다. 또한 임계값이 너무 큰 경우에는 전체가 배경으로 흡수된다. 반대로 임계값이 작으면 그룹의 크기는 작아지고, 그룹의 개수는 많아지게 된다.

마지막으로 RGB 채널 선택에서 배경이 가장 큰 것과 배경이 가장 작은 것을 분리한다. 그리고 유사도를 측정하기 위해서 Zernike Moment 처리를 거친 결과와 그리고 원 영상에서 얻은 Color Moment 를 참조하여 최종 유사한 이미지를 얻을 수 있다. 전체 처리 과정 중에서 RGB 채널 선택 과정은 다음과 같다.

### 3.1 RGB 채널 선택

RGB 채널은 각각의 세 가지 이미지들 중에서 최적의 이미지를 찾는 과정이다.

먼저, 그림 3.2 은 RGB 채널 선택의 세부적인 처리 내용이다.

<p>입력 : 320x240 인 3 개 이미지(Red, Green, 그리고 Blue) 출력 : 실수 값을 가지는 세 가지의 특징</p> <p>1 단계 : 기준치보다 작은 영역보다 작은 이미지 흡수 if(찾아낸 영역의 크기 &lt; 6.25%) /* 320x240 영상의 16 분의 1 */ 찾아낸 영역을 주변 영역에 흡수</p> <p>2 단계 : RGB 채널 선택 if(배경이 가장 큰 이미지를 선택) big_image = Zernike_Moment (배경이 가장 큰 이미지) if(배경이 가장 작은 이미지를 선택) small_image = Zernike_Moment (배경이 가장 작은 이미지) zernike_diff = big_image - small_image color_OR = Color_Moment (배경이 가장 큰 이미지 OR 원 영상) color_XOR = Color_Moment (배경이 가장 큰 이미지 XOR 원 영상)</p> <p>그림 3.2 RGB 채널 선택</p>
--

1 단계에서는 RGB 이미지들 중 아주 작은 이미지들이 많은 문제를 야기시킨다. 왜냐하면, 일정 크기보다 작은 이미지는 피사체인지 배경인지 알 수 없기

대문이다. 따라서, 일정 크기인 16 분의 1(6.25%)보다 작은 이미지의 경우는 피사체 혹은 배경으로 흡수되는 것이 실험 결과 좋은 성능을 나타내는 것으로 확인되었다. 다시 말해서, 흡수되는 방식은 작은 이미지가 피사체에 붙어 있다면, 피사체에 흡수되고, 그 외에 이미지들은 배경에 흡수된다.

2 단계 RGB 채널 선택에서는 유사도 측정에 필요한 세 가지 특징을 추출해야 한다. 세가지 특징 중에서 첫 번째 특징 정보는 배경이 가장 큰 것과 배경이 가장 작은 이미지를 찾아내며, 각각의 이미지는 Zernike Moment 처리를 거쳐서 이들간에 차이를 구한다. 두 번째와 세 번째 특징 정보는 배경이 가장 큰 이미지와 원 영상을 비교 연산하여 피사체의 색상과 배경 색상을 추출한다.

### 3.2 유사도 측정

대용량의 영상 데이터베이스를 검색하기 위해서는 간단하면서도 효율적인 유사도 척도(distance measure)가 정의되어야 한다. 일반적으로 영상간의 유사도를 얻기 위한 척도로 각 영상의 특징 값들 사이의 거리(distance)를 구하여 비교한다. 본 시스템에서는 RGB 채널에서 추출된 세 가지 특징 정보를 이용하여 유사도를 측정한다. 유사도 측정은 유클리언트 거리를 계산하여 가장 근접한 이미지들을 찾아낼 수 있으며, 이를 위해서 가중치를 사용한다.

가중치  $w_1, w_2, w_3$  은 특징간의 상대적인 가중치로 정의되며 그 합은 1 이다. 유사도 결정은 사용하는 여러 실험 영상의 피사체에 대한 테스트를 거쳐 가장 효과적인 것으로 결정하였다. 여기서  $w_1$  은 모양 특징에 대한 가중치이고,  $w_2$  와  $w_3$  은 색상에 대한 가중치로서  $w_1$  를 보다 크게 할 경우 영상내의 모양특징에 비중을 두는 경우가 되며, 그 반대의 경우는 색상 정보에 비중을 고려하는 것이다.

### 4. 실험 및 분석

본 논문에서 제안한 방법을 실험하기 위한 시스템 구현 환경으로 하드웨어 시스템은 펜티엄 4 1.8G 을 사용하며, 운영 체제는 Windows XP 환경에서 JAVA 1.4, 그리고 라이브러리는 JAI(Java Advanced Imaging) API 를 이용하여 구현하였다. 사용된 이미지는 Stanford[6]에서 제공하는 1000 개와 10000 개 이미지 모델이며, 각각의 해상도는 384\*256 과 128\*85 이다.

표 4.1 은 기존에 제안된 방법을 Wavelet 을 사용한 것인가에 여부에 따라서 결과가 나온다. 그렇지만 본 시스템에서 제안하는 시스템의 성능은 표 4.2 에서 나타난다. 표 4.2 에서처럼 세 가지 특징을 가지고 있으며, 각각의 특징들을 따로 분류하여 측정된 것과, 세 가지 모두를 결합한 방법으로 성능을 측정한다.

표 4.1. Wavelet 사용에 따른 검색 성능 비교

실험방법 \ 성능평가	Wavelet 을 사용하지 않음		Wavelet 을 사용함	
	Recall	Precision	Recall	Precision
Invariant 모멘트	0.65	0.60	0.73	0.69
Zernike 모멘트	0.67	0.63	0.75	0.71
Color 모멘트	0.80	0.62	0.82	0.64

표 4.2. 세가지 특징에 따른 검색 성능 비교

실험방법 \ 성능평가	Wavelet 을 사용함	
	Recall	Precision
첫 번째 특징	0.66	0.58
두 번째 특징(피사체 색상)	0.87	0.81
세 번째 특징(배경 색상)	0.83	0.79
세 가지 특징 결합	0.92	0.86

생성된 세 가지 특징들을 이용하여 실험한 결과가 그림 4.1 이다. 그림 4.1(a)는 Query 를 사용할 때 사용된 원본 영상을 의미하며, 그림 4.1(b)는 3 가지 특징(피사체 색상, 배경 색상, 그리고 배경 크기의 차이)을 모두 사용한 검색 결과이다.



(a)



(b)

그림 4.1. 말 이미지를 이용하여 Query 한 영상들. (a)는 원본 영상, (b)는 3 가지 특징(피사체 색상, 배경 색상, 배경 크기의 차이)을 모두 사용한 검색 결과.

## 5. 결론 및 향후 과제

본 논문에서 제안한 시스템은 지금까지 다루지 못했던 RGB 와 Wavelet 을 사용하여 통합하였으며, 실

제 내용 기반 시스템에 어려움을 겪은 다양한 피사체에 대한 대처 방안으로, 배경이 큰 것과 배경이 작은 이미지를 찾아서 보다 효율적인 검색이 가능하도록 제안하였다.

본 시스템에서는 효율적인 내용기반 검색 시스템을 구축하기 위한 다섯 단계를 제안하였으며, 첫 번째로 전처리 단계이며, 두 번째 단계는 RGB 분할한 이미지를 이용한 Wavelet 의 적용이며, 세 번째 단계는 이미지를 세분화하여 통합시키는 단계, 네 번째 단계에는 세 가지 특징을 추출하며, 마지막 단계에는 이들 세가지 특징을 이용하여 유사도를 측정한다.

이와 같은 특징 정보들로 질의 영상과 데이터베이스 영상간의 가중치 기반의 유사도 척도를 정의하여 유사도를 측정 비교함으로써 영상을 검색할 수 있도록 하였다. 그렇지만, 배경과 피사체의 구분이 모호한 경우에는, 검색하기 어렵다는 문제가 있다. 그뿐만 아니라, 피사체의 크기가 배경보다 큰 경우에는 배경을 찾아내기 어렵다는 문제가 발생한다. 이러한 문제를 해결하기 위한 연구가 계속 되어야 할 것이다.

## 참고문헌

- [1] Shinn-Ying Ho; Kual-Zheng lee, "An efficient evolutionary image segmentation algorithm", Proceedings of the 2001 Congress on IEEE, pp.1327-1334, 2001
- [2] Qian Huang, Byron Dom, David Steele, Jon Ashley, and Wayne Niblack. "Foreground/Background segmentation of color images by integration of multiple cues." In Proceedings of 1995 IEEE Conference on Image Processing, pages 1246-1249, Washington D.C, U.S.A., 1995
- [3] Nagasaka A. & Tanaka Y. "Automatic Video Indexing and Full-Video Search for Object Appearance", Visual Database System II, IFIP, Elsevier Science Publishers B.V., pp. 113-127, Oct. 1993.
- [4] D. Wang and S. N. Srihari, "Classification of Newspaper Image Blocks Using Texture Analysis", Computer Vision, Graphics and Image Processing, vol. 47, pp. 327-352, 1989.
- [5] A. Jian, Fundamentals of digital image processing. Prentice Hall, 1989
- [6] Stanford University, Content-based Image Retrieval Project, <http://www-db.stanford.edu/IMAGE/>