이미지의 주요 색상 정보들을 이용한 실시간 객체 검출 방법

박재한, 고재한, 박경욱, 백승호, 백문홍*
*한국생산기술연구원 로봇기술개발본부
e-mail : { hans1024, jaehanko, kwpark, shbaeg, mhbaeg}@kitech.re.kr

A Robust Real-time Object Detection Method using Dominant Colors in Images

Kyung-Wook Park, Jaehan Koh, Jae-Han Park, Seung-Ho Baeg, Moon-Hong Baeg*
*Sensing and Perception Research Group
Korea Institute of Industrial Technology (KITECH)

요 약

자동으로 이미지 안에 존재하는 객체들을 인식하는 문제는 내용 기반 이미지 검색이나 로봇 비전과 같은 다양한 분야들에서 매우 중요한 문제이다. 이 문제를 해결하기 위하여 본 논문에서는 객체의 주요 색상 정보들을 이용하여 실시간으로 이미지 안의 객체들을 인식하는 알고리즘을 제안한다. 본 논문에서 제안하는 방법의 전체적인 구조는 다음과 같다. 처음에 MPEG-7 색상 정보 기술자들 중 하나인 주요 색상 정보 기술자를 이용하여 객체의 주요 색상 정보들을 추출한다. 이 때 이 정보는 가우시안 색상 모델링을 통하여 빛이나 그림자와 같은 외부 환경 조건에 좀 더 강인한 색상 정보로 변환된다. 다음으로 변환된 색상 정보들을 기반으로 주요 객체와 입력 이미지와의 픽셀 값 차이를 계산하고, 임계값 이상의 값을 가지는 픽셀들을 제거한다. 마지막으로 입력 이미지에서 제거되지 않은 픽셀들을 기반으로 하나의 영역을 생성한다. 결론으로서, 본 논문에서는 제안된 방법에 대한 실험 평가들을 수행 및 분석하고 몇몇 한계점들에 대해서 알아본다. 또한 이 문제들을 해결하기 위한 앞으로의 연구 계획에 대해서 기술한다.

1. 서론

지난 10 년 동안, 인터넷과 무선 장치의 급속한 발전으로 인하여 이용 가능한 멀티미디어 컨텐츠가 수와 양 측면에서 급격하게 증가하였다. 현재 많은 사용자들과 애플리케이션들이 인터넷을 통하여 원하는 멀티미디어 컨텐츠들을 검색하고 브라우징하는 것이 가능하게 되었다.

현재의 이미지 검색 기술들은 크게 키워드 기반 방식과 내용 기반 방식으로 분류될 수 있다. 키워드 기반 방식에서는 이미지에 대한 주석을 생성하기 위해서 해당 분야의 전문가들이 직접 이미지를 보고 제한된 키워드들을 이용하여 주석을 달아야 한다 [1, 2]. 비록 이 방식이 의미 검색의 가능성을 제공하기는 하지만 해당 분야의 전문가들이 직접 이미지들에 대한 키워드들을 달아야 하므로 대량의 이미지들에 대해서 주석들을 생성하는 것은 매우 많은 시간과 노력을 필요로 한다. 또한, 이미지를 바라보는 전문가의 관점에 따라 서로 다른 키워드들이 동일 이미지들에 대해서 생성될 수 있다.

키워드 기반 방식과 다르게 내용 기반 방식 시스템들은 반-자동으로 추출되고 저장되는 색상 또는 무늬와 같은 시각 정보들의 시각적 유사성을 기반으로 사용자가 원하는 이미지들을 검색한다 [3, 4]. 비록 이 방식이 키워드 기반 방식보다 시간 소모가 적지만, 문제는 시각적 유사성이 반드시 의미적 유사성을 의미하지는 않는다는 것이다. 즉, 시각 정보와 의미 정보 사이에 '의미 불일치' 문제가 존재할 수도 있다.

많은 연구자들이 자동 객체 인식 문제를 해결하기 위하여 다양한 방법들을 연구하고 있지만 여전히 이 문제는 해결하기 어려운 문제들 중 하나로 남아있다.

[5]의 연구에서 저자들은 자동으로 객체를 인식하기

위하여 통계 모델링 방식을 이용한다. 특히 이미지에 대해서 좀 더 풍부한 표현이 가능하도록 하기 위해서 개념 지향 객체 (concept-sensitive sailent object)를 이용한다. 그러나 제안된 알고리즘에서 최적 성능을 보증하기 위해서는 충분한 학습 데이터 집합이 요구되지만 이것들은 새로운 도메인에 쉽게 적응될 수 없다는 문제점이 있다.

[6]의 연구에서 저자들은 이미지의 객체 정보를 기술하기 위하여 객체 온톨로지와 중간 단계 기술 값을 이용하는 방법을 제안한다. 처음에 이미지는 K-Means-with-connectivity-constraint 알고리즘을 이용하여 다수의 영역들로 분할되고, 이 때 각 영역에 대한 시각 정보들이 자동으로 추출된다. 그 다음으로 추출된 시각 정보들은 사람이 인식할 수 있는 중간 단계 기술 값들로 사상된다. 이러한 중간 단계 기술 값들은 객체 온톨로지의 속성 값들로서 저장된다. 여기서 객체 온톨로지는 특정 객체에 대한 명세서이다. 본 관련 연구에서는 온톨로지를 이용함으로써 좀 더 정확하고 풍부한 의미 정보 표현이 가능하게 되었지만 객체 온톨로지는 오직 객체와 시각 정보들 사이에 포함 관계만을 다루기 때문에 좀 더 상세한 기술을 위해서는 객체와 시각 정보들 사이의 관계뿐만 아니라 객체와 객체의 관계, 그리고 시각 정보와 시각 정보 사이의 관계에 대해서도 정의할 필요가 있다.

내용기반 이미지 검색이나 로봇 비전에 활용되기 위한 객체 검출 문제에서는 의미 불일치 문제뿐만 아니라 이외에도 환경 조건의 제약을 고려해야 한다. 다시 말해서, 실제 환경에서의 객체 검출은 광원의 불균일성에 따른 조도 변화, 카메라와 객체 간의 거리에 따른 스케일 변화, 그리고 공간 상에서 객체의 방향에 따른 입력 이미지의 변화 등이 중요 변수이다. 따라서 이 제약 조건들에 강인한 객체 검출 기법을 개발하는 문제가 매우 중요하다.

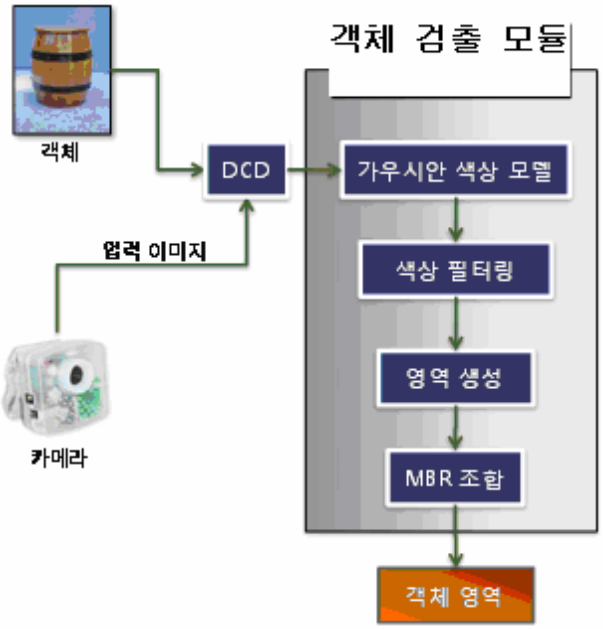
본 논문에서는 자동으로 이미지 안의 객체들을 검출하기 위해서 MPEG-7 시각 정보 기술자 들 중 하나인 주요 색상 정보 기술자 (Dominant Color Descriptor; DCD)를 이용하여 객체의 주요 색상 정보들을 추출하고, 다음으로 색상 정보가 조도 변화에 좀 더 강인할 수 있도록 가우시안 색상 모델 (Gaussian Color Model)로 변환한다. 여기서 시각 정보로서 색상 정보를 이용하는 이유는 다양한 시각 정보들 중에서 색상 정보는 에지나 모양 등과 같은 다른 시각 정보들에 비해서 위에서 언급한 방향에 따른 입력 이미지 변화와 스케일 변화에 덜 민감하기 때문이다 [10]. 반면에 이것은 조도 변화에 매우 민감한 특징을 갖는데, 이는 가우시안 색상 모델을 이용하여 해결한다. 마지막으로 변환된 색상을 기반으로 입력 이미지에서 해당 객체를 자동으로 검출하는 효율적인 방법을 제안한다.

본 논문의 전체적인 구성은 다음과 같다. 2 장에서는 본 연구에서 제안하는 방법의 전체적인 구조와 제안된 방법에 대해서 상세히 기술하며, 3 장에서는 본 알고리즘에 대한 성능 평가를 실시한다. 또한 연구를 통하여 얻은 결과들과 앞으로의 연구 계획에 대해서 기술한다.

2. 실시간 객체 인식

앞서 언급했듯이, 본 장에서는 제안하는 방법의 전체적인 구조와 이에 대해서 간략하게 설명하고, 세부적으로 각각의 모듈들에 대해서 상세히 기술한다.

(그림 1)은 본 논문에서 제안하는 방법의 전체적인 구조를 보여준다. 그림에서 보는 바와 같이, 객체 인식 모듈은 크게 4 단계로 구성된다. 처음에 가우시안 색상 모델은 색상 정보 기술자(DCD)로부터 추출된 객체의 주요 색상 정보 $DC_{0\sim1}$ 들을 그림자나 하이라이트와 같은 외부 조건들에 좀 더 강인하도록 만들기 위해서 이용된다. 다음으로 색상 필터링 단계에서는 $DC_{0\sim1}$ 을 기반으로 입력 이미지에서 $DC_{0\sim1}$ 과 유사한 색상 정보들을 가지는 픽셀들만을 추출한다. 이를 위해서 본 논문에서는 유클리디안 거리 함수를 이용한다. 또한, 추출된 픽셀들을 하나 이상의 영역들로 구성하기 위하여 8-connectivity labeling 알고리즘을 수행한다. 마지막으로 생성된 영역들 중에서 겹치는 영역들을 다시 하나의 영역으로 조합하고, 이 영역을 해당 객체가 존재하는 영역으로서 인식한다.



(그림 1) 전체적인 객체 영역 검출 과정

2.1 주요 색상 정보 기술자

직관적으로, 하나의 객체는 보통 1~3 종류의 주요 색상들로 이루어진다. 예를 들어, 축구공은 검정색과 하얀색으로 이루어질 수 있다.

이러한 특성을 이용하기 위하여, 본 논문에서는 주요 색상 정보 기술자(DCD)를 사용한다. MPEG-7 색상 정보 기술자 들 중 하나인 주요 색상 정보 기술자는 이미지 또는 특정 이미지 영역 안에 대표 색상들을 추출하고, 이를 명세하기 위한 간단한 기술서를 제공한다 [7].

주요 색상 정보 기술자를 통해서 생성되는 명세서는 다음과 같은 형식으로 표현된다.

$$F = \{(c_i, p_i, v_i), s\}, \ (i = 1, 2, ..., N)$$

여기서 N 은 주요 색상들의 갯수를 의미하며 각 주요 색상 정보 값 c_i 는 해당 색상 공간(여기서는 RGB 색상 공간)의 벡터이다. 그리고 p_i 는 c_i 값을 가지는 이미지 안에 픽셀들의 백분율을 나타내며, v 는 하나의 대표 색상을 나타내는 클러스터 안에 존재하는 픽셀들의 색상 값들의 분산을 나타낸다. 마지막으로 공간 응집도 s 는 이미지 안에서 대표 색상의 전체 공간 상의 동질성을 의미한다.

본 연구에서는 계산상의 효율성을 위하여 N 개의 대표 색상들이 아닌 가장 높은 p_i 를 차지하는 2 개의 대표 색상들 $DC_{0\sim1}$ 만을 이용한다.

2.2 가우시안 색상 모델

일반적으로 RGB 색상 공간은 광도나 그늘 같은 외부 환경 조건에 매우 민감하다. 따라서 색상 정보를 이용한 객체 인식에 있어서 RGB 색상 공간을 그대로 이용하는 것은 한계가 있다. 따라서 이러한 외부 환경으로부터의 영향을 줄이기 위하여 RGB 색상 모델은 CIE L*a*b, HSV, 그리고 YCbCr 같은 서로 다른 색상 모델로 변환될 필요가 있다. 본 연구에서는 이를 위하여 [8]에서 제안된 가우시안 색상 모델 (Gaussian Color Model)을 이용한다.

[8]의 연구에서 저자들은 가우시안 스케일 공간 이론을 기반으로 공간 상의 색상 정보를 측정하기 위한 프레임워크를 제안하며, 이를 위하여 하나의 통합 모델 안에서 색상과 공간 정보를 이용하는 가우시안 색상 모델을 고려한다. 특히 이 색상 모델은 그림자나 하이라이트와 같은 이미지의 명암부에 덜 민감한 특징을 가진다.

RGB 색상 공간을 가우시안 색상 모델로 변환하기 위한 최적의 선형 변환으로서 저자들은 아래와 같은 행렬 계수들을 제안하며, 본 연구에서도 이 계수들을 이용하여 카메라로부터 입력받은 이미지들과 객체의 $DC_{0\sim1}$ 을 가우시안 색상 모델로 변환한다.

$$\begin{bmatrix} \hat{E} \\ \hat{E}_\lambda \\ \hat{E}_{\lambda\lambda} \end{bmatrix} = \begin{pmatrix} -0.48 & 1.2 & 0.28 \\ 0.48 & 0 & -0.4 \\ 1.18 & -1.3 & 0 \end{pmatrix} \begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{Z} \end{bmatrix}$$

$$= \begin{pmatrix} -0.48 & 1.2 & 0.28 \\ 0.48 & 0 & -0.4 \\ 1.18 & -1.3 & 0 \end{pmatrix} \begin{pmatrix} 0.62 & 0.11 & 0.19 \\ 0.3 & 0.56 & 0.05 \\ -0.01 & 0.03 & 1.11 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$= \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

여기서 $\hat{E}, \hat{E}_\lambda$, 그리고 $\hat{E}_{\lambda\lambda}$ 는 각각 스펙트럼의 세기, 이에 대한 1 차 미분계수, 그리고 2 차 미분계수를 의미한다.

2.3 색상 필터링

본 장에서는 등식 (1.1)에 의하여 가우시안 색상 모델을 통하여 변환된 입력 이미지 I 의 픽셀들 중에서 $DC_{0\sim1}$ 과 유사한 색상 값들을 가지는 픽셀들만을 추출한다.

$$s = \begin{cases} 1 & if \ \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \geq \theta \\ 0 & otherwise \end{cases} \qquad (1.1)$$

여기서 θ 는 임계값을 나타내며, p_i 와 q_i 는 각각 $DC_{0\sim1}$ 과 I 의 공간 상의 색상 정보들을 의미한다. 실험적으로 θ 가 0.152 일 때, 실험 결과가 제일 좋았다.

결과적으로, 색상 필터링 과정을 통하여 I 는 각각의 $DC_{0\sim1}$ 에 대응하는 두 개의 이진 이미지 $I_{0\sim1}$ 이 생성된다.

2.4 영역 생성

객체 영역을 식별하기 위해서는 두 개의 이진 이미지 $I_{0\sim1}$ 에 대해서 색상 필터링 과정을 통하여 1 로 사상된 픽셀들에 대한 영역들을 생성할 필요가 있다. 이를 위해, 본 연구에서는 [9]의 연구에서 제안된 알고리즘을 이용한다.

[9]의 연구에서 저자들은 영역들을 생성하기 위하여, 처음에 블럽(blob)이라고 불리는 8-connect 컴포넌트들을 추출한다. 다음으로 최소/최대 영역을 임계값들로 하여 처음에 추출된 블럽들 중에서 해당 범위에 속하지 않는 블럽들을 제거한다.

2.5 MBR (Minimum Bounding Rectangle) 조합

객체 영역 인식을 위한 마지막 단계로서 영역 생성 과정을 통하여 추출된 블럽들 중에서 공간 상에서 서로 교차하는 블럽들을 하나의 블럽으로 조합한다. 다시 말해서, 각각 $I_{0\sim1}$ 에 속하는 블럽들 A 와 B 들 중에서 서로 겹치는 블럽들을 하나의 최종 블럽 $Rect$ 로 조합함으로써 객체 영역 인식을 수행한다.

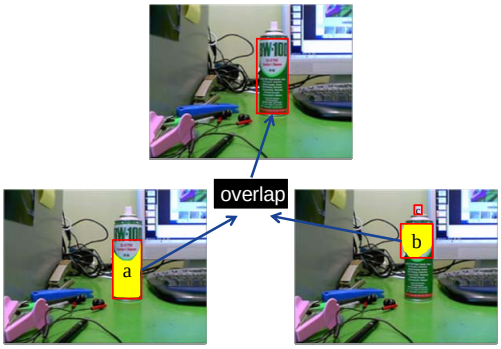
아래 등식 (1.2)는 이를 위한 등식을 나타낸다.

```
IF
 ((Bxs <= Axs || Bxs <= Axe)&&(Bxe >= Axe || Bxe >= Axs)
 &&
 (Bys <= Ays || Bys <= Aye)&&(Bye >= Aye || Bye >= Ays));
THEN
 Rectx = min(Axs, Bxs);
 Recty = min(Ays, Bys);
 Rectwidth = abs(Rectx - Rectxe);
 Rectheight = abs(Recty - Rectye);
```

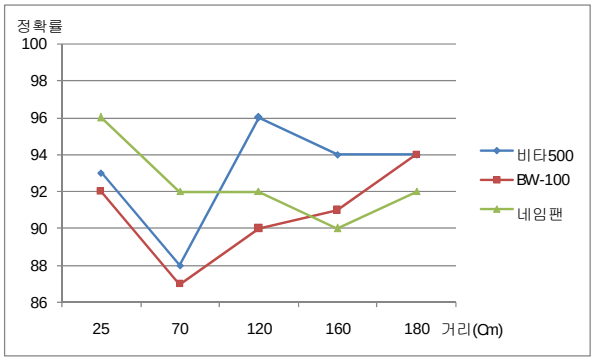
(1.2)

여기서 각각의 xs, xe 는 x 축 상에서의 시작과 끝점의 위치를 나타내며 ys 와 ye 또한 이와 동일하다.

(그림 2)는 MBR 조합에 대한 예제 그림으로서 3 개의 블럽 a, b, c 중에서 a 와 b 가 공간 상에서 겹치므로 이를 하나의 블럽으로 조합한다.



(그림 2) 세 블럽들 a, b, c 에 대한 조합



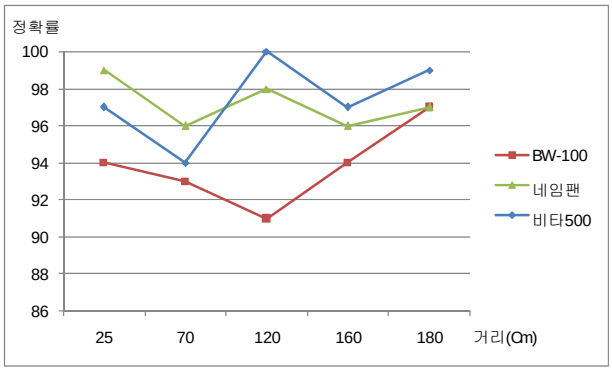
(그림 4) 거리 변화에 따른 객체 검출률(L2)

3. 실험 평가

본 연구에서는 실시간 객체 인식에 대한 성능 평가로서 일반 USB 카메라를 이용하여 이미지들을 입력받았으며, 시험 객체들로서 주요 색상들이 확연히 드러나는 총 3 개(즉, 스프레이의 일종인 BW-100, 비타500, 그리고 네임펜)의 객체들을 선택하였다.

본 연구에서는 두 가지 측면(즉, 조도와 거리 변화)에서 성능을 평가하였다. 즉, 세기가 다른 두 광원 환경에서 카메라와 객체 간의 거리 변화에 따라 제안된 알고리즘이 얼마나 강인한지를 측정하였다. 위 두 가지 측정 기준 선택 이유는 앞서 1 장에서 언급했듯이 색상 정보가 다른 시각 정보(즉, 에지 또는 모양)들에 비해서 조도 변화에 민감하기 때문이다.

(그림 3)과 (그림 4)는 서로 다른 조도를 가지는 환경에서 거리에 따른 객체 영역의 검출 정확성을 보여주며, (그림 3)의 조도(L1)를 (그림 4)보다 높게 설정하였다. 본 실험에서는 각 거리마다 100 장의 이미지들을 추출하였으며, 이 이미지들을 기반으로 추출된 블럽들 중에서 시험 객체와 일치하는 블럽들의 수를 측정하였다. 그림(3)에서 보는 바와 같이 제안된 알고리즘이 우수한 성능 결과를 보여주었다. 또한 '네임펜'을 제외한 두 객체들의 경우, 이미지 스케일 변화로 인하여, 거리가 180cm 일 때 25cm 보다 나은 결과를 보여준다.



(그림 3) 거리 변화에 따른 객체 검출률(L1)

(그림 4)는 (그림 3)의 조도 세기보다 낮은 환경(L2)에서의 검출 결과를 보여준다. (그림 4)에서 보는 바와 같이, 조명이 어두워지면 검출률이 다소 낮아지는 경향을 보이나 성능 면에서 보면 로봇 환경에 적용하기에 크게 지장이 없는 수준이다.

4. 결론

본 연구에서는 실시간 객체 검출을 위하여 객체의 주요 색상 정보들을 이용하였다. 특히, MPEG-7 표준에 정의된 주요 색상 정보 기술자를 이용함으로써 계산상의 효율성을 고려하였으며, 실험 평가를 통하여 여러 객체들에 대해서 만족할 만한 객체 검출 성공률을 확인하였다. 또한, 조도가 낮아짐으로써 성능의 변화가 있기는 하지만 로봇이 운영되는 조도 변화가 심하지 않은 환경에서 사용 가능한 만족스러운 결과라 할 수 있다. 물론 앞으로도 조도 변화에 좀 더 강인한 방법들에 대해서 연구할 필요가 있으며, 색상이 아닌 무늬, 모양 등의 다른 시각 정보들에 대해서도 고려할 필요가 있다.

참고문헌

[1] W. E. Mackay, "EVA: An experimental video annotator for symbolic analysis of video data", SIGCHI Bulletin, Vol. 21 (1989) 68-71

[2] Eitetsu Oomoto and Katsumi Tanaka, "OVID: Design and Implementation of a Video-Object Database System", IEEE TKDE, Vol. 5 (1993) 629-643

[3] John R. Smith and Shih-Fu Chang, "VisualSEEK: a fully automated content-based image query system", ACM Multimedia 96 (1996)

[4] Chad Carson, Megan Thomas, Serge Belongie, Joseph M. Hellerstein and Jitendra Malik, "Blobworld: A System for Region-Based Image Indexing and Retrieval", Third Intl. Conference on Visual Information Systems, (1999)

[5] Jianping Fan, Yuli Gao, Hangzai Luo and Guangyou Xu, "Statistical modeling and conceptualization of natural images", Pattern Recognition (2005) 865-885

[6] Vasileios Mezaris et al., "Region-based Image Retrieval using an Object Ontology and Relevance Feedback", EURASIP JASP, (2002)

[7] BS Manjunath, Philippe Salembier and Thomas Sikora, " Introduction to MPEG-7", (2002)

[8] Jan-Mark Geusebroek and Rein van den Boomgaard and Arnold W.M. Smeulders and Hugo Geerts, "Color Invariance", IEEE Trans. Pattern Analysis and Machine Intelligence, (2001)

[9] W. Synder and A. Cowart, "An iterative approach to region growing", IEEE Transactions on Pattern Analysis and Machine Intelligence, (1983)

[10] M. Swain and D. Ballad, "Color Indexing", Intl. J. of Computer Vision, (1991) 21-27