

WordNet 기반 개념적 이미지 주석 시스템 설계

조미영*, 최준호, 김판구
조선대학교 전자계산학과

Design of Conceptual Image Annotation System Using WordNet

Mi-Young Cho, Jun-Ho Choi, Pan-Koo Kim

Dept. of Computer Science, Chosun University

E-mail : irune80@hanmail.net

요약

이미지검색을 위해서 객체의 시각적인 특징에 대한 저차원의 특징 정보를 추출하고 이미지에 의미를 부여하기 위하여 주석을다는 것이 일반적이다. 하지만 주석 기반 검색에서는 주석으로 달아 놓은 단어와 정확한 매칭이 없다면 찾을 수가 없다. 이러한 문제를 해결하기 위해 재질의나 질의어 확장과 같은 기법을 써서 문제를 해결해 왔으나 여전히 개념적 매칭이 아닌 스트링 매칭의 문제를 안고 있다고 볼 수 있다.

이에 본 논문에서는 이미지 관련 Text에서 단어를 추출한 후 추출된 단어들간의 개념 관계를 WordNet을 이용하여 표현한 주석 시스템을 제안한다. 이 시스템은 단순 스트링 매칭이 아닌 개념적 매칭에 의한 개념 기반 검색을 지원할 수 있다.

1. 서론

이미지는 멀티미디어에서 가장 많이 사용되는 자료형으로 다양한 분야에서 많은 이미지가 생성되고 있다. 이러한 많은 이미지를 중에서 원하는 이미지를 어떻게 빠르고 정확하게 찾을 것인가 하는 물음이 제기 되었고, 이에 따라 여러 방법들이 현재 연구되어 오고 있다. 대표적 검색 시스템으로는 query by examples에 의한 Photobook, query by sketches에 의한 QBIC 등이 있다.

기존의 이미지 검색은 크게 내용 기반 검색과 주석 기반 검색으로 분류된다. 내용 기반 검색은 color, texture, edge 등과 같은 저차원의 특징정보를 자동으로 추출하여 색인한 다음, 사용자 질의로 제시되는 이미지의 특징들과 매칭함으로써 관련 이미지를 검색한다. 이 시스템은 이미지를 특징 정보들만으로 색인하고 검색하기 때문에 사용자의 개념 질의를 처리할 수 없다는 단점을 가지고 있다. 이를 보완하기 위해 모든 이미지에 대해 주석을 전문가가 직접 기술하는 주석 기반 검색을 이용한다.

주석 기반 검색은 텍스트 정보 검색 기술을 이용하여 이미지를 색인하고 검색할 수 있다. 그러나, 이런 주석 기반 검색은 주석으로 달아 놓은 단어와 정확한 매칭이 없다면 찾을 수가 없다. 이는 Thesaurus와 같은 용어사전을 사용해 용어 재쓰기(rewriting)기법, 질의어 확장(query expanding), Relevance Feedback과 같은 기법을 써서 문제를 해결해 왔다. 이런 방법은 시스템의 정확도에 상당한 문제가 있으며, 여전히 개념적 매칭이 아니 스트링 매칭의 문제를 안고 있다.

논문에서는 이미지 관련 Text에서 단어를 추출한 후 추출된 단어들간의 개념 관계를 WordNet을 이용하여 표현한 주석 시스템을 제안한다.

논문의 구성은 2장에서 이미지에 대한 주요 특징 정보에 대해 알아보고, 3장에서는 이미지 주석 시스템을 제안한다. 4장에서는 핵심어 추출기, 유사도 측정기에 의한 Semantic-feature정보를 5장에서는 객체 추출기, 중요도 계산기에 의한 visual-feature정보를 추출한다. 6장에서는 주석 데이터 생성 예를 보여 주

고 마지막 7장에서 결론을 맺는다.

2. 이미지에 대한 주요 특징 정보

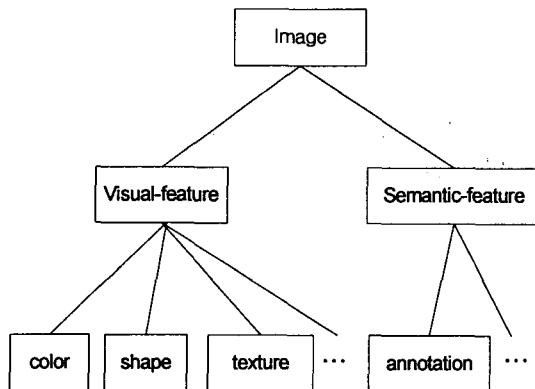


그림 1. Image feature

이미지에 대한 정보로는 크게 color, shape, texture 등의 눈에 보이는 시각적인 특징과 annotation 등의 의미적 특징으로 나눌 수 있다. 이러한 특징을 이용한 이미지 주석 시스템은 크게 3가지가 있다.

첫 번째는 단순한 이미지 파일의 시각적인 특징만을 추출하여 주석 처리한 것으로 이러한 시스템은 특징들을 자동으로 추출 할 수는 있지만 이미지가 내포하고 있는 의미적인 내용을 표현하지 못한다. 따라서 query by examples, query by sketches 등과 같은 단순한 질의밖에 하지 못한다.

두 번째는 의미적 특징만을 이용한 이미지 주석 시스템은 키워드, 추상화된 단어나 텍스트 형태의 의미 정보 등과 같이 자동 인식이 어려운 좀더 구체적인 요소를 수동으로 추출해 낼 수 있다. 하지만 이 방법은 서술 정보에 사용되는 어휘가 사용자 질의의 표현 어휘와 정확히 일치하지 않을 수 있을 뿐 아니라 서술자의 주관적 견해를 배제하지 못한다.

마지막으로 세 번째는 위의 두 방식을 혼용한 시각적인 특징과 의미적 특징을 모두 추출하는 방식이다. 이는 이미지 처리 기법을 통하여 몇 가지 특징을 추출해 내고, 자동으로 추출이 불가능한 서술 정보를 수동으로 기술해 준다.

본 논문에서는 세 번째 방식으로 주석을 달되 객관성을 주기 위해 이미지에 대한 관련 Text가 있는 HTML문서를 분석해 이미지 주석을 달고자 한다.

3. 이미지 주석 시스템 제안

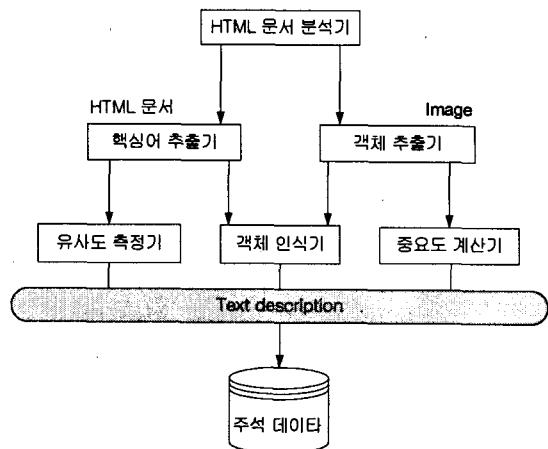


그림 2. 시스템의 전체 구조

이미지 주석 시스템은 크게 HTML문서 분석기, 핵심어 추출기, 객체 추출기와 주석 데이터를 생성하기 위한 유사도 측정기, 객체 인식기, 중요도 계산기로 구성되며 구조는 그림 2와 같다.

HTML문서 분석기는 HTML문서를 분석하여 이미지와 HTML문서를 추출하고, 이를 각각을 객체 추출기와 핵심어 추출기로 넘겨준다. 핵심어 추출기에서는 HTML문서에서 이미지 관련 핵심 단어를 추출하고 추출된 핵심 단어간의 유사도를 측정한다. 객체 추출기에서는 웨이블릿 변환을 이용하여 객체의 형태를 추출한 후 객체의 크기와 이미지 중심에서 떨어진 거리를 이용하여 객체의 중요도를 계산한다. 객체 인식기에서는 추출된 객체와 핵심어로 이미지내의 하나의 객체로 인식시킨다. 마지막으로 유사도 측정기, 객체 인식기, 중요도 계산기에 의해 나온 데이터를 Text description으로 나타낸 후 주석 데이터로 저장한다.

4. Semantic-feature 추출

4.1 핵심어 추출기

일반적인 문서상에서 핵심어를 추출하는 방법은 단어의 빈도수를 고려하여 추출하는 TF·IDF 방법이 많이 사용되고 있다. 이는 문서에서의 빈도수가 높은 단어가 핵심어가 아니라 단어가 들어있는 문서들 중 비례적으로 빈도수가 높은 단어가 그 문서의 핵심어가 된다. 하지만 HTML문서에서는 단순히 빈도수만을 고려해서는 이미지 관련 핵심어를 추출할 수 없다.

이에 본 논문에서는 HTML문서 분석기에 의해 추출된 HTML문서에서 이미지와 관련된 핵심 단어를

추출하기 위해 다음과 같은 Tag를 중심으로 HTML 문서의 핵심어를 추출한다.

```
<a href=" "> _____ ① </a>  
<img src= ... alt=" _____ ② " >  
<title> _____ ③ </title>
```

일반적으로 특정 문서 P 내에서 핵심어 k 에 대한 평가함수 Eval 은 다음과 같이 정의할 수 있다.

$$Eval(P) = TF(P, k)$$

$TF(P,k)$: 문서 P 에 k 가 나타나는 빈도수

하지만, 이는 특정 키워드에 대한 단순 빈도수에 해당하며 단어의 핵심적인 역할에 대한 빈도는 파악할 수 없다. 따라서, HTML의 특정 Tag를 고려한 평가함수를 다음과 같이 정의하였다.

$$Eval'(P) = \sum TV(P, k, t)$$

$TV(P, k, t)$: 문서 P 에 k 가 태그 t 를 사용할 때의 빈도수

4.2 단어간의 유사도 측정기

이 절에서는 하나의 이미지 내의 핵심 단어들 간의 개념적 관계를 WordNet을 이용하여 유사도를 측정하는 유사도 측정기기에 대해 설명한다.

WordNet의 기본 골격은 어휘개념으로 동의어 집합(synonym set, synset)으로 표현되고 있으며, 이 동의어 집합사이의 상·하위개념 관계는 계층 구조로서 표현된다.

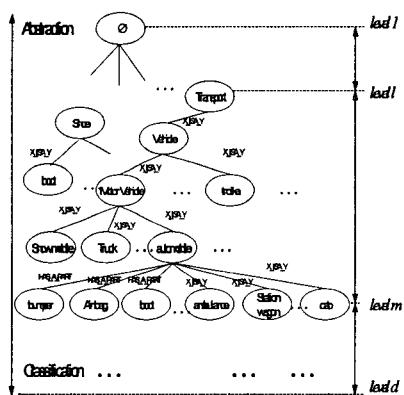


그림 3. “Automobile”의 WordNet 구조

WordNet의 명사 25개 범주 안에 대부분의 일반 명사들이 모두 포함될 수 있다는 분석 하에 본 연구에서는 의미지표 테이블을 구축 시 WordNet의 명사 25개 범주를 사용한다. 의미지표 테이블은 WordNet

의 명사 25개 범주에서 각 용용 도메인에 따라 의미 지표 레벨을 선택하여 사용할 수 있도록 2단계의 의미지표 레벨로 구축하였으며 1-level은 19개, 2-level은 162개로 구성되어 있다.

그림 3에서 알 수 있듯이 각 레벨간에 구성된 관계는 여러가지가 존재하며, 이들간의 연관성 측정을 위한 척도개발이 필요하다. 본 연구에서는 그림 4와 같이 각 개념들간의 Travel Cost를 고려하는 방법을 사용하였다.

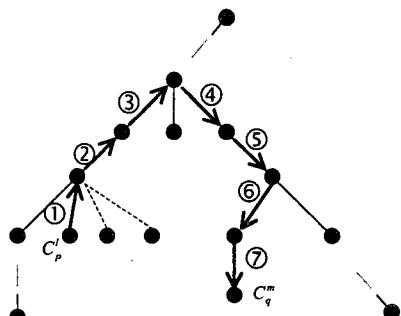
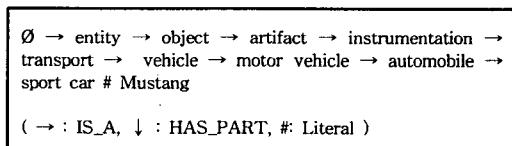


그림 4. Weighted Graph Travel Cost

또한, 검색을 위한 다의성 문제를 해결하기 위해 추가 정보로 Concept Type을 이용한다. 예를 들어, 키워드 검색 방식에서는 'Mustang'이라는 단어 자체에 대한 의미가 '스포츠카'인지 '말'의 종류인지는 분류할 수가 없다. 이를 위해서는 검색자의 추가적인 정보가 입력되어야 한다. 이러한 추가 입력 정보를 Concept Type을 이용한다면 검색 시 주석 처리 단계에서 검색자는 검색 결과에 대한 적당한 개념의 범주를 선택해야 한다. 검색자가 'Mustang'의 개념을 '스포츠카'로 정한다면 키워드 인덱싱에서 root로부터 '스포츠카'로의 개념적 패스로 검색을 하게 된다.



```

graph TD
    transport[...transport] --> vehicle[vehicle]
    vehicle --> motorVehicle[motor vehicle]
    motorVehicle --> automobile[automobile]
    motorVehicle --> ambulance[ambulance]
    automobile --> footwear[...footwear]
    automobile --> shoe[shoe]
    footwear --> boot[boot]
    shoe --> boot
    motorVehicle --> boot
  
```

그림 5. 색인된 Term과 Path

키워드를 검색하기 위해서는 DB 내용의 분류와 인덱스 구조가 생성이 되어야 하는데, 본 연구에서는 Term과 Path를 포함한 인덱스 구조를 개념적 계층 구조를 통해 검색하도록 하였다. 검색은 개념의 확장과 축소의 두 연산을 기반으로 한다.



그림 6. 개념적 유사도 측정

유사 비교에서의 거리는 $C_p^l = "automobile\#boot"$ 와 $C_q^m = "footware\#boot"$ 사이의 경로 검색을 계산한다. 유사도 측정의 정의에 의하여 $S^s(\text{automobile}\#\text{boot}, \text{footware}\#\text{boot})$ 는 $(91.5+92+93+94+95+96)+(96+95+94) = 328.5$ 의 값을 산출할 수 있다.

5. Visual-feature 추출

5.1 객체 추출기

웨이블릿 변환은 웨이블릿이라 불리는 기저함수를 이동 및 확장하여 주파수 영역별로 다른 해상도를 갖게 할 수 있다. 이러한 웨이블릿 변환은 다양한 분야에서 사용되고 있으며 근래에 들어서 이미지의 처리 분야에서 좋은 효과를 주고 있다. 일반적인 수식은 다음과 같다.

$$WT_{f(a,b)} = \int_{-\infty}^{+\infty} \phi((t-b)/a) f(t) dt$$

HTML 문서에서 추출된 이미지으로부터 웨이블릿 변환을 수행하면 저주파 부대역과 방향(수평, 수직, 대각선)성분을 갖는 3개의 고주파 부대역을 생성시킬 수 있다. 각각의 부대역들은 다해상도 분해에 의해 압축되어 검색 효율이 향상되고, 같은 해상도와 위치의 영상정보를 소유한 고주파 부대역(LH, HL, HH)의 조합으로 윤곽선을 얻을 수 있다. 추출된 윤곽선은 객체의 윤곽선과 배경을 포함하게 되는데, 본 논문에서의 객체의 추출이기에 배경은 잡음으로 간주하여 제거한다. 이를 위해서는 에지 정보를 구성하는 고주파 부대역의 웨이블릿 계수값을 제곱하여 영상의 평균에너지 값보다 작은 값은 제거함으로써 배경을 분리해낼 수 있다.

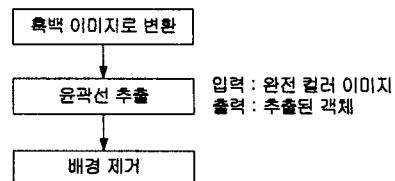


그림 7. 객체 추출 과정

5.2 객체 중요도 계산기

객체 중요도는 이미지내의 객체의 중요성을 나타내기 위한 것이다. 객체의 중요성을 나타내기 위한 다른 많은 특징들도 있으나 본 논문에서는 객체의 중요도를 객체의 크기(area)와 그것의 위치(closeness to the center)로 표현한다. 물론 이미지 내에 작은 객체가 큰 객체보다 더 큰 중요도를 가질 수 있다. 하지만 본 논문에서는 같은 이미지내에서 큰 객체는 작은 객체보다 더 큰 중요도를 갖는다고 가정한다.

객체 크기에 의한 중요성은 먼저 이미지 전체의 크기를 구하고 앞 절에서 설명한 웨이블릿을 이용하여 추출된 각 객체의 크기를 계산한다. 객체 중요성을 이미지 전체 크기에 대한 각 객체 크기의 비로 표현한다.

객체의 위치에 따른 중요성을 표현하기 위해 이미지를 5개의 똑같은 크기 영역으로 나눈다. 일반적으로 이미지의 중앙 부분에 있는 객체가 중요하다고 볼 수 있으므로 중앙 영역에 높은 가중치를 준다.

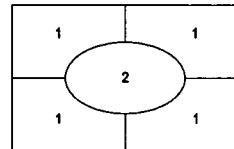


그림 8. 영역에 따른 가중치

객체의 중요도는 객체 크기, 위치에 의해 결정한다 했으므로

$$object(s_s, s_l)$$

으로 표현한다.

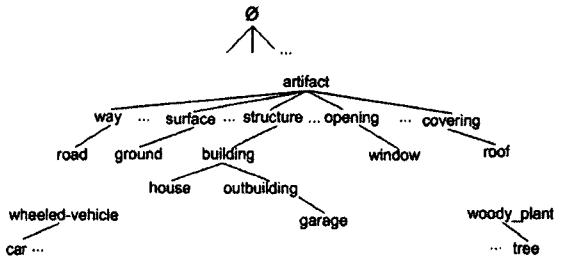
s_s - size에 의한 signification

s_l - location에 의한 signification

6. 이미지 주석 데이터 생성

6.1 객체 인식기

이 절에서는 5.1절에서 추출한 객체의 인식을 위해 4.1절에서 추출한 핵심어를 이용하여 라벨링 하는 객체 인식기에 대해 설명한다.



HTML 문서에 포함된 영상 안에서의 객체와 문서 내의 특정 키워드가 추출되면 영상의 객체를 인식시키기 위해 레이블링 작업이 필요하다.

본 논문에서는 레이블링 작업에 있어서 수작업이 가해지도록 설계하였다. 이는 모든 영상에 대해 수작업이 가해지는 어려움이 있지만, 정형화되지 않은 영상에 대해 정확한 인식을 할 수 있는 장점이 있다. 이를 위해 자동으로 추출된 객체와 핵심어에 대한 시각적인 레이블링 처리가 이루어지도록 하였다.

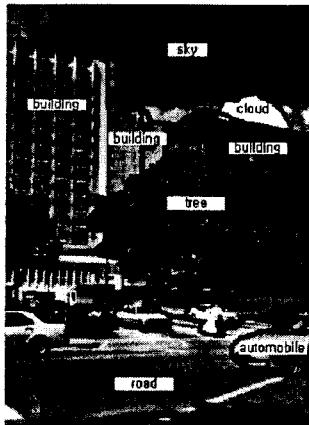


그림 9. 객체의 레이블링 결과

6.2 이미지 주석 데이터 생성의 예

주석 데이터 생성 예를 보이기 위해 그림 10과 같은 이미지가 포함된 웹문서가 있다고 가정한다.

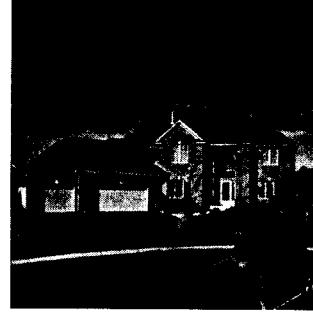


그림 10. sample 이미지

그림 10을 포함한 웹문서에서 추출한 핵심어로 tree, road, house, car 등이 있다면 추출된 핵심어를 WordNet을 이용하여 그림 11과 같은 트리 구조로 만든다. 그리고 이 트리 구조를 이용하여 핵심어간의 유사도를 측정한다.

그림 11. 트리 구조로 표현된 핵심어

영역 분할(그림 12)과 객체 크기에 의한 중요도를 계산한 후 수치로 표현한다.

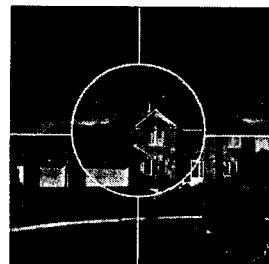


그림 12. 중요도 계산을 위한 영역 분할

추출한 객체와 핵심어간의 레이블링 작업을 통해 객체를 인식시킨다.(그림 13-a) 그리고 마지막으로 핵심어간 유사도 측정, 객체 중요도를 이용해 다음과 같은 주석 데이터(그림 13-b)를 생성한다.

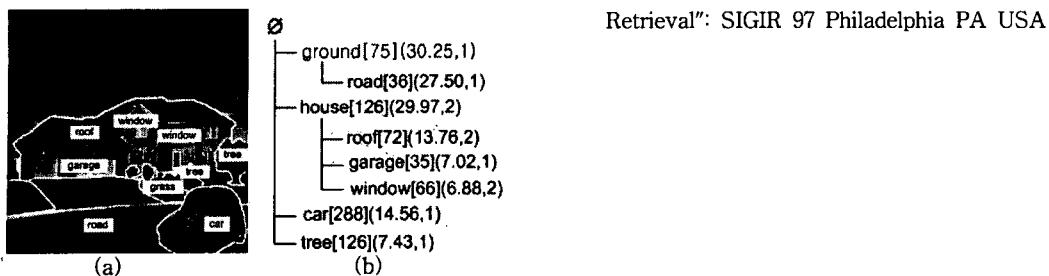


그림 13. 객체 인식으로 인한 주석 데이터 생성

7. 결론

본 논문에서는 ontology의 한 형태인 WordNet을 이용하여 이미지 내에 존재하는 객체들과 이미지 관련 핵심어들 간의 개념 관계를 표현하였다. 기존의 내용 기반 검색이나 주석 기반 검색은 개념적인 해석과 이를 기반으로 하는 검색이 이루어지지 못했다. 이에 본 논문에서 제시한 이미지 데이터를 활용한다면 개념 기반 검색이 가능 할 것이다.

향후 연구에서는 HTML 문서에서 특정 Tag에서만 핵심어를 추출한 경우 많은 이미지 관련 핵심어를 추출할 수 없었으므로 웹 문서에서 이미지 관련 text 추출에 대한 더 많은 연구가 필요하다.

[참고문헌]

- [1] Y.C. Park, F. Golshani, S. Panchanathan, "Conceptualization and Ontology: Tools for Efficient Storage and Retrieval of Semantic Visual Information", *Internet Multimedia Management Systems Conference*, Boston, MA, November 2000.
- [2] Biederman, I. "Recognition-by-components: A theory of human image understanding", *Psychological Review* 94:115-147
- [3] J. Z. Wang, J. Li, D. Chan, G. Wiederhold, "Semantics-sensitive Retrieval for Digital Picture Libraries" D-Lib. Magazine, Vol. 5, No. 11, November 1999.
- [4] Craig Dillon, Terry Caelli "Learning Image Annotation: The CITE System," journal of the VIDERE, Vol.1, No.2, 1998
- [5] Y. Alp Aslandogan, Chuck Thier, Clement T. Yu, Jon Zou, Naphtail Rishe "Using Semantic Contents and WordNet in Image