2002년 한국정보처리학회 춘계학술발표논문집 제9권 제1호

이미지 주석 시스템을 위한 의미 정보 모델링

최준호*, 곽효승*, 김원필*, 김판구*
*조선대학교 전자계산학과
e-mail : spica@hitel.net

Semantic Information Modeling for Image Annotation System

Jun-Ho Choi*, Hyo-Seung Kwak*, Won-Pil Kim*, Pan-Koo Kim*
*Dept. of Computer Science, Chosun University

요 약

의미 기반 영상 검색은 Color, Texture, Region 정보, Spatial Color Distribution등의 저차원 특징 정보와 이미지 데이터에 의미를 부여하기 위해 주석 처리하는 것이 일반적이다. 그리고 부여된 키워드나 시소러스와 같은 어휘 사전을 이용하여 의미기반 정보검색을 수행하고 있지만, 기존의 키워드기반 텍스트 정보검색의 한계를 벗어나지 못하는 문제를 야기 시킨다.

이에 본 논문에서는 시각 데이터에 존재하는 객체들과 그 객체 사이의 개념관계를 Ontology의 한 형태인 WordNet을 이용하여 의미 정보로 표현할 수 있도록 한다. 이를 활용하면 영상 데이터의 자동 주석 시스템이나 검색 시스템에서 인간이 인식하는 개념적인 사고방식에 더욱 접근할 수 있는 결과물을 얻을 수 있을 것이다.

1. 서론

최근 멀티미디어 관련 기술의 발전과 인터넷 사용의 증가로 영상의 사용이 증가함에 따라 영상을 효과적으로 검색하는 방법이 주요 관심분야로 대두되고 있다. 하지만, 지금까지 내용 기반 이미지 및 비디오 정보 검색에 있어서 완전한 내용인식에는 그 범위가 미치지 못하고 있고, 주로 Color, Size, Texture와 Shape등에 기초로 하기 때문에 시각적 특징을 벗어나 의미적 내용 인식까지는 많은 연구가 필요하다고 볼 수 있다. 기존 연구에서 시각 데이터의 개념적 분류, 사람의 인식, Indoor/Outdoor 장면 분류 등에 대한 의미 인식의 접근이 연구되어 왔지만, 아직까지는 시각 데이터가 가지는 의미적 요소들을 정확히 표현하지는 못하고 있다.

이에 본 연구에서는 이미지나 비디오와 같은 시각 데이터에 내포된 의미적(semantical) 요소를 표현해 주고, 그 내용을 검색하기 위한 개념 모델을 제시하고자 한다. 최근까지 연구되어온 내용을 살펴보면, 이미지 데이터의 경우 의미기반 정보검색을 위해서는 기존의 Color, Texture, Region 정보, Spatial Color Distribution등의 저차원(low-level) 특징 정보와 이미지 데이터에 의미(semantic)를 부여하기 위해 주석 처리하는 것이 일반적이다. 그리고 그 부여된 키워드를 가지고 Thesaurus와 같은 어휘 사전을 이용하여 의미기반 정보검색을 수행하고 있지만, 이런 경우 기존의 키워드기반 텍스트 정보검색의 한계를 벗어나지 못하는 문제가 발생될 수 있다. 따라서 본 논문에서는 WordNet 어휘 사전을 확장한 개념적(conceptual) 어휘 체계를 갖는 대형 Ontology를 기반으로 하여 영상 데이터 내의 객체 인식과 추출된 객체간의 관계를 정의할 수 있는 방법을 제시하고자 한다.

2. 영상 데이터의 객체 추출 및 레이블링

본 장에서는 영상 데이터의 객체를 추출하여 추출된 객체를 인식할 수 있는 방안에 대해 논의하고자 한다.

2.1 영상 데이터의 객체 추출

영상에서 객체를 추출하기 위해서 영상에 대해 웨이블릿 변환을 사용한다. 이를 통해 영상의 형태 특징을 추출하면 이를 영상 내의 하나의 객체로 인식시킬 수 있을 것이다.

2.1.1 Wavelet Transform

웨이블릿 변환은 웨이블릿이라 불리는 기저함수를 이동(translation) 및 확장(scaling)하여 주파수 영역별로 다른 해상도를 갖게할 수 있다. 웨이블릿 변환의 일반적인 수식은 다음과 같다.

$$WT_{f(a,b)} = \int_{\infty}^{-\infty} \varphi\left((t-b)/a\right) f(t)dt$$

2.1.2 형태 특징 추출

영상으로부터 웨이블릿 변환을 수행하면 저주파 부대역과 방향(수평, 수직, 대각선) 성분을 갖는 3개의 고주파 부대역을 생성시킬 수 있다. 각각의 부대역들은 다해상도 분해에 의해 압축되어 검색효율이 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 목적은 객체의 추출이기에 배경은 잡음으로 간주하여 제거한다. 이를 위해서는 에지정보를 구성하는 고주파 부대역의 웨이블릿 계수값을 제곱하여 영상의 평균 에너지 값보다 작은 값은 제거함으로써 배경을 분리해낼 수 있다.

(a) 원 영상 (b) 윤곽선 추출

그림 1. 형태 추출

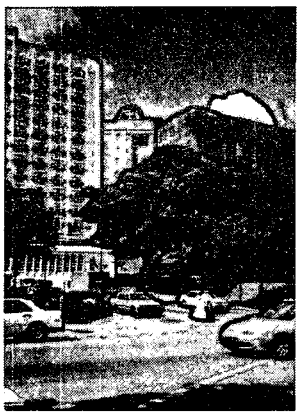

그림 2. 객체의 추출

2.2 HTML 문서상의 핵심어 추출

영상 데이터에서 추출된 각 객체에 레이블을 할당하여 이를 토대로 개념 인식의 기초로 삼는다. 이는 이미지 처리의 현 수준에서 사람이 인식하는 정도의 인식은 어려운 일이므로 본 논문에서는 영상과 관련된 단어를 이미지의 링크 정보가 포함된 HTML 문서 수집하여 이를 인덱싱 처리하여 각 객체에 레이블을 부여할 수 있도록 하였다.

문서상에서의 핵심어를 추출하는 방법은 단어의 빈도수를 고려하여 추출하는 TFIDF 방법이 많이 사용되고 있는데, 이는 문서에서의 빈도수가 높은 단어가 핵심어가 아니라 단어가 들어있는 문서들 중 비례적으로 빈도수가 높은 단어가 그 문서의 핵심어가 된다. 따라서, 메타 데이터의 특징을 고려하여 HTML에서 핵심어를 추출하기 위해서는 HTML의 특정 Tag에 가중치를 부여하도록 한다. 본 논문에서는 HTML 문서상의 핵심어를 추출하기 위해서 하이퍼링크된 텍스트와 이미지 관련 Tag, 그리고 메타 Tag를 중심으로 문서의 핵심어를 추출한다.

```
<a href=" ">_____ⓐ_____</a>
<img src= ··· alt="_____ⓑ_____">
<title>_____ⓒ_____</title>
```

일반적으로 특정 문서 P내에서 핵심어 k에 대한 평가함수 Eval은 다음과 같이 정의할 수 있다.

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

TF(P,k) : 문서 P에 k가 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 P에 k가 태그 t를 사용할때의 빈도수

2.3 객체 영역의 레이블링

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.

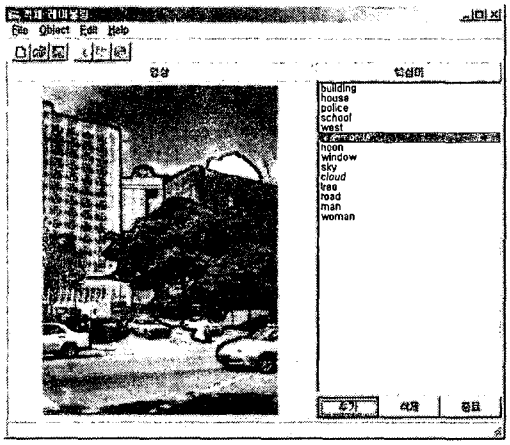

그림 3. 객체 레이블링 모듈 실행

다음 그림은 원 영상에 대해 객체의 레이블링 처리가 된 결과이다.

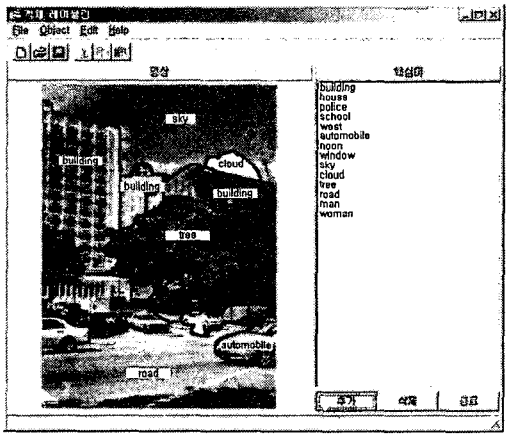

그림 4. 객체의 레이블링 결과

3. WordNet을 이용한 의미 정보 생성

3.1 WordNet의 개요

WordNet의 기본 골격은 어휘개념으로 동의어 집합(synonym set, synset)으로 표현되고 있으며, 이 동의어 집합사이의 상·하위개념 관계는 계층 구조로서 표현된다. WordNet은 명사뿐만 아니라 동사, 형용사, 부사의 엔트리에 IS_A 계층구조를 가지고 있는데 동사는 크게 15개의 범주로 분류하고 명사는 크게 25개의 범주로 분류하고 있다.

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

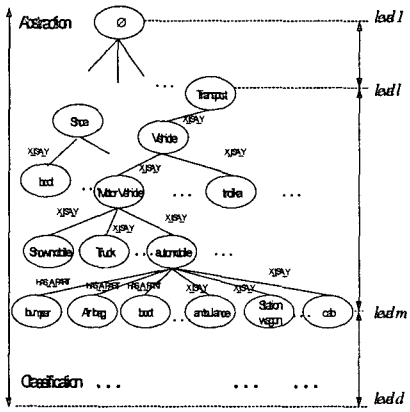

그림 5. "Automobile"의 WordNet 구조

3.2 단어 간의 유사도 측정

본 절에서는 하나의 영상 내의 각 객체에 매칭된 객체 레이블을 조합하여 영상의 의미 부여에 활용할 수 있는 방안을 제시한다.

그림 5에서 알 수 있듯이 각 레벨간에 구성된 관계는 여러가지가 존재하며, 이들간의 연관성 측정을 위한 척도개발이 필요하다. 본 연구에서는 그림 6과 같이 각 개념들간의 Travel Cost를 고려하는 방법을 사용하였다.

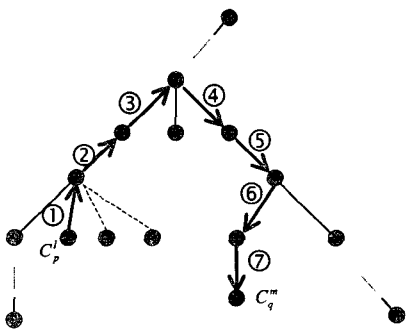

그림 7. Weighted Graph Travel Cost

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.

Ø → entity → object → artifact → instrumentation → transport → vehicle → motor vehicle → automobile → sport car # Mustang

(→ : IS_A, ↓ : HAS_PART, #: Literal)

..transport→vehicle →motor vehicle →automobile → ambulance

..footwear →shoe→ boot

..transport→vehicle →motor vehicle →automobile ↓ boot

그림 8. 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.

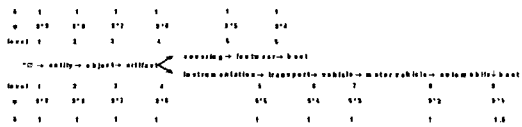

그림 9. 개념적 유사도 측정

유사 비교에서의 거리는 C^{t}_{p} ="automobile#boot" 와 C^{m}_{q} ="footware#boot" 사이의 경로 검색을 계산한다. 유사도 측정의 정의에 의하여 S^{s} (automobile#boot, footware#boot)는 (91.5+92+93+94+95+96)+(96+95+94) = 328.5의 값을 산출할 수 있다.

3.3 의미 정보 생성

본 논문에서 제시한 객체의 추출과 각 객체 사이의 의미 관계를 다음과 같이 XML DTD 구조로 정의할 수 있다.

```
<!ELEMENT visual-object (visual-object-name,char*)>
<!ELEMENT visual-object-name(#DATA)>
<!ELEMENT char (label?, cost?, similarity?, distance*)>
<!ELEMENT label(#DATA)>
<!ELEMENT cost(#DATA)>
<!ELEMENT similarity(#DATA)>
<!ELEMENT distance(#DATA)>
```

또한, 이미지의 시각적인 특징과 객체 추출을 통해 추출된 객체의 의미적인 특징의 연결을 위해 이미지와 의미 객체간의 매핑구조는 다음과 같다.

```
<!ELEMENT mapDB(map*)>
<!ELEMENT map(image?, visual*, semantic*)>
<!ELEMENT image EMPTY>
<!ATTLIST image src CDATA #REQUIRED size CDATA
    #IMPLIED object CDATA #IMPLIED>
```

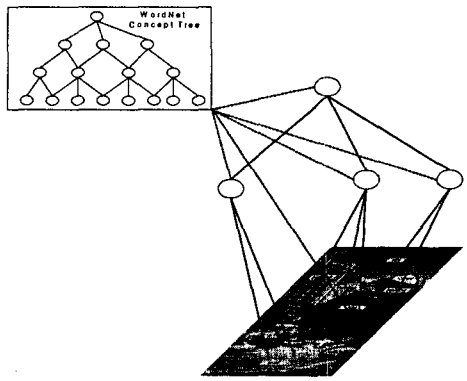

그림 10. 추출된 객체와 WordNet과의 개념 관계

4. 결론 및 향후 연구과제

본 논문에서는 영상 데이터 내에 존재하는 객체들과 그 객체 사이의 개념관계를 Ontology의 한 형태인 WordNet을 이용하여 의미 정보를 정의하였다. 기존의 이미지 검색 시스템은 영상의 저차원 정보만을 이용하거나 임의의 주석을 이용하여 검색하였기에 해당 영상이 가지고 있는 개념적인 해석과 이를 기반으로하는 검색이 이루어지지 못했다. 이에 본 논문에서 제시하는 영상 데이터의 의미 정보를 활용한다면 영상 데이터의 자동 주석 시스템이나 검색 시스템에서 인간이 인식하는 개념적인 사고방식에 더욱 접근할 수 있는 결과물을 얻을 수 있을 것으로 판단된다.

또한, 향후 연구에는 개념적인 인식률을 높이기 위해 WordNet에 존재하는 여러 계층별 단어들의 개념적 유사도 측정에서 보다 효율적이고 정확한 계산 방법을 위한 연구가 필요할 것이다.

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.