

비주얼 검색을 위한 위키피디아 기반의 질의어 추출

김종우[†], 조수선^{**}

Keyword Selection for Visual Search based on Wikipedia

Jongwoo Kim[†], Soosun Cho^{**}

ABSTRACT

The mobile visual search service uses a query image to acquire linkage information through pre-constructed DB search. From the standpoint of this purpose, it would be more useful if you could perform a search on a web-based keyword search system instead of a pre-built DB search. In this paper, we propose a representative query extraction algorithm to be used as a keyword on a web-based search system. To do this, we use image classification labels generated by the CNN (Convolutional Neural Network) algorithm based on Deep Learning, which has a remarkable performance in image recognition. In the query extraction algorithm, dictionary meaningful words are extracted using Wikipedia, and hierarchical categories are constructed using WordNet. The performance of the proposed algorithm is evaluated by measuring the system response time.

Key words: Mobile Visual Search, Keyword Generation, Wikipedia, WordNet, Convolutional Neural Network, Tag Classification,

1. 서 론

스마트폰 보급 확대와 네트워크 발달에 힘입어 영상 데이터의 폭발적인 증가와 함께 모바일기기를 통한 영상정보의 검색 또한 급속히 증가하고 있다. 따라서 더욱 편리하고 정확한 검색 방법에 대한 수요 또한 늘어나고 있다[1]. 특히 영상 기반의 정보검색은 사용자가 명칭이나 키워드를 모를 경우에도 검색의 편의성을 만족시킬 수 있어 그 효과가 크다. 모바일 비주얼 검색 기술은 검색하고자 하는 대상을 촬영하거나 촬영된 이미지를 활용하여 미리 구축된 영상 DB 검색을 통해 촬영 대상과 연계된 정보를 얻게 해준다[2]. (예를 들어 상품의 경우, 검색하려는 상품을 촬영하고 이를 DB에서 검색하여 상품명, 가격,

기능 및 판매처 등 부가정보를 얻을 수 있다. 혹은 거리의 건물을 촬영하고, 건물을 인식하여 건물 내의 정보나 주변정보를 제공받을 수 있다.)

이러한 MVS 기술은 이미지를 활용하여 미리 구축된 DB 검색을 통한 연계 정보 획득이라는 목적에 비추어 볼 때, 크게 영상 인식과 검색의 단계로 구성된다고 할 수 있다. 이미지를 이용하여 기 구축된 DB를 검색하는 것보다 동일한 입력 이미지를 이용하여 웹기반의 텍스트 검색을 할 수 있다면 좀 더 활용도가 높을 것이다. 예를 들어 산행 길에 발견한 들꽃을 촬영한 이미지를 이용하여 그 식물의 이름을 얻을 수 있다면 이를 검색창에 바로 입력되게 하여 웹 검색 결과를 확인할 수 있을 것이다. 이와 같은 경우에는 영상 인식과 영상 인식 후 제공되는 라벨들을 활

* Corresponding Author : Soosun Cho, Address: (16106) Uiwang, Gyeonggi, Korea, TEL : +82-31-460-0584, FAX : +82-31-462-8208, E-mail : sscho@ut.ac.kr

Receipt date : Jul. 10, 2018, Approval date : Jul. 25, 2018
[†] Dept. of Computer Science and Engineering., Inha University (E-mail : bestkjw@hotmail.com)

^{**} Dept. of Computer Science & Information Eng., Korea National University of Transportation

* This work was supported by the Technology development Program(201601180001) funded by the Ministry of SMEs and Startups(MSS, Korea)

용하여 검색 키워드를 추출하는 기술이 필수적으로 요구된다. 기존의 검색의 키워드 입력 방식과는 달리, 검색하고자 하는 대상을 스마트폰과 같은 모바일 기기의 카메라로 촬영하면, 해당 촬영 이미지가 사용자 입장에서 검색 키워드와 동일한 역할을 할 수 있도록 이미지에서 적절한 검색 키워드를 추출하여 사용할 수 있도록 해야 하기 때문이다.

본 논문에서는 이미지 인식분야에서 뛰어난 성능을 보여 각광을 받고 있는 딥러닝(Deep learning) 기반의 CNN(Convolutional Neural Network)을 이용하여 생성된 분류 라벨들로부터 검색어로 사용될 대표 질의어를 추출하는 방법을 소개한다. 본 논문의 구성은 다음과 같다. 2절에서는 관련 연구로 CNN 기법을 사용하여 이미지에 의미 정보를 자동으로 태깅하기 위한 다양한 연구들과 함께 위키피디아 온톨로지 구조를 이용한 키워드 매칭을 소개한다. 3절에서는 제안하는 검색 질의어 추출 알고리즘을 설명한다. 여기에서는 제안된 알고리즘의 전체적인 개요와 시스템 구현에 사용된 여러 가지 휴리스틱들을 소개한다. 그리고 4절에서는 제안된 알고리즘으로 실험한 결과를 소개하고 5절에서 결론을 맺는다.

2. 관련연구

본 연구 전체 시스템의 전형적인 실행 시나리오는 1) 모바일 기기로 촬영된 질의 이미지는 회선신경망 기술인 CNN의 이미지 인식 과정을 통해 자동 분류되며, 2) 생성된 다수의 분류 라벨들을 기반으로 대표 키워드를 선별한 후, 3) 기존 검색엔진의 검색 창에 대표 키워드를 질의어로 입력해주고 검색 결과를 출력해주는 방식이다. 본 논문에서는 시나리오의 두 번째 과정에 해당하는 대표 키워드 선별 방법에서 사용된 알고리즘을 소개하고자 한다. 이를 위해서 관련 연구로서 먼저 시나리오의 첫 번째 과정에 해당하는 CNN 기법을 사용한 이미지 의미 정보 자동 태깅에 관한 연구를 소개하고, 본격적인 주제와 관련된 연구로서 위키피디아를 이용한 응용 기술을 소개한다.

2.1 CNN 기법을 사용한 이미지 의미 정보 자동 태깅

2012년 ILSVRC(Imagenet Large Scale Visual Recognition Challenge)라는 세계적인 이미지 인식 경연 대회에서 토론토 대학의 Hinton 교수가 이끄는 슈퍼비전 팀이 회선처리(Convolution)를 이용하는

신경망인 CNN을 이용하여 탁월한 성과를 보인 이후로 영상 인식 분야에서 CNN에 관련한 연구가 활발하게 진행되어 왔다. 관련 연구[3-5]에서는 기본적인 CNN구조를 사용하여 다양한 방법으로 이미지 분석을 시도하고 있는데, 이를 위해 미리 수집된 정제된 이미지 데이터 셋으로 CNN을 훈련시킨다. 일반적으로 PASCAL Visual Object Classes, MIT Indoor Scene, ImageNet, Caltech 101 등 컴퓨터 비전 분야의 실험을 위한 정제된 이미지 데이터 셋이 사용되고 있다. 이들 연구에서는 이미지에 라벨이 붙여진 데이터 셋으로 CNN을 훈련시킨 후 동일 데이터 셋에서 분류 실험을 한다. 연구의 실험에서는 훈련과 분류에 사용된 데이터 셋에 따라 다른 결과를 보이고 있다. PASCAL의 이미지 셋으로 실험한 결과는 평균 72.6%의 이미지 자동 태깅 정확도를 보였고, MIT Indoor 이미지 셋에서는 평균 59.6%의 정확도를 보이고 있다. 이것은 기존의 태깅 방법에 비해 확실히 더 나은 결과임이 분명하다.

연구[6]에서는 ImageNet 데이터셋을 기반으로 학습된 CNN 모델들을 이용하여 실제 소셜 네트워크상의 이미지, 인스타그램의 이미지들을 분류해보고 모델별 성능을 비교하였다. ImageNet 2012 데이터셋을 트레이닝셋으로 하여 Pretrain된 모델을 이용하였다. ILSVRC 2012에서 우승한 AlexNet과 2015년에 우승한 ResNet 모델을 이용한 결과 12개의 분류 카테고리에 대하여 AlexNet 보다 ResNet에서 더 나은 결과를 얻었다. 본 논문에서는 이와 같은 ImageNet 분류 라벨의 결과를 이용하여 검색 키워드로 사용할 수 있을 만큼 의미적 대표성을 지닌 텍스트를 찾아내는 것을 목표로 한다.

2.2 위키피디아 온톨로지 구조를 활용한 키워드 매칭

위키피디아는 전 세계의 다양한 사용자가 자유롭게 등록 및 수정이 가능한 백과사전이다. 하지만, 데이터의 품질을 위해 등록된 정보는 엄격하게 관리하여 내용을 검사하고 상호비판을 통해 신뢰할 수 있는 품질의 서비스를 제공하고 있다. 이러한 강점 때문에 2001년 서비스를 시작한 이후 2016년 5월 기준, 영어판으로만 532만여 개에 이르는 방대한 정보를 제공하고 있다. 품질이 관리된 대량의 정보라는 특징 때문에 시맨틱 웹, 지식베이스, 자연어 처리 등 다양한 분야에서 활발하게 위키피디아를 이용하고 있다.

연구[7]에서는 위키피디아의 링크기반 의미 연관성을 활용하여 검색어 또는 다른 태그들과의 의미 연관성이 높은 태그를 해당 이미지의 대표 태그로 판단하는 방법을 제안하였다. 연구[8]에서는 개체명 사진 구축을 위하여 위키피디아를 이용하였다. 위키피디아 분류정보로 구성된 가상 문서를 개체명 범주당 하나씩 생성하고 잘 알려진 정보검색 모델인 BM25를 이용하여 위키피디아 엔트리와 가상문서 사이의 유사도를 계산한 후 마지막으로 유사도를 바탕으로 각 위키피디아 엔트리를 개체명 범주로 분류하였다. 연구[9]에서는 대부분의 인터넷 사용자들의 질의어가 매우 빈약하고, 짧아서 사용자의 정보 의도를 명확하게 표현할 수 없다는 문제를 해결하기 위해 위키피디아를 외부 자원으로 이용하여 초기 질의어에 대한 중의성을 해소하고, 질의어 개선 및 확장을 통하여 사용자의 의도를 충족시켜줄 수 있는 방법을 제안하고 있다.

이와 같은 관련 연구의 기술들을 본 연구 전체 시스템의 실행 시나리오에 적용하여 살펴보면, 시나리오의 첫 번째 단계인 ‘모바일 기기로 촬영된 질의 이미지의 CNN 이미지 인식 및 분류’에서는 입력된 질의 이미지를 설명하는 CNN의 분류 라벨, 즉 태그들이 산출된다. 두 번째 단계인 ‘생성된 다수의 분류 라벨을 이용한 대표 키워드 선별’에서는 첫 번째 단계에서 생성된 질의 이미지의 각 라벨 목록을 하나의 그룹으로 정의하고, 대표 단어 추출을 위해 이 중 Top 3에 해당하는 세 개의 태그 그룹을 선정한다. 선정된 그룹의 각 태그들을 이용하여 위키피디아(Wikipedia)를 이용한 사전적 의미의 주제어 추출, 워드넷을 이용한 태그 간 상하연결 관계를 나타내는 카테고리 추출하고, 추출된 카테고리의 단어 간 연관 관계 평가를 거쳐, 대표 질의어를 포함하는 연관어를 추천하는 과정을 본 논문에서 소개한다.

3. 위키피디아 기반의 대표 키워드 추출 방법

3.1 태그 그룹의 선정

대표 키워드 추출을 위해 사용되는 태그 그룹으로 연구[7]에서 위키피디아의 링크기반 의미 연관성을 활용하여 검색어 또는 다른 태그들과의 의미 연관성이 높은 태그를 해당 이미지의 대표 태그로 판단하는 방법을 이용하여 Top 3의 결과를 얻는다. 이때 하나

의 이미지의 태그 목록을 하나의 태그 그룹으로 할 경우, 우리는 Top 3 이미지에 대한 3개의 태그 그룹을 이용하여 대표 키워드 추출을 수행한다. 첫 번째 단계는 태그 그룹의 선정으로 Top 3의 정확도를 바탕으로 기준값에 따라 각 그룹의 태그 목록을 하나의 태그 목록으로 통합한다. Fig. 1은 키워드 추출을 위한 태그 그룹을 선정하는 과정을 나타낸다.

각 그룹의 정확도를 g_1, g_2, g_3 로 하고 기준값을 $\alpha, \dot{\alpha}, \ddot{\alpha}, \tilde{\alpha}$ 로 한다. $\alpha < \tilde{\alpha}$ 이고, $\alpha > \dot{\alpha} > \ddot{\alpha}$ 로 한다. Fig. 1에서 각 조건비교를 만족할 경우 그룹1, 또는 그룹1과2, 또는 그룹 1,2,3을 사용하여 하나의 태그 그룹으로 선정한다. 본 논문의 실험에서는 $\alpha=0.6, \dot{\alpha}=0.4, \ddot{\alpha}=0.2, \tilde{\alpha}=0.7$ 을 사용하였다. 이 기준은 Top 3를 선정하는 방법에서 Top 4에서 검색 결과 정확도를 나타내는 비교표에서 68.8%~91.7%를 보여주고 있다. 이 값을 참고하여 α 의 각 값을 임의 선정하였다. Top 3의 그룹의 정확도 합이 0.7보다 작을 때 이미지를 다시 선정하도록 하며, Top 1이 0.6보다 클 때 그룹 1을 선정하고, 이후 0.2의 차이를 두어 그룹 2, 3을 포함하도록 하였다. 즉 선정된 이미지의 그룹을 하나의 태그 목록으로 통합할 경우 검색하는 이미지를 설명하기 위한 태그로서 최소 60%대의 정확도를 보여 줄 수 있을 때 대표 키워드 추출을 수행하게 되는 것이다. 반복되는 이미지 선정에서 정확도 합이 70%를 못 넘을 때는 새로운 이미지가 요구되어야 한다.

3.2 위키피디아를 이용한 사전적 의미의 주제어 추출

어떤 낱말이 지닌 가장 기본적이고 객관적인 의미를 사전적 의미라고 한다. 3.1에서 선정된 태그 목록

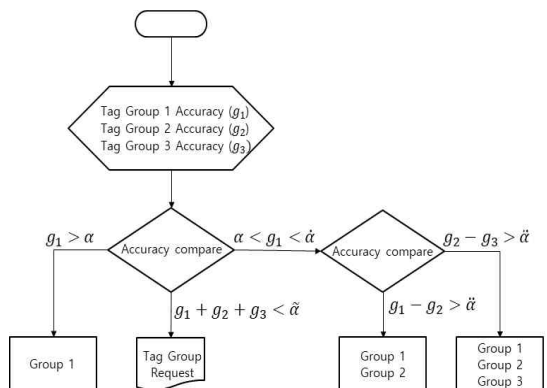


Fig. 1. Tag group selection process.

은 일반 사용자가 붙이는 의미로서 사전적 의미를 지니지 않는 단어를 사용하는 경우가 많다. 따라서 객관적인 단어의 평가를 위해서 각 태그가 지니는 주제어를 추출하는 것이 다음 단계이다. 위키피디아는 누구나 자유롭게 글을 쓸 수 있는 사용자 참여의 온라인 백과사전이다. 누구나 자유롭게 글을 쓸 수 있는 점에서 부정확한 내용이 포함될 수 있으나, 다양한 방면의 지식이 참여하여 끊임없이 갱신되므로 대표적인 참고자료가 되고 있다. 다양한 사용자가 많이 사용하는 단어가 포함된다는 특징에서 태그 목록을 대표하는 사전적 의미의 주제어를 추출하는 데 있어 적합한 참조 문서가 된다.

태그 목록의 단어를 위키피디아에서 검색할 경우 연관된 문서를 검색해 준다. 이때 하나 이상의 문서가 검색되지만 검색되지 않는 경우도 발생한다. 이는 태그 자체가 사전적 의미의 단어가 아니기 때문이라고 판단하여 이 단어를 목록에서 제외한다. 검색된 문서의 제목은 태그를 대표하는 주제로 사용한다. Fig. 2는 이와 같은 방법을 통해 태그의 주제어를 추출한 결과이다. Fig. 2의 (a)의 이미지를 통해 얻어지는 태그는 (b)와 같다. 각각의 태그로 위키피디아에서 문서를 검색하면 (c)와 같은 사전적 의미가 있는 주제어를 얻게 된다. 이 과정에서 얻는 (c)의 결과는 태그 목록을 대표하는 사전적 의미의 단어 그룹으로 표준화된 태그로 정의된다.

3.3 태그 간 상하연결 관계 카테고리 추출

위키피디아 문서를 통해 얻어낸 표준화된 태그 목록의 단어들의 상위어, 하위어의 관계 네트워크를 추출하여 단어 간에 연결 관계를 평가할 수 있는 카테고리 추출 방법으로 워드넷을 이용한다. 워드넷은 서로 다른 문법적 범칙에 따라 명사, 동사, 형용사, 부사들을 구분한다. 워드넷은 모든 단어의 동의어 집합

또는 연어 관계를 포함하여 ‘신셋(synset: 동의어 집합)’을 구성한다. 대부분의 신셋은 다른 동의어의 집합과 몇 개의 의미적 관계로 연결되어 있다. 이러한 관계들은 단어의 유형에 따라 다르며, 명사의 경우 상위어(hypernym), 하위어(hyponym), 등위어(coordinate term), 전체어(holonym), 부분어(meronym)에 대한 의미적 관계가 있다. 본 연구에서는 명사를 기반으로 하는 사전적 의미의 단어를 사용하므로 워드넷의 명사적 의미적 관계를 통해 단어의 상하연결 관계 카테고리를 하나의 네트워크로 표현한다. Fig. 3은 워드넷의 상위어, 하위어 연결 관계를 하나의 카테고리 네트워크로 표현한 것이다. Fig. 3의 (a)에서 태그 1~4까지의 상하연결 관계 카테고리는 모두 최상위는 엔트리이다. 따라서 각 태그의 카테고리는 Fig. 3의 (b)와 같이 엔트리를 중심으로 하나의 카테고리 네트워크로 통합 연결할 수 있다.

3.4 단어 간 연관 관계 평가 및 질의어 추천

위키피디아 문서를 통해 얻어낸 표준화된 태그 목록의 단어들을 Fig. 3과 같은 카테고리 네트워크로 표현한 후 각 태그에 대한 연관 관계를 평가한다. 본 연구에서 연관 관계의 평가는 태그와 연관된 상위어의 연결 빈도를 가중치로 부여하는 일반적 빈도분석 방법을 사용한다. Fig. 2의 (b)에서 카테고리 네트워크에서 상위어 c2는 태그 2와 3에서 중복되며, 상위어 d2는 태그 1과 2, 3에서 중복된다. 상위어의 발생 빈도의 가중치를 부여하면 Tag 2, Tag 3이 추천 질의어 대상이 된다. Tag 2, Tag 3는 상위어 c2에서 통합된다. Tag 2와 Tag 3의 최종 R(Recommendation)은 수식(1)과 같다. 식에서 α, β 는 본 연구에서는 1로 정하였다.

$$R_{tag} = \text{Appearance Frequency of Tag} * \alpha + \text{Distance with Hypernym} * \beta \quad (1)$$



Fig. 2. Tag extraction of dictionary meaning of tag.

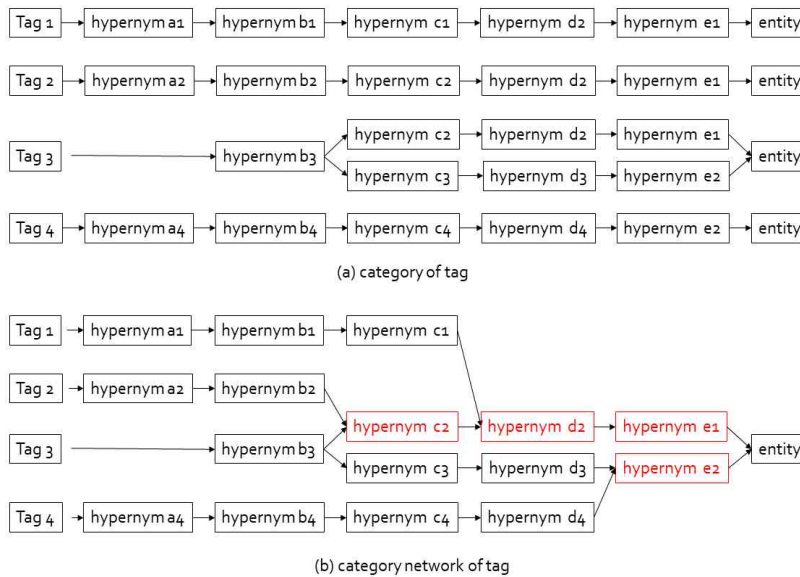


Fig. 3. category network of tag.

4. 구현 및 평가

4.1 시스템의 구현

본 연구의 평가를 위한 Keyword Finder를 구현하였다. Keyword Finder의 프로세스 아키텍처는 Fig. 4와 같으며, 앞에서 설명한 연구 내용의 순차적 처리 과정을 보여준다. 실시간 태그 분석은 입력된 태그 그룹에서 필요한 태그를 추출 및 위키피디아를 이용한 사전적 의미의 주제어 추출 과정이 된다. 카테고리

리 생성은 사전적 의미의 주제어를 지닌 태그 간 상하연결 관계 카테고리 추출이다. 이렇게 추출된 카테고리는 하나의 통합된 카테고리 네트워크로 형성되고 카테고리 관계 정의 프로세스에서 태그의 추천율을 평가하여 사용자에게 대표 질의어로 제시한다. 사용자에게 의해 질의어로 사용된 태그는 이후 추천율을 평가하는데 반영된다. Fig. 5는 Keyword Finder를 이용한 모바일 비주얼 검색 서비스의 구현 화면이다. Fig. 6는 Keyword Finder의 결과를 보여 주고 있

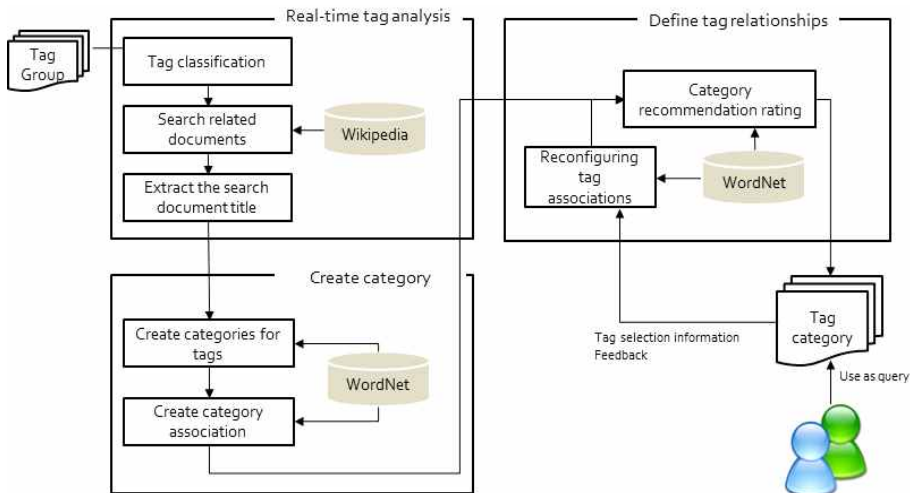


Fig. 4. Process architecture of Keyword Finder.

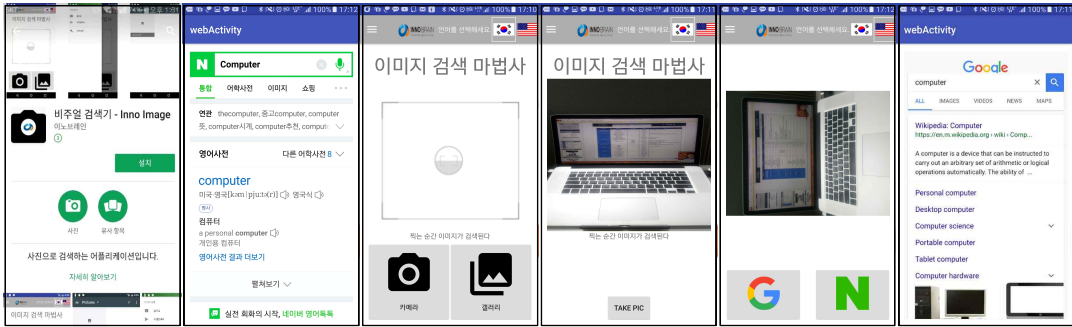


Fig. 5. Mobile Visual Search Service.

{“tags”:[“Cougar”, “Puma”, “Catamount”, “mountain lion”, “Painter”, “Panther”, “Felis concolor”], “accuracy”: 92.3}

(a)

```

Cougar[Cougar], Puma[Puma], Catamount[Cougar], mountain lion[Cougar], Felis concolor[cougar]
SID-02127963-N {cougar, puma, catamount, mountain_lion, painter, panther, Felis_concolor}
->SID-02127275-N {wildcat}
->SID-02124272-N {cat, true_cat}
->SID-02123649-N {feline, felid}
->SID-02077948-N {carnivore}
->SID-01889397-N {placental, placental_mammal, eutherian, eutherian_mammal}
->SID-01864419-N {mammal, mammalian}
->SID-01474323-N {vertebrate, craniate}
->SID-01468898-N {chordate}
->SID-00015568-N {animal, animate_being, beast, brute, creature, fauna}
->SID-00004475-N {organism, being}
->SID-00004258-N {living_thing, animate_thing}
->SID-00003553-N {whole, unit}
->SID-00002684-N {object, physical_object}
->SID-00001930-N {physical_entity}
->SID-00001740-N {entity}
    
```

(b)

```

Panther[Panther]
SID-02131577-N {jaguar, panther, Panthera_onca, Felis_onca}
->SID-02130460-N {big_cat, cat}
->SID-02123649-N {feline, felid}
->SID-02077948-N {carnivore}
->SID-01889397-N {placental, placental_mammal, eutherian, eutherian_mammal}
->SID-01864419-N {mammal, mammalian}
->SID-01474323-N {vertebrate, craniate}
->SID-01468898-N {chordate}
->SID-00015568-N {animal, animate_being, beast, brute, creature, fauna}
->SID-00004475-N {organism, being}
->SID-00004258-N {living_thing, animate_thing}
->SID-00003553-N {whole, unit}
->SID-00002684-N {object, physical_object}
->SID-00001930-N {physical_entity}
->SID-00001740-N {entity}
    
```

(c)

```

Painter[Painting]
SID-03882197-N {painting, picture}
->SID-03458753-N {graphic_art}
->SID-02746552-N {art, fine_art}
->SID-03133774-N {creation}
->SID-00022119-N {artifact, artefact}
->SID-00003553-N {whole, unit}
->SID-00002684-N {object, physical_object}
->SID-00001930-N {physical_entity}
->SID-00001740-N {entity}
    
```

(d)

Fig. 6. The process sample of Keyword Finder. (a)Input Data, (b) “Cougar” Result, (c) “Panther” Result, (d) “Painter” Result,



Fig. 7. Experimental image set, scissors, cup, clock, mouse, keyboard from left.

Table 1. Experimental systems

Specifications	
CPU	Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz
RAM	64GB
SSD	250GB
HDD	4TB
Video	NVIDIA GeForce GTX 1080
OS	Ubuntu 16.04.1 LTS(4.4.0-31-generic)
Etc.	CUDA 8.0.61 cuDNN 5.1 JDK 1.8.0_131 Python 2.7.12

다. 태그 그룹으로 Fig. 6의 (a)와 같은 “Cougar”, “Puma”, “Catamount”, “mountain lion”, “Painter”, “Panther”, “Felis concolor”가 입력 데이터이고, 각 태그의 처리 결과 사전적 의미의 태그로 “Cougar”, “Panther”, “Painter” 제시 되었고, (b), (c), (d)의 카테고리리를 추출하였다. 최종 추천은 Fig. 6의 (b) “Cougar”의 결과가 선정되었으며, 대표 질의어는 Cougar 가 된다.

4.2 성능 평가

본 연구에서 성능 평가로 출력 결과의 정확성은

Table 2. The performance for of Keyword Finder

	File Name	Category	Keyword	Image Size(MB)	Process Time	Start Time	End Time	category network
1	/VSKMP/AITF/image/scissors_pair_of_scissors.jpg	[scissors, pair_of_scissors]	scissors	1.9	2.7171sec [1.4301sec (s/mb)]	2017-05-14 23:40:08.50 918498	2017-05-14 23:40:11.22 6328427	{{["scissors", "pair_of_scissors"], ["edge_tool"], ["cutter", "cutlery", "cutting_tool"], ["cutting_implementation"], ["tool"], ["implementation"], ["instrumentality", "instrumentation"], ["artifact", "artefact"], ["whole", "unit"], ["object", "physical_object"], ["physical_entity"], ["entity"]]}, [{"scissors", "pair_of_scissors"}, {"compound_lever"}, {"lever"}, {"bar"}, {"implementation"}, {"instrumentality", "instrumentation"}, {"artifact", "artefact"}, {"whole", "unit"}, {"object", "physical_object"}, {"physical_entity"}, {"entity"}]}
2	/VSKMP/AITF/image/cup.jpg	[cup]	Cup	1.69	2.8173sec [1.6670sec (s/mb)]	2017-05-14 23:40:08.50 9031362	2017-05-14 23:40:11.32 6379615	{{["cup"], ["crocery", "dishware"], ["tableware"], ["ware"], ["article"], ["artifact", "artefact"], ["whole", "unit"], ["object", "physical_object"], ["physical_entity"], ["entity"]]}, [{"cup"}, {"container"}, {"instrumentality", "instrumentation"}, {"artifact", "artefact"}, {"whole", "unit"}, {"object", "physical_object"}, {"physical_entity"}, {"entity"}]}
3	/VSKMP/AITF/image/clock.jpg	[clock]	Clock	2.25	2.9177sec [1.2968sec (s/mb)]	2017-05-14 23:40:08.50 8949862	2017-05-14 23:40:11.42 6641424	{{["clock"], ["timepiece", "timekeeper", "horologe"], ["measuring_instrument", "measuring_system", "measuring_device"], ["instrument"], ["device"], ["instrumentality", "instrumentation"], ["artifact", "artefact"], ["whole", "unit"], ["object", "physical_object"], ["physical_entity"], ["entity"]]}
4	/VSKMP/AITF/image/mouse_computer_mouse.jpg	[mouse]	Mouse	3.89	3.1163sec [0.8011sec (s/mb)]	2017-05-14 23:40:08.50 9155117	2017-05-14 23:40:11.62 5420362	{{["mouse"], ["rodent", "gnawer"], ["placental_mammal", "mammal", "mammalian"], ["vertebrate", "craniate"], ["chordate"], ["animal", "animate_being", "beast", "brute", "creature", "fauna"], ["organism", "being"], ["living_thing", "animate_thing"], ["whole", "unit"], ["object", "physical_object"], ["physical_entity"], ["entity"]]}
5	/VSKMP/AITF/image/computer_keyboard_keypad.jpg	[computer_keyboard, keypad]	Computer keyboard	4.77	3.3231sec [0.6967sec (s/mb)]	2017-05-14 23:40:08.50 8980097	2017-05-14 23:40:12.83 2053258	{{["computer_keyboard", "keypad"], ["keyboard"], ["device"], ["instrumentality", "instrumentation"], ["artifact", "artefact"], ["whole", "unit"], ["object", "physical_object"], ["physical_entity"], ["entity"]]}, [{"computer_keyboard", "keypad"}, {"data_input_device", "input_device"}, {"peripheral", "computer_peripheral", "peripheral_device"}, {"electronic_equipment"}, {"equipment"}, {"instrumentality", "instrumentation"}, {"artifact", "artefact"}, {"whole", "unit"}, {"object", "physical_object"}, {"physical_entity"}, {"entity"}]}

평가하지 않으며, 시스템의 처리 속도의 성능 평가만 수행한다. 본 연구의 개발 목적은 이미지를 사용한 키워드 검색을 수행하는 Fig. 5와 같은 모바일 비주얼 검색 서비스 구현을 목적으로 처리 속도가 중요한 성능 평가 요소가 된다. 또한 객관적인 태그 정보를 가지고 있는 이미지가 실험 데이터로 사용되었으므로 모든 태그가 대표성을 가지고 있는 태그이다. 따라서 추천되는 키워드 또한 태그의 일부로 항상 대표성을 갖고 있으므로 실험 결과의 정확성을 측정하는 기준이 될 수 없다.

처리 속도 평가를 위한 시스템은 Table 1과 같다. Intel 기반 PC에서 Ubuntu 16.04.1 OS를 이용하였고, 프로그램은 자바 기반으로 구현 되었다. 실험을 위한 이미지는 Fig. 7과 같다. 각 이미지에 대한 크기는 Table 2의 이미지 사이즈(Image Size) 항목에 표기되어 있다. Table 2는 이미지를 분석하여 질의어로 사용될 대표 카테고리 및 키워드를 추출하는데 소요되는 시간을 측정한 결과이다. 측정 결과 이미지의 크기에 따라 소요되는 시간은 증가하지만 이미지의 크기에 비례하지는 않는 것을 확인할 수 있었다.

5. 결 론

본 논문에서는 이미지 인식분야에서 뛰어난 성능을 보여 각광을 받고 있는 딥러닝(Deep learning) 기반의 CNN(Convolutional Neural Network)을 이용하여 생성된 분류 라벨들로부터 검색어로 사용될 대표 질의어를 추출하는 알고리즘을 제안하였다. 선행 조건으로 ‘모바일 기기로 촬영된 질의 이미지의 CNN 이미지 인식 및 분류’에 의해 각 이미지에 부여된 태그들을 하나의 그룹으로 하여, Top 3 이미지에 해당하는 세 개의 태그 그룹을 입력받는다. 이렇게 입력된 태그 그룹에서 평가를 위한 태그 그룹을 선정하고, 위키피디아를 이용한 사전적 의미의 주제어 추출을 수행한다. 태그의 사전적 의미의 주제어는 워드넷을 이용하여 태그 간 상하연결 관계를 설명하는 카테고리로 만들어진다. 이렇게 만들어진 카테고리는 하나의 통합된 카테고리 네트워크로 구축되고 위에서 제안한 연관 관계 평가 및 질의어 추천 알고리즘에 따라 태그의 추천율을 평가하여 사용자에게 대표 질의어로 제시된다. 이와 같은 연구 내용을 바탕으로 Keyword Finder를 구현하여 모바일 비주얼 검색 서비스에 효과적으로 활용되었다. 모바일 비주얼

검색 서비스 구현을 위한 일부분으로서 본 연구에서는 시스템의 응답시간을 줄이는데 노력을 집중하였다. 때문에 시스템 성능 평가에서 사전적 의미의 단어 추출 및 질의어 추천율에 대한 평가는 수행되지 못하였다. 이후 연구에서는 객관적인 평가 기준을 설계하여 다양한 사용자를 대상으로 한 실험을 통한 사전적 의미의 단어 추출 및 질의어 추천에 대한 성능평가 수행이 필요하다고 판단된다. 또한 사전적 의미의 단어 추출 방법에 대한 연구를 확대하고, 질의어 추천을 위한 태그 간 상하 관계뿐만 아니라 의미 관계에 대한 평가를 확대 적용할 계획이다.

REFERENCE

- [1] J. Bughin, L. Corb, O. Nottebohm, M. Chui, B.D.M. Barbatet, R. Said, et al., *The Impact of Internet Technologies: Search*, McKinsey and Company, New York, 2011.
- [2] S.J. Lee, K.D. Lee, S.I. Na, S.K. Je, D.U. Jung, W.G. Oh, et al., “Technology and Standardization Trend of Mobile Visual Search,” *Electronics and Telecommunications Trends*, Vol. 29, No. 1, pp. 61-71, 2014
- [3] P. Fischer, A. Dosovitskiy, and T. Brox, “Descriptor Matching with Convolutional Neural Networks: a Comparison to SIFT,” *eprint arXiv:1405.5769*, 2014.
- [4] J. Wu, Y. Yu, C. Huang, and K. Yu, “Deep Multiple Instance Learning for Image Classification and Auto-Annotation,” *Proceeding of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3460-3469, 2015.
- [5] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, pp. 1097-1105, 2012.
- [6] N.W. Jeong and S.S. Cho, “Instagram Image Classification with Convolution Neural Networks,” *Proceeding of Korean Society for Internet Information The 8th International Conference on Internet*, pp. 2249-2252, 2016.

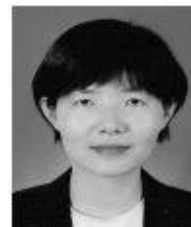
- [7] S.J. Lee and S.S. Cho, "Tagged Web Image Retrieval Re-ranking with Wikipedia-based Semantic Relatedness," *Journal of Korea Multimedia Society*, Vol. 14, No. 11, pp. 1491-1499, 2011.
- [8] Y.K. Song, S.W. Jeong, and H.S. Kim, "A Semi-automatic Construction Method of a Named Entity Dictionary Based on Wikipedia," *Journal of Korean Institute of Scientists and Engineers*, Vol. 42, No. 11, pp. 1397-1403, 2015.
- [9] S.H. Kim, S.J. Bae, and Y.J. Ko, "Word Sense Disambiguation and Feature Extraction for Query Expansion Method Based on Korean Wikipedia," *Journal of Korean Institute of Scientists and Engineers : Computing Practices and Letters*, Vol. 17, No. 3, pp. 185-189, 2011.



김 종 우

2005년 3월~2007년 2월 충주대학교 컴퓨터과학 공학사
2009년 3월~2013년 2월 인하대학교 대학원 컴퓨터공학 공학박사수료
2014년 7월~2017년 6월 전자부품연구원 네트워크융합센터 연구원

2018년 3월~현재 ㈜이노비즈 제어팀
관심분야: 시멘틱웹, 빅데이터, 인공지능



조 수 선

1983년 3월~1987년 2월 서울대학교 계산통계학과 이학사
1987년 3월~1989년 2월 서울대학교 대학원 계산통계학과 이학석사
2000년 3월~2004년 2월 충남대학교 대학원 컴퓨터과학과 이학박사

1994년 6월~2004년 3월 한국전자통신연구원 선임연구원
2006년 7월~2007년 7월 미시간대학교(앤아버) 통계학과 방문연구원
2004년 4월~현재 한국교통대학교 컴퓨터정보공학과 교수
관심분야: 웹 콘텐츠 마이닝, 모바일 웹, 기계학습