

# UCC 검색 기술 동향

□ 노용만, 민현석, 이시형, Wesley De Neve / 한국정보통신대학교

## 1. 서론

### 1. UCC 의 출현 및 확산

최근 여러 분야에서 논의되고 있는 사용자 참여, 공유, 개방의 개념은 멀티미디어 정보 서비스 분야에도 예외는 아니다. 이러한 개념은 미래 학자 엘빈 토플러가 '제 3의 물결'에서 언급한 프로슈머(producer + consumer) 개념과 맥을 같이하는 것으로, 생산자와 소비자의 경계가 허물어지면서 소비자가 제품 개발/유통에 참여하는 '생산적 소비자'가 등장하고 있다. 이런 양상이 멀티미디어 콘텐츠 분야에서는 최근 급속히 증가하고 있는 사용자 생성 콘텐츠(UCC)로써 나타나고 있다.

UCC는 과거에 비해 상대적으로 낮은 가격으로 제공되는 네트워크 대역폭, 풍부한 저장 공간 그리고 빠른 데이터 처리 속도 등의 기술 발전과 더불어

확산되고 있다. 또한 전문가가 아닌 일반 사용자가 보다 쉽게 콘텐츠를 생성, 가공, 공유하고 출판하게 하는 콘텐츠 저작 기술의 보급과 사용자 스스로가 자신의 창작을 외부로 드러내려고 하는 욕구가 이러한 UCC의 확산을 촉진시키고 있다. 이러한 UCC는 기존의 '패키징된(packaged) 콘텐츠'와는 달리 '대화형 콘텐츠'라는 특징을 가지고 있다. UCC의 확산은 콘텐츠의 생성과 소비 형태에 큰 변화를 주고 있으며 UCC는 점차 무시할 수 없는 중요한 경제 미디어가 되고 있다. 이는 향후 멀티미디어 관련 산업의 사업 전략에도 많은 변화를 줄 것이다.

UCC는 현재 인터넷 상의 사회적 네트워킹을 통하여 주로 소비되고 있다. 대표적인 UCC 영상 서비스인 'Flickr'에는 현재 약 850만 이상의 사용자가 등록되어 있으며, 1일에 약 200만장 이상의 사진이 등록되고 있고, 1초에 12,000장 이상의 사진들이 사용자에게 의해 검색되고 있다. 또한 대표적인 UCC 동

영상 서비스인 'YouTube'에는 하루 5만개 이상의 동영상이 등록되며, 이는 미국 내의 모든 방송사들이 제작할 수 있는 콘텐츠의 양을 훨씬 앞지르는 수치이다. 또한 900만명 이상의 하루 평균 방문자 수는 TV 방송사의 하루 시청자 수와 대등하다 [1][2]. 이렇게 웹에 배포되는 UCC 동영상은 현재 5십만 테라바이트로 추정되고 있으나, 2011년 이는 4천8백만 테라바이트로 증가할 것이라 예상되고 있다 [3].

이러한 UCC의 급속한 증가는 UCC의 효율적인 소비와 관리를 위한 검색, 브라우징, 인덱싱 그리고 필터링에 대한 요구를 증가시키고 있다. 최근 쥘피터 리서치사의 'iProspect 검색 마케터 사회적 네트워크 서베이'에 의하면 최근의 검색 시장의 거의 절반 이상이 멀티미디어 UCC 시장으로 인하여 성장하고 있으며, 향후 멀티미디어 UCC 검색에 대한 수요가 크게 증가할 것으로 예상하고 있다. UCC의 확산은 과거 콘텐츠의 생성과 소비 형태에 큰 변화를 주고 있으며, 이는 향후 멀티미디어 관련 기업이나 해당 기업의 사업 전략에도 많은 변화를 줄 것으로 예상된다 [4].

## 2. UCC 소비 환경 및 검색

최근 모바일 멀티미디어 단말이 보편화되면서 사용자들은 언제 어디서나 멀티미디어 콘텐츠를 생산하고 소비할 수 있게 되었다. 즉, 휴대전화, PDA, UMPC, 디지털 카메라/캠코더 등을 통하여 장소나 시간에 구애받지 않고 멀티미디어 콘텐츠의 획득 및 소비가 가능하게 되었다. 이를 통하여 'anytime-anywhere accessibility'가 실현되고 있으며 멀티미디어 콘텐츠의 생산/유통/소비 패러다임의 변화를 가속하고 있다. 또한 WiFi, WiMAX, UMB,

WIBRO 등의 광대역 인터넷 및 통신망의 기반 시설 구축 및 뉴미디어의 등장으로, 멀티미디어의 실시간 스트리밍 소비가 가능해졌다 [5]. 이러한 환경을 바탕으로 OSMU (one source-multi use) 방식으로 콘텐츠를 다양한 미디어에서 이용할 수 있는 디지털 컨버전스 환경이 도래하고 있다. 이러한 디지털 컨버전스 환경에서 UCC는 웹 뿐 아니라 IPTV, DMB 등의 방송에서도 서비스될 수 있다. 디지털 컨버전스 환경의 도래로 인하여 UCC 검색은 단지 웹에서 뿐만 아니라 방송 분야에서도 활용이 크게 요구되고 있는 상황이다 [6][7].

다양한 생산자와 소비자 환경 및 디지털 컨버전스 환경에서 멀티미디어를 소비하기 위해서 UCC 검색은 사용 환경에 적응을 해야 한다. 즉, UCC 검색 기술은 방송/통신 융합의 디지털 컨버전스 및 유비쿼터스 환경의 다양한 응용 플랫폼, 모바일 통신 단말에서의 멀티미디어 검색 지원, 시맨틱-웹 (Semantic-Web)/Web 2.0 출현, UCC를 통한 콘텐츠 생산 및 사회적 멀티미디어 공유 확산, IPTV/D-TV/DMB 등의 방송 멀티미디어 콘텐츠의 증가 등의 다양한 요구사항을 반영하여야 한다 [8][9].

현재 UCC 검색 시스템은 UCC 주변의 텍스트나 사용자 태그에만 의존하고 있다. 그러나 텍스트에 비해 사진/동영상 등의 멀티미디어는 데이터 구조가 복잡하고, 다른 데이터들과 결합되어 통합된 형태로 존재하며, 다양한 의미 정보를 포함하고 있기 때문에, 텍스트 기반 검색만으로는 늘어만 가는 멀티미디어를 효과적으로 검색하는데 많은 한계가 있다. 또한 사용자가 정보와 지식을 생산하고 배포하고 접근하고 다시 사용하는 방식으로 변화하면서 UCC의 메타데이터 생성 과정에서의 사용자 참여를 잠재적으로 증가시키고 있는데, 사용자의 일관되지

못한 메타데이터 생성은 텍스트만을 기반한 UCC 검색을 힘들게 한다. 구글(Google)이나 야후(Yahoo) 등에서 사용되고 있는 현재의 검색 엔진 기술로는 태그 등으로 비디오 클립이나 영상을 검색하고 있지만, 내용이 방대한 UCC와 같은 멀티미디어 검색은 불가능하다. 과거에도 내용 기반 비주얼 특징 정보를 이용한 멀티미디어 검색 기술 연구가 지속적으로 이루어져 왔으나 멀티미디어의 내용 기반 특징만으로는 콘텐츠의 무수히 많은 의미 정보를 인식할 수 없다 [10].

이에 본 논문에서는 UCC의 소비에 대한 동향을 알아보고, UCC 소비의 핵심인 검색에 필요한 요소 기술을 살펴 보고자 한다. 2장에서는 디지털 컨버전스 시대의 UCC 서비스 및 폭소노미 태깅을 살펴보고, 3장에서는 UCC 검색의 필요성과 검색을 위한 요소 기술들을 살펴보고, 4장에서는 향후 전망을 언급하고 본 고의 결론을 맺는다.

## II. 디지털 컨버전스 시대의 UCC 서비스

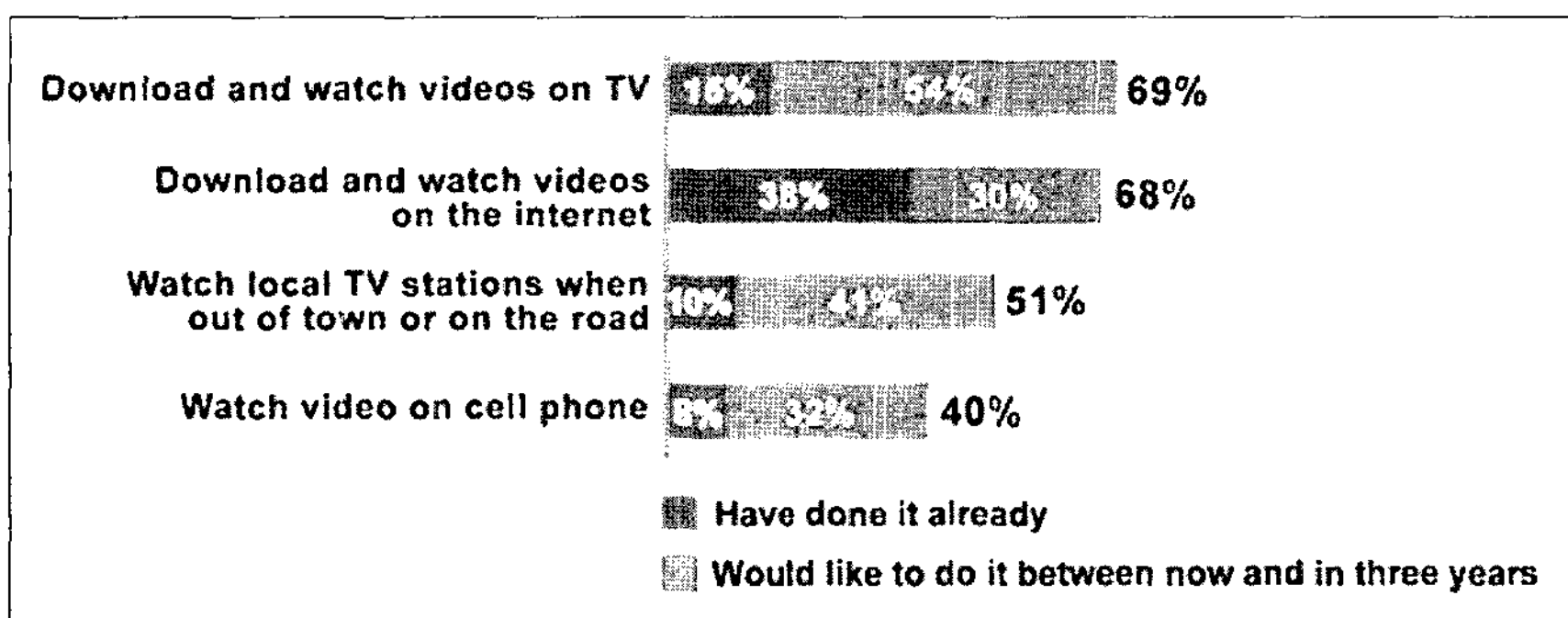
최근 다양한 모바일 멀티미디어 단말의 보급과

멀티미디어 콘텐츠 생산 기술의 확산으로 장소나 시간에 구애받지 않고 멀티미디어 콘텐츠의 획득 및 소비가 가능하게 되었다. 이를 통하여 멀티미디어 콘텐츠의 생산/유통/소비 패러다임의 변화를 가속하고 있다. 이러한 변화는 UCC 서비스에 대해서도 많은 변화를 가져오고 있다.

### 1. 융합환경에서 UCC 서비스

디지털 컨버전스 환경의 도래로 인하여 IPTV, DMB가 출현하였으며, 이는 기존 네트워크와 방송의 서비스의 융합 뿐 아니라 웹 기술과 방송의 융합을 이끌고 있다. 또한 시멘틱 웹, 집단 지성 등으로 대표되는 '웹2.0'에 대한 논의가 활발히 진행되면서 사용자 참여, 공유, 개방의 개념이 도입된 인터넷 기반 TV 서비스에 대한 논의가 활발하게 되고 있다. 이미 'TV 2.0', 'Broadcast 2.0'이란 용어가 등장하고 있고, TV 서비스가 인터넷과 같은 개방된 네트워크로 기반할 때 가질 수 있는 장점들이 새롭게 부각되고 있다. 이러한 융합 환경은 제한적인 환경에서 제공되던 UCC 서비스가 다양한 매체를 통해 더욱 다양한 방법으로 서비스될 수 있게 되었다.

이러한 융합 환경에서 사용자는 동영상 콘텐츠를



〈그림 1〉 향후 다양한 매체를 통한 비디오 콘텐츠 소비에 대한 조사 [11]

기존의 제한된 환경이 아니라 다양한 환경을 통해서 소비하려고 한다. 그림 1은 향후 자신의 비디오 소비 환경에 대한 의견을 조사한 결과이다. 많은 사람들이 보다 다양한 환경에서 비디오 콘텐츠를 소비하려는 경향을 알 수 있다. 이러한 경향을 반영하듯 가장 대표적인 UCC 동영상 서비스인 'YouTube'는 자신들의 서비스를 AppleTV와 모바일 단말을 통해 서비스할 수 있게 하였다. 이러한 디지털 컨버전스 시대에 다양한 매체를 통한 UCC 서비스가 제공됨에 따라 웹 뿐만이 아니라 다양한 매체를 통한 방송 분야에서도 UCC 서비스가 중요해졌다.

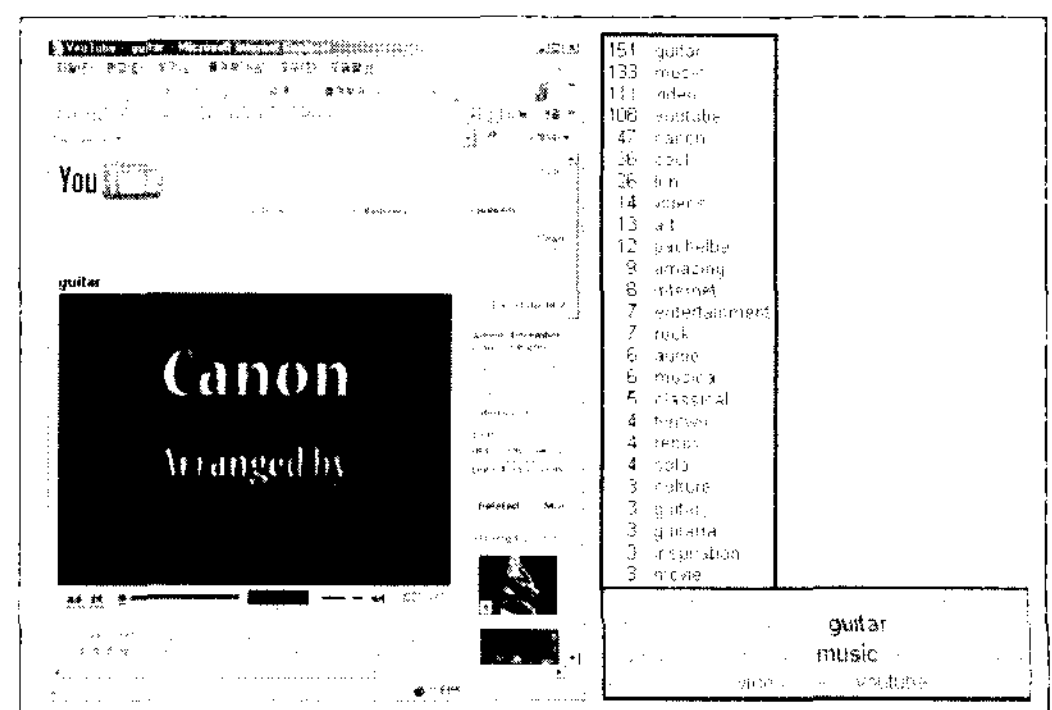
또한 지난 2007년 3월 OECD가 발간한 UCC의 급속한 성장, 개발 잠재적 영향력과 이에 따른 정책 방안을 제시하는 'Participative Web: User Created Content' 라는 제목의 보고서에 따르면, 현재 UCC는 개인이 촬영한 사진/동영상 등 엔터테인먼트 콘텐츠에 한정된 경향이 강하나, 앞으로 교육/출판/쇼핑/방송 등으로 멀티미디어 UCC의 분야가 크게 확대될 것으로 기대됨에 따라 효과적인 UCC 검색 기술에 대한 요구가 더욱 증대될 것이라 예상된다 [11].

## 2. UCC 태깅: 폭소노미

보통 멀티미디어 콘텐츠의 메타데이터는 지정된 전문가나 멀티미디어 시스템에 의해서 제작된다. 그리고 이렇게 생성된 메타데이터만을 기반으로 멀티미디어를 색인하고 검색한다. 그러나 UCC는 대부분 지정된 전문가가 아닌 사용자 스스로가 UCC 관한 메타데이터를 생성한다. 사용자는 사회적 네트워크와 온라인 자원을 통하여 자신의 UCC만이 아니라 다른 사용자에게 의해 제작된 UCC에도 영향

을 끼친다. 사용자는 자신의 UCC를 제작, 배포할 뿐만 아니라, 다른 사용자의 UCC에 대한 자신의 의견을 밝히고 UCC에 관한 비평 혹은 설명을 할 수 있다. 이렇게 여러 사용자들에 의해 협력적으로 메타데이터를 생성하고 관리하는 방법을 폭소노미(Folksonomy)라고 한다. 최근 Flickr나 Rojo 등과 같은 사회적 네트워크를 통하여 널리 알려진 폭소노미는 "Folk(people)+order+nomos(law)"의 합성어로 사용자가 자유롭게 선택한 태그들을 통해 정보를 체계화시키는 "참여에 의한 분류법"으로 협력태깅(collaborative tagging) 기술로도 불린다.

폭소노미는 기존 텍소노미에 비해 체계적이지 않고 무작위적으로 생성되지만, 사용자들에 자발적 참여에 의해 정보가 체계화된다는 특징을 갖는다. 이렇게 전문가들에 의해 작성된 다른 검색 시스템의 정보와는 사용자가 그들의 UCC를 분류하고 부여 설명할 수 있도록 하는 특징은 전문가 뿐 아니라 일반 사용자의 시스템에 즉시 참여할 수 있게 한다. 그림 2는 유튜브에 올린 UCC에 대하여 사용자들이 많이 단 태그를 나타낸 예이다. 그림 2에 왼쪽에 나타난 하나의 UCC를 기술하기 위하여 많은 사람들이 단 다양한 태그의 빈도수를 우측 그림에서 나타내고 있다. 이러한 폭소노미를 통하여 사용자는



(그림 2) UCC 태깅 (YouTube and Del.icio.us)

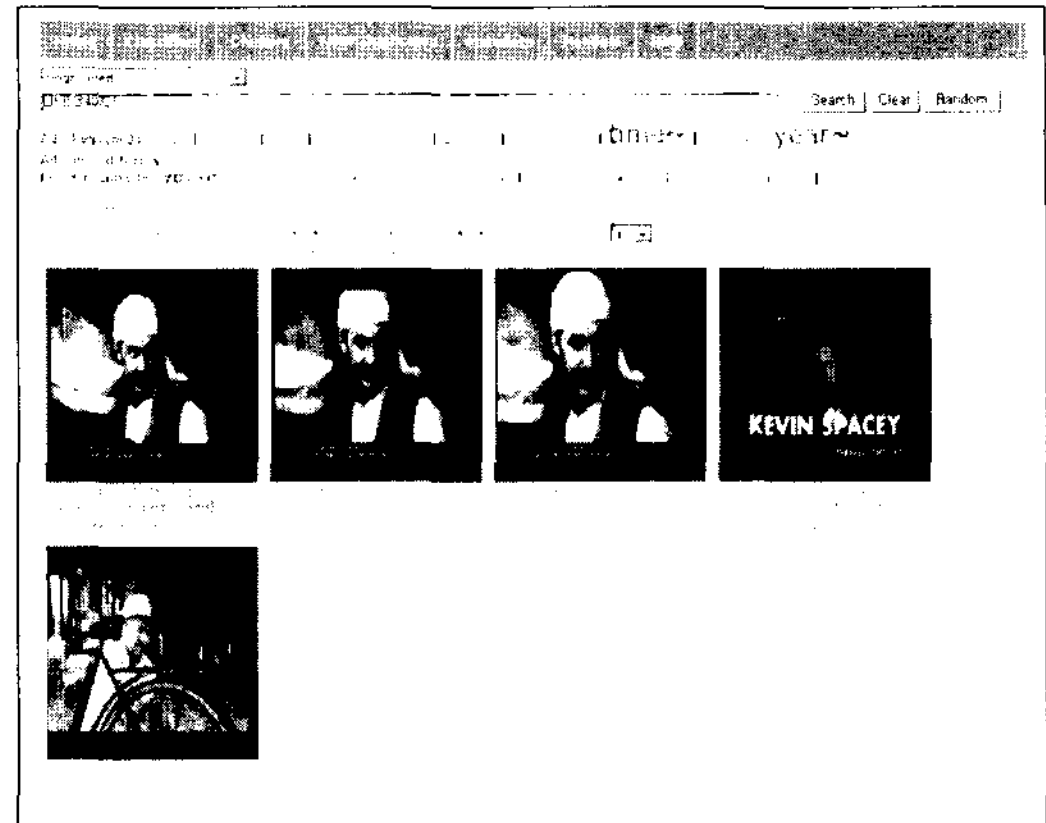
UCC가 사람들에 의하여 주로 어떠한 태그들로 부여되었는지 알 수 있다. 이렇듯 폭소노미는 UCC를 나타내는 하나의 중요한 정보로 부각되고 있다.

그러나 이러한 폭소노미는 사용자들에 의해 자유롭게 정해졌기 때문에 정해진 태그들에는 불명확함이 존재한다. 사용자는 자신의 UCC를 자신이 선호하는 태그들로 표현하기 때문에 많은 다른 사용자들에 의해 같은 태그라도 다른 방식으로 사용될 수 있다. 현재 UCC 검색 서비스는 사용자에 의해 생성된 태그 정보를 기반으로 서비스되고 있는데 UCC 환경에서는 이러한 폭소노미에 의해 정해진 태그의 모호성은 UCC 검색을 위해 해결해야 하는 문제로 인식되고 있다.

### III. UCC 검색을 위한 요소 기술

UCC는 현재 Flickr, MySpace, FaceBook, YouTube와 같은 인터넷 상의 사회적 네트워킹을 통하여 주로 소비되고 있다. 그러나 현재 YouTube와 같은 상용 UCC 검색 시스템은 사용자에 의해 입력된 제목과 태그와 같은 메타데이터를 이용하여 검색하고 있다. 그러나 이러한 텍스트 기반 정보들은 멀티미디어 UCC를 표현하는데 한계가 있다. 멀티미디어 콘텐츠는 데이터 구조가 복잡하고, 다른 데이터들과 결합되어 통합된 형태로 존재하므로 텍스트만으로 멀티미디어 콘텐츠를 표현할 수 없다. 이런 한계 때문에 텍스트 기반으로 UCC 검색이 충분한 성능을 발휘하지 못한다.

1990년대부터 이러한 텍스트 기반 멀티미디어 콘텐츠 검색의 단점을 해결하기 위하여 멀티미디어 콘텐츠의 검색과 색인을 위해서 콘텐츠의 내용 특성에 기반을 둔 연구가 활발하게 진행되었다. 현재



〈그림 3〉 IBM Marvel [15]

이러한 연구를 바탕으로 개발된 CueVideo, Marvel, WebSEEk, VideoQ, MediaMill 등의 많은 동영상 멀티미디어 검색 프로토타입 시스템이 개발되었다 [14]. 그림 3은 내용기반 검색 시스템의 예인 IBM의 Marvel을 보여주고 있다. 그림에서 볼 수 있듯이 사용자는 입력된 콘텐츠와 비슷한 시각적 특징을 가진 콘텐츠를 검색할 수 있다.

한편 콘텐츠의 내용 특징은 주로 콘텐츠의 시각적 특징에 기반을 둔다. 따라서 영상의 픽셀정보에 기반을 두어 특징이 추출된다. 이런 픽셀 기반의 특징은 실제 사람이 인지하는 것과 비교하여 다음과 같은 두 가지 차이가 있다. 하나는 감각차이 (sensory gap)이라는 것이고, 그 다음은 의미차이 (semantic gap)이라는 것이다. 감각 차이는 동일한 물체를 찍은 영상이라도 그 물체를 얻는 상태, 예를 들어 똑바로 물체를 찍을 때와 거꾸로 물체를 찍을 때는 픽셀 정보에 의존하면 물체의 특성이 다르게 나타날 수 있다. 이것을 감각차이라고 한다. 또 다른 차이인 의미차이는 사람이 동일한 영상을 보고도 서로 다르게 느끼는 차이를 말한다. 이는 콘텐츠의 시각적 내용에서 추출한 정보와 사용자가 주어진 상황에서 콘텐츠를 인식하는 정보의 차이에 의해



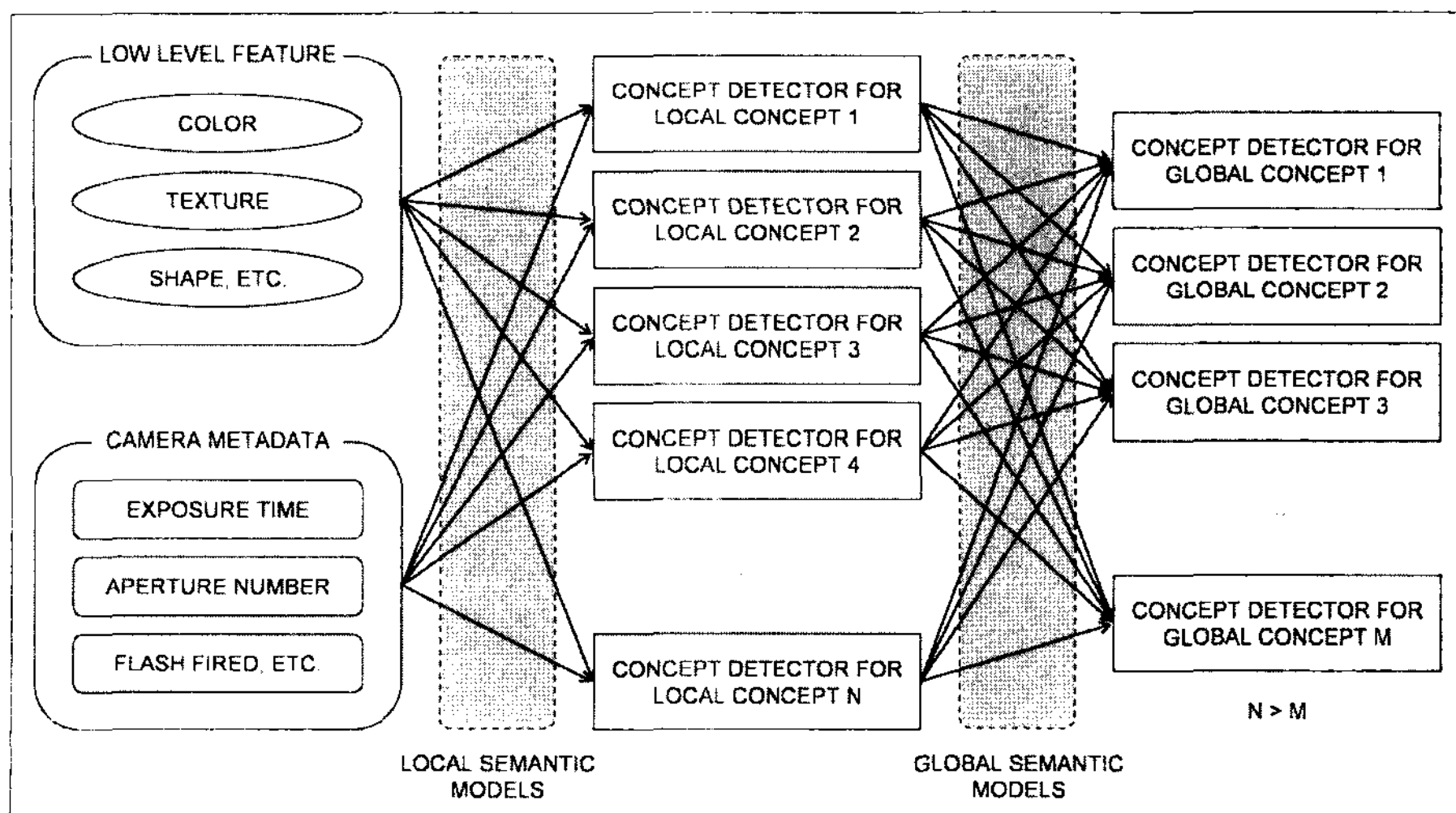
발생하는 의미 차이가 존재하기 때문이다 [16]. 멀티미디어 특징을 추출하는데, 이러한 감각차이와 의미 차이를 줄이기 위하여 많은 요소 기술들이 연구되고 있다. 이에 본고는 최근 많은 주목을 받고 있는 진보된 검색의 핵심 기술인, 의미추출 기술, 주석추천 기술, 정황 정보 기술에 대해 살펴본다.

### 1. UCC 검색을 위한 의미 추출 기술

위와 같이 멀티미디어 콘텐츠 검색에서 나타나는 의미갭을 줄이기 위하여 많은 연구들이 수행되고 있다. 한 예로 멀티미디어 콘텐츠의 시각적 의미들을 표현할 수 있는 의미 (concept)들을 모델링하는 것이다. 이러한 의미들을 멀티미디어 콘텐츠 검색에 사용하기 위해서는 자동으로 콘텐츠에 존재하는 의미를 추출하는 기술이 필요하게 된다. 기존 멀티미디어 콘텐츠에서 자동으로 의미를 추출하는 기술은 주로 멀티미디어 콘텐츠의 색상, 촉감과 같은 하위 레벨 특징 (low-level features)들을 바탕으로 멀

티미디어 콘텐츠 내에 존재하는 의미를 추출하였다. 최근에는 하위 레벨 특징 외에도 카메라에 존재하는 카메라 메타데이터들도 멀티미디어 콘텐츠에서 의미를 추출하는데 사용되었다 (그림 4) [17].

이러한 멀티미디어 콘텐츠에서의 의미 추출을 논의하는 대표적인 포럼은 TRECVID이다 [18]. TRECVID는 NIST에서 주관하는 멀티미디어 검색 관련 포럼으로 멀티미디어 콘텐츠의 검색을 위하여 동영상 콘텐츠의 구조 분석, 의미 정보 추출 등을 연구하고 있다. 특히 TRECVID의 'high level feature' 부문에서는 멀티미디어 콘텐츠 내에서 의미를 찾는 연구를 담당하고 있다. TRECVID high level feature에서는 매년 대상이 되는 멀티미디어 콘텐츠와 그 콘텐츠에 존재하는 의미 그룹을 제시하여 참가 연구 단체의 성능을 평가하고 발표한다. 대부분의 TRECVID 참가단체들은 콘텐츠의 다양한 시각적 정보를 바탕으로 Support vector machine (SVM)과 같은 분류기를 학습시키는 방법으로 의미를 추출하였다 [19]. 표 1은 TRECVID에



<그림 4> 하위 레벨 특징과 카메라 정보를 이용한 의미 추출 방법 [17]

〈표 1〉 TRECVID high-level feature 에서 사용된 의미들

TRECVID	의미의 수	의 미
2002	10	Outdoors, indoors, face, people, cityscape, landscape, text overlay, speech, instrumental sound, monologue
2003	17	Outdoors, new subject face, people, building, vegetation, animal, female, road vehicle, aircraft, news subject monologue, non-studio setting, sporting event, weather news, zoom in, physical violence, person x
2004	10	Boat/ship, Madeleine Albright, Bill Clinton, beach, basket scored, airplane takeoff, people walking/running, physical violence, road
2005~2007	39	Sports, Entertainment, Weather, Court, Office, Meeting, Studio, Outdoor, Building, Desert, Vegetation, Mountain, Road, Sky, Snow, Urban, Waterscape/Waterfront, Crowd, Face, Person, Government-Leader, Corporate-Leader, Police/Security, Military, Prisoner, Animal, Computer/TV-screen, Flag-US, Airplane, Car, Bus, Truck, Boat/Ship, Walking/Running, People-Marching, Explosion/Fire, Natural-Disaster, Maps, Charts

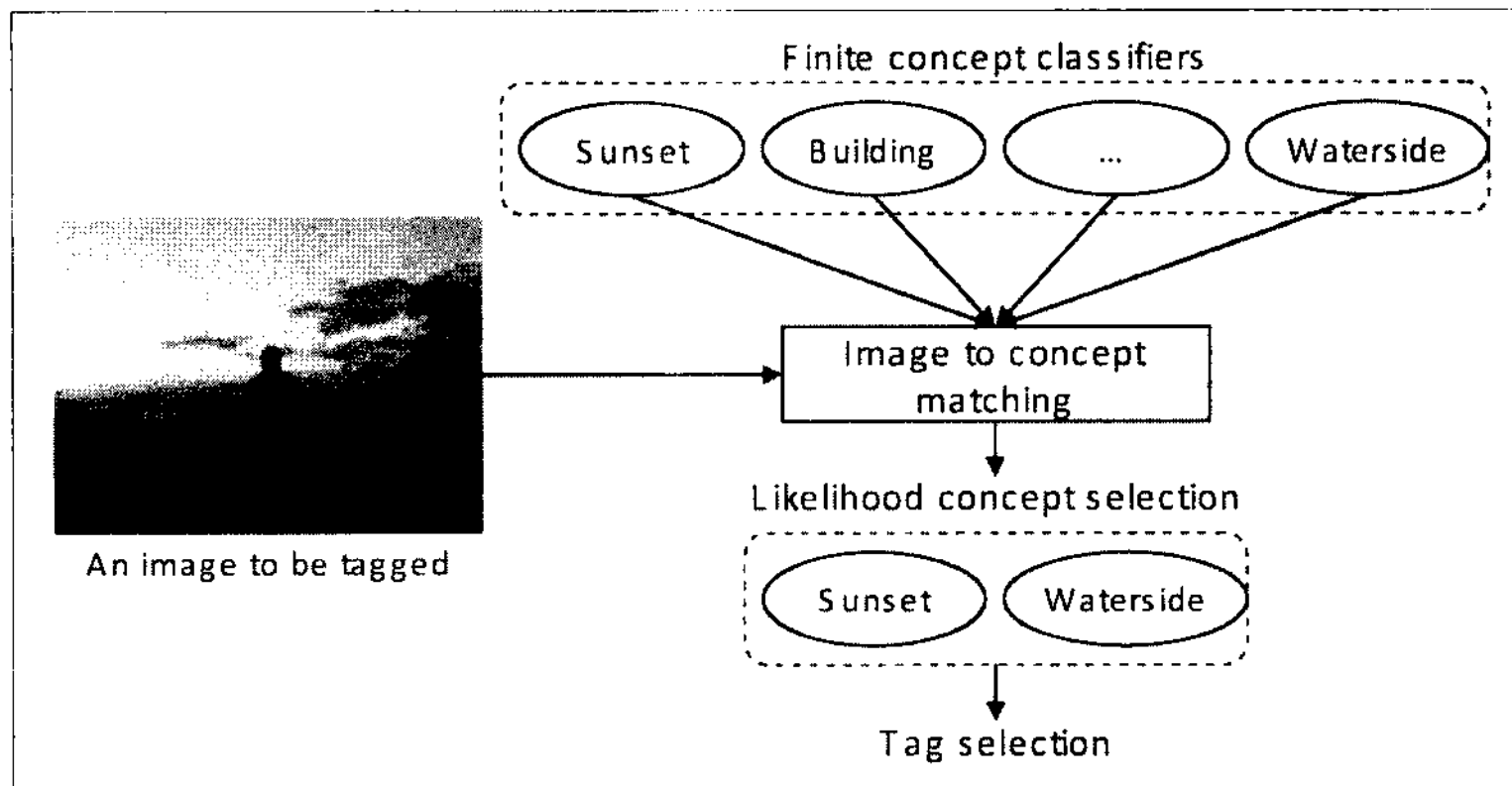
서 사용하는 의미들을 연도별로 나타낸 것이다.

TRECVID-2005부터는 LSCOM (Large-Scale Concept Ontology for Multimedia)에서 선정한 의미들을 사용하고 있다 [20]. 멀티미디어 콘텐츠의 의미 추출은 색상, 촉감과 같은 하위 레벨 특징 (low-level features)들을 바탕으로 하기 때문에 추출할 수 있는 의미의 수는 제한적이고, 의미의 수가 너무 많아지면 검출 성능이 떨어지고 알고리즘의 복잡도가 증가한다. 그렇기 때문에 추출하는 의미의 수와 종류는 의미 추출 기술에 중요한 요인으로 작용한다. LSCOM에서는 멀티미디어 콘텐츠를 위한 의미들의 표준을 선정하기 위한 연구를 진행하고 있다. 제한적인 수의 의미를 선택하기 위하여 LSCOM에서는 의미의 유용성 (utility), 범위 (coverage), 타당성 (feasibility), 관측성 (observability)의 4가지 조건을 사용한다. 유용성은 실제 사용자들이 사용하는 상황과 쿼리를 지원할 수 있는 의미를 선정하기 위한 조건이다. 또한 범위는 의미들이 포함하는 의미 공간(semantic space)을 의미하는데, 이는 제한된 수의 의미를 선택하여

야 하므로 사용자가 더 관심을 가지는 의미 공간을 포함한 의미를 선정하기 위한 조건이다. 타당성은 자동적으로 추출 가능한가를 나타낸다. '도전', '행복'과 같은 의미들은 자동적으로 콘텐츠를 분석하여 추출할 수 없다. 마지막으로 관측성은 멀티미디어 콘텐츠에서 자주 관찰될 수 있는지를 의미한다. 더 자주 관찰되는 의미를 우선적으로 선정하기 위한 조건이다. 이렇게 선정된 LSCOM 의미 집합의 중간 단계인 LSCOM-Lite annotation은 39개의 의미를 포함하고 있다.

## 2. UCC 태깅을 위한 주석 추천 기술

위에서 언급된 것과 같이 오늘날의 UCC 사이트들은 사용자가 태그와 같은 키워드들의 조합으로 자신이 제작한 콘텐츠 뿐 아니라 다른 사용자들의 콘텐츠도 기술할 수 있게 함으로써, 사용자들의 메타데이터 (User created metadata: UCM)를 생성할 수 있게 한다. 이렇게 생성된 UCM은 복잡한 구조를 가진 멀티미디어 UCC를 간단히 기술하여 UCC



〈그림 5〉 의미기반 태그 추천 과정

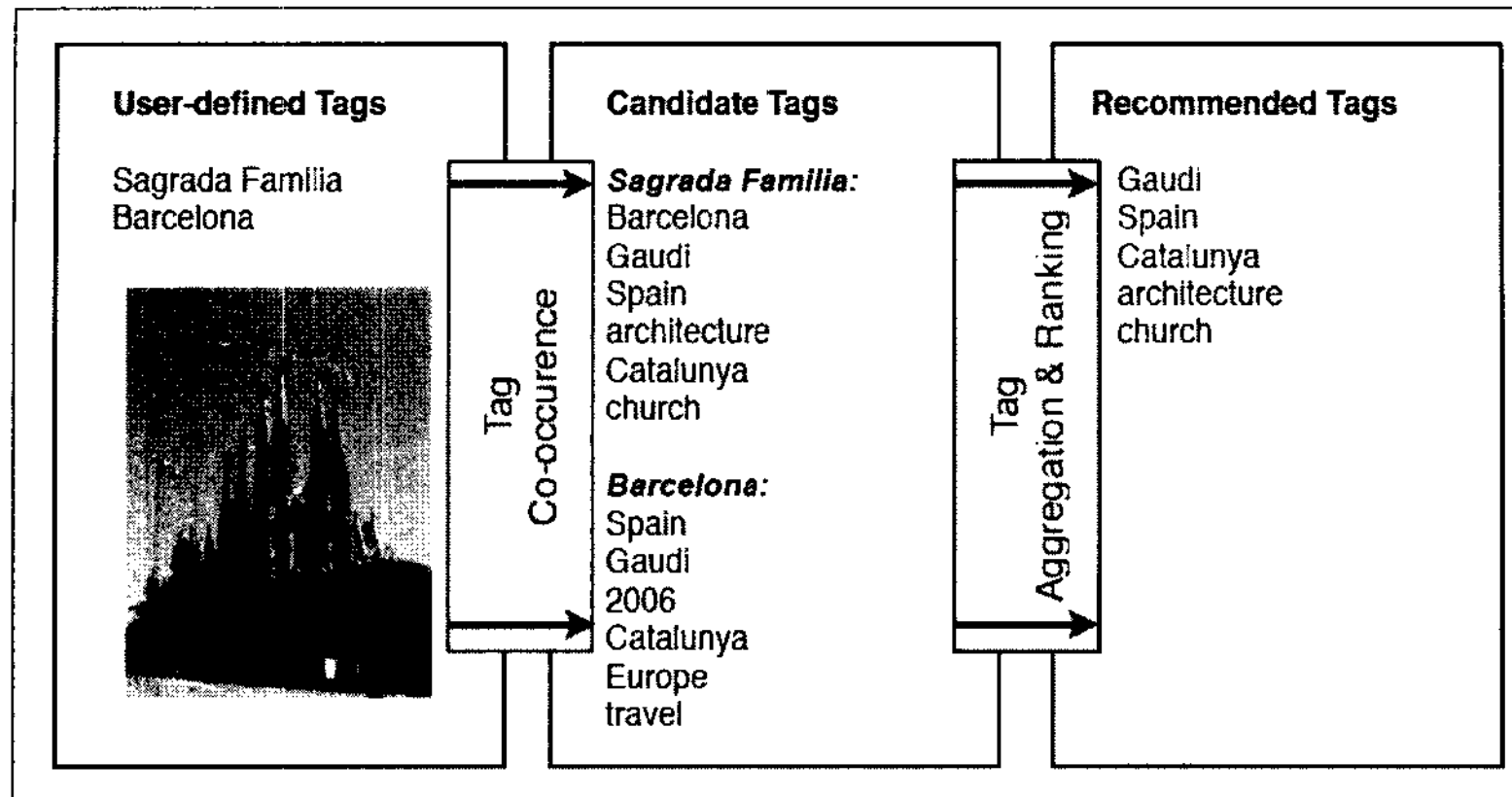
검색이 효율적으로 이루어지게 한다. 그러나 일반적으로 사용자는 자신의 UCC를 최선을 다하여 기술하지 않고, 자신들의 관심에 따라 몇 개의 태그를 남기는 정도로 그치는 것으로 알려져 있다 [21]. 또한 현재의 사용자 생성 태그들은 콘텐츠를 정확히 표현하기에는 모호하고 주관적인 경우가 많다. 이러한 문제점을 해결하기 위해서 콘텐츠 분석을 통하여 자동적으로 생성되는 태그를 지원함으로써 UCM을 풍부하게 하는 방법이 연구되고 있다 [22]-[25]. 특히 [24],[25]에서 query-by-semantic-example이라는 향상된 의미기반 태깅 기술이 소개되었다. 이 기술은 기존 query-by-visual-example과 의미기반 검색 기술을 결합시킴으로써 얻어진다. 그림 5는 이러한 의미기반 태깅 과정을 나타내는 그림으로 의미들을 판단할 수 있는 학습된 분류기를 사용한다. 이러한 분류기들을 바탕으로 영상-의미 매칭 방법을 사용하여 영상과 가장 유사한 의미들을 영상을 위한 태그들로 선택한다.

그러나 대부분의 기존 의미기반 인덱싱 방법들은 제한된 수의 분류기 모델을 바탕으로 한다. 이러한 모델을 이용한 인덱싱에서는 미리 학습되지 않은 개념들은 연관된 의미 정보의 부족으로 예측이 힘

들어진다. 그러므로, 알지 못하는 개념의 분류 문제는 모델을 이용한 멀티미디어 검색에서 문제가 되어왔다. 주관적인 의미에서부터 객관적인 의미, 추상적인 의미에서부터 명확한 의미들까지 UCC는 다양한 종류와 수준의 의미를 담고 있기 때문에 UCC에서는 알지 못하는 의미에 대한 문제가 더 중요해졌다.

이러한 의미의 제한을 해결하기 위하여 사용자들에 의해 기술된 폭소노미를 사용하는 방법이 연구되고 있다. [26]에서는 사용자의 이미지와 유사한 이미지 집합과 관련된 태그들을 바탕으로 사용자의 이미지를 위한 태그를 선택하는 기술이 소개되었다. 또한 [27]에서는 사용자의 태그와 비슷한 태그가 달린 이미지들에 관련된 태그를 추가 태그로 추천하는 기술이 소개되었다. 그림 6은 태그 추천 시스템으로 사용자가 정의한 태그를 시작으로 그 태그와 같이 나타나는 태그를 예비 태그들 (중간 모듈)을 만든다. 이후에 이 예비 태그들로부터 중복으로 많이 나오는 것을 추려서 최종 추천 태그를 만들 수 있는 것이다. 이때 예비 태그들은 다른 사용자들이 비슷한 영상에 붙인 태그들을 이용하는 것으로 폭소노미를 활용하여 추천 태그를 얻은 것이라 할 수 있다.





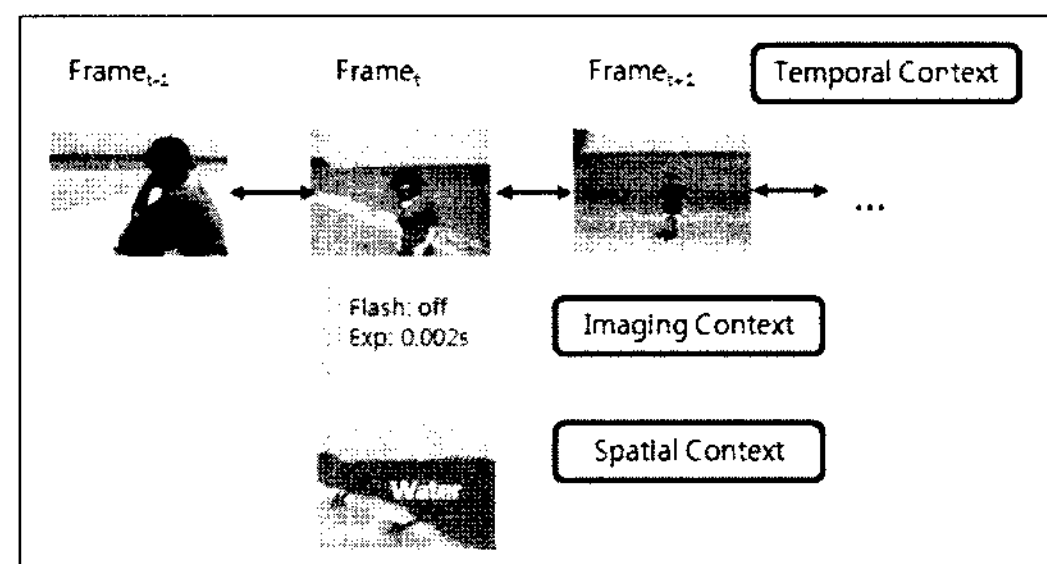
〈그림 6〉 태그 추천 시스템 [27]

### 3. UCC 검색을 위한 정황(context) 정보 기술

디지털 컨버전스 환경의 도래로 인하여 UCC의 생산/소비 환경이 다양해졌다. UCC 서비스 환경이 다양해짐에 따라, 사용자는 자신의 정황에 맞는 UCC 콘텐츠를 검색하여 소비할 수 있어야 한다. 그리고 다양한 UCC 생성 정황 정보는 UCC 검색을 효과적으로 향상시킬 수 있는 정보로 사용될 수 있다. 최근 여러 연구에서 영상 정보 분석을 위해 영상 내의 객체들의 위치 관계와 같은 공간적 정황 정보, 카메라 정보와 같은 영상 정황 정보와 같은 정황 정보를 하위 레벨 특징이나 의미 정보와 연합하여 사용하는 연구를 진행하였다 [28]. 그림 7은 비디오의 정황 정보를 나타낸 예로서 시간 컨텍스트 정보(temporal context), 공간 컨텍스트 정보(spatial context), 영상 생성 컨텍스트 정보(imaging context)를 나타낸다. 시간 컨텍스트 정보는 이웃한 프레임들간의 관계에서 획득할 수 있는 정보로써, 이웃한 프레임들을 바탕으로 더 효율적으로 콘텐츠를 분석할 수 있게 한다. 공간 컨텍스트 정보는 영상

내 객체들의 위치 관계를 통해 얻을 수 있다. 아래 그림에서 노란색 객체 위에 파란색 객체가 존재하는 공간적 컨텍스트 정보를 바탕으로 모래와 물이란 의미 정보를 생성하는데 사용할 수 있다. 또한 카메라 메타데이터와 같은 영상 생성 컨텍스트는 콘텐츠 생성 환경을 유추할 수 있는 정보로 활용될 수 있다.

이러한 정황 정보는 UCC 관리 및 검색을 위해 사용하기 위해서는 이러한 UCC의 생성/소비를 위한 정황 정보 메타데이터를 자동적으로 생성하여야 한다. [29]에서는 풍부한 정황 정보를 생성하기 위하여 블로그나 앨범에 존재하는 비슷한 정황 정보를



〈그림 7〉 콘텐츠 정보 인식을 위한 정황 정보들

가진 콘텐츠의 정황 정보를 추가적으로 사용자가 정황 정보 메타데이터를 생성하는데 사용할 수 있도록 제공하여 사용자로 하여금 더 풍부한 정황 정보를 제공하는 시스템을 제안하였다.

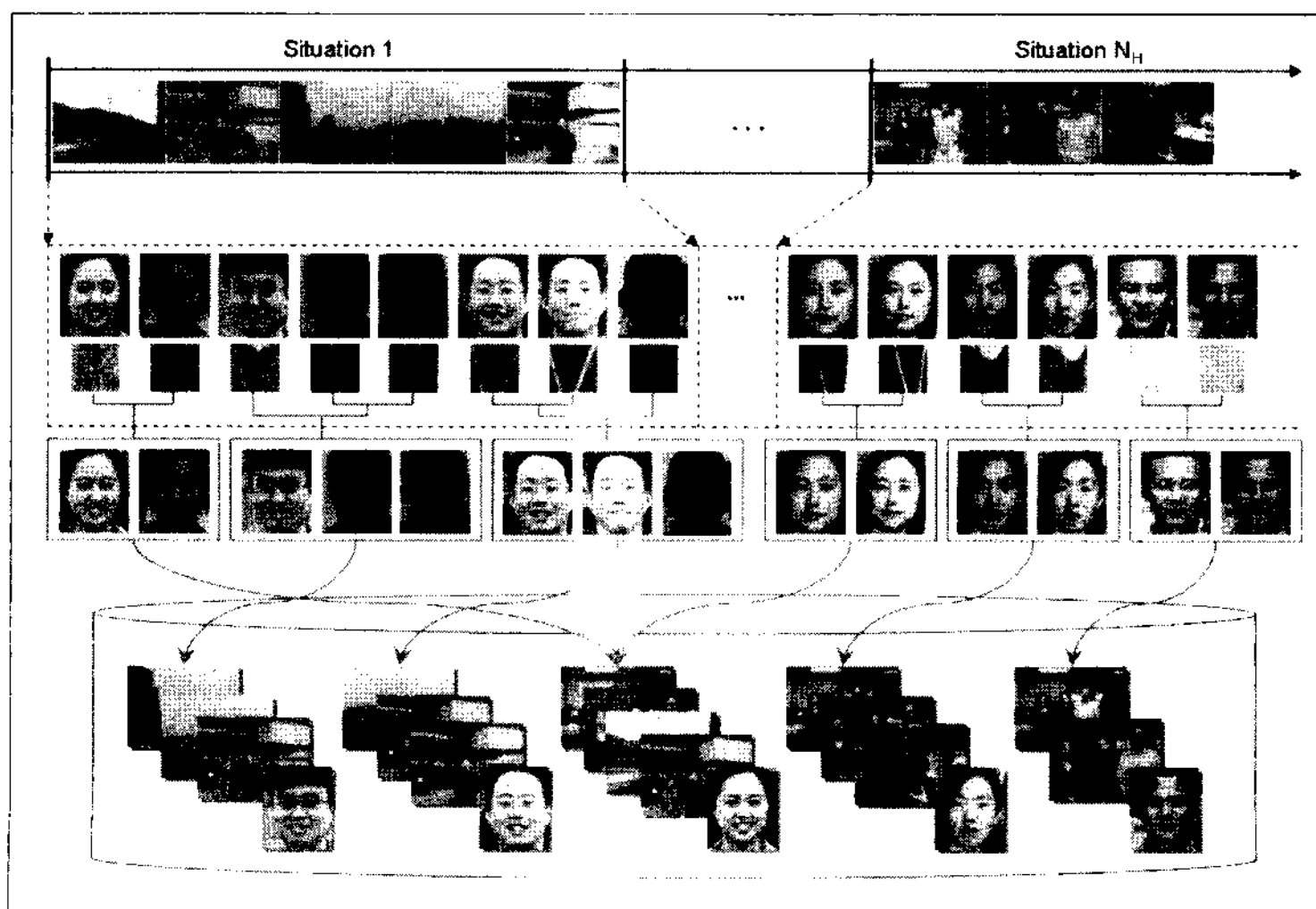
이러한 정황 정보는 기존 고려되었던 소비/생성 환경에 대한 정황 정보 외에도 콘텐츠에 담기는 내용의 정황 정보도 포함한다. 특히, 콘텐츠에 등장하는 인물에 대한 정보는 콘텐츠의 색인 및 검색을 위하여 중요한 정황 정보로 사용된다. [30]에서는 이러한 인물에 대한 정보를 사용하기 위하여 인물의 얼굴 정보 외에도 인물이 입고 있는 옷에 대한 정황 정보를 사용하여 인물의 얼굴 정보만을 사용하던 기존 시스템을 개선하였다(그림 8).

이러한 정황 정보는 UCC 검색을 위하여 중요한 정보로 사용될 수 있다. 또한 기존 의미 추출 등의 기술에 활용될 수 있는 정보이다. 그렇기 때문에 다양한 정황 정보를 효율적으로 생성하기 위한 연구가 계속 진행될 것으로 예상된다.

## IV. 결론

시멘틱 웹의 특징으로도 대변되는 사용자 참여, 공유, 개방의 개념과 네트워크 기술 및 멀티미디어 콘텐츠 기술의 발전으로 UCC는 현재 하나의 중요한 사회 현상이며, 또한 영향력 있는 하나의 미디어로 자리잡고 있다. 앞으로 방송과 통신이 융합하는 디지털 컨버전스 시대에는 UCC는 더욱 확산될 것이고, 앞으로 더욱 넓은 분야에 걸쳐 나타날 것이다. 이러한 상황에서 UCC 검색은 방송 통신 환경에서도 중요한 서비스가 될 것이라 전망되고 있다. 또한 이러한 서비스는 우리나라가 일궈온 초고속 인터넷 인프라를 기반으로 디지털 컨버전스 환경에서 빠르게 선점하여 세계적인 경쟁력을 갖도록 해야 하는 분야이다.

이러한 UCC 환경에서 기존 검색 서비스는 주로 텍스트 기반의 검색 시스템에 의해 제공되어 왔다. 그러나 텍스트로 UCC를 기술하는 데에는 한계가



〈그림 8〉 콘텐츠 내의 인물에 정황 정보를 활용한 시스템 [30]

존재하였다. 이러한 한계를 극복하기 위하여 내용 기반 검색이 개발되었으나 콘텐츠의 내용 특징만으로는 검색 성능을 보장할 수 없다는 문제를 해결하지 못하고 있다. 이는 콘텐츠의 시각적 내용만을 통해 추출한 정보와 콘텐츠의 정보간의 발생하는 감각 차이와 의미 차이가 존재하기 때문이다. 이러한 의미 차이를 극복하기 위하여 UCC에서 의미를 추출하는 기술, UCC를 위한 주석을 추천해 주는 기

술, 그리고 UCC의 콘텍스트를 인지하는 기술 등이 연구되고 있다. 그러나 아직은 UCC가 가지는 다양한 의미들을 지원하진 못하고 있다. 이러한 한계를 극복하고 UCC의 의미 정보, 콘텍스트 정보를 이용하여 UCC 검색 서비스에 대한 연구가 필요하다. 이러한 기술들의 개발은 UCC 검색을 위한 원천 기술을 확보하고 차세대 UCC 검색 시장을 선도할 수 있는 중요한 연구 과제이다.

## 참고 문헌

- [1] "Trend marker YouTube," 2006.
- [2] "7 things you should know about YouTube," 2006.
- [3] G. Ireland, L. Ward, "Transcoding Internet and Mobile Video: Solutions for the Long Tail," IDC, 2007.
- [4] "iProspect Search Marketer Social Networking Survey," 2007.
- [5] M. Perry, K. O'hara, A. Sellen, B. Brown, R. Harper, "Dealing with mobility: understanding access anytime, anywhere," ACM Transactions on Computer-Human Interaction, Vol. 8, Issue. 4, pp. 323-347, 2001.
- [6] D. Best, "Web 2.0 Next Big Thing or Next Big Internet Bubble?," technische universiteit eindhoven, lecture web information systems, 2006.
- [7] M. Davis, "Semantic Wave 2007, Industry Road map to Web 3.0," tutorial of semantic technology conference, 2007.
- [8] H. Holone, G. Misund and H. Holmstedt, "Users Are Doing It For Themselves: Pedestrian Navigation With User Generated Content," NGMAST 2007, pp. 91099, 2007.
- [9] M. Tkalcic, J. Tasic, "P2P for MajorEvents: A Hybrid Multimedia Distribution Platform," The International Conference on "Computer as a Tool", pp. 974-977, 2007.
- [10] T. A. S. Coelho, P. P. Calado, L. V. Souza, B. Ribeiro-Neto, R. Muntz, "Image retrieval using multiple evidence ranking," IEEE Transactions on Knowledge and Data Engineering, Vol. 16, pp. 408 - 417, 2004
- [11] "Accenture Global Digital Home Consumer Survey," 2006.
- [12] Graham Vickery, Sacha Wunsch-Vincent, "Participative Web And User-Created Content: Web 2.0 Wikis and Social Networking," Organization for Economic, 2007.
- [13] M. Ames, and M. Naaman, "Why We Tag: Motivations for Annotation in Mobile and Online Media". ACM CHI 2007, pp. 971-980, 2007
- [14] R. C. Veltkamp, M. Tanase, "Content-Based Image Retrieval Systems: A Survey," 2002.
- [15] IBM Multimedia analysis and retrieval system, "http://mp7.watson.ibm.com".
- [16] A. W. M. Smeulders et al., "Content based image retrieval at the end of the early years," IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 12, pp. 1349-1380, Dec. 2000.
- [17] Seungji Yang, Sang-Kyun Kim, Yonh Man Ro, "Semantic Home Photo Categorization," IEEE Trans. Circuits and Systems for Video Technology, vol. 17, no. 3, pp. 324-335, 2007.
- [18] P. Over et al., "TRECVID 2006-An Overview," in Proc. TRECVID'06 Video Retrieval Evaluation, Gaithersburg, MD, 2006.
- [19] W. Kraaij et al., "TRECVID-2006 High-Level Feature task: Overview," in Proc. TRECVID'06 Video Retrieval Evaluation, Gaithersburg, MD, 2006.

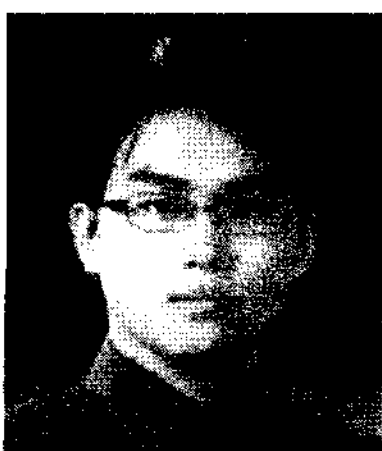
- [20] Naphade, et al., "Large-Scale Concept Ontology for Multimedia," IEEE MultiMedia, vol. 13, no. 3, pp. 86-91, 2006.
- [21] M. Tkalcić and J. Tkalcić, "Convergence of Web 2.0 and Semantic Web: A Semantic Tagging and Searching System for Creating and Searching Blogs," IEEE Int'l Conf. on Semantic Computing, pp. 201-208, Sep. 2007.
- [22] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans. on Pattern Anal. and Mach. Intelligence, Vol. 22, No. 12, pp. 1349-80, 2000.
- [23] R. Jain R, "Photo Retrieval: Multimedia's Chance to Solve a Real Problem for Real People," IEEE Multimedia, Vol. 14, No. 3, pp.111-112, July-Sept. 2007.
- [24] N. Rasiwasia, P.J. Moreno, N. Vasconcelos, "Bridging the Gap: Query by Semantic Example," IEEE Trans. on Multimedia, Vol. 9, No. 5, pp. 923-938, Aug. 2007.
- [25] G. Carneiro, A.B. Chan, P.J. Moreno, and N. Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," IEEE Trans. on IEEE Trans. Pattern Anal. and Mach. Intelligence, Vol. 29, No. 3, pp. 394-410, Mar. 2007.
- [26] R. Datta, W. Ge, J. Li., J.Z. Wang, "Toward Bridging the Annotation-Retrieval Gap in Image Search, IEEE Multimedia," Vol. 14, No. 3, pp.24-35, July-Sept. 2007
- [27] Borkur Sigurbjornsson, and Roelof van Zwol, "Flickr tag recommendation based on collective knowledge," WWW 2008, pp 327-336. 2007.
- [28] L. Jiebo, M. Boutell, C. Brown, "Pictures are not taken in a vacuum - an overview of exploiting context for semantic scene content understanding," IEEE Signal Processing Magazine, Vol. 23, No. 2, pp. 101-114, 2006.
- [29] S. Sarin et al, "On automatic contextual metadata generation for personal digital photographs," IEEE Conf. on Advanced Communication Technology, pp. 66-71, 2007.
- [30] S. Yang, K. S. Seo, S. Kim, Y. M. Ro, J. Kim, and Y. Seo, "Automatic Photo Indexing Based on Person Identity," LNCS, 3768, pp. 877-888, 2005

## 필자소개



### 노용만

- 1992년: KAIST 전기공학과 박사
- 1986년: Columbia University 연구원
- 1996년: University of California Berkeley 연구원
- 2007년: University of Toronto 방문교수
- 현재: 한국정보통신대학교 정교수
- 주관심분야: 이미지/비디오 처리 및 분석, 시맨틱 인덱싱, 비디오 적응변환, 미디어 QoS, MPEG



### 민현석

- 2005년: 아주대학교 정보및컴퓨터공학과 학사
- 2008년: 한국정보통신대학교 공학부 석사
- 2008년 ~ 현재: 한국정보통신대학교 공학부 박사과정
- 주관심분야: 이미지/비디오 인덱싱, 이미지/비디오 처리 및 분석

## 필자 소개



### 이시형

- 2005년 : 경희대학교 전자정보학부 학사
- 2007년 : 한국정보통신대학교 공학부 석사
- 2007년 ~ 현재 : 한국정보통신대학교 공학부 박사과정
- 주관심분야 : 이미지/비디오 처리 및 분석, 멀티미디어 응용 시스템



### Wesley De Neve

- 2002년 : M.Sc., Computer Science, Ghent University (Belgium)
- 2007년 : Ph.D., Computer Science Engineering, Ghent University (Belgium)
- 2007년 ~ 현재 : Postdoctoral Researcher, Ghent University-IBBT-ICU
- 주관심분야 : Video coding and adaptation, GPU-based video processing, efficient XML processing